Proc. Natl. Acad. Sci. USA
Vol. 75, No. 11, pp. 5631–5635, November 1978
Genetics

Polymorphism of DNA sequence adjacent to human β-globin structural gene: Relationship to sickle mutation

(sickle cell anemia/cDNA/restriction endonuclease/gene mapping)

YUET WAI KAN*† AND ANDRÉE M. DOZY*

* Hematology Service, and † Howard Hughes Medical Institute Laboratory, San Francisco General Hospital, San Francisco, California 94110; and Departments of * Medicine and † Laboratory Medicine, University of California, San Francisco, California 94143

Communicated by Choh Hao Li, August 21, 1978

ABSTRACT **Restriction endonuclease mapping of the human globin genes revealed a genetic variation in a *Hpa* I recognition site about 5000 nucleotides from the 3′ end of the β-globin structural gene. Instead of a normal 7.6-kilobase (kb) fragment which contains the β-globin structural gene, 7.0-kb and 13.0-kb variants were detected. Both variants were found in people of African origin and were not detected in Asians or Caucasians. The 13.0-kb variant is frequently associated with the sickle hemoglobin mutation and may be useful for the prediction of the sickle cell gene in prenatal diagnosis. Polymorphism in a restriction enzyme site could be considered as a new class of genetic marker and may offer a new approach to linkage analysis and anthropological studies.**

Polymorphism in structural genes in the human is a well-known phenomenon which has been utilized for many types of genetic analyses. Usually, the normal and variant genes or gene products are identified by structural studies, functional assays, or immunological methods. Recently, Southern (1) introduced a new method for analysis of DNA consisting of restriction endonuclease digestion of the genomic DNA, electrophoretic separation of the DNA fragments, and identification of the structural genes in these fragments by hybridization analysis. This method has been used to study the organization of many eukaryotic genomes, including the human globin genes (2, 3).

During study of the human globin genes by the Southern method, we detected genetic variations of a restriction endonuclease site close to the human β-globin structural gene. The variations were found in people of African origin, and one variant was commonly associated with the β-globin structural mutation, hemoglobin S. This type of polymorphism may be useful for linkage analysis, prenatal diagnosis, or anthropological studies.

MATERIALS AND METHODS

Preparation of Cellular DNA. DNA was prepared from leukocytes, placentas, and cultured fibroblasts according to described methods (4, 5), except that 100 μg of proteinase K per ml was used to digest the cells at 55°C, and the phenol-extracted DNA was extensively dialyzed against 1 mM Tris·HCl, pH 7.5. The subjects studied included 46 black individuals [15 with normal hemoglobin (type AA), 16 with sickle cell trait (AS), and 15 with sickle cell anemia (SS)] and 27 nonblack individuals [12 Caucasians and 15 Asians, all with normal hemoglobin (AA)]. To help identify and order the fragments containing the globin genes on the restriction endonuclease patterns, we prepared DNAs from patients with hereditary persistence of fetal hemoglobin (HPFH) (δ- and β-globin gene deletion) (6), homozygous α thalassemia associated with hydrops fetalis (α-globin gene deletion) (4, 7), or homozygous hemoglobin Lepore (δ-β-globin fusion gene) (8).

Digestion of DNA with Restriction Endonucleases. Ten micrograms of human DNA and 1 μg of λ DNA, added as an internal size marker, were digested for 4 hr at 37°C with 1.25 units of *Eco*RI or *Hpa* I per μg of DNA. The buffer for *Eco*RI digestion was 100 mM Tris·HCl, pH 7.5/50 mM NaCl/6 mM $MgCl_2$/6 mM 2-mercaptoethanol; for *Hpa* I digestion it was 6 mM Tris·HCl, pH 7.5/26 mM NaCl/6 mM $MgCl_2$/6 mM 2-mercaptoethanol. The samples were precipitated in alcohol, dried, and resuspended in 30 μl of 5 mM Tris, pH 7.5/0.1 mM EDTA.

Electrophoresis and Identification of Globin Genes. Samples (10 μg) of digested DNAs were applied to 6-mm-thick horizontal 0.8% agarose (SeaKem) gels in a buffer (pH 8.05) containing 0.04 M Tris·acetate, 0.02 M Na acetate, 0.018 M NaCl, and 0.02 M EDTA and were electrophoresed at 50 V for 14 hr. The gels were stained for 30 min in ethidium bromide (10 μg/ml in H_2O) and photographed under ultraviolet light. The DNAs were transferred to nitrocellulose filters with 0.90 M NaCl/0.09 M Na citrate for 40 hr (1). The filters were dried in a vacuum oven at 80°C for 2 hr and presoaked in 2–4 ml of 50% (vol/vol) formamide/0.45 M NaCl/0.045 M Na citrate containing 200 μg of yeast tRNA and 200 μg of denatured salmon sperm DNA per ml and 1% (vol/vol) Denhardt's solution (9). The filters were wrapped in adhesive polyethylene (Saran-Wrap) and incubated at 41°C for 16 hr. [^{32}P]cDNA (1.5×10^6 cpm) in 1.5 ml of the same buffer was added and the filters were rewrapped and hybridized at 41°C for 3 days. The filters were washed once in 0.3 M NaCl/0.03 M Na citrate containing 1% Denhardt's solution at room temperature for 60 min and twice in 15 mM NaCl/1.5 mM Na citrate/0.1% sodium dodecyl sulfate for 90 min at 50°C and then rinsed twice in 15 mM NaCl/1.5 mM Na citrate/0.1% sodium dodecyl sulfate and four times in 15 mM NaCl/1.5 mM Na citrate at room temperature. The dried filters were autoradiographed with Lightning Plus intensifying screens at −80°C for 1–3 days (10).

Globin cDNA Preparation. mRNA from the reticulocytes of a patient with pyruvate kinase deficiency was purified twice by oligo(dT)-cellulose chromatography, followed by sucrose density gradient centrifugation (5, 11). The 10S peak was collected and used for cDNA preparation as described (11), with [α-^{32}P]dCTP (350 Ci/mmol) as the radioactive precursor. To orientate the globin gene on the restriction fragment, a probe rich in 5′-end sequences was used for hybridization of the filter. This probe was made by using an *Hae* III fragment of single-stranded cDNA corresponding to the middle of the coding region of β-globin mRNA (11) as primer for reverse transcription of β-globin mRNA.

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: HPFH, hereditary persistence of fetal hemoglobin; kb, kilobases.

RESULTS

When normal human globin DNA was digested with *Eco*RI and probed with a mixture of α- and β-globin cDNAs, five major fragments were seen (Fig. 1). The largest fragment, about 21 kilobases (kb) in length, was absent in the DNA from homozygous α-thalassemia (hydrops fetalis), in which all the α-globin genes are deleted. This fragment contained both α-globin structural loci (3). Four fragments, 5.5, 3.7, 2.1, and 1.6 kb in length, were present in normal and hydrops DNA and absent in HPFH DNA, in which the β- and δ-globin structural genes are deleted. These four fragments were derived from the β- and δ-globin structural genes, both of which contain an *Eco*RI site in the coding sequences (2, 12).

When normal DNA was digested with the enzyme *Hpa* I, the two α-structural gene loci were located in two fragments, 14.5 and 4.2 kb in length, both of which were absent in the hydrops DNA. Three fragments, 7.6, 1.8, and 1.3 kb in length and present in the normal and hydrops DNA and absent in HPFH DNA, contained the β- and δ-globin structural genes. Double digestion and analysis of hemoglobin Lepore DNA revealed that the 7.6-kb band contained the β-globin gene and the 1.8- and 1.3-kb bands, the δ-globin gene (see below).

When we studied the DNAs from a number of black individuals, we found variations in the 7.6-kb fragments in the *Hpa* I digestion patterns, but the *Eco*RI patterns remained similar (Fig 2). Some DNA samples contained a 7.0-kb fragment and others, a 13.0-kb fragment. The variants were inherited in a Mendelian manner from parents to offspring and were also found among siblings (unpublished data). The following variations from the normal 7.6/7.6 kb patterns were found: 7.6/7.0, 7.0/13.0, 7.6/13.0, and 13.0/13.0 kb patterns.

Initially we observed the 13.0/13.0 kb pattern in the DNA from a patient with sickle cell anemia. To investigate the relationship of the *Hpa* I variant with the hemoglobin S genotype, we analyzed DNAs from black individuals with or without sickle hemoglobinopathies (Table 1). The normal 7.6 kb-fragment and the 7.0-kb variant were the usual patterns seen in individuals with the hemoglobin A genotype. Only 1 of 15 such individuals had the 7.6/13.0 kb pattern. In contrast, the 13.0-kb variant was most often associated with the hemoglobin S genotype. Of 15 patients with sickle cell anemia, the 13.0/13.0 kb pattern was found in 11 and the 7.6/13.0 pattern, in 4. Thus, the frequency of association of the 13.0-kb fragment with the normal hemoglobin A gene is 0.03 in blacks, and with the S gene it is 0.87 in patients with sickle cell anemia. In 16 individuals with sickle cell trait (AS), 9 had the 7.6/13.0 kb pattern, 1 had the 7.0/13.0 kb pattern, and 6 had either the 7.6/7.6 kb or 7.6/7.0 kb pattern, a frequency of the 13.0-kb band of 0.31 in the AS genotype. In contrast, in the Caucasian and Asian populations, we observed only the 7.6/7.6 kb pattern in 27 individuals studied; neither the 7.0-kb nor the 13.0-kb pattern was seen.

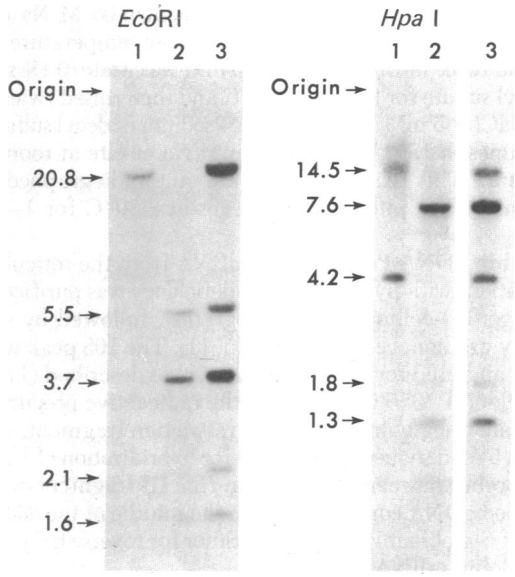

FIG. 1. Autoradiograms of *Eco*RI and *Hpa* I restriction endonuclease digestion patterns of human DNA. α- and β-Globin cDNAs were used as probes. The numbers indicate lengths in kilobases. Lanes: 1, HPFH; 2, homozygous α-thalassemia; 3, normal.

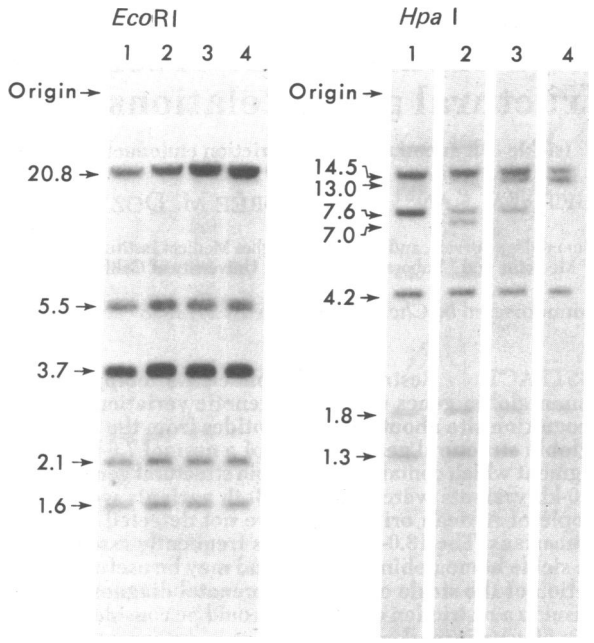

FIG. 2. Autoradiograms of *Eco*RI and *Hpa* I restriction endonuclease digestion patterns of DNA of black individuals with different hemoglobin types. Lanes: 1 and 2, AA; 3, AS; 4, SS. The four types of *Hpa* I fragments containing β-globin gene shown are the 7.6/7.6, 7.6/7.0, 7.6/13.0, and 13.0/13.0 kb patterns, respectively.

To identify the origin of the 7.6-kb fragment and its variants, we compared the *Eco*RI, *Hpa* I, and the double restriction enzyme digestion patterns of the DNAs from a normal individual with the 7.6/7.6 kb pattern, a patient with hemoglobin S with the 13.0/13.0 kb pattern, and one with hemoglobin Lepore (Fig. 3). Digestion of hemoglobin Lepore DNA with *Eco*RI yielded only the 3.7- and 2.1-kb non-α bands, instead of the four in the normal. Of the two, the 2.1-kb band hybridized with the 5′ probe and hence contained the 5′ δ gene. Therefore, the 3.7-kb band must contain the 3′ β globin gene. The 5′ β-probe also hybridized with the 5.5-kb *Eco*RI band of the normal DNA. Thus, we could order the 5.5-, 3.7-, 2.1-, and 1.6-kb *Eco*RI fragments as 5′ β, 3′ β, 5′ δ, and 3′ δ fragments, respectively. This order is in agreement with that of Mears *et al.* (2).

In the *Hpa* I digestion of normal DNA, the three fragments of β and δ origin were 7.6, 1.8, and 1.3 kb in length. Double digestion with *Hpa* I and *Eco*RI yielded four non-α fragments of 3.7, 2.0, 1.8, and 1.2 kb. The 5′ probe hybridized with the 7.6-, 1.8-, and 2.0-kb fragments. Because the *Eco*RI 3.7-kb fragment, which contains the 3′ part of the β-globin gene, remained unchanged in size on the double digestion, the *Hpa* I sites must be outside of the *Eco*RI sites of this fragment. The only non-α *Hpa* I fragment larger than 3.7 kb was the 7.6-kb fragment, which also hybridized with the 5′ probe. Hence, the

Table 1. Relationship between *Hpa* I fragments and hemoglobin genotype

Hemoglobin genotype	*Hpa* I β-globin gene fragment					Total	Frequency of 13.0 kb fragment
	7.6/7.6	7.6/7.0	7.0/13.0	7.6/13.0	13.0/13.0		
Black:							
AA	8	6	0	1	0	15	0.03
AS	5	1	1	9	0	16	0.31
SS	0	0	0	4	11	15	0.87
Caucasian:							
AA	12	0	0	0	0	12	0
Asian:							
AA	15	0	0	0	0	15	0

7.6-kb fragment must contain the whole β-globin gene. The 1.8-kb fragment which hybridized with the 5′ probe must contain the 5′ end of the δ gene, and thus the 3′ δ gene is located in the 1.3-kb band. These assignments were compatible with the restriction pattern obtained in hemoglobin Lepore, with which *Hpa* I digestion yielded only one non-α fragment, 7.3 kb in length, and double digestion yielded a 5′ δ fragment of 1.9 kb and a 3′ β fragment of 3.7 kb. Fig. 4 shows the map of these sites and the sizes of these fragments.

Double digestion with *Eco*RI and *Hpa* I of DNA from a patient with sickle cell anemia with the 13.0/13.0 kb pattern and of normal DNA with the 7.6/7.6 kb pattern abolished the differences seen in the *Hpa* I digestion alone (compare lanes 7 and 8 and lanes 4 and 5 in Fig. 3). Thus, the 13.0-kb fragment is produced by a variation of an *Hpa* I site surrounding the β-globin gene and outside of the *Eco*RI site. Of the two such sites, a variation in the *Hpa* I site on the 5′ side of the β-globin gene is unlikely because shifting this site by 5.4 kb in a 5′ direction to produce a 13.0-kb fragment would affect the digestion pattern of the δ gene, which is about 6 kb from the β-globin gene (2). It would also produce a large 5′ β-globin fragment instead of the 2.0-kb fragment found in double digestion. Hence, the *Hpa* I site affected in the change of the β-globin gene fragment from 7.6 to 13.0 kb is on the 3′ side of the fragment, about 5 kb from the 3′ end of the β-globin structural gene. Shifting this in a 3′ direction by 5.4 kb would not affect the *Eco*RI or the double digestion pattern. Likewise, we found that the double digestion pattern of the DNA with the 7.6/7.0 kb pattern is identical to that of the normal 7.6/7.6 kb pattern (gels not shown). In this case, the *Hpa* I site on the 3′ end is shifted 0.6 kb toward the β-globin gene.

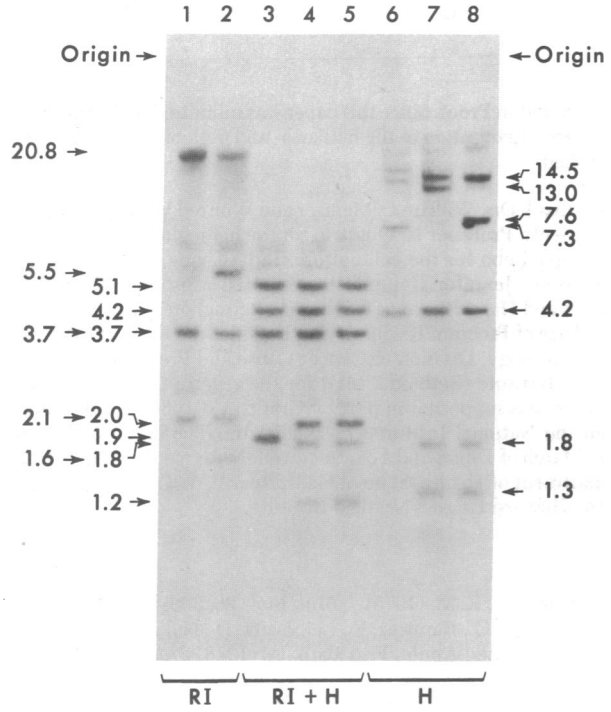

FIG. 3. Autoradiogram of *Eco*RI (RI), double digests (RI and H), and *Hpa* I (H) digests of human DNAs from hemoglobin Lepore (lanes 1, 3, and 6), hemoglobin AA (lanes 2, 5, and 8), and hemoglobin SS (lanes 4 and 7). The vertical columns of numbers indicate the sizes of fragments in the three types of digests. The 5.1- and 4.2-kb bands in the double digest were absent in hydrops DNA (data not shown) and hence contain the two α-globin structural loci. Rehybridization of the filter with a 5′ β-probe showed that the following bands contained 5′ sequences: *Eco*RI digest, 5.5, 2.1; *Hpa* I digest, 13.0 (SS), 7.6, 7.3 (Lepore), 1.8; double digests, 2.0, 1.8, 1.9 (Lepore).

DISCUSSION

In these studies, we detected polymorphism in the restriction endonuclease digestion pattern of a human DNA fragment that contains the β-globin structural gene. This variation is due to an alteration in a *Hpa* I recognition site about 5 kb from the 3′ end of the β-globin structural gene and produces fragments either 7.0 or 13.0 kb long instead of the normal 7.6-kb fragment (Fig. 5). The mechanism producing the variations is not known. It could arise from mutations that create or abolish *Hpa* I recognition sites, delete some sequences, or insert additional sequences. The variants are inherited; homozygous and heterozygous patterns could be recognized. The human non-α-globin genes are believed to be arranged in the order of $G\gamma - A\gamma - \delta - \beta$ from the 5′ to the 3′ end (13). The variable *Hpa* I site located 3′ to this gene complex could be in an intergenic region or a yet unidentified structural gene.

The association of this variation with the sickle cell gene is of great interest. In black individuals with the hemoglobin A genotype, the fragment containing the β-globin gene is predominantly 7.6 or 7.0 kb long. The 13.0-kb fragment is found in only 3% of the hemoglobin A genotype and in 87% of the hemoglobin S genotype. The frequency of the 13.0-kb fragment, 0.31, in the AS genotype is lower than that expected from the findings with the SS genotype. If substantiated with a larger population, this may indicate a selection process affecting the patients with sickle cell disease.

Classical genetic linkage studies have been used to predict certain abnormalities. For example, the secretion of ABH blood group substances in amniotic fluid has been used for the prediction of myotonic muscular dystrophy (14, 15). Restriction endonuclease mapping may provide a new class of genetic markers for linkage studies. In biochemical terms, a sequence 5000 nucleotides away from the β-globin structural gene may be considered very closely linked to that gene. For example, the

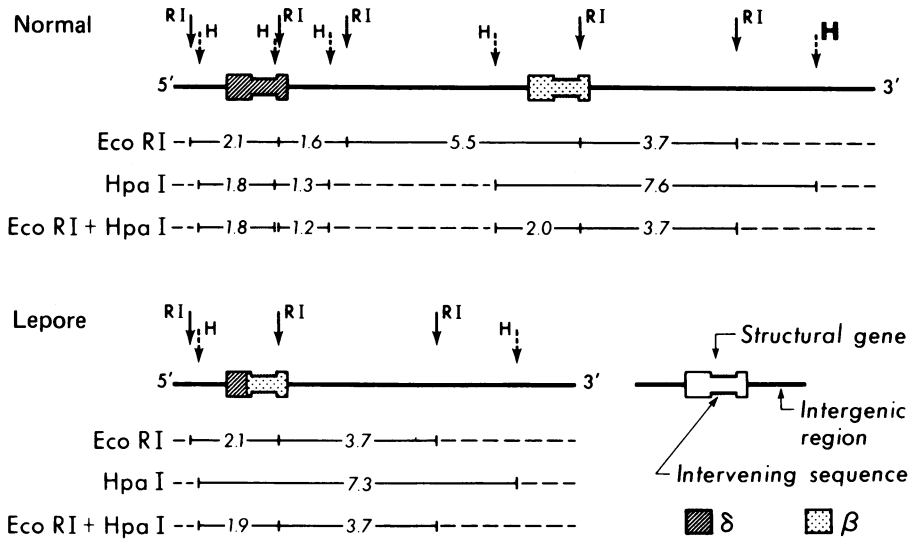

FIG. 4. Map of the *Eco*RI and *Hpa* I sites in the region of the δ- and β-globin structural genes in normal and hemoglobin Lepore DNA. The symbols representing the globin genes are diagrammatic and do not include all the intervening sequences that may be present in the genes. Alteration in the *Hpa* I site on the 3' side of the normal DNA map (marked by the bold **H**) is responsible for the variant 7.0- and 13.0-kb β-globin fragments.

β- and δ-globin structural genes are known to be genetically closely linked because analysis of 14 families with δ- and β-globin structural mutants revealed no recombinants in 76 offspring (16). Recent restriction enzyme mapping showed that the δ- and β-globin structural genes are about 6000 nucleotides apart (2).

The frequency of recombination between the variable *Hpa* I site and the β-globin structural gene needs to be established by studies of large families. If the incidence of recombination between them is low, the association may provide a new approach to the prenatal diagnosis of sickle cell anemia. With parents who carry the 13.0-kb band with the sickle hemoglobin gene, it may be possible to detect sickle cell anemia in the fetus by restriction enzyme analysis of DNA from amniotic fluid cells. Amniotic fluid cells have already been used successfully for the prenatal diagnosis of α-thalassemia by cDNA·DNA hybridization studies (17, 18). Such an approach would avoid the current method which relies on fetal blood sampling, a more involved procedure that carries a higher risk to the fetus than does amniocentesis (19, 20).

Polymorphism of restriction endonuclease patterns may be utilized for other hereditary diseases. If the DNA of a structural gene affected in a hereditary disease is available for hybridization analysis, one can use a number of restriction enzymes to find polymorphism of a restriction site adjacent to this gene. Such polymorphism could be used to predict the occurrence of a normal or abnormal gene. With the current interest in cloning of the human genome, structural genes encoding for other proteins will be found for such studies.

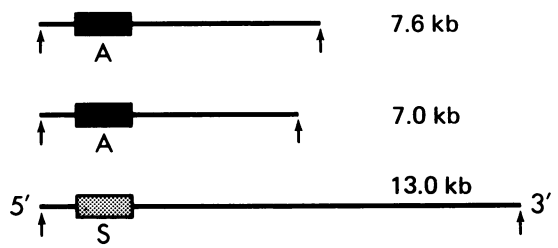

FIG. 5. Diagram of the three types of *Hpa* I fragments. Arrows, *Hpa* I sites; A and S, normal and sickle β-globin genes, respectively.

The preponderance of the *Hpa* I variants in people of African origin suggests that restriction endonuclease mapping, in addition to the conventional method of analysis of structural protein variants, may also be useful for anthropological studies. The fact that enough DNA could be obtained from a few milliliters of peripheral blood makes this method practical for large-scale field studies. Within Africa, the distribution of the 7.0- and 13.0-kb variants in different regions is of great interest.

Note Added in Proof. Since this paper was submitted, two articles (21, 22) on the organization of the human δ- and β-globin genes have been published.

We thank Drs. William C. Mentzer and Winfred Wang for obtaining the patients, Professor N. Quattrin for the hemoglobin Lepore DNA, Dr. Roger Lebo for the cell culture, Judy Chang for many helpful suggestions, Jennifer Gampell for editorial assistance, Drs. Steven Hughes and Harold Varmus for the gel electrophoresis techniques, and the Office of Program Resources and Logistics, Viral Cancer Program, Viral Oncology, Division of Cancer Cause and Prevention, National Cancer Institute (Bethesda, MD) for the reverse transcriptase. This research was supported in part by Grants AM 16666 and HL 20985 from the National Institutes of Health and The National Foundation-March of Dimes and a contract from Maternal and Child Health, Department of Health, State of California. Y.W.K. is an Investigator of the Howard Hughes Medical Institute.

1. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
2. Mears, J. G., Ramirez, F., Leibowitz, D., Nakamura, F., Bloom, A., Konotey-Ahulu, F. & Bank, A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1222–1226.
3. Orkin, S. H. (1978) *Clin. Res.* **26**, 507A.
4. Taylor, J. M., Dozy, A., Kan, Y. W., Varmus, H. E., Lie-Injo, L. E., Ganesan, J. & Todd, D. (1974) *Nature (London)* **251**, 392–393.
5. Kan, Y. W., Dozy, A. M., Trecartin, R. & Todd, D. (1977) *N. Eng. J. Med.* **297**, 1080–1084.
6. Kan, Y. W., Holland, J. P., Dozy, A. M., Charache, S. & Kazazian, H. H. (1975) *Nature (London)* **258**, 162–163.
7. Dozy, A. M., Kabisch, H., Baker, J., Koenig, H. M., Kurachi, S., Stamatoyannopoulos, G., Todd, D. & Kan, Y. W. (1977) *Hemoglobin* **1**, 539–546.

8. Quattrin, N., Bianchi, P., Cimino, R., De Rosa, L., Dini, E. & Ventruto, V. (1967) *Acta Haematol.* **37**, 266.
9. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641–646.
10. Swanstrom, R. & Shank, P. R. (1978) *Anal. Biochem.* **86**, 184–192.
11. Chang, J. C., Temple, G. F., Poon, R., Neumann, K. H. & Kan, Y. W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5145–5149.
12. Orkin, S. H. (1978) *J. Biol. Chem.* **253**, 12–15.
13. Schroeder, W. A. & Huisman, T. H. J. (1974) *Ann. N. Y. Acad. Sci.* **241**, 70–79.
14. Schrott, H. G., Karp, L. & Omenn, G. S. (1973) *Clin. Genet.* **4**, 38–45.
15. Schrott, H. G. & Omenn, G. S. (1975) *Neurology* **25**, 789–791.
16. Stamatoyannopoulos, G., Weitkamp, L. R., Kotsakis, P. & Akrivakis, A. (1977) *Hemoglobin* **1**, 561–570.
17. Kan, Y. W., Golbus, M. S. & Dozy, A. M. (1976) *N. Engl. J. Med.* **295**, 1165–1167.
18. Koenig, H. M., Vedvick, T. S., Dozy, A. M., Golbus, M. S. & Kan, Y. W. (1978) *J. Pediatr.* **92**, 278–281.
19. Kan, Y. W. (1977) *Prog. Hematol.* **10**, 91–105.
20. Alter, B. P. & Nathan, D. G. (1978) *Clin. Haematol.* **7**, 195–216.
21. Mears, J. D., Ramirez, F., Liebowitz, D. & Bank, A. (1978) *Cell* **15**, 15–24.
22. Flavell, R. A., Kooter, J. M., De Boer, E., Little, P. F. R. & Williamson, R. (1978) *Cell* **15**, 25–42.