Biochem. J. (1990) 271, 157–160 (Printed in Great Britain)

Identification of a single base change in a new human mutant glucose-6-phosphate dehydrogenase gene by polymerase-chain-reaction amplification of the entire coding region from genomic DNA

Vincenzo POGGI* Margaret TOWN, Nicholas S. FOULKES and Lucio LUZZATTO†
Department of Haematology, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0NN, U.K.

We report the characterization at the molecular level of a mutant glucose-6-phosphate dehydrogenase (G6PD) gene in a Greek boy who presented with a chronic non-spherocytic haemolytic anaemia. In order to identify the mutation from a small amount of patient material, we adopted an approach which by-passes the need to construct a library by using the polymerase chain reaction. The entire coding region was amplified in eight sections, with genomic DNA as template. The DNA fragments were then cloned in an M13 vector and sequenced. The only difference from the sequence of normal G6PD was a T → G substitution at nucleotide position 648 in exon 7, which predicts a substitution of leucine for phenylalanine at amino acid position 216. This mutation creates a new recognition site for the restriction nuclease *BalI*. We confirmed the presence of the mutation in the DNA of the patient's mother, who was found to be heterozygous for the new *BalI* site. This is the first transversion among the point mutations thus far reported in the human G6PD gene.

INTRODUCTION

Glucose-6-phosphate dehydrogenase (G6PD) is a 'housekeeping' enzyme which catalyses the first step in the hexose monophosphate pathway and provides reductive potential, in the form of NADPH, required for a variety of biosynthetic reactions and for the detoxification of H_2O_2. G6PD deficiency is an X-chromosome-linked inherited trait which affects more than 400 million people worldwide. Over 300 genetic variants of G6PD have been identified, with a wide range of clinical manifestations; the most severe type is chronic non-spherocytic haemolytic anaemia (see [1] for a recent review).

The G6PD gene is located in the q28 region of the X chromosome, and it consists of 13 exons and 12 introns distributed over approx. 20 kb of genomic DNA [2]. The exons range in size from 38 bp (exon 3) to 695 bp (exon 13). All introns are smaller than 300 bp, except intron 2, which is about 11 kb in size. The coding region is in exons 2–13, while exon 1, the first 8 bp of exon 2 and all but the first 88 bp of exon 13 form the untranslated portions of the G6PD mRNA.

To date, eleven different variants of G6PD have been characterized at the molecular level [3–8] by cloning the mutant genes using cDNA or genomic libraries. In all cases point mutations within the coding region were identified, predicting individual amino acid substitutions.

In order to work out in detail how naturally occurring point mutations cause biochemical changes and clinical manifestations, it will be necessary to sequence a larger number of G6PD variants. Therefore we were interested in developing a relatively rapid method that could be generally applicable, and naturally we turned to the polymerase chain reaction (PCR), first introduced by Saiki *et al.* [9]. Since variants are frequently identified from limited amounts of patient material, and RNA may be difficult to obtain, we report here how a point mutation in the G6PD gene can be identified by amplifying its entire coding region, using for template as little as 10 μg of the patient's genomic DNA.

MATERIALS AND METHODS

Subjects

The patient was a 4-year-old Greek boy who presented with severe neonatal jaundice, followed by chronic anaemia, requiring multiple blood transfusions. A diagnosis of chronic non-spherocytic haemolytic anaemia associated with G6PD deficiency was made. The patient's mother had no clinical abnormality. Biochemical data have shown that this patient has a new G6PD variant that has been designated 'G6PD Harilaou' [10].

Cultures and DNA extraction

Monolayer fibroblast cultures were grown using Dulbecco's modified Eagle's medium (Gibco) with 10% foetal-calf serum (Sera-labs) at 37 °C in a humidified atmosphere of CO_2/air (1:19). Cells were harvested and DNA was extracted using a method described by Sykes [11].

DNA amplification

Eight regions of the G6PD gene, covering all the coding exons, were amplified by the PCR reaction [9] (see Fig. 1). In order to design the PCR primers required we needed sequence information on G6PD introns. Some of this information has been published by Martini *et al.* [2] and by Viglietto *et al.* [12]. Additional data have been obtained in this laboratory (T. Vulliamy, unpublished work) from human genomic G6PD clones previously described [2]. The sequence of each of the 16 oligonucleotides (synthesized on a Biosearch Cyclone instrument) is shown in Table 1, and their position relative to the coding regions of the gene is seen in Fig. 1. PCR reactions were performed using *Thermus aquaticus* DNA polymerase (*Taq* polymerase) supplied by Perkin Elmer Cetus, according to the manufacturer's recommendations. Each reaction included 1.0 μg of genomic DNA, 0.5 μg of each amplimer and 1.0 unit of enzyme in a final volume of 100 μl. Temperature cycling was automated using a home-made apparatus previously described [13]. Each cycle consisted of con-

Abbreviations used: PCR, polymerase chain reaction; G6PD, glucose-6-phosphate dehydrogenase.

* On leave from Dipartimento di Pediatria, II Facolta di Medicina e Chirurgia, Universita degli Studi di Napoli, Napoli, Italy.
† To whom correspondence and reprint requests should be sent.

Table 1. Sequences of the amplimers and sizes of the fragments that they amplify

Exons amplified	Amplimer symbols and sequences	Expected amplified fragment size (bp)
2	δ 5′ CTCTAGAAAGGGGCTAACTTCTCAA 3′ φ 5′ GGAATTCCTGGCTTTTAAGATTGGG 3′	241
3+4	η 5′ CAGCCACTTCTAACCAC 3′ P_4 5′ AACCAGGCTGGGGGAG 3′	308
5	A 5′ AAGAGAGGGGCTGACAT 3′ M 5′ AGGCGGGAAGGGAGGGC 3′	301
6+7	B 5′ ACTCCCCGAAGAGGGGT 3′ J 5′ CCAGCCTCCCAGGAGAGA 3′	542
8	L 5′ GGAGCTAAGGCGAGCTC 3′ I 5′ CATGCTCTTGGGGACTG 3′	164
9	D 5′ CAAGGAGCCCATTCTCT 3′ R 5′ TGCCTTGCTGGGCCTCG 3′	253
10+11	E 5′ CTGAGAGAGCTGGTGCT 3′ Y 5′ CATAGCCCACAGGTATGC 3′	497
11+12+13	F 5′ CTTCAACCCCGAGGAGT 3′ G 5′ GGGAAGGAGGGTGGCCGTG 3′	657

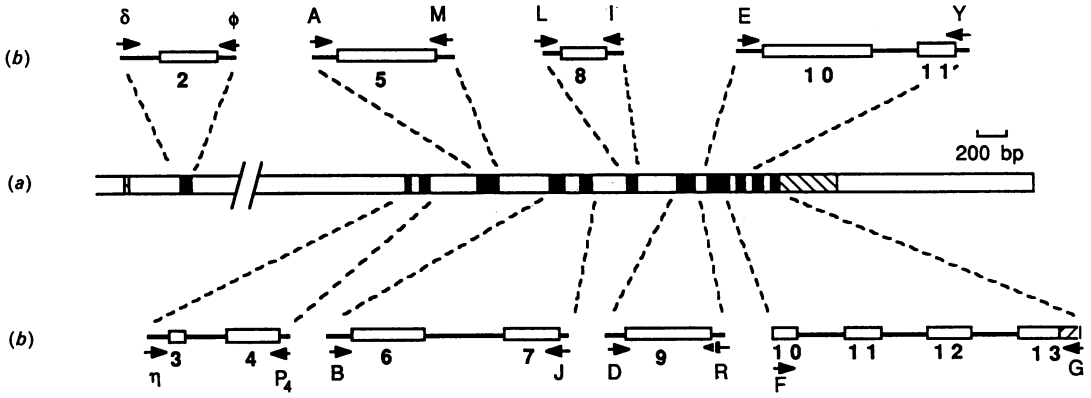

Fig. 1. PCR amplification of the human G6PD gene

(*a*) Schematic representation of the exon/intron organization. Filled segments represent exons (black = coding sequences; hatched = non-coding sequences) and open bars represent introns. A break between the second and third exons indicates the 11 kb second intron. (*b*) Expanded views of the eight amplified segments of the gene. Black bars are introns and open boxes are exons. Exons are numbered according to Martini *et al.* [2]. Letters denote individual oligonucleotide amplimers; arrowheads show the direction of primed DNA synthesis. Sequences of each oligonucleotide and the sizes of each amplified fragment are summarized in Table 1.

secutive 2 min incubations at 95, 48 and 67 °C. After 30 cycles, a 10 μl portion of each reaction mixture was analysed by electrophoresis on a 1.5%-agarose gel. This gel was Southern-blotted, and the filter (Amersham Hybond) was hybridized to a radioactively labelled G6PD cDNA probe (a 1.8 kb *Bam*HI–*Xho*I fragment isolated from the clone pGd.T5B [14]).

Subcloning and sequencing

The remaining 90 μl of the amplification reactions were phenol-extracted, and the DNA was purified through Sephacryl S-300 spinning columns. Portions (7.0 μl) of the eluates were treated with T4 polynucleotide kinase in a final reaction volume of 10 μl. After inactivation of the kinase by heating, 4.0, 2.0 and 1.0 μl portions of these reaction mixtures were ligated to 20 ng of an M13mp18 vector (cleaved with *Sma*I and dephosphorylated) in 10 μl ligation reactions. The recombinant DNA was used to transform competent JM101 *Escherichia coli* cells. Plaques which hybridized to the G6PD cDNA probe were used to prepare single-stranded DNA, and this was sequenced using the Sequenase system (USB, Cleveland, OH, U.S.A.). Both the M13 universal primer and the set of 16 amplimers were used as primers (see Fig. 1).

*Bal*I analysis of genomic DNA

Portions (10 μg) of genomic DNA were digested with 10 units of *Bal*I restriction enzyme (Amersham) at 37 °C for 48 h using buffer and conditions recommended by the manufacturer. The digests were gel-electrophoresed and then Southern-transferred to a nylon membrane. This membrane was hybridized to a probe consisting of a 264 bp fragment obtained by *Pst*I digestion of pGd. T5B. This fragment includes the 3′ end of exon 6, all of exon 7 and the 5′ end of exon 8.

RESULTS

PCR amplification

Eight pairs of oligonucleotide primers were used in PCR reactions to amplify sections of the G6PD gene from genomic DNA. Portions of the reaction mixtures were analysed by agarose-gel electrophoresis (Fig. 2*a*). Lanes 5 and 8 have single

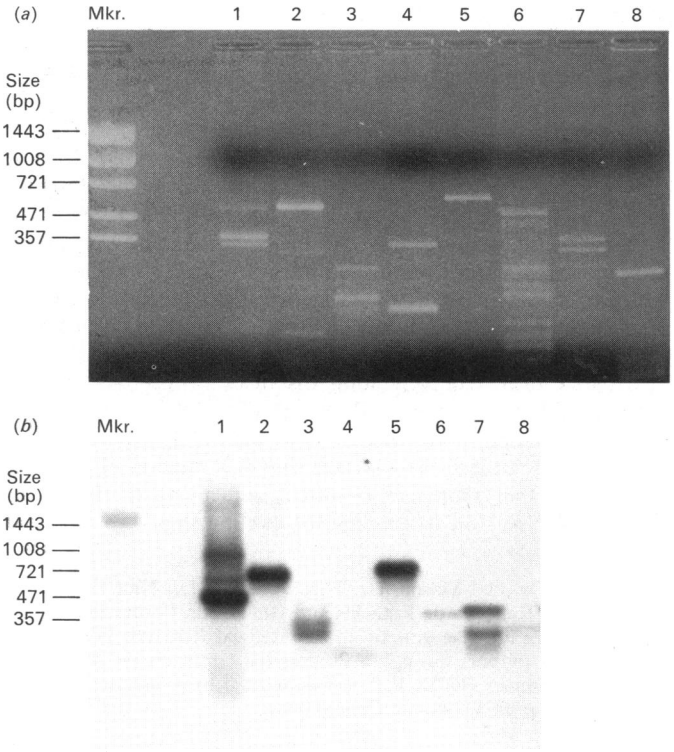

Fig. 2. Analysis of the PCR products

(*a*) Aliquots of the eight amplification reactions were applied to a 1.5%-agarose gel, which was stained with ethidium bromide. Lane 1, exons 10, 11, 12 and 13; lane 2, exons 10 and 11; lane 3, exon 9; lane 4, exon 8; lane 5, exons 6 and 7; lane 6, exon 5; lane 7, exons 3 and 4; lane 8, exon 2. The marker lane ('Mkr.') contains pEMBL8 DNA digested with *Taq*I. (*b*) Autoradiograph of a Southern blot from the gel shown in (*a*). The filter was hybridized to a 1.8 kb cDNA probe that extends from a *Bam*HI site in exon 2 to an *Xho*I site in the 3′ non-coding sequence. Lane numbers correspond to those in (*a*).

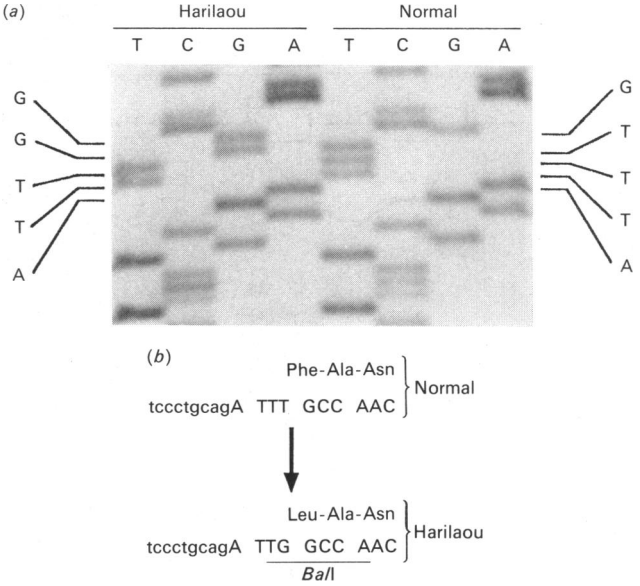

Fig. 3. Point mutation in the G6PD Harilaou gene

(*a*) Sequencing gel obtained using oligonucleotide B (Fig. 1 and Table 1) as primer on an M13 clone obtained from the amplified fragment containing exons 6 and 7. Sequencing templates were M13 single-stranded DNAs obtained from normal (G6PD B) and G6PD Harilaou clones. (*b*) Nucleotide sequences at the 5′ boundary of exon 7 in normal G6PD B and in G6PD Harilaou. Lower-case letters denote intron sequence; upper-case letters denote exon sequences and are divided into the coding triplets. Below the Harilaou sequence the new *Bal*I site created by the point mutation is shown. The Phe → Leu replacement is seen to affect the first amino acid encoded by exon 7.

bands of the predicted size. Lanes 2 and 4 have more than one band, but the band of the predicted size is the major species. In lanes 1, 3, 6 and 7, multiple bands are present, and the band of predicted size is a minor species. In order to confirm that all the reactions had amplified the desired G6PD sequences, this gel was Southern-blotted, and the filter was hybridized to a G6PD cDNA probe (Fig. 2*b*). In all tracks, at least one band hybridized to the probe. Only a single band hybridized in lanes 2, 4, 5, 6 and 8. We infer that the extra bands in lanes 2 and 4 probably result from the amplimers priming at sites outside the G6PD gene. In each of lanes 1, 3 and 7, several bands are seen on the autoradiograph (Fig. 2*b*), one of which is of the predicted size. Possibly these extra bands represent the single-stranded forms of the amplified fragments (as shown in another system by Gyllensten & Erlich [15]); alternatively the amplimers may have annealed to similar sequences within the G6PD gene itself.

Subcloning and sequencing

DNA from all amplification reactions was cloned into the M13mp19 phage vector, and the plaques were tested by filter hybridization with the G6PD cDNA probe. We found considerable variation in the number of G6PD recombinants obtained from one sample to another. In general the proportion of positive plaques was higher in the cases where fewer 'spurious' bands were observed (see above). Upon sequencing, we also noted that certain fragments were cloned preferentially in one of the two possible orientations. All amplified fragments were sequenced on each strand and from at least two different clones. The exon sequences were then compared with the published cDNA sequence of the normal G6PD B gene [14]. The only difference found was one base substitution in exon 7: T → G at position 648 (see Fig. 3). This mutation predicts an amino acid substitution of leucine for phenylalanine at position 216 in the protein.

Proof of mutation in genomic DNA

We were able to verify the presence of the G6PD Harilaou mutation in genomic DNA because it creates a recognition site for the restriction enzyme *Bal*I (see Fig. 4). In normal DNA a single fragment of 1.02 kb is detected. In DNA from the patient's mother, three bands were seen: one was normal and the other two were about 0.56 and 0.46 kb, confirming that she is a heterozygote for the Harilaou mutation.

DISCUSSION

The G6PD Harilaou mutation

The new mutation found in our patient provides further proof for the genetic heterogeneity of G6PD deficiency at the molecular level. We do not know how enzyme deficiency is caused by this particular amino acid substitution, but we note that it is only 11 residues away from the lysine-205 residue essential for enzyme activity [16] and which may be at, or near, the glucose 6-phosphate-binding site [17]. We also note that phenylalanine-216 is a conserved residue between *Drosophila* and human G6PD [18].

This is only one of few amino acid replacements in G6PD associated with chronic non-spherocytic haemolytic anaemia and the one that causes the most severe loss of enzyme activity

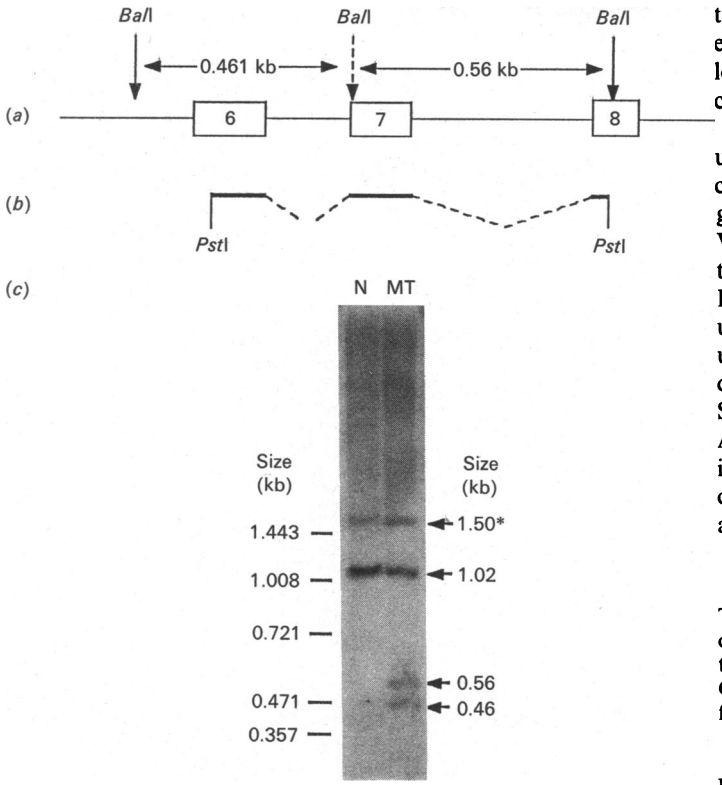

Fig. 4. The G6PD Harilaou mutation at the genomic level

(*a*) Schematic representation of G6PD exons 6, 7 and 8. The central *BalI* site (dotted arrow) is the location of the Harilaou mutation. The size of the 5′ fragment in the Harilaou allele is 461 bp by sequence analysis; the sizes of the 3′ fragment and the normal B fragment are 0.56 and 1.02 kb respectively (from [2]). (*b*) cDNA *PstI* fragment used to probe genomic Southern blots. Dotted lines denote the splicing of intron sequences. (*c*) Southern-blot analysis of genomic DNA digested with *BalI*. Portions (10 μg) of DNA from the mother of the patient (MT) and from a normal male subject (N) were digested with *BalI* and applied to a 1.5%-agarose gel. After transfer, the filter was hybridized with the 264 bp *PstI* cDNA probe. On the left of the autoradiograph are shown the positions of DNA markers of exactly known size (pEMBL8 digested with *TaqI*). On the right hand side are the sizes of the genomic *BalI* fragments as measured by graphic interpolation. As seen from the map, the 1.02, 0.56 and 0.46 kb fragments correspond to those expected; the 1.50 kb fragment (*) must result from partial digestion (a *BalI* site is known to exist about 0.5 kb upstream of that shown in intron 5). The mother has both the normal band (marked 1.02 kb) and the two extra bands expected; therefore she is a heterozygote.

in nucleated cells (less than 5% of normal activity in fibroblasts [10]). It is also the first example of a transversion mutation in this gene, in which over ten transition mutations have already been reported [3–8].

Use of the PCR to identify mutations

Here we describe the use of a PCR-based technique to clone and sequence rapidly all of the coding exons of a mutant G6PD gene. It has been stated that the use of *Taq* polymerase may involve an error rate of 0.25% [9] in the amplified product. In order to avoid being misled by such errors, for each fragment we have sequenced several M13 clones in both orientations. In fact, we have found only a single point mutation, and its existence was confirmed in genomic DNA.

The variability we have observed in the number of fragments produced in our reactions depends presumably on the design of the oligonucleotide amplimers. It is likely that the occurrence of extra amplified fragments could be reduced by synthesizing longer oligonucleotides or by further optimizing the PCR conditions.

Our data show that amplification of coding sequences of genes using the PCR technique is a powerful alternative to the conventional approach of constructing libraries in the study of genetic abnormalities. This approach has been used recently by Vrieling *et al.* [19] to amplify the entire mouse hypoxanthine:guanine phosphoribosyltransferase coding region using RNA as template. Here we have shown that this strategy can be used successfully with genomic DNA; we regard it as particularly useful for genes such as G6PD that have small coding exons distributed over relatively long distances of genomic DNA. Small quantities of genomic DNA are sufficient for the reactions. Amplification and subcloning are rapid once temperature cycling is automated. In fact, we envisage that this method of analysis could be further simplified by direct sequencing of the PCR-amplified DNA, thus by-passing the cloning step.

We thank Dr. Phil Mason, Dr. Tom Vulliamy, Dr. Nick Cross, Dr. Tim Cox, Dr. Raffaella de Franchis and Dr. Letizia Foroni for helpful discussions. We are also grateful to the Medical Research Council for their financial support through a Programme Grant on the Molecular Genetics of Human G6PD. V.P. was supported by a research contract from the European Economic Community.

REFERENCES

1. Luzzatto, L. & Mehta, A. (1989) in The Metabolic Basis of Inherited Disease, 6th edn. (Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D., eds.), pp. 2237–2265, McGraw-Hill, New York
2. Martini, G., Toniolo, D., Vulliamy, T. J., Luzzatto, L., Dono, R., Viglietto, G., Paonessa, G., D'Urso, M. & Persico, M. G. (1986) EMBO J. **5**, 1849–1855
3. Takizawa, T., Miwa, J. & Yoshida, A. (1987) Genomics **1**, 228–231
4. Hirono, A. & Beutler, E. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 3951–3954
5. Vulliamy, T. J., D'Urso, M., Battistuzzi, G., Estrada, M., Foulkes, N. S., Martini, G., Calabro, V., Poggi, V., Giordano, R., Town, M., Luzzatto, L. & Persico, M. G. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 5171–5175
6. DeVita, G., Alcalay, M., Sampietro, M., Cappellini, N., Fiorelli, G. & Toniolo, D. (1989) Am. J. Hum. Genet. **44**, 233–240
7. Beutler, E. (1989) Blood **73**, 1397–1401
8. Vulliamy, T. J., Wanachiwanawin, W., Mason, P. J. & Luzzatto, L. (1989) Nucleic Acids Res. **17**, 5868
9. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) Science **239**, 487–491
10. Town, M., Athanasiou-Metaxa, M. & Luzzatto, L. (1990) Somat. Cell Mol. Genet. **16**, 97–108
11. Sykes, B. C. (1983) Lancet **ii**, 787–788
12. Viglietto, G., Montanaro, V., Calabro, V., Vallone, D., D'Urso, M., Persico, M. G. & Battistuzzi, G. (1990) Ann. Hum. Genet. **54**, 1–15
13. Foulkes, N. S., Pandolfi de Rinaldis, P. P., Macdonnell, J., Cross, N. C. P. & Luzzatto, L. (1988) Nucleic Acids Res. **16**, 5687–5688
14. Persico, M. G., Viglietto, G., Martini, G., Toniolo, D., Paonessa, G., Moscatelli, C., Dono, R., Vulliamy, T. J., Luzzatto, L. & D'Urso, M. (1986) Nucleic Acids Res. **14**, 2511–2522, 7822
15. Gyllensten, U. B. & Erlich, H. A. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 7652–7656
16. Mason, P. J., Vulliamy, T. J., Foulkes, N. S., Town, M., Haidar, B. & Luzzatto, L. (1988) Eur. J. Biochem. **178**, 109–113
17. Camardella, L., Caruso, C., Rutigliano, B., Romano, M., DiPrisco, G. & Descalzi-Cancedda, F. (1987) Eur. J. Biochem. **171**, 485–489
18. Fouts, D., Ganguly, R., Gutierrez, A. G., Lucchesi, J. C. & Manning, J. E. (1988) Gene **63**, 261–275
19. Vrieling, H., Simons, J. W. & Van Zeeland, A. A. (1988) Mutat. Res. **198**, 107–113

Received 6 February 1990/24 April 1990; accepted 3 May 1990