JCI The Journal of Clinical Investigation

Alternative splicing of human glucose-6-phosphate dehydrogenase messenger RNA in different tissues.

A Hirono, E Beutler

J Clin Invest. 1989;83(1):343-346. https://doi.org/10.1172/JCI113881.

Research Article

Different forms of glucose-6-phosphate dehydrogenase (G-6-PD) have been described in different tissues. Moreover, the directly determined amino acid sequence amino end of the red cell enzyme does not exactly match the sequence deduced from cDNA isolated from HeLa cells or lymphoblasts. We have therefore investigated the sequence of cDNA from sperm, granulocytes, reticulocytes, brain, placenta, liver, lymphoblastoid cells, and cultured fibroblasts. A novel human cDNA, which has extra 138 bases coding 46 amino acids, was isolated from a lymphoblastoid cell library. Sequencing of genomic DNA amplified by the polymerase chain reaction (PCR) revealed that the extra sequence was derived from the 3'-end of intron 7 by alternative splicing. This longer form of mRNA was also detected in sperm and granulocytes. Sequence analysis using PCR-amplified cDNA revealed that the 5'-end of the coding sequence of G6PD mRNA in reticulocytes is identical to those in other tissues.

Alternative Splicing of Human Glucose-6-Phosphate Dehydrogenase Messenger RNA in Different Tissues

Akira Hirono and Ernest Beutler
Department of Basic and Clinical Research, Research Institute of Scripps Clinic, La Jolla, California 92037

Abstract

Different forms of glucose-6-phosphate dehydrogenase (G-6-PD) have been described in different tissues. Moreover, the directly determined amino acid sequence amino end of the red cell enzyme does not exactly match the sequence deduced from cDNA isolated from HeLa cells or lymphoblasts. We have therefore investigated the sequence of cDNA from sperm, granulocytes, reticulocytes, brain, placenta, liver, lymphoblastoid cells, and cultured fibroblasts.

A novel human cDNA, which has extra 138 bases coding 46 amino acids, was isolated from a lymphoblastoid cell library. Sequencing of genomic DNA amplified by the polymerase chain reaction (PCR) revealed that the extra sequence was derived from the 3'-end of intron 7 by alternative splicing. This longer form of mRNA was also detected in sperm and granulocytes. Sequence analysis using PCR-amplified cDNA revealed that the 5'-end of the coding sequence of G6PD mRNA in reticulocytes is identical to those in other tissues.

Introduction

Recently the structure of human glucose-6-phosphate dehydrogenase (G6PD[1]; D-glucose-6-phosphate: $NADP^+$ oxidoreductase, EC 1.1.1 49) has been extensively elucidated. The complete amino acid sequence (1), nucleotide sequence of cDNA (1, 2) and partial genomic DNA sequence (3) have already been published. Very recently single base substitutions have been disclosed in several G6PD variants (4–6).

Although G6PD is a single copy X-linked gene, some variability of the enzyme in various tissues has been noted. A G6PD isozyme has been described in human fetal brain (7) and some microheterogeneity has been noted in other tissues (8, 9). Such differences have generally been ascribed to post-translational modification of the enzyme. However, recent study of G6PD cDNA by molecular cloning has suggested the existence of multiple forms of human G6PD mRNA that may produce multiple forms of enzyme protein. We isolated two different forms of G6PD cDNA from human EBV-transformed lymphoblastoid cells (5). Moreover, discrepancies between the amino acid sequences of the amino end of normal red cell G6PD determined directly from peptides (1) and that deduced from the nucleotide sequence (2) might suggest the existence of further different forms of G6PD in the human red cell. These different forms of G6PD appear to be generated by the alternative splicing of the same gene transcript (5).

Recent development of a DNA amplification technique using the polymerase chain reaction (PCR) (10) has made it possible to detect different forms of mRNA with great sensitivity. We present the evidence that previously described two different forms of cDNA (5) are generated by alternative splicing of same gene transcript. The structure of the 5'-end of the coding region of human reticulocyte G6PD mRNA is also reported.

Methods

Enzymes and chemicals. Avian myeloblastosis virus reverse transcriptase was obtained from Molecular Genetic Resources. Other enzymes were from New England Biolabs, Beverly, MA. Oligo(dT)-cellulose and RNase H were from Pharmacia Fine Chemicals (Piscataway, NJ) and ^{32}P-labeled nucleotides were from New England Nuclear Research Products (Boston, MA). Other reagents used were of analytical grade.

cDNA libraries. The following cDNA libraries from various human tissues were used. All were constructed by a standard method (11). (*a*) EBV-transformed lymphoblastoid cells with G6PD A(−), in λGT10 (5), (*b*) EBV-transformed lymphoblastoid cells cultured from a patient with G6PD Riverside (12), in λGT10, (*c*) normal human fibroblast, in λGT11, (*d*) normal adult cortex, in pGEM4, a gift from Dr. Gabriel H. Travis (Research Institute of Scripps Clinic), (*e*) normal liver, in λGT11, (*f*) normal placenta, in λGT11.

Preparation of poly(A)+RNA and cDNA synthesis. Granulocytes and sperm were obtained from a normal male. Granulocytes were purified by 2% dextran-saline sedimentation combined with Ficoll barrier gravity sedimentation. Reticulocytes were obtained from a female patient with hemolytic anemia whose peripheral reticulocyte count was 23%. Whole blood was purified by filtration through an α-cellulose-microcrystalline cellulose column (13). Contamination by leukocytes in the final preparation was estimated as 1–14,000 reticulocytes. Total RNA was extracted by the guanidine thiocyanate method (14) and poly(A)+RNA was prepared by oligo(dT)-cellulose column chromatography (15). First and second cDNA strand synthesis was carried out by the method of Lapeyre and Amalric (16). A synthetic 17-mer oligonucleotide (Primer 8 or 10) corresponding to a unique sequence in G6PD cDNA was used as the primer (Table I).

DNA amplification. The first-strand cDNA synthesized from mRNA, double-stranded cDNA in libraries, and genomic DNA containing the target sequence were amplified by PCR (10), with Taq

Address reprint requests to Dr. Beutler, Department of Basic and Clinical Research, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, CA 92037.

Received for publication 15 August 1988 and in revised form 4 October 1988.

1. *Abbreviations used in this paper:* G6PD, D-glucose 6-phosphate dehydrogenase; PCR, polymerase chain reaction.

J. Clin. Invest.

0021-9738/89/01/0343/04 $2.00
Volume 83, January 1989, 343–346

Table I. Sequence of Oligonucleotide Primers

Primer	Sequence	cDNA Nucleotide numbers
4	5'-GGGCTCACTCTGTTTGC	258–242*
6	5'-CAGGTAGAAGAGGCGGT	420–404*
7	5'-CATCATCGTGGAGAAGC	498–514
8	5'-CACCATCTCCTTGCCCA	624–608*
9	5'-CGCCTGCGTTATCCTCA	691–707
10	5'-TGGTTCTGCATCACGTC	788–772*

* Prime synthesis of the bottom strand.

polymerase (17) according to a protocol by Chehab et al. (18) using primers in Table I. To determine the nucleotide sequence of reticulocyte G6PD cDNA, 100 ng of blunt-ended, double-stranded cDNA were circularized intramolecularly before PCR by incubating at 4°C for 14 h with 800 U of T4 DNA ligase in 10 μl of 50 mM Tris-HCl (pH 8.0), 7 mM $MgCl_2$, 1 mM dithiothreitol, and 0.5 mM ATP. The amplified DNA segments were separated by 5 or 8% polyacrylamide gel electrophoresis and visualized by ethidium bromide or silver stain (19). For nucleotide sequencing, amplified DNA was removed from the gel and sequenced by the chemical cleavage method (20) and/or the dideoxynucleotide chain terminating method (21) with slight modification (22).

Results

Nucleotide sequence of intron 7. The entire sequence of intron 7 from genomic DNA of G6PD B(+) and G6PD A(−) males was amplified using primers 9 and 10 (Table I). The amplified 463bp-segments were separated by 5% PAGE and sequenced (Fig. 1). Except for the 174th base, which was T in G6PD A(−), the nucleotide sequences of both amplified DNAs were identical and consistent with the published partial sequence (3). The 138 bases in the 3'-end of intron 7 were identical to the extra sequence found in a G6PD A(−) cDNA clone A-7 (5). A 3'-AG acceptor with a consensus splicing sequence (23) was succeeded immediately by this sequence. This strongly suggests that the extra sequence in the cDNA clone A-7 is derived from intron 7 by alternative splicing. The amino acid sequence by these 138 bases is characterized by its richness in glycine (21.7% vs. 6.4% in the whole G6PD sequence) and serine (17.3% vs. 4.7%).

```
GTGAGAGCTC TTCCTCTCTC CTGGGAGGCG GGCACAGGGT GGCAGAGCCA      50

GTCACCCTGC AGGGCTACTC TTCCCTATCT TGGGGGAGCT CCTCCTCACC     100

CTGCAGTTCA AAACCTAAGT GTCTGAGCTA TCAGACCGGG CTGGAAAGGC     150

TGGACCCCTA CACAGCCAAG CACCCCACGG TTTTATGATT CAGTGATAGC     200
                         *

ATCACCATGT CCTTCCTTGA TTTAAG [GGGA CCT GGA AGA CAA GGG     245
                              Gly Pro Gly Arg Gln Gly

GGA TCA GGA AGT GAG TCT TGC AGC TTG TCA CTA GGA AGC CTT    297
Gly Ser Gly Ser Glu Ser Cys Ser Leu Ser Leu Gly Ser Leu

GTT TGG GGT CCC CAT GCC CTT GAA CCA GGT GAA CAG GGC GGG    349
Val Trp Gly Pro His Ala Leu Glu Pro Gly Glu Gln Gly Gly

GAG CTA AGG CGA GCT CTG GCC TCT TCC GTC CGC AG]           384
Glu Leu Arg Arg Ala Leu Ala Ser Ser Val Pro Ar(g)
```

Figure 1. Nucleotide sequence of intron 7 of normal human G6PD gene. The deduced amino acids in extra coding sequences are shown below. The nucleotide sequence in brackets is an extra sequence found in a G6PD cDNA clone. The amino acid sequence that appears to be compatible with the PEST region (25) is underlined. *A base change C → T found in a G6PD A(−) subject.

Tissue distribution of different forms of G6PD mRNA. To examine the tissue distribution of the different forms of G6PD mRNA, corresponding DNA sequence was amplified by PCR using primers 9 and 10 from first-stranded cDNA prepared from granulocytes, reticulocytes, normal lymphoblastoid cells and sperm or double-stranded cDNA in cDNA libraries from liver, placenta, fibroblasts, and lymphoblastoid cells. Each amplified segment was separated by 8% polyacrylamide gel electrophoresis (Fig. 2). As summarized in Table II, in all the tissues examined a 99-bp fragment derived from normal splicing was dominant. Additional 237-bp fragments derived from a different splicing were detected only in samples from granulocytes, sperm, and lymphoblastoid cells. The 237-bp band of normal lymphoblastoid cell sample was very faint. Approximation by densitometry showed that the content of the 237-bp splicing product represented < 3% of the 99-bp splicing product.

Structure of the coding region of reticulocyte G6PD cDNA. To examine the existence of different forms of mRNA in human reticulocytes, cDNA corresponding to the 5'-half of the coding sequence was synthesized using 5'-phosphorylated primer 8 and circularized by T4 DNA ligase (see Methods).

Under these conditions some concatemers should have been formed as well as circular forms. Fig. 3 shows the location of sequences corresponding to each primer. The target DNA sequence was first amplified using primers 6 and 7. The length of amplified segments ranged from 40 to 1,500 bp (data not shown). The fragments of 800–1,000 bp, which might contain enough long target sequence, were removed from a gel and used as templates for re-amplification using 5'-labeled primer 4 and unlabeled primer 7. Amplified DNA was separated by 5% PAGE and well-isolated bands on the autoradiogram were removed from the gel. The sequencing of these 5'-labeled DNAs by the method of Maxam and Gilbert (20) revealed that several amplified DNA contained entire coding sequence 5' to the site of primer 4 and the nucleotide sequence was identical with a published sequence by Persico et al. (2).

Discussion

Alternative splicing of gene transcripts is now recognized as one of the most important mechanisms of regulating expression of proteins developmentally or tissue specifically. It is considered to be fairly common and more than 50 genes are

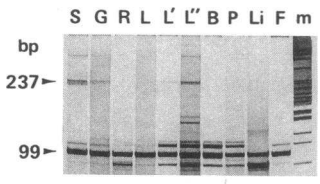

Figure 2. Tissue distribution of two forms of human G6PD mRNA. DNA sequences were amplified from G6PD cDNA from various tissues using primers 9 and 10. Each DNA segment was separated by 8% PAGE and visualized by silver staining. Lanes: *S*, sperm; *G*, granulocyte; *R*, reticulocyte; *L*, lymphoblastoid cell with normal G6PD B(+); *L'*, lymphoblastoid cell with G6PD Riverside; *L''*, lymphoblastoid cell with G6PD A(−); *B*, brain; *P*, placenta; *Li*, liver; *F*, fibroblast; and *m*, Msp I digest of pBR322 DNA for a size marker.

Table II. Summary of Alternative Splicing of Human G6PD mRNA in Various Tissues

Tissues	Type of G6PD	99 bp-band	237 bp-band
Sperm	B(+)	++	+
Granulocytes	B(+)	++	+
Reticulocytes	B(+)	++	−
Lymphoblastoid cells	B(+)	++	+
Lymphoblastoid cells	Riverside	++	+
Lymphoblastoid cells	A(−)	++	+
Brain	B(+)	++	−
Placenta	B(+)	++	−
Liver	B(+)	++	−
Fibroblasts	B(+)	++	−

Alternative splicing of G6PD mRNA.

known to generate multiple forms of protein using alternative splicing (24). Previously we reported a G6PD cDNA clone that has extra 138 bases coding 46 amino acids (5). In the present study we confirmed that this extra coding sequence was derived from intron 7 of G6PD gene and was preceded by a 3'-AG acceptor with a consensus splicing sequence (23). These findings strongly suggest that two types of mRNA with different coding sequences were generated by an alternative splicing of the 7th intron of the G6PD gene.

Although the longer form of G6PD cDNA clone was first found in a cDNA library from a Black male with G6PD A(−), the present study of its tissue distribution revealed that this type of mRNA or cDNA was not specifically related to G6PD A(−) variant nor to lymphoblastoid cells. It was detected in granulocytes and sperm as well as lymphoblastoid cells. Persico et al. (2) performed Northern blot analysis for G6PD mRNA from various tissues and reported the existence of only single species of G6PD mRNA. However, the extremely low content of the longer form mRNA might not allow it to be detected by conventional technique other than the DNA amplification by PCR.

The mRNA with extra coding sequence should generate an enzyme protein with 561 amino acids whose molecular weight is ~ 9% higher than the common G6PD monomer. Although this difference in molecular weight might make it possible to separate these two types of G6PD protein by SDS-PAGE, there has so far been no report of detecting the longer form of G6PD protein. We have carried out SDS-PAGE of leucocyte G6PD followed by immunoblotting analysis. However, only the common form of G6PD was detected (unpublished observation).

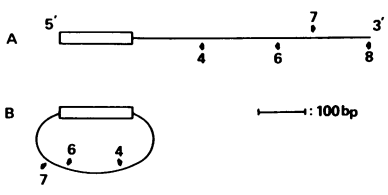

Figure 3. DNA amplification and sequencing strategy of human reticulocyte G6PD cDNA. (*A*) Location of target sequence (*open bar*) and each primer. Direction of nucleotide chain extension is indicated by arrows. (*B*) Circular form.

What is the function of the longer form G6PD? Considering the very low content of the mRNA, it might rather be regarded as a nonfunctional remnant of cell differentiation process than an active functioning enzyme even if it is translated normally. According to the PEST hypothesis proposed by Rogers et al. (25), the amino acid sequence of proteins with short intracellular half-lives contain one or more regions rich in proline, glutamic acid, serine, and threonine (PEST region). The serine rich region in the extra portion of the long G6PD (4th Arg to 25th His in Fig. 1) appears to be compatible with the PEST region. It is well known that in some tissues the switching of isozyme pattern occurs during maturation or malignant transformation of cells (26, 27). It is possible that similar switching of G6PD isozyme occurs in some tissues, and the longer G6PD is the more unstable form dominant in immature cells.

The discrepancy between the amino acid sequence in amino end of human red cell G6PD protein (1) and that deduced from cDNA sequence (2, 5) is an unsolved puzzle. Since the nucleotide sequence was determined by using cDNA from tissues other than erythroid cells, it seemed possible that a different form of G6PD mRNA existed in those cells. The existence of a red cell-specific G6PD is reasonable considering the unique feature of red cell metabolism. One possible mechanism of generating these different mRNA might be an alternative splicing of exon 2 in G6PD. We obtained amplified DNA segments corresponding to exon 2 using the PCR technique and revealed that the nucleotide sequence was identical to those of other tissue G6PD cDNA. This may suggest that at least at the reticulocyte stage mRNA for a red cell specific G6PD is not a product of alternative splicing of exon 2.

Acknowledgments

This is publication No. 5395BCR from the Research Institute of Scripps Clinic, La Jolla, CA.

This work is supported in part by grant HL-25552 from the National Institutes of Health.

References

1. Takizawa, T., I. Y. Huang, T. Ikuta, and A. Yoshida. 1986. Human glucose-6-phosphate dehydrogenase: Primary structure and cDNA cloning. *Proc. Natl. Acad. Sci. USA.* 83:4157–4161.

2. Persico, M. G., G. Viglietto, G. Martino, D. Toniolo, G. Paonessa, C. Moscatelli, R. Dono, T. Vulliamy, L. Luzzatto, and M. D'Urso. 1986. Isolation of human glucose-6-phosphate dehydrogenase (G6PD) cDNA clones: Primary structure of the protein and unusual 5'non-coding region. *Nucleic Acids Res.* 14:2511–2522, 7822.

3. Martini, G., D. Toniolo, T. Vulliamy, L. Luzzatto, R. Dono, G. Viglietto, G. Paonessa, M. D'Urso, and M. G. Persico. 1986. Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:1849–1855.

4. Takizawa, T., Y. Yoneyama, S. Miwa, and A. Yoshida. 1987. A single nucleotide base transition is the basis of the common human glucose- 6-phosphate dehydrogenase variant A(+). *Genomics.* 1:228–231.

5. Hirono, A., and E. Beutler. 1988. Molecular cloning and nucleotide sequence of cDNA for human glucose-6-phosphate dehydrogenase variant A(−). *Proc. Natl. Acad. Sci. USA.* 85:3951–3954.

6. Vulliamy, T. J., 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. 1988. Diverse point mutations in the human glucose 6-phosphate dehydrogenase gene cause enzyme defi-

ciency and mild or severe hemolytic anemia. *Proc. Natl. Acad. Sci. USA.* 85:5171–5175.

7. Toncheva, D., T. Evres, and M. Tzoneva. 1982. G6PD in immature and mature human brain: Electrophoretic and enzyme kinetic studies. *Hum. Hered.* 32:193–196.

8. Kahn, A., P. Boivin, M. Vibert, D. Cottreau, and J.-C. Dreyfus. 1974. Post-translational modifications of human glucose-6- phosphate dehydrogenase. *Biochimie.* 56:1395–1407.

9. Kahn, A., O. Bertrand, D. Cottreau, P. Boivin, and J.-C. Dreyfus. 1977. Evidence for structural differences between human glucose-6-phosphate dehydrogenase purified from leukocytes and erythrocytes. *Biochem. Biophys. Res. Commun.* 77:65–72.

10. Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science (Wash. DC).* 230:1350–1354.

11. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *In* Molecular Cloning. A Laboratory Manual. 213–246. Cold Spring Harbor Laboratory, New York.

12. Hirono, A., L. Forman, and E. Beutler. 1988. Enzymatic diagnosis in non-spherocytic hemolytic anemia. *Medicine (Baltimore).* 67:110–117.

13. Beutler, E., C. West, and K. G. Blume. 1976. The removal of leukocytes and platelets from whole blood. *J. Lab. Clin. Med.* 88:328–333.

14. Chirgwin, J. M., A. E. Pryzbyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active RNA from sources enriched in ribonuclease. *Biochemistry.* 18:5294–5298.

15. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

16. Lapeyre, B., and F. Amalric. 1985. A powerful method for the preparation of cDNA libraries: isolation of cDNA encoding a 100-kDal nuclear protein. *Gene.* 37:215–220.

17. Kogan, S. C., M. Doherty, and J. Gitschier. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. Application to hemophilia A. *N. Engl. J. Med.* 317:985–990.

18. Chehab, F. F., M. Doherty, S. Cai, Y. W. Kan, S. Cooper, and E. M. Rubin. 1987. Detection of sickle cell anaemia and thalassaemias. *Nature (Lond.).* 329:293–294.

19. Perbal, B. 1984. Separation of DNA fragments by electrophoresis. *In* A Practical Guide to Molecular Cloning. John Wiley & Sons, New York. 196–207.

20. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* 74:560–564.

21. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

22. Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* 14:1319–1324.

23. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459–472.

24. Breitbart, R. E., A. Andreadis, and B. Nadal-Ginard. 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* 56:467–495.

25. Rogers, S., R. Wells, and M. Rechsteiner. 1986. Amino acid sequences common to rapidly degraded proteins. The PEST hypothesis. *Science (Wash. DC).* 234:364–368.

26. Takegawa, S., H. Fujii, and S. Miwa. 1983. Change of pyruvate kinase isozymes from M2- to L-type during development of the red cell. *Br. J. Haematol.* 54:467–474.

27. Markert, C. L. 1975. Biology of isozymes. *In* Isozymes. C. L. Markert, editor. Academic Press, New York. 1–9.