Sequence variations in the first exon of \(\alpha\)-galactosidase A

Joanna P Davies, Bryan G Winchester, Sue Malcolm

Abstract

The \(\alpha\)-galactosidase A gene (GALA), which is deficient in males with Anderson-Fabry disease, is shown to be remarkably polymorphic in the 5' untranslated region. GALA contains seven exons. The first exon contains 60 bp of 5' untranslated sequence before the methionine initiation codon. Single strand conformation polymorphism (SSCP) screening has shown three polymorphic variants from the published sequence within the 60 base pairs. The sequence changes involved are C to T at -10, G to A at -12 (which removes an MspI site), and G to A at -30 (which removes a SacII site). The combined frequency of these is 10%. A further insertion-deletion polymorphism is detected by SSCP of a 400 bp fragment including exon 3. Both polymorphisms can be easily detected using small polycrylamide gels and ethidium bromide staining. Nine of 20 women were informative for one of these polymorphisms and this simple SSCP analysis should be of great assistance in family studies of Anderson-Fabry disease. Such a high level of polymorphism has not been previously reported in the 5' untranslated region of a human gene and is unusual in any such short stretch of DNA.

(J Med Genet 1993;30:658–63)

\(\alpha\)-galactosidase (EC 3.2.1.22) is an X linked lysosomal hydrolase involved in the catabolism of glycoconjugates. It catalyses the hydrolysis of \(\alpha\)-galactosidic linkages at the non-reducing terminals of naturally occurring glycosphingolipids and glycoproteins, including blood group B substance, and of synthetic substrates. A detergent is required for the hydrolysis of glycosphingolipid substrates in vitro whereas in vivo this function is achieved by sialic acid activator protein 1 (SAP-1, another lysosomal protein). \(\alpha\)-galactosidase is synthesised as a precursor which undergoes post-translational modification en route to lysosomes from its site of synthesis on the rough endoplasmic reticulum. The mature native enzyme has a molecular weight of approximately 100,000 and is composed of two apparently identical subunits containing 7 to 15% carbohydrate.

In Fabry disease, a genetic deficiency of \(\alpha\)-galactosidase leads to the excretion in the urine and the accumulation within lysosomes of a variety of tissues of glycosphingolipids containing terminal \(\alpha\)-galactose residues. The characteristic clinical symptoms in affected males include angiookeratoma, pain in the extremities, renal failure, and heart and cardiovascular disease. These can be attributed to the progressive accumulation of storage products within the vascular endothelium. Death most often occurs around the age of 40 years.

The definitive diagnosis is the demonstration of a deficiency of \(\alpha\)-galactosidase in leucocytes or plasma. The range of \(\alpha\)-galactosidase activity in heterozygous females is very wide, overlapping that in both affected males and normal controls. This makes enzymic detection of carriers very difficult. Most heterozygotes are asymptomatic but may show some clinical symptoms, such as corneal opacity or angiokeratoma in later life. In a few heterozygotes the symptoms have been reported to be as severe as in affected males. This variability in biochemical and clinical expression owing to random X inactivation emphasises the need for molecular genetic tests for heterozygosity.

\(\alpha\)-galactosidase A has been cloned and the full length cDNA sequences reported. The gene contains seven exons extending over 12 kb. The complete nucleotide sequence has been reported. \(\alpha\)-galactosidase has a 60 base pair 5' non-translated region, but is unusual in having no 3' non-translated region. The polyadenylation signal is in the coding region 12 nt from the termination codon, which is followed by the poly(A) tract. The \(\alpha\)-galactosidase gene has been assigned to the long arm of the X chromosome (Xq22) using somatic cell hybrids and linkage analysis.

Only two RFLPs have been reported for GALA. An NcoI polymorphism is located 3' to the gene, approximately 10 kb downstream from the termination codon, with allele frequencies 0.87/0.13. A rare SacI polymorphism with allele frequencies 0.92/0.08 is located in intron 4. Closely linked polymorphic DNA sequences have also been used in carrier detection for Fabry disease, using markers around Xq22, such as DXS17 (a TaqI polymorphism) and DXS88 (BglII RFLP). However, few specific single base changes within the gene have been identified and out of 12 point mutations reported, only two are reported as non-disease causing. At exon 1, codon 8, an A to G change is silent in a triplet coding for leucine, and in exon 2, codon 66 a Glu to Gln base change is non-causative for Fabry's disease as shown by expression of this mutant in CD5 cells. The others are described as disease causing either because whole cDNA or exon sequencing has shown no other mutations or because expression studies show that the mutation causes a reduction in enzyme activity.

No specific nucleotide variation...
changes have been described in the 5' untranslated region of the gene.

SSCP is a convenient method for screening stretches of DNA for sequence changes. It is based on the principle that denaturation and rapid cooling, single stranded DNA molecules will adopt unique sequence dependent conformations which can be separated by electrophoresis in native polyacrylamide gels.

Materials and methods

RADIONLABELLED SSCP ANALYSIS OF EXONS 1 AND 3

DNA samples were prepared from anticoagulated whole blood (in EDTA) in a process involving the salting out of cellular proteins by dehydration and precipitation with saturated ammonium acetate. Red blood cells were lysed by adding water to 5 to 10 ml of whole blood to give a final volume of 50 ml and centrifuging at 2300 rpm for 20 minutes. The nuclear pellet was then washed with 25 ml of 0.1% NP40 and centrifuged for a further 20 minutes. The nuclei were lysed and proteinase K digested as previously described. Then 1 ml of saturated ammonium acetate (9-6 mol/l) was added and the sample centrifuged for 20 minutes to precipitate protein. The DNA was precipitated from the supernatant by the addition of 2 volumes of absolute ethanol and was redissolved in 1 ml of 10 mmol/l Tris, 1 mmol/l EDTA solution.

Oligonucleotide primers were designed using the PRIMER program provided by the HGMP resource centre based at MRC Clinical Research Centre, Harrow, UK. The exon 1 primers were 5'GTCCCCAGTTGGCAGAGAAAC3' and 5'AAAGGGGAAGGAGTACCCGAAG, located from nucleotide position 1054-1073 and 1403-1384. The exon 3 primers were 5'GATGATTTTGGGTTTGTG3' and 5'GATTTGGTTCCTTGTGGCAGCC3' located at 7123-7142 and 7522-7503. Primers were synthesised on an Applied Biosystems 381A DNA synthesiser and 5'-biotinylated was carried out during oligonucleotide synthesis using DMT-biotin-C-6-PA(CRB) according to the manufacturer's instructions.

PCR was conducted in a total volume of 50 ml and each reaction mixture contained genomic DNA, 40 pmol of each primer, 0.2 mmol/l dATP, ddGTP, dTTP, 0.02 mmol/l dCTP, 1 μCi of 32P dCTP, in 1-5 mmol/l MgCl2, 50 mmol/l KCl, 10 mmol/l Tris-HCl (pH 8.8), 1% Triton X-100. Mineral oil (50 μl) was added to each reaction mixture before an initial 94°C denaturing step, followed by the addition of 1 unit of Taq Polymerase (Promega) at 58°C. Amplification was carried out by 30 cycles of elongation at 72°C (one minute), denaturation at 94°C (30 seconds), and annealing at 58°C (30 seconds). The amplification was completed by a final 10 minute elongation step at 72°C. The products (350 and 400 base pairs for exons 1 and 3 respectively) were checked by running 5 μl of each product on a 2% agarose gel with ethidium bromide (1 mg/ml) staining.

A volume of 5 μl of PCR product was diluted by the addition of 40 μl of 0.1% SDS, 10 mmol/l EDTA solution and 2 μl of diluted product was mixed with 2 μl of 95% formamide, 0.05% xylene cyanol, 0.05% bromphenol blue, 20 mmol/l EDTA solution (Sequenase kit stop solution, USB). The samples were then denatured by heating at 94°C for three minutes and placed on ice before loading onto a 0.4 mm thick 30 × 40 cm, 6% polyacrylamide gel (Protogel: 30% acrylamide to 0.8% bisacrylamide, National Diagnostics), with 5% glycerol.

Electrophoresis was carried out in 0.09 mmol/l Tris-borate, 0.002 mmol/l EDTA (pH 8.3) buffer (TBE), at 340 V at room temperature for approximately 15 hours, until the xylene cyanol dye had migrated 31 cm. The gel was dried onto 3MM Whatman paper under vacuum at 80°C and exposed (without intensifying screens) to x ray film (Kodak, IBI Ltd) for one to five days at −70°C.

NON-RADIONLABELLED SSCP

PCR was conducted in a volume of 50 μl as described above but using unlabelled dNTPs (0.2 mmol/l of each). A volume of 15 μl of each product was mixed with 5 μl of loading buffer (15% sucrose, 0.05% xylene cyanol) before denaturing (three minutes at 94°C) and placing on ice.

The entire volume of each sample was loaded onto a 1 mm thick, 16 × 20 cm polyacrylamide gel and electrophoresis was carried out using the Protein II gel system (BioRad, UK) in TBE buffer. For exon 1 a 6% polyacrylamide gel (Protogel: 30% acrylamide, 0.8% bisacrylamide, National Diagnostics) with 5% glycerol was run at 100 V overnight. For exon 3, the gel was 12% polyacrylamide (Protogel), without glycerol run at 200 V overnight.

SEQUENCE ANALYSIS

PCR amplification of exon 1 was performed as previously described, except that 5 pmol of each primer was used in 100 μl total reaction mixture volume and unlabelled dNTPs were used (0.2 mmol/l of each). The 3' primer was biotinylated at its 5' end allowing production of single stranded DNA using magnetic streptavidin coated beads (Dynal UK). Sequencing was carried out as follows: 50 μl of PCR product was mixed with 30 μl of magnetic Dynal W280 streptavidin beads and incubated for five minutes allowing the double stranded DNA to bind via a streptavidin-biotin bond. The supernatant was removed while the PCR product was attracted to a magnet (Promega) and the beads (with DNA) were then washed twice with 100 μl of TES (10 mmol/l Tris-HCl, pH 8.0, 1 mmol/l EDTA, 0.1 mmol/l NaCl). After a five minute denaturing step with 0.15 mol/l NaOH at room temperature, the non-biotinylated sense strand was removed and discarded, leaving the single stranded, antisense DNA bound to the bead and magnet. This was again washed twice with...
gels and polymorphisms were products of variants. SSCP
6% pattern I
Figure are 4 (M respectivly. radiolabelled PCR DNA non-labelled 660 lPl, Gibco with temperature; with ethidium 15°C run at 1P3, in (P1, (1 kb
polyacrylamide 5% 100 are shown. Exon 3 polymorphism and stained
(Tes
was
amid, National Diagnostics), 8·3 mol/l urea, denaturing gel. Electrophoresis was performed at 55 W for about two hours in TBE buffer using the BRL S2 sequencing gel apparatus.
The gel was fixed in 10% methanol and 10% acetic acid solution for 10 to 30 minutes before blotting and drying as described previously for the SSCP gels.

RESTRICTION ENZYME ANALYSIS
PCR amplification of exon 1 was carried out as previously described for non-radioactive SSCP analysis. A volume of 10 µl of PCR product was digested in a total volume of 20 µl, using MspI or SacII and the appropriate enzyme buffer solution (NBL). The entire sample was analysed using a 3% (2% agarose, 1% NuSieve agarose) for MspI or 4% (2% agarose, 2% NuSieve agarose) for SacII. The gel was stained with ethidium bromide (1 mg/ml).

Results
SSCP ANALYSIS OF EXONS 1 AND 3
Exon 1 of α-galactosidase A was amplified from DNA of 61 unrelated normal males and 20 normal females. Four band patterns were observed using SSCP analysis (fig 1). DNA from a subject with each of the three minor band patterns was sequenced.

Two SSCP patterns were observed in PCR amplified material from exon 3 (fig 2). Both polymorphic systems could be visualised using 16 × 20 cm polyacrylamide gels stained with ethidium bromide.

SEQUENCE ANALYSIS OF POLYMORPHIC VARIANTS
Three single base changes were observed. These were G to A at nucleotide 30 before the ATG initiation codon (1P3), G to A at −12 (1P1), and C to T at −10 (1P2) (fig 3).
1P1 removes a restriction enzyme site for MspI. This can be detected by MspI digestion of PCR products. Fig 4 illustrates the MspI

Figure 3 Sequence analysis of exon 1 SSCP variants. Direct sequencing of exon 1 polymorphisms using Dynal bead technology (Dynal, UK) and the Sequenase II kit (USB). Samples were run on a 6% polyacrylamide denaturing (8·3 mol/l urea) gel at 55 W. Sequencing gels for the sense strand of DNA from normal males with polymorphisms 1P1 and 1P2 are shown.
Sequence variations in the first exon of α-galactosidase A

site variant segregating in a family with Anderson-Fabry disease. 1P3 removes a SacII site and segregation of this polymorphism is illustrated in a normal family in fig 5.

The polymorphism in the amplified product from exon 3 was shown to be an insertion-deletion polymorphism by BstNI digestion of the PCR product. Predicted band sizes were 39 bp, 31 bp, 144 bp, and 186 bp. The 31 bp fragment, which falls 80 bp into intron 2, was reduced in size in subjects carrying the polymorphism. This was confirmed by Hindl digestion of the fragment. DNA sequencing showed a deletion of 7 base pairs between nucleotides 7184 and 7191 (CCCCCAGCC) presumably owing to replication stuttering across the stretch of Cs. It is not possible to determine the exact breakpoints because of the Cs at either end.

**POLYMORPHISM FREQUENCY**

One hundred and one X chromosomes were analysed, 61 from males and 40 from females. The frequencies are presented in the table. The combined frequencies of the three polymorphisms in exon 1 were 10/101 (10%). The polymorphism in exon 3 was present in 11/101 chromosomes (12%).

Of the 20 females analysed, five were heterozygous for one or more of the exon 1 polymorphisms and six were heterozygous for the exon 3 polymorphism. Nine of the women were informative using either exon 1 or 3 SSCP polymorphisms, indicating that these variants would be of use in the genetic analysis of Anderson-Fabry disease or in mapping studies.

**ANALYSIS OF A FAMILY WITH ANDERSON–FABRY DISEASE**

SSCP analysis of exons 1 and 3 was carried out in a family with two males affected with

<table>
<thead>
<tr>
<th>Frequency of X chromosome variants</th>
<th>Exon 1</th>
<th>Exon 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variant</td>
<td>Frequency</td>
<td>Variant</td>
</tr>
<tr>
<td>1P1</td>
<td>3</td>
<td>3P1</td>
</tr>
<tr>
<td>1P2</td>
<td>5</td>
<td>Normal</td>
</tr>
<tr>
<td>1P3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>101</td>
</tr>
</tbody>
</table>
Anderson-Fabry disease (fig 6). Two variants were found to be segregating in the family. The carrier mother was heterozygous for 1P1 and 3P1. Her affected brother and son each carried the mutant 3P1 allele but not the 1P1 allele. Her daughter has inherited the unaffected allele (with the 1P1 polymorphism but not 3P1) and is not a carrier of Anderson-Fabry disease.

Discussion
We describe a multiallelic polymorphic system within the α-galactosidase A gene. It can be readily visualised using a single PCR amplification and SSCP analysis, with either radioactive or ethidium bromide staining. A second two allele system, also detected by SSCP, increases the informativeness to a clinically useful level.

The variation is remarkable in that three point mutations are found within the 60 base pairs of 5' non-translated sequence. With the increased emphasis on sequencing of mutant genes, an increasing number of polymorphic base changes are being found within coding regions. These may involve single base silent substitutions or apparently neutral amino acid changes. It has been suggested that 3' non-translated regions may provide a rich source of polymorphic variation. These results suggest that 5' non-translated regions could also provide very useful polymorphic variation. Few such variants have been described, a striking exception being the highly variable stretch of (CGG) repeats within the 5' non-translated region of FMR-1 which, when expanded, gives rise to the fragile X syndrome.

A short region of 9 bp immediately before the AUG initiation codon has been shown to be important for ribosome initiation. These results imply that sequences further away from the initiation site may have relatively little, if any, functional effect.

Traditional restriction fragment length polymorphisms depend on base changes occurring within the recognition site for a restriction enzyme. SSCP analysis allows the detection of single base changes occurring outside restriction sites and can increase greatly the number of polymorphisms available. The factors controlling single strand conformation, and thus the best conditions to be used for detection, are not understood. It is noteworthy that three base changes occur within 20 bases of each other (two within 2 base pairs) but are all readily detected and each gives a distinctive pattern. This shows that the factors controlling the preferred conformation must be subtle.

Very few polymorphisms have previously been reported within the α-galactosidase A gene, making DNA studies in families with Anderson-Fabry disease difficult. Biochemical detection of female heterozygotes by the measurement of the α-galactosidase activity, usually relative to an unaffected lysosomal enzyme such as β-D-galactosidase or N-acetyl-β-D-hexosaminidase, in plasma, leukocytes, or fibroblasts is unreliable because of the random inactivation of the X chromosome. It is possible to show the cellular mosaicism associated with females heterozygous for an X linked disorder by assaying α-galactosidase in cloned skin fibroblasts. However, this is a very time consuming and technically difficult procedure. Alternatively, α-galactosidase may be measured in individual hair follicles. The detection of hair roots with normal, negligible, and intermediate levels of activity from a female is diagnostic for a heterozygote. This is also a time consuming procedure and does not give complete discrimination, especially when the activities are low with fine hair. Therefore, molecular genetic analysis will provide an important additional detection method. The polymorphisms described here are easy to use and sufficiently frequent to be of clinical use in carrier detection. The example shown in fig 6, where family member 5 can be excluded from being a carrier, illustrates this.

In the past, the majority of SSCP detection methods have used radiolabelled PCR pro-
Sequence variations in the first exon of α-galactosidase A

products. However, from the results shown here it is clear that single base changes can be effectively distinguished, even when changes occur in close proximity to each other (such as polymorphisms IP1 and IP2) without the need for radiolabelling. Results can be obtained immediately after running a polyacrylamide gel by ethidium bromide staining and on a small gel system, thus increasing the speed and safety of SSCP as a mutation detection method.

JD is a Wellcome prize student. We gratefully acknowledge financial support from the Research Trust for Metabolic Diseases in Children and the Fabry’s Disease Research Fund. We wish to thank Paul Rutland for oligonucleotide synthesis.