

Molecular cloning and nucleotide sequence of cDNA for human glucose-6-phosphate dehydrogenase variant A(-)

(DNA sequencing/mutation/polymerase chain reaction/alternative splicing)

AKIRA HIRONO AND ERNEST BEUTLER

Research Institute of Scripps Clinic, Department of Basic and Clinical Research, 10666 North Torrey Pines Road, La Jolla, CA 92037

Contributed by Ernest Beutler, February 16, 1988

ABSTRACT Glucose-6-phosphate dehydrogenase (G6PD; D-glucose-6-phosphate:NADP⁺ oxidoreductase, EC 1.1.1.49) A(-) is a common variant in Blacks that causes sensitivity to drug- and infection-induced hemolytic anemia. A cDNA library was constructed from Epstein-Barr virus-transformed lymphoblastoid cells from a male who was G6PD A(-). One of four cDNA clones isolated contained a sequence not found in the other clones nor in the published cDNA sequence. Consisting of 138 bases and coding 46 amino acids, this segment of cDNA apparently is derived from the alternative splicing involving the 3' end of intron 7. Comparison of the remaining sequences of these clones with the published sequence revealed three nucleotide substitutions: C³³ → G, G²⁰² → A, and A³⁷⁶ → G. Each change produces a new restriction site. Genomic DNA from five G6PD A(-) individuals was amplified by the polymerase chain reaction. The base substitution at position 376, identical to the substitution that has been reported in G6PD A(+), was present in all G6PD A(-) samples and none of the control G6PD B(+) samples examined. The substitution at position 202 was found in four of the five G6PD A(-) samples and no normal control sample. At position 33 guanine was found in all G6PD A(-) samples and seven G6PD B(+) control samples and is, presumably, the usual nucleotide found at this position. The finding of the same mutation in G6PD A(-) as is found in G6PD A(+) strongly suggests that the G6PD A(-) mutation arose in an individual with G6PD A(+), adding another mutation that causes the *in vivo* instability of this enzyme protein.

Glucose-6-phosphate dehydrogenase (G6PD; D-glucose-6-phosphate:NADP⁺ oxidoreductase, EC 1.1.1.49) is an X chromosome-linked enzyme that, in humans, is subject to enormous genetic variability. The normal enzyme is designated G6PD B(+). In West Africa and among American Blacks two mutant enzymes are very prevalent, G6PD A(+) and G6PD A(-). Both of these enzymes have the same electrophoretic mobility in standard systems, and 25 years ago Porter *et al.* (1) suggested that both were due to the same structural mutant, differing only in a closely linked regulatory gene. At the time we pointed out that this probably would not prove to be the case (2) because of emerging evidence that subtle differences existed between the residual enzyme in G6PD A(-) and G6PD A(+) individuals. More recently Babalola *et al.* (3) conducted detailed kinetic studies of G6PD A(+) and G6PD A(-) and suggested that the latter variant might have been derived from the former.

Yet the very similar, or identical, electrophoretic properties of these two common variants and the strong correlation that exists between the frequency of the *Gd^A* and *Gd^{A-}* genes in African populations (4) (the locus for these alleles is now designated *G6PD*) has remained a tantalizing puzzle that

could not be approached directly until development of the modern techniques of molecular cloning. We now describe the molecular cloning and sequence analysis of cDNA for G6PD A(-)* and show that this common variant has been derived from G6PD A(+) as a result of an additional, superimposed mutation. Moreover, we find that, presumably as a result of alternative splicing, there is considerable heterogeneity among different G6PD cDNAs. An abstract of some of these investigations has been published (5).

MATERIALS AND METHODS

Materials. Avian myeloblastosis virus reverse transcriptase was obtained from Molecular Genetic Resources (Tampa, FL). Oligo(dT)₁₂₋₁₈, oligo(dT)-cellulose and *EcoRI* linkers were from Pharmacia. ³²P-labeled nucleotides were purchased from New England Nuclear. *Taq* DNA polymerase was obtained from Perkin Elmer Cetus (Norwalk, CT), and restriction endonucleases and other enzymes were from New England Biolabs. Other reagents used were of analytical grade.

Construction of cDNA Library. An Epstein-Barr virus-transformed lymphoblastoid cell line was established from peripheral lymphocytes from a Black male that is G6PD A(-). Authenticity of the G6PD variant was verified by G6PD activity assay and starch gel electrophoresis (6, 7). Two × 10⁸ cells were harvested, and the total RNA was extracted by the guanidine thiocyanate method (8). Poly(A)⁺ RNA was prepared by oligo(dT)-cellulose column chromatography (9).

Beginning with 20 μg of poly(A)⁺ RNA, cDNA was prepared by a standard method (10) slightly modified by adding *Escherichia coli* DNA ligase during second-strand synthesis (11). Long cDNAs were selected by fractionation on a Sepharose 4B (Pharmacia) column and then ligated to *EcoRI*-cleaved, dephosphorylated λgt10 DNA (Promega Biotech, Madison, WI). *In vitro* packaging was done using commercially available packaging extracts, Gigapack Plus (Stratagene, La Jolla, CA).

Isolation of cDNA Clones and DNA Sequencing. Approximately 1.0 × 10⁶ recombinant phages from the unamplified library were screened by plaque hybridization (12) with a 900-base-pair (bp) cDNA probe and a synthetic oligonucleotide probe (17-mer). The cDNA probe (which contains 620 bp of coding sequence) was obtained from a fibroblast cDNA library from a G6PD B(+) individual using a 20-mer oligonucleotide probe. Positive phage clones were amplified by a large-scale liquid culture (10) and digested with *EcoRI*. The inserts were purified by electrophoresis and subsequently

Abbreviations: G6PD, glucose-6-phosphate dehydrogenase; A(-), A(+), and B(+), variants of G6PD.

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03231).

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.

inserted into the *EcoRI*-cleaved Bluescript M13⁺ vector (Stratagene).

Nucleotide sequences were determined by the dideoxynucleotide chain-termination method (13) using 7-deaza-2'-deoxyguanosine triphosphate (7-deaza-2'-dGTP) instead of dGTP (14). The circular single-stranded DNA serving as template of the reaction was prepared by denaturation of double-stranded DNA by alkali. In addition to primers for each T3 and T7 promoter site in Bluescript DNA, 19 different 17-mer oligonucleotides corresponding to various unique sequences in G6PD cDNA were used as primers. The sequence obtained was compared with the published normal human G6PD cDNA sequence (15, 16). The position of a base was indicated by the numbering system in the sequence published by Persico *et al.* (16) (Table 1).

Analysis of Genomic DNA. Genomic DNA from five G6PD A(-) and seven normal males was isolated from white blood cells (17). Four of these G6PD A(-) subjects were American Blacks and one was Mexican. The diagnosis of G6PD A(-) was based upon enzyme activity, electrophoretic mobility, and ethnic origin. Appropriate DNA sequences were amplified by the polymerase chain reaction (18) with *Taq* DNA polymerase (19) according to a protocol by Chehab *et al.* (20). The synthetic oligonucleotide primers used for polymerase chain reaction are summarized in Table 2. For sequencing some oligonucleotide primers were labeled with ³²P at the 5'-end by means of the polynucleotide kinase reaction. The amplified DNA samples were digested with the appropriate restriction endonuclease according to the supplier's conditions. The digested and undigested DNA fragments were separated by 10 or 12% polyacrylamide gel electrophoresis and visualized by silver staining (21) or autoradiography. The authenticity of amplified DNA sequences was verified by the direct sequencing by the method of Maxam and Gilbert (22).

RESULTS

Isolation of cDNA Clones. A cDNA library prepared from poly(A)⁺ RNA derived from lymphoblastoid cells with G6PD A(-) transformed by Epstein-Barr virus was screened using a cDNA probe. Eighteen positive clones were obtained in the first screening. Two successive screenings were done with the same probe to isolate homogeneous positive plaques. Finally, 15 positive clones were isolated. Of these 15 positive clones, four clones hybridized with an oligonucleotide probe corresponding to the midpoint sequence of coding region of normal human G6PD cDNA. The insert sizes of these four clones designated A-7, A-19, A-12, and A-15 were estimated at 2.3, 2.2, 1.9, and 1.5 kilobase pairs (kbp), respectively, by agarose gel electrophoresis.

Nucleotide Sequence. All of the four positive clones were subcloned into Bluescript M13⁺ vector, and the nucleotide sequence of both strands was determined as shown in Fig. 1. The longest clone, A-7, had an entire coding and 3' noncoding region and a short 5' noncoding sequence that consisted of 14 nucleotides. Clone A-19 also contained the entire coding and a short 5' noncoding sequence with a 3' noncoding region slightly shorter than that of clone A-7. Both A-12 and A-15 clones contained only a partial coding sequence. Surprisingly, the coding regions of A-7 and A-19 clones were found to be of different lengths. Although clone A-19, consisting of

Table 2. Sequence of synthesized oligonucleotide primers

Primer	Sequence
1	5' ACAGCGTCATGGCAGAGCAGGTGGC
2	5' AAAAGCTCTTCCCGCAGGATCCCCG
3	5' GTGGCTGTTCGGGATGGCCTTCTG
4	5' CTTGAGAAGGGCTCACTCTGTTTG
5	5' CAGTACGATGATGCAGC
6	5' CAGGTAGAAGAGGCGGT

1548 bases, was exactly the same length as that previously recorded for the normal G6PD cDNA (16), clone A-7 had an extra 138 bases located just after G⁷⁷⁰. An open reading frame was maintained through this extra sequence, coding 46 additional amino acids and extending part way into the intron (23) (Fig. 2). Neither of the two clones, A-12 and A-15, possesses such extra bases.

The sequences of 3' noncoding regions of the four cDNA clones are almost identical with that presented by Persico *et al.* (16).

When compared with the published sequence (16) the coding region of A-7 had three base substitutions, located at nucleotide positions 33, 202, and 376. As summarized in Table 1 each results in an amino acid substitution.

Analysis of Genomic DNA. To verify that the altered sequence in G6PD A(-) was a consequence of the base substitutions in the genomic DNA and to examine additional subjects, genomic DNA was amplified and examined by restriction endonuclease analysis and/or direct sequencing. Positions 202 and 376 were examined in DNA samples from five G6PD A(-) and five control G6PD B(+) individuals (two Whites, two Blacks, and one Oriental). Position 33 was examined in DNA from these and from two additional Whites.

The 109-bp DNA sequence containing the substitution G²⁰² → A was amplified with 5'-labeled primer 3 and nonradioactive primer 4 (Table 2) and digested with *Nla* III. Fig. 3A shows a typical autoradiogram of this analysis: The amplified DNA from two G6PD A(-) subjects was cleaved to 46-bp and 63-bp fragments (the latter is not visible because of its lack of labeled 5' end). The existence of this base substitution was also confirmed in three G6PD A(-) subjects by direct sequencing of the amplified DNA. Three of the four Black G6PD A(-) subjects and the Mexican G6PD A(-) subject were found to have this substitution.

For the analysis of the base substitution A³⁷⁶ → G, 90-bp DNA sequence was amplified with nonradioactive primers 5 and 6 (Table 2). After digestion with *Fok* I, resultant fragments were separated by 10% polyacrylamide gel electrophoresis and visualized by silver staining (Fig. 3B). The 90-bp amplified fragments from G6PD A(-) DNA were cleaved to 58-bp and 32-bp fragments that were not seen in normal DNA samples (the latter is not visible).

The base substitution C³³ → G was analyzed using 70-bp DNA amplified with 5' labeled primer 1 and nonradioactive primer 2. The amplified fragments were digested with *Bst*NI, and the resultant fragments were separated by 12% polyacrylamide gel electrophoresis and visualized by autoradio-

Table 1. Base substitutions found in G6PD A(-) cDNA

Position*	Base substitution	Amino acid substitution	Restriction site created	Interpretation
33	C → G	His → Gln	<i>Bst</i> NI	Normal or polymorphic
202	G → A	Val → Met	<i>Nla</i> III	Deficiency mutation
376	A → G	Asn → Asp	<i>Fok</i> I	A(+) mutation

*Bases are numbered as by Persico *et al.* (16).

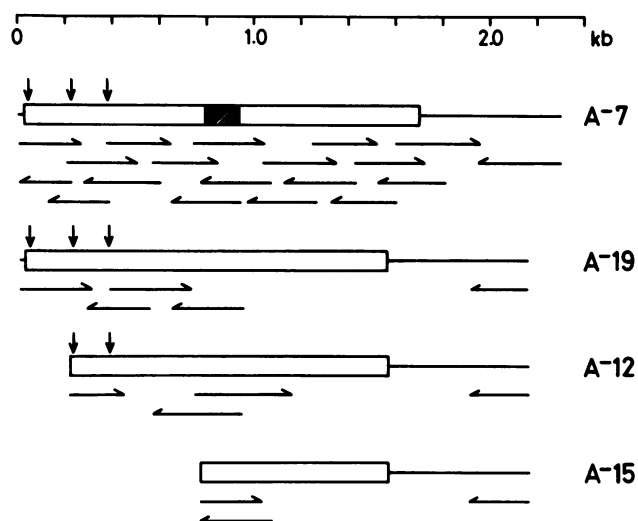


FIG. 1. Sequencing strategy of G6PD A(-) cDNA clones. The open bars represent the coding sequence. Extra sequence in clone A-7 is indicated by the hatched box. Horizontal arrows, direction and extent of sequencing; vertical arrows, position of each base substitution.

graphy (Fig. 3C). Every amplified 70-bp fragment was cleaved to 39-bp and 31-bp fragments (the latter is not visible). The existence of this base substitution in the 70-bp fragment was confirmed by direct sequencing. This result indicated that the base substitution C³³ → G is not specific to G6PD A(-) but rather is the usual, normal nucleotide at this position.

DISCUSSION

The structure of normal human G6PD has been extensively studied in recent years. Complete amino acid sequence (15) and nucleotide sequence of both cDNA (15, 16) and of a portion of genomic DNA (23) have already been elucidated. Very little is known about the structure of G6PD variants. However, Yoshida (24) and Takizawa *et al.* (25, 26) have determined the molecular abnormality of G6PD A(+) as a single-amino acid substitution of Asn → Asp due to a single-base change of adenine to guanine.

In the present study, we isolated G6PD A(-) cDNA and found three base substitutions in the coding sequence when compared with published data (16). Great care must be taken in interpreting an altered sequence in cDNA because irrelevant genetic polymorphism or cloning artifacts could cause a base change. Therefore additional genomic DNA samples from five unrelated G6PD A(-) subjects were examined, and we found that the base substitution, A³⁷⁶ → G, was present in all five G6PD A(-) individuals and G²⁰² → A was present in four of five G6PD A(-) individuals. Both substitutions were absent from all normal G6PD B(+) DNA. These findings show these two base substitutions to be the primary cause of abnormal properties of this variant enzyme. Of these two base substitutions A³⁷⁶ → G is of particular interest because this substitution is identical to that found in G6PD

```

G GGA CCT GGA AGA CAA GGG GGA TCA GGA AGT GAG TCT TGC AGC TTG TCA CTA
(Ar)g Gly Pro Gly Arg Gln Gly Gly Ser Gly Ser Glu Ser Cys Ser Leu Ser Leu

GGA AGC CTT GTT TGG GGT CCC CAT GCC CTT GAA CCA GGT GAA CAG GGC GGG
Gly Ser Leu Val Trp Gly Pro His Ala Leu Glu Pro Gly Glu Gln Gly Gly

GAG CTA AGG CGA GCT CTG GCC TCT TCC GTC CCC AG
Glu Leu Arg Arg Ala Leu Ala Ser Ser Val Pro Ar(g)

```

FIG. 2. Nucleotide sequence of the extra portion of clone A-7. Deduced amino acids are shown below. Underlined sequence is identical with the published partial sequence of the 3' end of intron 7 (23).

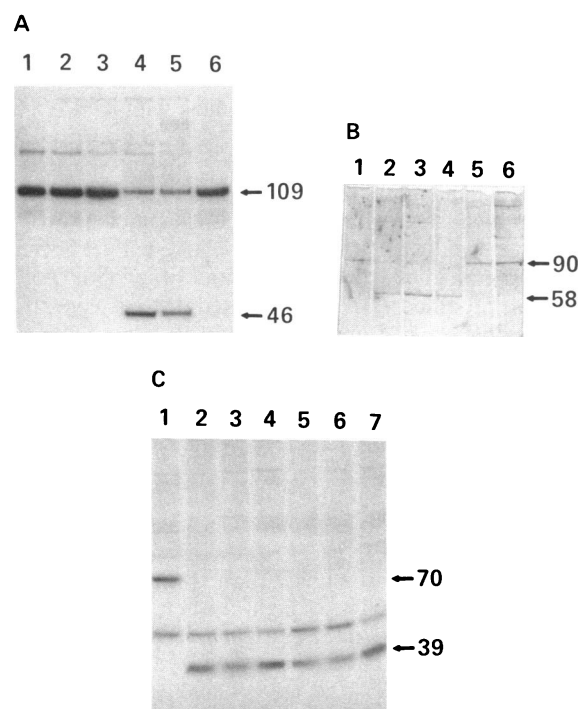


FIG. 3. Restriction endonuclease digestion of amplified genomic sequence. DNA samples from G6PD A(-) and normal individuals were amplified by the polymerase chain reaction using 5'-labeled or nonradioactive primers (Table 2) and digested with appropriate restriction endonucleases. (A) The base substitution G²⁰² → A. The amplified DNA (109 bp) was cleaved by *Nla* III to 46-bp and 63-bp (not visible) fragments in two G6PD A(-) samples (lanes 4 and 5). One G6PD A(-) DNA (lane 6) and normal DNA samples (lanes 2 and 3) remained uncut. Lane 1, undigested DNA sample (12% polyacrylamide gel, autoradiogram). (B) The base substitution A³⁷⁶ → G. The amplified DNA (90 bp) was cleaved by *Fok* I to 58-bp and 32-bp (not visible) fragments in G6PD A(-) samples (lanes 2-4). Lane 1, normal White sample; lanes 5 and 6, normal Black samples (10% acrylamide gel, silver stain). (C) The base substitution C³³ → G. The amplified DNA (70 bp) was cleaved by *Bst*NI to 39-bp and 31-bp (not visible) fragments in both G6PD A(-) (lanes 5 and 6) and normal DNA samples (lanes 2-4 and 7). Lane 1, undigested DNA sample. The 48-bp bands that are seen on all the lanes are artifacts derived from the self-hybridizing of primers.

A(+) (25). Because the substitution in G6PD A(+) would explain its abnormal fast mobility on electrophoresis, the identical amino acid substitution, Asn → Asp, is compatible with like electrophoretic behavior of G6PD A(-).

The other base substitution G²⁰² → A, with the amino acid substitution Val → Met presumably accounts for the other feature of G6PD A(-) — its decreased activity in red cells. Because the low enzyme activity of A(-) variant is considered to be due to low stability in red cells (27, 28), this mutation probably causes the instability of the enzyme protein. The expected very slight change of net charge of a G6PD molecule caused by this amino acid change might correlate with the similar electrophoretic mobilities of G6PD A(+) and G6PD A(-). That G6PD A(+) and G6PD A(-)

contain the same mutation strongly suggests that the latter occurred in a carrier of the former mutation as predicted by Babalola *et al.* (3).

The existence of a Black individual with G6PD A(-) encoded by genes in which the A³⁷⁶ → G substitution is present but the G²⁰² → A mutation is absent suggests the possibility of this individual having undergone another mutation that caused instability and thus deficiency of the electrophoretically rapid enzyme. We have not yet been able to verify the biochemical characteristics of the G6PD of this individual. Modiano *et al.* (29) reported the existence of two types of G6PD B(+) and G6PD A(+), respectively, and speculated that two types of G6PD A(-) might also exist. Probably the mutation in G6PD A(+) (A³⁷⁶ → G) has an ancient origin, and several different superimposed mutations could have subsequently occurred in the G6PD A(+) population.

The third base substitution found in G6PD A(-) cDNA, C³³ → G, was also found in all other DNA samples, both from G6PD A(-) and G6PD B(+). Although there is still the possibility that this base substitution is related to a widely distributed polymorphism, it might, rather, be derived from the base alteration in Persico's original cDNA sequence (16) due to a polymorphic base substitution or a cloning artifact.

Discrepancy between the amino acid sequences of the amino end of normal G6PD determined directly from peptides (15) and that deduced from nucleotide sequence (16) is an unsolved puzzle. Another discrepancy is also recognized in the 3' noncoding sequence of cDNAs reported from two laboratories: 3' noncoding sequence reported by Takizawa *et al.* (15) is more than two times longer than that reported by Persico *et al.* (16). Considering that the reported sequences were determined from different tissues, these heterogeneities are probably due to alternative splicing of G6PD mRNA in different tissues. The existence of additional bases coding an extra 46 amino acids in clone A-7 is of great interest in considering such alternative splicing. Because this extra sequence is located precisely between the sequences derived from exons 7 and 8 and the nucleotide sequence of the 3' end is identical with the published partial sequence of intron 7 (23), we concluded that the extra bases are derived from the 3' end of that intron. Existence of the extra sequence by such alternative splicing in only one cDNA clone of four different clones suggests that this phenomenon is unrelated to the abnormal function of G6PD A(-) variant; more likely, it represents the existence of multiple forms of G6PD molecule in lymphoblastoid cells. More detailed studies of this alternative splicing will be reported elsewhere.

Note Added in Proof. It has now been possible for Dr. Josef Prchal (University of Alabama, Birmingham, AL) to obtain an additional blood sample from the southern United States Black male whose DNA was normal at nucleotide 202. We confirmed the sequence data on a new DNA sample from this subject, and Dr. Prchal again showed that enzyme deficiency was present and that the residual red cell enzyme was electrophoretically indistinguishable from G6PD A(-). However this individual's G6PD differed from G6PD B(+) and G6PD A(-) in its deamino NADP utilization, pH-activity curve, and thermal stability. Meanwhile, we have examined four other putative G6PD A(-) DNA samples, and all have had the nucleotide G²⁰² → A substitution. In view of these additional findings it is possible that the one subject that did not conform to the nucleotide pattern found

in the other G6PD A(-) individuals represented a sporadic mutation occurring in a G6PD A(+) gene; it is not yet clear whether G6PD A(-) is actually polymorphic at the DNA sequence level.

We thank Dr. Josef Prchal for making available to us DNA from some G6PD A(-) individuals. This work is supported by National Institutes of Health Grants HL-25552 and RR00833. This communication is publication 5147BCR from the Research Institute of Scripps Clinic.

- Porter, I. H., Boyer, S. H., Watson-Williams, E. J., Adam, A., Szeinberg, A. & Siniscalco, M. (1964) *Lancet* i, 895-899.
- Beutler, E. & Sparkes, R. S. (1964) *Lancet* ii, 95.
- Babalola, A. O. G., Beetlestone, J. G. & Luzzatto, L. (1976) *J. Biol. Chem.* **251**, 2993-3002.
- Luzzatto, L. (1973) *Isr. J. Med. Sci.* **9**, 1181-1194.
- Hirono, A. & Beutler, E. (1988) *Clin. Res.* **36**, 561A (abstr.).
- Betke, K., Beutler, E., Brewer, G. J., Kirkman, H. N., Luzzatto, L., Motulsky, A. G., Ramot, B. & Siniscalco, M. (1967) *WHO Tech. Rep. Ser. No. 366* (WHO, Geneva).
- Beutler, E., Mathai, C. K. & Smith, J. E. (1968) *Blood* **31**, 131-150.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294-5299.
- Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408-1412.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 76-85, 213-246.
- Lapeyre, B. & Amalric, F. (1985) *Gene* **37**, 215-220.
- Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180-182.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
- Mizusawa, S., Nishimura, S. & Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319-1324.
- Takizawa, T., Huang, I. Y., Ikuta, T. & Yoshida, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4157-4161.
- Persico, M. G., Viglietto, G., Martini, G., Toniolo, D., Paonessa, G., Moscatelli, C., Dono, R., Vulliamy, T., Luzzatto, L. & D'Urso, M. (1986) *Nucleic Acids Res.* **14**, 2511-2522, 7822.
- Goossens, M. & Kan, Y. W. (1981) *Methods Enzymol.* **76**, 805-817.
- Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350-1354.
- Kogan, S. C., Doherty, M. & Gitschier, J. (1987) *N. Engl. J. Med.* **317**, 985-990.
- Chehab, F. F., Doherty, M., Cai, S., Kan, Y. W., Cooper, S. & Rubin, E. M. (1987) *Nature (London)* **329**, 293-294.
- Perbal, B. (1984) in *A Practical Guide to Molecular Cloning* (Wiley, New York), pp. 196-207.
- Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560-564.
- Martini, G., Toniolo, D., Vulliamy, T., Luzzatto, L., Dono, R., Viglietto, G., Paonessa, G., D'Urso, M. & Persico, M. G. (1986) *EMBO J.* **5**, 1849-1855.
- Yoshida, A. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 835-840.
- Takizawa, T., Yoneyama, Y., Miwa, S. & Yoshida, A. (1987) *Genomics* **1**, 228-231.
- Takizawa, T. & Yoshida, A. (1987) *Am. J. Hum. Genet.* **41**, A241 (abstr.).
- Luzzatto, L. & Allan, N.C. (1965) *Biochem. Biophys. Res. Commun.* **21**, 547-554.
- Piomelli, S., Corash, L. M., Davenport, D. D., Miraglia, J. & Amorosi, E. L. (1968) *J. Clin. Invest.* **47**, 940-948.
- Modiano, G., Battistuzzi, G., Esan, G. J. F., Testa, U. & Luzzatto, L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 852-856.