A glucose-6-phosphate dehydrogenase (G6PD) splice site consensus sequence mutation associated with G6PD enzyme deficiency

Sean Sanders, Darrin P. Smith, Geraldine A. Thomas, E. Dillwyn Williams

Department of Histopathology, Level 5, Lab Block, Addenbrooke’s Hospital, Hills Road, Cambridge, CB2 2QQ, UK

CRC Human Cancer Genetics Research Group, Level 3, Lab Block, Box 238, Addenbrooke’s Hospital, Hills Road, Cambridge, CB2 2QQ, UK

Received 17 September 1996; accepted 10 October 1996

Abstract

A glucose-6-phosphate dehydrogenase (G6PD) deficient strain of mouse (GPDX) which was developed using the ethylating agent ethylnitrosourea (ENU) has been used to study clonality in epithelial tissues. While the biochemical defect has been quantified, the genetic basis of the deficiency is unknown. The G6PD gene is composed of 13 exons. Exon 1 is not translated, and the ATG start site is near the 5’ end of exon 2. Direct sequencing of the exonic regions of the gene from GPDX, C3H, 101, C57BL/6 and BALB/c mice was carried out. The coding region, in which (with a single exception) all mutations found to cause G6PD deficiency in man are situated, showed identical sequences in three of the four strains studied (101 coding region sequence was not examined). However, the G6PD gene in the GPDX mouse showed a single base difference from the other four strains and from the published mouse G6PD sequence (BALB/c) in the 5’ splice site consensus sequence at the 3’ end of exon 1, part of the untranslated region. The difference was confirmed in four different GPDX mice. This mutation was of the type (A to T transversion) that is known to be induced by ENU; its effect is likely to be exerted through a defect in transcription, splicing or translation, leading to a reduction in protein levels. By Western blot we have found a marked decrease in the G6PD protein levels in the GPDX mouse, with the C3H = GPDX heterozygote showing a lesser decrease. Recently, an increasing number of mutations in the untranslated regions of genes have been found which have effects on protein levels. We believe that the reduced enzyme activity in the GPDX mouse is due to the mutation in the 5’ untranslated region (UTR), and that similar mutations may be relevant in other inherited conditions.

Keywords: Glucose-6-phosphate dehydrogenase; Mutation; Splice site; Ethylnitrosourea; Mouse

1. Introduction

Over 400 phenotypic variants of human G6PD deficiency have been described [1,2]. Of these, approximately 300 have been typed by protein biochemistry using WHO guidelines [3]. The remaining 100 were found using alternative methods.
More recently, G6PD gene defects in humans have been examined by direct sequencing; 97 mutations or combinations of mutations have so far been described [2,4,5] of which the majority are missense mutations. Five deletion mutations have been described, three 3 base pair deletions ('G6PD Sunderland', ‘Urayasu’ and ‘Tsukui’) [4,6], a 24 base pair deletion ('G6PD Nara') [7] and a 6 base pair deletion ('G6PD Stoney Brook') [5]. All these deletions are multiples of three so that a frameshift, which is thought to be lethal, does not result. A single nonsense mutation has been found, 'G6PD Georgia', in which a tyrosine codon TAC is mutated to a stop TAA [5]. This type of mutation, which in this case results in the loss of 87 C-terminal amino acids, was thought to be lethal, but the fact that the patient, who had no known affected relatives, was a heterozygote may explain her survival [5].

With only one exception, all mutations have been found in the coding region (exons 2 through 13) of the gene. The exception is ‘G6PD Vansdorf’, which involves the deletion of the invariant AG dinucleotide at the 3' acceptor splice site at the intron 10/exon 11 boundary [5]. The exact effect of this mutation was not determined, but it is thought to cause either exon skipping or incorrect splicing [5]. The phenotype of this mutation is non-spherocytic haemolytic anaemia [2]. Whether the scarcity of mutations in non-coding regions is due to their rarity or the fact that these regions have not been examined is not clear from the published literature.

The GPDX strain was created in 1984 by Charles and Pretsch [8] as a potential model for the human deficiency. It was developed by treating male (101/E1 × C3H/E1)F1 mice with ethylnitrosourea (ENU), a well-known ethylating agent, and mating these animals with untreated ‘test-stock’ females [9]. One of the offspring showed a decreased blood G6PD level [10]. Stock from this animal showed erythrocyte G6PD activity of 20, 60 and 15% in hemizygotes, heterozygotes and homozygotes, respectively [10].

We have made use of this mouse strain to provide a model in which the clonality of tissue architecture and tumours can be studied [11–14]. However, the underlying genetic cause of this deficiency has not yet been elucidated.

The aim of this study was to identify the causative mutation in the GPDX strain by sequencing, and to investigate the means by which this mutation produces the GPDX phenotype.

2. Materials and methods

2.1. Animals

For PCR and sequencing, liver tissue was taken from the following animals: one male BALB/c, one male C3H, one female C57BL/6, one male 101, 3 males and one female GPDX, and a single female C3H × GPDX heterozygote. For Western blot analysis liver tissue from 2 male C3H, 3 male and 1 female GPDX, and 1 female C3H × GPDX heterozygote was used.

2.2. Extraction of DNA

A small piece of liver was homogenised in a sterilised glass homogeniser in 1 ml DNA Extraction Buffer [10 mM Tris-HCl (Sigma), pH 8.0; 100 mM ethylenediaminetetraacetic acid (Sigma), pH 8.0; 0.5% sodium dodecyl sulphate (lauryl sulphate; SDS; Sigma); and 20 μg/ml RNase A (Sigma; 10 mg/ml), added just before use]. Proteinase K (Sigma; 100 μl of 1 mg/ml solution) was added, and the samples mixed and incubated in a water bath for 3 h at 50°C. DNA was then extracted using a standard protocol [15] and the final product dissolved in TE pH 7.6–8.0; 10 mM Tris, 1 mM EDTA). DNA concentration and purity was determined, and samples were diluted 1:10 in preparation for PCR.

2.3. PCR

All primers were 20mers with a 50% GC content, and were commercially synthesised by R&D Systems (Abington, UK) and HPLC purified. The 5’ primers were 5’-biotinylated to aid with later separation for sequencing (see below). The other primers were not biotinylated. Primers used are listed in Table 1.

A total reaction volume of 50 μl was used: 2–3 μl DNA, 5 μl of each of the two primers (2 μM), 5 μl 10× Promega Buffer (500 mM KCl, 15 mM MgCl₂, 100 mM Tris-HCl, pH 9.0, and 1.0% Triton
X-100, 5 μl 2 mM dNTPs, 0.5 μl Taq (Promega; 5 U/μl), made up to 50 μl with water. All sample were covered with a drop (~50 μl) sterile oil (Sigma). For each PCR a ‘no DNA’ control was included. A ‘false’ hot start at 80°C for ~15 s was used, followed by 35 cycles of 95°C for 30 s, 55°C for 60 s and 72°C for 2 min. A final step of 5 min at 72°C allowed for completion of partial polymerisa-

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequencea</th>
<th>Position</th>
<th>Typeb</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>TATCCTACCATCTCTGGGCT</td>
<td>Exon 3/4</td>
<td>5' B</td>
</tr>
<tr>
<td>A2</td>
<td>CTTTGACCTTCTCATCAGG</td>
<td>Exon 8/9</td>
<td>3'</td>
</tr>
<tr>
<td>A3</td>
<td>GATGGCCCTCTACCTGAAGA</td>
<td>Exon 4</td>
<td>5'</td>
</tr>
<tr>
<td>A4</td>
<td>TCACTCGAACCCTGATGAGC</td>
<td>Exon 8</td>
<td>3' S</td>
</tr>
<tr>
<td>B1</td>
<td>CAGGGATGTTCATGCAAAC</td>
<td>Exon 7/8</td>
<td>5' B</td>
</tr>
<tr>
<td>B2</td>
<td>GCTGCGCATATACATAGGGGA</td>
<td>Exon 12/13</td>
<td>3'</td>
</tr>
<tr>
<td>B4</td>
<td>CTGGGCGTTTCTGAPCAA</td>
<td>Exon 12</td>
<td>3' S</td>
</tr>
<tr>
<td>C1</td>
<td>ATGACTGACAGAGCTGGAA</td>
<td>Exon 5/6</td>
<td>5' B</td>
</tr>
<tr>
<td>C2</td>
<td>TTGGTTGGAAGATGTCACCTG</td>
<td>Exon 10</td>
<td>3'</td>
</tr>
<tr>
<td>C4</td>
<td>TCGGAACCTCAAGCTACATT</td>
<td>Exon 10</td>
<td>3' S</td>
</tr>
<tr>
<td>D1</td>
<td>TCTGTAACGTGTCTTGGCAG</td>
<td>Exon 1</td>
<td>5' B</td>
</tr>
<tr>
<td>D2</td>
<td>AAGGGCTTTCACCACTGGAT</td>
<td>Exon 6</td>
<td>3'</td>
</tr>
<tr>
<td>D4</td>
<td>TCCACGCTGTGTCTGAATCAT</td>
<td>Exon 5/6</td>
<td>3' S</td>
</tr>
<tr>
<td>E1</td>
<td>GACTGTGTGGAGCTAGCAGA</td>
<td>Promoter</td>
<td>5' B</td>
</tr>
<tr>
<td>E2</td>
<td>GTCAAGCCGTAAGCTGATGTC</td>
<td>Promoter</td>
<td>3'</td>
</tr>
<tr>
<td>E4</td>
<td>ATGACACACTCCTTTATCTCT</td>
<td>Promoter</td>
<td>3'</td>
</tr>
<tr>
<td>E5</td>
<td>AGATTAGCGGCACTCAAGGG</td>
<td>Promoter</td>
<td>3'</td>
</tr>
<tr>
<td>E6</td>
<td>GCTTATGGAGCTGGAGAGAG</td>
<td>Promoter</td>
<td>3'</td>
</tr>
<tr>
<td>E7</td>
<td>CTTCCTAGCTGCTAGTTG</td>
<td>Promoter</td>
<td>3'</td>
</tr>
<tr>
<td>E8</td>
<td>TTATGGGGACCTGTCTCAGA</td>
<td>Promoter</td>
<td>3'</td>
</tr>
<tr>
<td>α-E7</td>
<td>CAAACTAGACGCTTGGGAAAG</td>
<td>Promoter</td>
<td>3'</td>
</tr>
<tr>
<td>F1</td>
<td>AAGTGAGACTGCGCTTCCGA</td>
<td>Exon 10</td>
<td>5' B</td>
</tr>
<tr>
<td>F2</td>
<td>GCAGCATAGGCAAAATGGC</td>
<td>3' non-coding</td>
<td>3'</td>
</tr>
<tr>
<td>F4</td>
<td>TCCACGCCGAGAACGTAGAA</td>
<td>3' non-coding</td>
<td>3' S</td>
</tr>
</tbody>
</table>

Table 1
Primers used for sequencing of GPDX gene

---

a All sequences read in 5’ to 3’ direction.
b B, 5’-biotinylated; S, used as sequencing primer.
tion products. Samples were removed from under oil and stored at $-20^\circ$C.

Products were checked by running on a 1.5% agarose gel at a constant 50 V for 45 min. Loading buffer (xylene cyanol and bromophenol blue; 0.5 $\mu$L) was added to 4 $\mu$L of sample and the entire 4.5 $\mu$L was loaded onto the gel. Marker was 200 ng ØX 174 RF DNA/HaeIII digest (HT Biotechnology Ltd., Cambridge, UK; 2 $\mu$L of 1:2.5 dilution in bromophenol blue and xylene cyanol loading buffer). Ethidium bromide (EtBr; 2 $\mu$g/ml) was added to the running buffer (Tris/borate/EDTA) to visualise the bands under UV light.

2.4. Sequencing

PCR was performed with the 5' primers which had been 5'-biotinylated. This allowed the PCR product to be manipulated using Dynal® M-280 streptavidin beads and a Dynal® Magnetic Particle Concentrator (MPC), following the protocols supplied with the product.

For direct sequencing, the protocol supplied with the Sequenase™ Version 2.0 DNA Sequencing kit (USB) was followed. Samples were run on 40 cm sequencing gels for between 1.3 and 5.5 h at a constant 40 W. All samples were denatured at 75°C for > 3 min prior to loading 5 $\mu$L, making use of the particle concentrator. Gels were dried under vacuum at 80°C for 1 h, loaded into an X-ray cassette and exposed for between 1 and 7 days using Fuji RX X-ray film.

2.5. Extraction of protein

A small piece of liver ($\sim 0.3$ g) was added to a chilled 1 ml glass homogeniser containing 1 ml lysis buffer (66 mm Hepes, pH 7.4; 20 mM NaCl; 1% glycerol; 1% Triton X-100; 2 mM MgCl2; 6 mM EDTA, pH 8.0; 10 $\mu$g/ml aprotinin; 2 mM phenylmethylsulphonyl fluoride; 1 mM benzamidine). Tissue was homogenised for 1–2 min, sonicated on ice twice for approximately 10 s and spun at 4°C, 13 000 x $g$ for 10 min. The supernatant was removed to a fresh tube and the protein concentration determined at 750 nm using a Bio-Rad Protein Assay kit. The samples were then diluted in water and 2 $\times$ gel loading buffer (120 mM Tris, pH 6.8; 4% SDS; 20% glycerol; 500 mM DTT; 0.005% bromophenol blue) to yield a final protein concentration of 0.5 mg/ml.

2.6. Western blot

Samples were run on a 10% polyacrylamide gel with a 3.5% stack for 75 min at 200 V. The protein

![Fig. 1. Polyacrylamide sequencing gel showing normal (BALB/c), heterozygote and GPDX sequences around the mutation site. Arrows indicate the site of the mutation (cf. Fig. 2). Intron sequences shown in lowercase, exon in uppercase, and sequence reads 5’-3’ from top to bottom.](image)
was blotted onto nitrocellulose (Hybaid) at 25 V for 30 min. Filters were stained with Ponceau-S and the positions of the markers (High Molecular Weight Markers; Sigma) noted. Blocking took place for 3 h at room temperature in 6 ml of block solution (5% non-fat milk powder, 3% BSA, 0.5% Tween-20) per filter.

Incubation in primary antibody was carried out overnight at 4°C and the filters washed. Two different primary antibodies were used: PM2 rabbit anti-human G6PD antibody (a gift from Dr. Philip Mason), used at 1:200, and RK rabbit anti-rat G6PD antibody (a gift from Dr. Rolf Kletzien), used at 1:400. As a control for loading, a mouse monoclonal anti-MAP kinase was used at 1:500 (Upstate). Secondary antibody (HRP-linked) incubation was performed at room temperature for 1 h, followed by washing. For the PM2 and RK antibodies, a swine anti-rabbit secondary was used, while for the anti-MAP kinase antibody, a sheep anti-mouse was used.

Antibody was visualised using a chemiluminescence system (Pierce) and performed according to

---

**Fig. 2.** Schema of 5′ end of the G6PD gene. The putative TATA box sequence is indicated and the 5′ splice site is underlined, with the splice site consensus sequence below [27], and the sequence in all other mouse strains examined above. Intronic or non-coding sequences are shown in lowercase, exonic in uppercase. The GPDX splice site mutation is in boldface type. The ATG start in exon 2 is indicated.

---

**Fig. 3.** Intron/exon boundaries sequenced. Intronic sequences in lowercase, exonic in uppercase.
the protocol. The signal using the RK antibody was stronger, possibly due to higher specificity, so a range of exposures between 20 s and 3 min was used.

3. Results

Through direct sequencing of PCR-amplified genomic DNA, we have found a single base difference (A-T to T-A) in the G6PD sequence in the GPDX strain compared to four others: BALB/c, C57BL/6, C3H and 101 (the latter two being the GPDX background strains; Fig. 1). This mutation is in the untranslated 3' end of exon 1, in the 5' splice site consensus sequence (Fig. 2). No other differences were found in the entire coding region of three of the strains (101 strain not examined), except for a single silent mutation in the C57BL/6 strain (C718 to A, coding for Gly221; data not shown). The base difference was confirmed in 3 other GPDX animals, and in a C3H × GPDX heterozygote. The intron/exon borders flanking exon 6, the 3' end of exon 10, and flanking exon 11 (Fig. 3), as well as part of the 3' untranslated region were examined for another purpose, and ruled out other splice site mutations or mutations affecting mRNA stability in these regions.

Western blot analysis (Fig. 4) of protein extracted from C3H and GPDX mice detected a specific protein band at ~58 kDa, in the correct range for the mouse G6PD enzyme. This same band was seen with two different antibodies, one raised against human (PM2) and the other against rat (RK) G6PD. The gel clearly shows a large decrease in the amount of G6PD protein in the GPDX mouse (Fig. 4A and B, lane 2) relative to the C3H (lane 1), when equal amounts of total protein were used, while the C3H × GPDX heterozygote showed an intermediate level of protein (Fig. 4A, lane 3). This experiment was repeated with extracts from different animals of the same strain, and identical results were seen (data not shown). The amount of protein loaded was controlled for using a mouse monoclonal anti-MAP kinase antibody (Fig. 4B and D).

This provides evidence that the lower G6PD activity in GPDX mice shown by histochemistry [11] and biochemistry [16] is not due to an active site mutation (and therefore maintained protein levels with decreased activity), but rather to lower protein levels. This is the expected finding for a mutation causing a decrease in transcription, splicing or translation.

4. Discussion

We have identified a mutation in the 5' untranslated sequence of the G6PD gene from a glucose-6-phosphate dehydrogenase deficient mouse strain (GPDX) by direct sequencing. This transversion mutation (A-T to T-A) was in the penultimate base of exon 1 (Fig. 3). It is compatible with the type of mutation induced by ENU, the alkylating base of males [8]. We have not formally excluded the
unlikely possibility that mutation arose by chance in the test-stock strain. However, ENU has been shown to induce A to T transversions in mouse germline cells [17–20], and it is highly probable that it induced the mutation observed.

Western blot analysis showed severely lowered levels of G6PD in the GPDX strain (Fig. 4A and C, lane 2) and intermediate levels in the C3H × GPDX heterozygote (Fig. 4A, lane 3). This result was confirmed using two different antibodies and multiple animals, and suggested that the GPDX phenotype was due to a reduced level of G6PD enzyme rather than a decrease in protein activity caused by a missense mutation in the coding sequence. An alternative but unlikely explanation was a missense mutation in the coding sequence. An alteration in the protein sequence. The observed phenotype must therefore have been due to lower protein levels.

We excluded a coding region mutation by demonstrating exact identity of the entire G6PD coding sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3. One exception of sequence from four GPDX mice with that of 3 other strains, and with the published mouse coding sequence; intron sequences were not determined, except in the regions shown in Fig. 3.

Work done on the relative lifetimes of spliceosome complexes bound to normal and mutant splice sites supports a model in which the usage of a certain splice site is influenced by the dissociation rate of the complex [30,37]. So, a mutated donor splice site could influence the dissociation (or association) rate of the spliceosome components, leading to a relatively lower splicing rate.

Another complicating factor to take into account is the position of the base change. Being at the end of the first exon, the machinery with which it interacts might well be different from that used at the other downstream splice sites. Ohno et al. [38] showed that the 5' cap structure found in all mRNAs appears to play a role in the splicing of the first intron [39].

The −2 position of the 5' splice site is evidently an important site. The evidence indicates that the consensus A at this position provides the most efficient splicing [24,29]. The presence of a T results in lower levels of splicing [29]. In addition, an A residue is found significantly more often in donor sites examined from both mouse and human genes than a T [27,40].

It is highly unlikely that intronic mutations in the invariant dinucleotides resulting in exon skipping or intron inclusion are present, as no difference in mRNA transcript size was seen in the GPDX strain after RT-PCR. Five further intron-exon boundaries were sequenced for another purpose, and all showed complete identity in all strains (Fig. 3). An additional mutation in an intron affecting mRNA stability remains a theoretical possibility, but the mutation we have identified is of the type induced by ENU, is restricted to the GPDX strain, and is in a highly conserved splice site region where a mutation would be expected to influence splicing efficiency.
In summary, we have identified a single base at the 3’ end of exon 1 present in all GPDX mice examined which differs from the sequence found in four other mouse strains; it is the only such difference. The entire coding region from three of the four strains (101 not examined), as well as partial intron sequences and part of the 3’ UTR, was examined. We believe that this mutation causes a decrease in the amount of enzyme through a splicing defect. We have demonstrated, using Western blot analysis, that there is a decrease in G6PD protein in mice carrying the GPDX mutation.

Of the 400 known phenotypes of human G6PD deficiency, only 97 have been accounted for on a genetic level [4]. The demonstration of a splice site mutation associated with low protein levels emphasises the importance of examining the untranslated regions of the gene in patients with G6PD deficiency for possible causative mutations. Missense mutations in highly conserved regions can usually be designated causative with a good amount of certainty. However, the interpretation of mutations in less well conserved regions should be carried out with more caution, and both coding and non-coding regions examined to avoid misinterpretation of simple polymorphisms as causative. Splice site mutations of the type found here may well be important in other heritable diseases in which the underlying genetic defect has not yet been found.

Acknowledgements

This work has been supported by grants from the Wellcome Trust (SS) and the BBSRC (GAT). We are very grateful to both Dr. Philip Mason and Dr. Rolf Kletzen for the kind gifts of G6PD antibodies. We would also like to acknowledge Mr. Chris Burton for photography.

References


