Nucleic Acids Research, 1993, Vol. 21, No. 5 1265–1272

Comparative analysis of the locus control region of the rabbit β-like globin gene cluster: HS3 increases transient expression of an embryonic ε-globin gene

Ross Hardison, Jia Xu, John Jackson, James Mansberger, Olga Selifonova, Brent Grotch, Jessica Biesecker, Hania Petrykowska and Webb Miller[1]

Departments of Molecular and Cell Biology and [1]Computer Science, Center for Gene Regulation and Institute for Molecular Evolutionary Genetics, 206 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

Received October 23, 1992; Revised and Accepted January 26, 1993 GenBank accession nos L05833, L05835

ABSTRACT

The rabbit homolog to the locus control region (LCR) of the human β-like globin gene cluster was isolated, and long segments containing the DNase I hypersensitive sites (HS) were sequenced. The order and spacing of HS4, HS3, HS2 and HS1 are conserved between rabbit and human. Alignment of these sequences with their homologs from human, goat, and mouse shows that very long segments of DNA match between species, for over a thousand base pairs on either side of the previously identified functional cores, indicating that some important functions are found outside the cores. The activity of rabbit HS2 and HS3 was tested by attaching each to a novel reporter gene constructed by inserting the luciferase coding region into the rabbit ε-globin gene. In contrast to previous reports showing no effect of human or mouse HS3 on transient expression, both the rabbit HS2 and HS3 DNA fragments separately increased transient expression from the ε-luciferase hybrid gene and expression from stably integrated constructs in K562 erythroleukemia cells.

INTRODUCTION

The locus control region (LCR) is a large segment of DNA located at the 5′ end of the β-like globin gene cluster that exerts profound effects on the expression of these genes. A set of developmentally stable, DNase hypersensitive sites are found both 5′ and 3′ to the gene cluster (1, 2), and are called 5′HS1–5 and 3′HS1 (3). Recent data show that although all these sites are formed in erythroid cells, only 5′HS3 is completely erythroid-specific (4). This report is concerned only with the 5′ hypersensitive sites, so for brevity they will be referred to as HS1–HS4.

The LCR exerts a variety of effects on linked globin genes. When added to a human β-globin gene, the LCR confers high-level, position-independent, copy-number dependent expression in transgenic mice (5). Examination of mutant chromosomes from patients with Hispanic γδβ-thalassemia demonstrates that the LCR is required for expression of the β-like globin genes. Each of the globin genes on this chromosome is intact, yet none are expressed in these patients, who are missing the portion of the LCR containing HS2–4 (6). This deletion places the globin genes in a 'closed' chromatin conformation and delays their time of replication (7). The LCR will also confer high-level, inducible expression of linked β-like globin genes in stably transfected MEL cells, in a manner independent of the position of integration (8, 9, 10). Thus the LCR appears to carry at least three identifiable functions: opening of a chromatin domain (assayed by position-independent, copy number-dependent expression of linked genes), enhancement (assayed by high level of expression in transgenic mice or in transiently transfected cultured cells) and induction (assayed in stably transfected cells). Additionally, the LCR exerts complex effects on developmental regulation of the globin genes (recently reviewed by Orkin (3)).

In an effort to simplify the analysis of the LCR, individual hypersensitive sites have been examined for functional properties. HS2 will enhance the transient expression of a β-globin gene in unintegrated constructs after transfection of cultured erythroid cells (11, 12). This effect is independent of the orientation of the HS2 fragment, characteristic of a classic enhancer effect. Only fragments of the human LCR carrying HS2 have been shown to be effective in this assay (11, 13). DNA fragments containing HS2 are sufficient to confer position-independent expression of the human β-globin gene in transgenic mice (14, 15, 16). Interaction of the nuclear factor NFE2, which recognizes a site similar to that of AP1, with HS2 is required but not sufficient for enhancement of expression of linked genes (17, 18).

DNA fragments containing HS3 can also confer position-independent expression of the human β-globin gene in transgenic mice (19). In fact, HS3 has the strongest effect among the four individual hypersensitive sites examined in this assay (20). It is also active when integrated in MEL cells, allowing induction of linked genes (10, 21, 22). Surprisingly, no effect of either human

or mouse HS3 has been observed on either enhancement or activation after integration in K562 cells, a human erythroleukemia cell line (13, 22).

DNA fragments containing HS4 will confer position-independent expression of the human β-globin gene in transgenic mice, at a level comparable to that obtained with HS2 (20). However, little effect has been observed from HS1 alone, although it may act in concert with other hypersensitive sites to confer full LCR function (20).

In order to obtain greater insight into the structure of the LCR and to serve as a guide to its functional dissection, the homologs to the LCR are being examined in other species. Segments of the β-like globin gene cluster containing HS1, HS2 and HS3 have been isolated and sequenced from both goat (23, 24) and mouse (22, 25). Previous analysis of the rabbit β-like globin gene cluster began in the region homologous to HS1 (26, 27, 28). This paper reports the isolation of chromosomal DNA further 5′, containing HS4, HS3 and HS2. Comparisons with other mammals show that the order and spacing of the hypersensitive sites are well-conserved, and the sequence conservation extends far beyond the functional cores. We also find that rabbit HS2 and HS3 can stimulate expression in K562 cells of a novel reporter gene in which a luciferase gene is fused into the coding region of the rabbit ε-globin gene.

MATERIALS AND METHODS

Isolation of the rabbit β-globin LCR

The cloned library of rabbit genomic DNA (29) was screened by hybridization with a labeled probe from the 5′ end of clone λRβ′G3 (30) that is single-copy DNA (26) and homologous to human HS1 (27, 28). Clone λRβG8.3 was isolated in this screen (Fig. 1). Clones extending further 5′ were isolated by re-screening with an EcoRI to HindIII fragment from λRβG8.3 to obtain clones λRβG41.1 and λRβG33.1 (Fig. 1). Maps of the clones were constructed by partially digesting with restriction endonucleases, annealing with radioactive oligonucleotides complementary to either *cos*L or *cos*R, and resolving the digestion products on an agarose gel (31). Restriction fragments containing the homologs to specific hypersensitive sites of humans were identified by hybridization with radioactive probes from the human LCR (10).

Sequence determination

Restriction fragments containing the homologs to the human hypersensitive site regions were subcloned into one of the Bluescript family of phagemid vectors (Stratagene), and the sequence determined by the dideoxynucleotide chain-termination method (32) on double-stranded templates. The DNA sequences were obtained either from both strands (for most of the sequence) or multiple times from the same strand. The sequences are deposited in GenBank under accession number L05833 for the HS3 to HS2 region and L05835 for the HS4 region.

Sequence comparisons

Sequences of the LCR regions were aligned between pairs of species using the program *sim*, which produces optimal local alignments in a space-efficient manner (33). The parameters used for most comparisons are match = 1, mismatch =-1, gap open penalty = 6, gap extension penalty = 0.2 per position in the gap. However, since rodents apparently diverged from other mammals before most eutherians diverged from each other, and additionally are diverging at a more rapid rate (34), pairwise alignments involving mouse were done at a lower gap open penalty of 5; other parameters were the same as in other comparisons. Only those local alignments whose similarity scores exceed a certain threshold τ were examined, where the probability is 0.05 of finding a gap-free alignment scoring at least τ between two random sequences matching the test sequences in length and nucleotide composition (35). The program *laps* (36) generates plots showing the positions of the aligning segments, together with sequence features along the axes. The percent identity (number of matching pairs between consecutive gaps in the alignment divided by the number of positions in that gap-free alignment) is plotted along the same coordinates as the abscissa of the *laps* plot to provide information about the strength of the alignment. Blocks that consistently align between human, rabbit, goat and mouse (or just three of the four species) were computed from the pairwise alignments by a new version of the program *pab* (37, 38) and plotted in the same format as a *laps* diagram (Figure 3B). Multiple alignments over longer stretches (Figure 4) were computed by a new program called *yama* using the same scores for matches, mismatches, and gaps as in the pairwise *sim* alignments.

Construction of the ε-luciferase reporter gene

A 2638 bp PstI fragment containing the rabbit ε-globin gene (39) was cloned into the pBlueScriptIKS+ vector to make the construct pBSε. This plasmid was cut at the HindIII site in exon1 and a SalI site in the 3′ polylinker to remove intron1, exon2, intron2, exon3, and the 3′ flank of the ε-globin gene. The coding portion of the luciferase cDNA (40) was excised from pGEMluc (supplied by Promega) at the HindIII site in the 5′ polylinker and the SalI site in the 3′ polylinker, and fused in frame at the HindIII site in exon1 of the ε-globin gene to make an intermediate construct, pBS5′ε-luc. pBSε was cut at a BamHI site in exon2, blunt-ended, and cut again at a KpnI site in the 3′ polylinker to excise 19 bp of exon2 along with all of intron2, exon3, and the 3′ flank of the rabbit ε-globin gene. This BamHI to KpnI fragment was ligated to pBS5′ε-luc, opened in the 3′ polylinker at XhoI (blunted-ended) and KpnI, to generate pBSε-luc. This hybrid construct contains 581 bp of the 5′ flank and the first 85 bp of exon1 of the rabbit ε-globin gene, fused in frame with the coding portion plus 60 bp past the termination codon of luciferase. The remaining part of the rabbit ε-globin gene (BamH1 to KpnI) constitutes a long 3′ untranslated region; it provides a splicable intron and polyadenylation signals followed by 504 bp of 3′ flank. Thus this reporter gene maintains the genomic context of the embryonic ε-globin gene. Only intron1 and most of exon2 of the rabbit ε-globin gene are missing from the construct. The rabbit HS3 region (a 450 bp StuI to ScaI fragment extending from positions 1158 to 1608 in the rabbit sequence) was added in both orientations 5′ to the ε-luciferase gene to make plasmid rHS3rev-ε-luc (reverse of the normal genomic orientation) and rHS3nat-ε-luc (native genomic orientation). Likewise two different restriction fragments containing rabbit HS2, a 1.0 kb PstI fragment with the HS2 core in the middle and a 2.2 kb HindIII fragment with the HS2 core at the end proximal to the ε-globin gene, were inserted 5′ to the ε-luciferase gene to make rHS2(PP1.0)rev-ε-luc and rHS2(HH2.2)rev-ε-luc, respectively. Derivatives of ε-luciferase were also constructed containing the human HS2 fragment (376 bp HindIII to XbaI fragment, nucleotides 8486 to 8861) and the human HS3 fragment (a 778 bp PstI to PvuII fragment, nucleotides 4344 to 5122).

Transfection of mammalian cells

K562 cells were maintained in RPMI plus 10% bovine calf serum, 1% penicillin/streptomycin, and 25 mM HEPES pH 7.3. For transient expression assays, 5×10^6 K562 cells in 0.5 ml cold TBS (140 mM NaCl, 5 mM KCl, 25 mM Tricine, pH 7.4, 0.5 mM $MgCl_2$, 0.7 mM $CaCl_2$) were transfected using a Promega X-Cell 450 electroporator at 500 μFarad and 450 V for 500 ms. DNA for transfections included varying amounts of test DNA, 2.5 μg of pB19CAT (or 10 μg pRSV *lac*Z), and additional pBlueScript carrier DNA to make a constant amount of 50 μg DNA per transfection. The plasmid pB19CAT (from the laboratory of A.Nienhuis) contains the gene for chloramphenicol acetyl transferase (CAT) driven by a B19 parvovirus promoter; this serves as a control for transfection efficiency. Alternatively, the *E.coli* gene for β-galactosidase driven by an RSV long terminal repeat, pRSV *lac*Z (41), was used as a co-transfection control. HeLa cells were transfected with 10 μg luciferase test construct plus 10 μg pRSV *lacZ* by the Ca phosphate precipitation method (42). Cells were harvested 48 h after transfection, washed with a buffered salt solution, pelleted and lysed in 100 μl of a solution containing 25 mM Tris phosphate pH 7.8, 2 mM dithiothreitol, 2 mM 1,2-diaminocyclohexane-N,N,N′,N′-tetraacetic acid, 1.1 M glycerol, 16 mM Triton X-100 and 1 mg bovine serum albumin/ml. 5 μl of each extract was assayed for luciferase activity at 23°C for 10 s in 100 μl assay solution containing 20 mM Tricine, 1.07 mM $(MgCO_3)_4Mg(OH)_2$, 2.67 mM $MgSO_4$, 0.1 mM EDTA, 33.3 mM dithiothreitol, 270 μM coenzyme A, 470 μM luciferin, and 530 μM ATP, using a Berthold Lumat LB9501 luminometer. As appropriate, 20 μl of each extract was also assayed for chloramphenicol acetyltransferase (43) or β-galactosidase activity (41), and the luciferase activity (in relative light units, or RLU) was divided by either the fraction of chloramphenicol acetylated to obtain a luciferase/CAT ratio or by the A_{420} of the product *o*-nitrophenol to obtain a luciferase/β-gal ratio.

To produce stable transformants, 1×10^7 K562 cells in 1 ml cold TBS were transfected with 90 μg of the test plasmid (pBSεluc, with or without HS2 or HS3) or a negative control (pGL2basic from Promega) together with 10 μg pM5Neo as a selectable marker, using the same electroporation conditions as for the transfections described above. The test plasmids and pGL2basic were linearized with KpnI and pM5Neo was linearized with SspI. The plasmid pM5Neo contains the Tn5 neomycin resistance gene driven by the long terminal repeat of the myeloproliferative sarcoma virus; it confers resistance to the drug G418 in hematopoietic cells (44). 24 h after electroporation of the cells, G418 was added to a final concentration of 1.2 mg/ml. Clones of stable transformants were isolated from soft agar cultures; extracts were prepared from 3×10^6 cells and assayed for luciferase activity as described above.

RESULTS

Isolation and mapping of the rabbit β-globin LCR

The rabbit homolog to the human LCR was isolated by 'walking' 5′ to the previously isolated clones (29, 30). The maps of the new clones are presented in Fig. 1, along with the positions of regions that match the cores of the hypersensitive sites in the human LCR. The positions of the hypersensitive sites are almost identical between human and rabbit (Fig. 1), as are HS3, 2 and 1 between human and goat (23), whereas the spacing between HS3, 2 and 1 is longer in mouse (22). No homolog to human HS4 has been reported to date in goat (24), yet it is present in rabbit at the same position as in human.

Sequence comparisons in the HS3 to HS2 region of the LCR

The sequence of the 6.6 kb Eco RI fragment spanning the rabbit regions that hybridize to human HS3 and HS2 was determined (Fig. 1). This rabbit sequence aligns closely with the human sequence, with the expected breaks resulting from insertion of retroposon-type short repeats (Fig. 2). As shown on the lower panel, the percent identity of these local alignments is high, even in the regions far removed from the functional cores for HS2 (16) and HS3 (19). These are more extensive matches than had been previously detected in pairwise comparisons between human and goat (23) or human and mouse (22, 25). A portion of the HS2 core has one of the highest percentages of identical nucleotides, and a region of about 1500 bp around the HS3 core scores as high as the core itself (Fig. 2).

The extensive, high-scoring alignments between rabbit and human in the HS3 to HS2 region may indicate strong sequence conservation throughout the region. To investigate this further, pairwise alignments were made between four available sequences—those from human, rabbit, goat and mouse. The plots in Fig. 3A show long alignments in all pairwise comparisons. Much of the goat sequence aligns with human and rabbit sequences, but less so with the mouse. However, the region

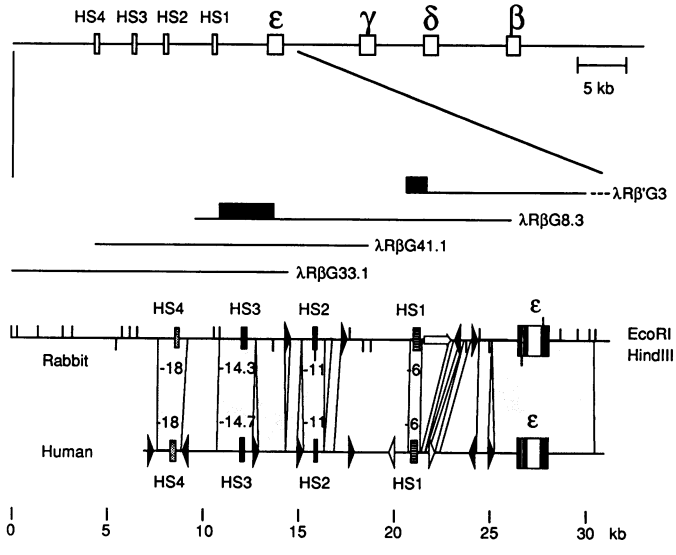

Figure 1. Map of the LCR of the rabbit β-like globin gene cluster. A schematic view of the rabbit gene cluster is shown on the top line, and an expanded diagram of the 5′ end is aligned with the homologous region in human. The positions of cloned DNAs extending 5′ to the originally isolated clones are shown above the expanded view. The probes used to screen the rabbit library are shown as dotted boxes on the maps of the clones. The DNase hypersensitive sites in the human LCR, and the matching regions in rabbits, are shown as boxes with distinctive fills. Segments of sequence that align between the two species are connected by lightly shaded areas. Short interspersed Alu repeats in humans and C repeats in rabbits are shown as filled triangles pointing in the direction of their 3′ A-rich tails; and long interspersed L1 repeats are shown as unfilled triangles or open arrows. The exons of the ε-globin gene are shown as filled boxes and introns as unfilled boxes. The sequence of a DNA segment containing rabbit HS4 (GenBank L05835) aligns with the human sequence (54). The unshaded regions between HS4 and HS3 and between HS2 and HS1 have not been sequenced in rabbit.

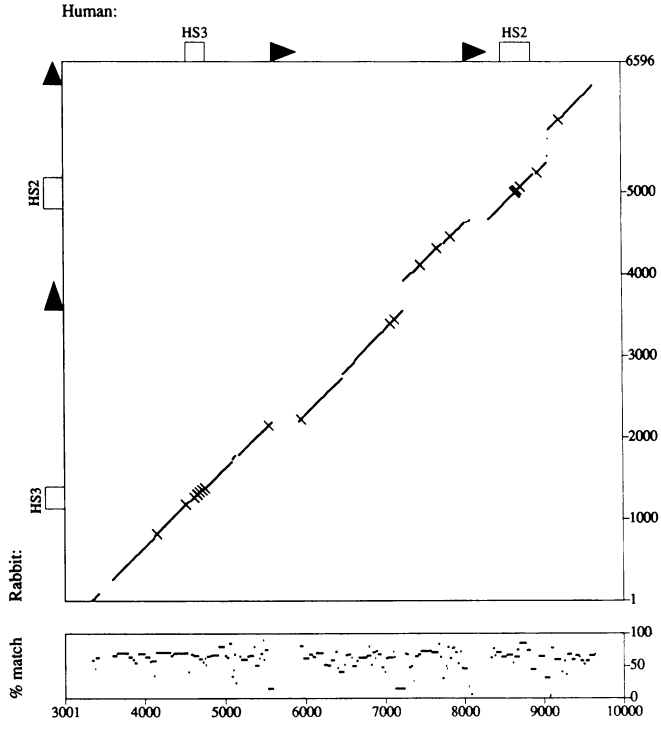

Figure 2. Positions of aligning sequences in the HS3 to HS2 region between human and rabbit. Local alignments between the two sequences were computed and displayed as described in Materials and Methods. The positions of aligning segments that match the binding sites for known transcription factors are indicated by the short lines perpendicular to the plotted diagonal. The cores of the hypersensitive sites (16, 19) are shown as open boxes on the axes. The lower panel plots the percent identity in each segment between gaps in the alignment, as a function of the position in the human sequence. The human sequence is from Li et al. (55), and the rabbit sequence is from this report (GenBank accession number L05833).

approximately between positions 3500 to 5000 in the goat sequence does not match with human, rabbit or mouse sequences. This unique sequence in goat is bounded by two NlaI repeats, suggesting that perhaps it transposed in a mechanism involving these retroposons in a replacement-type recombination (45). Much of the mouse sequence aligns with human. A long interspersed L1Md repeat between HS3 and HS2 of mouse causes a major disruption in the alignments with human and rabbit, as do several insertions of short repeats in this region. The L1Md insertions account for much of the additional DNA between HS3 and HS2 that is unique to mouse, as previously noted (22). The region just 3′ to the long L1Md in mouse does not align extensively with rabbit at these criteria, although it does with human. In many of the comparisons, the region 3′ to HS2 shows limited alignments, indicating that it may not be as well conserved as other sequences in this area.

A more stringent test for conservation is to find sequence segments whose pairwise alignments can be combined in a consistent manner to create a multiply aligned block (37, 38). These consistently aligning blocks are displayed in Fig. 3B in the coordinates of the rabbit−human comparisons. The human−rabbit−goat plot shows those human−rabbit aligning blocks that are also found in the rabbit−goat and goat−human pairwise alignments. Surprisingly, about 3500 bp around HS3 aligns consistently in all three pairwise comparisons, and about 1000 bp aligns around HS2. The major disruption in the human−rabbit−goat comparison is the nonaligning region between the goat NlaI repeats mentioned above. The three-way comparison among human−rabbit−mouse shows consistently aligning blocks extending far 3′ to HS3 and 5′ to HS2. In the most stringent test, the six pairwise comparisons among human, rabbit, goat and mouse still show consistently aligning blocks extending for about 1000 bp each around both HS3 and HS2.

These consistently aligning blocks are much more conserved than are the regions between globin genes, which will match in pairwise alignments but are largely subject to neutral evolution (28). For example, almost all of the pairwise matches between human and rabbit seen in the δ-β intergenic region of the β-like globin gene cluster are lost when filtered with rabbit-mouse and mouse-human comparisons. The comparison of the HS3-HS2 region among these three species produced 57 consistently aligning blocks of total length 1345 bp (plotted in Fig. 3B), whereas under identical conditions (sequence length and program parameters) a three-way comparison in the δ-β-globin intergenic region produced only 7 consistently aligning blocks of total length 112 bp. Each of the 9 longest consistently aligning blocks in the HS3-HS2 region had a score that was extremely unlikely to occur by chance, even in isolation. For example, the probability that 3 random sequences of comparable length and nucleotide composition have an aligning block scoring no lower than the second highest-scoring HS3-HS2 block was computed as 2×10^{-14} (35). In contrast, the two lowest probabilities computed for blocks in the δ-β-globin intergenic region were 0.25 and 0.74. Thus the matches in the HS3-HS2 region are highly significant and are much more extensive than those previously seen in intergenic regions. This strongly indicates that these matches in the HS3 to HS2 region result from sequence conservation and do not represent matches remaining after drift from a common ancestor.

Invariant sequences in the HS3 and HS2 regions

A multiple alignment of the human, rabbit, goat and mouse sequences in these well-conserved regions reveals several invariant sequences, i.e. the subregions of consistently aligning blocks that are absolutely identical (Fig. 4). The alignments in the HS2 region (Fig. 4A) show almost no sequence variation in the previously characterized binding sites for NFE2, C-ACC-binding protein or GATA1 (16). Additionally, the nucleotide string between positions 8700 and 8708 in human (CC-AGATGTT) is also invariant (Fig. 4A). Two NFE-2 binding sites are in human and mouse, but three are in goat and rabbit. The repeating AT motif from 8884 to 8937 in human and additional GATA1 sites 3′ to it, noted by Talbot et al. (16), are not conserved.

The alignments in the HS3 region reveal an invariant NFE2 site at positions 4511−4519 in the human sequence (Fig. 4B). Invariant sequence blocks in the functional core defined by Philipsen et al. (19) (extending from nucleotides 4552 to 4776 in the human sequence) are largely confined to four GATA1 sites and one CACC site. Additional sequences invariant among all four species are found over 300 bp 3′ to the HS3 core (plotted in Fig. 3B).

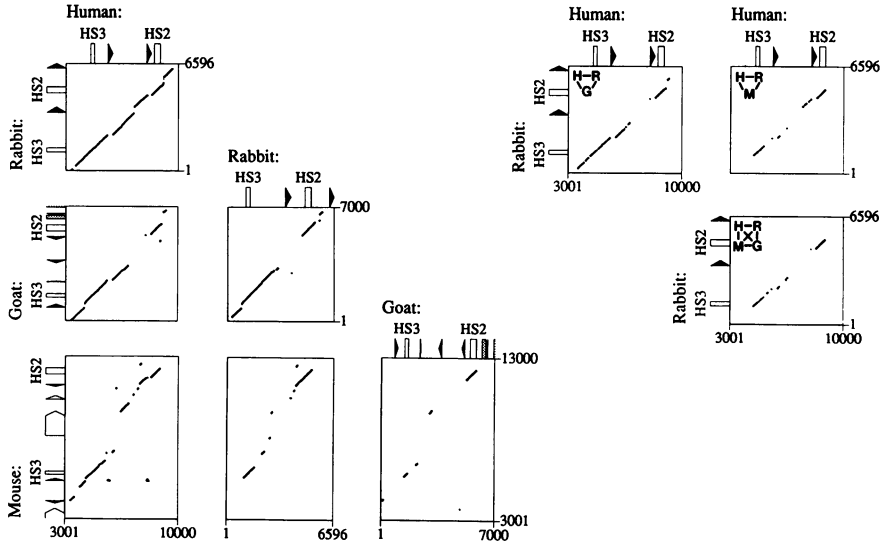

Figure 3. Comparisons of the HS3 to HS2 region among four species. (A, left group) Plots of the pairwise alignments among the available sequences were generated as in Fig. 2. Additional short interspersed repeats denoted by filled triangles are NlaI repeats in goats (24) and B1 or B2 repeats in mouse. The long interspersed L1Md repeats in mouse are shown as unfilled boxed arrows, pointing in the direction of their A-rich 3′ ends, and the D repeats in goats (24) are shown as stippled boxes. The goat sequence is from Li et al. (24) and the mouse sequence is from Moon and Ley (25) and Hug et al. (22). (B, right group) Plots of consistently aligning sequence blocks among multiple sequences, in each case plotted along the axes of the human−rabbit pairwise comparison. For example, in the human−rabbit−goat comparison (H−R−G), the human−rabbit pairwise alignment was compared with the rabbit-goat alignment and the goat-human alignment to find consistent matches.

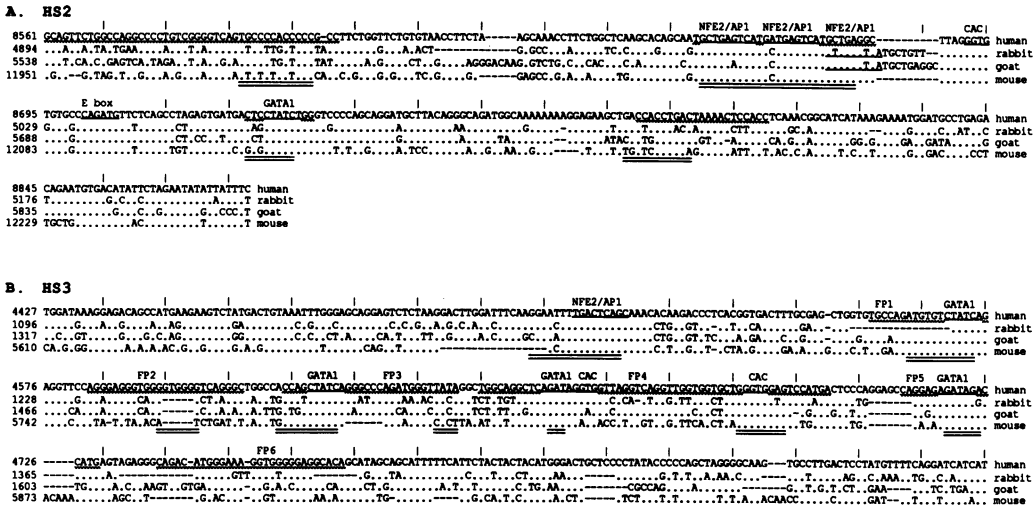

Figure 4. Multiple alignments in the HS2 **(A)** and HS3 **(B)** regions of the LCR. Dots are used for nucleotides that are identical to that in human, and dashes show gaps in the alignment. Sequences that are footprinted *in vitro* (16, 19) have a dotted underline, and sequences footprinted *in vivo* in human (50, 51, 53) are doubly underlined. Binding sites for known proteins within the invariant blocks are singly underlined and labeled. The additional NFE2/AP1 sites in goat and rabbit HS2 are also underlined.

Activation by rabbit HS2 and HS3 in transfected K562 cells

In order to show that the rabbit homolog to the human β-globin LCR would function in erythroid cells, a sensitive assay system for globin gene expression in K562 cells was devised. These human erythroleukemia cells express embryonic and fetal globin genes, so the rabbit embryonic ε-globin gene was chosen as a heterologous globin gene for transfection. This gene is transiently expressed after introduction into K562 cells by electroporation (46). The internal and 3′ flanking region of mammalian ε-globin genes are well-conserved (26, 27), suggesting that they may contain regulatory elements. We wished to add a reporter gene whose product could be rapidly assayed in a sensitive manner, but without disrupting potential regulatory elements. Thus intron 1 and most of exon 2 of the rabbit ε-globin gene were replaced by the coding region for firefly luciferase (Fig. 5), fusing exon 1 of the globin gene in frame with the luciferase coding region.

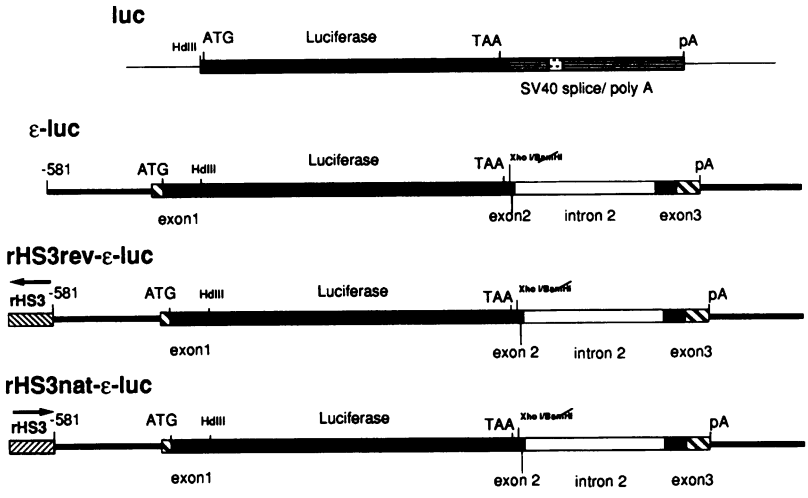

Figure 5. The ε-luciferase reporter gene and derivatives. The coding region of the firefly luciferase gene (dark stippled box), from the plasmid pGEMluc, was used to replace intron 1 and most of exon 2 of the rabbit ε-globin gene by an in-frame fusion. The rabbit HS3 region (a 450 bp StuI to ScaI fragment, denoted by the light stippled box) was added 5′ to the the ε-luciferase gene in both orientations. DNA from the rabbit ε-globin gene is shown as thick lines (flanking regions), filled boxes (polypeptide coding portions of the exons), striped boxes (untranslated portions of exons) and unfilled boxes (intron2). Translation initiates at the AUG of the ε-globin mRNA and terminates at the UAA at the end of the luciferase coding region. A promoterless luciferase gene with SV40 signals for splicing and polyadenylation, pGL2basic from Promega (labeled luc in the figure), served as a negative control.

Expression of this ε-luciferase gene should produce a hybrid protein, with the first 28 amino acids of rabbit ε-globin gene attached to the N terminus of luciferase via a linker of 12 amino acids encoded by the multiple cloning sites of pGEMluc. The ε-luciferase reporter gene is transiently expressed after electroporation into K562 cells in a manner dependent on the DNA concentration, whereas a promoterless luciferase construct produced no activity (Fig. 6A).

The effect of the rabbit HS3 region was tested by inserting a 450 bp StuI to ScaI fragment, corresponding to positions 4489 to 4993 in the human sequence (Fig. 4B), 5′ to the ε-luciferase gene in the reverse orientation to that found on the chromosome (Fig. 5). At each DNA concentration, the amount of luciferase activity, corrected for transfection efficiency by co-transfection with a CAT reporter plasmid, was greater for the rHS3rev-ε-luciferase construct than for the parental ε-luciferase construct (Fig. 6A). The increase in expression ranged from 5 to 10 fold depending on the DNA concentration in the transfection. This increase is seen with both orientations of the HS3 fragment, although the activity with the reverse orientation tends to be somewhat higher than with the native orientation (Fig. 6B). A similar result is obtained with the human HS3 fragment (Fig. 6B). The human HS2 DNA fragment will give an increase in expression of about 15 fold, and the rabbit HS2 will produce a similar activation (data not shown). Thus the rabbit HS3 can increase expression of the ε-luciferase reporter gene in unintegrated constructs in K562 cells, but not as much as HS2. Neither HS3 nor HS2 will activate expression of ε-luciferase in nonerythroid HeLa cells (Table 1).

The strongest effect of both HS2 and HS3 on the expression of ε-luciferase is seen after stable integration into the genome. Clones of K562 cells stably transfected with ε-luciferase plasmids, with or without rabbit HS2 and HS3 fragments, were isolated and tested for luciferase activity. The averaged values in Table 1 show that the ε-luciferase reporter was active when integrated into the K562 genome, but inclusion of the rabbit LCR fragments increased expression about 250-fold for HS3 and 540-fold for HS2.

DISCUSSION

The sequencing and comparisons of LCRs from several different mammalian species should be a helpful guide to the functional dissection of this complex regulatory element. The clones containing the rabbit LCR cover regions homologous to HS1, HS2, HS3 and HS4 in human, similar to those previously reported for goat and mouse (22, 23, 25, 47). Although HS4 was not in the goat clones analyzed, it is likely that a homolog is in the region further 5′. The spatial orientation of the hypersensitive sites in the LCRs from these four species is the same, and despite the insertions of several short repeats and deletions of other sequences, the distance between hypersensitive sites is remarkably similar in human, rabbit and goat (Fig. 1; (23)). The distance between HS2 and HS3 is larger in mouse, largely caused by the insertion of L1Md long interspersed repeats (Fig. 3A; (22)). The distance between HS3 and HS4 is polymorphic in mouse, with two alleles having distances of 5 and 6.5 kb (47), both of which are longer than the spacing in humans and rabbits, 3.3 and 3.7 kb, respectively (Fig. 1). These observations suggest that the LCR may be most effective with a certain spacing between the hypersensitive sites. Granted that micro-LCRs containing short fragments with the hypersensitive sites in juxtaposition are powerful *cis*-acting elements in several assays (10, 20), full LCR function may be dependent on some critical, albeit flexible, spacing parameters.

The regions between the hypersensitive site cores appear to be under selective pressure. Pairwise alignments show that the sequences match with high similarity scores for very long distances (thousands of bp) around the core regions. Many of these LCR segments align consistently in multiple species (Fig. 3B), in contrast to the bulk of the intergenic matches within the gene clusters, which have been ascribed to matches that remain after neutral evolution from a common ancestor (28). Thus the matches in the LCR segments are stronger than those found elsewhere in the intergenic regions of the gene cluster, arguing strongly that they are being selected for some function. However, many of the invariant segments, which are candidates for sequence-specific binding proteins, are clustered in the cores of

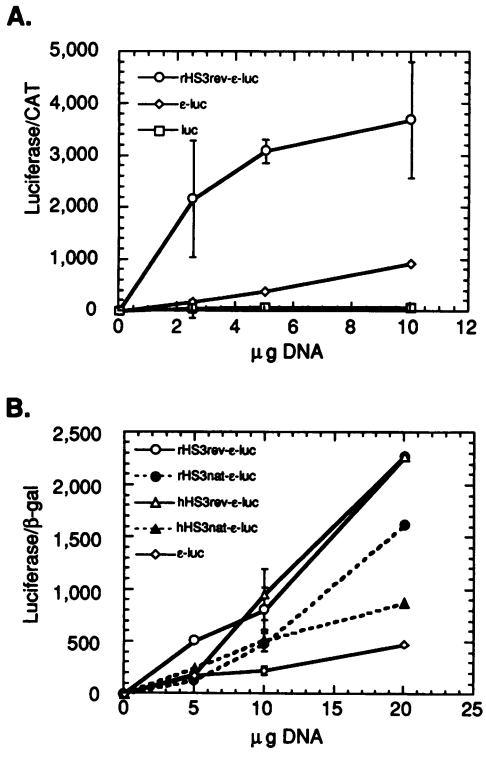

Figure 6. Effect of the rabbit HS3 fragment on expression of the ε-luciferase gene in K562 cells. **(A)** Transient expression of ε-luciferase with and without HS3. Three different amounts of the test plasmids were transfected in triplicate into K562 cells, along with the plasmid pB19CAT (17) as a control for transfection efficiency. Luciferase activity was normalized to the CAT activity, and the averages (± standard deviation) are plotted as a function of amount of transfecting DNA. The diamonds are values for the parental reporter ε-luciferase (ε-luc), circles are for rHS3rev-ε-luciferase, and the squares are for the negative control, the promoterless plasmid pGL2basic (luc). **(B)** Comparison of orientation and species of origin of HS3 in transient expression assays. K562 cells were transfected with ε-luciferase constructs containing rabbit or human HS3 fragments in either orientation. Assays using 10 μg test plasmid were done in triplicate; those with 5 and 20 μg were single transfections. Luciferase activity in relative light units (RLU) was corrected for transfection efficiency by dividing by the β-galactosidase activity (A_{420}) to obtain the Luciferase/β-gal numbers plotted on the ordinate. Abbreviations are rHS3 = rabbit HS3, hHS3 = human HS3, nat = native orientation, rev = reverse orientation.

the hypersensitive sites (Fig. 4). This suggests that some of these conserved regions outside the core may be acting by a mechanism not dependent on precise recognition of unique sequences, such as binding to a nuclear matrix or establishing some altered chromatin structure.

Segments that do not vary in sequence between species have been referred to as 'phylogenetic footprints' (48) and frequently correspond to binding sites for sequence-specific nuclear proteins (49). In many cases the invariant segments of the LCRs correlate well with the *in vitro* and *in vivo* footprinting results on the human hypersensitive sites, but several specific differences are notable. In HS2, the repeated NFE2 sites and GATA1 site identified as footprints *in vitro* (16) and *in vivo* (50, 51) are highly conserved in all four species (Fig. 4A). However, *in vivo* footprints observed around CACC motifs at positions 8592–8606 (51) and 8787–8797 (50) in the human sequence are not in well-conserved regions. In contrast, an additional CACC motif (GGTG at positions 8691–8694) and a sequence CCAGATGTT, located at positions 8700–8708, are both invariant in the four species. No *in vivo* footprints have been observed in this latter segment

Table 1. Effect of rabbit LCR fragments on expression of ε-luciferase

Construct	HeLa cells transient	K562 cells stable transformants	
	Luc/β-gal	(RLU) Luciferase	fold increase
mock	0		
pGL2control (SV40-luc)	3480 ± 2800		
ε-luc	33 ± 28	563	1
rHS3rev-ε-luc	127 ± 72	143,000	254
rHS2(PP1.0)rev-ε-luc	217 ± 303		
rHS2(HH2.2)rev-ε-luc		306,000	543
hHS2nat-ε-luc		473,000	840
pGL2basic (luc)		31	

For the stable transformants, the averaged luciferase activity from 3 to 11 individual clones containing each construct is presented.

(50, 51), but it matches the consensus sequence for an E box, i.e. CANNTG (52) (Fig. 4A). The entire region between the repeated NFE2 sites and the GATA1 site is very highly conserved, and likely plays an important role in HS2 function.

For HS3, Philipsen et al. (19) mapped six *in vitro* footprints in the functional core. Of these, the GATA-1 sites in footprints 1, 3, 4 and 5 are virtually invariant in the four species (Fig. 4B), and these are also contacted by proteins *in vivo* (53). However, of the GGTGG (or CACC) motifs previously noted in the human sequence, only the one in footprint 4 (positions 4687–4690 in human) is invariant, and again this is contacted by proteins *in vivo* (53). Although footprint 2 contains several GGTGG motifs in the human sequence (19) and is footprinted *in vivo* (53), the least variant sequence in this region is GAGGG (positions 4586–4590 in human), and even this is missing in mouse (Fig. 4B). *In vitro* footprint 6 (19) is not contacted *in vivo* (53), nor is it well-conserved. Just 5′ to the functional core of HS3 is a match with a NFE2 site (19) that is invariant among the available sequences (Fig. 4B, (22)). This position is also contacted *in vivo* (53). This NFE2 site and more distal invariant sequences are notable candidates for additional functional elements, although they are not required for position-independent expression (19).

In K562 cells, either the rabbit HS2 or HS3 fragments will stimulate expression of an ε-luciferase reporter gene, both from unintegrated plasmids and after selection for stably integrated plasmids (Fig. 6, Table 1). This is expected from the ability of the homologous sequences in human to confer high-level, position-independent expression of the human β-globin gene in transgenic mice (20) and in MEL cells (19). The increased expression in either orientation shows that HS3 has properties of a classical enhancer. However, earlier studies had indicated that only HS2 would act as an enhancer without integration (11), and no effect of either the human or mouse HS3 was seen in several different expression assays in K562 cells, although these same constructs were active in MEL cells (13, 22). These several experiments are not directly comparable. One major difference is the reporter genes used. We have developed a luciferase reporter that not only is expressed from the rabbit ε-globin gene promoter, but it also retains intronic and flanking sequences that could contain regulatory elements or sequences needed for interaction with the LCR. Previous experiments relied on the ability of the LCR fragments to affect the SV40 promoter (11), or used a test construct with the promoter region of the human γ-globin gene but no internal or flanking sequences (13, 22). Other potentially important factors include the exact fragment selected to cover HS3, since conserved sequences are found far

from the core. Our results show that the rabbit DNA described here is not only structurally homologous to the human LCR, but both HS3 and HS2 have a positive effect on linked globin genes. Further analysis is required to test the ability of the rabbit LCR fragments to establish high-level expression independently of the position of integration and dependent on the copy number. Such experiments are in progress.

ACKNOWLEDGEMENTS

We thank Drs W.Forrester, E.Epner, and M.Groudine for the 'miniLAR' plasmid with human HS1−4, Drs T.Shimada and A.Nienhuis for the pB19CAT plasmid, Drs J.Thompson and R.Raghow for the pRSV *lacZ* plasmid, Dr T.Ley for providing the sequence and manuscript on mouse HS3 prior to publication, and J.Newman, L.Bursey, and B.Bour for constucting and testing the plasmid rHS2(PP1.0)rev-ϵ-luc. This work was supported by PHS grants RO1 DK27635 and RO1 HL44491, an RCDA KO4 DK01589 to R.C.H., and grant RO1 LM05110 to W.M.

REFERENCES

1. Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384−6388.
2. Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1359−1363.
3. Orkin, S. (1990) *Cell*, **63**, 665−672.
4. Dhar, V., Nandi, A., Schildkraut, C.L. and Skoultchi, A.I. (1990) *Mol. Cell. Biol.*, **10**, 4324−4333.
5. Grosveld, F., van Assendelft, G.B., Greaves, D. and Kollias, G. (1987) *Cell*, **51**, 975−985.
6. Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7470−7474.
7. Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T. and Groudine, M. (1990) *Genes Develop.*, **4**, 1637−1649.
8. van Assendelft, G.B., Hanscombe, O., Grosveld, F. and Greaves, D. (1989) *Cell*, **56**, 969−77.
9. Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. and Greaves, D.R. (1989) *Nature*, **338**, 352−355.
10. Forrester, W.C., Novak, U., Gelinas, R. and Groudine, M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5439−5443.
11. Tuan, D., Abelovich, A., Lee-Oldham, M. and Lee, D. (1987) In G. Stamatoyannopoulos and A. W. Nienhuis (ed.), Developmental Control of Globin Gene Expression. A. R. Liss, Inc., New York. pp. 211−220.
12. Tuan, D., Solomon, W., London, I. and Lee, D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2554−2558.
13. Moon, A.M. and Ley, T.J. (1991) *Blood*, **77**, 2272−2284.
14. Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D. and Brinster, R.L. (1989) *Genes Devel.*, **3**, 314−323.
15. Curtin, P.T., Liu, D., Liu, W., Chang, J.C. and Kan, Y.W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7082−7086.
16. Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) *EMBO J.*, **9**, 2169−2178.
17. Ney, P., Sorrentino, B., McDonagh, K. and Nienhuis, A. (1990) *Genes Devel.*, **4**, 993−1006.
18. Moi, P. and Kan, Y.W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9000−9004.
19. Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990) *EMBO J.*, **9**, 2159−2167.
20. Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) *Nucl. Acids Res.*, **18**, 3503−3508.
21. Collis, P., Antoniou, M. and Grosveld, F. (1990) *EMBO J.*, **9**, 233−240.
22. Hug, B.A., Moon, A.M. and Ley, T.J. (1992) *Nucl. Acids Res.*, **21**, 5771−5778.
23. Li, Q., Zhou, B., Powers, P., Enver, T. and Stamatoyannopoulos, G. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8207−8211.
24. Li, Q., Zhou, B., Powers, P., Enver, T. and Stamatoyannopoulos, G. (1991) *Genomics*, **9**, 488−499.
25. Moon, A.M. and Ley, T.J. (1990) *Proc. Natl. Acad. Sci., USA*, **87**, 7693−7697.
26. Margot, J.B., Demers, G.W. and Hardison, R.C. (1989) *J. Mol. Biol.*, **205**, 15−40.
27. Hardison, R.C. (1991) In R. K. Selander, T. S. Whittam and A. G. Clark (ed.), Evolution at the Molecular Level. Sinauer Associates, Inc., Sunderland, MA. pp. 272−289.
28. Hardison, R. and Miller, W. (1993) *Mol. Biol. Evol.*, in press.
29. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) *Cell*, **15**, 687−701.
30. Lacy, E., Hardison, R.C., Quon, D. and Maniatis, T. (1979) *Cell*, **18**, 1273−1283.
31. Rackwitz, H.-R., Zehetner, G., Frishchauf, A.-M. and Lehrach, H. (1984) *Gene*, **30**, 195−200.
32. Sanger, F.S., Nicklen, S. and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 503−517.
33. Huang, X., Hardison, R. and Miller, W. (1990) *Computer Appl. Biosci.*, **6**, 373−381.
34. Li, W.-H., Gouy, M., Sharp, P., O'hUigin, C. and Yang, Y.-W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6703−6707.
35. Karlin, S. and Altschul, S.F. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2264−2268.
36. Schwartz, S., Miller, W., Yang, C.-M. and Hardison, R. (1991) *Nucl. Acids Res.*, **19**, 4663−4667.
37. Boguski, M.S., Hardison, R.C., Schwartz, S. and Miller, W. (1992) *The New Biologist*, **4**, 247−260.
38. Miller, W. (1993) *Computer Appl. Biosciences*, **9**, in press.
39. Hardison, R.C. (1983) *J. Biol. Chem.*, **258**, 8739−8744.
40. de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. and Subramani, S. (1987) *Mol. Cell. Biol.*, **7**, 725−737.
41. Thompson, J.P., Simkevich, C.P., Holness, M.A., Kang, A.H. and Raghow, R. (1991) *J. Biol. Chem.*, **266**, 2549−2556.
42. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T. and Silverstein, S. (1979) *Cell*, **16**, 777−785.
43. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044−1051.
44. Laker, C., Stocking, C., Bergholz, U., Hess, N., De Lamarter, J.F. and Ostertag, W. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8458−8462.
45. Zelnick, C.R., Burks, D.J. and Duncan, C.H. (1987) *Nucleic Acids Res.*, **15**, 10437−10453.
46. Price, D.K. (1992) Ph.D. thesis, The Pennsylvania State University, pp. 127−172.
47. Jimenez, G., Gale, K.B. and Enver, T. (1992) *Nucl. Acids Res.*, **20**, 5797−5803.
48. Tagle, D.A., Koop, B.F., Goodman, M., Slightom, J., Hess, D.L. and Jones, R.T. (1988) *J. Mol. Biol.*, **203**, 7469−7480.
49. Gumucio, D.L., Heilstedt-Williamson, H., Gray, T.A., Tarle, S.A., Shelton, D.A., Tagle, D., Slightom, J., Goodman, M. and Collins, F.S. (1992) *Mol. Cell. Biol.*, **12**, 4919−4929.
50. Reddy, P.M.S. and Shen, C.-K.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 8676−8680.
51. Ikuta, T. and Kan, Y.W. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10188−10192.
52. Blackwell, T.K. and Weintraub, H. (1990) *Science*, **250**, 1104−1110.
53. Strauss, E.C. and Orkin, S.H. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5809−5813.
54. Chao, K.-M., Hardison, R.C. and Miller, W. (1993) *Computer Appl. Biosciences*, **9**, in press.
55. Li, Q., Powers, P.A. and Smithies, O. (1985) *J. Biol. Chem.*, **260**, 14901−14910.