

Human glucose-6-phosphate dehydrogenase: Primary structure and cDNA cloning

(oligonucleotide probe/amino acid sequence/X chromosome)

TAKENORI TAKIZAWA, I-YIH HUANG*, TOHRU IKUTA, AND AKIRA YOSHIDA†

Department of Biochemical Genetics, Beckman Research Institute of the City of Hope, Duarte, CA 91010

Communicated by Arno G. Motulsky, February 10, 1986

ABSTRACT The X-chromosome-linked glucose-6-phosphate dehydrogenase (D-glucose-6-phosphate:NADP⁺ oxidoreductase, EC 1.1.1.49) of humans and other mammals consists of a subunit with a molecular weight of about 58,000. The enzyme plays a key role in the generation of NADPH, particularly in matured erythrocytes, and the genetic deficiency of the enzyme is associated with chronic and drug- or food-induced hemolytic anemia in humans. The enzyme was purified to homogeneity from human erythrocytes. The complete amino acid sequence of the subunit, consisting of 531 amino acid residues, was determined by automated and manual Edman degradation of tryptic, chymotryptic, thermolytic, and cyanogen bromide peptides obtained from the enzyme. Based on the amino acid sequence data thus obtained, a 41-mer oligonucleotide with unique sequence was prepared. Two cDNA libraries constructed in phage λ gt11—i.e., a human liver cDNA library and a human hepatoma Li-7 cDNA library—were screened with the synthetic nucleotide probe. Two positive clones, λ G6PD-19 and λ G6PD-25, were obtained from the hepatoma library. λ G6PD-19 contained an insertion of 2.0 kilobase pairs (kbp), and encoded 204 amino acid residues that were completely compatible with the COOH-terminal portion of the enzyme. The insertion of the clone had a 3' noncoding region of 1.36 kbp. The other clone, λ G6PD-25, had an insertion of 1.8 kbp and encoded 362 amino acid residues of G6PD. Southern blot analysis of DNA samples obtained from cells with and without the human X chromosome indicated that the cDNA hybridizes with a sequence in the X chromosome.

Glucose-6-phosphate dehydrogenase (G6PD; D-glucose-6-phosphate:NADP⁺ oxidoreductase, EC 1.1.1.49) plays a key role in the production of ribose 5-phosphate and the generation of NADPH in the hexose monophosphate shunt. Since this pathway is the only NADPH-generation process in mature erythrocytes, which lack the citric acid cycle, a genetic deficiency of G6PD is often associated with adverse physiological effects. In fact, deficiency of G6PD is the most common genetic cause for chronic and drug- or food-induced hemolytic anemia in human beings. The variations found in G6PD deficiency are the most prevalent and heterogeneous of known human enzyme abnormalities. About 300 variants, which are distinguishable by electrophoretic mobility and kinetic properties, have been found (references in ref. 1). Several G6PD variants associated with erythrocyte enzyme deficiency exist in very high frequencies in certain populations of malaria-endemic areas.

The gene for G6PD is located at the q28 position of the X chromosome in human beings (2). G6PD has often been used as a marker for the study of pathogenesis of tumors and arteriosclerosis and for the study of mechanisms of X chromosome inactivation.

In addition to the X-linked G6PD, an autosomal isozyme, commonly designated hexose-6-phosphate dehydrogenase, which oxidizes other hexose 6-phosphates as well as glucose 6-phosphate and uses NAD as well as NADP as a coenzyme, exists in microsomes of humans and other mammals (3, 4). The existence of another G6PD isozyme in fetal brain has been suggested (5).

The protomer of G6PD in humans and other organisms has a molecular weight of about 58,000 (6).

For further study of molecular abnormalities of G6PD variants at the molecular level, determination of primary structure and cloning of cDNA for the enzyme would have particular importance. It has been reported that cDNA for human G6PD has been obtained (7). Thus far, however, compatibility between the nucleotide sequence of the cDNA cloned and the amino acid sequence of G6PD has not been established, and the nature of this cDNA clone is not clear. This paper reports G6PD cDNA whose nucleotide sequence completely matches the amino acid sequence of the X-linked human G6PD.

MATERIALS AND METHODS

Amino Acid Sequence Analysis of Human G6PD. G6PD was purified to homogeneity from out-dated and fresh human blood (total of about 1400 liters) by a slight modification of the previous method (8). Most of the blood came from blood banks in the Boston area, and about one-third was supplied by the Amsterdam Central Blood Bank. Considering the ethnic origins of people in these areas, more than 90% of the blood donors should be normal Gd B⁺, and the rest could be variant Gd A⁺ and Gd A⁻, with a smaller number Gd Mediterranean and other variants. The final enzyme preparation indicates a single enzyme band that corresponds to the normal B⁺ enzyme. Thus, the enzyme preparation represents normal human G6PD.

The enzyme protein was reduced, S-carboxymethylated, and digested by trypsin. The tryptic peptides were isolated by combinations of ion-exchange column chromatography, gel filtration, two-dimensional paper electrophoresis and chromatography, and reverse-phase HPLC. The S-carboxymethylated enzyme was also subjected to cyanogen bromide treatment. Cyanogen bromide peptides thus produced were separated by combinations of gel filtration, chromatography, and HPLC. Individual cyanogen bromide peptides were digested by trypsin, chymotrypsin, and thermolysin, and these peptide fragments were isolated. Amino acid sequences of isolated peptides were determined by a combination of automated and manual Edman degradation (9).

Oligonucleotide Probes. Synthetic oligonucleotides were synthesized by a solid-phase phosphotriester method. A

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: G6PD, glucose-6-phosphate dehydrogenase; kbp, kilobase pair(s).

*Present address: Smith-Kline Beckman, Palo Alto, CA 91010.

†To whom reprint requests should be addressed.

41-mer nucleotide with a unique sequence, which corresponded to the amino acid sequence of human G6PD, was used as a probe (Fig. 1). The synthesized probe was labeled at the 5' end with [γ - 32 P]ATP (5000 Ci/mmol; 1 Ci = 37 GBq; ICN) and T4 polynucleotide kinase (Bethesda Research Laboratories) by the standard method (10). Several relatively short mixed nucleotides (septadecamer to dodecamer) corresponding to the amino acid sequence were also prepared. These short probes were found to be not useful, due to less specific hybridization.

cDNA Library. The human liver cDNA library, constructed by inserting the cDNA copies of poly(A)⁺ mRNA from human liver into the *Eco*RI site of bacteriophage vector λ gt11 (11), was provided by S. L. C. Woo (Howard Hughes Medical Institute, Houston, TX). The human hepatoma cDNA library, constructed by inserting the cDNA copies from human hepatoma cell line Li-7 (12), was provided by J. R. DeWet (University of California, San Diego, CA). We found that the hepatoma cell line Li-7 contained about 30 times more G6PD than usual human livers—i.e., 1.1 mg and 0.037 mg of G6PD per g of soluble protein in these two tissues, respectively.

Screening of cDNA Libraries. Approximately 5×10^5 recombinant phage plaques from each of the two libraries were screened by hybridization with the 41-mer nucleotide probe. Briefly, $1\text{--}1.5 \times 10^4$ recombinant bacteriophage λ gt11 particles were plated on a 9-cm agar plate with *Escherichia coli* Y1090 and incubated at 42°C for 4 hr and at 37°C for 4 hr. Each plate was overlaid with a nitrocellulose filter (BA 85, 0.45 μ m; Schleicher & Schuell). The filter was treated with 0.5 M NaOH for 20 min, water for 5 min, 1 M Tris-HCl, pH 7.5, for 20 min, and $6 \times$ SSC ($1 \times$ SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.0) for 20 min, and baked for 2 hr at 70°C. The filter was treated with the prehybridization solution consisting of $6 \times$ SSC, $10 \times$ Denhardt's solution, 0.2% sodium dodecyl sulfate, and salmon sperm DNA (100 μ g/ml) at 60°C for 4 hr, and bound DNA was hybridized with the radiolabeled oligonucleotide probe (5×10^5 cpm/ml) in the prehybridization solution at 60°C for 16 hr. Hybridization-positive plaques, detected by autoradiograph, were picked up from the mother agar plate. Successive screenings were carried out, using fewer and fewer plaques at each step until well-isolated phage plaques were cloned. The libraries were also screened by hybridization with the shorter (17- to 20-mer) probes at lower temperature and with a specific rabbit antibody against human G6PD as a probe by the previous method (13).

Restriction Endonuclease Maps and DNA Sequence Analysis. Recombinant phage DNA was prepared from a large-scale liquid culture by a modification of the procedure described by Maniatis *et al.* (14). The inserted cDNA was separated from the *Eco*RI-digested λ DNA by gel electrophoresis followed by electroelution (15, 16). Restriction mapping of the isolated DNA was performed by single or double endonuclease digestion of the DNA. Restriction enzymes (Bethesda Research Laboratories and Boehringer Mannheim) were used under the conditions recommended by the suppliers.

The restriction fragments were subcloned in phage M13-mp18 and -mp19 (17, 18). Since *Acc* I exhibits degenerate

restriction specificity, producing several different termini, ends of the *Acc* I fragments of λ G6PD-19 were made blunt by treatment with the DNA polymerase Klenow fragment and inserted into the *Sma* I site of M13-mp18. The DNA sequence was determined by the dideoxynucleotide chain termination method (19).

Southern Hybridization Analysis. DNA samples from Chinese hamster and the human-Chinese hamster fibroblast cell line CF 74-5, which contains the human X chromosome but lacks other human chromosomes, were provided by T. Mohandas (Harbor UCLA Medical Center, Torrance, CA). Human DNA was prepared from fresh human blood by a previously described method (20). DNA samples, about 10 μ g each, were completely digested by *Eco*RI, *Pst* I, *Sst* I, *Sal* I, *Taq* I, *Msp* I, *Hind*III, *Sma* I, and *Bam*HI, and DNA fragments were separated by agarose gel electrophoresis and transferred onto a nitrocellulose filter as described by Southern (21). Hybridization was carried out in a solution containing 50% (vol/vol) formamide, $3 \times$ SSC, $1 \times$ Denhardt's solution, 5% dextran sulfate, and salmon sperm DNA (100 μ g/ml) at 42°C overnight, using 32 P-labeled G6PD cDNA (insert of λ G6PD-19) as a probe. Hybridization profiles of *Pst* I digests most clearly distinguished human DNA bands and rodent DNA bands. Therefore, *Pst* I was used for determination of chromosomal assignment of the cDNA clone.

RESULTS

Amino Acid Sequence. The complete amino acid sequence of human G6PD, determined by protein sequencing, is shown in Fig. 2. The structure was established based on the sequence determination of tryptic, chymotryptic, and thermolytic peptides obtained from the *S*-carboxymethylated whole protein and from cyanogen bromide fragments of G6PD. The experimental details of the amino acid sequence determination will be reported elsewhere.

Isolation of cDNA Clone. The two cDNA libraries—i.e., the human liver cDNA library and the human hepatoma Li-7 library—were screened with the 41-mer synthetic nucleotide probe. Twelve positive clones were obtained from the hepatoma Li-7 cDNA library and six positive clones were obtained from the liver cDNA library. Insert sizes of these recombinant phage clones, estimated by *Eco*RI digestion, ranged from 0.6 to 2.0 kilobase pairs (kbp). Clone λ G6PD-19, which carried an insert of 2.0 kbp, and clone λ G6PD, which carried an insert of 1.8 kbp, were subjected to nucleotide sequence determination.

Restriction Endonuclease Map and Nucleotide Sequence. Restriction endonuclease cleavage maps of the two clones and nucleotide sequencing strategies are illustrated in Fig. 3. Several other endonucleases, such as *Bam*HI, *Bgl* II, *Hind*III, *Pvu* II, etc., were also used to establish the map. *Pvu* II cleavage sites exist in the DNA insert. The restriction patterns of λ G6PD-19 and λ G6PD-25 were identical in their overlapping regions (Fig. 3).

The nucleotide sequence of cDNA for G6PD derived from the two clones is shown in Fig. 4. The DNA sequence was verified by the data generated from both strands. The overlapping sequence of λ G6PD-19 and λ G6PD-25 span 2.44

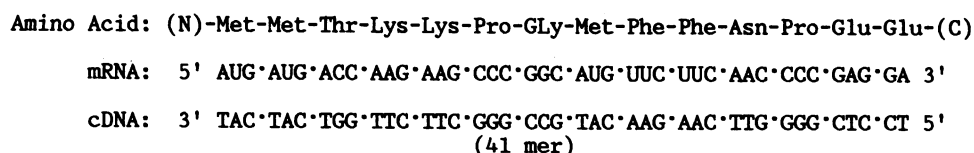


FIG. 1. Synthetic oligonucleotide probe. The probe consisted of 41 nucleotide residues corresponding to the amino acid sequence of the human G6PD.

NH₂-Pro-Arg-Ile-Asp-Ala-Asp-Leu-Lys-Leu-Asp-Phe-Lys-Asp-Val-Leu-Leu-Arg-Pro-Lys-Ser- (20)
 Ser-Leu-Lys-Ser-Arg-Ala-Glu-Val-Asp-Leu-Glu-Arg-Thr-Phe-Thr-Phe-Arg-Asn-Ser-Lys- (40)
 Glu-Thr-Tyr-Ser-Gly-Ile-Pro-Ile-Ile-Val-Ala-Asp-Met-Gly-Ala-Ser-Gly-Asp-Leu-Ala- (60)
 Lys-Lys-Lys-Ile-Tyr-Pro-Thr-Ile-Trp-Trp-Leu-Phe-Arg-Asp-Gly-Leu-Leu-Pro-Glu-Asn- (80)
 Thr-Phe-Ile-Val-Gly-Tyr-Ala-Arg-Ser-Arg-Leu-Thr-Val-Ala-Asp-Ile-Asp-Lys-Gln-Ser- (100)
 Glu-Pro-Phe-Phe-Lys-Ala-Thr-Pro-Glu-Glu-Lys-Leu-Lys-Leu-Glu-Asp-Phe-Phe-Ala-Arg- (120)
 Asn-Ser-Tyr-Val-Ala-Gly-Gln-Tyr-Asp-Asp-Ala-Ala-Ser-Tyr-Gln-Arg-Leu-Asn-Ser-His- (140)
 Met-Asn-Ala-Leu-His-Leu-Gly-Ser-Gln-Ala-Asn-Arg-Leu-Phe-Tyr-Leu-Ala-Leu-Pro-Pro- (160)
 Thr-Val-Tyr-Glu-Ala-Val-Thr-Lys-Asn-Ile-His-Glu-Ser-Cys-Met-Ser-Gln-Ile-Gly-Trp- (180)
 Asn-Arg-Ile-Ile-Val-Glu-Lys-Pro-Phe-Gly-Arg-Asp-Leu-Gln-Ser-Ser-Asp-Arg-Leu-Ser- (200)
 Asn-His-Ile-Ser-Ser-Leu-Phe-Arg-Glu-Asp-Gln-Ile-Tyr-Arg-Ile-Asp-His-Tyr-Leu-Gly- (220)
 Lys-Glu-Met-Val-Gln-Asn-Leu-Met-Val-Leu-Arg-Phe-Ala-Asn-Arg-Ile-Phe-Gly-Pro-Ile- (240)
 Trp-Asn-Arg-Asp-Asn-Ile-Ala-Cys-Val-Ile-Leu-Thr-Phe-Lys-Glu-Pro-Phe-Gly-Thr-Glu- (260)
 Gly-Arg-Gly-Gly-Tyr-Phe-Asp-Glu-Phe-Gly-Ile-Ile-Arg-Asp-Val-Met-Gln-Asn-His-Leu- (280)
 Leu-Gln-Met-Leu-Cys-Leu-Val-Ala-Met-Glu-Lys-Pro-Ala-Ser-Thr-Asn-Ser-Asp-Asp-Val- (300)
 Arg-Asp-Glu-Lys-Val-Lys-Val-Leu-Lys-Cys-Ile-Ser-Glu-Val-Gln-Ala-Asn-Asn-Val-Val- (320)
 Leu-Gly-Gln-Tyr-Val-Gly-Asn-Pro-Asp-Gly-Glu-Gly-Glu-Ala-Thr-Lys-Gly-Tyr-Leu-Asp- (340)
 Asp-Pro-Thr-Val-Pro-Arg-Gly-Ser-Thr-Thr-Ala-Thr-Phe-Ala-Ala-Val-Val-Leu-Tyr-Val- (360)
 Glu-Asn-Glu-Arg-Trp-Asp-Gly-Val-Pro-Phe-Ile-Leu-Arg-Cys-Gly-Lys-Ala-Leu-Asn-Glu- (380)
 Arg-Lys-Ala-Glu-Val-Arg-Leu-Gln-Phe-His-Asp-Val-Ala-Gly-Asp-Ile-Phe-His-Gln-Gln- (400)
 Cys-Lys-Arg-Asn-Glu-Leu-Val-Ile-Arg-Val-Gln-Pro-Asn-Glu-Ala-Val-Tyr-Thr-Lys-Met- (420)
Met-Thr-Lys-Lys-Pro-Gly-Met-Phe-Phe-Asn-Pro-Glu-Glu-Ser-Glu-Leu-Asp-Leu-Thr-Tyr- (440)
Gly-Asn-Arg-Tyr-Lys-Asn-Val-Lys-Leu-Pro-Glu-Pro-Tyr-Glu-Arg-Leu-Ile-Leu-Asp-Val- (460)
Phe-Cys-Gly-Ser-Gln-Met-His-Phe-Val-Arg-Ser-Asp-Glu-Leu-Arg-Glu-Ala-Trp-Arg-Ile- (480)
Phe-Thr-Pro-Leu-Leu-His-Gln-Ile-Glu-Leu-Glu-Lys-Pro-Lys-Pro-Ile-Pro-Tyr-Ile-Tyr- (500)
Gly-Ser-Arg-Gly-Pro-Thr-Glu-Ala-Asp-Glu-Leu-Met-Lys-Arg-Val-Gly-Phe-Gln-Tyr-Glu- (520)
Gly-Thr-Tyr-Lys-Trp-Val-Asn-Pro-His-Lys-Leu-COOH (531)

FIG. 2. Complete amino acid sequence of normal human G6PD. Solid underlining indicates the region that corresponds to the synthetic nucleotide probe. The residue numbers from the NH₂ terminus are in parentheses.

kbp and encode 362 amino acid residues, which are compatible with the amino acid sequence of human G6PD (Fig. 4). Although the insert of λG6PD-19 contains a 1.36-kbp-long 3' noncoding region, it does not contain a poly(A) sequence or the AATAAA polyadenylation signal sequence.

Chromosomal Assignment. Southern hybridization analysis of DNA samples prepared from human, Chinese hamster, and the human-Chinese hamster hybrid fibroblast cell with human X chromosome indicated that the nucleotide sequence

hybridizable with λG6PD-19 probe exists in the human X chromosome (Fig. 5).

DISCUSSION

A tentative or partial amino acid sequence of human G6PD was previously reported (ref. 22; I.-Y.H. and A.Y., personal communication cited in ref. 23). In comparison with the complete amino acid sequence shown in Fig. 2, the previous

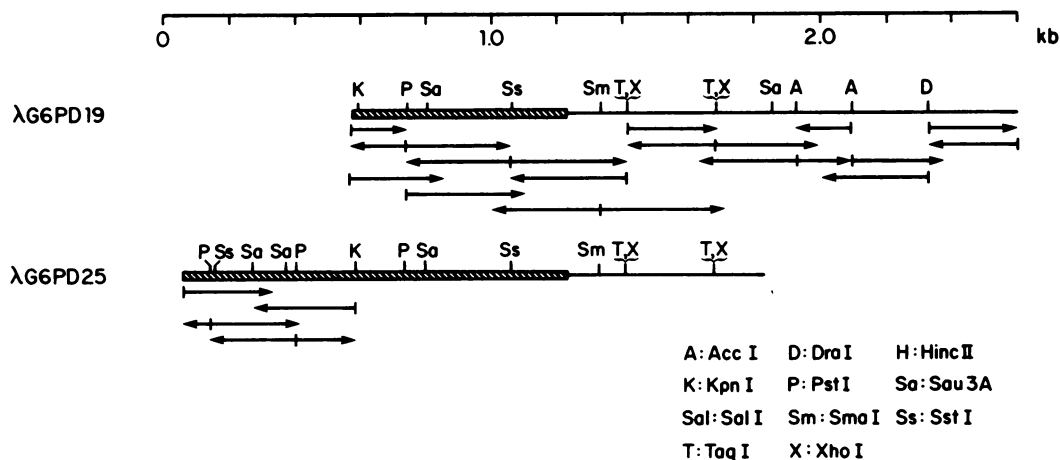


FIG. 3. Restriction map of λG6PD-19 and -25 and sequence determination strategy. Horizontal arrows indicate direction and extent of sequencing. Hatched boxes indicate the coding region.

Eco RI
 GAATTCCTTCACGAGTCCTGCATGAGCCAGATAGGCTGGAACCGCATCATCGTGGAGAAGCCCTTCGGGAGG (66)
 IleHisGluSerCysMetSerGlnIleGlyTrpAsnArgIleIleValGluLysProPheGlyArg
 GACCTGCAGAGCTCTGACCGGCTGTCCAACCACATCTCCTCCCTGTTCCGTGAGGACCAGATCTACCGCATCGAC (141)
 AspLeuGlnSerSerAspArgLeuSerAsnHisIleSerSerLeuPheArgGluAspGlnIleTyrArgIleAsp
 CACTACCTGGGCAAGGAGATGGTGCAGAACCTCATGGTGTGAGATTTGCCAACAGGATCTTCGGCCCATCTGG (216)
 HisTyrLeuGlyLysGluMetValGlnAsnLeuMetValLeuArgPheAlaAsnArgIlePheGlyProIleTrp
 AACCGGACAAACATCGCCTGCGTTATCCTCACCTTCAAGGAGCCCTTTGGCACTGAGGGTCCGGGGGCTATTTT (291)
 AsnArgAspAsnIleAlaCysValIleLeuThrPheLysGluProPheGlyThrGluGlyArgGlyGlyTrpPhe
 GATGAATTTGGGATCATCCGGGACGTGATGCAGAACCTACTGCAGATGCTGTGTCTGGTGGCCATGGAGAAG (366)
 AspGluPheGlyIleIleArgAspValMetGlnAsnHisLeuLeuGlnMetLeuCysLeuValAlaMetGluLys
 CCCGCTCCACCAACTCAGATGACGTCCGTGATGAGAAGTCAAGGTGTTGAAATGCATCTCAGAGGTGCAGGCC (441)
 ProAlaSerThrAsnSerAspValArgAspGluLysValLysValLeuLysCysIleSerGluValGlnAla
 AACAAATGTGGTCTGGGCCAGTACGTGGGGAACCCCGATGGAGAGGGCGAGGCCACCAAAGGTACTGGACGAC (516)
 AsnAsnValValLeuGlyGlnTyrValGlyAsnProAspGlyGluGlyGluAlaThrLysGlyTyrLeuAspAsp
 CCCACGGTCCCCCGGGTCCACCACCGCCACTTTTGCAGCCGTCTCTATGTGGAGAATGAGAGGTGGGAT (591)
 ProThrValProArgGlySerThrThrAlaThrPheAlaAlaValValLeuTyrValGluAsnGluArgTrpAsp
 GGGGTGCCCTTCATCTCGCTGCGCAAGGCCCTGAACGAGCGCAAGGCCGAGGTGAGGCTGCAGTTCCATGAT (666)
 GlyValProPheIleLeuArgCysGlyLysAlaLeuAsnGluArgLysAlaGluValArgLeuGlnPheHisAsp
 GTGGCCGGGACATCTCCACCAGCAGTCAAGCGCAACGAGCTGGTGTCCGCTGCAGCCCAACGAGGCCGTG (741)
 ValAlaGlyAspIlePheHisGlnGlnCysLysArgAsnGluLeuValIleArgValGlnProAsnGluAlaVal
 TACACCAAGATGATGACCAAGAAGCCGGGCATGTTCTTCAACCCGAGGAGTCCGAGCTGGACCTGACCTACGGC (816)
 TyrThrLysMetMetThrLysLysProGlyMetPhePheAsnProGluGluSerGluLeuAspLeuThrTyrGly
 AACAGATAACAAGAAGCTGAAGCTCCCTGAGCCCTACGAGCGCCTCATCTGGACGTCTTCTGCGGGAGCCAGATG (891)
 AsnArgTyrLysAsnValLysLeuProGluProTyrGluArgLeuIleLeuAspValPheCysGlySerGlnMet
 CACTTCGTGCGCAGCGACGAGCTCCGTGAGGCCTGGCGTATTTTACCCCACTGCTGCACCAGATTGAGCTGGAG (966)
 HisPheValArgSerAspGluLeuArgGluAlaTrpArgIlePheThrProLeuLeuHisGlnIleGluLeuGlu
 AAGCCCAAGCCATCCCCTATATTTATGGCAGCCGAGGCCCAAGGAGGAGGAGCTGATGAAGAGAGTGGGT (1041)
 LysProLysProTyrIleTyrGlySerArgGlyTrpGluAlaAspGluLeuMetLysArgValGly
 TTCCAGTATGAGGGCACCTACAAGTGGTGAACCCCAAGCTCTGAGCCCTGGGCACCCACCTCCACCCCGC (1116)
 PheGlnTyrGluGlyThrTyrLysTrpValAsnProHisLysLeu***
 CACGGCCACCTCCTTCCCGCCGCGACCCGAGTCCGGGAGGACTCCGGGACCATTGACCTCAGCTGCACATTCCT (1191)
 GGCCCGGGCTCTGGCACCTTGGCCGCCCTCGCTGTGCTACTACCCGAGCCAGCTACATTCCTCAGCTGCCA (1266)
 AGCACTCGAGACCATCTGGCCCTCCAGACCCTGCCTGAGCCAGGAGCTGAGTCACTCCTCCTCACTCACTCCA (1341)
 GCCCAACAGAAGGAGGAGGAGGGCGCCATTGCTGTGCCAGAGCTTATTGGCCACTGGGTCTCACTCCTGAG (1416)
 TGGGCGCAGGGTGGAGGGGAGGCAAGGGGGAGGAAAGGGCGAGACCCACCTGAGAGAATCTGCTGTGCCTT (1491)
 GCCCGCCAGCCTCAGTGCCACTTGACATTCCTTGTCCACCAGCAACATCTCGAGCCCCCTGGATGTCCCTGTCC (1566)
 ACCAACTCTGCACCTCCATGGCCACCCCGTCCACCCGTAGGCAGCCTCTCTGCTATAAGAAAAGCAGACGAGCA (1641)
 GCTGGGACCCTCCCAACCTCAATGCCCTGAAGGATTTTGGCCACTGGGCAACTCAGAAATACTTCGATCTCCC (1716)
 AAGATAAAGAGGCAAGCAGCAAAAGCTGACCTATTGACGTCTGTTTCATAGTTACCACTTACGCGAGTAGACAGAAC (1791)
 TCGGCTTTTCAGAAAATAGGTGTCAAGTCCACTTTATAAGAACCTTTTTTTCTAAAAATAAGATAAAAGGTGGCTT (1866)
 TGCAATTTCTGATTAACGACTGTGTCTTTGTCACTCTGCTTAACTTTAGGAGTATCCATTCTGTGATTGTAG (1941)
 ACTTTTGTGATATCTTCTCTGGAAGAATATCATTTCTTTCTTGAAGGGTGGTTTACTAGAATATTCAAATCA (2016)
 ATCATGAAAGCGTTACTATTTTGAAGTCTTAAAGTTTCTAAAAATTAACCTCACATCCCTTCTGTTAGGGTCTT (2091)
 TCAGAATATCTTTTATAAACAGAACATTGAAAGTCAATTTGCTACATGATTTTGTGTGTGAAGGACATA (2166)
 CCAGTTTAAATCAATTAATTGAAAACATCATATAAGCCCAACTTTGTTTGGAGGAAGACCGAGGTTGAGGT (2241)
 TTTTCTTCTGTATAAGCAGCTACTGACAAAATGTAGAGGCCATCAACCGTCAAACCATTTGTTTATATCGC (2316)
 AGAGGAGACGGATGTGTAATAACTGCATTGCTTTTTTTTTCAGTTTGTATAACCTCTAATCTCCGTTTTCATGA (2391)
 TACGCTTTTGTAGAAACATTAATTGATGTTTGAAGCAAGTGTGTATGGGAATTC
Eco RI

FIG. 4. Nucleotide sequence of cDNA and deduced amino acid sequence for human G6PD. The termination codon is indicated by ***. The numbers in parentheses are counted from the first coding codon.

sequence includes sequence errors and several misalignments of cyanogen bromide peptides and proteolytic peptides.

A cDNA expression library is commonly screened with a specific antibody against a protein whose cDNA is to be cloned. Positive clones thus obtained can be screened by synthetic oligonucleotide probes for further confirmation. Utilizing this approach, we previously obtained cDNA clones for several human enzymes (13, 24, 25). For cloning of G6PD cDNA, we initially screened the two cDNA expression libraries with a specific rabbit antibody against human G6PD and relatively short synthetic nucleotide probes. Several positive clones were obtained, but all of them were found to be false positive by partial nucleotide sequencing. To avoid the problem of false-positive cloning, a 41-mer synthetic

nucleotide corresponding to the amino acid sequence of human G6PD was used as a probe in the screening. The unique nucleotide sequence was derived from the presumed common codon usage in eukaryotes (26). Long nucleotides with unique sequences as screening probes were found to be useful in cDNA cloning by previous investigators (27, 28).

The G6PD content in erythrocytes and other tissues is very low—i.e., 1:50,000 in erythrocytes and 1:30,000 in soluble liver proteins—and G6PD mRNA is expected to be proportionally low. For cloning G6PD cDNA, the use of a cDNA library originating from human hepatoma Li-7, which expressed G6PD far more strongly than the liver, would be a logical choice. Indeed, clones λ G6PD-19 and -25 were obtained from this library.

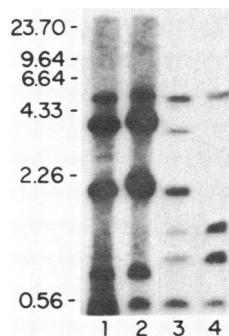


FIG. 5. Southern blot hybridization of DNA samples from the cells with and without human X chromosome. DNA samples were digested by *Pst* I and subjected to hybridization. Numbers on the left are kbp. Lanes: 1, human female DNA; 2, human male DNA; 3, DNA from a human-Chinese hamster hybrid cell line (F74-5) with human X chromosome only; 4, Chinese hamster DNA.

Cloning of cDNA for human G6PD was previously reported (7). However, the cDNA reported (pGD6405) appears to lack the coding sequence (29, 30), and its authenticity was not verified.

The nucleotide sequences of the present clones, λ G6PD-19 and -25, are perfectly compatible with the amino acid sequence of normal human G6PD. The clones are not full-length and encode only 362 amino acid residues, which constitute about 70% of the G6PD molecule. The polyadenylation signal and a poly(A) segment are absent in the 3' noncoding region of 1.36 kbp, and they could be located far distal from the termination signal.

Judging from the compatibility with the amino acid sequence and from hybridization with the human X chromosome, one can conclude that the clones obtained are for the X-linked G6PD and not for autosomal hexose-6-phosphate dehydrogenase or other isozymes.

We are indebted to Dr. Savio L. C. Woo and Dr. J. R. DeWet for providing us with the cDNA libraries and to Dr. T. Mohandas for providing us with the DNA samples. This work was supported by U.S. Public Health Service Grant HL-29515.

1. McKusick, V. A. (1983) in *Mendelian Inheritance in Man* (Johns Hopkins Press, Baltimore), 6th Ed., pp. 1012-1129.
2. Pai, G. S., Sprenkle, J. A., Do, T. T., Mareni, C. E. & Migeon, B. R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2810-2813.
3. Show, C. R. & Barto, E. (1965) *Science* **148**, 1099-1100.
4. Beutler, E. & Morrison, M. (1967) *J. Biol. Chem.* **242**, 5289-5293.
5. Toncheva, D., Evres, T. & Tzoneva, M. (1982) *Hum. Hered.* **32**, 193-196.

6. Shreve, D. S. & Levy, H. R. (1977) *Biochem. Biophys. Res. Commun.* **78**, 1369-1375.
7. Persico, M. G., Toniolo, D., Nobile, C., D'Urso, M. & Luzzatto, L. (1981) *Nature (London)* **294**, 778-780.
8. Yoshida, A. (1966) *J. Biol. Chem.* **241**, 4966-4976.
9. Huang, I.-Y., Rubinfien, E. & Yoshida, A. (1980) *J. Biol. Chem.* **255**, 6408-6411.
10. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
11. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194-1198.
12. DeWet, J. R., Fukushima, H., Dewji, N. N., Wilcox, E., O'Brien, J. S. & Helinski, D. R. (1984) *DNA* **3**, 437-447.
13. Ikuta, T., Fujiyoshi, T., Kurachi, K. & Yoshida, A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2703-2707.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 76-85.
15. McDonnell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119-146.
16. Yang, R. C. A., Lis, J. & Wu, R. (1979) *Methods Enzymol.* **68**, 176-182.
17. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101-106.
18. Cohen, S. N., Chang, A. C. Y. & Hsu, L. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110-2114.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
20. Bell, G. I., Karam, J. H. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5759-5763.
21. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517.
22. Yoshida, A. (1977) *Acta Biol. Med. Ger.* **36**, 689-701.
23. Beutler, E. (1980) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Wyngaarden, J. B. & Fredrickson, D. S. (McGraw-Hill, New York), 4th Ed., p. 1629.
24. Hsu, L. C., Tani, K., Fujiyoshi, T., Kurachi, K. & Yoshida, A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3771-3775.
25. Ikuta, T., Szeto, S. & Yoshida, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 634-638.
26. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) *Nucleic Acids Res.* **9**, r43-r74.
27. Wood, W. I., Capon, D. J., Simonsen, C. C., Eaton, D. L., Gitschier, J., Keyt, B., Seeburg, P. H., Smith, D. H., Hollingshead, P., Wion, K. L., Delwart, E., Tuddenham, E. G. D., Vehar, G. A. & Lawn, R. M. (1984) *Nature (London)* **312**, 330-337.
28. Lathe, R. (1985) *J. Mol. Biol.* **183**, 1-12.
29. Luzzatto, L. & Battistuzzi, G. (1984) *Adv. Hum. Genet.* **14**, 217-329.
30. Toniolo, D., Persico, M. G., Battistuzzi, G. & Luzzatto, L. (1984) *Mol. Biol. Med.* **2**, 89-103.