Edith Cowan University
Research Online

Theses : Honours Theses

2000

# Molecular investigations in the role of the GALK1 Gene in Galactokinase Deficiency

Michael L. Hunter
*Edith Cowan University*

Follow this and additional works at: https://ro.ecu.edu.au/theses_hons

Part of the Genetic Phenomena Commons

Recommended Citation
Hunter, M. L. (2000). *Molecular investigations in the role of the GALK1 Gene in Galactokinase Deficiency*. https://ro.ecu.edu.au/theses_hons/344

This Thesis is posted at Research Online.
https://ro.ecu.edu.au/theses_hons/344

// Edith Cowan University

## Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.

- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).

- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

# USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

# Molecular Investigations into the Role of the *GALK1* Gene in Galactokinase Deficiency

by

Michael Hunter

A Thesis Submitted in Partial Fulfilment of the
Requirements for the Award of
Bachelor of Science, Honours (Human Biology).

At the Faculty of Communications, Health and Science, Edith Cowan University,
Joondalup.

Date of submission: 6th November, 2000

## ABSTRACT

Galactokinase deficiency is an autosomal-recessive inborn error of galactose metabolism whose major clinical manifestation is the development of cataracts during the first months of life. This metabolic disorder is caused by defects in the first enzyme of the Leloir pathway, galactokinase, encoded by the gene *GALK1* on chromosome 17q24. Despite the identification of a number of conserved domains in *GALK1*, understanding of the functional significance of these regions and the molecular basis of the disorder is limited. This is largely due to the rarity of the disease and the fact that the small number of *GALK1* mutations identified to-date are confined to individual families, thus precluding extensive genotype/phenotype correlations.

The current investigation involved an analysis of the *GALK1* gene in nine patients with biochemical phenotypes indicative of classical galactokinase deficiency. Sequencing of the entire coding region, flanking intronic sequence and both the 5' UTR and 3'UTR of *GALK1* using Dye Terminator chemistry and the ABI377 DNA Analyzer, revealed four novel mutations in two unrelated patients with galactokinase deficiency. Three of these were amino acid substitutions: 1569C→T in exon 2 (R68C), 7093C→T in exon 6 (T288M) and 7538G→C in exon 8 (A384P). In addition, a single base deletion was found in exon 5 (2833delC), predicted to result in a shift of the reading frame and a premature termination codon at position 263. Sequence analysis in a third patient detected a 563C→A transversion (P28T), previously identified as a common mutation in six Bulgarian Gypsy families (Kalaydjieva et al., 1999). The remaining six patients, belonging to socially and geographically dispersed Gypsy groups from Spain, Hungary and Bulgaria were screened for P28T using a PCR-based AvaI restriction assay. All were homozygous for P28T. Haplotype analysis established a common origin of the mutation, providing further

evidence that P28T is a founder mutation among patients with Gypsy ethnicity and signaling the mutation is more widespread than initial indications suggest.

To determine the carrier frequency and distribution of the P28T mutation, 227 unrelated individuals originating from various Gypsy groups in Europe, were screened using the AvaI assay. Carrier frequency in the Vlax Roma was calculated at 2.7% and 1.8% in Spanish Gypsies. These results indicate that the prevalence of the P28T mutation warrants further investigation preceeding the development of newborn screening strategies and dietary intervention in these groups in a bid to prevent infantile blindness. The findings also allow the identification of a large sample of carriers suitable for future investigations into genotype/phenotype correlations.

Sequence analysis of the *GALK1* gene in two Afro/Hispanic patients presenting with a rare atypical phenotype with metabolic symptoms of galactokinase deficiency and normal RBC galactokinase activity, revealed no plausible disease-causing mutations. In fact, lack of similarity in the *GALK1* gene in both patients, despite sharing a common ethnic background, suggested exclusion of *GALK1* in this phenotype.

## DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material.

Signature_

Date 6th NOVEMBER 2000

## ACKNOWLEDGMENTS

I wish to extend thanks to a number of people and organizations who have made this research possible. First and foremost I would like to thank my supervisor Assoc./Prof. Luba Kalaydjieva, whose guidance and contribution of knowledge and expertise towards the topic was an inspiration. Also, I would like to thank my co-supervisor Dr Dora Angelicheva, for helpful technical assistance, suggestions and support in the lab.

I would also like to acknowledge the contribution of a number of collaborating institutions who provided patient samples and data. These include the Genetic Service, Children's Hospital and Department of Pediatrics at Harvard Medical School, Boston, USA, the University Children's Hospital in Zurich, Switzerland, the Pediatric Clinic at the University of Szeged Medical School, Hungary and the Institut Bioquimica Clinica in Barcelona, Spain. Funding for the research was provided by the Wellcome Trust (Grant 058898/Z/99/GEK/SRD).

A big thankyou also to the entire team at the Centre for Human Genetics, Edith Cowan University, Joondalup for providing a friendly, supportive and stimulating environment in which to work and learn. Special mention also to Rebecca Gooding for her assistance with sequencing and Marie McCluskey for her helpful suggestions along the way.

Finally, special mention must also go to my family, and my partner and best-friend Naomi, whose encouragement, understanding, support and patience made surviving my honours year and completing this thesis possible.

TABLE OF CONTENTS


TITLE PAGE i

ABSTRACT ii

DECLARATION iv

ACKNOWLEDGMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER 1 INTRODUCTION 1

1.1 Overview 1

1.2 Significance of the study 5
- 1.2.1 Adding knowledge to the structure and function of GALK1 5
- 1.2.2 Understanding the mechanisms of spread of common founder mutation(s) in *GALK1* and identification of high-risk groups 5
- 1.2.3 Possible links to presenile cataract formation 5

1.3 Research aims 6

1.4 Operational terms and definitions 7

CHAPTER 2 LITERATURE REVIEW 8

2.1 Galactokinase 8
- 2.1.1 Galactokinase and the Leloir pathway 8
- 2.1.2 Age and tissue-related differences in galactokinase expression 9
- 2.1.3 Differences in the physical properties of the enzyme 11
- 2.1.4 Allelic variation and genotype/phenotype differences 12

2.2 Galactokinase Deficiency 13
- 2.2.1 Clinical synopsis and symptoms 13
- 2.2.2 Epidemiology 14
- 2.2.3 Possible role of galactokinase deficiency in presenile cataract formation 16

2.3 The *GALK1* gene: Localization and Structure 17
- 2.3.1 Localization 17
- 2.3.2 Characterization 18
- 2.3.3 5'UTR and promoter region 21

2.4 Mutations in *GALK1* associated with enzyme deficiency 23

2.6 Summary 28

CHAPTER 3 IDENTIFICATION OF *GALK1* MUTATIONS IN PATIENTS WITH GALACTOKINASE DEFICIENCY 29

3.1 SUBJECTS 29

3.2 METHODS 31
3.2.1 Experimental Design 31
3.2.2 DNA Quantification 31
3.2.3 PCR Amplification 32
3.2.4 BigDye Terminator Cycle Sequencing 35
3.2.5 Ethanol/Sodium Precipitation 36
3.2.6 Sequencing Gel Preparation 36
3.2.7 Electrophoretic Conditions 37
3.2.8 Web-Based Databases Utilized in the *GALK1* Analysis 37
3.2.9 Mutation Detection Techniques 39

3.3 RESULTS 43
3.3.1 Patient 1 43
3.3.2 Patient 2 50
3.3.3 Patients homozygous for the P28T mutation 56
3.3.4 Non-disease causing nucleotide variations 59

3.4 DISCUSSION 60
3.4.1 Nature of the newly identified mutations 60
3.4.2 Nucleotide variations 62
3.4.3 General discussion of mutation in the GALK1 gene 62
3.4.4 Future Directions 65

CHAPTER 4 ANALYSIS OF *GALK1* IN PATIENTS WITH THE ATYPICAL PHENOTYPE 66

4.1 INTRODUCTION 66

4.2 SUBJECTS 66

4.3 METHODS 67

4.4 RESULTS 68
4.4.1 Sequence analysis of the GALK1 gene 68

4.5 DISCUSSION 71
4.5.1 Sequence analysis 72
4.5.2 Other aberrations to galactose metabolism 73
4.5.3 Conclusions 75

CHAPTER 5 ANALYSIS OF THE ORIGINS, DISTRIBUTION AND FREQUENCY OF A COMMON *GALK1* MUTATION CAUSING GALACTOKINASE DEFICIENCY 77

5.1 INTRODUCTION 77

5.2 SUBJECTS 78
5.2.1 Haplotype analysis determining the origin of the P28T mutation 78
5.2.2 Population Screening for the P28T mutation 79

5.3 METHODS 81
5.3.1 Haplotype analysis 81
5.3.2 Population screening for the P28T mutation 83

5.4 RESULTS 84
5.4.1 Haplotype analysis 84
5.4.2 Population screening and Carrier Frequency of P28T 87

5.5 DISCUSSION 88
5.5.1 Haplctype analysis 88
5.5.2 Population screening and Carrier Frequency of P28T 88

CHAPTER 6 SUMMARY, CONCLUSIONS AND FUTURE DIRECTIONS 89

REFERENCES 91

APPENDIX A 96

## LIST OF TABLES


| | | Page |
|---|---|---|
| Table 2.1 | Galactokinase Properties in Various Tissue | 10 |
| Table 2.2 | Galactokinase Properties in Adult and Fetal Erythrocytes | 11 |
| Table 2.3 | Variation in Incidence and Allele Frequency | 15 |
| Table 2.4 | Mutations Identified in *GALK1* | 25 |
| Table 3.1 | Biochemical Findings of Study Sample | 29 |
| Table 3.2 | PCR Conditions and Protocols for *GALK1* | 33 |
| Table 3.3 | *GALK1* Mutations in Patients with Galactokinase Deficiency | 43 |
| Table 3.4 | Nucleotide Differences Between Published *GALK1* Sequences | 59 |
| Table 4.1 | Biochemical Findings of Study Sample | 66 |
| Table 4.2 | *GALK1* Nucleotide Differences in Atypical Patients | 68 |
| Table 5.1 | Individuals with the P28T Mutation | 79 |
| Table 5.2 | Individuals Screened for the P28T Mutation | 80 |
| Table 5.3 | PCR Conditions and Protocols for Microsatellite Markers | 81 |
| Table 5.4 | Population Screening for P28T Carrier Frequency | 87 |

# LIST OF FIGURES


| | | Page |
|---|---|---|
| Figure 1.1 | Cataract in child with galactokinase deficiency | 2 |
| Figure 2.1 | The role of galactokinase in the Leloir pathway | 8 |
| Figure 2.2 | Functionally important domains and conserved regions | 20 |
| Figure 2.3 | *GALK1* 5'UTR | 22 |
| Figure 2.4 | Distribution of mutations in *GALK1* | 24 |
| Figure 3.1 | 1569C→T transition in *GALK1* exon 2 | 44 |
| Figure 3.2 | Residue change at codon 68 | 44 |
| Figure 3.3 | Molecular comparison between arginine and cysteine | 44 |
| Figure 3.4 | Secondary-structural prediction for R68C | 45 |
| Figure 3.5 | AlwNI restriction digest | 46 |
| Figure 3.6 | 7538G→C transversion in *GALK1* exon 8 | 47 |
| Figure 3.7 | Residue change at codon 384 | 47 |
| Figure 3.8 | Molecular comparison between alanine and proline | 48 |
| Figure 3.9 | Secondary-structural prediction for A384P | 48 |
| Figure 3.10 | BbvI restriction digest | 49 |
| Figure 3.11 | 7093C→T transition in *GALK1* exon 6 | 50 |
| Figure 3.12 | Residue change at codon 288 | 51 |
| Figure 3.13 | Molecular comparison between threonine and methoinine | 51 |
| Figure 3.14 | Secondary-structural prediction of T288M | 51 |
| Figure 3.15 | Galactokinase fingerprint motif four | 52 |
| Figure 3.16 | NlaIII restriction digest | 53 |
| Figure 3.17 | 2833delC deletion in *GALK1* exon 5 | 54 |

| | | |
|---|---|---|
| Figure 3.18 | Frameshift at codon 237 | 54 |
| Figure 3.19 | Novel protein sequence in exon 5 | 55 |
| Figure 3.20 | HDA and SSCP detection of 2833delC | 55 |
| Figure 3.21 | 563C→A transversion in *GALK1* exon 1 | 57 |
| Figure 3.22 | AvaI restriction digest | 58 |
| Figure 3.23 | Distribution of mutations in *GALK1* | 64 |
| | | |
| Figure 4.1 | 7087G→A transition in *GALK1* exon 6 | 69 |
| Figure 4.2 | Residue change at codon 286 | 69 |
| Figure 4.3 | Molecular comparison between arginine and glutamine | 70 |
| Figure 4.4 | Secondary-structural prediction of R286Q | 70 |
| Figure 4.5 | Galactokinase fingerprint motif four | 71 |
| | | |
| Figure 5.1 | Pedigree of Spanish Gypsy family | 79 |
| Figure 5.2 | Pedigrees and marker haplotypes for patients with P28T | 85 |
| Figure 5.3 | Marker haplotypes in the 17q24 region | 86 |

# CHAPTER 1 INTRODUCTION

## 1.1 Overview

Galactokinase deficiency (Gitzelmann, 1965; Gitzelmann, 1967) is an autosomal recessive inborn error of the first step of the Leloir metabolic pathway through which galactose enters glycolysis (Frey, 1996). Unlike the severe and debilitating symptoms associated with other forms of galactosemia (Segal & Berry, 1995), the only confirmed symptom of galactokinase deficiency is the development of cataracts (Figure 1.1). In homozygotes, these occur within the first weeks to months of life (Levy, Krill, & Beutler, 1972; Beutler et al., 1973; Stambolian, 1988). There is also evidence to suggest that heterozygote carriers of a deficiency causing allele may be at risk of developing presenile cataracts (Elman, Miller, & Matalon, 1986; Stambolian et al., 1986). The progression to sight impairment is caused by the accumulation of an alternative product of galactose metabolism, called galactitol, which accumulates within the lens of the eye, ultimately leading to the rupture and opacification of lens fibers. Fortunately, with early detection and intervention with a galactose-restricted diet, cataract formation in affected individuals can be prevented and the pathological changes in the lens can be reversed in some cases (Olambiwonnu, McVie, Ng, Frasier, & Donnell, 1974).

Elucidating the molecular basis underlying galactokinase deficiency began in 1995 with the cloning of the gene *GALK1* (Stambolian et al., 1995) and its mapping to the chromosomal location 17q24 (Stambolian et al., 1995; Kalaydjieva et al., 1999). Despite this, only a limited amount of knowledge exists in regards to the functional organization of *GALK1* and conclusive information about the genetic regulation of galactokinase expression is yet to be fully obtained. Investigations of the molecular basis of galactokinase deficiency can contribute such information. However, these studies are hampered by the low frequency of the disorder and by the fact that the small number of identified mutations

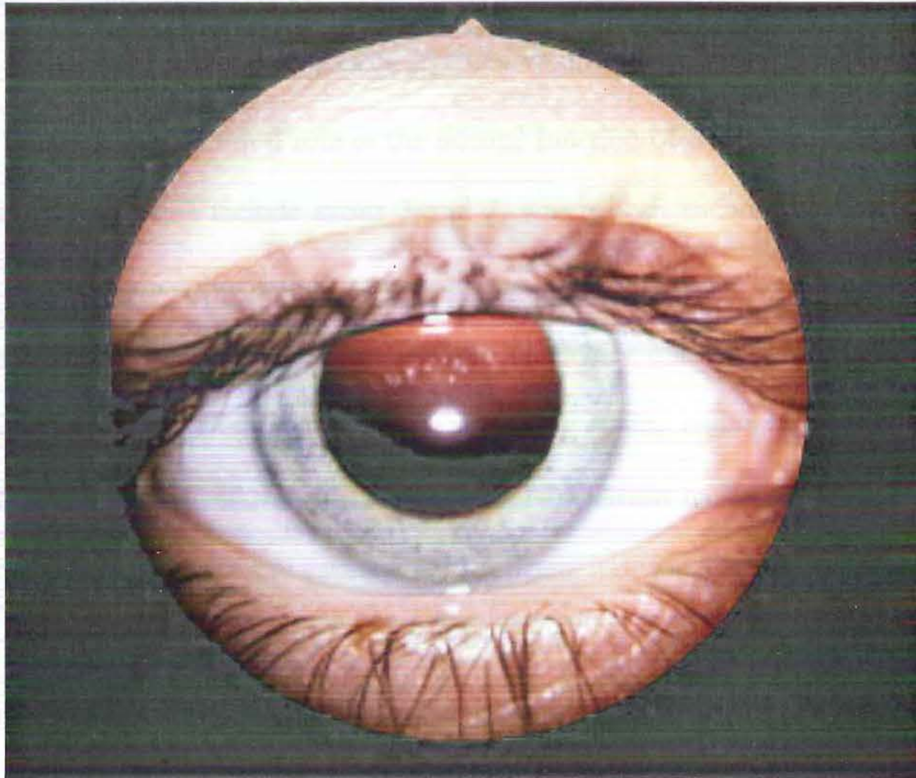


**Figure 1.1** "Oil-droplet"cataract in three-month-old child with galactokinase deficiency. (Source: Campolattaro, 2000)

are confined to single families, thus precluding reliable genotype-phenotype correlations.

What is known is that the *GALK1* gene spans approximately 7.3kb of DNA and is arranged into 8 exons that code for a mRNA transcript of 1.35kb (Bergsma et al., 1996). Exon 1 encodes a highly conserved signature sequence common to galactokinases in a wide range of organisms. The function of this shared region of homology remains unknown, but mutations within this sequence have been associated with the deficiency phenotype (Stambolian et al., 1995; Kalaydjieva et al., 1999; Kolosha et al., 2000) suggesting that this area of the gene has a critical role in the normal function of the encoded enzyme. Other highly conserved areas include exons 3 and 7, which both encode ATP binding motifs essential for the phosphorylating activity of galactokinase. Again, mutations within these exons have been identified and linked to the deficiency phenotype (Asada et al., 1999; Kolosha et al., 2000). The function of the remaining coding regions of *GALK1* and how they affect the activity of the galactokinase protein remain to be determined. Mutation analysis of DNA samples from galactokinase-deficient individuals and comparison with the wild-type allele, offers a useful tool for elucidating the role that regions within the gene play in the structure and function of the encoded enzyme.

An opportunity to add to the known mutations of *GALK1* and the possibility of gaining an understanding of the regulatory mechanism of the gene has arisen as a result of collaboration between the Centre for Human Genetics at Edith Cowan University, the Department of Pediatrics at Harvard University Medical School, USA, the University Children's Hospital in Zurich, Switzerland, the Department of Pediatrics at the Medical School in Szeged, Hungary and the Institut Bioquimica Clinica in Barcelona, Spain. The aim is to sequence the *GALK1* gene in patients who exhibit biochemical findings typical of galactokinase deficiency. This analysis may provide new information about the molecular pathology underlying the condition and about the function of the encoded protein domains.

A number of these patients are of Gypsy origins. Analysis of the *GALK1* gene in such individuals will provide information on the homo/heterogeneity of the molecular basis of galactokinase deficiency in this population, known to have a high frequency of the disorder. The identification of a common mutation, P28T (Kalaydjieva et al., 1999) will allow a study of the history of the mutation. It will also open the way to the detection of large numbers of genetically homogenous affected individuals and carriers which, in turn, will allow future studies of genotype-phenotype correlation and of the role of *GALK1* in presenile cataracts.

In addition to these individuals, a new sample of patients has been identified who share elevated plasma galactose levels and galactitol production analogous to the deficiency phenotype, yet whose biochemical assays reveal no discernible loss of function to any of the red blood cell enzymes involved in the Leloir pathway. These clinical observations question existing concepts of galactose metabolism, in particular casting doubt upon the presumption that measurement of erythrocyte enzyme expression reflects adequately the phenotype. The possibility that multiple forms of the galactokinase enzyme exist is supported by age and tissue-related differences in enzyme expression levels (Ng, Donnel, & Bergren, 1965; Shin-Buerhring, Beier, Tan, Osang, & Schaub, 1977), variation in kinetic properties between galactokinase from adult and fetal/newborn sources (Cuatrecasus & Segal, 1965; Mathai & Beutler, 1967; Magnani, Cucchiarini, Stocchi, Dacha, & Fornaini, 1982b) and discrepancies between biochemical phenotype and genotype in some affected individuals (Kalaydjieva et al., 1999).

Whether these observations can be attributed to two forms of galactokinase encoded by separate genes, or reflect post-translational modifications, or alternative transcripts of the same gene, remains to be seen. The study of the *GALK1* gene in patients with metabolic evidence of galactokinase deficiency yet normal RBC enzyme activity, may provide an answer to some of these questions.

## 1.2 Significance of the study

The study of the molecular basis of galactokinase deficiency is important for a number of reasons including;

### *1.2.1 Adding knowledge to the structure and function of GALK1*

An important outcome from this investigation is the possibility of adding to the limited knowledge concerning the molecular organization and function of *GALK1*. While some aspects of the coding region of *GALK1* are known, the role of other regions within the gene and how they determine the structure and hence function of the enzyme remain relatively unknown. The identification of mutations and correlation with the disease phenotype will provide information into the importance of these genomic areas to normal galactokinase expression.

### *1.2.2 Understanding the mechanisms of spread of common founder mutation(s) in GALK1 and identification of high-risk groups*

There is evidence suggesting wide ethnic variation in the incidence of galactokinase deficiency, with a higher frequency among subjects of Gypsy origin. Investigations into the molecular epidemiology of the disorder will provide an insight into the nature and mechanisms of spread of the underlying mutations, and will identify the Gypsy populations where early neonatal detection and dietary treatment will prevent childhood blindness.

### *1.2.3 Possible links to presenile cataract formation*

Further benefits derived from the understanding of galactokinase expression and functional domains would not only be limited to homozygote sufferers. There is also evidence to suggest that heterozygote carriers of a deficiency-causing allele may be at risk of developing presenile cataracts (Stambolian et al., 1986). When we consider that heterozygote frequency for galactokinase deficiency has been

estimated to fall within the ranges of approximately 1 in 100 (Mayes & Guthrie, 1968) to 1 in 300 individuals (Tedesco et al., 1975), the relationship between carriers and presenile cataract formation may point towards partial galactokinase deficiency as a significant contributor to the overall frequency of cataract cases in the wider population. Thus, understanding the genetic regulation of galactokinase expression would have great implications in the development of intervention strategies, in particular the extent that pre-clinical testing and dietary measures would need to be taken in susceptible individuals.

## 1.3 Research aims

The aims of the research consist of the following:

i. To study the molecular basis of classical galactokinase deficiency through;
   a) Identification of novel mutations and polymorphism in the *GALK1* gene.
   b) Analysis of the history and frequency of common mutation(s) in the *GALK1* gene.

ii. To investigate the role of *GALK1* in the development of an atypical phenotype with metabolic symptoms of galactokinase deficiency and normal RBC activity.

## 1.4 Operational terms and definitions

| | |
|---|---|
| A, C, T, G | Adenine, Cytosine, Thymine, Guanine |
| ABI | Applied Biosystems Incorporated |
| ATP | Adenosine triphosphate |
| Bp | Base pairs |
| cDNA | Complimentary deoxyribonucleic acid |
| CpG | Dinucleotide with a cytosine at the 5' end connected by a phosphodiester bond to a guanine at the 3' end. Mutational hotspot due the susceptibility of cytosine to undergo methylation and deamination to thymine. |
| Da | Daltons |
| ddNTPs | Dideoxynucleotide triphosphates |
| DNA | Deoxyribonucleic acid |
| dNTPs | Deoxynucleotide triphosphates |
| *GALK1* | Galactokinase gene |
| kb | Kilobases |
| $K_M$ | Michaelis constant |
| mRNA | Messenger ribonucleic acid |
| $MgCl_2$ | Magnesium chloride |
| mM | Millimolar ($10^{-3}$) |
| nM | Nanomole ($10^{-9}$) |
| PCR | Polymerase chain reaction |
| RBCs | Red Blood Cells |
| RNA | Ribonucleic acid |
| 5' UTR | 5' Untranslated region |
| 3' UTR | 3' Untranslated region |

# CHAPTER 2 LITERATURE REVIEW

## 2.1 Galactokinase

### *2.1.1 Galactokinase and the Leloir pathway*

The function of the Leloir pathway of metabolism is to convert galactose into glucose-6-phosphate in preparation for utilization in glycolysis, glycogenesis or incorporation into normal cellular complex carbohydrates (Frey, 1996). In mammals, the main dietary source of galactose is the hydrolytic breakdown of milk-sugar into the monosaccharides glucose and galactose. Galactose then enters the Leloir pathway where it undergoes phosphorylation by galactokinase to form galactose-1-phosphate (Figure 2.1).

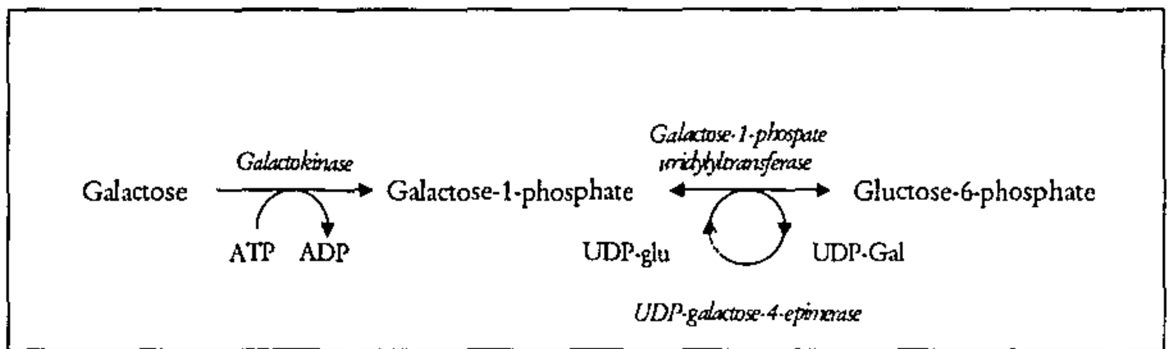



**Figure 2.1** Role of galactokinase in the Leloir pathway

The function and expression of this enzyme has been characterized in a number of organisms, including bacteria (Wilson & Hogness, 1969), mice (Ai et al., 1995), rat (Cuatrecasus & Segal, 1965), porcine liver (Ballard, 1966), as well as various human tissues including erythrocytes, leukocytes, placenta (Srivastava, Blume, Van Loon, & Beutler, 1972) and fibroblasts (Friedman, Yarkin, & Merril, 1974). The most significant galactose conversion occurs within the liver, where expression levels of galactokinase are approximately 50 times higher than those in other tissue sources (Shin-Buerhring, Beier, Tan, Osang, & Schaub, 1977). Uptake of galactose from the bloodstream by the liver is a highly efficient process (Goresky, Bach, & Nadeau, 1973) mediated by two factors, firstly,

the movement of galactose across the hepatocyte membrane by the presence of GLUT2 transporters (Burchell, 1994) and secondly, its conversion to galactose-1-phosphate by galactokinase. Such is the efficiency of this process, that galactose has been used to measure hepatic function and blood flow in physiological studies (Henderson & Hanna, 1983). The other significant site of net removal and utilization of galactose is in the erythrocyte. Here, galactose is used as an energy source via its conversion to glucose-6-phosphate and entry into the glycolytic pathway. Although galactokinase from both erythrocytes and liver catalyzes the same reaction, the utilization of the end metabolic product differs. It is not surprising therefore that a number of differences exist between the levels of activity and kinetic properties of the liver and RBC enzyme, some of which have been interpreted as evidence of the existence of multiple forms of the enzyme (Cuatrecasus & Segal, 1965; Ng et al., 1965; Mathai & Beutler 1967).

### *2.1.2 Age and tissue-related differences in galactokinase expression*

Measurement of galactokinase in individuals of various ages suggests that enzyme activity is considerably higher during infancy than later in life (Ng et al., 1965). In fact, up to a period of three years after birth, galactokinase activity is nearly 3 times higher than that of older children and adults. This may in part reflect changes from the early nutritional intake of galactose derived from breast milk to a more varied diet in adulthood. The high expression of galactokinase in infancy may reflect the important role galactose may have as an energy source in early development and growth.

Galactokinase is found to differ between fetal liver, kidney, lung, spleen, muscle and brain in regards to substrate affinity as well as duration and degree of expression (Shin-Buerhring et al., 1977). Enzyme activity in liver increases in a linear progression with gestational age, reaching its maximal peak by 28 weeks. Expression within the brain however, remains constant for the first 5 months of prenatal development before increasing

to preterm levels. These observations suggest that independent or tissue-specific regulation of galactokinase expression may be occurring between these tissue sources based upon the substrate needs of each tissue type.

Other tissue differences are noted when galactokinase isolated from placenta is compared to both the liver and erythrocyte enzyme (Table 2.1). These differences refer to kinetic properties of the enzyme, as well as to the physical characteristics of the protein.

**Table 2.1**

Galactokinase Properties in Various Tissue

| | Placenta[a] | Erythrocyte[b] | Liver[c] |
|---|---|---|---|
| Molecular weight | 58 000 | 53 000 - 57 000 | Not determined |
| Km galactose | 250μM | 100-150 μM | Fetal/newborn>adult |
| Km ATP | 100μM | 200-500μM | Fetal/newborn>adult |
| Inhibition by G-1-P | High | nil | Fetal/newborn>adult |

[a] Srivastava et al.. (1972)
[b] Blume and Beutler (1971)
[c] Cuatrecasas and Segal (1965)
Note. Liver properties measured in rat

Other differences are noted in the functional role of galactokinase. It has been suggested that galactokinase in liver can act in a rate-limiting manner that prevents the toxic accumulation of galactose-1-phosphate, implicated in the severe pathological symptoms observed in other defects of galactose metabolism (Segal & Berry, 1995). Evidence supporting this function comes from sustained galactose loading in rat liver (Rogers, Bovee, Saunders, & Segal, 1989) and human liver-derived cell lines (Davit-Spraul, Pourci, Soni, & Lemonnier, 1994), where galactokinase activity declines in response to chronic overloading as a result of the inhibition by galactose-1-phosphate. This rate-limiting role is not evident in human erythrocytes (Mathai & Beutler, 1967) suggesting that this function is limited to the liver form of the enzyme. The rate-limiting role of liver galactokinase is age-dependent (Cuatrecasus & Segal, 1965), with fetal galactokinase inhibited by galactose-1-phosphate more than the adult enzyme. Other differences noted in

this investigation include greater stability and a galactose $K_M$ approximately 4 times higher in fetal galactokinase compared to the adult enzyme.

Further evidence of age-related differences in kinetic properties comes from comparative studies between adult and fetal erythrocytes. Mathai and Beutler (1967) found slight but statistically significant evidence for a fetal RBC galactokinase with higher $K_M$ for galactose than that derived from adult sources. This observation remains when the age-related differences in erythrocyte composition are accounted for (Magnani et al., 1982b), such as the separation of both fetal and adult red blood cells into fractions of different mean ages by density gradient ultracentrifugation prior to measuring galactokinase activity. This technique offers a more accurate indication of enzyme activity by accounting for the inverse relationship between erythrocyte life-span and enzyme levels. Further findings by Magnani et al. (1982b) include a fetal erythrocyte galactokinase with specific activity three times higher than the adult enzyme (Table 2.2) and displaying considerably greater thermostability than the galactokinase derived from adult erythrocytes.

**Table 2.2**

Properties of Galactokinase in Adult and Fetal Erythrocytes

| | Erythrocyte galactokinase[a] | |
|---|---|---|
| | **Fetal** | **Adult** |
| Molecular weight | 48 000 | 48 000 |
| Specific activity | 39.5 mU/g Hb | 12.7 mU/g Hb |
| $K_M$ galactose | 215 μM | 113 μM |
| $K_M$ ATP | 400 μM | 320 μM |

[a] Magnani et al. (1982b)

### *2.1.3 Differences in the physical properties of the enzyme*

Differences in the physical properties of galactokinase are also known to exist. Srivastava et al. (1972) describe a placental galactokinase with a larger molecular weight than that derived from adult erythrocytes. Tedesco et al. (1973) provide a description of erythrocyte galactokinase isolated from one individual that has higher electrophoretic

mobility and higher enzyme activity than that found in erythrocytes from adult controls. The kinetic properties are similar to placental galactokinase, suggesting a persistent expression of a fetal form of the enzyme.

Conclusive information about the differences in molecular structure that would account for these observations is lacking. An early investigation into the physical makeup of galactokinase derived from erythrocytes suggested a dimeric structure consisting of two monomers of molecular weights 25 000Da to 27 000Da (Blume & Beutler, 1971). A succeeding study, using a much more effective technique of protein purification, found no evidence for a dimeric structure in either erythrocytes or placental galactokinase (Stambolian, Scarpino-Myers, & Harris, 1985). In fact, the latter study reports the monomeric galactokinase from both sources as having a molecular weight of approximately 38 000Da, significantly lower than previously thought (Srivastava et al., 1972). This, again, contrasts with an analysis of cDNA derived from osteoclastoma and T-cell cDNA libraries (Stambolian et al., 1995): based upon the nucleotide sequence, the molecular weight of the encoded galactokinase from these sources is predicted to be 42 271 Da.

*2.1.4 Allelic variation and genotype/phenotype differences*

Allelic variations from the recognized wild-type $GK_A/GK_A$ genotype have been identified and point towards polymorphism between populations. The Philadelphia variant, $GK_P$ (Tedesco et al., 1977) is found in high frequency amongst the African/American population (Soni, Brivet, Blanc, Jaeger, & Lemonnier, 1988) and causes reduced galactokinase activity in RBCs, yet normal levels of enzyme expression and function in leukocytes. Another suggested allelic variation is the Urbino variant (Magnani, Cucchiarini, Dacha, & Fornaini, 1982a). This galactokinase has reduced catalytic properties, decreased *in-vivo* stability and a lower $K_M$ for $ATPMg^{2-}$ than the enzyme encoded by both the wild-type, $GK_A$ and some deficiency-causing; $GK_G$ alleles. It is possible that the degree and range of allelic variants may in fact be more widespread than those already characterized.

Other genotype/phenotype discrepancies have been detected that may support heterogeneity in galactokinase expression. An interesting observation in a kindred studied by Kalaydjieva et al. (1999) was the biochemical findings of an individual homozygous for the P28T mutation who displayed intermediate erythrocyte galactokinase activity analogous to a heterozygote carrier. This particular individual reported a history of juvenile cataracts suggesting a lack of enzyme activity at an early age. Why this person is now expressing galactokinase despite being homozygous for the deficiency causing mutation is unclear. It has been suggested that this intermediate activity may represent expression of an adult form of the enzyme. Whether this expression is due to the existence of another gene or age-related differences in the expression of the *GALK1* encoded protein is unknown.

It remains unclear whether the age and tissue-related differences in levels of activity, kinetic properties and physical characteristics are the product of differences in the regulation of expression and/or post-translational modification of a single gene product or whether they result from the existence of more than one galactokinase gene.

## 2.2 Galactokinase Deficiency

### *2.2.1 Clinical synopsis and symptoms*

In 1933 Fanconi, as cited in Gitzelmann (1965), first described "galactose diabetes", a term used to denote an aberration of galactose metabolism, whose only symptom was the development of early-onset cataracts. The exact enzyme defect was not confirmed until Gitzelmann (1965) measured erythrocyte galactokinase in a patient first diagnosed by Fanconi, and found negligible enzyme activity. It is now recognized that deficiency of galactokinase can cause the formation of juvenile cataracts in homozygotes.

Cataracts remain the only confirmed symptom of the disorder, despite isolated reports of neurological impairment (Pickering & Howell, 1972) hepatosplenomegaly (Thalhammer, Gitzelmann, & Pantlitschko, 1968), neurofibromatosis (Gitzelmann, 1965) and pseudotumor cerebri (Litman, Kanter & Finberg, 1975).

Biochemical assays in galactokinase-deficient individuals reveal elevated plasma levels of galactose, called hypergalactosemia, the presence of unmetabolized galactose in the urine, known as galactosuria, and the presence of a sugar polyl in both urine and tissue called galactitol (Gitzelmann & Hansen, 1980). It is galactitol, the product of an alternative pathway of galactose metabolism, that is implicated in causing the degenerative changes to the lens of the eye. The pathological progression to cataract formation in humans occurs due to the influence of a number of interacting processes initiated by high intake of dietary galactose. When galactose levels are elevated, aldose reductase within the lens of the eye catalyzes the conversion of galactose to galactitol. Galactitol cannot be further metabolized by the body and as a result accumulates within the lens. This causes osmotic changes in the optic fluid, causing the disruption of cell membranes and ultimately the opacification of lens fibres. This mechanism is evident in galactokinase-deficient strains of mice transfected with the human aldose reductase (*hAR*) gene (Ai et al., 2000). These transgenic mice, when fed a high galactose diet rapidly develop cataracts. This is in contrast to wild-type galactokinase-deficient mice where aldose reductase activity within the lens is negligible. These observations suggest that it is the high intake of galactose, in conjunction with a deficiency in galactokinase, on a background of aldose reductase expression in the lens, that results in the development of cataracts.

### *2.2.2 Epidemiology*

The range and incidence of galactokinase deficiency (Table 2.3) varies widely between populations. Across USA, Japan and Western Europe newborn screening has only detected 6 cases amongst 6 million births (Levy, 1980). By contrast, data from Austria, Germany and Bulgaria suggest a higher incidence ranging from approximately 1: 52 000 to 1: 157 000 births (Gitzelmann & Hansen, 1980; Kalaydjieva et al., 1999). The highest frequency has been found among the Vlax Roma, with one affected homozygote per 1 in 1600-2500

births. The high incidence has been attributed to a founder mutation segregating in this isolate (Kalaydjieva et al., 1999).

**Table 2.3**
Variation in Incidence and Allele Frequency

| | Incidence/Frequency | Region/Population | Source |
|---|---|---|---|
| Homozygotes | | | |
| | 1 : 1 000 000 | USA, Asia, Europe | Levy (1980) |
| | 1 : 157 000 | Germany | Gitzelmann & Hansen (1980) |
| | 1 : 153 000 | Austria | Gitzelmann & Hansen (1980) |
| | 1 : 52 000 | Bulgaria | Kalaydjieva et al.. (1999) |
| | 1 : 1600 - 1: 2500 | Vlax Roma | Kalaydjieva et al.. (1999) |
| Heterozygotes | | | |
| | 1 : 50 | Caucasian | Gitzelmann (1967) |
| | 1 : 107 | Caucasian | Mayes & Guthrie (1968) |
| | 1 : 304 | Caucasian/Nth American | Tedesco et al.. (1975) |

Early studies that measure erythrocyte galactokinase activity in adults suggest that the heterozygote carrier frequency of a deficiency allele falls in the range between 1 in 50 (Gitzelmann, 1967) to 1 in approximately 100 individuals (Mayes & Guthrie, 1968). The predicted frequency of affected births from these estimates is considerably higher than that actually detected by newborn screening programs.

The above studies are based on the detection of heterozygous carriers by measurement of RBC enzyme activity, with some inherent problems. On one hand, enzyme activity fluctuates widely and is an unreliable indicator of carrier status. Furthermore, the results can be affected by variations in methodology, increasing further the error range. In addition, it is unclear to what extent RBC activity reflects the organism's ability to metabolize galactose.

It is possible that other factors may be involved in this discrepancy, including miscalculation of heterozygote frequency due to the presence of allelic variants of galactokinase. The aforementioned Philadelphia variant (Section 2.1.4) is one such allele in the African/American population that causes a reduction in erythrocyte activity yet normal expression in other tissue sources (Tedesco et al., 1973; Tedesco et al., 1975). No detrimental effects are associated with this genotype. It is possible that early studies that relied upon erythrocyte assay to determine galactokinase activity may have inadvertently measured similar allelic variants unknown at the time of the study, therefore overestimating heterozygote frequencies for the deficiency allele.

Other considerations that may account for these discrepancies include the possibility of the existence of two forms of the galactokinase enzyme, one expressed during early development and childhood, the other during adulthood. Therefore if heterozygote frequencies have been calculated from adult samples based upon erythrocyte expression alone, the enzyme measured may not be the galactokinase involved in the infantile deficiency. Given the differences in galactokinase properties and expression patterns as a function of age (Section 2.1.2), this second explanation appears plausible.

### *2.2.3 Possible role of galactokinase deficiency in presenile cataract formation*

The pathological progression to cataract formation may not be restricted to homozygote sufferers. It has been suggested that heterozygote carriers of a deficiency allele may be at risk of developing cataracts during the 2$^{nd}$ to 5$^{th}$ decade of life (Stambolian et al., 1986). Other studies find no such relationship (Magnani, Cucchiarini, Stocchi, & Dacha, 1983). The reasons for these discrepant findings are unclear. One possible explanation is differences in galactose content in the diet, where a high intake can contribute to cataract formation. This may be due to the fact that erythrocyte galactokinase assay of individuals

carrying only one functioning allele reveal galactokinase activity at approximately 50% of normal (Sitzmann & Kaloud, 1976). Thus, sustained levels of galactose in the diet may overload the net conversion of galactose to galactose-1-phosphate, in turn promoting the formation of galactitol and with it the disruptions to lens integrity.

The effects of increased dietary intake of galactose coupled with reduced galactokinase activity and its role in cataract formation is also supported in an investigation of maternal enzyme levels during pregnancy (Harley, Mutton, Irvine, & Gupta, 1974). This study found that galactokinase was significantly reduced in mothers whose children have congenital cataracts, despite these children having normal erythrocyte enzyme expression levels and activity. This suggests that marginal maternal deficiency of galactokinase coupled with an increase in lactose intake during pregnancy may exceed the maternal ability to effectively metabolize galactose.

The proposed link between galactokinase deficiency and presenile cataract formation has never received solid support. Mutation detection would offer a much more direct approach to the identification of carriers, however this has been hampered by the low incidence of the disorder and the private nature of most of the *GALK1* mutations identified thus far (Section 2.4). It is the through the identification of populations at high risk, where a single common founder mutation occurs at a high frequency that will allow a large number of carriers to be identified. The relationship between partial galactokinase deficiency and presenile cataract could thus be investigated.

## 2.3 The *GALK1* gene: Localization and Structure

### *2.3.1 Localization*

The localization of the gene encoding galactokinase on the long arm of chromosome 17 was first determined using somatic cell hybridization by Elsevier et al. (1974). In 1995, physical mapping using fluorescent *in-situ* hybridization (FISH) placed the *GALK1* gene on

band 17q24 (Stambolian et al., 1995). Further refinement of this position to within a 1 cM distance was achieved using linkage analysis and radiation hybrid mapping by Kalaydjieva et al. (1999). The earlier identification in 1992, of a cDNA from a gene called *GK2* which shared similar characteristics to galactokinase, questioned the exact chromosomal assignment of the galactokinase gene (Lee, Peterson, Calman, Herskowitz, & O'Donnell, 1992). However a succeeding study (Pastuszak, O'Donnell & Elbein, 1996) found that although the protein encoded by *GK2* could phosphorylate galactose, this could only occur when the substrate was present in millimolar concentrations. Instead, *GK2*'s major role is to phosphorylate N-acetylgalactosamine, a component of complex carbohydrates found in eukaryotic cells.

### *2.3.2 Characterization*

The isolation of *GALK1* cDNA by Stambolian et al. (1995) was initially based on comparisons of the sequences of peptide fragments derived from human placental galactokinase, to a number of cDNA libraries and expressed sequence tags (ESTs). Significant similarities were found in two human cDNA clones. The analysis of the nucleotide sequence of these cDNAs predicted a mRNA product of approximately 1350 nucleotides. Further studies revealed the full-length genomic organization of *GALK1* (Bergsma et al., 1996). The gene spans approximately 7.3kb of DNA and is made up of 8 exons coding for a mRNA transcript of 1358 nucleotides, in accordance with the transcript size isolated earlier by Stambolian et al. (1995).

Despite these analyses, understanding of the functional organization of the *GALK1* gene is limited to only a small number of domains (Figure 2.2). Exon 1 contains a highly conserved signature sequence common to a wide range of organisms including bacteria, yeast (Stambolian et al., 1995) mice (Ai et al., 1995) and humans. The function of this shared region of homology remains unknown but a number of mutations within this domain have been associated with the deficiency phenotype (Stambolian et al., 1995;

Kalaydjieva et al., 1999; Kolosha et al., 2000), suggesting that this area of the gene has a critical role in the normal functioning of the encoded enzyme. Other conserved areas include exons 3 and 7, both encoding ATP-binding motifs essential for the phosphorylating activity of galactokinase.

The other remaining area of the human *GALK1* gene characterized, includes a 72bp sequence of the last exon which has been found to encode the carboxyl terminus of the galactokinase protein (Bergsma, et al., 1996).

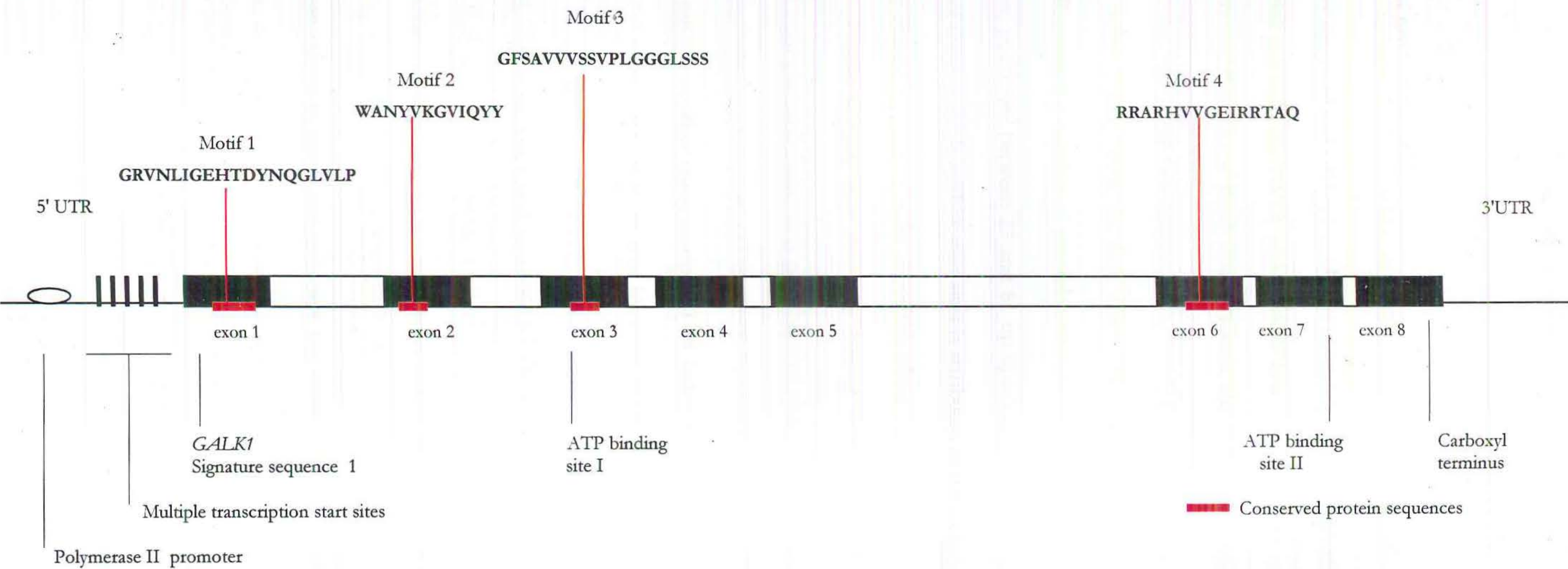



**Figure 2.2** Functionally important domains and conserved protein fingerprint motifs.

### *2.3.3 5'UTR and promoter region*

The promoter region of *GALK1* is found approximately 370bp upstream of exon 1 (Bergsma et al., 1996) and lacks the TATA and CCAAT box motifs thought necessary for accurate mRNA initiation (Strachan & Read, 1998). Instead, the 5'UTR is characterized by a polymerase II promoter with a high GC content, indicative of potential binding sites for the Sp1 transcription factor (Tjian, 1995). In fact, three sites for Sp1 are recognized. In addition, the 5'UTR contains a number of possible transcription start sites downstream of the promoter region, positioned between 21 and 61 bp upstream of the first exon (Figure 2.3). The presence of these multiple mRNA start sites is attributed to the lack of TATA and CCAAT motifs. An interesting point to make about this observation is that Bergsma et al.'s (1996) description of these multiple start sites was derived from the 5' ends of mRNA transcripts isolated from placental tissue. It is possible that these start sites are unique for the galactokinase produced in this tissue and may not be indicative of the transcriptional origin of *GALK1* mRNA in other tissue sources. This factor may partly account for the different physical and kinetic properties evident between placental and liver forms of the enzyme.

Discrepancies exist in the nucleotide sequence of the non-coding regions of *GALK1* reported by Asada et al. (1999) and by Bergsma et al. (1996). Three nucleotide differences have been found in the 5'UTR of *GALK1* among a small sample of Japanese and Caucasian subjects representing both wild type and deficiency phenotypes. This area, although not transcribed, has been shown to affect and influence the transcriptional activities of genes (Cooper, Krawczak, & Antonarakis, 1995). These differences may reflect polymorphisms, however given the limited number of samples analyzed, this possibility is unclear. Other discrepancies noted by Asada et al. (1999) include several nucleotide re-arrangements in

introns 1, 2 and 5. A 20bp tandem repeat was also found in intron 5 that had not been previously detected by Bergsma et al. (1996). As with the differences noted in the 5'UTR, it could not be confirmed whether these are the result of polymorphisms unique to the population studied or in fact reflect errors in the published sequence.

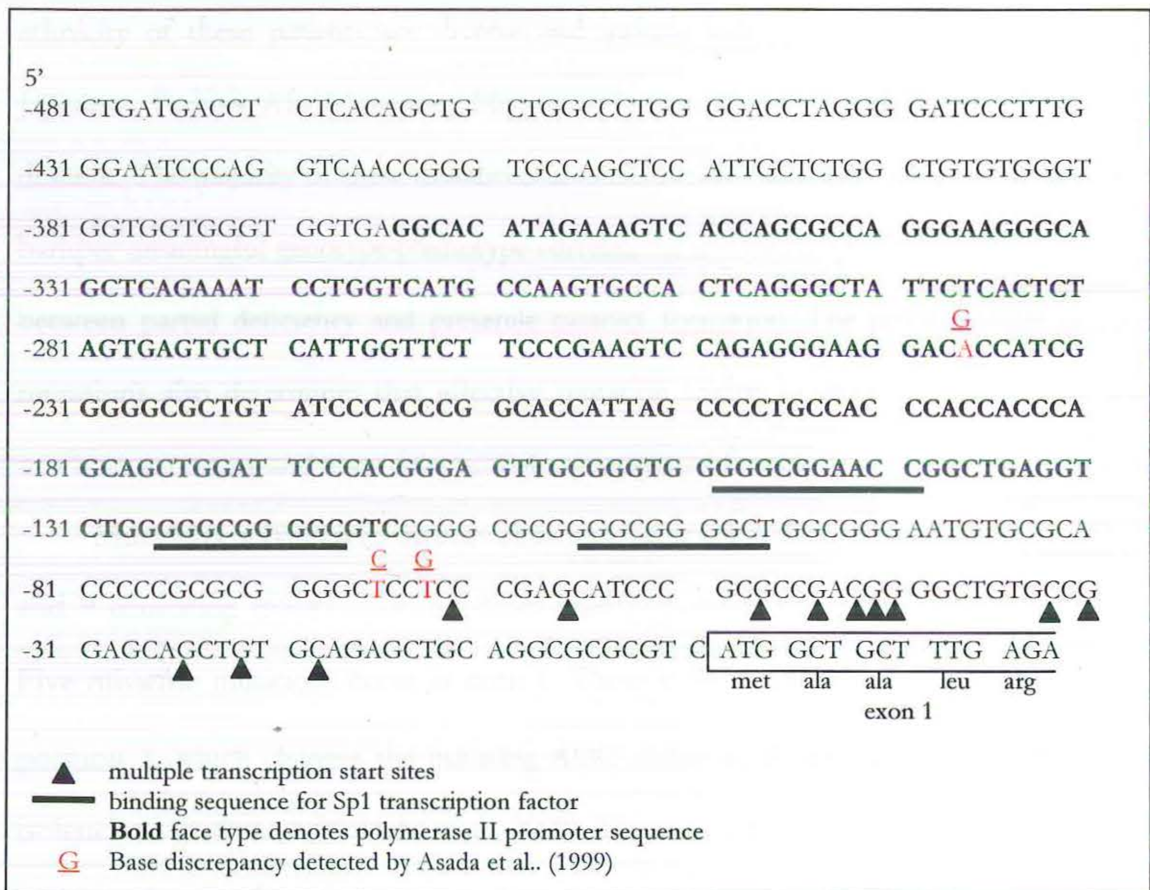



**Figure 2.3** *GALK1* 5'UTR[a]

[a] Adapted from Bergsma et al.. (1996)

## 2.4 Mutations in *GALK1* associated with enzyme deficiency

Due to the low incidence of galactokinase deficiency, only a small number of patients have been available for mutation analysis, therefore the number of mutations identified and characterized in the *GALK1* gene remains small (Figure 2.4). To-date, only 20 mutations have been reported among 32 individuals with galactokinase deficiency (Stambolian et al., 1995; Kalaydjieva et al., 1999; Asada et al., 1999 & Kolosha et al., 2000). The origins and ethnicity of these patients are diverse and include individuals of European/American, Japanese, Turkish, Afro/American, Hispanic, German, Swiss, Costa Rican and Vlax Gypsy descent. The majority of these mutations are confined to individual families and as a result hamper meaningful genotype-phenotype correlations and also impede research into the link between partial deficiency and presenile cataract formation. The private nature of these mutations also determines that effective mutation testing in most populations for both newborn screening and mass detection of carriers is not feasible.

Mutations published to-date are shown in Table 2.4 and include 9 missense mutations and 9 truncating mutations arising from insertions, deletions and nonsense stop signals. Five missense mutations occur in exon 1. These include a G→T transition at nucleotide position 3, which changes the initiating AUG codon to AUU, causing a methionine to isoleucine substitution (Kolosha et al., 2000). This causes translation to begin at the second AUG codon resulting in a galactokinase protein lacking the first 54 amino acids. The second missense mutation in exon 1 is a 82C→A transversion causing a proline to threonine amino-acid change at codon 28. This substitution was identified as the founder mutation causing galactokinase deficiency within the Vlax Roma population (Kalaydjieva et al., 1999). This missense substitution occurs in close proximity to a 94G→A transition that results in a V32M change, reported in an unrelated sample (Stambolian et al., 1995).

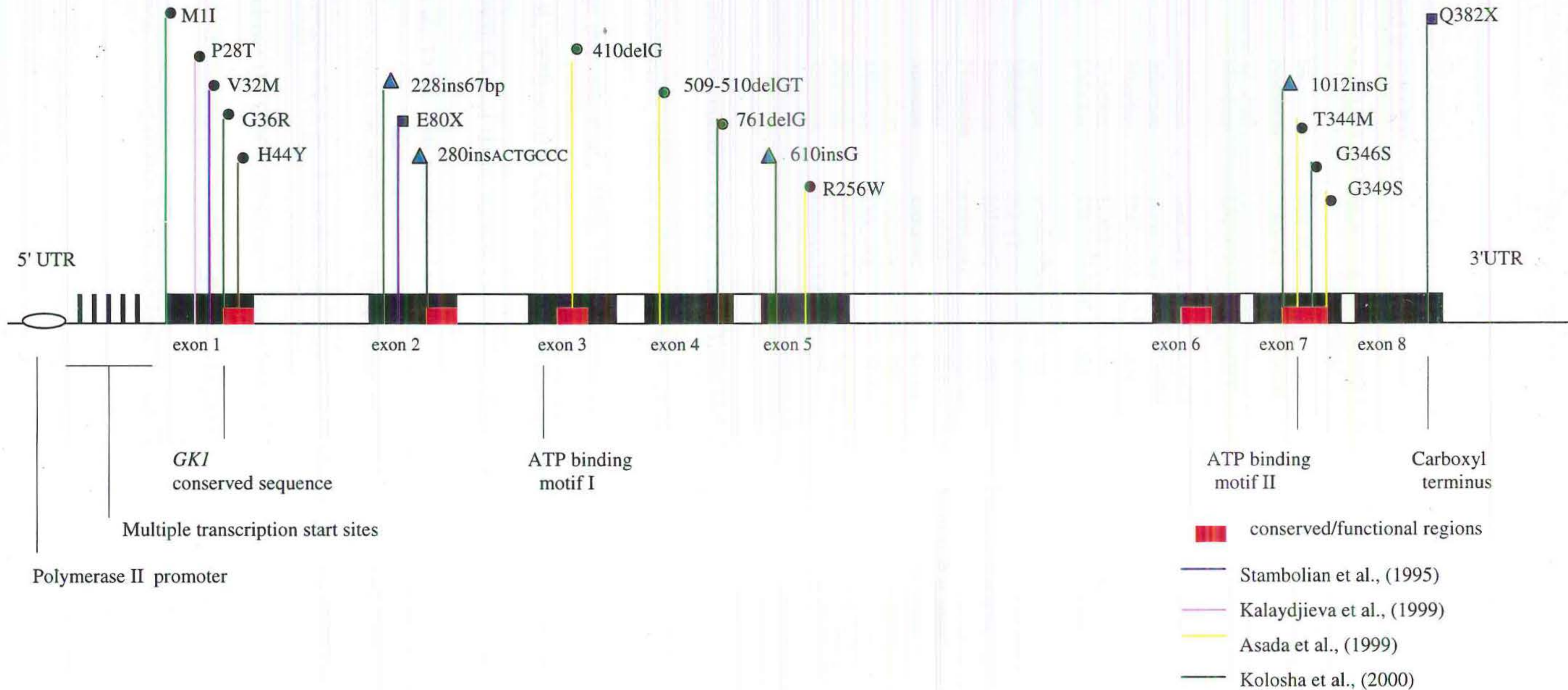



**Figure 2.4** Distribution of mutations in *GALK1*.

**Table 2.4**

Mutations Identified in *GALK1*

| | Mutation type | Nucleotide | Amino acid | Result |
|---|---|---|---|---|
| Stambolian et al. (1995) | Missense | 94A→G | V32M | Disrupts loop-beta turn motif |
| | Nonsense | 238G→T | Q80X | Stop codon→ truncated protein |
| Kalaydjieva et al. (1999) | Missense | 82C→A | P28T | Disrupts loop-beta turn motif |
| Asada et al. (1999) | Missense | 766C→T | R256W | Not characterized |
| | Missense | 1031C→G | T344M | Occurs at ATP binding site II |
| | Missense | 1045G→A | G349S | Occurs at ATP binding site II |
| | Deletion | 410delG | 163 | In frame nonsense codon |
| | Deletion | 509-510delGT | 201 | In frame nonsense codon |
| Kolosha et al. (2000) | Insertion | 228ins 67bp | 77 | Frameshift at amino acid 77 |
| | Missense | 3G→T | M1I | Protein lacks first 54 amino acids |
| | Insertion | 1012insG | 338 | Frameshift at amino acid 338 |
| | Nonsense | 1144C→T | Q382X | Stop codon→ truncated protein |
| | Deletion | 761delG | 254 | Frameshift at amino acid 254 |
| | Missense | 106G→C | G36R | Occurs near conserved region |
| | Insertion | 610insG | 204 | Stop codon→ truncated protein |
| | Missense | 130C→T | H44Y | Occurs near conserved region |
| | Missense | 1036G→A | G346S | Occurs at ATP binding site II |
| | Insertion | 280insACTGCCC | 94 | Frameshift at amino acid 94 |

Two other missense mutations occur in exon 1 and include a 106G→C transversion causing a G36R substitution, while the other is a 130C→T transition resulting in a H44Y amino acid change (Kolosha et al., 2000). Three missense mutations, R256W, T344M and G348S (Asada et al., 1999) occur at CpG nucleotides, which are known mutational hotspots (Cooper et al., 1995). One of these, R256W caused by a 766C→T transition, is found in exon 5 (Asada et al., 1999). This mutation does not occur within a recognized conserved motif and its effects upon the structure of the protein were not investigated. Both the 1031C→G transversion which produces the T344M amino acid change and the 1045G→A transition which causes the G349S substitution, occur in the ATP-binding site sequence in exon 7. An additional missense mutation has also been found to occur in this region and is a G346S substitution resulting from a 1036G→A base change (Kolosha et al., 2000).

A number of truncating mutations have also been reported in the *GALK1* gene. Two of these, E80X and Q382X, are the result of nucleotide changes. E80X results from a 238G→T transversion in exon 2, which creates a termination codon producing a truncated galactokinase protein of only 79 amino acids (Stambolian et al., 1995). The Q382X nonsense mutation results from a 1144C→T transition in exon 8, which is known to encode the carboxyl terminus of the enzyme. As a result of this premature termination codon, the enzyme was predicted to lack a number of C-terminal amino acids. Subsequent expression analysis in oocytes suggested that low measured galactokinase activity was attributed to the 1144C→T substitution. This finding suggests that the carboxyterminus of the protein is essential for normal galactokinase function.

The remaining truncating mutations result from 3 deletions, 410delC, 509-510delGT, (Asada et al., 1999) and 761delG (Kolosha et al., 2000) and 3 insertions, 610insG, 280insACTGCCC and 1012insG (Kolosha et al., 2000). All result in frameshifts that produce a truncated protein. The duplication of a 67bp sequence of the 5' end of exon 2 and 5bps of the adjoining intron (288ins67bp) found in one individual was found to cause an additional 5' splice acceptor site (Kolosha et al., 2000). This insertion results in the removal of the normal 5' end of exon 2 during splicing, producing a partial and thus nonfunctional galactokinase transcript.

Mutations characterized within the *GALK1* gene have pointed towards the functional significance that particular regions may have upon the encoded protein. The clustering of nucleotide aberrations within highly conserved areas of the *GALK1* gene or recognized functional motifs, highlight the importance that each region has in the normal activity of the encoded galactokinase. Five missense mutations, M1I, P28T, V32M, G36R and H44Y occur in exon 1 and of these, both G36R and H44Y reside within the highly conserved galactokinase signature domain. Their effects upon the secondary structure of the protein

are not reported, however it is likely that they interfere with an unknown function that seems critical for normal galactokinase activity. The two missense mutations, P28T and V32M, occur in close proximity to this signature sequence. In fact, computer models of the protein configuration caused by both P28T and V32M, reveal similar secondary structural changes that interfere with a loop-beta turn motif immediately preceding the conserved domain (Kalaydjieva et al., 1999). The fact that the majority of the missense mutations found in the *GALK1* gene cluster within exon 1, which encodes the highly conserved fingerprint motif, suggest this region plays a crucial role in the normal activity of the enzyme.

Another clustering of three missense mutations and one insertion occurs in the second ATP-binding motif in exon 7. Three of these, T344M, G346 and G349S, occur within a 5 codon span, suggesting a small region exists within this domain that is critical for normal function. The fact that these are missense mutations, may reflect a critical conformational dependence in order for normal ATP binding to occur in the encoded protein.

To-date, only three mutations (P28T, Q382X and G349S) have been reported to occur in multiple individuals with galactokinase deficiency. The G349S substitution has been found in the heterozygous state in two galactokinase-deficient individuals. These patients differ in ethnic origin: one is German (Kolosha et al., 2000) and the other Japanese (Asada et al., 1999), suggesting this mutation is a reccurence of the amino acid substitution at the same codon. The remaining two mutations, P28T and Q382X appear to be founder mutations. Although the common origin of P28T has been confirmed by haplotype analysis in six Gypsy families from Bulgaria (Kalaydjieva et al., 1999), it is unclear whether the same mutation accounts for all affected individuals among the European Gypsies. No haplotype analysis has been conducted among the Costa Rican/Europeans with Q382X, therefore a founder effect can only be assumed.

## 2.6 Summary

Clearly, the lack of conclusive data about both the functional and structural characteristics of the *GALK1* gene, as well as the influence of age, tissue specificity and allelic variation upon the encoded enzyme product, allows only a limited insight to the complete process of galactose metabolism. One way towards understanding this process is further analysis of the *GALK1* gene. Any mutations found during such investigations will help the progression towards a complete characterization of the galactokinase gene and galactose metabolism. Furthermore, the lack of conclusive information regarding the frequency and distribution of common mutations in the *GALK1* gene needs to be addressed to allow further research into genotype/phenotype correlations. Not only will this aid in the development of newborn screening strategies in high risk groups and prevention of infantile blindness, but will also allow the link between partial deficiency and presenile cataract development to be fully investigated.

# CHAPTER 3 IDENTIFICATION OF *GALK1* MUTATIONS IN PATIENTS WITH GALACTOKINASE DEFICIENCY

## 3.1 SUBJECTS

Patients included 9 individuals with galactokinase deficiency. Most had been detected as hypergalactosemic by newborn screening (Table 3.1). The exception was a Gypsy child from Bulgaria, who had undergone surgery for infantile cataracts. The patients are on galactose-restricted diet and are symptom-free. Galactokinase deficiency (< 5% of the lower limit of the normal range) was confirmed by enzyme activity measurements in red blood cells (RBC). All patients have normal galactose-1-phosphate uridyl transferase and UDP-epimerase RBC activity.

**Table 3.1**

Biochemical Findings of Study Sample

| | | | | RBC Activity | | |
|---|---|---|---|---|---|---|
| Patient | Hypergalactosemia | Galactosuria | Galactokinase | G-1-P uridyl transferase | UDP-epimerase | Origin |
| 1[a] | + | + | Deficient | Normal | Normal | European/America |
| 2[b] | + | + | Deficient | Normal | Normal | Turkish |
| 3[a] | + | + | Deficient | Normal | Normal | European/America |
| 4[c] | + | + | Deficient | Normal | Normal | Spanish Gypsy |
| 5[c] | + | + | Deficient | Normal | Normal | Spanish Gypsy |
| 6[d] | + | + | Deficient | Normal | Normal | Hungarian Gypsy |
| 7[d] | + | + | Deficient | Normal | Normal | Hungarian Gypsy |
| 8[d] | + | + | Deficient | Normal | Normal | Hungarian Gypsy |
| 9[e] | + | + | Deficient | Normal | Normal | Bulgarian Gypsy |

[a] Dept of Pediatrics, Harvard University Medical School, Boston, USA
[b] University Children's Hospital, Zurich, Switzerland
[c] Institut Bioquimica Clinica, Barcelona, Spain
[d] Dept. of Pediatrics, Medical School, Szeged, Hungary
[e] Bulgaria

The DNA samples of these patients have been provided for analysis by the Genetic Service, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, USA, the University Children's Hospital, Zurich, Switzerland, the Institut Bioquimica Clinica, Barcelona, Spain and the Department of Pediatrics, Medical School,

Szeged, Hungary. DNA from three individuals without galactokinase deficiency were used as controls and were sequenced in parallel with patient DNA. These samples were chosen from a DNA bank maintained within the Centre for Human Genetics at Edith Cowan University.

Consent for DNA analysis was obtained from the parents of all participants in the study. The investigation complies with the guidelines of the Edith Cowan University Ethics Committee.

## 3.2 METHODS

### *3.2.1 Experimental Design*

The strategy used in the current investigation included:

a) Detection of mutations:
- by sequence analysis of the entire *GALK1* coding region, flanking intronic sequences and both the 5' and 3' UTR.
- by direct testing of the P28T mutation in individuals of Gypsy ethnicity.

b) Independent confirmation of identified variants using other methods of mutation detection:
- confirmatory analysis of the original DNA sample using an independent PCR amplification and a different method.
- screening a panel of 50 Caucasian controls (100 chromosomes).

c) Computer analysis:
- comparisons of sequences.
- analysis of the nature of the sequence variants.
- predicted effects on protein structure/function.

### *3.2.2 DNA Quantification*

The concentration of the provided DNA samples was determined using a Beckman DU® 640 UV spectrophotometer. This is achieved by reading the absorbance of a DNA sample in $dH_2O$ at a wavelength of 260nm. Concentration of each sample can be calculated using the fact that one optical density unit at this wavelength equates to 50μg/ml of DNA (Sambrook, Fritsch, & Maniatis, 1989). The ratio between the readings at 260nm (DNA) and 280nm (phenolic rings) provides an estimate of the purity of the nucleic acid. Knowing the starting concentration of each of the DNA samples allows the accurate dilution to a suitable working solution. For the purposes of this study, a concentration of 20ng/μl of DNA was employed.

### *3.2.3 PCR Amplification*

Amplification of *GALK1* was performed using the polymerase chain reaction (PCR). This procedure has revolutionized molecular investigations of the genome, allowing great specificity and ease in obtaining a large number of copies of template DNA from even the smallest concentration of starting material. A standard PCR reaction involves the addition of a heat-stable DNA polymerase, dNTPs, sequence specific primers, buffer, $dH_2O$ and $MgCl_2$, to a sample of DNA. Successive temperature cycling causes denaturing of DNA to single-stranded templates, then a period of cooling to allow the annealing of oligonucleotide primers, followed by another temperature change that allows activation of the polymerase. The activated polymerase catalyses synthesis of new DNA strands, utilizing the provided dNTPs in the reaction. With continued cycling, an exponential increase in DNA templates is achieved. A number of parameters may influence the yield of the desired target product and can be adjusted accordingly to obtain optimal amplification. Some of these include concentration, quality and the sequence of starting DNA, reagents, cycling times, temperature and the addition of solvents such as DMSO for fragments with high denaturation temperatures.

Primers previously used for PCR amplification of *GALK1* were utilized for obtaining the sequence of the coding regions of the gene and at least 100 bp of flanking intronic sequences (Table 3.2). Primers for the entire 5'UTR were designed from the published *GALK1* sequence (Genbank Accession # L76927). Due to suspected errors in the published 3'UTR sequence, primers for this region were selected from the Homo sapiens chromosome 17 clone RP11-474I111 (Genbank Accession # AC019214). Optimal cycling conditions and parameters were determined for each fragment using a series of $MgCl_2$ titrations (1mM, 1.5mM, 2mM & 2.5mM) and DMSO concentrations (1-5%) and run upon a PTC-200 gradient thermal cycler (MJ Research Inc.). PCR amplification of DNA

**Table 3.2**

PCR Primer Sequences and Conditions for *GALK1*

| *GALK1* Region | Primer sequence[a] | Nucleotide position[b] | PCR Conditions | Additives | Notes |
|---|---|---|---|---|---|
| 5'UTR (F) | 5'-CTGATGACCTCTCACAGCTGC-3' | 1-21 | 55°C for 35 cycles | 1.5 mM $MgCl_2$ | - |
| 5'UTR (R) | 5'-CGAAGTCCAGAGGGAAGGAC-3' | 224-243 | | | |
| 5'UTR (F) | 5'-CTATTCTCACTCTAGTGAGTGC-3' | 188-209 | 63°C for 35 cycles | 1.5 $MgCl_2$ | - |
| 5'UTR (R) | 5'-CGTCATGGCTGCTTTGAGACAG-3' | 473-499 | | | |
| 1 (F) | 5'-CCCGAGCATCCCGCGCCGAC-3' | 419-438 | 63°C for 35 cycles | 4% DMSO, 1.25mM $MgCl_2$ | - |
| 1 (R) | 5'-GACAGGCTGTTCCCCACGT-3' | 800-818 | | | |
| 2 (F) | 5'-ACTGTGGAGGCATCAGAACC-3' | 1436-1455 | TD 63°C→56°C | 4% DMSO, 1.5mM $MgCl_2$, | Manual hotstart |
| 2 (R) | 5'-CACAGAGCCCATTCATTTGTCTGA-3' | 1781-1805 | | | |
| 3 (F) | 5'-CCTAGAAACAGTTGCTAGGC-3' | 2071-2090 | TD 63°C→55°C | 4% DMSO, 1.5mM $MgCl_2$ | - |
| 3 (R) | 5'-GTGGTCCCAGCTTCTACTAT-3' | 2354-2373 | | | |
| 4 (F) | 5'-AGTGTCATTGAAGCCACTGC-3' | 2408-2427 | TD 63°C→57°C | 1.5 mM $MgCl_2$ | - |
| 4 (R) | 5'-CAAGCACACGCTTGGCCTCGT-3' | 2686-2706 | | | |
| 5 (F) | 5'-AGCAGCTCCTGGGTGGAGTGT-3' | 2647-2667 | TD 63°C→55°C | 4% DMSO, 1.5mM $MgCl_2$, | Manual hotstart |
| 5 (R) | 5'-CTCAGTGTGGCCTTGACCT-3' | 2971-2990 | | | |
| 6 (F) | 5'-ATCACCGCCTGCTGGTCCTC-3' | 6945-6964 | TD 63°C→55°C | 4% DMSO, 1.0mM $MgCl_2$, | Manual hotstart |
| 6 (R) | 5'- CCCAGGCCCACCCCTTCAATA-3' | 7217-7238 | | | |
| 7 (F) | 5'-CTCATGGTGGAGAGCCACCG-3' | 7146-7165 | 65°C for 30 cycles | 4% DMSO, 1.0mM $MgCl_2$ | Manual hotstart |
| 7 (R) | 5'-CTCAAGCAGCCGATGGAGCC-3' | 7530-7549 | | | |
| 8 (F) | 5'-ACGCCATGCGGCACATCCAG-3' | 7394-7413 | TD 63°C→55°C | 4% DMSO, 1.5mM $MgCl_2$ | - |
| 8 (R) | 5'-CTCTGTGCCCGGTGCCATCTT-3' | 7624-7644 | | | |
| 3' UTR (F) | 5'-GCTGTGCTTGTGAGGCACC-3' | 7555-7573 | 63°C for 30 cycles | 4% DMSO, 1.5mM $MgCl_2$ | - |
| 3' UTR (R) | 5'-CTTCCAAACTTGACCGCACCCT-3' | 8059-8081 | | | |

[a] Adapted from Kalaydjieva et al (1999)

[b] Nucleotide numbering based upon Genbank accession #L76927

according to the optimized conditions was carried out in 50µl reactions. The conditions are detailed in Table 3.2. Both the 96 well PTC-100™ Programmable Thermal Cycler (MJ Research Inc.) and the Geneamp PCR system 2400 (Applied Biosystems) were used for PCR cycling.

To check the successful amplification of each genomic target, all PCR products were size-separated by gel electrophoresis. 2.5% agarose minigels were prepared by adding 2.5 grams of DNA grade agarose powder to 100mls of 1xTAE buffer. The solution was heated in a microwave oven until agarose had dissolved, and placed in a dry incubator to cool to 60 degrees C. Prior to pouring, 3.5µl of ethidium bromide was added to each agarose solution to allow UV visualization of PCR products. The agarose solution was poured into Perspex casting trays to a depth of approximately 5mm, well combs were inserted and the gel allowed to set at room temperature under a fumehood for 30 minutes before use. Combs were then removed and the gel was placed into an electrophoresis tank (Biorad Mini-Sub® Cell GT), 1xTAE buffer was added to cover the gel to a depth of approximately 2-3mm. 5µl of each PCR product were mixed with 1µl of loading buffer (0.25% bromo blue, 0.25% xylene cyanole, 15% ficoll) and loaded into single lanes of the gel. A 1µl aliquot of pUC19/Hpa II (PE Biosystems) DNA molecular marker was loaded into the left-hand side of each gel to give an indication as to the size of each PCR product. The pUC19/HpaII ladder gives a series of bands of known nucleotide length ranging from 26 to 501 bps. This allows a visual size estimation of the PCR product to be obtained and compared with the expected fragment length. Each gel underwent electrophoresis at 80V for approximately 30 minutes and was then removed from the electrophoresis tank and visualized on a Hoefer Mighty Bright™ UV Illuminator. Images were recorded using a Kodak® DC120 digital camera and Kodak® Digital Science™ software.

After confirming sufficient amplification, PCR products were cleaned of unincorporated dNTPs, primer dimers and amplification primers, using the QIAquick

purification kit (QIAGEN, Australia). This entailed the addition of 5 volumes of PB buffer to each PCR product. Each sample was inverted to mix, then pipetted into a separate QIAquick spin column. DNA was bound to the QIAquick membrane by centrifugation at 15 000 rpm for 1 minute in an Eppendorf 5417c centrifuge. The flow-through was discarded before the addition of 750µl of PE wash buffer and further centrifugation at 15 000 rpm for 1 minute. The flow-through from this step was discarded and an additional spin under the same conditions was performed to remove any residual PE buffer. Columns were then placed into sterile, labeled 1.5ml Eppendorf tubes and the bound PCR product eluted with the addition of 30µl of $dH_2O$ and a final spin at 15 000 rpm for 1 minute. Eluted DNA was then stored at -20 degrees C until further analysis.

### *3.2.4 BigDye Terminator Cycle Sequencing*

Clean PCR products were sequenced using the ABI Prism™ Big Dye Terminator Cycle Sequencing Ready Reaction Kit. This procedure involves the incorporation of a fluorescent-labeled dideoxynucleotide at the 3' end of DNA strands. The dideoxynucleotide prevents further elongation of the template DNA and as its addition is mediated by competition with its corresponding unlabelled dNTP added to the same reaction, a series of fragments of different sizes are produced.

Each reaction was prepared on ice and consisted of 1-2µl of target DNA (20ng/µl), 4µl of BigDye Terminator, 1µl of forward or reverse primer (20ng/µl) (Table 3.2) and $dH_2O$ to a final volume of 10µl. Reactions were then placed in a 96 well PTC-100™ Programmable Thermal Cycler (MJ Research Inc.) under the following cycling parameters: rapid thermal ramp to 96 degrees C, held at this temperature for 10 seconds, a rapid thermal ramp down to 50 degrees C held for 5 seconds. This was followed by rapid ramping to 60 degrees C and held for 4 minutes. This cycling was repeated 24 more times.

### *3.2.5 Ethanol/Sodium Precipitation*

Sequencing reactions were cleaned using a precipitation reaction designed to remove excess nucleotides and primers that would otherwise interfere with the accuracy of the sequence signal. This entailed the addition of 1µl of 3M sodium acetate (pH 4.6) and 25µl of 95% ethanol to each sequencing reaction. Tubes were inverted to mix, then left at room temperature for 15 minutes to allow the precipitation of extension products before centrifugation at 3000 x g for 30 minutes in a Hermle Z823 centrifuge. Open tubes were then inverted on a tissue folded to fit the size of the carry tray and then spun at 700 x g for 1 minute to remove any residual ethanol. 75µl of 70% EtOH was then added to each tube and again tubes were inverted to mix before centrifugation at maximum speed for 10 minutes. Open tubes were then placed upside down on a clean folded tissue and spun at 700 x g for an additional minute before being left to air dry for approximately 10 minutes to ensure complete evaporation of residual ethanol.

### *3.2.6 Sequencing Gel Preparation*

Sequencing gels were prepared by adding 5 ml of filtered 10XTBE to 18g of urea in a 50mL glass graduated cylinder. 4.75mL of LongRanger Gel solution was added to this mix before the addition of $dH_2O$ to a final volume of 45mL. The top of the cylinder was covered with parafilm and inverted to dissolve the remaining urea. The final volume was brought to 50ml with $dH_2O$. The solution was filtered using 0.2µm filter paper before degassing for 10 minutes in a vacuum. Prior to pouring the gel, 250µl of 10% APS (ammonium persulphate) and 35µl of TEMED (N',N',N',N'-tetra-methylethylenediamine) were added. The solution was poured using a 50 ml syringe, the loading inserted and the gel assembly clamped. The gel was left to set for approximately 2 hours before fitting to an ABI Prism ™ 377 DNA Analyzer.

### *3.2.7 Electrophoretic Conditions*

Prior to loading, 4μl of a 5:1 EDTA (50mM) + Blue dextran (50mg/ml)/ Formamide loading solution were added to each precipitated reaction pellet. The reactions were denatured for 3 minutes at 94 degrees C. Products were then snap-chilled on ice to ensure they remained in a single-stranded conformation. 1μl of each sequencing reaction were pipetted into alternate single lanes of the gel, pre-warmed to 51 degrees C. Samples were run for 2 minutes before sequencer was paused and the remaining samples loaded in adjacent empty lanes. The electrophoresis was run for 11 hours at 2400V. Sequencing data were downloaded to an iMAC-Macintosh computer and analyzed using the Sequence Navigator™ (PE Biosytems) software.

### *3.2.8 Web-Based Databases Utilized in the GALK1 Analysis*

A number of web-based genomic and protein databases and tools have been developed to cope with the increasing amount of sequence data being deposited in the public domain. A number of these were utilized in the current investigation.

#### *3.2.8.1 Nucleotide and protein sequences*

The nucleotide and the amino-acid sequences of the Homo sapiens *GALK1* gene and galactokinase protein were obtained from the National Centre for Biological Information (NCBI) available at [http://www.ncbi.nlm.nih.gov]. Genbank accession numbers L76927, AFO84935 and U26401 were utilized for nucleotide comparisons. Protein sequence and codon numbering were derived from Genbank Accession # PS00106.

#### *3.2.8.2 Residue homology*

The BLAST alignment program, available at [http://www.ncbi.nlm.nih.gov/BLAST/], was utilized to establish whether mutations occurred at evolutionary conserved residues. It is assumed that conserved residues are under positive selection and are thus maintained across species due to their functional significance for the encoded

protein. The BLASTP alignment compares characterized protein sequences deposited in the public databases to a query sequence. In this study the human galactokinase protein sequence was compared with the galactokinase protein sequences of approximately 20 organisms.

*3.2.8.3 Restriction endonuclease selection*

The WebCutter program available at [http://www.firstmarket.com/cutter/cut2.html] was used to determine whether the detected nucleotide changes produced or abolished restriction endonuclease recognition sites. The WebCutter software utilizes a database called REBASE available at [http://www.rebase.neb.com]. The database is a compilation of characterized restriction endonucleases and associated recognition sites that is updated on a regular basis (Roberts & Macelis, 2000).

*3.2.8.4 Motif detection and alignment*

The web-based motif alignment tool SPRINT, available at [http://www.bioinf.man.ac.uk/dbbrowser/sprint/], was used to calculate and identify significant regions of homology between the human galactokinase protein and the sequences of galactokinase proteins from other organisms. The purpose of the motif alignments is to compare known or similar motifs in other proteins deposited in the public databases and obtain an indication as to the function of a protein or the family to which a protein sequence belongs.

*3.2.8.5 Secondary Structural Prediction*

Computer modeling of the predicted structure of the mutant protein was done with the PHDsec algorithm (Rost & Sander, 1994) available at [http://dodo.cpmc.columbia.edu/predictprotein/submit_def.html]. This algorithm calculates the conformational probability that a specific residue will result in the formation of a helix, loop or extended beta-sheet in the secondary folding of the protein.

### *3.2.9 Mutation Detection Techniques*

Nucleotide changes were also independently detected on a second PCR product derived from genomic DNA of each patient by using various mutation identification methods, including restriction endonuclease assays, single strand conformational polymorphism (SSCP) and heteroduplex analysis (HDA). The principles and methods for each of these techniques are described below.

#### *3.2.9.1 Restriction Endonuclease Assays*

Nucleotide changes were detected using restriction endonucleases. Isolated from prokaryotic organisms, these enzymes cut DNA at specific recognition sites. Their natural function is to remove foreign DNA sequences from the host genome. They also provide a useful tool for mutation detection if the nucleotide change in question abolishes or creates a recognition site differing from the wild-type sequence. Four restriction enzymes were utilized in the current analysis: NlaIII (New England Biolabs cat# RO125S), AlwNI (New England Biolabs cat# RO514S), BbvI (New England Biolabs cat# RO173S) and AvaI (New England Biolabs cat# RO152S). For the BbvI assay, 10µl of product (~100ng/µl) was digested with 1µl of enzyme (2 units/µl), 2µl of 1x NE buffer and 7µl of $dH_2O$ in a final reaction volume of 20µl. The protocol for the AvaI, NlaIII and AlwNI assays differed by the addition of 2µl of 1 x BSA (bovine serum albumen). All digests were incubated at 37 degrees C for 2hours. Products from the AvaI, BbvI and NlaIII digest were then run upon a 4% agarose/EtBr gel at 60V for 1 hour and visualized on a Hoefer Mighty Bright™ UV Illuminator and the image was recorded using a Kodak® DC120 digital camera and Kodak® Digital Science™ software. The AlwNI digest products were run on 12% PAG gels in 1 X TBE buffer for 1 hour at 200V. Gels were then silver stained according to the protocol in Section 3.2.9.2

*3.2.9.2 SSCP and HDA Mutation Detection*

Single-strand conformation polymorphism (SSCP) is a mutation detection technique which relies on the fact that single base changes and deletions may cause changes to the primary folding of each DNA strand resulting in aberrant electrophoretic patterns (Orita, Suzuki, Sekiya, & Hayashi, 1989).

Heteroduplex (HD) analysis involves the detection of similar shifts in mobility of mismatched duplexes of DNA strands, caused by single base substitutions and particularly insertions and deletions (White, Carvalho, O'brien, & Dean, 1992)

SSCP analysis involves denaturation of a double-stranded PCR product with formamide and heat and are snap-chilled on ice before electrophoresis on a nondenaturing polyacrylamide gel. For HD analysis, PCR products are denatured without formamide and are allowed to cool slowly so as to form both matched (homoduplex) and mismatched (heteroduplex) DNA duplexes before electrophoresis.

For both SSCP and HD analysis, a 140 bp product of exon 5 was amplified using forward primer 5'-CCAGGTCCTTGGAGACCAGC-3' and reverse primer 5'-CAAGGAAAGCCTCCGGGAGG-3'. Cycling conditions for this fragment included initial denaturation at 94 degrees C for 5 minutes, followed by 16 cycles using a touch-down regime of denaturing at 94 degrees C for 30 sec, annealing from 63 degrees C to 55 degrees C (0.5 degrees C increments) for 30 sec and extension at 72 degrees C for 30 sec. This was followed by 20 cycles of 94 degrees C for 30 sec, 55 degrees C for 30 sec and 72 degrees C for 30 sec before final extension at 72 degrees C for 7 minutes.

PCR products for SSCP analysis were prepared by adding 2µl of formamide loading buffer (98% formamide, 0.025% xylene cyanol & 0.025% bromophenol blue) and 2µl of $dH_2O$ to 2µl of each PCR reaction. Samples were denatured for 3 minutes

at 94 degrees C before being snap-chilled on ice prior to loading. Heteroduplexes were generated by independently amplifying the same exonic fragment and then subjecting 3μl of PCR product to post-amplification denaturation for 3 minutes at 94 degrees C. Samples were then allowed to cool to room temperature over a two hour duration. Prior to loading, 3μl of ficoll loading buffer was added to each tube.

Products were run on 12% PAG (29:1 bis-acrylamide) non-denaturing gels in a ice-cooled buffer consisting of 1 x TBE and 5% glycerol. Electrophoresis was carried out in a Mini-Protean® Cell II tank (Biorad) at 200V for 2hrs.

Gels were stained using a silver staining technique, which offers a higher degree of sensitivity than ethidium bromide in detecting and visualising DNA on thin gels. After electrophoresis, gels were soaked in 10% ethanol solution and gently agitated for 3 minutes. This solution was drained, and sufficient 1% nitric acid was added to cover the gel(s). Again, gels were left to soak for 3 minutes before the solution was drained and three washes with $dH_2O$ were performed. Gels were then soaked under gentle agitation in 0.1% silver nitrate solution for 10 minutes in a light-protected tray. The solution was discarded and gels were washed three times in $dH_2O$. A solution made up of 25g of sodium carbonate dissolved in 250ml of $dH_2O$ plus 300μl of formaldehyde was used to develop the DNA bands. Gels were soaked for approximately 1 minute until the solution changed colour. The developing solution was then drained and a fresh measure was added. The gels were left to soak under gentle agitation until the bands were seen. The remaining solution was discarded and the developed gels were rinsed three times in $dH_2O$ before fixation in a 10% acetic acid solution for 10 minutes. A final rinse with $dH_2O$ was followed by a 10 minute soak in 1% glycerol to prevent cracking of the gel during the drying process. Gels were placed between two drying film sheets and clamped within a Promega gel drying rack and left in the fumehood to dry overnight. Dried gels were photographed

on a light box using a Kodak® DC120 digital camera and Kodak® Digital Science™ software.

## 3.3 RESULTS

Analysis of the *GALK1* gene in nine patients detected one previously identified disease-causing mutation and 4 novel mutations that are likely to cause the galactokinase deficiency phenotype (Table 3.3).

**Table 3.3**

*GALK1* Mutations in Patients with Galactokinase Deficiency.

| Patient | RBC GK activity | Nucleotide change[a] | Amino acid change | Exo | Genotype | Ethnic Origin |
|---|---|---|---|---|---|---|
| 1 | Deficient | 1569 C→T | R68C | 2 | Heterozygous | European/American |
| | | 7538G→C | A384P | 8 | Heterozygous | |
| 2 | Deficient | 7093C→T | T288M | 6 | Heterozygous | European/American |
| | | 2833delC | Frameshift after Y236 | 5 | Heterozygous | |
| 3 | Deficient | 563C→A | P28T | 1 | Homozygous | Turkish |
| 4 | Deficient | 563C→A | P28T | 1 | Homozygous | Spanish Gypsy |
| 5 | Deficient | 563C→A | P28T | 1 | Homozygous | Spanish Gypsy |
| 6 | Deficient | 563C→A | P28T | 1 | Homozygous | Hungarian Gypsy |
| 7 | Deficient | 563C→A | P28T | 1 | Homozygous | Hungarian Gypsy |
| 8 | Deficient | 563C→A | P28T | 1 | Homozygous | Hungarian Gypsy |
| 9 | Deficient | 563C→A | P28T | 1 | Homozygous | Bulgarian Gypsy |

[a] Nucleotide numbering according Genbank Accession # L76927.

### *3.3.1 Patient 1*

Patient 1 is a compound heterozygote for two nucleotide changes that result in the novel missense mutations, R68C and A384P.

**R68C**

Sequence analysis revealed a 1569C→T transition in exon 2 (Figure 3.1). The 1569C→T transition is predicted to result in the replacement of arginine by cysteine at codon 68 (Figure 3.2). The arginine residue at this position is strictly conserved in the galactokinase protein of over twenty different species checked, including the mouse. The mutation is a drastic change from a charged to an uncharged amino acid residue with a sulfhydryl side-chain (Figure 3.3), however the computer predictions did not suggest significant alterations in the secondary structure of the mutant protein (Figure 3.4).

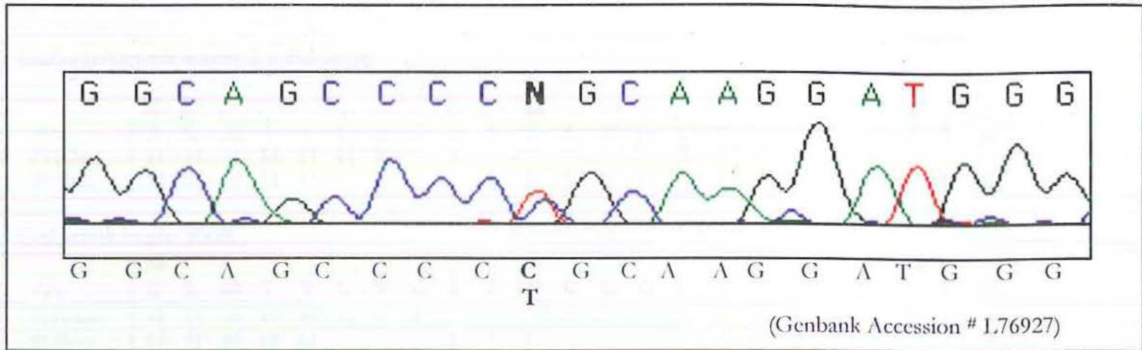



**Figure 3.1** 1569C→T transition in *GALK1* exon 2 causing R68C.

| **Wildtype** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Codon** | 65 | 66 | 67 | **68** | 69 | 70 | 71 | 72 |
| **Amino acid** | G | S | P | **R** | K | D | G | L |
| **Nucleotide** | GGC | AGC | CCC | **C**GC | AAG | GAT | GGG | CTG |
| **R68C** | | | | | | | | |
| **Codon** | 65 | 66 | 67 | **68** | 69 | 70 | 71 | 72 |
| **Amino acid** | G | S | P | **C** | K | D | G | L |
| **Nucleotide** | GGC | AGC | CCC | **T**GC | AAG | GAT | GGG | CTG |

**Figure 3.2** Residue change at codon 68 caused by the 1569C→T transition in exon 2. Arginine is replaced with cysteine.

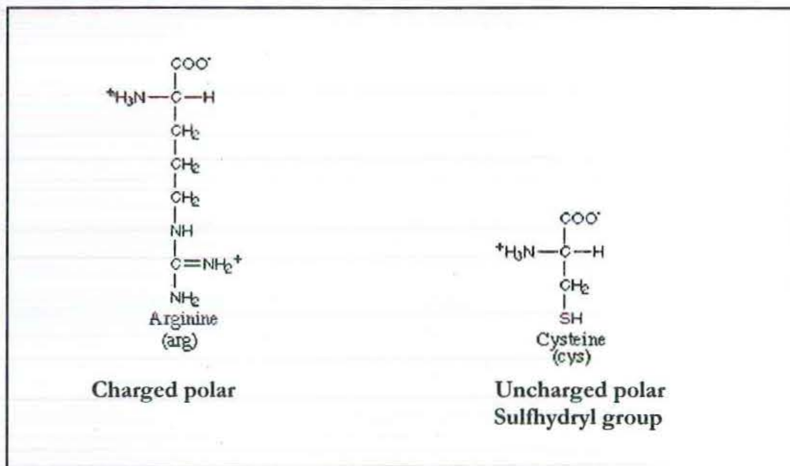



**Figure 3.3** R68C - molecular comparisons between the arginine and cysteine residues.

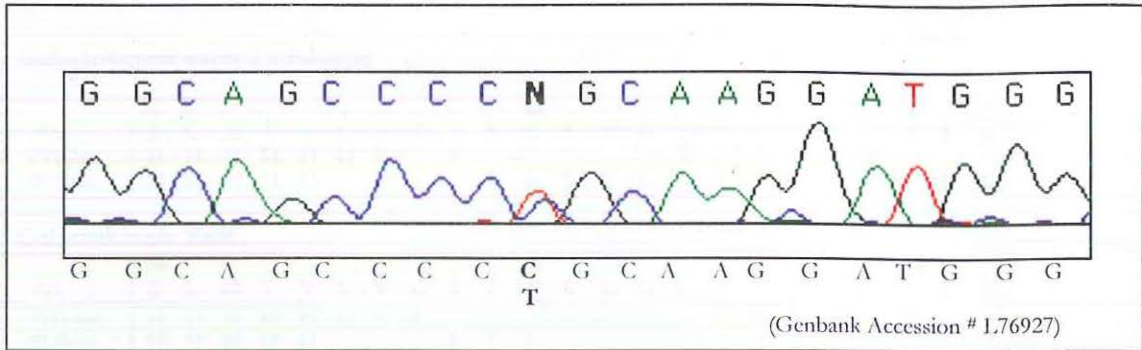



**Figure 3.1** 1569C→T transition in *GALK1* exon 2 causing R68C.

| **Wildtype** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Codon** | 65 | 66 | 67 | **68** | 69 | 70 | 71 | 72 |
| **Amino acid** | G | S | P | **R** | K | D | G | L |
| **Nucleotide** | GGC | AGC | CCC | **C**GC | AAG | GAT | GGG | CTG |
| **R68C** | | | | | | | | |
| **Codon** | 65 | 66 | 67 | **68** | 69 | 70 | 71 | 72 |
| **Amino acid** | G | S | P | **C** | K | D | G | L |
| **Nucleotide** | GGC | AGC | CCC | **T**GC | AAG | GAT | GGG | CTG |

**Figure 3.2** Residue change at codon 68 caused by the 1569C→T transition in exon 2. Arginine is replaced with cysteine.

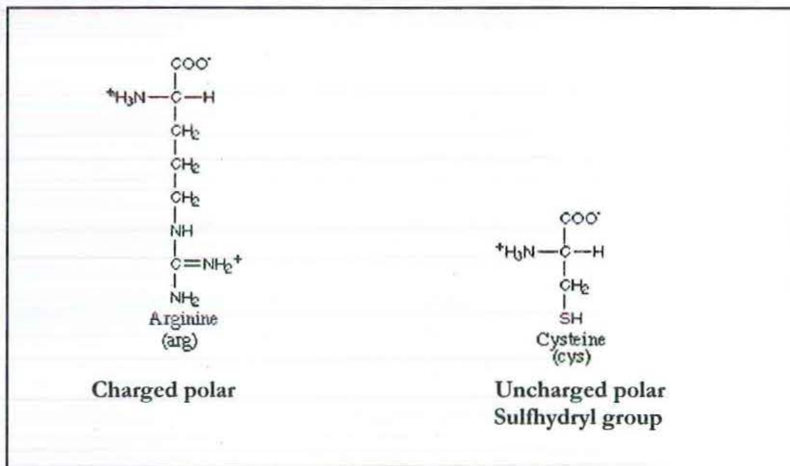



**Figure 3.3** R68C - molecular comparisons between the arginine and cysteine residues.

**Galactokinase exon 2 wild-type**

| | 58 | . | . | . | . | , | . | . | . | . | 68 | . | . | . | . | , | . | . | . | . | 78 | . | . |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | E | L | M | T | V | L | V | G | S | P | R | K | D | G | L | V | S | L | L | T | T | S | E |
| PHDsec | H | H | H | H | H | H | E | | | | | | | | E | E | E | E | E | E | | | |
| SUBsec | H | H | H | H | H | . | . | . | . | L | L | L | L | L | . | E | E | E | E | . | . | L | L |

**Galactokinase R68C**

| | 58 | . | . | . | . | , | . | . | . | . | 68 | . | . | . | . | , | . | . | . | . | 78 | . | . |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | E | L | M | T | V | L | V | G | S | P | C | K | D | G | L | V | S | L | L | T | T | S | E |
| PHDsec | H | H | H | H | H | H | E | E | | | | | | | | E | E | E | E | E | E | | |
| SUBsec | H | H | H | H | H | . | . | . | L | L | L | . | . | . | . | E | E | E | E | E | . | . | . |

**Fig ure 3.4** Secondary-structure prediction for the wild-type exon two galactokinase sequence and the R68C missense mutation using the PHDsec algorithm. C68 in boldtype face. E = extended sheet (beta sheet); H = helix; L = loop. The prediction is meaningful for all residues with an expected average correlation >.69.

The 1569C→T transition (R68C) creates an Alw NI restriction site marked by the sequence CAGN$_3$/CTG (Figure 3.5). The PCR-based restriction assay, used as a screening test, confirmed the presence of the mutation in patient 1 and failed to detect the mutation among 100 control chromosomes.

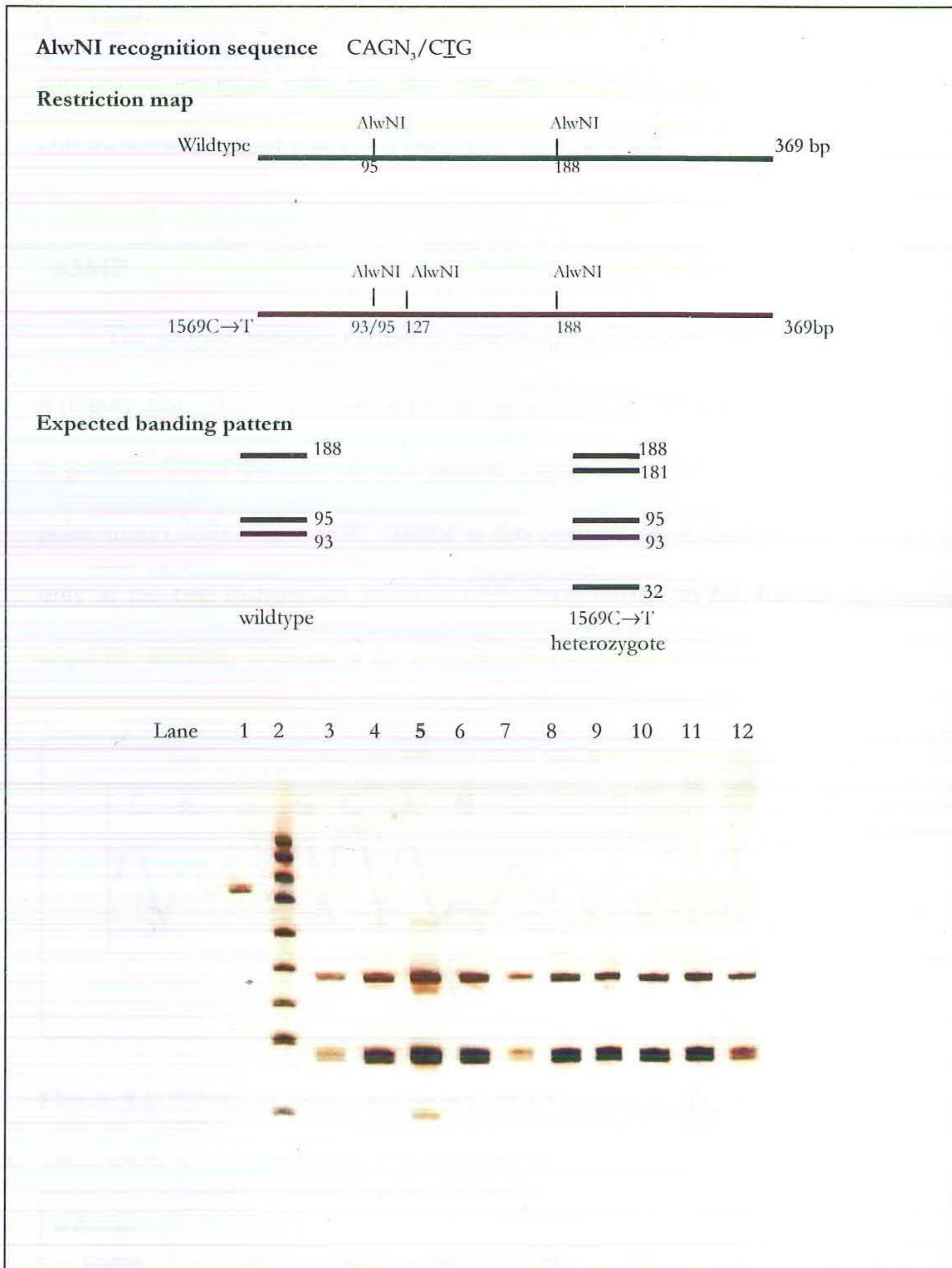



**Figure 3.5** AlwNI restriction digest detecting 1569C→T transition in exon 2 on silver-stained 12% PAG. Lane 1: Undigested exon 2 patient 1. Lane 2: pUC19/HpaII Ladder. Lane 3,4, 6-12: Unaffected. Lane 5: Patient 1 digest.

The highly conserved nature of the amino acid residue, as well as the drastic nature of the substitution coupled with the fact that the mutation did not occur among control chromosomes, suggest that it it is likely that R68C is a disease-causing mutation.

**A384P**

The second mutation carried by patient 1 is a 7538G→C transversion found in exon 8 (Figure 3.6). The 7538G→C base change leads to the substitution of proline for alanine at position 384 of the galactokinase protein (Figure 3.7). Both alanine and proline are non-polar amino acids (Figure 3.8). Alanine at this position is not highly conserved, it is present only in the two mammalian galactokinases characterized so far, human and mouse, and sequence similarity is found in the proteins of seven additional species.

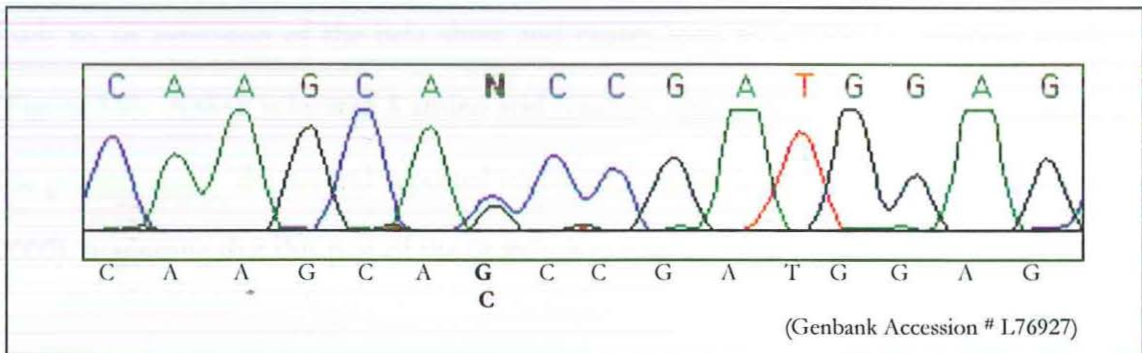



**Figure 3.6** 7538G→C transversion in *GALK1* exon 8 causing A384P.

| **Wildtype** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Codon** | 381 | 382 | 383 | **384** | 385 | 386 | 387 | 388 |
| **Amino acid** | S | Q | A | **A** | D | G | A | K |
| **Nucleotide** | TCT | CAA | GCA | **G**CC | GAT | GGA | GCC | AAG |
| **A384P** | | | | | | | | |
| **Codon** | 381 | 382 | 383 | **384** | 385 | 386 | 387 | 388 |
| **Amino acid** | S | Q | A | **P** | D | G | A | K |
| **Nucleotide** | TCT | CAA | GCA | **C**CC | GAT | GGA | GCC | AAG |

**Figure 3.7** Residue change at codon 384 caused by the 7538G→C transversion. Wildtype alanine is replaced with proline.

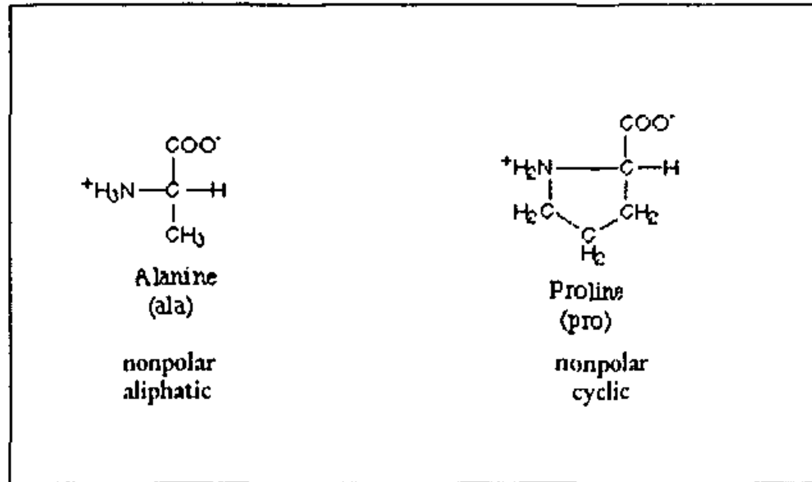



**Figure 3.8** A384P- molecular comparisons between the alanine and proline residues

The substitution of proline for alanine is likely to induce significant topological changes, as predicted by the secondary structural analysis. The helix formed between the F378 and A384 residues in the wild-type protein is abolished by the mutation, which also leads to an extension of the beta sheet and causes loop formation by adjacent residues (Figure 3.9). A384P is located 8 amino acid residues upstream of the carboxyl terminus of the protein and is the second reported mutation in exon 8, after Q382X (Kolosha et al., 2000), suggesting that this part of the protein is essential for galactokinase function.

**Galactokinase exon 8 wild-type**

| | 370 | . | . | . | . | . | . | . | . | . | 380 | . | . | . | . | . | . | . | . | . | 390 | . | . |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | E | H | Y | G | G | T | A | T | F | Y | L | S | Q | A | A | D | G | A | K | V | L | C | L |
| PHDsec | H | H | | | | | E | E | H | H | H | H | H | H | H | | | | E | E | E | E | |
| SUBsec | H | H | . | L | L | L | . | . | . | . | . | . | . | . | . | . | L | L | . | . | . | . | L |

**Galactokinase A384P**

| | 370 | . | . | . | . | . | . | . | . | . | 380 | . | . | . | . | . | . | . | . | . | 390 | . | . |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | E | H | Y | G | G | T | A | T | F | Y | L | S | Q | A | **P** | D | G | A | K | V | L | C | L |
| PHDsec | H | H | H | | | | E | E | E | E | E | E | | | | | | | E | E | E | E | |
| SUBsec | H | H | H | . | . | . | E | E | E | E | E | . | . | L | L | L | L | . | . | . | . | . | L |

**Figure 3.9** Secondary-structure prediction for the wild-type exon eight galactokinase sequence and the A384P missense mutation using the PHDsec algorithm. P384 in boldtype face. E = extended sheet (beta sheet); H = helix; L = loop. The prediction is meaningful for all residues with an expected average correlation > .69.

The 7538G→C transversion (A384P) abolishes a BbvI endonuclease restriction site marked by the sequence 5' GCA_GCN$_8$ 3' and 3' CGTC_GN$_{12}$ 5'. To confirm that the transversion does not occur in the wild-type population, a Bbv I restriction assay was performed on the PCR product encompassing the entire exon 8 sequence from both the proband and 50 unaffected controls (Figure 3.10). The 7538G→C transversion was not detected among the control samples assayed.

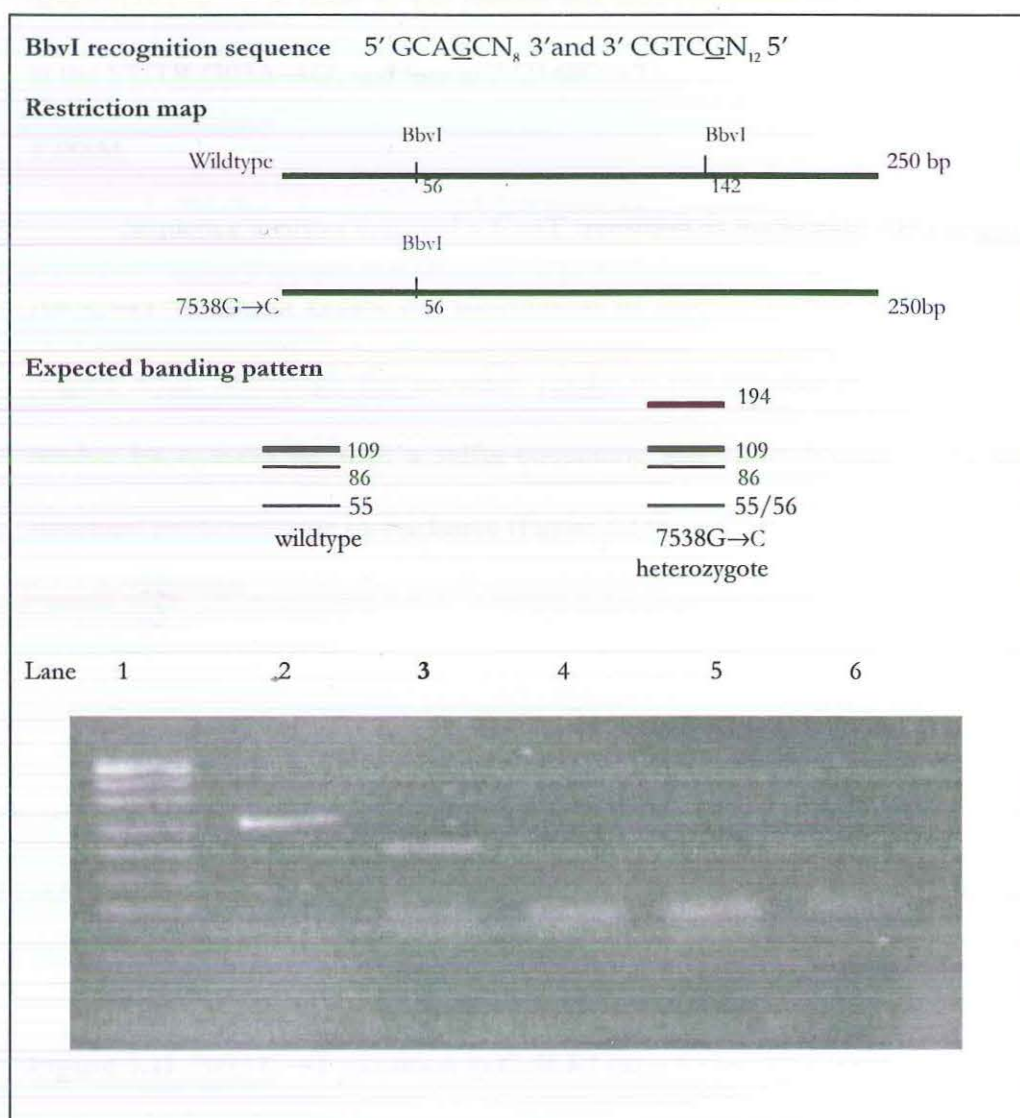



**Figure 3.10** Bbv I Restriction Digest of Exon 8 PCR Product. Lane 1: pUC19/Hpa II Ladder. Lane 2: Undigested exon 8 of patient 1. Lane 3: patient 1 digest. Lanes 4,5,6: Unaffected controls.

It is likely that A384P is a disease-causing mutation based upon the fact that it induces topological changes in the secondary structure of the protein, resulting from the change from an aliphatic (alanine) residue to a cyclic (proline) residue. This is supported by failure to detect it among 100 control chromosomes.

### *3.3.2 Patient 2*

Patient 2 is heterozygous for a nucleotide transistion in exon 6 and a single base deletion occurring in exon 5. The patient was also heterozygous for nucleotide transitions in the 5'UTR (303A→G) and intron 2 (2148C→T).

**T288M**

Sequence analysis detected a C→T transition at nucleotide 7093 (Figure 3.11). The 7093C→T transition causes the substitution of methionine for threonine at codon 288 (Figure 3.12). Although the mutation results in the replacement of a polar amino acid residue by a nonpolar with a sulfur-containing side-chain (Figure 3.13), the secondary structural predictions are inconclusive (Figure 3.14).

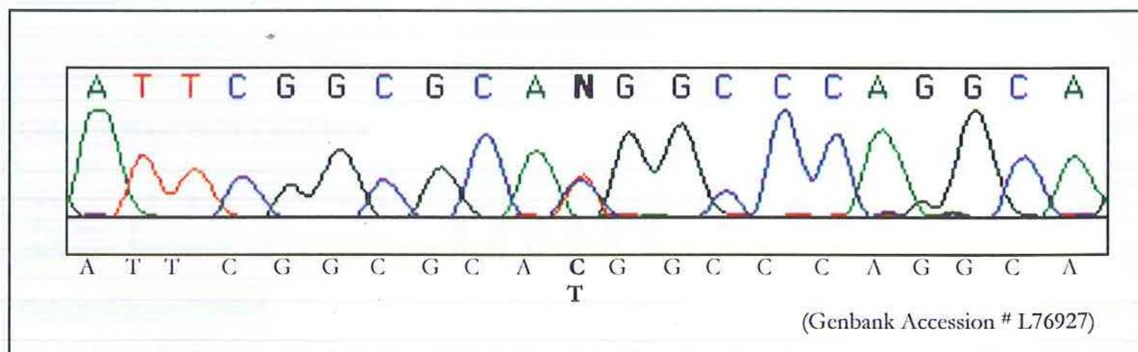



**Figure 3.11** 7093 C→T transition in *GALK1* exon 6 causing T288M.

| **Wildtype** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Codon | 285 | 286 | 287 | **288** | 289 | 290 | 291 | 292 |
| Amino acid | I | R | R | T | A | Q | A | A |
| Nucleotide | ATT | CGG | CGC | ACG | GCC | CAG | GCA | GCG |
| **A384P** | | | | | | | | |
| Codon | 285 | 286 | 287 | **288** | 289 | 290 | 291 | 292 |
| Amino acid | I | R | R | M | A | Q | A | A |
| Nucleotide | ATT | CGG | CGC | ATG | GCC | CAG | GCA | GCG |

**Figure 3.12** Residue change at codon 288 caused by the 7093C→T transition. Wildtype threonine is replaced with methionine.

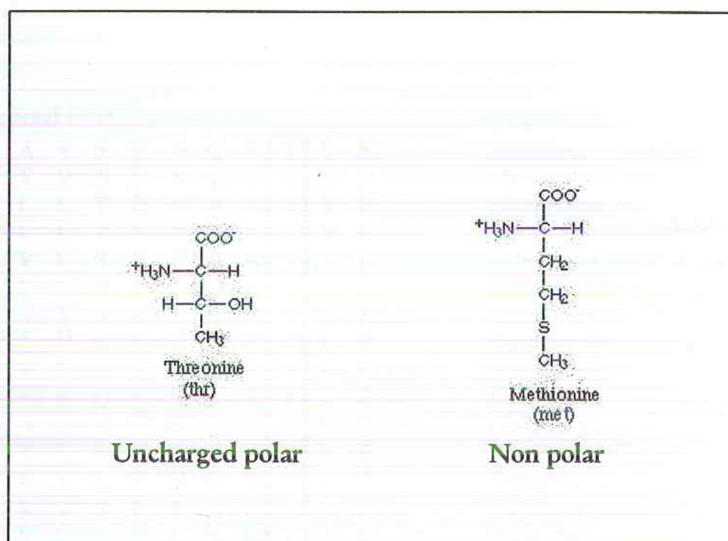



**Figure 3.13** T288M – molecular comparisons between the threonine and methionine residues

**Galactokinase exon 6 wild-type**

| | . | . | . | . | , | . | . | . | . | 280 | . | . | . | . | , | . | . | . | . | 290 | . | . | . |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | S | K | E | G | F | R | R | A | R | H | V | V | G | E | I | R | R | T | A | Q | A | A | A |
| PHDsec | | | | | | | | | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |
| SUBsec | L | L | L | L | L | . | . | . | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |

**Galactokinase T288M**

| | . | . | . | . | . | , | . | . | . | 280 | . | . | . | . | , | . | . | . | . | 290 | . | . | . |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | S | K | E | G | F | R | R | A | R | H | V | V | G | E | I | R | R | **M** | A | Q | A | A | A |
| PHDsec | | | | | | E | E | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |
| SUBsec | L | L | L | L | L | . | . | . | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |

**Figure 3.14** Secondary-structure prediction for the wild-type exon six galactokinase sequence and the T288M missense mutation using the PHDsec algorithm. M288 in boldtype face. E = extended sheet (beta sheet); H = helix; L = loop. The prediction is meaningful for all residues with an expected average correlation > .69.

T288M is the first mutation identified in exon 6 of *GALK1* and occurs in a region identified by the web-based motif alignment tool, SPRINT (available http://www.bioinf.man.ac.uk/dbbrowser/sprint/) as encoding a conserved galactokinase fingerprint motif sequence. As shown by the sequence alignment of this galactokinase motif (Figure 3.15), the threonine residue at position 288 is conserved in nine out of twenty galactokinases. The residue at this position in all but one of the remaining galactokinases is valine, which does not differ drastically in terms of side-chain structure from threonine.

| **Galactokinase motif four** | | | | | | | | | | | | | | | **Organism** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| K | R | A | R | H | A | V | S | E | N | Q | R | T | L | K | *Streptomyces lividans* |
| K | R | A | R | H | V | V | T | E | N | Q | R | V | L | D | *Haemophilus influenzae* |
| K | R | V | R | H | I | L | T | E | N | A | R | T | V | E | *Escherichia coli* |
| K | R | V | R | H | I | L | T | E | N | A | R | T | V | E | *Bacteriophage lambda* |
| K | R | V | R | H | V | L | S | E | N | A | R | T | V | E | *Salmonella typhimurium* |
| K | R | A | K | H | I | I | S | E | N | Q | R | V | L | E | *Actinobacillus pleuropneumoniae* |
| Q | R | A | K | H | V | Y | S | E | S | L | R | V | L | K | *Saccharomyces cerevisiae* |
| K | R | A | R | H | A | V | S | E | N | Q | R | T | L | R | *Lactobacillus helveticus* |
| Q | R | A | K | H | V | Y | S | E | A | L | R | V | L | K | *Kluyveromyces lactis* |
| Q | R | A | K | H | V | Y | Q | E | S | L | R | V | L | E | *Candida albicans* |
| E | R | A | L | H | V | Y | R | E | S | L | R | V | L | K | *Candida parapsilosis* |
| R | R | A | K | H | A | V | Y | E | N | H | R | A | I | K | *Bacillus subtilis* |
| K | R | A | R | H | V | V | L | E | N | Q | R | T | L | Q | *Streptococcus thermophilus* |
| Q | R | A | K | H | V | Y | S | E | S | L | R | V | L | K | *Saccharomyces cerevisiae (GAL3)* |
| R | R | A | R | H | V | V | G | E | I | R | R | T | A | Q | *Homo sapiens* |
| R | R | A | R | H | V | V | S | E | I | R | R | T | A | Q | *Mus musculus* |
| K | R | A | R | H | A | V | F | E | N | Q | R | T | L | K | *Lactobacillus casei* |
| K | R | A | Q | H | V | L | E | E | N | E | R | V | L | K | *Thermotoga neapolitana* |
| K | F | F | G | Y | I | V | R | E | N | A | R | V | L | E | *Pyrococcus horikoshii* |
| R | L | V | R | H | V | V | T | E | D | E | R | V | E | R | *Streptomyces lividans* |

**Figure 3.15** Galactokinase fingerprint motif four. Bold highlight showing amino acid variation analogous to residue 288 of the human galactokinase protein at which the T288M substitution is found. Methionine at this residue is not shared by any of the galactokinases that share the fourth galactokinase fingerprint motif. Motif alignment by SPRINT (available http://www.bioinf.man.ac.uk/dbbrowser/sprint/)

The 7093C→T change creates a Nla III endonuclease restriction site by forming the recognition sequence CA<u>T</u>G/ (Figure 3.16). Exon 6 of the affected proband was amplified by PCR using previously described primers and conditions (Table 3.2). The mutation was tested in the panel of control individuals using the restriction assay. It was not found among 100 normal chromosomes.

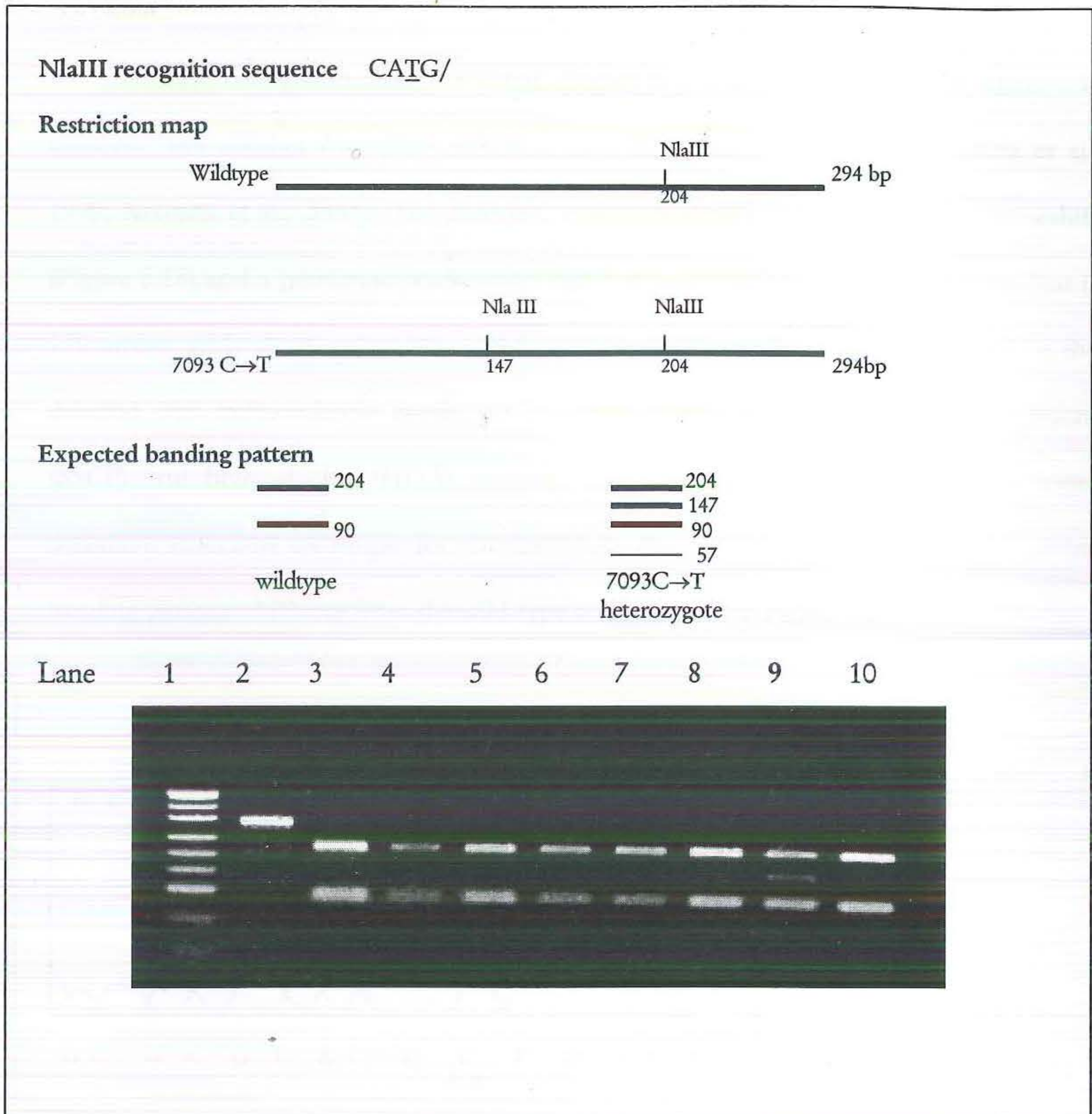



**Figure 3.16** Nla III Restriction Digest of 7093 C→T (T288M) in exon 6 PCR Product. Lane 1: pUC19/Hpa II Ladder. Lane 2. Undigested exon 6 of proband. Lanes 3-8 & 10: Unaffected controls. Lane 9: Patient 2 digest.

It is likely T288M is a disease-causing mutation based on the drastic nature of the substitution, the observation that it occurs within a fingerprint motif sequence at a residue that shows a moderate degree of conservation and the failure to detect it among 100 control chromosomes.

**2833delC**

A single base-pair deletion, 2833delC (Figure 3.17), was found in exon 5, where one missense and another truncating mutation have been reported previously (Asada et al., 1999; Kolosha et al., 2000). The 2833delC mutation is predicted to result in a frameshift (Figure 3.18) and a premature termination signal at codon 263, leading to a protein that is 130 amino acids shorter than the wild-type enzyme (Figure 3.19). The presence of the deletion was independently confirmed by single strand conformational polymorphism (SSCP) and heteroduplex (HDA) analysis. The SSCP analysis proved to be a more definitive detection technique for the 2833delC deletion as evidenced by the aberrant banding patterns differing from the wild-type controls seen in Figure 3.20.

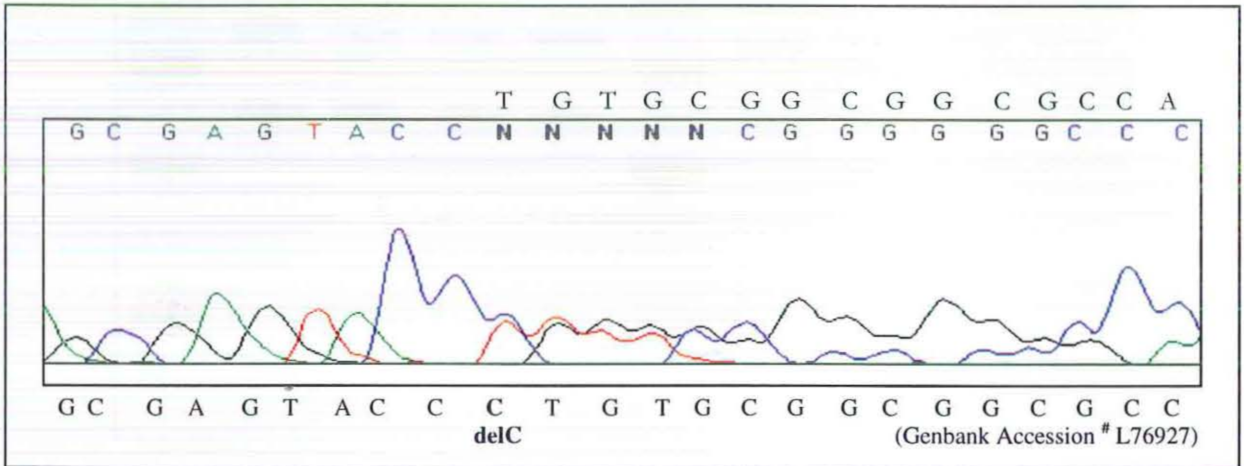



**Figure 3.17** 2833delC in exon 5 resulting in a shift of the reading frame at codon 237

**Wildtype**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Codon** | 233 | 234 | 235 | 236 | **237** | 238 | 239 | 240 |
| **Amino acid** | S | S | E | Y | **P** | V | R | R |
| **Nucleotide** | TCC | AGC | GAG | TAC | **CCT** | GTG | CGG | CGG |

**2833delC**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Codon** | 233 | 234 | 235 | 236 | **237** | 238 | 239 | 240 |
| **Amino acid** | S | S | E | Y | **L** | C | G | G |
| **Nucleotide** | TCC | AGC | GAG | TAC | **CTG** | TGC | GGC | GGC |

**Figure 3.18** Frameshift at codon 237 caused by the 2833delC deletion. Wildtype proline is replaced by leucine and novel protein sequence.

SLETSLVPLSDPKLAVLITNSNVRHSLASSEYPVRRRQCEEVARALGKESLREVQLEELE

SLETSLVPLSDPKLAVLITNSNVRHSLASSEY**LCGGANVKKWPGRWARKASGRYNWKS***

**Figure 3.19** Protein sequence of *GALK1* exon 5 wild-type (upper sequence) and novel protein sequence (lower sequence) and premature stop signal resulting from the frameshift caused by 2833delC.

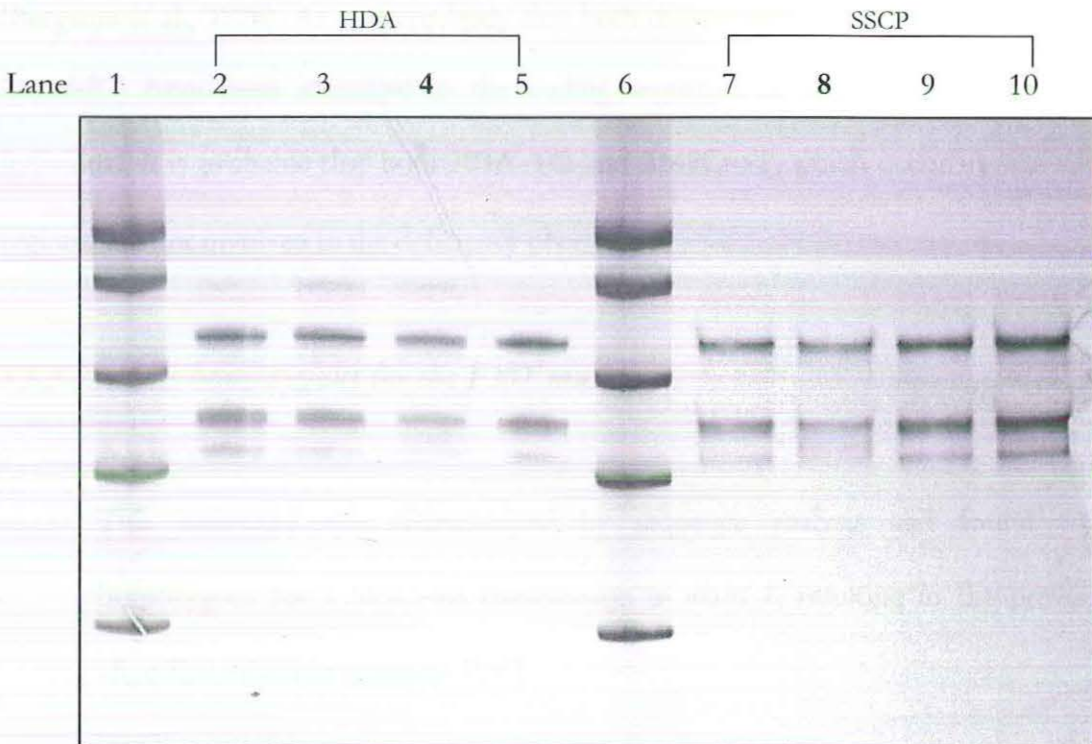
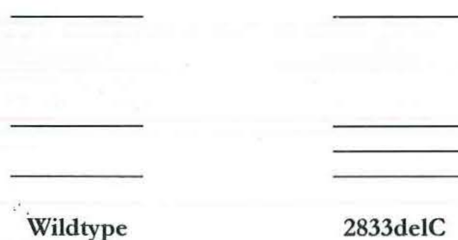



**Figure 3.20** Detection of the 2833delC in *GALK1* exon 5 by heteroduplex analysis HDA and single strand conformational polymorphism SSCP. Products were run on 12% PAG at 200V for ~2hrs. Lane 1: pUC19/HpaI marker. Lanes 2,3 & 5: Unaffected controls. Lane 4: Patient 2 HDA Lane 6: pUC19/HpaI marker. Lanes 7, 9 & 10: Unaffected controls. Lane 8: Patient 2 SSCP.

Schematic representation of aberrant banding pattern as detected by SSCP



**Other Nucleotide changes**

Patient 2 is heterozygous for a 2148C→T transition in intron 2. This change is located 34 bp upstream of the 5' end of exon 3 and 3 bp upstream of a putative eukaryotic consensus branch point sequence, YNRAY (Rautmann & Breathnach, 1985, Chiara et al., 1996). Patient two is also heterozygous for a 303A→G transition in the 5'UTR. This base change occurs within the 250 bp putative eukaryotic polymerase II promoter region (Bergsma et al., 1996). As it seems likely that both disease-causing mutations (T288M and 2833delC) have been identified in the coding sequence of the *GALK1* gene in this individual, it is probable that both 303A→G and 2148C→T, which occur in non-coding regions, are not involved in the deficiency phenotype.

### *3.3.3 Patients homozygous for the P28T mutation*

#### *3.3.3.1 Patient Three*

This individual was characterized by sequence analysis and found to be homozygous for a 563C→A transversion in exon 1, resulting in the previously identified missense mutation P28T.

**P28T**

Sequencing revealed a 563C→A transversion on both the forward and reverse strands of DNA. The mother of this patient is heterozygous for this nucleotide change (Figure 3.21). This nucleotide transversion results in the substitution of proline for threonine at codon 28 and has been previously shown to induce topological changes in the secondary structure of the protein (Kalaydjieva et al., 1999). The P28T mutation occurs in close proximity to the galactokinase signature sequence in exon 1 signaling the importance of this region to normal enzyme function.

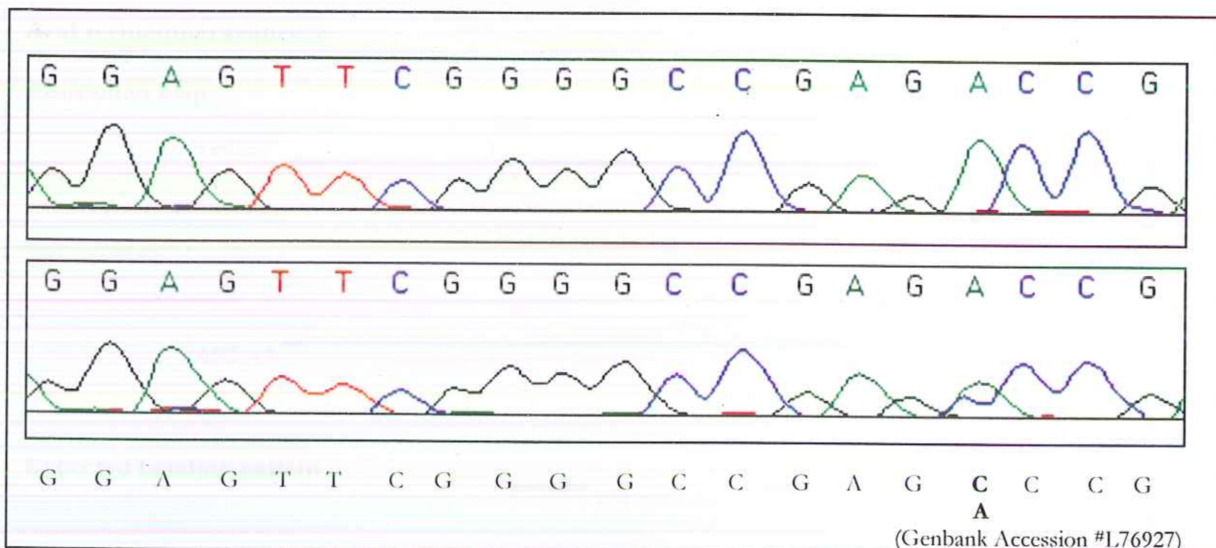



**Figure 3.21** 563C→A resulting in the P28T substitution. Upper sequence shows homozygous affected patient. Lower sequence shows mother of patient is heterozygous for same nucleotide transversion.

The 563C→A transversion abolishes an AvaI restriction site and was also confirmed using this method (Figure 3.22). In addition, both mother and child share a 7746C→T base substitution in the 3'UTR. As with the 563C→A transversion, patient 3 is homozygous for this change and the mother is heterozygous. It was not determined whether this change occurs in other carriers and homozygotes for the P28T mutation.

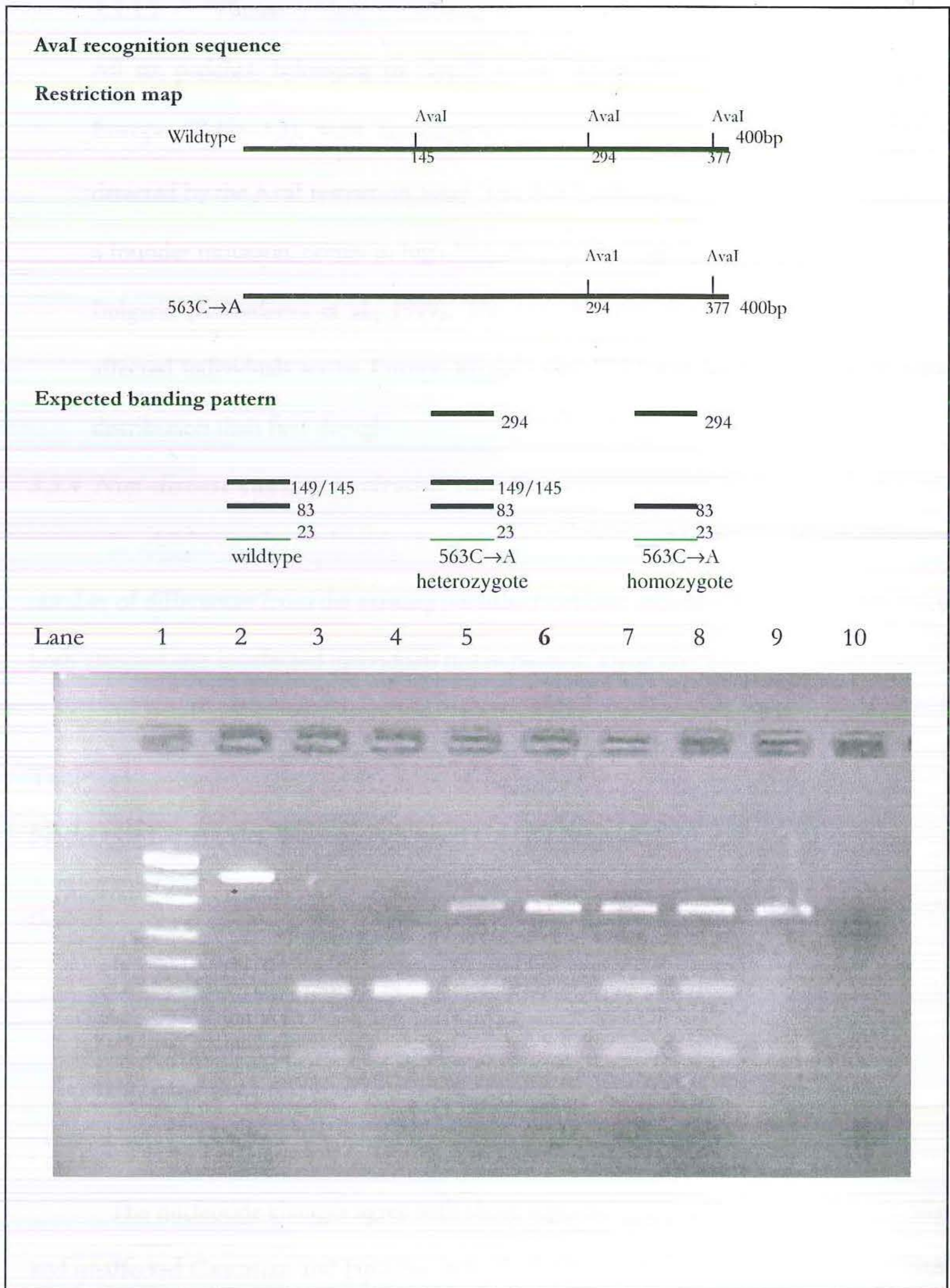



**Figure 3.22** Ava I restriction endonuclease digest of 563C→A transversion. Lane 1: pUC19/HpaII marker. Lane 2: Undigested exon product. Lanes 3 & 4: Unaffected controls. Lane 5: Mother of patient 3. Lane 6: Patient 3. Lanes 7 & 8: heterozygous controls. Lane 9: homozygous control. Lane 10: negative control.

#### 3.3.3.2 *Patients of Gypsy ethnic background*

All six patients, belonging to Gypsy groups geographically dispersed throughout Europe (Table 3.3), were homozygous for the 563C→A transversion (P28T) as detected by the AvaI restriction assay. The P28T substitution, previously described as a founder mutation, occurs in high frequency in the Vlax Roma Gypsy population in Bulgaria (Kalaydjieva et al., 1999). The fact that this mutation is shared amongst affected individuals across Europe suggests that P28T may have a more widespread distribution than first thought.

### *3.3.4 Non-disease causing nucleotide variations*

In addition to the nucleotide changes that were detected in the sample of patients, a number of differences from the existing published *GALK1* sequence were also identified in both affected and unaffected individuals run in parallel. These are detailed in Table 3.4.

**Table 3.4**

Nucleotide Differences Between Published *GALK1* Sequences.

| Nucleotide # | Region | Nucleotide and Source | | |
|---|---|---|---|---|
| Genbank Accession #L76927 | | Bergsma et al. (1996)[a] | Asada et al. (1999)[b] (n=7) | This study (n=9) |
| 415 | 5UTR | C | T | T |
| 418 | 5UTR | C | T | T |
| 694 | intron 1 | C | T | T |
| 2149 | intron 2 | C | G | G |
| 2151 | intron 2 | A | C | C |
| 2172 | exon 3 | C | delC | delC |

[a] Genbank accession # L76927
[b] Genbank accession # AF084935

The nucleotide changes agree with those reported in a previous study of both affected and unaffected Caucasian and Japanese individuals (Asada et al., 1999). This suggests that the changes are probably errors in the original sequence (Bergsma et al., 1996) rather than true sequence variants.

## 3.4 DISCUSSION

The sequence analysis of the entire *GALK1* coding region, flanking intronic sequence and both the 5' and 3' UTR in three patients with galactokinase deficiency revealed 3 novel missense amino acid substitutions (R68C, A384P, T288M) and a single base deletion (2833delC) which are likely to be disease-causing mutations. In addition one previously characterized missense mutation (P28T) was also detected.

### *3.4.1 Nature of the newly identified mutations*

The R68C substitution in exon 2 occurs at a highly conserved residue that is preserved in the galactokinase protein sequence of a wide number of organisms including the mouse, yeast and bacteria. This high degree of evolutionary conservation, coupled with the nature of the amino acid substitution, suggests that this mutation is responsible for an interruption to normal galactokinase activity. As the substitution does not occur in a recognized motif sequence, the role that this region of the gene has in the normal activity of the enzyme is unclear. R68C is the fourth mutation described in exon 2, the remaining three being truncating mutations (Stambolian et al., 1995; Kolosha et al., 2000).

The A384P change in exon 8 occurs at a residue that is conserved in the murine protein sequence and shares a homologous residue among bacteria and yeast. This substitution also occurs 2 residues downstream of a previously identified Q382X mutation implicated in causing a reduction in galactokinase activity (Kolosha et al., 2000). The close proximity of these mutations signifies that the carboxyterminus of the protein encoded by exon 8 plays an important role in the normal function of the galactokinase protein. It is likely that the A384P substitution disrupts this function by inducing topological changes, as noted by the secondary structural predictions.

The fact that R68C and A384P were not detected among 100 normal chromosomes suggests that these amino-acid substitutions are the cause of the galactokinase deficiency phenotype in patient 1.

The T288M substitution is the first reported mutation to occur in exon 6 of the *GALK1* gene. This amino-acid change is located within the fourth galactokinase fingerprint motif whose function has yet to be characterized. The threonine residue at this position is conserved in the galactokinase protein of 7 organisms, including the mouse. The nature of the substitution (polar to nonpolar) also suggests that this amino acid change causes a reduction of galactokinase activity by interfering with a critical function provided by the fourth fingerprint sequence.

The 2833delC deletion in exon 5 would produce a truncated non-functional galactokinase protein that is significantly shorter than the wild-type enzyme. When coupled with the T288M missense mutation, it is likely that galactokinase activity would be reduced to levels conducive to the deficiency phenotype, as is the case in the biochemical description of patient 2. The intronic 2149T→C transition also carried by this patient may potentially effect the transcriptional processing of exon three of the *GALK1* gene by interfering with a putative branchpoint motif consensus sequence (Rautmann & Breathnach, 1985, Chiara et al., 1996). Similarly the 303A→G transition found within the promoter region of the 5'UTR may also affect expression. However without functional expression studies both these possibilities are unclear. Evidence will be presented in Chapter 4 that these nucleotide changes are not a cause of reduced erythrocyte galactokinase activity as the 2149T→C transversion is shared, in the homozygous state, by an individual with normal erythrocyte galactokinase activity. The same individual is also heterozygous for the 303A→G transition.

The detection of the previously characterized P28T mutation in patient 3 suggests that the incidence of this particular mutation may occur in groups other than the Roma population. If in fact Patient 3 is of true Turkish extraction as reported, this finding may signal the possibility that the P28T mutation may have a more widespread distribution across populations in Europe than first thought. Supporting this, was the detection of the P28T mutation in eight patients belonging to geographically dispersed Gypsy groups throughout Bulgaria, Hungary and Spain. Further analysis as to whether the P28T mutation found in these patients is the same founder mutation common to the Vlax Roma or reflects a recurrence of the substitution at the same codon, forms the basis for the investigation in Chapter 5.

### *3.4.2 Nucleotide variations*

It is likely that the non-disease associated nucleotide variations from the published *GALK1* sequence detected in this study reflect errors in the sequence deposited in the Genbank database (Genbank accession# L76927). Support for this claim comes from the fact that the same nucleotides were also detected previously among both Caucasian and Japanese samples (Asada et al., 1999). The current investigation has detected these base changes among European/American, Afro-Hispanic and Turkish individuals, signaling that it is likely that these differences are sequence errors rather than population-specific polymorphisms. No information as to the ethnic origin of the individuals used in the generation of the L76927 *GALK1* sequence has been provided (Bergsma et al., 1996). Further analysis using a more comprehensive cross-section of population groups would rebuke or confirm definitively whether these changes are errors or represent true polymorphism.

### *3.4.3 General discussion of mutation in the GALK1 gene*

A total of twenty *GALK1* mutations associated with the galactokinase deficiency phenotype have been reported to-date (Stambolian et al., 1995; Asada et al., 1999;

Kalaydjieva et al., 1999; Kolosha et al., 2000). Except for two founder mutations, namely P28T which occurs among the Vlax Roma (Kalaydjieva et al., 1999) and Q382X found in patients of Costa Rican/European descent (Kolosha et al., 2000), the remaining aberrations are private mutations, confined to individual families. Similar to most individuals with galactokinase deficiency described previously, two patients in this analysis were also found to be compound heterozygotes, adding four novel molecular defects to the allelic heterogeneity of *GALK1* mutations (Figure 3.23).

The lack of other changes in the coding sequence of the *GALK1* gene, the nature of the detected sequence variants and their predicted effect on the protein, lead to the suggestion that these are disease-causing mutations. Including the mutations identified in this analysis, the overall distribution to-date is 14 amino acid substitutions and 10 truncating mutations, with exons 1, 2 and 7 harbouring 13 of the known mutations. The majority of missense mutations cluster in close proximity to the galactokinase signature sequence in exon 1 and the second ATP-binding domain in exon 7, highlighting the functional significance of these regions for the enzyme activity.

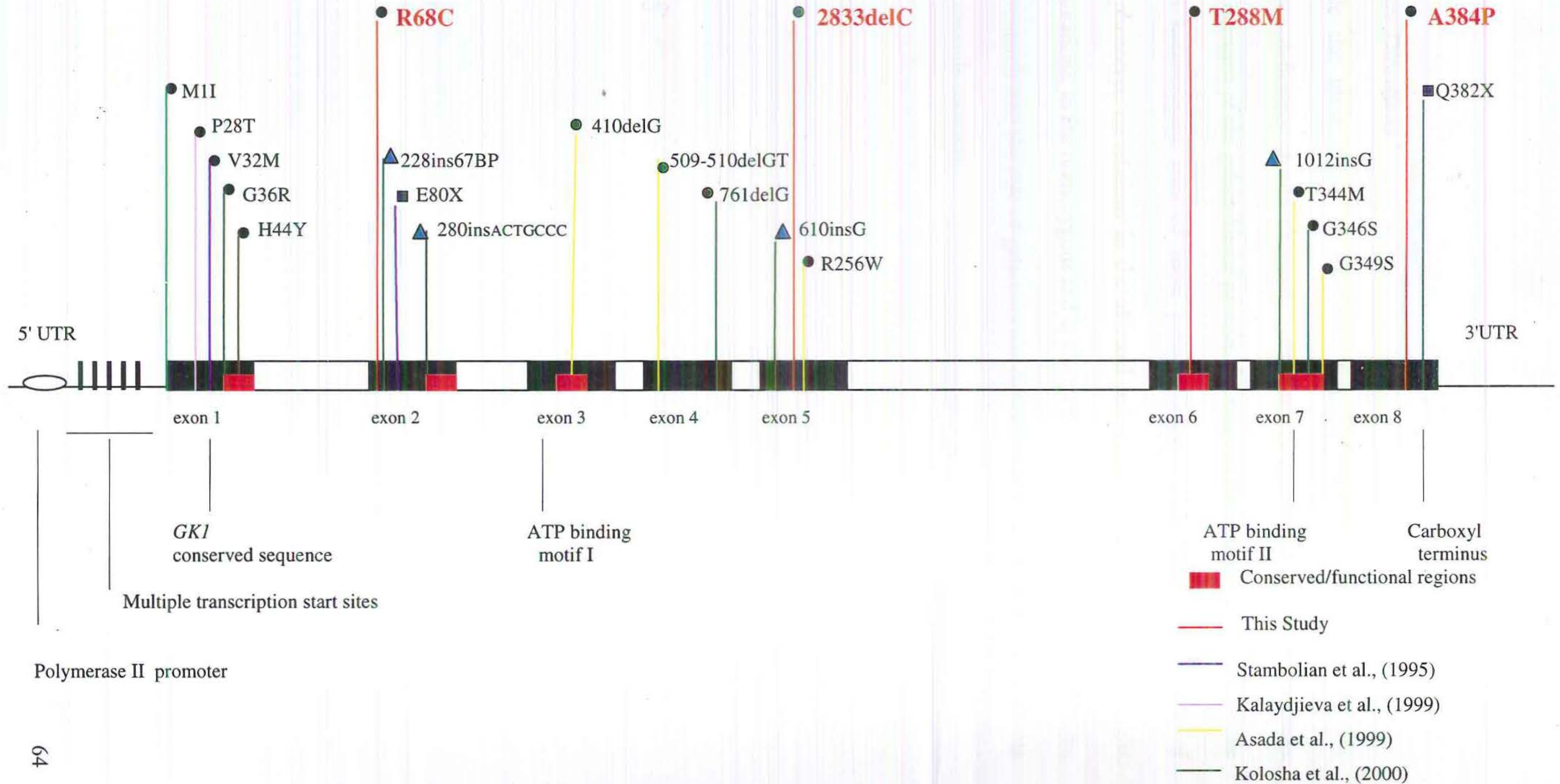



Figure 3.23 Distribution of mutations in *GALK1*.

### *3.4.4 Future Directions*

While the identification of additional *GALK1* mutations in individuals with galactokinase deficiency should ultimately contribute to understanding the functionally important domains of the galactokinase protein, the rarity of the individual defects and the compound heterozygous state of most patients, suggest that meaningful studies of genotype/phenotype correlations in the affected individuals are not feasible. It is the founder mutations, in the homozygous and in the carrier state, that will allow such studies, as well as research into the role of galactokinase deficiency as a factor predisposing to the common presenile cataracts.

# CHAPTER 4 ANALYSIS OF *GALK1* IN PATIENTS WITH THE ATYPICAL PHENOTYPE

## 4.1 INTRODUCTION

The *GALK1* gene was analyzed in two patients sharing a common ethnic background (Afro/Hispanic) and presenting with a rare, newly identified phenotype characterized by increased blood galactose, lack of production of galactose-1-phosphate, but no detectable reduction in erythrocyte galactokinase, galactose-1-phosphate uridyl transferase or UDP galactose-4-epimerase activity. Due to the role that galactokinase plays in the metabolism of galactose, it was considered that the newly identified phenotype may result from a defect in the tissue specific expression of *GALK1*, in particular a liver form of the enzyme. Because of the common ethnic background of the patients and the fact that they share this rare phenotype, it was hypothesized that a disease-causing mutation in the *GALK1* gene common to both patients, would be present in the homozygous form.

## 4.2 SUBJECTS

The study included two patients, who had been identified by newborn screening as hypergalactosemic. The biochemical phenotype of these patients is summarized in Table 4.1. The two families are unaware of any relatedness, however both are of Afro-Hispanic descent. The aims of the study were explained to the families and informed consent for participation was obtained from the parents according to the ethical guidelines of the institutions involved.

**Table 4.1**

Biochemical Findings of Study Sample

| | | | | RBC Activity | |
|---|---|---|---|---|---|
| Patient | Hypergalactosemia | Gal-1-P | Galactokinase | G-1-P uridyl transferase | UDP-epimerase |
| *A*[a] | + | -ve | Normal | Normal | Normal |
| *B*[a] | + | -ve | Normal | Normal | Normal |

[a] Dept of Pediatrics, Harvard University School of Medicine

## 4.3 METHODS

DNA was extracted from whole venous blood by the Genetic Service, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, USA.

The coding regions of the galactokinase *GALK1* gene and flanking intronic sequences were amplified according to protocols and conditions as described (Chapter 3, Table 3.2). PCR primers for the GK1 5' untranslated region (5'UTR) were designed from the published *GALK1* sequence (Genbank Accession # L76927) as follows: forward primer 5'-CTGATGACCTCTCACAGCTGC-3' and reverse 5' - CTGTCTCAAAGCAGCCATCACG-3'. Due to errors in the published GK1 sequence (Genbank Accession # L76927), amplification primers for the 3' UTR were chosen from the sequence of Homo sapiens chromosome 17 clone RP11-474I111 (AC019214): forward 5'-GCTGTGCTTGTGAGGCACC-3' and reverse 5'-CTTCCAAACTTGACCGCACCCT-3'. PCR amplification was performed in a total volume of 50 μl, containing 20ηg of genomic DNA, 1x PCR Buffer, 1.5mM $MgCl_2$, 1.25mM dNTPs, 4% DMSO, 2mM of each primer and 0.5 U of Taq. Cycling conditions included initial denaturation at 94°C for 5 minutes, followed by 35 cycles of 94°C for 30 sec, annealing at 63°C for 30 sec and extension at 72°C for 45 sec, and a final extension at 72°C for 7 minutes. All PCR products were cleaned using Qiagen purification columns.

Cycle sequencing of *GALK1* was performed using the ABI Prism Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit. The products were run on the ABI Prism 377 Automated DNA Analyzer. The Sequence Navigator (PE Biosystems) software was used for sequence analysis. Patients' sequence data were compared to unaffected controls run in parallel and to the published *GALK1* sequence. Searches of the public domain genomic databases were performed using the BLAST program

[available http://www.ncbi.nlm.nih.gov/BLAST/]. Computer modeling of predicted protein structure was done with the PHDsec (Rost & Sander, 1994) software [available http://dodo.cpmc.columbia.edu/predictprotein/submit_def.html].

## 4.4 RESULTS

### *4.4.1 Sequence analysis of the GALK1 gene*

Analysis of the entire coding sequence, flanking intronic sequences and the 5' and 3' untranslated regions of the two patients identified five nucleotide substitutions differing from the wild-type *GALK1* sequence. These changes are shown in Table 4.2.

**Table 4.2**

*GALK1* Nucleotide Differences in Atypical Patients

| | Nucleotide[a] | Location | Genotype |
|---|---|---|---|
| Patient *A* | 303A→G | 5' UTR | Heterozygous |
| | 2149T→C | Intron 2 | Homozygous |
| | 7087G→A | Exon 6 | Heterozygous |
| | 7456C→T | Intron 7 | Heterozygous |
| Patient *B* | 1556G→A | Exon 2 | Heterozygous |

[a] Nucleotide numbering according to Genbank Accession # L76927

The A→G transition at nucleotide position 303 is located in the *GALK1* 5' UTR, within the 250 bp putative eukaryotic polymerase II promoter region (Bergsma et al., 1996). Patient *A* is heterozygous for this sequence variant.

Patient *A* is also heterozygous for a C→T transition at position 7456 in intron 7. This is in close proximity to the 5' end of exon 8 of the *GALK1* gene, however it is not located within a recognized consensus branch splice site sequence or the splice acceptor site. The 2149T→C nucleotide change in intron 2 is located 34 bp upstream of the 5' end of exon 3 and 3 bp upstream of a putative eukaryotic consensus branch point sequence, YNRAY (Chiara et al., 1996; Rautmann & Breathnach, 1985). Patient *A* was homozygous

for this intronic substitution. The same sequence variant was also found in the heterozygous state in another patient, with typical galactokinase deficiency, where two disease-causing mutations (T288M and 2833delC) have been identified (Chapter 3, Section 3.4.2).

The only change found in the coding sequence of *GALK1* in patient *A*, was the G→A transition at nucleotide position 7087 (Figure 4.1) which results in the substitution of glutamine for arginine at codon 286 (R286Q) in exon 6 (Figure 4.2). Although arginine and glutamine differ in terms of side-chain charge, both are hydrophilic amino acids (Figure 4.3), suggesting that the substitution is unlikely to affect secondary protein structure. Indeed, secondary-structural analysis using the PHDsec algorithm (Rost & Sander, 1994) predicts structural identity and no interference with the helix formed by this fingerprint motif (Figure 4.4).

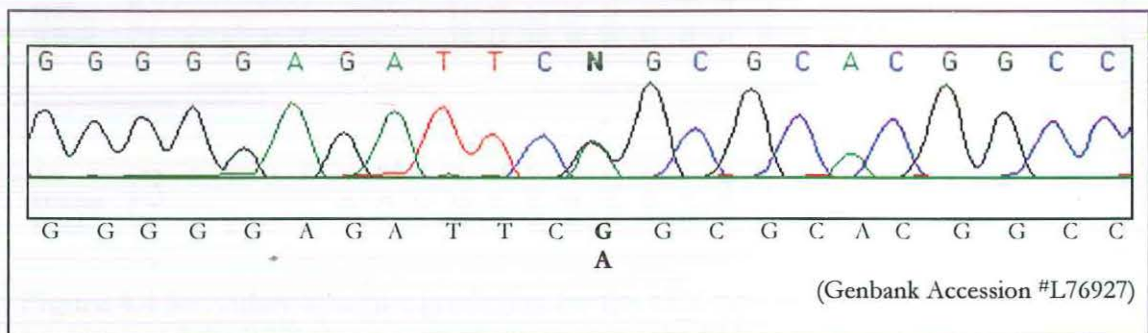



**Figure 4.1** 7087G→A transition causing a R286Q substitution in exon six.

| **Wildtype** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Codon** | 283 | 284 | 285 | **286** | 287 | 288 | 289 | 290 |
| **Amino acid** | G | E | I | **R** | R | T | A | Q |
| **Nucleotide** | GGG | GAG | ATT | **CGG** | CGC | ACG | GCC | CAG |
| **R286Q** | | | | | | | | |
| **Codon** | 283 | 284 | 285 | **286** | 287 | 288 | 289 | 290 |
| **Amino acid** | G | E | I | **Q** | R | T | A | Q |
| **Nucleotide** | GGG | GAG | ATT | **CAG** | CGC | ACG | GCC | CAG |

**Figure 4.2** Residue change at codon 286 caused by the 7087G→A transition. Wildtype arginine is replaced with glutamine.

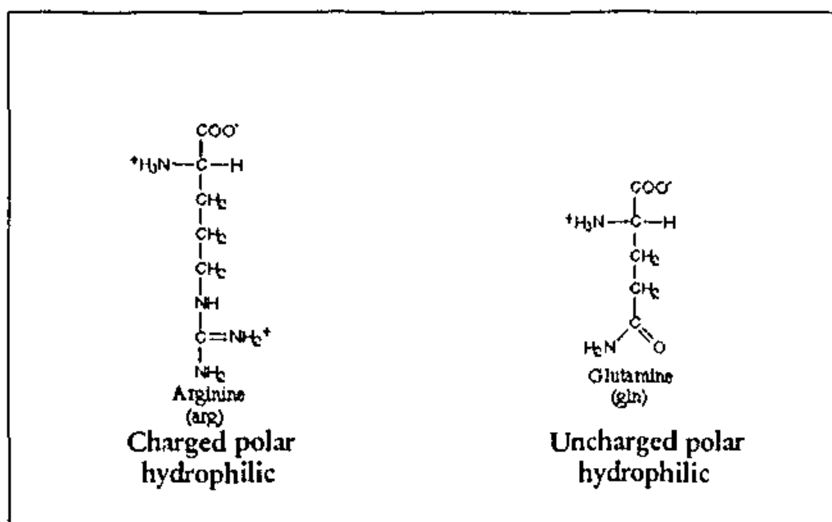



**Figure 4.3** R286Q – molecular comparison between arginine and glutamine residues

**Galactokinase exon 6 wild-type**

| | | | | | | | | | | 280 | | | | | | | | | | 290 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | S | K | E | G | F | R | R | A | R | H | V | V | G | E | I | R | R | T | A | Q | A | A | A |
| PHDsec | | | | | | | | | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |
| SUBsec | L | L | L | L | L | . | . | . | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |

**Galactokinase R286Q**

| | | | | | | | | | | 280 | | | | | | | | | | 290 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AA | S | K | E | G | F | R | R | A | R | H | V | V | G | E | I | Q | R | T | A | Q | A | A | A |
| PHDsec | | | | | | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |
| SUBsec | L | L | L | L | L | . | . | . | H | H | H | H | H | H | H | H | H | H | H | H | H | H | H |

**Figure 4.4** Secondary-structure prediction for the wild-type exon six galactokinase sequence and the R286Q missense mutation using the PHDsec algorithm. Q286 in boldtype face. E = extended sheet (beta sheet); H = helix; L = loop. The prediction is meaningful for all residues with an expected average correlation > .69.

This R286Q missense mutation occurs within the fourth, evolutionary conserved galactokinase fingerprint sequence (RRARHVVGEIRRTAQ). However, a BLAST homology search comparing published galactokinase sequences of diverse species revealed that this particular position is highly variable: while an arginine residue occurs in the murine and human galactokinase, glutamine is the common amino acid at position 286, found in 6 out of 20 galactokinases which share the fingerprint motif (Figure 4.5).

| Galactokinase motif four | | | | | | | | | | | | | | | Organism |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| K | R | A | R | H | A | V | S | E | N | Q | R | T | L | K | *Streptomyces lividans* |
| K | R | A | R | H | V | V | T | E | N | Q | R | V | L | D | *Haemophilus influenzae* |
| K | R | V | R | H | I | L | T | E | N | A | R | T | V | E | *Escherichia coli* |
| K | R | V | R | H | I | L | T | E | N | A | R | T | V | E | *Bacteriophage lambda* |
| K | R | V | R | H | V | L | S | E | N | A | R | T | V | E | *Salmonella typhimurium* |
| K | R | A | K | H | I | I | S | E | N | Q | R | V | L | E | *Actinobacillus pleuropneumoniae* |
| | R | A | K | H | V | Y | S | E | S | L | R | V | L | K | *Saccharomyces cerevisiae* |
| K | R | A | R | H | A | V | S | E | N | Q | R | T | L | R | *Lactobacillus helveticus* |
| | R | A | K | H | V | Y | S | E | A | L | R | V | L | K | *Kluyveromyces lactis* |
| | R | A | K | H | V | Y | Q | E | S | L | R | V | L | E | *Candida albicans* |
| E | R | A | L | H | V | Y | R | E | S | L | R | V | L | K | *Candida parapsilosis* |
| R | R | A | K | H | A | V | Y | E | N | H | R | A | I | K | *Bacillus subtilis* |
| K | R | A | R | H | V | V | L | E | N | Q | R | T | L | Q | *Streptococcus thermophilus* |
| | R | A | K | H | V | Y | S | E | S | L | R | V | L | K | *Saccharomyces cerevisiae (GAL3)* |
| **R** | **R** | **A** | **R** | **H** | **V** | **V** | **G** | **E** | **I** | **R** | **R** | **T** | **A** | **Q** | ***Homo sapiens*** |
| R | R | A | R | H | V | V | S | E | I | R | R | T | A | Q | *Mus musculus* |
| K | R | A | R | H | A | V | F | E | N | Q | R | T | L | K | *Lactobacillus casei* |
| K | R | A | Q | H | V | L | E | E | N | E | R | V | L | K | *Thermotoga neapolitana* |
| K | F | F | G | Y | I | V | R | E | N | A | R | V | L | E | *Pyrococcus horikoshii* |
| R | L | V | R | H | V | V | T | E | D | E | R | V | E | R | *Streptomyces lividans* |

**Figure 4.5** Galactokinase fingerprint motif four. Bold highlight showing amino acid variation analogous to residue 286 of the human galactokinase protein at which the R286Q substitution is found. Glutamine (Q) occurs in six of the galactokinases that share the fourth galactokinase fingerprint motif. Motif alignment by SPRINT (available http://www.bioinf.man.ac.uk/dbbrowser/sprint/).

The only difference found between *GALK1* in patient *B* and the wild-type sequence was a G→A transition at nucleotide 1556 in exon 2. This change results in a silent mutation at codon 63 (L63L).

Comparisons of the *GALK1* sequence in both patients and unaffected controls with the published wild-type sequence (Genbank Accession # L76927) revealed a number of nucleotide discrepancies previously reported by Asada et al. (1999) that were shared in all samples analyzed. These changes include a 415C→T and 418C→T in the 5'UTR, a 694C→T in intron 1, as well as a 2130C→G and 2132A→C in intron 2 (Chapter 3, Table 3.4).

## 4.5 DISCUSSION

The unique characteristics of this condition, namely a markedly increased blood galactose concentration, lack of galactose-1-phosphate and normal RBC galactokinase activity, pointed to a rare disorder of the first step of galactose metabolism. The metabolic evidence of galactokinase deficiency, together with the normal activity of this enzyme in peripheral blood cells, suggested a defect in tissue-specific galactokinase expression. The rarity of the disorder and the common ethnic background of the affected individuals suggested a common molecular defect. This study was therefore based on the hypothesis that the two patients will be homozygous for the same disease-causing founder mutation in the galactokinase gene *GALK1*.

### *4.5.1 Sequence analysis*

Approximately 3.1kb of the *GALK1* sequence were investigated, including all coding regions, at least 100 bp of the flanking intronic sequences and the 5′ and 3′ UTR. This analysis failed to reveal any sequence variants that are plausible disease-causing mutations. Within the coding sequence, the 7087G→A transition identified in patient 1 resulted in the non-drastic glutamine for arginine substitution at a non-conserved position in exon 6. In fact, this substitution reverted to the glutamine residue which is more common in the galactokinase protein of different species. These findings, together with the lack of effect on secondary structure predictions, lead to the conclusion that R286Q is a polymorphism that is unlikely to cause the disease. The single exonic change found in patient 2, was a silent mutation at codon 63, i.e. again a neutral polymorphism.

Patient 1 also displayed three additional nucleotide substitutions in the intronic sequence of the *GALK1* gene. The 7456C→T transition in intron 7 does not affect a known consensus splice motif, nor does it seem to activate a cryptic splice site and thus appears to be a polymorphic variant. The other two nucleotide substitutions, at positions 303 in the 5'UTR and 2149 in intron 2, could potentially affect the level of expression of

*GALK1* or the splicing of the primary transcript. The 303A→G change occurs in the putative polymerase II promoter sequence (Bergsma et al., 1996), whereas 2149T→C is located in close proximity to a putative eukaryotic branch point motif (Chiara et al., 1996). While expression studies are needed to investigate the hypothetical effect of these changes on *GALK1* expression, their likely role as neutral polymorphisms is supported by the fact that the same intronic substitutions have also been found to occur in a patient with typical galactokinase in whom both disease-causing mutations have been identified (Chapter 3, Section 3.4.2).

The most significant finding of this analysis was the lack of similarity between the two patients. Both were heterozygous for nucleotide variation in *GALK1* and none of the nucleotide variants were shared between the two patients. The findings indicate that these two individuals, presenting with the same rare phenotype and originating from the same ethnic background, carry different copies of the known galactokinase gene. The results provide strong evidence against a role of *GALK1* in this condition.

#### *4.5.2 Other aberrations to galactose metabolism*

Although it is likely that this investigation excluded *GALK1* as the responsible gene, the possibility remains that both patients share an aberration in another gene involved in galactose metabolism. Supporting this possibility is evidence pointing towards the existence of multiple forms of the enzyme including both age-related and tissue differences in galactokinase expression levels (Ng, Donnel, & Bergren, 1965; Shin-Buerhring, Beier, Tan, Osang, & Schaub, 1977) and variation in kinetic and functional properties (Cuatrecasus & Segal, 1965; Mathai & Beutler, 1967; Magnani et al., 1982b). Perhaps the most compelling evidence for multiple forms of the enzyme comes from clinical data of an adult who is homozygous for a known deficiency-causing mutation yet has intermediate erythrocyte galactokinase activity analogous to a heterozygote carrier (Kalaydjieva et al., 1999). Whether

this discrepancy and the reported differences in the properties of galactokinase are due to the presence of two different genes or reflect post-translational modification of the same protein remains to be characterized. Identifying the mechanism(s) responsible for these observations may aid in determining the cause of the persistently high levels of blood galactose that is associated with this new phenotype.

Alternatively, future directions for determining the genetic basis for this new phenotype include an analysis of other genes or processes that may play a role in galactose utilization. One likely candidate is the SLC2A2 gene (Genbank Accession # 4557850) which encodes a hepatic facilitative hexose transporter (GLUT-2). The cellular uptake of galactose is made possible by the GLUT-2 transport mechanism capable of carrying both glucose and galactose across the hepatocyte membrane (Burchell, 1994). The fact that such receptors display different degrees of specificity for the two hexoses, suggests either inherent differences in the binding sites of the transport protein, or may reflect variation in the efficiency of binding at the same site. Both pharmacological (Ibu & Short, 1986) and physiological (Fiete, Brownell, & Baenziger, 1983) evidence indicates that galactose uptake at these receptors can be modulated, possibly signaling inherent differences in the substrate binding domain of the transporter (Wu, Fritz, & Powers, 1998). Mutations to a galactose-binding site would limit liver uptake, leading to an accumulation of galactose in the blood. If the sugar was present in high concentrations, the erythrocyte capacity to successfully phosphorylate galactose may become saturated, leading to further accumulation of galactose and subsequent initiation of the galactitol pathway. The production of a pathological level of galactitol would result due to the wide distribution of the galactose reducing enzyme, aldose reductase, found in a number of tissue sources including the lens of the eye (Segal & Berry, 1995). A defect to the hepatic galactose transport mechanism may possibly lead to an increase in blood galactose analogous to the biochemical

description of the patients in the current study. Both possibilities should be considered for further analysis in determining the molecular aberration responsible for this phenotype.

Other pathologies affecting galactose metabolism should also be considered. Two reports (Jakobs et al., 1988; Budde, Gusek-Schneider, Junemann, Jansen, & Shin, 1999) detail another metabolic disorder affecting the galactose pathway called Shin-Jakob's disease. In this condition, subjects present with increased blood galactitol levels and cataracts without any identifiable defect to any of the three enzymes in the Leloir pathway. Jakobs et al. (1988) suggest these findings may be explained by three possibilities including an aberration to a fourth unknown galactose enzyme, a variant of a known enzyme or an increase in activity of the enzyme, aldose reductase. The molecular defect underlying these observations has yet to be elucidated.

Segal (1998) suggests that the hypergalactosemia seen in some galactosemic and galactokinase-deficient sufferers, despite a restriction of galactose from the diet, may be attributed to endogenous sources of galactose. This source includes the recycling of UDP-galactose from cell membranes. If the enzymatic mechanism for this salvage process was over-expressed, it would cause a rise in blood galactose levels that may overload the functional capacity of galactokinase.

### *4.5.3 Conclusions*

Although this analysis has excluded mutations in the known promoter region within the 5'UTR of the *GALK1* gene as well as the entire coding regions, flanking intron regions and the 3'UTR, functional expression studies and characterization of mRNA transcripts from both patients should be carried out in the future. Such investigations may identify differences in expression of galactokinase shared by these patients. These differences may possibly reflect the use of different splicing mechanisms or alternatively reflect the use of tissue-specific *GALK1* promoters not identified in our current knowledge of the gene. Testing for mutations in the SLC2A2 gene, which encodes the hepatic glucose/galactose

transporter, or linkage analysis are other strategies that should be used in future studies to determine the cause of this phenotype.

# CHAPTER 5 ANALYSIS OF THE ORIGINS, DISTRIBUTION AND FREQUENCY OF A COMMON *GALK1* MUTATION CAUSING GALACTOKINASE DEFICIENCY

## 5.1 INTRODUCTION

The P28T mutation has previously been identified as the common mutation causing galactokinase deficiency in six Romani families from Bulgaria (Kalaydjieva et al., 1999). These families originate from three socially distinct but genetically related groups; Kalderas, Rudari and Thracian Tinkers, whose migratory patterns can be traced back to the northern part of the Balkans. Indeed the Vlax Roma, as they are collectively known, show a high number of affected births, with incidence in the range of 1:1600 to 1:2500 births and based on the screening of 130 individuals, a predicted carrier rate of the P28T mutation, in the vicinity of 4%-5%.

The fact that the current investigation (Chapter 3, Section 3.3.3) has identified a number of patients from Gypsy groups geographically dispersed throughout Europe and socially and culturally distinct from the Vlax Roma, suggests that the distribution of the P28T mutation is more widespread than first thought. Establishing that the P28T mutation in these groups is shared common by descent and is not a recurrence of the same mutation at the same codon, will provide information into the genetic origins and relatedness between such groups and will allow a study of the history of the mutation and the historical demography of the affected groups. At the same time, the identification of populations at high risk of infantile and perhaps presenile blindness means that screening strategies and early intervention can be utilized to prevent the progression to sight impairment caused by galactokinase deficiency in these groups.

Two strategies need to be employed to achieve this. This first is establishing the common origin of the P28T mutation. Previous haplotyping analysis in Bulgarian Romani families has identified a conserved haplotype surrounding the P28T mutation (Kalaydjieva

et al., 1999). The disease allele has been shown to be in strong linkage disequilibrium with allele 2 of the marker D17S1839, located at a small physical distance to *GALK1*. Identifying similar conserved regions of homozygosity will point towards a common origin of the P28T mutation. Evidence of historical recombinations in the diverse Gypsy groups will allow further studies of the history of the mutation.

Establishing carrier rates by screening unrelated individuals for the P28T mutation will allow future prevention programs for infantile blindness such as newborn screening and dietary intervention to be designed, and their cost-efficiency estimated. Furthermore it will allow future studies investigating the link to presenile cataracts to be conducted, based on the large numbers of carriers of the same mutation where other factors, eg. nutrition, can be taken into account.

## 5.2 SUBJECTS

### *5.2.1 Haplotype analysis determining the origin of the P28T mutation*

The study sample included 13 subjects (Table 5.1): six have been identified as homozygous for the P28T mutation (Chapter 3, Section 3.3.3) and seven unaffected family members. These patients derive from geographically dispersed regions throughout Europe including, Turkey, Spain and Hungary and all but the Turkish patient report Gypsy origins. All were small nuclear families, except for the consanguineous Spanish Gypsy kindred shown in Figure 5.1. In addition, control P28T homozygous, heterozygous and unaffected individuals derived from Bulgarian families were run in parallel for each of the markers.

Table 5.1

Individuals with the P29T Mutation

| Individual | P28T Genotype | Origin |
|---|---|---|
| 1[a] | Homozygous | Turkish |
| 2[a] | Heterozygous | Turkish |
| 3[b] | Heterozygous | Spanish Gypsy |
| 4[b] | Heterozygous | Spanish Gypsy |
| 5[b] | Homozygous | Spanish Gypsy |
| 6[b] | Heterozygous | Spanish Gypsy |
| 7[b] | Homozygous | Spanish Gypsy |
| 8[b] | Heterozygous | Hungarian Gypsy |
| 9[b] | Homozygous | Hungarian Gypsy |
| 10[b] | Homozygous | Hungarian Gypsy |
| 11[b] | Heterozygous | Hungarian Gypsy |
| 12[b] | Heterozygous | Hungarian Gypsy |
| 13[b] | Homozygous | Hungarian Gypsy |

[a] University Children's Hospital, Zurich, Switzerland
[b] Institut Bioquimica Clinica, Barcelona, Spain
[c] Dept. of Paediatrics, University Medical School, Szeged, Hungary
⊏ denotes kindred

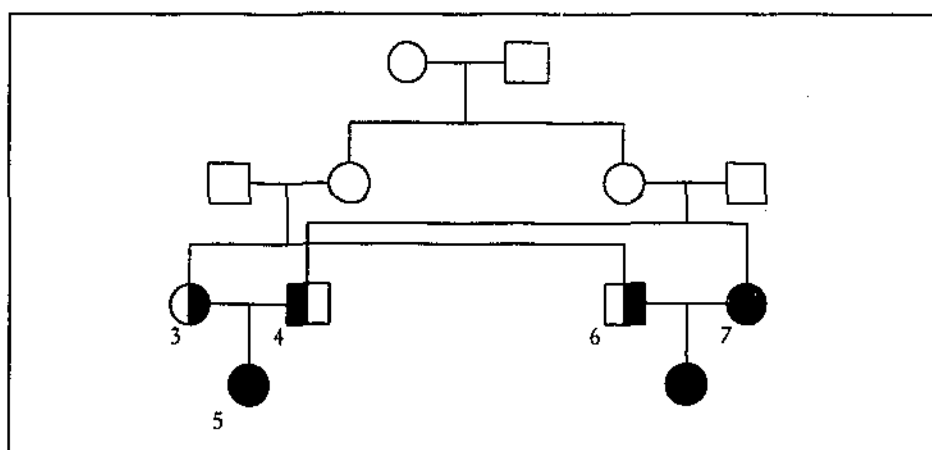



**Figure 5.1** Pedigree of consanguineous Spanish Gypsy family. Individuals whose DNA was available for analysis are numbered.

### *5.2.2 Population Screening for the P28T mutation*

The population distribution and frequency of the P28T mutation was analyzed in a sample of 227 individuals from socially and geographically diverse Gypsy groups. 91 of these belong to the Vlax Roma groups: 45 Rudari and 46 Lom; 36 are derived from the

Xoroxane Gypsies; and 39 belong to other Gypsy groups from Bulgaria. In addition, 55 Spanish Gypsies and 6 Lithuanian Gypsies were also screened for the P28T mutation. These data were combined with the results of previous screening carried out at the Centre for Human Genetics, Edith Cowan University, of 144 individuals from Rudari, Lom, Kalderas and Xoroxane Gypsy groups (Table 5.2).

**Table 5.2**

Individuals Screened for the P28T Mutation

| Population | Previously screened n | This Study n | Total n |
|---|---|---|---|
| **Vlax groups** | 135 | 91 | **226** |
| Rudari | 79 | 45 | 124 |
| Lom | 13 | 46 | 59 |
| Kalderas | 43 | - | 43 |
| **Xoroxone** | 9 | 36 | **45** |
| **Others** | - | 39 | **39** |
| **Spanish** | - | 55 | **55** |
| **Lithuanian** | - | 6 | **6** |

## 5.3 METHODS

### *5.3.1 Haplotype analysis*

#### *5.3.1.1 Microsatellite markers*

Six microsatellite markers (Table 5.4) spanning a 9cM region surrounding the *GALK1* gene were chosen from a selection of markers used in a previous study to map the disease gene (Kalaydjieva et al., 1999). PCR amplification protocols and conditions are detailed in Table 5.4. Each reaction consisted of 20ηg of DNA, 0.2mM each primer, 0.25mM each dNTP, optimal $MgCl_2$ concentration (Table 5.4), 1 x PCR buffer, 0.2U of Taq polymerase and $dH_2O$ to a final volume of 10µl. Radioactive labeling was done through incorporation of 0.1µl of $\alpha^{32}P$-dCTP (GeneWorks cat #ADC-32U/L) in the product during PCR amplification.

**Table 5.3**

PCR Conditions and Protocols for Microsatellite Markers

| Marker | Size | PCR Conditions | Additives |
|---|---|---|---|
| D17S1797 | 183-193bp | TD 63°C →55°C | 1.5mM $MgCl_2$ |
| D17S1602 | 197-207bp | 58°C for 35 cycles | 1.5mM $MgCl_2$ |
| D17S1864 | 183-209bp | TD62°C → 52°C | 1.5mM $MgCl_2$ |
| D17S929 | 217-229bp | TD 63°C →55°C | 1.5mM $MgCl_2$ |
| D17S1839 | 237-245bp | TD 63°C →55°C | 1.5mM $MgCl_2$ |
| D17S1817 | 183-209bp | 62°C for 35 cycles | 1.0mM $MgCl_2$ |

#### *5.3.1.2 Electrophoresis on polyacrylamide gel*

PCR products were electrophoresed on 6% polyacrylamide gels, prepared by adding 420.4g of (7M) Urea and 150mL of 40% 19:1 acrylamide/bisacrylamide to 500ml of $dH_2O$. This solution was deionized for 30 minutes using mixed bead resin. The solution was then filtered and 100mL of 10x TBE buffer was added and

the final volume made up to 1L with $dH_2O$. Prior to pouring, 130µl of 12% APS (ammonium persulphate) and 130µl of TEMED (N',N',N',N'-tetramethylethylenediamine) were added to 60mL of polyacrylamide solution, mixed, then injected between two cleaned and clamped glass plates. Well formers were inserted and the gels were left to set for 2 hours before fitting to the Hoefer Poker Face II SE1600 Sequencing Series apparatus. Upper and lower chambers were filled with 1 x TBE buffer and the gels were pre-warmed for 30 minutes at 1400V. The well former was removed and a loading comb inserted before each well space was rinsed using a transfer pipette. 10µl of formamide loading buffer (98% formamide, 10mM NaOH, 0.1% bromophenol blue & 0.1% xylene cyanol) was added to each PCR reaction and denatured for 5 minutes at 94 degrees C. Products were placed on ice and 2µl of each sample was loaded into single lanes of the gel. The gel was electrophoresed at 1400V for 2 hours. Plates were removed from the apparatus and the outer glass plate was lifted off using a plastic wedge. Approximately 200mL of fixer solution (10% methanol, 10% glacial acetic acid in $dH_2O$) was poured over the gel 3 times over a duration of 10 minutes. The fixer was then drained and a piece of chromatography paper was laid over the gel. The paper and gel were lifted of the glass plate, covered with plastic wrap, and allowed to dry on a Savant Slab Gel Dryer at 60 degrees C for 1 hour. Dried gels were exposed on Cronex® Medical X-Ray film (Sterling Diagnostic Imaging) in film cassettes for approximately 14hrs at room temperature. The films were developed on a AGFA® CURIX 60 developing machine.

*5.3.1.3 Haplotype analysis*

Allele scoring was done manually, using a lightbox. Alleles were numbered sequentially from top to bottom. To ensure compatibility of allele assignment with

the previous study, several individuals from the original Bulgarian Gypsy families (Kalaydjieva et al., 1999) were genotyped in parallel. Haplotypes were constructed manually. In case parental information was missing (One of the Hungarian Gypsy subjects), phasing was done by assignment of the most common haplotype to one of the chromosomes.

*5.3.2 Population screening for the P28T mutation*

The 563C→ A transversion (P28T) abolishes an AvaI restriction site and thus a PCR-based restriction assay can be used for screening purposes. A 450bp DNA fragment spanning the region between nucleotides 358 and 808 in the GALK1 sequence (Genbank Accession # L76927) was amplified from genomic DNA and FTA cards by use of primers 5′-GGGGCGTCCGGGCGCGGGGC-3′ and 5′-GACAGGCTGTTCCCCACGT-3′. PCR volumes were made up to 20μl and consisted of 20ng of starting DNA for genomic samples, or a 1mm x 1mm section of FTA card. In addition, the PCR mixture contained 1.25mM of $MgCl_2$, 2mM of each dNTPS, 1 x PCR buffer, 4% DMSO, 0.5U of Taq polymerase and $dH_2O$. PCR conditions included initial denaturation at 94 degrees C for 5 minutes, followed by 35 cycles of 94 degrees C for 30 seconds, 63 degrees C for 30 seconds and 72 degrees C for 40 seconds. This was followed by an additional extension step at 72 degrees C for 7 minutes. PCR products were sodium/ethanol precipitated as described in Section 3.3.4 and resuspended in 5μl of $dH_2O$. Products were cut with 0.3μl of the restriction endonuclease, AvaI [10 000U/ml] (New England Biolabs cat# RO152S), 1.5μl of NE buffer, 1.5μl of BSA and $dH_2O$ in a final volume of 12.5μl. Products were incubated at 37 degrees C for 3 hours before electrophoresis upon 4% agarose/EtBr gel at 60V for 1 hour. Digests were visualized on a Hoefer

Mighty Bright™ UV Illuminator and the image was recorded using a Kodak® DC120 digital camera and Kodak® Digital Science™ software.

Carrier frequency was determined by counting the number of affected alleles detected and compared with carrier frequencies calculated from a similar screening strategy carried out previously at the Centre For Human Genetics, E.C.U.

## 5.4 Results

### *5.4.1 Haplotype analysis*

Haplotype analysis revealed complete homozygosity for marker D17S1839 for mutant chromosomes across Europe (Figure 5.2 & 5.3). Allele 2 of this marker was present in all chromosomes with the P28T mutation and has previously been shown to be in strong linkage disequilibrium with the P28T mutation among six Bulgarian Gypsy families (Kalaydjieva et al., 1999). The fact that the Hungarian Gypsy patients, Spanish Gypsy patients and the Turkish patient are also homozygous for this allele in conjunction with P28T, points towards a common founder mutation.

A highly conserved haplotype common to the haplotype seen in the Bulgarian Gypsy families, was carried by 5 chromosomes in 4 Hungarian Gypsies and 1 Spanish Gypsy (Figure 5.2). The alleles seen in the rest provides evidence of different historical recombination which would be expected, given that the sample of patients analyzed originate from dispersed groups.

Allele 2 of the D17S1839 marker was not found in any of the Bulgarian, Hungarian and Turkish control individuals, not carrying the P28T mutation. This provides further evidence that this particular allele is in strong linkage disequilibrium with the disease mutation.

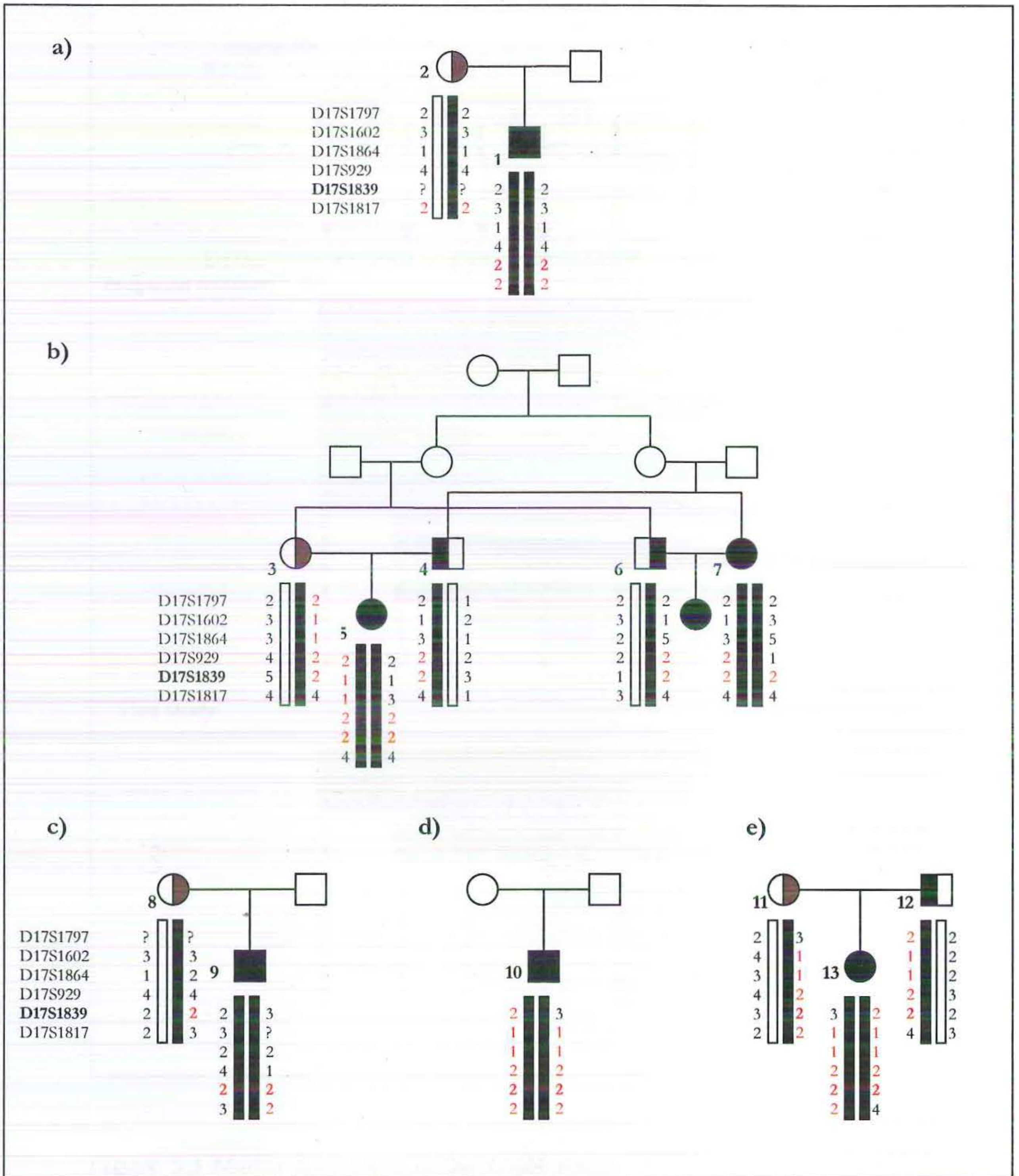



**Figure 5.2** Pedigrees and marker haplotypes for patients with P28T. a) Turkish. b) Consanguineous Spanish Gypsy kindred. c-e) Hungarian Gypsy. Red highlight indicates conserved haplotype. Bold red highlight indicates conserved allele 2 of marker D17S1839.

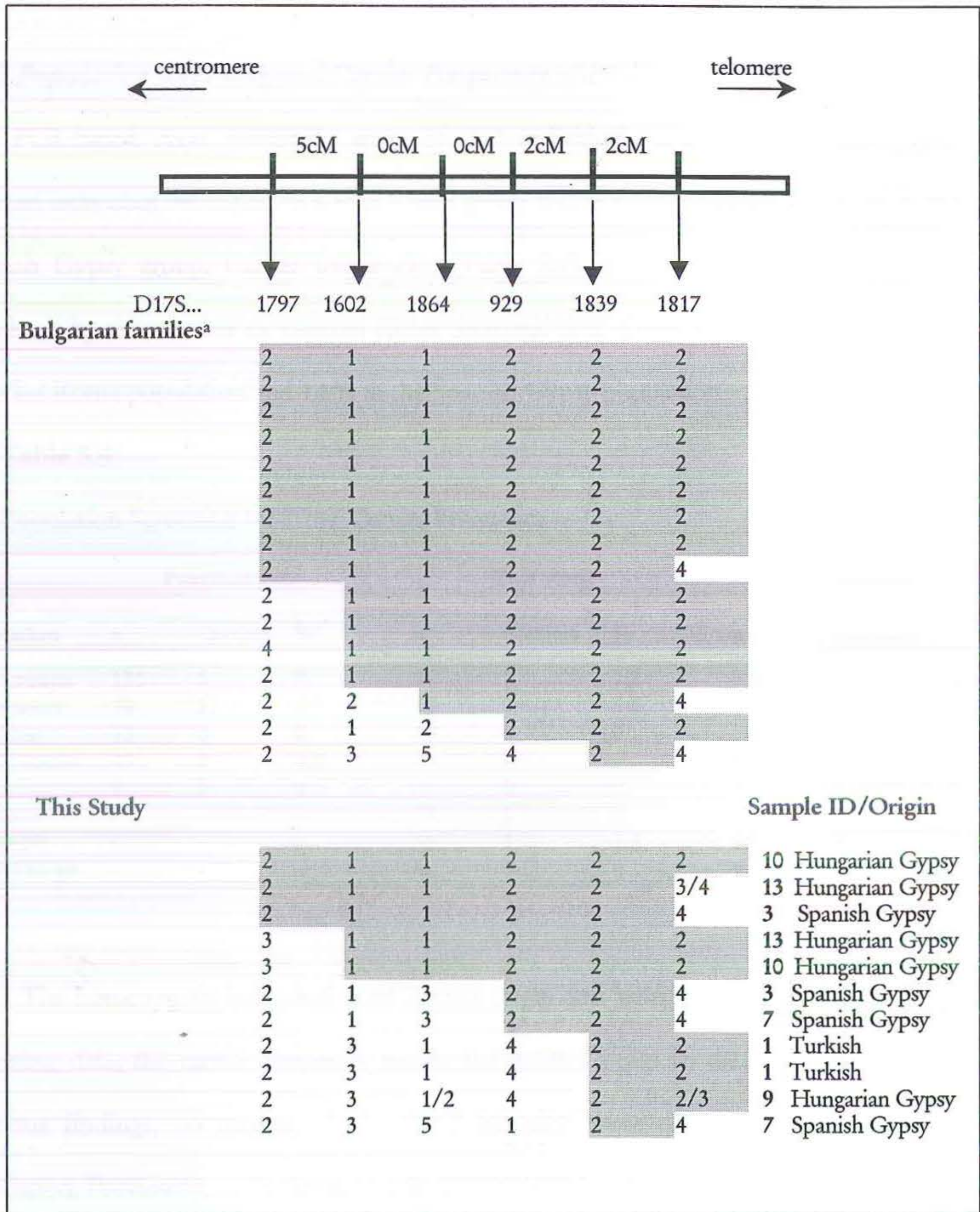



**Figure 5.3** Marker haplotypes in the 17q24 region of disease chromosomes from Bulgarian Gypsy groups [a](Kalaydjieva et al., 1999) (upper map) and current study sample consisting of Spanish, Hungarian and Turkish Gypsies (lower map). Allele 2 of marker D17S1839 is shared in all chromosomes with the mutation.

### *5.4.2 Population screening and Carrier Frequency of P28T*

The PCR-based AvaI restriction assay of 227 individuals detected one homozygous affected individual belonging to a Vlax Roma group and one heterozygous carrier from the Spanish Gypsy group. Carrier frequencies (Table 5.4) in both these populations were calculated by the number of affected alleles detected, thus giving a carrier rate of 2.2% in the Vlax Roma population and 1.8% in the Spanish Gypsy population.

**Table 5.4**

Population Screening for P28T Carrier Frequency

| | Previous data | | | This study | | | Overall | | |
|---|---|---|---|---|---|---|---|---|---|
| **Gypsy Population** | n | Carriers | % | n | Carriers | % | **n** | **Carriers** | **%** |
| **Vlax groups** | 135 | 4 | 3 | 91 | 2 | 2.2 | **226** | 6 | **2.7** |
| Rudari | 79 | 2 | 2.5 | 45 | 2 | 4.4 | 124 | 4 | 3.2 |
| Lom | 13 | 0 | 0 | 46 | 0 | 0 | 59 | 0 | 0 |
| Kalderas | 43 | 2 | 4.6 | - | - | - | 43 | 2 | 4.6 |
| | 9 | 0 | 0 | 36 | 0 | 0 | **45** | **0** | **0** |
| **Others** | - | - | - | 39 | 0 | 0 | **39** | **0** | **0** |
| **Spanish** | - | - | - | 55 | 1 | 1.8 | **55** | **1** | **1.8** |
| **Lithuanian** | - | - | - | 6 | 0 | 0 | **6** | **0** | **0** |

The homozygous individual is of Rudari origin and when combined with previous screening data, the carrier frequency within the Rudari group to date is 3.2%. As with previous findings, no carriers of the P28T mutation were detected among the Lom population. Previously, carrier frequency in the Kalderas has been found at 4.6%. Although individuals from this group were not available for screening in the current investigation, it seems that both the Rudari and Kalderas groups are major contributors to the overall carrier frequency of the P28T mutation seen in the Vlax Roma, which from the current analysis stands at 2.7%. No carriers were detected in the Xoroxone group nor within the small number of Lithuanian Gypsies analyzed.

## 5.5 DISCUSSION

### *5.5.1 Haplotypes*

It is evident that the P28T mutation detected in Bulgarian, Spanish and Hungarian Gypsy patients as well as the Turkish patient is the founder mutation reported in six Bulgarian families (Kalaydjieva et al., 1999). The identification of a shared region of homozygosity surrounding the D17S1839 marker in all patients with the P28T mutation signifies a common origin. The fact that allele 2 of the D17S1839 marker was not seen in individuals without the P28T mutation also lends further support that the disease allele is in strong linkage disequilibrium with this allele.

It is interesting to note that the haplotype analysis has revealed that the P28T mutation carried by the Turkish patient, is in fact the same founder mutation common to Bulgarian, Hungarian and Spanish Gypsy groups in the study. It is likely that the ethnicity of this patient and mother is obscured perhaps to avoid the racial and social persecution and prejudices historically aimed at the Gypsy population. The high degree of polymorphisms in the haplotype pattern evident in this patient, as compared to other patients, may reflect historical recombinations resulting from the divergence of Turkish groups away from a common Gypsy origin or alternatively may indicate admixture.

### *5.5.2 Population screening and Carrier Frequency of P28T*

The detection of a heterozygote carrier in the population screening of the Spanish Gypsy group, combined with the findings of the common founder mutation in patients of Spanish and Hungarian Gypsy descent, highlights the widespread distribution of the P28T allele. This information is not only important for the development of intervention strategies targeting groups at risk of infantile and presenile blindness but also paves the way for anthropological studies perhaps investigating the historical origins and diversification of the many Gypsy groups.

### *5.5.3 Conclusions and future directions*

Both the detection of a heterozygote carrier through screening of the Spanish Gypsy population and the haplotype analysis of a Spanish Gypsy kindred suggest that the incidence of the P28T mutation is widespread throughout Europe. Additional data need to be collected for the Spanish Gypsy population, before newborn screening is initiated. For the time being, galactokinase deficiency should be considered in any patient of Gypsy ethnicity with infantile cataracts and the P28T mutation provides an easy test. The mutation test can be incorporated in the existing newborn screening in Hungary and Bulgaria to replace the current lengthy procedure of diagnosing galactokinase deficiency by metabolic and enzyme assays. Finally, large numbers of adult carriers can be identified by either cascade testing in the families of affected children or by random population screening. These subjects can then be studied further to clarify the proposed link between partial galactokinase deficiency and presenile cataracts.

# CHAPTER 6 SUMMARY, CONCLUSIONS AND FUTURE DIRECTIONS

Conclusive information regarding the molecular basis of galactokinase deficiency and galactose metabolism is far from complete. The mechanisms underlying the discrepancies that exist between galactokinase activity and properties in relation to factors such as age, tissue and allelic heterogeneity, are still unclear. Atypical clinical observations further confound existing ideas about the function and role of galactokinase in galactose metabolism. This is evident from the clinical descriptions provided in Chapter 4 of patients with a rare galactokinase deficiency phenotype, characterized by increased blood galactose levels, yet normal enzyme activity in RBCs, as well as the observations of an adult with intermediate galactokinase activity despite being homozygous for a known deficiency-causing mutation (Kalaydjieva et al., 1999). Perhaps these observations signal the existence of multiple forms of the enzyme or even multiple mechanisms and/or pathways of galactose metabolism. Whether these are the result of an enzyme encoded by separate genes or reflect differences in post-translational modifications and/or alternative regulation of expression of the same gene product, remain to be clarified. Future steps which need to be taken to address these possibilities, include extensive mRNA comparisons, enzymological characterization, or haplotype analysis. The outcomes from these investigations may not only complete our understanding of galactose metabolism but also benefit the treatment of patients. For instance, recognizing heterogeneity in expression or activity as a function of age would have implications as to the extent and duration of dietary intervention in affected individuals.

The current investigation has added 4 novel molecular defects: R68C, T288M, A384P and 2833delC to the allelic heterogeneity of *GALK1* mutations. Despite the lack of expression analysis in this study, it is likely these are disease-causing mutations due to the nature of the mutations, the fact that they occur in conserved residues and did not occur

among 100 normal chromosomes. While the identification of additional *GALK1* mutations in individuals with galactokinase deficiency should ultimately contribute to understanding the functionally important domains of the galactokinase protein and the molecular basis of the disorder, the rarity of the individual defects and the compound heterozygous state of most patients, suggest that meaningful studies of genotype/phenotype correlations in the affected individuals are not feasible. It is the founder mutations, in the homozygous and in the carrier state, that will allow such studies, as well as research into the role of partial galactokinase activity as a factor predisposing to presenile cataracts.

The findings of a common origin, carrier frequency and distribution of one such founder mutation P28T, amongst dispersed Gypsy populations, have implications for the development of newborn screening programs and dietary intervention to prevent the incidence of infantile blindness in susceptible groups. The results also pave the way for the future identification of large sample numbers of both homozygotes and carriers in order for extensive genotype/phenotype correlations to be investigated. This will ultimately contribute to the complete characterization of the role galactokinase plays in galactose metabolism and the molecular basis of the deficiency phenotype.

## References

Ai, Y., Zheng, Z., O'brien-Jenkins, A., Bernard, D. J., Wynshaw-Boris, T., Ning, C., Reynolds, R., Segal, S., Huang, K., & Stambolian, D. (2000). A mouse model of galactose-induced cataracts. *Human Molecular Genetics, 9*(12), 1821-1827.

Ai, Y. J., Jenkins, N. A., Copeland, N. G., Gilbert, D. J., Bergsma, D. J., & Stambolian, D. (1995). Mouse galactokinase - isolation, characterization, and location on chromosome 11. *Genome Research, 5*(1), 53-59.

Asada, M., Okano, Y., Imamura, T., Suyama, I., Hase, Y., & Isshiki, G. (1999). Molecular characterization of galactokinase deficiency in Japanese patients. *Journal of Human Genetics, 44*(6), 377-382.

Ballard, F. J. (1966). Purification and properties of galactokinase from pig liver. *Journal of Biochemistry, 98*, 347-351.

Bergsma, D. J., Ai, Y., Skach, W. R., Nesburn, K., Anoia, E., Van Horn, S., & Stambolian, D. (1996). Fine structure of the human galactokinase GALK1 gene. *Genome Research, 6*(10), 980-5.

Beutler, E., Matsumoto, F., Kuhl, W., Krill, A., Levy, N., Sparkes, R., & Degnan, M. (1973). Galactokinase deficiency as a cause of cataracts. *New England Journal of Medicine, 288*(23), 1203-6.

Blume, K. G., & Beutler, E. (1971). Purification and properties of galactokinase from human red blood cells. *The Journal of Biological Chemistry, 246*(21), 6507-6510.

Budde, M., Gusek-Schneider, G. C., Junemann, A., Jansen, F., & Shin, Y. S. (1999). Familial cataract and increased plasma galactitol levels without known enzyme defect [German]. *Klinische Monatsblatter fur Augenheilkunde, 215*(4), 255-257.

Burchell, A. (1994). Glucose transport across the hepatic membrane. In D. Keppler & J. K. (Eds.), *Transport in the liver* (pp. 59-72). United kingdom: Klumer Academic Publishers.

Campolattaro, B. N. (2000). Congenital and infantile cataracts. *The New York Eye and Ear Infirmary* [on-line]. Available WWW: http://www.nyee.edu/diglib/pedriatri/text/part1.htm [2000, April 10].

Chiara, M. D., Gozani, O., Bennett, M., Champion-Arnaud, P., Palandjian, L., & Reed, R. (1996). Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of the spliceosome assembly. *Molecular and Cellular Biology, 16*(7), 3317-3326.

Cooper, D. N., Krawczak, M., & Antonarakis, S. T. (1995). The nature and mechanisms of human gene mutation. In C. S. Scriver, A. L. Beaudet, W. S. Sly, & D. Valle (Eds.), *The metabolic and molecular bases of inherited disease* (7 ed., Vol. 1, pp. 259-291). New York: McGraw-Hill Inc.

Cuatrecasus, P., & Segal, S. (1965). Mammalian galactokinase: developmental and adaptive characteristics in the rat liver. *The Journal of Biological Chemistry, 240*(6), 2382-2388.

Davit-Spraul, A., Pourci, M. L., Soni, T., & Lemonnier, A. (1994). Metabolic effects of galactose on human HepG2 hepatoblastoma cells. *Metabolism: Clinical & Experimental, 43*(8), 945-52.

Elman, M. J., Miller, M. T., & Matalon, R. (1986). Galactokinase activity in patients with idiopathic cataracts. *Ophthalmology, 93*(2), 210-215.

Elsevier, S. M., Kucherlapati, R. S., Nichols, E. A., Creagan, R. P., Biles, R. E., Ruddle, F. H., Willecke, K., & McDougall, J. K. (1974). Assignment of the gene for galactokinase to human chromosome 17 and its regional localisation to band q21-22. *Nature, 251*, 633-636.

Fiete, D., Brownell, M. D., & Baenziger, J. U. (1983). Evidence for transmembrane modulation of the ligand-binding site of the hepatocyte galactose/N-acetylgalactosamine-specific receptor. *Journal of Biological Chemistry, 258*(2), 817-23.

I ey, P. A. (1996). The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB Journal, 10*(4), 461-70.

Fri dman, T. B., Yarkin, R. J., & Merril, C. R. (1974). Galactose and glucose metabolism in galactokinase deficient, galactose-1-P-uridyl transferase deficient and normal human fibroblasts. *Journal of Cell Physiology, 85*, 569-577.

Gitzelmann, R. (1965). Deficiency of erythrocyte galactokinase in a patient with galactose diabetes. *The Lancet, 2*, 670-671.

Gitzelmann, R. (1967). Hereditary galactokinase deficiency, a newly recognized cause of juvenile cataracts. *Pediatric Research, 1*, 14-23.

Gitzelmann, R., & Hansen, R. G. (1980). Galactose metabolism, hereditary defects and their clinical significance. In D. Burman, J. B. Holton, & C. A. Pennock (Eds.), *Inherited disorders of carbohydrate metabolism.* (pp. 61-101). UK: MTP Press Limited.

Goresky, C. A., Bach, G. G., & Nadeau, B. E. (1973). On the uptake of materials by the intact liver; the transport and net removal of galactose. *The Journal of Clinical Investigation, 52*, 991-1009.

Harley, J. D., Mutton, P., Irvine, S., & Gupta, J. D. (1974). Maternal enzymes of galactose metabolism and the "inexplicable" cataract. *The Lancet, 2*, 259-261.

Henderson, M. J., & Hanna, S. S. (1983). Effective liver blood flow: determination by galactose clearance. *The Canadian Journal of Surgery, 26*(2), 129-132.

Ibu, J. O., & Short, A. H. (1986). The inhibitory effect of phlorhizin and phloretin on hexose transport in the liver. *Scandinavian Journal of Gastroenterology - Supplement, 124*, 75-81.

Jakobs, C., Douwes, A. C., Kok, R., de Jong, A., Endres, W., & Shin, Y. S. (1988). Elevated plasma galactitol levels in patients with congenital cataracts without apparent enzyme defect, *European Journal of Pediatrics*, 147, 446.

Kalaydjieva, L., Perez-Lezaun, A., Angelicheva, D., Onengut, S., Dye, D., Bosshard, N. U., Jordanova, A., Savov, A., Yanakiev, P., Kremensky, I., Radeva, B., Hallmayer, J., Markov, A., Nedkova, V., Tournev, I., Aneva, L., & Gitzelmann, R. (1999). A founder mutation in the GK1 gene is responsible for galactokinase deficiency in Roma (Gypsies). *American Journal of Human Genetics, 65*(5), 1299-1307.

Kolosha, K., Anoia, E., de Cespedes, C., Gitzelmann, R., Shih, L., Casco, T., Saborio, M., Trejos, R., Buist, N., Tedesco, T., Skach, W., Mitelmann, O., Ledee, D., Huang, K., & Stambolian, D. (2000). Novel mutations in 13 probands with galactokinase deficiency. *Human Mutation, 15*(5), 447-453.

Lee, R. T., Peterson, C. L., Calman, A. F., Herskowitz, I., & O'Donnell, J. J. (1992). Cloning of a human galactokinase gene (GK2) on chromosome 15 by complementation in yeast. *Proceedings of the National Academy of Sciences of the United States of America, 89*(22), 10887-91.

Levy, H. L. (1980). Screening for galactosemia. In D. Burman, J. B. Holton, & C. A. Pennock (Eds.), *Inherited disorders of carbohydrate metabolism.* (pp. 133). UK: MTP Press Ltd.

Levy, N. S., Krill, A. E., & Beutler, E. (1972). Galactokinase deficiency and cataracts. *American Journal of Ophthalmology, 74*(1), 41-48.

Litman, N., Kanter, A. I., & Finberg, L. (1975). Galactokinase deficiency presenting as pseudotumor cerebri. *Journal of Pediatrics, 86*(3), 410-412.

Magnani, M., Cucchiarini, L., Dacha, M., & Fornaini, G. (1982a). A new variant of galactokinase. *Human Hereditary, 32*, 329-334.

Magnani, M., Cucchiarini, L., Stocchi, V., Dacha, M., & Fornaini, G. (1982b). Adult and fetal galactokinases in human red blood cells. *Mechanisms of Ageing and Development, 18*, 215-223.

Magnani, M., Cucchiarini, L., Stocchi, V., & Dacha, M. (1983). Red blood cell galactokinase activity and presenile cataracts. *Enzyme, 29*(1), 58-60.

Mathai, C. K., & Beutler, E. (1967). Biochemical characteristics of galactokinase from adult and fetal human red blood cells. *Enzymologia, 33*(4), 224-230.

Mayes, J. S., & Guthrie, R. (1968). Detection of heterozygotes for galactokinase deficiency in a human population. *Biochemical Genetics, 2*(3), 219-30.

Ng, W. G., Donnel, G. N., & Bergren, W. R. (1965). Galactokinase activity in human erythrocytes of individuals at different ages. *Journal of Laboratory and Clinical Medicine, 66*, 115-121.

Olambiwonnu, N. O., McVie, R., Ng, W. G., Frasier, S. D., & Donnell, G. N. (1974). Galactokinase deficiency in twins: clinical and biochemical studies. *Pediatrics, 53*(3), 314-8.

Orita, M., Suzuki, Y., Sekiya, T., & Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics, 5*, 874-879.

Pastuszak, I., O'Donnell, J., & Elbein, A. I. (1996). Identification of the GalNAc kinase amino acid sequence. *Journal of Biological Chemistry, 271*(39), 23653-23656.

Pickering, W. R., & Howell, R. R. (1972). Galactokinase deficiency: clinical and biochemical findings in a new kindred. *Journal of Pediatrics, 81*(1), 50-5.

Rautmann, G., & Breathnach, R. (1985). A role for branchpoints in splicing *in vivo*. *Nature, 315*, 430-432.

Roberts, R. J., & Macelis, D. (2000). REBASE-restriction enzymes and methylases. *Nucleic Acids Research, 28*, 306-307.

Rogers, S. R., Bovee, B. W., Saunders, S. L., & Segal, S. (1989). Galactose as a regulatory factor of its own metabolism by rat liver. *Metabolism, 28*(8), 810-815.

Rost, B., & Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins, 19*, 55-72

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual.* New York: Cold Spring Harbor Laboratory Press.

Segal, S. (1998). Galactosemia today: the enigma and the challenge. *Journal of Inherited Metabolic Disease, 21*(5), 455-471.

Segal, S., & Berry, G. (1995). Disorders of Galactose Metabolism. In C. Scriver, A. Beaudet, W. Sly, & D. Valle (Eds.), *The metabolic and molecular bases of inherited disease* (7 ed., pp. 967-1000). New York: McGraw-Hill.

Shin-Buerhring, Y. S., Beier, A., Tan, A., Osang, M., & Schaub, J. (1977). The activity of galactose-1-phosphate uridyltransferase and galactokinase in human fetal organs. *Pediatric Research, 11*, 1003-1009.

Sitzmann, F. C., & Kaloud, H. (1976). Biokinetics of galactose in the homozygotes and heterozygotes of both forms of galactosemia. *Clinica Chimica Acta, 72*(3), 343-51.

Soni, T., Brivet, M., Blanc, M., Jaeger, G., & Lemonnier, A. (1988). Screening of the Philadelphia variant of galactokinase in racially unmixed black Africans: first results. *American Journal of Human genetics, 42*, 96-103.

Srivastava, S. K., Blume, K. G., Van Loon, C., & Beutler, E. (1972). Purification and kinetic properties of galactokinase from human placenta. *Archives of Biochemistry and Biophysics, 150*, 191-198.

Stambolian, D. (1988). Galactose and cataract. *Survey of Ophthalmology, 32*(5), 333-349.

Stambolian, D., Ai, Y., Sidjanin, D., Nesburn, K., Sathe, G., Rosenberg, M., & Bergsma, D. J. (1986). Cataracts in patients heterozygous for galactokinase deficiency. *Investigations in Ophthalmology and Visual Sciences, 27*, 429-433.

Stambolian, D., Ai, Y., Sidjanin, D., Nesburn, K., Sathe, G., Rosenberg, M., & Bergsma, D. J. (1995). Cloning of the galactokinase cDNA and the identification of mutations in two families with cataracts. *Nature Genetics, 10*, 307-317.

Stambolian, D., Scarpino-Myers, V., & Harris, H. (1985). Purification of human galactokinase and evidence for its existence as a monomer form. *Biochimica et Biophysica Acta, 831*, 306-312.

Strachan, T., & Read, A. P. (1998). *Human Molecular Genetics*. New York: John Wiley & Sons.

Tedesco, T. A., Miller, K., Rabin, P., Diamond, R., Rawnsley, E., & Mellman, W. J. (1973). A variant of human galactokinase with elevated activity. *Pediatric Research, 7*, 394.

Tedesco, T. A., Miller, K. L., Rawnsley, B. E., Adams, M. C., Markus, H. B., Orkwiszewski, K. G., & Mellman, W. J. (1977). The Philadelphia variant of galactokinase. *American Journal of Human genetics, 29*, 240-247.

Tedesco, T. A., Miller, K. L., Rawnsley, B. E., Mennuti, M. T., Spielman, R. S., & Mellman, W. J. (1975). Human erythrocyte galactokinase and galactose-phosphate uridylyltransferase: a population study. *American Journal of Human genetics, 27*, 737-747.

Thalhammer, O., Gitzelmann, R., & Pantlitschko, M. (1968). Hypergalactosemia and galactosuria due to galactokinase deficiency in a newborn. *Pediatrics, 42*(3), 441-5.

Tjian, R. (1995). Molecular machines that control genes. *Scientific American, 272*(2), 38-45.

White, M. B., Carvalho, M., O'brien, S. J., & Dean, M. (1992). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics, 12*, 301-306.

Wilson, D., & Hogness, D. (1969). The enzymes of the galactose operon in E. coli III. The size and composition of galactokinase. *Journal of Biological Chemistry, 244*(8), 2137-2142.

Wu, L., Fritz, J. D., & Powers, A. C. (1998). Different functional domains of GLUT2 glucose transporter are required for glucose affinity and substrate specificity. *Endocrinology, 139*(10), 4205-12.

# Appendix A

## Human Mutation

## Mutation In Brief Report

HUMAN MUTATION Mutation in Brief #394 (2000) Online

*MUTATION IN BRIEF*

# Novel Mutations in the *GALK1* Gene in Patients With Galactokinase Deficiency

**Michael Hunter[1], Dora Angelicheva[1], Harvey L. Levy[3], Sigfried M. Pueschel[4], and Luba Kalaydjieva[1,2]***

[1]*Centre for Human Genetics, Edith Cowan University, Perth, Western Australia and* [2]*Western Australian Institute for Medical Research, Perth, Western Australia;* [3]*Genetic Service, Children's Hospital and Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA;* [4]*Rhode Island Hospital and Brown University Medical School, Providence, Rhode Island, USA*

*Correspondence to: Luba Kalaydjieva, Centre for Human Genetics, Edith Cowan University, 100 Joondalup Drive, Western Australia, 6027; E-mail: l.kalaydjieva@ecu.edu.au

Grant sponsor: The Wellcome Trust; Contract grant number: 058898/Z/99/GEK/SRD

*Communicated by Haig Kazazian*

**Galactokinase deficiency is an inborn error of galactose metabolism whose major clinical manifestation is the development of cataracts during the first months of life. Only 20 mutations have been reported to date and understanding of the functionally important domains of the galactokinase protein is still limited. Here we report four novel mutations in *GALK1* that were identified in two unrelated patients with galactokinase deficiency. Three of these were amino acid substitutions: 1569C→T in exon 2 (R68C); 7093C→T in exon 6 (T288M) and 7538G→C in exon 8 (A384P). In addition, a single base-pair deletion was found in exon 5 (2833delC), predicted to result in a shift of the reading frame and a premature termination codon at position 263. Some differences with the *GALK1* sequence deposited in Genbank are also reported.** © 2000 Wiley-Liss, Inc.

KEY WORDS: Galactokinase deficiency, *GALK1*, galactosemia, galactose metabolism, cataracts

## INTRODUCTION

Galactokinase (E.C.2.7.1.6) catalyses the conversion of galactose to galactose-1-phosphate, the first step of the Leloir metabolic pathway through which galactose enters glycolysis. Galactokinase deficiency (MIM# 230200) is an autosomal-recessive disorder whose metabolic consequences include a dramatic elevation of blood galactose concentration, galactosuria, lack of or greatly diminished production of galactose-1-phospate and activation of an alternative pathway of galactose metabolism, leading to the production of galactitol (Gitzelmann, 1965). The clinical phenotype is confined to the development of cataracts in affected infants during the first months or even weeks of life, as a result of the accumulation of galactitol in the lens of the eye and subsequent osmotic disruption of the lens fibers. Galactokinase deficiency can be diagnosed in the newborn period by hypergalactosemia screening and its consequences are fully preventable by galactose-restricted diet. It has also been suggested that partial deficiency of galactokinase activity in heterozygous carriers may contribute to susceptibility to presenile cataracts (Stambolian et al., 1986).

Received 29 September 2000; Revised manuscript accepted 13 November 2000.

© 2000 WILEY-LISS, INC.

The gene encoding human galactokinase, *GALK1*, was cloned in 1995 and mapped by *in-situ* hybridization to the long arm of chromosome 17 (Stambolian et al., 1995). Linkage analysis in affected families and radiation hybrid mapping have subsequently been used to refine its position to a 1 cM region on 17q24 (Kalaydjieva et al., 1999). The gene spans approximately 7.3kb of genomic DNA and consists of eight exons encoding a protein of 392 amino acid residues (Bergsma et al., 1996). Apart from two putative ATP-binding motifs encoded by exons 3 and 7, understanding of the structure of the galactokinase protein and its functionally important domains is still poor, partly due to the limited number of studies addressing the genetic basis of galactokinase deficiency.

Here we report four novel probable disease-causing mutations identified in two unrelated Caucasian individuals with galactokinase deficiency.

## METHODS

The affected individuals, one female and one male now aged 4 and 17 years, were detected as hypergalactosemic by newborn screening. Galactokinase deficiency (< 5% of the lower limit of the normal range) was confirmed by enzyme activity measurements in red blood cells (RBC). The patients are on galactose-restricted diet and are symptom-free.

The search for mutations in *GALK1* was performed on genomic DNA, by PCR amplification and direct sequencing of both strands of the 5' and 3' UTR, all coding regions of the gene and at least 100 bp of flanking intronic sequences. The primers used for the analysis of the untranslated regions were: for the 5'UTR 5'-CTGATGACCTCTCACAGCTGC-3' and 5'-CTGTCTCAAAGCAGCCATCACG-3', and for the 3'UTR 5-GCTGTGCTTGTGAGGCACC-3' and 5'- CTTCCAAA CTTGACCGCACCCT-3'. Exons and flanking intronic sequences were analysed as described (Kalaydjieva et al., 1999).

PCR reactions were made up to 50 µl and consisted of 20ηg of genomic DNA, 1x PCR Buffer, 1.5mM $MgCl_2$, 1.25mM dNTPs, 4% DMSO, 2mM of each primer and 0.5 U of Taq. Cycling conditions included initial denaturation at 94°C for 5 minutes, 35 cycles of 94°C for 30 sec, 63°C for 30 sec and 72°C for 45 sec, and final extension at 72°C for 7 minutes. Sequencing was performed using the ABI Prism Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit and the reactions were run on an ABI Prism 377 automated DNA Analyzer. Data processing was done using Sequence Navigator™ (PE Biosystems), where the patient sequences were compared to unaffected controls run in parallel, and to the published *GALK1* sequence (Genbank Accession # L76927). Searches of the public domain genomic databases were performed using the BLAST program [available at http://www.ncbi.nlm.nih.gov/ BLAST/]. Computer modeling of the predicted structure of the mutant protein was done with the PHDsec software (Rost & Sander, 1994) available at [http://dodo.cpmc.columbia.edu/ predictprotein/submit_def.html].

A panel of 50 unrelated adult healthy Caucasian controls was tested for the presence of the putative disease-causing mutations by PCR-based restriction assays.

The nucleotide numbering provided below is according to Genbank Accession # L76927.

## RESULTS

Both individuals were found to be compound heterozygotes for different molecular defects, none of which has been described previously: patient 1 presented with two missense mutations, R68C and A384P, and patient 2 carried one missense, T288M, and one frameshift mutation, 2833delC.

### R68C

A 1569C→T transition in exon 2 is predicted to result in the replacement of arginine by cysteine at codon 68. R68C is the fourth mutation described in exon 2, the remaining three being truncating mutations (Stambolian et al., 1995; Kolosha et al., 2000). The arginine residue at this position is strictly conserved in the galactokinase protein of over twenty different species checked, including the mouse. The mutation is a drastic change from a charged to an uncharged amino acid residue with a sulfhydryl side-chain, however the computer predictions did not suggest significant alterations in the secondary structure of the mutant protein. The 1569C→T transition creates an Alw NI restriction site. The PCR-based restriction assay, used as a screening test, failed to detect the mutation among 100 control chromosomes.

### A384P

The 7538G→C transversion found in exon 8 leads to the substitution of proline for alanine at position 384. A384P is located 8 amino acid residues upstream of the carboxyl terminus of the protein and is the second reported mutation in exon 8, after Q382X (Kolosha et al., 2000), suggesting that the carboxyterminal part of the protein is essential for galactokinase function. Alanine at this position is not highly conserved, it is present only in the two

mammalian galactokinases characterised so far, human and mouse, and sequence similarity is found in the proteins of seven additional species. The substitution of proline for alanine is likely to induce significant topological changes, as predicted by the secondary structural analysis. The helix formed between the F378 and A384 residues in the wild-type protein is abolished by the mutation, which also leads to an extension of the beta sheet and causes loop formation by adjacent residues (Figure 1). The mutation abolishes a Bbv I restriction site. The restriction assay failed to detect A348P among the 50 control individuals.

**Galactokinase, wild-type sequence**

```
         370 . . . . . . . . . . 380 . . . . . . . . . . 390 . .
AA     | E H Y G G T A T F Y L S Q A A D G A K V L C L |
PHDsec | H H         E E H H H H H H H       E E E E   |
SUBsec | H H . L L L . . . . . . . . . . L L . . . . L |
```

**Galactokinase A384P**

```
         370 . . . . . . . . . . 380 . . . . . . . . . . 390 . .
AA     | E H Y G G T A T F Y L S Q A P D G A K V L C L |
PHDsec | H H H       E E E E E E             E E E E   |
SUBsec | H H H . . . E E E E E . . L L L L . . . . . L |
```

**Figure 1** Secondary-structure prediction for the wild-type exon eight galactokinase sequence and the A384P missense mutation using the PHDsec algorithm. P384 in boldtype face. E = extended sheet (beta sheet); H = helix; L = loop. The prediction is meaningful for all residues with an expected average correlation > .69.

**T288M**

A 7093C→T transition causes the substitution of methionine for threonine at codon 288. T288M is the first mutation identified in exon 6 of *GALK1*, encoding a conserved galactokinase fingerprint motif whose function is currently unknown. As shown by the sequence alignment of this galactokinase motif (Figure 2), the threonine residue at position 288 is conserved in nine out of twenty galactokinases. The mutation results in the replacement of a polar amino acid residue by a nonpolar with a sulfur-containing sidechain. Secondary structural predictions are inconclusive. The mutation was tested in the panel of control individuals using a restriction assay (Nla III site abolished by the mutation). It was not found among 100 normal chromosomes.

**2833delC**

A single base-pair deletion, 2833delC, was found in exon 5, where one missense and another truncating mutation have been reported previously (Asada et al., 1999; Kolosha et al., 2000). The 2833delC mutation is predicted to result in a frameshift and a premature termination signal at codon 263, leading to a protein that is 130 amino acids shorter than the wild-type enzyme. The presence of the deletion was independently confirmed by single strand conformational polymorphism (SSCP) and heteroduplex (HDA) analysis.

In addition, we have detected a number of differences to the sequence published under Genbank Accession # L76927 (Table 1). They were found in both affected individuals and unaffected controls (n=9) of Caucasian and Afro/Hispanic descent in the present study, as well as in the Japanese and Caucasian individuals reported by Asada et al. (1999).

**Table 1. Nucleotide Differences from Published *GALK1* Sequence GenBank Accession # L76927**

| Region | Position (L76927) | Nucleotide (L76927) | Nucleotide (this study) n=9 | Nucleotide (Asada et al 1999) n=7 |
|---|---|---|---|---|
| 5'UTR | 415 | C | T | T |
| 5'UTR | 418 | C | T | T |
| Intron 1 | 694 | C | T | T |
| Intron 2 | 2149 | C | G | G |

| Intron 2 | 2151 | A | C | C |
|---|---|---|---|---|

## DISCUSSION

The incidence of galactokinase deficiency, as reported by newborn screening programs in Europe, the United States and Japan, ranges between 1:150,000 and 1:1,000,000 (Levy, 1980; Gitzelmann & Hansen, 1980). The low frequency of the disorder has imposed limitations on research into the genetic causes of galactokinase deficiency, resulting in a small number of reported mutations in *GALK1*. Moreover, the small number of definite carriers that can be identified by mutation detection precludes studies addressing the proposed role of partial galactokinase deficiency in the development of the common forms of presenile cataracts.

| Galactokinase motif four | Species |
|---|---|
| K R A R H A V S E N Q R **T** L K | *Streptomyces lividans* |
| K R A R H V V T E N Q R **V** L D | *Haemophilus influenzae* |
| K R V R H I L T E N A R **T** V E | *Escherichia coli* |
| K R V R H I L T E N A R **T** V E | *Bacteriophage lambda* |
| K R V R H V L S E N A R **T** V E | *Salmonella typhimurium* |
| K R A K H I I S E N Q R **V** L E | *Actinobacillus pleuropneumoniae* |
| Q R A K H V Y S E S L R **V** L K | *Saccharomyces cerevisiae* |
| K R A R H A V S E N Q R **T** L R | *Lactobacillus helveticus* |
| Q R A K H V Y S E A L R **V** L K | *Kluyveromyces lactis* |
| Q R A K H V Y Q E S L R **V** L E | *Candida albicans* |
| E R A L H V Y R E S L R **V** L K | *Candida parapsilosis* |
| R R A K H A V Y E N H R **A** I K | *Bacillus subtilis* |
| K R A R H V V L E N Q R **T** L Q | *Streptococcus thermophilus* |
| Q R A K H V Y S E S L R **V** L K | *Saccharomyces cerevisiae (GAL3)* |
| R R A R H V V G E I R R **T** A Q | *Homo sapiens* |
| R R A R H V V S E I R R **T** A Q | *Mus musculus* |
| K R A R H A V F E N Q R **T** L K | *Lactobacillus casei* |
| K R A Q H V L E E N E R **V** L K | *Thermotoga neapolitana* |
| K F F G Y I V R E N A R **V** L E | *Pyrococcus horikoshii* |
| R L V R H V V T E D E R **V** E R | *Streptomyces lividans* |

**Figure 2.** Galactokinase fingerprint motif four. Bold highlight shows amino acid variation analogous to residue 288 of the human galactokinase protein at which the T288M substitution is found. Threonine at this residue is conserved in 9 of 20 organisms. Methionine at this residue is not shared by any of the galactokinases that share the fourth galactokinase fingerprint motif. Motif alignment by SPRINT (available at http://www.bioinf.man.ac.uk/ dbbrowser/sprint/).

A total of twenty *GALK1* mutations associated with the galactokinase deficiency phenotype have been identified to-date (Stambolian et al., 1995; Asada et al., 1999; Kalaydjieva et al., 1999; Kolosha et al., 2000). Except for two founder mutations, namely P28T which occurs among the Vlax Roma (Kalaydjieva et al., 1999) and Q382X found in patients of Costa Rican/European descent (Kolosha et al., 2000), the remaining aberrations are private mutations, confined to individual families. Similar to most individuals with galactokinase deficiency described previously, our patients were also found to be compound heterozygotes, adding four novel molecular defects to the allelic heterogeneity of *GALK1* mutations. The lack of other changes in the *GALK1* gene, the nature of the detected sequence variants and their predicted effect on the protein, lead us to suggest that these are disease-causing mutations. Including the mutations identified in this study, the overall distribution to-date is 14 amino acid substitutions and 10 truncating mutations, with exons 1, 2 and 7 harbouring 13 of the known mutations. The majority of missense mutations cluster in close proximity to the galactokinase signature sequence in exon 1 and the

second ATP-binding domain in exon 7, highlighting the functional significance of these regions for the enzyme activity.

While the identification of additional *GALK1* mutations in individuals with galactokinase deficiency should ultimately contribute to understanding the functionally important domains of the galactokinase protein, the rarity of the individual defects and the compound heterozygous state of most patients, suggest that meaningful studies of genotype/phenotype correlations in the affected individuals are not feasible. It is the founder mutations, in the homozygous and in the carrier state, that will allow such studies, as well as research into the role of galactokinase deficiency as a factor predisposing to the common presenile cataracts.

## REFERENCES


Asada M, Okano Y, Imamura T, Suyama I, Hase Y, Isshiki G. 1999. Molecular characterization of galactokinase deficiency in Japanese patients. J Hum Genet 44:377-382.

Bergsma DJ, Ai Y, Skach WR, Nesburn K, Anoia E, Van Horn S, Stambolian D. 1996. Fine structure of the human galactokinase GALK1 gene. Genome Res 6:980-5.

Gitzelmann, R. 1965. Deficiency of erythrocyte galactokinase in a patient with galactose diabetes. Lancet 2:670-671.

Gitzelmann R, Hansen, RG. 1980. Galactose metabolism, hereditary defects and their clinical significance. In Burman D, Holton JB, Pennock CA, editors. Inherited disorders of carbohydrate metabolism.UK: MTP Press Limited. p 61-101.

Kalaydjieva L, Perez-Lezaun A, Angelicheva D, Onengut S, Dye D, Bosshard NU, Jordanova A, Savov A, Yanakiev P, Kremensky I, Radeva B, Hallmayer J, Markov A, Nedkova V, Tournev I, Aneva L, Gitzelmann R. 1999. A founder mutation in the GK1 gene is responsible for galactokinase deficiency in Roma (Gypsies). Am J Hum Genet 65:1299-1307.

Kolosha K, Anoia E, de Cespedes C, Gitzelmann R, Shih L, Casco T, Saborio M, Trejos R, Buist N, Tedesco T, Skach W, Mitelmann O, Ledee D, Huang K, Stambolian D. 2000. Novel mutations in 13 probands with galactokinase deficiency. Hum Mutat 15:447-453.

Levy HL. 1980. Screening for galactosemia. In Burman D, Holton JB, Pennock CA, editors. Inherited disorders of carbohydrate metabolism. UK: MTP Press Ltd. p 133.

Rost B, Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19:55-72.

Stambolian D, Scarpino-Myers V, Eagle RC, Hodes B, Harris H. 1986. Cataracts in patients heterozygous for galactokinase deficiency. Invest Ophthal Vis Sci 27:429-433.

Stambolian D, Ai Y, Sidjanin D, Nesburn K, Sathe G, Rosenberg M, Bergsma DJ. 1995. Cloning of the galactokinase cDNA and the identification of mutations in two families with cataracts. Nat Genet 10:307-317.