EXPRESSION VARIATION
IN
LYSOSOMAL STORAGE DISORDER GENES

by
Lyndel Ann Mason
Bachelor of Applied Science (Honours, 1A)
PhD Candidate

Co-operative Research Centre for Diagnostics (CDx)
Queensland University of Technology (QUT)
Brisbane, 4001. Australia.

A thesis submitted for the degree of Doctor of Philosophy of the Queensland University of Technology.

2005
KEYWORDS
Gaucher disease (GD), glucocerebrosidase gene (GBA), glucocerebrosidase (GBA), metachromatic leucodystrophy (MLD), arylsulphatase A gene (ARSA), arylsulphatase A (ASA), single nucleotide polymorphism (SNP), polymerase chain reaction (PCR), promoter, transcriptional regulation, translational regulation, lysosomal storage disorders (LSDs)
ABSTRACT

Metachromatic leukodystrophy (MLD) and Gaucher disease (GD) are caused by a deficiency of arylsulphatase A (ASA) and β-glucocerebrosidase (GBA), respectively. They are lysosomal storage disorders with a heterogeneous clinical spectrum encompassing visceral, skeletal and neurologic involvement resulting in high morbidity and mortality. The overall aim of this study is to elucidate the genetic component/s of high ASA and GBA enzyme activity in normal healthy individuals with the ultimate goal of using this information to produce greater protein activity from a recombinant protein.

A wide variation in ASA and GBA enzyme activity levels has been observed in the normal population. The first objective of this project was to identify and characterise single nucleotide polymorphisms (SNPs) in the arylsulphatase A (ARSA) and glucocerebrosidase (GBA) genes that are responsible for determining the levels of expressed enzyme activity in the normal population. The second objective was to assess the contribution of transcriptional regulation and TCP80 mediated translational control to normal enzyme variation. TCP80, a translational control protein that interacts with the GBA coding region, is a splice variant of the interleukin binding factor 3 (ILF3) gene.

Ten samples from individuals with high ASA activity and twenty samples from individuals with high GBA activity were screened for polymorphisms via denaturing high pressure liquid chromatography (dHPLC) and sequencing. The frequency of these polymorphisms in the normal population was determined using dot-blot hybridisation. Fifteen ARSA polymorphisms (4 promoter, 5 coding, 5 intronic and 1 poly(A) signal) and two GBA polymorphisms (1 intronic and 1 in 3′-UTR) were identified. Two low frequency ASA polymorphisms (2723A>G, W193C) were found to be correlated with low activity, while another low frequency ASA polymorphism (1101+123C>T) was found to be correlated with high activity in a population of 113 individuals.

Real time PCR was used to measure mRNA levels of GBA, ASA and ILF3 along with enzyme activity levels of GBA and ASA in two cell types (leucocytes and skin fibroblasts) from four healthy individuals and seven cell lines (HL60, THP1, Huh7, U118, SW1353, Hep G2, and B-cells). Transcriptional control was evident for all three genes with GBA mRNA levels varying over 30 fold, ASA mRNA levels varying over seven fold and ILF3
levels varying more than 24 fold. The 5′-flanking region of GBA was investigated for the 
cis-elements responsible for tissue-specific expression. However, it was not possible to
demonstrate that the cis-element region was influencing GBA expression.

Translational efficiency was measured using the magnitude of the mRNA:enzyme activity
ratio as an indicator. GBA translational inefficiency was most pronounced in B cells which
require four times more mRNA molecules than hepatocytes (Hep G2) and over 25 times
more mRNA molecules than chondrocytes (SW1353) to produce one unit of GBA
enzyme activity. Except in B-cells, GBA translational efficiency appears to increase as
ILF3 mRNA levels decrease. The tissue-specific variation observed in the protein levels
of the ILF3 splice variants, TCP80 and DRBP76, may play a role.

The correlation of several low frequency SNPs with low ASA enzyme activity or high
ASA activity indicates a role in determining the distribution of enzyme activity levels in
the normal population. However, there do not appear to be any common high activity
polymorphisms. Knowledge of the exact mechanisms responsible for the observed
transcriptional and translational control of these lysosomal genes will greatly enhance the
understanding of genotype-phenotype correlation and the contribution of genetic variants
to natural variation.
CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION ............................................................................................................................2

1.2 THE LYSOsome ....................................................................................................................................3

1.2.1 The Endosomal-Lysosomal System .............................................................................................3

1.2.2 Synthesis and Trafficking of Lysosomal Enzymes (the ‘synthetic’ pathway) .........................6

1.2.3 Classification of Lysosomal Storage Disorders .........................................................................7

1.3 MOLECULAR GENETICS OF LYSOSOMAL STORAGE DISORDERS........................................9

1.3.1 Clinical Presentation .................................................................................................................9

1.4 METACHROMATIC LEUCODYSTROPHY (OMIM 250100) ......................................................10

1.4.1 The Arylsulphatase A Gene (ARSA) (OMIM 607574) ...............................................................12

1.5 GAUCHER DISEASE (OMIM 230800, 230900, 231000) .............................................................14

1.5.1 The Glucocerebrosidase Gene (GBA) (OMIM 606463) .............................................................14

1.6 GENOTYPE PHENOTYPE CORRELATION .................................................................................17

1.6.1 Normal Levels of Enzyme ........................................................................................................18

1.6.2 High Activity Polymorphisms in ARSA and GBA? .................................................................22

1.6.3 Other Possible Causes of Variation of Expression in Individuals ...........................................22

1.7 EXPRESSION OF THE GBA GENE .......................................................................................23

1.7.1 GBA Transcription ..................................................................................................................24

1.7.2 GBA Translation ......................................................................................................................31

1.7.2.1 The Translational Control Protein (TCP80) .........................................................................32

1.7.2.2 The Interleukin Enhancer Binding Factor Gene (ILF3) .........................................................34
1.8 EXPRESSION OF THE ARSA GENE .................................................................................... 38
1.9 SUMMARY AND RELEVANCE TO PROJECT ................................................................... 39
1.10 SPECIFIC AIMS ......................................................................................................................... 40

CHAPTER 2
MATERIALS AND METHODS
2.1 INTRODUCTION .......................................................................................................................... 44
2.2 GENERAL REAGENTS AND CHEMICALS ........................................................................ 44
2.3 LYSOSOMAL ENZYME ACTIVITY ASSAYS ........................................................................ 44
   2.3.1 Isolation of Peripheral Blood Leucocytes ........................................................................ 44
   2.3.2 Preparation of a Cell Pellet for Enzyme Assay ............................................................... 44
   2.3.3 Assay of Arylsulphatase A Enzyme Activity in Human Cells ........................................ 45
   2.3.4 Assay of β-Glucosidase Enzyme Activity in Human Cells ............................................. 45
   2.3.5 Assay of α-Galactosidase Enzyme Activity in Human Cells .......................................... 46
2.4 ASSAY OF PROTEIN CONCENTRATION IN CELL EXTRACT ........................................ 47
2.5 DNA PREPARATION .................................................................................................................... 47
   2.5.1 Genomic DNA Extraction .................................................................................................. 47
2.6 THE POLYMERASE CHAIN REACTION (PCR) .................................................................. 47
   2.6.1 Electrophoresis of PCR amplicons .................................................................................... 48
2.7 DNA SEQUENCING .................................................................................................................... 48
2.8 STATISTICAL ANALYSIS ........................................................................................................... 49
2.9 CELL TYPES USED ....................................................................................................................... 49
2.10 TISSUE CULTURE ....................................................................................................................... 49
   2.10.1 Cell Counting .................................................................................................................... 50
2.11 USE OF CELLS ............................................................................................................................ 50
   2.11.1 Enzyme Activity and mRNA Quantification ................................................................. 50
      2.11.1.1 Preparation of Each Flask of Cultured Cells for Assay of Enzyme Activity and
      RNA Extraction ....................................................................................................................... 51
2.12 QUANTIFICATION OF mRNA .............................................................................................. 51
   2.12.1 Total RNA Preparation ..................................................................................................... 51
      2.12.1.1 Removal of Genomic Contamination ....................................................................... 52
   2.12.2 Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) ..................................... 53
   2.12.3 The Real-Time Polymerase Chain Reaction ................................................................. 53
      2.12.3.1 Design of Real-Time PCR Primers ............................................................................. 53
2.12.3.2 Standard Curve Generation ........................................................................................................55
2.12.3.3 Real-Time PCR Method ..............................................................................................................56

2.13 CLONING ........................................................................................................................................56
2.13.1 Preparation of Bacterial Fosmid DNA .......................................................................................56
2.13.2 Preparation of the pGL3-Basic Reporter Vector for the Ligation of GBA Promoter Constructs ...............................................................................................................................57
2.13.3 Preparation of the GBA Promoter PCR Products for Ligation into pGL3 Reporter Vectors ........................................................................................................................................58
2.13.4 Ligation of PCR Amplicon into the pGL3 Reporter Vector .......................................................58
2.13.5 Transformation of pGL3-Basic or pGL3-Promoter Vectors into JM109 High Efficiency E. coli Competent Cells by Heat Shock ...........................................................................58
2.13.6 Plating of Transformation Cultures onto LB/Ampicillin Plates ..............................................59
2.13.7 Identification of Positive Clones .................................................................................................59
2.13.8 Extraction of Plasmid DNA Containing Inserts From the Bacterial Cell Pellets ..................59

2.14 REPORTER GENE ASSAYS ON THE GBA PROMOTER .........................................................59
2.14.1 Transfection of GBA pGL3 Promoter Constructs Into Attached Cell Lines .......................60
2.14.2 Reporter Gene Assays .................................................................................................................61

CHAPTER 3
EXPRESSION VARIATION OF HUMAN GLUCOCEREBROSIDASE AND ARYSULPHATASE A AND HIGH ACTIVITY SNPs

3.1 INTRODUCTION .................................................................................................................................64

3.2 MATERIALS AND METHODS ........................................................................................................65
3.2.1 Design of Amplicons for ARSA Gene ..........................................................................................66
3.2.2 Design of Amplicons for the GBA Gene .....................................................................................68
3.2.3 Polymorphism Detection Using Denaturing High Performance Liquid Chromatography (DHPLC) ..........................................................................................................................71
3.2.4 Dot Blot Hybridisation ................................................................................................................73

3.3 RESULTS ..........................................................................................................................................75
3.3.1 Samples .......................................................................................................................................75
3.3.2 Analysis of ASA Polymorphisms in the Sample Population .............................................78
  3.3.2.1 Statistical Analysis of ASA Polymorphism Data.................................................................86
3.3.3 Analysis of the GBA Gene........................................................................................................94
  3.3.3.1 Development of the Whole Gene Amplification Protocol for GBA and Internal
GBA Fragment PCRs .........................................................................................................................94
  3.3.3.2 Analysis of GBA Fragments ..................................................................................................99
3.3.4 Does a high activity allele exist? .............................................................................................99
3.3.5 Investigation of Environmental Contribution to Enzyme Activity Level in
Individuals ........................................................................................................................................100
  3.3.5.1 Quantitation of ASA and GBA mRNA Levels in Leucocytes and Skin Fibroblasts
in Four Individuals ........................................................................................................................104
3.4 DISCUSSION...............................................................................................................................111
3.5 CONCLUSION ............................................................................................................................114

CHAPTER 4
TRANSCRIPTIONAL REGULATION OF THE GLUCOCEREBROSIDASE
AND ARYLSULPHATASE A GENES
4.1 INTRODUCTION .......................................................................................................................116
4.2 MATERIALS AND METHODS ..............................................................................................118
  4.2.1 Promoter Construct Design ...............................................................................................118
  4.2.2 Acquisition of Fosmid and Preparation of DNA ..........................................................118
  4.2.3 PCR Amplification of the GBA 5′-UTR ...............................................................................120
  4.2.4 Cloning PCR Amplicons into pGL3-Basic Luciferase Vector ..............................................120
  4.2.5 Transfection of the GBA Constructs into Cell Lines and Analysis of Luciferase
Activity ..............................................................................................................................................120
4.3 RESULTS ......................................................................................................................................123
  4.3.1 Choosing Cell Lines in which to Investigate Regulation ..................................................123
  4.3.2 Variation in mRNA Levels .................................................................................................123
  4.3.3 GBA 5′-UTR Analysis Strategy ........................................................................................127
  4.3.3.1 Identification of Restriction Sites for the Cloning of GBA Constructs Into pGL3-Basic
Luciferase Vector ...........................................................................................................................127
  4.3.3.2 Amplification of GBA Constructs from Fosmid DNA .......................................................132
  4.3.3.3 Cloning of GBA Promoter Constructs Into the pGL3-Basic Luciferase Vector ..........132
  4.3.3.4 Sequencing of pGL3-Basic Luciferase Constructs .............................................................132
4.3.4 Analysis of the GBA Promoter Activity .................................................................133
  4.3.4.1 Comparison of Basal Expression Levels in Four Cell Lines .........................134
  4.3.4.2 Effect of Serum Stimulation in Hep G2 Cells ......................................................134
  4.3.4.3 Effect of Serum Stimulation in All Four Cell Lines .............................................136
  4.3.4.4 Repeat Serum Stimulation Experiments ...............................................................138

4.3.5 Bioinformatic Analysis of GBA 5′-UTR .................................................................140
  4.3.5.1 Expression of Putative Cis-Element Region .......................................................141

4.4 DISCUSSION ...............................................................................................................145

4.5 CONCLUSION ..............................................................................................................147

CHAPTER 5
TRANSLATIONAL REGULATION OF THE GLUCOCEREBROSIDASE
AND ARYLSULPHATASE A GENES

5.1 INTRODUCTION .........................................................................................................150

5.2 MATERIALS AND METHODS ....................................................................................151
  5.2.1 Western Blot Analysis of Proteins ...........................................................151
    5.2.1.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) ....151
    5.2.1.2 Western Blot Analysis .................................................................................152
  5.2.2 TNT Reticulocyte Coupled Transcription/Translation Reaction ...................154
    5.2.2.1 Preparation of GBA cDNA Plasmid DNA ..................................................154
    5.2.2.2 The Coupled Transcription/Translation Reaction .......................................154

5.3 RESULTS .....................................................................................................................154
  5.3.1 Variation in Enzyme Activity Levels in Nine Human Cell Types .......................154
    5.3.1.1 Efficiency of Translation ............................................................................157
  5.3.2 Analysis of the Relationship Between TCP80 and GBA ....................................157
    5.3.2.1 Primer Design for TCP80 mRNA Quantification ........................................157
    5.3.2.2 Quantification of ILF3 mRNA Levels .........................................................159
  5.3.3 The Translational Efficiency Ratio (TER) ...........................................................164
    5.3.3.1 Translational Efficiency of GBA and ASA in Nine Human Cell Types ........164
    5.3.3.2 Translational Efficiency of GBA and ASA in Individuals ............................164
  5.3.4 Evaluation of In Vitro Translational Efficiency of GBA in Human Cell Lines ....168
    5.3.4.1 The GBA cDNA Clone ..............................................................................168
    5.3.4.2 Measurement of GBA Enzyme Activity of the In Vitro Translation Product ...168
  5.3.5 Measurement of TCP80/ILF3 Protein Levels .....................................................169
LIST OF FIGURES

Figure 1.1   Lysosomal Enzyme Production
Figure 1.2   Lysosomal Catabolism of Some of the Sphingolipids
Figure 1.3   Progressive Deterioration in Juvenile Metachromatic Leucodystrophy (MLD)
Figure 1.4   The Arylsulphatase A Gene (ARSA)
Figure 1.5   The Glucocerebrosidase Gene (GBA)
Figure 1.6   Population Distribution of ASA and GBA Activity Levels in Leucocytes
Figure 1.7   Deletion Analysis of the GBA Promoter
Figure 1.8   The GBA Promoter Region
Figure 1.9   Cytoplasmic Protein (TCP80) Binding to the GBA RNA-coding Region
Figure 1.10  Alternative Splicing of the ILF3 Gene Creates Five Transcripts
Figure 3.1   Arylsulphatase A Fragments for PCR
Figure 3.2   α-glucocerebrosidase Gene Fragments for PCR
Figure 3.3   Distribution of Leucocyte ASA Levels in the Sample Population
Figure 3.4   Distribution of Leucocyte GBA Levels in the Sample Population
Figure 3.5   Linkage Disequilibrium Analysis of the ASA SNPs
Figure 3.6   Mean ASA Enzyme Activity of Each Genotype in the Sample Population
Figure 3.7   The Gene Arrangement at Chromosome 1q21
Figure 3.8   Development of a ‘Whole Gene’ PCR for Specific Amplification of GBA Gene
Figure 3.9   The GBA-Specific Exon 9 Confirmatory Fragment
Figure 3.10  Lysosomal Enzyme Activity Levels in Four Individuals Over Time
Figure 3.11  Lysosomal Enzyme Expression in Leucocytes
Figure 3.12  Lysosomal Enzyme Expression in Skin Fibroblasts
Figure 3.13  Lysosomal Enzyme Expression in Skin Fibroblasts (method 2)
Figure 4.1   The GBA Fosmid
Figure 4.2   The Level of ASA mRNA in Nine Human Cell Types
Figure 4.3   The Level of GBA mRNA in Nine Human Cell Types
Figure 4.4   The pGL3-GBA Plasmids – The DNA Sequence
Figure 4.5   The pGL3-GBA Plasmids – The Constructs
Figure 4.6   The Expression of the Nine GBA Promoter Constructs in Four Human Cell Lines
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7</td>
<td>The Effect of Serum Stimulation on the Expression of GBA Promoter Constructs</td>
</tr>
<tr>
<td>4.8</td>
<td>The Effect of Serum Stimulation on Four Human Cell Lines</td>
</tr>
<tr>
<td>4.9</td>
<td>Analysis GBA 5’-Flanking Genomic Sequence by CISTER</td>
</tr>
<tr>
<td>4.10</td>
<td>Expression of pGL3-4700 and Putative Enhancer Region</td>
</tr>
<tr>
<td>5.1</td>
<td>The DRBP76 Antibody</td>
</tr>
<tr>
<td>5.2</td>
<td>GBA and ASA mRNA and Enzyme Activity Levels in Nine Cell Types</td>
</tr>
<tr>
<td>5.3</td>
<td>Alignment of the Coding Sequences of the ILF3 Splice Variants</td>
</tr>
<tr>
<td>5.4</td>
<td>ILF3 mRNA Levels in Nine Cell Types</td>
</tr>
<tr>
<td>5.5</td>
<td>The Translational Efficiency Ratio for GBA and ASA Compared to ILF3 mRNA Levels in Nine Cell Types</td>
</tr>
<tr>
<td>5.6</td>
<td>The Translational Efficiency Ratio Compared to ILF3 mRNA Levels in Skin Fibroblasts from Four Individuals</td>
</tr>
<tr>
<td>5.7</td>
<td>The Translational Efficiency Ratio Compared to ILF3 mRNA Levels in Leucocytes from Four Individuals</td>
</tr>
<tr>
<td>5.8</td>
<td>ClustalW Multiple Sequence Alignment of ILF3 Splice Variants</td>
</tr>
<tr>
<td>5.9</td>
<td>Western Blot Analysis of ILF3 Isoform Proteins</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1 Primers for Real-Time PCR
Table 3.1 Primer Sequences, Annealing Temperatures and Product Sizes for PCR of the Arylsulphatase A Gene
Table 3.2 Primer Sequences, Annealing Temperatures and Product Sizes for PCR of the Glucocerebrosidase Gene
Table 3.3 Optimum Partial Denaturation Temperatures for PCR Amplicons of the Glucocerebrosidase Gene
Table 3.4 Allele Specific Oligonucleotides for ARSA Used in the Dot-Blot Hybridisation
Table 3.5 The ASA Genotypes of Individuals in the Sample Population
Table 3.6 Statistical Analysis of the ASA SNPs
Table 3.7 Forward Linear Regression Analysis of ASA SNPs
Table 3.8 Enzyme Activity Levels for ASA, GBA and α-GAL in Four Individuals Over Three Consecutive Weeks
Table 4.1 Primers for pGL3-GBA Constructs
Table 4.2 The Levels of ASA and GBA mRNA in Nine Human Cell Lines
Table 5.1 ASA and GBA Enzyme Activity Levels in Nine Cell Types
Table 5.2 ILF3 Splice Variants
Table 5.3 Level of ILF3 mRNA in Nine Cell Types
Table 5.4 Western Blot Analysis of ILF3 Isoform Proteins
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASA</td>
<td>arylsulphatase A</td>
</tr>
<tr>
<td>ARSA</td>
<td>arylsulphatase A gene</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CO₂</td>
<td>carbon dioxide</td>
</tr>
<tr>
<td>DMEM</td>
<td>dulbecco’s modified eagles medium</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTPs</td>
<td>equimolar mixture of dATP, dCTP, dGTP, dTTP</td>
</tr>
<tr>
<td>EB</td>
<td>ethidium bromide</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diamine tetra acetate</td>
</tr>
<tr>
<td>FCS</td>
<td>fetal calf serum</td>
</tr>
<tr>
<td>g</td>
<td>gravity</td>
</tr>
<tr>
<td>GBA</td>
<td>glucocerebrosidase gene</td>
</tr>
<tr>
<td>GBA</td>
<td>β-glucocerebrosidase or glucocerebrosidase</td>
</tr>
<tr>
<td>GD</td>
<td>Gaucher disease</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>ILF3</td>
<td>interleukin binding factor 3</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>L</td>
<td>litres</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple cloning site</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>mg</td>
<td>milligrams</td>
</tr>
<tr>
<td>ml</td>
<td>millilitres</td>
</tr>
<tr>
<td>MLD</td>
<td>metachromatic leucodystrophy</td>
</tr>
<tr>
<td>nmol</td>
<td>nanomole</td>
</tr>
<tr>
<td>mol</td>
<td>moles</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>mol/L</td>
<td>moles per litre</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>ng</td>
<td>nanograms</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>RE</td>
<td>restriction enzyme</td>
</tr>
<tr>
<td>RER</td>
<td>rough endoplasmic reticulum</td>
</tr>
<tr>
<td>RF</td>
<td>relative fluorescence</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>rotations per minute</td>
</tr>
<tr>
<td>RPMI</td>
<td>Roswell Park Memorial Institute Media 1640</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcription-polymerase chain reaction</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>sec</td>
<td>seconds</td>
</tr>
<tr>
<td>SF</td>
<td>skin fibroblast</td>
</tr>
<tr>
<td>SOP</td>
<td>standard operating procedures</td>
</tr>
<tr>
<td>SSC</td>
<td>sodium chloride, sodium citrate buffer</td>
</tr>
<tr>
<td>TCP80</td>
<td>translational control protein 80</td>
</tr>
<tr>
<td>TER</td>
<td>translational efficiency ratio</td>
</tr>
<tr>
<td>TSS</td>
<td>transcription start site</td>
</tr>
<tr>
<td>U</td>
<td>unit</td>
</tr>
<tr>
<td>µg</td>
<td>micrograms</td>
</tr>
<tr>
<td>µl</td>
<td>microlitres</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>ultra-violet</td>
</tr>
<tr>
<td>4MU</td>
<td>4-methyllumbiferone</td>
</tr>
</tbody>
</table>
STATEMENT OF ORIGINAL AUTHORSHIP

The work contained in this thesis has not been previously submitted for a degree or a diploma at any other higher education institute. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signed:

Lyndel Ann Mason
B. App. Sci. (Hons).
ACKNOWLEDGEMENTS

Over the past five years, many people have taken part in my journey through my PhD. It has been interesting, challenging, memorable, life-changing and so much more rewarding because of their participation.

Firstly, I would like to very gratefully acknowledge the help, encouragement, patience and understanding of my supervisor, Associate Professor C. Phillip Morris who taught me the essential scientific skill of critical thinking, by example. Without him and his sound advice, finding my way through this project would have been impossible.

I would also like to thank my associate supervisor, Associate Professor Angela van Daal, for her help throughout my journey but especially for the critical reading of this thesis and her support towards the end of my project.

Many people provided valuable samples and scientific expertise for my research. I am very grateful to Paul Nelson, Michael Fietz and their laboratory associates from the Lysosomal Diseases Research Unit at Women’s and Children’s Hospital in Adelaide for samples that they took the time to prepare and send, for the enzyme assay methods and advice on their execution, and their support via phone and email whenever it was needed. I would also like to thank my ‘normal healthy’ volunteers, Kim, Levi and Phil, for their time and their blood (literally). The suspension cell lines that were happily donated by Queensland Institute of Medical Research (QIMR) and Mater Medical Research Institute (MMRI) definitely added that extra dimension to my project. For their help with statistical analysis, I am indebted to Dr. Dale Nyholt (QIMR), Anna Coussens and Ian Hughes. A meeting with Dr. Greg Grabowski at the Society for Inherited Metabolic Disorders 2005 Annual Meeting initiated his offer of the TCP30 antibody which arrived home before I did and was crucial to finalising my project.

The research for this thesis took place within and was financially supported by the Cooperative Research Centre for Diagnostics (CDx). For the top-up scholarship that made financial survival easier and for the conference and travel support that facilitated valuable contacts and lifelong memories, thank you. I am indebted to my many CDx colleagues (Flavia, Jo, Shea, Anna, Erin, Trudi, John, Chris, Levi, Alex, Justin, Ian and
others who have already moved on) for your friendship and helpfulness and for providing a stimulating and fun environment in which to learn and grow. No laboratory can operate efficiently without people who do the essential but less rewarding tasks and for their happy and willing assistance. I also thank the laboratory and administrative support staff of the Science Research Centre.

My friends, Marianne and Janet, were invaluable for the revival of my sanity with our weekend chats.

My final thanks go to my family. For the great love of learning that is in me, I thank my Mum and Dad. It is truly one of the greatest gifts you could have given me beyond your love. A work of this magnitude has required enormous support, understanding, love, encouragement and sacrifice from my family. Clive, thank you for being there and for never asking why I had to do it or why it was taking so long. I couldn’t have done this without you. For my three sons, Michael, Glenn and Daniel, you make everything worthwhile.
CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW
1.1 INTRODUCTION

Gregor Mendel's 1866 paper on plant hybridization formed the basis for the modern study of genetics. By the beginning of the twentieth century, the pathways for the breakdown of certain amino acids had been worked out and so it was in 1909 that the concept of an ‘inborn error of metabolism’ was proposed by Garrod who perceived a relationship between genes and enzymes to explain the rare condition, alkaptonuria. To date, hundreds of genetic diseases caused by the deficiency of a single enzyme have been described (Scriver, 2001).

The first description of a lysosomal storage disorder was that of Tay-Sachs disease in 1881 although De Duve and his co-workers did not introduce the concept of lysosomes, being intracellular organelles, rich in hydrolytic enzymes until 1955 (De Duve et al., 1955). In 1963, Hers defined the characteristics of lysosomal storage diseases when he described the occurrence of the lysosomal enzyme acid maltase (acid $\alpha$-glucosidase) in normal human tissues as well as the absence of this enzyme in the liver, heart, and skeletal muscles of patients with Pompe’s disease (Hers, 1963).

Lysosomal storage disorders (LSD) are inborn errors of metabolism, caused by defects in one or more catabolic enzymes in the lysosome, resulting in the accumulation of undegraded natural substrate within the lysosome and perturbation of complex cell signalling mechanisms. This leads to an increase in the size and number of lysosomes in the cell, organelle disruption along with distortion of cell architecture and function. The ensuing biochemical changes, cellular damage and tissue responses are responsible for development of the symptoms seen in these disorders.

Over forty lysosomal storage diseases have been described. The disorders are mostly severe with a heterogeneous clinical spectrum encompassing visceral, skeletal and neurologic involvement leading to high morbidity and mortality. Progression is inevitable although the rate of progression is quite variable. The majority are caused by deficiency of a single lysosomal enzyme; others, for which I-cell is the paradigm, result from pleiotropic loss of several lysosomal enzymes because of an underlying defect in some common protein (Neufeld, 1991). Although individually rare, except among certain ethnic groups, LSDs have been estimated to have an overall prevalence of about one in 7000-8000 births (Meikle et al., 1999).
Most of the genes encoding the lysosomal enzymes have been cloned, and animal models have been obtained for most diseases. In the last few decades, LSDs have been used as models for the development of molecular and cellular therapies for inherited metabolic diseases. Studies in preclinical in vitro systems and animal models have allowed the development of bone marrow transplantation, substrate deprivation, enzyme replacement therapy and gene transfer methods as therapeutic options for several LSDs. These approaches have had quite variable rates of success while for cystinosis, pharmacological intervention with the use of cysteamine to facilitate removal of stored cystine has been very successful. Gaucher disease (GD), the most common LSD, and metachromatic leukodystrophy (MLD) are inherited as autosomal recessive traits as are most metabolic disorders. These two disorders have been selected as model diseases for this project as GD is the most successful and best characterized example of enzyme replacement therapy used to treat a LSD, while MLD is a good target for gene therapy and provides a simple and well characterized model to test some of our hypotheses about the role that single nucleotide polymorphisms (SNPs) play in the level of enzyme expression in the population and the implications for diagnosis and treatment.

1.2 THE LYSOSOME

1.2.1 The Endosomal-Lysosomal System

The lysosome is just one component of a series of seemingly unconnected intracellular organelles, collectively known as the endosomal-lysosomal system or vacuolar apparatus (De Duve and Wattiaux, 1966). As seen in Figure 1.1, the early endosome, the late endosome and the lysosome form a chain responsible for the catabolism of naturally endogenous and exogenous macromolecules and the subsequent recycling of their constituent monomeric components. The lysosome is characterised by the presence of a membrane, a low internal pH and vesicles containing many hydrolytic enzymes (Vellodi, 2005).
Figure 1.1 Lysosomal Enzyme Production

Plate A depicts an overview of lysosomal enzyme production from the synthesis of the inactive precursor enzyme in the rough endoplasmic reticulum (RER), the passage through the lumen of the endoplasmic reticulum (ER) to the Golgi apparatus and the movement to the lysosome after acquiring an address. Plate B depicts the phosphorylation of mannose residues on lysosomal enzymes. This two-step reaction is essential before a soluble lysosomal enzyme can acquire a mannose 6-phosphate (M6-P) ligand. Phosphorylation occurs via modification of the N-linked side chain by a phosphotransferase in the cis-Golgi and the subsequent action of a phosphodiesterase to remove the N-acetylglucosamine (GlcNAc) group, leaving a phosphorylated mannose residue on the lysosomal enzyme. Plate C depicts the mannose-6-phosphate (M6P) pathway, the major route for targeting lysosomal enzymes. Proteins are transported from the endoplasmic reticulum via membrane bound vesicles to the cis-Golgi apparatus for phosphorylation of mannose residues. Mistargeted enzymes are secreted from the trans-Golgi apparatus in secretory vesicles for exocytosis at the cell surface. In the trans-Golgi apparatus, phosphorylated enzymes bind to M6P receptor and are directed into vesicles coated with clathrin. Following degradation of the clathrin coating, the uncoated transport vesicle fuses with the late endosome. The enzyme is dephosphorylated and the M6P receptor returned to the Golgi before incorporation into another transport vesicle that takes the enzyme to the lysosome. Secreted phosphorylated lysosomal enzyme may also be endocytosed and transported to the lysosome via a similar pathway.
The membrane contains transport systems that carry particles between lumen and cytosol, and an electrogenic proton pump called the vacuolar proton pump, or V-type H\(^+\)-ATPase to acidify the internal lysosomal environment (Arai et al., 1993). Substrate breakdown occurs within the lysosome. It is also capable of secreting its contents after fusion with the plasma membrane (Luzio et al., 2000). In another important role for lysosomes, elevation in intracellular Ca\(^{2+}\) triggers fusion of lysosomes with the plasma membrane facilitating plasma membrane resealing. That is, the calcium-regulated exocytosis of lysosomes is very important for membrane repair of the cell (Reddy et al., 2001).

1.2.2 Synthesis and Trafficking of Lysosomal Enzymes (the ‘synthetic’ pathway)

Lysosomal enzymes are glycoproteins that are synthesised on the rough endoplasmic reticulum (ER) and are initially inactive. With the help of N-terminal signal sequences, the enzymes are translated through the ER membrane into the lumen of the ER where there is N-glycosylation at the asparagine residues within an Asn-X-Thr/Ser consensus sequence. However, not all potential N-glycosylation amino acid sequence sites are used. The signal sequence is trimmed before transportation to the Golgi compartment in membrane bound vesicles. Only the soluble proteins destined for the lysosome acquire a mannose 6-phosphate (M6-P) ligand (Kaplan et al., 1977) via sequential modification of the N-linked side chain by phosphorylation by a phosphotransferase (Reitman and Kornfeld, 1981, Waheed et al., 1981) and the action of a diesterase (Varki and Kornfeld, 1981, Waheed et al., 1981). If this essential ligand is not acquired, these soluble lysosomal enzymes are mistargeted and cannot enter the lysosome. Therefore, substrate breakdown cannot occur.

In general, the acquisition of the M6-P ligand separates glycoproteins that are destined for the lysosome from the secretory glycoproteins. The function of the M6-P ligand is to bind soluble protein to a membrane receptor. However, GBA is atypical in that it is not a transmembrane protein but is associated with the lysosomal membrane and although it does not acquire the M6-P ligand, it does undergo N-glycosylation and is targeted to the lysosome (Aerts et al., 1988). The precise mechanism by which this occurs is unknown, although several other lysosomal enzymes appear to use a similar process. Acid phosphatase is another membrane-associated protein that lacks M6-P residues and it is transported to the lysosome in a membrane bound form, immediately after reaching the trans-Golgii region (Tanaka et al., 1990).
The mannose 6-phosphate structure mediates binding to one of the two mannose 6-phosphate receptors, 300 kDa and 46 kDa (Chao et al., 1990). The two receptors appear to have little in common apart from a similar function and a 145 amino acid region of cysteine rich repeat units. However, this cysteine rich region does not define the mannose 6-phosphate binding capacity. It is thought that the two receptors bind overlapping yet distinct subsets of lysosomal enzymes. Association with the receptor occurs in the late Golgi compartment and the proteins are directed to the lysosome (Zhu and Conner, 1994). The enzyme-receptor complex exits the late Golgi compartment in clathrin-coated vesicles and dissociation occurs, facilitated by the low pH of the acidic pre-lysosomal (or endosome) compartment (Gonzalez-Noriega et al., 1980). The system is not fully efficient, and a fraction of the newly made enzyme is secreted (Neufeld, 1991).

The lysosomal enzyme moves into the lysosome and the receptor is recycled to the plasma membrane or to the Golgi to pick up another ligand. Within the lysosome, enzymes undergo maturation including dephosphorylation of mannose 6-phosphate, proteolysis, folding and aggregation. The acidic environment of the lysosome provides an optimal environment for the action of the lysosomal enzymes on their substrates.

1.2.3 Classification of Lysosomal Storage Disorders

Lysosomal disorders are classified according to the catabolic pathway affected and the deficient enzyme. For example, disorders associated with defects in lysosomal enzyme proteins are classified into the sphingolipidoses (Figure 1.2), cholesterol ester and triglyceride storage diseases, glycogen storage disease, glycoprotein storage diseases and mucopolysaccharidoses. Defective lysosomal enzyme phosphorylation and transport, defective sphingolipid activator proteins, defective cathepsin A and multiple enzyme deficiencies may cause other lysosomal disorders.

Sphingolipidosis are caused by defective catabolism of sphingolipids which are a component of all membranes but are particularly abundant in the myelin sheath. In Figure 1.2, the lysosomal enzymes included in this study (arylsulphatase A, β-glucocerbrosidase and α-galactosidase) and the sphingolipidosis for which they are responsible are indicated in red beside the arrow indicating the catabolic pathway which is interrupted as a result of the enzyme deficiency.
Lysosomal Catabolism of Some of the Sphingolipids

Figure 1.2 Figure 1.2  Lysosomal Catabolism of Some of the Sphingolipids

LSDs are classified according to the catabolic pathway affected and the type of enzymatic defect and/or stored substrate product. Sphingolipidoses are caused by defective catabolism of sphingolipids which are a component of all membranes but are particularly abundant in the myelin sheath. The lysosomal enzymes included in this study and the sphingolipidosis for which they are responsible are outlined in red beside the arrow indicating the catabolic pathway which is interrupted as a result of the enzyme deficiency.

Adapted from Vellodi, A., 2005
1.3 MOLECULAR GENETICS OF LYSOSOMAL STORAGE DISORDERS

All lysosomal storage disorders (LSDs) are single gene disorders and, with three exceptions, they are all autosomal recessive disorders. The exceptions are Hunter syndrome (mucopolysaccharidosis (MPS) type II), Fabry disease and the recently described Danon disease, which is inherited in an X-linked dominant manner (Sugie et al., 2003). The successful isolation of the genes encoding specific lysosomal enzymes and continuing investigations into the catalytic capacity of mutant enzymes have significantly increased our understanding of these disorders.

1.3.1 Clinical Presentation

A detailed description of all the known storage disorders is beyond the scope of this introduction but excellent reviews can be found in The Metabolic and Molecular Bases of Inherited Disease by Scriver et al. (2001).

For enzyme disorders, in general, the level of residual enzyme activity is a significant determinant of the severity of the phenotype of an affected individual. An early age of onset and severe disease presentation is generally associated with lower residual enzyme activity. It has been proposed that there was a ‘critical threshold’ of enzyme activity (Conzelmann and Sandhoff, 1983). This is a level sufficient to deal with the substrate flux but below this level there will be accumulation of substrate. It has also been demonstrated that small changes in residual enzyme activity can have a profound effect on the rate of accumulation of the substrate. However, it may not be the global level of activity, but rather the activity in specific compartments that is crucial.

Lysosomal enzymes and lysosomal enzyme deficiency can be found in every type of cell except the mature erythrocyte that lacks subcellular organelles. Hence, it is the tissue specificity of the substrate rather than of the enzyme that determines the tissues affected in lysosomal enzyme deficiency diseases (Neufeld, 1991). The consequences of lysosomal dysfunction can be widespread. These include bone deformities, joint stiffness, loss of skills such as speech and learning, behaviour problems, mental retardation, sight and hearing impairment, respiratory infection, heart disease and enlarged organs such as liver and spleen. In most cases, symptoms are progressive and eventually fatal.

The type of therapy considered is determined by which of the three predominant cell types are involved – neurological, mesenchymal and reticuloendothelial. Most LSDs have some
neurological involvement. However, in some disorders, it is present in all individuals e.g. MLD, while within disorders such as MPS 1 and GD, there are distinct groups, MPS 1-S and Type 1 respectively, that do not develop neurological disease. Mesenchymal involvement is universal in all the mucopolysaccharidoses and is also present in some Gaucher patients. Reticuloendothelial involvement is a hallmark of the sphingolipidoses e.g. Gaucher, Niemann-Pick and Fabry’s diseases. This group of disorders tends to respond best to therapy, as the reticuloendothelial cells are far more amenable to treatment than mesenchymal cells and neurones (Vellodi, 2005).

1.4 METACHROMATIC LEUCODYSTROPHY (OMIM 250100)

MLD is a fatal, autosomal recessive lysosomal storage disorder for which three major clinical classifications have been described, based on the age of onset and severity of disease: late-infantile, juvenile, and adult. As a lysosomal hydrolase, arylsulphatase A (ASA) is found in all tissues, although cerebroside sulphate, its primary substrate, is a sphingolipid abundant in myelin. Therefore, the intralysosomal storage caused by deficiency of ASA mainly affects oligodendrocytes (Kolodny and Fluharty, 1995) resulting in the various neurological symptoms, behavioural abnormalities and dementia, which characterise MLD. In Figure 1.3, the progressive nature of MLD is depicted in a young boy between the ages of five and nine years. While strong genotype-phenotype correlation exists, the presence of intragenic polymorphisms has been shown to modulate the expression of disease symptoms for a particular ARSA genotype (Harvey et al., 1998).

ASA pseudodeficiency (ASA-PD) is an apparent enzyme deficiency occurring in about 1% of healthy individuals (Nelson et al., 1991). The PD allele was one of the earliest characterised examples of polymorphisms that affect enzyme activity. The ASA deficiency in MLD and the reduced levels of ASA observed in PD are caused by allelic variants of the same gene (Chang and Davidson, 1983). In the heterozygous state, both result in a reduction in enzyme activity to approximately 50 percent of normal (Kolodny and Fluharty, 1995). However, based on enzyme levels alone, it is not possible to distinguish heterozygotes from homozygous normal individuals due to the large variation of ASA levels in the normal population. Therefore, molecular analysis must be used to demonstrate the presence of MLD or the PD allele.
Figure 1.3 Progressive Deterioration in Juvenile Metachromatic Leucodystrophy (MLD)

Photo A depicts the boy as a normal 5½ yr old. Photo B shows ataxia and difficulties in standing alone soon after diagnosis at 6-1/2 yrs. In photo C, the patient is bedridden at almost 9 yrs old. He can’t talk, is tube fed, but does recognise his family.
1.4.1 The Arylsulphatase A Gene (ARSA) (OMIM 607574)

ASA or cerebroside-3-sulphate 3-sulphohydrolase (EC 3.1.6.8), is one of more than ten evolutionarily related sulphatases found in human cells. The ARSA gene (Figure 1.4), located on the end of the long arm of chromosome 22, is about 3.2 kb. It contains eight exons (103 to 320 nucleotides in size) (Kreysing et al., 1990), and produces a mature enzyme of 489 amino acids.

By 1998, more than 40 mutations had been shown to cause MLD (Gieselmann et al., 1998) and many ASA polymorphisms had been described. The Human Gene Mutation Database (HGMD) now lists 105 mutations associated with MLD.

It has been shown that several polymorphisms reduce the amount of ASA activity in in vitro expression constructs (Harvey, 1996). This finding appears consistent with the severity of the disease seen in the MLD patients in which these polymorphisms occur. Hence, it was concluded that some polymorphisms contribute to the low activity seen in these disease alleles. More recently, Regis et al. (2002) reported that multiple mutations cause a greater reduction in ASA activity than do the corresponding single mutants. The total deficiency is generally likely to correspond to the sum of the reductions attributed to each mutation. Hence, each mutation may contribute to ASA activity reduction, and therefore, to the degree of disease severity. Pseudodeficiency mutations could therefore play a role in affecting clinical phenotype when they occur on the same allele as a disease-causing mutation (Regis et al., 2002).

However, there have also been reports of polymorphisms that may increase enzyme activity, such as the R496H polymorphism. The enzyme specific activity in a H496H individual was greater than that of any of the R496R individuals and R496H heterozygotes had intermediate enzyme levels, suggesting that the mutation may actually enhance activity compared to wild type (Ricketts et al., 1998).
The Arylsulphatase A Gene (ARSA)

Figure 1.4 The Arylsulphatase A Gene (ARSA)

The ARSA gene is located on chromosome 22. It is approximately 3.2 kb long and contains 8 exons that encode a mature enzyme of 489 amino acids.
1.5 GAUCHER DISEASE (OMIM 230800, 230900, 231000)

Mutations in the glucocerebrosidase (GBA) gene are responsible for the development of Gaucher disease (GD). This occurs when a deficiency of the lysosomal hydrolase, glucocerebrosidase (EC.3.2.1.45) results in the accumulation of glucocerebrosides (Beutler and Grabowski, 1995) primarily in the cells of the reticuloendothelial system. Spleen, liver and bone manifestations are therefore common in GD. The signs, symptoms, severity and rate of progression are quite variable. Neurological disease also occurs although this is rare. GD is the most common LSD with reported general population frequencies between 1 in 50,000 and 1 in 100,000. Although pan-ethnic, the highest frequency of 1 in 500-1000 births is encountered is in the Ashkenazi Jewish population (Grabowski et al., 1990). Estimated carrier frequencies are 1 in 500-1000 for the general population and 1 in 10-18 for the Ashkenazi Jewish population.

GD phenotypes have been arbitrarily classified into three types, based on the nature of effects on central nervous system (CNS), although Sidransky (Sidransky, 2004) suggests that GD is more correctly characterized as a continuum of phenotypes as there is clinical heterogeneity within each group. Type 1 or the non-neuronopathic variant is the most common, has no CNS involvement and accounts for about 94% of GD patients (Charrow et al., 2000). Clinical presentation may range from asymptomatic to fatal. Type 2 or infantile variant affects only about 1% of GD patients. There is severe CNS involvement and death in childhood with only minimal clinical heterogeneity. Type 3, affecting approximately 5% GD patients, has an adolescent or early adult onset with mild CNS involvement. It has a less heterogeneous presentation than Type 1 patients and is found with increased frequency in the Swedish (Norrbottnian) population. The most frequent mutation, N370S, accounts for 75% of alleles in Ashkenazi Jewish patients (Tsuji et al., 1988, Zimran et al., 1989, Theophilus et al., 1989) although it only accounts for 53% of GD alleles in the general population (Charrow et al., 2000). This mutation is not seen in patients with neuronopathic disease indicating that the residual enzyme activity is sufficient to protect against the most severe manifestations of the disease.

1.5.1 The Glucocerebrosidase Gene (GBA) (OMIM 606463)

The GBA enzyme (EC.3.2.1.45) is a 55 kDa protein for which the x-ray structure has been reported (Dvir et al., 2003). It is produced by the GBA gene located on chromosome 1 at 1q21 (Ginns et al., 1985). The gene is about 8 kb, consists of 11 exons and encodes a mature enzyme of 497 amino acids (Figure 1.5). It has two functional
ATG start codons (located in exons 2 and 3) that produce signal peptides, which differ in their hydrophobicity (Sorge et al., 1987). At least two GBA mRNAs, 2.6 and 2.2 kb in size, are the result of polyadenylation at different sites (Graves et al., 1986, Reiner et al., 1988b).

On the same chromosome, just 16 kb downstream from the functional GBA gene is a pseudogene (GBAP), approximately 5 kb in length (Horowitz et al., 1989, Zimran et al., 1990). The GBAP promoter was found to demonstrate activity when attached to a reporter gene (Reiner and Horowitz, 1988a) and initially this level of activity was found to be negligible (Horowitz et al., 1989). However, it was subsequently found to be consistently transcribed and this is sometimes at a level comparable to the active gene although a functional protein is not produced (Sorge et al., 1990). The two genes are tightly linked and there is 96% homology between GBA and GBAP. Hence, it is very important for any genetic analysis to distinguish between the two genes with several reports of PCR based methods that achieved this (Tayebi et al., 1996) (Finckh et al., 1998, Zhao et al., 2003).

According to The Human Gene Mutation Database (HGMD), 188 GBA mutations have been identified, including exonic missense and nonsense mutations, splice junction mutations, deletions and insertions of one or more nucleotides, and complex alleles resulting from gene conversion or recombination.
The Glucocerebrosidase Gene (GBA)

The GBA gene is located on chromosome 1 at 1q21. It is approximately 8 kb long and contains 11 exons that encode a mature enzyme of 497 amino acids and produce a protein of 55 kDa. Utilisation of two polyadenylation sites results in production of mRNAs of 2.2 kb and 2.6 kb in size.
1.6 GENOTYPE PHENOTYPE CORRELATION

Within the lysosomal storage disorders, there is a general relationship between the inherited mutations and the disease manifestations, that is, the most severe phenotype occurs when no functional enzyme is produced, while milder expression of phenotype occurs with mutations that do not completely abolish enzyme function. However, many unexplained discrepancies remain and mutations alone seem unable to completely predict the phenotypic expression. Phenotypic variation of a simple Mendelian condition (such as GD or MLD) within kindreds emphasizes the influence of modulating factors. These factors may include specific modifier genes (Rozmahel et al., 1996), the genetic background in general and the influence of the environment (Wolf, 1997).

Within both GD and MLD, there are examples of good genotype phenotype correlation. Just four genotypes (N370S/?, N370S/N370S, N370S/L444P, N370S/c.84insG) are found in the majority of GD type 1 patients. N370S appears to be protective of neurological consequences and most individuals with an N370S or S370S genotype have a mild clinical course and many are asymptomatic. The genotypes L444P and P444P are found in the majority of patients with GD type 3. However, there is still much unexplained clinical variability and this occurs to a greater extent in GD type 1, while GD type 2 shows the least variation.

Of the 105 mutations associated with MLD, two (459+1G>A and P426L) are the most frequent, each accounting for about 25% of MLD alleles (Polten et al., 1991, De Duve et al., 1955). ARSA alleles comprise quite a spectrum of mutations, ranging biochemically from a lack of enzyme activity to medium or even high residual activities and to a clinical spectrum ranging from severe MLD to PD without clinical findings (Berger et al., 1999). A strong correlation between genotype and phenotype is seen in MLD patients although there are some cases where this does not hold (Polten et al., 1991, Draghia et al., 1997).

It is proposed that at least some of the variation seen in GD and MLD is due to the presence of intragenic polymorphisms that modulate the expression of disease symptoms for a particular genotype. Recently, the E326K polymorphism found in GBA was described as a ‘modifier variant’ as it has never been found alone in a Gaucher disease-causing allele but does appear to decrease the activity of the allele on which it is present (Montfort et al., 2004).
1.6.1 Normal Levels of Enzyme

Within the normal population, there may be many variants that play a role in defining the observed range of enzyme activity. A histogram of the ASA activity in 894 individuals presenting to the Women’s and Children’s Hospital in Adelaide shows that there is a skewed distribution of activity around the mean (Figure 1.6a) (Mason, 2001). The distribution also shows that there is more than ten-fold variation in the leucocyte ASA enzyme activity between individuals at the extremes of the normal range (0.5 to >6.0 nmol/min/mg cell protein). The GBA activity in 701 individuals was also assessed and shows a similar distribution with more than five-fold variation between individuals with high and low activity (normal range: 600 to 3200 pmol/min/mg cell protein in leucocytes) (Figure 1.6b) (Mason, 2001).

In the population distribution of GBA enzyme levels shown in Figure 1.6b, there are a significant number of individuals with leucocyte GBA levels that are well above the population mean also creating a long ‘high activity tail’. A similar distribution is seen in cultured fibroblasts indicating that whole body and environmental effects are unlikely to be responsible for the observed variation in GBA activity. The original hypothesis on which this project was based, is that some of the variation seen in lysosomal enzyme activity levels in the normal population is due to polymorphic sequence variants and that some of these alter the primary amino acid sequence of the enzyme.

What causes some individuals within the normal population to have very high lysosomal enzyme activity levels when only a fraction of this amount is required? This wide variation in enzyme activity levels is tolerated as even high enzyme levels are still relatively low compared to most metabolic enzymes and higher levels of enzyme simply clears substrate at a faster rate. In both MLD and PD, it has been clearly demonstrated that naturally occurring polymorphisms within ARS-A alter the protein sequence and result in altered enzyme activity. The report of the E326K polymorphism supports this conclusion for GBA too. These variants may be the source of much of the variation observed in the normal population. However, for any single polymorphism to have a major contribution to the overall distribution of ASA or GBA levels in the population, it should occur relatively frequently.
Common to all forms of therapy (enzyme replacement, bone marrow transplantation and gene therapy) for lysosomal disorders is the need to deliver sufficient enzyme to the tissues of pathology. This level is determined by variables such as the targeting of the enzyme, tissue vascularisation, the blood brain barrier and ultimately, the circulating level of enzyme activity. A number of studies in vitro have suggested that only small amounts of additional lysosomal enzyme activity in deficient cells may be biologically effective and could correct the metabolic abnormality (Neufeld, 1991). Therefore, it would be of benefit to have an enzyme with higher activity levels so that whatever does reach the tissue will be more effective. The additional benefit of a reduction in cost of treatment may also be possible. The estimated cost of replacement enzyme for a typical adult with Gaucher disease can be as much as US$550 000 per year depending on the dosage (Clarke et al., 2001). Hence, any increase in enzyme activity that leads to a decrease in the required amount of enzyme replacement therapy would also be quite likely to reduce treatment costs.

It has been shown that there are polymorphisms within the ARSA gene that decrease the level of enzyme activity especially in MLD patients but also in PD. Hence, it was thought that polymorphisms may exist in the normal population, particularly within those individuals with very high enzyme activity levels, that may be used to enhance the level of enzyme activity obtained from an ASA or GBA construct. In support of this idea, it is known that polymorphisms exist within several other genes including the catechol O-methyl transferase (COMT) gene, the CYP2D6 gene, the promoter of the matrix metalloproteinase-1 (MMP-1) gene, the monoamine oxidase (MAO-A) gene and the microsomal epoxide hydrolase gene that result in or are related to increased enzyme activity (David et al., 2002, Bertilsson et al., 2002, Noll et al., 2001, Costa-Mallen et al., 2000, Harrison et al., 1999).

If high ASA activity sequence variants exist, they would be useful in the search for improved therapy for MLD. There is a need for expression of increased levels of total enzyme activity for both enzyme replacement therapy and gene therapy of MLD and other lysosomal storage disorders. This may be achieved by use of improved viral promoters in expression constructs or by expressing ASA protein sequence variants that result in higher levels of enzyme activity from a given quantity of expressed protein. Therefore, it is important to define the contribution of any polymorphisms that may increase the level of enzyme activity.
Figure 1.6 Population Distribution of ASA and GBA Activity Levels in Leucocytes

**Graph A** shows the population distribution of ASA activity levels in leucocytes. The normal range in the healthy population is 0.5 to >6.0 nmol/min/mg cell protein in leucocytes. Individuals with ASA levels <0.5 nmol/min/mg have metachromatic leucodystrophy. The mean ASA enzyme activity level is 2.62 nmol/min/mg protein and standard deviation is 1.03. **Graph B** shows the population distribution of GBA activity levels in leucocytes. The normal range in the healthy population is 600 to 3200 pmol/min/mg cell protein in leucocytes. Individuals with GBA levels below 500 pmol/min/mg have Gaucher disease. The mean GBA enzyme activity level is 1141 pmol/min/mg protein and the standard deviation is 376.
A. Leucocyte Arylsulphate A

ASA activity (nmol/min/mg protein)

B. Leucocyte β-glucocerebrosidase

GBA activity (pmol/min/mg protein)

high activity individuals
While some prior research has been done to define modifying polymorphisms within MLD and PD individuals, the first study involving individuals with high levels of ASA activity was done in this laboratory (Mason, 2001). The hypothesis was that polymorphisms in the ARSA gene can significantly affect enzyme activity and in part, result in the observed distribution of enzyme activity in the population.

1.6.2 High Activity Polymorphisms in ARSA and GBA?
Fourteen previously described polymorphisms were detected during screening of normal population samples for ‘high activity’ polymorphisms in ARSA (Mason, 2001). Fifty-two normal population samples were genotyped for ten of these polymorphisms, including four promoter (-789G>A, -720C>T, -498A>G, -351G>C), four coding (W193C, N350S, 1143C>T, T391S), one intronic (1102-32T>C) and a polyadenylation variant (2723A>G). The results from this small group of high and low activity samples showed that three of the promoter variants had a weak association with high activity. In addition, as has been previously demonstrated, the polyadenylation variant and W193C are significantly associated with low activity ($P = 0.026$) (Mason, 2001).

It is clear that a larger number of samples needed to be analysed to better establish an association between high ASA enzyme activity and the promoter variants, and also to determine if there is any correlation between the other ASA polymorphisms and enzyme activity level. In addition, with the possibility of high activity variants being present in ARSA, it would also be beneficial to determine if there are any high activity variants in GBA. There is also a need to examine whether GBA variants are responsible for the variation seen in the GBA enzyme activity levels in individuals.

These initial data warrant further investigation to determine whether some of the variation seen in lysosomal enzyme activity levels in the normal population is due to intragenic polymorphic sequence variants. Should high activity polymorphisms exist, they could be used to generate more enzyme activity from a construct for gene therapy or production of a recombinant protein.

1.6.3 Other Possible Causes of Variation of Expression in Individuals
High activity polymorphisms may be solely responsible for the higher enzyme activity levels seen in some individuals but as genotype/phenotype correlation in affected
individuals can be quite variable, it is likely that variation in gene expression, the influence of other genes and environmental influences could also be involved. A thorough examination of the expression of the ARSA and GBA genes in the normal population has not been done and may assist in further elucidating the inconsistencies in phenotype-genotype correlation observed between affected individuals.

To determine the exact nature of the variation in the expression of the ARSA and GBA genes requires further investigation. The size of the population genotyped for ASA polymorphisms needs to be increased and a significant correlation of genotype or haplotype with enzyme activity level must be determined to warrant continuing this line of investigation. For GBA, the tail of ‘high activity’ individuals is more extended, providing further hope of developing a very high activity recombinant enzyme. Hence, molecular analysis of the GBA gene will form the primary basis of my project. GD is the most common LSD and enzyme replacement and gene therapies are further developed than for other LSDs. The ability to increase enzyme activity is therefore likely to be useful immediately.

1.7 EXPRESSION OF THE GBA GENE

The GBA gene is often described in the literature as a ‘housekeeping’ gene due to its ubiquitous expression, but the expression of GBA would appear to be controlled at the transcriptional, translational and post-translational level. The only significant post-translational modification is the occupation of the first of five glycosylation sites that is essential to GBA activity (Berg-Fussman et al., 1993). While these glycosylation sites are necessary for the correct targeting of enzyme to the lysosomal compartment of the cell, they play little role in control of GBA expression and the final cellular phenotype. In general, protein expression levels are governed by the transcription rates of a gene, efficiency of nuclear export and mRNA localization (Lipshitz and Smibert, 2000), transcript stability (Belasco and Brawerman, 1993), translational regulation (Sonenberg et al., 2001) and protein degradation (Bochtler et al., 1999, Voges et al., 1999, Kirschner, 1999). For the GBA gene, it is the transcriptional and translational controls that have been found to have the major influence on mRNA, protein and enzyme activity levels. It has also been suggested that expression may be influenced by modifier gene(s) acting in cis, as has been proposed for the flanking genes at the GBA locus (Winfield et al., 1997) but no concrete evidence of this has been produced.
1.7.1 GBA Transcription

GBA has several notable features that distinguish it from the classical ‘housekeeping’ gene. The GBA promoter has been shown to possess two ‘TATA’ and ‘CAAT’ boxes (Reiner et al., 1988b), while housekeeping genes usually lack these and possess Sp1 binding sites (Dynan et al., 1986). In addition, the expression of the GBA gene has been found to be predominantly tissue-specific (Reiner and Horowitz, 1988a) which is also at variance with the term ‘housekeeping’ gene, but potentially consistent with manifestation of GD in a limited number of specific cell types and organs. That is, GBA is expressed ubiquitously but higher levels are also found in some tissues and cells. Tissue-, cell-, and developmental stage-specific variations in the expression of murine GBA have also been observed (Ponce et al., 2001).

GBA mRNA levels are lowest in B-cells and macrophages, higher in placenta, skin fibroblasts and promyelocytic lines and highest in epithelial cells (Reiner and Horowitz, 1988a). Skin fibroblast mRNA levels are ten times that of B cell mRNA levels, even within GD individuals (Reiner and Horowitz, 1988a). More recently, developmental and tissue-specific GBA mRNA expression has been documented especially within regions of the brain. It has also been shown that general visceral expression is much lower and ubiquitous except for higher levels in the epidermis of the skin (Ponce et al., 2001). This specificity appears to be promoter-directed as the levels of CAT expression are positively correlated with the amount of endogenous GBA mRNA in corresponding cell types (Reiner and Horowitz, 1988a).

Early findings indicated that enzymatic activity present in different cell types generally correlated well with GBA mRNA levels (Reiner and Horowitz, 1988a). However, it has since been shown that this correlation is present only in some cell types. The expression of the GBA gene is regulated with mRNA levels varying over 50-fold and activity levels over 60-fold in several human cell lines (Doll and Smith, 1993). Fibroblasts and brain derived cell lines had very high activity, lymphoblasts had very low activity, and the other cell types (epithelial cells, monocytes, histiocytes and promyelocytes) had intermediate levels (Doll and Smith, 1993). It appears that mRNA levels play a major role in regulating GBA activity but other factors are almost certainly also involved.
Three possible patterns of GBA expression have been identified (Doll and Smith, 1993):

a) a direct relationship between GBA enzyme activity and mRNA levels (epithelial, lymphoblasts, histiocytes, glioblastoma, astrocytoma);

b) a relationship between GBA enzyme and mRNA levels with enzyme activity levels being six-fold higher than expected when compared with the first group (promyelocytes, neuroglioma, fibroblasts);

c) a high level of mRNA but only an intermediate level of enzyme activity (monocytes).

The correlation between GBA enzyme activity and mRNA levels as seen in the expression pattern (a) identified by Doll and Smith (1993) is in agreement with the previous findings of Reiner and Horowitz (1988a). The two remaining expression patterns (b and c) indicate that mechanisms such as regulation of protein synthesis and degradation almost certainly play a significant role in the post-transcriptional regulation of GBA activity levels.

Tissue-specific expression of another lysosomal enzyme β-glucuronidase has been shown to result from changes in translational efficiency (Bracey and Paigen, 1987). Hence, it is apparent that both transcriptional and post-transcriptional controls of GBA expression are highly probable.

The existence of a feedback mechanism has been investigated to determine its role in regulation but the question remains unresolved. Reiner and Horowitz (1988a) found that GBA mRNA levels in Gaucher-versus non-Gaucher-derived cells were always more abundant in the Gaucher-derived cells than in the corresponding non-Gaucher counterparts suggesting sensitivity to the levels of actual enzymatic activity. However, earlier investigations had evaluated mRNA levels in HeLa cells and fibroblasts from normal and GD individuals and found equal concentration from all sources, in addition to normal amounts of the three poly (A)$^+$ mRNAs (5.6 kb, 2.5 kb, 2.0 kb) in fibroblasts from several subtypes and variants of GD (Graves et al., 1986).

More recent work also found similar levels of mRNA in both Gaucher and non-Gaucher individuals (Xu and Grabowski, 1998); (Doll and Smith, 1993), and that mRNA levels in cultured cells were unaffected by the GBA-specific inhibitor, conduritol-B-epoxide, indicating that regulation of GBA mRNA levels is unlikely to explain the feedback regulation of GBA activity observed in vivo in response to substrate load (Doll and Smith, 1993). Part of the answer to this problem may lie in the ability of some cells, such as
fibroblasts, to avoid substrate accumulation by relegating undegradable glucosylceramide into an anabolic compartment where it is converted into more highly glycosylated glycospingolipids (Saito and Rosenberg, 1985), while other cells such as macrophages lack this pathway.

All studies of the GBA promoter have confined the search for elements controlling transcription to about 622 bp of 5′-untranslated region (UTR) and flanking region (-354 to +254). This region was shown by structural and functional criteria to contain the minimal promoter element (Reiner et al., 1988b). Doll et al. (1995) used deletion constructs within this region and measured relative amounts of mRNA and CAT activity in five cell lines in which it had been previously shown that a direct relationship existed between GBA activity and mRNA levels. They found that elements within the 622 bp region examined specified differential tissue expression (Doll et al., 1995). However, the CAT activity from the lymphoblast cell line was a little higher than expected, indicating there may be more elements outside this region. However, an extended 5′-flanking sequence has not been investigated. There is approximately 6 kb between GBA and C1orf2, the gene immediately upstream of GBA and 4.7 kb of this sequence was analysed for further regulatory elements during this project.

The minimal promoter element was defined by Doll et al. (1995) who found that functional promoter activity is maintained in all cell lines with a construct containing only 70 bp upstream of the mRNA start site. In addition, activity is reduced to less than 10% in all cells when exon one sequences downstream of position +136 are removed while removal of sequences upstream of position –128 results in strong enhancement of activity for all human cells tested (Doll et al., 1995). Doll et al. (1995) concluded that all cell lines showed the same general pattern of activity and used a lesser number of cell lines for further evaluation. However, when compared with the epithelial and glioblastoma cell lines, the lymphoblast and astrocytoma cell lines show a significant reduction in activity with removal of additional upstream sequence between –128 and –70, and strong enhancement is seen in the lymphoblast cell line with removal of the downstream sequence from position +166. Further investigation of the lymphoblast cell line certainly appears warranted especially as lymphoblast GBA mRNA and activity is very low but Doll et al. (1995) concluded that differential expression of GBA is a result of subtle changes in
the magnitude of the effect of the different elements rather than the presence of elements active in only certain cell types (Doll et al., 1995).

Doll et al. (1995) initially investigated promoter activity in five different human cell lines and a murine cell line and concluded that the same regions generally act similarly in all cell types. Therefore, two cell lines; human epithelial and glioblastoma cells were selected for further analysis and again produced similar results. However, if the initial results are re-examined as has been done by depicting them graphically (Figure 1.7), it can be seen that the initial results for human epithelial and lymphoblast cells display differences for the p4.1-CAT and pHph-CAT constructs. More informative results may have been obtained if these two cell lines had been chosen for further analysis. It was also found that multiple sequence elements (both inhibitory and stimulatory) influence expression with the elements upstream of the TATA box being dispensable, while those within exon one were essential for reporter gene expression with at least two of these elements regulating mRNA levels and a subset of them likely to act as transcriptional enhancers (Doll et al., 1995). There is one element capable of binding a factor present in HeLa and glioblastoma cells but this factor was not identified and nor were any of the structural elements definitely identified.

Some of the transcription factors (TF) that regulate GBA expression have been identified (Figure 1.8). They include OCTA binding protein, AP-1, PEA3 and CAAT binding protein (Moran et al., 1997) and are found in many cell types, possibly explaining the ubiquitous expression of the GBA gene. An E box sequence was identified and the presence of B-cell specific OCT-2 was also indicated but no further investigation was done into the role of these elements. The AP1 site is located in downstream stimulatory region 4 (DSR4), the only identified transcription factor binding site to fall within one of the many inhibitory and stimulatory regions such as identified by Doll et al. (1995). OBP binds to the CAAT box and the other TF binding sites occur around the transcription start site. An attempt at defining the relative contribution of each of the identified transcription factors in SV80 cells found that mutations in the OBP, AP1, or PEA3 sites reduced activity to approximately 21% of normal, while loss of the CBP site had less effect, reducing activity to about 60% of normal (Moran et al., 1997).
Figure 1.7 Deletion Analysis of the GBA Promoter

The relative CAT activity of GBA deletion CAT constructs as reported by Doll et al. (1995) in Figure 1 has been re-analysed and presented graphically. CAT activity is expressed as the percentage of the value obtained for clone p622-CAT (= 100%). A series of deletions were made of p622-CAT by taking advantage of conveniently located restriction sites in the 5'-UTR and exon one.
Figure 1.8 The GBA Promoter Region

This 622 bp, Sac I to Sac I, fragment of the GBA promoter was analysed by Reiner and Horowitz (1988a), Doll et al. (1995) and Moran et al. (1997). It lacks SP1-binding consensus sequences but has two CAAT boxes and two TATA boxes. Transcription initiates at multiple sites (+1, -36, -11 and +61). Doll et al. (1995) identified several regulatory regions both upstream (upstream stimulatory region (USR), upstream inhibitory region (UIR)) and downstream (downstream stimulatory region (DSR), downstream inhibitory region (DIR)) of the +1 transcription initiation site: UIR1 (-354 to –318), USR1 (-220 to –204), UIR2 (-204 to –174), DSR5 (+10 to +56), DSR4 (+56 to +79), DSR3 (+85 to +135), DSR2 (+136 to +166), DIR1 (+166 to +184) and DSR1 (+184 to +224). Some regions in exon 1 were found to be essential for transcription. Differential expression was found to be the result of subtle changes in the magnitude of the effect of different elements rather than the presence of elements active in only certain cell types. The transcription factor binding sites: OBP (-99 to –92), E box, PEA3 (-7 to –4), CBP (-1 to +4) and AP-1 (+70 to + 76), were identified by Moran et al. (1997) who suggested that the availability of these transcription factors is one of the factors that dictates the level of transcription in different tissues.
On the basis of the intensity of the complexes formed between the appropriate promoter fragment and nuclear extract, it was found that the amount of OBP and AP1 was higher in transformed cells than primary fibroblasts and higher in Gaucher skin fibroblasts than non-Gaucher skin fibroblasts but comparable with the level in the amniotic, foetal fibroblasts (Moran et al., 1997). OBP and AP1 are gene activators involved in many processes such as development, differentiation, proliferation and transformation, while synergistic co-operation of OBP with CBP (Hinkley and Perry, 1992) and PEA3 with AP1 and other transcription factors (Kim et al., 1990, Oikarinen et al., 1987, Wasylyk et al., 1990) has been reported. Several factors including CBP bind to CCAAT sequences and may be involved in constitutive and differential expression. It was concluded that the availability of these transcriptional regulatory proteins is one of the factors that dictates the level of transcription in different tissues (Moran et al., 1997). However, there has been no definitive explanation of why or how these factors work to vary GBA mRNA levels between tissues.

It is clear that many questions still remain to be answered about transcription of the GBA gene. As indicated by Moran et al. (1997), availability of transcription factors is just one of the influences affecting the activity of the GBA promoter. Despite identification of nine inhibitory and stimulatory regions by Doll et al. (1995), only one of these has been shown to include a transcription factor binding site. All investigators have analysed the same region and yet there is very limited consensus between these studies. All studies appear to simply be addressing the structural components rather than carefully studying the elements that might be involved in tissue specific regulation. While the 622 bp region used by Doll et al. (1995) and others appears to be sufficient to specify differences in GBA expression in some cells, it remains unclear whether the region is sufficient to specify tissue-specific GBA mRNA levels in all cell types. Although this region contains some regulatory elements, it would seem prudent to do a more careful study of tissue regulation of the GBA promoter including a larger portion of the 5′−flanking region. Hence, it is proposed that a significant proportion of the sequence between GBA and C1orf2 be analysed for enhancer or inhibitor regions.

In addition to the transcriptional control, that clearly plays a role in determining GBA mRNA and enzyme activity levels, in some tissues there are anomalies where six times more enzyme activity is produced from an equivalent amount of mRNA. This possibly
indicates the presence of translational control of GBA mRNA. Therefore, one of the aims of my project is to investigate translation of the GBA gene.

1.7.2 GBA Translation

Translational control is a widespread mechanism for regulating gene expression. Within mammalian cells, translation may be regulated by the mRNA cap site accessibility, initiator codon and surrounding context (Kozak, 1991), cytoplasmic factors (Standart and Jackson, 1994), the 40S ribosomal subunit (Kozak, 1989), the coding region (Cao and Geballe, 1996) and the 3′-UTR and poly(A) region (Adam et al., 1986). These components can interact to alter the translational efficiency of cytoplasmic mRNAs and the resultant final rate of protein synthesis. Within the GBA gene, it is the initiator codon and surrounding context, and a cytoplasmic factor interaction with the coding sequence that have been found to play a role to date.

The GBA gene contains two in-frame start sites. In vitro studies of the translation of the GBA protein, involving oligonucleotide mutagenesis to remove one or the other, found that either ATG can function independently giving rise to either a 38 or 19 amino acid leader sequence (Sorge et al., 1987). However, the start site that is used in vivo remains unknown. Kozak’s ‘scanning theory’ suggests that the first ATG is most likely to be the preferred translation initiation site (Kozak, 1989) and although the sequences surrounding the second ATG match Kozak’s consensus sequence quite well (Kozak, 1984), it is possible that the second ATG is not used at all. The additional finding that protein synthesised from the first ATG is translocated more readily through the endoplasmic reticulum than its counterpart directed by the second ATG, resulting in 30% greater activity (Pasmanik-Chor et al., 1996) appears to add support to this hypothesis.

Variation in GBA translational efficiency was initially observed by Doll and Smith (1993) who found tissue-specific variation in GBA mRNA/activity ratios. When the over-expression of GBA in mammalian cells and Sf9 insect cells was evaluated by quantitative rather than the qualitative means used previously (Krall et al., 1996, Xu et al., 1995), a large discrepancy between the amounts of mRNA and GBA protein, that is, a 55- to 135-fold inefficiency, was observed between mammalian cell lines which was not evident in Sf9 insect cells (Xu and Grabowski, 1998). Cell-specific mistransferring, misfolding, transduction inefficiency and feedback inhibition were excluded as causes and lysosomal distribution and active site function were found to be normal.
The *in vitro* translational efficiency of endogenous GBA mRNA derived from human fibroblasts appeared to be twice as efficient as the *in vitro* translational efficiency of mRNA from transgenic sources leading to the suggestion that this may be due to 5′ and/or 3′ sequence differences. However, sequence differences would not account for the large translational inefficiency observed. Translation inhibition both in transgenic and in non-transgenic mammalian cells was found to be due to an 80 kDa heat-labile cytoplasmic protein that binds to the GBA mRNA coding region and is absent in Sf9 insect cells (Xu and Grabowski, 1998). This protein may be responsible for some or all of the tissue-specific variation in GBA mRNA/activity observed by Doll *et al.* (1995).

1.7.2.1 *The Translational Control Protein (TCP80)*

TCP80, an inhibitory translational control protein, was found to be responsible for the translational inefficiency observed in mammalian cells (Xu *et al.*, 2000). It binds within a region of 184 nucleotides near the 5′ end of the mature GBA mRNA. This *Hind* III-*Pst* I segment between nucleotides 246 and 429 is in the coding region and encompasses most of exon three and 17 bp of exon four (Figure 1.9). Higher levels of TCP80 were found in heart, skeletal muscle and pancreas than in lung, liver and brain tissues (Xu and Grabowski, 1999) and differences in TCP80 expression were also observed in a variety of human cells. This finding of tissue-specific variation in TCP80 levels is consistent with the suggestion that TCP80 may be at least partly responsible for tissue-specific variation in GBA translational efficiency.

*Trans*-proteins mostly act as negative regulators (translational repressors) primarily by controlling mRNA entry into polysomes (Standart and Jackson, 1994) and only a few of these proteins modulate RNA translation by interaction with RNA coding regions. The suggested mechanism of action for TCP80 is the prevention of engagement of GBA mRNAs with polysomes as it has been shown that TCP80 does not interfere with the continued translation of GBA mRNA that has already engaged polysomes (Xu and Grabowski, 1999). Hence, modulation of translation by TCP80 results in a change in the cytoplasmic distribution of GBA mRNAs that become segregated into two groups, active and inactive. The active group are those already engaged with polysomes being translated and the inactive group are bound to TCP80 and unavailable for initiation of translation.
**Figure 1.9** Cytoplasmic Protein (TCP80) Binding to the GBA RNA-coding Region

TCP80 binds within a 184 nucleotide *Hind* III-*Pst* I (H-P) segment overlapping exons two and three in GBA mRNA.

Adapted from Xu and Grabowski (1999)
TCP80 was found to interact with a number of other RNAs raising the possibility of competition for TCP80 binding. Significant homology of TCP80 with an M-phase phosphoprotein (MPP4; 99%) and an IL-2 enhancer binding protein (NF90; 96%) was observed (Xu and Grabowski, 1999). More recent studies indicate that TCP80, NF90 and probably MPP4 appear to be alternative splice variants from the Interleukin Enhancer Binding Factor 3 (ILF3) gene (Duchange et al., 2000) and may be members of an expanding family of dsRNA binding proteins that have significant homology of their N-terminus but differ significantly in their C-terminal regions (Reichman et al., 2002).

1.7.2.2 The Interleukin Enhancer Binding Factor Gene (ILF3)

The ILF3 gene that has been mapped to chromosome 19 (Marcoulatos et al., 1998), spans 38 kb and contains 21 exons (Duchange et al., 2000). According to OMIM, alternative transcripts previously described in the literature include MPP4 (M phase phosphoprotein), DRBP76 (Double-stranded RNA-binding protein), NF90 (Nuclear factor of activated T cells), and NFAR (Nuclear factor associated with double-stranded RNA). In the case of NF90, MPP4 and NFAR, the cDNAs have also been found to encode proteins of different sizes (Kao et al., 1994, Matsumoto-Taniura et al., 1996, Saunders et al., 2001).

The first full-length cDNA encoding ILF3 (referred to as DRBP76) was isolated by Patel et al. (1999) with the prediction of a 702 amino acid protein with a bipartite nuclear localization signal, two dsRNA binding domains, an arginine/glycine rich domain (RG2) and multiple potential phosphorylation sites. Further elucidation of ILF3 structure was provided by Duchange et al. (2000) with the identification of a major alternative splicing site in exon 17 responsible for the sequence divergence in the 3′ part of the transcripts and another alternative splicing event at a site between the two double stranded RNA binding motifs leading to the additional presence in some cases of a four amino acids NVKQ peptide (Figure 1.10). There is also utilisation of three distinct polyadenylation signals. The exons 12 to 15 encode the two double-stranded RNA binding domains (Duchange et al., 2000).
Figure 1.10 Alternative Splicing of the *ILF3* Gene Creates Five Transcripts

The *ILF3* gene consists of 21 exons represented by the numbered boxes. The translation start and three putative polyadenylation signals are indicated. The black boxes are the exons corresponding to the double-stranded RNA binding domain. The cDNAs for each alternate transcript is shown beneath the genomic map. An alternative splicing event in exon 14 results in the insertion of four additional amino acids. A major alternative splicing event in exon 17 at position 2330 results in the five transcripts that are depicted with the position of the stop codons indicated. TCP80 is most analogous to NF90 or *ilf3-d*. 
ILF3 and DRBP76, two major products of the ILF3 gene have been described (Figure 1.10) (Duchange et al., 2000). In addition, it is proposed that NF90 (Kao et al., 1994) differs at several nucleotides and may be a polymorphic variant or a mutant form of ILF3 (Duchange et al., 2000), while Tang et al. (2000) states that, although the first 600 residues are virtually identical, ILF3 and NF90 are distinct protein species (Kao et al., 1994, Buasas et al., 1999, Corthésy and Kao, 1994) with molecular weights of 110 kDa and 90 kDa, respectively. They can co-exist in the same cell. The cDNA sequence of MPP4 (Matsumoto-Taniura et al., 1996) terminates upstream of exon 17 and could be either ILF3 or DRBP76, while in TCP80, sequence divergence occurs after exon 17 and the coding sequence for the 77 amino acid specific C-terminal domain is in exon 21 (Duchange et al., 2000).

It would appear that the transcripts subsequently described are in fact splice variants of ILF3 although the actual role of the individual variants is unclear. The cDNA sequence of TCP80 is 96% identical to NF90 (Kao et al., 1994) and 99% identical to an M-phase phosphoprotein (Matsumoto-Taniura et al., 1996). It is possible that some or all of this variation is due to sequencing irregularities. NF90 has been shown to participate in cell division and cell cycle regulation (Kuang et al., 1989), transcriptional activation (Kao et al., 1994) and associate with both double stranded nucleic acid dependent protein kinases (Rice et al., 1989, Langland et al., 1999, Thomis et al., 1992, Chou et al., 1995, Patel et al., 1999, Ting et al., 1998). PKR, a double stranded RNA dependent protein kinase, modulates eIF-2α phosphorylation to regulate translation (Langland et al., 1999, Thomis et al., 1992, Chou et al., 1995, Patel et al., 1999). NF90 shares a consensus dsRNA binding sequence with PKR and other dsRNA binding proteins (Langland et al., 1999) and is phosphorylated by PKR (Langland et al., 1999, Thomis et al., 1992, Chou et al., 1995, Patel et al., 1999). Two NFAR variants (90 kDa and 110 kDa) were described and found to facilitate dsRNA regulated gene expression at the post-transcriptional level but the NFAR gene was also mapped to the ILF3 locus (Saunders et al., 2001).

M phase phosphoprotein (Matsumoto-Taniura et al., 1996) is the shortest of the transcripts described and it remains to be fully characterised. MPP4 is also reported to be involved in cell cycling. Although involvement in translational control had not been demonstrated previously, the two predicted double-stranded RNA-binding domains are present in a region of 100% homology between NF90, MPP4 and TCP80. All the
transcripts described are, in fact, part of a family of dsRNA binding proteins that contain one or more dsRNA binding motifs (dsRBM) (Fierro-Monti and Mathews, 2000) and have significant homology in their N-terminus but differ significantly in their C-terminal regions. ILF3 homologues in other species include ILF3 and p74 (rat); ILF3 and SPNR (mouse) and 4F:1 and 4 F.2 (Xenopus laevis) (Aoki and Zhao, 1998, Bass et al., 1994, Buaas et al., 1999, Coolidge and Patton, 2000, Matsumoto-Taniura et al., 1996, Patel et al., 1999, Schumacher and Lee, 1995, Tang et al., 2000, Xu and Grabowski, 1999).

To gain a better understanding of the mode of action of TCP80, a 681 bp fragment (TCP30) that included the two RNA-binding domains, and an N-terminal fragment (TCP50) were expressed and found not to inhibit translation either individually or together. Interestingly, this provided insight into the mechanism of action of TCP80 as both TCP30 and TCP80 bind to GBA mRNA but it is only the intact TCP80 that is capable of inhibiting translation (Xu and Grabowski, 1999).

Two forms of 80-90 kDa and 110 kDa have been reported for both MPP4 and TCP80 although the minor form was not present in all tissues (Xu and Grabowski, 1999). It is possible that the tissue-specific expression of TCP80 leads to the tissue-specific expression pattern observed for GBA. Expression of TCP80 was found to be highest in B-cells, neuroblastoma and embryonic kidney cells, while tissue levels were highest in heart, skeletal muscle and pancreas (Xu and Grabowski, 1999). Although GBA activity levels in B-cells are known to be very low and studies in Hep G2 cells have found that protein kinase C (PKC) phosphorylation of TCP80 enhances its translational inhibitory function (Xu and Grabowski, 2005), comparative studies of both GBA and TCP80 mRNA, protein or activity levels in a range of cell types needs to be done to further elucidate their relationship.

TCP80 has been shown to bind to at least seven other mRNAs. A lack of any obvious correlation between TCP80 and GBA activity levels may therefore indicate a more complex relationship with potential physiologic competition for TCP80 binding and modulation of translation by these other mRNAs. TCP80 appears to have significant specificity for binding particular mRNAs and does not function as a general translational suppressor (Xu et al., 2000). As the exact nature of this binding specificity has not been determined and as the possibility of the binding of the other ILF3 splice variants to GBA
has not been excluded, the prospect of regulation by the other ILF3 variants on a tissue-specific, developmental or cell-cycle basis needs to be considered. It seems likely that the different COOH-terminal regions of the ILF3 splice variants are responsible for the differences in the proteins.

It may be possible to produce a GBA gene construct for use in gene therapy. This construct would be selectively mutated, possibly through alternative codon usage, to prevent the binding of TCP80 without affecting activity level. There would be further indication for doing this if an inverse correlation is found between GBA and TCP80 levels. Alleviation of the large translational inhibition could lead to secretion of very high levels of GBA in large excess of that currently available.

As has been discussed earlier, polymorphisms in ARSA can affect enzyme activity and gene function. Screening the TCP80 gene for SNPs could give an insight into the possibility that any SNPs present may affect the RNA binding or translation inhibition functions culminating in higher GBA activity in some individuals.

1.8 EXPRESSION OF THE ARSA GENE

Very little has been published on the regulation of ASA expression probably because, along with most of the lysosomal enzymes, it is considered a ‘housekeeping gene’ and is expressed in all tissues examined (Kreysing et al., 1990). The upstream region of ARSA does not include a TATA box but does have a very high GC content with potential SP 1 binding sites, an arrangement common to acid phosphatase (Geier et al., 1989), α-glucosidase (Martiniuk et al., 1990), the β subunit of β-hexosaminidase (Neote et al., 1988) and other housekeeping genes. The ASA promoter region includes two transcription initiation sites between –367 and –387, a CpG island and four potential Sp 1 binding sites. Three different mRNA species of 2.1, 3.7 and 4.8 kb are transcribed from the gene and probably arise from the use of different polyadenylation signals (Kreysing et al., 1990). Consistent with this, no reports of transcriptional regulation have been published.

Translational regulation of ASA mRNA has been reported in mouse testis. There is an increase in ASA mRNA during spermatogenesis but this is not accompanied by an increase in enzyme activity or ASA polypeptides in whole testis or isolated spermatocytes and spermatids (Kreysing et al., 1994). A protein specifically expressed in the testis binds
to the 5′- and 3′-UTRs of ASA mRNA. The same protein binds both ends of the mRNA and it is suggested that the nature of the translational control may vary depending on whether it is the 5′- or 3′- end of the mRNA involved. Binding of the protein to the 5′- end of ASA mRNA may suppress translation, whereas binding of the same protein to the 5′- end may cause stabilisation of the mRNA (Kreysing et al., 1994). An example of this has also been reported for intracellular control of iron levels (Casey et al., 1988, Hentze et al., 1987). As ASA transcription does not appear to be increased, high levels of ASA mRNA may be due to a stabilization of the mRNA (Kreysing et al., 1994).

1.9 SUMMARY AND RELEVANCE TO PROJECT

GD and MLD are lysosomal storage diseases characterised by a remarkable degree of heterogeneity. Genotypic variation is seen between individuals with the same clinical manifestations and phenotypic variation is seen between siblings with the same mutational genotype. Across the population of unaffected individuals, the range in GBA and ASA enzyme activity levels varies five- to ten-fold. The role of polymorphisms in modifying gene expression has been demonstrated. The initial hypothesis of this project was that some of the variation seen in lysosomal enzyme activity levels in the normal population is due to polymorphic sequence variants and some of these alter the primary amino acid sequence of the enzyme. These variants may be the source of much of the variation observed in the normal population. However, for any polymorphism to have a major contribution to the overall distribution of ASA or GBA levels in the population, it should occur relatively frequently. Should high activity polymorphisms exist, they could be used to express more enzyme activity from a construct for gene therapy or production of a recombinant protein.

Regulation of GBA expression at the transcriptional and post-transcriptional levels has been observed. The tissue specific expression of GBA is consistent with its promoter structure and three different expression patterns have been described (Doll et al., 1995). A number of structural (inhibitory or stimulatory) elements were identified (Moran et al., 1997) but the reasons for tissue specificity and exactly how it is achieved remains to be elucidated. The translational control protein (TCP80) inhibits GBA via interaction with the coding region, thereby affecting the efficiency of translation and the final level of enzyme activity available. The overall objective of this study is to investigate the
expression of GBA and ARSA to further understand the determinants of enzyme activity and how it may be increased.

Further elucidation of the mechanisms of control of GBA expression should clarify the determinants of an individual’s GBA enzyme activity level while also increasing understanding of the pathogenesis of GD. This may have important implications for diagnosis and treatments, including enzyme replacement and gene therapies.

1.10 SPECIFIC AIMS
The overall aim of this project was to investigate the causes of variation in lysosomal enzyme activity levels and more specifically, the cause of high enzyme activity levels in some individuals. Information gained from the identification and characterisation of high activity, natural sequence variants could be used to improve diagnosis, phenotype prediction and to produce greater protein activity from a recombinant protein expressed in vivo for gene therapy and in vitro for enzyme replacement therapy.

The specific aims of this thesis were:

1. (a) For ASA, to genotype the sample population for the ASA polymorphisms had been identified previously. For GBA, to screen individuals with high GBA enzyme activity for polymorphisms and then genotype the sample population for these GBA polymorphisms. To assess the correlation of these polymorphisms with enzyme activity level to identify high activity polymorphisms. If high activity polymorphisms were found to exist, this information was to be used to express more enzyme activity from a construct for gene therapy or production of a recombinant protein.

(b) To investigate the contribution of non-genetic factors to the variation in enzyme activity levels.

(c) To determine whether there is variation in the mRNA levels of lysosomal enzymes between normal individuals.

2. To characterise the expression of GBA and ARSA in human cell lines and identify sources of transcriptional regulation in the GBA gene.
3. To ascertain the translational efficiency of *GBA* and *ARSA* in human cell lines and identify any correlation in *GBA* expression with TCP80/ILF3 production.

The significance of this research is that knowledge of the contribution of genetic variants and expression control to the natural variation seen in these genes will lead to advances in disease diagnosis, better understanding of genotype-phenotype correlation, enhanced knowledge of molecular pathogenesis of the disease and improved therapeutic outcomes.
CHAPTER 2

MATERIALS AND METHODS
2.1 INTRODUCTION
General materials and methods used in this thesis are outlined in detail below. In the
individual chapters, specific materials and methods relevant to each chapter are described.

2.2 GENERAL REAGENTS AND CHEMICALS
All general reagents and chemicals of analytical grade were obtained from Ajax Chemicals
(Melbourne, Australia), BDH Chemicals (Kilsyth, Australia) or Sigma Chemical Company
(Castle Hill, Australia), unless otherwise stated.

2.3 LYSOSOMAL ENZYME ACTIVITY ASSAYS
2.3.1 Isolation of Peripheral Blood Leucocytes
Leucocytes were separated from erythrocytes by differential sedimentation at unit gravity
in a dextran/saline solution. The cells were washed briefly in 0.2% saline to haemolyse
the red cells, and then brought back to isotonicity with 1.8% saline (Standard Operating
Procedures, Women’s and Children’s Hospital, Carey and Nelson, 2001). Approximately
4-5 ml of freshly collected whole EDTA blood was aliquoted into a 15 ml Falcon tube
and centrifuged at 2100 x g for 5 min at 10°C. The plasma was removed carefully and
discarded, leaving the buffy coat undisturbed. Normal saline (0.9%) was added to give a
final volume of 8 ml followed by 2 ml dextran/saline (dextran, MW 150-200 x 10³, BDH,
0.5g/L; NaCl 0.007g/L in de-ionised water). The tubes were mixed carefully and bubbles
removed. The uncapped tubes were left to stand at room temperature for approximately
45 min to 1 h. The leucocyte containing supernatant was transferred to a fresh 15 ml
Falcon tube and centrifuged at 1500 x g for about 2 min at 10°C. To wash the pellet, it
was resuspended in 4 ml of 0.2% saline and mixed by pipetting with a plastic Pasteur
pipette. After precisely 1 min, 3.2 ml of 1.8% saline was added and mixed gently before
centrifuging at 1500 x g for 2 min at 10°C. The pellet was washed again to obtain a white
cell pellet free of red cells. All supernatant was removed and the pellet frozen at -80°C
until assayed which was usually within one month. At least two pellets per sample were
prepared. The QUT Human Research Ethics Committee approved these studies
(approval number: 1080/1H).

2.3.2 Preparation of a Cell Pellet for Enzyme Assay
The cell pellet (Materials and Methods 2.3.1 and 2.11.1.1) was resuspended in 1.0 ml of
0.1% (w/v) Triton X-100 (BDH). It was freeze-thawed 6 times by alternating the tube, at
approximately 3 min intervals, between a dry-ice/ethanol bath and a room temperature
water bath. These extracts were centrifuged at 2100 x g for 5 min at 10°C and the supernatant put into a clean tube in an ice bath for assay.

2.3.3 Assay of Arylsulphatase A Enzyme Activity in Human Cells

All cell extracts (Materials and Methods 2.3.2) were tested in duplicate. A blank was prepared for each test. Cell extract (100 µl) was aliquoted into the ‘test’ tube and 500 µl p-nitrocatechol sulphate substrate (0.02056 g/L sodium pyrophosphate, 3.116 g/L p-nitrocatechol sulphate, 100 g/L NaCl in 0.5 M acetate buffer: pH 5.0) was added to both ‘test’ and ‘blank’ tubes at timed 10 s intervals. Both tubes were incubated at 37°C for exactly 60 min. At timed 10 s intervals, 0.5 ml of 1.0 M NaOH was added to ‘test’ and ‘blank’ tubes. After mixing well, 100 µl leucocyte extract was added to the ‘blank’ tube.

Duplicate standards were prepared by mixing 50 µl of 500 µM p-nitrocatechol standard, 50 µl of water, 500 µl of 11.4 mM p-nitrocatechol sulphate substrate (Sigma) and 500 µl of 1.0 M NaOH. The absorbance of each tube was read at 515 nm in a 1.0 cm light path cuvette on a spectrophotometer (Pharmacia LKB Ultrospec II).

The blank reading was subtracted from each tube and the arylsulphatase A enzyme activity calculated as nanomoles per minute per milligram of protein (nmol/min/mg protein) using the following calculation.

\[ \text{ASA enzyme activity}^* = \frac{\text{Absorbance}}{0.0113 \times \text{time (min)} \times \text{vol (ml)} \times \text{protein (µg/ml)}} \times 1000. \]

*Based on the molar extinction coefficient for p-nitrocatechol of 12400, 1 nmol/ml will produce an absorbance change of 0.0113 in the assay volume of 1.1 ml.

(Baum et al., 1959)

2.3.4 Assay of β-Glucosidase Enzyme Activity in Human Cells

All cell extracts were tested in duplicate and duplicate blanks were prepared with each test run. Cell extract (100 µl) was aliquoted into the ‘test’ tube and 100 µl of 0.9% saline was added into the ‘blank’ tube. At timed 10 s intervals, 100 µl of 20 mM 4-methylumbelliferone-β-glucoside substrate (9.8 g/L glutathione, 2.0 g/L sodium taurocholate, 6.76 g/L 4-MU-β-glucopyranoside, 18.62 g/L potassium chloride in citrate-phosphate buffer: pH 4.5) was added and the tubes were incubated at 37°C for exactly 60
min. The reaction was stopped by the addition of 1.5 ml of glycine buffer (15.02 g/L glycine, 13.25 g/L sodium carbonate, 6.38 g/L NaOH: pH 10.7). Duplicate 4-methylumbelliferone standards (4MU) (2.84 nmol) were prepared by mixing 20 µl of 4MU standard (142 mM), 180 µl of distilled water and 1.5 ml of glycine buffer (pH 10.7). The relative fluorescence of each tube was read on the Perkin-Elmer LS-50B spectrofluorimeter with the excitation wavelength/slit width at 366 nm/2.5 nm, and the emission wavelength/slit width at 446 nm/2.5 nm. The relative fluorescence readings (RF) were corrected by subtracting the blank reading from each tube and glucocerebrosidase enzyme activity calculated as picomoles per minute per milligram of protein using the following equation.

\[
\text{GBA enzyme activity}^* \left( \text{pmol/min/mg protein} \right) = \frac{RF}{10} \times \frac{1}{\text{time (min)}} \times \frac{1000}{\text{vol (µl)}} \times \frac{1000}{\text{protein (µg/ml)}} \times \frac{28.4}{\text{std RF value}} \times 1000
\]

*20 µl of the 4-methylumbelliferone (4MU) standard (0.142 mol/l) contains 2.84 nmol 4MU.

(Kolodny and Mumford, 1976)

2.3.5 Assay of α-Galactosidase Enzyme Activity in Human Cells

All cell extracts were tested in duplicate and duplicate blanks were prepared with each test run. Cell extract (50 µl) plus 50 µl of 0.9% saline were aliquoted into the ‘test’ tube and 100 µl of 0.9% saline was aliquoted into the ‘blank’ tube. At timed 10 s intervals, 100 µl of 10 mM 4-methylumbelliferone-α-galactoside substrate (3.375 g/L 4-MU-α-galactoside in 0.15 M acetate buffer: pH 4.5) was added and the tubes were incubated at 37°C for exactly 60 min. The reaction was stopped by the addition of 1.5 ml of glycine buffer (15.02 g/L glycine, 13.25 g/L sodium carbonate, 6.38 g/L NaOH: pH 10.7). Duplicate 4-methylumbelliferone standards (4MU) (2.84 nmol) were prepared by mixing 20 µl of 4MU standard (142 mM), 180 µl of distilled water and 1.5 ml of glycine buffer (pH 10.7). The relative fluorescence of each tube was read on the Perkin-Elmer LS-50B spectrofluorimeter with the excitation wavelength/slit width at 366 nm/2.5 nm, and the emission wavelength/slit width at 446 nm/2.5 nm. The relative fluorescence readings (RF) were corrected by subtracting the blank reading from each tube and α-galactosidase enzyme activity calculated as nanomoles per minute per milligram of protein using the following equation.
\[ \alpha\text{-GAL enzyme activity}^* \ (\text{nmol/min/mg protein}) = \text{RF} \times \frac{1}{10} \times \frac{1}{\text{time (min)}} \times \frac{1}{\text{vol (ml)}} \times \frac{1000}{\text{protein (\mu g/ml)}} \times \frac{28.4}{\text{std RF value}}. \]

*20 \mu l of the 4-methylumbelliferone (4MU) standard (0.142 mol/l) contains 2.84 nmol 4MU.

(Desnick et al., 1973)

2.4 ASSAY OF PROTEIN CONCENTRATION IN CELL EXTRACT

The Bio-Rad Protein Assay (Bio-Rad Laboratories, Regents Park, Australia), which is based on the method of Bradford, was used to assay protein concentration in each leucocyte extract. The standard procedure for microtitre plates was used. A series of eleven bovine serum albumin standards from 0.0 mg/ml to 0.5 mg/ml was prepared. Standards were assayed in duplicate and cell extracts were assayed in triplicate. Briefly, one part of the dye reagent concentrate was diluted with four parts of distilled water, 10 \mu l of each standard or cell extract was pipetted into the appropriate wells followed by addition of 200 \mu l of diluted dye reagent to each well and the extract and the reagent were mixed thoroughly. The microtitre plate was incubated at room temperature for 50-55 min and absorbance at 595 nm read on the Biomek plate reader (Beckman, Sydney, Australia). Cell extracts that had protein concentrations greater than 0.5 mg/ml were diluted with 0.1\% (w/v) Triton X-100 and re-assayed.

2.5 DNA PREPARATION

2.5.1 Genomic DNA Extraction

Genomic DNA was extracted from the leucocyte pellets obtained from the Women’s and Children’s Hospital, Adelaide using the DNeasy Tissue Kit (QIAGEN, Doncaster, Australia) as per the manufacturers instructions for animal tissues. DNA samples were normalised by dilutions between 1:10 and 1:100, dependent upon DNA concentration of individual samples.

2.6 THE POLYMERASE CHAIN REACTION (PCR)

All PCR reactions were performed in a MJ Research PTC-200 Peltier thermal cycler (Bresatec, South Australia). Using 0.2 ml thin-walled PCR tubes, the PCR reactions were carried out in a final volume of 20 \mu l or 50 \mu l containing the reagents as listed and a negative control was included. The amplification parameters that were used are as outlined in each experiment.
2.6.1 Electrophoresis of PCR amplicons

To separate PCR amplicons by molecular size, electrophoresis of all PCR amplicons was carried out on 0.8%, 1% or 2% agarose gels. The agarose was dissolved in 1 X TBE buffer (90 mM Tris-borate, 2 mM EDTA) by microwaving on the high setting for 1-2 min followed by cooling to approximately 50°C. Ethidium bromide (100 µg/L) was added and the gel left to set for about 30 min before the PCR amplicons were loaded. To monitor the movement of amplicons in the gel system, the PCR products (5 µl or 10 µl) were mixed with 2 µl loading dye (0.25% bromophenol blue, 30% glycerol). This dye runs at approximately 40 bp and therefore did not obscure the amplicons being tested. The electrophoresis was carried out at 120 V in a Biorad Minigel System (Biorad, Sydney, Australia) for 40 min and the gel was visualised using a UV transilluminator (λ = 302 nm) and the image captured on a Geldoc system (UVP, England).

2.7 DNA SEQUENCING

Sequencing of PCR amplicons required purification of the amplicon using QIAquick (QIAGEN, Germany) PCR purification kit according to the manufacturer's instructions. Sequencing reactions were performed using 2–5 µl purified PCR product, 3.2 pmol primer (forward or reverse) and 4 µl ABI Dye 2 Terminator mix in a total volume of 20 µl. Sequencing of DNA products was performed by mixing approximately 500 ng plasmid DNA miniprep sample with 1 µl of BigDye 3 Terminator mix (Applied Biosystems, Melbourne, Australia) and 3.2 pmol of gene specific primers (see individual experiments) in a final volume of 12 µl of nuclease-free water.

Amplification was performed in an MJ Research PTC-200 Peltier thermal cycler for 25 cycles of 94°C for 30 s, 50°C for 15 s, 72°C for 4 min. Each sample was precipitated with the addition of a 1/10 volume of sodium acetate (3 M; pH 5.2) and 2.5 volumes of 100% ethanol followed by incubation for 15 min at room temperature. The sample was centrifuged at 14000 x g for 20 min and the supernatant removed. The DNA pellet was washed in 70% ethanol and centrifuged again at 14000 x g for 5 min. The supernatant was removed and the DNA pellet was dried. Automated fluorescent sequencing using the ABI BigDye Terminator was performed at the Australian Genome Research Facility, University of Queensland, Brisbane, Australia. Sequence alignments were performed.

2.8 STATISTICAL ANALYSIS
Statistical analysis of the ARSA genotype data was performed using a number of different programs to compare and expand on the analysis. Within the SPSS statistical program, the functions used were the compare-means one-way ANOVA to generate the mean ASA activity level for each genotype, one-sample Kolmogorov-Smirnov test to assess the normality of the data, linear regression to enter the individual SNPs to assess correlation with enzyme activity, and the forward linear regression to analyse all the SNP data together to determine which SNPs were most correlated with enzyme activity. The Haploview program (Barrett et al., 2005) was used to look at linkage disequilibrium (LD) for the ASA SNPs. QTPHASE is one of the programs in UNPHASED which is a suite of programs for association analysis of multilocus haplotypes from unphased genotype data (Dudbridge, 2003). This was used as a more sophisticated analysis to compare results with the SPSS analysis. SNPSpD was used to establish to experiment wide significance thresholds and do association analysis for comparison with the other programs (Nyholt, 2004).

2.9 CELL TYPES USED
The following cell lines; glioblastoma (U118), chondrocyte (SW1353), hepatoma (Hep G2), hepatoma (Huh7), lymphocyte (B-cells), monocyte (THP1), promyelocyte (HL60) and 4 individual skin fibroblast lines (SF3351, SF3542, SF3693, SF3785), were obtained from the Queensland Institute of Medical Research (QIMR), Mater Medical Research Institute (MMRI), Queensland University of Technology (QUT) and Women’s and Children’s Hospital (WCH).

Fresh leucocytes were also isolated from freshly collected whole EDTA blood (Materials and Methods 2.3.1) that was collected from four volunteers as required.

2.10 TISSUE CULTURE
All cell lines used were cultured at 37°C in 5% CO₂ in an IR Sensor Incubator (Sanyo, Quantum Scientific, Brisbane, Australia). The SW1353 cell line was cultured in Dulbecco’s Modified Eagle’s Medium (DMEM) (Gibco BRL Life Technologies, Melbourne, Australia) with 10% foetal calf serum (FCS) (Life Technologies)
supplemented with 1% Antibiotic-Antimycotic (Life Technologies). All other cell lines were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco BRL Life Technologies, Melbourne, Australia) with 10% FCS (Life Technologies) supplemented with 1% Antibiotic-Antimycotic (Life Technologies). Culture medium was changed at two to three day intervals and cell morphology and viability was checked via microscopic examination and regular mycoplasma testing.

2.10.1 Cell Counting

Cells were grown to confluency and detachment facilitated by rinsing with 2-3 ml of trypsin EDTA (Gibco BRL Life Technologies, Melbourne, Australia). A further 2 ml trypsin EDTA (Life Technologies) was added, detached cells washed in fresh growth medium and centrifuged at 1000 x g for 5 min in a bench top centrifuge. Following aspiration of the wash medium, a single cell suspension was prepared in the appropriate growth medium, 10 µl of the cell suspension was added to one side of the haemocytometer cell counting chamber and four square areas counted and averaged. The average cell count was multiplied by the dilution factor and by 10⁴ to determine cells per ml.

2.11 USE OF CELLS

2.11.1 Enzyme Activity and mRNA Quantification

To quantify GBA or ASA enzyme activity and mRNA levels of a cell line, 6 T80-cm² culture flasks per cell line were prepared. Culture medium was changed after 48-72 h. Three flasks were assayed after 72-96 h culture and the remaining flasks were assayed 48 h later. For each flask for each cell line, GBA and ASA enzyme activity levels were assayed (Materials and Methods 2.3.3 and 2.3.4) and GBA, ASA and ILF3 mRNA levels measured (Materials and Methods 2.12). To measure mRNA levels, total RNA from each cell type (Materials and Methods 2.12.1) was reverse transcribed into complementary DNA (cDNA) using the 1st Strand cDNA Synthesis Kit for reverse transcriptase-PCR (AMV) (Materials and Methods 2.12.2). Real-time PCR was used to perform absolute quantification of GBA, ASA and ILF3 mRNA levels in all cells. Analysis of results and production of charts was done in Microsoft Excel.
2.11.1.1 Preparation of Each Flask of Cultured Cells for Assay of Enzyme Activity and RNA Extraction

To collect attached cells (SW1353, Hep G2, Huh7, U118, SF) for enzyme assay, the lower portion of each T80-cm² culture flask of near confluent cells was scraped with a sterile cell scraper and the cells rinsed into solution with fresh medium. The detached cell suspension was transferred to a 15 ml falcon tube and centrifuged at 400 x g for 5 min. The supernatant was discarded and the cells placed on ice for the minimum amount of time until preparation of the cell pellet for enzyme assay (Materials and Methods 2.3.2), usually within one hour.

The remaining cells were collected for total RNA extraction by direct addition of 2 ml of TRIZOL reagent to the T-80cm² culture flask with gentle agitation. The lysed cells were mixed well, and 1 ml aliquots placed into 1.5 ml eppendorf tubes and frozen immediately on dry ice. These TRIZOL/cell preparations were stored at -80°C until required for total RNA preparation (Materials and Methods 2.12.1).

To collect suspension cells (HL60, THP1, B-cells) for enzyme assay and total RNA extraction, the cells from a T80-cm² culture flask were transferred into two 15 ml Falcon tubes and centrifuged. The supernatant was removed from both tubes and the cells to be used for enzyme assay were placed on ice. The cells in the remaining tube were lysed in 2 ml of TRIZOL reagent and 1 ml aliquots placed into 1.5 ml eppendorf tubes and frozen immediately on dry ice. These TRIZOL/cell preparations were stored at -80°C until required for total RNA preparation (Materials and Methods 2.12.1).

2.12 QUANTIFICATION OF mRNA

2.12.1 Total RNA Preparation

Total RNA was extracted all cell lines using TRIZOL (Life Technologies) as instructed by the manufacturer's protocol. Confluent T-80cm² culture flasks were used and cells were collected in TRIZOL reagent, mixed well to lyse the pellet followed by incubation for 5 min at room temperature. To separate the RNA, chloroform (0.2 ml/ml TRIZOL) was added, the sample shaken vigorously for 15 s, incubated at room temperature for 15 min and centrifuged at 12000 x g for 15 min at 4°C. Isopropanol (0.5 ml/ml TRIZOL) was used to precipitate the RNA from the aqueous phase after it was removed. The sample was incubated at room temperature for 10 min and centrifuged at 12000 x g for 15 min at...
4°C. The RNA pellet was washed once with 75% ethanol (1 ml/ml TRIZOL) by vortexing and centrifuging at 7500 x g for 5 min at 4°C. The ethanol was removed and the RNA pellet was air dried for 10 min and redissolved in 30 µl sterile Tris-borate-EDTA (TBE) (108 g/L Tris, 55 g/L boric acid, 9.3 g/L EDTA).

Total RNA from leucocytes was extracted using the PAXgene blood RNA kit as instructed by the manufacturer’s protocol. Blood was collected in the PAXgene blood RNA tube and incubated overnight at room temperature to increase yield. The tube was centrifuged for 10 min at 5000 x g. The supernatant was removed and 5 ml RNase-free water added to the pellet, which was vortexed to resuspend and then centrifuged for 10 min at 5000 x g. The pellet was resuspended in 360 µl of buffer BR1. After transferring the suspension into a 1.5 ml eppendorf tube, 300 µl of buffer BR2 and 40 µl proteinase K were added. The tube contents were vortexed, incubated for 10 min at 55°C with occasional mixing and then centrifuged for 3 min at maximum speed in a microcentrifuge (Hermle 2160M). The supernatant was transferred to a fresh 1.5 ml eppendorf tube and mixed with 350 µl of 100% ethanol. The solution was transferred to a PAXgene column, centrifuged for 1 min at 8000 x g and the eluate was discarded. Sequential additions of 700 µl of buffer BR3 and 2 x 500 µl of buffer BR4 were applied to the column followed by centrifugation for 1 min at 8000 x g. An additional centrifugation step was also used to minimise buffer BR4 carryover. Total RNA was eluted from the PAXgene column with two additions of 40 µl buffer BR5 each followed by centrifugation for 1 min at 8000 x g. The eluate was heated for 5 min at 65°C and chilled on ice.

The RNA from both cell lines and leucocytes was electrophoresed on a 1% agarose to ascertain the integrity of the sample and the concentration was measured by spectrophotometric analysis (Pharmacia LKB Ultrospec II) where 1 OD at A_{260} is equal to approximately 40 µg RNA. To monitor the purity of the sample, the 260 nm/280 nm ratio was recorded. A ratio approaching 2.0 indicates a pure sample that is free from protein and solvent contamination.

2.12.1.1 Removal of Genomic Contamination

Samples that showed evidence of genomic DNA contamination underwent DNase treatment to remove it. RQ1 DNase (1 U/1 µg contaminant DNA) was added to RNA
sample and incubated for 15 min at 37°C. RNA was extracted with 1 volume of TE-
saturated phenol:chloroform:isoamyl alcohol (25:24:1; pH 4.5) by vortexing for 1 min and
centrifuging for 2 min at 12000 x g.

2.12.2 Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)
One µg of total RNA was reverse transcribed into complementary DNA (cDNA) using
the 1st Strand cDNA Synthesis Kit for RT-PCR (AMV) as instructed by the manufacturer
(Roche Applied Science, Castle Hill, Australia). Total RNA (1 µg) was mixed with
reaction buffer (1x), magnesium chloride (MgCl₂) (5 mM), deoxynucleotide mix (dATP,
dTTP, dCTP, dGTP; 1 mM), a random primer p(dN)₆ (0.096 A₂₆₀ units; 3.84 µg), RNase
inhibitor (60 units), AMV Reverse Transcriptase (24 units) and sterile water to make a 24
µl reaction volume. The reaction was incubated at 25°C for 10 min and 42°C for 60 min.
To inactivate the enzyme, the reaction was heated to 99°C before cooling to 4°C and
storage at -20°C.

2.12.3 The Real-Time Polymerase Chain Reaction
The ABI real-time PCR system (Applied Biosystems, Scoresby, Australia) is based on the
detection and quantification of SYBR Green, which is a fluorescent dye that binds to
double-stranded DNA but not to single-stranded DNA. Reactions are characterized by
the point in time during cycling when amplification of a PCR product is first detected
rather than the amount of PCR product accumulated after a fixed number of cycles. The
higher the starting copy number of the nucleic acid target, the sooner a significant increase
in fluorescence is observed. It is used to quantify PCR amplification as it occurs, allowing
accurate and precise measurements to be made in the exponential phase of the PCR
reaction.

2.12.3.1 Design of Real-Time PCR Primers
Primer Express is a primer design program from Applied Biosystems that facilitates
design of primers for real-time PCR. Primers for GBA, ARSA and ILF3 were designed
using Primer Express (Applied Biosystems) and also analysed using Netprimer from
The primer sequences and fragment length can be seen in Table 2.1.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer Name</th>
<th>Primer 5’ to 3’</th>
<th>Product Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S</td>
<td>18SF</td>
<td>5’-TTCGGAACCTGAGGCCATGAT-3’</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>18SR</td>
<td>5’-CGAACCTCCGACTTTCGTTCT-3’</td>
<td></td>
</tr>
<tr>
<td>ARSA</td>
<td>ASA141F</td>
<td>5’-ACCCCAGAGGTGCTGCAA-3’</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>ASA199R</td>
<td>5’-GCCTCTAACTGGGAGCTGAG-3’</td>
<td></td>
</tr>
<tr>
<td>GBA</td>
<td>GBA43F</td>
<td>5’-GGGCTTCTGGAGACAAATCTC-3’</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>GBA100R</td>
<td>5’-GCAGCGCCACAGGATGAGG-3’</td>
<td></td>
</tr>
<tr>
<td>ILF3</td>
<td>ILF3F1</td>
<td>5’-GCAGGGAGGATTTTG-3’</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>ILF3R1</td>
<td>5’-CAGAGTTGCTCGGTACCATA-3’</td>
<td></td>
</tr>
</tbody>
</table>

F = forward and R = reverse primer sequences. The size of each fragment is shown in base pairs (bp).
The \textit{GBA} primers (Table 2.1) amplify both \textit{GBA} and \textit{GBAP} mRNA. However, a representative group of total RNA preparations was checked for pseudogene contamination with primers specific for GBAP. The level of GBAP mRNA was found to be less than ten percent and in most cases, it was much less than five percent.

\textbf{2.12.3.2 Standard Curve Generation}

Standard curves were prepared for each gene under study for absolute RNA quantification. The cycle number or cycle threshold (Ct) is inversely proportional to the starting amount of target cDNA; therefore the number of target gene copies can be extrapolated from a standard curve equation.

To prepare a standard curve, each gene of interest was amplified from human genomic DNA using the appropriate primers in a standard PCR reaction. The product was precipitated with the addition of a 1/10 volume of sodium acetate (3 M; pH 5.2) and 3.0 volumes of ice-cold 100% ethanol and incubation at 4°C for 30 min. Following this, the sample was centrifuged at 14000 \(x\) g for 20 min and the supernatant removed. The DNA pellet was washed in 70% ethanol and centrifuged at 14000 \(x\) g for 5 min. The supernatant was removed and the DNA pellet was resuspended in sterile distilled water. The amplicon was electrophoresed on a 2% agarose gel to ascertain the integrity of the sample and the concentration determined by spectrophotometric analysis (Pharmacia LKB Ultrospec II) where 1 OD at A\textsubscript{260} is equal to approximately 50 \(\mu\)g DNA. To monitor the purity of the sample, the 260 nm/280 nm ratio was recorded. A ratio approaching 1.8 indicates a pure sample that is free from protein and solvent contamination.

The number of copies is calculated on the basis of the molecular weight of each individual gene amplicon. A serial dilution series of this reference amplicon was used to generate a gene specific standard curve that extended above and below the expected abundance of the target RNA. Real-time PCR was performed on both the experimental samples and the relevant standard curve serial dilutions. An absolute value for target abundance in each experimental sample was extrapolated from the standard curve generated from the serial dilutions.
2.12.3.3 Real-Time PCR Method

To enable absolute quantification of each target mRNA species, a standard curve for the gene of interest (Materials and Methods 2.12.3.2), an endogenous control (ribosomal 18S RNA) and negative controls were run in each experiment. 18S ribosomal RNA was used as an endogenous control for normalisation of the differences in the amount of total RNA added to each reaction and correct for sample to sample variations in real-time PCR efficiency.

For each sample of cDNA, triplicate tubes were prepared for the gene of interest and the 18S ribosomal RNA that was used to normalise the results. In addition, a standard curve for the gene of interest (Materials and Methods 2.12.3.2), five replicates of 18S ribosomal RNA reaction from the control cDNA reaction, and negative reactions for the gene of interest and 18S were run on each plate. Each real-time PCR reaction contained 11 µl of SYBR Green Master Mix, 100-200 pmol of forward and reverse primers and 0.88 µl cDNA reaction (Materials and Methods 2.12.2) in a total volume of 22 µl. Each test reaction was run in triplicate. The default thermal cycling parameters used were 50°C for 2 min, 95°C for 10 min, then 40 cycles of 95°C for 15 s and 60°C for 1 min. The results for each run were analysed at the threshold at which the lowest average standard deviation of cycle time occurred for the sample and 18S reactions in that run. All sample 18S values were normalised to the control cDNA 18S values and the absolute quantity of mRNA of the gene of interest calculated for each sample based on the normalised 18S value for that sample. For each cell type, replicates were averaged and standard error calculated. All analysis was done in Microsoft Excel.

2.13 CLONING

In order to study the regulatory regions of the GBA promoter, a number of GBA promoter PCR amplicons were cloned into the pGL3-basic and pGL3-promoter luciferase reporter vectors (Promega) (Chapter 5), transformed into high efficiency JM109 competent E. coli cells (Life Technologies) and screened for by an internal amplicon specific PCR and sequencing.

2.13.1 Preparation of Bacterial Fosmid DNA

The fosmid, G248P86705B7, was used as template to amplify the GBA 5’-UTR region to avoid contamination with the GBA pseudogene (GBAP) sequence. The fosmid was
provided as a stab from The Wellcome Trust Sanger Institute, Cambridge, United Kingdom. Using a sterile loop, the stab was sampled and streaked onto a Luria Bertani (LB)/chloramphenicol agar plate (containing 10 g Bacto-tryptone, 5 g Bacto-yeast extract, 5 g NaCl and 15 g agar per litre plus a final concentration of chloramphenicol of 25 µg/ml) and incubated at 37°C overnight. A single isolated colony was picked, placed into 10 ml of LB/chloramphenicol broth (containing 10 g Bacto-tryptone, 5 g Bacto-yeast extract and 5 g NaCl per litre plus a final concentration of chloramphenicol of 25 µg/ml) and incubated for 6-8 h at 37°C with shaking at 225 rpm. This was followed by addition of 2.5 ml of the young culture into 250 ml LB/chloramphenicol broth and incubated overnight at 37°C with shaking at 225 rpm.

The fosmid DNA was extracted using the QIAGEN Midiprep protocol for very low-copy plasmid/very low-copy cosmid. Briefly, the culture was centrifuged at 6000 x g for 15 min at 4°C and the pellet resuspended in 20 ml of buffer P1. Buffer P2 (20 ml) was added and the mixture inverted six times and left to stand at room temperature for 5 min. Buffer P3 (20 ml) was added and the mixture inverted six times, left on ice for 30 min, and then centrifuged at 20000 x g for 30 min at 4°C. The supernatant was removed promptly and centrifuged at 20000 x g for 15 min at 4°C. The supernatant was again collected and the DNA precipitated with the addition of 170 ml isopropanol at room temperature followed by centrifugation at 15000 x g for 30 min at 4°C. The DNA pellet was dissolved in 500 µl TE (pH 7.0) and 4.5 ml of QBT buffer added. The DNA solution was applied to an equilibrated QIAGEN-tip 100 and allowed to enter the resin. Following washing with 2 x 10 ml of buffer QC, the DNA was eluted with 5 ml of buffer QF. The DNA was re-precipitated with isopropanol at room temperature and centrifugation, followed by washing with 70% ethanol at room temperature and centrifugation at 15000 x g for 10 min. The pellet was air dried and re-dissolved in sterile de-ionised water. The concentration was determined spectrophotometrically, where 1 OD at 260 nm was equated to approximately 50 µg/ml of DNA.

2.13.2 Preparation of the pGL3-Basic Reporter Vector For the Ligation of GBA Promoter Constructs

To prepare the pGL3-basic vector for ligation of the GBA promoter constructs, the vector was digested as follows. A 20 µl reaction consisting of 2 µg of pGL3-basic vector DNA, 2 µl of R* buffer, both 1 µl of HindIII and 1 µl of Xho I or 1 µl of HindIII
(Progen, Brisbane, Australia) was incubated at 37°C for 60 min. One µl of digested sample and control (uncut) were electrophoresed to determine the efficiency of the restriction digest. The pGL3-basic vector/HindIII digest reaction was de-phosphorylated by addition of 1 µl of calf intestinal phosphatase (CIP, Roche), 10 µl of 10 x CIP buffer in a 50 µl reaction at 37°C for 60 min. The reaction was stopped by heating the sample to 65°C for approximately 60 min. The sample was then purified through a QIAGEN DNA column as described previously. DNA sequencing was performed as described in 2.7 with GBA specific primers (Chapter 5).

2.13.3 Preparation of the GBA Promoter PCR Products For Ligation Into pGL3 Reporter Vectors
To prepare the PCR products (Materials and Methods 4.2.3) amplified by the “proof reading” enzyme, Pfx (Invitrogen Life Technologies, Melbourne, Australia), for cloning into the pGL3-basic or pGL3-promoter vectors, the amplicon was digested as follows. A 50 µl reaction consisting of 15 µl GBA promoter amplicon PCR reaction, 5 µl of R+ buffer, both 1 µl of HindIII and 1 µl of Xho I or 1 µl of HindIII (Progen, Brisbane, Australia) was incubated at 37°C for 60 min. The sample was then purified through a QIAGEN DNA column as previously described.

2.13.4 Ligation of PCR Amplicon Into the pGL3 Reporter Vector
The PCR products that had been digested and purified were cloned into the pGL3 reporter vector (Promega) by following the manufacturer’s instructions. Briefly, the purified PCR amplicon, 1 µl 10 x ligation buffer, 1 µl pGL3-basic vector (30 ng) or 1 µl pGL3-promoter vector (30 ng), 1 µl T4 DNA ligase (3 Weiss units/µl), and sterile water to make a total volume of 10 µl, were mixed and incubated overnight at 4°C. The amount of purified PCR amplicon added was calculated such that the insert:vector molar ratio was maintained at 3:1.

2.13.5 Transformation of pGL3-Basic or pGL3-Promoter Vectors into JM109 High Efficiency E. coli Competent Cells by Heat Shock
Ten µl of the mixture containing ligated pGL3-basic or pGL3-promoter vector (containing the PCR amplicon of interest) plus 50 µl of JM109 bacterial cells (Life Technologies) were incubated on ice for 30 min. The JM109 bacterial cells were then transformed by heat-shocking at 42°C for 45 s. This was followed by incubation in 800 µl
of SOC medium containing 2.0 g Bacto-tryptone, 0.5 g Bacto-yeast extract, 1 ml of 1 M NaCl, 1 ml of 2 M MgCl$_2$ and 1 ml of 2 M glucose and shaking at 225 rpm for 60 min at 37°C.

2.13.6 Plating of Transformation Cultures onto LB/Ampicillin Plates
The incubated transformation culture was centrifuged at 14000 x g for 5 min and most of the resulting supernatant was removed. The bacterial pellet was resuspended in about 100 µl of supernatant and spread out onto LB/ampicillin agar plates (final concentration of ampicillin 100 µg/ml) and incubated at 37°C overnight.

2.13.7 Identification of Positive Clones
The pGL3-basic vector does not contain the lacZ gene that is involved in lactose metabolism; therefore, use of a blue/white selection process was not possible. Instead, an internal insert specific PCR (for the PrC fragment (Materials and Methods 3.2.2.1) was performed on multiple colonies and the PCR product electrophoresed on a 2% agarose gel for 30 min. Presence of a band for the 190 bp fragment in both the positive control (B7 fosmid DNA used) and the colony PCR is indicative of the presence of a promoter amplicon insert in the clone. The positive clones were picked and cultured in 5 ml of LB /ampicillin broth (final concentration of ampicillin 100 µg/ml) for 16 h at 37°C with shaking at 225 rpm. The samples were then centrifuged for 5 min at 3000 x g to collect the bacterial cell pellets and prepared for plasmid extractions as described below (Materials and Methods 2.13.8). A screening PCR could not be performed for potential clones in which the amplicon insert was too small to contain the PrC fragment. Several clones were therefore cultured and plasmid extracted as outlined below, followed by direct sequencing.

2.13.8 Extraction of Plasmid DNA Containing Inserts From the Bacterial Cell Pellets
The QIAprep Spin Miniprep Kit was used as outlined by the manufacturer (QIAGEN) to extract plasmid DNA.

2.14 REPORTER GENE ASSAYS ON THE GBA PROMOTER
In order to study the regulatory regions of the GBA promoter, a number of GBA promoter PCR amplicons were cloned into the pGL3-basic or pGL3-promoter luciferase reporter vector (Promega) (Chapter 5). These constructs were then transfected into the
U118, HepG2, Huh7 and SW1353 cell lines and assayed for reporter signals as described below.

2.14.1 Transfection of GBA pGL3 Promoter Constructs Into Attached Cell Lines

The U118, HepG2 and Huh7 cell lines were transfected with the pGL3-GBA promoter constructs and assayed for GBA promoter activity. Twenty-four well plates were used and the cell lines were seeded at approximately $2.0 \times 10^5$ cells per well in Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco BRL Life Technologies, Melbourne, Australia) with 10% FCS supplemented with 1% Antibiotic-Antimycotic (Life Technologies) and incubated at 37°C with 5% CO$_2$ in an IR Sensor Incubator (Sanyo, Quantum Scientific, Brisbane, Australia) until the cells were 90% confluent. Two µg of GBA pGL3-basic or GBA pGL3-promoter vector plus 10 ng of control Renilla vector (pRL-TK, Promega) was diluted in 50 µl of Opti-MEM-1 reduced serum medium (Life Technologies) and 3 µl of Lipofectamine 2000 (LF2000, Life Technologies) in 50 µl of Opti-MEM–1 was added and mixed. The sample was left at room temperature for approximately 30 min.

The SW1353 cell line was also transfected with the pGL3-GBA promoter constructs and assayed for GBA promoter activity. The SW1353 cells were seeded into 24-well plates at approximately $1.5 \times 10^5$ cells per well in Dulbecco’s Modified Eagle’s Medium (DMEM) (Gibco BRL Life Technologies, Melbourne, Australia) with 10% FCS supplemented with 1% Antibiotic-Antimycotic (Life Technologies) and incubated at 37°C with 5% CO$_2$ in an IR Sensor Incubator (Sanyo, Quantum Scientific, Brisbane, Australia) until 50-80% confluent. To 100 µl of Opti-MEM-1, 0.6 µl Fugene 6 (Roche) was added slowly and incubated at room temperature for 5 min. Two hundred nanograms of pGL3-GBA plasmid plus 10 ng of control Renilla vector (pRL-TK, Promega) was also added and a second incubation of 45 min at room temperature followed.

DNA/transfection reagent mix (100 µl) was added to each well and incubated at 37°C for 20-24 h. A pGL3 control vector (2 µg) that contains a constitutively active SV40 promoter was also transfected into the cells by the above methods. After 24 h, the cells were assayed for luciferase production (Materials and Methods 2.14.2) or fresh growth medium added and the samples incubated for a further 24 h. After 48 h, for samples
undergoing two-day serum stimulation, 1 ml of fresh growth medium was added and the samples incubated for a further 24 h.

2.14.2 Reporter Gene Assays
At the end of the required incubation period, the cells were rinsed with 1 x phosphate buffered saline (PBS) (pH 7.4) before 100 µl of passive lysis buffer (PLB, Promega) was added to each well and the cells were lysed by gentle shaking at room temperature for 15 min. Then, 20 µl of the PLB lysate from each well was aliquoted into a 96-well Optiwell plate (Perkin Elmer OptiPlate™-96). The plate was inserted into the Polar Star Optima luminometer (BMG Labtech) which was programmed to add 100 µl of LARII (Luciferase assay reagent, Promega) and measure the luciferase activity followed by addition of 100 µl of Stop & Glo reagent (Promega) and measurement of the Renilla luciferase activity for each well. The Renilla vector was used as an internal transfection control. The level of luciferase activity is normalised by presenting it as a ratio to the level of Renilla luciferase activity.
CHAPTER 3

EXPRESSION VARIATION OF HUMAN
GLUCOCEREBROSIDASE AND ARYLSULPHATASE A
AND HIGH ACTIVITY SNPS
3.1 INTRODUCTION

It is well established that single nucleotide polymorphisms (SNPs) can affect gene expression and enzyme activity (Harrison et al., 1999). For example, there is a common TATA box SNP in the CYP2A6 gene that causes large inter-individual differences in levels of P450 enzyme and the CYP2D6*17 allele causes altered substrate specificity, while the PD allele in ARSA causes ASA pseudo-deficiency (ASA-PD) (Fukami et al., 2004, Gieselmann et al., 1989). As described in Chapter 1, there are a significant number of individuals with high ASA and GBA enzyme activity, that is, those with levels that are well above the population mean. It was hypothesised that some of the variation seen in lysosomal enzyme activity levels in the normal population is due to polymorphic sequence variants.

This chapter will describe the screening of the ARSA and GBA genes in individuals with high enzyme activity to identify SNPs that result in high levels of expressed enzyme activity. In previous work, the analysis of the ARSA gene in 52 individuals was performed and fourteen previously described polymorphisms were detected (four promoter, five coding, four intronic and one in the poly(A) signal) (Mason, 2001). Three of the promoter variants (-351C>G, -498A>G and -720C>T) appeared to identify alleles associated with high activity, that is, there were significant associations of certain genotypes with high activity samples. One of the intronic polymorphisms (1101+123C>T) seemed to be more frequent in high activity samples but no further investigations were done. The polyadenylation variant (2723A>G), -789G>A and W193C polymorphisms were associated with low activity (Mason, 2001).

It was clear that a larger sample size was necessary to help establish if there was any significant correlation of the variants with high or low enzyme activity levels and that more in-depth statistical analysis should also be performed. This work was completed as part of this PhD project.

Aside from polymorphisms, it is possible that other non-genetic factors may affect lysosomal enzyme activity levels. In an attempt to determine the extent of the effect of these other factors, a pilot study was done measuring GBA, ASA and α-galactosidase (α-GAL) enzyme activity levels in four individuals over three consecutive weeks.
3.2 MATERIALS AND METHODS
The materials and methods specific to this chapter are described in detail below. General materials and methods used in this thesis are outlined in detail in Chapter 2.

3.2.1 Design of Amplicons for ARSA Gene
Ten fragments of less than 550 bp each were designed to encompass all 8 exon structures and adjacent splice sites and the promoter region of ARSA. In addition, two fragments, ASA Part 1 and ASA Part 2 were designed to encompass exons 1 to 4 and exons 4 to 8, respectively (Figure 3.1).

3.2.1.1 ASA Fragment PCR
All PCR reactions were carried out as described in Material and Methods 2.6.

PCR reactions for fragments 1 to 8 contained 1-5 µL of DNA template, 3 mM MgCl₂, 20 mM Tris-HCl pH 8.4, 50 mM KCl, 0.2 mM dNTPs, 0.5 µM forward and reverse primers (Table 3.1) and 1 unit Platinum Taq DNA polymerase (Life Technologies) per 50 µL reaction volume. PCR reactions for promoter region fragments, fragment 9 and parts 1 and 2 contained 1-5 µL of DNA template, 1.75 mM MgCl₂ (part 2) or 2.25 mM MgCl₂ plus detergents (part 1), 20 mM Tris-HCl pH 8.4, 100 mM KCl, 0.4 mM dNTPs, 0.5 µM forward and reverse primers, and 2.5 units of Expand Long Template enzyme mix (Roche) per 50 µL reaction volume.

Thermal cycling conditions for fragments less than 600 bp were 96°C for 5 min, followed by at least 30 cycles of 96°C for 30 s, annealing at various temperatures for 30 s (Table 3.1), 72°C for 30 s, then 72°C for 10 min followed by a ramp from 95°C to 65°C over 30 min. Thermal cycling conditions for ASA parts 1 and 2 were conducted following instructions for the Expand Long Template PCR system (Roche) with 95°C denaturation, annealing at various temperatures (Table 3.1) and initial elongation times of 60 s for part 1 and 70 s for part 2.
The Arylsulphatase A Gene (ARSA)

Ten fragments of less than 550 bp each were designed to cover all ASA exons, the promoter and 3'-UTR regions. Two fragments were also designed to amplify the gene in two parts for use in dot-blot hybridisation experiments.

Figure 3.1 Arylsulphatase A Fragments for PCR

Ten fragments of less than 550 bp each were designed to cover all ASA exons, the promoter and 3'-UTR regions. Two fragments were also designed to amplify the gene in two parts for use in dot-blot hybridisation experiments.
### Table 3.1 Primer Sequences, Annealing Temperatures and Product Sizes for PCR of the Arylsulphatase A Gene

<table>
<thead>
<tr>
<th>Fragment Name</th>
<th>Primer Name</th>
<th>Primer 5′ to 3′</th>
<th>Annealing Temp °C</th>
<th>Product Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASA Pr1</td>
<td>Pr1F</td>
<td>5′-GCACCCAGCCACAGCTTACAGCA-3′</td>
<td>74.6</td>
<td>549</td>
</tr>
<tr>
<td></td>
<td>Pr1R</td>
<td>5′-GAAGCCGCTAGAGGGGAGCCACAGAGGA-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA Pr2</td>
<td>Pr2F</td>
<td>5′-GTTCGCTAGAGGGGAGCCACAGACA-3′</td>
<td>69.0</td>
<td>533</td>
</tr>
<tr>
<td></td>
<td>Pr2R</td>
<td>5′-GAAGAGCTAGAGGGGAGCCACAGAGGA-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 1</td>
<td>1F</td>
<td>5′-AGACAGCAGCAGAAAGTGTCGCTTCC-3′</td>
<td>68.0</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>1R</td>
<td>5′-AGACAGCAGCAGAAAGTGTCGCTTCC-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 2</td>
<td>2F</td>
<td>5′-CCTTTCCTCCATGATTCGAGCAG-3′</td>
<td>70.2</td>
<td>379</td>
</tr>
<tr>
<td></td>
<td>2R</td>
<td>5′-GAAGGCTGAGGGGAGCCACAGAGGA-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 3</td>
<td>3F(4)</td>
<td>5′-CGAGCCCTACCTCACAGAGAT-3′</td>
<td>67.7</td>
<td>524</td>
</tr>
<tr>
<td></td>
<td>3R(1)</td>
<td>5′-GGGGCCAGAGCTAGAGGGGAGCCACAGAGGA-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 4</td>
<td>4F</td>
<td>5′-CTCTCACAGGTAGTTCTTCGAGCCC-3′</td>
<td>70.2</td>
<td>326</td>
</tr>
<tr>
<td></td>
<td>4R</td>
<td>5′-AAGCAGCAGCAGAAAGTGTCGCTTCC-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 5/6</td>
<td>5/6F</td>
<td>5′-ATGGCTCATGAGCGCTCTCTGTG-3′</td>
<td>68.0</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>5/6R</td>
<td>5′-GGGGGCAAGAGCTAGAGGGGAGCCACAGAGGA-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 7</td>
<td>7F</td>
<td>5′-GCTGCTGGCCACAGGCAAGGTAGG-3′</td>
<td>68.0</td>
<td>412</td>
</tr>
<tr>
<td></td>
<td>7R</td>
<td>5′-TCAGCAGCAAGCTAGAGGGGAGCCACAGAGGA-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 8</td>
<td>8F</td>
<td>5′-CTGCCCTCCTCTCTCGCTGCTG-3′</td>
<td>69.0</td>
<td>419</td>
</tr>
<tr>
<td></td>
<td>8R</td>
<td>5′-AAGCAGCAGCAGAAAGTGTCGCTTCC-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 9</td>
<td>9F</td>
<td>5′-ATGCTGATGAGCTTCCTCAGCTGG-3′</td>
<td>60.0</td>
<td>166</td>
</tr>
<tr>
<td>Poly(A)</td>
<td>ON3,9; c</td>
<td>5′-TTCCTCATCCTACGACAGG-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA Part 1</td>
<td>1F</td>
<td>5′-AGGGGTGAGGGGAGCCACAGAGGA-3′</td>
<td>67.0</td>
<td>1319</td>
</tr>
<tr>
<td></td>
<td>4R</td>
<td>5′-AAGCAGCAGCAGAAAGTGTCGCTTCC-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA Part 2</td>
<td>4F</td>
<td>5′-CTCCAGCTTACTCATAGGAGCCCTCAGAGGA-3′</td>
<td>65.0</td>
<td>1741</td>
</tr>
<tr>
<td></td>
<td>8R</td>
<td>5′-AAGCAGCAGCAGAAAGTGTCGCTTCC-3′</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA Int7</td>
<td>Int 7F</td>
<td>5′-TTCCTCTTCTACCCGCTCTCTACCC-3′</td>
<td>65.3</td>
<td>299</td>
</tr>
<tr>
<td></td>
<td>Int 7R</td>
<td>5′-TTCCTCTTCTACCCGCTCTCTACCC-3′</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F = forward and R = reverse primer sequences. The size of each fragment is shown in base pairs (bp). a(Harvey, 1996) b(Nelson et al., 1991) cGieselmann et al. (1994)
3.2.2 Design of Amplicons for the GBA Gene

A nested PCR approach was used for this gene due to the presence of the highly homologous, transcribed GBA pseudogene (GBAP) just 16 kb downstream from the functional GBA gene. Whole gene amplification was achieved using allele specific PCR and a further set of nested primers were designed to encompass all eleven exon structures and adjacent splice sites and the promoter region (Figure 3.2).

3.2.2.1 GBA Fragment PCR

All PCR reactions were carried out as described in Material and Methods 2.6.

The ‘whole GBA gene’ PCR reactions contained 100 ng of Human Genomic DNA (Promega), 1 x Buffer 2 containing 2.25 mM MgCl₂, 350 µM dNTPs, 500 nM (2.5 µl of 20 µM solution) forward and reverse primers (Table 3.2), and 7.5 units of Expand Long Template enzyme mix (Roche) in a 50 µL reaction volume.

Thermal cycling conditions for the whole gene fragment were 85°C for 5 min, 94°C for 3 min followed by ten cycles of denaturation at 94°C for 10 s, annealing at 55.4°C for 30 s and extension at 68°C for 5 min, followed by five cycles of denaturation and annealing as in the initial ten cycles and then an additional 10 s per cycle for extension at 68°C, and a final extension step at 68°C for 3 min.

The PCR reactions for GBA fragments PrC, 1, 2, 4, 5, 6, 7, 8, 10, 11a, 11b contained approximately 50-100 ng of the whole gene PCR reaction for each sample, 1 x Buffer 2 containing 2.25 mM MgCl₂, 350 µM of dNTPs, 250 nM (1.25 µl of 20 µM solution) of forward and reverse primers (Table 3.2) and 3.75 units of Expand Long Template enzyme mix (Roche) in a 50 µL reaction volume. The PCR reaction for GBA fragment 9 was the same as previously described except, instead of Buffer 2, Buffer 3 containing 2.25 mM MgCl₂ plus detergents was used. The PCR for GBA fragment 3 contained approximately 50 ng of the whole gene PCR reaction for each sample, 1.5 mM MgCl₂, 20 mM Tris-HCl pH 8.4, 50 mM KCl, 0.2 mM dNTPs, 0.5 µM forward and reverse primers (Table 3.2) and 1 unit Platinum Taq DNA polymerase (Life Technologies) per 50 µL reaction volume.
The Glucocerebrosidase Gene (GBA)

One fragment was designed to specifically amplify the whole *GBA* gene (7.2 kb). Thirteen internal fragments were designed to cover all GBA exons, the promoter and 3’-UTR regions.

**Figure 3.2 β-glucocerebrosidase Gene Fragments for PCR**

Not to scale
Table 3.2 Primer Sequences, Annealing Temperatures and Product Sizes for PCR of the Glucocerebrosidase Gene

<table>
<thead>
<tr>
<th>Fragment Name</th>
<th>Primer Name</th>
<th>Primer 5' to 3'</th>
<th>Annealing Temp °C</th>
<th>Product Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole GBA Gene</td>
<td>PR-BF R0</td>
<td>5'-ATCTTGGTATGTGTCATATTATATGC-3'</td>
<td>55.4</td>
<td>7250</td>
</tr>
<tr>
<td>Prom C</td>
<td>PrC-F Pr-R</td>
<td>5'-TTGACCTTTTGACATTTCATATGC-3' 5'-CGATGGAATTGGGATCAT-3'</td>
<td>56.4</td>
<td>272</td>
</tr>
<tr>
<td>GBA 1</td>
<td>F1 R1</td>
<td>5'-CTCTCCTAAGTGTGTCACCCACACATGC-3' 5'-CAGATGCAAGGATTCGAAG-3'</td>
<td>60.0</td>
<td>385</td>
</tr>
<tr>
<td>GBA 2</td>
<td>F2 R2</td>
<td>5'-TGCCAGGAGATAGTTGAG-3' 5'-GCCAGGCAACAGAAGTAAGA-3'</td>
<td>60.0</td>
<td>350</td>
</tr>
<tr>
<td>GBA 3</td>
<td>F3 R3</td>
<td>5'-CTAGGATTACAGGTGTGAGC-3' 5'-GACATCCACAGGAATAAGG-3'</td>
<td>58.0</td>
<td>448</td>
</tr>
<tr>
<td>GBA 4</td>
<td>F4 R4</td>
<td>5'-AGAGGTTAATGTTGTAGTGCAG-3' 5'-GACAGAAGCTGGCAAGTGAG-3'</td>
<td>60.0</td>
<td>331</td>
</tr>
<tr>
<td>GBA 5</td>
<td>F5 R5</td>
<td>5'-CAGACACTATTAGGACTGGCAAG-3' 5'-AGCAGACCTACCTACAGTTTC-3'</td>
<td>64.3</td>
<td>298</td>
</tr>
<tr>
<td>GBA 6</td>
<td>F6 R6</td>
<td>5'-TGGAGGCTAATGGCTGAA-3' 5'-GTGGAGGGTGAGGGACACAGA-3'</td>
<td>64.3</td>
<td>338</td>
</tr>
<tr>
<td>GBA 7</td>
<td>F7-2 R7</td>
<td>5'-TCAAGTGATCCACTGCTCAGTCTG-3' 5'-CTCTACTTGAATTTGGAAGCTATTC-3'</td>
<td>67.6</td>
<td>385</td>
</tr>
<tr>
<td>GBA 8</td>
<td>F8 R8</td>
<td>5'-TCCTCCGTCACCTAAACTC-3' 5'-TGCTCAAAAGGGCAGCTAAG-3'</td>
<td>61.4</td>
<td>440</td>
</tr>
<tr>
<td>GBA 9</td>
<td>F9 R9</td>
<td>5'-CCCTCCTGCTGCTCAG-3' 5'-TCCCTCCTGCTGCTGCTG-3'</td>
<td>56.4</td>
<td>302</td>
</tr>
<tr>
<td>GBA 10</td>
<td>F10 R10,9</td>
<td>5'-GGTGAATCTCTCTAGATGAGGCTTC-3' 5'-TGCTCTGACGCTACCTTCC-3'</td>
<td>63.3</td>
<td>227</td>
</tr>
<tr>
<td>GBA 11a</td>
<td>F11a R11a</td>
<td>5'-GAGGGCAATGGGACTGGCTC-3' 5'-CCAGTTTCAAAAGCAAGCA-3'</td>
<td>61.4</td>
<td>471</td>
</tr>
<tr>
<td>GBA 11b</td>
<td>F11b-2 R11b-2</td>
<td>5'-GCAATGTTTTGGGATGACTC-3' 5'-TGCCCTCCTGGTGTA-3'</td>
<td>55.4</td>
<td>481</td>
</tr>
</tbody>
</table>

F = forward and R = reverse primer sequences. The size of each fragment is shown in base pairs (bp). * Finckh et al. (1998)
Thermal cycling conditions for all fragments except fragment 3 were 95°C for 3 min, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at various temperatures (Table 3.2 for individual fragment annealing temperatures) for 20 s and extension at 68°C for 40 s, and a final extension step at 68°C for 5 min. Thermal cycling conditions for fragment 3 were 95°C for 5 min, followed by 30 cycles of denaturation at 95°C for 40 s, annealing at 58°C for 40 s and extension at 68°C for 40 s, and a final extension step at 68°C for 5 min.

3.2.3 Polymorphism Detection Using Denaturing High Performance Liquid Chromatography (DHPLC)

A Prostar Helix™ system (Varian, Australia) was used for fragment analysis. Buffer composition was according to manufacturer’s specifications. Analysis of 5 µL of PCR product was carried out using a flow rate of 0.9 ml/min over a 9 min period. Optimal conditions for double stranded DNA analysis were found using the universal methods for control of pump flow, as recommended by the manufacturer, and varying only the temperature of partial denaturation for each fragment. The optimum partial denaturation temperature ($T_m$) is sequence dependent and was calculated using a sequence analysis algorithm (http://insertion.Stanford.edu.au/melt.html).

A melting curve was constructed using a temperature gradient of 1°C increments over a range of plus and minus 3°C from the calculated $T_m$. The $T_m$ for each fragment was determined when a significant shift (approximately 25%) in retention time was observed. A control, pUC18 Hae III, was run under non-denaturing conditions (50°C) at the start and end of each run to ensure sufficient sensitivity for heteroduplex detection had been retained throughout each run.

Optimum partial denaturation temperature was determined for each of the GBA fragments (Table 3.3).
Table 3.3  Optimum Partial Denaturation Temperatures for PCR Amplicons of the Glucocerebrosidase Gene

<table>
<thead>
<tr>
<th>Fragment Name</th>
<th>Optimum Partial Denaturation Temp °C</th>
<th>Fragment Name</th>
<th>Optimum Partial Denaturation Temp °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prom C</td>
<td>54, 59</td>
<td>GBA 7</td>
<td>61</td>
</tr>
<tr>
<td>GBA 1</td>
<td>59</td>
<td>GBA 8</td>
<td>61</td>
</tr>
<tr>
<td>GBA 2</td>
<td>60</td>
<td>GBA 9</td>
<td>61</td>
</tr>
<tr>
<td>GBA 3</td>
<td>56, 61</td>
<td>GBA 10</td>
<td>61</td>
</tr>
<tr>
<td>GBA 4</td>
<td>55, 60</td>
<td>GBA 11a</td>
<td>61</td>
</tr>
<tr>
<td>GBA 5</td>
<td>60</td>
<td>GBA 11b</td>
<td>62</td>
</tr>
<tr>
<td>GBA 6</td>
<td>61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2.4 Dot Blot Hybridisation

3.2.4.1 Allele Specific Oligonucleotide (ASO) probes for the ARSA gene

Oligonucleotide probes were designed from sequence data of the ARSA gene for all polymorphisms detected in coding and promoter regions (Table 3.4). The probes were designed to ensure a mismatch near the middle of the oligonucleotide, a $T_m^*$ preferably between 50°C and 70°C in 2 x sodium chloride, sodium citrate solution (SSC; 0.3 M NaCl, 0.03 M sodium citrate: pH 7.0), and approximately matched the $T_m$ for both reference and variant sequences. Oligonucleotides were 5’-end labelled with a fluorescein tag (GeneWorks Pty Ltd, Adelaide).

* ($T_m = 81.5 + 16.6 \log [Na] + 41(G+C)/l-500/l-0.62[F]$ where $l =$ total bases, $[Na] =$ monovalent cation molar concentration, $[F] =$ formamide molar concentration)

3.2.4.2 Dot-Blot Hybridisation Procedure

Two PCR reactions covering all ASA exons, part 1 and part 2, were designed to increase the efficiency of ASO screening. The PCR product from each sample was prepared in a solution with a final concentration of 0.4 M NaOH and 10 mM EDTA, heated to 100°C for 10 min, cooled to 2°C and centrifuged for 5 s. Each sample preparation was applied in two x 2 µl aliquots to each piece of Hybond-N+ nylon filter (Amersham Pharmacia Biotech, USA) and allowed to dry. DNA concentration was approximately 100 ng per dot.

Pre-hybridisation of blots was carried out at 45°C (except for 1143C>T filters which were pre-hybridised at 56°C) in standard hybridisation solution (Roche) with 100 µg/ml denatured herring sperm DNA for 1 h with agitation. Approximately 50 ng of 5’-end labelled oligonucleotide probe was added to the pre-hybridisation solution and filters were hybridised for approximately 3 h at the same temperature used for pre-hybridisation.

Hybridised filters were rinsed twice in 2 x SSC for 10 min at room temperature, followed by a wash in 2 x SSC, 0.1% sodium dodecyl sulphate (SDS) at the critical temperature (Table 3.4) for 5-10 min. Detection was achieved using the CDP-Star product with anti-fluorescein alkaline phosphatase conjugate (Roche).
Table 3.4 Allele Specific Oligonucleotides for ARSA Used in the Dot-blot Hybridisation

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>ASO Type</th>
<th>Sequence 5’ to 3’</th>
<th>Critical Temp°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>N350S a</td>
<td>Reference</td>
<td>5’-AAGGTGACATGGGCAGTG-3’</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-AAGGTGACATGGGCAGTG-3’</td>
<td></td>
</tr>
<tr>
<td>T391S b</td>
<td>Reference</td>
<td>5’-GTGCAGCTGGGAAAGTA-3’</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-GTGCAGCTGGGAAAGTA-3’</td>
<td></td>
</tr>
<tr>
<td>PD a</td>
<td>Reference</td>
<td>5’-CTGGTTTACGTTAT-3’</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-CTGGTTTACGTTAT-3’</td>
<td></td>
</tr>
<tr>
<td>1143C→T</td>
<td>Reference</td>
<td>5’-CTACCGAGTGAGTCCGT-3’</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-CTACCGAGTGAGTCCGT-3’</td>
<td></td>
</tr>
<tr>
<td>W193C</td>
<td>Reference</td>
<td>5’-GCCGCCCTGCTGCCCGGA-3’</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-GCCGCCCTGCTGCCCGGA-3’</td>
<td></td>
</tr>
<tr>
<td>R496H</td>
<td>Reference</td>
<td>5’-CCCCCGCCAGCT-3’</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-CCCCCGCCAGCT-3’</td>
<td></td>
</tr>
<tr>
<td>1101+123C→T</td>
<td>Reference</td>
<td>5’-CAAGTTGGGCCACCCT-3’</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-CAAGTTGGGCCACCCT-3’</td>
<td></td>
</tr>
<tr>
<td>1102-32T→C</td>
<td>Reference</td>
<td>5’-GGGCTCTGCCGCC-3’</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-GGGCTCTGCCGCC-3’</td>
<td></td>
</tr>
<tr>
<td>-720C→T</td>
<td>Reference</td>
<td>5’-ATAACGCTGGGCT-3’</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-ATAACGCTGGGCT-3’</td>
<td></td>
</tr>
<tr>
<td>-498A→G</td>
<td>Reference</td>
<td>5’-GCGCACTGGCCG-3’</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-GCGCACTGGCCG-3’</td>
<td></td>
</tr>
<tr>
<td>-789G→A</td>
<td>Reference</td>
<td>5’-TGAAGGAAAGAGCCCG-3’</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-TGAAGGAAAGAGCCCG-3’</td>
<td></td>
</tr>
<tr>
<td>-351G→C</td>
<td>Reference</td>
<td>5’-GGTACCAGCCTCC-3’</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Variant</td>
<td>5’-GGTACCAGCCTCC-3’</td>
<td></td>
</tr>
</tbody>
</table>

All ASO oligonucleotides were 5’-end-labelled with fluorescein.

a(Gieselmann et al., 1989), b(Polten et al., 1991)
3.3 RESULTS

3.3.1 Samples

A total of 113 samples from healthy individuals with known ASA and GBA activity levels were obtained from the Women’s and Children’s Hospital, Adelaide (WCH). These leucocyte pellets had ASA/GBA levels across the normal range (GBA: 600-3200 pmol/min/mg cell protein for leucocytes; ASA 0.5->6.0 nmol/min/mg cell protein for leucocytes). The samples were collected and assayed as part of the diagnostic service provided by The National Referral Laboratory for lysosomal storage disorders, which receives up to 1000 peripheral blood samples per year for investigation. As part of their testing regime, ASA activity was measured in sonicated leucocytes using the artificial substrate p-nitrocatechol-sulphate (Materials and Methods 2.3.3) and GBA activity was measured using 4-methylumbelliferone-β-glucosidase substrate (Materials and Methods 2.3.4). The unused portion of the leucocyte samples was frozen and stored for use in future studies.

Analysis of the ARSA gene in 52 of these samples had been previously performed with individuals having ASA activity above 4.5 nmol/min/mg protein being classified as having high ASA activity and individuals having ASA activity between 0.5 and 2.0 nmol/min/mg protein being classified as having low normal activity (Figure 3.3, hatched bars). The remaining 61 samples with ASA enzyme activity levels from 2.1 to 4.5 nmol/min/mg protein (see maroon bars in Figure 3.3) were analysed for ten of the fourteen polymorphisms identified previously. Excluded were three low frequency polymorphisms (678+33C>T, 1102-75C>A, 1143C>T) because they did not cause a change in an amino acid, limited genotyping was available and it was thought they were unlikely to be significant. Also excluded was the coding polymorphism (R496H) which has a very low frequency and is thought to be associated with low ASA activity (Ricketts et al., 1998). For the GBA gene, twenty samples with the highest GBA activity levels (above 2500 pmol/min/mg cell protein, see blue bars in Figure 3.4) were selected to comprise the high activity test group and these were screened for high activity variants. The remainder of the samples represent a randomly selected population group.
The distribution of leucocyte ASA enzyme activity levels in the sample population of normal individuals (N = 113) is presented. The mean ASA enzyme activity level is 3.0 nmol/min/mg protein and standard deviation is 1.13. This mean is slightly higher than that observed for a larger population (N = 894) presented in Figure 1.6 which reported a mean ASA enzyme activity level of 2.62 nmol/min/mg protein and a standard deviation of 1.03. Samples from the high and low activity groups (hatched bars) (N=52) had been screened for polymorphisms in an earlier study and the remaining samples (maroon bars) (N=61) were analysed during this PhD project. All samples were analysed in the same laboratory.
Figure 3.4 Distribution of Leucocyte GBA Levels in the Sample Population

The distribution of GBA enzyme activity levels in the sample population of normal individuals (N = 108) is presented. The mean GBA enzyme activity level is 1687 pmol/min/mg protein and the standard deviation is 689. The mean is slightly higher than that observed for the larger population (N = 701) presented in Figure 1.6b which reported a mean GBA enzyme activity level of 1141 pmol/min/mg protein and a standard deviation of 376. Samples from the high activity group (blue bars) (N=21) were analysed during this PhD project and the remaining samples (hatched bars) were not analysed.
3.3.2 Analysis of ASA Polymorphisms in the Sample Population

It is known that the statistical significance of any data correlation is more robust when conducted in a larger population and, as a further 61 samples were available the investigation of the frequency of ten ASA polymorphisms was continued in this additional group of samples.

Dot-blot hybridisation (section 3.2.4) was used, as for the first 52 samples, to identify the presence of ten polymorphisms: the four 5′-flanking polymorphisms (-789G>A, -720C>T, -498A>G, -351G>C), three non-synonymous coding polymorphisms (W193C, T391S, N350S), the poly(A) variant (2723A>G) and the intronic polymorphisms (1101+123C>T, 1102-32T>C) in the individual samples SA to SX, MA to MX and DA to DV. The W193C, T391S and pseudo-deficiency allele (N350S and 2723A>G) polymorphisms were included in the group for further study because previous research was published identifying their contribution to reduction in arylsulphatase A activity (Harvey et al., 1998). The 5′-flanking and the 1102-32T>C (intron 6) polymorphisms were included because the evidence for correlation of 5′-flanking polymorphisms with enzyme activity level was inconclusive and the frequency of this intronic polymorphism appeared abnormally high in the research to date. An additional inclusion was the 1101+123C>T polymorphism as it appeared to be present in greater frequency in high activity samples. It is present in samples AB, AH and AG and these three samples have the highest ASA enzyme activity levels (5.4 to 6.3 nmol/min/mg protein) in the sample group.

An additional polymorphism (1204+20C>G) was identified in intron 7 in some samples. This polymorphism was identified after an investigation into the reasons for inconsistent results between the dot-blot hybridisation and sequencing of some samples. An additional fragment using primers Int7F and Int7R was designed to use for sequencing problematic samples. The inconsistency arose as the larger fragment (ASA2) was used for the dot-blot hybridisation experiments but fragment 7 primers were used for sequencing. The polymorphism occurred within the reverse primer (7R) sequence used for fragment 7 which lead to incomplete priming of the reverse strand in the sequencing reaction and incorrect sequencing results.
In total, fifteen polymorphisms were identified in the ARSA gene: four promoter (-789G>A, -720C>T, -498A>G, -351G>C), five coding (W193C, T391S, 1143C>T, N350S, R496H), one poly(A) (2723A>G) and five intronic (678+33C>T, 1101+123C>T, 1102-75C>A, 1102-32C>T, 1204+20C>G). Table 3.5 lists the genotype and ASA enzyme activity level of all samples. The SNPs are also numbered for easy reference during the statistical analysis.
Table 3.5  The ARSA Genotypes of Individuals in the Sample Population

The genotype of 113 individuals was determined at up to 15 SNP loci in the ARSA gene. The frequency of each SNP in the sample population is presented. The sample code, ASA enzyme activity (nmol/min/mg cell protein) and the ARSA genotype of each individual is also presented. The genotype is recorded as 11 for homozygous common alleles, 12 for heterozygous and 22 for homozygous variant alleles. Where genotyping was not performed a “-” indicates that results are not available.
<table>
<thead>
<tr>
<th>Sample Code</th>
<th>ASA Enzyme Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-789 G&gt;A</td>
</tr>
<tr>
<td></td>
<td>-720 C&gt;T</td>
</tr>
<tr>
<td></td>
<td>-496 A&gt;G</td>
</tr>
<tr>
<td></td>
<td>-351 G&gt;C</td>
</tr>
<tr>
<td></td>
<td>W193X</td>
</tr>
<tr>
<td></td>
<td>678+33 C&gt;T</td>
</tr>
<tr>
<td></td>
<td>N396S</td>
</tr>
<tr>
<td></td>
<td>1101+121 C&gt;T</td>
</tr>
<tr>
<td></td>
<td>1102-75 C&gt;A</td>
</tr>
<tr>
<td></td>
<td>1102-327 C&gt;T</td>
</tr>
<tr>
<td></td>
<td>1143 C&gt;T</td>
</tr>
<tr>
<td></td>
<td>1204+20 C&gt;G</td>
</tr>
<tr>
<td></td>
<td>R406H</td>
</tr>
<tr>
<td></td>
<td>2723 A&gt;G</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNP No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>0.9</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>LB</td>
<td>1.0</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LC</td>
<td>1.1</td>
<td>22</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>22</td>
<td>12</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LD</td>
<td>1.1</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LE</td>
<td>1.1</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LF</td>
<td>1.4</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LG</td>
<td>1.4</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LH</td>
<td>1.5</td>
<td>12</td>
<td>22</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>22</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LJ</td>
<td>1.5</td>
<td>22</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LK</td>
<td>1.7</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LL</td>
<td>1.7</td>
<td>22</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>22</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LM</td>
<td>1.8</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LO</td>
<td>1.8</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>-</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>LP</td>
<td>1.8</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>LQ</td>
<td>1.9</td>
<td>11</td>
<td>22</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LR</td>
<td>1.9</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>22</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>LT</td>
<td>1.9</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>Sample Code</td>
<td>ASA Enzyme Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LV</td>
<td>2.0 22 11 11 11 11 12 11 11 12 12 11 11 11 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LY</td>
<td>2.0 12 12 11 11 11 11 11 - 11 11 11 12 12 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DA</td>
<td>2.1 11 12 11 11 11 11 12 - 11 11 11 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DB</td>
<td>2.1 11 11 11 11 11 11 11 - 11 11 11 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC</td>
<td>2.1 12 11 11 11 11 11 11 - 11 12 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE</td>
<td>2.2 12 11 11 11 11 11 11 - 12 11 11 11 11 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF</td>
<td>2.2 11 12 12 11 11 11 11 - 12 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DG</td>
<td>2.2 12 11 11 11 11 11 11 - 12 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH</td>
<td>2.2 12 12 11 11 11 11 11 - 12 11 11 11 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DJ</td>
<td>2.2 11 22 11 11 11 11 12 - 11 11 11 11 11 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DL</td>
<td>2.2 11 12 11 11 11 11 12 - 11 11 11 11 11 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DM</td>
<td>2.2 11 12 11 11 11 11 12 - 11 11 11 11 11 11 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DP</td>
<td>2.3 11 12 12 11 11 11 11 - 11 11 11 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQ</td>
<td>2.3 11 12 12 11 11 11 12 - 11 11 11 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR</td>
<td>2.3 11 12 12 11 11 11 11 - 11 11 11 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS</td>
<td>2.4 11 11 11 11 11 11 11 - 11 11 11 11 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>2.4 22 11 11 11 11 11 11 - 11 11 11 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>2.5 11 11 11 11 22 11 11 - 11 11 11 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MB</td>
<td>2.5 12 12 11 11 11 11 11 - 11 11 11 12 - 12 - 12 - 12 - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Code</td>
<td>ASA Enzyme Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>2.5 11 11 11 11 11 11 11 12 12 - 11 11 - 11 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ME</td>
<td>2.5 11 22 11 11 11 11 - 11 11 - 11 11 - 12 12 - 11 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>2.5 11 11 11 11 11 11 - 11 11 - 11 11 - 11 11 - 12 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MG</td>
<td>2.5 12 12 11 11 11 11 - 11 11 - 11 11 - 12 12 - 12 12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MJ</td>
<td>2.6 11 12 11 12 11 11 - 11 11 - 11 11 - 12 12 - 11 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK</td>
<td>2.6 12 11 11 11 11 11 - 12 12 - 11 11 - 11 11 - 11 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>2.6 11 22 11 11 11 11 - 11 11 - 11 11 - 22 - 12 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AJ</td>
<td>2.6 12 12 12 11 11 11 12 12 11 11 11 11 12 - 11 12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>2.7 12 12 12 12 11 11 12 12 11 11 11 11 12 - 11 12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>2.7 11 11 11 11 12 12 - 11 11 - 11 11 - 11 11 - 12 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP</td>
<td>2.7 12 11 11 11 11 11 - 11 11 - 11 11 - 12 12 - 11 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>2.8 11 22 11 11 11 12 - 11 11 - 11 11 - 22 - 12 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>2.8 11 12 11 11 11 11 - 11 11 - 11 11 - 12 12 - 12 12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MT</td>
<td>2.8 11 22 12 11 12 - 11 11 - 11 11 - 11 11 - 22 - 11 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MX</td>
<td>2.9 11 12 11 11 12 - 11 12 - 11 12 - 11 12 - 11 12 - 11 11</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MY</td>
<td>2.9 12 11 11 11 11 11 - 11 12 - 12 12 - 11 12 - 12 12</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MZ</td>
<td>2.9 12 11 11 11 11 11 - 12 11 - 12 12 - 12 12 - 12 12</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB</td>
<td>3.0 12 12 11 11 11 11 - 11 11 - 12 12 - 12 12 - 12 12</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Code</td>
<td>ASA Enzyme Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>3.0 12 11 11 12 11 - 11 12 - 12 - 11 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>3.1 11 22 11 11 11 - 11 11 - 11 - 22 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td>3.1 12 11 11 11 11 - 11 12 - 12 - 11 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SJ</td>
<td>3.1 12 12 11 11 11 - 11 12 - 12 - 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>3.2 11 12 11 11 11 - 11 12 - 11 - 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SL</td>
<td>3.2 11 12 11 11 11 - 11 11 - 11 - 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM</td>
<td>3.2 11 22 11 11 11 - 11 11 - 11 - 22 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO</td>
<td>3.2 11 22 11 11 11 - 11 11 - 11 - 22 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>3.2 11 12 11 12 12 - 11 11 - 11 - 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SQ</td>
<td>3.3 22 11 11 11 11 - 11 11 - 12 - 12 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>3.3 12 12 11 11 11 - 11 11 - 12 - 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>3.4 12 11 11 11 11 - 11 11 - 22 - 11 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>3.5 11 22 11 11 11 - 11 11 - 11 - 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>3.5 11 22 11 11 11 - 11 11 - 11 - 22 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SX</td>
<td>3.5 12 12 11 11 11 - 11 11 - 12 - 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA</td>
<td>3.7 11 12 11 11 11 - 11 11 11 12 11 12 12 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HB</td>
<td>3.7 11 11 11 11 11 - 11 11 - 22 12 12 12 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HC</td>
<td>3.8 11 12 11 12 11 - 11 11 - 11 11 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD</td>
<td>3.8 11 12 11 12 12 - 11 11 - 11 11 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>3.9 11 22 11 11 11 - 11 11 - 11 11 22 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF</td>
<td>3.9 11 12 11 11 11 - 11 12 - 11 11 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HH</td>
<td>4.0 11 12 11 11 11 - 11 11 - 12 11 12 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HJ</td>
<td>4.1 11 22 11 11 11 - 11 11 - 11 12 22 - - 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Code</td>
<td>ASA Enzyme Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HK</td>
<td>4.1 11 12 11 12 11 - 11 11 - 11 11 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HL</td>
<td>4.1 12 12 11 11 11 - 11 11 - 12 11 12 12 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>4.1 11 12 12 11 11 - 11 11 - 11 11 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HO</td>
<td>4.2 12 12 11 11 11 - 11 11 - 12 11 11 12 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP</td>
<td>4.2 22 11 11 11 11 - 12 11 - 11 11 11 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HQ</td>
<td>4.3 11 22 12 11 11 - 11 11 - 11 11 22 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AF</td>
<td>4.5 11 11 11 11 11 - 11 11 12 11 12 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>4.7 11 12 11 11 11 - 11 11 - 11 11 22 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM</td>
<td>4.8 12 11 11 12 11 - 12 11 11 11 11 11 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AK</td>
<td>4.9 11 22 11 11 11 - 11 11 11 11 12 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AL</td>
<td>5.0 12 12 11 11 11 - 11 11 11 11 22 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS</td>
<td>5.0 22 11 11 11 11 - 11 11 - 12 11 11 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HT</td>
<td>5.1 11 12 11 11 11 - 11 11 - 12 11 12 12 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HV</td>
<td>5.2 12 12 11 11 11 - 12 11 - 11 11 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>5.2 12 11 11 12 11 - 11 11 11 11 11 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>5.3 11 12 12 11 11 12 11 11 11 11 12 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>5.4 12 11 11 11 11 - 12 12 11 11 11 11 - 11 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3.2.1 Statistical Analysis of ASA Polymorphism Data

The correlation between ASA SNPs and ASA enzyme activity was determined. The three SNPs for which limited genotyping data were available (678+33C>T, 1102-75C>A, 1204+20C>G) were excluded from the analysis because it was considered the little additional information would be gained by their inclusion. The remaining twelve SNPs for which extensive genotyping data were available did not differ from the normal Hardy-Weinberg equilibrium (Table 3.6a). A one-sample Kolmogorov-Smirnov test was also performed to assess the suitability of the ARSA genotype data for further analysis (Table 3.6b). The significance value of $P = 0.144$ indicates that the observed distribution resembles the theoretical (normal) distribution. These two measures indicate that the ASA SNP data can be analysed with procedures that assume normality.

Pairwise Linkage Disequilibrium (LD) was analysed with the Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD) (Nyholt, 2004) and Haploview (Barrett et al., 2005) programs (Figure 3.5). Strong linkage was suggested between -789G>A and -720C>T, -720C>T and 1102-32T>C, -720C>T and T391S, N350S and 2723A>G, and -789G>A and 1102-32T>C in the SNPSpD analysis ($r = -0.37, 0.44, 0.77, 0.75, -0.37$ respectively) (Figure 3.5a). This was substantiated by the Haploview analysis for the first four SNP combinations (Figure 3.5b). Strong linkage between 2723A>G and N350S ($r = 0.75$) was expected as these SNPs comprise the PD allele but linkage between the other SNP combinations has not been reported previously.

Haplotyping was done using HAP (Halperin and Eskin, 2004) but did not show any additional effects related to the cis or trans effect. The analysis also did not reveal any additional strong haplotype associations except for the N350S and 2723A>G linkage known to be responsible for the PD allele.

Linear regression (LR) analysis was used to analyse the individual SNPs for correlation with enzyme activity. Three SNPs, 2723A>G, N350S and W193C show significance ($P < 0.05$) and the intronic SNP, 1101+123C>T, is close to significance ($P = 0.055$) (Table 3.6a). To aid with the interpretation of these results, the twelve SNPs were analysed together using a forward linear regression (Table 3.7). This analysis first entered 2723A>G into the model equation indicating that it is the most significantly associated with ASA levels. However, 1101+123C>T was also entered into the model equation that
already contained 2723A>G, thus indicating an additional independent effect at or near 1101+123C>T. Although W193C and N350S are individually nominally significant ($P = 0.017$ and 0.046 respectively) (Table 3.6a), they have not been entered into the model equation once 2723A>G is entered. This appears to indicate that their individual association is likely due to a correlation with 2723A>G, which is known to be true for N350S ($r = 0.75$) but not for W193C ($r = 0.03$).
A

ASA SNP Analysis Results

<table>
<thead>
<tr>
<th>SNP No.</th>
<th>ASA SNP</th>
<th>Frequency</th>
<th>HWpval</th>
<th>LR P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-789G&gt;A</td>
<td>0.235</td>
<td>0.082</td>
<td>0.464</td>
</tr>
<tr>
<td>2</td>
<td>-720C&gt;T</td>
<td>0.403</td>
<td>0.196</td>
<td>0.571</td>
</tr>
<tr>
<td>3</td>
<td>-498A&gt;G</td>
<td>0.049</td>
<td>1</td>
<td>0.639</td>
</tr>
<tr>
<td>4</td>
<td>-351G&gt;C</td>
<td>0.075</td>
<td>0.956</td>
<td>0.317</td>
</tr>
<tr>
<td>5</td>
<td>W193C (G&gt;T)</td>
<td>0.075</td>
<td>1</td>
<td>0.017</td>
</tr>
<tr>
<td>7</td>
<td>N350S (A&gt;G)</td>
<td>0.084</td>
<td>0.876</td>
<td>0.046</td>
</tr>
<tr>
<td>8</td>
<td>1101+123C&gt;T</td>
<td>0.080</td>
<td>0.959</td>
<td>0.055</td>
</tr>
<tr>
<td>10</td>
<td>1102-32T&gt;C</td>
<td>0.221</td>
<td>0.261</td>
<td>0.224</td>
</tr>
<tr>
<td>11</td>
<td>1143C&gt;T</td>
<td>0.058</td>
<td>1</td>
<td>0.533</td>
</tr>
<tr>
<td>12</td>
<td>T391S (C&gt;G)</td>
<td>0.496</td>
<td>0.423</td>
<td>0.898</td>
</tr>
<tr>
<td>14</td>
<td>R496H (G&gt;A)</td>
<td>0.010</td>
<td>1</td>
<td>0.123</td>
</tr>
<tr>
<td>15</td>
<td>2723A&gt;G</td>
<td>0.049</td>
<td>1</td>
<td>0.000055</td>
</tr>
</tbody>
</table>

B

One-Sample Kolmogorov-Smirnov Test

<table>
<thead>
<tr>
<th>ASA No.</th>
<th>ASA</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>113</td>
</tr>
<tr>
<td>Normal</td>
<td>Mean 2.965</td>
</tr>
<tr>
<td>Parameters(a,b)</td>
<td>Std. Deviation 1.1285</td>
</tr>
<tr>
<td>Most Extreme</td>
<td>Absolute .108</td>
</tr>
<tr>
<td>Differences</td>
<td>Positive .108</td>
</tr>
<tr>
<td></td>
<td>Negative -.054</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov Z</td>
<td>1.148</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.144</td>
</tr>
</tbody>
</table>

a Test distribution is Normal.
b Calculated from data.

Table 3.6 Statistical Analysis of the ASA SNPs

(A) The ASA SNPs are listed together with the SNP number used in some of the statistical analysis output, their frequency, HWpval (the probability that deviation from Hardy-Weinberg equilibrium could be explained by chance) and the LR P value (p value result from the linear regression (LR) analysis of the individual SNPs). (B) The output from the one-sample Kolmogorov-Smirnov test done in SPSS showing the final result of 0.144 indicating that the ASA distribution is normal.
**SNPSpD: Pairwise LD Correlations for ASA SNPs**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-0.37</td>
<td>-0.15</td>
<td>-0.19</td>
<td>0.28</td>
<td>0.07</td>
<td>-0.37</td>
<td>0.12</td>
<td>-0.3</td>
<td>0.15</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.37</td>
<td>1</td>
<td>0.28</td>
<td>-0.23</td>
<td>0.29</td>
<td>-0.25</td>
<td>-0.18</td>
<td>0.44</td>
<td>-0.11</td>
<td>0.77</td>
<td>-0.08</td>
<td>-0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.19</td>
<td>0.28</td>
<td>1</td>
<td>-0.06</td>
<td>-0.06</td>
<td>0.01</td>
<td>0.02</td>
<td>0.12</td>
<td>-0.06</td>
<td>0.23</td>
<td>-0.02</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.1</td>
<td>-0.23</td>
<td>-0.06</td>
<td>1</td>
<td>-0.04</td>
<td>-0.06</td>
<td>0.02</td>
<td>0.15</td>
<td>-0.07</td>
<td>-0.22</td>
<td>-0.03</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.19</td>
<td>0.29</td>
<td>-0.06</td>
<td>-0.04</td>
<td>1</td>
<td>-0.06</td>
<td>-0.08</td>
<td>0.15</td>
<td>-0.06</td>
<td>0.29</td>
<td>-0.02</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.28</td>
<td>-0.25</td>
<td>0.01</td>
<td>-0.06</td>
<td>-0.06</td>
<td>1</td>
<td>0</td>
<td>0.16</td>
<td>-0.1</td>
<td>-0.3</td>
<td>0.24</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.07</td>
<td>-0.18</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.08</td>
<td>0</td>
<td>1</td>
<td>0.16</td>
<td>-0.06</td>
<td>-0.24</td>
<td>-0.02</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.37</td>
<td>0.44</td>
<td>0.12</td>
<td>0.15</td>
<td>0.15</td>
<td>0.16</td>
<td>0.16</td>
<td>1</td>
<td>-0.2</td>
<td>0.44</td>
<td>0.05</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.12</td>
<td>-0.11</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.06</td>
<td>-0.1</td>
<td>-0.06</td>
<td>-0.2</td>
<td>1</td>
<td>0.25</td>
<td>-0.02</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-0.3</td>
<td>0.77</td>
<td>0.23</td>
<td>-0.22</td>
<td>0.29</td>
<td>-0.3</td>
<td>-0.24</td>
<td>0.44</td>
<td>0.25</td>
<td>1</td>
<td>0.1</td>
<td>-0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0.15</td>
<td>-0.08</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.24</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.02</td>
<td>0.1</td>
<td>1</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.22</td>
<td>-0.19</td>
<td>0.1</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.75</td>
<td>-0.06</td>
<td>0.12</td>
<td>-0.02</td>
<td>-0.22</td>
<td>0.3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.5 Linkage Disequilibrium Analysis of the ASA SNPs**

Pairwise Linkage Disequilibrium (LD) was analysed with the Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD) and Haploview programs. (A) SNPSpD generates a matrix of LD correlations. Values generated are in the range of -1 to +1 and values close to +/-1 indicate LD. The highlighted numbers are referred to in the text. (B) The Haploview output shows the SNP combinations that demonstrate strong linkage, recombination or are uninformative. The numbers within the squares represent D'. Where D' = 1, it is not shown. D' = 0.70 is shown as 70.
## Forward Linear Regression Analysis

### Variables Entered/Removed(a)

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables Entered</th>
<th>Variables Removed</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SNP15</td>
<td></td>
<td>Forward (Criterion: Probability-of-F-to-enter &lt;= .050)</td>
</tr>
<tr>
<td>2</td>
<td>SNP15, SNP8</td>
<td></td>
<td>Forward (Criterion: Probability-of-F-to-enter &lt;= .050)</td>
</tr>
</tbody>
</table>

a  Dependent Variable: ASA

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.529(a)</td>
<td>.280</td>
<td>.265</td>
<td>1.3314</td>
</tr>
<tr>
<td>2</td>
<td>.615(b)</td>
<td>.379</td>
<td>.353</td>
<td>1.2489</td>
</tr>
</tbody>
</table>

a  Predictors: (Constant), SNP15  
b  Predictors: (Constant), SNP15, SNP8

### ANOVA(c)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>34.402</td>
<td>1</td>
<td>34.402</td>
<td>19.408</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>88.630</td>
<td>50</td>
<td>1.773</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>123.032</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Regression</td>
<td>46.603</td>
<td>2</td>
<td>23.302</td>
<td>14.939</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>76.429</td>
<td>49</td>
<td>1.560</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>123.032</td>
<td>51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a  Predictors: (Constant), SNP15  
b  Predictors: (Constant), SNP15, SNP8  
c  Dependent Variable: ASA

### Coefficients(a)

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>3.674</td>
<td>.205</td>
<td>17.883</td>
</tr>
<tr>
<td></td>
<td>SNP15</td>
<td>-2.064</td>
<td>.468</td>
<td>-.529</td>
</tr>
<tr>
<td>2</td>
<td>(Constant)</td>
<td>3.476</td>
<td>.205</td>
<td>16.928</td>
</tr>
<tr>
<td></td>
<td>SNP15</td>
<td>-1.866</td>
<td>.445</td>
<td>-.478</td>
</tr>
<tr>
<td></td>
<td>SNP8</td>
<td>1.664</td>
<td>.595</td>
<td>.319</td>
</tr>
</tbody>
</table>

a  Dependent Variable: ASA
### Table 3.7 Forward Linear Regression Analysis of ASA SNPs

This series of tables shows the criteria used for the SNPs entered into the SPSS forward linear regression equation (Variables Entered/Removed), the correlation value (R) for the predictors in each model (Model Summary), the significance for each of the two proposed models (ANOVA), the significance values for each SNP within the two proposed models (Coefficients), and the significance values for each of the excluded SNPs (Excluded Variables). The values in red are the notable results in the table or are discussed in the text.

<table>
<thead>
<tr>
<th>Model</th>
<th>SNP</th>
<th>Beta In</th>
<th>t</th>
<th>Sig.</th>
<th>Partial Correlation</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SNP1</td>
<td>-.093(a)</td>
<td>-.737</td>
<td>.465</td>
<td>-.105</td>
<td>.917</td>
</tr>
<tr>
<td></td>
<td>SNP2</td>
<td>-.001(a)</td>
<td>-.005</td>
<td>.996</td>
<td>-.001</td>
<td>.961</td>
</tr>
<tr>
<td></td>
<td>SNP3</td>
<td>.208(a)</td>
<td>1.741</td>
<td>.088</td>
<td>.241</td>
<td>.970</td>
</tr>
<tr>
<td></td>
<td>SNP4</td>
<td>.200(a)</td>
<td>1.695</td>
<td>.096</td>
<td>.235</td>
<td>.998</td>
</tr>
<tr>
<td></td>
<td>SNP5</td>
<td>-.243(a)</td>
<td>-2.076</td>
<td>.043</td>
<td>.284</td>
<td>.983</td>
</tr>
<tr>
<td></td>
<td>SNP7</td>
<td>.203(a)</td>
<td>1.087</td>
<td>.283</td>
<td>.153</td>
<td>.413</td>
</tr>
<tr>
<td></td>
<td>SNP8</td>
<td>.319(a)</td>
<td>2.797</td>
<td>.007</td>
<td>.371</td>
<td>.975</td>
</tr>
<tr>
<td></td>
<td>SNP10</td>
<td>.204(a)</td>
<td>1.721</td>
<td>.092</td>
<td>.239</td>
<td>.985</td>
</tr>
<tr>
<td></td>
<td>SNP11</td>
<td>-.101(a)</td>
<td>-.839</td>
<td>.406</td>
<td>-.119</td>
<td>.999</td>
</tr>
<tr>
<td></td>
<td>SNP12</td>
<td>-.060(a)</td>
<td>-.494</td>
<td>.624</td>
<td>-.070</td>
<td>.979</td>
</tr>
<tr>
<td>2</td>
<td>SNP14</td>
<td>-.070(a)</td>
<td>-.558</td>
<td>.579</td>
<td>-.079</td>
<td>.918</td>
</tr>
<tr>
<td></td>
<td>SNP1</td>
<td>-.050(b)</td>
<td>-.417</td>
<td>.678</td>
<td>-.060</td>
<td>.901</td>
</tr>
<tr>
<td></td>
<td>SNP2</td>
<td>.038(b)</td>
<td>.323</td>
<td>.748</td>
<td>.047</td>
<td>.948</td>
</tr>
<tr>
<td></td>
<td>SNP3</td>
<td>.165(b)</td>
<td>1.441</td>
<td>.156</td>
<td>.204</td>
<td>.950</td>
</tr>
<tr>
<td></td>
<td>SNP4</td>
<td>.183(b)</td>
<td>1.650</td>
<td>.105</td>
<td>.232</td>
<td>.994</td>
</tr>
<tr>
<td></td>
<td>SNP5</td>
<td>-.214(b)</td>
<td>-1.926</td>
<td>.060</td>
<td>-.268</td>
<td>.974</td>
</tr>
<tr>
<td></td>
<td>SNP7</td>
<td>.159(b)</td>
<td>.901</td>
<td>.372</td>
<td>.129</td>
<td>.409</td>
</tr>
<tr>
<td></td>
<td>SNP10</td>
<td>.145(b)</td>
<td>1.255</td>
<td>.215</td>
<td>.178</td>
<td>.942</td>
</tr>
<tr>
<td></td>
<td>SNP11</td>
<td>-.063(b)</td>
<td>-.552</td>
<td>.583</td>
<td>-.079</td>
<td>.984</td>
</tr>
<tr>
<td></td>
<td>SNP12</td>
<td>.015(b)</td>
<td>.127</td>
<td>.899</td>
<td>.018</td>
<td>.925</td>
</tr>
<tr>
<td></td>
<td>SNP14</td>
<td>-.070(b)</td>
<td>-.595</td>
<td>.554</td>
<td>-.086</td>
<td>.918</td>
</tr>
</tbody>
</table>

a Predictors in the Model: (Constant), SNP15
b Predictors in the Model: (Constant), SNP15, SNP8
c Dependent Variable: ASA
Furthermore, it can be seen in model 2 in the Forward Linear Regression – Excluded Variables table (Table 3.7) that W193C is close to significance ($P = 0.06$). The possibility of increasing the sample size was considered but as the power of the analysis is very high (0.98), the question of whether W193C is responsible for an additional independent effect is unlikely to be elucidated in this way.

Multiple testing was corrected for using the SNPSpD approach (Nyholt, 2004). Analysis of twelve SNPs is equivalent to approximately 10 independent SNPs indicating that the experiment wide significance threshold ($\alpha$) is $P = 0.0051$ ($P = 0.05/10$). This approach also identified 2723A>G as significant ($P < 0.005$) and when the analysis is conditioned on 2723A>G, 1101+123C>T was not significant ($P = 0.24$), but W193C is significant ($P = 0.0051$, empirical $P = 0.002$). This provides evidence for an independent effect at or near W193C.

More sophisticated analysis was performed using QTPHASE (Dudbridge, 2003) which identified only 2723A>G as significant (asymptotic $P = 0.0000176$, permuted empirical $P = 0.000999$). This SNP reaches experiment wide significance after 1000 permutations (empirical $p = 0.0012$). If W193C and/or 1101+123C>T are responsible, or linked to a polymorphism responsible for an independent effect on enzyme activity, functional studies may be required as the probability that a real difference might have been missed in the current analyses is less than two percent.

The mean ASA enzyme activity for each genotype at each locus in the sample population was calculated (Figure 3.6). 2723A>G (frequency = 0.049) is known to be responsible for low ASA enzyme activity with the average ASA activity for heterozygous individuals in this sample population being reduced to about half of that for homozygous reference individuals. W193C (frequency = 0.075) also appears to be associated with low ASA enzyme activity with heterozygous individuals having about 75% of the ASA enzyme activity of a W193W individual. However, 1101+123C>T (frequency = 0.080) heterozygous individuals have higher ASA enzyme activity levels than homozygous C individuals indicating this SNP may be associated with higher ASA levels. A sample population that included 1101+123 T homozygous individuals would be useful in elucidating this situation.
Figure 3.6 Mean ASA Enzyme Activity of Each Genotype in the Sample Population

The mean ASA enzyme activity for each genotype for the twelve SNP loci was compared using the Compare Means, One-Way ANOVA function in the SPSS statistical analysis program. The statistical significance of the result was calculated using linear regression. The same twelve SNPs were included in the haplotype analysis. The red star indicates the SNPs (W193C, 1102+123C>T, 2723A>G) that are significantly correlated with ASA enzyme activity in one or more of the analyses performed.
3.3.3 Analysis of the GBA Gene

The same approach used for the ARSA gene, was also used to analyse the GBA gene, but this was more challenging for several reasons. The primary one was the presence of the glucocerebrosidase pseudogene (GBAP) that is 96% homologous to GBA and is only 16 kb downstream on chromosome 1. Unless specific for GBA, PCRs would also amplify GBAP and any polymorphism detected could not be attributed to the gene or pseudogene. A whole gene PCR was needed that would specifically amplify GBA and exclude GBAP.

The GBA locus includes both the GBA and MTX pseudogenes (GBAP and MTXP) (Winfield et al., 1997). These resulted from duplication of a region which extends from approximately 5.5 kb 5' to GBA to exon 2 of MTX, with the consequence being the significant homology that has been noted between the GBAP and the GBA gene (Figure 3.7).

3.3.3.1 Development of the 'Whole Gene' Amplification Protocol for GBA and Internal GBA Fragment PCRs

Finckh et al. (1998) described “a protocol which allows the selective analysis of a PCR-amplified 7.1 kb genomic GBA-fragment encompassing the entire GBA coding region.” The same primers (F1, R0), PCR reaction mix and thermal cycling conditions as stated in the paper were reproduced. However, instead of the required 7.2 kb band, only smearing was seen after electrophoresis of the product on an agarose gel (data not shown). Two new forward primers (PrA-F, PrB-F) and a new reverse primer (R0-B) were designed and all six forward and reverse primer combinations were tried with a hot-start protocol but the same smeared result was obtained (data not shown). A temperature gradient to ascertain correct annealing temperature also produced similar results (data not shown). All three buffers available in the Expand Long Template PCR Kit (Roche) containing 1.75 mM MgCl₂, 2.25 mM MgCl₂, 2.25 mM MgCl₂ plus detergents were tried with all six primer combinations but smearing only increased (data not shown). Changing the number of cycles with two primer sets was also trialled (Figure 3.8, Gel 1). As less smearing was visible, it was decided to assess these whole gene PCR reaction products for specificity of GBA amplification.
The GBA locus at chromosome 1q21 contains seven genes within 70 kb including GBAP and the metaxin pseudogene (MTXP). Metaxin (MTX) and MTXP are transcribed in the opposite direction to GBA and the remaining genes at the locus. The duplication that resulted in the pseudogenes includes the region approximately 5.5 kb 5’ to GBA to exon 2 of MTX. GBAP shows 96% homology with GBA. The hashed bar between THBS3 and MTX indicates the existence of a promoter sequence that is common to both genes.
Figure 3.8 Development of a ‘Whole Gene’ PCR for Specific Amplification of GBA Gene

**Gel 1:** Two primer sets (shown as A and B) were used to amplify 50 ng human genomic DNA for 15, 21 and 25 cycles. Amplification of the GBA gene can be seen as a 7 kb band although contamination with GBAP product cannot be excluded at this point. **Gel 2:** In this exon 9 confirmatory PCR, a single band of 302 bp confirms the presence of the GBA gene and absence of the pseudogene that appears as a 247 bp fragment. In lane 2 (red arrow), there is a single 302 bp band indicating the original whole gene PCR contained GBA gene product only. All other lanes have 2 bands indicating the presence of both gene and pseudogene amplification product in the original whole gene PCR. **Gel 3:** Primer set B was used for a 15 cycle PCR of 3 samples. Specific amplification of the GBA gene in 3 samples is demonstrated with the confirmatory PCR (lanes 1-3). The robustness of the method is confirmed as specificity has been maintained despite doubling the template concentration for the same 3 samples (lanes 4-6).
A reduced number of cycles were used as non-specific priming increases with cycle number. The band of interest is at about 7 kb but it could not be determined whether this was the gene or pseudogene. One of the obvious points of difference between the gene and the pseudogene is a 55 bp deletion in exon 9 of the pseudogene (Figure 3.9).

The PCR for exon 9 using primers (F9; R10,9) as described by Finckh et al. (1998) was successfully reproduced. In exon 9 of GBAP, there is a 55 bp deletion. The result is that when human genomic DNA is used as the template, two fragments are produced: a 774 bp fragment from GBA and a 719 bp fragment from GBAP. The optimum annealing temperature was identified as 63.3°C instead of 69°C as published by Finckh et al. (1998). However, a fragment of less than 500 bp is required for DHPLC analysis and the authors did not explain the advantages, if any, of using the 774 bp exon 9 fragment as the confirmatory PCR. Therefore, as it is easier to detect a 55 bp difference between smaller fragments the F9, R9 primer combination for exon 9 was used. This PCR produces fragments of 302 bp from GBA and 247 bp from GBAP that are more easily differentiated by gel electrophoresis (Figure 3.8, Gel 2). That is, only one 302 bp band will be seen if the DNA template such as that from a GBA-specific ‘whole gene’ amplification is used but both bands will be seen if the pseudogene is also present.

In Figure 3.8, the six lanes in Gel 2 represent the internal exon 9 confirmatory PCRs performed on the products of the six PCRs seen in the photo of Gel 1. Five have produced double bands (302 bp and 247 bp) indicating presence of both GBA and GBAP, while in the second lane (red arrow) only a 302 bp band can be seen. This process was trialled on three samples and results can be seen in the photo of Gel 3 (Figure 3.8). The double band is only present in the positive control (genomic DNA). It can also be seen that despite increasing the amount of whole gene PCR used as template in the internal PCR (Figure 3.8, Gel 3, lanes 4-6), a second band of 247 bp cannot be seen indicating the development of a successful protocol for specific ‘whole gene amplification’ of GBA.
Figure 3.9 The GBA-Specific Exon 9 Confirmatory Fragment

Within exon 9 of the GBA4, there is a 55 bp deletion that differentiates it from GBA. GBA4 is represented in black and GBA4P in blue. The yellow highlighted sequence is the 55 bp that are not present in GBA4P. The underlined sequences are the forward and reverse primers for the exon 9 fragment. The GBA exon 9 fragment is 302 bp and the GBAP exon 9 fragment is 247 bp.
Non-specific amplification was demonstrated in five of the six different whole gene PCRs that were trialled (Figure 3.8, Gel 2). Increasing the primer concentration, introducing a hot-start step, minor modifications to the thermal cycling conditions and determining the optimum annealing temperature contributed to optimizing the parameters used for the one apparently successful ‘whole gene amplification’ PCR for GBA. However, the most important optimization step was the reduction of the number of PCR amplification cycles to fifteen as doing more than fifteen cycles was found to introduce non-specific priming.

Using this procedure (Materials and Methods 2.6.2 and 2.6.2.1), the entire GBA gene was amplified specifically from twenty high GBA activity samples. The exon 9 PCR (Materials and Methods 2.6.2 and 2.6.2.1) was used to verify the specificity of the reaction and the specific ‘whole GBA gene’ PCR reaction product was diluted for use as template for all thirteen internal fragment PCRs (Materials and Methods 2.6.2 and 2.6.2.1), thereby excluding the possibility of detecting GBA polymorphisms.

Twenty ‘high activity’ samples were screened for polymorphisms by DHPLC and about 25% of fragments were sequenced to confirm genotype. Only two polymorphisms were found, one in intron 4 (454+47G>A) and one in the 3′-UTR (1608+168T>C).

### 3.3.3.2 Analysis of GBA Fragments

The thirteen fragments covering all exons and the 5′–flanking region were screened for polymorphisms at the empirically determined partial denaturation temperature on the DHPLC (Varian). In addition, about 25% of fragments were sequenced with an average of three fragments per sample. Despite this, only two polymorphisms were found; a G>A polymorphism in intron 4 (454+47G>A) and a T>C polymorphism in the 3′-UTR (1608+168T>C). The intronic polymorphism has been previously described and does not appear to be within a cryptic splice site. The 3′-UTR polymorphism is over 500 bp beyond the major polyadenylation signal, so it is unlikely that either polymorphism would be affecting gene expression or mRNA stability, therefore no further analysis was done on these polymorphisms and no further samples were analysed.

### 3.3.4 Does a high activity allele exist?

If a high activity allele/s exist/s for the ARSA and GBA genes, they are rare. Any alleles having a marked effect on enzyme levels, such that they play a major role in determining
the overall distribution of enzyme levels in the population, would have a significant frequency. Therefore, there was no further analysis of these genes for high activity SNPs. Nonetheless, the hypothesis that some of the variation seen in lysosomal enzyme activity levels in the normal population is due to polymorphic sequence variants and that some of these do alter the primary amino acid sequence of the enzyme has been established but only with respect to polymorphic sequence variants associated with low activity. It has previously been demonstrated that not only can the PD allele and other ‘harmless’ polymorphisms contribute to increased disease severity in MLD, they can also combine with other polymorphisms to actually cause disease in the absence of recognised mutations. That is, polymorphisms are partly responsible for the low activity seen in both disease alleles and normal alleles.

3.3.5 Investigation of Environmental Contribution to Enzyme Activity Level in Individuals

To determine the feasibility of doing a longitudinal and/or traditional twin study to investigate the level of environmental contribution to an individual’s enzyme activity level, a pilot study was undertaken. The enzyme assays were set up according to the Standard Operating Procedures (Material and Methods 2.3) obtained from the National Reference Laboratory for lysosomal storage disorders. In addition to β-glucocerebrosidase and arylsulphatase A, it was decided to include a third lysosomal enzyme that could act as a reference or control if needed. α-galactosidase was chosen as the additional assay. A deficiency of α-galactosidase causes Fabry disease, a disorder for which enzyme replacement has recently become available.

Enzyme assays for all three enzymes were performed on the same day over three consecutive weeks on four healthy volunteers to ascertain the level of variation in enzyme activity in each individual over that time period. The four volunteers consisted of a male (M1) and female (F1) between 45 and 50 years, and a male (M2) and female (F2) between 20 and 25 years. The enzyme levels obtained for each individual over the time period can be seen in Table 3.8 and are presented graphically in Figure 3.10. To ascertain inter-assay variation, a sample was collected prior to the trial from one of the participants and run as the control with each assay. The co-efficient of variation (CV) for the control is a measure of inter-assay variation and this can be compared to the CV for each individual (Table 3.8).
Table 3.8 Enzyme Activity Levels for ASA, GBA and α-GAL in Four Individuals Over Three Consecutive Weeks

<table>
<thead>
<tr>
<th>Sample</th>
<th>ASA (Week 1)</th>
<th>ASA (Week 2)</th>
<th>ASA (Week 3)</th>
<th>ASA Mean</th>
<th>ASA CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>5.4</td>
<td>8.1</td>
<td>4.5</td>
<td>6.9</td>
<td>0.3</td>
</tr>
<tr>
<td>F2</td>
<td>4.6</td>
<td>5.4</td>
<td>4.2</td>
<td>3.5</td>
<td>0.1</td>
</tr>
<tr>
<td>M1</td>
<td>3.6</td>
<td>4.3</td>
<td>3.3</td>
<td>4.0</td>
<td>0.1</td>
</tr>
<tr>
<td>M2</td>
<td>4.8</td>
<td>4.3</td>
<td>4.3</td>
<td>4.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Control</td>
<td>5.4</td>
<td>6.2</td>
<td>5.3</td>
<td>5.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>GBA (Week 1)</th>
<th>GBA (Week 2)</th>
<th>GBA (Week 3)</th>
<th>GBA Mean</th>
<th>GBA CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>201</td>
<td>307</td>
<td>191</td>
<td>233</td>
<td>0.3</td>
</tr>
<tr>
<td>F2</td>
<td>202</td>
<td>181</td>
<td>204</td>
<td>195</td>
<td>0.1</td>
</tr>
<tr>
<td>M1</td>
<td>167</td>
<td>196</td>
<td>170</td>
<td>178</td>
<td>0.1</td>
</tr>
<tr>
<td>M2</td>
<td>180</td>
<td>154</td>
<td>160</td>
<td>165</td>
<td>0.1</td>
</tr>
<tr>
<td>Control</td>
<td>206</td>
<td>264</td>
<td>252</td>
<td>240</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample</th>
<th>α-GAL (Week 1)</th>
<th>α-GAL (Week 2)</th>
<th>α-GAL (Week 3)</th>
<th>α-GAL Mean</th>
<th>α-GAL CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>1.6</td>
<td>2.7</td>
<td>1.6</td>
<td>2.0</td>
<td>0.3</td>
</tr>
<tr>
<td>F2</td>
<td>1.4</td>
<td>1.6</td>
<td>1.4</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>M1</td>
<td>1.4</td>
<td>1.9</td>
<td>1.4</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>M2</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Control</td>
<td>1.5</td>
<td>1.9</td>
<td>1.7</td>
<td>1.7</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The ASA, GBA and α-GAL enzyme activity levels in nmol/min/mg protein, pmol/min/mg protein and nmol/min/mg protein respectively, were measured in four individuals over three consecutive weeks. The same control sample was run with each assay on each occasion. The mean enzyme activity and coefficient of variation (CV) for each enzyme in each individual and the control has been calculated.
Figure 3.10  Lysosomal Enzyme Activity Levels in Four Individuals Over Time

Arylsulphatase A (ASA), glucocerebrosidase (GBA) and α-galactosidase (α-Gal) levels were assayed in two female volunteers (F1, F2) and two male volunteers (M1, M2) over three consecutive weeks. The control (Ctrl) consisted of single collection of blood taken from an individual prior to the testing period and run each week as a control to assess inter-assay variation.
Arylsulphatase A Levels in Individuals over Time

Glucocerebrosidase Levels in Individuals over Time

α-Galactosidase Levels in Individuals over Time
Enzyme activity levels appear to show the least variability in the younger individuals with the greatest variability being seen in the levels of all enzymes for individual F1 (Figure 3.10). Hormonal influences are known to cause variation in arylsulphatase A levels with ASA activity highest at ovulatory phase and lowest at early follicular phase and in post menopausal women (Oner et al., 1994). Other lysosomal enzymes may also be affected. However, it appears that age may have more influence on variation in these lysosomal enzyme activity levels than gender as more variability in activity levels is seen in both of the older individuals than in the two younger ones. The control CV for each assay was 0.1 and for the individuals, eight of twelve CV results were the same or lower (Table 3.8). Therefore, although this study involved a small number of individuals over a short time period, it can be concluded that intra-individual variation falls mostly within the range of inter-assay variation.

As this study was very labour intensive, time consuming and mostly revealed only minor changes in enzyme levels within an individual, it was concluded that there would be limited benefit from a larger study. However, it may still be useful to know whether the variation seen in these enzyme levels between individuals is also reflected at the mRNA level. It would clearly be easier to see if changes in enzyme activity levels are reflected in changes in mRNA levels in high and low activity individuals. Therefore, four skin fibroblast cell lines, from high and low activity individuals, were sourced from the Women’s and Children’s Hospital in Adelaide. In addition, a further sample of blood was collected from each of the four volunteers used in the previous study.

3.3.5.1 Quantitation of ASA and GBA mRNA Levels in Leucocytes and Skin Fibroblasts in Four Individuals

The enzyme activity and mRNA levels of ASA and GBA (Materials and Methods 2.3.3 and 2.3.4) were measured in leucocytes from a single sample of blood. The results are presented in Figure 3.11. It appears possible that in three of the individuals – F1, F2 and M2, there may be a trend to direct correlation between mRNA levels and enzyme activity levels. However, the individual, M1, has very high mRNA levels but these have not translated to higher enzyme activity levels. This could be caused by translational inefficiency or experimental error but, as these are the results of a single sample, they would need to be replicated to be significant.
The enzyme activity and mRNA levels of ASA and GBA were also measured in the four skin fibroblast cell lines obtained from the Women’s and Children’s Hospital, Adelaide. The results of these cell lines (SF3351, SF3542, SF3693, SF3785) are presented in Figure 3.12. These results are the mean of six replicates per cell line. Although only four cell lines have been assayed, there appears to be a trend towards direct correlation between ASA mRNA and enzyme activity in all cell lines. This also appears to be the case for three of the four cell lines for GBA mRNA and enzyme activity. However for SF3785, the enzyme activity results obtained during this study are lower than expected based on mRNA levels. This skin fibroblast cell line was selected by WCH as a high GBA activity cell line (46 nmol/min/mg protein) and was expected to have a GBA enzyme activity level similar to SF3542 (42 nmol/min/mg protein) and higher than SF3351 and SF3693 (20 nmol/min/mg protein and 16 nmol/min/mg protein respectively) (Figure 3.13). A different method was used to assay glucocerebroside enzyme activity levels in this study than at WCH where a natural substrate and radioimmunoassay are used. This accounts for differences in the absolute values obtained for each skin fibroblast cell line, although the relative values obtained should be independent of method. In addition to the factors that affect enzyme activity which are under investigation in this study, experimental factors such as variation due to cell culture conditions may be the cause of the discrepancy.

The variation seen in enzyme activity levels between individuals is also reflected at the mRNA level for both enzymes and the expression pattern of each enzyme appears similar in a given individual. That is, the individual with the lowest ASA mRNA and enzyme activity also has the lowest GBA mRNA and activity in these cell lines. There are also cell specific differences in the expression of these enzymes with ASA expression being about four-fold higher in skin fibroblasts than leucocytes and GBA expression being at least 10-fold higher in skin fibroblasts.
Figure 3.11 Lysosomal Enzyme Expression in Leucocytes

The enzyme activity and mRNA levels for arylsulphatase A (ASA) and glucocerebrosidase (GBA) were measured in the leucocytes of two females (F1, F2) and two males (M1, M2). The individuals, F1 and M1 are aged 45 to 50 years and the individuals F2 and M2 are aged 20 to 25 years. These are the results of a single sample from each individual.
Arylsulphatase A Expression in Leucocytes

Glucocerebrosidase Expression in Leucocytes
Figure 3.12 Lysosomal Enzyme Expression in Skin Fibroblasts

The mean ASA and GBA enzyme activity and mRNA levels in four skin fibroblast cell lines (SF3351, SF3542, SF3693, SF3785) from four individuals. The results for each cell line are the average of six replicates.
Arylsulphatase A Expression in Skin Fibroblasts

Glucocerebrosidase Expression in Skin Fibroblasts
Figure 3.13 Lysosomal Enzyme Expression in Skin Fibroblasts (method 2)

The GBA enzyme activity levels as supplied by WCH are plotted with the mRNA levels as in the first graph. Of note is that SF3785, had a higher GBA enzyme activity level when initially tested at WCH. The subsequent lower result may be the cause of some apparent change in culture conditions and is unlikely to be the result of experimental error as six replicates were done.
3.4 DISCUSSION

In this chapter, the search for high activity polymorphisms in the ARSA and GBA genes is described followed by an investigation of the causes of the variation seen in ARSA and GBA gene expression in individuals.

Fifteen polymorphisms were detected in the ARSA gene in the sample population (N = 113). Twelve of these were validated for inclusion in further statistical analysis. Analysis of the individual ASA SNPs for association with enzyme activity revealed significant values for 2723A>G, N350S and W193C while 1101+123C>T was close to significant. All analyses performed on the ASA SNPs as a group confirmed the significant correlation 2723A>G has with low ASA enzyme activity. In addition, an intronic SNP (1101+123C>T) and W193C were identified in separate analyses as having an effect on enzyme activity independent of 2723A>G. The effect may be at or near the identified loci. The linear regression analyses that identified 1101+123C>T was close to including W193C. Testing of a larger number of samples or functional studies will be required to elucidate the association of W193C and 1101+123C>T. All three polymorphisms are low frequency (<0.1).

Strong linkage disequilibrium between 2723A>G and N350S was observed. This was expected because these two polymorphisms comprise the PD allele that is known to cause ASA pseudodeficiency (Gieselmann et al., 1989). Although N350S appeared to be significantly associated with enzyme activity in the individual SNP analysis, this association was not significant in any of the combined tests. However, an N350S heterozygote is expected to have approximately 79% of normal ASA activity (Harvey et al., 1998). N350S may not have been identified as associated with low enzyme activity in the combined tests as with a frequency of just 8.4%, the sample population was not large enough to produce a statistically valid association for a reduction of 20% of ASA activity. Furthermore, the effect is additive (to that of 2723A>G) rather than independent when they occur together. It is also likely that the method used to measure ASA activity did not distinguish between correctly and incorrectly targeted ASA and simply measured total ASA activity. This is because the N350S polymorphism removes a glycosylated asparagine residue (Herz and Bach, 1984) and this may affect the targeting and the catalytic activity of ASA. In vitro expression of the N350S ASA enzyme demonstrated that only about 45% reaches the lysosome (Harvey et al., 1998).
Linkage disequilibrium analysis found strong linkage between \(-720\text{C}>\text{T}\) and \(1102\text{-32T}>\text{C}\), \(-720\text{C}>\text{T}\) and \(-789\text{G}>\text{A}\), and \(-720\text{C}>\text{T}\) and \(\text{T391S}\) although no association with an increase or decrease in enzyme activity was identified. None of the 5'-flanking or 5'-UTR polymorphisms show any association with high or low ASA enzyme activity despite an indication previously that there may have been an association (Mason, 2001). The analysis of a larger number of samples has proved to be valuable in clarifying this situation.

In the \textit{GBA} gene, only two polymorphisms were detected in high activity individuals and neither of these polymorphisms appear likely to be associated with high enzyme activity. This result is consistent with the recent finding that the vast majority of \textit{GBA} mutations affect the global conformation of the enzyme leading to decreases in catalytic power (Liou \textit{et al.}, 2006) and the reported existence of only two major haplotypes and two minor haplotypes for \textit{GBA} (Beutler \textit{et al.}, 1992), although it was expected that a larger number of polymorphisms would have been superimposed on the basic haplotypes.

It appears that no common high activity polymorphisms exist for \textit{GBA} or \textit{ARSA}. Even if further analysis showed that the \(1101+123\text{C}>\text{T}\) polymorphism was significantly associated with high ASA enzyme activity and functional studies confirmed this, with a frequency of less than ten percent, the overall effect it is having on the population distribution of enzyme activity levels is small. Any alleles having a marked effect on enzyme levels, such that they play a major role in determining the overall distribution of enzyme levels in the population, would have a significant frequency. Therefore, the continuation of the search for high activity SNPs in these genes was not justifiable.

In spite of this, the hypothesis has not been disproved. Correlation was found for the \textit{ARSA} gene between certain polymorphisms and low activity or high activity, and these will influence the distribution of enzyme activity levels. However, the question of what causes the wide distribution in activity levels and especially the high activity levels seen in some of the normal population remains unanswered. This is especially true for the \textit{GBA} gene which seems remarkably conserved, at least in the high activity individuals as no low GBA activity individuals were screened.
The screening of the GBA gene was completed relatively quickly because of the development of a simple yet very robust ‘whole gene PCR’ method. The protocol described by Finckh et al. (1998) was trialled unsuccessfully. However, the protocol developed for this project was used successfully on 113 population samples with confirmation of specificity demonstrated through use of a confirmatory PCR. This gave a clear result in a shorter time because a smaller exon 9 fragment PCR was used.

It is clear that the investigation into the cause/s of variation in enzyme activity levels between individuals needed to be widened to further evaluate other genetic and/or environmental influences. A longitudinal study involving many individuals including twins and siblings could potentially provide valuable information but this would be time consuming, costly and difficult to co-ordinate within the required time frame. In addition, although the enzyme assays were set up according to the Standard Operating Procedures provided by the WCH, they were only performed for the purposes of this research and with the limited number of specimens that needed testing for this study. Therefore, evidence of their accuracy, precision and reliability would be required before investing the required resources for such a large study. It would also have been necessary to collect and develop a number of control specimens to ensure consistency between assays. It was with these factors in mind that the pilot study was undertaken and although too small to be of any real value, this study of intra-individual variation is still useful. If correct, it would suggest that the mean GBA and ASA enzyme activity for an individual is predominantly genetically determined. There will be differences between the mean enzyme activity levels of individuals within the range of normal enzyme activity and variation around that mean may also be caused by environmental factors.

Further investigation of the regulation of the ARSA and GBA genes is warranted as variation was seen in the mRNA levels between individuals and between the two cell lines assayed. The mRNA and enzyme activity levels of both ASA and GBA are much lower in leucocytes than in skin fibroblasts. It also appears possible that there is a direct correlation between mRNA and enzyme activity levels for both ASA and GBA, i.e. higher mRNA levels lead to higher enzyme activity levels. However, this apparent correlation did not hold in all cases for GBA so regulation at the post-transcriptional level should also be investigated.
3.5 CONCLUSION
For the ARSA gene, there is strong evidence of an independent effect on enzyme activity originating at or near three ASA SNPs. The correlation of at least two of these SNPs with low ASA activity will influence the distribution of ASA enzyme activity levels. However, a larger sample size or functional studies are required to fully elucidate these effects. There is also the potential for one of the ASA SNPs to be associated with higher ASA enzyme activity. Although GBA has only two common haplotypes, the lack of additional polymorphisms in the GBA gene was somewhat surprising and suggests significant conservation of this gene. In this study, it has been shown that there are no common polymorphisms associated with high enzyme activity for the ARSA and GBA genes. The variation in enzyme activity levels in individuals was also reflected at the mRNA level and cell specific differences were also observed. If the variation in enzyme activity levels between individuals is predominantly genetically determined as indicated by the small pilot study, an investigation of transcriptional and translational regulation and the causes of cell specific differences is warranted.
CHAPTER 4

TRANSCRIPTIONAL REGULATION OF THE
GLUCOCEREBROSIDASE AND ARYLSULPHATASE A
GENES
4.1 INTRODUCTION

Measurement of enzyme activity is an end-point measurement of gene expression. To fully understand the factors that influence gene expression such as transcription, translation and post-translational processes, separate investigations of each are required. Transcription, the first major step in the process, will be influenced by the strength of the elements within the promoter along with the presence or absence of enhancer and inhibitor sequences. Overall control of transcription facilitates spatial, temporal, tissue and cell-specific expression.

Both *ARSA* and *GBA* genes have previously been described as ‘housekeeping genes’ as the ASA and GBA enzymes appear to be ubiquitously expressed in human tissues. However, the promoters of these two genes differ significantly. The ASA promoter conforms most closely with that of other ‘housekeeping’ genes as it does not have the typical TATA and CAAT box sequences but does include several potential Sp1 binding sites (Kreysing *et al.*, 1990), while the GBA promoter has two CAAT and two TATA boxes (Reiner *et al.*, 1988b). Consistent with its promoter structure, the expression of GBA has been found to be predominantly tissue-specific (Reiner and Horowitz, 1988a). That is, it is expressed in all tissues but levels do vary. However, there have been no previous studies of transcriptional regulation of ASA expression.

Evidence of the regulation of GBA expression has been accumulating. Early reports that GBA activity differed according to the tissue of origin were confirmed by several authors (Beutler *et al.*, 1976, Stephens *et al.*, 1981, Takahashi *et al.*, 1990). GBA mRNA levels were also found to vary from very low in B-cells, to high in skin fibroblasts and higher in epithelial cells (Reiner and Horowitz, 1988a) indicating that there is regulation at the level of transcription. This variation was further investigated and up to 50-fold variation in mRNA levels and up to 60-fold variation in enzyme activity levels was reported (Doll and Smith, 1993). Evidence of this variation can be seen in the two cell types, skin fibroblasts and leucocytes, assayed in Chapter 3, where at least four-fold differences were seen in GBA mRNA levels and ten-fold difference in enzyme activity levels were noted.

Three different expression patterns for GBA were identified by Doll and Smith (1993) wherein mRNA and enzyme activity levels were directly correlated (group A), or enzyme activity levels were comparatively increased or decreased (groups B and C). As
transcriptional regulation of mRNA levels appeared to be the most likely source of the variation observed in the group A cell lines. Further analysis by Doll et al. (1995) identified the regions of the promoter that are required for functional activity in these cell lines. It was concluded that elements within the 622 bp GBA fragment that was investigated, played a role in determining differential expression. It was further proposed that it was subtle changes in the magnitude of the effect of these elements and not the activity of different elements in different cell types that controlled the differential expression of GBA (Doll et al., 1995). Doll and Smith (1993) also reported a 50-fold variation in GBA mRNA levels suggesting that the subtle changes in the effects of the inhibitory and stimulatory elements are more than subtle. The most important regions for controlling promoter activity were in exon one although none of the inhibitory or stimulatory elements were specifically identified as binding to any known transcription factors (Doll et al., 1995).

Moran et al. (1997) identified several transcription factors including OCTA binding protein, AP-1, PEA3 and CAAT binding protein that participate in regulating the expression of GBA. It was proposed that differential expression was caused by different levels of these transcription factors. The AP-1 binding site is the only site that falls within any of the inhibitory or stimulatory elements identified by Doll et al. (1995). The same region of GBA has been analysed in all investigations to date and several structural elements have been identified. However, no clear picture has emerged of how the GBA gene is regulated at the level of transcription, how tissue specificity is achieved nor why variations in enzyme activity levels exist between individuals.

To further elucidate this problem, a thorough investigation of the control of GBA expression was commenced, initially, by the characterization of GBA expression in several cell lines that were likely to provide examples of the 3 different expression patterns as described by Doll and Smith (1993). In addition, as the GBA promoter is atypical for a ‘housekeeping’ gene, the promoter and 5′-flanking region were analysed to more fully elucidate transcriptional regulation of the GBA gene. ASA mRNA levels were also measured in the same cell lines as used for investigation of GBA expression. This provided information on transcriptional regulation of the ARSA gene and a reference against which the GBA mRNA information may be compared to further elucidate the source of the observed variation.
4.2 MATERIALS AND METHODS

The materials and methods specific to this chapter are described in detail below. General materials and methods used in this thesis are outlined in detail in Chapter 2.

4.2.1 Promoter Construct Design

The region between \textit{GBA} and the gene immediately 5′ of \textit{GBA}, \textit{C1orf2} (also called \textit{Cote 1}) was analysed. This region of approximately 6.5 kb was interrogated for restriction enzyme (RE) sites that were compatible with the RE cloning site in the pGL3-Basic Luciferase reporter gene vector. Any RE sites in the Luciferase vector that were also located in the \textit{C1orf2} to \textit{GBA} region were not suitable to use for cloning. It was also important that the promoter or 5′–UTR region to be analysed, be inserted in the direction of transcription. The Web-Cutter software (http://www.firstmarket.com/cutter/cut2.html) was used to analyse the \textit{Cote 1} to \textit{GBA} region for restriction sites. This was done with the default settings to identify enzymes that did not cut or that cut only once but not within the region of interest. The pGL3 Luciferase vectors used in this study contain a multiple cloning site (MCS) that facilitates the insertion of a specific DNA sequence to be analysed.

4.2.2 Acquisition of Fosmid and Preparation of DNA

To ensure the accurate replication of the 5′–UTR region of the \textit{GBA} gene, a fosmid that did not include the \textit{GBA} pseudogene, was sourced (Figure 4.1). The fosmid, G248P86705B7, encompassing a genome sequence of 37,456 bp from chromosome 1:152,013,233 to 152,050,688, was obtained from The Wellcome Trust Sanger Institute, Cambridge, United Kingdom. The preparation of fosmid DNA was performed as described in Chapter 2 (Materials and Methods 2.13.1).
Figure 4.1 The GBA Fosmid

The fosmid, G248P86705B7, encompassing a genome sequence of 37,456 bp from Chromosome 1:152,013,233 to 152,050,688, was obtained from The Wellcome Trust Sanger Institute, Cambridge, United Kingdom. It included the *GBA* gene but not the *GBA* pseudogene ensuring accurate replication of the 5′-UTR region of the *GBA* gene.
4.2.3 PCR Amplification of the GBA 5′-UTR

To avoid introducing errors via PCR, significant efforts were made to generate constructs using restriction digests of the fosmid DNA but this approach proved to be unsuccessful due to problems with the ligation of vector and insert. PCR was performed basically as described in Chapter 2 (Materials and Methods 2.6) using the “proof-reading” enzyme, Pfx and primers that incorporated the restriction site. Table 4.1 lists the constructs, the forward primers and the common reverse primer. The underlined region is the restriction site. Thermal cycling conditions were 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 15 sec, annealing at 64-68°C for 30 sec, extension at 68°C for 30-150 sec. All PCR amplicons were electrophoresed on a 0.8%, 1% or 2% agarose TBE-gel to ascertain quality and were purified through a Roche PCR Purification system as outlined by the manufacturer (Roche).

4.2.4 Cloning PCR Amplicons into pGL3-Basic Luciferase Vector

PCR amplicons were cloned into the pGL3-Basic vector (Promega) and transformed into JM109 E.coli cells (Life Technologies) as outlined in Chapter 2 (Materials and Methods 2.13). Preliminary screening using insert-specific PCR was used to identify vectors containing an insert. Purification of these vectors was performed as described in Chapter 2 (Materials and Methods 2.13.8). Isolated clones were sequenced in both directions with pGL3-Basic Luciferase primers that flank the multiple cloning site (MCS) i.e. (RVprimer3-5′-CTAGCAAAATAGGCTGTCCC-3′ and Rvprimer4-5′-GACGATAGTCATGCCCCGCG-3′).

4.2.5 Transfection of the GBA Constructs into Cell Lines and Analysis of Luciferase Activity

The transfection of the GBA promoter constructs into the attached cell lines (SW1353, U118, Hep G2 and Huh7) and the measurement of luciferase activity was performed as outlined in Chapter 2 (Materials and Methods 2.14). Briefly, the U118, Hep G2 and Huh7 cell lines were seeded at approximately 2 x 10^5 cells in 24-well plates in RPMI (Gibco, BRL Life Technologies, Melbourne, Australia). After 20 to 24 hours incubation in 5% CO₂ at 37°C, 2 µg of pGL3-GBA DNA construct or 2 µg of pGL3-Basic control DNA vector, was diluted in 100 µl of Opti-MEM-I Reduced Serum Medium (Life Technologies) with 3 µl of Lipofectamine 2000 (LF2000, Life Technologies) and added to the cells.
Table 4.1 Primers for pGL3-GBA Constructs

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5’ to 3’</th>
<th>Restriction Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBAPCRR2</td>
<td>5’-gggAAgCTTACTAAACAAAAACAAggATgC-3’</td>
<td>HindIII</td>
</tr>
<tr>
<td>GBACAATXHOF</td>
<td>5’-gggCTCgAgCAAATCTgTgTTCTAggCTCT-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA328XHOF</td>
<td>5’-gggCTCgAgTCAAAAAAgAAgTgTgCCTgC-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA538SACF</td>
<td>5’-gggCTCgAgTggTCCCTCCTCATCTCAgA-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA742F</td>
<td>5’-gggCTCgAgAAggAAgAACAgTTATTACCA-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA1365F</td>
<td>5’-gggCTCgAgAAAgAAACAAAAGAgATggCTg-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA1956F</td>
<td>5’-gggCTCgAgTCAAggAACCACTCAGAgAACAgA-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA2474F</td>
<td>5’-gggCTCgAgCCgTCTCTACAAAAATAAACAg-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA3559XHOF</td>
<td>5’-gggCTCgAgACTTTAggAAgggCTCTg-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA4700Nhe1F</td>
<td>5’-gggCTAggCCAgCAATgAAACCCT-3’</td>
<td>NheI</td>
</tr>
<tr>
<td>MIDIXHO1R</td>
<td>5’-gggCTCgAgTgggCgggACTgAgACT-3’</td>
<td>XhoI</td>
</tr>
<tr>
<td>GBA6519H</td>
<td>5’-gggAAgCTTTgACAAAAAgAAACCTgC-3’</td>
<td>HindIII</td>
</tr>
</tbody>
</table>

The common reverse primer, GBAPCRR2, was used with all other forward primers to amplify fragments of the lengths indicated in their name. In addition, a 1313 bp fragment was amplified using the forward and reverse primers, GBA4700Nhe1F and MIDIXHO1R respectively. The restriction sites (underlined) were used for cutting and ligating the fragment into the appropriate pGL3 vector.
To monitor transfection efficiency, 10 ng of control *Renilla* vector (pRL-TK, Promega) was co-transfected with the pGL3-GBA constructs or the pGL3-Basic control vector.

After 24 hr, the transfection medium was removed. If serum stimulation was required, 1 ml of complete RPMI (Life Technologies) or complete DMEM (Life Technologies) was added and incubation in 5% CO$_2$ at 37°C continued for 24 hr. This step was repeated if 48 hr serum stimulation was required. Following removal of the growth medium, 100 µl of passive lysis buffer (PLB, Promega) was added. The luciferase activity was read on a luminometer following addition of 100 LARII (Luciferase Assay Reagent, Promega) to a 20 µl aliquot of the lysate, and *Renilla* luciferase activity was measured following addition of 100 µl of Stop and Glo Reagent (Promega).
4.3 RESULTS

4.3.1 Choosing Cell Lines in which to Investigate Regulation

Previous work by Doll and Smith (1993) identified three expression patterns of GBA, so it was appropriate to investigate at least one cell line that was representative of each expression pattern. Therefore, human cell lines from tissues that have previously been shown to exhibit transcriptional and/or post-transcriptional regulation of GBA were established. Cell lines in which transcriptional regulation of GBA had been demonstrated included the neuronal cell lines, glioblastoma (U-87MG) and astrocytoma (CCF-STTG1), with high mRNA levels, lymphoblasts (GM621A) with low mRNA levels and macrophages (U937) with intermediate mRNA levels (Doll and Smith, 1993). The human glioblastoma cell line (U118 MG) and EBV-transformed B-cells were obtained as representatives of this group. Translational control has been observed in fibroblasts (GM5659B), promyelocytes (HL60) and monocytes (THP-1) (Doll and Smith, 1993). The skin fibroblast lines (SF) from four individuals (as used in Chapter 3) along with HL60 and THP-1 lines were considered representative of this group. Human embryo kidney cells are reported to have high levels of TCP80, the translational control protein that inhibits GBA translation. This cell line was not available. Additional cultured cell lines, the hepatoma cell lines (Hep G2 and Huh7) and chondrosarcoma cell line (SW1353) were included as the visceral organs (liver and spleen) and bone are affected tissues in type 1 GD.

All cell lines were maintained in appropriate media (Materials and Methods 2.12).

4.3.2 Variation in mRNA Levels

Regulation of transcription is observed as variation in mRNA levels between cell types or tissues. Table 4.2 presents the levels of ASA and GBA mRNA measured in a panel of nine human cell types by quantitative real time PCR (Materials and Methods 2.12.3). The final values are the mean of at least four replicates. For skin fibroblasts and leucocytes, replicates were done for the cell sample for each individual and the results of the four individuals averaged to obtain results for that cell type. Figures 4.2 and 4.3 depict the mRNA levels of ASA and GBA respectively in each of the nine cell types.
Table 4.2  The Levels of ASA and GBA mRNA in Nine Human Cell Lines

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>mRNA molecules/µg RNA</th>
<th>ASA</th>
<th>GBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leucocytes</td>
<td>1.15 x 10^5 +/- 2.4 x 10^4</td>
<td>1.30 x 10^5 +/- 3.3 x 10^4</td>
<td></td>
</tr>
<tr>
<td>THP1</td>
<td>1.40 x 10^5 +/- 1.2 x 10^4</td>
<td>7.67 x 10^5 +/- 1.5 x 10^5</td>
<td></td>
</tr>
<tr>
<td>HL60</td>
<td>1.46 x 10^5 +/- 1.8 x 10^4</td>
<td>7.98 x 10^5 +/- 2.2 x 10^5</td>
<td></td>
</tr>
<tr>
<td>B-cells</td>
<td>5.60 x 10^5 +/- 9.8 x 10^4</td>
<td>9.42 x 10^5 +/- 1.1 x 10^5</td>
<td></td>
</tr>
<tr>
<td>Skin Fibroblasts</td>
<td>4.00 x 10^5 +/- 2.9 x 10^4</td>
<td>1.38 x 10^6 +/- 9.1 x 10^4</td>
<td></td>
</tr>
<tr>
<td>U118</td>
<td>3.27 x 10^5 +/- 2.0 x 10^4</td>
<td>1.51 x 10^6 +/- 1.8 x 10^5</td>
<td></td>
</tr>
<tr>
<td>HuH7</td>
<td>3.15 x 10^5 +/- 8.7 x 10^4</td>
<td>1.69 x 10^6 +/- 2.7 x 10^5</td>
<td></td>
</tr>
<tr>
<td>Hep G2</td>
<td>2.55 x 10^5 +/- 3.8 x 10^4</td>
<td>1.77 x 10^6 +/- 2.4 x 10^5</td>
<td></td>
</tr>
<tr>
<td>SW1353</td>
<td>8.25 x 10^5 +/- 1.6 x 10^5</td>
<td>4.07 x 10^6 +/- 5.3 x 10^5</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.2 The Level of ASA mRNA in Nine Human Cell Types

The level of ASA mRNA was measured by real-time PCR in nine human cell types: leucocytes, monocytes (THP1), promyelocytes (HL60), hepatocytes (Hep G2 and Huh7), glioblastoma (U118), skin fibroblasts, B-cells and chondrocytes (SW1353). Each of these results is the mean of four to six replicates.
Figure 4.3 The Level of GBA mRNA in Nine Human Cell Types

The level of GBA mRNA was measured by real-time PCR in nine human cell types: leucocytes, monocytes (THP1), promyelocytes (HL60), hepatocytes (Hep G2 and Huh7), glioblastoma (U118), skin fibroblasts, B-cells and chondrocytes (SW1353). Each of these results is the mean of four to six replicates.
The mRNA levels of the lysosomal enzymes, ASA and GBA, are lowest in leucocytes and highest in chondrocytes (SW1353), varying seven-fold for ASA from $1.15 \times 10^5$ to $8.25 \times 10^5$ molecules of mRNA and varying more than thirty-fold for GBA from $1.30 \times 10^5$ to $4.07 \times 10^6$ molecules of mRNA. This wide variation seen in GBA levels suggests that GBA is the more highly regulated of these genes.

The comparison of the variation of mRNA levels for GBA and ASA may give some indication as to whether the observed regulation is affecting all lysosomal enzymes or is more gene specific, that is, affecting only GBA or ARSA. For example, in Figures 4.2 and 4.3, B-cells and skin fibroblasts can be seen to have higher levels of ASA mRNA but moderate levels of GBA mRNA production when compared to the other cell types. Having established the relative levels of GBA and ASA mRNA transcriptional regulation in various cell types, further analysis of the 5′-UTR was undertaken to search for tissue-specific elements.

4.3.3 GBA 5′-UTR Analysis Strategy

To analyse the GBA 5′-UTR, a series of fragments were designed to include incrementally larger distances upstream from the GBA cap site. There is a distance of approximately 6.5 kb to the nearest gene 5′ of GBA, C1orf2, that could potentially be investigated. Initially, approximately 3.5 kb was analysed with plans to increase the length of 5′-UTR sequence to be investigated as required. A region of 622 bp encompassing the GBA promoter and between the two Sac1 sites, had previously been investigated (Doll and Smith, 1993, Moran et al., 1997). Figure 4.4 and Figure 4.5a show the position of the constructs relative to the GBA gene.

4.3.3.1 Identification of Restriction Sites for the Cloning of GBA Constructs Into pGL3-Basic Luciferase Vector

To identify potential regions for the cloning of the GBA constructs into the pGL3-Basic luciferase vector, a restriction map was used to identify two restriction sites that cut the pGL3-Basic vector once each, and would either not cut the GBA sequence or would cut it only once at a suitable site. A restriction map analysis was performed using the Web Cutter software (www.firstmarket.com/cutter/cut2.html). This analysis showed that HindIII did not cut the GBA sequence and that there was a XhoI site approximately 5.5 kb upstream from the GBA cap site.
The sequence between \textit{C1orf2} and \textit{GBA} from AL713999 human DNA sequence from clone RP11-263K19 on chromosome 1 is annotated to display the structural features and the primers used to generate the GBA inserts for the pGL3-GBA constructs. The first \textit{GBA} exon is in bold and underlined, the cap site and first ATG for \textit{GBA} and the \textit{C1orf2} poly(A) signal are in red, the primers are in blue and annotated immediately to the right of the line.
Figure 4.5 The pGL3-GBA Plasmids – The Constructs

(A) A graphical representation of the GBA inserts is presented drawn to scale to show the length and positioning of each of the inserts with reference to the C1orf2 to GBA sequence can be appreciated. (B) Miniprep DNA from each of the pGL3-GBA plasmids was digested with the appropriate restriction enzymes to ensure an insert of the correct size was present. Two plasmids of each size were analysed. The uncut plasmid and the cut plasmid were electrophoresed in adjacent lanes on a 1% gel. Odd-numbered lanes contain the uncut plasmid and the even-numbered lanes contain the same plasmid cut with appropriate restriction enzymes with the size of the insert indicated. The size of the pGL3-Basic vector is 4818 bp.
A  GBA 5′-Flanking Region Inserts

Numbers indicate length of insert in base pairs.

X = XhoI
H = HindIII
N = NheI

\[
\begin{align*}
&258 \quad X \quad H \\
&328 \quad X \quad H \\
&358 \quad X \quad H \\
&742 \quad X \quad H \\
&1365 \quad N \quad H \\
&1956 \quad N \quad H \\
&2474 \quad N \quad H \\
&4700 \quad N \quad H \\
&6519 \quad H \\
&1313 \quad X \\
\end{align*}
\]

\[\text{Clorf2} \quad \text{Poly(A)} \quad \text{GBA} \quad \text{Cap site}\]

B  The pGL3-GBA Plasmids – Uncut and Post-digest.

Gel 1

Gel 2
It was necessary for the \textit{GBA} sequence to be inserted into the pGL3-Basic vector in the correct orientation, that is, in a 5′ to 3′ direction relative to the 5′ to 3′ direction of the luciferase gene. As the \textit{XhoI} and \textit{HindIII} sites were present in the multiple cloning site of the pGL3-Basic vector, the GBA constructs would be in the correct orientation if PCR primers were designed to include the \textit{XhoI} site at the 5′ end of the forward primer and the \textit{HindIII} site at the 5′ end of the reverse primer (Figure 4.5a). For constructs that were larger than 5.5 kb and included the native \textit{XhoI} site, a \textit{HindIII} site was used on both primers, thereby enabling a non-directional cloning strategy.

\subsection*{4.3.3.2 Amplification of GBA Constructs from Fosmid DNA}

Using the \textit{Pfx} proof-reading enzyme and gene specific primers containing the restriction sites \textit{XhoI} and \textit{HindIII}, constructs containing up to 3559 bp of the GBA promoter and 5′-flanking region were amplified using B7 fosmid DNA (Materials and Methods 4.2.3) to ensure there was no contamination with the \textit{GBA} pseudogene sequence. All products were electrophoresed on 0.8% to 2% agarose gels (Figure 4.5b).

\subsection*{4.3.3.3 Cloning of GBA Promoter Constructs Into the pGL3-Basic Luciferase Vector}

Each PCR product was purified using the Roche PCR Purification kit (Roche), double digested with \textit{XhoI} and \textit{HindIII} restriction enzymes, purified again with the Roche PCR Purification kit (Roche), and quantitated via electrophoresis on a 1-2% agarose gel against a quantitative DNA marker or a known quantity of cut pGL3-Basic vector. The purified GBA constructs were cloned into purified pGL3-Basic vector that had been previously digested with \textit{XhoI} and \textit{HindIII} restriction enzymes. Clones were screened via insert-specific PCR followed by a restriction digest of the pGL3 construct DNA with \textit{XhoI} and \textit{HindIII} restriction enzymes and electrophoresis of the product to check for the liberated insert.

\subsection*{4.3.3.4 Sequencing of pGL3-Basic Luciferase Constructs}

To ensure no errors were generated during the PCR or cloning procedures, all GBA constructs were sequenced completely in the forward direction and for 600-800 bp in the reverse direction from the anchor primer, to verify authenticity and analysed by BLAST (\url{www.ncbi.nlm.nih.gov}) against the human genome. All constructs were identical to a chromosome 1 contig, AL713999 Human DNA sequence from clone RP11-263K19, known to contain the \textit{GBA} genomic sequence. The constructs were named: pGL3-
CAAT, pGL3-328, pGL3-538, pGL3-742, pGL3-1365, pGL3-1956, pGL3-2474, pGL3-3559. An additional construct pGL3-1485 is a *Hind*III to *Hind*III construct, a by-product of the PCR and ligation of the product amplified with the GBA6519H forward and reverse anchor primers. The 258 bp fragment used for pGL3-CAAT contained the minimal promoter element as described by Doll *et al.* (1995).

### 4.3.4 Analysis of the GBA Promoter Activity

To analyse GBA promoter activity, the nine pGL3-Basic constructs were co-transfected with *Renilla* vector, pRL-TK (Promega) as an internal control, into cell lines exhibiting transcriptional control; SW1353, U118, HepG2, and Huh7. The synthetic *Renilla* Luciferase Reporter vector is widely used to normalize experimental variations including transfection efficiency in transient transfection experiments. The pGL3-control vector containing an SV40 promoter and enhancer sequences was also used to monitor transfection efficiency. A no-insert control was used to measure background luminescence. Transfection conditions were optimised for each cell line/transfection reagent combination. Four replicates of each test and control were done in each experiment.

Many unsuccessful attempts were also made to transfect the suspension cell lines, HL60, THP1 and B-cells, but these cell lines are notoriously difficult to transfect. Five transfection reagents - Lipofectamine (Life Technologies), Lipofectamine 2000 (Life Technologies), DMRIE C (Life Technologies), Fugene 6 (Roche) and ExGen 500 (Progen, Archerfield, Australia) were tried. Electroporation appears to be the only technique that may produce sufficient transfection efficiency to be useful for these cells (see [http://www.amaxa.com/celllinelist.html](http://www.amaxa.com/celllinelist.html)).
4.3.4.1 Comparison of Basal Expression Levels in Four Cell Lines

The basal expression levels of the nine GBA promoter constructs were compared to those of the control in each of the four cell lines, U118, HepG2, SW1353 and Huh7 (Figure 4.6a). The expression level of the shortest construct, pGL3-CAAT, containing the minimal GBA promoter element, ranges from 4 percent of the control in U118 cells to fifty percent of the control in HepG2 and SW1353 cells, and almost seventy percent of the control in Huh7 cells. The pGL3-328 construct has a similar expression level but the expression level gradually reduces as the length of the construct increases until it is about two to four-fold less for the longest construct, pGL3-3559, at less than two percent of the control in U118 cells, about thirteen percent of the control in HepG2 and SW1353 cells and approximately twenty percent of the control in Huh7 cells.

The strength of the GBA promoter appears to range from strong in Huh7 cells to very weak in U118 cells although, comparing the expression of the GBA promoter to that of the luciferase gene in different cell types may give an erroneous impression. When the expression of each construct is compared to the expression of the minimal promoter element in that cell line, all constructs of the same length have a very similar expression level as seen in Figure 4.6b.

4.3.4.2 Effect of Serum Stimulation in Hep G2 Cells

Serum inhibits the transfection efficiency of most transfection reagents. Therefore, all previous transfections were done in the reduced serum medium, OptiMEM (Life Technologies) and the luciferase assay performed without changing the medium to the usual complete growth medium. It was considered possible that only a basal level of expression was being observed due to the absence of growth factors or nutrients normally present in the serum. Hence, the effect of serum stimulation for 24 and 48 hours was investigated for a reduced number of constructs in the Hep G2 cell line.

The Hep G2 cell line was chosen because it transfected more reliably and efficiently than the other cell lines. The constructs were chosen as representative of constructs of a similar length. For example, the basal expression level of pGL3-328 is similar to that of pGL3-CAAT and as pGL3-CAAT is the smallest construct and contains the minimal promoter element, it was more prudent to exclude pGL3-328.
Figure 4.6  The Expression of the Nine GBA Promoter Constructs in Four Human Cell Lines

Each of the nine pGL3-GBA promoter constructs and the pGL3-control vector were transfected into four human cell lines: glioblastoma (U118), hepatocytes (Hep G2 and Huh7), and chondrocytes (SW1353). The pGL3-GBA promoter inserts ranged from 258 bp to 3559 bp. A luciferase assay was performed using Promega’s Dual-Glo Luciferase assay system.  (A) Expression of the GBA constructs within each cell line is compared to the expression of the pGL3-control. The pGL3-control luciferase vector was set at 100% and the no insert control was set at 0%. Note that the results for the U118 cell line have been increased 10-fold to enable more effective comparison between cell lines.  (B) Expression of the GBA constructs is compared to the expression of the pGL3-CAAT construct (set at 100%) within that cell line. The pGL3-CAAT contains the minimal GBA promoter element.
The constructs pGL3-1484 and pGL3-2474 were excluded because their expression levels were similar to pGL3-1365 and pGL3-1956, respectively (Figure 4.6). The experiment was set up in triplicate so that one set of transfections could be assayed without the addition of serum, while the second and third sets were assayed after serum stimulation for 24 hr and 48 hr, respectively.

The effect of the extended growth period with serum stimulation was to progressively increase the Luciferase:Renilla ratio of seven for the control with no serum stimulation to almost twice that (13.6) with 24 hr serum stimulation and to 22 after 48 hr serum stimulation (Figure 4.7a). A similar pattern was seen for the GBA constructs. However, when the expression of the GBA constructs is compared with the control for each of the scenarios, the relative expression of the constructs is the same (data not shown) and hence the serum stimulation has had no significant effect on any one construct. Despite this, it was considered possible due to the tissue specific nature of GBA expression, that the effect of serum stimulation may be different in different cell lines and maybe the observed increase in the Luciferase to Renilla ratio would accentuate the effect.

4.3.4.3 Effect of Serum Stimulation in All Four Cell Lines

All of the GBA constructs were transfected into each of the cell lines, U118, Hep G2, Huh7 and SW1353, and serum stimulated for 24 hr. When the expression level of each GBA construct is compared to that of the control, the familiar pattern of expression decreasing with increasing length of the GBA construct is seen (Figures 4.7b). As seen without serum stimulation, the lowest overall expression level is seen in U118 cells and the highest in Huh7 cells. However, the SW1353 cells are displaying significantly decreased expression with serum stimulation as the expression of all the pGL3-GBA constructs has fallen to less than 20% of the pGL3-control vector expression (Figure 4.7b and Figure 4.6a). While this may be a real phenomenon, the SW1353 cells do not transfect very efficiently and the raw data from the replicates are quite variable compared with the data for the other cell lines. Therefore, it is likely that this apparent decrease in expression with serum stimulation is an experimental aberration caused by a significant decrease in the transfection efficiency of the SW1353 cells used for this experiment.
Figure 4.7  The Effect of Serum Stimulation on the Expression of GBA Promoter Constructs

(A) Six pGL3-GBA constructs were transfected into hepatocytes (HepG2) in three replicate experiments. Luciferase assays were performed after incubation without the addition of serum and after a further 24 hr or 48 hr incubation with serum. The Luciferase:Renilla ratio for each construct at each stage is depicted. The expression level for all constructs increases with length of serum stimulation although all constructs appear to be similarly affected. (B) Nine pGL3-GBA constructs were transfected into four human cell lines and incubated for an additional 24 hr with serum. The results for all nine GBA constructs are calculated as a proportion of the control vector expression set at 100%.
When the basal levels of expression are compared with the expression levels after serum stimulation for 24 hr for each of the constructs in each of the cell lines, it can be seen in Figure 4.7b that except for SW1353 cells, serum stimulation appears to have increased the expression levels as a proportion of the control vector expression, for all constructs in U118, Hep G2 and Huh7 cell lines. Hence it appears most likely that the serum is providing overall enhancement of transcription rather than stimulating any particular cis-element or genomic DNA region. However, when construct expression is compared to that of the minimal promoter construct, pGL3-CAAT, the serum stimulated expression of the pGL3-1365 and pGL3-1956 constructs in U118 cells (Figure 4.8a) seems to be about twenty percent higher than expected. As a result, it was decided to repeat the serum stimulation experiment for a small number of relevant constructs with serum stimulation over 24 h and 48 h.

4.3.4.4 Repeat Serum Stimulation Experiments

Six GBA constructs, pGL3-CAAT, pGL3-328, pGL3-742, pGL3-1365, pGL3-1484 and pGL3-1956 were transfected into U118 and Hep G2 cells and serum stimulated for 24 hr or 48 hr. The expression levels for each of these constructs is compared to those of the pGL3-CAAT construct containing the minimal promoter element. In addition, these basal expression levels along with the results of the two 24 hr serum stimulation experiments and one 48 hr stimulation experiment are compared in Figure 4.8b. Serum stimulation for 48 hr of the pGL3-1956 construct in U118 cells has apparently produced significantly increased expression although, examination of the raw data (not shown) reveals that the luciferase values are much the same for the other constructs but there has been a significant decrease in the Renilla values to create an increased ratio. It is unlikely that this result is real as there are no results of any significance for the other constructs and hence, is most likely to be caused by an experimental problem such as a pipetting error. The trend in the expression of the constructs pGL3-742 to pGL3-1956 is similar within each cell line and between the U118 and Hep G2 cell lines, while the expression of the pGL3-328 construct is very similar to pGL3-CAAT.
Figure 4.8  The Effect of Serum Stimulation on Four Human Cell Lines

The effect of serum stimulation on pGL3-GBA constructs was investigated in four human cell lines.  

(A) All nine pGL3-GBA constructs were transfected into four human cell lines (U118, HepG2, Huh7, SW1353) in duplicate experiments and luciferase assays were performed before and after serum stimulation.  

(B) The experiment was repeated in U118 and HepG2 cells using a reduced number of constructs and luciferase assays were performed after serum stimulation for both 24 h and 48 h. The red arrows indicate results that are discussed in the text.
4.3.5 Bioinformatic Analysis of GBA 5′-UTR

A distance of 3.5 kb of the GBA promoter and 5′-UTR has been analysed, and while there may be tissue specific differences in the basal level of expression, there do not appear to be any tissue specific differences in the expression of any particular regions of the GBA 5′-UTR sequence investigated. If there is an element that defines tissue specificity within this region, it has not been identified by the current experimental protocol. A bioinformatics search for any cis-elements within the C1orf2 to GBA region was therefore carried out so that a more targeted approach could be taken.

The bioinformatics program, CISTER (http://zlab.bu.edu/~mfrith/cister.shtml), is a Cis-element Cluster Finder that predicts regulatory regions in DNA sequences by searching for clusters of cis-elements (Frith et al., 2001). Using CISTER, a 7320 bp sequence of genomic DNA between the C1orf2 and GBA genes was interrogated for cis-elements. This region corresponds to the sequence between bases 56873 and 63873 of the AL713999 human DNA sequence from clone RP11-263K19 on chromosome 1. The output from this program is displayed in Figure 4.9. This region has predicted a cis-element cluster region in an 809 bp sequence between positions 59696 and 60505 of the sequence (Figure 4.9a). The coloured lines represent the cluster of transcription factors in a common region with the length of the lines representing the probability of cis-motifs in these positions and the potential to bind the relevant trans-acting factor. The black curve represents the overall probability that these motifs will occur in a cluster and the strand on which they occur is also indicated.

In Figure 4.9b, the predicted cis-elements have been mapped onto the genomic sequence. The predicted CCAAT element between bases 60490 and 60505 is the element closest to the GBA gene and overlaps the GBA3559XHO forward primer by 5 bp. That is, the entire region up to but not including this cis-element region had already been analysed (Figure 4.4).

The predicted cis-elements are listed in Figure 4.9c. The motif that occurs in the middle of the cis-element cluster and also has the highest probability (0.54) is CRE, a cyclic AMP response element or a transcription factor complex that is a pleiotropic activator and participates in the induction of a wide variety of cellular and viral genes.
Within 100 bp 5′ of CRE are three potential Sp1 elements with two of these having the next highest probabilities of 0.44 and 0.32 respectively. Because of the large number of elements in this region, it was decided to investigate an additional 1.3 kb of the 5′-flanking region, that is, a total of approximately 4.7 kb from the GBA cap site.

4.3.5.1 Expression of Putative Cis-Element Region
Several GBA constructs were designed to investigate the putative cis-element region. Firstly, a long construct, pGL3-4700, was designed using the GBA4700F primer and the common anchor primer, GBAPCRR2. The expression of this construct was compared with that of the shortest and longest constructs, pGL3-CAAT and pGL3-3559 respectively, that had previously been investigated. In Figure 4.10a, the expression of these three constructs in four cell lines is depicted. The trend observed in previous expression experiments is for the level of expression to decrease as the length of the construct increases and the low level of expression observed for the pGL3-4700 construct is consistent with this trend.

To investigate the existence of enhancer or inhibitor sequences within the cis-element region, a 1313 bp fragment was amplified using the GBA4700F and GBAMIDIR primers and the pGL3-prom1313 construct was created with the insertion of the 1313 bp fragment into the pGL3-promoter vector. A pGL3-CAAT1313 construct was also created in a similar manner with the 1313 bp fragment being inserted immediately 5′ to the 258 bp fragment containing the GBA minimal promoter region so the effect of the cis-element region on expression of the GBA promoter could be seen. When the expression of pGL3-prom1313 is compared to pGL3-prom in four human cell lines, an increase in expression of between 1.5 and 3.5-fold can be seen (Figure 4.10b) indicating the possible existence of enhancers in the cis-element region. However, the expression levels of pGL3-CAAT1313 and pGL3-CAAT are similar in at least three of the cell lines. In SW1353, pGL3-CAAT1313 expression appears to be decreased to about half that of pGL3-CAAT. These are the results from a single set of experiments so further experiments would be needed to confirm these trends. This is especially true for SW1353 as transfection efficiency is quite variable in this cell line (data not shown).
A  Output of CISTER for Clotl2 to GBA Region

B  Possible Functional Cis-Elements Mapped on 1313bp Fragment

59341 cacaagggact ttgcacacct ctgttttttcc agcatgaaac ccttgccccc tatactgtgt 4700F
59401 gctatcacct atatcaggt tttaagcccc ctctccaccat tccagttgac acccaatccc
59461 catgcaagtc attctatcac atagcatgtg ttttttttgt ttttttttgt cattttgtct
59521 tttttgagaa caggctcttg cttgtaacct caggctggag tggcccc tatatctggt 4700F
59581 tctacgcaac ctctacatct ctggtctcct gcaattctcc tcgcctgagc tcccgagtag
59641 ctgggttacct cggcggctct ccccatgccc gcccagctaa attttgtatt tttagaagag GATA
59701 gctcggctct cctaaaaggt ttgcattact cgcagttcct gcacattacc Ap1
59761 gctcggctct cctaaaaggt ttgcattact cgcagttcct gcacattacc Ap1
59821 ctatgtttct cctatgtaata ctatctaatct ttttttttgt ttttttttgt cattttgtct
59881 cctagccctt ccagctcagt attcttctag cttctctctc gctctctgct cggctctcct Ap1
59941 agtagtcct caaaccatgt ctttggaatt actaaaagtc aggctaggt gctggccc Myf
60001 gccaggggaa ccttttggcc cttggtggccc tccaggaag aca ggga Sp1 Sp1
60061 ggagggaa ctgctcagct ccccttcccct cttggtggccc tccaggaag aca ggga Sp1 Sp1
60121 atatccaccc aggccaggg attgtagg accggcgcct ccgaggggt attgtagg Sp1 Myf
60181 gcactgttgg ttgcagtgag gcagctgttt cctgctagc Myf
60241 tcgctgctct ttttattacc ctgctattac ctgctattac ctgctattac CRE Ap1
60301 ccagggtttt cttgtgtttt cttgtgtttt cttgtgtttt cttgtgtttt cttgtgtttt CRE Ap1
60361 ttgcagtcct ttgcagtcct ctttttttttt ttttttttttt ttttttttttt ctttttttttt CCAAT 3559F
60421 ttgcagtcct ttgcagtcct ctttttttttt ttttttttttt ttttttttttt ctttttttttt CCAAT 3559F
60481 ttgcagtcct ttgcagtcct ctttttttttt ttttttttttt ttttttttttt ctttttttttt CCAAT 3559F
60541 ttgcagtcct ttgcagtcct ctttttttttt ttttttttttt ttttttttttt ctttttttttt CCAAT 3559F
60601 ttgctgctct cttgtgtttt ctttttttttt ttttttttttt ttttttttttt ctttttttttt CCAAT 3559F
60661 gtacctgctct ccggcgcctt gcagctctct ccgtggttgt tgtgtgttgt tgtgtgttgt MIDIR

cis-elements:  TATA  Sp1  CRE  ERE  NF-1  E2F  MEF-2  Sp1  CCAAT  AP-1
Table 4.9  Position, Sequence and Probability of Possible Functional Cis-Elements.

<table>
<thead>
<tr>
<th>Type</th>
<th>Position</th>
<th>Strand</th>
<th>Sequence</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRE</td>
<td>60080-60091</td>
<td>-</td>
<td>gaggggtcagg</td>
<td>0.54</td>
</tr>
<tr>
<td>Sp1</td>
<td>59980-59992</td>
<td>+</td>
<td>gcagggtcccc</td>
<td>0.44</td>
</tr>
<tr>
<td>Sp1</td>
<td>60058-60070</td>
<td>+</td>
<td>ggtggagagca</td>
<td>0.32</td>
</tr>
<tr>
<td>Ets</td>
<td>60245-60255</td>
<td>+</td>
<td>aagaggaagtt</td>
<td>0.29</td>
</tr>
<tr>
<td>Ets</td>
<td>60134-60144</td>
<td>+</td>
<td>ctcaggaagg</td>
<td>0.26</td>
</tr>
<tr>
<td>Myf</td>
<td>59997-60008</td>
<td>+</td>
<td>aaacagcttcag</td>
<td>0.21</td>
</tr>
<tr>
<td>GATA</td>
<td>59696-59708</td>
<td>+</td>
<td>aagagataggtt</td>
<td>0.14</td>
</tr>
<tr>
<td>Sp1</td>
<td>60054-60066</td>
<td>+</td>
<td>gggagataggggag</td>
<td>0.14</td>
</tr>
<tr>
<td>AP-1</td>
<td>60107-60117</td>
<td>-</td>
<td>aataatcagc</td>
<td>0.12</td>
</tr>
<tr>
<td>CCAAT</td>
<td>60490-60505</td>
<td>+</td>
<td>cttgaccaatgagact</td>
<td>0.12</td>
</tr>
<tr>
<td>SRF</td>
<td>59881-59893</td>
<td>-</td>
<td>cccatagagtt</td>
<td>0.11</td>
</tr>
<tr>
<td>AP-1</td>
<td>59793-59803</td>
<td>+</td>
<td>tgtgacctcacc</td>
<td>0.11</td>
</tr>
<tr>
<td>Myf</td>
<td>60200-60211</td>
<td>+</td>
<td>gggcagttcctg</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Figure 4.9  Analysis of GBA 5’-Flanking Genomic Sequence by CISTER

(A) The 7320 bp region between bases 56873 and 63873 of the AL713999 human DNA sequence from clone RP11-263K19 on chromosome 1 was analysed for potential cis-acting motifs using CISTER. In the scan and plot, the coloured lines represent the cluster of transcription factors at a common region: for example, a CRE site (dark green) and a number of Sp1 (royal blue) and Ets (olive green) sites were identified between positions 59696 and 60505. In addition, these lines represent the probability of cis-motifs in these positions and the potential to bind trans-acting factors such as Sp1. The black curve corresponds to the overall probability that these motifs will occur in a cluster. Lines in the upper half of the plot represent the direct strand whilst lines in the bottom half of the plot represent the complementary strand. (B) The cis-elements identified in the CISTER scan are mapped on the genomic sequence. The names of the motifs that occur within each line of sequence are indicated at the end of the line. The CCAAT motif overlaps the GBA3559F primer. The GBA4700F and GBAMIDIR primers were used for the 1313 bp fragment. (C) The cis-elements are listed in order of decreasing probability along with their sequence, position and strand on which they occur.
Figure 4.10 Expression of pGL3-4700 and Putative Enhancer Region

(A) The expression of the pGL3-4700 construct is compared to the expression of the pGL3-CAAT and pGL3-3559 constructs in four human cell lines, HepG2, Huh7, SW1353 and U118. The results are expressed as a percent of the pGL3-CAAT which was set at 100% for each cell line. The pGL3-CAAT construct contains the GBA minimal promoter element. (B) The effect of the 1313 bp fragment containing the cis-element region on the expression of the pGL3-promoter and pGL3-CAAT plasmids is shown as a ratio. The ratio of the expression levels of the pGL3-promoter and pGL3-prom1313 plasmids and the pGL3-CAAT and pGL3-CAAT1313 plasmids in four human cell lines is shown.
4.4 DISCUSSION

In Chapter 3, variability in GBA mRNA levels was demonstrated in leucocytes and skin fibroblasts. In this chapter, the tissue specific expression of GBA has been confirmed with over thirty-fold variation in GBA mRNA levels observed in human cells. Real time PCR was used to quantify mRNA levels in nine human cell types including hepatocytes (Hep G2 and Huh7), glioblastoma (U118), chondrocytes (SW1353), promyelocytes (HL60), monocytes (THP1) and B-cells. The lowest levels of GBA mRNA were seen in leucocytes and the highest levels were seen in chondrocytes. The levels of ASA mRNA varied only seven-fold in the same nine cell types indicating that the $G_{BA}$ gene is more highly regulated.

The finding that $G_{BA}$ is highly regulated and expressed in a tissue-specific manner is consistent with that of Doll and Smith (1993) but on closer examination, there are significant differences. The promyelocyte line (HL60) and the monocyte line (THP1) were included in both studies but Doll and Smith (1993) reported a 26-fold difference in the GBA mRNA levels for these two cell lines while in the present study, the mRNA levels were almost the same. Although some variation in results is to be expected especially when different techniques are used, the magnitude of this variation is too large to be accounted for in this way. In the work of Doll and Smith (1993), the neuronal cell lines (U87-MG and CCF-STTG) have the highest GBA mRNA levels, (being about 6-fold higher than skin fibroblasts (GM5659)), while in this study, the neuronal cell line (U118) and skin fibroblasts have similar levels of mRNA that represent intermediate levels in the GBA mRNA range seen in this group of cell lines.

The relative differences in the GBA and ASA mRNA levels in some cell lines is noteworthy and indicates that, aside from the overall regulation of lysosomal genes, there are almost certainly individual modes of regulation for these two genes. In Hep G2 cells, the levels of GBA mRNA are high and but the levels of ASA mRNA are comparatively low, while in B-cells, the situation is reversed with high levels of ASA mRNA but low levels of GBA mRNA. The levels of mRNA for both genes are lower overall in the suspension cell lines (leucocytes, THP1, HL60, B-cells) than in the attached cell lines although, B-cells are the exception with a relatively high level of ASA mRNA.

To search for elements defining the tissue specificity of GBA, analysis of approximately 3.5 kb of the GBA promoter and 5’-UTR was undertaken. Nine GBA promoter
constructs, ranging in size from 258 bp to 3559 bp were fused to a synthetic luciferase reporter gene. These were transfected into the four attached cell lines, U118, Hep G2, Huh7 and SW1353. Basal levels of expression were observed with expression decreasing as the length of construct increased. The pGL3-258 construct containing the minimal promoter element had the highest expression level whilst the pGL3-3559 construct had the lowest expression level. Highest overall levels of expression were seen in Huh7 cells with the lowest levels, well over ten-fold lower, in U118 cells. One may have expected SW1353 cells to have the highest expression as this cell line had GBA mRNA levels more than twice as high as the other three cell lines used for transfection. However, SW1353 cells were more difficult to transfect and a different transfection reagent was used so results may not be directly comparable with the other cell lines.

Serum stimulation of the transfected Hep G2 cells produced an overall increase in the level of expression although the same trend of the expression decreasing with increasing length of the construct was still observed. It appears therefore that there are no elements within 3.5 kb upstream of GBA that are responsive to factors within serum.

The likelihood of finding tissue specific elements was probably reduced by the inability to transfect all of the cell lines used initially. There are only small differences in the GBA mRNA levels of the U118, Huh7 and Hep G2 cell lines. Although the GBA mRNA levels in SW1353 are over two-fold higher than the other attached cell lines, the ability to transfect the suspension cell lines may have proved very useful because their average GBA mRNA level was less than a quarter of the SW1353 GBA mRNA levels.

As there does not appear to be any tissue specific response elements within the sequence investigated, bioinformatics analysis using CISTER, a \textit{cis}-element cluster search program (Frith \textit{et al.}, 2001), was undertaken. The entire DNA sequence between the \textit{C1orf2} and \textit{GBA} genes was searched and an 818 bp region containing a putative \textit{cis}-element cluster was identified. This DNA sequence is immediately adjacent to the 3.5 kb previously analysed so the analysis was extended to include this region.

Expression of this putative \textit{cis}-element region in the luciferase promoter vector suggested the possible existence of enhancers in the region as an increase in expression of between 1.5- and 3.5-fold was seen. However, no change in expression level was seen in three of the cell lines when the same region was expressed in the pGL3-CAAT plasmid. In
addition, expression of the entire 4.7 kb sequence was less than for the 3.5 kb sequence, a
pattern that was consistent with earlier findings for sequences of increased length.

It appears most likely that enhancers within the *cis*-element region are not acting on the
*GBA* gene or that inhibitor elements also exist that are negating the effect. The
probability that some or all of the elements within the identified region are acting on
another gene seems relatively high with a predicted probability of greater than 0.6 that
these motifs occur in a cluster. Time constraints prevented further analysis of the
individual elements within the *cis*-element region but this is required to establish the target
gene affected by these elements.

It is interesting to note that the GBA promoter region and the elements within it have not
been specifically identified by the CISTER program and appear as a very small bump in
the black curve directly below the 7000 position in the sequence. This may indicate that
the cluster of elements comprising the GBA promoter is relatively small and conversely
that there is a relatively large number of elements in the putative *cis*-element region
identified. Therefore, this region is likely to play a significant but as yet unidentified role
in gene expression.

The tissue specific regulation of the transcription of *GBA* needs to be further elucidated
and, as some *trans*-acting proteins exert their effects through 3′-UTRs (Adam et al., 1986),
investigation of this region could be worthwhile. However, variation in enzyme activity
levels has also been observed and this does not always parallel the variation in mRNA
levels. That is, there is regulation of GBA translation and a translational control protein
(TCP80) that interacts with *GBA* has been identified (Xu and Grabowski, 1999). In
Chapter 5, the translational regulation of *GBA* will be investigated.

4.5 CONCLUSION

In this study, GBA mRNA levels were found to vary over thirty-fold compared to
variation in ASA mRNA levels of only seven-fold in human cells. While overall
regulation of lysosomal enzymes almost certainly exists, it is clear that regulation of
individual lysosomal genes also occurs and that *GBA* is highly regulated. It is also likely
that *ARSA* is individually regulated but to a much lesser extent. Analysis of the 5′-UTR
of *GBA* for *cis*-elements that could be responsible for the tissue-specific expression of
GBA revealed a large cis-element cluster approximately 4 kb upstream. Although capable of promoter regulation, it did not appear to act on GBA. However, closer analysis of this region or the 3′-UTR could be undertaken. As both transcriptional and translational regulation will contribute to the final GBA enzyme activity level, the translational regulation of GBA will be investigated in the next chapter.
CHAPTER 5

TRANSLATIONAL REGULATION OF THE
GLUCOCEREBROSIDASE AND ARYLSULPHATASE A
GENES
5.1 INTRODUCTION

Translational regulation of expression has been reported for several lysosomal enzymes including GBA, ASA, β-glucuronidase (GUS) and aspartic protease (mLAP) (Xu and Grabowski, 1998, Kreysing et al., 1994, Bracey and Paigen, 1987, Cho and Raikhel, 1992). In each case, the regulation occurs by different mechanisms including RNA binding proteins, changes in translational yield and negative control by steroids. As translational control plays an important role in determining the final levels of enzyme activity for these proteins, further investigation of its role in GBA and ASA expression will be examined in this chapter.

Variation in GBA enzyme activity levels between individuals was investigated and in Chapter 3, it was shown that enzyme activity levels vary between different cell types such as leucocytes and skin fibroblasts. Other researchers have also reported variation between cell lines, with Doll and Smith (1993) finding that there was a sixty-fold variation in the enzyme activity levels of the cell lines included in their study. In addition, Doll and Smith (1993) found mRNA levels varied over fifty-fold. This variation in GBA mRNA levels was confirmed in this study as it was found that the levels varied over thirty-fold in nine human cell types. Variation in GBA mRNA levels between individuals was also seen in the leucocytes and fibroblasts from different individuals.

The transcriptional regulation of GBA was investigated in Chapter 4 because a direct correlation between GBA mRNA and enzyme activity levels was reported in some cell lines, such as those in group A (glioblastoma, U-87MG; histiocytes, U937; astrocytoma, CCF-STTG1; and lymphoblasts, GM621A) (Doll and Smith, 1993) and possibly the skin fibroblasts investigated in Chapter 3. However, in other cell lines such as those in groups B and C (promyelocytes, HL60; neuroglioma, H4; fibroblasts, GM5659; monocytes, THP1) (Doll and Smith, 1993) and maybe one of the leucocyte samples (M1) investigated in Chapter 3, the level of GBA enzyme activity is quite disproportionate to the mRNA level indicating the likelihood of post-transcriptional regulation.

A translational control protein (TCP80) has been identified and may be responsible for some or all of the tissue-specific variation in GBA mRNA/activity observed in this project and previously by Doll et al. (1995). TCP80 is an 80 kDa cytoplasmic protein that inhibits GBA mRNA translation by binding the RNA coding regions (Xu and Grabowski,
and preventing the engagement of GBA mRNA with polysomes (Xu and Grabowski, 1999). TCP80 has two RNA binding domains and was found to be very similar to several other RNA binding proteins including NF90, ILF3, DRBP76 and MPP4. These proteins have been found to be splice variants of the Interleukin Enhancer Binding Factor gene (ILF3), which has been mapped to chromosome 19 (Duchange et al., 2000). There is also the possibility of competition for binding of TCP80 as it was found to interact with other RNAs.

Translational control of ASA expression in mouse testis has been reported wherein a protein specifically expressed in testis binds to the 5′- and 3′-UTR of ASA mRNA (Kreysing et al., 1990). An increase in ASA mRNA levels in germ cells at specific stages was found not to be accompanied by a similar increase in enzyme activity. While the exact mechanism of its action is uncertain, it seems likely that it plays an important role in the process of murine spermatogenesis. There have been no reports of investigations into translational control of human ASA expression.

To further elucidate the cause of the variation seen in GBA and ASA enzyme activity levels, the enzyme activity levels of the cell lines investigated in Chapter 4 were measured. The GBA translational efficiency for each cell type was determined by comparison of the mRNA level to the enzyme activity level. In addition, the levels of TCP80/ILF3 were measured in these cell types to ascertain how it might be affecting the final levels of GBA (or ASA) enzyme activity.

5.2 MATERIALS AND METHODS

The materials and methods specific to this chapter are described in detail below. General materials and methods used in this thesis are outlined in detail in Chapter 2.

5.2.1 Western Blot Analysis of Proteins

5.2.1.1 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Ten percent resolving gels were used to separate total protein. Total protein extract (10 µg) from each cell type was heated for 10 min at 90°C in loading buffer (0.25 M Tris-HCl, pH 6.8, 3% SDS, 10% glycerol, 20 mM β-mercaptoethanol, 0.01% bromophenol blue) and electrophoresed in running buffer (0.025 M Tris, 0.25 M glycine, 0.1% w/v SDS) on a Protean II minigel apparatus (Biorad, Sydney, Australia). The SDS-PAGE gels were run
at 100 V until the loading dye moved into the lower gel (about 20 min) and then at 200 V for 90 min. To ascertain molecular sizes of the resulting bands, a pre-stained molecular weight protein marker (low range, Biorad) was used.

5.2.1.2 Western Blot Analysis

A transfer blot apparatus (BioRad) was used to transfer the separated proteins within the SDS-PAGE gel to a nitrocellulose membrane (Amersham Pharmacia Biotech). The nitrocellulose membrane and SDS-PAGE gel were equilibrated in 3-[cyclohexylamino]-1-propane-sulfonic acid (CAPS) buffer (0.01M/L CAPS, pH 11, 10% ethanol, 0.0001% w/v SDS) at room temperature for 5 min. The protein samples were then electrotransferred to the nitrocellulose membrane in CAPS buffer at 4°C for 2 h using constant current of 200 mA. The membrane was rinsed in PBS (14 mM NaCl, 3 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄; pH 6.8). After drying overnight, the membrane was washed well in PBS (pH 6.8) (4 x 5 min) and then blocked in casein blocking buffer (1% casein in PBS; pH 6.8) for 2 h.

The primary antibodies, rabbit anti-TCP30 (gift from Dr. Greg Grabowski) and mouse anti-DRBP76 monoclonal antibody (Figure 5.1) (BD Transduction Laboratories, Sydney, Australia), were diluted 1:1500 and 1:1000 respectively, in casein blocking buffer and incubated with the membrane at room temperature for 1 h. The loading control monoclonal primary antibody, mouse anti-GAPDH (Imgenex, California, USA) was also diluted 1:3000 in each of the primary antibody solutions. The membrane was washed in PBS (pH 6.8) plus 0.1% Tween 80 (4 x 5 min) and incubated with the appropriate secondary antibody solution. The secondary antibodies, Alexa Fluor® 680 goat anti-rabbit IgG (Molecular Probes, Mount Waverley, Australia) and/or IRDye™ 800 goat anti-mouse IgG were both diluted 1:20,000 in casein blocking buffer. The membrane was washed in PBS (pH 6.8) plus 0.1% Tween 80 (4 x 5 min) before being rinsed in PBS (pH 6.8) for 5 min.

The damp membrane was visualised using the Li-Cor Odyssey imager (John Morris Scientific). The membrane was scanned at the default intensity of 5.0 for both channels in addition to various combinations of channel intensity to achieve optimum detection results.
Figure 5.1 The DRBP76 Antibody

(A) The DRBP 76 antibody (BD Transduction Laboratories) is a mouse IgG1 isotype that contains a bi-partite nuclear localisation signal at amino acids 369-373 and 386-394, two double-stranded RNA binding domains in the C-terminal half and a C-terminal RG2 domain (a nuclear localisation signal) that is present in many RNA binding proteins.  

(B) The DRBP76 antibody is a monoclonal antibody that has been shown to detect protein species at 90 kDa in HeLa cells.

(Adapted from BD Transduction Laboratories DRBP76 product information)
5.2.2 TNT Reticulocyte Coupled Transcription/Translation Reaction

5.2.2.1 Preparation of GBA cDNA Plasmid DNA

The agar stab was sampled and streaked onto LB/ampicillin agar plates and incubated at 37°C overnight. A single colony was picked and cultured in 5 ml of LB/ampicillin broth for 16 h at 37°C with shaking at 225 rpm. Following this, the culture was centrifuged at 3000 x g for 5 min to collect the bacterial cell pellet and then prepared for plasmid extraction. The extraction of plasmid DNA was performed using the QIAprep Spin Miniprep kit as outlined by the manufacturer (QIAGEN).

5.2.2.2 The Coupled Transcription/Translation Reaction

The TNT Quick Coupled Transcription/Translation System (L1171; Promega) was used for in vitro translation of GBA cDNA as outlined by the manufacturer. Briefly, a TNT lysate reaction was prepared containing 40 µl of TNT quick master mix, 1 µl of 1 mM methionine, 2 µl of GBA plasmid cDNA template (0.5 µg/µl) and nuclease free water in a total volume of 50 µl. The reaction was incubated at 30°C for 90 min. The translation product of TNT lysate reaction was made up to 100 µl with sterile 0.9% NaCl as required for the enzyme assay. The results of the translation were analysed with measurement of the GBA enzyme activity and the assay was performed as previously described (Materials and Methods 2.3.4).

5.3 RESULTS

5.3.1 Variation in Enzyme Activity Levels in Nine Human Cell Types

In Chapter 4, the results of the measurement of mRNA levels in several human cell lines were presented. Concurrent measurements of GBA and ASA enzyme activity were also performed against the artificial substrates using the Standard Operating Procedures provided by the Women’s and Children’s Hospital, Adelaide (Materials and Methods 2.3). Table 5.1 presents the levels of GBA and ASA enzyme activity levels in nine human cell types. The final values are the mean of at least four replicates. Figures 5.2a and 5.2b depict both the enzyme activity and mRNA levels of each of the genes respectively in each of the nine cell types.
<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Average GBA Enzyme Activity (pmol/min/mg protein)</th>
<th>Average ASA Enzyme Activity (nmol/min/mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-cells</td>
<td>49 +/- 6.6</td>
<td>0.9 +/- 0.1</td>
</tr>
<tr>
<td>Leucocytes</td>
<td>172 +/- 23.2</td>
<td>4.6 +/- 0.6</td>
</tr>
<tr>
<td>HL60</td>
<td>361 +/- 27.1</td>
<td>1.7 +/- 0.3</td>
</tr>
<tr>
<td>HepG2</td>
<td>417 +/- 64.6</td>
<td>0.3 +/- 0.1</td>
</tr>
<tr>
<td>U118</td>
<td>757 +/- 32.6</td>
<td>0.8 +/- 0.2</td>
</tr>
<tr>
<td>THP1</td>
<td>922 +/- 102</td>
<td>1.4 +/- 0.3</td>
</tr>
<tr>
<td>Huh7</td>
<td>996 +/- 137</td>
<td>1.6 +/- 0.3</td>
</tr>
<tr>
<td>Skin Fibroblasts</td>
<td>1659 +/- 154</td>
<td>9.4 +/- 0.8</td>
</tr>
<tr>
<td>SW1353</td>
<td>5936 +/- 427</td>
<td>13.0 +/- 1.5</td>
</tr>
</tbody>
</table>
Figure 5.2 GBA and ASA mRNA and Enzyme Activity Levels in Nine Cell Types

(A) The level of GBA mRNA and enzyme activity was measured in nine human cell types: leucocytes, monocytes (THP1), promyelocytes (HL60), hepatocytes (Hep G2 and Huh7), glioblastoma (U118), skin fibroblasts, B-cells and chondrocytes (SW1353). Each of these results is the mean of four to six replicates. (B) The level of ASA mRNA and enzyme activity was measured in nine human cell types: leucocytes, monocytes (THP1), promyelocytes (HL60), hepatocytes (Hep G2 and Huh7), glioblastoma (U118), skin fibroblasts, B-cells and chondrocytes (SW1353). Each of these results is the mean of four to six replicates.
The enzyme activity levels of GBA are lowest in B-cells and highest in chondrocytes (SW1353), varying 120-fold from $4.9 \times 10^1$ pmol/min/mg protein to $5.9 \times 10^3$ pmol/min/mg protein. The ASA enzyme activity levels are also highest in chondrocytes (SW1353) but lowest in hepatocytes (Hep G2) cells. They vary over forty-fold from $2.6 \times 10^1$ nmol/min/mg protein to $1.3 \times 10^1$ nmol/min/mg protein.

GBA shows greater variation in enzyme activity levels than ARSA which would suggest it is the more highly regulated of these two genes. Comparison of the enzyme activity levels of both GBA and ASA reveals that, while skin fibroblasts (SF) and chondrocytes (SW1353) have high enzyme activity levels, leucocytes have very low levels of GBA activity but moderately high levels of ASA enzyme activity. That is, levels of the lysosomal enzymes vary independently and changes in the level of one enzyme are not necessarily reflected in the level of another lysosomal enzyme.

5.3.1.1 Efficiency of Translation

When an enzyme is expressed, if there is no post-transcriptional regulation, one expects to see a direct relationship between the levels of mRNA and enzyme activity such that higher levels of mRNA would lead to higher levels of enzyme activity. As can be seen in Figures 5.2a and 5.2b, this is clearly not the case for either GBA or ASA in some cell lines. For example, both hepatocyte cell lines have higher levels of GBA mRNA but lower levels of enzyme activity than skin fibroblasts. Similarly for ASA, B-cells have higher levels of mRNA but much lower levels of enzyme activity than skin fibroblasts. For GBA, the variation in enzyme activity levels (120-fold) is four times greater than the variation in mRNA levels (30-fold) and for ASA, the variation in enzyme activity levels (50-fold) is also greater than the variation in mRNA levels (7-fold). Hence for both GBA and ASA, the presence of post-transcriptional regulation is indicated. As this regulation can occur in various ways, this project focuses on translational regulation because of the previous reports of the translational control of GBA.

5.3.2 Analysis of the Relationship Between TCP80 and GBA

An inhibitory translational control protein (TCP80) that binds within the GBA coding region has been reported in human cells (Xu and Grabowski, 1998). Based on this, the relationship between the mRNA/protein levels of TCP80 and the GBA activity levels in tissues was investigated. It was planned to quantify the levels of TCP80 mRNA using real time PCR and if a suitable antibody was available, the levels of TCP80 protein would also
be quantified via ELISA or western blot. This data would be correlated with GBA expression data to ascertain if the variation in GBA enzyme activity levels could be solely attributable to variation in TCP80 levels. However, the research plan was modified as further information about TCP80 became known.

5.3.2.1 Primer Design for TCP80 mRNA Quantification

Since TCP80 was first identified as an inhibitor of GBA translation, it has been found to be a splice variant of *ILF3*. Five splice variants of *ILF3* have been identified: DRBP76α (*ilf3*-a), DRBP76 (*ilf3*-b), DRBP76δ (*ilf3*-c), NF90/ILF3 (*ilf3*-d) and ILF3 (*ilf3*-e) (Figure 1.10) (Duchange *et al.*, 2000). To quantify the mRNA levels for TCP80 alone, the primers would have to specifically amplify TCP80 exclusively of the other ILF3 splice variants. This task proved to be intractable as although the amino acid sequences vary after the splice site at position 2330, there was insufficient variability in the nucleotide sequences to ensure specific amplification of TCP80 alone. The ILF3 splice variants are part of a large family of double-stranded-RNA binding proteins that contain a double-stranded-RNA binding motif (dsRBM) (Fierro-Monti and Mathews, 2000). Double-stranded-RNA binding proteins play critical roles in several aspects of cellular metabolism, including transcriptional activation, translational control, and mRNA processing and localisation (Reichman *et al.*, 2002). It is possible that although TCP80 is the only ILF3 splice variant that has been identified to date as interacting with *GBA*, the other ILF3 splice variants also play a role in translational control of *GBA*. Hence, it was decided to design primers that would quantify as many ILF3 splice variants as possible.

Table 5.2 lists the splice variants of *ILF3*, the National Centre for Biotechnology Information (NCBI) reference sequence numbers and the length of the mRNA and protein sequences. TCP80 has been included with NF90 (Kao *et al.*, 1994) as the cDNA sequences are nearly identical (96%) (Xu and Grabowski, 1999) and they have identical functions (Xu *et al.*, 2000). Although M-phase phosphoprotein MPP4 is reported to be identical (99%) to TCP80 (Xu *et al.*, 2000), it has not been included as the sequence does not include a stop codon and is likely to be incomplete. It is therefore not possible to assess whether the entire sequence is identical to TCP80.
In Figure 5.3, a ClustalW alignment of the NCBI reference sequences for the ILF3 splice variants and TCP80 demonstrates that the ILF3 fragment designed does indeed amplify all sequences. However for NF90, the reverse primer (ILF3R1) will bind in the 3′-UTR because the sequence contains an extra two nucleotides (TC). These occur 276 bases prior to the alternative splicing position causing a frameshift and subsequent premature termination of the NF90 (Figure 5.3). The 86 bp ILF3 fragment was amplified in the forward direction with ILF3F1 (5'-gCgggCgAggATTTgg-3') and in the reverse direction with ILF3R1 (5'-CAgAgTTgCCTCCgTACCCATA-3').

5.3.2.2 Quantification of ILF3 mRNA Levels
The level of ILF3 mRNA was measured in a panel of nine human cell types and the leucocyte and skin fibroblast cells from individuals by quantitative real time PCR (Materials and Methods 2.12.3). This was done concurrently with previous measurements of ASA and GBA mRNA and also enzyme activity levels. The final values are the mean of at least four replicates. For skin fibroblasts and leucocytes, replicates were done for the cell sample for each individual and the results of the four individuals averaged to obtain results for that cell type. Figure 5.4 depicts the ILF3 mRNA levels in nine human cell types. The mRNA levels are lowest in leucocytes and highest in hepatocytes (Hep G2), varying 24-fold from $2.6 \times 10^5$ to $6.3 \times 10^6$ molecules of ILF3 mRNA (Table 5.2).
<table>
<thead>
<tr>
<th>ILF3 Splice Variant</th>
<th>NCBI Nucleotide Reference</th>
<th>Length of Coding Sequence (nt)</th>
<th>NCBI Protein ID</th>
<th>Length of Protein (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ilf3-a (DRBP76α)</td>
<td>AJ271744</td>
<td>1697</td>
<td>CAC01404</td>
<td>564*</td>
</tr>
<tr>
<td>ilf3-b (DRBP76)</td>
<td>AJ271745</td>
<td>2109</td>
<td>CAC01405</td>
<td>702</td>
</tr>
<tr>
<td>ilf3-c (DRBP76δ)</td>
<td>AJ271746</td>
<td>2073</td>
<td>CAC01406</td>
<td>690</td>
</tr>
<tr>
<td>ilf3-d (NF90/ILF3)(TCP80)</td>
<td>U10324/AF141870</td>
<td>2016/2295</td>
<td>AAA20994/AAD37575</td>
<td>671**/764**</td>
</tr>
<tr>
<td>ilf3-e (ILF3)</td>
<td>AJ271747</td>
<td>1776</td>
<td>CAC01407</td>
<td>591*</td>
</tr>
</tbody>
</table>

* 5’ end has not been fully characterised i.e. does not start with an ATG codon
** 3’ end of mRNA sequence does not include a polyA signal
nt = nucleotides
aa = amino acids
Figure 5.3  Alignment of the Coding Sequences of the ILF3 Splice Variants

A ClustalW multiple sequence alignment was performed with the NCBI reference sequences for each of the ILF3 splice variants and NF90 (Table 5.2). Identical nucleotides are indicated by an asterisk (*). The length of the nucleotide sequence for each sequence is recorded at the end of each line. The fragment amplified for mRNA quantification of ILF3 is indicated in blue with the forward and reverse primers indicated in bold and underlined. The yellow highlighted vertical line prior to the last nucleotide in the fifth row indicates the point of alternative splicing in exon 17. The NF90 coding sequence has a 2-nt insertion as seen in the sequence U10324 in row 1. This leads to a frameshift resulting in termination of the amino acid sequence with the TGA stop codon indicated in red in row 4.
Figure 5.4 ILF3 mRNA Levels in Nine Cell Types

The level of ILF3 mRNA and enzyme activity was measured in nine human cell types: leucocytes, monocytes (THP1), promyelocytes (HL60), hepatocytes (Hep G2 and Huh7), glioblastoma (U118), skin fibroblasts, B-cells and chondrocytes (SW1353). Each of these results is the mean of four to six replicates.
<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Molecules of ILF3 mRNA/µg RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leucocytes</td>
<td>$2.62 \times 10^5 \pm 1.11 \times 10^5$</td>
</tr>
<tr>
<td>Skin Fibroblasts</td>
<td>$2.47 \times 10^6 \pm 1.42 \times 10^5$</td>
</tr>
<tr>
<td>SW1353</td>
<td>$2.95 \times 10^6 \pm 3.94 \times 10^5$</td>
</tr>
<tr>
<td>B-cells</td>
<td>$3.39 \times 10^6 \pm 4.49 \times 10^5$</td>
</tr>
<tr>
<td>THP1</td>
<td>$5.03 \times 10^6 \pm 1.42 \times 10^5$</td>
</tr>
<tr>
<td>HL60</td>
<td>$5.25 \times 10^6 \pm 5.17 \times 10^5$</td>
</tr>
<tr>
<td>Huh7</td>
<td>$5.59 \times 10^6 \pm 7.72 \times 10^5$</td>
</tr>
<tr>
<td>U118</td>
<td>$5.68 \times 10^6 \pm 4.89 \times 10^5$</td>
</tr>
<tr>
<td>Hep G2</td>
<td>$6.30 \times 10^6 \pm 1.29 \times 10^6$</td>
</tr>
</tbody>
</table>
5.3.3 The Translational Efficiency Ratio (TER)

To better assess the efficiency of the translation of the lysosomal enzymes GBA and ASA, a translational efficiency ratio (TER) was estimated for each of the enzymes, GBA and ASA, in each of the nine cell types. The TER was determined by calculating the ratio between molecules of mRNA and enzyme activity. That is, translation is most efficient when the ratio is small and a rising ratio indicates decreasing efficiency. Furthermore, these results are compared to the ILF3 mRNA level in each cell type and in individuals.

5.3.3.1 Translational Efficiency of GBA and ASA in Nine Human Cell Types

In Figure 5.5a, the possible effect of ILF3 on GBA expression is examined. B-cells have the highest TER. This indicates that GBA translational inefficiency is most pronounced in B-cells which require four times more mRNA molecules than hepatocytes (Hep G2) and over 25 times more mRNA molecules than chondrocytes (SW1353) to produce one unit of GBA enzyme activity.

In Figure 5.5b, the possible effect of ILF3 on ASA expression is considered. Once again, B-cells have quite a high TER but the highest TER is in Hep G2 cells. Excluding B-cells, the overall trend for both lysosomal enzymes is for the TER to rise as the levels of ILF3 mRNA rise (r[ASA] = 0.61, r[GBA] = 0.68). That is, translational efficiency appears to lower in cell types that have higher levels of ILF3 mRNA.

5.3.3.2 Translational Efficiency of GBA and ASA in Individuals

In Figures 5.6 and 5.7, the translational efficiency ratios for GBA and ASA respectively, are compared with ILF3 mRNA levels in skin fibroblasts and leucocytes from four individuals. As with the comparison of nine different cell types, there is a trend towards a higher translational efficiency ratio when ILF3 mRNA levels are higher in leucocytes (SF r[ASA] = -0.09, r[GBA] = 0.26; leucocytes r[ASA] = 0.99, r[GBA] = 0.99). The skin fibroblast lines, especially SF3542, does not appear to fit proposed trend but there is not much variation in the ILF3 mRNA levels between the skin fibroblast cell types.
Figure 5.5  The Translational Efficiency Ratio for GBA and ASA Compared to ILF3 mRNA Levels in Nine Cell Types

The translational efficiency ratio (TER) is the ratio of mRNA level to enzyme activity level. A low level indicates efficient translation and a high level indicates inefficient translation. (A) The GBA TER is compared to ILF3 mRNA levels in nine human cell types: leucocytes, monocytes (THP1), promyelocytes (HL60), hepatocytes (Hep G2 and Huh7), glioblastoma (U118), skin fibroblasts, B-cells and chondrocytes (SW1353). (B) The ASA TER is compared to ILF3 mRNA levels in the same nine human cell types.

Effect of ILF3 on GBA Expression

Effect of ILF3 on ASA Expression
Figure 5.6  The Translational Efficiency Ratio Compared to ILF3 mRNA Levels in Skin Fibroblasts from Four Individuals

(A) The GBA TER is compared to ILF3 mRNA levels in four skin fibroblast cell lines from four individuals: SF3351, SF3542, SF3693, SF3785. (B) The ASA TER is compared to ILF3 mRNA levels in the same four skin fibroblast cell lines.
Figure 5.7  The Translational Efficiency Ratio Compared to ILF3 mRNA Levels in Leucocytes from Four Individuals

(A) The GBA TER is compared to ILF3 mRNA levels in leucocytes from four individuals: F1, F2, M1, M2.  (B) The ASA TER is compared to ILF3 mRNA levels in leucocytes from the same four individuals.
5.3.4 Evaluation of In Vitro Translational Efficiency of GBA in Human Cell Lines

The TNT Quick Coupled Reticulocyte Lysate System (Promega) is a single tube, coupled transcription/translation system in which approximately 1 µg of plasmid DNA template is used to produce the translation product. Hence, a GBA cDNA clone was required to prepare template DNA. Duplicate reaction tubes with and without the addition of cytoplasmic extract from each cell line were prepared. A comparison of in vivo and in vitro results was required for the evaluation of the existence of TCP80/ILF3 or any further translational control factors in the cytoplasm of each cell type.

5.3.4.1 The GBA cDNA Clone

The GBA cDNA clone (Clone ID: IRATp970A021D; entry:2899915 (IMAGE)) from cDNA-Library 970 (IRAT MGC Human verified full length amp cDNA) was obtained from RZPD Deutsches Ressourcenzentrum fur Genomforschung GmbH (Berlin, Germany). The GBA cDNA sequence (Accession ID: BC003356) of 2279 bp is contained in the vector pCMV-SPORT6 and in the orientation that uses the SP6 RNA polymerase promoter.

5.3.4.2 Measurement of GBA Enzyme Activity of the In Vitro Translation Product

The GBA cDNA plasmid template was prepared (Materials and Methods 2.15.1) and quantified. Using 1 µg of plasmid DNA template, a TNT lysate reaction was incubated (Materials and Methods 2.15.2) and the GBA enzyme activity of the translation product measured (Materials and Methods 2.3.4). The one hour incubation at 37°C denatured the rabbit reticulocyte lysate resulting in a brown precipitate. The GBA enzyme activity was measured with and without removal of this precipitate and also with and without the freeze/thaw step that is used in preparation of the cell lysate. No GBA enzyme activity could be detected.

This in vitro translation system was expected to produce 150 to 300 ng of GBA protein as per the kit protocol. However, this could not be quantified using the Bio-Rad protein assay (Bio-Rad Laboratories, Regents Park, Australia) (Materials and Methods 2.4), as was done with the cell extracts, because the 50 µl TNT lysate reaction also contains 10 mg of rabbit reticulocyte protein. Although GBA enzyme activity is normally expressed per milligram of total protein in the cell lysate, it is not known how much GBA protein is required to produce that activity and hence, whether the amount of GBA protein
produced in the *in vitro* translation system is a sufficient quantity from which to measure GBA activity.

### 5.3.5 Measurement of TCP80/ILF3 Protein Levels

To more accurately assess the relationship between GBA translation and ILF3, the measurement of ILF3 protein levels is required but no commercial antibody for ILF3 was available. Two alternate antibodies were obtained – an anti-DRBP76 antibody (Transduction Laboratories, BD Biosciences, Pharmagen, North Ryde, Australia) (Figure 5.1) and an anti-TCP30 antibody (a gift from Dr. G. Grabowski) (Xu and Grabowski, 1999). The relationship between the ILF3 splice variants was examined to establish how many and which splice variants the DRBP76 and TCP30 antibodies would detect.

The DRBP76 antibody was produced using a region of 104 amino acids (indicated in blue in Figure 5.8) as the immunogen. It would appear that this is the C-terminal region of *ilf3-b* or DRBP76 (AJ271745) (also see Figure 1.10). However, as the specific details of the binding sequence remains proprietary information, it is not possible to determine exactly how many of the ILF3 splice variants it will detect. All amino acids, except the last nine (SDFFTDCYG) of the 104 amino acid region, are common to all other variants excluding NF90 (U10324). Therefore, unless the specificity of the antibody is determined by this SDFFTDCYG amino acid sequence, the DRBP76 antibody will detect all ILF3 splice variants including TCP80 but not NF90.

The TCP30 antibody was a generous gift from Dr. G Grabowski. This polyclonal antibody was raised against a region of 227 amino acids that includes the two double-stranded RNA binding motifs. As can be seen in Figure 5.8, this region (highlighted yellow) is homologous to all of the ILF3 splice variants except NF90 and DRBP76α (*ilf3-d* and *ilf3-a*; U10324 and AF271744 respectively). Despite uncertainty about the specificity of these antibodies, it was decided to use them for the measurement of ILF3 splice variant protein levels because they could still provide useful information and no other antibodies were available.
Figure 5.8 ClustalW Multiple Sequence Alignment of ILF3 Splice Variants

A ClustalW multiple sequence alignment was performed with the NCBI sequences for each of the ILF3 splice variants and TCP80. Identical amino acids are indicated by an asterisk (*). The length of the amino acid sequence for each sequence is recorded at the end of each line. Two double-strand RNA-binding domains are present in the mRNA transcript and these are indicated by the underlining between amino acids 419 to 464 and 535 to 604. The underlined amino acid sequence (660 to 688) corresponds to the ILF3 fragment amplified for quantification of ILF3 mRNA. The vertical line after amino acid 692 indicates the point of alternative splicing in exon 17. The 104 amino acid sequence indicated in blue corresponds to the immunogen used to produce the DRBP76 antibody. The AF141870 sequence highlighted in yellow is the region from which polyclonal TCP30 antibodies were raised.
5.3.5.1 Western Blot Analysis of ILF3 Isoform Protein Levels

Western blots (section 5.2.2.2) to detect the ILF3 isoform proteins were performed using the cell pellets (Materials and Methods 2.3.2) prepared for all cell types on which mRNA quantification and enzyme activity levels had been done: SW1353, U118, Hep G2, HL60, THP1, B-cells, SF3351, SF3542, SF3693, SF3785 and leucocytes from individuals, F1, F2, M1, and M2. The GAPDH monoclonal antibody (Imgenex, San Diego, California) was used as the loading control to confirm that approximately equal amounts of total cell protein (10 µg) were loaded.

In each cell type, a similar pattern of bands was obtained with both antibodies although in some cell types, more bands have been detected with the TCP30 antibody, probably because it is polyclonal (Figure 5.9). The bands detected fell into three size categories (Table 5.4). A 110 kDa isoform was detected in B-cells, Hep G2 and SF3351 with both antibodies and also in THP1, Huh7 and SW1353 with the DRBP76 antibody. However, in all cases, this was at lower levels than other isoforms for the same cell type. For all cell types in which ILF3 isoforms were present, there were up to three bands between 93 and 103 kDa and in most cases, these were seen with both antibodies. An additional band at 77 kDa was detected in some cell types with the TCP30 antibody alone. The 88 kDa band seen in the DRBP76 blots is the result of non-specific binding by the GAPDH antibody. In some cell types, no ILF3 isoforms were detected.

Two forms of MPP4 and TCP80 have been described in previous reports, a major form at 90 kDa and a minor form at 110 kDa (Matsumoto-Taniura et al., 1996, Xu and Grabowski, 1999). It is likely that it is these two forms that can be seen at 110 kDa and between 93 and 103 kDa on these western blots. In some cell types, up to three bands can be seen between 93 and 103 kDa and these may represent other ILF3 isoforms or a single isoform at various stages of post-translational processing. NF90, which has a predicted molecular weight of 73.3 kDa may account for the third band detected by the TCP30 antibody at 77 kDa.
5.3.5.2 Assessment of ILF3 mRNA and Protein Levels

The ILF3 protein bands detected in each cell type varied in quantity, molecular weight and number. However, comparison of the western blots probed with the DRBP76 antibody shows that the highest levels of the DRBP76 isoforms were seen in THP1, Hep G2 and U118 with moderate levels in B-cells and three of the skin fibroblast lines, and low levels in Huh7. There were no detectable DRBP76 isoforms in SW1353, HL60, leucocytes (F1, F2, M1, M2), and one of the skin fibroblast lines, SF3785. The TCP80 isoforms were highest in B-cells and Hep G2, moderate in Huh7, U118, THP1 and three of the skin fibroblasts (SF3351, SF3542, SF3693), low in SW1353 and almost undetectable in SF3785. There were no TCP80 isoforms in HL60 or leucocytes (F1, F2, M1, M2).

The levels of ILF3 isoform protein seen in leucocytes (nil), Hep G2 (high), skin fibroblasts and SW1353 (low) are consistent with the observed levels of mRNA (Figure 5.5). For Huh7, U118 and THP1, the protein levels vary depending on the antibody used but are potentially relatively comparable. However, B-cells appear to have moderately high levels of ILF3 isoform protein but low levels of ILF3 mRNA, while HL60 have moderately high levels of mRNA but no detectable ILF3 isoform proteins.

Variation between individuals is seen in the skin fibroblasts where both the 110 kDa and 90 kDa forms are seen in SF3351, while in SF3542 and SF3693, only the 90 kDa protein is seen. However, this does appear to be at a slightly lower molecular weight than seen in SF3351. In SF3785, very little if any ILF3 isoform proteins have been detected. This was unexpected as there is very little variation in ILF3 mRNA levels in skin fibroblasts. In leucocytes, ILF3 mRNA levels are overall very low and results for all individuals have been recorded as no ILF3 isoform proteins detected. However, it is interesting to note that the ILF3 mRNA levels for M1 are about three-fold higher than for the other individuals and close examination of the western blot for M1 reveals a very faint band at 98 kDa (red arrow, Figure 5.9). Limited sample availability prevented repeating of this experiment for leucocytes.
Figure 5.9 Western Blot Analysis of ILF3 Isoform Proteins

Western analysis using two antibodies, DRBP76 and TCP30, was performed on several human cell types as indicated above each blot. Three blots were run for each antibody. Hep G2 cells were run on each blot as quality control. A GAPDH loading control antibody was used to monitor consistent loading of 10 µg protein in each lane. The SDS-PAGE standards (107 kDa, 81 kDa, 48.7 kDa) can be seen on the TCP30 blots. The GAPDH loading control and molecular weight SDS-PAGE standards are visualised at 800 nm and 680 nm respectively and were run on all six blots. The production of images for this thesis was done at a single wavelength so only the GAPDH loading control or the molecular weight SDS-PAGE standards can be seen on each blot although both were present on all blots. The red arrow indicates the faint band in sample M1 referred to in the text.
B cells  SF3351  SF3542  SF3693  SF3785  Hep G2  B cells  SF3351  SF3542  SF3693  SF3785  Hep G2

kDa

81

48.7

107

103
to

93

77

TCP30

DRBP76

GAPDH loading control

Non-specific binding of GAPDH at 88 kDa

Hep G2  Huh7  SW1353  U118  HL60  THP1

Hep G2  Huh7  SW1353  U118  HL60  THP1

Non-specific binding of GAPDH at 88 kDa

Hep G2  F2  M2  F1  M1  Huh7

Hep G2  F2  M2  F1  M1  Huh7

Non-specific binding of GAPDH at 88 kDa

GAPDH loading control
Western blot analysis using two antibodies, DRBP76 and TCP30, was performed on several human cell types as listed. The bands detected by the DRBP76 antibody were in three size categories: 110 kDa, 103 to 93 kDa and 88 kDa. Where more than one band was detected in the 103 to 93 kDa range, the size of each is listed. The latter category, recorded in grey, is non-specific binding by the GAPDH antibody. The bands detected by the TCP30 antibody were also in three size categories: 110 kDa, 103 to 93 kDa and 77 kDa. All bands detected in the 103 to 93 kDa range are listed.
## DRBP76 Antibody

<table>
<thead>
<tr>
<th>Average Band Size (kDa)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H160</td>
<td>112</td>
<td>nd</td>
<td>107</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>110</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>THP1</td>
<td>101</td>
<td>102</td>
<td>100</td>
<td>98</td>
<td>102</td>
<td>98</td>
<td>100</td>
<td>98</td>
<td>98</td>
<td>95</td>
<td>95</td>
<td>96</td>
<td>99</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Hep G2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huh7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW1353</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3351</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3693</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3785</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3693</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3785</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH Control</td>
<td>34</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>34</td>
<td>35</td>
<td>35</td>
<td>34</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## TCP30 Antibody

<table>
<thead>
<tr>
<th>Average Band Size (kDa)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H160</td>
<td>112</td>
<td>nd</td>
<td>110</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>111</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>THP1</td>
<td>101</td>
<td>103</td>
<td>99</td>
<td>101</td>
<td>102</td>
<td>98</td>
<td>102</td>
<td>98</td>
<td>97</td>
<td>100</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>Hep G2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huh7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW1353</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3351</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3542</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3693</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3785</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3693</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF3785</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH Control</td>
<td>34</td>
<td>43</td>
<td>43</td>
<td>34</td>
<td>35</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*nd = no bands detected*
5.4 DISCUSSION

In this chapter, the variation in both GBA and ASA enzyme activity levels was found to be much greater than the variation in mRNA levels. For GBA, the 120-fold variation in enzyme activity levels was also much greater than previously reported by Doll and Smith (1993) who found that GBA enzyme activity varied over 60-fold compared to a variation in mRNA levels of 50-fold. ASA enzyme activity levels varied over 40-fold. The levels of these lysosomal enzymes were found to vary independently and furthermore, the relationship between the mRNA and enzyme activity levels is not direct. That is, higher mRNA levels did not necessarily lead to higher enzyme activity levels indicating that there is post-transcriptional regulation of GBA and ASA expression in at least some of the cell types examined.

Post-transcriptional regulation may consist of both translational and post-translational regulation. Factors affecting translation may include regulatory motifs, response elements, rate of initiation, mRNA secondary structure and polyadenylation while the rate of protein degradation and allosteric modulation are modes of post-translational control. The activity of an enzyme is the result of an equilibrium between the rates of synthesis and degradation of an enzyme. A translational efficiency ratio (TER), that is, a ratio of the mRNA level to the enzyme activity level for each cell type was used to assess and compare the extent of the observed post-transcriptional regulation. This ratio does not differentiate between translational and post-translational regulation.

A translational control protein (TCP80) has been reported to inhibit GBA translation in human cells (Xu and Grabowski, 1998). Cell and tissue levels of TCP80 were found to vary widely from high in embryonic kidney cells and B-cells to low in human intestinal smooth muscle cells (Xu and Grabowski, 1999). However, no data have been published on the extent to which GBA translation is inhibited, how it affects the final level of GBA enzyme activity and whether the magnitude of its effect is the same in all cell types. TCP80 was found to have significant homology to the \textit{ILF3} gene which has five known splice variants (Duchange \textit{et al.}, 2000) with NF90 being one of them. Xu \textit{et al.} (2000) use the term TCP80/NF90 to refer to TCP80 and although the cDNA sequences have 96\% homology (Kao \textit{et al.}, 1994), a two base pair insertion leads to a frameshift resulting in significant differences in the C-terminal amino acid sequences of the two proteins. It
seems more probable that there are in fact, six splice variants of \textit{ILF3} with NF90 and TCP80 being two different isoforms.

It would be preferable to measure the cellular levels of TCP80 and the other ILF3 isoforms individually. However, the significant level of homology between the ILF3 isoforms makes it very difficult to differentiate between them but also possible that isoforms other than TCP80 could be involved in the translational control of \textit{GBA}. Therefore, the levels of ILF3 mRNA in nine human cell types and the ILF3 isoforms detected by DRBP76 and TCP30 antibodies were measured to determine the contribution of ILF3/TCP80 to variation in GBA enzyme activity levels.

ILF3 mRNA levels were found to vary 24-fold from low in leucocytes to high in Hep G2 cells. Comparison of the TER for GBA and ASA to the ILF3 mRNA levels revealed similar patterns and in fact overall, surprisingly, the correlation for ASA was more consistent than for GBA. With the exception of B-cells, there is a trend towards the TER being higher in cells that have higher levels of ILF3 mRNA. That is, GBA and ASA translation appears to become less efficient as ILF3 mRNA levels rise. This association may also apply to leucocytes from individuals since the highest GBA and ASA TERs were seen in the individual with the highest ILF3 mRNA levels. There does not appear to be any correlation between ILF3 mRNA levels and TER in skin fibroblasts but the variation in ILF3 mRNA levels for skin fibroblasts was minimal.

Both \textit{GBA} and \textit{ARSA} were found to have comparatively low translational efficiency in B-cells. That is, between 4 and 25 times more GBA molecules must be produced in B-cells than in the other cell types examined in this study. Clearly, there are factors other than ILF3 inhibiting \textit{GBA} and \textit{ARSA} translation in these cells. These cells would be ideal to use for identifying additional factors involved in post-transcriptional control of \textit{GBA} and \textit{ARSA} if technical difficulties encountered in transfecting these cells could be overcome in a timely manner. For \textit{ARSA}, Hep G2 cells have both the highest ILF3 mRNA levels and highest TER. The translational repression of \textit{ARSA} in mouse testis is due to an RNA-binding protein that is absent from liver and brain (Kreysing \textit{et al.}, 1994), hence ILF3 could not be responsible for this previously described regulation.
To further analyse the relationship between ILF3 and the translational efficiency of GBA and ARSA, the TCP30 and DRBP76 antibodies were used to ascertain the presence of ILF3 isoforms and also to gauge their level in nine human cell types. The sequences of the proteins used to raise these antibodies were compared to each other and the sequences of all ILF3 isoforms. The TCP30 antibody was raised against a sequence that is common to all ILF3 isoforms except NF90 to which it differs by just four amino acids. The sequence used to raise the DRBP76 antibody is significantly different from the corresponding sequence for NF90 but homologous to the other ILF3 isoforms except for the nine C-terminal amino acids. While factors other than sequence define specificity of an antibody, it is not feasible to exclude the possibility that these antibodies detect more than one ILF3 isoform without precise knowledge of the epitopes involved. This information is not available for either antibody. Clearly, both antibodies have detected multiple bands in the target molecular weight range and it is proposed that, at 110 kDa and in the 93 to 103 kDa range, the same isoform/s has been detected by both antibodies.

There is clearly variation in both the number and quantity of the ILF3 isoforms present in different cell types and very interestingly, also between individuals in a single cell type. The 110 kDa isoform was found in six cell types and is always at lower levels than other isoforms in that cell type. This is consistent with its description as the ‘minor’ isoform and it is likely to be the ILF3 (ilf3-e) isoform which, although not completely characterised, has the longest 3'-tail. The NF90 isoform has been detected with the TCP30 antibody alone as predicted by an alignment of the sequences used to produce the antibodies. NF90 can be found in B-cells, Hep G2 and Huh7 with lesser amounts in SW1353 and U118 cells. It is absent from HL60, leucocytes and skin fibroblasts. One or more bands in the 93 to 103 kDa range can be found in all cell types in which an ILF3 isoform was detected. These bands probably represent some or all of the DRBP76, DRBP76α, DRBP76δ and TCP80 isoforms.

ILF3 mRNA levels were found to vary 24-fold and ILF3 isoform levels also show great variation. The estimated total amount of ILF3 isoforms present in each cell type appears relatively consistent with the corresponding ILF3 mRNA level in most cell types e.g. high levels in Hep G2 cells and very low levels in leucocytes. The exceptions are HL60 and B-cells. HL60 cells have moderately high ILF3 mRNA levels but no ILF3 isoforms were detected, while the quantity of ILF3 isoforms in B-cells appears to be higher than
expected when compared to the level of mRNA. However, as the GBA and ASA TERs for B-cells are so high, the level of ILF3 isoforms may in fact may be consistent with this effect. The absence of ILF3 isoforms in HL60 and their apparently increased level in B-cells when compared to mRNA levels would appear to indicate post-transcriptional regulation of ILF3. Alternatively, the possibility of variation in the rate of degradation of the enzyme between different cell types and individuals also warrants investigation.

The pattern of ILF3 isoform bands appears to be quite different in most of the cell types. More importantly, the differences in the isoform patterns for four individual skin fibroblast lines and possibly the M1 leucocyte sample may provide clues to variation in GBA enzyme activity levels between individuals. It could be most informative to investigate this situation further in the future.

5.5 CONCLUSION
This study has shown that the translational efficiency of both GBA and ARSA appears to be correlated to ILF3 levels. In most of the cell types investigated, the levels of the ILF3 isoforms appeared consistent with ILF3 mRNA levels. However, in B-cells, the level of ILF3 isoforms appears to be higher than expected when compared to the mRNA levels, potentially explaining some of the low level of post-transcriptional efficiency observed in these cells. This, together with the absence of ILF3 isoforms in HL60 cells despite a moderately high level of mRNA, indicates the likelihood of post-transcriptional regulation of ILF3 itself. The finding that ARSA is subject to post-transcriptional regulation in cells other than spermatocytes is novel, as is the involvement of ILF3 in this regulation. The pattern of ILF3 isoforms varied between cell types and maybe between individuals. Further investigation of the role ILF3 plays in the observed variation in enzyme activity levels between individuals is required.
CHAPTER 6

GENERAL DISCUSSION

AND

FINAL CONCLUSIONS
6.1 INTRODUCTION

Defining genotype phenotype correlation in many disorders is very challenging. If the mechanisms of development were such that every change in genotype resulted in a different phenotype and every different phenotype was the consequence of a difference in genotype, the study of the origin of organic variation would be greatly simplified. Given a knowledge of the phenotype, the underlying causal genotype could be unambiguously inferred and *vice versa*. However, the correlation between genotype and phenotype arises from four sources: (1) the relation between the DNA sequence and the chemical structure of proteins, (2) relations between the products of transcription and translation of the information coded in the genome, (3) the dependence of development and physiology on both the genotype of the organism and the temporal sequence of environments in which the organism develops and functions, and (4) the stochastic variations of molecular processes within cells (Lewontin, 2004). It is the first two of these that have been investigated for the genes in this project.

The two disorders selected as model diseases for this project were Gaucher disease because it is the most successful and best characterized example of enzyme replacement therapy used to treat a lysosomal storage disorder, and metachromatic leucodystrophy because it is a good target for gene therapy. Metachromatic leucodystrophy also provided a simple and well characterized model to test some of our hypotheses about the role that single nucleotide polymorphisms (SNPs) play in the level of enzyme expression in the population and the implications for diagnosis and treatment. Many studies of the genotype phenotype relationship within the lysosomal storage disorders including Gaucher disease and metachromatic leucodystrophy have been carried out (Beutler, 1997, Beutler, 2001, Regis et al., 2002, Theophilus et al., 1989, Tsuji et al., 1988, Zhao et al., 2003, Zimran et al., 1989, Kappler et al., 1991). Broad genotype phenotype correlation exists within Gaucher disease and metachromatic leukodystrophy and each disorder has three clinical variants based on age of onset and clinical severity. Many of the causative mutations are associated solely with one clinical variant and a most notable example of this is N370S which is always associated with non-neuronopathic Gaucher disease. However, genotypic heterogeneity is encountered among patients with similar clinical presentations and patients with the same genotype will not necessarily share the same phenotype, thereby causing difficulties in diagnosis and assessment of prognosis.
The proposal for this project evolved from the observation of the wide range in lysosomal enzyme activity levels present in the normal population. A very low level of enzyme activity is sufficient to remain asymptomatic as illustrated by individuals with pseudodeficiencies such as ASA-PD. If only a few percent of the average enzyme activity is sufficient, it is puzzling that some people have extremely high levels of enzyme activity. This project began with an examination of the causes of variation in the expression of human GBA and ASA in unaffected ‘normal’ individuals followed by investigation of the regulation of the expression of these genes in different cell types.

6.2 VARIATION IN LYOSOMAL ENZYME ACTIVITY

Although unaffected individuals are frequently included as controls, no separate studies of GBA and ASA expression in the normal population have previously been done. The ARSA and GBA genes were screened for high activity polymorphisms so that information gained from the identification and characterisation of high activity, natural sequence variants could be used to produce greater protein activity from a recombinant protein expressed in vivo for gene therapy and in vitro for enzyme replacement therapy and to improve diagnosis, phenotype prediction. However, no common high activity polymorphisms appear to exist in these genes.

Polymorphisms associated with low ASA activity were identified within the ARSA gene. The poly(A) polymorphism (2723A>G) was found to be significantly associated with low ASA activity as was already known (Gieselmann et al., 1989) and independent effects on enzyme activity originating at or near two additional polymorphisms (W193C and 1101+123C>T) were also found. While W193C appears likely to be associated with low ASA activity, the finding of the association of 1101+123C>T or a linked site to high ASA activity is novel. A larger sample population (which was not available for this project) or functional studies would be required to confirm this finding. If this polymorphism was found to increase enzyme activity even marginally, it would be useful for gene therapy purposes as a number of studies in vitro have suggested that only small amounts of additional lysosomal enzyme activity in deficient cells may be biologically effective and could correct the metabolic abnormality which leads to intracellular accumulation of undegraded storage products (Neufeld, 1991, Bou-Gharios et al., 1993).
The \( GBA \) gene appears to be remarkably well conserved and although only two major haplotypes have been reported (Beutler et al., 1992), it was anticipated that a greater number of polymorphisms would have been identified. A simple yet robust ‘whole \( GBA \) gene’ PCR was developed that enabled amplification of the \( GBA \) gene but not the nearby highly homologous \( GBA \) pseudogene. It is possible that the presence of the pseudogene or the importance of the biological role of \( GBA \) could result in selective pressure e.g. genes involved in cell structure or motility show a signature of negative or purifying selection (Bustamante et al., 2005), thereby reducing the number of polymorphisms that occur in \( GBA \). The lack of genetic difference implicates unlinked regulatory genes or environmental factors as the cause of the observed variation in individual \( GBA \) enzyme activity levels.

The study of enzyme activity variation in individuals over time was very limited but indicated that intra-assay variation fell mostly within the range of inter-assay variation. This observation together with the findings that enzyme activity levels often mirror mRNA levels appears to limit the likelihood of environmental influence. In addition, reports of the transcriptional regulation of \( GBA \) and the translational regulation of both \( GBA \) and \( ARSA \) appeared consistent with the theory of a predominantly genetic cause for the observed variation. Because of the previous research done on \( GBA \) gene regulation, the lack of genetic variation and the greater relevance to therapeutic applications, the remainder of the project concentrated on elucidation of the mechanisms for control of \( GBA \) expression. The parallel investigation of \( ARSA \), initially thought to be a ‘housekeeping’ gene, was done to allow comparison of the \( GBA \) results to those of another lysosomal enzyme.

### 6.3 GENE REGULATION

This research has shown that \( GBA \) and \( ARSA \) are independently regulated and ubiquitously expressed but with tissue specific higher levels of enzyme.

#### 6.3.1 \( GBA \)

\( GBA \) expression was found to be more highly regulated than \( ASA \) expression. The \( GBA \) mRNA and enzyme activity levels were measured in nine mammalian cell lines and \( GBA \) mRNA levels were found to vary thirty-fold while enzyme activity levels varied over 120-fold. This supports the findings of Doll and Smith (1993), although the variation in
enzyme activity levels found in this study was much greater. There are some differences in the findings for individual cell lines which provide the opportunity for further investigation. In this thesis, the mRNA levels of HL60 and THP1 were similar while Doll and Smith (1993) reported a large divergence in the levels, and the neuronal and skin fibroblast results also appeared to vary between the studies. A simple explanation may be that exactly the same cell lines were not used in each study but additional causes may be experimental or environmental variance. The ability to transfet the suspension cell lines included in this study may well have provided some of the answers if cell specific differences in *GBA* gene regulation had been identified.

The search for tissue specific elements in the region between *GBA* and *C1orf2* met with limited success. The putative *cis*-element region identified has certainly showed the potential to enhance promoter expression by up to 3.5-fold but it was not possible to demonstrate that it was acting on GBA in the cell lines in which it was tested. It is possible that there may be both stimulatory and inhibitory *cis*-elements with the region that could be identified with the use of deletion constructs but time did not permit the continuation of this line of investigation. If this putative *cis*-element region is not acting on GBA, then identification of the target gene will be both challenging and interesting.

It is disappointing and perplexing that no GBA regulatory elements were identified especially in light of the conserved nature of the GBA gene. However, there are still many ways in which mRNA levels can be regulated. These include the interaction of *trans* factors directly or indirectly within the gene or 3′-UTR, mRNA stability and interference due to convergent transcription of adjacent genes e.g. *MTX* pseudogene.

TCP80, a splice variant of *ILF3*, is reported to bind to the GBA coding region and cause translational inefficiency in mammalian cells (Xu and Grabowski, 1998). Five splice variants of *ILF3* have been reported (Duchange *et al.*, 2000) and it has been proposed that TCP80 has the greatest similarity with NF90 (*ilf3-d*). However, after comparison of all five ILF3 splice variant cDNA and amino acid sequences with that of TCP80, it would appear that TCP80 should be considered the sixth splice variant as there are significant differences at the amino acid level between NF90 and TCP80. There is 100% homology in the sequences of the RNA binding motifs for all splice variants and so the possibility that any or all of the splice variants can bind to GBA cannot be excluded. Furthermore,
the DRBP76 and TCP30 antibodies used to quantify the level of ILF3 appeared to bind to bands of the same size indicating the probable binding of the same proteins with one exception. NF90, a 77 kDa protein, was only detected with the TCP30 antibody as predicted by the sequence comparison of the immunogens used.

It was postulated that there may be a relationship between the level of TCP80 and GBA enzyme activity. The translational regulation of GBA was analysed in nine mammalian cell types with the use of the translational efficiency ratio. ILF3 mRNA levels were measured due to the difficulty in differentiating between the ILF3 splice variants and the levels of all proteins detected with both the DRBP76 and TCP30 antibodies compared to the GBA translational efficiency ratio. This study reports a trend towards greater inefficiency of GBA translation with rising ILF3 levels and the possibility that ILF3 splice variants other than TCP80 may also be involved. It may ultimately be informative to assess the ILF3 levels in GD patients especially in the tissues of pathology with the view to developing a therapeutic that inhibits the action of ILF3 on GBA, thereby facilitating the production of greater *in vivo* enzyme activity levels.

6.3.2 ARSA

The finding that ASA mRNA levels vary at least seven-fold while ASA enzyme activity levels vary over forty-fold is novel. It is expected that ‘housekeeping’ genes such as ARSA are expressed ubiquitously and with little variation. The cause of the variation in mRNA levels was not investigated but very high levels of ASA mRNA have previously been reported in mouse testis (Kreysing *et al.*, 1994). *In vitro* run-on assays gave no indication that the transcription of the ARSA gene in spermatogenesis was enhanced, so it was concluded that the high levels of ASA mRNA were solely due to stabilisation of the mRNA as the translationally active mRNA is extremely unstable, hence the increase in ASA mRNA was not accompanied by an increase in ASA enzyme activity (Kreysing *et al.*, 1994). The mRNA stabilisation was carried out by a testis-specific RNA binding protein that binds to both the 5′-UTR and 3′-UTR so it is possible that mRNA instability may be the cause of the ASA mRNA variation observed in mammalian cells. However, any protein involved would need to be expressed in many more cell types due to the variety of cells in which the ASA mRNA variation is reported.
The cell types selected for this project were based on previous work done on GBA transcription, the tissues of pathology in GD and those cell types available in our laboratory. Even though little consideration was given to the cell types in which it would be best to examine ASA expression, the selection made has been equally useful for the ASA investigations. The inclusion of mammalian gonadal cell lines would be considered for any further work.

The wide variation in ASA enzyme activity levels in mammalian cells was a little unexpected but even more unexpected and novel was the correlation between ASA translational efficiency and ILF3 expression. Clearly, there is a need for further investigation of this finding. The possibility that the correlation is purely coincidental for both genes has been considered but seems unlikely given the relatively large number of cell types used and the known relationship between GBA and TCP80.

6.3.3  ILF3

The ILF3 gene appears to be regulated at both the transcriptional and post-transcriptional levels. ILF3 mRNA levels were found to vary over twenty-fold and, although the protein levels of the splice variants were not quantified, there is clear evidence of variation between cell types and possibly even between individuals. Evidence of post-transcriptional regulation is seen in HL60 cells in which no splice variant proteins were detected despite moderate levels of ILF3 mRNA.

6.4  INDIVIDUAL VARIATION

Investigation of the variation in enzyme activity levels in individuals has shown that there is variation between individuals in the GBA, ASA and ILF3 mRNA levels in two different cell types and that the GBA and ASA translational efficiency ratio in leucocytes is correlated with ILF3 levels ($r = 0.99$). This suggests that GBA and ASA enzyme activity levels may be related to the expression of ILF3 in an individual.

The activity levels of these lysosomal enzymes in an individual appears likely to be the result of the level of transcriptional regulation and its determinants, the factors that influence the expression of the ILF3 gene, the final expression level of the ILF3 gene itself, post-translational modifications in addition to the influence of polymorphic variants within the lysosomal gene.
6.5 FUTURE STUDIES
This study has provided insights into the expression of GBA and ASA but also indicates
the need for further studies on the genetic causes of the variation in enzyme activity levels
between individuals that are not covered in this thesis. Furthermore, the transcriptional
control mechanisms for the GBA and ARSA genes, the roles of the ILF3 splice variants
in the expression of GBA and ASA, the additional GBA and ASA regulatory mechanisms
evident in B-cells, the control of ILF3 expression and the identification of the role of the
\textit{cis}-element region upstream from GBA and its target gene require investigation.

6.6 FINAL CONCLUSIONS
The outcomes of this project are as follows:

1. The development of a simple yet robust PCR method for specifically
amplifying the entire GBA gene.
2. There are no common polymorphisms associated with high enzyme activity
for the ARSA and GBA genes. However, for the ARSA gene, there is
correlation between certain polymorphisms and low activity and these will
influence the distribution of enzyme activity levels. The question of what
causes the wide distribution in activity levels and especially the high activity
levels remains unanswered. The GBA gene seems remarkably conserved.
3. The variation in enzyme activity levels in individuals is reflected at the
mRNA level and the TER is correlated with ILF3 mRNA levels.
4. GBA and ARSA are regulated independently and subject to both
transcriptional and translational regulation in numerous cell types with GBA
being more highly regulated.
5. A \textit{cis}-element cluster region that is capable of promoter regulation was
identified 4 kb upstream of GBA but it was not possible to demonstrate
that it acted on GBA.
6. The level of translational efficiency of GBA and ARSA is correlated with
the level of ILF3 expression.
7. ILF3 appears to be regulated at both the transcriptional and post-
transcriptional level.
8. In HL60 and B-cells, additional mechanisms of GBA and ASA expression
regulation are evident.
6.7 SUMMARY

Overall, this thesis has further elucidated the genetic mechanisms for the control of expression of GBA and ASA. The significant results of this investigation can be applied to future research into lysosomal gene regulation and the contribution of genetic variants to natural variation. This will lead to a greater understanding of the molecular pathogenesis of lysosomal storage disorders and improve diagnosis, the ability to predict disease severity and treatment.
CHAPTER 7

REFERENCES


