Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 2882–2885, May 1987
Genetics

The same β-globin gene mutation is present on nine different β-thalassemia chromosomes in a Sardinian population

(polymorphic restriction endonuclease sites/haplotypes/oligonucleotide probes/crossing over/gene conversion)

MARIO PIRASTU*†, RENZO GALANELLO†, MARIE A. DOHERTY*, TERESA TUVERI‡, ANTONIO CAO‡, AND YUET WAI KAN*

*Howard Hughes Medical Institute Laboratory, and Division of Genetics and Molecular Hematology of the Department of Medicine, University of California, San Francisco, CA 94143; †Istituto di Ricerca sulle Talassemie e Anemie Mediterranee, Consiglio Nationale delle Richerche, Cagliari, Sardinia, Italy; and ‡Istituto di Clinica et Biologia del Etá Evolutiva, Universitá degli Studie di Cagliari, Cagliari, Sardinia, Italy

Contributed by Yuet Wai Kan, December 31, 1986

ABSTRACT **The predominant β-thalassemia in Sardinia is the β^0 type in which no β-globin chains are synthesized in the homozygous state. We determined the β-thalassemia mutations in this population by the oligonucleotide-probe method and defined the chromosome haplotypes on which the mutation resides. The same $\beta^{39(CAG\rightarrow TAG)}$ nonsense mutation was found on nine different chromosome haplotypes. Although this mutation may have arisen more than once, the multiple haplotypes could also be generated by crossing over and gene conversion events. These findings underscore the frequency of mutational events in the β-globin gene region.**

The DNA sequences along the human β-globin cluster are highly polymorphic; over 20 polymorphic restriction endonuclease sites along this 60-kilobase (kb) region of DNA have been described (1). The polymorphic restriction sites serve as useful genetic markers, and analyses for the presence or absence of these restriction sites generate chromosome haplotypes (2). Chromosome haplotypes have provided effective strategies for defining various thalassemia lesions (3) and for prenatal diagnosis of β-thalassemia (4–8). In addition, polymorphic restriction sites have been used as markers for tracing the origin and migration of mutant genes (9–13). As with other genetic markers, the haplotypes have characteristic distributions in a given population (3, 14–16). Mutations of the β-globin gene occurred on chromosomes with these haplotypes, and hence, each mutation became strongly linked to the haplotype from which it arose (3). However, as more β-thalassemia lesions are defined, it becomes apparent that more than one thalassemic mutation can be associated with one haplotype, and conversely, a single mutation may be found on several chromosome haplotypes. In this study we describe an extreme example of heterogeneity in chromosome haplotypes found in Sardinian β-thalassemia.

Previous studies demonstrated that the predominant β-thalassemia in southern Sardinia is the β^0 type in which no β-globin chains are produced from the affected chromosome (17, 18). We have now identified the β^0-thalassemia lesion in affected Sardinian individuals and compared the chromosome haplotypes to those in unaffected Sardinian subjects. The results revealed that a single β-thalassemia lesion resides on nine different chromosomes. Although the same nucleotide mutation could have arisen more than once on several occasions, the multiple haplotypes could also be generated by crossing over and gene conversion events.

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.

MATERIALS AND METHODS

Subjects. The haplotypes of the polymorphic sites along the β-globin gene cluster were analyzed initially in 50 Sardinian patients affected with homozygous β^0-thalassemia. Normal chromosome haplotypes were determined either from normal individuals or derived from the nonthalassemic chromosomes from parents of the homozygously affected subjects. Haplotype analysis was also performed on a family with the Sardinian $^A\gamma$–$\delta\beta$ thalassemia mutation (19, 20) and on a patient with triplicated γ-globin genes due to the addition of an extra $^G\gamma$-globin gene (M.P. and A.C., unpublished data).

Restriction Endonuclease Analysis. DNA was prepared from peripheral white blood cells as described (21). The following polymorphic restriction sites were determined in all patients and controls: the *Hin*cII 5′ to the ε-globin gene, the two *Hin*dIII sites at the intervening sequence of the $^G\gamma$- and $^A\gamma$-globin genes, the two *Hin*cII sites at the ψβ-globin gene and 3′ to it, the *Ava* II site at the β-globin gene, and the *Bam*HI site 3′ to it. Additional sites determined in patients with homozygous β-thalassemia included the *Hgi*AI site in exon I of the β-globin gene and the *Hpa* I, *Hin*dIII, and *Rsa* I sites 3′ to the β-globin gene. The locations of these sites are summarized in Fig. 1.

Oligonucleotide Probes. The presence of the $\beta^{39(CAG\rightarrow TAG)}$ nonsense mutation on the chromosomes of the β^0-thalassemic patients was analyzed by the oligonucleotide method (22). Two nonadecamer probes, 5′-CCTTGGACCCAGAGGTTCT-3′ and 3′-GGAACCTGGATCTCCAAGA-5′, were synthesized and were complementary to each other except for the underlined nucleotides, which correspond to the first nucleotide of the β-globin position 39 codon. The former probe (β^A) corresponds to the coding strand of the normal β-globin gene at the position corresponding to amino acid numbers 36–42. The latter probe (β^{th}) corresponds to the sequence of the noncoding strand of the β^0-thalassemia gene at the same position. Ten micrograms of genomic DNA was digested with *Bam*HI, and the DNA fragments were separated by agarose gel electrophoresis. The gels were dried, hybridized with the oligonucleotides that had been ^{32}P-labeled at the 5′ termini by phosphorylation, washed at 55°C, and autoradiographed. Two parallel gels were run and hybridized with the β^A and β^{th} probes. The experimental procedures have been described in detail (23).

RESULTS

Chromosome Haplotypes in 100 Normal and 100 β-Thalassemia Chromosomes. Table 1 shows the distribution of the chromosomes containing the various polymorphic restriction endonuclease sites in the β-globin cluster. More heterogeneity was found in the haplotypes of the normal population than

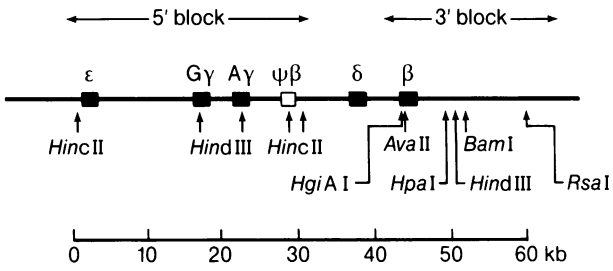

FIG. 1. The β-globin gene cluster and the location of the polymorphic restriction endonuclease sites studied. The sites in the upper row were examined in all patients and controls; those in the lower row were defined for the β-thalassemia chromosomes only. The sites are divided into 5′ and 3′ blocks as indicated.

in those from the thalassemic patients. For purposes of analysis, these sites were divided into two blocks: a 5′ block consisting of sites associated with ε-, $^{G}\gamma$-, $^{A}\gamma$-, and ψβ-globin genes, and a 3′ block that includes the *Ava* II sites at the β-globin gene and the *Bam*HI site 3′ to it. Five different patterns were associated with the 5′ block, and three patterns were associated with the 3′ block. Of the 15 possible haplotypes that could result from the combination of these two blocks, 13 were found in the 100 normal chromosomes, but only 5 in the 100 β-thalassemia chromosomes. Of these 5, both the *Ava* II site at the β-globin gene and the *Bam*HI site 3′ to it were present, suggesting that this region is similar in all chromosomes.

Definition of the β-Thalassemia Mutation in the β-Thalassemia Chromosome. Of the five haplotypes containing the β-thalassemia chromosome, haplotypes 1a and 3a (Table 1) have previously been characterized as $\beta^{39(CAG\rightarrow TAG)}$ nonsense mutation by molecular cloning and sequence analysis (3, 24). We used oligonucleotide probes to determine whether the other haplotypes contained the same mutation. Two oligonucleotides corresponding in sequence either to the normal β-globin gene or to the β^{39} nonsense mutation served as hybridization probes for DNA digested with *Bam*HI and fractionated by agarose gel electrophoresis. Under stringent hybridization and washing conditions, DNA from normal chromosomes hybridized only with the β^{A}, whereas DNA from thalassemic chromosomes hybridized to the β^{th} probe (23). This analysis revealed that DNAs from the patients who are heterozygous for β-thalassemia hybridized to both probes, whereas DNAs from the homozygous patients hybridized only to the β^{th} probe, irrespective of which one of the five haplotypes was associated with the β-thalassemia lesion (Fig. 2). These results demonstrate that all five haplotypes contain the same β^{39} nonsense mutation.

Extended Analysis of the β-Thalassemia Chromosome. The similarity of the *Ava* II and *Bam*HI sites in all five β-thalassemia chromosomes suggests that the DNA sequences at this region of the β-globin gene cluster are also similar. We therefore examined four additional sites at the β-globin gene and 3′ to it: *Hgi*AI at exon I of the β-globin gene, *Hpa* I, *Hin*dIII, and *Rsa* I 3′ to the β-globin gene and extended the haplotype analysis to additional patients with the β^{39} nonsense mutation. The results of these studies raised the number of chromosome haplotypes that harbored the β^{39} nonsense mutation to nine (Table 2). The six polymorphic sites at the 3′ block, which include sites at the β-globin gene and 3′ to it, were identical in the nine haplotypes with two exceptions: (*i*) the *Rsa* I site was present in an individual with haplotype 1 (designated 1′ in Table 2); (*ii*) one chromosome with haplotype 2 (designated 2″) did not have the *Hgi*AI and *Ava* II sites that were present in all the other β-thalassemia chromosomes.

In the 5′ block two other variations were found. One chromosome that harbored the Sardinian nondeletion $^{A}\gamma$–δβ thalassemia mutation also contained a β^{39} nonsense mutation. This chromosome (2′) had all the other characteristics of haplotype 2. Another chromosome (5′) containing triplicated γ-globin genes due to the addition of a $^{G}\gamma$-globin gene also harbored the β^{39} nonsense mutation. Otherwise, the restriction sites of this chromosome resembled haplotype 5.

DISCUSSION

In this study we examined the chromosome haplotypes of Sardinian patients with β^{0}-thalassemia and found that nine different chromosomes harbored the same β^{39} nonsense mutation. There are three possible explanations for this remarkable degree of heterogeneity of haplotype distribution. (*i*) The β^{39} nonsense mutation in Sardinia arose more than once, although the chances of the same mutation occurring nine times in a single population are extremely remote. (*ii*) The β^{39} mutation occurred only once, and the new chromosome haplotypes were generated by recombination events. (*iii*) The diversity results from a combination of multiple mutations and recombination events.

Antonarakis *et al.* (2) described nonrandom distribution of restriction sites along the β-globin gene cluster and divided these sites into two blocks—one encompassing the ε-, $^{G}\gamma$-,

Table 1. Haplotypes of polymorphic restriction endonuclease sites in the β-globin gene cluster in Sardinia

Haplotype	*Hin*cII 5′ε	*Hin*dIII $^{G}\gamma$	*Hin*dIII $^{A}\gamma$	*Hin*cII ψβ	*Hin*cII 3′ ψβ	*Ava* II β	*Bam*HI 3′ β	Frequency, % Normal	Frequency, % β-thalassemia	Mediterranean haplotype*
1a	−	+	+	−	+	+	+	6	64	II
b						−	+	12	0	
c						+	−	4	0	
2a	+	−	−	−	−	+	+	32	22	I
b						−	+	12	0	VII
c						+	−	14	0	V
3a	−	+	−	+	+	+	+	10	4	IX
b						−	+	3	0	IV
c						+	−	2	0	III
4a	−	+	−	−	+	+	+	1	4	
b						−	+	0	0	
c						+	−	1	0	
5a	−	+	+	−	−	+	+	0	6	
b						−	+	1	0	VI
c						+	−	2	0	

One hundred chromosomes each from normal and β^{0} thalassemic subjects were studied.
*These Mediterranean thalassemia haplotypes are classified according to Orkin *et al.* (3).

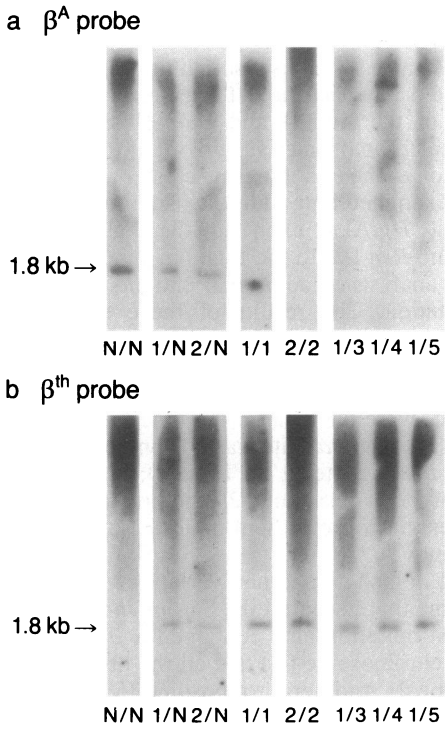

FIG. 2. Autoradiographs of gels used to detect the $\beta^{39(CAG\rightarrow TAG)}$ nonsense mutation. N denotes normal chromosomes, and the numbers designate β-thalassemic chromosomes according to their respective haplotypes shown in Table 1. (*a*) Gels hybridized with the normal (β^A) probe and (*b*) gels hybridized with the β-thalassemia (β^{th}) probe. The β^{39} codon resides in the 1.8-kb *Bam*HI fragment.

$^A\gamma$-, and $\psi\beta$-globin genes, and the other at the β-globin gene and 3′ to it (Fig. 1). The nonrandom association of these two blocks of sites indicates that a hot spot for recombination exists somewhere between the $\psi\beta$- and β-globin genes (Fig. 3*a*). Most of the multiple haplotypes found in Sardinian β-thalassemia can be explained by such recombination events because the restriction sites at the β-globin gene and 3′ to it are identical. The differences in the restriction sites in the 5′ block could then be generated by homologous recombination events that link the mutant β-globin gene to different restriction sites in the 5′ block. Because haplotype 1 [Mediterranean haplotype II of Orkin *et al.* (3)] is the most common, the original mutation probably occurred on a parental chromosome containing this haplotype.

We noted two exceptions to the uniformity of restriction sites in the 3′ block. In chromosome 1′ (Table 2), the single difference was the presence of the *Rsa* I site, which could have arisen from a recombination event between the *Bam*HI and *Rsa* I sites. Indeed in their study of the inheritance of β-globin haplotypes in a nonthalassemic family, Gerhard and coworkers detected a *de novo* recombination event that changed this *Rsa* I site (26).

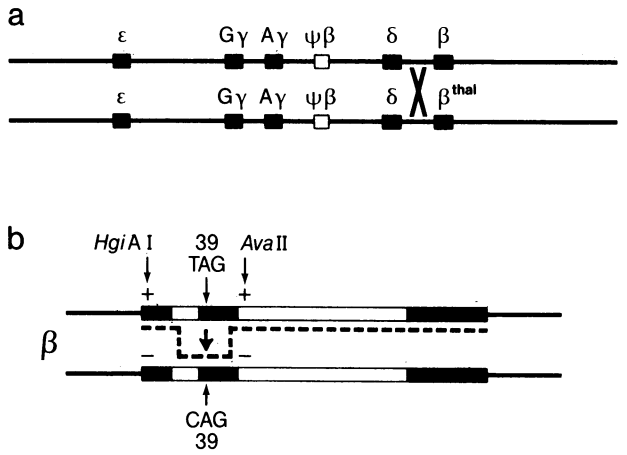

FIG. 3. The two types of recombination events that can alter the chromosome haplotypes. (*a*) Crossing over between the β- and $\psi\beta$-globin genes; (*b*) gene conversion that transfers a β^{39}TAG mutation from a β-globin gene that contains the *Hgi*AI and *Ava* II sites to a normal chromosome that lacks these two sites.

In another chromosome (2″ in Table 2), the *Hgi*AI and *Ava* II sites were absent rather than present, as they were in all the other chromosomes. Interestingly, the β^{39} mutation lies in between these two restriction enzyme sites. This haplotype could have been generated by a new mutation, or, alternatively, a gene conversion event could have transferred a β^{39} mutation to a chromosome that lacks both the *Hgi*AI and *Ava* II sites (Fig. 3*b*). The distance between the *Hgi*AI and *Ava* II sites, ≈300 base pairs, defines the limits within which gene conversion might have occurred. Because this haplotype (2a in Table 1) is common in the nonthalassemic Sardinian population and examples of gene conversion events in the human globin genes have been described (27, 28), such an event could conceivably have generated haplotype 2″.

Two other changes in the 5′ block are also of interest. In haplotype 2′, a chromosome 2 haplotype is associated with increased $^A\gamma$ production. The $^A\gamma$-globin gene from this chromosome has been cloned and sequenced and a cytosine→thymine point mutation was found 196 nucleotides upstream from the $^A\gamma$-globin gene (25). The increased $^A\gamma$ production is attributable to this mutation as similar point mutations have also been found in other examples of nondeletion $\delta\beta$ thalassemia and hereditary persistence of fetal hemoglobin (HPFH). The same mutation at position −196 was also found in China and in southern Italy in chromosomes not associated

Table 2. Haplotypes of β^{39} nonsense mutation

Haplotype*	*Hin*cII 5′ ε	*Hin*dIII $^G\gamma$	*Hin*dIII $^A\gamma$	*Hin*cII ψβ	*Hin*cII 3′ ψβ	*Hgi*AI β	*Ava* II β	*Hpa* I 3′ β	*Hin*dIII 3′ β	*Bam*HI 3′ β	*Rsa* I 3′ β
1	−	+	+	−	−	+	+	+	+	+	−
1′	−	+	+	−	−	+	+	+	+	+	[+]
2	+	−	−	−	−	+	+	+	+	+	−
2′	+	−	[↑]†	−	−	+	+	+	+	+	−
2″	+	−	−	−	−	[−]	[−]	+	+	+	−
3	−	+	−	+	+	+	+	+	+	+	−
4	−	+	−	−	+	+	+	+	+	+	−
5	−	+	+	−	−	+	+	+	+	+	−
5′	− [3γ]‡	+	+	−	−	+	+	+	+	+	−

The β^{39} codon is situated between the *Hgi*AI and the *Ava* II sites.

*The subtype within each haplotype is denoted by ′ or ″, and the differences between the subtypes are indicated within the brackets [ ].

†$^A\gamma$-globin synthesis is increased due to the mutation at position −196 (25). [3γ] refers to triplicate γ-globin gene.

with the β^{39} mutation of the β-globin gene (29, 30). Thus, chromosome 2′ could have arisen from the crossing over of a β thalassemia chromosome with another chromosome harboring the $^{A}\gamma$ mutation at position −196. A chromosome carrying such a double mutation could be expected to impart distinct selective advantages, because the β-thalassemia lesion would protect against malaria, while the increased γ-globin production would ameliorate the severity of the β-thalassemia gene. A similar mechanism could account for chromosome 5′, which is identical to chromosome 5 except for the triple γ loci produced by the addition of the $^{G}\gamma$-globin gene. Such a chromosome has been found in other populations—also without the presence of the β^{39} nonsense mutation (30).

The finding of nine different haplotypes in a single population underscores the frequency of mutational events in the β-globin gene region and perhaps in the human genome in general. However, it is difficult to determine precisely which type of mutational event created this diversity. The simplest explanation is that a single mutation arose initially on chromosome haplotype 1, and subsequent crossovers between the 5′ and 3′ blocks (Fig. 3*a*) produced six other chromosomes (2, 2′, 3, 4, 5, 5′). Chromosome 1′ was then generated by the crossing over between the *Bam*HI and *Rsa* I sites, and chromosome 2″ was generated by a gene conversion event between a β-thalassemia chromosome and a normal chromosome that lacks the *Ava* II and *Hgi*AI sites on either side of the β^{39} position (Fig. 3*b*). Other explanations are also possible. For example, the difference in the *Hin*cII sites at the ψβ gene between haplotypes 3 and 4 could have been generated by a gene conversion event. Additional diversity could have arisen via *de novo* β^{39} nonsense mutations. We recently detected the identical cytosine→thymine mutation at codon 39 that arose spontaneously in a family of northern European origin, suggesting that this position of the β-globin gene is a hot spot for mutation (31).

These findings in the Sardinian population also indicate that mutational events occur at a rapid rate in the human genome. It is generally believed that β-thalassemia assumes a high frequency because the gene provides protection against malaria (32). Hence the diversity in the Sardinian β-thalassemia haplotypes must have occurred over the past few thousand years with the introduction of malaria.

We thank Dr. R. Bruce Wallace for the oligonucleotide probes and J. Gampell for editorial comments. This work was supported in part by grants from the National Institutes of Health (DK16666), UNICO National Inc., the Assessorato Igiene e Sanitá Regione Autonoma della Sardegna, Program of Preventive Medicine Consiglio Nationale della Richerche, and Consiglio Nationale delle Richerche Institute for the Study of Thalassemia and Mediterranean Anemias. Y.W.K. is an Investigator of the Howard Hughes Medical Institute.

1. Bunn, H. F. & Forget, B. G. (1986) *Hemoglobin: Molecular, Genetic and Clinical Aspects* (Saunders, Philadelphia), p. 187.
2. Antonarakis, S. E., Boehm, C. D., Giardinia, P. J. V. & Kazazian, H. H., Jr. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 137–141.
3. Orkin, S. H., Kazazian, H. H., Jr., Antonarakis, S. E,. Goff, S. C., Boehm, C. D., Waber, P. G. & Giardinia, P. J. V. (1982) *Nature (London)* **296,** 627–631.
4. Kan, Y. W., Lee, K. Y., Furbetta, M., Angius, A. & Cao, A. (1980) *N. Engl. J. Med.* **302,** 185–188.
5. Kazazian, H. H., Jr., Phillips, A. J., Jr., Boehm, C. D., Vik, T. A., Mahoney, M. J. & Richey, A. K. (1980) *Blood* **56,** 926–930.
6. Little, P. F. R., Annison, G., Darling, S., Williamson, R., Camba, L. & Modell, B. (1980) *Nature (London)* **285,** 144–147.
7. Boehm, C. D., Antonarakis, S. E., Phillips, J. A., III, Stetten, G. & Kazazian, H. H., Jr. (1983) *N. Engl. J. Med.* **308,** 1054–1058.
8. Chan, V., Leung, N. K., Chan, T. K., Ghosh, A., Kan, Y. W. & Todd, D. (1984) *Brit. Med. J.* **289,** 947–948.
9. Kan, Y. W. & Dozy, A. M. (1980) *Science* **209,** 388–391.
10. Antonarakis, S. E., Boehm, C. D., Sergeant, G. R., Theisen, C. E., Dover, G. J. & Kazazian, H. H., Jr. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 853–856.
11. Pagnier, J., Mears, J. G., Dunda-Belkhodja, O., Schaefer-Rego, K. E., Beldjord, C., Nagel, R. L. & Labie, D. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1771–1773.
12. Antonarakis, S. E., Orkin, S. H., Kazazian, H. H., Jr., Goff, S. C., Boehm, C. D., Waber, P. G., Sexton, J. P., Ostrer, H., Fairbanks, V. F. & Chakravarti, A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6608–6611.
13. Kulozik, A. E., Wainscoat, J. S., Sergeant, G. R., Kar, B. C., Al-Alwamy, B., Essan, G. J. F., Falusi, A. G., Haque, S. K., Hilali, A. M., Kate, S., Ranasinghe, W. A. E. P. & Weatherall, D. J. (1986) *Am. J. Hum. Genet.* **39,** 239–244.
14. Pirastu, M., Loi, A., Rosatelli, C., Saglio, G., Camaschella, C., Serra, A. & Cao, A. (1985) *Bull. Mol. Biol. Med.* **10,** 327–333.
15. Chan, V., Chan, T. K., Cheng, M. Y., Leung, N. K., Kan, Y. W. & Todd, D. (1986) *Hum. Genet.* **73,** 23–26.
16. Chehab, F. F., Der Kaloustian, V., Khouri, F. P., Deeb, S. S. & Kan, Y. W. (1987) *Blood*, in press.
17. Cao, A. (1981) *J. Med. Genet.* **18,** 196–199.
18. Rosatelli, C., Leoni, G. B., Tuveri, T., Scalas, M. T., DiTucci, A. & Cao, A. (1987) *J. Med. Genet.*, in press.
19. Pirastu, M., del Senno, L., Conconi, F., Vullo, C. & Kan, Y. W. (1984) *Nature (London)* **307,** 76.
20. Guida, S., Giglioni, B., Comi, P., Ottolenghis, S., Camaschella, C. & Saglio, G. (1984) *EMBO J.* **3,** 785–787.
21. Goossens, M., & Kan, Y. W. (1981) *Methods Enzymol.* **76,** 805–817.
22. Conner, B. J., Reyes, A. A., Morin, C., Itakura, K., Teplitz, R. L. & Wallace, R. B. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 278–282.
23. Pirastu, M., Kan, Y. W., Cao, A., Conner, B. J., Teplitz, R. L. & Wallace, R. B. (1983) *N. Engl. J. Med.* **309,** 284–287.
24. Trecartin, R. F., Liebhaber, S. A., Chang, J. C., Lee, K. Y., Kan, Y. W., Furbetta, M., Angius, A. & Cao, A. (1981) *J. Clin. Invest.* **68,** 1012–1017.
25. Giglino, B., Casini, C., Mantovani, R., Merli, S., Comi, P., Ottolenghi, S., Saglio, G., Camaschella, C. & Mazza, U. (1984) *EMBO J.* **3,** 2641–2645.
26. Gerhard, D. S., Kidd, K. K., Kidd, J. R., Egeland, J. A. & Housman, D. E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7875–7879.
27. Shen, S. H., Slightom, J. L. & Smithies, O. (1981) *Cell* **26,** 191–203.
28. Liebhaber, S. A., Goossens, M. & Kan, Y. W. (1981) *Nature (London)* **290,** 26–29.
29. Gelinas, R., Bender, M., Lotshaw, C., Waber, P., Kazazian, H. H., Jr., & Stamatoyannopoulos, G. (1986) *Blood* **67,** 1777–1779.
30. Horano, T., Huisman, T. H., Zeng, Y. T. & Huang, S. Z. (1985) *Am. J. Hematol.* **19,** 349–353.
31. Chehab, F. F., Honig, G. R. & Kan, Y. W. (1986) *Lancet* **i,** 3–5.
32. Weatherall, D. J. & Clegg, J. B. (1981) *The Thalassaemia Syndromes* (Blackwell, Oxford, UK), 3rd Ed.