

Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population

Ernest Beutler and Terri Gelbart

Pyruvate kinase (PK) deficiency is the most common cause of hereditary nonspherocytic hemolytic anemia. The prevalence of this deficiency is unknown, though some estimates have been made based on the frequency of low red cell PK activity in the population. An additional 20 patients with hereditary nonspherocytic hemolytic anemia caused by PK deficiency have been genotyped. One previously unreported mutation 1153C→T (R385W) was encountered. The relative frequency of PK mutations in patients with hemolytic anemia caused by PK deficiency was calculated from the 18 white patients re-

ported here and from 102 patients previously reported in the literature. DNA samples from 3785 subjects from different ethnic groups have been screened for the 4 more frequently encountered mutations—c.1456 C→T(1456T), c.1468 C→T(1468T), c.1484 C→T(1484T), and c.1529 G→A (1529A)—by allele-specific oligonucleotide hybridization. Among white patients the frequency of the 1456T mutation was 3.50×10^{-3} ; that of the 1529A mutation was 2.03×10^{-3} . Among African Americans the frequency of the 1456T mutation was 3.90×10^{-3} . The only mutation found in the limited number of

Asians tested was 1468T at a frequency of 7.94×10^{-3} . Based on the gene frequency of the 1529A mutation in the white population and on its relative abundance in patients with hemolytic anemia caused by PK deficiency, the prevalence of PK deficiency is estimated at 51 cases per million white population. This number would be increased by inbreeding and decreased by failure of patients with PK deficiency to survive. (Blood. 2000;95:3585-3588)

© 2000 by The American Society of Hematology

Introduction

Pyruvate kinase (PK) deficiency is probably the most common cause of hereditary nonspherocytic hemolytic anemia.¹ The true prevalence of this disorder is unknown because diagnoses are made not only in specialized centers throughout the world but also in commercial and hospital laboratories.

The red cells of heterozygotes for PK deficiency generally have approximately half the normal level of enzyme activity. Based on the lowered erythrocyte enzyme activity of heterozygotes, a few efforts have been made to obtain some information regarding the prevalence of this disorder. Assays of red cell PK in 214 normal subjects led Blume et al² to conclude that approximately 1.4% of the German subjects were heterozygous for the deficiency. Mohrenweiser³ reported that among 697 newborns, 84% of whom were white, 1 (0.15%) manifested red cell PK activity that was more than 3 standard deviations below the mean. In a subsequent study, however, he suggested that the frequency of heterozygous subjects was 18 of 1736 white newborns.⁴ However, because red cell PK activity varies widely from person to person, there is bound to be overlap between normal and heterozygous levels.

Detection of the existence of mutations based on the DNA analysis is more robust than is the estimation of enzyme levels. Although it is not feasible to perform complete sequence analysis on thousands of persons, it is possible to determine the frequency of those mutations that are the most common causes of PK deficiency. Based on the known frequency of such mutations in patients with the disease, it is possible to extrapolate from the general population

to obtain some estimate of the number of persons in the population who may have the PK-deficient genotype.

We now report the genotypes of an additional 20 patients with PK deficiency. Based on the accumulated information of the occurrence of mutations in patients with PK deficiency and a survey of more than 3500 normal subjects, we estimate the prevalence of PK deficiency in the white population at large.

Patients and methods

Patients with PK deficiency had documented hemolytic anemia and deficiency of erythrocyte PK activity when assayed by standard methods as modified from previously published methods.⁵ When the more common mutations were not detected by restriction analysis or direct sequencing, the entire coding region was sequenced as described previously.⁶

DNA samples from 3785 anonymous persons, identified only by ethnic origin, were examined for 4 PK mutations: c.1456 C→T(1456T), c.1468 C→T(1468T), c.1484 C→T(1484T), and c.1529 G→A (1529A). These samples were derived from several different sources. Approximately 3500 samples were obtained from patients attending a health-screening clinic. Ethnic origin was based on self-identification. Approximately 300 additional samples from African Americans, identified as such by the phlebotomist, were from discarded diagnostic samples. The segment of DNA containing these mutations was amplified by means of the polymerase chain reaction (PCR) with the following 2 primers: sense 5'-CTCGTTCACCACTTCTTGC-3' and antisense 5'-GAG-GCAAGGCCCTTTGAGTG-3'. The PCR mixture contained 34 mmol/L

From the Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA.

Submitted November 29, 1999; accepted January 21, 2000.

Supported by National Institutes of Health grants HL25552 and RR00833 and the Stein Endowment Fund.

Reprints: Ernest Beutler, Department of Molecular and Experimental Medi-

cine, The Scripps Research Institute, 10550 North Torrey Pines Rd, La Jolla, CA 92037.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2000 by The American Society of Hematology

Tris-HCl, pH 8.8, 8.3 mmol/L ammonium sulfate, 1.5 mmol/L MgCl₂, 85 µg/mL bovine serum albumin, 0.2 mmol/L each dATP, dCTP, dGTP, and dTTP, 120 ng of each oligonucleotide primer listed above, 200 ng genomic DNA, and 1 U Taq polymerase. After a 4-minute denaturation at 98°C, 30 cycles of PCR at 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds were performed.

The amplified DNA was spotted in duplicate on Nytran SuPerCharge membranes (Schleicher and Schuell, Keene, NH) and probed with the wild-type and mutant probes for each mutation studied (Table 1). All filters contained controls of normal, heterozygous, and homozygous samples. All positive results were confirmed by restriction analysis or direct sequencing.

Results

Patient studies

The mutations found in the 20 patients with PK deficiency are summarized in Table 2. In one patient, only 1 of the 2 mutations was identified. It is possible that the missing mutation, denoted "?," was not found because it was not in the coding region of the gene or because of technical reasons. One mutation that has not been reported previously is 1153T. The deduced amino acid change for this mutation is arginine to tryptophan at amino acid 385 (R385W). This new mutation was found with the 1456T mutation in a 1-year-old boy. He was born at full term and had a bilirubin level of 14 mg/dL. His reticulocyte count was 28%, and his hematocrit level was 32%. He did well after exchange transfusion. At 3 months of age he was found to have a hemoglobin level of 5 g/dL and a reticulocyte count of 9.9%. He is transfused every 3 to 4 weeks. His parents are of Irish and mixed European ancestry.

Population studies

Heterozygotes for each of the 4 mutations for which the population was screened were detected in the population survey except for the 1484T mutation, but there were no homozygotes. The results are shown in Table 3. The gene frequency for the 1529A mutation was 0.00203 ± 0.0006 (mean \pm 1 SE) in the white population; that of the 1456T mutation was 0.00350 ± 0.0008 in the white population and 0.0039 ± 0.0031 in the African American population. The 1468T mutation was found only among Asians. In the small group of 126 persons examined, it was encountered twice, giving a gene frequency of 0.00794 ± 0.00559 . The 1484T mutation was not found in any of the subjects studied.

Discussion

Estimating the prevalence of uncommon autosomal recessive diseases is a difficult challenge. For example, numerous attempts have been made to estimate the prevalence of Gaucher disease in the Jewish population. One approach has been to attempt to identify all cases in a target population by surveying physicians, hospitals, or both.⁷ This method depends on accurate and complete ascertainment of cases and thus tends to underestimate the true incidence of the disease. A second means of estimating prevalence in the population is to attempt to identify heterozygotes by

Table 2. PK mutations found in the present series

Patient designation	Ethnic origin	Pyruvate kinase IU/gHb V _{max}	Genotype
1	US White	2.3	1529A/370C
2	US White	4.3	1493G/?
3	US White	7.0	1529A/401A
4	US White	5.6	1529A/1529A
5	US White	3.4	1654A/721T
6	US White	7.4	1456T/307Cdel
7	US White	4.8	1456T/1153T*
8	US White	1.6	1529A/1529A
9	US White	2.4	1529A/1484T
10	US White	6.7	1529A/1276T
11	US White	1.6	1529A/1529A
12	US White	3.8	1529A/1529A
13	US White	5.1	1529A/476T
14	US White	4.6	1436A/1151T
15	US White	4.1	1574Gins/721T
16	US White	4.7	1484T/1378A
17	Portuguese	Hydrops	993A/1022A
18	Brazilian	9.9	1151T/1151T
19	Arabic	5.3	1456T/1168A
20	Asian	2.7	1456T/943A

*Newly described mutation.

measuring the gene product, which tends to be half the normal level in heterozygotes.^{8,9} However, there is always overlap between heterozygous and normal values and sometimes between those of heterozygous and homozygous subjects. Correcting for this overlap introduces a large error and, indeed, is often ignored, giving inaccurate estimates. The actual documentation of mutations at the DNA level is the most accurate way of identifying heterozygotes, but, unless complete sequence analysis is carried out, only the mutations that are selected for detection will be identified, and correction must be made for patients with other mutations.

In the case of PK deficiency, the only method that has been used to estimate population prevalence is that of detecting heterozygotes on the basis of red cell enzyme activity.²⁻⁴ Three of the most prevalent mutations in patients with PK deficiency are 1529A, 1456T, and 1468T; 1529A is most common in northern and central Europe,^{6,10} 1456T is most common in southern Europe,¹¹ and 1468T is most common in Asia. Each of these mutations is found in the context of its own haplotype, arguing that each has a unique origin. Each is in the C domain of the enzyme, involved in the intersubunit contact of the homotetramer catalytic unit.¹²

If the gene frequencies of the mutations that cause a disease are known, its prevalence may be estimated by applying the Hardy-Weinberg equilibrium. When this is done, the assumption is made that the penetrance of the clinical disorder is 100%. How well the clinical expression approaches this ideal can be estimated by comparing the relative frequency of the mutations that are encountered in the general population with the frequency in the patient population. We have previously studied the ratio of various Gaucher disease mutations in the Jewish population with Gaucher disease on the one hand and in the general population on the other.

Table 1. Oligonucleotide probes for allele-specific oligonucleotide hybridization

Mutation	Normal probe (wash temp)	Mutant probe (wash temp)
1456 C→T	5'-TTCTGTCT <u>CGGT</u> ACCGA-3' (50C)	5'-TTCTGTCT <u>TGGT</u> ACCGA-3' (48C)
1468 C→T	5'-TACCGACCT <u>CGGG</u> CAG-3' (52C)	5'-TACCGACCT <u>TGGG</u> CAG-3' (50C)
1484 C→T	5'-AGTCATTG <u>CTGT</u> CACCC-3' (52C)	5'-AGTCATTG <u>TTGT</u> CACCC-3' (50C)
1529 G→A	5'-CTTATGCC <u>GAGG</u> AGTCT-3' (50C)	5'-CTTATGCC <u>AAGG</u> AGTCT-3' (48C)

The nucleotides that are varied are shown in underlined bold font.

Table 3. PK mutant alleles found in 3785 persons of various ethnic groups

	White (non-Hispanic)	African American	Asian	Hispanic	Native American	Pacific Islanders	Unknown	Total
Alleles examined	5430	772	252	616	22	42	436	7570
Mutation								
1456T	19 (3.50 × 10 ⁻³)	3 (3.90 × 10 ⁻³)						22
1468T			2 (7.94 × 10 ⁻³)					2
1484T								
1529A	11 (2.03 × 10 ⁻³)							11

The number of heterozygotes found is followed by the gene frequency in parentheses.

Here we found marked overrepresentation of the 1226G mutation in the general population compared with the patient population.¹³ This implied that many of the persons who carry this particular mutation never came to medical attention as patients with Gaucher disease, and this is, indeed, the case.

A similar picture emerges in the case of PK deficiency. Table 4 summarizes the frequency of the 1529A mutation among patients with hemolytic anemia who were of European, non-Gypsy ancestry. (Gypsies have been excluded because they have a unique PK deletion¹²). Among this population, the 1529A mutation is the most common, accounting for 28.3% of all the deficiency-producing alleles; its frequency in the general white population is 0.00203. The 1456T mutation, on the other hand, represents only 15.4% of the disease-producing alleles in the patient population, but its frequency in the general white population, at 0.00350, is appreciably higher than that of the 1529A. It is of interest, in this respect, that the 1456T mutation is only rarely found in the homozygous state in patients with hemolytic anemia and that when the homozygous state has been documented, the anemia is very mild or does not exist at all.¹⁴ Further attesting to the relatively mild nature of this mutation, there is a tendency for it to be found together with null (nonsense) mutations rather than missense mutations.

However, within these limitations, population data can be used to estimate the prevalence of PK deficiency in the white population. The approach that we have adopted is to select an "index mutation." The frequency of this mutation in the patient population is taken to be p_i , and its frequency in the general population under investigation is taken to be g_i . The index mutation should have penetrance that approaches 100%, and it should be one of the more prevalent mutations in the patient population. In the case of PK deficiency, the 1529A mutation has been selected as the index mutation. It is the most common mutation among northern Europeans with PK deficiency, and its penetrance seems to be high because most homozygotes for this mutation have severe hemolytic anemia. Moreover, we know of no instances of siblings found in family studies to be homozygous for this mutation who did not also have hemolytic anemia. If the gene frequency of the index mutation, 1529A mutation, in the general population is g_i and the frequency of all other PK mutations is g_o , then the frequency of all homozygotes and compound heterozygotes for PK deficiency mutations is $g_i^2 +$

$2 g_i g_o + g_o^2$, where g_i^2 is the population frequency of homozygotes for the 1529A mutation, $2 g_i g_o$ is that of compound heterozygotes of 1529A mutations and other mutations, and g_o^2 is that of homozygotes and compound heterozygotes of all other mutations. In this study we found the value of g_i to be 0.00203, but how do we obtain the value of the other mutations, g_o ? If we assume that the representation of all mutations in the patient population is the same as in the general population, then the ratio of the index mutation to all other mutations in the general population will be $g_i/g_o = p_i/p_o$. Therefore, $g_o = p_o \times g_i/p_i$ where p_i is the frequency of the index mutation in the patient population and p_o that of all other PK mutations in the patient population. The fractional frequency of the occurrence of the 1529A mutation in the patient population, p_i , was found to be 0.283 (Table 4). The value of p_o , representing all other mutations in the patient population, is then $1 - 0.283$ or 0.717, and the value of g_o , the frequency of mutations other than 1529A in the general population, is then calculated to be 0.00514. Using the values of 0.00203 for g_i and 0.00514 for g_o , the value of the expression $g_i^2 + 2 g_i g_o + g_o^2$ is $4.12 \times 10^{-6} + 20.9 \times 10^{-6} + 26.4 \times 10^{-6}$, or 51 per million white population, approximately 10 000 patients in the United States. The standard error of the estimate of 51 per million can be assessed with the delta method,¹⁸ which uses a multivariate Taylor series to incorporate the sampling variability in both g_i and p_i into an overall estimate of the variance of g_i^2/p_i^2 because $g_i^2 + 2 g_i g_o + g_o^2$ reduces to this ratio. As expected from the relatively small numbers, this method provides a the standard error is 32.5 per million. This estimate would be increased by inbreeding, as in the Pennsylvania Amish community and in Gypsies, and would be decreased by failure of PK-deficient patients to survive. The number of patients actually identified seems to be smaller. In the past 25 years we have diagnosed PK deficiency in 201 patients. Because we do not know what percentage of all diagnoses in the United States were made in our laboratory, it is not possible to calculate the number of patients actually diagnosed in the United States. However, because there is a limited number of laboratories that carry out these assays, it appears that the actual cases of PK deficiency diagnosed falls far short of the 10 000 case estimate.

As we have pointed out above, the assumption that the distribution of mutations in the patient population is the same as in the general population is not entirely accurate; mild mutations such as 1456T are underrepresented among patients. Nevertheless, the approach we have used has the effect of correcting for those mutations that are underrepresented in the patient population because the estimate of the frequency in the general population depends on representation in the patient population. It thus has the effect of providing a fairly accurate picture of the patients in whom the disease actually develops.

Table 4. Alleles from European patients with clinical PK deficiency who have the 1456T or 1529A common mutations

Reference	Population	1456T (%)	1529A (%)
12	US White	5/50 (10)	25/50 (50)
10	European	6/58 (10.3)	23/58 (39.7)
15	German	0/10 (00.0)	2/10 (20.0)
11	Italian	7/30 (23.3)	3/30 (10.0)
16	Spanish	7/24 (29.1)	0/24 (00.0)
14	Portuguese	7/18 (38.9)	0/18 (00.0)
17	Southern Italy	3/14 (21.4)	2/14 (14.3)
Present study	White	2/36 (5.6)	13/36 (36.1)
Total		37/240 (15.4)	68/240 (28.3)

Acknowledgment

The authors thank Dr James Koziol for his help in calculating the estimate of error.

References

- Miwa S, Fujii H. Molecular basis of erythroenzymopathies associated with hereditary hemolytic anemia: tabulation of mutant enzymes. *Am J Hematol*. 1996;51:122-132.
- Blume KG, Löhr GW, Praetsch O, Rüdiger HW. Beitrag zur Populationsgenetik der Pyruvatkinase menschlicher Erythrocyten. *Humangenetik*. 1968; 6:261-265.
- Mohrenweiser HW. Frequency of enzyme deficiency variants in erythrocytes of newborn infants. *Proc Natl Acad Sci U S A*. 1981;78:5046-5050.
- Mohrenweiser HW. Functional hemizyosity in the human genome: direct estimate from twelve erythrocyte enzyme loci. *Hum Genet*. 1987;77: 241-245.
- Beutler E. *Red Cell Metabolism: A Manual of Biochemical Methods*. New York: Grune & Stratton; 1984.
- Baronciani L, Beutler E. Analysis of pyruvate kinase-deficiency mutations that produce non-spherocytic hemolytic anemia. *Proc Natl Acad Sci U S A*. 1993;90:4324-4327.
- Fried K. Gaucher's disease among the Jews of Israel. *Bull Res Council Israel. Proc Fourth Meeting of the Israel Genetics Circle*. 1958;7B:213.
- Grabowski GA, Dinur T, Gatt S, Desnick RJ. Gaucher type 1 (Ashkenazi) disease: considerations for heterozygote detection and prenatal diagnosis. In: Desnick RJ, Gatt S, Grabowski GA, eds. *Gaucher Disease: A Century of Delineation and Research*. New York, NY: Alan R. Liss; 1982:573-595.
- Matoth Y, Chazan S, Cnaan A, Gelernter I, Klibansky C. Frequency of carriers of chronic (type I) Gaucher disease in Ashkenazi Jews. *Am J Med Genet*. 1987;27:561-565.
- Lenzner C, Nürnberg P, Jacobasch G, Gerth C, Thiele BJ. Molecular analysis of 29 pyruvate kinase-deficient patients from central Europe with hereditary hemolytic anemia. *Blood*. 1997;89: 1793-1799.
- Zanella A, Bianchi P, Baronciani L, et al. Molecular characterization of PK-LR gene in pyruvate kinase-deficient Italian patients. *Blood*. 1997;89: 3847-3852.
- Baronciani L, Beutler E. Molecular study of pyruvate kinase deficient patients with hereditary non-spherocytic hemolytic anemia. *J Clin Invest*. 1995;95:1702-1709.
- Beutler E. Gaucher disease: new molecular approaches to diagnosis and treatment. *Science*. 1992;256:794-799.
- Manco L, Ribeiro ML, Almeida H, Freitas O, Abade A, Tamagnini G. PK-LR gene mutations in pyruvate kinase deficient Portuguese patients. *Br J Haematol*. 1999;105:591-595.
- Lakomek M, Huppke P, Neubauer B, Pekrun A, Winkler H, Schröter W. Mutations in the R-type pyruvate kinase gene and altered enzyme kinetic properties in patients with hemolytic anemia due to pyruvate kinase deficiency. *Ann Hematol*. 1994;69:253-260.
- Zarza R, Alvarez R, Pujades A, et al. Molecular characterization of the PK-LR gene in pyruvate kinase deficient Spanish patients. *Br J Haematol*. 1998;103:377-382.
- Pastore L, Morte RD, Frisso G, et al. Novel mutations and structural implications in R-type pyruvate kinase-deficient patients from Southern Italy. *Hum Mutat*. 1998;11:127-134.
- Rao CR. *Linear Statistical Inference and Its Applications*. New York: John Wiley & Sons; 1975.