Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 3177–3181, April 1993
Genetics

Lethal thalassemia after insertional disruption of the mouse major adult β-globin gene

(homologous recombination/gene targeting/locus control region)

W. RONALD SHEHEE, PAULA OLIVER, AND OLIVER SMITHIES*

Department of Pathology, University of North Carolina, 701 Brinkhous-Bullitt CB7525, Chapel Hill, NC 27599-7525

Contributed by Oliver Smithies, December 22, 1992

ABSTRACT **Thalassemias are hereditary anemias caused by mutations that disturb the normal 1:1 balance of α- and β-globin chains that form hemoglobin. We have disrupted the major adult β-globin gene (*b1*) in mouse embryonic stem cells by using homologous recombination to insert selectable sequences into the gene. Mice homozygous for this insertional disruption of the *b1* gene ($Hbb^{th\text{-}2}/Hbb^{th\text{-}2}$) are severely anemic and die perinatally. In contrast, ≈60% of mice homozygous for deletion of the same gene ($Hbb^{th\text{-}1}/Hbb^{th\text{-}1}$) survive to adulthood and are much less anemic [Skow, L. C., Burkhart, B. A., Johnson, F. M., Popp, R. A., Goldberg, S. Z., Anderson, W. F., Barnett, L. B. & Lewis, S. E. (1983) *Cell* 34, 1043–1052]. These different phenotypes have implications for the control of β-globin gene expression.**

The α- and β-globin polypeptide chains of hemoglobin are encoded by multigene clusters that are generally conserved throughout mammalian evolution. Within the clusters, the individual genes are regulated in both a temporal and tissue-specific manner. There are five functional genes in the human β-globin gene cluster: ε, an embryonic globin gene expressed primarily in yolk sac-derived cells from 3 to 8 weeks of gestation; $^{G}\gamma$ and $^{A}\gamma$, fetal globin genes expressed primarily in fetal liver-derived cells from 6 weeks of gestation to ≈3 mo after birth; and δ and β, adult globin genes expressed primarily in bone marrow-derived cells starting shortly before birth and persisting throughout adult life. The β gene is responsible for 97–98% of adult β-globin, and the δ gene is responsible for ≈2–3%. There are four functional genes in the mouse β-globin gene cluster: *bh1*, an early embryonic globin gene expressed primarily in yolk sac-derived cells from 9.5 to 12.5 days of gestation; *y*, a late embryonic globin gene expressed primarily in fetal liver-derived cells from 11.5 to 16.5 days of gestation; and *b1* (β^{major}) and *b2* (β^{minor}), adult globin genes first expressed at 9.5 days of gestation in yolk sac-derived cells, then expressed in fetal liver-derived cells, then expressed in spleen, and finally expressed in bone marrow-derived cells throughout adult life (1). The *b1* gene is responsible for ≈80% of adult β-globin, and *b2* is responsible for ≈20% (2). The net synthesis of α- and β-type globin polypeptide chains is normally balanced at 1:1.

As a step toward generating mice unable to synthesize any mouse adult β-globins, we have used homologous recombination in embryonic stem (ES) cells to disrupt the mouse major adult β-globin gene (*b1*) by inserting into its second exon a bacterial gene (*neo*) that confers resistance to the drug G418. Mice homozygous for this insertional disruption of the *b1* gene die perinatally due to severe anemia. This thalassemic phenotype is much more severe than expected because mice homozygous for deletion of the same gene can survive (3). We propose a model suggesting how the differences in the two phenotypes can be explained based on competition of genes in the globin cluster for interaction with the locus control region (LCR) (4–7) and for translational factors.

MATERIALS AND METHODS

Cells and Electroporation. Gene targeting was done with E14TG2a ES cells isolated from inbred strain 129 mice (8). The cells were cultured on embryonic fibroblast feeder layers in Dulbecco's modified Eagle's medium (DMEM)/15% fetal bovine serum/0.1 mM 2-mercaptoethanol/2 mM glutamine. The targeting plasmid (see Fig. 1*B*) was introduced into the ES cells by electroporation by discharging a 200-μF capacitor charged to 300 V through a 5-mm-long chamber with a 100-mm² cross-section. Cells (2×10^7) were used with a DNA concentration of 5 nM. After electroporation, cells were cultured at a density of either 5×10^6 or 2×10^6 on 100-mm-diameter dishes and were exposed to 200 mg of G418 (GIBCO) per ml and 2 mM ganciclovir (Syntex, Palo Alto, CA).

Screening for Homologous Recombinants. One portion of each colony surviving exposure to the two drugs was subcultured; another was pooled with five other colonies and analyzed by PCR as described (9), except that 0.5 μg of DNA and 1 unit of *Taq* DNA polymerase were used in each reaction. The PCR mixtures were amplified for 35 cycles with 45 sec of denaturing at 92°C followed by annealing and extension at 65°C for 10 min. One primer, 5′-GCTAACCA-GATTTGTGAGCTCAGGG-3′, was specific for sequences in the target gene *b1*, and the other, 5′-TGGCGGACCGC-TATCAGGAC-3′, was specific for sequences in the *neo* gene. After PCR, 20 μl of the reaction mixture was electrophoresed on an agarose gel, transferred to nylon membranes, and hybridized to a ^{32}P-labeled *b1*-specific probe (see Fig. 1*C*) designed to detect a 1.6-kb band expected when these two primers are used to amplify DNA from a correctly targeted cell. Three pools gave a positive PCR signal. Individual colonies from these pools were then tested by PCR, and the positive colonies so identified were further analyzed by Southern blots of digested genomic DNA.

Preparation of Blood. Blood from weanling mice, or day 18.5 embryos, was collected in heparinized hematocrit tubes. Blood smears made for some animals were stained in Wright's stain. For other animals, the tubes were sealed and centrifuged, and hematocrits were measured; the resulting packed red blood cells were then removed, washed once with phosphate-buffered saline, and lysed in cystamine-containing buffer as described (10). Samples containing approximately equal amounts of hemoglobin were electrophoresed on cellulose acetate membranes (Helena Laboratories).

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: LCR, locus control region; neo, neomycin; tk, thymidine kinase; ES, embryonic stem.

*To whom reprint requests should be addressed.

RESULTS AND DISCUSSION

Gene Targeting. The targeting plasmid and gene targeting scheme are illustrated in Fig. 1. The plasmid contains 6.2 kb of DNA, from a strain BALB/c mouse, that is homologous to the *b1* locus. A 1.1-kb fragment from pMC1POLA (Stratagene) capable of conferring resistance to G418 is inserted at a *Bam*HI site in exon 2 of the *b1* gene. The homologous segments were 0.8 kb on the 5′ side and 5.4 kb on the 3′ side of the *neo* gene. Copies of the thymidine kinase (tk) gene from herpes simplex virus 1 are inserted both 5′ and 3′ of the homologous segments so that positive–negative selection could be used (11). After electroporation with the linearized targeting plasmid, 144 colonies were picked from a total of 604 that were resistant to both G418 and ganciclovir. PCR tests of these colonies indicated that three of the 144 colonies were correctly targeted. This number corresponds to a frequency of one targeted colony for every 1.5×10^6 starting cells. The selection with ganciclovir reduced the number of surviving cells to about one-tenth of those surviving with G418 alone.

The three independently targeted cell lines identified by PCR were confirmed by Southern blots of digested genomic DNA. Fig. 2 shows a Southern blot of DNA from the starting cell line and from one of the targeted cell lines after digestion with four restriction enzymes. The *Eco*RI digest shows an endogenous band of 7.3 kb (also present in DNA from the starting cell line) and a band of 2.0 kb, which is the size expected from a correctly targeted chromosome. The *Bgl* II digest gives a 5.4-kb band (corresponding to the endogenous locus) and the expected 6.5-kb band derived from a targeted locus. The *Pst* I digest shows a 2.2-kb endogenous band and a 1.95-kb band expected for targeting. The *Sph* I pattern is also as expected: 20 kb for the endogenous locus and 4.7 kb for the targeted locus. This and other blots were subsequently hybridized to a probe specific for the *neo* gene. The results demonstrated that the only copy of the targeting plasmid in the ES cell genomes is at the targeted locus (data not shown).

Phenotypes. One of the three targeted ES cell lines was used to generate chimeric animals by injection into host blastocysts. Of seven chimeras tested, three males transmitted the ES-derived strain 129 genome to their offspring after being mated to C57BL/6J females. The F_1 offspring that are heterozygous for the disrupted *b1* gene (+/−) appear normal. In contrast, homozygous mutants (−/−) obtained after mating heterozygotes are very pale and die either *in utero* close to term or just after birth. Of five −/− mice that were alive at birth and were cared for by their mothers, none survived >4 hr. The phenotype of the mutant homozygotes is readily apparent as a marked pallor at day 16.5 of gestation, the time at which the mouse embryonic genes are no longer expressed, and the embryo becomes dependent on its adult genes (1). The hematocrits of −/− mice at 18.5 days of gestation are extremely low (8.3 ± 3.4%; $n = 7$), reflecting their severe anemia. Heterozygotes have hematocrits (33 ± 4.8%; $n = 7$) indistinguishable from normal homozygotes (34 ± 4.5%; $n = 4$).

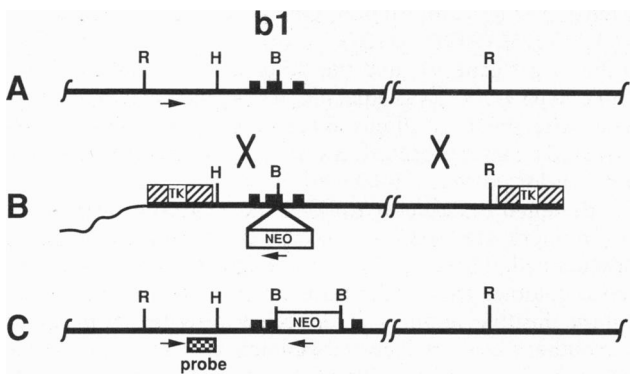

FIG. 1. Targeting construct and homologous recombination. Line A represents the normal endogenous *Hbb*^{d} *b1* locus. Line B represents the linearized targeting plasmid. Line C represents the targeted *b1* locus. A 1.1-kb fragment (NEO in open box) from pMC1POLA (Stratagene) capable of conferring resistance to G418 was inserted at a *Bam*HI (B) site in exon 2 of the *b1* gene. Exons of the *b1* gene are shown as black boxes. Large X's denote regions where crossing-over could occur within the homologous segments. PCR primers are indicated by arrows. A fragment containing the tk gene (TK in a hatched box) from herpes simplex virus 1 was inserted both 5′ and 3′ of the homologous segments. A thin wavy line indicates vector sequences. The probe for hybridization of Southern blots of PCR-amplified fragments and of genomic DNA digests is shown as a stippled box.

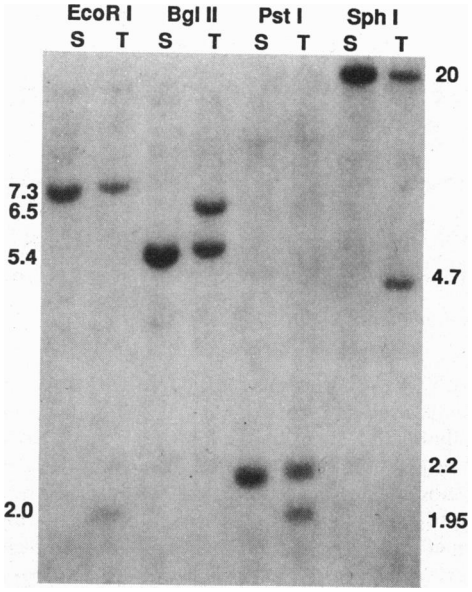

FIG. 2. Southern blots of DNA from the starting and targeted cell lines. The blots of DNA from the starting cell line E14TG2a (S) and a targeted cell line (T) after digestion with four enzymes. The location of the probe used is shown in Fig. 1. Sizes of bands are given in kb.

The lethal thalassemia seen in our disruption homozygotes is not detectably affected by genetic heterogeneity, as judged by observing the same phenotype in F_2, F_3, and F_4 disruption homozygotes derived from the F_1 animals initially obtained by breeding the strain 129-derived male chimeras to C57BL females. (Each of these homozygotes has an essentially random sampling of 129-derived and C57BL-derived alleles at any locus at which these two inbred strains differ.) The insensitivity of the disrupted mutant phenotype to this limited amount of genetic heterogeniety is also illustrated by our observing the same phenotype when the mutation was present in otherwise completely inbred strain 129 animals.

Fig. 3 shows blood smears from normal/mutant heterozygous (+/−) and homozygous mutant (−/−) embryos at 18.5 days of gestation. The mutant (−/−) smear shows characteristics typical of severe thalassemia: few normally rounded red cells, evidence of hypochromia, poikilocytosis, and presence of nucleated red blood cell precursors.

Fig. 4 shows a comparison of the hemoglobins produced by normal (+/+), heterozygous (+/−), and mutant homozygous (−/−) embryos at 18.5 days of gestation. The hemoglobin produced by the −/− mice shows no detectable protein corresponding to the *b1* gene, although the *b2* gene is clearly active.

Comparison with a Deletion Thalassemia in Mice. Some features of the thalassemic phenotype of mice homozygous for the insertionally disrupted *b1* gene (which we propose to

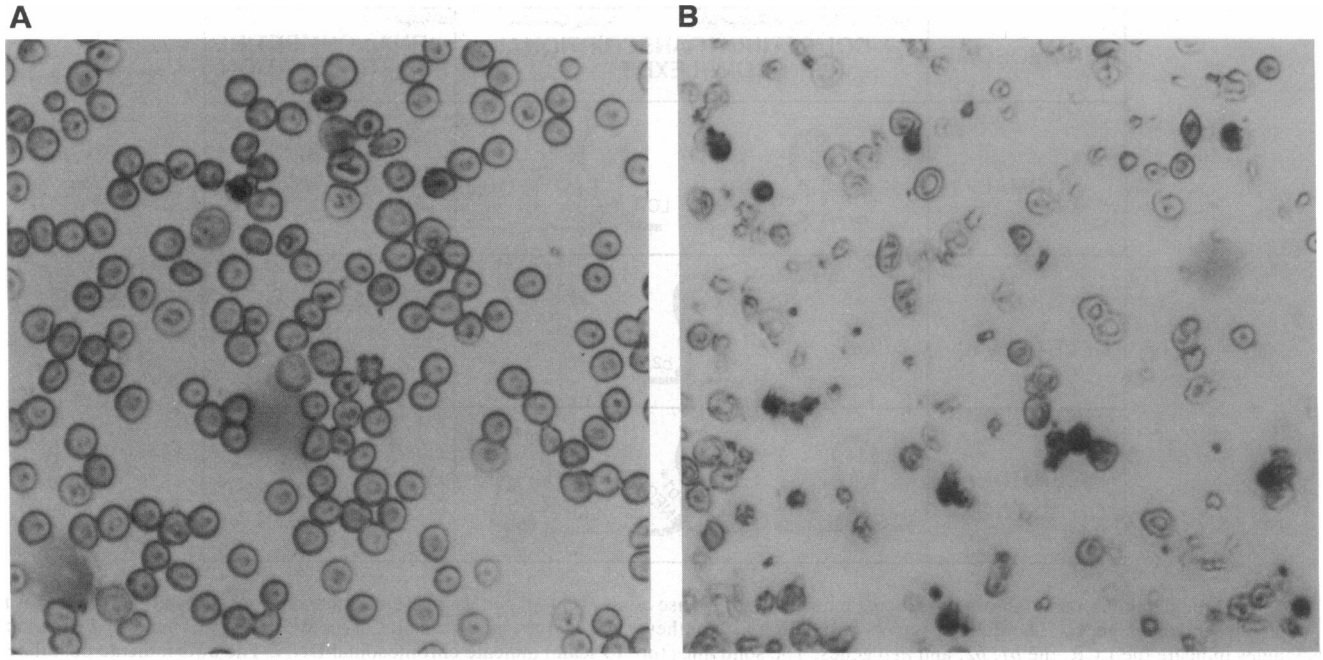

FIG. 3. Blood smears. Smears of whole blood from embryos at 18.5 days of gestation from a normal/mutant heterozygote (+/−) (*A*) and a mutant homozygote (−/−) (*B*) stained with Wright's stain.

call Hbb^{th-2}) are similar to those seen in the thalassemic mice that result from homozygosity for a naturally arising 3709-bp deletion of the *b1* gene (which was named Hbb^{th-1}), in which all of its coding regions and 5′ regulatory sequences are missing (3, 12). Both are small and pale at birth, have reduced hematocrits, and have blood smears characteristic of thalassemia. Like the disruption mutant, the phenotype of the deletion homozygous animals is generally insensitive to genetic background, as judged by finding essentially the same phenotype in mice derived from crosses between mouse strains DBA and C57BL, and in mice derived after backcrossing to C57BL. However, the insertionally disrupted gene is lethal in the homozygous state, whereas more than half of the deletion homozygotes survive to adulthood (3).

Curcio *et al.* (13) have suggested that in the deletion animals there is a "compensatory" increase of 3- to 4-fold in hemoglobin tetramers containing *b2*-derived protein compared to normal mice. The authors report that the ratio of *b2*-derived mRNA to α-globin mRNA is increased from 0.2 in normal animals to 0.3 in deletion homozygotes; thus, part of the compensation is probably due to a relative increase in b2 mRNA levels. They also report that the ratio of *b2*-derived polypeptide chain synthesis to α-globin polypeptide chain synthesis is increased from 0.2 in normal animals to 0.75 in the deletion homozygotes; thus another part of the compensation must occur during the translation of the messages, and the authors present evidence that this is likely due to competition for translational factors.

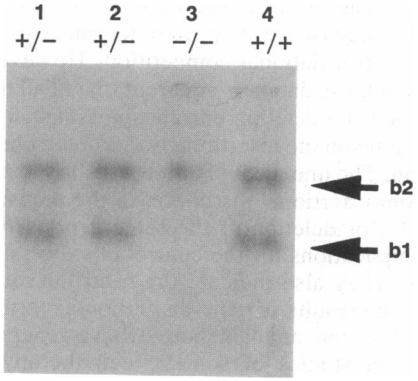

FIG. 4. Cellulose acetate electrophoresis of mouse hemoglobins. The *b1* genotypes of the animals from which the samples were prepared are indicated, as are the positions of hemoglobins containing either *b1*- or *b2*-derived proteins.

Preliminary comparisons have been made in animals heterozygous for one of the mutant chromosomes and for a normal Hbb^s chromosome of the levels of *b2* gene-derived polypeptides from the Hbb^d *b1*-deletion chromosome, and from the Hbb^d *b1*-disruption chromosome, and also from a normal Hbb^d chromosome (J. B. Whitney III, personal communication). The *b2* gene in the Hbb^d *b1*-deletion chromosome produces 58.4% (±1.1%) as much protein as a normal Hbb^s chromosome with two genes ($b1^s$ and $b2^s$). The *b2* gene in the Hbb^d *b1*-disruption chromosome produces 43.1% (±2.5%). The *b2* gene in a normal Hbb^d chromosome produces 32.8% (±3.5%). These results are in accord with the two mutant phenotypes, in that the amount of b2 product is greater with the *b1*-deletion than with the *b1*-disruption chromosome.

Comparison with Human Thalassemias. The different effects on *b2*-globin gene expression of deletion versus disruption of the *b1* gene are remarkably similar to the effects that various mutations in the adult human β-globin gene have on the neighboring adult β-type globin gene, δ (for reviews, see refs. 14 and 15). In normal humans, adult hemoglobin consists of 97–98% hemoglobin A ($\alpha_2\beta_2$) and 2–3% hemoglobin A_2 ($\alpha_2\delta_2$). All β-thalassemic patients with normal δ-globin genes have elevated levels of hemoglobin A_2. Patients who are heterozygotes for a point mutation or small deletion in the β-globin gene that does not remove its 5′ regulatory sequences show hemoglobin A_2 levels in the range of 3.5–6.8%. Patients heterozygous for deletional mutations that remove part or all of the 5′ regulatory sequences of the β-globin gene have unusually high levels of hemoglobin A_2 (7–12%). In one study, by Codrington *et al.* (16), a direct comparison with the mouse data is possible because the authors could distinguish proteins derived from the δ gene in cis and in trans to the mutated β-globin gene. In a mutation removing the β-globin promoter (a 1393-bp deletion), the expression of the δ gene in

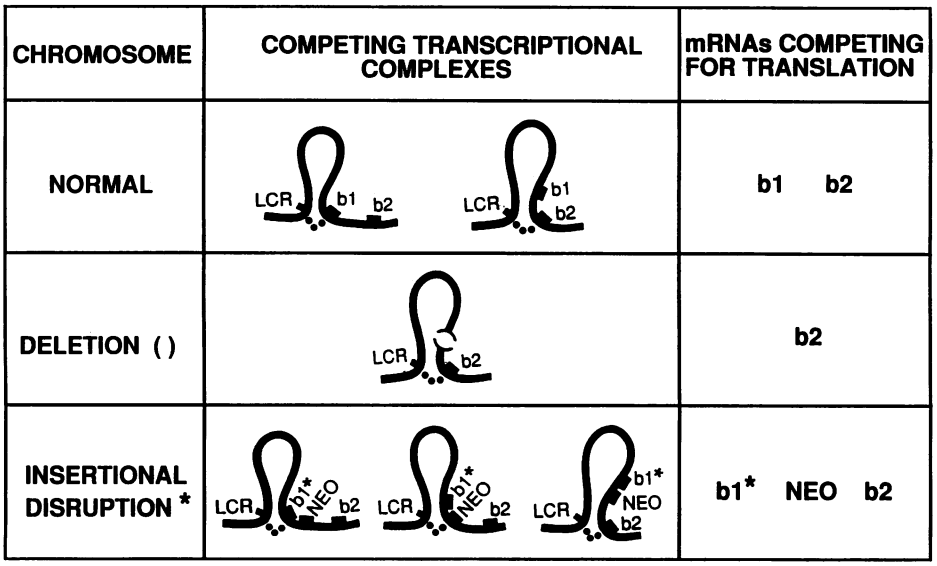

FIG. 5. Model summarizing factors influencing expression of mouse adult β-globin genes. Three chromosomes are identified by the status of their respective *b1* genes. The deleted region is shown by parentheses ( ), and the insertionally disrupted *b1* gene is labeled *b1**. Black rectangles indicate the LCR, the *b1*, *b2*, and *neo* genes. The solid lines (not to scale) indicate chromosomal DNA. The dots indicate postulated interactions between the LCR and promoters in the adult β-globin gene region.

cis was increased much more (5.5% of total hemoglobin) than that of the δ gene in trans (2.5%). In contrast, a thalassemic mutation not affecting the promoter (a frameshift mutation at codon 47) increased the expression equally of both the cis (2.4%) and trans (2.5%) δ-globin genes. The similarity of effects that these various mutations in one adult β-globin gene have on the levels of expression of the neighboring adult β-globin gene suggests that they are the consequence of the same molecular mechanisms in humans and mice.

Expression in Cultured Cells. Related effects have also been observed in gene-targeting experiments in which selectable genes have been inserted into the human β-globin locus in mouse erythroleukemia cells carrying a human chromosome 11. In two cases in which tests were made, expression of the inserted gene was inducible by agents that normally induce β-globin synthesis in mouse erythroleukemia cells (17, 18). This result suggests that the selectable gene has fallen under the influence of elements that normally control β-globin synthesis.

Summary Model. Fig. 5 presents a model summarizing factors that could account for the phenotypic differences between the deletion (Hbb^{th-1}) and disruption (Hbb^{th-2}) animals, for the similar observations in humans, and for the mouse erythroleukemia cell data. This model is based on current views of the way in which the LCR controls expression of genes in the β-globin gene cluster (4–7, 19, 20). The LCR, which lies upstream of the β-globin gene cluster, was originally identified as five regions containing DNase I-hypersensitive sites, the sequences of which are highly conserved in various mammalian species throughout evolution (21, 22), including the mouse (23). The LCR is thought to stimulate β-globin expression as a result of its interaction (probably mediated by proteins) with the promoters of specific β-globin genes. The LCR also has the ability to interact with and stimulate nonglobin promoters, including a *tk* promoter driving a *neo* gene (20). We suggest in our model that deletion of the *b1* gene and its 5′ regulatory sequences allows the promoter of the *b2* gene to interact with the LCR without competition from the promoter of a *b1* gene. Transcription of the *b2* gene would thus be enhanced, and net synthesis of *b2*-derived polypeptides would be increased. In contrast, insertional disruption of the *b1* gene by homologous recombination does not remove any promoter sequences; rather, it inserts an additional promoter (the *tk* promoter driving the *neo* gene). Thus, the *b2* gene on the disrupted chromosome has two promoters competing with it for the upstream LCR, and its transcription is expected to be less than on the deletion chromosome and, possibly, less than on the normal chromosome. In the gene-targeted mouse erythroleukemia cells, interaction of the LCR with the promoter of the selectable gene could explain the selectable gene having acquired the ability to be induced.

Translational "compensation" may also be a factor (13). In the deletion animals, *b2*-globin mRNA is the only β-globin-related mRNA competing with α-globin mRNAs for translational factors. In the insertional disruption animals, the b2-derived mRNA also has to compete with mRNAs transcribed from the *b1* and *neo* gene promoters, neither of which codes for a functional β-globin polypeptide. Such translational competition would further exacerbate the thalassemia caused by the insertion relative to that caused by the deletion.

We stress that, in the absence of direct experiments, the relative importances of the various factors summarized in the model cannot be specified. Clearly, it will be necessary to carry out a detailed analysis of the transcriptional efficacies of the disrupted *b1* gene, the *tk*-driven *neo* gene, and the unaltered *b2* gene. The abundances and stabilities of their mRNAs will likewise need to be determined to assess the importance of translational competition. However, even in the absence of these data it is apparent that at all these levels the *b2* gene on the deletion chromosome has an advantage over the *b2* gene on the insertionally disrupted chromosome.

Conclusion. The unexpected differences in the phenotypes resulting from insertional disruption of the mouse *b1*-globin gene (Hbb^{th-2}) or deletion of the same gene (Hbb^{th-1}) have important implications for the control of expression of clustered genes. They also indicate the need for caution when interpreting the results of gene-disruption experiments. The severely thalassemic animals that we have generated should be of value in studies of somatic cell therapy and as a background for breeding with human β sickle transgenes.

We thank Paulette Catalano for technical assistance and J. B. Whitney III and T. Townes for helpful discussions and critical reading of the manuscript. This work was supported by National

Institutes of Health Award HL37001. W.R.S. was supported by Research Training Grant HL08303.

1. Whitelaw, E., Tsai, S.-F., Hogben, P. & Orkin, S. H. (1990) *Mol. Cell. Biol.* **10,** 6596–6606.
2. Whitney, J. B., III (1977) *Cell* **12,** 863–871.
3. Skow, L. C., Burkhart, B. A., Johnson, F. M., Popp, R. A., Goldberg, S. Z., Anderson, W. F., Barnett, L. B. & Lewis, S. E. (1983) *Cell* **34,** 1043–1052.
4. Choi, O.-R. & Engel, J. D. (1988) *Cell* **9,** 17–26.
5. Enver, T., Raich, N., Ebens, A. J., Papayannopoulou, T., Costantini, F. & Stamatoyannopoulos, G. (1990) *Nature (London)* **344,** 309–313.
6. Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L. & Townes, T. M. (1990) *Genes Dev.* **4,** 380–389.
7. Townes, T. M. & Behringer, R. R. (1990) *Trends Genet.* **6,** 219–223.
8. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. (1987) *Nature (London)* **326,** 292–295.
9. Kim, H.-S. & Smithies, O. (1987) *Nucleic Acids Res.* **16,** 8887–8903.
10. Whitney, J. B., III (1978) *Biochem. Genet.* **16,** 667–672.
11. Mansour, L. S., Thomas, K. R. & Capecchi, M. R. (1988) *Nature (London)* **336,** 348–352.
12. Goldberg, S. Z., Kuebbing, D., Trauber, D., Schafer, M. P., Lewis, S. E., Popp, R. A. & Anderson, W. F. (1986) *J. Biol. Chem.* **261,** 12368–12374.
13. Curcio, M. J., Kantoff, P., Schafer, M. P., Anderson, W. F. & Safer, B. (1986) *J. Biol. Chem.* **261,** 16126–16132.
14. Steinberg, M. H., Coleman, M. B. & Adams, J. G., III (1982) *J. Lab. Clin. Med.* **100,** 548–557.
15. Steinberg, M. H. & Adams, J. G., III (1991) *Blood* **9,** 2165–2177.
16. Codrington, J. F., Li, H.-W., Kutlar, F., Gu, L.-H., Ramachandran, M. & Huisman, T. H. J. (1990) *Blood* **76,** 1246–1249.
17. Nandi, A. K., Roginski, R. S., Gregg, R. G., Smithies, O. & Skoultchi, A. I. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3845–3849.
18. Kim, C. G., Epner, E. M., Forrester, W. C. & Groudine, M. (1992) *Genes Dev.* **6,** 928–938.
19. Hanscombe, O., Vidal, M., Kaeda, J., Luzzatto, L., Greaves, D. & Grosveld, F. (1989) *Genes Dev.* **3,** 1572–1581.
20. Blom van Assendelft, G., Hanscombe, O., Grosveld, F. & Greaves, D. R. (1989) *Cell* **56,** 969–977.
21. Li, Q., Zhou, B., Powers, P., Enver, T. & Stamatoyannopoulos, G. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8207–8211.
22. Boguski, M. S., Hardison, R. C., Schwartz, S. & Miller, W. (1992) *New Biol.* **4,** 247–260.
23. Moon, A. M. & Ley, T. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 7693–7697.