The EMBO Journal Vol.2 No.5 pp.727–733, 1983

Evidence against a scanning model of RNA splicing

T.Kühne, B.Wieringa[1], J.Reiser and C.Weissmann*

Institut für Molekularbiologie I, Universität Zürich, 8093 Zürich, Switzerland

Communicated by C.Weissmann
Received on 21 February 1983

A cloned rabbit β-globin gene was modified by introducing a DNA fragment containing the 5′ splice region of the large intron upstream or downstream of its natural counterpart. Analogous constructions were carried out with the 3′ splice region. The genes were linked to SV40 DNA, transiently expressed in HeLa cells and the transcripts analyzed by S1 mapping. In all cases, the splice site further removed from the intron was utilized to the complete exclusion of its counterpart. This finding argues persuasively against a simple scanning model of RNA splicing, in which the splicing enzyme(s) attaches at the 'donor' spliced region and moves along the intron until it encounters the closest 'acceptor' splice region. A model compatible with the currently known facts is presented.

Key words: duplicated splice sites/cryptic splice sites/intron/S1 mapping/splicing

Introduction

Splicing involves the precise recognition of 5′ and 3′ splice sites at the exon-intron junctions of the primary RNA transcript, the cleavage of the RNA at these positions and the correct joining of the exons. Pre-mRNAs frequently contain more than one, occasionally as many as 51 introns (de Crombrugghe and Pastan, 1982; Mount, 1982). This raises the question as to the mechanism which ensures that only neighboring exons are joined and that the excision of one or more exons, along with flanking introns, is avoided (Mount, 1982; Lerner *et al.*, 1980; Lewin, 1980). One model [cf. review by Sharp (1981)] proposes that the splicing system first recognizes the donor splice site [it has not been established whether this is the 5′ or the 3′ splice site, despite current nomenclature (cf. Sharp, 1981)] and then moves along the intron until it encounters a suitable acceptor splice site. Alternatively, attachment could be within the intron, and scanning could occur in both directions. To test this model, we placed an additional copy of the 5′ splice region of the large intron of the rabbit β-globin (RβG) gene either upstream or downstream of the original 5′ splice region. In a second series of experiments the analogous operations were carried out with the 3′ splice region. We argued that if scanning started at a 5′ splice site, use would be made of the upstream copy of the pair of identical 3′ splice regions; if scanning proceeded from 3′ to 5′, then the downstream copy of the 5′ splice site splice region would be used, and if scanning were bidirectional, starting from within the intron, then the downstream copy of the 5′ and the upstream copy of the 3′ splice site would be utilized. Surprisingly, whenever there was a choice between equivalent splice sites, splicing occurred exclusively at the outermost splice sites, arguing persuasively against a simple scanning model.

[1]Present address: Rijksinstituut voor de Volksgezondheid, Antonie van Leeuwenhoeklaan 9, 3720 BA Bilthoven, The Netherlands.

*To whom reprint requests should be sent.

Results

Splicing of β-globin transcripts containing a duplication of the 5′ splice region of the large intron

We wished to determine the effect on splicing of a duplicate 5′ splice region, placed either upstream or downstream of the original counterpart. We had found that neither deletion of the *Mbo*II fragment (position 603–1043) from the large intron of the RβG gene (Dierks *et al.*, 1981b) nor deletion of the segment between position 438 and 477 in the second exon, impaired the accuracy or efficiency of splicing (Wieringa *et al.*, 1982). Therefore, it seemed likely that the segment between the *Alu*I site at position 467 and the *Mbo*II site at position 603 would contain all sequence elements required for a functional 5′ splice region.

The 136-bp *Alu*I-*Mbo*II fragment was provided with *Xba*I linkers and inserted in the correct orientation into *Xba*I sites introduced either upstream (position 467) or downstream (position 718) of the original 5′ splice region. The *Xba*I sites had been generated by introducing an 8-bp *Xba*I linker into appropriate *Alu*I sites (Figure 1). The insertion of linkers did not detectably affect splicing (data not shown). All RβG genes were in the expression vector shown in Figure 2 (Wieringa *et al.*, 1983), consisting of pBR327, a fragment of SV40 containing the origin of replication and the enhancer sequence (Banerji *et al.*, 1981; Moreau *et al.*, 1981), and to provide an internal reference RNA, the mouse β-globin (MβG) gene. DNA was introduced into HeLa cells by the calcium phosphate method, and RNA was prepared 48 h later.

Rabbit and mouse specific β-globin transcripts were analyzed by S1 mapping. Hybridization with a uniformly labelled probe derived from wild-type RβG DNA (Wieringa *et al.*, 1982, 1983) showed that in all cases the signals due to exons I, II and III were present at about the same level (data not shown). The uniformly labelled probe was expected to give a normal signal for exon II regardless of where the splice occurred. Therefore, to determine which of the two 5′ splice sites had been joined to the 3′ splice site, we carried out S1 analyses using terminally labelled probes (Weaver and Weissmann, 1979) derived from the homologous DNA. In each case the probe was a minus strand fragment, 3′-^{32}P-labelled at the *Taq*I site (position 308) in the second exon. RNA from cells transformed with the globin gene containing the 5′ splice region duplication in position 467, i.e., upstream of the original 5′ splice region, protected a fragment of 195 bp (Figure 3, panel A); this is the distance from the *Taq*I site across the *Xba*I linker to the upstream 5′ splice site (cf. Figure 4). No 347-bp signal was seen, showing that no molecules had been spliced only at the original (downstream) 5′ splice site. When the duplicate 5′ splice region was placed downstream of its original counterpart, the RNA gave a 187-bp signal, corresponding to the distance from the *Taq*I site to the original 5′ splice site. This experiment showed that when a choice existed between the two 5′ splice regions, the

more upstream copy was used exclusively.

Figure 3, panel A and Figure 4 show an additional experiment with RβG DNA in which the *Xba*I-linkered 136-bp 5′ splice region replaced the intron segment between position 501 (6 bp downstream of the original 5′ splice site) and position 718 (construction to be described elsewhere). In this case, S1 mapping with the homologous probe gave two fragments (Figure 3), a major one of 228 bp, due to splicing at the downstream duplicate copy of the 5′ splice site, and a minor one of 187 bp, due to splicing at the original 5′ splice site. This experiment suggested that when the original (the upstream) 5′ splice region was shortened close to the splice site, its efficiency as a substrate for the splicing machinery was strongly reduced and the downstream copy of the 5′ splice region was used more efficiently. In any event it is clear from these results that the 136-bp fragment contains all the information required for 5′ splicing, even when it is located in an alien environment.

Splicing of β-globin transcripts containing a duplication of the 3′ splice region

A similar set of experiments as above was performed with the 3′ splice region (Figure 3, panels B, C and D; Figure 4). A 143-bp *Bsp*I fragment extending from position 1024 (in the large intron) to 1167 (in the third exon) (cf. Figure 1) was thought to contain all sequence elements required for recognition as a 3′ splice region. This conclusion was reached because: (a) deletion of an intron fragment from position 603 to 1043 did not affect splicing (Dierks *et al.*, 1981b) and (b) a *Bam*HI (position 477)-*Bgl*II (position 1196) fragment containing the large intron and 18 bp of 5′- and 133 bp of 3′-flanking DNA, when inserted into the *Bgl*II site of an intronless RβG gene (D.Valenzuela, G.McAllister, B.Wieringa, unpublished results) allowed efficient splicing of the large intron. The 143-bp *Bsp*I fragment was provided with *Xba*I linkers and inserted into each of the four *Xba*I sites indicated in Figure 4. These had been generated by introducing an *Xba*I linker into the *Alu*I sites at positions 467, 718 and 1067, respectively, and into the blunted *Eco*RI site in position 1117. In all cases, the introduction of the *Xba*I linker did not affect splicing (data not shown). It should be noted that insertion of the *Xba*I linker (lower case letters) at position 1067 (precisely into the 3′ splice junction) reconstituted the sequence

1067 1068

CAG/CTC (----CAGctctagagCTC----)

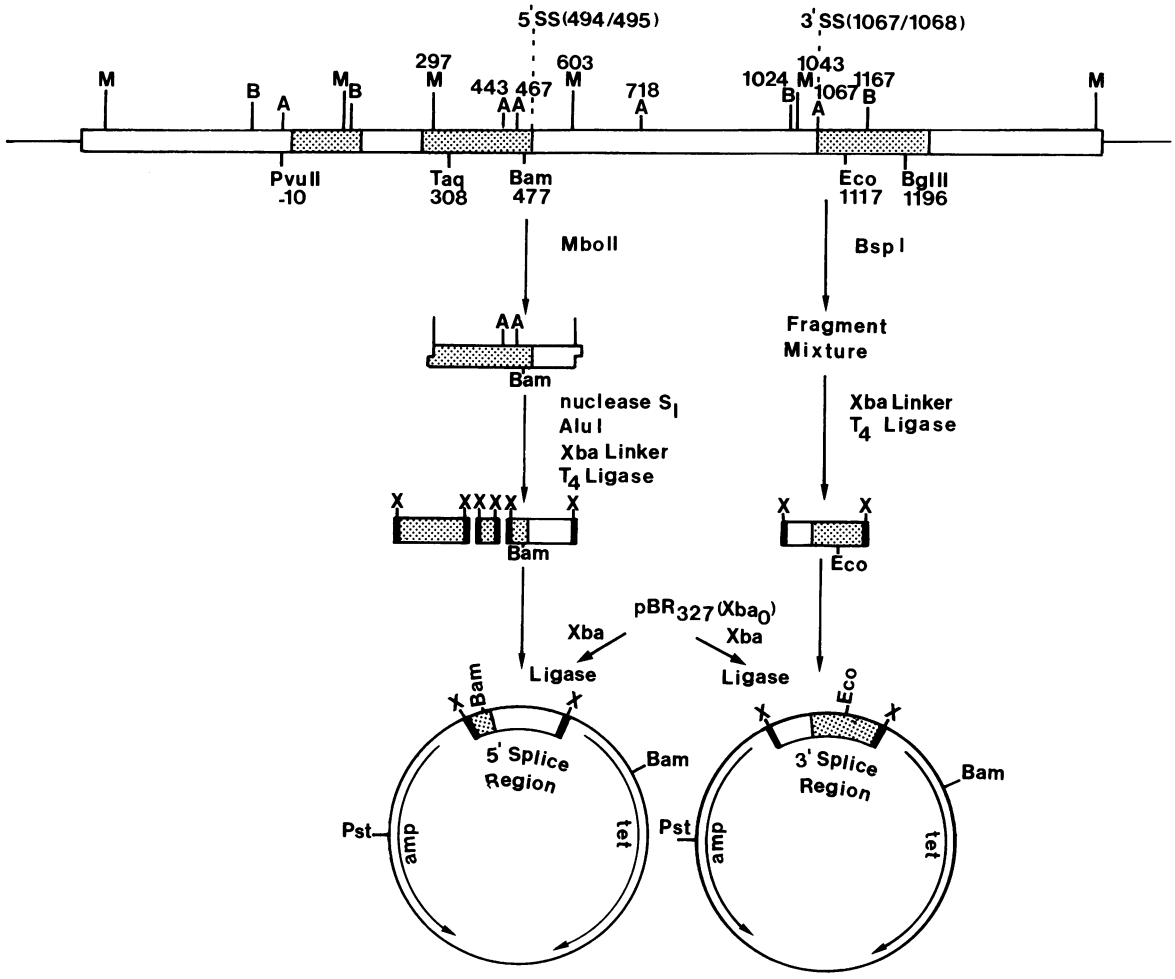

Fig. 1. Strategy for cloning 5′ and 3′ splice site regions of the large intron of RβG DNA. **Left**: pBR322/RβGΔ425B DNA was cut with *Mbo*II, the 306-bp fragment containing the 5′ splice region isolated and blunted with S1 nuclease. After cleaving with *Alu*I, *Xba*I linkers were added and the linkered fragment mixture was ligated to *Xba*I-cleaved pBR327(Xba0). The hybrid DNA was cloned and plasmids containing the 5′ splice region were identified. **Right**: pBR327/RβGΔ425B DNA was cut with *Bsp*I. *Xba*I linkers were joined to the fragments and the 143-bp fragment containing the 3′ splice region was isolated and cloned as above. Shaded areas, exon sequences; white areas, introns; black bars, *Xba*I linkers; lines, pBR sequences; M, *Mbo*II; B, *Bsp*I; A, *Alu*I; X, *Xba*I; amp, ampicillinase resistance; tet, tetracycline resistance. For experimental details, see Materials and methods.

found at the intron-exon junction of wild-type RβG DNA.

In a first experiment (Figure 4), the 3′ splice region was inserted into position 467, within the second exon. We had previously demonstrated the existence of cryptic 5′ splice sites at positions 359/360, 442/443 and 498/499 (Wieringa *et al.*, 1982, 1983); if indeed the 'portable' 3′ splice region were functional, we might expect to find a splice extending from one of the upstream cryptic 5′ splice sites (or perhaps the 5′ splice site of the small intron) to the new 3′ splice site. S1 mapping of the cognate RNA with a uniformly labelled minus strand probe from wild-type RβG DNA gave a 133-bp signal corresponding to the first and third exons, at a level ~50% of wild-type, but no signal for the second exon (Figure 3, panel C). This could mean that an RNA section had been spliced out of the second exon, and that the residual exon segments were too short to give rise to protected probe fragments. Alternatively, the bulk of the second exon along with the small intron could have been spliced out. To determine the 3′ boundary of the spliced-out region we used as probe a *Bam*HI-*Pvu*II fragment in which the *Bam*HI terminus of the minus strand was 5′-^{32}P-labelled. As shown in Figure 3, panel D, this probe yielded a fragment of ~121 bp, which is equivalent to the sum of the distance from the *Bam*HI site (position 477) to position 467 (where the *Xba* linker was inserted), 8 bp of *Xba*I linker and 100 bp of the 3′ splice region fragment (from its 3′ end to the 3′ splice site). Thus, a splice had occurred at the site normally used in the inserted fragment. To determine the 5′ boundary of the spliced-out region we used a *Taq*I-*Bgl*II probe prepared from wild-type RβG DNA in which the *Taq*I terminus (position 308) of the minus strand was 3′-^{32}P-labelled. As shown in Figure 3 (panel B), a signal of 52 bp was obtained, showing that splicing had occurred around position 360 of RβG DNA, i.e., at the location of one of the cryptic splice sites (359/360) identified earlier (Wieringa *et al.*, 1983). A second cryptic splice site further downstream (442/443) was not utilized at a detectable level. In a control experiment, S1 mapping of RNA transcribed from RβG DNA with an *Xba*I linker in position 718, yielded the expected signal of 187 bp (Figure 3, panel B). As judged from the relative intensities of the signals obtained in upstream mapping from the *Bam*HI site and downstream mapping from the *Taq*I site, other modes of splicing could only play a minor role. This experiment shows that the 143-bp 3′ splice region fragment contains the information required to direct efficient splicing when inserted into an abnormal position.

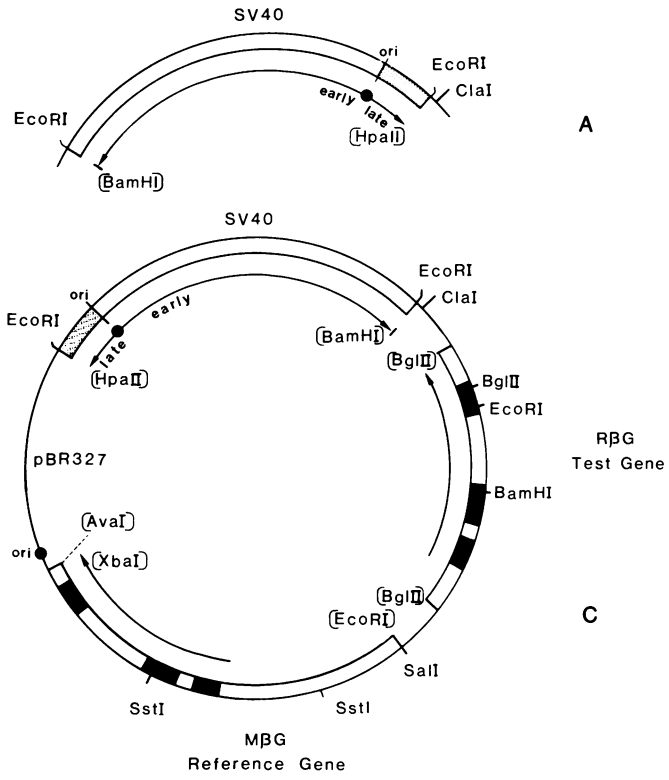

Fig. 2. Expression plasmid for assaying RβG transcription and RNA splicing. Construction and properties of the pBR327/SV40/MβG/RβG expression plasmid are detailed in Wieringa *et al.* (1983). The 'test gene', a wild-type or modified RβG gene, was introduced into the vector as a *Sal*I-*Cla*I fragment. The 'reference gene' is the 3.1-kb *Eco*RI-*Xba*I fragment of the MβG DNA isolated by Leder and his colleagues (van Ooyen *et al.*, 1979; Tilghman *et al.*, 1977). The arrows indicate directions of transcription. A and C differ in the orientation of the SV40 fragment.

The other locations tested (mentioned above) for insertion of the 3′ splice region were upstream, downstream, or just at the authentic 3′ splice site (Figure 4). As mentioned above, insertion of the *Xba*I linker into the *Alu*I site, which coincides with the original 3′ splice sites restored three nucleotides of the sequence originally following the 3′ splice site, and did not impair splicing. Thus, insertion of the 3′ splice region fragment into this *Xba*I site in fact constituted a downstream duplication (Figure 4).

The RNA derived from the three constructions, when S1 mapped with a uniformly labelled wild-type chromosomal minus strand probe extending from the *Pst*I to the *Bgl*II site, gave rise to the signals diagnostic for exon I (145 bp), exon II (222 bp) and exon III (133 bp) (Figure 3, panel C) at about the levels observed for wild-type RβG DNA. Although the RNAs have insertions and form loops when hybridized to the wild-type DNA probe, the probe is not cleaved opposite the loops under the S1 digestion conditions used (Favaloro *et al.*, 1980). This assay showed that the level of spliced transcripts was similar for all constructions, and that no gross abnormalities in splicing occurred.

To determine which of the two 3′ splice sites was used, each RNA was S1 mapped with the homologous *Bam*HI-*Bgl*II fragment 5′-^{32}P-labelled at the *Bgl*II terminus of the minus strand. The RNA from the construction with the duplicate 3′ splice region inserted upstream of the authentic 3′ splice site (nominal position 718) gave a single signal of ~133 nucleotides (Figure 3, panel D). As shown in Figure 4, this means that splicing occurred exclusively at the authentic 3′ splice site. When the duplicate splice regions were inserted downstream of the authentic 3′ splice site, at positions (nominally) 1067 or 1120, S1 mapping gave fragments of ~241 and 192 nucleotides, respectively, showing that in both cases only the duplicate splice sites were utilized. Thus, in all three cases the more downstream of the 3′ splice sites was chosen.

Discussion

We had reasoned that if the scanning model of splicing held true, then the more downstream of a pair of equivalent 5′ splice regions would be utilized if scanning proceeded from 3′ to 5′, and conversely, the more upstream of a pair of 3′ splice regions would be preferred if scanning proceeded from 5′ to 3′. Surprisingly, in all constructions examined, it was always the more distal (relative to the intron) of a pair of

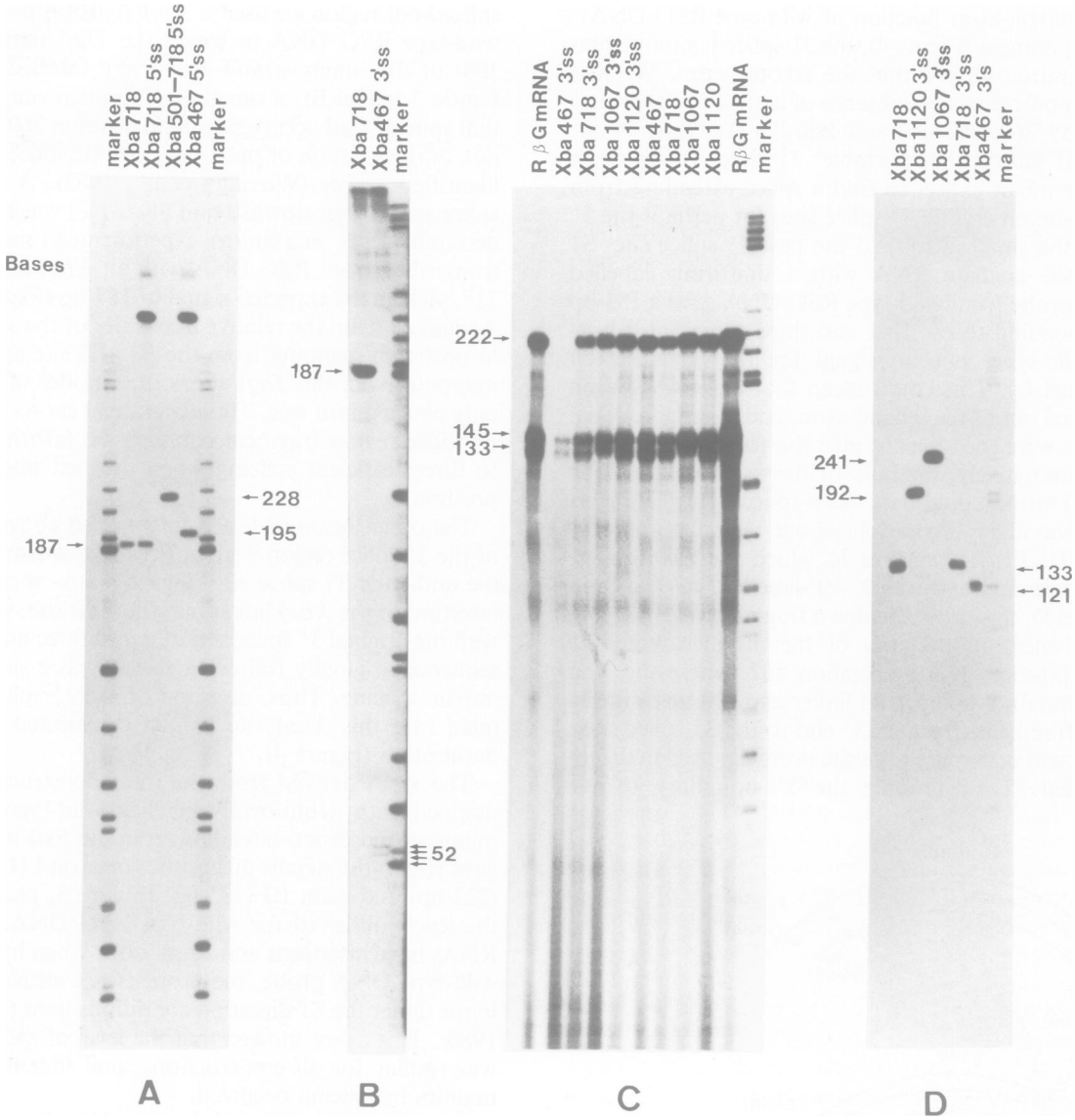

Fig. 3. S1 mapping of splice sites utilized in RβG transcripts containing duplicate splice regions. **Panel A**: mapping of the 5′ splice site utilized in RβG RNA with duplicate 5′ splice regions. Nuclease S1 mapping was carried out with 15 μg of total cellular RNA from HeLa cells transformed with the expression vector (shown in Figure 2) containing the various altered RβG genes. In each case, RNA was hybridized to a homologous *Taq*I-*Bgl*II fragment (van Ooyen *et al.*, 1979), 3′-^{32}P-labelled at the *Taq*I site. Sample designations are above each lane of the autoradiogram. XbaN, *Xba*I linker inserted in position N; XbaN 5′ss, XbaN 3′ss, 5′ splice region fragment and 3′ splice region fragment, respectively, inserted in the *Xba*I site introduced in position N; Xba501-718 5′ss, 5′ splice region fragment replacing intron segment from position 501 to 718 (see also Figure 4). Marker, ^{32}P-labelled *Bsp*I fragments of pBR322. **Panel B**: Xba467 3′ss has a duplicate 3′ splice region in the second exon (position 467). To determine whether splicing occurred at this position and to which site it extended, 15 μg RNA were mapped with a 888-bp *Taq*I-*Bgl*II probe, 3′-^{32}P-labelled at the *Taq*I terminus. **Panel C:** mapping of RβG transcripts with a uniformly labelled minus strand probe. Uniformly ^{32}P-labelled RβG minus strand DNA extending from the *Bgl*II site (position 1197) to position −425 was prepared as described (Wieringa *et al.*, 1982), hybridized to 15 μg of each of the various RNAs indicated and treated with nuclease S1 and DNase I. Protected signals correspond to exon II (222 nucleotides), exon I (145 nucleotides) and part of exon III (133 nucleotides) of RβG RNA. RβG mRNA (**left lane**, 200 ng; **right lane**, 400 ng) was from reticulocytes. Designation of samples is as described under **panel A**. **Panel D**: mapping of 3′ splice sites in RβG RNAs with 5′-^{32}P-labelled probes. Each sample (15 μg total cellular RNA) was mapped with the homologous *Bgl*II-*Bam*HI fragment (first four lanes) or *Bam*HI-*Pvu*II fragment (fifth lane) 5′-^{32}P-labelled at the *Bgl*II and *Bam*HI sites, respectively (see Figure 4). Samples are designated as above.

equivalent splice regions that was utilized. Only in one case (cf. Figure 4), where the distal splice region was severely truncated by the insertion of the proximal duplicate, was the proximal splice site used with great preference. These results argue strongly against a simple scanning model; they also raise the question as to why a duplicate splice region in a distal unnatural location is preferred to its authentic counterpart. Several explanations can be considered. (i) The initial splice may occur between the closest 5′ and 3′ splice sites, and the new junction may reconstitute a new splice site which can then splice to the more distal splice site. This is quite unlikely, because the new junction formed by splicing out the large intron (Figure 4) had no resemblance to a splice region consensus sequence (Mount, 1982). It could be argued that cryptic (Wieringa *et al.*, 1982, 1983; Dierks *et al.*, 1981b) 5′ and 3′ splice sites within the large intron became functional after insertion of an 'unpaired' splice region and that, in a first step, an internal pair of splice sites (for example, the natural 5′ splice site and a cryptic 3′ splice site) were utilized exclusively, generating a splicing intermediate which was subsequently spliced at a pair of external splice sites (in this example, a duplicate upstream 5′ splice site and the natural

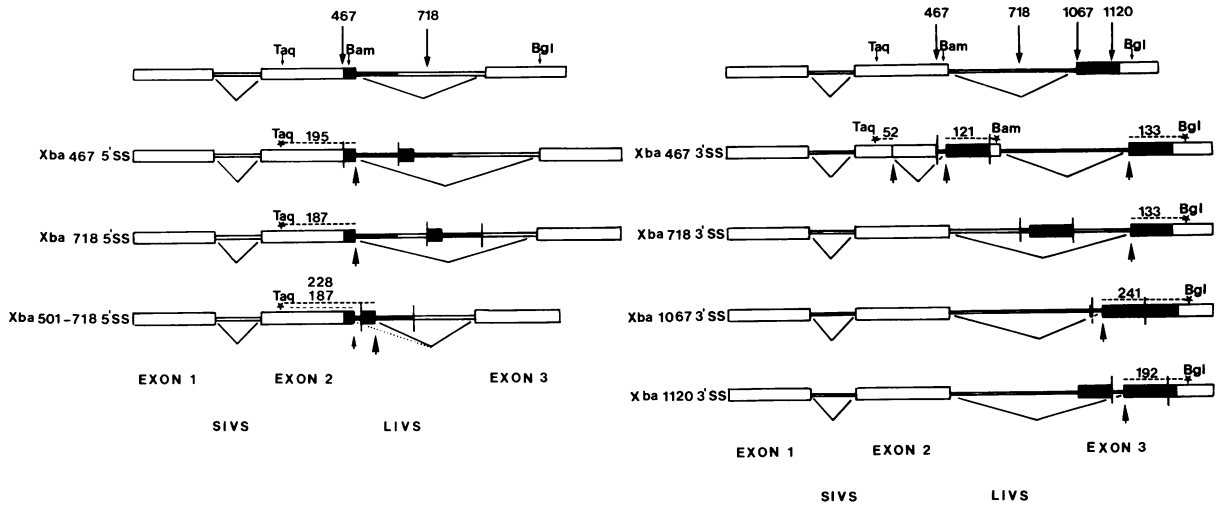

Fig. 4. The effect of supernumerary splice sites in different positions of the RβG gene. Schematic diagrams of the splicing events in RβG transcripts that contain either a duplicate 5′ splice region or a duplicate 3′ splice region of the large intron. The 5′ (136 bp) and 3′ (143 bp) splice regions used for duplication are indicated in black. The *Xba*I linkers flanking the duplicated splice regions are given by vertical lines. The dashed lines above each RNA indicate the fragment protected by S1 mapping, the star represents the ^{32}P-labelled end of the probe and the arrowhead the splice site that was mapped. Regions removed from the primary transcripts by splicing events are indicated under each RNA by an angled line.

3′ splice site). It is however difficult to explain why the addition of a duplicate splice region distal to a natural splice site would cause the latter to splice to a cryptic splice site rather than to the natural one. If, nevertheless, such a mechanism were operative, a scanning mechanism proceeding from within the intron towards the splice sites would remain a possibility. However, we have generated deletions in the large β-globin intron which removed either the middle two thirds of the intron (position 604–1044), the first half (position 501–719, leaving only six nucleotides next to the 5′ splice site) or the last half (position 719–1053, leaving only 15 nucleotides next to the 3′ splice site); as accurate and efficient splicing still occurred (B.Wieringa, unpublished results), it seems unlikely that an internal intron region is essential for splicing. (ii) The splicing enzyme attaches to the RNA far upstream and downstream of the intron, respectively, and splices at the first pair of splice sites encountered – a scanning mechanism opposite to the one discussed above. Clearly, such a model raises more questions than it answers in the case of RNAs containing more than one intron. (iii) Both the duplicated 5′ and 3′ splice regions were shown to be active in splicing even when placed in an alien environment. Nevertheless, what appears to be a self-contained splice region may in fact still be influenced by more distant sequences, and a duplicate splice region fragment may separate the authentic splice region from these sequences, or affect it adversely by its presence. For instance, in the constructions *Xba* 467 5′ ss and *Xba* 1120 3′ ss (Figure 4) the inserted splice regions regain the complete exon sequences which normally flank them (including the *Xba*I linkers, which were shown not to disturb splicing) and might therefore be used in preference to the truncated, authentic splice regions. Previous experiments, in which deletions or insertions were introduced in regions flanking an intron, had not shown an impairment of splicing, perhaps because a small decrease in splicing efficiency cannot be readily picked up unless measured in a competition type experiment. (iv) The splice regions and their duplicates used in our experiments have the same affinity for the splicing enzyme; however, the position of the splice region within the three-dimensional structure determines its reactivity; the fact

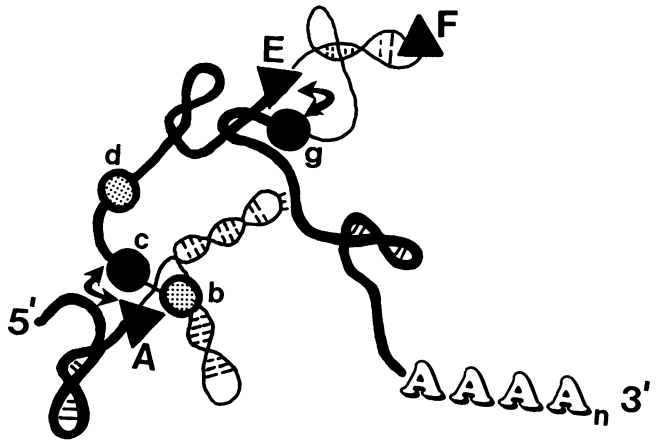

Fig. 5. A model for splice site selection. The model is explained using a hypothetical pre-mRNA as example. Heavy lines, exons; light lines, introns; stacked lines, hydrogen bonds; triangles, '5′ splice enzyme'; circles, '3′ splice enzyme'; light shading, weak binding; heavy shading, strong binding. Double-headed arrow indicates preferred splicing interaction: when the two complexes join, the splicing enzymes are activated and splicing ensues. Under normal conditions, splicing occurs between A(5′) and c(3′), and E(5′) and g(3′), respectively. Splicing is rare or non-existent between A(5′) and b(3′) or d(3′) because the complex b is unstable and therefore occurs infrequently, and d is unfavourably located compared with c. If c is mutationally inactivated, A(5′) may splice at a lower rate to b(3′) or d(3′), which are thereby revealed as cryptic splice sites. Although the sequence of F(5′) is typical for a 5′ splice region, it is so unfavourably placed that it is never involved in a splicing reaction even if the authentic splice site at E is inactivated, and thus does not constitute a cryptic splice site.

that in our experiments the more distal splice region was preferred may be incidental as regards its linear position. Khoury *et al.* (1979) have suggested earlier that the secondary or tertiary RNA structure may influence the location and efficiency of a splicing event. (v) Both the proximal and the distal copy of a pair of duplicate splice sites are used; however, the transcripts resulting from the use of the proximal copy are rapidly degraded, perhaps because a segment of intron RNA is retained. Even if this explanation is correct, it is unlikely that a major fraction of the primary transcripts is spliced at the proximal splice site and destroyed, because the overall

transcript level is not noticeably reduced by the insertion of a duplicate splice region. Thus, the argument against a processive scanning mechanism is not affected.

A model for splice site selection

A pre-mRNA may contain many introns and many potential or cryptic splice sites (Wieringa *et al.*, 1982), besides the sites normally used. Nonetheless, in most cases splicing seems to occur predominantly, if not exclusively, between a single pair of appropriate 5′ and 3′ splice sites. We assume (cf. Figure 5) that splice site usage is influenced by two factors: (a) the nucleotide sequence of the splice regions, which determines the binding affinity for the splicing enzymes; this affinity may be modulated by secondary and tertiary structure as regards the availability of the region for binding, (b) the relative position of a pair of 5′ and 3′ splice regions, which determines the probability of the splicing reaction occurring between them. The frequency of a splice occurring at a particular pair of sites would thus depend both on the stability of the enzyme splice region complexes and their spatial relationship.

In Figure 5 we make the *ad hoc* assumption that the splicing system consists of two enzymes, one binding to the 5′ and the other to the 3′ splice regions, and that the enzymes bound to their substrates remain enzymatically inactive until a 5′ and a 3′ binding complex collide. Alternatively, the splicing enzyme could be a single particle, with the capacity of binding both a 5′ and a 3′ splice region. It has been suggested (Lerner *et al.*, 1980; Rogers and Wall, 1980) that U1 RNA, as part of a nucleoprotein, forms hydrogen bonds to the 5′ and 3′ ends of an intron and thereby lines up the sequences that are to be spliced. Davies *et al.* (1982) have proposed a model for the splicing of fungal mitochondrial transcripts in which a segment of intron hydrogen bonds to and aligns the exon ends that are to be joined, and a set of base pairings within the intron is essential to maintain an appropriate conformation. In our model there is neither a requirement for a unique conformation of the pre-mRNA, nor for an RNA to line up the exons; it is possible that recognition of splice regions is mediated by nucleoproteins.

Within the frame-work of this model, alternative splicing as found in a mutant β-globin RNA (Busslinger *et al.*, 1981), calcitonin (Amara *et al.*, 1982), growth hormone (de Noto *et al.*, 1981) and IgD (Moore *et al.*, 1981) mRNA may come about because two 5′ (or 3′) splice sites have comparable splicing efficiencies; lesser affinity for the splicing enzyme can be compensated by a more favorable spatial position, and *vice versa*. Preference for one or the other alternative splice site may be imparted by proteins or factors that change the spatial structure of the RNA or cover the splice regions. Cryptic splice sites are normally not used because they do not compete effectively with the true splice sites; they are revealed when a true splice site is deleted or its affinity diminished by mutations, or if a new splice region is generated or introduced in the vicinity (Orkin *et al.*, 1982).

The model predicts that it should be possible to change the splicing pattern of a pre-mRNA by modifying its spatial structure by appropriately placed deletions or insertions. In this connection, it is interesting to note that natural mutations which prevented correct splicing of one intron of a pre-mRNA slowed down splicing of another (Busslinger *et al.*, 1981; Treisman *et al.*, 1982; Benyajati *et al.*, 1982). Perhaps, at least in some cases, the conformational change of the RNA which results after splicing of one intron facilitates the splicing of other introns.

Materials and methods

Plasmids

RβG DNA was cloned as a partial (2.1-kb) *Bgl*II fragment in the *Bam*HI site of pBR322 [pBR322/RβGΔ425B, (Dierks *et al.*, 1981a)] or of pBR327 (pBR327/RβGΔ425B, unpublished). The expression vector pBR327/SV40/MβG/RβG was constructed as described (Wieringa *et al.*, 1983). Plasmid DNAs were cloned and propagated in *Escherichia coli* HB101 and purified as described in Wieringa *et al.* (1983).

Enzymes and nucleotides

Restriction enzymes were from New England Biolabs, except for *Eco*RI and *Cla*I (Boehringer) and were used as recommended by the manufacturers. Polynucleotide kinase, T4 DNA ligase and exonuclease III were from New England Biolabs, nuclease S1 from P-L Biochemicals and calf intestine alkaline phosphatase from Boehringer. Labelled [α-^{32}P]dCTP was from Amersham, and [γ-^{32}P]ATP was prepared according to Johnson and Walseth (1979).

Acceptor plasmid for XbaI fragments

pBR327 (Soberon *et al.*, 1981) was linearized with *Eco*RI, blunted by filling in with T4 DNA polymerase (Challberg and Englund, 1980), dephosphorylated and purified by electrophoresis through a 1% low gel temperature (LGT) agarose gel (Biorad). About 20 ng (0.06 pmol) of this DNA in agarose (Frischauf *et al.*, 1980) and 2.5 pmol of 5′-phosphorylated *Xba*I linker ($^{CTCTAGAG}_{GAGATCTC}$, Collaborative Research) were ligated in a 10 μl reaction mixture using 400 units of T4 DNA ligase (Maniatis *et al.*, 1982). The hybrids were cloned and plasmids with an *Xba*I site flanked by two reconstituted *Eco*RI sites were identified [pBR327(Xba0)].

Insertion of XbaI linkers into AluI and EcoRI sites in RβG DNA

(A) As a first step towards these constructions, the *Eco*RI site in the pBR327 moiety of pBR327/RβGΔ425B was obliterated, while retaining the site in the RβG moiety. pBR327/RβGΔ425B was linearized by partial *Eco*RI cleavage, dephosphorylated and blunted with nuclease S1. Full-length linear DNA was purified on a 0.6% LGT agarose gel and circularized with T4 DNA ligase. After cloning in *E. coli* HB101, a plasmid with a single *Eco*RI site in the RβG moiety was identified (pBR327Eco^R/RβGΔ425B).

(B) Insertion of *Xba*I linkers into the *Alu*I sites at positions 467, 718 or 1067 of RβG DNA. pBR327Eco^R/RβGΔ425 (18 μg) was linearized by partial *Alu*I digestion and incubated with a 3-fold molar excess of 5′-phosphorylated *Xba*I linker and T4 DNA ligase. The mixture was digested with *Xba*I and electrophoresed through a 0.8% LGT agarose gel. Full-length linears were circularized with T4 DNA ligase and transfected into *E. coli* HB101. Approximately 1000 colonies were pooled and plasmid DNA was prepared; >90% of this DNA was cleavable by *Xba*I. To obtain a plasmid with an *Xba*I linker in the *Alu*I site at position 467, the plasmid pool was cleaved with *Pvu*II (position −10) and *Bam*HI (position 477), and the 490-bp fragment joined to the large *Pvu*II-*Bam*HI fragment of pBR327/RβGΔ425. The resulting DNA was cloned and the desired plasmid identified by restriction mapping with *Xba*I and *Eco*RI. To obtain plasmids with *Xba*I linkers in the *Alu*I positions 718 or 1067, the 640-bp *Bam*HI-*Eco*RI fragments from the plasmid pool were joined to the large *Bam*HI-*Eco*RI fragment of pBR327Eco^R/RβGΔ425B and the desired plasmids identified by restriction mapping with *Bam*HI, *Eco*RI and *Xba*I.

(C) Insertion of an *Xba*I linker at the *Eco*RI site (position 1117) of the RβG gene. pBR327Eco^R/RβGΔ425B was cut with *Eco*RI, dephosphorylated and the ends filled in with Klenow DNA polymerase I (Wartell and Reznikoff, 1980). The DNA was ligated to 5′-phosphorylated *Xba*I linkers and digested with *Xba*I. Full length linear DNA was isolated from a 0.8% LGT agarose gel and circularized with T4 DNA ligase.

Introduction of RβG fragments containing XbaI linkers into the expression plasmid

*Sal*I-*Bgl*II fragments (containing 275 bp of pBR and most of the RβG DNA) were isolated from each of the four plasmids containing a single *Xba*I linker at a unique position and ligated (Frischauf *et al.*, 1980) together with the *Cla*I-*Bgl*II fragment (containing the 3′-untranslated region of RβG DNA) to the large fragment of *Cla*I-*Sal*I cleaved pBR327/SV40/MβG/RβG expression vector DNA (see Figure 2 and Wieringa *et al.*, 1983). This yielded plasmids pBR327/SV40/MβG/RβG(Xba467),-/RβG(Xba718), -RβG-(Xba1067) and -/RβG(Xba1120).

Introduction of the 136-bp AluI-MboII fragment containing the 5′ splice region of the large intron into the XbaI-substituted RβG expression plasmids

About 1.5 μg of the 306-bp *Mbo*II (van Ooyen *et al.*, 1979) fragment (position 297 – 603) of pBR322/RβGΔ425B were purified as described (Weaver and Weissmann, 1979). The fragment was blunted with S1 nuclease, purified and cleaved with *Alu*I (at positions 443 and 467). About 30 ng (0.9 pmol ends) of the digest were ligated to 7.5 pmol of the 5′-phosphorylated *Xba*I linker described above, cleaved with *Xba*I and purified. Acceptor plasmid pBR327(Xba0) was cleaved with *Xba*I, dephosphorylated and linear DNA (30 ng) isolated on a 1% LGT agarose gel. About 20 ng of the *Alu-Mbo*II fragment mixture bearing *Xba*I ends were joined to the acceptor DNA in agarose (Frischauf *et al.*, 1980). The DNA was cloned and plasmids having the *Bam*HI site [position 477 in the RβG gene (van Ooyen *et al.*, 1979)] indicative for the fragment containing the 5′ splice region were further characterized. *Xba*I cleavage released the 136-bp *Alu*I (position 467)-*Mbo*II position (603) fragment, now flanked by *Xba*I sites (total length, 148 bp). The *Xba*I fragment containing the 5′ splice region was inserted into the *Xba*I-cleaved, dephosphorylated plasmids pBR327/SV40/MβG/RβG(Xba467), -/RβG(Xba718) and -/RβG(Xba502-718) (see text and Figure 1). Plasmids containing the fragment in the correct orientation were identified by restriction analysis with *Bam*HI and *Xba*I.

Cloning of the 143-bp fragment containing the 3′ splice region

Plasmid pBR327/RβGΔ425 (0.15 pmol) was cleaved with *Bsp*I (a gift from A.Kiss), purified and ligated to 5′-phosphorylated *Xba*I linkers [13 pmol (Maniatis *et al.*, 1982)]. The DNA was cleaved with *Bsp*I and *Xba*I and the fragments were separated on a 2.5% LGT agarose gel. The 143-bp *Bsp* fragment (position 1024 – 1167) flanked by *Xba*I sites (total length 155 bp) was ligated to the *Xba*I-cleaved, phosphatase-treated pBR327(Xba0) acceptor plasmid and cloned. Plasmid DNA containing an additional *Eco*RI site, due to the RβG DNA fragment, was identified and characterized. The 143-bp *Xba*I fragment containing the splice region was excised and inserted into the *Xba*I sites of the plasmids pBR327/SV40/MβG/RβG(Xba467), -/RβG(Xba718), -/RβG(Xba1067) and -/RβG/(Xba1120). Plasmids containing the fragment in the correct orientation were identified by restriction analysis with *Eco*RI and *Xba*I.

Transient transformation of HeLa cells

Plasmids were introduced into HeLa cells as described (Wieringa *et al.*, 1982, 1983). Total RNA was prepared 48 h after DNA addition (Auffray and Rougeon, 1980).

S1 mapping

Uniformly labelled probes were prepared by cleaving pBR327/SV40/MβG/RβG DNA with *Sal*I, treating with exonuclease III and repairing with DNA polymerase I using [α-^{32}P]dCTP and the other dNTPs, as described (Wieringa *et al.*, 1982). End-labelled probes were prepared either by 5′-terminal labelling with polynucleotide kinase and [γ-^{32}P]ATP (Maxam and Gilbert, 1977), or by filling in 3′ ends with Klenow DNA polymerase I using [α-^{32}P]dCTP and the three other dNTPs (Wartell and Reznikoff, 1980). Probes were electrophoresed through LGT agarose gels and recovered as described (Wieslander, 1979). Probe (0.01 pmol, 3 x 10^5 – 2 x 10^6 c.p.m./pmol) and RNA sample (10 – 25 μg of total RNA) were hybridized in 10 μl Pipes-formamide buffer (Maniatis *et al.*, 1982; Weaver and Weissmann, 1979) under paraffin oil in an Eppendorf tube for 16 h at 49°C. After digestion with nuclease S1 (150 units/ml) in a 250 μl reaction mixture for 45 min at 30°C, protected probes were analyzed on a 6.5% acrylamide-urea sequencing gel (Weaver and Weissmann, 1979; Berk and Sharp, 1977). When uniformly labeled probe was used (Wieringa *et al.*, 1982, 1983), incubation with DNase I prior to S1 digestion was sometimes performed to destroy interfering DNA-DNA hybrids. Detailed information about this procedure is available on request.

Acknowledgements

We thank J.Fujisawa and P.Dierks for [γ-^{32}P]ATP, and P.Dierks for useful comments on the manuscript. This work was supported by the Schweizerische Nationalfonds (5.481.79) and the Kanton of Zürich.

References

Amara,S.G., Jonas,V., Rosenfeld,M.G., Ong,E. and Evans,R.M. (1982) *Nature,* **298**, 240-264.
Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.,* **107**, 303-314.
Banerji,J., Rusconi,S. and Schaffner,W. (1981) *Cell,* **27**, 299-308.
Benyajati,C., Place,A.R., Wang,N., Pentz,E. and Sofer,W. (1982) *Nucleic Acids Res.,* **10**, 7261-7272.
Berk,A.J. and Sharp,P.A. (1977) *Cell,* **12**, 721-732.
Busslinger,M., Moschonas,N. and Flavell,R.A. (1981) *Cell,* **27**, 289-298.
Challberg,M.D. and Englund,E.T. (1980) in Grossmann,L. and Moldave,K. (eds.), *Methods in Enzymology,* Vol. **65**, Academic Press, NY, pp. 39-43.
Davies,R.W., Waring,R.B., Ray,J.A., Brown,T.A. and Scazzocchio,C. (1982) *Nature,* **300**, 719-724.
de Crombrugghe,B. and Pastan,I. (1982) *Trends Biochem. Sci.,* **7**, 11-13.
de Noto,F.M., Moore,D.D. and Goodman,H.M. (1981) *Nucleic Acids Res.,* **9**, 3719-3730.
Dierks,P., van Ooyen,A., Mantei,N. and Weissmann,C. (1981a) *Proc. Natl. Acad. Sci. USA,* **78**, 1411-1415.
Dierks,P., Wieringa,B., Marti,D., Reiser,J., van Ooyen,A., Meyer,F., Weber,H. and Weissmann,C. (1981b) in Brown,D.D. and Fox,C.F. (eds.), *Developmental Biology Using Purified Genes, ICN-UCLA Symposium on Molecular and Cellular Biology,* Vol. **23**, Academic Press, NY, pp. 347-366.
Favaloro,J., Treisman,R. and Kamen,R. (1980) in Grossmann,L. and Moldave,K. (eds.), *Methods in Enzymology,* Vol. **65**, Academic Press, NY, pp. 718-749.
Frischauf,A.M., Garoff,H. and Lehrach,H. (1980) *Nucleic Acids Res.,* **8**, 5541-5549.
Johnson,R.A. and Walseth,T.E. (1979) *Adv. Cyclic Nucleotide Res.,* **10**, 135-167.
Khoury,G., Gruss,P., Dhar,R. and Lai,C.-J. (1979) *Cell,* **18**, 85-92.
Lerner,M.R., Boyle,J.A., Mount,S.M., Wolin,S.L. and Steitz,J.A. (1980) *Nature,* **283**, 220-224.
Lewin,B. (1980) *Cell,* **22**, 324-326.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) in *Molecular Cloning,* published by Cold Spring Harbor Laboratory Press, NY.
Maxam,A.N. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 560-564.
Moore,K.W., Rogers,J., Hunkapiller,T., Early,P., Nottenburg,C., Weissman,I., Bazin,H., Wall,R. and Hood,L.E. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 1800-1804.
Moreau,P., Hen,R., Wasylyk,B., Everett,R., Gaub,M.P. and Chambon,P. (1981) *Nucleic Acids Res.,* **9**, 6047-6068.
Mount,S.M. (1982) *Nucleic Acids Res.,* **19**, 459-471.
Orkin,S.H., Kazazian,H.H.,Jr., Antonarakis,S.E., Ostrer,H., Goff,S.C. and Sexton,J.P. (1982) *Nature,* **300**, 768-769.
Rogers,J. and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 1877-1879.
Sharp,P.A. (1981) *Cell,* **23**, 643-646.
Soberon,X., Covarrubias,L. and Bolivar,F. (1981) *Gene,* **9**, 287-305.
Tilghman,S.M., Tiemeier,C.D., Polsky,F., Edgell,M.H., Seidman,J.G., Leder,A., Enquist,L.W., Norman,B. and Leder,P. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 4406-4410.
Treisman,R., Proudfoot,N.J., Shander,M. and Maniatis,T. (1982) *Cell,* **29**, 903-911.
van Ooyen,A., van den Berg,J., Mantei,N. and Weissmann,C. (1979) *Science (Wash.),* **206**, 337-344.
Wartell,R.M. and Reznikoff,W.S. (1980) *Gene,* **9**, 307-319.
Weaver,R. and Weissmann,C. (1979) *Nucleic Acids Res.,* **6**, 1175-1193.
Wieringa,B., Meyer,F., Reiser,J. and Weissmann,C. (1982) in O'Malley, B.W. (ed.), *Gene Regulation,* Academic Press, NY, pp. 65-85.
Wieringa,B., Meyer,F., Reiser,J. and Weissmann,C. (1983) *Nature,* **301**, 38-43.
Wieslander,L. (1979) *Anal. Biochem.,* **98**, 305-309.