

# Structure and function of the murine $\beta$ -globin locus control region 5' HS-3

Bruce A.Hug, Anne M.Moon and Timothy J.Ley\*

Division of Hematology/Oncology, Departments of Medicine and Genetics, Jewish Hospital at Washington University Medical Center, 216 So. Kingshighway Boulevard, St Louis, MO 63110, USA

Received July 1, 1992, Accepted August 17, 1992

EMBL accession no. Z13985

## ABSTRACT

We previously identified the murine homologue of the human  $\beta$ -globin Locus Control Region (LCR) 5' HS-2. The  $\lambda$  clone containing murine 5' HS-2 extends approximately 12 kb upstream from this site; here, we report the sequence of this entire upstream region. The murine homologue of 5' HS-3 is located approximately 16.0 kb upstream from the mouse  $\epsilon\gamma$ -globin gene, but no region homologous to human 5' HS-4 was present in our clone. Using a reporter system consisting of a human  $\gamma$ -globin promoter driving the neomycin phosphotransferase gene ( $\gamma$ -neo), we tested murine LCR fragments extending from -21 to -9 kb (with respect to the  $\epsilon\gamma$ -globin gene cap site) for activity in classical enhancer and integration site assays in K562 and MEL cells. 5' HS-2 behaved as a powerful enhancer and increased the number of productive integration events (as measured by a colony assay) in both K562 and MEL cells. 5' HS-3 had no activity in K562 cells or in transiently transfected MEL cells, but was nearly as active as 5' HS-2 in the MEL cell colony assay. Two additional tests confirmed the identification of murine 5' HS-3: first, a DNA fragment containing 5' HS-3 confers copy number-dependent, integration-site independent inducibility on a linked  $\beta$ -globin gene in the MEL cell environment. Secondly, a strong DNaseI hypersensitive site maps to the location of the 5' HS-3 functional core in chromatin derived from MEL cells. Collectively, these data suggest that we have identified the murine homologue of human 5' HS-3, and that this site is functional when integrated into the chromatin of MEL cells but not K562 cells. 5' HS-3 may therefore contain information that contributes to the development-specific expression of the  $\beta$ -like globin genes.

## INTRODUCTION

The human  $\beta$ -globin Locus Control Region is comprised of a series of erythroid-specific, developmentally stable DNase I hypersensitive sites located 6–18 kb upstream from the human  $\beta$ -like globin gene cluster (1–4). This region is required for

function of the linked  $\beta$ -globin gene cluster, since a 30 kb deletion that removes most of the LCR inactivates all of the  $\beta$ -like globin genes linked in *cis* (5), presumably by causing the chromatin domain of the linked  $\beta$ -like globin genes to 'close' in erythroid cells (6).

The  $\beta$ -globin LCR has many powerful functions in transgenic mice and in tissue culture models of globin gene expression. The LCR confers integration site-independent, copy number-dependent expression on linked globin genes in transgenic mice (7–15) and in erythroid tissue culture cells (8, 11, 12, 14, 16–18). The LCR confers inducibility on linked globin and TK promoters in mouse erythroleukemia cells (8, 11, 14, 16–18), and likewise confers hemin inducibility on many linked promoters in K562 erythroleukemia cells (19–24). Finally, the LCR seems to increase the probability that chromatin integration events will permit active expression of linked promoters, perhaps by isolating them from negative regulatory influences that lie nearby (7–15, 22, 24).

Many LCR properties are contained within the individual DNase I hypersensitive sites (HS) that collectively comprise the LCR. For example, human 5' HS-2 (located 10.9 kb upstream from the  $\epsilon$ -globin gene cap site) activates linked globin genes in transgenic mice (9, 10, 12–15), behaves as a transient enhancer in K562 cells (19–26), confers inducibility on linked globin genes in K562 and MEL cells (12, 17, 18–24), and increases productive integration events in K562 cell colony assays (22, 24). Human 5' HS-3 alone is also active in transgenic mice (10, 11) and in assays where it is stably integrated into the chromatin of MEL cells (11, 16, 18). However, this site does not act as a transient enhancer in K562 cells (24, 25), and it is inactive in K562 cell chromatin (20, 24). Fundamental differences must therefore exist between the activities and functions of 5' HS-2 and 5' HS-3.

To further define the critical DNA sequences required for LCR functions, we previously cloned mouse genomic sequences homologous to human 5' HS-2 (22), and we discovered that the homologous sequences were located just upstream from the mouse  $\beta$ -like globin gene cluster in a position very similar to that of human 5' HS-2. The DNA sequence of a functional HS-2 core was precisely conserved between the mouse and human species. We also reported that murine and human 5' HS-2 have

\* To whom correspondence should be addressed

identical functional properties in all K562 cell assays (22). At the same time, Li *et al.* (27, 28) identified the goat  $\beta$ -globin LCR sequences: 5' HS-1, HS-2, and HS-3 were spatially organized in the same way as the human LCR, and were each found to contain conserved sequences corresponding to the functional cores of each site.

To further explore the critical DNA sequences required for the functions of 5' HS-3, we decided to complete our sequence analysis of a  $\lambda$  clone that extended from 7 to 21 kb upstream from the murine  $\epsilon\gamma$ -globin gene. We found that this fragment contains the murine homologue of 5' HS-3; no region corresponding to 5' HS-4 could be identified. In a series of assays, we have shown that murine 5' HS-3 appears to have all of the same functional properties as human 5' HS-3, and that a strong DNaseI hypersensitive site maps to the functional core of murine 5' HS-3 in MEL cells. Differences in the activity profiles of 5' HS-3 in an embryonic-fetal erythroid cell line (K562) and in an adult erythroid cell line (MEL) suggest that this site may be involved in the developmental regulation of the  $\beta$ -like globin genes.

## MATERIALS AND METHODS

### DNA sequence analysis

The  $\lambda$ mLAR clone (22) was obtained from a Stratagene  $\lambda$  Fix genomic library prepared from 3T3 cell DNA (these cells are derived from a Swiss-Webster strain of *Mus musculus*); the entire clone was sequenced from positions -21754 to -9155 with respect to the  $\epsilon\gamma$ -globin gene cap site (29). Both strands were sequenced with the dideoxy chain termination method using Sequenase 2 (USB). Sequence analysis was performed with Microgenie<sup>R</sup> software (Beckman).

### Tissue culture

K562 cells and MEL cells (a gift from Elliot Epner, William Forrester, and Mark Groudine) were maintained in modified IMEM containing 5% fetal calf serum (GIBCO), 5% Control Process Serum Replacement-4 (Sigma Chemical Company, St. Louis, MO), and 10  $\mu$ g/ml gentamicin at 37°C in 5% CO<sub>2</sub> with 100% humidity.

### Expression vectors and plasmid constructions

The  $\gamma$ -neo expression vector contains the  $\Delta\gamma$ -globin promoter (positions -299 to +36) driving the gene for neomycin phosphotransferase in a pUC9-based plasmid (30). The  $\gamma$ -neo transcription unit was isolated from this vector as a 2.7 kb BamHI fragment.

The murine LCR 5'HS-2 EcoRI fragment (mLCR 1.1) extends from positions -10442 to -9287 with respect to the murine  $\epsilon\gamma$  gene transcription start site, and contains the region of homology with human LCR 5'HS-2 (22; in our previous publication, the coordinates of this fragment were incorrectly reported as -10,509 to -9353). This fragment was subcloned into the EcoRI site of pUC9. The  $\gamma$ -neo/mLCR 1.1 ( $\gamma$ -neo 9) plasmid was generated by cleaving the pUC9/mLCR 1.1 plasmid with BamHI, and inserting the  $\gamma$ -neo transcription unit.

Fragments 1-8 (see Fig. 5) contain murine genomic sequences located upstream from mLCR 1.1 in the original murine  $\lambda$  mLCR clone previously described (22). Fragments were isolated as described below and were subcloned into appropriately digested pUC9 vectors. The resulting plasmids were linearized with BamHI and the  $\gamma$ -neo transcription unit was inserted.

**Fragment 1.** This 2.3 kb BamHI fragment extending from positions -21401 to -19121 was obtained directly from  $\lambda$  mLCR and subcloned into pUC9.

**Fragment 2.**  $\lambda$  mLCR was cleaved with BamHI and a 508 bp fragment (positions -13149 to -12641) was isolated. This entire fragment is part of an L1md-2 repetitive element (29) that occurs twice in the upstream region. The DNA sequence from -19121 to -18619 (Fragment 2\*) is identical to this sequence except for 13 single nucleotide changes.

**Fragments 3-6.** A 5.5 kb fragment was isolated from  $\lambda$  mLCR and subcloned into pUC9 cleaved with BamHI to generate the plasmid p778. p778 was secondarily cleaved with HindIII which released fragments 3 (positions -18618 to -18085), 4 (positions -18084 to -17600), and 5 (positions -17599 to -15561). Fragment 6 (positions -15560 to -13149) remained attached to pUC9; this fragment was gel purified and religated to itself.

**Fragment 7.** Plasmid 650 contains a ~5.6 kb BamHI/SalI murine genomic fragment (3' end at ~ -7.5 kb with respect to  $\epsilon\gamma$ ) isolated from the original  $\lambda$  mLCR clone. Plasmid 650 was cleaved with BamHI and PvuII and a 1.1 kb fragment extending from positions -12640 to -11565 was isolated.

**Fragment 8.** Plasmid 650 was cleaved with PvuII and EcoRI and the fragment extending from positions -11564 to -10443 was isolated.

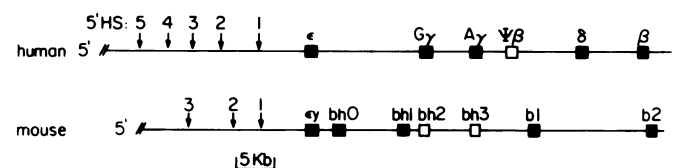
The human  $\beta$ -globin gene reporter construct contains a 4.4 kb PstI fragment that includes the human  $\beta$ -globin gene (with 2.2 kb of 5' flanking sequence and 0.5 kb of 3' flanking sequence) inserted into the pUC9 PstI site. The murine HS-3/human  $\beta$ -globin construct also contains a 2.0 kb HindIII fragment (extending from positions -17600 to -15561 upstream from the  $\epsilon\gamma$ -globin cap site) oriented upstream and tandem to the human  $\beta$ -globin gene. Constructs were linearized at a SalI site downstream from the  $\beta$ -globin gene prior to transfection.

### Transient transfection of K562 and MEL cells

K562 cells were transiently transfected exactly as described, with electroporation parameters of 800  $\mu$ F and 250 volts (22, 30). MEL cells were transiently transfected using the same protocol, but at 165 volts and 600  $\mu$ F.

### K562 and MEL cell colony assays

The K562 cell colony assay was performed by plating transfected K562 cells in complete media containing 1 mg/ml G418 and 0.3% BactoAgar (Difco 0140-01) exactly as previously described (22).



**Figure 1.** Map of the human and mouse  $\beta$ -globin clusters and Locus Control Regions. The functional human or mouse globin genes are shown as black boxes. Open boxes denote pseudogenes. The locations of the DNase I hypersensitive sites that comprise the human and mouse  $\beta$ -globin Locus Control Regions are shown (5' HS-1 through 5 and 5' HS-1 through 3, respectively).

The MEL cell colony assay was performed by electroporating approximately  $7.5 \times 10^6$  MEL cells (in late log phase growth) with 10  $\mu\text{g}$  of  $\gamma$ -neo DNA (or a molar equivalent of test constructions) at 120 V and 600  $\mu\text{F}$ . Transfected cells were transferred to 5 ml of complete media for 16 hours at 37°C in 5%  $\text{CO}_2$  and 100% humidity. 1.5 ml of cells were then diluted 1:10 with complete media containing 1 mg/ml G418. 100  $\mu\text{l}$  of diluted cells were transferred to each well of a 96 well plate. Individual colonies were counted after 14 days. Plating densities were adjusted so that no more than 40–50% of the wells contained colonies using the  $\gamma$ -neo/mLCR 9 (HS-2) construct.

#### Stable transfection of MEL cells with human $\beta$ -globin reporter constructs

MEL cells were stably cotransfected with a reporter construct and RSV-neo by electroporation (120 V, 600  $\mu\text{F}$ ) as described (30). Individual clones surviving in complete media containing 1 mg/ml G418 were isolated and expanded.

The presence of the human  $\beta$ -globin test gene was established by Southern blot analysis. DNA was prepared from individual clones, digested to completion with BamHI, electrophoresed on 1% agarose gels, and transferred to Genescreen Plus as described (31). Blots were hybridized consecutively with a 1.9 kb BamHI fragment derived from the 5' end of human  $\beta$ -globin gene, followed by a 1.7 kb EcoRI–BamHI fragment from the 5' end of mouse  $\beta$ -major globin gene. Human  $\beta$ -globin gene copy number was determined by densitometry and normalized to the mouse  $\beta$ -major globin gene signal.

#### S1 nuclease protection assays

RNA preparation, probe preparation, hybridizations, and S1 digestions were performed exactly as previously described (30).

#### Chromatin structure

DNA was prepared as described (32) with the following exceptions: nuclei from approximately  $10^8$  MEL cells were prepared by suspending cells in 4 ml RSB (10 mM Tris 7.4, 10 mM NaCl, 3 mM  $\text{MgCl}_2$ ); 1 ml of 0.5% NP-40/RSB was then slowly added to lyse the cells. Nuclei were harvested by centrifugation for 10 min at 500 $\times$ G, and were treated with DNaseI as previously described (32). DNaseI-treated DNA preparations were digested to completion with EcoRI, electrophoresed on 1% agarose gels and transferred to Genescreen Plus (New England Nuclear). Blots were hybridized with  $10^7$  cpm of a random primer  $^{32}\text{P}$ -labeled, 359 bp mouse LCR probe (an SspI–BglII fragment corresponding to bp –14239 to –13880). Blots were washed twice at 23° in  $2 \times \text{SSC}$  ( $1 \times \text{SSC} = 0.15 \text{ M NaCl}/0.015 \text{ M sodium citrate, pH 7.0}$ ), twice at 60° in  $2 \times \text{SSC}/1.0\% \text{ SDS}$ , and twice at 23° in  $0.1 \times \text{SSC}$ . Blots were autoradiographed at –70°C for 3 days.

## RESULTS

#### DNA sequence analysis of the $\lambda$ mLCR clone

We previously obtained a  $\lambda$  clone from a murine genomic DNA library using a probe derived from human 5' HS-2 (22). Using restriction analyses, we determined that the  $\sim 14$  kb insert extended from  $\sim -7$  kb to  $\sim -21$  kb upstream from the  $\epsilon$ -globin gene transcription initiation site. Initial DNA sequence analysis revealed that this clone contained the murine homologue of 5' HS-2, as previously reported (22). The relationship of the upstream sequences to the mouse  $\beta$ -globin cluster and a

comparison to the human  $\beta$ -like globin gene cluster are shown in Figure 1.

To identify the murine homologues of LCR 5' HS-3 and HS-4, we decided to sequence the entire  $\lambda$  mLCR clone. Since human 5' HS-3 is located approximately 14.7 kb upstream from the human  $\epsilon$ -globin gene, and since 5' HS-4 is approximately 18.0 kb upstream, we felt that it was quite likely that both homologous sites would be present on our clone. However, the DNA sequence analysis shown in dot matrix form (Figure 2) suggests that only 5' HS-3 is present.

The murine homologue of 5' HS-3 is located approximately 16.0 kb upstream from the  $\epsilon$ -globin gene. The region of homology with the functional human core extends from bp 5678 to 5870, corresponding to positions –16077 to –15885 upstream from the  $\epsilon$ -globin gene cap site. Since murine 5' HS2 is located at approximately 9.7 kb upstream from the  $\epsilon$  gene, the spacing between these two sites is approximately 6.3 kb, a distance that is substantially greater than the 3.8 kb that separates human 5' HS-3 (–14.7 kb) and 5' HS-2 (–10.9 kb) (1–3). The distance between human 5' HS-3 and HS-4 (–18 kb) is 3.3 kb. Since our DNA sequence extends approximately 5.7 kb upstream from 5' HS-3, and since no region with strong homology to human 5' HS-4 exists in our sequence, the distance between murine 5' HS-3 and 4 must be considerably greater in the mouse than in humans.

The murine LCR region was found to contain a large number of repetitive elements that may contribute to differences in the spacing between the mouse and human homologues of the LCR. For example, an L1 repeat extends from position  $\sim 2000$  to position  $\sim 3680$ , and a second L1 repeat extends between positions 8050 to 9750. While screening for probes to be used in Southern analyses of the LCR, many more highly repetitive areas were discovered. Three DNA probes derived from the region between –21401 and –19853 were repetitive, as were probes derived from the region between 18085 to –16286, –15217 to –14702, –14474 to –13880, and –11314 to –10917. A total of nine different probes derived from these various regions were moderately to highly repetitive in the mouse genome.

The dot matrix analysis shown in Figure 2 identifies the positions of the functional cores of the human 5' hypersensitive sites. Although the functional cores are highly conserved between mouse and human, additional regions of DNA homology outside of these cores are also conserved. However, the regions within and nearest to the functional cores of the major hypersensitive sites are most conserved regions within this large stretch of DNA.

To further assess the conservation of critical DNA sequences within 5' HS-3, we compared the DNA sequences of the human, mouse, and goat in a 'phylogenetic footprinting' analysis (33) (Figure 3). The functional core of human 5' HS-3 defined by Philipsen *et al.* (11) is enclosed by brackets. Within this region, five motifs are precisely conserved among the human, mouse, and goat species. Each of these motifs lies within a region that was footprinted *in vitro* by extracts derived from MEL cells (11); the footprinted sites are designated FP-1 through FP-6. Four of the conserved sites are GATA-1 sites, and one is a CACC/GT motif, a putative binding site for SP-1 or TEF2. Six potential CACC/GT motifs are present in the human HS-3 core, but only one (at position 5843) is conserved in both the mouse and the goat. Of note, an AP-1/NF-E2 consensus motif that lies just upstream from the functional core is highly conserved among all three species.

### 5' HS-3 is active when integrated into the chromatin of MEL cells but not K562 cells

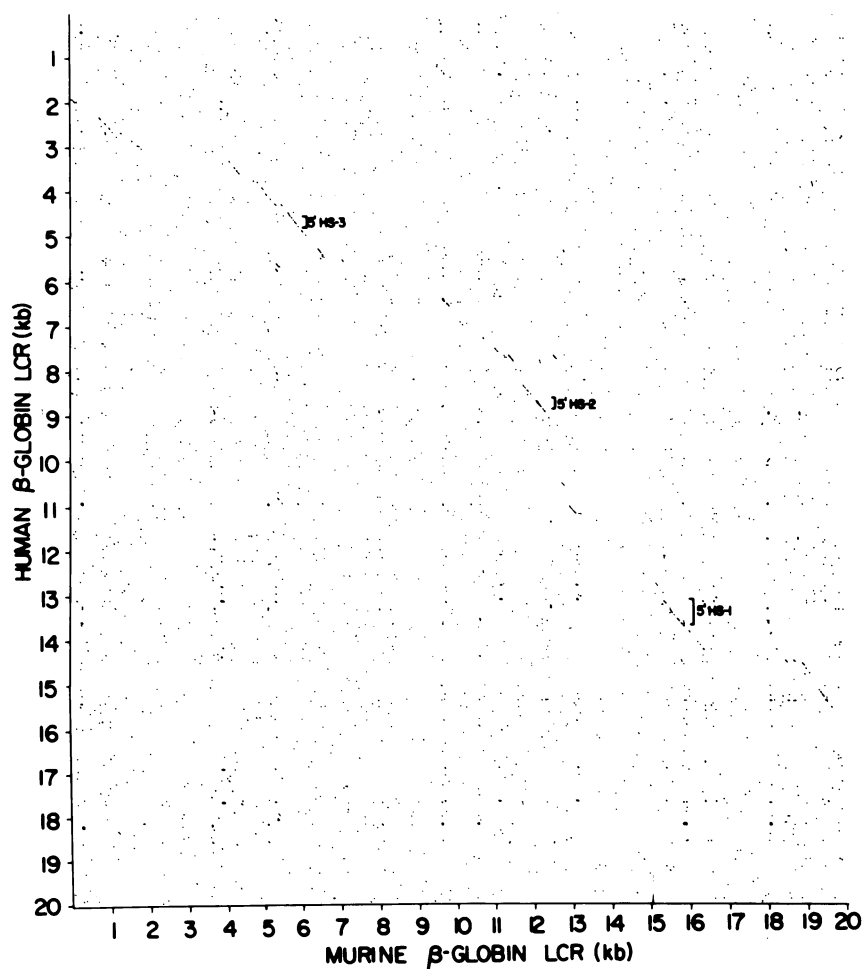
Since regions outside of the functional cores of 5' HS-2 and 5' HS-3 are also conserved between mouse and human, we decided to systematically test all of the DNA from position -21401 to position -9287 in functional assays (Figure 4). A series of nine DNA fragments obtained from this region were isolated and subcloned into pUC9; a 2.7kb BamHI fragment containing the  $\gamma$ -neo transcription unit was inserted into each plasmid. The  $\gamma$ -neo transcription unit is in a variety of positions and orientations with respect to each test fragment, but 5' HS-2 has been shown to function in all orientations and positions in transient and stable assays performed in K562 cells (22, 24). Each of these plasmids was transfected into K562 (embryonic/fetal erythroid) or MEL (adult erythroid) cells, and colony assays were performed. The number of G418-resistant clones per  $10^6$  transfected cells per microgram of transfected  $\gamma$ -neo DNA is summarized in Table 1. We had previously shown that human and mouse 5' HS-2 increase the number of G418 resistant clones obtained with the  $\gamma$ -neo expression plasmid by 50–100 fold (22, 24). Since the amount of  $\gamma$ -neo mRNA obtained from pools of clones with or without

5' HS-2 is identical (22, 24), we have suggested that 5' HS-2 acts to increase the probability that each integration event is 'productive'. As shown in Table 1, none of the DNA fragments except #9, containing 5' HS-2, substantially alters the number of colonies obtained with  $\gamma$ -neo in K562 cells. Fragment 5, which contains 5' HS-3, did not alter the number of colonies obtained. This data is consistent with previous experiments demonstrating that human 5' HS-3 has no activity in the K562 cell colony assay (24).

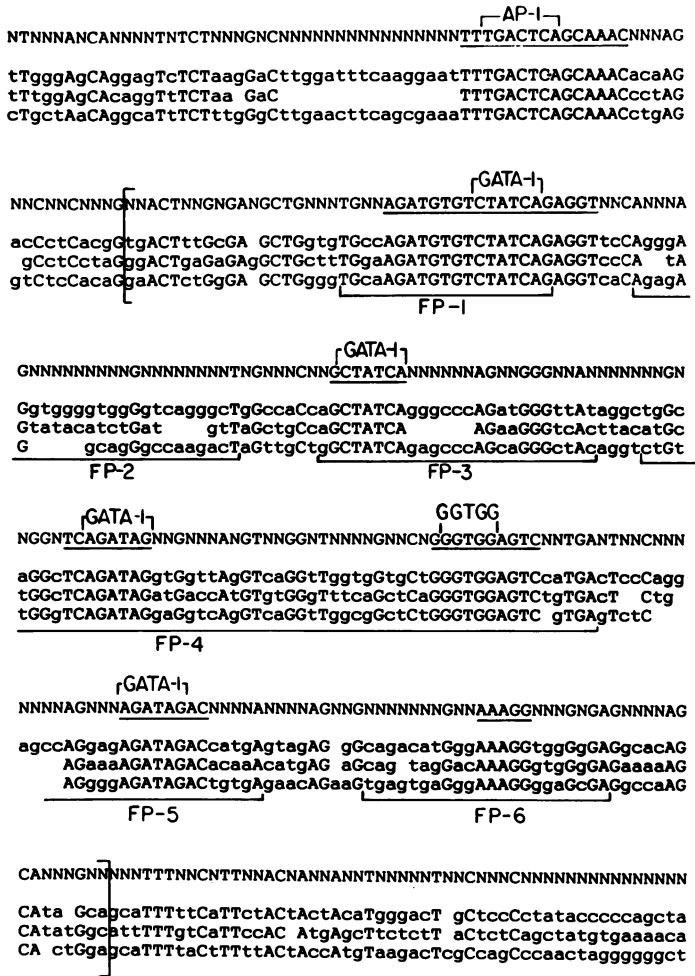
When the same set of plasmids were used in colony assay experiments in mouse erythroleukemia cells, plasmids containing either 5' HS-2 or 5' HS-3 demonstrated substantial colony assay activity over  $\gamma$ -neo alone (Table 1). No other LCR fragments demonstrated significant colony assay activity. The colony assay effect with 5' HS-3 is similar but slightly less than that of 5' HS-2; nonetheless, the effect is significant, and highly reproducible.

### 5' HS-3 is inactive in transient enhancer assays in both K562 and MEL cells

The same plasmids used in the colony assay described above were next used to perform transient transfection assays in both K562 and MEL cells. As shown in Figure 5, 5' HS-2 significantly



**Figure 2.** Dot matrix analysis of the region upstream from the human and mouse  $\beta$ -globin gene clusters. The analysis was performed with Microgenie<sup>R</sup> software. Each dot represents a position where there is 70% identity over a stretch of 20 bp. The locations of the functional cores and DNase I hypersensitive sites corresponding to human 5' HS-3, HS-2, and HS-1 are shown. On the X-axis, the location of the murine  $\epsilon$ -globin gene cap site is 21755. The location of the human  $\epsilon$ -globin gene cap site (on the Y axis) is 19504. The EMBL accession number for the upstream mouse LCR sequence is Z13985.



**Figure 3.** Sequences comprising the functional core of 5' HS-3 from the human (second line), mouse (third line), and goat (fourth line) are shown. The top line of each grouping is a consensus sequence from the three species, comprising a 'phylogenetic footprint' (33). All regions conserved among all three species that are 5 bp or greater in length are underlined. The location of the human 5' HS-3 functional core, as defined by Philipsen *et al.* (11), is bracketed. The locations of *in vitro* footprints from this region, as described by Philipsen *et al.* (11), are indicated below the sequence and designated FP-1 through FP-6. The locations of consensus sequences for transcription factors that bind to AP-1, GATA-1, and CACC/GT (GGTGG) sequences are indicated.

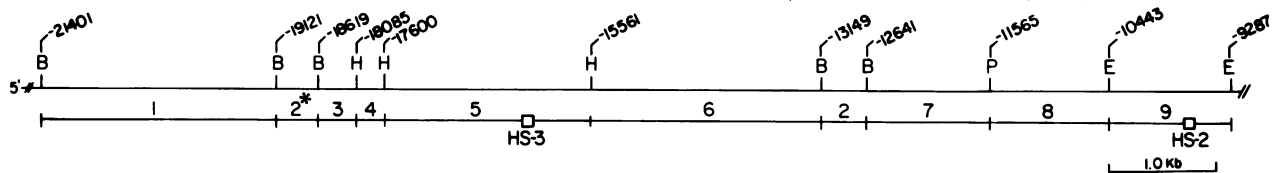
enhances the abundance of correctly initiated  $\gamma$ -neo mRNA obtained from transiently transfected K562 cells (panel A, compare lanes 1 and 2). However, 5' HS-3 has little or no activity in the same assay (panel A, lane 3). None of the other  $\gamma$ -neo test plasmids containing other regions of the mLCR enhanced  $\gamma$ -neo expression in K562 cells (data not shown).

The same transient transfection experiment was also performed in mouse erythroleukemia cells. Several different MEL cell lines were tested for their ability to be transiently transfected, but only one, a line maintained by William Forrester, Elliot Epner and Mark Groudine, was reproducibly transfectable by electroporation. The optimal electroporation conditions for transient transfections are listed in the Materials and Methods section. Again, 5' HS-2 substantially increased the level of  $\gamma$ -neo mRNA in transiently transfected MEL cells (Figure 5, panel B; compare lanes 1 and 2), but 5' HS-3 (lane 3) had little or no activity.

**5' HS-3 confers integration site-independent inducibility on a linked human  $\beta$ -globin gene in MEL cells**

Previous studies have shown that human 5' HS-2, 3, or 4, when linked in cis with a human  $\beta$ -globin gene, each confers copy number-dependent, integration site-independent, inducible expression on linked human  $\beta$ -globin genes in mouse erythroleukemia cells (8, 11, 12, 14, 17, 18). To determine whether murine 5' HS-3 has the same activity, we stably co-transfected MEL cells with RSV-neo and a human  $\beta$ -globin gene with or without a DNA fragment containing murine 5' HS-3, and selected clones based on the presence of the human  $\beta$ -globin gene. As shown in Figure 6, we isolated four independent clones containing the human  $\beta$ -globin gene alone. In all of these clones, the mouse  $\beta$ -major globin gene was inducible with DMSO (bottom panel). In clones # 1 and 6, constitutive levels of human  $\beta$ -globin mRNA were present, but these levels were not increased with successful DMSO inductions. Clones # 3 and 7 exhibited no expression of the  $\beta$ -globin gene with or without DMSO.

We isolated five independent clones with the human  $\beta$ -globin gene linked to murine 5' HS-3. Clones 1 and 3 had relatively high copy numbers of the human  $\beta$ -globin gene, but 7, 8, and 9 were all low copy number clones. All of the clones exhibited inducible expression of the mouse  $\beta$ -major globin gene. In the high copy number clones (1 and 3) substantial amounts of human  $\beta$ -globin mRNA were present in uninduced cells. In all five

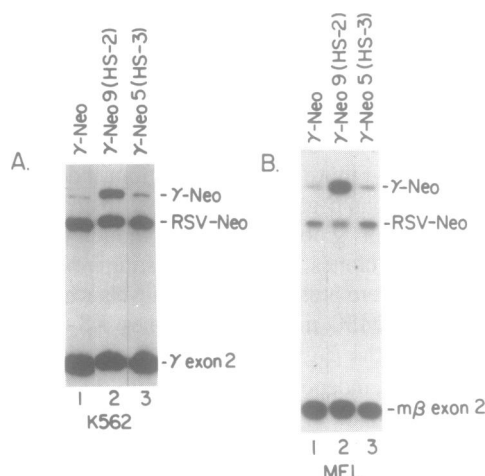


**Figure 4.** Map of DNA fragments used to test murine LCR functions. The DNA region contained within the  $\lambda$  mLCR clone is shown on the top line. The coordinates of various restriction sites, with respect to the  $\epsilon\gamma$ -globin gene cap site, are shown. The sites shown are selected and do not represent all the known sites for each enzyme. B = BamHI, H = HindIII, and E = EcoRI. On the lower line, the DNA fragments used for the functional analyses are shown. The region designated '2' is part of an L1 repeat and is nearly identical between positions -19121 to -18619 and positions -13149 and -12641. The actual fragment used was derived from the downstream BamHI fragment (-13149 to -12641); the upstream fragment was not specifically subcloned and tested, so it is designated '2\*'. The locations of the conserved cores of 5' HS-3 and 5' HS-2 are shown as open boxes. The numbering system of this figure is based on the location of the cap site of the  $\epsilon\gamma$ -globin gene, as reported by Sheehee *et al.* (29); our previously reported coordinates for the location of 5' HS-2 (22) were slightly displaced due to an editorial error.

**Table I.** K562/MEL cell colony assays

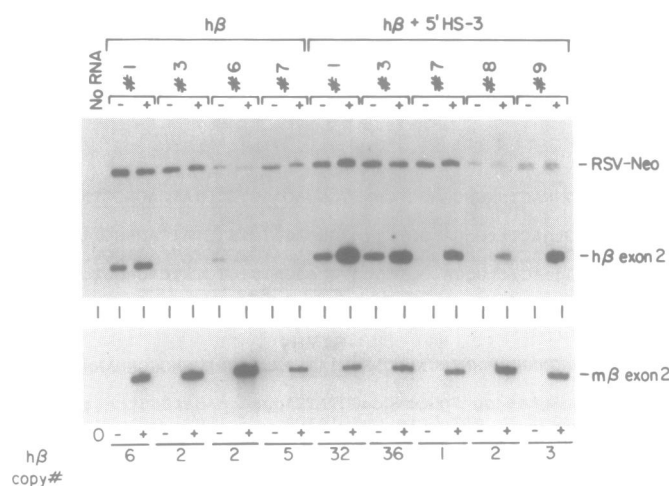
Test vector	# colonies per 10 <sup>6</sup> transfected cells per $\mu$ g transfected $\gamma$ -neo DNA (mean $\pm$ SD)			
	K562	N	MEL	N
None	< 1	6	< 1	4
$\gamma$ -neo	4 $\pm$ 2	6	22 $\pm$ 16	5
$\gamma$ -neo/mLCR 1	3 $\pm$ 1	4	31 $\pm$ 49	3
$\gamma$ -neo/mLCR 2	3 $\pm$ 2	4	21 $\pm$ 16	3
$\gamma$ -neo/mLCR 3	3 $\pm$ 1	2	21 $\pm$ 8	3
$\gamma$ -neo/mLCR 4	4 $\pm$ 2	2	41 $\pm$ 37	6
$\gamma$ -neo/mLCR 5 (HS-3)	2 $\pm$ 2	2	189 $\pm$ 89	6
$\gamma$ -neo/mLCR 6	2 $\pm$ 2	4	6 $\pm$ 0	3
$\gamma$ -neo/mLCR 7	3 $\pm$ 1	4	48 $\pm$ 26	3
$\gamma$ -neo/mLCR 8	4 $\pm$ 2	4	33 $\pm$ 28	3
$\gamma$ -neo/mLCR 9 (HS-2)	508 $\pm$ 180	5	252 $\pm$ 150	5

For each experiment,  $\sim 7.5 \times 10^6$  K562 or MEL cells in late log phase growth were electroporated with 10  $\mu$ g of the supercoiled pUC9-based  $\gamma$ -neo plasmid or an equimolar amount of LCR containing plasmids. Transfection and plating conditions are described in Materials and Methods. Numbers refer to the mLCR fragments shown in Figure 4. N = number of independent experiments.



**Figure 5.** Transient transfection analysis of murine LCR fragments in K562 and MEL cells. Murine LCR fragments 9 and 5 (5' HS-2 and 5' HS-3, respectively) were linked with a  $\gamma$ -neo reporter construct and transiently transfected into K562 or MEL cells, as described in Materials and Methods. RNA was harvested from the cells 16 hours after transfection, and was hybridized simultaneously with a 5' end-labeled  $\gamma$ -neo probe and a 5' end-labeled probe that detects correctly spliced  $\gamma$ -globin mRNA (for K562 cells) or mouse  $\beta$ -major globin mRNA (for MEL cells). S1 nuclease protection assays were then performed. The positions of correctly initiated  $\gamma$ -neo mRNA, RSV-neo mRNA (transfection control), and endogenous  $\gamma$ -globin exon 2 (K562, panel A; a control for RNA quality and content) or mouse  $\beta$ -major globin mRNA (MEL, panel B) are designated at the right side of the figure. Note that the  $\gamma$ -neo signal in lanes 2 ( $\gamma$ -neo 9, with 5' HS-2) is markedly enhanced over  $\gamma$ -neo alone (lanes 1). 5' HS-3 does not significantly alter expression of the  $\gamma$ -neo reporter plasmid (lanes 3). None of the other mouse LCR fragments had enhancer activity in either cell line (data not shown).

clones, the human  $\beta$ -globin gene appeared to be inducible with DMSO treatment. Murine 5' HS-3 therefore appears to function the same as human 5' HS-3 when integrated into in MEL cell chromatin.



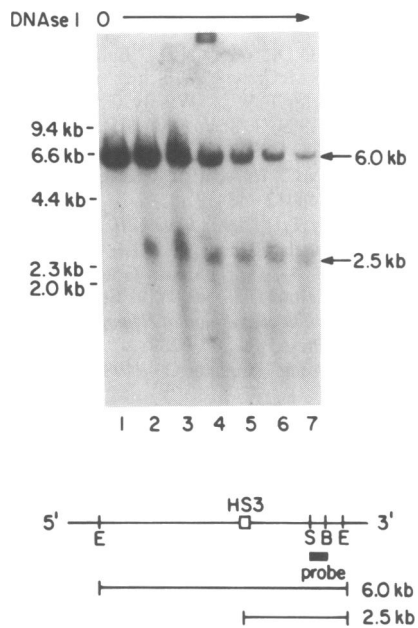
**Figure 6.** Analysis of human  $\beta$ -globin gene expression in MEL cell clones. pUC9 based plasmids containing the human  $\beta$ -globin gene without (H $\beta$ ) or with murine 5' HS-3 (h $\beta$  + 5' HS-3) were stably co-transfected into MEL cells with RSV-neo. G418 resistant clones containing the human  $\beta$ -globin gene were identified by Southern blot analysis. These clones were then expanded, and treated for five days with 1.0% DMSO. Ten micrograms of RNA derived from uninduced (-) or induced (+) MEL cells was hybridized with a probe that detects RSV-neo mRNA and a second probe end-labeled at the BamHI site located in exon 2 of the human  $\beta$ -globin gene. The locations of probe bands protected by RSV-neo mRNA, and of correctly spliced human  $\beta$  exon 2 mRNA, are shown on the right. In a separate analysis, 2 mg of mRNA from the same cell lines was hybridized with a probe 5' end-labeled at the BamHI site in mouse  $\beta$ -major globin exon 2, which detects correctly spliced mouse  $\beta$ -major globin mRNA (bottom panel). The DNA copy number of the human  $\beta$ -globin gene in each clone is indicated at the bottom of the figure. In four clones containing the human  $\beta$ -globin gene alone (# 1, 3, 6, and 7), there is no consistent relationship between copy number and constitutive expression of human  $\beta$ -globin mRNA. The human  $\beta$ -globin gene is not inducible in any clone. In contrast, all clones containing the human  $\beta$ -globin gene linked with 5' HS-3 exhibit inducible expression of the human  $\beta$ -globin gene. In two high copy number clones (# 1 and 3), the constitutive level of human  $\beta$ -globin mRNA is markedly elevated. Both clones exhibit inducible expression of the human  $\beta$ -globin gene; a separate analysis revealed that the level of induction was similar to that of the mouse  $\beta$ -globin gene when less mRNA was used (data not shown). The human  $\beta$ -globin probe must therefore be saturated in the induced lanes from clones # 1 and 3. Note that 5' HS-3 does not confer DMSO inducibility on RSV-neo.

### MEL cell chromatin contains a DNase I hypersensitive site at the conserved region of 5' HS-3

We obtained nuclei from uninduced MEL cells, treated them with increasing concentrations of DNase I, purified DNA, digested it to completion with EcoRI, and performed Southern blot analysis with a unique probe lying just downstream from HS-3. The results of this analysis, depicted in Figure 7, reveal a strong DNaseI hypersensitive site that maps to the conserved core of 5' HS-3.

### DISCUSSION

By sequencing a genomic  $\lambda$  clone extending 7 to 21 kb upstream from the mouse  $\beta$ -like globin gene cluster, we have identified conserved sequences that correspond to the functional cores of mouse Locus Control Regions 5' HS-2 and HS-3. The upstream region contains a large amount of repetitive DNA, but the spatial organization of the hypersensitive sites is conserved among the human, goat, and mouse species. In a series of functional assays, we have shown that murine 5' HS-3 actively influences linked promoters when it is integrated into the chromatin of MEL cells,



**Figure 7.** DNase I hypersensitivity analysis of MEL cell chromatin. Nuclei from uninduced MEL cells were prepared and treated with increasing amounts of DNase I. DNA was purified, cleaved to completion with EcoRI, electrophoresed on a 1% agarose gel, and transferred to Genescreen Plus. The blot was hybridized with a 359 bp SspI–BglII probe; the BglII site that defines the 3' end of the probe is located 452 bp upstream from the 3' end of the EcoRI fragment. The diagram at the bottom of the figure shows the position of key restriction sites (E = EcoRI, S = SspI, and B = BglII), and the position of the functional core of 5' HS-3. The location of the parental 6.0 kb EcoRI band is shown, and the position of the DNaseI dependent subband is indicated. The DNaseI hypersensitive site maps to the functional core of 5' HS-3.

but it is not a transient enhancer in MEL cells; HS-3 does not have any measurable activity in K562 cells. Finally, we have shown that the DNA sequences that comprise the core of murine 5' HS-3 are also DNase I hypersensitive in MEL cell chromatin. Collectively, these data suggest that we have identified the functional murine homologue of 5' HS-3.

DNA sequence analysis of  $\beta$ -globin LCR regions has revealed that the spatial organization of 5' HS-1, 2, and 3 is conserved among the human (1–3), mouse (24, 29, and this report), goat (27, 28), and rabbit (R. Hardison, personal communication). However, differences in the spacing of the three sites exist among the three species. Most of these spacing differences are due to the presence of a large number of repetitive sequences within each of the loci. In the human, these repetitive sequences are predominantly of the Alu type (34). In the goat, the repeats are of the D and NLA types (27, 28); in the mouse, L1 repeats flank each of the 5' hypersensitive sites. The spacing between human 5' HS-2 and HS-3 (3.8 kb) is substantially less than the distance between the mouse 5' HS-2 and HS-3 (6.3 kb). Similarly, if the mouse does have a functional 5' HS-4, it is substantially further upstream from 5' HS-3 than its human counterpart: in the human, 5' HS-4 lies 3.3 kb upstream from 5' HS-3. However, we have sequenced 5.7 kb upstream from murine 5' HS-3, and have not yet encountered the murine homologue of this site. Since human 5' HS-4 does function in the murine erythroid environment (10, 17, 18), we suspect that this site is most likely displaced further upstream.

The conservation of critical DNA sequences within the 5' HS-3 core is remarkable and similar in extent to the conservation found in 5' HS-2 (22, 27). An AP-1/NF-E2 site, four GATA-1 motifs, and one CACC/GT motif are all highly conserved among all three species (Figure 4). Philipsen *et al.* (11) defined a 225 bp core fragment of 5' HS-3 that conferred copy number-dependent, integration site-independent expression of linked human  $\beta$ -globin genes in transgenic mice. Using extracts derived from the nuclei of MEL cells, they demonstrated extensive DNaseI footprints of the 5' HS-3 region *in vitro*. These footprints, designated FP-1 through FP-6 (see Figure 4), suggested the presence of four GATA-1 binding sites and three sites containing the motif CACC/GT (a binding site for SP1 and/or the CACC box binding protein, TEF2). More recently, Strauss and Orkin (35) have performed *in vivo* DNase footprinting studies of the human 5' HS-3 region as it exists on human chromosome 11 in the MEL cell environment (HU11), the K562 environment, and a non-erythroid environment (HepG2 cells). Their studies revealed that the HS-3 region was not footprinted in HepG2 nor in K562 cells; Ikuta and Kan (36) similarly showed that 5' HS-3 was not footprinted *in vivo* in K562 cells. In contrast, the 5' HS-3 region displayed limited *in vivo* footprints in HU11 cells, and the footprints became more complex upon DMSO induction. All four consensus GATA motifs were footprinted in uninduced and induced HU11 cells. Although six CACC/GT motifs were footprinted *in vitro*, only two of these elements were contacted in HU11 cells *in vivo*. Of these two sites, one is absolutely conserved among humans, mice, and goats, but the other is unique to the human 5' HS-3. Finally, our sequence analysis has demonstrated a conserved AP-1/NF-E2 binding site just upstream from the functional core defined by Philipsen *et al.* (11). Strauss and Orkin (35) demonstrated that this site was *in vivo* footprinted in HU11 cells, and that DMSO induction caused additional contacts to appear. Again, this AP-1/NF-E2 site was not footprinted in HepG2 cells. Collectively, these data suggest that essentially all of the 5' HS-3 motifs that are absolutely conserved in sequence and position among human, mice, and goats are engaged by proteins *in vivo*. The techniques of phylogenetic footprinting (33) and *in vivo* DNase I footprinting are therefore at least somewhat complementary, and may be useful for identifying critical DNA sequence elements for functional mutagenesis studies.

Even though murine 5' HS-3 exhibits powerful influences on linked promoters when stably integrated into the chromatin of MEL cells, this site has no detectable activity when transiently transfected into MEL or K562 cells; these data suggest that 5' HS-3 is not a classical enhancer. Furthermore, although 5' HS-3 is active in the chromatin of MEL cells, it exhibits no activity in the chromatin of K562 cells (20, 24). These functional observations are supported by the studies of Dhar *et al.* (37) and of Strauss and Orkin (35), who demonstrated that the organization of chromatin in the 5' HS-3 region is quite different in K562 vs. MEL cells. Collectively, these data suggest that the activity of 5' HS-3 may be influenced by the type of erythroid environment present in the test cell. Additional experiments will be required to determine whether 5' HS-3 does in fact play a role in the developmental regulation of the  $\beta$ -like globin genes.

The identification of the functional murine homologue of 5' HS-3 has provided us with reagents that will be useful for further delineation of the mechanisms by which the LCR functions. Using mouse tissues and developing murine hematopoietic cells, we will

be able to define the events that lead to chromatin reorganization in the murine LCR region during hematopoietic and/or erythroid development. Furthermore, the identification of murine 5' HS-2 and 3 will allow us to perform targeted disruption studies of these sites in embryonic stem cells and in MEL cells (38). These studies should allow us to begin to determine the relative contributions of 5' HS-2 and HS-3 to the overall organization of the chromatin structure of the  $\beta$ -like globin genes during erythroid development.

## ACKNOWLEDGMENTS

The authors wish to thank members of the Ley Laboratory for stimulating discussions, Kathleen Maloney for expert technical assistance, and Diana Coleman-Peters for expert preparation of the manuscript. We thank Erich Strauss and Stuart Orkin, and Ross Hardison for sharing unpublished data with us. This work was supported by NIH DK 38682 and CA 49712.

## REFERENCES

- Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6384–6388.
- Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1359–1363.
- Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987) *Nucleic Acids Res.* **15**, 10159–10177.
- Ley, T.J. (1991) *Blood* **77**, 1146–1152.
- Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7470–7474.
- Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990) *Genes Dev.* **4**, 1637–1649.
- Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) *Cell* **51**, 975–985.
- Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. and Greaves, D.R. (1989) *Nature* **338**, 352–355.
- 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.
- Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) *Nucleic Acids Res.* **18**, 3503–3508.
- Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990) *EMBO J.* **9**, 2159–2167.
- Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) *EMBO J.* **9**, 2169–2178.
- Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1626–1630.
- Talbot, D. and Grosveld, F. (1991) *EMBO J.* **10**, 1391–1398.
- Liu, D., Chang, J.C., Moi, P., Liu, W., Kan, Y.W. and Curtin, P.T. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3899–3903.
- Forrester, W.C., Novak, U., Gelinis, R. and Groudine, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5439–5443.
- Collis, P., Antoniou, M. and Grosveld, F. (1990) *EMBO J.* **9**, 233–240.
- Novak, U., Harris, E.A.S., Forrester, W., Groudine, M. and Gelinis, R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3386–3390.
- Ney, P.A., Sorrentino, B.P., McDonagh, K.T. and Nienhuis, A.W. (1990) *Genes Dev.* **4**, 993–1006.
- Sorrentino, B., Ney, P., Bodine, D. and Nienhuis, A.W. (1990) *Nucleic Acids Res.* **18**, 2721–2731.
- Moi, P. and Kan, Y.W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9000–9004.
- Moon, A.M. and Ley, T.J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7693–7697.
- Ney, P.A., Sorrentino, B.P., Lowrey, C.H. and Nienhuis, A.W. (1990) *Nucleic Acids Res.* **18**, 6011–6017.
- Moon, A.M. and Ley, T.J. (1991) *Blood* **77**, 2272–2284.
- Tuan, D., Abeliovich, A., Lee-Oldham, M. and Lee, D. (1987) In Stamatoyannopoulos, G. and Nienhuis, A.W. (eds), *Developmental Control of Globin Gene Expression*. Alan R. Liss, Inc., New York, pp. 211–220.
- Tuan, D.Y.H., Solomon, W.B., London, I.M. and Lee, D.P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2554–2558.
- Li, Q., Zhou, B., Powers, P., Enver, T. and Stamatoyannopoulos, G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8207–8211.
- Li, Q., Zhou, B., Powers, P., Enver, T. and Stamatoyannopoulos, G. (1991) *Genomics* **9**, 488–499.
- Shehee, W.R., Loeb, D.D., Adey, N.B., Burton, F.H., Casavant, N.C., Cole, P., Davies, C.J., McGraw, R.A., Schichman, S.A., Severynse, D.M., Voliva, C.F., Weyter, F.W., Wisely, G.B., Edgell, M.H. and Hutchinson, C.A. III (1989) *J. Mol. Biol.* **205**, 41–62.
- Ulrich, M. and Ley, T.J. (1990) *Blood* **75**, 990–999.
- Hohn, P.A., Popescu, N.C., Hanson, R.D., Salvesen, G. and Ley, T.J. (1989) *J. Biol. Chem.* **264**, 13412–13419.
- Hanson, R.D. and Ley, T.J. (1990) *Mol. Cell. Biol.* **10**, 5655–5662.
- Gumucio, D.L., Blanchard-McQuate, K.L., Heilstedt-Williamson, H., Tagle, D.A., Gray, T.A., Tarle, S.A., Gragowski, L., Goodman, M., Slightom, J.L. and Collins, F.S. (1991) In Stamatoyannopoulos, G., and Nienhuis, A.W. (eds), *The Regulation of Hemoglobin Switching*. The Johns Hopkins University Press, pp. 277–289.
- Li, Q., Powers, P.A. and Smithies, O. (1985) *J. Biol. Chem.* **260**, 14901–14910.
- Strauss, E.C. and Orkin, S.H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5809–5813.
- Ikuta, T. and Kan, Y.W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10188–10192.
- Dhar, V., Nandi, A., Schildkraut, C.L. and Skoultschi, A.I. (1990) *Mol. Cell. Biol.* **10**, 4324–4333.
- Kim, C.G., Epner, E.M., Forrester, W.C. and Groudine, M. (1992) *Genes Dev.* **6**, 928–938.