The EMBO Journal vol.3 no.9 pp.2021 – 2028, 1984

Synthetic donor and acceptor splice sites function in an RNA polymerase B (II) transcription unit

G.Rautmann[1], H.W.D.Matthes[1,2], M.J.Gait[2] and R.Breathnach[1]

[1]Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11, rue Humann, 67085 Strasbourg Cédex, France, and [2]MRC, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Communicated by P.Chambon

We have synthesised a 32-bp oligonucleotide containing sequences conforming to the consensus sequences for donor and acceptor splice sites. The oligonucleotide has been inserted into an RNA polymerase B (II) transcription unit and the resulting recombinant used to study the splicing mechanism. Our findings are as follows: (i) the synthetic sites function when separated by several different prokaryotic or eukaryotic DNA fragments providing bulk intron sequence, (ii) intron size need not be greater than 29 bp, (iii) an AG dinucleotide 11 bp upstream from the invariant AG of an acceptor splice site renders the latter non-functional, and (iv) sequence changes distant from splice sites can affect the efficiency of their utilisation.

Key words: splice site/RNA polymerase B (II)/transcription unit/consensus sequence

Introduction

While the mechanisms which ensure the accuracy and specificity of the splicing of RNA polymerase B (II) primary transcripts remain unclear, it has been known for some time that donor and acceptor splice sites can always be derived from consensus sequences (Figure 1) by changing only a few base pairs, and that the almost invariant GT (donor) and AG (acceptor) residues mark the splice points (Breathnach *et al.*, 1978; Breathnach and Chambon, 1981; Mount, 1982). Mutations within the consensus sequence area can destroy splice sites (see e.g., Treisman *et al.*, 1983; Mount and Steitz, 1983). On the other hand, the bulk of the intron sequences separating donor and acceptor splice sites can be deleted without any apparent effect on splicing (Dierks *et al.*, 1981; Volckaert *et al.*, 1979). These results suggest that sequences designed to resemble the consensus sequences for donor and acceptor splice sites should function as such, irrespective of the nature of the intron sequences separating them. We show below that this is the case, thus reinforcing the importance of the consensus sequences for splicing. However, we also show that sequences > 35 bp away from a splice site can affect the efficiency of its utilisation, demonstrating that similarity to the consensus sequences alone is insufficient to guarantee splice-site function.

Results

Construction of a plasmid for studying splicing

We have synthesized a double-stranded oligonucleotide of 32 bp carrying sequences closely related to the consensus sequences for donor and acceptor splice sites (Figure 1). To enable us to test *in vivo* the splicing capability of the synthetic splice sites, the oligonucleotide was introduced into the transcription unit of a rabbit β-globin gene carried by pEP1 (Figure 2). This plasmid contains an SV40 enhancer sequence and a rabbit β-globin gene; when introduced into eukaryotic cells in culture, it gives rise to correctly spliced globin mRNA

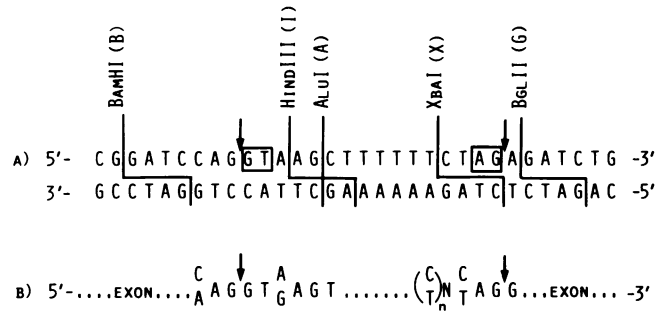

Fig. 1. (A) Sequence of the synthetic oligonucleotide containing donor and acceptor splice sites. Restriction enzyme sites (and their abbreviations) discussed in the text are shown. Vertical arrows indicate the splice points as given by the GT-AG rule (Breathnach *et al.*, 1978). **(B)** Consensus sequences for donor and acceptor splice sites from Mount (1982). N = any nucleotide.

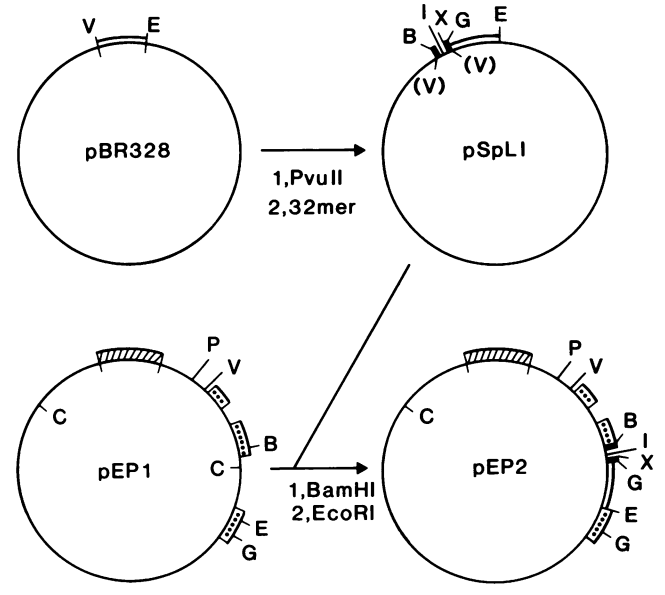

Fig. 2. Introduction of the splicing oligonucleotide into a rabbit β-globin gene transcription unit (see text and Materials and methods). The 100-bp *Pvu*II-*Eco*RI (V-E) fragment of pBR328 present in nearly all other constructions is indicated by double lines on the pBR328 map and elsewhere. The hatched box is the SV40 enhancer sequence (covering nucleotides 399 – 5171 in the BBB numbering system). The dotted boxes are the rabbit β-globin exons. The filled-in boxes correspond to the exon sequences of the splicing oligonucleotide (Figure 1A). Restriction site abbreviations are as in Figure 1A. In addition, C = *Hin*cII; V = *Pvu*II; E = *Eco*RI; sites in parentheses were destroyed during the recombinant construction.

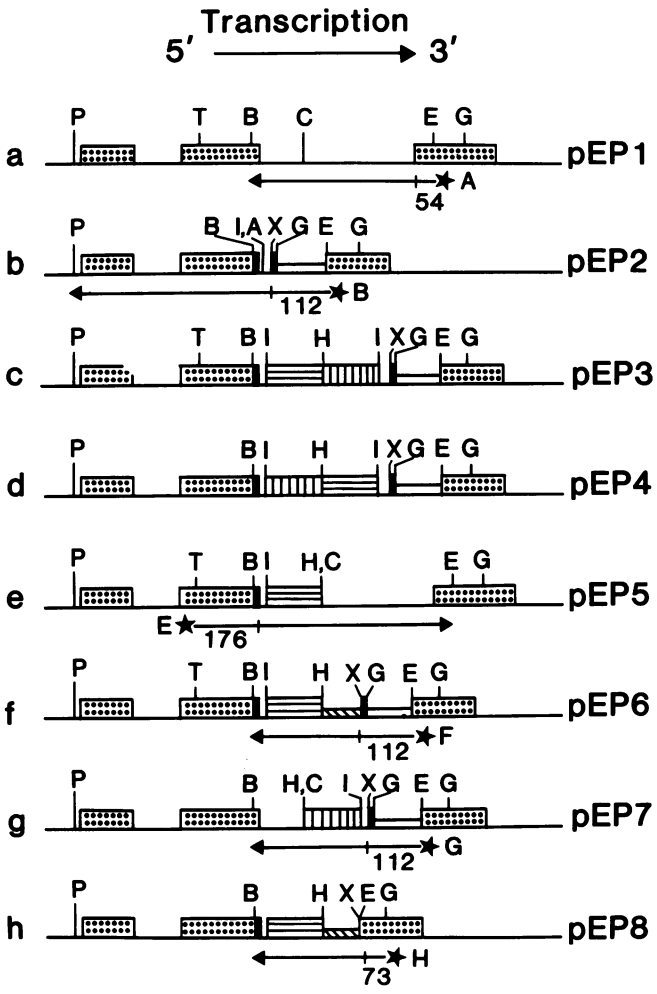

Fig. 3. Transcription units present in plasmids used to study splicing (see Materials and methods, and text). Probes A, B, E, F, G and H prepared from some of these plasmids are indicated under the relevant transcription unit, an asterisk marking the labelled end and an arrowhead the other end. A vertical line marks the extent of probe which would be protected by correctly spliced RNA, and the size of the protected fragment is given in nucleotides. Symbols are as in the legend to Figures 1 and 2. In addition the vertically and horizontally shaded boxes correspond respectively to a *Hind*III-*Hpa*I and a *Hpa*I-*Hind*III fragment of SV40 DNA which together make up the SV40 *Hind*III fragment covering bp 3476–4002 in the BBB numbering system. The diagonally shaded box corresponds to chicken ovalbumin gene intron sequence (see Materials and methods, and text). T = *Taq*I, B = *Bam*HI; H = *Hpa*I; A = *Alu*I; X = *Xba*I; G = *Bgl*II; E = *Eco*RI; P = *Pst*I; C = *Hinc*II; I = *Hind*III; V = *Pvu*II.

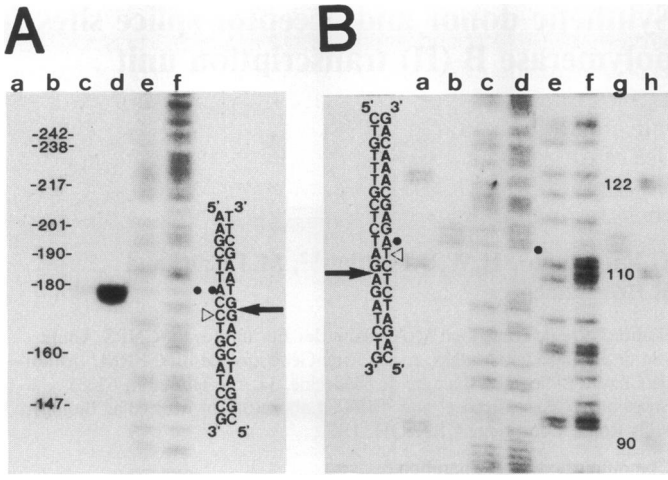

Fig. 4. (A) Nuclease S1 mapping experiment demonstrating use of the synthetic donor splice site. RNA (50 μg) from HeLa cells mock-transfected **(lane a)** or transfected with pEP3 **(lane c)** or pEP5 **(lane d)** was hybridized to the single-stranded 3′ end-labelled probe E (Figure 3e), hybrids were digested with nuclease S1 and electrophoresed on an 8% polyacrylamide 7 M urea denaturing gel. **Lane b:** migration of marker fragments from an *Msp*II digest of pBR322. **Lanes e, f:** A + C and A + G cleavages of probe E, respectively. A dot against **lane f** shows an A cleavage product with which the protected probe fragments co-migrate. Next to **lane f** is shown the sequence of probe E in this area and of its complementary strand, the relevant A residue again shown dotted. A C residue (arrowhead) marks the extent of probe E which should be protected had the donor splice site (arrow on complementary strand) been used correctly (see text). **(B)** Nuclease S1 mapping experiment demonstrating use of an ovalbumin acceptor splice site. RNA (50 μg) from HeLa cells transfected with pEP6 **(lanes b, g)** was hybridized to the single-stranded 5′ end-labelled probe F (Figure 3f), nuclease S1-digested and electrophoresed on a denaturing 8% polyacrylamide 7 M urea gel. **Lanes a, h:** marker bands from an *Msp*II digest of pBR322, size shown in bp against lane **h**. **Lanes c–f:** A+C, A+G, C and C+T cleavages of probe F, respectively. A dot against **lane d** shows an A cleavage product with which the fastest-moving protected probe fragment co-migrates. Next to **lane a** is shown the sequence of probe F in this region and of its complementary strand, the relevant A residue again shown dotted. A T residue (arrowhead) marks the extent of probe F which should be protected had the acceptor splice site (arrow on complementary strand) been used correctly (see text).

(not shown). In practice, the oligonucleotide was first cloned into the *Pvu*II site of the plasmid pBR328 (Figure 2) to make pSpL1. A 132-bp *Bam*HI-*Eco*RI fragment which can be excised from this plasmid contains an ~100-bp *Pvu*II-*Eco*RI fragment of pBR328 fused to the oligonucleotide at its acceptor splice site end. This *Bam*HI-*Eco*RI fragment containing the synthetic splice sites was used to replace a *Bam*HI-*Eco*RI fragment of the rabbit β-globin gene of pEP1. This latter fragment carries a natural donor and acceptor splice site. Experiments have shown that it contains all the information needed for splicing of its intron transcript (O'Hare, 1981; Breathnach and Harris, 1983). The resulting plasmid pEP2 (Figures 2 and 3b) contains the splicing oligonucleotide within the β-globin gene transcription unit.

The synthetic donor splice site is functional

pEP2 (Figure 3b) was introduced into HeLa cells by the calcium phosphate transfection procedure. Cytoplasmic RNA was harvested 72 h later and hybridized to a single-stranded 707 nucleotide *Eco*RI-*Pst*I fragment of pEP2, 5′ end-labelled at its *Eco*RI terminus (probe B, Figure 3b). Hybrids were digested with nuclease S1 before electrophoresis on a denaturing gel. Use of the synthetic acceptor splice site should result in a protected fragment of probe of 112 nucleotides (Figure 3b, distance from *Eco*RI site to the acceptor splice site). In fact, no protected fragments were found (data not shown), suggesting that the synthetic acceptor splice site was not used. We speculated that this might be due to an insufficient separation of the synthetic donor and acceptor splice sites in pEP2, and therefore constructed two new recombinants pEP3 and pEP4 (Figure 3c,d) by introducing a 526-bp *Hind*III fragment of SV40 into the *Hind*III site of pEP2. In these recombinants the size of the synthetic intron is 542 bp. However, repeating the nuclease S1 mapping experiment with these recombinants gave the same negative result as pEP2 (data not shown). We decided therefore to test the synthetic donor and acceptor splice sites separately.

Plasmids were constructed where the synthetic donor splice site of pEP3 was opposed to an acceptor splice site from the rabbit β-globin gene or a chicken ovalbumin gene (pEP5 and pEP6, Figure 3e and f, respectively). These plasmids and

pEP3 were introduced into HeLa cells, cytoplasmic RNA harvested 72 h later and hybridized to a single-stranded *Taq*I-*Eco*RI fragment (297 nucleotides) of pEP5, 3′ end-labelled at its *Taq*I extremity (Figure 3e, probe E). Hybrids were treated with nuclease S1 and electrophoresed on a denaturing gel. Use of the synthetic donor splice site should result in a protected fragment of probe of 176 nucleotides. This result was obtained with all three plasmids (Figure 4A, lanes c and d; result for pEP6 not shown). Spliced RNA using the synthetic donor splice site according to the GT-AG rule should protect probe E up to the C residue shown by an arrowhead. In fact, the protected probe migrates slightly faster than an A residue two nucleotides higher (dots). However, comparison of the mobilities of nuclease-generated and chemically-generated fragments is complicated by the fact that their structure is not identical. Protected fragments obtained using a 3′ end-labelled probe migrate slightly slower than expected relative to a sequencing ladder. The reverse is true for protected fragments obtained using a 5′ end-labelled probe. A further complication lies in the fact that the nuclease S1 and sequencing samples will not have identical salt contents. All in all, the results shown in Figure 4A are consistent with correct use of the synthetic donor splice site.

Nuclease S1 mapping of pEP5 and pEP6 RNA with single-stranded *Bam*HI-*Eco*RI fragments of pEP1 or pEP6 5′ end-labelled at the *Eco*RI extremities (Figure 3a and f, probes A and F) resulted in protected fragments of 54 and 112 nucleotides, respectively. This result is shown in Figure 4B, lanes b and g for pEP6 (the pEP5 result is not shown). Splicing of the synthetic donor splice site 5′-AG↓GTAAGCTT-3′ to the ovalbumin acceptor splice site 5′-TCTAG↓AGATCT-3′ will produce the sequence 5′-AGAGATCT-3′ (we show all sequences as DNA). Such spliced RNA will protect probe F up to the T residue marked by an open arrowhead in Figure 4B. In fact two protected fragments are observed (Figure 4B, lanes b and g). A similar effect is seen in our hands with a natural β-globin acceptor (see Figure 7A, f−j), and may be ascribed to incomplete nuclease S1 digestion. In this case, the shorter protected fragment corresponds to the splice point. For the ovalbumin case, the shorter fragment co-migrates with an A residue immediately following the T residue (see dots). Taking into account the differences in migration as discussed above, this result is consistent with the notion that the synthetic donor splice site is spliced to the globin or ovalbumin acceptor splice site in the respective RNAs. On the other hand, while the synthetic donor splice site has been used in pEP3 RNA (Figure 4A) it has apparently not been spliced to the synthetic acceptor splice site. We have searched in pEP3 for an acceptor splice site downstream from the synthetic acceptor splice site using a single-stranded *Bgl*II-*Bam*HI fragment of pEP8 5′ end-labelled at a *Bgl*II terminus (Figure 3h, probe H). When this probe was used for nuclease S1 mapping with pE3 RNA, groups of protected fragments of ~85−88 (I) and ~76 (II) nucleotides were observed (Figure 5). The first of these groups of fragments corresponds to RNAs covering the entire extent of the probe common to these recombinants, i.e., from the *Bgl*II site up to the *Eco*RI site, a distance of 85 nucleotides. As discussed above, these fragments show some heterogeneity in size, due to under-digestion by nuclease S1. The second group of fragments corresponds to use of an acceptor splice site lying in globin exon sequences just downstream from the *Eco*RI site. This acceptor splice site corresponds to the sequence 5′-GAATTCACTCCTCAG↓GTG-3′ in the globin

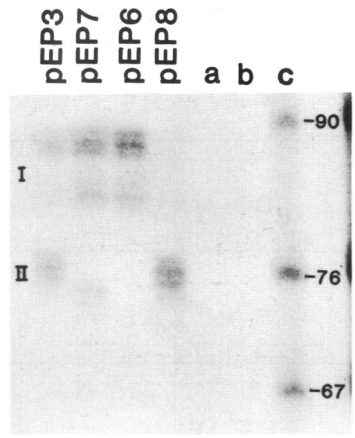

Fig. 5. Evidence for a cryptic rabbit β-globin gene acceptor splice site. RNA (50 μg) from HeLa cells transfected with pEP3, pEP7, pEP6, pEP8 or not transfected (**lanes a, b**) was hybridized to single-stranded probe H (Figure 3h), the hybrids treated with nuclease S1 and electrophoresed on an 8% polyacrylamide 7 M urea denaturing gel. **Lane c:** marker bands from a *Msp*II digest of pBR322. I and II refer to groups of bands discussed in the text.

genes. This site is not used in pEP6 RNA, where the synthetic donor is spliced to the ovalbumin acceptor splice site (Figure 5). However, if the distance between the ovalbumin acceptor and this cryptic globin acceptor splice site is reduced by deletion of an *Xba*I-*Eco*RI fragment from pEP6 to generate pEP8, the cryptic globin acceptor splice site is now used to the exclusion of the ovalbumin acceptor (Figure 5). The relevant sequence generated is 5′-TCTAG↓AATTCACT CCTCAG↓GTG-3′, where the invariant AG dinucleotide of the ovalbumin acceptor site is underlined and that of the cryptic globin acceptor site boxed.

An AG 11 nucleotides upstream from the AG of an acceptor renders the latter non-functional

Since the synthetic donor splice site can be spliced to both 'natural' and 'cryptic' acceptor splice site, the results obtained with pEP2 and pEP3 discussed above imply that the synthetic acceptor splice site cannot be spliced in its present form. This was confirmed by construction of pEP7 (Figure 3g), where the synthetic acceptor splice site is opposed to a rabbit β-globin gene donor splice site. This plasmid was introduced into HeLa cells and cytoplasmic RNA analysed by nuclease S1 mapping using a single-stranded *Bam*HI-*Eco*RI fragment of pEP7, 5′ end-labelled at its *Eco*RI extremity as probe (probe G, Figure 3g). No protected fragment of probe was observed (data not shown). When probe H (Figure 3H) was used, we found evidence for use of the cryptic globin acceptor splice site discussed above (Figure 5). Thus, the synthetic acceptor site is not functional.

The presence of a second AG dinucleotide 11 nucleotides upstream from the invariant AG of the acceptor splice site in the constructions pEP2, pEP3, pEP4 and pEP7 (see Figures 1A and 3) might explain why the synthetic acceptor splice site is not functional. To test this hypothesis we constructed pEP9, pEP10 and pEP11 (Figure 6, a−c). In pEP10 and pEP9, the synthetic donor splice site is opposed to a synthetic acceptor splice site of the form 5′-GTTCTTTTTTCTAG↓A-3′ [derived from cleavage of the oligonucleotide at the *Alu*I site (Figure 1A) and ligation of the *Alu*I terminus to a *Hpa*I terminus GTT] instead of 5′-AGCTTTTTTTCTAG↓A-3′ as in pEP3, pEP4 and pEP7. The body of the intron is a 272- or 282-bp fragment of SV40 DNA (pEP10 or pEP9 respec-

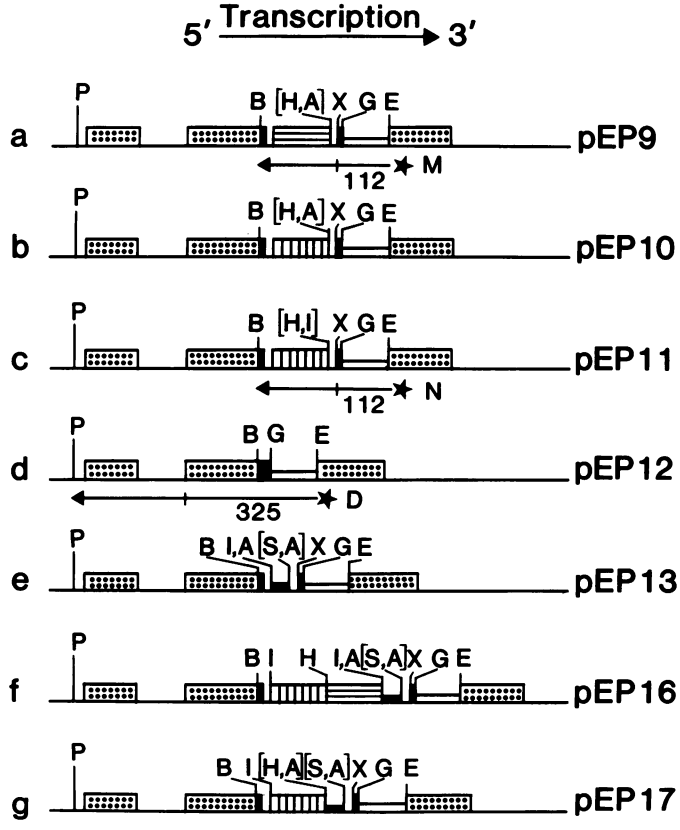

Fig. 6. Transcription units present in plasmids used to study splicing (see Materials and methods, and text). Symbols and abbreviations are as in the legend to Figure 3. In addition, S = *Sma*I. Bracketed sites have been destroyed during the cloning process and are included to show the origin of various fragments. The horizontal narrow filled-in box between the *Hind*III and [*Sma*I/*Alu*] sites of pEP13, pEP16 and pEP17 represent a 27-bp fragment of pUC8 (see Materials and methods, and text).

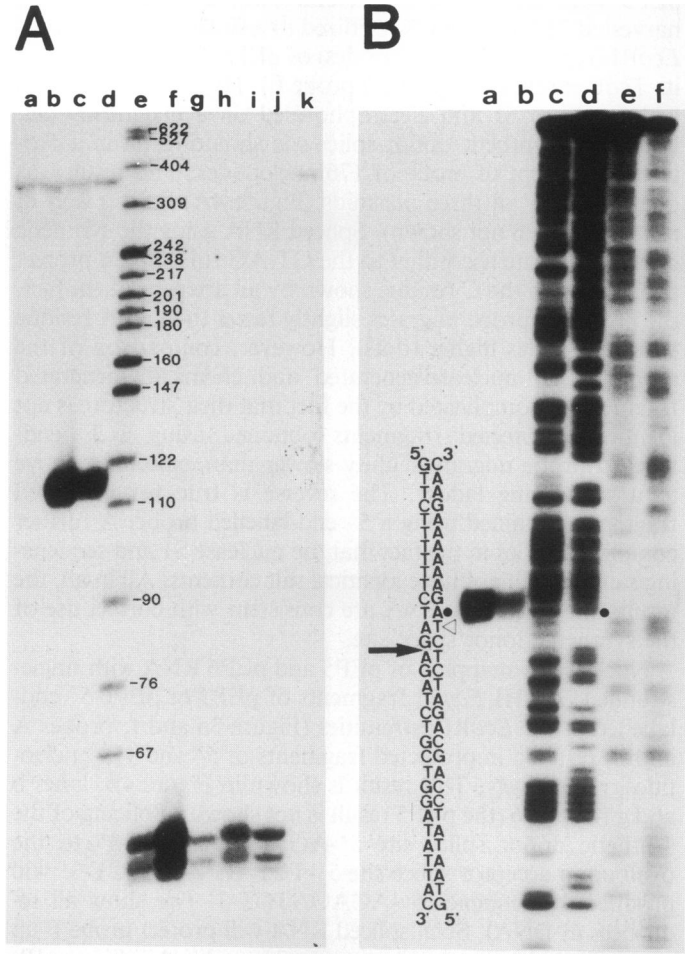

Fig. 7. (A) Masking of an acceptor splice site by an upstream dinucleotide AG. RNA (50 μg) from HeLa cells transfected with a 1:4 mixture of pEP1 and either pEP11 **(lanes a, b** and **i, j)** or pEP10 **(lanes c, d** and **f, g)** or pEP1 alone **(lane h)** was hybridized with single-stranded probe N (Figure 6c) **(lanes a–d)**, or probe A (Figure 3a) **(lanes f–j)**, treated with nuclease S1 and electrophoresed on a denaturing 8% polyacrylamide 7 M urea gel. **Lane e:** marker bands from an *Msp*II digest of pBR322. **Lane k:** an aliquot of probe N. **(B)** Nuclease S1 mapping experiment demonstrating use of the synthetic acceptor splice site. RNA (50 μg) from HeLa cells transfected with pEP9 **(lane a)** or pEP10 **(lane b)** was hybridized to the single-stranded 5′ end-labelled probe M (Figure 6a), treated with nuclease S1 and electrophoresed on an 8% polyacrylamide 7 M urea gel. **Lanes c–f:** A+C, A+G, C and C+T cleavages of probe M, respectively. An A residue with which the protected probe fragments almost co-migrate is shown by a dot against lane **d**. The sequence of probe M in this region and of its complementary strand is shown against lane **a**. The relevant A residue is shown dotted. A T residue (arrowhead) marks the extent of probe M which should be protected had the acceptor splice site (arrow on complementary strand) been used correctly (see text).

tively). The sole difference between pEP10 and pEP11 is that in the latter the acceptor splice site is of the form 5′-GTTAGCTTTTTTCTAG$^{\downarrow}$A-3′ [derived from *Hind*III cleavage and polymerase repair of the oligonucleotide (Figure 1A) before ligation to a *Hpa*I terminus GTT], i.e., it has the AG dinucleotide 11 bp upstream from the acceptor. Otherwise the intron is of exactly the same structure as that in pEP10. Each of these three plasmids was mixed with pEP1, introduced into HeLa cells and RNA was analyzed by nuclease S1 mapping using different probes. The first probes used were 384- or 394-nucleotide single-stranded *Bam*HI-*Eco*RI fragments of pEP11 and pEP9 respectively, 5′ end-labelled at their *Eco*RI extremities (Figure 6, probes N and M). In all these cases, use of the synthetic acceptor splice site should result in a protected fragment of 112 nucleotides. This result was found for RNA obtained using pEP10 and pEP9, but not for pEP11 RNA, where no protected fragment was observed (see Figure 7A, lanes c and d for pEP10, and lanes a and b for pEP11 RNA; data for pEP9 RNA are not shown). The second probe used was a 649-nucleotide single-stranded *Bam*HI-*Eco*RI fragment of pEP1, 5′ end-labelled at its *Eco*RI terminus (Figure 3a, probe A). This probe will only detect pEP1-derived RNA giving a protected fragment of 54 nucleotides (distance between the *Eco*RI site in pEP1 and the upstream globin acceptor splice site) and thus acts as a control for the efficiency of the different transfections. The results shown in Figure 7A (lanes f and g for pEP10; lane h for pEP1; lanes i and j for pEP11) demonstrate that the difference observed between pEP10 and pEP11 RNAs cannot be due to different transfection efficiencies. We conclude that an AG dinucleotide 11 nucleotides upstream from an acceptor splice site will render the latter non-functional.

Attempts to construct an analogue of pEP11 for pEP9, led to the isolation of a recombinant pEP19, where the acceptor splice site is of the form 5′-GTTGCTTTTTTCTAG$^{\downarrow}$A-3′. Nuclease S1 mapping experiments have shown that this acceptor is used as efficiently as that present in pEP9, reinforcing the notion that it is the additional dinucleotide AG which renders the acceptor splice site of pEP11 non-functional (data not shown).

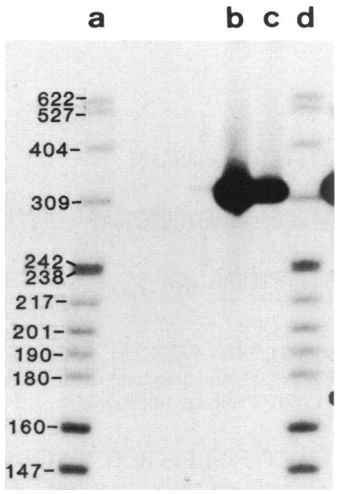

Fig. 8. Synthetic donor and acceptor splice sites can be spliced together. RNA (50 μg) from HeLa cells transfected with pEP10 (independent experiments are shown in **lanes b, c**) was hybridized with probe D (Figure 6d), treated with nuclease S1 and electrophoresed on an 8% polyacrylamide 7 M urea gel. **Lanes a** and **d**, marker bands from an *Msp*II digest of pBR322.

Synthetic donor and acceptor sites can be spliced together

We have shown above (Figure 4A) that the synthetic donor splice site is functional and that the splice point is defined by the invariant GT of the site. We compare in Figure 7B with a sequencing ladder of probe N, the results of a nuclease S1 mapping experiment using the same probe and pEP10 or pEP9 RNAs. The product of splicing synthetic donor and acceptor splice sites should have the sequence 5′-GGATCCAGAGATCT-3′ (Figure 1A) and should protect probe N up to the T residue shown by an arrowhead in Figure 7B. In fact the protected fragment migrates slightly slower than the A residue (dotted) immediately following. As discussed above, this result is nonetheless consistent with correct use of the synthetic acceptor splice site.

Further proof that synthetic donor and acceptor splice sites are spliced together correctly in pEP10 RNA was obtained by using a 'cDNA' probe. Digestion of pEP2 with *Bam*HI and *Xba*I followed by polymerase fill-in and re-ligation generates pEP12 with a sequence 5′-GGATCCTAGAGATCT-3′ (see Figure 1A). There is a single base insertion (underlined T) in this sequence relative to that given above for spliced RNA. A *Pst*I-*Eco*RI fragment of pEP12, 5′ end-labelled at the *Eco*RI terminus (probe D) can thus be used as a 'cDNA' probe for detecting splicing of synthetic donor and acceptor splice sites (nuclease S1 cuts duplexes with single base mismatches very inefficiently). Spliced RNA should protect the probe up to the β-globin gene acceptor splice site as shown in Figure 6d, a distance of 325 nucleotides. Nuclease S1 mapping of pEP10 RNA with probe D does indeed give a protected fragment of this size (Figure 8, lane b and c).

These experiments rule out the possibility that the splice sites are recognised by a nuclease activity which would cut at the donor and acceptor splice sites but that the cut molecules would not be recognised by a ligase activity and not be ligated together. This would be a possible interpretation of the results we obtained previously when testing the donor and acceptor separately, using probes labelled in exon sequence and extending into intron sequence. Using of the cDNA probe, however, assays directly for correctly cut and ligated species: only these can give a protected fragment of 325 nucleotides. Were cutting to occur at the donor and acceptor splice sites without re-ligation, the resulting species could still hybridise to the probe, but the only protected fragment observed after S1 nuclease treatment would be 112 nucleotides.

The efficiency of the splicing reaction is affected by the nature of the sequences separating the synthetic donor and acceptor splice sites

We have constructed a recombinant pEP13 (Figure 6e) in which the synthetic donor and acceptor splice sites are separated by a polylinker fragment from pUC8. The environment of the synthetic donor and acceptor is as shown in Figure 9. The size of the intron in pEP13 is 39 bp. Nuclease S1 mapping using probe D (Figure 6d) of RNA obtained after co-transfection of pEP13 and pEP1 into HeLa cells yields a protected fragment of 325 nucleotides (Figure 10A) as expected for splicing of synthetic donor and acceptor splice sites. However, the amount of spliced RNA obtained is greatly reduced relative to that obtained with pEP10 (Figure 10A). In these experiments and those described below we have measured the efficiency of the transfection by co-transfecting pEP1 and assaying for globin mRNA using probe A (Figure 3A), which generates a nuclease S1-resistant fragment of 54 nucleotides (Figure 10B). The effect of transfection efficiency on the amount of RNA produced using a given plasmid can thus be eliminated by comparing the relative intensities of bands in Figure 10A (probe D) with their counterparts in Figure 10B (probe A).

Increasing the size of the intron by introducing the 526-bp *Hind*III fragment of SV40 DNA present in pEP4 and pEP3 into the *Hind*III site of pEP13, in pEP16, (Figure 6f and 9D) or replacing the 14-bp *Hind*III-*Sal*I fragment of the pEP13 intron by the 621-bp *Hind*III-*Sal*I fragment of pBR322 to generate pEP15 (Figure 9C) did not increase the amount of spliced RNA obtained (Figure 10A and not shown, respectively). Nor did decreasing the intron size to 29 bp by *Hind*III + *Sal*I digestion of pEP13, polymerase fill-in and re-ligation to generate pEP14 (Figure 10A). However, elimination of a 282-bp *Hpa*I-*Hind*III fragment of pEP16 produced pEP17 (Figure 6g and 9E), capable of generating spliced RNA as efficiently as pEP10 (compare Figure 10A and B). The fragment deleted from pEP16 to generate pEP17 lies >200 bp away from the donor splice site, and 36 bp away from the acceptor splice site.

Discussion

We have synthesised an oligonucleotide containing sequences close to the consensus sequences for donor and acceptor splice sites (Figure 1) and shown that the synthetic sites can function when expressed in an RNA polymerase B transcription unit. The donor splice site is spliced to the acceptor splice site whether the bulk intron sequence separating them is derived from either of two fragments of SV40 DNA (neither one lying in SV40 intron sequences), or fragments of various prokaryotic plasmids. The bulk intron sequences do not appear to contribute any information essential for splicing, although their nature can affect the efficiency of the splicing reaction. This result underlines the importance in splicing of the several nucleotides surrounding an intron-exon junction which conform to the consensus sequences.

In one of our constructions (pEP6) the acceptor splice site

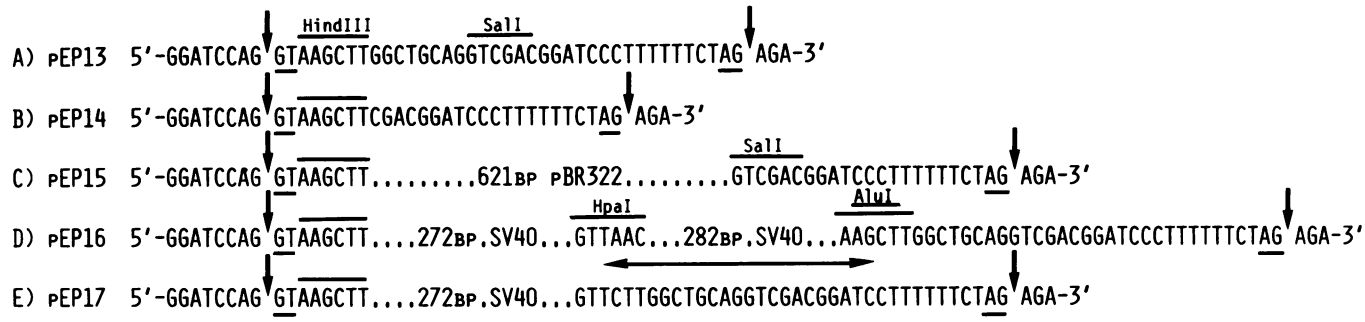

Fig. 9. Sequences of the various introns of **(A)** pEP13, **(B)** pEP14, **(C)** pEP15, **(D)** pEP16 and **(E)** pEP17. *Hind*III (AAGCTT), *Sal*I (GTCGAC), *Hpa*I (GTTAAC) and *Alu*I (AGCT) sites used for constructing different plasmids are shown overlined. In **D** the *Hpa*I-*Alu*I fragment deleted from pEP16 to generate pEP17 is shown by linked arrowheads. Vertical arrows mark splice-points. GT and AG of the GT-AG rule are underlined.

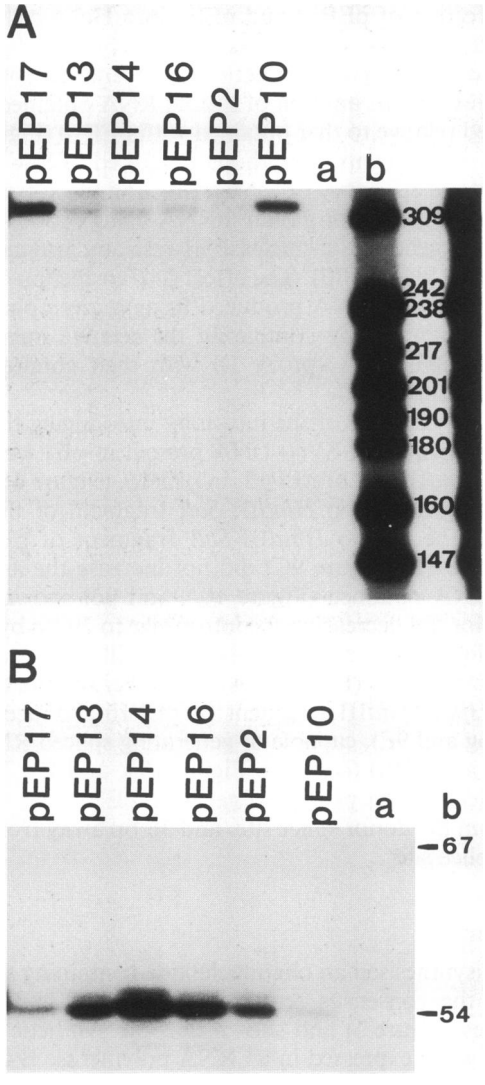

Fig. 10. (A) Nuclease S1 mapping experiment showing the effect of intron sequence on splicing efficiency. RNA (50 μg) from HeLa cells transfected with a mixture of pEP1 and either pEP17, pEP13, pEP14, pEP16, pEP2, pEP10 or not transfected **(lane a)** was hybridized with single-stranded probe D (Figure 6d), treated with nuclease S1, and electrophoresed on an 8% polyacrylamide 7 M urea gel. Marker bands from an *Msp*II digest of pBR322 are shown in **lane b. (B)** Nuclease S1 mapping experiment as in Figure 10A, except that probe A (Figure 3a) was used. **Lane b** shows the position of marker bands of 67 and 54 nucleotides.

sequence 5'-TTCTGCTGTTTGCTCTAG↓AG-3' lies ~120 nucleotides upstream from the sequence 5'-GAATTCACTC CTCAG↓GT-3' contributed from globin gene exon sequences. In this case only the former sequence functions as an acceptor splice site; if the former sequence is absent, however, the latter sequence can function as an acceptor splice site. When the distance between the two sequences is reduced (pEP8), producing the sequence 5'-TTCTGCTGTTTGCTC TAG↓AATTCACTCCTC[AG]↓GT-3', only the cryptic β-globin gene acceptor splice site is used. In yeast genes the sequence 5'-TACTAACA-3' 20–55 nucleotides upstream from the 3' splice site is vital for splicing and apparently activates a downstream AG to function as an acceptor, though not necessarily the first downstream AG (Langford *et al.*, 1984). It is tempting to speculate that the pyrimidine-rich sequence preceding the invariant AG dinucleotide marking the acceptor site in genes of higher eukaryotes fulfils a similar role (Mount and Steitz, 1983). This hypothesis provides a possible explanation for the above results. The cryptic globin gene acceptor splice site cannot compete with the pyrimidine-rich, activated acceptor 5'-TTCTGCTGTTTGCTCTAG↓ AG-3' when the two are separated by 120 nucleotides. Reducing the distance between them brings the activating sequence within reach of the cryptic rabbit β-globin gene acceptor splice site: competition is now between the sequences 5'-AG↓A-3' and 5'-AG↓G-3', and the latter globin sequence, which corresponds more closely to the consensus sequence (Figure 1B), is used exclusively. Once again, the importance of the 'consensus sequences' for splicing is underlined.

tions could be found where the distance between the invariant dinucleotide AG of the acceptor splice site and the first upstream AG was <14 nucleotides. We have determined the reason for this observation, by comparing the functioning as acceptor of the sequence 5'-GTTCTTTTTTCTAG↓A-3' and 5'-GTTAGCTTTTTTCTAG↓A-3' (pEP10 and pEP11, respectively). While the former sequence functions efficiently as an acceptor splice site, the latter does not function at all. An AG 11 nucleotides away can thus 'hide' a downstream acceptor splice site. However, the results obtained with pEP8 confirm that an AG 14 nucleotides upstream from an acceptor splice site does not show this effect.

The results we have obtained underline the importance of the consensus sequence nucleotides: synthetic sequences designed to resemble the consensus sequences will be spliced together even when separated by prokaryotic plasmid se-

quences. This result begs the question as to why the sequences close to the consensus sequences for donor and acceptor splice sites which can be found in eukaryotic genes elsewhere than at intron-exon junctions are not normally used. Our data cannot provide an answer to this question. However, it is interesting to consider the results we have obtained with a series of plasmids (Figures 6 and 9) in which the synthetic donor and acceptor sites are separated by different introns (see Figure 9). In pEP13 (Figure 6), the 39-bp intron sequence is derived from pUC8. In pEP14, the *Hind*III-*Sal*I fragment of pEP13 has been removed, leaving an intron of 29 bp. In pEP15, the *Hind*III-*Sal*I fragment of pEP13 has been replaced by the 621-bp *Hind*III-*Sal*I fragment of pBR322, giving an intron of 644 bp. In pEP16, the 526-bp *Hind*III fragment of SV40 DNA present in pEP4 has been introduced into the *Hind*III site of pEP13, giving an intron of 565 bp. When transfected into HeLa cells, all four plasmids give rise to identical spliced RNAs at the same efficiencies. Varying intron size over the range 29 bp to >600 bp does not affect the amount of spliced RNA produced. We have not investigated plasmids with smaller intron sizes, as the amount of spliced RNA generated using the series pEP13-16 is low (see below). The question as to what is the minimum size of an intron thus remains open. We are testing whether splicing efficiency can be increased in pEP14 by changing the acceptor splice site to the form 5′-AG^{t}G-3′, conforming more closely with the consensus sequence of Figure 1B. If so, further deletions of intron sequence from pEP14 can be undertaken to answer this question. However, when a *Hpa*I-*Alu*I fragment is deleted from the intron of pEP16, the resulting pEP17 will give rise to significantly more spliced RNA than any of the previous four plasmids under defined conditions. The endpoints of the deletion lie 299 nucleotides away from the donor splice site and 36 nucleotides away from the acceptor splice site, in other words in a region well away from the consensus sequence region. Making the assumption that the deletion is not operating by increasing the transcription rate, or stability of unspliced RNA in the nucleus, it seems that sequences distant from splice junctions can have an effect on the efficiency of their utilisation. This result joins the previous observations of Khoury *et al.* (1979) that deletions of SV40 DNA distant from the early messenger splice junctions can affect the relative efficiencies of utilisation of the large T donor splice site *versus* the small t donor splice site. Another relevant observation, made by Dobkin *et al.* (1983), is that of an abnormal splice in a human β-globin gene containing a mutation which does not lie at the normal or cryptic splice sites. This mutation, which changes a T to a G (underlined) in the intron sequence 5′-GAGGTAAGA-3′, apparently inactivates the natural β-globin acceptor splice site 271 bp downstream, and results in the use of a cryptic acceptor splice site. It seems that while conforming to the consensus sequence for a donor or acceptor splice site can be sufficient to confer on a nucleotide sequence the capacity to function as such, whether it actually does so or not depends on factors we do not yet understand.

Materials and methods

Oligonucleotide synthesis

The oligonucleotide P1, d(CGGATCCAGGTAAGCTTTTTTCT), was prepared by a batchwise polydimethylacrylamide phosphotriester method using dimers (Duckworth *et al.*, 1981). The oligonucleotide P2, d(CAGATCTCTAGAAAAAAGCTAC), was made using the column polydimethylacrylamide/kieselguhr phosphotriester method using dimers (Gait *et al.*, 1982). P1 and P2 were purified on 20% polyacrylamide 7 M urea gels, eluted and de-salted. They were annealed, and made fully double-stranded using reverse transcriptase and DNA polymerase. The resulting 32-mer (see Figure 1) was isolated by polyacrylamide gel electrophoresis. Full details for the various steps are available on request.

Plasmid construction

pEP1 (Figure 2) was obtained by cutting pSXβ+ (Banerji *et al.*, 1981) with *Hpa*I and *Sal*I, repairing with DNA polymerase I and re-ligating, thus eliminating a *Hpa*I-*Sal*I fragment from pSXβ+ and leaving a plasmid (pEP1) containing an SV40 enhancer (nucleotides 399−5171, BBB system; Tooze, 1982), a rabbit β-globin gene, and sequences of the plasmid pAT153. pSpL1 (Figure 2) was constructed by introducing the splicing oligonucleotide into the *Pvu*II site of pBR328 (Soberon *et al.*, 1980). pEP2 was obtained by exchanging the *Bam*HI-*Eco*RI fragment of pEP1 with a *Bam*HI-*Eco*RI fragment of pSpL1 (Figure 2). Introduction of a *Hind*III fragment of SV40 (coordinates 3476−4002, BBB system; Tooze, 1982) into the *Hind*III site of pEP2 in both orientations led to pEP3 and pEP4 (Figure 3c,d). Ligation of *Hinc*II digests of pEP1 and pEP3 led to pEP5 and pEP7 (*Hpa*I sites are subsets of *Hinc*II sites, Figure 3e and g). Replacement of the *Hpa*I-*Xba*I fragment of pEP3 (Figure 3c) with a 117-bp.*Hpa*I-*Xba*I fragment of the chicken ovalbumin gene (Benoist *et al.*, 1980) led to pEP6 (Figure 3f). The *Hpa*I site of the ovalbumin fragment lies in intron A of this gene. The *Xba*I site (TCTAGA) lies at the intron A-exon 1 junction, the AG of the *Xba*I site marking the end of intron A in the ovalbumin gene. Cutting pEP6 with *Xba*I + *Eco*RI, DNA polymerase I repair and re-ligation led to pEP8 (Figure 3h). For construction of pEP9 and pEP10 (Figure 6a,b), a 132-bp *Bam*HI-*Eco*RI fragment of pEP2 (Figure 3b) was isolated by sucrose gradient centrifugation, cut with *Alu*I, and ligated to a *Hpa*I + *Eco*RI digest of pEP3 or pEP4. For construction of pEP11 (Figure 6c), pEP2 (Figure 3b) was cut with *Hind*III, repaired with DNA polyermase I, recut with *Eco*RI after destruction of the polymerase by a 20-min incubation at 65°C, and ligated to a *Hpa*I + *Eco*RI digest of pEP4 (Figure 3d). pEP12 (Figure 6d) was obtained by cutting pEP2 (Figure 3b) with *Bam*HI + *Xba*I, DNA polymerase repair and re-ligation. pEP13 (Figure 6e) was constructed in a two-step procedure, by first replacing the *Hind*III-*Eco*RI fragment of pEP2 (Figure 3b) by a *Hind*III-*Eco*RI fragment of pUC8 (Messing and Vieira, 1982, see text) to produce an intermediate plasmid from which a *Sma*I-*Eco*RI fragment was subsequently removed and replaced by the *Alu*I-*Eco*RI fragment of pEP2 used in the preparation of pEP9 and pEP10 (see above). pEP16 (Figure 6f) was generated by introducing the SV40 *Hind*III fragment used for synthesis of pEP4/3 into the *Hind*III site of pEP13. For construction of pEP17 (Figure 6g) a 157-bp *Bam*HI-*Eco*RI fragment of pEP13 was isolated by sucrose gradient centrifugation, cut with *Alu*I and ligated to a *Hpa*I + *Eco*RI digest of pEP4.

Enzymes and techniques used in plasmid constructions were as described elsewhere (O'Hare, 1981; Rautmann *et al.*, 1982).

Nuclease S1 mapping

All techniques involved (transfection of HeLa cells, harvesting cytoplasmic RNA, 3′ and 5′ end-labelling of DNA, hybridization, nuclease S1 digestion and gel electrophoresis) were as described elsewhere (Banerji *et al.*, 1981; Hen *et al.*, 1982).

DNA sequencing

As described by Maxam and Gilbert (1980).

Acknowledgements

We are much indebted to Prof. P.Chambon for useful discussions and a critical reading of the manuscript. We thank M.C.Gesnel, M.Ackner and B.Augsburger for excellent technical assistance. H.W.D.Matthes was supported by an EMBO long term fellowship. This work was supported by grants from the CNRS, INSERM, the MIR (83 V 0092), the Fondation pour la Recherche Médicale and the Association pour le Développment de la Recherche sur le Cancer.

References

Banerji,J., Rusconi,S. and Schaffner,W. (1981) *Cell*, **27**, 299-308.
Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127-142.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Breathnach,R. and Harris,B. (1983) *Nucleic Acids Res.*, **11**, 7119-7136.
Breathnach,R., Benoist,C., O'Hare,K., Gannon,F. and Chambon,P. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4852-4857.
Dierks,P., Wieringa,B., Marti,D., Reiser,J., Van Oyen,A., Meyer,F., Weber,H. and Weissmann,C. (1981) in Brown,D.D. and Fox,C.F. (eds.), *ICN UCLA Symposium on Molecular and Cellular Biology*, **33**, Academic Press, NY.

Dobkin,C., Pergolizzi,R., Bahre,P. and Bank,A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1184-1188.
Duckworth,M.L., Gait,M.J., Godet,P., Hong,G.F., Singh,M. and Titmas, R.C. (1981) *Nucleic Acids Res.*, **9**, 1691-1706.
Gait,M.J., Matthes,H.W.D., Singh,M. and Titmas,R.C. (1982) *J. Chem. Soc. Chem. Commun.*, 37-40.
Hen,R., Sassone-Corsi,P., Corden,J., Gaub,M.P. and Chambon,P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7132-7136.
Khoury,G., Gruss,P., Dhar,R. and Lai,C.J. (1979) *Cell*, **18**, 85-92.
Langford,C.J., Klinz,F.J., Donath,C. and Gallwitz,D. (1984) *Cell*, **36**, 645-653.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 259-276.
Mount,S. (1982) *Nucleic Acids Res.*, **10**, 465-473.
Mount,S. and Steitz,J. (1983) *Nature*, **303**, 380-381.
O'Hare,K. (1981) *J. Mol. Biol.*, **151**, 203-210.
Rautmann,G., Glaichenhaus,N., Nahgashfar,Z., Breathnach,R. and Rassoulzadegan,M. (1982) *Virology*, **122**, 306-317.
Seif,I., Khoury,G. and Dhar,R. (1979) *Nucleic Acids Res.*, **6**, 3387-3398.
Soberon,X., Covarrubias,L. and Bolivar,F. (1980) *Gene*, **9**, 287-305.
Tooze,J., ed. (1982) *DNA Tumor Viruses*, 2nd Edition, Cold Spring Harbor Laboratory Press, NY.
Treisman,R., Orkin,S.H. and Maniatis,T. (1983) *Nature*, **302**, 591-596.
Volkaert,G., Feunten,J., Crawford,L.V., Berg,P. and Fiers,W. (1979) *J. Virol.*, **30**, 674-682.

Received on 6 June 1984; revised on 18 June 1984