Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 5171–5175, July 1988
Genetics

Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia

(enzymopathy/genetic variants/human genetics/cloning of mutants)

T. J. VULLIAMY*, M. D'URSO†, G. BATTISTUZZI‡, M. ESTRADA§, N. S. FOULKES*, G. MARTINI†, V. CALABRO‡, V. POGGI*, R. GIORDANO†¶, M. TOWN*, L. LUZZATTO*, AND M. G. PERSICO†

*Department of Haematology, Royal Postgraduate Medical School, Du Cane Road, London W12 OHS, Great Britain; †International Institute of Genetics and Biophysics, CP 3061, Naples, Italy; ‡Department of Genetics, University of Naples, via Mezzocannone, Naples, Italy; and §Instituto de Hematologia y Immunologia, Havana, Cuba

Communicated by Y. W. Kan, April 4, 1988 (received for review January 14, 1988)

ABSTRACT **Glucose-6-phosphate dehydrogenase (G6PD; EC 1.1.1.49) deficiency is a common genetic abnormality affecting an estimated 400 million people worldwide. Clinical and biochemical analyses have identified many variants exhibiting a range of phenotypes, which have been well characterized from the hematological point of view. However, until now, their precise molecular basis has remained unknown. We have cloned and sequenced seven mutant G6PD alleles. In the nondeficient polymorphic African variant G6PD A we have found a single point mutation. The other six mutants investigated were all associated with enzyme deficiency. In one of the commonest, G6PD Mediterranean, which is associated with favism among other clinical manifestations, a single amino acid replacement was found (serine → phenylalanine): it must be responsible for the decreased stability and the reduced catalytic efficiency of this enzyme. Single point mutations were also found in G6PD Metaponto (Southern Italy) and in G6PD Ilesha (Nigeria), which are asymptomatic, and in G6PD Chatham, which was observed in an Indian boy with neonatal jaundice. In G6PD "Matera," which is now known to be the same as G6PD A−, two separate point mutations were found, one of which is the same as in G6PD A. In G6PD Santiago, a *de novo* mutation (glycine → arginine) is associated with severe chronic hemolytic anemia. The mutations observed show a striking predominance of C → T transitions, with CG doublets involved in four of seven cases. Thus, diverse point mutations may account largely for the phenotypic heterogeneity of G6PD deficiency.**

Glucose-6-phosphate dehydrogenase (G6PD; EC 1.1.1.49) deficiency is a genetic abnormality associated with a range of clinical conditions (1). Some subjects are asymptomatic, whereas others suffer from neonatal jaundice, acute hemolytic anemia, or severe chronic nonspherocytic hemolytic anemia. Over 300 G6PD variants have been described by using clinical and biochemical criteria (2) and at least 100 of these have polymorphic frequencies in various populations (3). This makes G6PD deficiency probably the most common human enzymopathy, affecting an estimated 400 million people worldwide. There is abundant evidence that this widespread polymorphism results from relative resistance of heterozygotes to *Plasmodium falciparum* malaria (4–6). The fact that a large number of different deficient variants have been characterized suggests that a considerable variety of structural changes can cause abnormal enzyme activity.

G6PD is a housekeeping enzyme, present in all species so far tested (3). It is produced only at low levels in all cells and therefore purification of sufficient quantities for protein sequence analysis has been very difficult, or impossible in cases of severe deficiency. This has prevented the identification of the specific amino acid changes responsible for altered phenotypes. Recently, the gene that encodes human G6PD, which is located on chromosome Xq28, has been cloned and the coding sequence was determined (7–9), enabling us to study the molecular basis of G6PD deficiency. In this paper we describe the cloning and sequencing of seven mutant G6PD alleles, six of them from deficient subjects. We show that the wide range of clinical phenotypes associated with G6PD deficiency has arisen from different point mutations in the coding sequence of the *G6PD* gene.

MATERIALS AND METHODS

Biochemical Characterization. Of the seven G6PD variants described here, three have been previously published. For the other four, the following biochemical properties were studied by using methods recommended by the World Health Organization (WHO) (10): G6PD activity in erythrocytes, electrophoretic mobility in Tris/borate/EDTA buffer (pH 8.9), K_m^{G6P}, 2-deoxyglucose 6-phosphate % rate, K_m^{NADP}, and the thermostability at 56°C. K_i^{NADPH} was determined as in ref. 11, and the elution profile from DEAE-Sephadex was determined as in ref. 12. The K_m^{G6P} for G6PD Chatham was measured under the conditions described in ref. 13, and the value was normalized to WHO conditions by using the ratio of values obtained by the two methods with G6PD B.

Preparation and Screening of Genomic Libraries. DNA was prepared from whole blood or Epstein–Barr virus-transformed lymphocytes (14). Phage λ genomic libraries were constructed from this DNA after it had been digested with *Hin*dIII or *Eco*RI and size-fractionated on 0.5% agarose gels or 10–40% sucrose gradients. To obtain the entire G6PD coding sequence, two independent phage clones were isolated (see Fig. 1). (*i*) Exons I and II were cloned on a 12.5-kilobase (kb) *Eco*RI fragment into λGTWes or λEMBL4 (15). (*ii*) Exons III–XIII were cloned on either an 8.5-kb partial *Eco*RI fragment into λGTWes or a 17-kb *Hin*dIII fragment into λ2001 (16). λGTWes *Eco*RI arms were obtained from Bethesda Research Laboratories; λEMBL4 *Eco*RI arms were prepared by digesting λEMBL4 DNA with *Eco*RI and *Bam*HI and then purifying the arms by passage through Sephacryl S-300 columns followed by isopropyl alcohol precipitation. λ2001 *Hin*dIII arms were prepared in the same way, except that the DNA was digested with *Hin*dIII and

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.

Abbreviation: G6PD, glucose-6-phosphate dehydrogenase.

¶On leave of absence from Istituto Biochimica delle Proteine ed Enzimologia, Naples, Italy.

*Eco*RI. Ligated material was packaged by using Gigapack extracts (Northumbria Biologicals, U.K.) or freeze-thaw lysates and sonicated extracts (17). The libraries were screened by standard Benton–Davis filter hybridization (18) using the appropriate G6PD probes labeled by nick-translation (Amersham).

Sequencing. DNA from plaque-purified phage was subcloned into M13 phage for Sanger dideoxy sequencing (19) or into pUC18 for dideoxy sequencing of denatured plasmid DNA (20). Sequencing reactions were primed by using a set of oligonucleotides homologous to intron sequences near the intron–exon junctions (see Fig. 1). Exons I and II were subcloned on either a 2-kb *Pst* I fragment or a 6-kb *Hin*dIII/*Eco*RI fragment, whereas for exons III–XIII the 3.5-kb and 5-kb *Eco*RI fragments were subcloned.

Southern Blotting. Analysis of *Pst* I and *Hin*fI digestions of genomic DNA prepared from various individuals was carried out by Southern hybridization using conventional methods (17).

RESULTS

The G6PD variants studied are representative of a wide range of clinical and biochemical features. G6PD A is a nondeficient variant, widely distributed in Africa (21), with polymorphic gene frequencies reaching 0.25 in several populations (22). G6PD Mediterranean is one of the best-studied G6PD variants (23); it is polymorphic and is associated with acute hemolytic anemia, including favism. G6PD Ilesha is a sporadic variant from West Africa, found in a patient with sickle cell anemia (24). G6PD Metaponto is asymptomatic; since it was encountered in two unrelated people in Lucania (Southern Italy), it may be polymorphic in that region. G6PD Chatham was identified in a boy of Indian ancestry living in England who has had neonatal jaundice; its frequency in the Indian population as a whole has not been established. G6PD "Matera," which is in fact probably G6PD A− (see *Discussion*), is associated with favism; it was encountered in Lucania and it is polymorphic. G6PD Santiago de Cuba is a sporadic variant (see below) found in a boy from Cuba with chronic nonspherocytic hemolytic anemia.

The biochemical features of the G6PD variants we have investigated include varying degrees of deficiency with normal or abnormal electrophoretic mobility and, in some cases, distinctive changes in thermostability and enzyme kinetics (Table 1). For example, G6PD Mediterranean has a significantly increased affinity for G6P; G6PD Chatham has a slightly decreased affinity for G6P and a significantly reduced *in vitro* thermostability; G6PD Santiago de Cuba also has reduced *in vitro* thermostability and a markedly decreased affinity for NADP.

To identify changes in the G6PD protein responsible for these altered properties, we have obtained genomic clones of the *G6PD* gene from seven male individuals with variant enzyme. Sequencing of G6PD-positive clones was achieved by using a set of custom-designed oligonucleotides (Fig. 1); these permit the rapid sequencing of different clones and will allow the amplification of specific exons with the polymerase chain reaction technique. Comparison of the coding sequence of these G6PD clones with the previously published sequence of normal G6PD B revealed unique base changes (Table 2).

In six cases, single point mutations were identified. With the exception of G6PD A, we can infer these to be the cause of G6PD deficiency. In the seventh case, G6PD "Matera," two separate point mutations are present. One of these is the same as that found in the nondeficient variant, G6PD A, and so we infer that the other is responsible for enzyme deficiency. Six of the seven mutations are C → T transitions on either the coding or the noncoding strand, and in four of these the C precedes a G in a CG doublet. The amino acids changed by these mutations are spread throughout the G6PD protein sequence, as shown in Fig. 2. In each case, the inferred amino acid substitution is consistent with the observed change in electrophoretic mobility of the variant enzyme by comparison with the normal.

The G → A mutation in G6PD Santiago de Cuba creates a *Pst* I site that is not present in the normal G6PD B gene. This has enabled us to confirm that this mutation is present in the genomic DNA of the individual from which the library was prepared. Southern blot analysis of his DNA, using a probe for the 3′ end of the *G6PD* gene, revealed a 1.8-kb *Pst* I fragment that has not been seen in >100 similarly analyzed DNA samples, including one from the mother of the propositus (Fig. 3). By using a minisatellite probe (30), there is no suggestion that she is not the real mother (data not shown). Therefore we conclude that G6PD Santiago de Cuba must have arisen through a new mutation that took place either in the propositus early in embryogenesis or in his mother's germ cell line.

In G6PD Ilesha the G → A mutation destroys an *Hin*fI site present in the normal sequence. As above, Southern hybridization analysis of *Hin*fI digests of genomic DNA from this variant confirms the mutation (Fig. 4). In this case, the mother's DNA was not available, but we know from pedigree analysis (31) that this mutation, although sporadic, is not new in the propositus.

DISCUSSION

Identification of mutations that cause pathology is of interest not only in the understanding of genetic disease but also for the information that can be provided concerning the corre-

Table 1. Hematological and biochemical features of G6PD variants investigated

G6PD variant	Clinical expression	Enzyme activity in RBC, % of normal	Electrophoretic mobility, % of normal	K_m^{G6P}, μM	2dG6P, % rate	K_m^{NADP}, μM	K_i^{NADPH}, μM	DEAE-Sephadex peak at KCl, mM	Thermostability
B	None	100*	100	50	<4	3–5	9	230 ± 1	N
A	None	84	110	50	<4	3–5	7	220	N
Metaponto	None	14–39	90	47	5	3	12	231 ± 4	N
Ilesha	None	25	75	60	—	—	—	—	—
Chatham	NNJ	<2	100	60	14	—	—	—	↓
Mediterranean	AHA	0–7	100	23	50	1.2–1.6	16.0	230 ± 3	↓
Matera (A−)	AHA	10–23	110	47	7	2	13	230	N
Santiago de Cuba	CNSHA	5	80	50	<2	43	—	260	↓

Variants are listed in order of increasing severity of clinical expression. All donors were males. G6PD B is the human normal "wild type." Properties of published variants are from ref. 2 except for the following: data for G6PD Ilesha, ref. 24; activity quoted for G6PD A, ref. 25; 2-deoxyglucose 6-phosphate (2dG6P) % rate for Mediterranean, ref. 26; elution peaks off DEAE-Sephadex, ref. 12. NNJ, neonatal jaundice; AHA, acute hemolytic anemia; CNSHA, chronic nonspherocytic hemolytic anemia; RBC, erythrocytes.
*Normal G6PD activity in our laboratory at 30°C is 9 ± 3 international units/g of Hb.

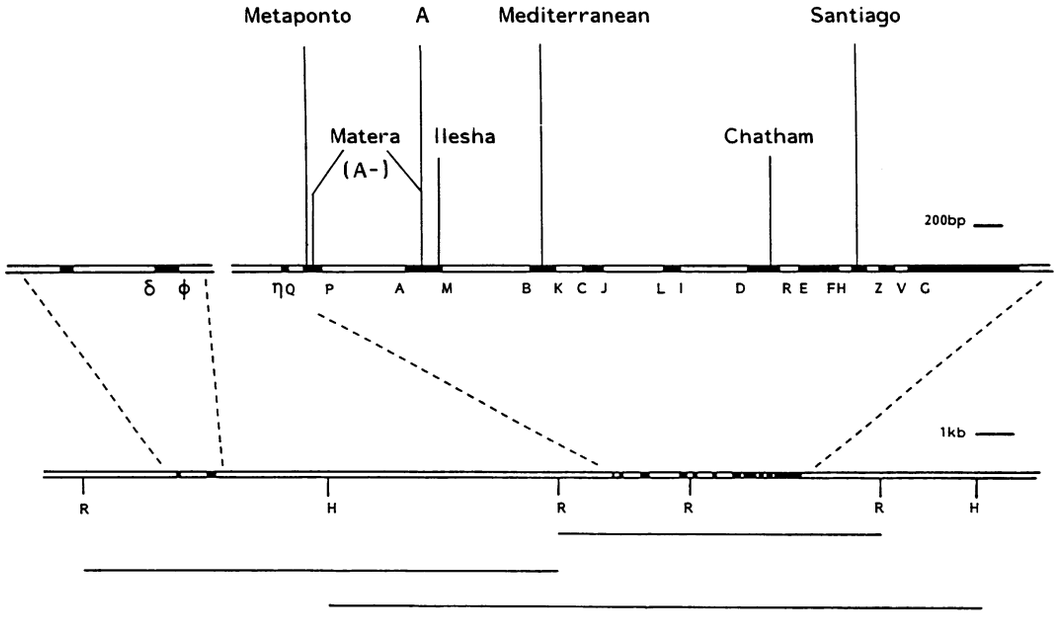

FIG. 1. Sketch of the human *G6PD* gene showing the positions of mutations found in seven genetic variants and our cloning and sequencing strategies. From the bottom upward: the horizontal lines represent the extent of the different phage clones we have used to isolate the *G6PD* gene from different individuals. Above is a map of the *G6PD* gene showing *Eco*RI sites (R) and *Hin*dIII sites (H), and the 13 exons are blacked in (from ref. 8). The two regions of coding sequence are expanded above, again with the exons in black and introns in white. Letters below the sketch show the locations of the 21 custom-synthesized oligonucleotides that we have used to prime sequence reactions into the coding sequences. The vertical bars and names above indicate the positions of the mutations found in the respective variants. bp, Base pairs.

sponding normal function of the respective gene product. The biochemistry of G6PD and the clinical manifestations of its deficiency have been studied extensively (2, 3), yet little is known about the protein at the molecular level. Recently, the primary sequence for the protein was determined through the isolation of full-length cDNA clones (7). Here we identify different point mutations in this coding sequence that are the cause of enzyme deficiency in different individuals. They provide direct proof of the genetic heterogeneity of G6PD deficiency, previously inferred from biochemical data. We demonstrate that different point mutations give rise to different clinical phenotypes. In addition these data provide an insight into which amino acids are important for G6PD enzyme stability and function.

Of course, assignment of specific function to different amino acids within the protein awaits elucidation of its three-dimensional structure, and this set of variants is too small for us to attempt an analysis of how amino acid replacements cause a reduction of G6PD activity by impairing catalytic function, making the protein unstable, or both. However, with respect to catalytic function, the variant with the most abnormal K_m^{G6P} (G6PD Mediterranean) has the amino acid replacement nearest lysine-205, which may be involved in G6P binding (28, 29). With respect to protein stability, the two variants with drastic amino acid changes (serine → phenylalanine in G6PD Mediterranean and glycine → arginine in G6PD Santiago de Cuba) have very low residual enzyme activity in circulating erythrocytes, presumably as a result of accelerated G6PD breakdown (23). They also have decreased *in vitro* thermostability, as does G6PD Chatham, where the amino acid replacement is alanine → threonine. The glycine → arginine replacement in G6PD Santiago de Cuba is located within a sequence of hydrophobic amino acids (32), which might make this replacement particularly damaging. Also, within our series of variants this is the only example that is associated with a severe clinical condition and it is a *de novo* mutation.

With respect to G6PD A, our data confirm the asparagine → aspartic acid amino acid replacement reported by Yoshida 20 years ago (33) and recently placed at position 142 (34). Although we give a position number of 126, we presume that these two different numbers represent the same amino acid due to a difference in the N-terminal portion (52 vs. 36 amino acids) of the two published protein sequences (7, 9). The N-terminal sequence shown here (Fig. 2) is that derived from

Table 2. Location of point mutations in seven different G6PD variants

G6PD variant	Exon number	Base position	Base change	Codon change	Amino acid position	Amino acid substitution
A	V	376	A → G	(G)AAT → GAT	126	Asn → Asp
Metaponto	IV	172	G → A	(G)GAT → AAT	58	Asp → Asn
Ilesha	V	466	G → A	(C)GAG → AAG	156	Glu → Lys
Chatham	IX	1003	G → A	(C)GCC → ACC	335	Ala → Thr
Mediterranean	VI	563	C → T	(C)TCC → TTC	188	Ser → Phe
Matera (A−)	IV	202	G → A	(C)GTG → ATG	68	Val → Met
	V	376	A → G	(G)AAT → GAT	126	Asn → Asp
Santiago de Cuba	XI	1339	G → A	(C)GGG → AGG	447	Gly → Arg

Base and amino acid position numbers refer to the revised published sequence (ref. 7, p. 7822). The letter in parentheses preceding the codon change is the nucleotide immediately 5′ on the coding strand; for Ilesha, Chatham, Matera, and Santiago de Cuba the mutated C on the noncoding strand is part of a CpG doublet. An additional mutation (1116, G → A) has recently been identified as being responsible for a *Pst* I restriction fragment length polymorphism in African populations (27). This mutation is in the coding sequence but it does not produce any amino acid substitution.

10 20
Met Ala Glu Gln Val Ala Leu Ser Arg Thr Gln Val Cys Gly Ile Leu Arg Glu Glu Leu
30 40
Phe Gln Gly Asp Ala Phe His Gln Ser Asp Thr His Ile Phe Ile Ile Met Gly Ala Ser
50 60
Gly Asp Leu Ala Lys Lys Lys Ile Tyr Pro Thr Ile Trp Trp Leu Phe Arg Asp Gly Leu
70 80
Leu Pro Glu Asn Thr Phe Ile Val Gly Tyr Ala Arg Ser Arg Leu Thr Val Ala Asp Ile
90 100
Arg Lys Gln Ser Glu Pro Phe Phe Lys Ala Thr Pro Glu Glu Lys Leu Lys Leu Glu Asp
110 120
Phe Phe Ala Arg Asn Ser Tyr Val Ala Gly Gln Tyr Asp Asp Ala Ala Ser Tyr Gln Arg
130 140
Leu Asn Ser His Met Asn Ala Leu His Leu Gly Ser Gln Ala Asn Arg Leu Phe Tyr Leu
150 160
Ala Leu Pro Pro Thr Val Tyr Glu Ala Val Thr Lys Asn Ile His Glu Ser Cys Met Ser
170 180
Gln Ile Gly Trp Asn Arg Ile Ile Val Glu Lys Pro Phe Gly Arg Asp Leu Gln Ser Ser
190 200
Asp Arg Leu Ser Asn His Ile Ser Ser Leu Phe Arg Glu Asp Gln Ile Tyr Arg Ile Asp
210 220
His Tyr Leu Gly <u>Lys</u> Glu Met Val Gln Asn Leu Met Val Leu Arg Phe Ala Asn Arg Ile
230 240
Phe Gly Pro Ile Trp Asn Arg Asp Asn Ile Ala Cys Val Ile Leu Thr Phe Lys Glu Pro
250 260
Phe Gly Thr Glu Gly Arg Gly Gly Tyr Phe Asp Glu Phe Gly Ile Ile Arg Asp Val Met
270 280
Gln Asn His Leu Leu Gln Met Leu Cys Leu Val Ala Met Glu Lys Pro Ala Ser Thr Asn
290 300
Ser Asp Asp Val Arg Asp Glu Lys Val Lys Val Leu Lys Cys Ile Ser Glu Val Gln Ala
310 320
Asn Asn Val Val Leu Gly Gln Tyr Val Gly Asn Pro Asp Gly Glu Gly Glu Ala Thr Lys
330 340
Gly Tyr Leu Asp Asp Pro Thr Val Pro Arg Gly Ser Thr Thr Ala Thr Phe Ala Ala Val
350 360
Val Leu Tyr Val Glu Asn Glu Arg Trp Asp Gly Val Pro Phe Ile Leu Arg Cys Gly Lys
370 380
Ala Leu Asn Glu Arg Lys Ala Glu Val Arg Leu Gln Phe His Asp Val Ala Gly Asp Ile
390 400
Phe His Gln Gln Cys Lys Arg Asn Glu Leu Val Ile Arg Val Gln Pro Asn Glu Ala Val
410 420
Tyr Thr Lys Met Met Thr Lys Lys Pro Gly Met Phe Phe Asn Pro Glu Glu Ser Glu Leu
430 440
Asp Leu Thr Tyr Gly Asn Arg Tyr Lys Asn Val Lys Leu Pro Asp Ala Tyr Glu Arg Leu
450 460
Ile Leu Asp Val Phe Cys Gly Ser Gln Met His Phe Val Arg Ser Asp Glu Leu Arg Glu
470 480
Ala Trp Arg Ile Phe Thr Pro Leu Leu His Gln Ile Glu Leu Glu Lys Pro Lys Pro Ile
490 500
Pro Tyr Ile Tyr Gly Ser Arg Gly Pro Thr Glu Ala Asp Glu Leu Met Lys Arg Val Gly
510
Phe Gln Tyr Glu Gly Thr Tyr Lys Trp Val Asn Pro His Lys Leu

FIG. 2. Amino acid sequence of G6PD and location of substitutions. Boxes show the positions of the different amino acids that are changed in the variants we have studied. The underlined lysine in position 205 is known to be reactive with pyridoxal phosphate and may be close to the G6P-binding site (28, 29). The sequence shown is the same as that published (ref. 7, p. 7822) except for glutamine at amino acid 11, which was histidine in the published sequence. This change has now been found in all individuals we have sequenced and is a sequencing error in the previous publication.

cDNA clones and now confirmed as a continuous sequence in genomic clones from seven different individuals, and we therefore believe it to be correct.

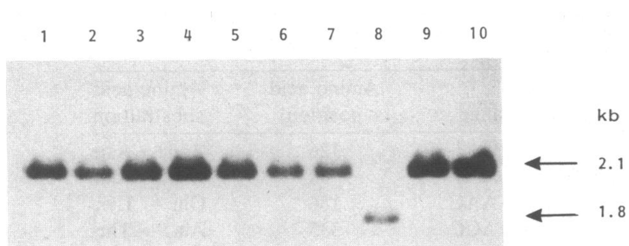

FIG. 3. G6PD Santiago de Cuba is a new mutation. Southern blot analysis was carried out by standard techniques on *Pst* I digests, using a probe to the 3′ end of the *G6PD* gene, p2.1 (ref. 27). The normal fragment of 2.1 kb is seen in five unrelated individuals (lanes 1–5), two maternal uncles (lanes 6 and 7), the mother (lane 9), and a sister (lane 10) of the individual with G6PD Santiago de Cuba. This individual (lane 8) shows a 1.8-kb fragment due to the presence of an extra *Pst* I site created by the new mutation.

The same mutation as in G6PD A is present in G6PD "Matera," along with another base change. The biochemical characteristics described here for G6PD "Matera" are very similar to those of G6PD A−. We now consider that G6PD

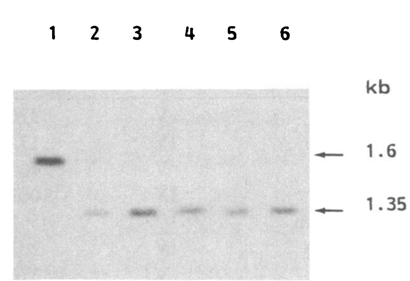

FIG. 4. G6PD Ilesha mutation is confirmed. Southern hybridization analysis with a G6PD cDNA probe to *Hin*fI digests of genomic DNA from G6PD Ilesha (lane 1) and five unrelated individuals (lanes 2–6) shows the expected change to a *Hin*fI fragment from 1.35 kb in normals to 1.6 kb in this variant due to a point mutation.

"Matera" is in fact A−, as we have learned recently that the amino acid replacements reported here for G6PD "Matera" are the same as those found in G6PD A− (35). The extent to which biochemically similar variants turn out to be genetically identical will become clear as more are defined at the molecular level. The finding of two mutational differences between G6PD B and G6PD A− did not come as a complete surprise, since this had been hypothesized some time ago on the grounds of biochemical and genetic considerations (13).

We note the striking predominance of C → T transitions in our series of mutations. Although, of course, the sample size is small, this finding is entirely consistent with previous observations on interspecies (36) and intraspecies (37) variation in globins and other genes. In four cases the mutated C residue is in a CpG doublet. Methylation and subsequent deamination of this C are thought to increase the probability of a C → T transition (38, 39), and there is growing evidence that this is indeed the case in other human genes as well (40).

The pattern of mutations causing human pathology has been extensively explored in a number of genes with cell-specific expression. For instance, with respect to globins, factor VIII, and steroid sulfatase there have been reports of point mutations and sizeable deletions (41–43). The latter are possible because complete loss of function of such genes is compatible with development in embryonic life and beyond. The findings reported here begin to define the pattern of mutations in a highly polymorphic housekeeping gene, in which complete loss of function would probably be lethal.

We thank Drs. G. Hayes and M. Cascone for supplying blood samples from their patients. We are grateful to Dr. Philip Mason, Dr. Letizia Foroni, and an anonymous reviewer for advice and help. Financial support was received from a Medical Research Council (England) Programme Grant; from the Consiglio Nazionale Ricerche (Italy) Progetti Finalizzati; and from a research contract from the European Economic Community.

1. Beutler, E. (1983) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. J., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), pp. 1629–1653.
2. Beutler, E. (1978) *Haemolytic Anaemia in Disorders of Red Cell Metabolism* (Plenum, London), pp. 23–167.
3. Luzzatto, L. & Battistuzzi, G. (1985) *Adv. Hum. Genet.* **14**, 217–329.
4. Allison, A. C. (1960) *Nature (London)* **186**, 531–532.
5. Motulsky, A. G. (1960) *Hum. Biol.* **32**, 28–32.
6. Luzzatto, L. (1979) *Blood* **54**, 961–976.
7. 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.
8. 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.
9. Takizawa, T., Huang, I.-Y., Ikuta, T. & Yoshida, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4157–4161.
10. Betke, K., Brewer, G. J., Kirkman, H. N., Luzzatto, L., Motulsky, A. G., Ramot, B. & Siniscalo, M. (1967) *W.H.O. Tech. Rep. Ser. 366.*
11. Luzzatto, L. (1967) *Biochim. Biophys. Acta* **146**, 18–25.
12. D'Urso, M., Battistuzzi, G. & Luzzatto, L. (1980) *Anal. Biochem.* **108**, 146–150.
13. Babalola, A. O. G., Beetlestone, J. G. & Luzzatto, L. (1976) *J. Biol. Chem.* **251**, 2992–3002.
14. Foroni, L., Foldi, J., Matutes, E., Catovsky, D., O'Connor, N. J., Baer, R., Forster, A., Rabbitts, T. H. & Luzzatto, L. (1987) *Br. J. Haematol.* **67**, 307–318.
15. Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170**, 827–842.
16. Karn, J., Matthes, H., Gait, M. J. & Brenner, S. (1984) *Gene* **32**, 217–224.
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
18. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–186.
19. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
20. Hattori, M., Hidaka, S. & Sakaki, Y. (1985) *Nucleic Acids Res.* **13**, 7813–7827.
21. Porter, I. H., Boyer, S. H., Watson-Williams, E. J., Adam, A., Szeinberg, A. & Siniscalco, M. (1964) *Lancet* **i**, 895–899.
22. Luzzatto, L. (1973) *Isr. J. Med. Sci.* **9**, 1181–1194.
23. Morelli, A., Benatti, U., Gaetani, G. F. & DeFlora, A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1979–1983.
24. Usanga, E. A., Bienzle, U., Cancedda, R., Fasuan, F. A., Ajayi, O. & Luzzatto, L. (1977) *Ann. Hum. Genet.* **40**, 279–286.
25. Battistuzzi, G., Esan, G. J. F., Fasuan, F. A., Modiano, G. & Luzzatto, L. (1977) *Am. J. Hum. Genet.* **29**, 31–36.
26. Ferraris, A. M., Giuntini, P., Galiano, S. & Goetani, G. F. (1981) *Am. J. Hum. Genet.* **33**, 307–313.
27. D'Urso, M., Luzzatto, L., Perroni, L., Ciccodicola, A., Gentile, G., Peluso, I., Persico, M. G., Pizella, T., Toniolo, D. & Vulliamy, T. J. (1988) *Am. J. Hum. Genet.* **42**, 735–740.
28. Jeffrey, J., Hobbs, L. & Jörnvall, H. (1985) *Biochemistry* **24**, 666–671.
29. Camardella, L., Caruso, C., Rutigliano, B., Romano, M., D'Prisoco, G. & Descalzicancedda, F. (1988) *Eur. J. Biochem.* **171**, 485–489.
30. Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) *Nature (London)* **314**, 67–73.
31. Luzzatto, L., Usanga, E. A., Bienzle, U., Esan, G. F. J. & Fasuan, F. A. (1979) *Science* **205**, 1418–1420.
32. Persico, M. G., Viglietto, G., Martini, G., Dono, R., D'Urso, M., Toniolo, D., Vulliamy, T. & Luzzatto, L. (1986) in *Glucose-6-Phosphate-Dehydrogenase*, eds. Yoshida, A. & Beutler, E. (Academic, New York), pp. 503–516.
33. Yoshida, A. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 835–841.
34. Takizawa, T. & Yoshida, A. (1987) *Am. J. Hum. Genet.* **41**, A241 (abstr.).
35. Hirono, A. & Beutler, E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3951–3954.
36. Vogel, F. & Kopun, M. J. (1977) *Mol. Evol.* **9**, 159–180.
37. Modiano, G., Battistuzzi, G. & Motulsky, A. G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1110–1114.
38. Bird, A. P. (1980) *Nucleic Acids Res.* **8**, 4091–4109.
39. Barker, D., Schafer, M. & White, R. (1984) *Cell* **36**, 131–138.
40. Youssoufian, H., Kazazian, H. H., Phillips, D. G., Aronis, S., Tsfitis, G., Brown, V. A. & Antonarakis, S. E. (1986) *Nature (London)* **324**, 380–382.
41. Weatherall, D. J. (1985) *The New Genetics and Clinical Practice* (Oxford Univ. Press, Oxford), pp. 66–86.
42. Youssoufian, H., Antonarakis, S. E., Aronis, S., Tsiftis, G., Phillips, D. G. & Kazazian, H. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3772–3776.
43. Ballabio, A., Parenti, G., Carrozzo, R., Sebastio, G., Andria, G., Buckle, V., Fraser, N., Craig, I., Rocchi, M., Romeo, G., Jobsis, A. C. & Persico, M. G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4519–4523.