Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 8207–8211, November 1990
Developmental Biology

β-Globin locus activation regions: Conservation of organization, structure, and function

(regulatory elements/transgenic mice/hemoglobin switching/hypersensitive sites)

QILIANG LI*†, BING ZHOU*, PATRICIA POWERS‡, TARIQ ENVER†, AND GEORGE STAMATOYANNOPOULOS†§

*Shanghai Institute of Biochemistry, Chinese Academy of Sciences, Shanghai 200031, China; ‡Waisman Institute, University of Wisconsin, Madison, WI 53705; and †Division of Medical Genetics, RG-25, University of Washington, Seattle, WA 98195

Communicated by Stanley M. Gartler, July 20, 1990 (received for review June 12, 1990)

ABSTRACT **The human β-globin locus activation region (LAR) comprises four erythroid-specific DNase I hypersensitive sites (I–IV) thought to be largely responsible for activating the β-globin domain and facilitating high-level erythroid-specific globin gene expression. We identified the goat β-globin LAR, determined 10.2 kilobases of its sequence, and demonstrated its function in transgenic mice. The human and goat LARs share 6.5 kilobases of homologous sequences that are as highly conserved as the ε-globin gene promoters. Furthermore, the overall spatial organization of the two LARs has been conserved. These results suggest that the functionally relevant regions of the LAR are large and that in addition to their primary structure, the spatial relationship of the conserved elements is important for LAR function.**

The β-globin locus is regulated by the locus activation region (LAR) or dominant control region, which is placed 6–20 kilobases (kb) upstream of the ε-globin gene (1) and contains four erythroid-specific DNase I hypersensitive sites (HSI to -IV) located at positions −6.1, −10.9, −14.7, and −18 relative to the canonical cap site of the ε-globin gene, respectively (2, 3). Support for the role of this region in locus activation is provided by a group of naturally occurring mutants in man in which deletions of the LAR result in total inactivation of the β-globin genes (1). The dominance of this genetic element is shown by its ability to confer high-level erythroid-specific expression on heterologous globin genes (4, 5) as well as housekeeping (6) and nonerythroid (7) genes.

The delineation of the cis-active elements of the LAR can be approached functionally by the analysis of *in vivo* and *in vitro* mutants. A dissection of the functional contribution of an individual HS has been attempted by linking the sites either individually or in various combinations to a β-globin gene and assessing its expression in transgenic mice (6, 8, 9) or mouse erythroleukemia cells (10, 11). The analyses of the individual sites reveal that the main transcriptional-enhancing activity resides within HSII and -III, some activity resides in HSIV but that HSI is essentially inactive. The precise nature of the sequence elements involved in transcriptional potentiation or the mechanism by which they act remains unknown.

It is, however, widely thought that the LAR is not simply an enhancer in the classical sense but represents a class of genetic regulatory elements, acting at the level of chromatin organization. In the absence of an assay for authentic LAR function, we approached the problem from an evolutionary standpoint. The functionally relevant regions of the LAR are presumably under constant evolutionary selection. By identifying conserved regions in evolutionary diverged species, one can delineate LAR sequences deemed functionally important by natural selection itself. Humans and goats have been diverging for roughly 80–90 million years. In this report we identify and characterize the LAR of the goat β-globin locus¶ and compare it with its human homolog.

MATERIAL AND METHODS

Phage λ clones Cl3 and Cl15 (12) containing the region immediately 5′ to the 5′-most ε-globin gene of the goat were used for subcloning and sequencing, as described (13–15). About 90% of the sequence was determined on both strands and the remainder was sequenced at least twice on the same strand. The sequence data were analyzed using the University of Wisconsin Genetics Computer Group program. For the studies in transgenic mice (16), the construct pGII(0.7)β containing a goat 0.68-kb *Nsi* I fragment (sequence coordinates 5331–6012), corresponding to the human HSII, linked to a 4.8-kb *Bgl* II fragment of the human β-globin gene (GenBank coordinates 60,628–65,610), was used. Preparation of DNA and RNA from transgenic mice was performed as described (17). Transgene copy number was determined by comparative scanning densitometry of Southern blots using human DNA as a standard. RNA (0.1 μg) was used in each RNase protection assay performed as described (16, 18). Data obtained from the scintillation counting of excised protection products were used for calculating the expression level of the human β-globin gene in transgenic mice. The data were corrected (*i*) for the uridine content of the protection products and (*ii*) for gene copy number, assuming the mouse possesses four α-globin genes. Analysis of globin chains was done as described (16, 18).

RESULTS

Identification of the Goat LAR. As in the human, the goat switches from embryonic to fetal to adult globin formation; however, the organization of the globin genes in the goat is considerably more complex, representing a triplication of a basic set of two ε-globin and two β-globin-like genes (12). Two overlapping clones (Cl3 and Cl15) (12) containing the 5′-most region of the goat β-globin cluster (on the 5′ side of the εI-globin gene) were digested with *Eco*RI, *Bam*HI, and *Kpn* I and the fragments obtained were hybridized with human probes covering the region from the second exon of the human ε-globin gene to ≈20 kb upstream (19). Three human plasmids (p5′ε6, p5′ε3, and p5′ε2) containing, respectively, HSI, -II, and -III showed hybridization to the goat fragments. There was no detected hybridization between the plasmid containing human HSIV and the goat clones used because the goat lacks HSIV, the goat HSIV is located

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LAR, locus activation region; HS, hypersensitive site(s); NFE, nuclear factor erythroid.

§To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M37648).

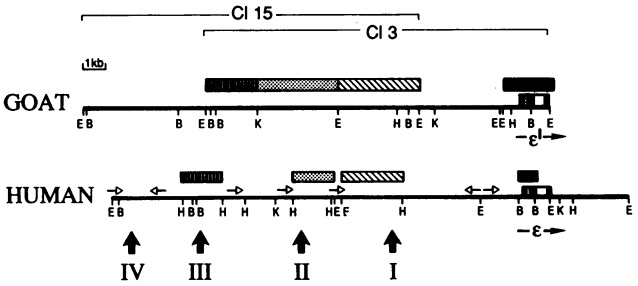

FIG. 1. Identification of the goat LAR. The restriction map at each of the regions is shown. E, *Eco*RI; B, *Bam*HI; K, *Kpn* I; H, *Hin*dIII. The location and transcriptional orientation of the ε-globin genes is shown by a solid horizontal arrow. Regions of homology as detected by cross-hybridization are shown above the restriction maps by boxes filled with the same pattern. The location of four DNase I HS are shown by solid vertical arrows. Small open arrows immediately above the human restriction map represent the *Alu* I repeats found within this sequence. The thin lines above the goat restriction map represent the goat bacteriophage λ clones Cl15 and Cl3.

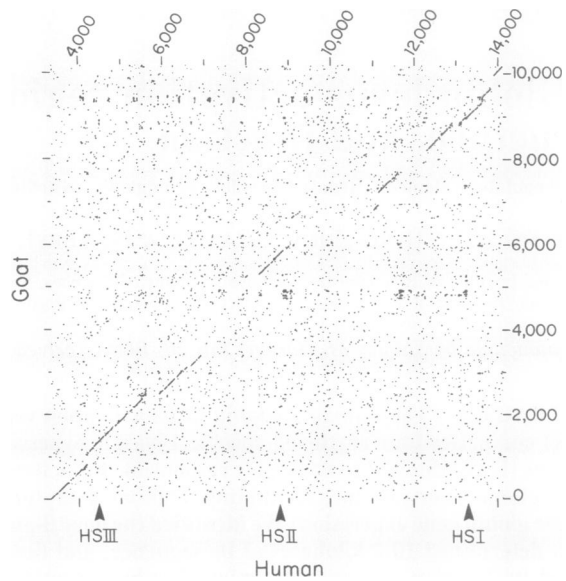

FIG. 2. Comparison of human and goat sequences. Goat and human DNA sequences were compared using the Compare program of the University of Wisconsin Genetics Computer Group. Dots were recorded whenever the two sequences used in the comparison had 14 or more identical nucleotides out of the 20 nucleotides compared. The approximate locations of human HSI, -II, and -III are shown. Human and goat sequence coordinates are as in the GenBank data base (accession nos. J00179 and M37648).

further upstream, or in humans and goats HSIV has significantly diverged. The HSI, -II, and -III-hybridizing fragments were colinear and spanned a region of ≈10 kb in the goat and 11 kb in the human. The extent of homology as detected by these two sets of hybridization is shown in Fig. 1. The sequence of the 10,194 base pairs (bp) of goat DNA hybridizing with the sequences upstream of the human ε-globin gene was determined.

Conservation of Sequence of Human and Goat LARs. The 10.2-kb goat sequence was compared to the 20 kb of DNA sequence from the region 5′ to the human ε-globin gene by using a dot matrix program (Fig. 2). Diagonal lines indicate colinear regions of homology. Interruptions in the diagonal line represent interruptions in the homologous sequence due to insertions, deletions, or local interruption in the conservation of these two sequences. Three large blocks of homology corresponding to human HSI, -II, and -III were detected. In these regions, the goat and human DNA sequences shared ≈68% identity. This level of identity, which is relatively uniform throughout sequences, is greater than that shared by the intervening sequences of the human and goat ε-globin genes (≈40%) and similar to that shared by the human and goat ε-globin gene promoter regions (positions −420 to +50), which are 68% homologous. Complete sequences of the LARs of other species are not available. However, previous comparison of the human β-globin locus to that of mouse (20) or rabbit (21) have revealed high homology of sequences corresponding to the region between HSI and the promoter of the ε-globin gene.

Conservation of Organization. A comparison of the organization of the goat and human sequences is diagrammed in Fig. 3. The blocks of conserved sequence are shown as hatched boxes. Most of the interruptions in the homology regions are due to the insertions of repetitive elements and nine *Nla* elements in the goat sequence and to three *Alu* repeats and one L1 repeat in the human sequences. However, the spatial relationships between the homologous regions in the human and goat LARs are remarkably similar. This conservation of the global organization of the LAR implies that the relative positioning of the conserved elements may be important for LAR function.

Conserved Sequence Motifs in DNase I HS Regions. We compared human and goat sequences to identify sequences that could be implicated to the function of LAR. We examined each HS region for the presence of highly conserved sequences and especially for conserved recognition sites of transcriptional factors.

The most conserved region (98% homology) between human and goat LAR lies near HSII (Fig. 4, positions 5650–5629 in goats and positions 8058–8717 in humans). Several other interesting sequence motifs are conserved in the HSII region. They include a 62-bp sequence that is present twice in the goat sequence (positions 1034–1095 and 5454–5522) and once

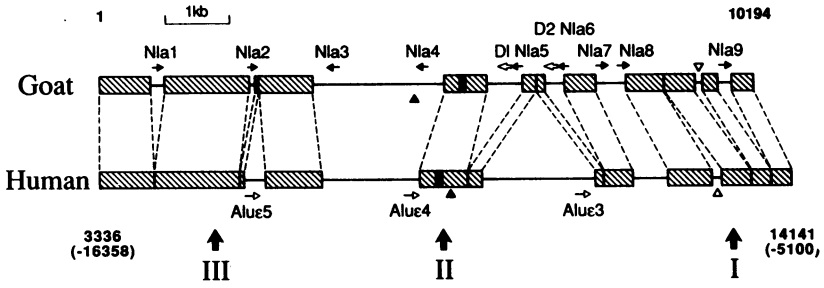

FIG. 3. Structural organization of the goat and human LAR sequences. Hatched boxes represent the conserved regions (conserved 68%) between the goat and the human, as determined by nucleotide sequence comparisons. The solid rectangle indicates the most highly conserved region (98%). Thin lines connecting the hatched boxes represent nonhomologous regions. ➝ , *Alu* repeats; ➝ , *Nla* repeats; ⇦⇨ , D repeats; ▲, alternative A+T-rich region; ▽, 120-bp purine-rich region in the goat that contains many AGAAGG repeats; △, *Kpn* I or L1 repeat in the human. The heavy vertical arrows point the approximate locations of DNase super-HSI, -II, and -III in the human sequence.

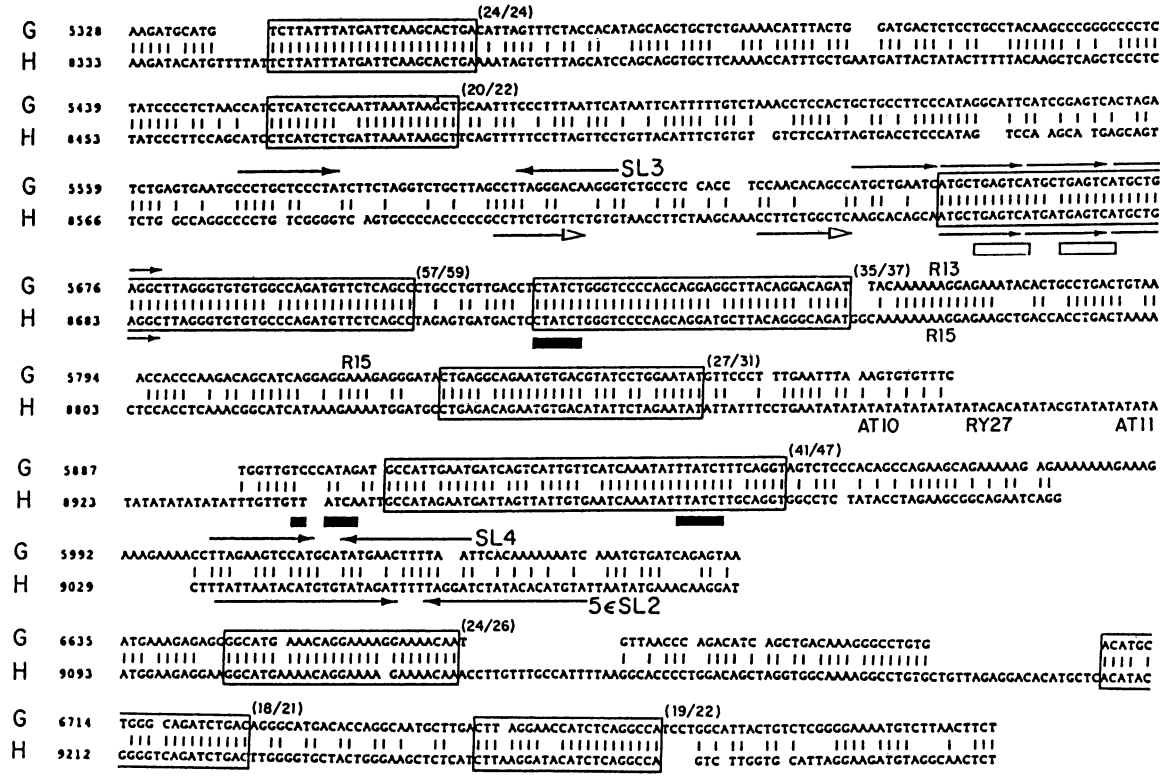

FIG. 4. Alignment of goat (G) and human (H) sequences in the locale of HSII. Regions with >85% homology in the stretches >20 bp are boxed. ■ and □, Potential binding sites in the human sequence for NFE-1 and for NFE-2, respectively; ➙, 10-bp tandem repeats that are present three times in human and four times in goat; ▷, short direct repeats; ➙⬅, potential stem and loop structures. The tracts of polypurine, alternating A-T and alternating purine-pyrimidine are signified by R, AT, and RY with the number after the letters representing the length of the tracks.

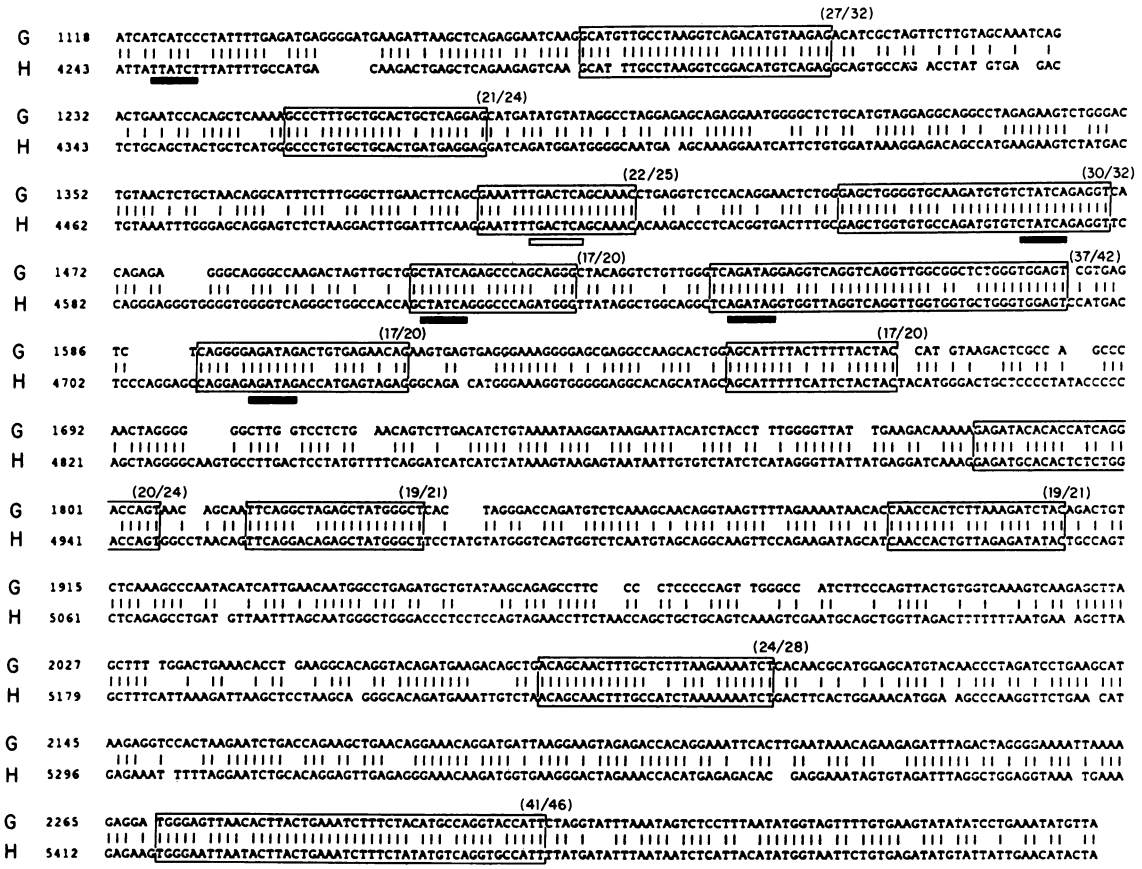

FIG. 5. Alignment of goat (G) and human (H) sequences in the locale of HSIII. Methods of analysis and symbols used are as in Fig. 4.

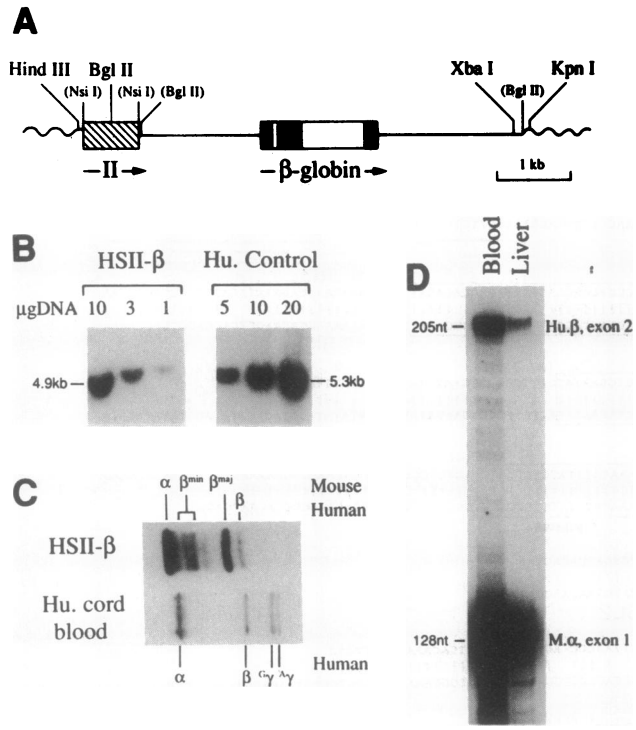

FIG. 6. Functional analysis of the goat LAR in transgenic mice. (*A*) Map of the goat HSII–human β-globin construct. Restriction sites destroyed during cloning are shown in parentheses. The eukaryotic insert used to prepare transgenic mice was excised using the *Hin*dIII and *Kpn* I sites in the polylinker of the plasmid (wavy line) vector. (*B*) DNA blot analysis of HSII–β-globin transgenic mice. The indicated quantities of DNA from the liver of an HSII–β-globin mouse were cleaved with *Bgl* II and *Xba* I (indicated in *A*), blotted, and hybridized using a probe derived from the intervening sequence II region of the human β-globin gene. *Bgl* II-cleaved human (Hu.) lymphocyte DNA was analyzed in parallel as a control. (*C*) Globin chain synthesis. Isoelectric focusing analysis of [^{3}H]leucine-labeled globin chains from the liver erythroblasts of HSII–β-globin transgenic mice. Labeled chains from human (Hu.) cord blood serve as a marker for the migration position of the human β-globin chain. (*D*) RNase protection. RNA (100-ng portions) derived from the liver and blood of an HSII–β-globin transgenic mouse was analyzed by RNase protection using probes specific for the human β- and mouse α-globin mRNAs. nt, Nucleotides; Hu., human; M., mouse.

in the human sequence (positions 8469–8534), purine-rich stretches, and potential stem-and-loop structures.

Comparison of the human and goat sequences in the vicinity of HSIII discloses 12 blocks of >85% homology (from 21 bp to 46 bp long), a nuclear factor erythroid (NFE) 2 site, and seven NFE-1 sites, four of which have been conserved (Fig. 5). The conserved NFE-1 and NFE-2 sites are clustered in a single stretch of DNA, some 180 bp long. This region contains the cis elements responsible for the enhancer function of HSIII (22).

The ATGCTGAGTC Motif in the Core of HSII. The 98% homologous region of HSII contains the consensus sequence ATGCTGAGTC, which is repeated four times in the goat and three times in man (Fig. 4). This decanucleotide also appears once in the enhancer of the human $^A\gamma$-globin gene, once 3′ to the human $^G\gamma$-globin gene, and once between a pair of *Alu* repeats on the 5′ side of the δ-globin gene. Triplication of this decanucleotide in human generates two NFE-2 sites and the quadruplication in goat generates three NFE-2 sites. The NFE-2 sites generated by the ATGCTGAGTC repeat have been shown to bind *jun/fos* and NFE-2 (23) and to confer inducibility on the β-globin gene promoter (24).

Goat HSII Enhances Human β-Globin Gene Expression in Transgenic Mice. To test whether the goat sequences possess

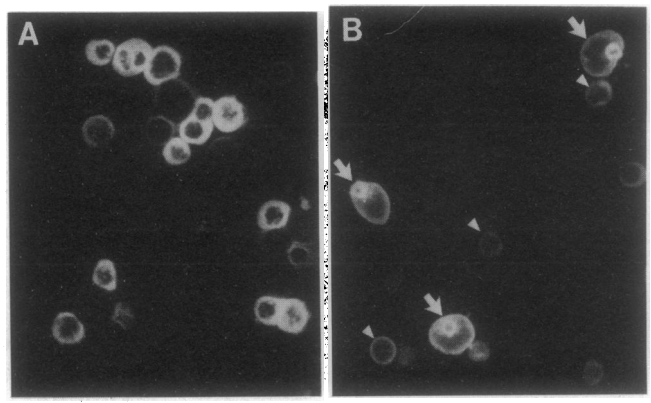

FIG. 7. Goat LAR overides the developmental regulation of the human β-globin gene. Cellular spreads from the d14 liver (*A*) and blood (*B*) of an HSII–β-globin transgenic animal were labeled by indirect immunofluorescence using an anti-β-globin monoclonal antibody. Note the staining of definitive fetal liver erythroblasts (*A*). Also note that in addition to the labeled enucleated definitive erythroid cells of the blood (some examples are indicated by short arrowheads in *B*), there is also labeling of the large nucleated yolk sac-derived embryonic erythroblasts indicated by large arrows (*B*).

LAR function, we made a construct in which a 0.7-kb *Nsi* I fragment from the region of the goat sequence homologous to human HSII was linked to a 4.8-kb *Bgl* II fragment containing the human β-globin gene (Fig. 6*A*). The eukaryotic insert from this construct was purified and used to prepare d15 transgenic fetuses. The copy number and fidelity of integration of the construct was assessed by Southern blot analysis of murine liver DNA by using a human β-globin-specific DNA probe. Three transgenic animals were identified in this manner. Two of them contained correctly integrated copies of the construct. In one of these mice the copy number of the transgene was very low (0.2 copy), the other had a transgene copy number of two (Fig. 6*B*). Human globin gene expression in this latter animal was documented at the protein level by isoelectric focusing of tritiated globin chains (Fig. 6*C*), at the RNA level by RNase protection of liver and blood RNA (Fig. 6*D*), and also at the cellular level by immunofluorescent staining of blood and liver cellular smears with human β-globin-chain-specific monoclonal antibodies (Fig. 7). The level of human β-globin expression in the fetal blood was quantitated by comparison to endogenous mouse α-globin expression and corrected for gene copy number. This analysis yielded human β- to mouse α-globin expression values of 16% at the protein level and 25% at the RNA level. In the absence of a linked LAR, human β-globin transgenes are expressed at 0.1–0.6% of the level of the endogenous mouse globin genes (8). The enhanced level of human β-globin expression we observed as a result of linkage to the goat HSII homologous region is comparable to the levels (13–63%) produced by linkage to a 1.9-kb human HSII-containing fragment (8). This indicates that function as well as the structure of the HSII region is conserved between goats and humans.

Goat HSII Overrides the Developmental Control of β-Globin Gene. Staining of liver and blood-derived cellular preparations with fluorescently conjugated anti-human β-globin-specific monoclonal antibodies revealed that both the adult globin-expressing definitive erythroblasts of the liver (Fig. 7*A*) and embryonic globin-expressing primitive erythroid cells (Fig. 7*B*) derived from the yolk sac contained human β-globin. In the absence of the LAR, human β-globin transgene expression is restricted to definitive erythroid cells (25, 26), whereas human LAR-β-globin constructs are expressed at all stages of murine ontogeny (18, 27). Thus the goat LAR

like the human LAR can override the developmental control of a linked β-globin gene.

DISCUSSION

Our comparison of the human and goat β-globin LARs shows extensive conservation of the sequence over 6.5 kb of DNA. In addition, the organization of the homologous regions is conserved. The implications of these observations are as follows. (*i*) The functionally relevant regions of HSI, -II, and -III are large. (*ii*) All these sites are functionally significant. (*iii*) The relative spacing of the sites plays an important role in LAR function.

Current studies on human LAR function rely on delineating the minimum LAR sequence elements capable of enhancing the transcription of linked reporter genes (6, 8, 10, 11). Thus these studies have shown that HSII and -III are the most active transcriptional potentiators. Although HSIV has some potentiating activity, HSI appears essentially inactive (10). These results raise something of a paradox. If, as implicit from the comparative analyses, the integrity of the LAR is important for function, how can individual HS regions, or for that matter small parts of them, perform so well in the functional assays? This paradox may stem from the apparently bifunctional nature of the LAR. (*i*) It seems to be involved in initiating the formation of active chromatin in the β-globin cluster. (*ii*) It appears to be the major transcriptional enhancer of the β-globin-like genes. The individual sites (or site fragments) may contain cis elements responsible for LAR enhancer function but not for domain activation, which may require the concerted action of several LAR elements.

It has been suggested that the LAR activates the domain through the generation of torsional stress through DNA supercoiling brought about by the loss of nucleosomes at the DNase I HSs and unfolding of the higher-order chromatin solenoid (28): i.e., the larger the nucleosome-free region, the more free supercoils, the greater the torsional stress, and thus presumably the greater the range of the LAR effect. The requirement *in vivo* for the LAR to maintain an active chromatin domain spanning a range of at least 80 kb may underlie the apparent size of its conserved elements. The nonhypersensitive regions located between the sites in both the goat and the human may fulfill some spatial constraint necessary for the efficient execution of these processes.

Whatever the mechanism of LAR action, our results show that it possesses a size and complexity of organization that distinguishes it from other regulatory elements described to date.

We are grateful to Drs. J. Lingrell and S. Shapiro for providing the λ clones, to Dr. Th. Papayannopoulou for help with monoclonal antibodies, to Dr. R. Perlmutter and the Transgenic Mouse Facility of the University of Washington for producing transgenic mice, and to Drs. O. Smithies, R. Patient, N. Raich, and A. Canin for useful discussions. This work was supported by grants from the National Science Foundation of China (3861011) and the National Institutes of Health (DK 31232 and HL 20899).

1. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. (1987) *Cell* **51**, 975–985.
2. Tuan, D., Solomon, W., Li, Q. & London, I. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6384–6388.
3. Forrester, W. C., Thompson, C., Elder, J. T. & Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1359–1363.
4. Ryan, T. M., Behringer, R. R., Townes, T. M., Palmiter, R. D. & Brinster, R. L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 37–41.
5. Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L. & Greaves, D. (1989) *Genes Dev.* **3**, 1572–1581.
6. Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. & Greaves, D. (1989) *Nature (London)* **338**, 352–355.
7. van Assendelft, G. B., Hanscombe, O., Grosveld, F. & Greaves, D. R. (1989) *Cell* **56**, 969–977.
8. Ryan, T. M., Behringer, R. R., Martin, N. C., Townes, T. M., Palmiter, R. D. & Brinster, R. L. (1989) *Genes Dev.* **3**, 314–323.
9. Curtin, P. T., Lin, D., Lin, W., Chang, J. C. & Kan, Y. W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7082–7086.
10. Collis, P., Antoniou, M. & Grosveld, F. (1990) *EMBO J.* **9**, 233–240.
11. Forrester, W. C., Novak, U., Gelinas, R. & Groudine, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5439–5443.
12. Townes, T. M., Fitzgerald, M. C. & Lingrel, J. B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6589–6593.
13. Li, Q. & Wu, G. (1987) *Gene* **56**, 245–252.
14. Zhou, B., Chen, N. & Li, Q. (1988) *Gene* **70**, 405–409.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Enver, T., Ebens, A. J., Forrester, W. C. & Stamatoyannopoulos, G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7033–7037.
17. Karlinsey, J., Stamatoyannopoulos, G. & Enver, T. (1989) *Anal. Biochem.* **180**, 303–306.
18. Enver, T., Raich, N., Ebens, A. J., Papayannopoulou, Th., Costantani, F. & Stamatoyannopoulos, G. (1990) *Nature (London)* **344**, 309–313.
19. Li, Q., Powers, P. A. & Smithies, O. C. (1985) *J. Biol. Chem.* **260**, 14901–14910.
20. Shehee, W. R., Loeb, D. D., Adey, N. B., Burton, F. H., Casavant, N. C., Cole, P., Davies, C. J., McGraw, R. A., Schichman, S. A., Severynse, D. M., Voliva, C. F., Weyter, F. W., Wisely, G. B., Edgell, M. H. & Hutchison, C. A. (1989) *J. Mol. Biol.* **205**, 41–62.
21. Margot, J. B., Demers, G. W. & Hardison, R. C. (1989) *J. Mol. Biol.* **205**, 15–40.
22. Philipsen, S., Talbot, D., Fraser, P. & Grosveld, F. (1990) *EMBO J.* **9**, 2159–2167.
23. Talbot, D., Philipsen, S., Fraser, P. & Grosveld, F. (1990) *EMBO J.* **9**, 2169–2178.
24. Ney, P. A., Sorrentino, B. P., McDonagh, K. T. & Nienhuis, A. W. (1990) *Genes Dev.* **4**, 993–1006.
25. Magram, J., Chada, K. & Costantini, F. (1985) *Nature (London)* **315**, 338–340.
26. Kollias, G., Wrighton, N., Hurst, J. & Grosveld, F. (1986) *Cell* **46**, 89–94.
27. Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L. & Townes, T. M. (1990) *Genes Dev.* **4**, 380–389.
28. Patient, R. K. & Allan, J. (1989) *J. Curr. Opin. Cell Biol.* **1**, 454–459.