THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 20, Issue of May 17, pp. 11871–11878, 1996

Printed in U.S.A.

Role of DNA Sequences Outside the Cores of DNase Hypersensitive Sites (HSs) in Functions of the β-Globin Locus Control Region

DOMAIN OPENING AND SYNERGISM BETWEEN HS2 AND HS3*

(Received for publication, November 2, 1995, and in revised form, January 31, 1996)

John D. Jackson‡§, Hanna Petrykowska‡, Sjaak Philipsen¶, Webb Miller§‖, and Ross Hardison‡§**

From the ‡Department of Biochemistry and Molecular Biology, the ‖Department of Computer Science and Engineering, and the §Center for Gene Regulation, The Pennsylvania State University, University Park, Pennsylvania 16802, and ¶MGC, Department of Cell Biology and Genetics, Erasmus University, Rotterdam, The Netherlands

The roles of each DNase hypersensitive site (HS), and the DNA sequences between them, in the activity of the locus control region of the mammalian β-globin gene domain were examined by placing human and rabbit restriction fragments containing the cores of HS2, HS3, HS4, and HS5, along with varying amounts of flanking DNA, upstream of a hybrid ϵ-globin-luciferase reporter gene and testing for effects on expression both prior to and after integration into the chromosomes of K562 cells, a human erythroid cell line. Prior to integration, fragments containing HS2 enhanced expression to the greatest extent, and the modest enhancement by some fragments containing HS3 correlated with the presence of a well-conserved binding site for AP1/NFE2. The stronger effects of larger locus control region DNA fragments in clones of stably transfected cells indicates a role for sequences outside the HS cores after integration into the genome. The strong effect of a 1.9-kilobase *Hin*dIII fragment containing HS3 after, but not prior to, integration argues for the presence of a chromatin domain-opening activity. Use of a rabbit DNA fragment containing both HS2 and HS3 demonstrated a synergistic interaction between the two HSs when their natural context and spacing are preserved.

The expression of the β-like globin gene cluster in birds and mammals is regulated both by proximal elements, such as promoters, and a distal locus control region, or LCR.[1] The LCR will greatly increase the level of expression of linked reporter genes in erythroid cells and will allow expression of constructs regardless of their position of integration in the genome of transgenic mice (1). Deletion of the LCR leaves the gene cluster in a chromatin conformation that is inaccessible to DNase I (2). Hence, several functions have been proposed for the LCR, including enhancement, insulation from position effects, and activation of a chromosomal domain (reviewed in Refs. 3 and 4).

The LCR is a large regulatory region covering about 20 kb of DNA at the 5′ end of the gene cluster. It is marked by 5 DNase hypersensitive sites, HS1–HS5 (5, 6). The number and spacing of these sites is very well conserved in mammals (7–11). The positions of DNase I cleavage have been mapped precisely, and the minimal sequences required for position-independent expression of a linked β-globin gene in transgenic mice have been established for HS2 (12–14), HS3 (15, 16), and HS4 (17, 18). These 200–400-bp minimal regions will be referred to as the cores of the HSs. The DNA sequences within the cores are very well conserved, especially for HS2. In addition, prominent conserved blocks of sequence are also found outside the cores of the HSs (7, 11, 19), but the functional significance of these sites is largely untested at present.

Numerous studies have examined the roles of individual HSs in various expression assays. HS1, -2, -3, and -4 will produce position-independent expression of linked globin genes in transgenic mice, each with a characteristic level of expression and a preferred developmental stage (20, 21). In transfected erythroid cells, HS2 will enhance expression of β-like globin genes both transiently from unintegrated constructs (22, 23) and after stable integration into a chromosome (12, 24). HS3 will enhance stable expression in MEL cells (9, 24, 25) and transiently in K562 cells (11). HS4 has a much stronger effect in transgenic mice than in stably transfected MEL cells (17) and will generate a DNase hypersensitive site in transgenic mice (18). HS5 has been implicated in insulation from some position effects (26, 27), analogous to the situation in the chicken β-like globin gene cluster (28). It is not yet clear how these individual activities lead to the full effect of the LCR, and it is likely that the HSs act together for some or perhaps all functions. Indeed, the quantitative effects of the individual HSs on expression in transgenic mice are markedly less than the effects of the entire LCR (*e.g.* Refs. 13, 17, 24, and 29), and Collis *et al.* (24) have shown that restriction fragments containing the cores of three HSs, but not two HSs, are needed for levels of expression in stably transfected cells approaching that of the entire LCR. The full activity of a specific function, such as domain opening or enhancement, may require some combination of HSs, possibly including DNA sequences between the HS cores.

In order to examine the effects of sequence elements both inside and outside the cores of the HSs on expression, restriction fragments of increasing lengths have been added to an ϵ-globin-luciferase hybrid gene and tested in the human cell line K562, which produces ϵ-, γ-, ζ-, and α-globins. Tests of expression of the reporter gene prior to integration (transient expression) reveal enhancement function, and measurement of levels of expression after integration into a chromosome (stable expression) should show the effects of enhancement, domain

* This work was supported by United States Public Health Service Grants DK27635 and LM05573 (to R. H.) and LM05110 (to W. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, 206 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802. Tel.: 814-863-0113; Fax: 814-863-7024; E-mail: rch8@psu.edu.

[1] The abbreviations used are: LCR, locus control region; kb, kilobase(s); bp, base pair(s); HS, hypersensitive site; RSV, Rous sarcoma virus; Tricine, *N*-tris(hydroxymethyl)methylglycine.

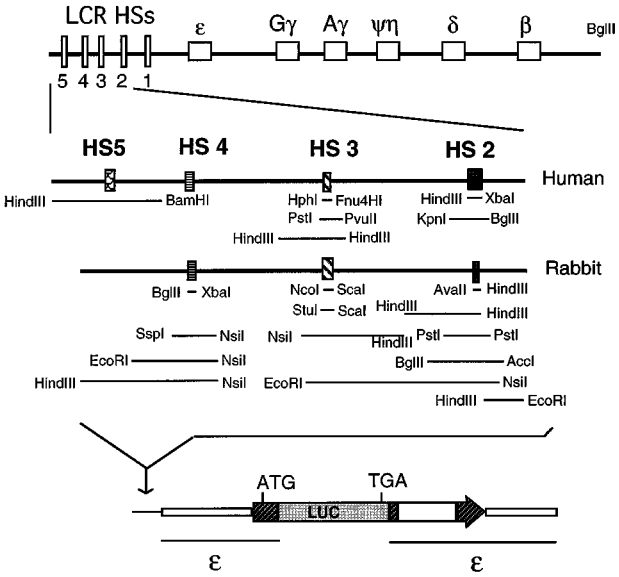

FIG. 1. **Map of LCR fragments and the ϵ-luciferase expression vector.** The *top line* is a schematic of the human β-globin gene cluster. The *next set of lines* shows the locations of the restriction fragments containing segments of the human and rabbit LCR, indicated underneath the thick lines showing the positions of the DNase HSs. Each of these fragments was placed 5′ to the ϵ-globin-luciferase hybrid reporter gene, diagrammed on the *bottom line*.

opening, and perhaps protection from some position effects. However, the use of a selectable marker in the stable transfection assays precludes the observation of strong negative position effects, since clones showing such negative effects would not express the drug resistance gene. Results reported here show that sequences outside the cores of HS2 and HS3 do increase the level of expression in stably integrated constructs. A large DNA fragment containing HS3 has a chromatin domain-opening activity. A striking synergism is seen between rabbit HS2 and HS3 when their natural spacing and LCR sequences between them are preserved, and this effect is dependent on sequences outside the cores. Only the HS5 region showed an appreciable effect on protection from position effects in these assays.

MATERIALS AND METHODS

DNA Constructs—Fragments from both human and rabbit β-globin LCRs were inserted 5′ to the hybrid ϵ-globin-luciferase gene in pBSϵ-luc (Fig. 1). This plasmid carries a *Pst*I fragment containing the rabbit ϵ-globin gene in which intron 1 through most of exon 2 are replaced by the coding region of luciferase. This expression vector contains 573 bp of 5′-flanking sequences, much of the internal sequences and 504 bp of 3′-flanking sequences. This ϵ-globin gene fragment is larger than one used in other studies (23) and contains both positive and negative *cis*-acting sequences (30–32). Four different versions of pBSϵ-luc have been used in this work. The first, pBSϵ-luc.1 (11), retained multiple cloning sites between exon 1 of the ϵ-globin gene and the luciferase coding region, which hampered the construction of new derivatives. Those cloning sites were removed and replaced with an *Nhe*I linker in the next generation of plasmids, pBSϵ-luc.2 and pBSϵ-luc.3. These plasmids differ only in the orientation of the ϵ-luciferase genes, with pBluescript (Stratagene) cloning sites *Sac*I through *Pst*I located 5′ to the gene in pBSϵ-luc.2 and *Kpn*I through *Pst*I located 5′ to the gene in pBSϵ-luc.3. Translation of the hybrid ϵ-luciferase mRNA from pBSϵ-luc.1 requires internal reinitiation at the start codon of luciferase, because a translational terminator is located in-frame at the 5′ junction between the ϵ-globin and luciferase genes. This termination codon is removed from pBSϵ-luc.2 and pBSϵ-luc.3, and thus mRNA from these constructs should be translated to make a hybrid ϵ-luciferase protein. The vector pBSϵ-luc.4 has almost all the pBluescript cloning sites 5′ to the ϵ-luciferase gene, but the 5′ flank is slightly truncated to 544 bp (begins at a *Nsi*I site) and the 3′ flank has 344 bp (ends at a *Bcl*I site). In all cases, the amount of activity from a given plasmid containing a fragment of the LCR was compared to the activity from the appropriate parental expression vector, pBSϵ-luc.1, -2, -3, or -4. The amount of luciferase activity in transfected cells differs very little among these different plasmids.

A map of the LCR indicating the positions of the fragments tested is shown in Fig. 1. Table I provides details about the restriction endonuclease cleavage sites used to add the LCR fragments to the expression vectors, along with an indication of the orientation of the insert in the expression plasmid. The sequenced regions of the rabbit β-globin LCR are very similar to those of the human (11), and Table I lists both the rabbit DNA positions and the aligning human sequence in GenBank™ file HUMHBB. The alignment allows one to refer to the LCR positions in terms of the human DNA sequence, even for DNA segments from other mammals (33).

Transfection of K562 Cells and Measurement of Expression—For tests of transient expression (prior to integration into the chromosome), 7×10^6 K562 cells were electroporated with 50 μg of total DNA that included increasing amounts (1.5–10 μg) of test DNA, 10 μg of pRSV-*lac*Z, and 30–38.5 μg of pBluescript as a carrier plasmid. The plasmid pRSV*lac*Z contains the *Escherichia coli* β-galactosidase gene driven by an RSV long terminal repeat and serves as a control for transfection efficiency (34). Each assay consisted of three independent transfections using a given DNA concentration of a particular construct, and most constructs were tested at a series of DNA concentrations. These triplicate assays were repeated multiple times, and the number of times they were repeated (n) is reported in the appropriate figures and tables. Transfected cells were cultured for 48 h in Dulbecco's modified Eagle's medium plus 10% bovine calf serum, 2% penicillin/streptomycin, and 0.5 μg/ml fungizone. Cells were harvested and lysed in 100 μl of 25 mM Tris-phosphate, pH 7.8, 2 mM dithiothreitol, 2 mM 1,2-diaminocyclohexane-*N*,*N*,*N*′,*N*′-tetraacetic acid, 1.1 mM glycerol, 16 mM Triton X-100, and 1 mg/ml bovine serum albumin (from Promega Biotech). A 10 μl sample of cell extract was added to 100 μl of a reagent (from Promega) containing 20 mM Tricine, 1 mM magnesium carbonate, 2.7 mM magnesium sulfate, 0.1 mM EDTA, 33 mM dithiothreitol, 270 μM coenzyme A, 470 μM luciferin, and 530 μM ATP at room temperature and assayed for luciferase activity for 10 s in a Berthold Lumat LB9501 luminometer. Using this reagent, the light output from the reaction is steady for at least 20 s. In addition, a 20-μl aliquot of the extract was assayed for β-galactosidase activity (35), and the luciferase activity (in relative light units per s) was divided by the A_{420} of the β-galactosidase reaction product, *o*-nitrophenol, to obtain a luciferase/β-galactosidase ratio. Results are reported as the arithmetic mean, with standard deviation and number of replicate assays.

To produce individual stably expressing clones of K562 cells, 1×10^7 cells were electroporated with 100 μg of total DNA containing 90 μg of linearized test plasmid and 10 μg of linearized pM5neo as a selectable marker. The plasmid pM5neo contains the Tn5 phosphotransferase gene driven by the long terminal repeat of the myeloproliferative sarcoma virus, thereby conferring resistance to the drug G418 in hematopoietic cells (36). G418 (final concentration of 1.2 mg/ml) was added 24 h after electroporation to begin drug selection. Individual stably expressing clones were plucked from soft agar 2 weeks after electroporation and grown another 4–6 weeks until there were enough cells to prepare both cell extract and genomic DNA. Transgene expression was determined using luciferase assays as described above. Transgene copy number was determined by hybridizing a luciferase DNA probe to Southern blots containing 10 μg of genomic DNA from each clone digested with *Pst*I plus *Eco*RV. Copy number standards containing the equivalent of 1, 10, and 100 copies of pBSϵ-luc in 10 μg of λ-DNA were included on each blot. For each clone, the expression per copy of the transfecting gene was calculated by dividing the luciferase activity by the number of copies of the ϵ-luciferase construct in the transfected cells. The test constructs are subject to strong positive position effects after integration, occasionally resulting in a small number of individual clones producing much more luciferase than do the rest of the clones. In order to reduce the impact of these outliers on the summary value for the group of clones, the geometric mean is reported in the appropriate tables and figures.

The luciferase activity is a reliable indicator of the amount of RNA produced from these reporter constructs. The amount of ϵ-luciferase RNA was measured by an RNase protection assay (37) in several of the stably transfected clones whose level of expression of luciferase ranged from 4000 to over 1×10^6 relative light units per s. A plot of the amount of RNA *versus* the luciferase activity shows a strong, linear correlation between the amount of RNA and the luciferase activity over a this very wide dynamic range. The slope of the line is 0.66, showing that the

TABLE I
Constructs tested in transfected cells

Corresponding positions between rabbit and human were determined by sequence alignment. Human positions in parentheses correspond to the nearest end point of an aligning segment in rabbit (due to the restriction endonuclease site being in a repetitive element in rabbit), and those in brackets are extrapolations from the alignment. Rabbit positions for HS3 and HS2 are from GenBank™ file RABGLOBCON (accession number L05833) and those for HS4 are from RABGLOBHSB (L05835). NA, sequence not available.

Restriction fragment	Orientation(s) tested	Nucleotide positions in RAB		Corresponding positions in HUMHBB		Version of pBSε-luc.*n*
rHS2 + HS3 5.6EN	Native	1	5558	[3300]	(9066)	3
rHS2 2.2HH	Reverse	2861	5084	6558	8750	1
rHS2 2.2BAc	Native	3798	6072	(7239)	9406	1
rHS2 1.0PP	Reverse	4276	5313	7626	9016	1
rHS2 0.1AH	Native	4968	5084	8642	8750	1, 3
hHS2 1.5KB	Native	4380	5890	7764	9218	4
hHS2 0.4HX	Native/reverse	4817	5191	8486	8860	1, 4
rHS3 2.9NH	Native/reverse	−268	2861	3079	6558	3, 4
rHS3 0.5StSc	Native/reverse	1158	1608	4489	4993	1, 3, 4
rHS3 0.4NcSc	Native	1201	1608	4532	4993	4
hHS3 1.9HH	Native/reverse	−79	1786	3266	5172	2, 4
hHS3 0.8PPv	Native/reverse	1015	1765	4344	5122	1, 4
hHS3 0.2HpF	Native/reverse	1218	1394	4550	4772	2, 4
rHS4 3.9HN	Native	−2700	1166	[−2170]	1782	2
rHS4 2.4EN	Native	−1400	1166	[−870]	1782	3
rHS4 1.2SsN	Native	−3	1166	[530]	1782	3
rHS4 0.2BX	Native/reverse	524	744	997	1211	1
hHS5 3.0HBa	Native	NA	NA	−2686	306	3
3′ to rHS2 1.5HE	Reverse	5084	6590	8750	(9641)	1

increase in luciferase activity can exceed that in the amount of RNA only by about 2-fold.

RESULTS

Enhancement of Transient Expression by Individual LCR HSs—The role of individual HSs from the β-globin LCR was initially addressed by placing short restriction fragments containing each HS core, or a subfragment of it, upstream of a hybrid ε-globin-luciferase reporter gene and assaying for transient luciferase activity in transfected K562 cells (prior to integration of the construct). The greatest effect on ε-luciferase expression was from those constructs containing HS2. As shown in Fig. 2, a rabbit 110-bp *Ava*II to *Hin*dIII fragment containing a highly conserved subregion of the HS2 core, corresponding to positions 8642 to 8750 in the human sequence, enhances ε-luciferase expression 17–30-fold at each DNA concentration tested. Similarly, the human HS2 core (the *Hin*dIII to *Xba*I fragment corresponding to positions 8486 to 8860 in the human sequence) enhanced ε-luciferase activity 13-fold (Fig. 3). This fragment was equally effective in the reverse genomic orientation (data not shown). The effects of conserved regions outside the HS cores were examined by testing larger genomic fragments. As summarized in Fig. 3, two different rabbit DNA fragments containing HS2 (with varying amounts of 5′ or 3′ adjacent sequences) and a human 1.5-kb *Kpn*I to *Bgl*II fragment gave about the same level of enhancement as did the HS2 core fragments. Thus, sequences flanking the HS2 core had no effect on enhancement in transient expression.

Some restriction fragments encompassing the HS3 core also enhanced transient ε-luciferase expression, although not as much as HS2. As shown previously (11), a rabbit 504-bp *Stu*I to *Sca*I fragment (corresponding to positions 4489 to 4993 in the human sequence) increased the expression of ε-luciferase an average of 6-fold at all DNA concentrations tested (Fig. 2). Since others have not observed an effect on transient expression by human or mouse fragments containing HS3 (22, 25), we examined this effect in more detail. Indeed, the data in Fig. 4 show that the human 225-bp *Hph*I to *Fnu*4HI that has been defined as the HS3 core (16) does not enhance expression in either orientation, nor does a larger 1.9-kb *Hin*dIII fragment, in agreement with previous results (22). However, the human

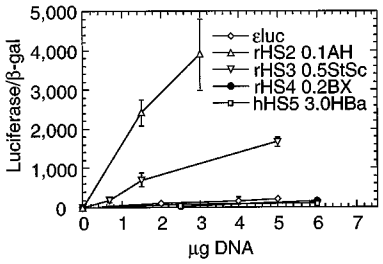

FIG. 2. **Effect of rabbit individual HS sites on transient expression of the ε-luciferase gene in K562 cells.** Increasing amounts of the test plasmid were transfected in triplicate into K562 cells, along with the plasmid pRSV*lacZ* as a control for transfection efficiency. Luciferase activity was normalized to the β-galactosidase activity, and the averages (± S.D.) are plotted as a function of amount of transfecting DNA. The *diamonds* are values for the parental reporter ε-luciferase (ε*luc*), *triangles* are for rabbit 0.1-kb *Ava*II to *Hin*dIII HS2-ε-luciferase, *inverted triangles* are for rabbit 0.5-kb *Stu*I to *Sca*I HS3-ε-luciferase, *filled circles* are for rabbit 0.2-kb *Bgl*II to *Xba*I HS4-ε-luciferase, and *squares* are for human 3.0 kb *Hin*dIII to *Bam*HI HS5-ε-luciferase.

0.8-kb *Pst*I to *Pvu*II fragment encompassing the HS3 core does produce a modest enhancement, similar to that of the rabbit *Stu*I to *Sca*I fragment. Additionally, a large 2.9-kb *Nsi*I to *Hin*dIII fragment from rabbit will also enhance expression in transiently transfected cells. For all three fragments that show enhancement in this assay, the effect is greater in the reverse genomic orientation. Although most DNA fragments showed similar effects in the different versions of the ε-luciferase reporter construct in various assays, these fragments containing HS3 placed into pBSε-luciferase.4 showed enhancement only at lower concentrations of DNA (≤5 μg/transfection), whereas the fragments in pBSε-luciferase.1 showed enhancement at higher DNA concentrations as well (11).

As shown on the map in Fig. 4, the rabbit and human DNA fragments that enhance expression contain a well-conserved sequence that matches the binding site for AP1, NFE2, and related proteins, whereas the 225-bp HS3 core *Hph*I to *Fnu*4HI fragment does not contain this site. This site was removed from the rabbit fragment containing HS3 by cutting with *Nco*I, and the resulting 400-bp *Nco*I to *Sca*I fragment no longer showed enhancement in transiently transfected K562 cells (Fig. 4),

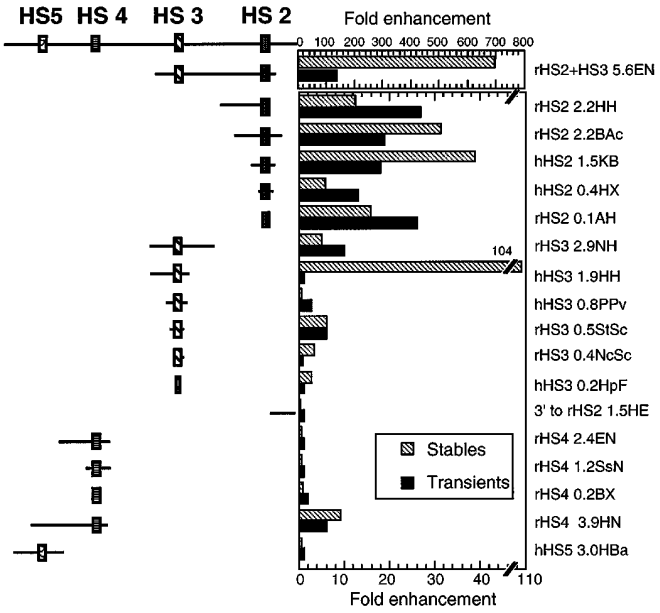

FIG. 3. **Summary of the effects of LCR fragments on the transient expression of the ϵ-luciferase gene in K562 cells.** The *bar graph* shows average fold enhancement over expression of the parental ϵ-luciferase. Average fold enhancement values for transiently transfected cells (arithmetic means) were obtained from multiple determinations at more than one concentration of test DNA. The average fold enhancement for sets of clones of stably transfected cells (geometric means) are derived from the data in Figs. 5 and 6. Each LCR fragment was placed 5′ to ϵ-luciferase in either the reverse or native genomic orientation and tested for level of expression. The effects of the 5.6-kb *Eco*RI to *Nsi*I fragment containing both HS3 and HS2 was much greater than the effects of other sites, and hence it is plotted on a scale different from the others. Restriction sites defining each fragment are as follows: *A* = *Ava*II, *Ac* = *Acc*I, *B* = *Bgl*II, *Ba* = *Bam*HI, *E* = *Eco*RI, *F* = *Fnu*4HI, *H* = *Hin*dIII, *Hp* = *Hph*I, *K* = *Kpn*I, *N* = *Nsi*I, *Nc* = *Nco*I, *P* = *Pst*I, *Pv* = *Pvu*II, *Sc* = *Sca*I, *Ss* = *Ssp*I, *St* = *Stu*I, and *X* = *Xba*I. *r* = rabbit, *h* = human.

supporting a possible role for AP1-like proteins in this ability to enhance expression transiently. The inability of the human 1.9-kb *Hin*dIII fragment to enhance in this assay suggests that a negative element may be located 5′ to the *Pst*I site at position 4344.

Tests of rabbit and human DNA fragments containing HS4 and HS5 showed no effect on ϵ-luciferase expression in these assays (Fig. 2). Use of larger DNA fragments up to 2.4 kb containing HS4 had no effect on enhancement (Fig. 3). In contrast, a *Hin*dIII to *Nsi*I fragment from the rabbit genome enhanced ϵ-luciferase expression 6-fold (Fig. 3). This fragment contains HS4 and was originally thought to contain the homolog to human HS5, based on hybridization data and the expected distance between these hypersensitive sites in rabbit. However, analysis of recently determined DNA sequences[2] shows that this region of the rabbit DNA does not match with human HS5. Thus, the small increase in expression achieved with this DNA fragment results from the inclusion of distal sequences that are not homologous to human.

Synergism by Combinations of HSs for Enhancement of Transient Expression—A rabbit 5.6-kb fragment containing both HS2 and HS3 was placed in the native genomic orientation 5′ to ϵ-luciferase; use of this genomic fragment preserves the DNA sequence and spacing between these HS cores. This fragment corresponds approximately to positions 3300 to 9066 in the human sequence, based on extrapolations from the aligning segments. As shown in Fig. 3, this fragment caused an average of 140-fold enhancement at all DNA concentrations tested. This is much larger than the effect seen with either HS2 or HS3 alone, and it indicates that HS2 and HS3 together (in their natural context) have a synergistic effect on ϵ-globin gene expression.

Enhancement by Individual HSs after Integration into the Genome—The effects of individual HSs on ϵ-luciferase expres-

[2] J. Slightom, personal communication.

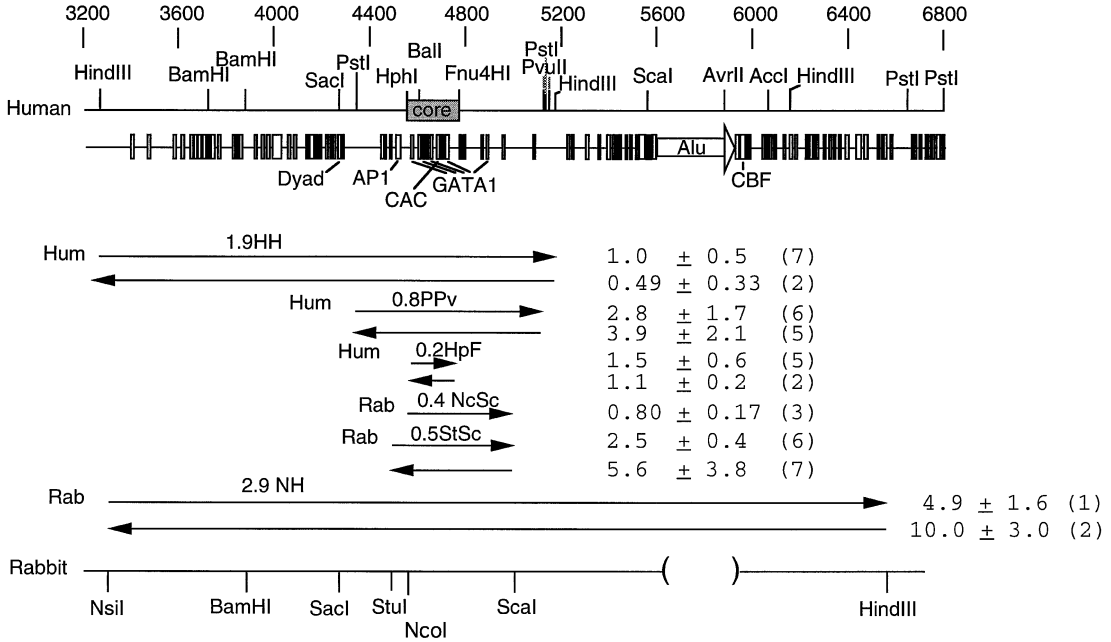

FIG. 4. **Finer dissection of the effect of restriction fragments containing HS3 on transient expression of the ϵ-luciferase in K562 cells.** Restriction maps are shown on the *top* and *bottom lines*. The *second line* shows conserved blocks, which are strings of six or more positions in the multiple alignment of sequences from human, galago, rabbit, goat, and mouse with no more than one mismatch per position (33), and an *Alu* repeat in human. Those conserved blocks that are known (GATA1 and AP1 or a relative) or proposed to be binding sites for proteins are indicated below the second line; *CBF* = CCAAT binding factor, *CAC* = proteins that bind to a CAC motif, such as Sp1, TEF2, and relatives, *Dyad* = a prominent conserved dyad. Some conserved blocks are so short on this scale that they appear as *thin filled boxes*. The positions and orientations of fragments tested are indicated along with the mean fold enhancement, ± S.D. (range for two experiments), with the number of experiments (most of which were triplicate determinations) in *parentheses*.

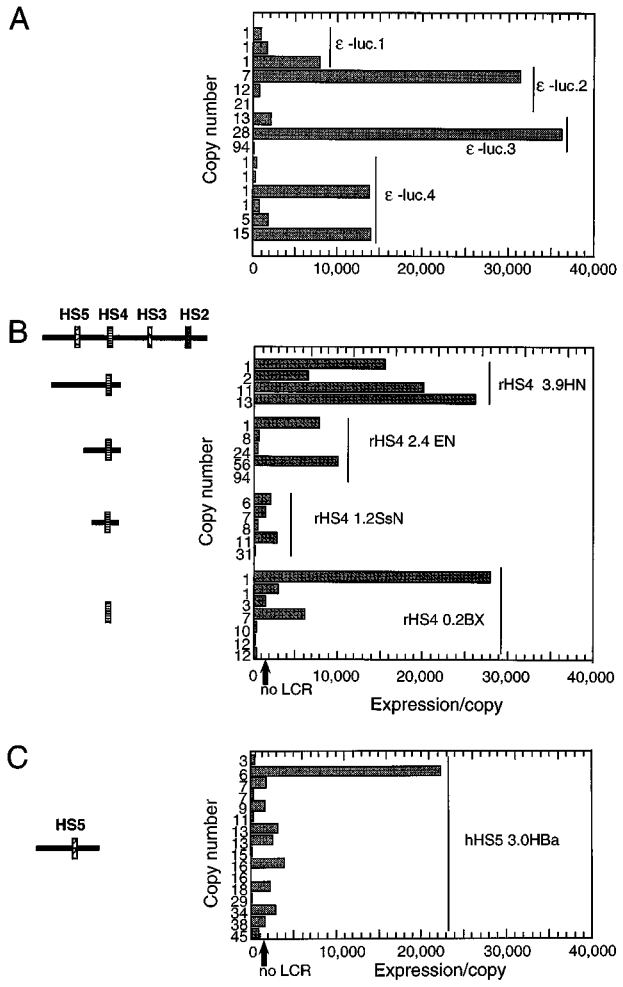

FIG. 5. **Expression of ϵ-luciferase in sets of stably expressing K562 clones and effects of fragments containing HS4 and HS5.** Individual G418-resistant clones were assayed for both luciferase activity and transgene copy number, and the luciferase activity per gene copy is plotted along with the gene copy number. Each set of clones is labeled according to the test construct used, along with a schematic map of the LCR and relevant fragments. *A*, 15 clones expressing the parental ϵ-luciferase, with clones from each of the four versions of the reporter construct. *B*, sets of clones with ϵ-luciferase linked to four different DNA fragments containing HS4. *C*, set of clones with ϵ-luciferase linked to a fragment containing the human HS5.

sion after integration were tested in stable transfection experiments. Figs. 5 and 6 show the results obtained from several stably expressing K562 clones transfected with the same constructs analyzed above. In all cases, the measured ϵ-luciferase activity was divided by the number of integrated copies of the construct to obtain the expression per copy of transfecting gene. The level of expression of ϵ-luciferase in the absence of an LCR fragment was variable, with 11 of 15 clones having very low expression but four showing appreciable expression per copy (Fig. 5*A*). These higher expressing clones likely reflect strong positive position effects. The average expression per copy (geometric mean) for all 15 clones (1650 relative light units s^{-1} $copy^{-1}$) was used as the base line for comparison with the activities of constructs containing LCR fragments.

The presence of fragments containing HS2 caused a substantial increase in the expression per copy of the transfecting gene (Fig. 6*A*; note the difference of scale from the plot in Fig. 5*A*). However, the level varied considerably among individual clones within each set, regardless of the size of the tested LCR fragment, showing that these constructs are still subject to some position effects. The largest increase in expression was obtained with the 1.5-kb *Kpn*I to *Bgl*II fragment from human; it was 2.5 to 7 times as effective as fragments containing only HS2 core sequences, *i.e.* the rabbit 0.1-kb *Ava*II to *Hin*dIII fragment and the human 0.4-kb *Hin*dIII to *Xba*I HS2 core fragment (summarized in Fig. 3). Likewise, the rabbit 2.2-kb *Bgl*II to *Acc*I fragment also produced a larger increase in expression than did the core fragments after integration. Although less of an effect was obtained with the rabbit 2.2-kb *Hin*dIII fragment, all three larger DNA fragments containing HS2 gave stronger enhancement than that obtained with the human HS2 core (0.4-kb *Hin*dIII to *Xba*I fragment). These data indicate that sequences adjacent to the core play a positive role after integration, such as in domain opening.

Several DNA fragments containing HS3 also produced an increase in expression in stably transfected cells, when compared to constructs lacking LCR fragments. Two of the tested fragments, the 0.5-kb *Stu*I to *Sca*I fragment and the 2.9-kb *Nsi*I to *Hin*dIII fragment from rabbit (corresponding to human positions 3079 to 6558) gave an increase in expression (Fig. 6*B*) comparable to the values obtained in transient transfections (Fig. 3). In contrast, the human HS3 core fragment showed a 4-fold enhancement in stably transfected cells, whereas it had no effect in transient transfections (Fig. 3). The larger 1.9-kb *Hin*dIII fragment containing human HS3 showed a dramatic increase in expression in virtually all clones (Fig. 6*B*), averaging over 100-fold, whereas it had no effect in transiently transfected cells (Fig. 3). The capacity of large fragments containing HS3 to greatly increase expression only after stable integration argues for an effect on domain opening when sequences flanking the core are present.

Most small or large restriction fragments containing HS4 had no effect on ϵ-luciferase expression in stably transfected clones; as seen in Fig. 5*B*, the profile of expression per copy is very similar to that of the parental vectors without an LCR fragment. However, the 3.9-kb *Nsi*I to *Hin*dIII fragment from rabbit gave an average 9-fold increase in expression.

Partial Insulation by HS5 after Integration into the Genome—The average level of expression from the set of clones transfected with hHS5-ϵ-luciferase was no greater than that for ϵ-luciferase (Fig. 5*C*). However, the variability in expression per copy from clone to clone was substantially decreased, with most of the clones showing a consistently low level of luciferase activity. This suggests that human HS5 may insulate against many positive position effects. One of the HS5 clones (with 6 integrated copies) had a much higher level of expression per copy, showing that HS5 does not fully insulate against all positive position effects.

Strong Synergism by Combinations of HSs after Integration into the Genome—Addition of the rabbit 5.6-kb fragment containing both HS2 and HS3 to ϵ-luciferase produced very high levels of expression in all clones tested (Fig. 6*C*, note the change of scale). The average level of expression was 700-fold greater than that of ϵ-luciferase alone (Fig. 3), which is significantly greater than the increase obtained with any fragments containing either HS2 or HS3 alone. Thus, the large DNA fragment, containing both HS3 and HS2 along with sequences between and flanking them, shows a synergistic effect for HS3 and HS2, both before and after integration (Fig. 3).

This synergism requires sequences outside the hypersensitive site cores. Removal of 1.2 kb from the 5′ end of this fragment dramatically reduces the activity seen in clones of stably transfected cells, as shown for the 4.4-kb *Nco*I to *Nsi*I fragment in Fig. 6*C*. These clones average a 34-fold increase over the expression of ϵ-luciferase alone, a level comparable to that seen with for large fragments containing only HS2 (Fig. 3). Thus, the DNA sequences flanking the HS3 core, including the

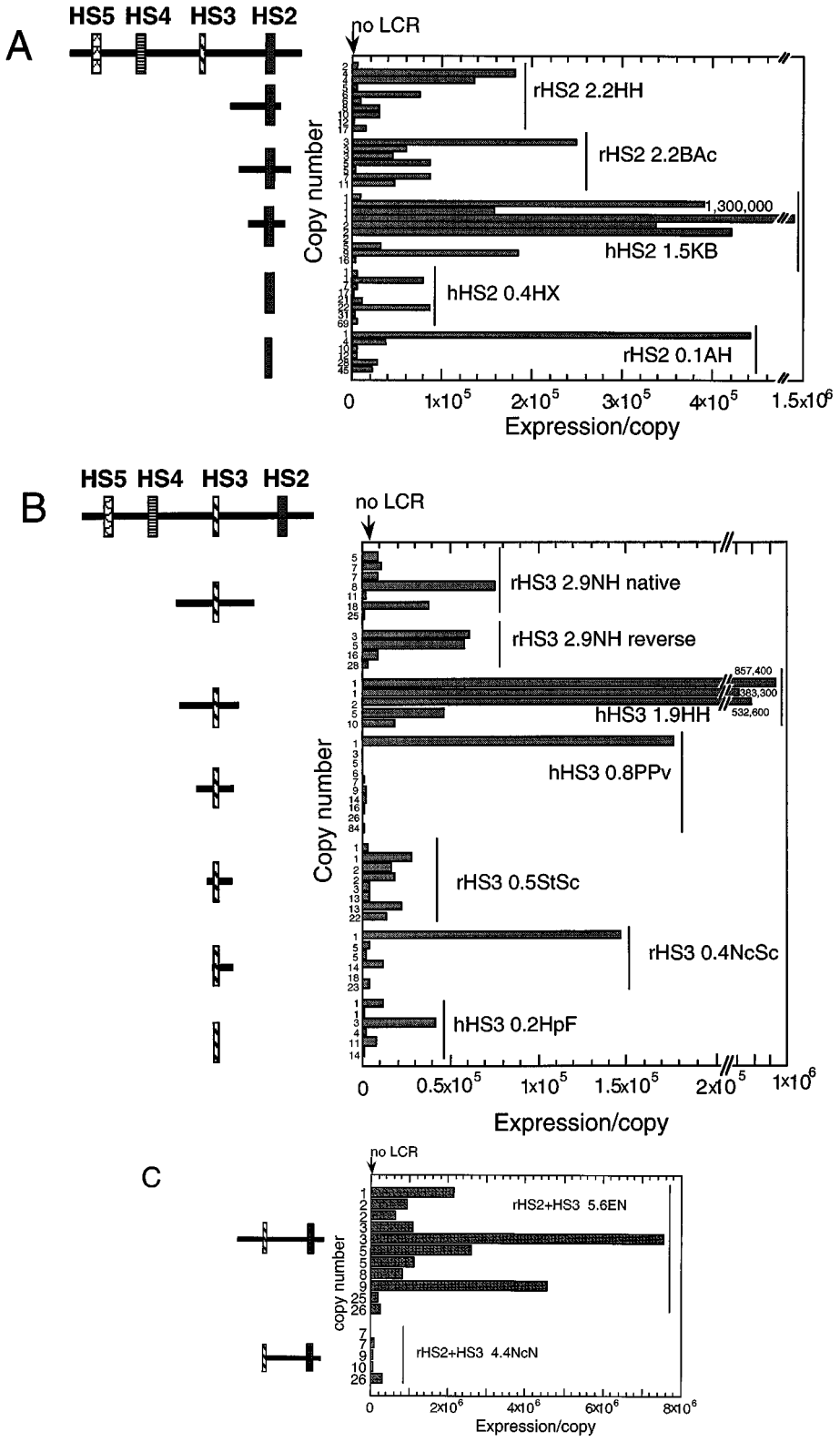

FIG. 6. **Expression of sets of clones of stably transfected K562 cells with ϵ-luciferase linked to fragments containing HS2, HS3, and HS2 plus HS3 together.** *A*, expression per copy for sets of clones with five different fragments containing HS2. *B*, sets of clones with the ϵ-luciferase gene linked to four different fragments containing HS3. *C*, sets of clones with the ϵ-luciferase gene linked to fragments containing both HS3 and HS2.

binding site for AP1-like proteins (Fig. 4), are needed for the synergism observed between rabbit HS2 and HS3.

DISCUSSION

Comparison of the sequences of the β-like globin gene LCRs from various mammals has shown that the cores of the HSs are well conserved (7, 8, 9, 10, 11), and indeed small fragments (200 to 400 bp) that constitute the core of HS2 (12), HS3 (15), and HS4 (17, 18) are active in some assays of LCR function. In addition, the number and distance between the HSs is almost invariant, and many blocks of sequence between the cores of the HSs are conserved. This was shown initially as blocks of sequences that consistently align in multiple pairwise comparisons among LCR sequences from four mammals (11) and is confirmed dramatically by the results of a simultaneous alignment of these sequences from five mammals: human, galago, rabbit, goat, and mouse (19, 31, 33). Even at a very stringent criterion for "conserved," such as strings of seven contiguous

invariant positions in the multiple alignment, conserved blocks of sequence are detected between the cores of the HSs, sometimes at a density equal to that within the cores.[3] These sequence comparisons implicate DNA segments between the HS cores in function of the LCR. However, much experimental work has focused on the HS cores singly or in combination. The data in this paper show that sequences between the HS cores have a strong positive effect on reporter constructs, especially after integration into the chromosome. These effects of flanking sequences are particularly notable for a domain opening activity of fragments containing HS3, for a synergistic interaction between rabbit HS2 and HS3, and for demonstrating a role for a highly conserved AP1 binding site close to HS3.

The very large increase in expression for the human 1.9-kb *Hin*dIII fragment containing HS3 dramatically illustrates the effect of sequences outside the core after integration. With the ϵ-globin luciferase reporter construct in K562 cells, little to no effect was observed with the 225-bp *Hph*I to *Fnu*4HI HS3 core fragment, whereas this core fragment is highly effective with the human β-globin gene as a reporter in transgenic mice and stably transfected MEL cells (15). However, the effect of the minimal HS3 core was reduced substantially when tested with an H2-K gene driven by the β-globin gene promoter (38). Thus, in some assays, the body of the β-globin gene may contribute to the high level of expression obtained with HS3 constructs. In our assays, the additional sequences in the 1.9-kb *Hin*dIII fragment are needed to see the effect of HS3 on the ϵ-globin gene in K562 cells. When assaying β-globin gene expression in transgenic mice, the human 1.9-kb *Hin*dIII fragment gives a higher level of expression than the 225-bp *Hph*I to *Fnu*4HI core fragment (15, 20), providing further evidence that sequences outside the core contribute to the activity of HS3.

This effect of the large 1.9-kb DNA fragment containing HS3 is seen only after integration of the constructs in stably transfected cells. The absence of an effect prior to integration shows that this DNA fragment does not function as a classical enhancer. Also, the expression (adjusted for copy number) varied among clones, arguing against a strong insulator effect. The most reasonable interpretation of these data is that the 1.9-kb fragment containing HS3 has *cis*-acting sequences needed for opening a chromatin domain at the site of integration. A different line of experiments also leads to the assignment of a dominant domain-opening activity for this DNA fragment. Constructs containing this 1.9-kb fragment with HS3 attached to a β-globin gene are active as single copy integrants in transgenic mice (39), whereas constructs with a HS2 fragment are effective only in multiple copies (40). The single copy integrants containing HS3, but not HS2, formed DNase hypersensitive sites in these experiments, demonstrating an opened chromatin domain. Such a domain opening activity has been inferred previously for an artificial construct composed of small DNA fragments containing the cores of HS1, HS2, and HS3 (8), and our experiments help localize at least part of this function to the region surrounding the HS3 core. Similarly, a domain-opening activity, but not enhancement, has been mapped to the region surrounding the HS2 core (13, 41).

The strongest effects on expression of the ϵ-luciferase reporter gene were obtained with a 5.6-kb rabbit DNA fragment containing both HS3 and HS2 along with all the sequences between them and some DNA flanking HS3 on the 5′ end and HS2 on the 3′ end. Thus, it contains all the conserved sequences from this region of the LCR. This combination of HS2 and HS3 in their natural sequence context and with the native spacing produces a very large increase in expression of the linked reporter gene, far greater than the effects obtained with individual sites. This synergism could reflect interactions between the proteins bound to the two sites, the effects of sequences between the HS cores, and/or the effects of an optimal spacing. The reduction in activity upon deletion of DNA 5′ to the HS3 core shows that flanking sequences are needed for this effect; it is not simply from interactions between proteins bound to the HS cores. Indeed, combinations of human HS2 and HS3 core fragments show no increase over the level of expression of the individual fragments.[4] Additional experiments in progress using other DNAs as spacers indicate that non-globin sequences will not substitute in these constructs,[4] further supporting the involvement of LCR sequences between the cores. In this context, it is interesting to note that constructs with HS2 (1.5-kb fragment) and HS3 (1.9-kb fragment) juxtaposed 5′ to the human β-globin gene had no additional effect, compared to the effects of individual sites, in stably transfected cells (24). These authors report that three HSs together (*e.g.* HS4-HS3-HS2 or HS4-HS3-HS1) were required to generate a substantially higher level of expression than the individual sites, but the DNA fragments used did not include all the sequences between the HSs. The synergistic effects between the pair of sites obtained in the present work suggests that spacing and/or sequences between the cores could be important.

A third example of the effects of sequences outside the HS cores is a well-conserved binding site for members of the AP1 family (such as NFE2). We find that the ability of some DNA fragments containing HS3 to enhance expression prior to integration correlates with the presence of this site, and constructs with this AP1 binding site closer to the target promoter (reverse genomic orientation) tend to show a stronger enhancement. In studies of HS3 effects on human β-globin gene expression after integration in MEL cells, addition of this AP1 binding site had no effect on the ability of the intact human HS3 core fragment to stimulate transcription, but it would increase the activity of subregions of the HS3 core (38). This AP1 binding site is in the region 5′ to the HS3 core that when deleted causes a loss of synergism between HS3 and HS2. Thus, several lines of evidence implicate this site in LCR function, although it is outside the minimal HS3 core (defined by its effect on the human β-globin gene). It should be emphasized that despite the modest enhancement seen in transient transfections, the strongest effect from DNA fragments containing HS3 is seen after integration.

The effects of sequences outside the HS cores and the ability of these cores, in the proper context, to interact synergistically, are consistent with a model for the LCR forming a large holocomplex that interacts alternately with promoters of β-like globin genes (42). Just as enhancers are composed of multiple sites (enhansons) for binding *trans*-activating proteins (43), full LCR function may be achieved by the several HSs together, perhaps in a context-dependent manner. Our data show that the HS3 and HS2 interactions require sequences outside the cores. Recent loss-of-function experiments show that deletion of the HS3 or HS4 core fragments from the human β-globin gene cluster in transgenic mice causes a catastrophic loss of expression of the β-like globin genes at all developmental stages (44), whereas gain-of-function experiments showed that individual HSs were adequate to drive a high level of expression of the target genes. One explanation is that the residual DNAs, outside the HS cores, elicit a dominant-negative phenotype on the

[3] The full multiple alignment can be searched at a variety of criteria for "conserved" at the Globin Gene Server on the World Wide Web at http://globin.cse.psu.edu. Output from the alignment is also available by E-mail to globin@groucho.cse.psu.edu.

[4] J. Jackson and R. Hardison, unpublished data.

rest of the locus (44). Thus, these recent experiments also support a role for sequences outside the cores in the function of the LCR.

Acknowledgments—We thank Dr. J. Slightom for providing a DNA clone with human HS5 and LCR DNA sequences prior to publication, Drs. M. Groudine and F. Grosveld for clones with segments of the human LCR, and Nynke Gillemans for help with clone construction.

REFERENCES

1. Grosveld, F., van Assendelft, G. B., Greaves D., and Kollias, G. (1987) *Cell* **51,** 975–985
2. Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990) *Genes & Dev.* **4,** 1637–1649
3. Townes, T., and Behringer, R. (1990) *Trends Genet.* **6,** 219–223
4. Grosveld, F., Antoniou, M., Berry, M., de Boer, E., Dillon, N., Ellis, J., Fraser, P., Hanscombe, O., Hurst, J., Imam, A., Lindenbaum, M., Philipsen, S., Pruzina, S., Strouboulis, J., Raguz-Bolognesi, S., and Talbot, D. (1993) *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **339,** 183–191
5. Tuan, D., Solomon, W., Li, Q., and London, I. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 6384–6388
6. Forrester, W. C., Thompson, C., Elder, J. T., and Groudine, M. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 1359–1363
7. Li, Q., Zhou, B., Powers, P., Enver, T., and Stamatoyannopoulos, G. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 8207–8211
8. Moon, A. M., and Ley, T. J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 7693–7697
9. Hug, B. A., Moon, A. M., and Ley, T. J. (1992) *Nucleic Acids Res.* **21,** 5771–5778
10. Jimenez, G., Gale, K. B., and Enver, T. (1992) *Nucleic Acids Res.* **20,** 5797–5803
11. Hardison, R., Xu, J., Jackson, J., Mansberger, J., Selifonova, O., Grotch, B., Biesecker, J., Petrykowska, H., and Miller, W. (1993) *Nucleic Acids Res.* **21,** 1265–1272
12. Talbot, D., Philipsen, S., Fraser, P., and Grosveld, F. (1990) *EMBO J.* **9,** 2169–2178
13. 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. U. S. A.* **88,** 1626–1630
14. Liu, D., Chang, J. C., Moi, P., Jiu, W., Kan, Y. W., and Curtin, P. T. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3899–3903
15. Philipsen, S., Talbot, D., Fraser, P., and Grosveld, F. (1990) *EMBO J.* **9,** 2159–2167
16. Philipsen, S., Pruzina, S., and Grosveld, F. (1993) *EMBO J.* **12,** 1077–1085
17. Pruzina, S., Hanscombe, O., Whyatt, D., Grosveld, F., and Philipsen, S. (1991) *Nucleic Acids Res.* **19,** 1413–1419
18. Lowrey, C. H., Bodine, D. M., and Nienhuis, A. W. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 1143–1147
19. Hardison, R., Elnitski, L., Goldstrohm, A., ElSherbini, A., Riemer, C., Schwartz, S., Stojanovic, N., and Miller, W. (1995) in *Molecular Biology of Hemoglobin Switching* (Stamatoyannopoulos, G., ed) pp. 405–426, Intercept Ltd, Andover, Hampshire, UK
20. Fraser, P., Hurst, J., Collis, P., and Grosveld, F. (1990) *Nucleic Acids Res.* **18,** 3503–3508
21. Fraser, P., Pruzina, S., Antoniou, M., and Grosveld, F. (1993) *Genes & Dev.* **7,** 106–113
22. Tuan, D., Abelovich, A., Lee-Oldham, M., and Lee, D. (1987) in *Developmental Control of Globin Gene Expression* (Stamatoyannopoulos, G., and Nienhuis, A. W., eds) pp. 211–220, A. R. Liss, Inc., New York
23. Tuan, D., Solomon, W., London, I., and Lee, D. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2554–2558
24. Collis, P., Antoniou, M., and Grosveld, F. (1990) *EMBO J.* **9,** 233–240
25. Moon, A. M., and Ley, T. J. (1991) *Blood* **77,** 2272–2284
26. Yu, Z., Bock, J., Slightom, J., and Villeponteau, B. (1994) *Gene (Amst,)* **139,** 139–145
27. Li, Q., and Stamatoyannopoulos, G. (1994) *Blood* **84,** 1399–1401
28. Chung, J. H., Whiteley, M., and Felsenfeld, G. (1993) *Cell* **74,** 505–514
29. 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
30. Cao, S. X., Gutman, P. D., Dave, H. P. G., and Schechter, A. N. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 5306–5309
31. Hardison, R., Chao, K.-M., Adamkiewicz, M., Price, D., Jackson, J., Zeigler, T., Stojanovic, N., and Miller, W. (1993) *DNA Sequence* **4,** 163–176
32. Trepicchio, W., Dyer, M., and Baron, M. (1993) *Mol. Cell. Biol.* **13,** 7457–7468
33. Hardison, R., Chao, K.-M., Schwartz, S., Stojanovic, N., Ganetsky, M., and Miller, W. (1994) *Genomics* **21,** 344–353
34. Thompson, J. P., Simkevich, C. P., Holness, M. A., Kang, A. H., and Raghow, R. (1991) *J. Biol. Chem.* **266,** 2549–2556
35. Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp. 352–355, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
36. Laker, C., Stocking, C., Bergholz, U., Hess, N., De Lamarter, J. F., and Ostertag, W. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8458–8462
37. James-Pederson, M., Yost, S., Shewchuk, B., Zeigler, T., Miller, R., and Hardison, R. (1995) *J. Biol. Chem.* **270,** 3965–3973
38. Pruzina, S., Antoniou, M., Hurst, J., Grosveld, F., and Philipsen, S. (1994) *Biochim. Biophys. Acta* **1219,** 351–360
39. Ellis, J., Tan-Un, K., Harper, A., Michalovich, D., Yannoutsos, N., Philipsen, S., and Grosveld, F. (1996) *EMBO J.* **15,** 562–568
40. Ellis, J., Talbot, D., Dillon, N., and Grosveld, F. (1993) *EMBO J.* **12,** 127–134
41. Caterina, J. J., Ciavatta, D. J., Donze, D., Behringer, R. R., and Townes, T. M. (1994) *Nucleic Acids Res.* **22,** 1006–1011
42. Wijgerde, M., Grosveld, F., and Fraser, P. (1995) *Nature* **377,** 209–213
43. Ondek, B., Gross, L., and Herr, W. (1988) *Nature* **333,** 40–45
44. Bungert, J., Dave, U., Lim, K.-C., Kieuw, K. H., Shavit, J. A., Liu, Q., and Engel, J. D. (1995) *Genes & Dev.* **9,** 3083–3096

Nucleic Acids, Protein Synthesis, and Molecular Genetics:

Role of DNA Sequences Outside the Cores of DNase Hypersensitive Sites (HSs) in Functions of the -Globin Locus Control Region: DOMAIN OPENING AND SYNERGISM BETWEEN HS2 AND HS3

John D. Jackson, Hanna Petrykowska, Sjaak Philipsen, Webb Miller and Ross Hardison

J. Biol. Chem. 1996, 271:11871-11878.
doi: 10.1074/jbc.271.20.11871