Specific transcription and RNA splicing defects in five cloned β-thalassaemia genes

Richard Treisman*, Stuart H. Orkin† & Tom Maniatis*

* Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA
† Children's Hospital Medical Center, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

Transcriptional analysis of five different cloned β-thalassaemia genes introduced into cultured mammalian cells revealed specific defects in transcription and RNA splicing. A single base change 87 base pairs to the 5′ side of the mRNA cap site significantly lowers the level of transcription and therefore appears to represent a promoter mutation. Three genes contain different single base changes in the first intervening sequence (IVS) 5′ splice site. One mutation, at IVS1 position 1, inactivates the splice site completely; the other two, at IVS1 positions 5 and 6, reduce its activity. Each mutation activates the same three cryptic splice sites. The fifth gene contains a single base change within IVS2 at position 745, which results in the formation of abnormal β-globin RNA that contains an extra exon.

THE human globin genes provide a good model for the investigation of genetic disease at the molecular level. Much is known about the structure and function of the globin proteins and the temporal regulation of their synthesis during development (reviewed in refs 1, 2). Both the human α- and β-globin gene clusters have been isolated and the DNA sequences of each gene determined. In addition, the structures of the major RNA transcripts of each gene are known[3–6]. This structural information is complemented by a wealth of clinical, genetic and molecular data concerning hereditary abnormalities of globin gene structure and expression[2,3,5,7]. Such abnormalities may alter the regulation of globin gene expression during development, leading to conditions such as the hereditary persistence of fetal haemoglobin (HPFH). Alternatively, the aberrant expression of individual α- or β-globin genes causes thalassaemias, severe anaemias characterized by imbalanced synthesis of α- and β-globin polypeptide chains.

The β-thalassaemias comprise a group of diseases in which the synthesis of normal β-globin polypeptide is either absent or reduced, designated β^0 and β^+ thalassaemias, respectively. The heterogeneity of these diseases at the clinical level suggested the existence of many different β-thalassaemia alleles, and this prediction has been confirmed by subsequent structural studies. Sequence analysis of β-globin genes from normal and thalassaemic individuals revealed that the β-thalassaemia phenotype is usually associated with point mutations or deletions of a few nucleotides; in rare cases, the third exon of the gene is deleted (reviewed in ref. 7). These mutations are superimposed on a complex pattern of sequence polymorphisms within and surrounding the β-globin gene[8]; in some cases, they may be detected directly in genomic DNA as a result of the alteration of restriction endonuclease recognition sites[8,9].

Two different approaches have been used to investigate the effect of thalassaemia mutations on globin gene expression: analysis of RNA synthesis in the erythroid cells of thalassaemia patients, or analysis of the expression of the cloned mutant gene after its introduction into cultured cells on a suitable vector. Both approaches have demonstrated abnormal globin RNA processing in cases of β^{10-17} and α^{18} thalassaemia. Here we analyse the expression of five different β-thalassaemia alleles[8] after their introduction into cultured cells on vectors derived from SV40[19] and bovine papillomavirus (BPV)[20].

β-Thalassaemia alleles analysed

We analysed the expression of five β-thalassaemia (β^{thal}) alleles of known DNA sequence after their introduction into cultured cells. In addition to various common sequence polymorphisms[8], these genes contain single base changes which may be responsible for the β-thalassaemia phenotype (Fig. 1*A*): (1) a C→G transversion 87 base pairs (bp) to the 5′ side of the mRNA cap site[8]; (2) a G→A transition at position 1 of the first intervening sequence (IVS*1*)[8]; (3) a G→C transversion at IVS*1* position 5 (S.H.O. and H.H. Kazazian, unpublished results); (4) a T→C transition at IVS*1* position 6[8]; and (5) a C→G transversion at IVS*2* position 745[8].

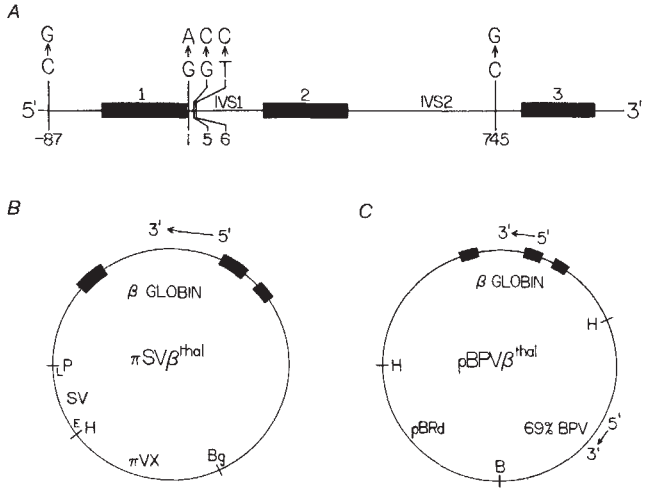

Fig. 1 ***A***, β-thalassaemia genes analysed in this study. The sequences of all genes except the IVS*1* position 5 mutant have been described elsewhere[8]. This mutant, which was isolated and sequenced using a previously described strategy[8], has sequence polymorphisms at exon 1 position 59 C→T, and IVS*2* positions 16 C→G, 74 G→T, and 666 T→C (S.H.O. and H.H. Kazazian, unpublished data). ***B***, $\pi SV\beta^{thal}$, the SV40-derived plasmids used for expression studies in HeLa cells. These plasmids carry a 3.7-kbp *Bgl*II+*Pst*I fragment containing each of the mutants or a normal β-globin gene inserted between *Bgl*II and *Pst*I sites in the miniplasmid (B. Seed, manuscript in preparation) vector πSVHPplac (P. Little, unpublished). SV, the SV40 *Pvu*II–*Hin*dIII fragment containing sequences necessary for efficient β-globin gene expression in HeLa cells[21]; πVX, the small tRNA *supF*-containing plasmid constructed by Seed (manuscript in preparation) ; P, *Pst*; Bg, *Bgl*II; H, *Hin*dIII (destroyed in construction of πSVHPplac). The reference plasmid πSVHPα2 was constructed by subcloning a 1.5-kbp *Pst*I fragment of human DNA containing the human α_1 globin gene into the *Pst*I site of πSVHPplac. ***C***, $pBPV\beta^{thal}$, bovine papillomavirus-derived expression plasmids. These plasmids comprise a 7.6-kbp *Hin*dIII fragment containing the β-globin gene inserted at the *Hin*dIII site of the vector pBPVH11, which carries the 69% subgenomic transforming *Bam*HI+*Hin*dIII fragment of bovine papillomavirus[20]. Polarity of the BPV and globin transcription units is indicated by arrows. H, *Hin*dIII; B, *Bam*HI.

0028-0836/83/150591—06$01.00

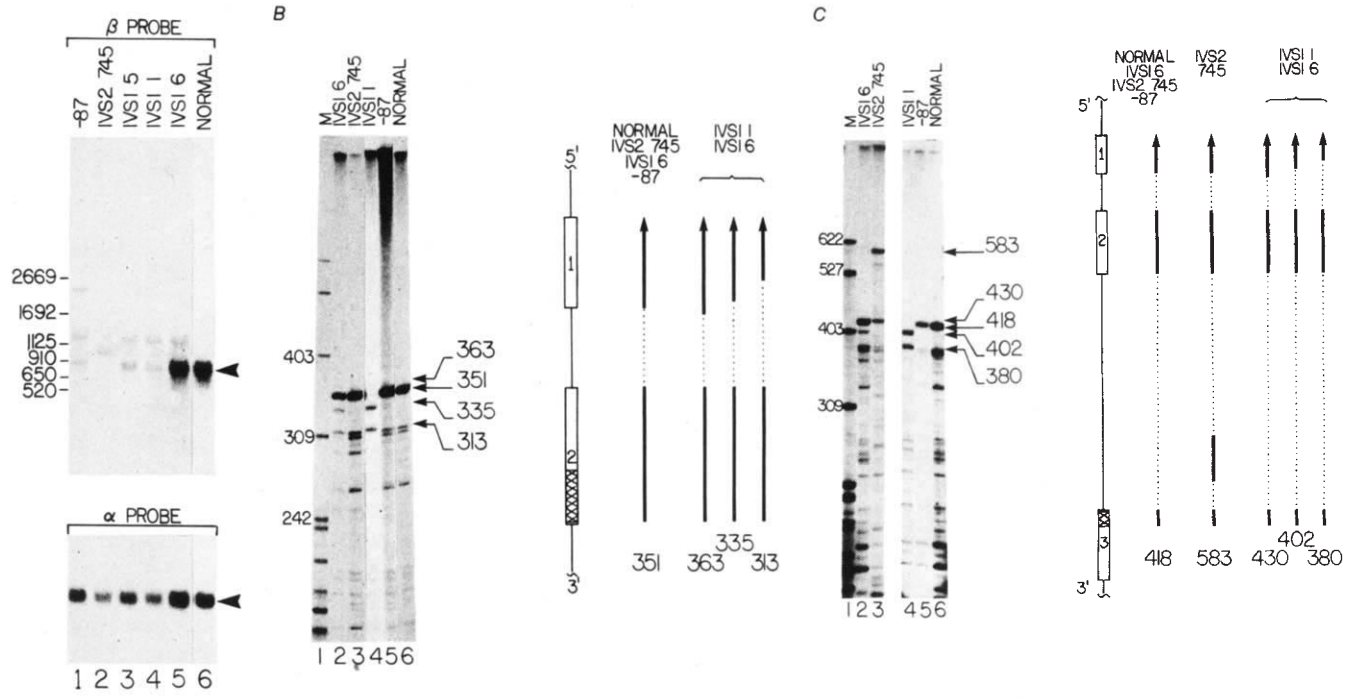

Fig. 2 ***A***, β-globin specific RNAs produced by the different $\pi SV\beta^{thal}$ plasmids in HeLa cells. 20 μg of total cellular RNA was fractionated on a 1.4% agarose gel containing formaldehyde, transferred to nitrocellulose and hybridized as described previously[47] with ^{32}P-labelled probes specific for either β (upper panel) or α (lower panel) globin transcripts produced by the co-transfected reference plasmid αSVHPα2. Lanes 1–6, RNA produced by the −87, IVS*2* position 745, IVS*1* position 5, IVS*1* position 1, IVS*1* position 6 and normal β-globin alleles, respectively. On long autoradiographic exposure a small amount of apparently normal RNA produced by the IVS*2* position 745 mutant was visible. The lengths of fragments of pBR322 DNA run in parallel as size markers are indicated. ***B***, Primer extension analysis of polyadenylated cytoplasmic RNA from BPV-globin transformed cell lines using an exon 2 primer (the antisense strand of the *Hae*III + *Bam*HI fragment, comprising nucleotides 210–287, 5′-end-labelled with ^{32}P). Experimental protocols were as previously described[19]. In each experiment, 3 μg RNA were used, except in the case of the IVS*1* position 1 mutant, where 6 μg were used. The cDNA products were displayed on a 4% polyacrylamide sequencing gel. The 363-nucleotide product is clearly visible on long autoradiographic exposure of the gel (data not shown). Lane 1, *Msp*I fragments of pBR322 DNA as size markers; lanes 2–5, the IVS*1* position 6, IVS*2* position 745, IVS*1* position 1 and −87 mutants, respectively; lane 6, the BPV-globin cell line β11C, which carries a normal β-globin gene[20]. A schematic representation of the cDNA synthesis is shown at the right: exons are shown as open boxes with the primer sequences cross-hatched; cDNA products generated in each experiment are shown as arrows. ***C***, Primer extension analysis of polyadenylated cytoplasmic RNA from BPV-globin transformed cell lines using an exon 3 primer (the antisense strand of the *Bst*NI + *Eco*RI fragment, comprising nucleotides 6–53, 5′-end-labelled with ^{32}P). Analyses were performed as above, and the various extension products shown schematically at the right. The 430-nucleotide product is clearly visible on long autoradiographic exposure of the gel (data not shown).

Vectors used to obtain efficient β-globin gene expression

Initially we used a transient assay to investigate the expression of each of the mutant genes, carried on plasmids of the type shown in Fig. 1*B* ($\pi SV\beta^{thal}$ plasmids). These plasmids carry a small segment of SV40 DNA required in *cis* for the efficient expression of cloned β-globin genes in HeLa cells[21]. HeLa cells were transfected[22,23] with equimolar amounts of each plasmid and a reference plasmid carrying the human α_1-globin gene (plasmid $\pi SVHP\alpha 2$; see legend to Fig. 1*B*) to provide an internal standard for transfection efficiency and RNA recovery. As a control, we performed parallel transfections with a plasmid carrying a normal β-globin gene[23,24]. RNA was analysed 36–50 h after transfection.

In an alternative approach to the analysis of β^{thal} gene expression, we generated permanent lines of morphologically transformed mouse cells using β-globin gene carrying derivatives of the BPV-derived vector pBPVH11[20]. Human β-globin genes carried on this vector are maintained extrachromosomally at about 30 copies per cell[20]. Plasmids carrying each mutant gene ($BPV\beta^{thal}$ plasmids; Fig. 1*C*) were used to transform mouse C127 cells and individual foci were picked and grown in mass culture without recloning. Analysis of the DNA in each cell line showed that the majority of the $BPV\beta^{thal}$ DNA was extrachromosomal and indistinguishable from the input DNA. A variable proportion (up to 30%) of the plasmids consisted of deletion variants lacking pBRd[20] sequences and sequences located to the 3′ side of the β-globin gene; however, the gene itself was unaffected (data not shown). The structure of the β-globin RNA produced in these cells is identical to that of RNA produced by globin genes introduced into HeLa cells on the SV40-derived vectors[20]; we therefore used $BPV\beta^{thal}$ cell lines as a convenient source of large amounts of β^{thal} globin RNA.

Transcriptional analysis of the mutant β-globin genes

We used a variety of standard RNA mapping techniques to characterize the RNA produced by the mutant genes. The amounts of β-globin-specific RNA produced in HeLa cells by the various mutant genes present were examined by RNA blotting and nuclease S_1 mapping, using the transcripts of the co-transfected α-globin plasmid as a reference standard. We did not attempt to measure the relative levels of RNA produced by the mutant genes in the $BPV\beta^{thal}$ cell lines, because these lines were not cloned and contained different proportions of nontransformed cells. The spliced structures of the various RNAs were analysed using the primer extension assays shown in Fig. 2. Short radiolabelled single-stranded DNA primers derived from either exon 2 (Fig. 2*B*) or exon 3 (Fig. 2*C*) were hybridized to the RNA and extended with unlabelled deoxynucleoside triphosphates using reverse transcriptase. Correctly

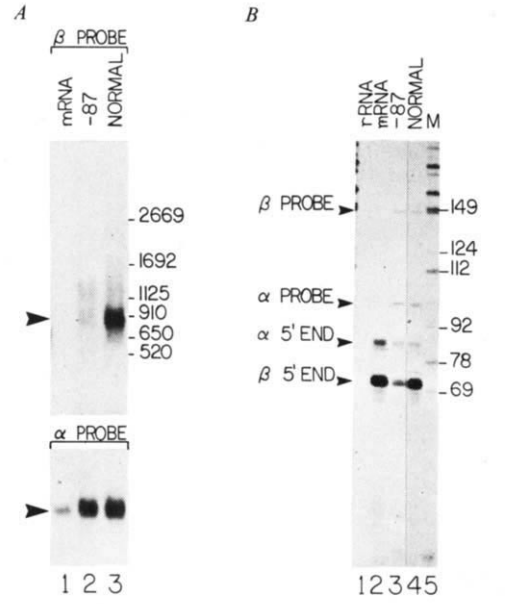

Fig. 3 Transcripts produced in HeLa cells by derivatives of the normal and −87 mutant $\pi SV\beta^{thal}$ plasmids containing only 128 nucleotides of β-globin 5′-flanking sequence. These plasmids, constructed by standard procedures, lack all β-globin sequences to the 5′ side of an *Rsa*I site at position −128. HeLa cells were transfected with each plasmid plus the reference plasmid $\pi SVHP\alpha 2$ which carries the human α_1 globin gene. ***A***, Total cellular RNA was analysed by RNA blotting as described in Fig. 2*A* legend. Hybridization with a β-globin specific probe is shown in the upper panel, and with an α-globin-specific probe in the lower panel. Lane 1, 20 μg untransfected HeLa cell RNA plus mRNA from human cord blood; lanes 2, 3, 20 μg of total cell RNA from HeLa cells transfected with the −87 mutant and the normal β-globin genes, respectively. The lengths of pBR322 fragments run in parallel as size markers are indicated. ***B***, 5′ end analysis of these RNAs by S_1 nuclease mapping[48]. Protocols were as previously described[19], using a mixture of 5′ end ^{32}P-labelled[48] single-stranded probes specific for β (*Mst*II + *Hae*III, nucleotides −76 to +70) and α (*Bst*NI, nucleotides −19 to +80) globin transcripts. Lane 1, 40 μg of total RNA from HeLa cells; lane 2, 40 μg of HeLa cell RNA plus mRNA from human cord blood; lanes 3, 4, 40 μg of total cell RNA from HeLa cells transfected with the −87 mutant and the normal β-globin gene, respectively; lane 5, marker fragments of pBR322 DNA. Nuclease-resistant fragments mapping the 5′ ends of β- and α-globin RNAs (70/71 and ~85 nucleotides respectively), and the intact probes, are indicated.

initiated and spliced β-globin mRNA should yield extension products 351 and 418 nucleotides in length from the exon 2 and exon 3 primers, respectively: these products were indeed generated when polyadenylated cytoplasmic RNA from the BPV-globin cell line β11C[20], which contains a normal human β-globin gene, was used (Fig. 2*B*, *C*, lanes 6). Cytoplasmic RNA from HeLa cells transfected with the normal gene also generated these products (data not shown). The predictions of the primer extension assays were confirmed using nuclease S_1 mapping and cDNA sequencing.

The −87 mutant: This β^{thal} gene produces correctly initiated and spliced β-globin RNA, but at a lower level than with the normal gene. Primer extension analysis of polyadenylated cytoplasmic RNA from the $BPV\beta^{thal}$ cell line V5B1, which carries this gene, is shown in Fig. 2. Primers derived from exons 2 or 3 generated extension products characteristic of correctly initiated and spliced β-globin RNA (351-nucleotide product, Fig. 2*B*, lane 4; 418-nucleotide product, Fig. 2*C*, lane 4). Nuclease S_1 mapping confirmed these results, and demonstrated the correct position of the polyadenylated 3′ end of these transcripts (data not shown). RNA blotting experiments demonstrated that this gene produces about 10-fold less RNA than the normal gene in HeLa cells (Fig. 2*A*, lane 1). The $\pi SV\beta^{thal}$ plasmids used in this experiment contained approximately 1.5 kilobase pairs (kbp) of uncharacterized β-globin 5′ flanking sequences. To demonstrate that the low level of transcription observed in the −87 mutant was not the result of an additional mutation in these sequences, we examined the transcription in HeLa cells of deleted derivatives of the −87 mutant and the normal β-globin gene containing only 128 nucleotides of 5′ flanking sequence. Both RNA blotting (Fig. 3*A*) and nuclease S_1 mapping (Fig. 3*B*) experiments demonstrated that in this case the −87 mutant again produces about 10-fold less correctly initiated RNA in HeLa cells than the normal β-globin gene. In addition, some larger RNAs were produced at low levels by both genes: we did not characterize these RNAs further. We conclude that the −87 mutation is responsible for the decreased level of correctly initiated RNA.

The IVS*1* position 1 mutant: This β^{thal} mutation completely inactivates the IVS*1* 5′ splice site, which leads to the utilization of three nearby cryptic (normally inactive) 5′ splice sites. RNA blot analysis showed that in HeLa cells this gene produces some RNA that co-migrates with RNA produced by the normal gene, but at a 10–20-fold lower level (compare Fig. 2*A*, lanes 4, 6). We analysed the spliced structures of the RNA produced by this gene in the $BPV\beta^{thal}$ cell line P5C2, using the primer extension assay (Fig. 2*B*, *C*). The exon 2 primer generated three abnormal products 313, 335 and 363 nucleotides in length rather than the expected 351-nucleotide product (Fig. 2*B*, lane 4). Nuclease S_1 mapping showed that the 5′ ends of the RNAs were normal, and that the IVS*1* 3′ splice site was active (data not shown); the RNAs that generate these products must therefore be spliced using novel IVS*1* 5′ splice sites. Nuclease S_1 mapping located these splice sites at GT dinucleotides at exon 1 positions 105 and 127 and IVS*1* position 13 (Fig. 4*A*, lane 6; Fig. 4*B*, lane 8). The exon 3 primer generated products of the lengths expected for RNAs in which IVS*2* is excised correctly, while IVS*1* is excised using the novel 5′ splice sites (380, 402 and 430 nucleotide products, Fig. 2*C*, lane 4). This conclusion was confirmed by nuclease S_1 mapping experiments (data not shown).

The IVS*1* position 5 and 6 mutants: Both of these β^{thal} genes produce normally spliced RNA and three abnormal RNAs identical to those produced by the IVS*1* position 1 mutant. An RNA blot analysis of RNA produced by these genes in HeLa cells is shown in Fig. 2*A*: the position 6 mutant produced almost normal amounts of RNA, and the position 5 mutant about half as much RNA as the normal β-globin gene (compare lanes 3, 5, 6). Primer extension analysis of the structure of the RNAs produced by the IVS*1* position 6 mutant in the $BPV\beta^{thal}$ cell line C3C1, which carries this gene, is shown in Fig. 2. With the exon 2 primer, a prominent 351-nucleotide product characteristic of correctly initiated and spliced RNA was generated; in addition, three new products identical to those generated by the IVS*1* position 1 mutant were observed (313, 335 and 363 nucleotide products, Fig. 2*B*, lane 2). Nuclease S_1 mapping experiments confirmed that these abnormal extension products are specified by RNAs spliced at the same cryptic 5′ splice sites as the IVS*1* position 1 mutant (compare Fig. 4*A*, lanes 4, 6; Fig. 4*B*, lanes 8, 9). Using the exon 3 primer, we demonstrated that IVS*2* was correctly excised from all transcripts, generating in addition to the normal 418-nucleotide extension product, products identical in size to those observed in the case of the IVS*1* position 1 mutant (compare Fig. 2*C*, lanes 2, 4). Nuclease S_1 mapping studies of RNA produced in HeLa cells by the IVS*1* position 5 mutant revealed that this gene uses the same IVS*1* 5′ splice sites as the IVS*1* position 6 mutant; however, in this case a greater proportion of the RNA is abnormally spliced (compare Fig. 4*B*, lanes 7, 9).

The IVS*2* position 745 mutant: This β^{thal} gene produces a major RNA product containing 165 nucleotides of IVS*2* inserted between exons 2 and 3 in addition to a small amount of normal β-globin RNA. RNA blotting analysis showed that the gene produces about half as much RNA as the normal β-globin

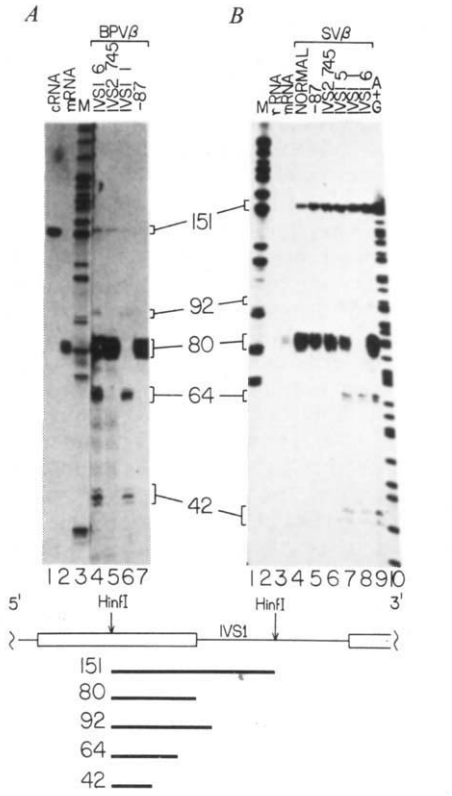

Fig. 4 Nuclease S_1 mapping[48] of the IVS*1* 5′ splice sites in the normal and mutant β-globin genes, using as probe the 3′-end ^{32}P-labelled antisense strand of the *Hin*fI fragment spanning nucleotides exon 1 62 to IVS*1* 68. Experimental protocols were as previously described[19]; the results are indicated schematically below the figure. ***A***, Analysis of RNA from BPVβ^{thal} cell lines. Lane 1, 40 μg of HeLa cell rRNA plus a continuous transcript of the β-globin gene synthesized *in vitro* (this serves as a control for nuclease sensitivity of continuous hybrids); lane 2, 40 μg of HeLa cell RNA plus mRNA from human cord blood; lane 3, *Msp*I cut pBR322 size markers; lanes 4–7, 3 μg of polyadenylated cytoplasmic RNA from BPVβ^{thal} cell lines carrying the IVS*1* position 6, IVS*2* position 745, IVS*1* position 1 and −87 β^{thal} alleles, respectively, plus 40 μg of HeLa cell rRNA. ***B***, Analysis of RNA produced by the various πSVβ^{thal} plasmids in HeLa cells. Lane 1, *Msp*I-cut pBR322 size markers; lane 2, 40 μg untransfected HeLa cell RNA; lane 3, 40 μg of HeLa cell RNA plus mRNA from human cord blood; lanes 4–9, 40 μg total cell RNA from HeLa cells transfected with SV40 plasmids carrying the normal, −87, IVS*2* position 745, IVS*1* position 5, IVS*1* position 1 and IVS*1* position 6 alleles, respectively; lane 10, partial chemical degradation products[49] of the probe run as homologous size markers, allowing mapping of each splice site to within one nucleotide.

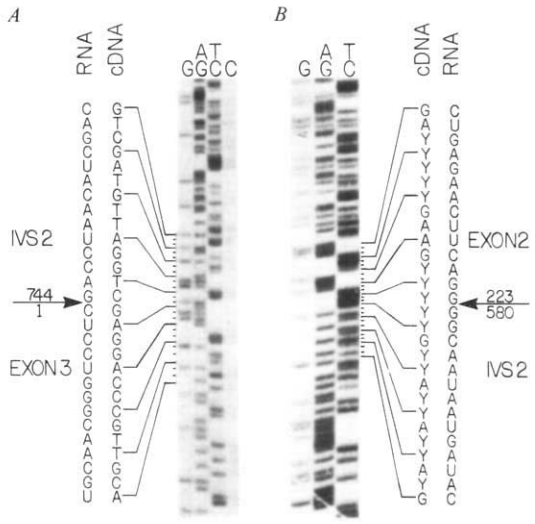

Fig. 5 Sequence analysis of the abnormal transcript produced by the IVS*2* position 745 mutant. The primer extension described in Fig. 2*C* was repeated on a preparative scale using 150 μg polyadenylated cytoplasmic RNA from the BPVβ^{thal} cell line G1C2. The 583-nucleotide cDNA product was subjected to partial chemical degradation[49] and the products fractionated on 8% (*a*) and 6% (*b*) polyacrylamide sequencing gels. ***A***, Sequence across the novel splice joining IVS*2* nucleotide 744 to exon 3. ***B***, Sequence across the novel splice joining exon 2 to IVS*2* nucleotide 580. See Fig. 7 for summary.

gene in HeLa cells (Fig. 2*A*, lane 2). Figure 2 shows primer extension analysis of polyadenylated cytoplasmic RNA from the BPVβ^{thal} transformed cell line G1C2, which carries this gene. Extension of the exon 2 primer generated a product of 351 nucleotides, indicating correct initiation and IVS*1* excision in all transcripts (Fig. 2*B*, lane 3). These conclusions were confirmed by S_1 nuclease mapping (data not shown). However, extension of the exon 3 primer yielded a predominant product of 583 nucleotides, in addition to a small amount of the 418-nucleotide product characteristic of normal β-globin RNA (Fig. 2*C*, lane 3). Nucleotide sequence analysis of the 583-nucleotide cDNA product (Fig. 5) showed that it was specified by an RNA containing IVS*2* nucleotides 580–744 inserted between exons 2 and 3. Thus, in addition to creating a 5′ splice site at IVS*2* nucleotide 745, this mutation results in the activation of a cryptic 3′ splice site at IVS*2* nucleotide 579 (see Fig. 6).

The −87 mutation creates a defective promoter

The C→G transversion 87 nucleotides 5′ to the mRNA cap site seems to be a naturally occurring promoter mutation: in our assay, this gene produces correctly spliced β-globin RNA but at a considerably lower level than the normal gene. Sequences in the −87 to −100 region of several genes are required for efficient RNA synthesis, including those for rabbit β-globin[25–27], herpes simplex virus thymidine kinase[28] and the SV40 early genes[29,30]. The −87 mutation lies within the sequence 5′ ACACCC 3′, which is conserved in both the mouse and rabbit β-globin genes[4,31] and is necessary for efficient transcription of the rabbit gene when carried on a polyomavirus vector[25]. Other sequences that have been shown to be required for efficient transcription of the rabbit β-globin gene carried on SV40- and polyomavirus-derived vectors include the conserved 'CCAAT' and 'TATA' elements[4,32] located 75 and 31 bp to the 5′ side of the mRNA cap site, respectively[25–27]. The recent detection of a β-thalassaemia mutation within the conserved 'TATA' homology[33] suggests that this sequence is also important for RNA synthesis *in vivo*, although the effect of this mutation on transcription has not been studied.

As the gene does produce a small amount of correctly spliced β-globin RNA in our assay, we predict that it represents a β$^+$-thalassaemia allele; however, no individual homozygous for the allele has yet been identified.

The IVS*1* 5′ splice site mutations differ in their effects on RNA splicing

The three β-thalassaemia mutations located at the IVS*1* 5′ splice site provide new information regarding the sequence requirements for splicing activity. The requirement for the GT dinucleotide conserved at virtually all 5′ splice sites[34–36] is well established[18,19,25,36–39], and the relatively high degree of

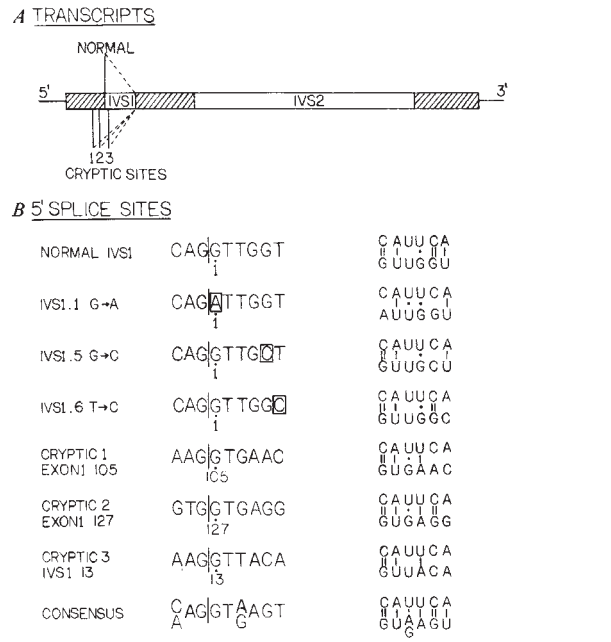

Fig. 6 ***A,*** Processing of the IVS*1* mutant transcripts. The β-globin gene is shown, with exons represented by hatched boxes and intervening sequences by open boxes; flanking sequences are shown as thin lines. The locations of the normal and cryptic IVS*1* 5′ splice sites are shown by vertical lines above and below the gene. Dashed lines indicate that each 5′ splice site is joined to the normal IVS*1* 3′ splice site. ***B,*** Splice site sequences. In the left column, sequences of the mutant IVS*1* and cryptic 5′ splice sites are compared with the normal IVS*1* 5′ splice site and the consensus sequence of Mount[50]. Vertical lines indicate IVS 5′ boundary according to the GT–AG rule[34]; the mutations in the IVS*1* 5′ splice site are boxed. In the right column are shown possible base pairing interactions of the first six nucleotides of each intervening sequence with U1 RNA, according to previous models[43,44].

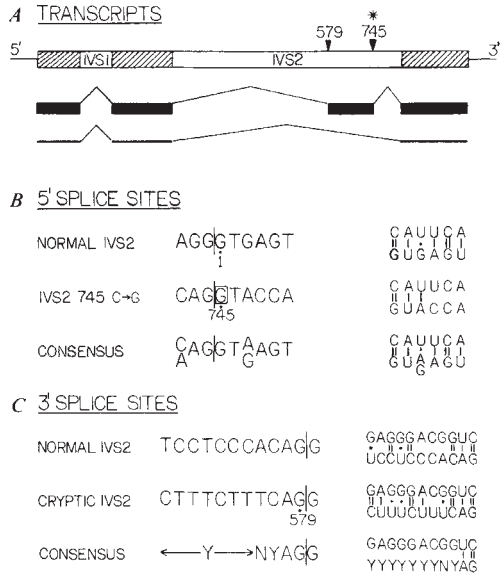

Fig. 7 ***A,*** Processing of the IVS*2* position 745 mutant transcripts. The β-globin gene is shown, with exons represented by hatched boxes and intervening sequences by open boxes; flanking sequences are shown as thin lines. The locations of the mutation (asterisk) and the cryptic 3′ splice site at IVS*2* nucleotide 579 are arrowed. Below are shown the two different RNAs produced by the gene in cultured cells: the exons are shown as solid blocks, with splices indicated by carets. ***B,*** 5′ splice site sequences. In the left column, the novel 5′ splice site is compared with the normal IVS*2* 5′ splice site and the consensus sequence of Mount[50]. Vertical lines indicate the boundary of the IVS according to the GT–AG rule[34]. The mutation at IVS*2* position 745 is boxed. In the right column possible base pairing between the first few nucleotides of the IVS and U1 RNA is shown, according to previous models[43,44]. ***C,*** 3′ splice site sequences. In the left column the cryptic 3′ splice site at IVS*2* position 579 is compared with the sequence of the normal IVS*2* 3′ splice site and the consensus sequence of Mount[50]. Vertical lines indicate the 3′ boundary of the IVS according to the GT–AG rule[34]. In the right column possible base pairing between the last nucleotides of the IVS and U1 RNA is shown, according to previous models[43,44].

sequence conservation at 5′ splice sites suggests that the surrounding sequences are important. However, the introduction of purine transitions at positions 3 and 4 of the rabbit β-globin gene IVS*2* has no effect on RNA splicing[36], while mutation of both nucleotides 5 and 6 of an adenovirus 5 IVS is sufficient to abolish splicing completely[40]. Our results demonstrate that the primary sequence at positions 5 and 6 of IVS*1* determines the extent to which the splicing apparatus chooses the correct 5′ site for IVS*1* splicing from several available potential 5′ splice sites.

The IVS*1* position 1 mutant does not produce any correctly spliced β-globin RNA in our assays. However, it does synthesize a small amount of aberrantly spliced RNA via the utilization of three cryptic IVS*1* 5′ splice sites, located at exon 1 positions 105 and 127, and at IVS*1* position 13 (Fig. 6*A*). Cryptic splice site activity resulting from the inactivation of a 5′ splice site has been observed in several other cases, including a β^0-thalassaemia allele[18,19,25,36]. The two IVS*1* mutations located at positions 5 (G→C) and 6 (T→C) reduce the efficiency with which the splicing machinery discriminates between the correct IVS*1* 5′ splice site and the three cryptic 5′ splice sites nearby that are used by the IVS*1* position 1 mutant. This results in the production of the same three abnormally spliced RNAs and in addition, varying amounts of normally spliced β-globin RNA. Interestingly, the cryptic splice site at nucleotide 127 is also activated, independently of the others, by two other mutations: the β^E mutant G→A change at exon 1 position 129 (ref. 41), and a silent T→A change at position 125 (ref. 42) (see Fig. 6*B*).

Two groups have proposed that the small nuclear RNA U1 acts to align 5′ and 3′ splice sites by base pairing[43,44]. The IVS*1* position 5 and 6 mutations both affect the potential of the IVS*1* 5′ splice site for base pairing with U1 RNA (Fig. 6*B*), unlike the rabbit β-globin gene IVS*2* mutations mentioned above[36], thereby providing a rationale for the differing effects of these mutations on splicing. However, although all the cryptic 5′ splice sites detected in our experiments can base pair with U1 RNA to some extent, not all sequences that can form base-paired structures of the proposed type with U1 RNA are cryptic 5′ splice sites. Similarly, although all the cryptic splice sites we observe are related to the previously defined consensus (Fig. 6*B*), not all consensus-related sequences form cryptic splice sites. It therefore seems likely that other factors besides primary sequence, such as higher order RNA structure or RNA–protein interactions, must influence splice site activity.

Both the IVS*1* position 5 and the IVS*1* position 6 mutations are β^+-thalassaemia alleles, as β^+-thalassaemia patients homozygous for each allele have been identified (S.H.O. and H.H. Kazazian, unpublished data). This is consistent with our observations that both these genes produce some correctly spliced β-globin RNA in our assays. Furthermore, the severity of the disease observed in these cases correlates with the amount of correctly spliced β-globin RNA produced in our assays: the position 5 mutation causes a more severe anaemia than the position 6 mutation (S.H.O. and H.H. Kazazian, unpublished data). It seems likely that the IVS*1* position 1 mutant represents a β^0 allele, as this gene produces no correctly spliced β-globin RNA in our assays.

The IVS2 position 745 mutation creates a new exon

The IVS2 position 745 mutation creates a new GT dinucleotide in a sequence context that produces a new 5′ splice site; this in turn leads to the activation of a cryptic 3′ splice site at IVS2 position 579. The principal RNA produced by this gene therefore contains an extra exon, comprising IVS2 nucleotides 580–744, inserted between exons 2 and 3; in addition, this gene produces a small amount of normally spliced β-globin RNA (Fig. 7A).

Figure 7B compares the sequence of the new 5′ splice site with the 5′ splice site consensus and the normal IVS2 5′ splice site sequences: it matches the consensus at six out of nine positions, compared with eight out of nine positions for the normal IVS2 5′ splice site, and its capacity for base pairing with U1 RNA is less (Fig. 7B). However, it, rather than the normal IVS2 5′ splice site, is paired with the IVS2 3′ splice site in the majority of the RNA produced by this gene. The cryptic 3′ splice site is highly homologous to the 3′ splice site consensus (Fig. 7B) and can base pair with U1 RNA (Fig. 7C), and yet is totally inactive in the normal gene. Furthermore, the pairing of splice sites active in this and other IVS2 mutants of the human[19] and rabbit[25,36] β-globin genes cannot be explained on the basis of a scanning model in which the splicing apparatus binds to one splice site and scans the RNA for a partner[45]. The dramatic effect of this and other single base changes on RNA splicing indicates that such mutations may be instrumental in the evolution of new proteins via the generation of new exons and exon combinations[46].

Our results predict that the IVS2 position 745 mutant represents a β^+-thalassaemia allele, because it does produce a small amount of correctly spliced β-globin RNA in our assays; verification of this awaits the identification of suitable β^+-thalassaemia patients.

Conclusions

The expression of a number of cloned mutant globin genes has previously been examined using cultured cell assays of the type described here[15,16,18,19,22,25-27,36,41,42]. An important question is whether the behaviour of the genes in these assays is an accurate reflection of their behaviour in the erythroid cell. The cell culture assay does provide a qualitatively accurate picture of the abnormal splicing events associated with various β-thalassaemia alleles, as several previous studies have established that the abnormal RNA splicing events observed in cultured cells are identical to those observed in erythroid cells from thalassaemia patients (refs 15–19, 22, 40, and R.T. and A. Oppenheim, unpublished results). In addition, we find that a β-thalassaemia allele with a single base change 87 bp 5′ to the mRNA cap site exhibits abnormally low transcriptional activity in a cultured cell assay, strongly suggesting that the assay also provides an accurate reflection of promoter activity *in vivo*.

Quantitative comparisons between the amounts of abnormal RNA produced in cultured cells and in the erythroid cells of patients may not be meaningful because it is likely that the rates of synthesis and turnover of these RNAs in the two cell types are quite different. For example, previous studies of the expression of several cloned β-thalassaemia alleles showed that in each case the mutant gene produced the same quantity of RNA as a normal β-globin gene when introduced into cultured cells, while the abnormal RNAs accumulated to only low levels *in vivo*[15-19,22]. In contrast, all the splicing mutants examined here consistently produced less RNA than the normal gene in our assays: this may be due either to the inefficient processing of these particular mutant precursor RNAs or to the instability of these abnormal RNAs in cultured cells.

In addition to increasing our understanding of the molecular basis of human genetic disease, study of the thalassaemias has led to several conclusions concerning the mechanism of RNA splicing; these can be summarized as follows. The GT and AG dinucleotides at 5′ and 3′ splice sites are necessary for splicing (IVS1 position 1; refs 18, 19, 25, 36, 37); mutations near GT (or AG?) dinucleotide may result in either a decrease (IVS1 positions 5, 6; ref. 40) or an increase[41,42] in RNA splicing at these sites. Single base changes that create GT or AG dinucleotides may generate splice sites if in an appropriate sequence context (IVS2 position 745; refs 16, 22). Lastly, mutations which interfere with normal splicing may also lead to the utilization of cryptic splice sites (IVS1 positions 1, 5, 6; IVS2 position 745; refs 18, 19, 25, 36), or generate novel exon combinations[19].

We thank Dan DiMaio for advice on the BPV system, Peter Little for the plasmid πSVHPplac, members of the Maniatis group for helpful discussions and Sophia Stern for assistance in preparation of the manuscript. R.T. is the recipient of an ICRF travelling fellowship. This work was supported by grants from the National Foundation-March of Dimes and the NIH to S.H.O. and T.M.

Received 3 December 1982; accepted 15 February 1983.

1. Bunn, H. F., Forget, B. G. & Ranney, H. M. *Human Hemoglobins* (Saunders, Philadelphia, 1977).
2. Weatherall, D. J. & Clegg, J. B. *The Thalassemia Syndromes* (Blackwell Scientific, Oxford, 1981).
3. Maniatis, T., Fritsch, E. F., Lauer, J. & Lawn, R. M. *A. Rev. Genet.* **14**, 145 (1980).
4. Efstratiadis, A. *et al. Cell* **21**, 653–688 (1980).
5. Orkin, S. H. & Nathan, D. G. *Adv. Hum. Genet.* **11**, 233–280 (1981)
6. Proudfoot, N. J., Gil, A. & Maniatis, T. *Cell* **31**, 553–563 (1982).
7. Weatherall, D. J. & Clegg, J. B. *Cell* **29**, 7–9 (1982).
8. Orkin, S. H. *et al. Nature* **296**, 627–631 (1982).
9. Baird, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 4218–4221 (1981).
10. Old, J. M. *et al. Cell* **14**, 289–298 (1978).
11. Benz, E. J. *et al. Cell* **14**, 299–312 (1978).
12. Maquat, L. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 4287–4291 (1980).
13. Kantor, J. A., Turner, P. H. & Neinhuis, A. W. *Cell* **21**, 149–157 (1980).
14. Maquat, L. E., Kinniburgh, A. J., Rachmilewitz, E. A. & Ross, J. *Cell* **27**, 543–553 (1981).
15. Moschonas, N. *et al. Nucleic Acids Res.* **9**, 4391–4401 (1981).
16. Fukumaki, Y. *et al. Cell* **28**, 585–593 (1982).
17. Ley, T. J., Aragon, N. P., Pepe, G. & Neinhuis, A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4775–4779 (1982).
18. Felber, B. K., Orkin, S. H. & Hamer, D. H. *Cell* **29**, 895–902 (1982).
19. Treisman, R. H., Proudfoot, N. J., Shander, M. & Maniatis, T. *Cell* **79**, 903–911 (1982).
20. DiMaio, D., Treisman, R. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4030–4034 (1982).
21. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299–308 (1981).
22. Busslinger, M., Moschonas, N. & Flavell, R. A. *Cell* **27**, 289–298 (1981).
23. Mellon, P., Parker, V., Gluzman, Y. & Maniatis, T. *Cell* **27**, 279–288 (1981).
24. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157–1174 (1978).
25. Dierks, P. *et al. ICN–UCLA Symp. molec. cell. Biol.* **23**, 347–366 (1981).
26. Grosveld, G. C., deBoer, E., Shewmaker, C. K. & Flavell, R. A. *Nature* **295**, 120–125 (1982).
27. Grosveld, G. C., Rosenthal, A. & Flavell, R. A. *Nucleic Acids Res.* **10**, 4951–4984 (1982).
28. McKnight, S. L. & Kingsbury, R. *Science* **217**, 316–324 (1982).
29. Benoist, C. & Chambon, P. *Nature* **290**, 304–310 (1981).
30. Fromm, M. & Berg, P. *J. molec. appl. Genet.* **1**, 457–481 (1982).
31. Lacy, E. & Maniatis, T. *Cell* **21**, 545–553 (1980).
32. Gannon, F. *et al. Nature* **278**, 428–434 (1979).
33. Poncz, M. *et al. J. biol. Chem.* **257**, 5994–5996 (1982).
34. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853–4857 (1978).
35. Dodgson, J. B. & Engel, J. D. *J. biol. Chem.* (submitted).
36. Wieringa, B., Meyer, F., Reiser, J. & Weissmann, C. *Nature* **301**, 38–43 (1983).
37. Montell, C., Fisher, E. F., Caruthers, M. H. & Berk, A. J. *Nature* **295**, 380–384 (1982).
38. Choi, E., Kuehl, M. & Wall, R. *Nature* **286**, 776–779 (1980).
39. Seidman, J. G. & Leder, P. *Nature* **286**, 779–783 (1980).
40. Solnick, D. *Nature* **291**, 508–510 (1981).
41. Orkin, S. H. *et al. Nature* **300**, 768–769 (1982).
42. Humphries, R. K. *et al. Blood* **60**, 54A (1982).
43. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* **283**, 220–224 (1980).
44. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1877–1879 (1980).
45. Sharp, P. A. *Cell* **23**, 643–646 (1981).
46. Gilbert, W. *Nature* **271**, 501 (1978).
47. Seed, B. *Genetic Engineering* Vol. 4 (eds Setlow, J. K. & Hollander, A.) 91–102 (Plenum, New York, 1982).
48. Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **6**, 1175–1193 (1979).
49. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
50. Mount, S. M. *Nucleic Acids Res.* **10**, 459–472 (1982).