The EMBO Journal vol.9 no.1 pp.233 – 240, 1990

Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression

Philip Collis, Michael Antoniou and Frank Grosveld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Communicated by F.Grosveld

The human β-globin dominant control region (DCR) was previously identified as a region from the 5′ end of the human β-globin locus which directs high level, site of integration-independent, copy number-dependent expression on a linked human β-globin gene in transgenic mice and stably transfected mouse erythroleukaemia (MEL) cells. We have now analysed the elements comprising the DCR by systematic deletion mutagenesis in stable MEL transfectants. We have identified two independent elements within the DNase I hypersensitive sites 2 and 3, containing fragments which direct strong transcriptional induciblity of a β-globin gene. Whilst the remaining two hypersensitive sites do not direct significant transcriptional induction, our data suggest that all four sites may be necessary for the fully regulated expression conferred by the DCR. We have also tested a number of β-globin minigene constructs under the control of the DCR to assess if any of the local sequences from the gene may be removed without loss of expression. We find that the 3′ enhancer may be removed without affecting expression, but there is an absolute requirement for the presence of the second intron, not related to the enhancer present in that intron.

Key words: deletion analysis/DNase I hypersensitive sites/ dominant control region/somatic gene therapy

Introduction

In recent years the mechanisms by which genes within the human β-like globin locus are expressed in a developmental stage- and tissue-specific manner have started to be delineated. Using transgenic mice and mouse erythroleukaemia (MEL) cells as models, the activation of the human β-globin gene during erythroid differentiation has been shown to be dependent upon a number of regulatory regions. In particular, erythroid-specific elements have been identified within the promoter of the gene which, together with two downstream enhancer elements, give regulated erythroid-specific expression (Behringer *et al.*, 1987; Kollias *et al.*, 1987; Trudel *et al.*, 1987; Antoniou *et al.*, 1988). Within these *cis*-acting regions a number of ubiquitous and erythroid-specific protein binding sites have been identified (deBoer *et al.*, 1988; Wall *et al.*, 1988), although the precise *in vivo* function of these sites remains to be proven.

The studies on these local *cis*-acting regulatory regions indicated that additional elements were required to obtain full expression of transfected and microinjected β-globin genes, which were prone to strong position effects in transgenic mice and MEL cells (Magram *et al.*, 1985; Townes *et al.*, 1985; Kollias *et al.*, 1986; Antoniou *et al.*, 1988). Recently, we identified a regulatory region at the 5′ end of the β-like globin locus, some 50–60 kb upstream from the β-globin gene. When this region is linked directly to a β-globin gene it specifies site of integration-independent, copy number-dependent, high level expression of the gene in transgenic mice (Grosveld *et al.*, 1987) and MEL cells (Blom van Assendelft *et al.*, 1989). This dominant control region (DCR) contains four tissue-specific DNase I hypersensitive sites spanning a region of some 20 kb (Tuan *et al.*, 1985; Forrester *et al.*, 1987; Grosveld *et al.*, 1987). We have subsequently reduced this fragment in size to a plasmid construct, retaining the full characteristics of the DCR (Talbot *et al.*, 1989). The DCR is also capable of inducing high level expression of α-globin (Hanscombe *et al.*, 1989; Ryan *et al.*, 1989), γ-globin (N.Dillon and F.Grosveld, unpublished data) and of heterologous genes such as the murine *Thy*1 gene and the Herpes thymidine kinase promoter linked to a $G418^R$ gene (*tk-neo*) (Blom van Assendelft *et al.*, 1989; Talbot *et al.*, 1989) in a tissue-specific manner.

The mechanism by which the DCR activates the β-globin locus is not clear and the nature of the functional elements within this controlling region have not been elucidated. Additionally it is important to identify redundant sequences within the DCR for the design of future gene therapy experiments. At the present time, gene therapy protocols designed around retroviral vector delivery systems appear to hold the most promise for treatment of haematopoietic disorders such as thalassaemias (Dzierzak *et al.*, 1988). It is thus vital to minimize the sequences required for predictably high level β-globin expression within a retrovirus construct.

In this paper we describe a functional analysis of the β-globin DCR to assess the contribution of each DNase I hypersensitive site to the full DCR activity. We also present data based on a series of mini gene constructs which were designed to see if any of the local non-coding regions of the β-globin gene could be removed in the presence of the DCR without affecting overall expression. The results from this deletion analysis now allow us to define the minimum elements which are required for the high level expression of a β-globin gene. This should permit the design of more efficient retroviral constructs for use in a somatic gene therapy protocol.

Results

Construction of a human β-globin dominant control region cassette

The original human β-globin minilocus was constructed with 21 kb of DNA from 5′ of the ε-globin gene containing four erythroid cell-specific DNase I hypersensitive sites and 12 kb of DNA 3′ of the human β-globin gene (Grosveld *et al.*, 1987). This reconstructed β-globin minilocus was sub-

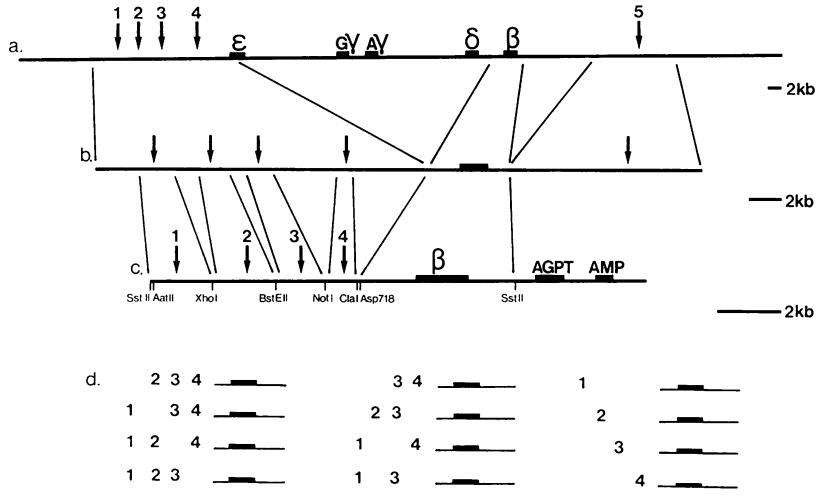

Fig. 1. Schematic illustration of the β-globin DCR constructs. (**a**) The human β-globin locus on chromosome 11. (**b**) The original cosmid β-globin minilocus as described in Grosveld *et al.* (1987) was derived using the 5′ and 3′ elements illustrated. (**c**) The β-globin microlocus cassette containing the four DNase I hypersensitive site fragments described in Talbot *et al.* (1989). See Materials and methods for detailed construction. The unique restriction enzyme sites flanking each hypersensitive site are illustrated. (**d**) Deletion variants of the microlocus. Numbers represent the hypersensitive site(s) present in each construct.

sequently reduced to a 6.5 kb controlling fragment containing the four 5′ DNase I hypersensitive sites and was shown to retain the full function of the minilocus vector (Talbot *et al.*, 1989). In order to dissect the function of this 5′ region in more detail, the four hypersensitive site fragments comprising the 6.5 kb region were recloned in a synthetic polylinker vector such that each hypersensitive site was flanked by unique restriction sites (Figure 1, see Materials and methods for detailed construction). From this microlocus cassette, deletion variants were constructed in which different combinations of hypersensitive sites were removed.

Expression analysis of the DCR deletion mutants identifies two separate transcriptional enhancing elements

Deletion mutants of the locus cassette (Figure 1d) were linearized at a *Pvu*I site in the vector, transfected into MEL cells and three independent stable populations selected with G418 for each construct. Total RNA from uninduced cells (not shown), and cells induced to erythroid differentiation with dimethyl sulphoxide (DMSO), was prepared for each population and analysed by quantitative S1 nuclease protection using mixed probes for 3′ human β-globin and mouse α-globin (Figure 2). DNA was also prepared from each population and analysed with a human β-globin probe to estimate the average transfected gene copy number (data not shown). The expression of human β-globin was then calculated as a ratio of the endogenous mouse globin genes, normalized per gene copy of the exogenous β-globin gene (Table I).

When constructs containing single hypersensitive sites (Figure 1d, lanes 1–4) were analyzed for expression of the human β-globin gene, constructs containing site 2 and site 3 demonstrated significant inducibility, whilst constructs containing site 1 and site 4 showed little transcriptional activation (Figure 2, lanes 1–4; Table I). This suggests that both site 2 and site 3 fragments contain transcriptional enhancing elements. However, in the complementary series of constructs, where site 2 or site 3 had been deleted from the full cassette (lanes 134 and 124, Figure 1d and Table I),

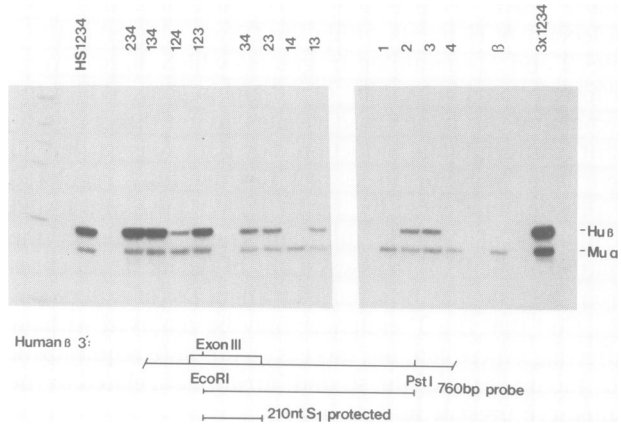

Fig. 2. S1 nuclease protection analysis of the β-globin DCR deletents. The constructs illutrated in Figure 1 were linearized with *Pvu*I or *Sca*I and transfected into C88 MEL cells by electroporation (see Materials and methods). Three populations were isolated for each construct after selection in G418 and analysed before and after induction in 2% DMSO for 4 days. 10 μg of RNA from each population was hybridized to a 760 bp *Eco*RI–*Pst*I 3′ human β-globin probe resulting in a 210 nucleotide protected fragment (illustrated) and a 260 bp *Bam*HI mouse α-globin second exon probe resulting in a 170 nucleotide protected fragment (Muα). The panel shows a representative induced population for each construct. The track labelled β represents RNA from C88 cells transfected with a construct containing the full β-globin gene fragment but lacking any DCR sequences. The 3 × 1234 lane contained 30 μg of induced RNA from the sample used in lane HS1234 to demonstrate that the assay was performed in probe excess. Quantitation was performed by Cerenkov counting each band excised from the gel, centrifuged to the bottom of an Eppendorf tube. A local background count was obtained for each sample by measuring a gel slice from immediately above the band of interest, which was subtracted from the actual count (see Table I).

there appeared to be no significant effect. Loss of sites 1 and 4 also appeared to have no effect on the level of expression of the induced β-globin genes (lanes 234 and 123). Constructs containing combinations of two of the hypersensitive sites gave varying levels of human β-globin induction; as expected, the construct containing sites 1 and 4 gave a very low level of expression. The construct

Table I. Copy number and expression levels of the β-globin DCR constructs

Construct		c.p.m. β^H	c.p.m. α^M	Copy number β^H	$^{\beta H}$Expn. gene copy
HS1234	a*	2087	559	8	1.9
	b	9966	2922	7	2.0
	c	8068	1903	7	2.4
HS234	a	250	123	4	2.0
	b	475	167	4	2.8
	c*	3514	594	7	3.4
HS134	a	1524	741	4	2.0
	b*	2476	629	9	1.8
	c	1383	685	5	1.6
HS124	a	2034	1615	2	2.5
	b*	531	524	3	1.4
	c	3061	1468	3	2.8
HS123	a	2680	927	8	1.4
	b*	1935	595	8	1.6
	c	1315	525	4	2.5
HS34	a*	668	405	4	1.6
	b	782	421	4	1.9
	c	411	156	ND	ND
HS23	a	1595	899	4	1.8
	b*	607	410	9	0.6
	c	1652	1358	7	0.7
HS14	a	30	974	8	0.0
	b	9	262	8	0.0
	c*	11	431	ND	<0.1
HS13	a	1018	1338	3	1.0
	b*	409	276	3	0.8
	c	2192	1523	8	0.7
HS1	a	48	258	3	0.2
	b	29	258	2	0.2
	c*	13	350	ND	<0.2
HS2	a*	533	381	4	1.4
	b	913	699	7	0.8
	c	2914	829	7	2.0
HS3	a	1163	1643	3	1.0
	b*	599	478	7	0.7
	c	2991	2991	7	0.6
HS4	a	48	261	ND	<0.7
	b*	33	425	ND	<0.3

Cerenkov counts for human β-globin (c.p.m. β^H) and mouse α-globin (c.p.m. α^M) were determined by excising and counting each band in Figure 2. These were corrected for background by subtraction of the counts in a gel slice immediately above each band. Average copy numbers of human β-globin (Copy number β^H) were determined for each population by Southern blotting and laser densitometry, using an endogenous mouse *Thy*1 loading control (data not shown). The human β-globin counts were then normalized using the mouse α-globin counts to correct for different levels of induction and loading on to the gel. They were subsequently divided by the copy number and multiplied by 4 to obtain an expression ratio (β^H Expn. gene copy) per α-gene (a total of 4α, two non-allelic genes per haploid genome). Populations marked with an asterisk are shown in Figures 2 and 3.
ND not determined.

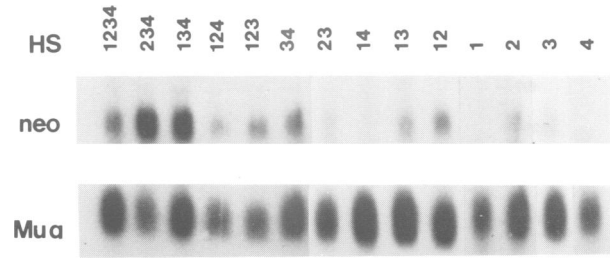

Fig. 3. Northern blot analysis of the level of **tk-neo** RNA induction of the DCR deletents. 10 μg of induced total RNA from one population of each construct illustrated in Figure 1 was run on a denaturing Northern gel and probed with *tk-neo*^{r} (580 bp *Sph*I−*Bgl*II) and mouse α-globin (260 bp *Bam*HI) probes (see Materials and methods). This figure is a composite of two separate experiments.

containing sites 2 and 3 provided significant inducibility, although this combination of sites did not produce a level of induction greater than that specified by each individual site, as might have been predicted (compare lanes 23 with 2 and 3). Other two-site constructs which contained site 3 (lanes 34 and 13) directed significant β-globin expression.

We conclude that site 2 and site 3 are critical for the transcriptional activation of the induced β-globin gene in this assay system, but it would appear that the enhancing activities of these two sites are not complementary when only these sites are present. Whilst sites 1 and 4 alone do not appear to contribute significantly to the induction of β-globin expression, they are required for the fully regulated activity and appear to act synergistically when sites 2 or 3 are present (compare lanes 23 to 234 or 123 and Table I). When only 2 or only 3 is present, both 1 and 4 are required; when 2 and 3 together are present, only 1 or only 4 is sufficient for full activity.

Induction of a linked heterologous promoter in the DCR mutants follows that observed with the β-globin gene

We have previously observed that a linked *tk-neo* selectable marker gene in the β minilocus and β microlocus is also induced in MEL cells (Blom van Assendelft *et al.*, 1989; Talbot *et al.*, 1989) and that this effect is solely dependent upon the DCR sequences (Talbot *et al.*, 1989). To see the effect on induction of the *tk-neo* gene by the deletion mutants described here, the levels of mRNA were analysed by a Northern blot (Figure 3, same populations as Figure 2). This showed a similar pattern of expression to that observed for the human β-globin gene; of the populations containing single hypersensitive sites, only sites 2 and 3 conferred high induction of *tk-neo* RNA, although at levels significantly reduced from the full construct. All constructs containing two of the four sites showed some induction, with the exception of 1,4 which contains sites 1 and 4 only. Constructs containing three of the four sites showed good inducibility, comparable to the full cassette after adjusting for the level of induction and copy number variation. We conclude from this that the individual hypersensitive sites show no difference in their ability to act upon a homologous (β-globin) or heterologous (*tk*) gene.

β-Globin mini gene constructs define an absolute requirement for the second intervening sequence

The data from the DCR deletion constructs in MEL cells demonstrate that in these constructs there is functional

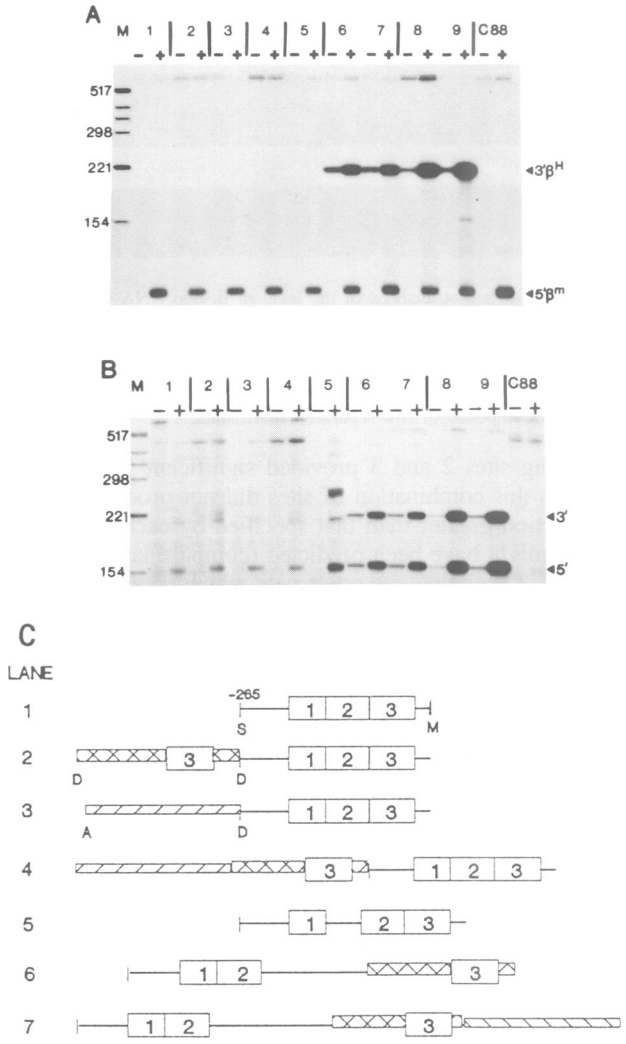

Fig. 4. Expression analysis of human β-globin minigenes in combination with the microlocus DCR. Various deletion mutant and hybrid gene constructs of human β-globin (**panel C**) were cloned downstream of the full DCR (see Figure 1). The resulting plasmids were linearized with *Pvu*I and used to generate stable transformed populations of MEL cells as described in Materials and methods. The pairs of lanes in **panels A** and **B** correspond to the constructs illustrated in **panel C. Lane 1**, intronless or cDNA gene from the *Sna*BI (S) site at −265 to the *Mnl*I (M) site 45 bp past poly(A) addition; **lane 2**, as **lane 1** but with the intragenic enhancer as a 620 bp *Dra*I (D) fragment spanning exon 3 linked at −265; **lane 3**, as **lane 2** but with the 3′ enhancer as a 680 bp *Acc*I (A) to *Dra*I (D) fragment in place of the intragenic enhancer; **lane 4**, cDNA construct in combination with both enhancer regions; **lane 5**, as **lane 1** but incorporating the small intron; **lanes 6** and **7** are constructs possessing the large intron without and with the 3′ enhancer respectively; **lanes 8** and **9** show expression from the complete gene in anti-sense and sense orientation respectively, with respect to the DCR (see Figure 1). The numbered boxes correspond to exons 1, 2 and 3. **Panel A**: S1 nuclease protection analysis with 2 μg RNA from pre-induced (−) and 4 day induced (+) cells, probed for mouse β^{maj}-globin ($5'\beta^{M}$) and human β globin ($3'\beta^{H}$) sequences (see Materials and methods). **Panel B**: **lanes 1−5**, 30 μg RNA and **lanes 6−9**, 2 μg RNA from pre-induced (−) and induced (+) cells analysed with 5′ and 3′ β^{H} probes. The average transfected gene copy number for the populations in **lanes 6** and **7** was the same as that in **lane 9**. RNA from untransfected cells (C88) acted as a negative control. The markers (M) are a *Hin*fI digest of pBR322. The specific activities of the probes in a given panel were equivalent.

redundancy; HS2 and HS3 each contain an enhancer element, which do not appear to function additively in their effect. This appears to be similar to the β-globin gene which in MEL cells contains, in addition to a promoter element, two erythroid-specific enhancers (Antoniou *et al.*, 1988). Thus, in addition to reducing the size of the DCR we sought to identify non-coding sequences within and around the β-globin gene which could be removed without affecting overall expression. A number of mini gene constructs were designed under the control of the full DCR, which lacked combinations of intron sequences and 3′ sequences including the 3′ enhancer element (Figure 4C). Constructs were linearized at a *Pvu*I site within the vector, transfected into MEL cells and stable populations selected with G418. Total RNA was prepared from uninduced and induced populations and analysed by S1 nuclease protection using mixed probes; either a human β-globin 3′ probe and a mouse β major 5′ probe to control for induction (Figure 4A) or human β-globin 3′ and 5′ probes to demonstrate correct initiation and termination of the transcript (Figure 4B). All of the populations shown had a copy number of 3−4 as determined by Southern blots (not shown).

The first construct tested was a β-globin cDNA lacking both introns and terminating at +45 past the poly(A) site, linked to the full DCR (Figure 4C, lane 1). This construct produced very low levels of β-globin mRNA when compared to the full β-globin gene (Figure 4B, lane 1; compare to lanes 8 and 9); a weak signal was observed in the induced population using 30 μg of RNA. No signal was observed using 2 μg of RNA (Figure 4A, lane 1), although this lane demonstrates good induction of the endogenous β major gene. Addition of the intragenic enhancer in the form of the second intron and third exon (Figure 4C, lane 2), the 3′ enhancer (Figure 4C, lane 3) or both enhancers (Figure 4C, lane 4) to the cDNA construct failed to increase the amount of β-globin mRNA produced after induction (Figure 4A and B), suggesting that the very low expression seen with the cDNA construct was not due to the requirement for a local enhancer element.

Clearly then, some function must be provided by the introns, in agreement with some reports in the literature which suggest that the presence of introns increases the levels of transcript of exogenous genes in transgenic mice and cell lines (Gruss *et al.*, 1979; Bender *et al.*, 1988; Brinster *et al.*, 1988; Buchman and Berg, 1988). To test this hypothesis we made constructs containing either the first or the second intron (Figure 4C, lanes 5 and 6). The construct containing only the first intron still failed to express significant amounts of induced mRNA (Figure 4A, lane 5) although when 30 μg of RNA was used, a signal was detectable (Figure 4B, lane 5). However, the 5′ signal is more intense than the 3′ signal, and corresponds in intensity to the signal of a larger protected fragment at ~260 nucleotides. This signal represents a protected RNA species extending past the poly(A) addition site to the end of the region of probe homology at +45 past the poly(A) site (the *Mnl*I site). This imples that the first intron-containing construct was able to be transcribed, but was unable to undergo correct polyadenylation, and results in less RNA compared to the full gene. When nuclear and cytoplasmic RNA were analysed almost all of the incorrectly polyadenylated RNA (>90%) was localized in the nucleus, while a small amount of correctly polyadenylated RNA was present in the cytoplasm (data not shown).

In contrast to the intron 1 containing construct, the IVS2 construct produced high levels of RNA, although at lower levels than the full gene (Figure 4A and B; lane 6, compared

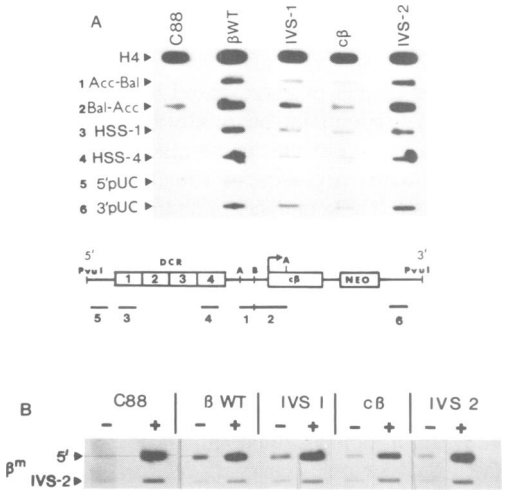

Fig. 5. Nuclear run-off analysis of DCR-β-globin mini-gene constructs. Nuclei were isolated from MEL cells transfected with the DCR β-globin mini gene constructs described in Figure 4, and used in a nuclear run-off transcription assay (Materials and methods). The nascent ^{32}P-labelled RNA isolated from the nuclei was hybridized to identical strips of nitrocellulose possessing denatured DNA probes. A number of probes (**panel A**, lower section) were prepared from various parts of the transfected construct in order to assess the presence of transcripts originating from different regions. Probe 1, 300 bp *Acc*I (A)−*Bal*I (B) fragment from the upstream promoter region of human β-globin; probe 2, 225 bp *Bal*I−*Acc*I fragment spanning the promoter and first exon of the gene; probe 3, 190 bp *Pst*I−*Acc*I fragment from DCR HSS1; probe 4, 300 bp *Sst*I−*Hin*dIII fragment from DCR HSS4; probe 5, 207 bp *Xmn*I−*Ssp*I fragment from the 5′ region of the pUC vector; probe 6, 180 bp *Pvu*II−*Afl*III fragment from the 3′ region of the pUC vector. The filter strips contained 1 μg of each probe. Control probes included were: H4, mouse histone H4 cloned in pBR322 and linearized with *Pvu*I (6 μg/slot); β^{M}-5′, 700 bp *Hin*dIII−*Nco*I fragment of 5′ half of mouse β^{maj}-globin gene; β^{M}-IVS-2, 600 bp *Bam*HI−*Pst*I fragment spanning the large intervening sequence (IVS2) of mouse β^{maj}-globin. The histone H4 probe acts as a control for loading, while the β^{maj}-globin probes assess the level of erythroid differentiation of the cells. **Panel A**: analysis of nuclei from 3 day induced MEL cell populations transfected with the following: C88, untransfected control; βWT, full wild-type β-globin gene (see Figure 1); IVS1-, β-globin gene containing only intron 1; cβ, intronless gene; IVS2, β-globin gene containing only intron 2. **Panel B**: nuclei from pre-induced (−) and 3 day induced (+) cells transfected with the same constructs as in **panel A**. (Note, the 5′β^{maj} probe shows some cross-hybridization with human sequences as indicated by the significant signal in the pre-induced lanes. The IVS2 probe is β^{maj}-specific.)

to lanes 8 and 9) and was correctly polyadenylated. Addition of the 3′ enhancer fragment to this construct (Figure 4C, lane 7) did not further stimulate the level of the induced mRNA (Figure 4A and B, lane 7). Lanes 8 and 9 represent constructs containing the full β-globin gene (4.9 kb *Bgl*II fragment) linked to the DCR in either the opposite orientation (lane 8) or the same orientation (lane 9), demonstrating that the full gene is not dependent upon its orientation with respect to the DCR (Talbot *et al.*, 1989).

These data suggest that the β-globin gene does not require the 3′ enhancer or the first intron sequence for high level expression. By contrast, the second intron appears to be essential for full expression, although it would seem unlikely that this is due to the requirement for the intragenic enhancer, since when this fragment is placed upstream of the coding sequence it does not restore expression (Figure 4A and B, lane 2).

To determine whether the differences observed in the levels of expression between the β minigene constructs were due to a transcriptional or post-transcriptional effect, nuclear run-offs were performed on isolated nuclei containing the β-globin cDNA, with or without IVS1 or IVS2 and wild-type β-globin gene constructs (Figure 5). Preliminary experiments suggested that a large amount of transcription was occurring within the DCR (data not shown) and so in addition to a probe spanning the first exon (*Bal*−*Acc*), a number of additional unique sequence probes were derived to determine any transcriptional readthrough 5′ of the β-globin cap site (*Acc*−*Bal*) and the 5′ and 3′ borders of the DCR (HSS1 and HSS4).

From the β-promoter fragment (*Acc*−*Bal*), it was apparent that there was some transcriptional readthrough, presumably from the DCR (Figure 5A). In addition, the *Bal*−*Acc* first exon fragment used to detect nascent transcripts from the β-globin promoter was subject to some cross-hybridization with the endogenous mouse gene (C88 untransfected lane, Figure 5). After subtracting both of these contributions from the signals detected by the *Bal*−*Acc* fragment, it was clear that there were differences in the level of transcription between the β minigene constructs; the IVS1 and cDNA construct signals became negligible, in contrast to the IVS2 and wild-type gene signals. The levels of RNA detected by this assay for these four constructs appears to mimic exactly the result obtained by S1 nuclease analysis (Figure 4), indicating that the different constructs have different transcription rates, resulting in different amounts of primary transcripts.

Discussion

There are two enhancer elements within the β-globin dominant control region

Previous work from this laboratory has demonstrated that small restriction fragments containing the four 5′ hypersensitive sites upstream of the ε-globin gene are sufficient to direct the full biological expression of the human β-globin gene in transgenic mice and MEL cells (Talbot *et al.*, 1989). It was not clear, however, if all of these sites are necessary for the full biological function or, indeed, what functional elements are contained within this upstream region. The data presented in this paper demonstrates that the β-globin DCR contains a number of functional elements, which may all be necessary for the full *in vivo* function of the DCR.

We have identified two elements within the DCR which promote strong transcriptional inducibility in MEL cells. These elements, within sites 2 and 3, each account for ~50% of the activity of the full β-globin cassette when assayed individually. The site 3 fragment has been demonstrated to have enhancer activity in a transient CAT assay system (Tuan *et al.*, 1989), no activity was detected for any of the other sites, including the site 2 fragment using this assay. Clearly, in stable transfections (Figure 2) both fragments show strong transcriptional enhancing activity, illustrating the difference between classical enhancer elements identified in transient enhancer-trap experiments and at least three of the elements within the DCR.

Although sites 2 and 3 contribute significant inducibility on a β-globin gene when tested individually, in this assay they do not appear to cooperate to give full inducibility in the absence of sites 1 and 4. We thus conclude that sites 1 and/or 4, which alone or in combination have a low activity, are nevertheless required for the full regulated

induction observed with the complete locus cassette. Thus, it may be that the site 1 and 4 sequences are involved in opening up the chromatin around the domain, so that the local regulatory promoter and enhancer sequences are then accessible to *trans*-acting factors. Full activation of the gene is only possible, however, in the presence of the powerful enhancing activities specified within sites 2 and 3.

The observation that the removal of site 2 or 3 does not lead to a significant reduction in the inducibility of the linked β-globin gene, raises the possibility that the DCR may contain regions of functional redundancy. This has been previously identified with a number of tissue-specific enhancers, e.g. the *Drosophila SGS4* gene (Jongens *et al.*, 1988) or immunoglobulin genes (e.g. Lenardo *et al.*, 1987), where several elements were identified within the enhancer which are involved in tissue specificity and it was found that any pairwise combination of elements was sufficient to direct expression indistinguishable from the full enhancer, although no single element could function in isolation. Clearly in the β-globin DCR, two separate transcription-inducing elements are present within sites 2 and 3 and it is possible that these may not both be necessary for full β-globin induction in the presence of sites 1 and 4.

Based on the data obtained from this cell culture system, we predict that site 2 and site 3 constructs would produce high levels of β-globin mRNA in transgenic mice, although these constructs may not express in a regulated, copy number-dependent manner. We would predict that combination with site 1 or site 4 should be sufficient to regulate position-independent, copy number-dependent expression, but that all four sites are necessary to produce the full level of human β expression observed with the microlocus.

Recently Ryan *et al.* (1989) demonstrated high levels of human β-globin expression in transgenic mice using sites 3 and 4 (HSI and II according to their nomenclature). However, although their data also demonstrate that sites 3 and 4 produce high levels of β-globin mRNA, they do not allow any conclusions about full expression and copy number dependence, since their work was based on 16 day embryos. Our observations using the DCR show that fully transgenic mice are severely anaemic by day 13.5 after fertilization (Talbot *et al.*, 1989; P.Fraser, unpublished data) and that such mice do not survive to day 16 (Hanscombe *et al.*, 1989). Survival at this gestation period therefore selects for mice which are mosaic for the transgene, and hence would not express the gene in a truly quantitative manner.

The β-globin second intron is vital for expression

Deletion analysis of the β-globin gene under the control of the full DCR has identified two important features. Firstly, our data demonstrate that removal of the β-globin 3′ enhancer does not lower the level of expression of the gene in MEL cells. This prompts the question of whether the 3′ enhancer is functionally important *in vivo*. Our observation may be an artefact due to the artificial proximity of the powerful enhancers residing within HS2 and 3 which can functionally replace the role of the 3′ enhancer. Thus, *in vivo* when these elements are some 60 kb away from the β-globin gene (at least as defined by primary structure) the 3′ enhancer truly is required for β-globin expression. However, to our knowledge there have been no reports in the literature of a β-thalassemia caused by mutation within the 3′ enhancer, making this the only local *cis*-acting element which has not been implicated as the causative mutation site for haemoglobinopathy. It is thus intriguing to speculate that this enhancer element may be redundant *in vivo*.

Secondly our data show that, while removal of the first intron sequence results in a decreased level of expression of the gene, the second intron is absolutely required for expression. The reason for this is not immediately apparent. By analogy with the argument outlined above for the 3′ enhancer, that the DCR enhancers should compensate for local enhancer loss, and by the observation that when placed upstream of the coding sequence the intragenic enhancer does not restore expression, it seems unlikely that the intragenic enhancer located at the 3′ end of the intron is required.

An alternative hypothesis, previously suggested by others is the *de facto* requirement for an intron (Gruss *et al.*, 1979; Bender *et al.*, 1988; Brinster *et al.*, 1988; Buchman and Berg, 1988). In particular, Buchman and Berg showed that recombinant SV40 viruses carrying rabbit β-globin cDNA failed to express the β-globin gene in cell lines in the absence of introns, but the addition of either IVS1 or IVS2 caused a substantial increase in mRNA production.

Our data, using a number of different intron-containing constructs, suggest that the two introns within the human β-globin gene may have different functions; the IVS1 construct, lacking IVS2, was able to undergo correctly initated transcription and splicing, as determined by 5′ S1 nuclease analysis (Figure 4B, lane 5), but produced an aberrant 3′ S1 signal which could be accounted for by a failure to polyadenylate. This implies that the large intron contains sequences which are important for polyadenylation of the transcript, or that the processing *per se* of the large intron is coupled to correct polyadenylation. By contrast, the construct containing only the large intron produced a normal β-globin transcription unit at a high steady-state level (Figure 4A and B, lane 6), although this level was 3-fold down on that observed for the full gene (compare Figure 4A and B, lanes 6 and 7 with lane 9). Thus, whilst the small intron does not appear to play an important role in the processing of the primary transcript, its removal does appear to reduce the steady-state RNA level, implying that there may be a splice requirement for correct transcription and processing, which is not completely fulfilled by the presence of the second intron splice.

The nuclear 'run-off' experiment is designed to look at the presence of nascent nuclear RNA. We performed a nuclear run-off analysis on the four constructs containing different intron combinations and observed that the level of RNA detected for each construct, corrected to reflect only correctly initiated RNA, mirrored the result obtained by S1 nuclease analysis. Interestingly, all transcription throughout the plasmid, including readthrough and upstream starts, was increased by the presence of IVS2, although transcription of the β-promoter increased most; the data therefore suggest that different transcription rates are responsible for the different mRNA levels. However, the run-off was performed for a 20 min time course, and it is not excluded that mRNAs lacking IVS2 may be unstable within this time and were thus undetectable. Reducing the time course would not necessarily yield a more definitive answer; even a 5 min assay could contain considerable undetectable degradation products from an unstable primary transcript which would give an incorrect result, whilst reducing the time course to 1−2 min would not allow sufficient incorporation to permit a quantitative analysis.

Based on the result obtained with the intron 1 construct,

which suggested that the second intron is specifically required for correct polyadenylation and production of cytoplasmic mRNA, it is also possible that the reduction in RNA levels observed with the intron minus constructs is due to instability and a failure to export the RNA from the nucleus. Clearly the second intron is important for polyadenylation to occur and a more detailed analysis of this intron is required to delinate the nature of this function.

Definition of a minimal β-globin domain for somatic gene therapy

One of the aims of this work was to delinate the minimal requirements for efficient expression of a β-globin gene in a retrovirus. Our data show that it is possible to obtain high level expression of a β-globin gene without all four of the DCR DNase I hypersensitive sites, although using smaller DCR constructs may require the screening of individual clones to identify the highest expressors. We have also demonstrated that the size of the β-globin gene may be reduced by removing the first intron and 3′ sequences without drastically reducing the level of expression. Using this data, we are now in a position to design smaller β-globin domains which should in turn lead to the development of more efficient β-globin retroviral constructs for use in gene addition experiments and ultimate gene therapy.

Materials and methods

Construction of the human β-globin microlocus cassette

To construct the full microlocus cassette three new polylinker vectors, LL1–LL3, were designed. These were derived from pUC18, from which nucleotides 2617–628 (*Aat*II–*Pvu*II) were removed and replaced with synthetic linker sequences. LL1 contained a polylinker sequence: 5′ *Sst*II-*Nae*I-*Aat*II-*Xba*I-*Xho*I-*Hin*dIII-*Bst*EII 3′; the LL2 polylinker consisted of: 5′ *Bst*EII-*Hinc*II-*Not*I-*Sma*I-*Cla*I 3′; the LL3 polylinker contained 5′ *Cla*I–*Bgl*II-*Kpn*I-*Bam*HI-*Mlu*I-*Sst*II-*Nar*I-*Pvu*II 3′. The hypersensitive site (HSS) fragments described in Talbot *et al.* (1989) were inserted into LL1 and LL2: the 2.1 kb *Bam*HI–*Xba*I HSS1 fragment was taken from a subclone donating an extra 3′ *Xba*I site and inserted into the LL1 *Xba*I site; the 1.9 kb *Hin*dIII HHS2 fragment was inserted into the LL1 *Hin*dIII site; the 1.5 kb *Asp*718–*Bgl*II HSS3 framgent was blunt ligated into the LL2 *Hin*dII site and the 1.0 kb partial *Sst*I–*Hin*dIII HHS4 fragment was blunted into the LL2 *Sma*I site. The 4.8 kb *Bgl*II fragment of the human β-globin gene was cloned into the *Bam*HI site of LL3 and a 2.0 kb partial *Nar*I *tk-neo*^{r} gene was inserted into the LL3 *Nar*I site.

The final microlocus cassette was put together as a three fragment ligation: a 4.6 kb LL1 *Pvu*I–*Bst*EII fragment containing part of the pUC vector sequence and HSS1 and 2; a 2.5 kb LL2 *Bst*EII–*Cla*I fragment containing HSS3 and 4; and an 8.3 kb LL3 *Cla*I–*Pvu*I fragment containing human β-globin, *tk-neo*^{r} and the remaining pUC vector sequences.

Deletion mutants of the microlocus were prepared by restriction and blunt end ligation between the appropriate unique sites flanking each HSS.

Construction of the human β-globin mini gene constructs

All β-globin minigene constructs were cloned at the *Cla*I–*Asp*718 sites of the full microlocus cassette lacking the 4.9 kb *Bgl*II β-globin fragment; the β-globin cDNA was an *Sna*BI–*Mnl*I fragment starting at −265 bp from the cap site to +45 bp past the poly(A) site, linkered with *Cla*I and *Asp*718. The 3′ and intragenic enhancers were added as *Cla*I linkered fragments at the *Sna*BI site; the 3′ enhancer fragment used was a 680 bp *Acc*I–*Dra*I fragment. The intragenic enhancer was isolated as a 620 bp *Dra*I fragment. The construct containing the first intron was made by exchanging the *Nco*I–*Bam*HI cDNA fragment spanning the first and second intron junction with the corresponding genomic fragment. Similarly, the second intron construct was prepared by exchanging the cDNA *Bam*HI–*Mnl*I fragment spanning the second and third intron junction to +45 past the poly(A) site with the corresponding genomic fragment. The 680 bp *Acc*I–*Dra*I 3′ enhancer fragment was linkered and inserted at the *Asp*718 site of the second intron construct.

Tissue culture and cell transfections

The MEL cell line C88 was maintained in standard αMEM plus 10% fetal calf serum tissue culture medium. Microlocus constructs (50–100 μg) were linearized at the *Pvu*I or *Sca*I sites and transfected into MEL cells by electroporation, as follows: cells in rapid, log phase growth ($3-5 \times 10^7$ cells per transfection) were washed and resuspended in 1 ml HEPES-buffered saline (Smithies *et al.*, 1985) containing the linearized plasmid. After incubation at room temperature for 10 min the cell–plasmid DNA mixture was electroporated with 1–2 pulses from a Hoefer 'Pro-Genitor' set to deliver 250 V for 10 ms. The cells were allowed to stand at room temperature for a further 5 min before being divided to generate three independent transfected populations as previously described (Antoniou *et al.*, 1988). MEL cells were induced to undergo erythroid differentiation by incubation in the presence of 2% (v/v) DMSO for 4 days.

RNA analysis

Total RNA extraction of transfected cell populations was performed using the lithium chloride–urea method of Auffray and Rougeon (1980), including a 2 min sonication step.

S1 nuclease protection analysis using human β-globin, mouse α-globin and mouse β^{maj} globin DNA probes was carried out as described previously (Kollias *et al.*, 1986; Antoniou *et al.*, 1988).

Northern blotting was performed as described by Krumlauf *et al.* (1987).

DNA analysis

Southern blotting was performed as described by Southern (1975).

Nuclear run-off analysis

Nuclei from pre-induced and 3 day induced MEL cell populations were prepared and subjected to nuclear 'run-off' transcription analysis as previously described (Antoniou *et al.*, 1988), except that the hybridization conditions of Wright and Bishop (1989) were employed.

Acknowledgements

We are grateful to Cora O'Carroll for the preparation of the manuscript. This work was supported by the Medical Research Council (UK).

References

Antoniou,M., deBoer,E., Habets,G. and Grosveld,F. (1988) *EMBO J.*, **7**, 377–384.

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.

Behringer,R.R., Hammer,R.E., Brinster,R.L., Palmiter,R.D. and Townes,T.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7056–7060.

Bender,M.A., Miller,A.D. and Gelinas,R.E. (1988) *Mol. Cell. Biol.*, **8**, 1725–1735.

Blom van Assendelft,M., Hanscombe,O., Grosveld,F. and Greaves,D.R. (1989) *Cell*, **56**, 969–977.

Brinster,R.L., Allen,J.M., Behringer,R.R., Gelinas,R.E. and Palmiter,R.D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 836–840.

Buchman,A.R. and Berg,P. (1988) *Mol. Cell. Biol.*, **8**, 4395–4405.

deBoer,E., Antoniou,M., Mignotte,V., Wall,L. and Grosveld,F. (1988) *EMBO J.*, **7**, 4203–4212.

Dzierzak,E., Papayannopoulou,T. and Mulligan,R. (1988) *Nature*, **331**, 35–41.

Forrester,W.C., Takegawa,S., Papayannopoulou,T., Stamatoyannopoulos,G. and Groudine,M. (1987) *Nucleic Acids Res.*, **15**, 10159–10177.

Grosveld,F., Blom van Assendelft,M., Greaves,D.R. and Kollias,G. (1987) *Cell*, **51**, 975–985.

Gruss,P., Lai,C.-J., Dhar,R. and Khoury,G. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4317–4321.

Hanscombe,O., Vidal,M., Kaeda,J., Luzzatto,L., Greaves,D.R. and Grosveld,F. (1989) *Genes Dev.*, **3**, 1572–1581.

Jongens,T.A., Fowler,T., Shermoen,A. and Beckendorf,S.K. (1988) *EMBO J.*, **7**, 2559–2567.

Kollias,G., Wrighton,N., Hurst,J. and Grosveld,F. (1986) *Cell*, **46**, 89–94.

Kollias,G., Hurst,J., deBoer,E. and Grosveld,F. (1987) *Nucleic Acids Res.*, **15**, 5739–5747.

Krumlauf,R., Holland,P., McVey,J. and Hogan,B. (1987) *Development*, **99**, 603–617.

Lenardo,M., Pierce,J.W. and Baltimore,D. (1987) *Science*, **236**, 1573–1577.

Magram,J., Chada,K. and Costantini,F. (1985) *Nature*, **315**, 338–340.

Ryan,T.M., Behringer,R.R., Martin,N.C., Townes,T.M., Palmiter,R.D. and Brinster,R.L. (1989) *Genes Dev.*, **3**, 314–323.

Smithies,O., Gregg,R.G., Boggs,S.S., Karalewski,M.A. and Kucherlapati,R.S. (1985) *Nature*, **317**, 230–234.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.

Talbot,D., Collis,P., Antoniou,M., Vidal,M., Grosveld,F. and Greaves,D.R. (1989) *Nature*, **338**, 352–353.
Townes,T.M., Lingrel,J.B., Chen,H.-Y., Brinster,R.L. and Palmiter,R.D. (1985) *EMBO J.*, **4**, 1715–1723.
Trudel,M., Magram,J., Bruckner,L. and Costantini,F. (1987) *Mol. Cell. Biol.*, **7**, 4024–4029.
Tuan,D., Solomon,W., Qiliang,L.S. and Irving,M.L. (1985) *Proc. Natl. Acad. Sci. USA*, **32**, 6384–6388.
Tuan,D.Y.H., Solomon,W.B., London,I.M. and Lee,D.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2554–2558.
Wall,L., deBoer,E. and Grosveld,F. (1988) *Genes Dev.*, **2**, 1089–1099.
Wright,S. and Bishop,J.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 505–509.

Received on August 4, 1989; revised on October 16, 1989