Phosphoglucomutase: Structural studies of the
PGM1 gene
and a search for PGM3 and PGM4

Jane H. Ives

April 1995

Thesis submitted for the degree of Doctor of Philosophy
in the University of London

MRC Human Biochemical Genetics Unit
The Galton Laboratory
University College London
University of London
Abstract

Phosphoglucomutase is an enzyme central to glycolysis and gluconeogenesis. There are three well characterised forms: PGM1, PGM2 and PGM3.

Genomic clones containing the 5' end of the PGM1 gene were isolated and characterised. The exon/intron boundaries of the first exon (designated 1A) were determined. 15kb of intron 1 have been mapped and 2.8kb of the flanking regions sequenced. The proximal promoter shows features of a 'housekeeping' promoter. Features characteristic of a CpG island have been identified in this region. An alternative first exon has been identified (1B). The two exons show similarities indicating a gene duplication event. Exon 1B is transcribed from a promoter in the first large intron of PGM1.1A. Expression studies of PGM1.1B by RNA-PCR show a limited tissue expression, with predominate expression in striated muscle.

Exons 1A and 1B were sequenced in DNA from eight individuals of known PGM1 protein phenotypes. No genetic variation in these sequences was encountered.

The PGM1 cDNA was used as a probe in a search for the PGM3 gene. A human chromosome 6 library was screened. Positive cosmids were assigned to chromosome positions by fluorescent in situ hybridisation, restriction mapped and sequenced in part. These studies led to the identification of a novel PGM on chromosome 9. PGM3 was not isolated by this procedure, suggesting that the sequences of PGM1 and PGM3 have diverged significantly.

Mammary gland cDNA libraries were screened using both anti-PGMI antibodies and PGM1 cDNA, to search for PGM4. Positive clones were all identified as PGM1. In addition the PGM1 type of DNA from recent mothers was compared to the typing of matched colostrum samples (by isoelectric focusing). Some similarities were seen between the colostrum and PGM1 isoform patterns. Together, these data suggest that PGM4 protein is a modified form of PGM1.
Acknowledgements

I would like to thank Dr. Yvonne Edwards, who supervised this work, for all her dedication, help and enthusiasm throughout the project. I would also like to thank Dr. David Whitehouse for his support and encouragement and both he and Prof. David Hopkinson for critical reading of this manuscript. Thanks also to everybody in the lab. for their varied help and support, with special mention to Dr. Jane Sowden for her friendship and endless patience, and Wendy Putt for help with everything technical. Thanks to Dr. Margaret Fox who supplied the FISH analysis, and Dr. Ruth March for the SSCP analysis, and a general thank you to lots of friends at the Galton who were supportive regardless of the situation including Katie, Julian, Una, Lisa, Felicity, Janine, Jenny, Rod and Alex. Enormous thanks go to Martine who is an excellent friend. A very special thank you goes to Nik, for all his love and support in the last year. Finally, thank you to my Mum, Graeme and Sarah for all their love and encouragement.

This thesis is dedicated to the memory of my father, William C. Ives.

This work was supported by a studentship from the Medical Research Council
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Contents</td>
<td>4</td>
</tr>
<tr>
<td>Figures</td>
<td>8</td>
</tr>
<tr>
<td>Tables</td>
<td>13</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>14</td>
</tr>
<tr>
<td>Chapter One Introduction</td>
<td>16</td>
</tr>
<tr>
<td>1.1 Phosphoglucomutase - the enzyme</td>
<td>16</td>
</tr>
<tr>
<td>1.1.1 The historical context</td>
<td>16</td>
</tr>
<tr>
<td>1.1.2 Purification of PGM protein</td>
<td>17</td>
</tr>
<tr>
<td>1.1.3 PGM isoforms</td>
<td>17</td>
</tr>
<tr>
<td>1.1.3.1 Structural and physical properties</td>
<td>18</td>
</tr>
<tr>
<td>1.1.3.2 Genetic properties</td>
<td>21</td>
</tr>
<tr>
<td>1.1.3.3 Chromosome assignments of ( PGM1 ), ( PGM2 ) and ( PGM3 )</td>
<td>26</td>
</tr>
<tr>
<td>1.1.3.4 PGM4 - is there a fourth locus?</td>
<td>26</td>
</tr>
<tr>
<td>1.1.4 Multiple loci</td>
<td>29</td>
</tr>
<tr>
<td>1.1.5 Protein structure</td>
<td>29</td>
</tr>
<tr>
<td>1.1.6 Immunological studies</td>
<td>32</td>
</tr>
<tr>
<td>1.2 Phosphoglucomutase 1 - the gene.</td>
<td>33</td>
</tr>
<tr>
<td>1.2.1 Isolation and characterisation of PGM1 cDNA</td>
<td>33</td>
</tr>
<tr>
<td>1.3 Clinical aspects of PGM</td>
<td>35</td>
</tr>
<tr>
<td>1.3.1 Deficiencies of enzymes involved in glycogen metabolism</td>
<td>35</td>
</tr>
<tr>
<td>1.3.2 PGM deficiency</td>
<td>36</td>
</tr>
<tr>
<td>1.4 Objectives</td>
<td>40</td>
</tr>
<tr>
<td>Chapter Two Materials and Methods</td>
<td>41</td>
</tr>
<tr>
<td>2.1 Materials</td>
<td>41</td>
</tr>
<tr>
<td>2.1.1 Standard reagents</td>
<td>41</td>
</tr>
<tr>
<td>2.1.2 Enzymes</td>
<td>41</td>
</tr>
<tr>
<td>2.1.3 Electrophoresis reagents</td>
<td>41</td>
</tr>
<tr>
<td>2.1.4 Miscellaneous</td>
<td>41</td>
</tr>
<tr>
<td>2.1.5 Common buffers</td>
<td>42</td>
</tr>
</tbody>
</table>
2.1.6 Microbiology media 42
2.1.7 Bacterial strains 43

2.2 Methods 43
2.2.1 DNA precipitation and assay of concentration 43
2.2.2 DNA modification 44
  2.2.2.1 Restriction enzyme digests 44
  2.2.2.2 Dephosphorylation 44
  2.2.2.3 Ligation 44
2.2.3 Agarose gel electrophoresis and recovery of DNA fragments 44
2.2.4 Preparation and transformation of competent *E. coli* 45
2.2.5 Large scale plasmid preparation 46
2.2.6 Mini-plasmid preparations 47
2.2.7 Large scale λ preparations 47
2.2.8 *In vivo* excision of pBluescript SK(-) from Uni-ZAP XR 48
2.2.9 Single-stranded DNA preparations 48
  2.2.9.1 Preparations from M13 48
  2.2.9.2 Preparations from pBluescript 49
2.2.10 Southern blotting 50
2.2.11 Preparation of ³²P labelled probes 50
2.2.12 cDNA library screening 51
  2.2.12.1 Screening with antibody probes 52
  2.2.12.2 Screening with cDNA probes 52
2.2.13 Screening the chromosome 6 library 53
2.2.14 Polymerase Chain Reaction (PCR) 54
2.2.15 'AT' direct cloning of unmodified PCR products 55
2.2.16 DNA sequencing 56
  2.2.16.1 Sequencing of single stranded DNA 56
  2.2.16.2 Sequencing of double-stranded templates 57
2.2.17 Isolation of RNA 58
2.2.18 Preparation of cDNA by Reverse Transcription and RT-PCR 58
2.2.19 RNAse Protection assay 59
2.2.20 Primer extension analysis 60
2.2.21 Isoelectric Focusing (IEF) 61
### Chapter Three  
**PGM1**  
3.1 The architecture of the 5' end of human PGM1  
3.2 Sequence analysis of the 5' end of PGM1  
3.3 The promoter region of human PGM1  
3.4 Exon 1B, an alternative first exon in PGM1: position and sequence  
  3.4.1 Evidence for a duplication of PGM1  
  3.4.2 The 5' extent of the PGM1.1B transcript  
  3.4.3 Expression of the PGM1.1B transcript  
3.5 Search for genetic variation in the 5' exons of PGM1  

### Chapter Four  
**PGM3**  
4.1 Isolation and analysis of cosmids from the sorted chromosome 6 library. 
  4.1.1 Sequence analysis of cosmid P1721 (6p23-24)  
  4.1.2 Sequence analysis of cosmid I238  
  4.1.3 Sequence analysis of cosmid L1347 (9p11, 9q12)  

### Chapter Five  
**PGM4**  
5.1 Isolation and analysis of cDNAs from mammary tissue cDNA libraries  
  5.1.1 A search for PGM4 cDNA using antibody probes  
  5.1.2 Isolation using human PGM1 cDNA as probe  
    5.1.2.1 Analysis of human recombinants  
    5.1.2.2 Analysis of mouse recombinants  
5.2 PCR amplification of human cDNA library  
5.3 IEF analysis of milk  

### Chapter Six  
**Discussion**  
6.1 PGM1 gene structure  
  6.1.1 Alternative first exons, function and significance  
  6.1.2 Exon 1A and the upstream promoter  
  6.1.3 Exon 1B and promoter region  

6
6.1.4 Intragenic duplication and the molecular basis of polymorphism 162

6.2 A search for PGM3 163
   6.2.1 Initial screening for PGM3 164

6.3 PGM4 167

References 171
## Figures

<table>
<thead>
<tr>
<th>Chapter One</th>
<th>Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 1.1</td>
<td>Diagram of PGM components of phenotypes PGM1 1, PGM1 2-1, PGM1 2, PGM2 1, PGM3 1, PGM3 2-1 and PGM3 2.</td>
<td>19</td>
</tr>
<tr>
<td>Fig. 1.2</td>
<td>PGM1 phenotypes on starch gel electrophoresis and further resolved by IEF.</td>
<td>22</td>
</tr>
<tr>
<td>Fig. 1.3</td>
<td>A proposed phylogeny relating 8 alleles in the PGM1 system.</td>
<td>25</td>
</tr>
</tbody>
</table>

| Chapter Three     | PGM1                                                                       |      |
|-------------------|                                                                            |      |
| Fig. 3.1          | The cloning site of λ2001                                                  | 64   |
| Fig. 3.2          | A Single and double digests of recombinant PGM1.5'                        | 65   |
|                   | B Autoradiograph following hybridisation of PGM1 cDNA to recombinant PGM1.5'. |      |
| Fig. 3.3          | Restriction map of PGM1.5'.                                                | 66   |
| Fig. 3.4          | A Restriction digests of XbaI/HindIII 2.8kb region of PGM1.5'             | 68   |
|                   | B Autoradiograph following hybridisation of 5' EcoRI/HindIII fragment of PGM1 cDNA to XbaI/HindIII 2.8kb region. |      |
|                   | C Restriction map of XbaI/HindIII 2.8kb region.                           |      |
| Fig. 3.5          | A Sequencing strategy of XbaI/HindIII 2.8kb region.                       | 69   |
|                   | B Summary of XbaI/HindIII 2.8kb sequence data.                            |      |
| Fig. 3.6          | Sequence of XbaI/HindIII 2.8kb region.                                    | 71   |
| Fig. 3.7          | Restriction map of PGM1.5' showing position and orientation of exon 1.    | 72   |
| Fig. 3.8          | Genomic clones which span the entire PGM1 gene.                           | 73   |
| Fig. 3.9          | Resolution of GC compression in intron 1                                   | 75   |
|                   | A Compressed sequence                                                      |      |
|                   | B Accurate sequence using dITP and ddITP mixes.                           |      |
| Fig. 3.10         | PGM1 exon 1 and flanking sequence                                          | 76   |
| Fig. 3.11         | A XbaI/HindIII 2.8kb fragment and position of exon 1                       | 77   |
|                   | B Plot of GpC distribution                                                 |      |
C Plot of CpG distribution
D Plot of HpaII restriction sites

Fig.3.12  
A Smal/HindI genomic fragment encompassing exon 1 used as a RNAse protection assay probe.  
B Multiple cloning site and positions of T7 and SP6 RNA polymerases promoters in plasmid pGEM-3Z (+).

Fig.3.13  
A Smal/HindI RNAse protection assay probe  
B RNAse protection assay of exon 1 with the Smal/HindI probe.

Fig.3.14  
A Comparison of the 5' ends of rabbit PGM1 and novel rabbit sequences.  
B PCR amplification of rabbit DNA with primers pgmppF and pgmppR

Fig.3.15  
A Genomic clones spanning the entire human PGM1 gene.  
B PCR amplification products of human genomic clones with primers pgmppF and pgmppR.

Fig.3.16  
A Single and double digests of PGM1 genomic clone λ3.  
B Autoradiograph of exon 1B PCR product hybridised to clone λ3.  
C Restriction map of clone λ3.

Fig.3.17  
A Position of exon 1B within the PGM1 genomic structure.  
B Sequence in and around exon 1B.

Fig.3.18  
Comparison of rabbit and human exon 1B and upstream sequences.

Fig.3.19  
A Nucleotide sequence comparison between human exons 1A and 1B.  
B Amino acid sequence comparison of human exons 1A and 1B.

Fig.3.20  
PGM1 genomic structure.

Fig.3.21  
A Position and orientation of primer used for exon 1B primer extension analysis.  
B Autoradiograph of exon 1B primer extension analysis.

Fig.3.23  
Position of PCR primers for the amplification of exons

9
1A, 1B and the exon 2/ exon 5 control product in reverse transcriptase analysis of expression.

Fig.3.24 Analysis of expression of exons 1A and 1B in a range of human tissues by reverse transcriptase PCR.

Fig.3.25 Reverse transcriptase PCR analysis of expression of exons 1A and 1B in rabbit muscle.

Chapter Four

Fig.4.1 Chromosome 6 cosmids Do446, L1347, P1721 and I238:
A Restriction enzyme digests
B Hybridisation to PGM1 cDNA.

Fig.4.2 Chromosome 6 cosmids C1011, C1847, C1853 and C1849:
A Restriction enzyme digests
B Hybridisation to PGM1 cDNA.

Fig.4.3 Psfl digests of cosmids D046, P1721 and C1011, and hybridisation to PGM1 cDNA.

Fig.4.4 Fluorescent in situ hybridisation of cosmids L1347 (Panel A) and P1721 (Panel B) to chromosome spreads from blood of a normal 46XX female.

Fig.4.5 Fluorescent in situ hybridisation of cosmids I238 (Panel A) and C1847 (Panel B) to chromosome spreads from blood of a normal 46XX female.

Fig.4.6 Fluorescent in situ hybridisation of cosmids C1853 (Panel A) and C1849 (Panel B) to chromosome spreads from blood of a normal 46XX female.

Fig.4.7 A Sequence of 489bp Sphl/HindI fragment from cosmid P1721.
B BESTFIT sequence comparison of P1721 sequence to PGM1 cDNA.

Fig.4.8 450bp of sequence from cosmid I238.

Fig.4.9 Sequence of the 900bp HindII fragment from cosmid L1347.

Fig.4.10 GAP computer comparison of L1347 sequence and human PGM1 cDNA sequence.
Fig.4.11  Comparison of amino acid sequence in the region of high homology between L1347 and exon 5 of human PGM1.

Fig.4.12  
A Sequence of PCR primers L1347F and L1347R. 
B L1347F/L1347R PCR amplifications of human and rodent genomic DNAs, and human chromosome 6 and chromosome 9 DNAs.

Chapter Five  PGM4

Fig.5.1  Illustration of the relative positions of PGM1 exons and protein domains, and the regions of the protein used to raise anti-PGM antibodies.

Fig.5.2  
A Position and sequence of λgt11 forward and reverse primers relative to the LacZ gene.

Fig.5.3  
A PCR amplification products of clones 1, 4, 5, 6 and 7. 
B Autoradiograph following hybridisation of PGM1 cDNA to clones 1, 4, 5, 6 and 7.

Fig.5.4  
A Sequence of clone MPL-1. 
B Sequence of clone MPL-5.

Fig.5.5  
A HindII and RsaI digests of clones H1, H2, H3, H4 and H5.
B Autoradiograph following hybridisation of PGM1 3' probe to clones H1, H2, H3, H4 and H5.

Fig.5.6  
Aligned restriction maps of PGM1 cDNA and clones H1, H2, H4 and H5.

Fig.5.7  
A Illustration of the 5' extent of clones H1, H2, H4 and H5.
B Position an sequence of the PGM1 3' sequencing primer.

Fig.5.8  
Proposed structure of clone H3 aligned to PGM1 exons 2 and 4 and intervening sequences.

Fig.5.9  
PstI, HindIII and HincII restriction sites in and around PGM1 exons 2 to 4, aligned to the PstI, HindIII and HincII sites in clone H3.

Fig.5.10  
5' sequence of mouse clone M1 aligned to the 5' untranslated and exon 1 sequence of human PGM1.

Fig.5.11  
Mouse clone M1 sequence aligned to rat and rabbit PGM and to human PGM1 sequences.
Fig.5.12 Sequence at the 5' end of mouse clone M2 aligned to human PGM1 sequence.

Fig.5.13 Position of PGM1 PCR primers Ex2, Ex3/4 and Mg, and expected and observed sizes of amplification products from the human mammary tissue library.

Fig.5.14 Sequence of Ex2/Mg/300 PCR product.

Fig.5.15 A Colostrum samples from mothers following isoelectric focussing and staining with a PGM activity overlay.

B Diagrammatic representation of stained bands shown in A.
Tables

Chapter One  Introduction
Table 1.1  Multiple loci in some of the glycolytic enzymes.  30
Table 1.2  Glycogen storage diseases.  37

Chapter Three  PGM1
Table 3.1  Summary of the number of samples collected from individuals of the ten common PGM1 phenotypes and numbers analysed by sequencing exon 1A and exon 1B.  101

Chapter Four  PGM3
Table 4.1  Summary of the chromosomal localisations of chromosome 6 cosmids L1347, P1721, I238, C1847, C1853 and C1849 determined by fluorescent in situ hybridisation.  110

Chapter Five  PGM4
Table 5.1  Summary of the screening results of the human nad mouse mammary tissue libraries with anti-rabbit PGM and anti-human PGM1 6' and 10' antibodies.  125
Table 5.2  Summary of screening of human and mouse libraries with PGM1 cDNA as probe.  131
Table 5.3  Summary of sizes of cDNA inserts of human breast tissue recombinant clones H1, H2, H3, H4 and H5, estimated by PCR.  132
Table 5.4  Summary of the SSCP analysis of PGM1 exon 4 and exon 8 from mothers' DNA samples, and the deduced PGM1 haplotypes of these individuals.  149

Chapter Six  Discussion
Table 6.1  Comparison of features of proximal and distal PGM1 promoters.  159
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>adenosine 5' triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>Ci</td>
<td>curie</td>
</tr>
<tr>
<td>CIP</td>
<td>calf intestinal phosphatase</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>DAB</td>
<td>diaminobenzidine</td>
</tr>
<tr>
<td>ddw</td>
<td>double distilled water</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethylpyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNAse</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleoside triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>IEF</td>
<td>isoelectric focusing</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>MMLV</td>
<td>Moloney murine leukaemia virus</td>
</tr>
<tr>
<td>MTT</td>
<td>(3-[4,5-dimethylthiazol-2-y]-2,5)-diphenyltetrazolium bromide</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>pfu</td>
<td>plaque forming unit</td>
</tr>
<tr>
<td>PGM</td>
<td>phosphoglucomutase</td>
</tr>
<tr>
<td>PGM</td>
<td>phosphoglucomutase gene</td>
</tr>
<tr>
<td>pipes</td>
<td>1,4-piperazinediethanesulphonic acid</td>
</tr>
<tr>
<td>PMS</td>
<td>phenazine methosulphate</td>
</tr>
<tr>
<td>psi</td>
<td>pounds per square inch</td>
</tr>
<tr>
<td>PSM</td>
<td>phage storage medium</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
</tbody>
</table>
RNAse: ribonuclease
RNAsin: ribonuclease inhibitor
rpm: revolutions per minute
SAP: shrimp alkaline phosphatase
SCD: systemic carnitine deficiency
SDS: sodium dodecyl sulphate
TEMED: N,N,N',N'-tetramethylethylenediamine
Tris: 2-amino-2-(hydroxymethyl)-propane-1,3-diol
UV: ultraviolet
X-gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Chapter One
Introduction

Phosphoglucomutase is an enzyme central to glycolysis and gluconeogenesis. In evolutionary terms it is an ancient protein, found in all orders of organisms from bacteria to mammals, reflecting its important position in a central metabolic pathway. PGM1 was one of the first polymorphic loci to be recognised. The high degree of genetic variation and the occurrence of multiple loci has made it the focus of genetic, forensic and molecular studies.

Phosphoglucomutase will be presented here in its historical context together with a discussion of its catalytic properties. Genetic studies will also be presented, revealing insight into the molecular basis of the polymorphism of PGM1, and ideas concerning the evolutionary mechanisms which gave rise to the PGM gene family.

1.1 Phosphoglucomutase - the enzyme

1.1.1 The historical context

The history of research into the glycolytic pathway can be traced to the turn of this century when interest was stimulated by the studies of mechanisms involved in alcoholic fermentation. Important contributions were made in 1905 by Harden and Young. They observed that inorganic phosphate disappeared in the course of fermentation and inferred that it was being incorporated into sugar phosphate. Two types of components in fermentations were recognised, and later identified as enzymes and cofactors (metal ions, ATP, ADP and coenzymes). Later, studies of muscle extracts showed that many of the individual reactions which were steps in lactic acid biosynthesis and alcoholic fermentations were the same and this finally led to the elucidation of the glycolytic pathway in 1940.

By the mid 1930's the existence of enzymes that catalyse an intramolecular transfer of phosphate, the phosphomutases, had been confirmed (Cori et al, 1938). A class of phosphomutases, the phosphoglucomutases (PGM), were found to catalyse the transfer of phosphate between the 1 and 6 positions of glucose. This interconversion provides an important link in carbohydrate metabolism, whereby the units produced by glycogen degradation (glucose 1-
phosphate) can be converted to glucose 6-phosphate and hence enter the glycolytic pathway.

In the deduction of the catalytic activity of PGM, Leloir et al. (1948) and Cardini et al. (1949) found that glucose 1-6 bisphosphate played an important part in the reaction. It acts as a type of cofactor for the enzyme (Ray and Roscilli, 1964) and is formed as an intermediate in catalysis when the phosphoryl group of the active enzyme molecule is transferred to the hydroxyl group at C6 of glucose 1-phosphate. The C1 phosphoryl group of this intermediate is then transferred to the active site of the enzyme resulting in glucose 6-phosphate and regeneration of the phosphoenzyme (Najar and Pullman, 1954). At equilibrium the reaction mixture contains 95% glucose 6-phosphate (Stryer, 1988). The phosphoryl group on PGM is slowly lost by hydrolysis, but is restored by phosphoryl transfer from glucose 1,6 bisphosphate, catalysed by phosphoglucokinase. An analogous catalytic activity is that of phosphoglyceromutase, which functions in glycolysis interconverting 3 phosphoglycerate to 2 phosphoglycerate. A phosphoenzyme intermediate also participates in this reaction.

1.1.2 Purification of PGM protein

Purified PGM protein was first isolated from rabbit muscle tissue extracts (Najar, 1955; Najar, 1962; Joshi et al. 1967; Ray and Peck, 1972) by a series of precipitations involving pH, heat, and addition of ammonium sulphate with subsequent purification on modified cellulose or Sephadex resins. Procedures for the isolation of PGM protein from a variety of other organisms, both prokaryotic and eukaryotic, including shark muscle (Hashimoto and Handler, 1966), human muscle (Joshi et al., 1967), rabbit liver, rat muscle, rat liver (Joshi et al., 1967; Hashimoto et al., 1967) and Escherichia coli (Joshi and Handler, 1964) were developed.

1.1.3 PGM isoforms

Studies of red cell lysates revealed the presence of several isozymes which could be distinguished by starch gel electrophoresis and could be shown to be genetically determined (Spencer, Hopkinson and Harris, 1964). Two sets of isozymes showed independent genetic variation. The least anodal set showed the highest frequency of variation and was designated PGM1. The next set towards the anode did not differ when samples of different PGM1 phenotypes
were compared implicating a second structural locus, *PGM2* (Hopkinson and Harris, 1965). Subsequent studies of mouse PGM isozymes also revealed several distinct molecular forms which could be divided into at least two zones of PGM components and the PGM phenotype was shown to be determined by at least two different genes (Shows et al, 1969).

While the isozymes attributable to these two loci made up the majority of PGM activity in human tissues, electrophoretic analysis of fibroblast cells revealed a third set of components (Hopkinson and Harris, 1968). The genetic variation in the electrophoretic patterns of these isozymes was independant of that identified for PGM1 and PGM2, thus implying the presence of a third locus, *PGM3*.

The relative positions of the three sets of isozymes as demonstrated by starch gel electrophoresis are illustrated in Fig.11. The sets of isozymes associated with each PGM isoform are composed of both primary polypeptide products (the least anodal bands) and more negatively charged secondary modification products: in addition to the primary product each isozyme set includes two or three secondary isozymes. The relative activities of the isozymes in a set determined by any one locus varies with the average age of the cell population sampled (Harris, 1980). For example, the primary polypeptide product is more prominent in extracts of cultured lymphocytoid cells, where there is more new protein, than in normal circulating red cells where the average age of the proteins is greater (Fisher and Harris, 1972). The *in vivo* mechanism of the post-translational modification generating the secondary isozymes is not known in detail, but might be due to deamidation of certain asparagine or glutamine residues in the molecule (Harris, 1980).

1.1.3.1 Structural and physical properties

The relative activities of the isozyme products of the three loci have been measured in a range of human tissues (McAlpine et al, 1970). In most tissues the PGM activity is predominately that of the PGM1 isozymes (85-95% of the total PGM activity). The remainder is 2-15% PGM2 activity and 1-2% PGM3 activity. The high proportion of PGM1 activity could be due to higher specific activity or a higher rate of synthesis of these isozymes. There are two cell types which show an exceptional distribution; in red cells, activity is provided approximately equally
Fig. 1.1
Diagram showing PGM components of phenotypes PGM1 1, PGM1 2-1, PGM1 2, PGM2 1, PGM3 1, PGM3 2-1 and PGM3 2 (modified from Hopkinson and Harris, 1968).
by PGM1 and PGM2 isozymes but no PGM3 isozymes are found, however, in fibroblasts, PGM3 isozymes are relatively more active and account for approximately 7% of the PGM activity.

It is interesting to compare the physical properties of the three loci products. The active isozymes are all monomeric. The sizes of the human PGM isozymes have been estimated by gel filtration and ultracentrifugation (Joshi et al., 1967; Santachiara and Modiano, 1969(a); McAlpine et al., 1970). These studies indicate that the isozymes determined by different alleles at the same locus, and those within a set determined by a single allele have the same molecular weight. In contrast, size measurements of the products of different loci, made by gel filtration, show each to be different (McAlpine et al., 1970). A consensus of the sizes presents the PGM2 isozymes at M.W. 61,000, significantly larger than the PGM3 isozymes (M.W. 53,000) which are slightly larger than the PGM1 isozymes (M.W. 51,000).

These structural differences are reflected in differences in thermostability (McAlpine et al., 1970). The PGM2 isozymes are the most stable (activity is hardly affected after 40 minutes at 60°C) while the PGM1 isozymes are less stable (weak after 20 minutes at 60°C and weaker but still detectable after 40 minutes at 60°C) and those of PGM3 the least stable (activity no longer detectable after 10 minutes at 60°C). However, there are no significant inter-allelic differences in thermostability in any one locus. Given that in vitro measurements of stability closely reflect the in vivo situation, then these stability differences between loci products may account for the relative proportions of each set of isozymes in different tissues. As discussed, in red cells a relatively much higher proportion of activity is derived from PGM2 isozymes than in other tissues. Circulating red cells are on average a few weeks old, they have no nucleus and so have no protein synthesis. It would be expected in this case that the most stable enzymes would predominate. Therefore in a normal blood sample where most of the cells are relatively old, the activities of PGM1 and PGM3 decline more rapidly than PGM2. In contrast PGM3 is most abundant in rapidly dividing fibroblasts.
1.1.3.2 Genetic properties

Starch gel electrophoresis studies revealed the existence of three isoforms of PGM, but in addition differences in the banding patterns were seen amongst individuals. As such PGM became one of the first genetic polymorphisms to be studied by electrophoresis. Each of the isoforms PGM1, PGM2 and PGM3 exhibits different degrees of polymorphism and varying frequencies of these forms in different populations.

The PGM1 locus is the most widely studied of the three PGM loci. The phenotypes were first separated using starch gel electrophoresis (Spencer et al., 1964). Four isozyme bands were identified: two bands associated with one allele designated PGM1*1 and two bands with a second allele, PGM1*2. In each case one of the bands is the primary gene product and the other is a secondary modification product. PGM1*1 is found at a frequency of 0.757 and PGM1*2 at a frequency of 0.242 (Harris and Hopkinson, 1976). In addition to these alleles, other rare electrophoretic variants have been observed. The alleles PGM1*3, PGM1*4 and PGM1*5 were identified by starch gel electrophoresis and observed in heterozygous combination with one of the common alleles (Hopkinson and Harris, 1965). Alleles PGM1*6 and PGM1*7 were also identified in heterozygous combination with both PGM1*1 and PGM1*2. The frequency of these alleles was found to be very low in 'English' populations. In each case (from a group of 2000 individuals) only one example of each of the alleles PGM1*3, PGM1*4, PGM1*5 and PGM1*7 was observed (Hopkinson and Harris, 1966). However, the PGM1*3 and PGM1*7 alleles occur with higher frequencies in Japanese populations (reviewed in Blake and Omoto, 1975) with PGM1*7 at 0.014 and PGM1*3 at 0.0015. The PGM1*3 allele determines two new isozyme bands with mobilities slower than the fastest PGM1*2 band and faster than the slowest PGM2 band, and the PGM1*7 type has two components, one at a similar position to the fastest component of PGM1*1 and the other faster than the fastest PGM1*2 band.

The development of isoelectric focusing allowed further investigation of protein polymorphisms. In this type of electrophoresis the thin layer polyacrylamide gel contains a pH gradient with the acid end of the gradient at the positive pole and the basic end at the negative pole. Charged proteins migrate in the gel to the isoelectric point, defined as the position in the pH gradient at which
Fig. 1.2
PGM1 phenotypes on starch gel electrophoresis and further resolved by isoelectric focusing, (modified from Carter et al, 1979).
contributions of negatively and positively charged groups exactly balance. Using this method a further interpretation of the PGM1 system was made by Bark et al. (1976), this is illustrated in Fig.1.2. The PGM1 1 phenotype was further resolved as three phenotypes designated PGM1 1+, PGM1 1-, PGM1 1+1-; the PGM1 2 phenotype was resolved as the three phenotypes designated PGM1 2+, PGM1 2- PGM1 2+2-; the PGM1 2-1 phenotype was resolved as four phenotypes designated PGM1 2-1+, PGM1 2+1-, PGM1 2-1- and PGM1 2+1+. This now gave a total of ten common PGM1 phenotypes explained by the occurrence of four common alleles $PGM1^*1+$, $PGM1^*1-$, $PGM1^*2+$ and $PGM1^*2-$. In each case the '+' refers to the more anodal band and the '-' to the more cathodal band. The new interpretation increased the value of PGM1 as a forensic marker in the identification of individuals. This can be expressed in terms of "discrimination power" which is the probability that two individuals would differ in phenotype (Sensabaugh, 1982). With the starch gel resolution of phenotypes the discriminating power is 0.56 (56 pairs out of 200 pairs of samples are discriminated) whereas with isoelectric focusing typing the discriminating power is 0.74 (74 pairs out of 200 are discriminated) (Bark et al, 1976).

All populations studied have the PGM1 1 and PGM1 2 groups of phenotypes (Hopkinson and Harris, 1966). The frequencies of the four common alleles $PGM1^*1+$, $PGM1^*1-$, $PGM1^*2+$ and $PGM1^*2-$ vary within populations (reviewed in Roychoudhury and Nei, 1988) however a study of four human populations showed that $PGM1^*1+$ is the most frequently occurring allele, with for example a frequency of 0.62 in Northern European populations (Carter et al, 1979). Variations in the frequencies of these four common alleles between different racial populations, although statistically significant (Carter et al, 1979), do not give discriminating power between races (Sensabaugh, 1982).

A mechanism for the evolution of the four common alleles was first proposed by Carter et al (1979) based on a consideration of the PGM1 polymorphism patterns revealed by isoelectric focusing. Isoelectric focusing typings of a range of fourteen higher primates showed an isozyme pattern the same as that of human $PGM1^*1+$, and as $PGM1^*1+$ is the most commonly occurring human allele it was proposed that it could be the ancestral $PGM1$ gene. In addition the observation was made that the difference in isoelectric points
between '+' and '-' for both the 1 and 2 alleles is similar. This key observation led to the novel suggestion that the other alleles could have evolved from PGM1*1+ by two independent mutations, one forming the '-' type and the other the '2' type, and a subsequent intragenic crossover between the different alleles to form the fourth '2-' type. The mechanism could be applied to the evolution of the other rarer allelic types and indeed this idea was expanded by Takahashi et al (1982) with particular reference to the PGM1*3 and PGM1*7 alleles. The proposed phylogeny is depicted in Fig1.3. They proposed that the four alleles PGM1*3+, PGM1*3-, PGM1*7+ and PGM1*7- could be explained by one more nucleotide substitution and a further three intragenic recombination events.

In summary, the model proposes a total of three nucleotide substitutions and four intragenic recombination events at two separate sites to form the eight common alleles. The involvement of intragenic reciprocal recombination in the creation of variation within a locus is rare. This type of exchange of genetic information causes no alteration in the overall size of the locus, relatively subtle protein variation and can be difficult to distinguish from more usual polymorphisms involving point mutations only. In contrast, unequal intragenic recombination leads to gain or loss of genetic material forming hybrid genes and therefore greater changes to the resultant protein seen, for example, in the globin genes coding for Lepore haemoglobin (Weatherall et al, 1989).

PGM2 is monomorphic in virtually all populations with individuals homozygous for the PGM2*1 allele. However, it was the recognition of some rare, independently segregating variants which led to the firm identification of PGM2 as a separate locus (Hopkinson and Harris, 1965). The PGM2*2 allele was found in black Africans living in England, Nigeria and South Africa and in these populations achieved polymorphic frequencies (Hopkinson and Harris, 1966). In addition other rare alleles found in a study of an English family were PGM2*4 and PGM2*5 (Parrington et al, 1968), and a distinct PGM2*5 allele was observed in South American Indians with a gene frequency of 9.8% (Geerdink et al, 1974). Rare phenotypes PGM2 4-1 and PGM2 4-2 have been reported in Indian and African Pygmies respectively (Santachiara-Benerecetti et al, 1972; Santachiara-Benerecetti and Modiano, 1969(b)). The PGM2*3 allele was found to occur at very low frequency in English populations (Hopkinson and Harris, 1966), in Negroes (Hopkinson, 1968) and Aboriginals where the PGM2 3-1 phenotype is at
Fig. 1.3
A proposed phylogeny relating eight alleles in the PGM1 system (modified from Takahashi et al., 1982).
polymorphic frequencies (Blake and Omoto, 1975). Other recognised alleles are PGM2*6, PGM2*8, PGM2*9 and PGM2*10 (reviewed in Blake and Omoto, 1975).

The genetics of the PGM3 polymorphism was investigated by typing the placentae from dizygotic twin births (Hopkinson and Harris, 1968). The frequencies and electrophoretic pattern of the phenotypes could be accounted for by two alleles: PGM3*1 and PGM3*2. The English population was found to be polymorphic for these two alleles with gene frequencies of 0.74 and 0.26 for PGM3*1 and PGM3*2, respectively. In a parallel study of the Nigerian population, the same alleles were represented at different frequencies of 0.34 and 0.66 for PGM3*1 and PGM3*2, respectively. All populations studied so far are polymorphic at this locus. European, Asian and North American gene frequencies are in a similar range to those first described for the English population, whereas the Australian frequencies are between the English and the Nigerian values (reviewed in Roychoudhury and Nei, 1988).

1.1.3.3 Chromosome assignments of PGM1, PGM2 and PGM3

Due to the polymorphic nature of the three PGM loci, it has been possible to assign them to chromosomes by isozyme and linkage analysis. Isozyme analysis in rodent-human somatic cell hybrids allowed PGM1 to be assigned to chromosome 1 (Harris, 1980) and the regional localisation was then refined by linkage analysis to 1p22 (Bruns and Sherman, 1989). In the CEPH consortium linkage map of human chromosome 1 (Dracopoli et al, 1991) multipoint analysis of 58 loci on chromosome 1 positioned PGM1 distal to the markers D1S22 and ACADM and proximal to D1S19. The same order was reported by Pakstis et al (1989) after linkage analysis of non-CEPH pedigrees but the shortest region of overlap of these markers implied that PGM1 must lie in 1p31. PGM2 was assigned to chromosome 4 at 4p14-q12, and PGM3 to chromosome 6 at 6q12 (McAlpine et al, 1990).

1.1.3.4 PGM4- is there a fourth PGM locus?

In 1982 Cantu and Ibarra made an electrophoretic study of human milk samples and found patterns of PGM activity which could be explained by the existence of a fourth PGM locus distinct from PGM1, PGM2 and PGM3. The samples studied were from 140 Mexican women who had given birth 3 to 10
Chapter 1

days previously. The electrophoresis pattern of each milk sample was compared with that of a sample of erythrocytes from the same woman. Eight milk specific phenotypes were distinguished and explained on the basis of a distinct PGM locus with four alleles PGM4*1, PGM4*2, PGM4*3 and PGM4*4. The PGM4-4 homozygote was not found and the *4 allele was only found in combination with *1 and *2 alleles. The gene frequencies of the alleles were studied and analysis showed good agreement with the Hardy-Weinberg expectation (Cantu and Ibarra, 1982).

Other studies were made of PGM4. For example, tests using a pentose sugar as substrate showed that PGM activity in milk did not show phosphopentomutase activity (characteristic of PGM2). Thermostability tests were conducted with samples heated to a range of temperatures before electrophoresis. These showed that after 15 minutes PGM4 activity disappeared at 60°C, PGM1 at 65°C and PGM2 at 70°C. This result is in accordance with previous estimates of thermostability (McAlpine et al., 1970), where after 10 minutes heating at 60°C both PGM1 and PGM2 isozymes were stable, and after 20 minutes at 60°C PGM1 isozymes had reduced but still detectable activity and PGM2 isozymes had intact activity. Thus, this result supports the idea of a distinct PGM protein, PGM4.

Cantu and Ibarra found that in non-lactating mammary gland extract, PGM activity was composed of PGM1 and PGM2 isozymes. In milk however they saw no evidence of these isozymes and so proposed a 'switch' from PGM1 and PGM2 gene expression to PGM4 gene expression, which is reversible and probably dependant on the hormonally controlled activation of lactation. Kvito and Weimer expanded on this proposal in their study of PGM activity in milk samples taken from a Brazilian population (Kvito and Weimer, 1990). Electrophoresis patterns in 652 colostrum samples (taken 24 to 48 hours after birth) were typed and milk samples from 175 of these individuals were typed approximately 17 days later. The same eight phenotypes were identified in this study as in Cantu and Ibarra's study. However, in the Brazilian study, the gene frequencies of the alleles seen in the colostrum samples did not fit the Hardy-Weinburg equilibrium, and furthermore, in 15% of the colostrum samples no PGM activity was detectable at all. In the milk samples examined later, 36% of women who had detectable patterns in colostrum showed a different phenotype in milk and these
changes were always from a proposed homozygote to a heterozygote phenotype. As in the Mexican study, gene frequencies of alleles in the milk samples did fit the Hardy-Weinburg expected values but the frequencies were dissimilar to those of the Mexican population.

To explain their observations Kvito and Weimer (1990) proposed either that there may be differences in enzyme activity at the beginning of lactation or that there may be variability in regulation of the locus. In the first proposal, low activity may result in false typing initially of the colostrum with higher activity later in lactation accounting for the apparent changes in phenotype. In the second proposal they suggest segregation of two regulatory alleles, one for early and one for late activation. The homozygote 'late' individual would have no PGM activity in colostrum, the homozygote 'early' would have the same phenotype in colostrum and milk, whereas the heterozygote would have a changing phenotype. A possible molecular model for the regulatory allele would be changes in the promoter and upstream regulatory regions which altered the binding of hormone receptor sites hence altering sensitivity to hormonal activation of lactation. From the results of their study there appears to be no association between structural allele status and regulatory allele status. In this case the model for variability in regulation of the locus becomes complex, with all combinations of late and early and all of the four recognised structural alleles necessary to explain the observed electrophoretic patterns. In view of this complexity, the first proposal of differential activity of the enzyme during lactation is perhaps more attractive and could be explained by environmental factors affecting the general health of the mother: in this context it would be of interest to investigate the activity levels of other enzymes in milk samples.

Further information regarding PGM4 comes from immunological analysis. An anti-rabbit muscle PGM antiserum which had been shown to be monospecific in its immunoreactivity to PGM1 isozymes (i.e. it does not cross react with PGM2 or PGM3) has been shown to immunoreact with PGM4 isozymes (Drago et al 1991). In terms of extending the PGM gene family this would suggest that PGM4 arose by a recent gene duplication where sequence homology and, as a result, antigenic determinants have been conserved (Drago et al 1991).
Confirmation of a distinct gene locus can often be provided by demonstrating inherited independent, genetic variation. However, since PGM4 shows expression only in milk, it is precluded from this type of study. In view of the lack of information about the heritability of the putative polymorphism, another interpretation of the data is that PGM4 is encoded by PGM1 and arises either by modification of the PGM1 mRNA or protein product. For example the PGM1 gene could undergo alternative splicing events, or PGM1 protein could be modified post-translationally.

1.1.4 Multiple loci

As described above there are at least three isoforms PGM1, PGM2 and PGM3 encoded by three separate gene loci. Gene families are a feature common to many proteins, including other enzymes which act in glycolysis. Isoforms of enzymes were initially grouped together on the basis of similar catalytic properties but structural studies of the proteins, and in most cases analysis at the level of the cDNA, have confirmed the evolutionary relationship of members of the gene family. Table 1.1 summarises details of multiple loci of some glycolytic enzymes.

It is intriguing that many of the enzymes in the glycolytic pathway show evidence of gene duplication. Exceptions are glucose phosphate isomerase, triose phosphate isomerase and phosphoglycerate kinase with only one isoform each. Glyceraldehyde 6-phosphate dehydrogenase has a single active locus, but many pseudogenes. As yet there is no evidence to suggest that the PGM gene family includes pseudogenes.

1.1.5 Protein structure

The complete amino acid sequence of rabbit muscle phosphoglucomutase was determined in 1983 by Ray et al. The authors were confident that the protein studied was one isoenzyme only and, since in rabbit muscle PGM1 is the most abundant isoenzyme, the purified protein analysed was thought to be PGM1. The purified protein was treated with cyanogen bromide which cleaved it into eleven peptide fragments. Each of the cyanogen bromide fragments were treated with hydroxylamine, iodosobenzoic acid, mild acid, cyanogen bromide in formic and heptafluorobutyric acids, Staphylococcus aureus V8 protease and trypsin; the products produced from these treatments were used to assemble the sequence.
<table>
<thead>
<tr>
<th>Enzyme and Locus</th>
<th>Subunit number</th>
<th>Size</th>
<th>Chromosomal location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hexokinase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HK1</td>
<td>1</td>
<td>107</td>
<td>10q22</td>
</tr>
<tr>
<td>HK2</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>HK3</td>
<td>1</td>
<td>116</td>
<td>*</td>
</tr>
<tr>
<td>HK4</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><strong>Phosphofructokinase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PFKL Liver</td>
<td>4</td>
<td>340</td>
<td>21q22.3</td>
</tr>
<tr>
<td>PFKM Muscle</td>
<td>4</td>
<td>340</td>
<td>1cen-q32</td>
</tr>
<tr>
<td>PFKP Platelet</td>
<td>4</td>
<td>340</td>
<td>10p15</td>
</tr>
<tr>
<td><strong>Aldolase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALDOA Muscle, red blood cells</td>
<td>4</td>
<td>40</td>
<td>16q22.1</td>
</tr>
<tr>
<td>ALDOB Liver, kidney, intestine</td>
<td>4</td>
<td>40</td>
<td>9q21.3-q22.2</td>
</tr>
<tr>
<td>ALDOC Nervous tissue</td>
<td>4</td>
<td>40</td>
<td>17</td>
</tr>
<tr>
<td><strong>Phosphoglyceromutase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGAM1 brain</td>
<td>2</td>
<td>30</td>
<td>10q24.3</td>
</tr>
<tr>
<td>PGAM2 muscle</td>
<td>2</td>
<td>30</td>
<td>7p13-p12</td>
</tr>
<tr>
<td><strong>Enolase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENO1</td>
<td>2</td>
<td>47</td>
<td>1p36</td>
</tr>
<tr>
<td>ENO2 Neuronal</td>
<td>2</td>
<td>47</td>
<td>12p13</td>
</tr>
<tr>
<td>ENO3 Muscle</td>
<td>2</td>
<td>47</td>
<td>17pter-p12</td>
</tr>
<tr>
<td><strong>Pyruvate kinase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKLR liver, red blood cells</td>
<td>4</td>
<td>60</td>
<td>1q21</td>
</tr>
<tr>
<td>PKM2 muscle</td>
<td>4</td>
<td>60</td>
<td>15q24-q25</td>
</tr>
<tr>
<td><strong>Lactate dehydrogenase</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDHA</td>
<td>4</td>
<td>35</td>
<td>11p15.1-p14</td>
</tr>
<tr>
<td>LDHB</td>
<td>4</td>
<td>35</td>
<td>12p12.2-p12.1</td>
</tr>
<tr>
<td>LDHC</td>
<td>4</td>
<td>35</td>
<td>11p15.5-p14.3</td>
</tr>
</tbody>
</table>

Table 1.1 Multiple loci in some of the glycolytic enzymes. * indicates undetermined data.
of each fragment. The fragments themselves were ordered by identifying peptides produced after a lysine-blocked trypsin digest. The sequence obtained was 561 residues long and the protein molecular weight was calculated as 61,609, considerably larger than in previous estimates (McAlpine et al., 1970 (a)). The active site of the enzyme was known to contain a phosphoryl goup (discussed in section 1.1.1), Ray et al. defined this as the phosphoserine residue, serine 116. There was found to be no obvious sequence similarities with the phosphoglycerate mutase from yeast (Ray et al., 1972) nor with alkaline phosphatase, a phosphate-transferring enzyme, (from Escherichia coli, Bradshaw et al., 1981). PGM was distinguished by its unusually long and frequently occurring stretches of hydrophobic residues.

The three dimensional structure was described by Lin et al. (1986) and further refined by Dai et al. in 1992. The PGM monomer contains 42 strands/helices arranged in four α/β domains. The appearance of the monomer is that of a thick, 'heart-shaped' box. There is an unusually deep crevice between the two lobes which involves 58 residues. The active site of the enzyme is buried at the bottom of the cleft at approximately the position where the four domains meet. The active-site cleft is large (volume of 4000-6000Å³) relative to the size of the normal substrate, compared to fructose bisphosphatase for example (which has a substrate of the same molecular weight) where the active site cleft is only 600Å³. The metal-ion binding loop (which binds Mg²⁺) is composed of the sequence -Asp-Gly-Asp-Gly-Asp- and is found in domain II at the very bottom of the active site cleft. The two lobes of PGM are formed by domains I and IV and are connected by domains II and III, giving a more robust connection than seen in other enzymes with a two-lobed structure such as phosphoglycerate kinase (Banks et al., 1979).

Dai et al. (1992) noted that only four other four-domain globular proteins have been reported (wheat germ agglutinin, aconitase, and the related glutathione reductase/lipoamide pair) and none of these has a domain-domain organisation like PGM. Furthermore the α/β architecture of PGM does not show structural similarities seen in some other glycolytic enzymes (Fothergill-Gilmore, 1986) and is different from three enzymes involved in reactions in glycolysis with direct bearing on PGM, hexokinase, glucose-6-phosphate isomerase and
phosphorylase. As mentioned previously no amino acid sequence similarities were seen between PGM and either phosphoglycerate mutase or alkaline phosphatase and this disimilarity is seen again when comparing their structures.

1.1.6 Immunological studies

Polyclonal antibodies were raised against PGM protein (Whitehouse et al, 1989) by immunising a sheep with a commercially made preparation of purified rabbit skeletal muscle PGM. An IgG fraction of the serum was prepared and was found to recognise both purified rabbit muscle PGM and PGM from a crude extract of rabbit muscle material, on both SDS and isoelectric focusing gels. The same immunoglobulins were shown to cross-react with a human protein after SDS gel electrophoresis of human tissue homogenates. The protein detected was determined to be PGM from its molecular size. Furthermore, the antibody detected the isozyme patterns usually detected by activity staining and characteristic of the PGM1 genetic polymorphism. The antiserum was found to be a more sensitive reagent than activity staining for the detection of rabbit PGM, but was less sensitive in detecting human PGM.

Drago et al (1991) investigated the cross-reactivity of the anti-rabbit muscle PGM antibodies to the other human PGM isozymes, PGM2 and PGM3. By incubating suitable tissue extracts and haemolysates with the IgG antibody fraction and then separating the extracts by starch gel electrophoresis, it was demonstrated that PGM2 and PGM3 activities were undiminished and thus PGM2 and PGM3 are probably not immunoreactive. This implies a large structural divergence between PGM1 and both PGM2 and PGM3 which is reflected in the earlier description of their differing physical properties. As described above (section 1.1.3.4) the antibodies did react with PGM4 isozymes, and this observation could be taken to support the model of a recent gene duplication with a high sequence conservation. However it also supports the idea that PGM4 arises by post-transcriptional or post-translational modification of PGM1 gene products.
1.2 PGM the gene

1.2.1 Isolation and characterisation of PGM1 cDNA

A method for cloning genes, using antibodies as probes for polypeptide products, was developed whereby large libraries of recombinant DNA in the phage expression vector λgt11 could be rapidly screened (Young and Davis, 1983 (a); Young and Davis, 1983 (b)). Clearly this was a powerful technique for the isolation of genes whose protein products were easily purified. Therefore, the development of anti-rabbit muscle PGM antibodies soon after the development of the first cDNA expression libraries made the cloning of the PGM1 gene a real possibility.

As described in section 1.1.6, the anti-rabbit muscle PGM antibodies were found to be a highly sensitive reagent in the detection of rabbit PGM but were less sensitive in the detection of human PGM1; it was subsequently found that the antibodies were not sufficiently sensitive to detect human PGM1 expressed in phage expression libraries. The approach therefore taken was to employ the antibodies to clone the rabbit PGM cDNA from a rabbit muscle cDNA library, and then to use the rabbit PGM1 cDNA to isolate the human PGM1 cDNA from a human muscle cDNA library (Whitehouse et al., 1992). Positive clones from the human muscle cDNA library and the original rabbit cDNAs were sequenced.

Human PGM1 cDNAs detected mRNA bands of approximately 2.3 kilobases after Northern Blot analysis of human skeletal muscle RNA. The full length of the human cDNA was 2320 base pairs, and of the rabbit cDNA 2279 base pairs. A comparison of the rabbit and human PGM nucleotide sequences showed 92% homology. From the sequence data each RNA was found to have an open reading frame of 1686 base pairs. The 5' untranslated region was 62 base pairs in the human mRNA and 39 base pairs in rabbit while the 3' untranslated regions were 572 and 554 base pairs in human and rabbit clones respectively. More recently the rat PGM1 cDNA has been cloned (Rivera et al., 1993). The nucleotide sequence of the rat cDNA showed 90% identity to rabbit and human PGM1 cDNAs, and 97% and 96% amino acid sequence identity respectively. Northern analyses determined the size of the rat PGM1 mRNA to be approximately 2.2 kilobases.
When translated into amino acids, the derived rabbit sequence agreed completely with the published rabbit muscle PGM amino acid sequence (Ray et al, 1983) and the human sequence showed 97% homology with the derived rabbit sequence (Whitehouse et al, 1992). The amino terminals of the two proteins are more conserved than the carboxyl terminals, including an eighty-three amino acid region of complete conservation spanning the active site region and including the active serine (residue 116). Nine of the eighteen amino acid substitutions between human and rabbit are in the final 145 amino acid residues; the residue changes give a higher positive charge to the rabbit protein and this is reflected in its more basic isoelectric point (Whitehouse et al, 1989). The molecular weight derived from the amino acid translation of the human cDNA is 61,300 which is in close agreement with the work of Ray et al (1983) and is considerably larger than the previous estimate by gel-filtration chromatography (McAlpine et al, 1970 (a)). In accordance with the findings of Ray et al (1983) that there were no protein similarities between rabbit muscle PGM and yeast phosphoglycerate mutase, Whitehouse et al (1992) could show no homologies between human PGM1 nucleotide sequence and that of human muscle-specific phosphoglycerate mutase (Shanske et al, 1987) and diphosphoglycerate mutase (Joulin et al, 1986).

The cloning of the PGM1 gene allowed direct chromosomal localisation of the locus, by in situ hybridisation, to 1p31 (Whitehouse et al, 1992).

The human cDNA was used as a probe in a Southern Blot analysis which demonstrated that PGM1 is conserved in vertebrates including primates (e.g. orangutan), birds (e.g. chicken) and amphibia (e.g. Xenopus) (Whitehouse et al, 1992). However, the PGM2 and PGM3 loci were not detected with the human PGM1 cDNA by either Southern Blot analysis nor by in situ hybridisation to chromosomes. This observation complements that of no cross-reactivity observed between anti-PGM1 antibodies and PGM2 and PGM3 (Whitehouse et al, 1989) as discussed in section 1.1.6.

The cloning of the human PGM1 cDNA has allowed the identification of three polymorphisms at the nucleotide level. Two of the polymorphisms are both diallelic polymorphic TaqI sites found by Southern analysis of DNA from
unrelated individuals, using the PGM1 cDNA as probe (Hollyoake et al., 1992). From family studies there was found to be complete linkage between the two sites, and between the sites and the +/- PGM1 protein polymorphism. Using a fragment of the cDNA from the 3' end as probe only one polymorphic TaqI site could be detected and this was mapped to a position approximately 1.3 kilobases downstream of the most 3' exon. A third polymorphism has been described on the basis of base pair changes at three sites spanning 81 base pairs in the 3' untranslated region of the PGM1 gene (March et al., 1993). The observed gene frequencies of four alleles fitted with the Hardy-Weinburg expectation. A comparison of the PGM1.3' phenotypes and PGM1 isozyme polymorphism showed a high level of linkage disequilibrium between the +/- protein polymorphism and the PGM1.3' polymorphism. The positioning of the TaqI and 3' untranslated polymorphisms and their tight linkage to the +/- PGM1 protein polymorphism strongly indicated that the position of the +/- mutation must lie towards the 3' end of the PGM1 gene. Furthermore, as the polymorphisms did not show association with the 2/1 protein polymorphism, these results provided support for the model of two distinct sites in the gene determining the 2/1 and +/- mutation positions separated by a region where intragenic recombination occurs (March et al., 1993).

1.3 Clinical aspects of PGM

1.3.1 Deficiencies of enzymes involved in glycogen metabolism

Glycogen is present in all animal cells and especially abundant in the liver and in skeletal muscle. Regulation of glycogen metabolism involves both homeostatic and hormonal mechanisms. Alterations in the amount and structure of glycogen can be caused by defects in enzymes and transporters involved in glycogen metabolism. Such defects are clinically described as diseases of glycogen storage, or the glycogenoses.

The first cases of patients with accumulation of glycogen in liver and kidney were recognised by von Gierke in 1929 (reviewed in Hers et al., 1989). In 1952 Cori and Cori suggested that such an accumulation may be due to enzyme deficiency and demonstrated a lack of glucose-6-phosphatase in some cases of glycogen storage (reviewed in Hers et al., 1989). Similarities between the von Gierke cases and those deficient in glucose-6-phosphatase lead to the term "von
Gierke's disease" for this type of deficiency, which was also termed Type I glycogen storage disease.

The causative defects in at least ten other glycogenoses have been identified, and the numerical classification has been extended to at least number VII. The affected organs and both the structure and amount of glycogen present has been investigated and the data for types I to VII is summarised in table 1.2 (Brown and Brown, 1968; Hers et al, 1989). The ways in which glycogen is stored and utilised in liver and muscle differs and is reflected in the clinical expressions of the different defects. In liver the glycogen is stored to provide glucose to the blood for the benefit of other tissues. In diseases types I, III and VI, where the liver is affected, the tissue cannot convert glycogen to glucose which leads to hepatomegaly (increase in the size of hepatocytes) and hypoglycaemia. In muscle the glycogen is there to provide a fuel for contraction, and defects cause milder symptoms limiting strenuous exercise. These symptoms are seen in types V, VII, III and a subgroup of type VI. In types II and IV the defect is not in degradation of glycogen but in its location and structure. Type II cases are identified by the presence of many membrane-bound, glycogen filled vacuoles in all tissues. Type IV cases are distinguished by glycogen with abnormally long branches found in many tissues but especially the liver and spleen.

The glycogenoses are inherited in an autosomal recessive fashion (reviewed in Hers et al, 1989). A complex exception is that of type VI which includes at least three genetic defects (Lederer et al, 1975): autosomal phosphorylase b kinase deficiency, liver phosphorylase deficiency, and X-linked phosphorylase b kinase deficiency (Huijing and Fernandes, 1969).

1.3.2 PGM deficiency

The first reported case of PGM deficiency was of a four year old boy who presented with abnormal gait (Thomson et al, 1963). The case history of the child revealed episodes of abnormal heartbeat in early childhood, with subsequent normal development until the age of 2 1/2 years. At this age he began to walk on his toes, but had no muscular pain and did not tire sooner than other children of his age. On examination he was found to have a generally slightly weak muscle tone for his age and both gastrocnemii showed marked shortening which tautened the Achilles tendons causing the unusual gait. Staining of muscle
<table>
<thead>
<tr>
<th>Type</th>
<th>Defective enzyme</th>
<th>Organ affected</th>
<th>Clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (von Gierke's disease)</td>
<td>Glucose 6-phosphatase</td>
<td>Liver and Kidney</td>
<td>Increased glycogen of normal structure. Enlargement of liver, hypoglycaemia, ketosis, hyperuricemia, hyperlipemia</td>
</tr>
<tr>
<td>II (Pompe's disease)</td>
<td>α-1,4-Glucosidase</td>
<td>All</td>
<td>Massive increase in glycogen of normal structure. Cardiorespiratory failure causes death usually before age 2.</td>
</tr>
<tr>
<td>III (Cori's disease)</td>
<td>Amylo-1,6-glucosidase (debranching enzyme)</td>
<td>Muscle and Liver</td>
<td>Increased glycogen, with short outer branches. Features as Type I but milder.</td>
</tr>
<tr>
<td>IV (Andersen's disease)</td>
<td>α-1,4-Glucan :α-1,4-glucan 6-glycosyl transferase (branching enzyme)</td>
<td>Liver and Spleen</td>
<td>Normal glycogen with very long outer branches. Cirrhosis of the liver: liver failure causes death usually before age 2.</td>
</tr>
<tr>
<td>V (McArdle's disease)</td>
<td>Muscle Phosphorylase</td>
<td>Muscle</td>
<td>Moderately increased glycogen of normal structure. Painful muscle cramps on strenuous activity.</td>
</tr>
<tr>
<td>VI (Hers' disease)</td>
<td>Liver Phosphorylase or Phosphorylase b kinase</td>
<td>Liver</td>
<td>Increased glycogen. Features like Type I, but milder.</td>
</tr>
<tr>
<td>VII</td>
<td>Phosphofructokinase</td>
<td>Muscle</td>
<td>Increased glycogen of normal structure. Features like Type V.</td>
</tr>
</tbody>
</table>

Table 1.2 Glycogen Storage Diseases
biopsies (by the periodic acid-Schiff method) revealed abnormally large amounts
of glycogen in the sarcoplasm and under the sarcolemma in the muscle fibres of
the gastrocnemius. Determination of glycogen content of the biopsies
demonstrated that the child had skeletal muscle glycogenosis. Glycogenoses
types I, II, III, IV and VI were excluded as candidates because symptoms of
hepatomegaly, cardiac enlargement and abnormal glycogen structure were not
observed.

In active muscle, anaerobic glycolysis provides most of the energy needed
and this is reflected in a rise in lactic acid concentration; the subject did not show
the sharp increase expected, which is indicative of defective muscle glycolysis.
The nature of the defects was examined by measuring the lactic acid produced
after anaerobic glycolysis by muscle homogenates in the presence of various
substrates. The defects seemed to involve the glycolytic pathway widely, but
PGM was notably severely deficient (no direct assay for PGM was made). At the
time of investigation the case was classified as type V but with adequate muscle
phosphorylase activity. However, muscle phosphorylase deficient cases are now
classed as type V, so this early case should be excluded from this grouping.
Thomson et al (1963) give an explanation for the sufficient glycolysis observed in
this case. Although the PGM activity is still adequate for provision of enough
glucose-1-phosphate for glycogen synthesis by the UDPG system, the need for
PGM in glycogen degradation is by-passed by amylo-1,6-glucosidase. This
enzyme degrades glycogen to free glucose, which is then converted to glucose-
6-phosphate by hexokinase (in the presence of ATP) which can enter the
glycolytic pathway.

A second case of PGM deficiency was reported by Illingworth and Brown
(1964) and Brown and Brown (1968). They describe a 17 month-old boy with
marked hepatomegaly. Assays of liver homogenate showed that phosphorylase,
amylo-1,6-glucosidasae, α-1,4-glucosidase and glucose 6-phosphate were at
normal levels. Further studies showed that the majority of the endogenous
glycogen, when incubated with added phosphate and phosphorylase, was
converted into α-glucose 1- phosphate and not into glucose as seen in normal
controls and other glycogenoses. This implicated a deficiency of PGM activity,
and a direct assay confirmed that PGM activity was less than 15% of the normal
liver level. Muscle from the same individual did not have excess glycogen storage, but PGM activity levels were found to be reduced in muscle, although not to the same extent as found in liver.

A third case of PGM deficiency was described by Sugie et al (1988). A 5 month-old boy with lethargy and poor weight gain was shown to have significantly low serum and muscle carnitine levels (60% and 10% of control means respectively) suggesting systemic carnitine deficiency syndrome (SCD). A muscle biopsy showed a slight accumulation of glycogen; an in vitro study on aerobic glycolysis confirmed a block after glucose-1-phosphate and before glucose-6-phosphate and direct measurement showed PGM to be at 11% of the control mean. Sugie et al (1988) propose that the specific deficiency of PGM is the primary abnormality in this case, and that PGM deficiency can be recognised as the cause of secondary systemic carnitine deficiency syndrome. The mechanism proposed to explain the carnitine deficiency is that impairment of the glycolytic pathway might enhance use of lipids to supplement the energy supply for muscle contraction. Excess acyl CoA would be formed in muscle cells causing a biased acyl CoA/CoA-SH ratio. Carnitine modulates this ratio such that excess acyl CoA would be converted to acyl carnitine and in the process deplete the carnitine levels.

The reported PGM deficiencies are heterogeneous in their clinical symptoms. Sugie et al (1988) proposed that as the isoforms of PGM are under separate genetic control that one or the other may be missing from human tissues as a congenital defect, although no attempt was made to distinguish between isoforms in that study. As discussed earlier (section 1.1.3.1) PGM activity in most tissues comes predominately from PGM1 (85-95%) with PGM2 the next active (2-15%). Therefore, a possible explanation of the case reported by Sugie et al (1988) where muscle PGM levels were 11% of the control mean, would be that PGM1 was absent but PGM2 present in this tissue. No mode of inheritance has been proposed for PGM deficiencies: the only discussion of family history is found in the first case where the only sib was found to be normal as were the parents and no similar condition had been previously described in the family (Thomson et al, 1963).
1.4 Objectives

When this project began, the human and rabbit PGM1 cDNAs had been cloned. The primary objective was to establish the molecular basis of the PGM1 genetic polymorphism with a view to testing the hypothesis that recombination was a major factor in the generation of different alleles.

In addition, the broad expression of PGM1 predicted that the gene was likely to be 'house-keeping' in nature, and therefore another objective was to identify the promoter region. This was part of a group based effort to establish the structure of the entire PGM1 gene.

The PGM1 cDNA provided a useful tool to try to recover other members of the PGM gene family and to investigate the nature of the PGM4 milk isozymes.

In brief my major aims were:
1. To take part in the resolution of the molecular basis of the PGM1 polymorphism by sequence analysis of DNA from individuals of known PGM1 phenotype.

2. To isolate clones which encoded the PGM1 promoter region and characterise the 5' region of the PGM1 gene by restriction mapping, sequence analysis and RNAse protection assays.

3. To attempt to isolate other members of the gene family in particular PGM3 using the PGM1 cDNA as probe on a sorted chromosome 6 library.

4. To search for PGM cDNAs in mammary gland libraries in order to provide some information regarding the nature of PGM4.
Chapter Two
Materials and Methods

2.1 Materials

2.1.1 Standard reagents

Analar grade standard reagents were supplied by Sigma, Fisons and BDH/Merck, Sharpe & Dome.

2.1.2 Enzymes

Restriction enzymes were supplied by Bethesda Research Laboratories (BRL), and Boehringer Mannheim. All other enzymes were from Boehringer Mannheim except Taq polymerase from Advanced Biotechnologies, and shrimp alkaline phosphatase and Sequenase from USB/Amersham.

2.1.3 Electrophoresis reagents

Agarose was from ICN, and Nusieve low melting point agarose from FMC Bioproducts. Acrylamide was from BIO-RAD; TEMED was from BDH. Ampholines for IEF were supplied by LKB Pharmacia.

2.1.4 Miscellaneous

Nick columns and dNTPs were supplied by Pharmacia. Wizard prep. kits were from Promega. Bovine Serum Albumin and antibiotics were supplied by Sigma. RNAse Inhibitor was from Boehringer Mannheim. Nitrocellulose membrane filters were from Schleicher and Schuell. Amersham supplied $^{32}$P-$\gamma$-ATP and $^{32}$P-$\alpha$-dCTP, ICN supplied $^{32}$P-$\alpha$-dCTP and $^{32}$P-$\alpha$-UTP and New England Nuclear (NEN) supplied $^{35}$S-$\alpha$-dATP. RNAzol was supplied by Biogenesis.
2.1.5 Common buffers

10 xTBE: 890mM Tris. HCl, 890mM Boric acid, 20mM EDTA
1xTE: 10mM Tris. HCl and 1mM EDTA, pH 7.5
20xSSC: 3.0M NaCl, 0.3M Na citrate, pH to 7.0 with citric acid
PBS: 137mM NaCl, 2mM KCl, 8mM Na₂HPO₄, 1.5mM KH₂PO₄
TBS (10x): 100mM Tris. HCl pH 7.5, 1.7M NaCl.
PSM: 100mM NaCl, 8mM MgSO₄, 50mM Tris. HCl pH 7.2, 0.01% gelatin
100xDenhardt's: 2% Ficoll, 2% polyvinylpyrrolidone, 2% BSA
'Phenol' refers to phenol equilibrated with TE pH 7.5
'Chloroform' refers to a 24:1 (v/v) mixture of chloroform and isoamyl alcohol
'Phenol/chloroform' refers to a 25:24:1 (v/v) mixture of equilibrated phenol, chloroform and isoamyl alcohol
'Polyacrylamide' refers to a 29:1 mixture of acrylamide and bis acrylamide

2.1.6 Microbiology media

Tryptone, Yeast Extract and Bacto agar from Difco.
L-broth (L⁻¹): 10g Tryptone, 5g Yeast Extract, 5g NaCl, 1g glucose
L-agar (L⁻¹): as L-broth but with glucose omitted and 14g Bacto agar added
L-agarose (L⁻¹): as L-broth but with glucose omitted and 7g agarose added
LM-broth (L⁻¹): 10g Tryptone, 5g Yeast Extract, 5g NaCl, 2g MgSO₄.7H₂O
LM-agar (L⁻¹): as LM-broth with 14g Bacto agar added
LM-agarose (L⁻¹): as LM-broth with 7g agarose added
2xYT (L⁻¹): 10g Tryptone, 10g Yeast Extract, 5g NaCl
H-agar (L⁻¹): 10g Tryptone, 8g NaCl, 12g Bacto agar
H-top agar (L⁻¹): 10g Tryptone, 8g NaCl, 8g Bacto agar
Super-broth (L⁻¹): 35g Tryptone, 20g Yeast Extract, 5g NaCl, 1ml 5M NaOH
M9 minimal salts (L-1): 6g Na₂HPO₄, 3g KH₂PO₄, 1g NH₄Cl, 0.5g NaCl
Glucose/minimal medium plates: add to 1 litre of sterile M9 salts: 1ml
1M MgSO₄, 1ml 0.1M CaCl₂, 1ml 1M Thiamine HCL, 10ml 20% (w/v)
glucose

2.1.7 Bacterial strains

JM101 (M13 host): supE thi (lac-proAB). F'[traD36 proAB⁺ lacF⁺ lacZM15], r+m⁺
RR1 (pUC host): supE44 hsdS20(rB⁺mB⁻) ara-14 proA2 lacY1 galK2
rpsh20 xyl-5 mtl-1
XL1Blue (host strain for pBluescript and Uni-ZAP): supE44 hsdR17 recA1
endA1 gyrA46 thi relA1 lac⁺ F'[proAB⁺ lacF⁺ lacZΔM15 Tn10(ter)][(hsdR)]
Y1090 (host strain for λgt11) SupF hsdR araD139 Δlon Δlac U169 rpsL
trpC22 :: Tn10(ter) pMC9
LE392 (host strain for λ2001) supE44 supF58 hsdR514 galK2 galT22
metB1 trpR55 lacY1

2.2 Methods

All solutions for nucleic acid work were sterilised by autoclaving (15psi,
121°C for 20 min) or by filter sterilisation (through 0.22µm pore size 'Acrodisc',
Gelman Sciences). Glassware and disposable plasticware was sterilised by
autoclaving (15psi, 121°C for 20 min). Disposable gloves were worn for all
experimentation. All autoradiography involving ³²P and ³⁵S was carried out by
exposure to X-ray film (Kodak), at room temperature for ³⁵S and with an
intensifying screen at -70°C for ³²P.

2.2.1 DNA precipitation and assay of concentration

DNA precipitations were performed by addition of 3M sodium acetate pH
5.5 to a final concentration of 0.3M and adding 2 volumes of 100% ethanol before
chilling at -20°C overnight or -70°C for 1 hour. Precipitated DNA was recovered
by centrifugation in a bench centrifuge at 13,000g and 4°C and washed in 70%
(v/v) ethanol before air-drying and resuspending in ddw or TE.
DNA was quantified by measurement of absorbance at 260nm; 1 OD unit=50µg·ml⁻¹ (the ratio of OD₂₆₀/OD₂₈₀ should be 1.8 for pure DNA).

2.2.2 DNA modification

2.2.2.1 Restriction enzyme digests. Digests were performed using the incubation buffers provided with the enzymes and according to the conditions recommended by the manufacturer.

2.2.2.2 Dephosphorylation. The terminal phosphate of linearised vector DNA was removed to prevent self-ligation in subcloning reactions. Typically 1µg of vector was linearised in a standard restriction enzyme digest and phosphatased with either calf intestinal alkaline phosphatase (CIP) (Boehringer Mannheim) or shrimp alkaline phosphatase (SAP) (USB/Amersham).

When using CIP the volume was made up to 43µl with ddw and incubated with 5µl CIP buffer and 1 unit of CIP for 15 min at 37°C, then for 15 min at 56°C. A further 1 unit of CIP was added and the two incubation stages repeated. The reaction was stopped by adding 40µl ddw, 10µl TE and SDS to 0.5% followed by incubation at 68°C for 15 min. The CIP was removed by a phenol extraction followed by a chloroform extraction and DNA recovered by ethanol precipitation and resuspended in TE at 50ng·µl⁻¹.

When using SAP the vector restriction enzyme digest was first performed in a 20µl total volume which was then directly incubated with 0.1µl SAP (at 5 units. µl⁻¹) for 15 min at 37°C. A further 0.1µl SAP was added and incubation for 30 min at 37°C was followed by heat inactivation for 5 min at 85°C.

2.2.2.3 Ligation. Insert DNA was ligated to appropriately cut and phosphatased vector in a 10µl reaction mix, containing 20-50ng vector and generally a 3-fold molar excess of insert DNA to vector DNA in 1xligase buffer (10x buffer is 660mM Tris-HCl, 50mM MgCl₂, 10mM DTT, 10mM ATP, pH 7.5) with 1µl (1unit. µl⁻¹) T4 DNA ligase enzyme. Sticky-ended ligations were incubated at 15°C overnight.

2.2.3 Agarose gel electrophoresis and recovery of DNA fragments
DNA was resolved in 0.8-2.0% agarose gels made up and run in 1xTBE buffer, and prepared in 20x20 cm, 14x11 cm or 8x10 cm flat-bed moulds. Gels were usually run at 2V. cm⁻¹ overnight or at up to 10V. cm⁻¹ for shorter periods. Gels prepared with Nusieve low melting point agarose were run in a 4°C temperature controlled room. DNA fragments were visualised by ethidium bromide staining (at a concentration of 100ng.ml⁻¹ in both gel and running buffer) and ultra-violet transillumination.

DNA fragments required for subcloning, sequencing or ³²P-labelling reactions were isolated from agarose gels by initially cutting the band from the gel under long-wave UV. Bands cut from standard agarose gels were placed inside 2 cm width boiled dialysis tubing with 300µl TE and electroeluted by placing the sealed tube in a tank of 1xTBE at 100V for 1-4 hr. The recovered solution was then ethanol precipitated. Bands cut from low melting point agarose gels were weighed and cleaned either by processing with GeneClean™ kit (BIO 101) or with Wizard™ PCR Preps DNA Purification columns (Promega) both according to the manufacturers instructions.

2.2.4 Preparation and transformation of competent E. Coli

Competent E.Coli JM101 and RR1 cells were prepared using the CaCl₂ procedure as described by Maniatis et al (1982).

JM101 were grown on a minimal medium plate to preserve the F' episome containing the gene for proline synthesis (including the lac gene used for blue/white screening of recombinants). A single JM101 colony was picked from an M9 plate into 10 ml of 2xYT and grown overnight at 37°C with shaking. 40 ml of 2xYT was inoculated with 1 ml of the overnight culture and grown at 37°C with shaking until the culture reached an OD₆₀₀ of 0.4-0.6. The culture was centrifuged at 2500 rpm at 4°C in a MSE Mistral 2000 centrifuge. The cell pellet was re-suspended in 20 ml of ice-cold, filter-sterilised 100mM CaCl₂, 10mM Tris pH 7.5, and kept on ice for 15 min. Cells were pelleted as before and resuspended in 4 ml of ice-cold 100mM CaCl₂, 10mM Tris pH 7.5 and stored on ice until transformed. To transform competent JM101s, 300µl of cells were mixed gently with 40-60ng of DNA and kept on ice for 30 min before a 90 sec heat shock at 42°C. Each sample was plated onto a 90 mm H-agar plate in 3 ml H-top agar with 25µl each of 2% IPTG and 2% X-Gal (in dimethyl formamide) and incubated at 37°C overnight. Under these conditions recombinant and non-
recombinant phage can be distinguished due to insertional inactivation of the B-galactosidase gene of M13. A plate of competent cells with no DNA was set up as a control for the cells and a phosphatased vector-only ligation was transformed to test the efficiency of the dephosphorylation process.

A 10ml L-broth culture was inoculated with a single RR1 colony and grown overnight with shaking at 37°C. 100ml of L-broth was inoculated with 1ml of the overnight culture and grown with shaking at 37°C until the OD550 was 0.28. The culture was kept on ice for 30 min then spun at 2500 rpm in a MSE Mistral 2000 centrifuge for 5 min. The pellet was resuspended in 40ml ice-cold, filter-sterilised 100mM CaCl2, 10mM Tris pH 7.5 and kept on ice for 20 min. Cells were re-spun as before and resuspended in 1ml ice-cold, filter-sterilised 100mM CaCl2, 10mM Tris pH 7.5. For each transformation, 100µl of competent cells were mixed with 40-60ng DNA, and after 20 min on ice were heat-shocked at 42°C for 2 min. 200µl of L-agar, warmed to 37°C, was added to each sample and incubated at 37°C for 15 min. The mixture was spread onto the surface of a 90 mm L-agar plate (with ampicillin at 100µg. ml⁻¹) with a glass loop. Plates were incubated overnight at 37°C.

2.2.5 Large scale plasmid preparations

Bacterial plasmids were isolated from large-scale, liquid cultures by the alkaline lysis method of Birnboim and Doly (1979) as modified by Ish-Horowitz and Burke (1981) and carried out exactly as described in Maniatis et al (1982). Supercoiled plasmid DNA was purified either on a CsCl-ethidium bromide gradient (based on the technique of Clewell and Helinski (1972)) or with a silica-based resin in the Wizard Maxipreps DNA Purification System (Promega). Following centrifugation of the DNA in the CsCl gradient (Sorvall Combi Ultracentrifuge), the supercoil band, visualised under UV light, was removed through the side of the centrifuge tube with a hypodermic syringe. Ethidium bromide was removed by repeated extractions with isopropanol and the DNA removed from solution by ethanol precipitations. The Wizard™ Maxi-Purification system was used as recommended by the manufacturer: the plasmid DNA is recovered in a final volume of 1.5ml of TE.
2.2.6 Mini-plasmid preparations

Glycerol stocks were prepared from 3ml overnight cultures grown with the appropriate antibiotic and broth (800μl of culture and 400μl 50% glycerol for each glycerol stock). This method was used also for mini M13 double stranded preparations where cultures are grown for 4.5 hr in 2xYT and no antibiotic. 1.5ml of the remaining culture was spun for 5 min at 13 000g in a microfuge. All supernatant was removed and the pellet suspended in 100μl of solution 1 (50mM glucose, 10mM EDTA, 25mM Tris/HCl pH 7.5). 200μl of solution 2 (1%SDS, 0.2MNaOH) was added, mixed gently and kept on ice for 5 min. 150μl of 3M sodium acetate pH 5.5 was added, mixed vigorously and kept on ice for 5 min. After spinning for 5 min at 13 000g in a microfuge, the supernatant was transferred to a new tube, extracted once with phenol, once with chloroform and precipitated after the addition of 600μl ethanol. Recovered DNA pellets were resuspended in 30μl TE with RNAse at 0.5μg. μl⁻¹.

2.2.7 Large scale λ preparations

A high titre stock was prepared from λ bacteriophage isolates by adsorbing a sufficient number of plaque forming units (pfu) to host bacteria to produce confluent lysis when plated out on 90mm L-agar plates. 2ml of PSM was added to the surface of the plate which was gently agitated overnight at 4°C. The PSM was removed from the plate, spun at 13000rpm for 10 min in a microfuge, and the supernatant removed and stored at 4°C.

For large scale preparations of λ2001 bacteriophage DNA, phage stocks were titred and 5x10⁸ pfu of each isolate were allowed to adsorb to 1x10¹⁰ E.Coli LE392 for 20 min at 37°C without shaking. The mix was added to 200ml of L-broth, 10mM MgCl₂ and incubated with shaking at 37°C for approximately 5 hr, until cell lysis was obvious with the appearance of cell debris and protein froth. Lysis was completed by addition of 2ml chloroform and shaking for 20 min at 37°C. 8g NaCl was added followed by DNAase to 1μg. ml⁻¹ and RNAase to 1μg. ml⁻¹ and the cultures were stood at room temperature for 1 hr before centrifugation at 10000 rpm (MSE Europa 24M centrifuge), for 10 min at 4°C. Phage were precipitated from the supernatant with 20g PEG 6000 (10%w/v) at 4°C overnight, collected by centrifugation at 10000 rpm, for 10 min at 4°C and resuspended in 10ml PSM. 10ml of chloroform was added to the phage suspension and the mixture stood on ice for 30-60min. CsCl (0.75g. ml⁻¹) was
Chapter 2

added to the aqueous layer which was transferred to a tube and spun in a vertical rotor at 35000 rpm (Sorvall Combi Ultracentrifuge) at 10°C for 24hr. The phage band was collected, by puncturing the centrifuge tube with a hyperdermic syringe, and dialysed against sterile 10mM NaCl, 50mM Tris pH 8.0, 10mM MgCl2 (2x 1L). The phage protein coat was removed by digestion with 50µg. ml⁻¹ Proteinase K, 20mM EDTA, 0.5% SDS and incubation for 1hr at 65°C. The solution was extracted twice with phenol, twice with chloroform and the DNA ethanol precipitated.

2.2.8 In vivo excision of pBluescript SK(-) from Uni-ZAP XR

This procedure allows the excision and recircularisation of any cloned insert contained within the lambda vector Uni-ZAP XR to form a phagemid containing the cloned insert. A 10ml L-broth culture of E.Coli XL1Blue with 100µl of a 1.5mg. ml⁻¹ tetracycline solution (in 50% (v/v) ethanol) was grown overnight at 37°C with shaking. 50ml of L-broth was inoculated with 0.5ml of the overnight culture and grown for approximately 7hr, until the OD₆₀₀ was 1. 200µl of this culture was mixed with 1µl MK1307 helper phage (8.63x10¹¹ pfu. ml⁻¹; Pharmacia) and 1x10⁵ pfu of Uni-ZAP XR phage stock. and incubated for 15 min at 37°C before the addition of 5ml 2xYT broth and incubation at 37°C overnight with shaking. The mix was heated at 70°C for 20 min. After centrifugation at 4000rpm (Eppendorf microfuge 5402) for 5 min, the supernatant, containing the rescued phage particles, was collected. 200µl of this phage stock was mixed with 200µl of XL1Blue culture (OD₆₀₀=1) and incubated for 15 min at 37°C. 5µl of this mix was mixed with 200µl of 2xYT and spread onto an L-agar plate with ampicillin at 100µg. ml⁻¹, and grown overnight at 37°C. Colonies were picked individually into 3ml L-broth and 100µg. ml⁻¹ ampicillin and grown overnight at 37°C with shaking. Mini-plasmid preparations were completed as described in 2.2.6.

2.2.9 Single-stranded DNA preparations

2.2.9.1 Preparations from M13

During the single stranded lytic phase of M13 bacteriophage, single-stranded DNA is extruded through the host cell wall and can be isolated from infected E. coli culture supernatant.
A culture of JM101 *E. coli* were grown overnight in 10ml of 2xYT. This culture was diluted 1:100 in 2ml 2xYT with either a single white plaque or 20μl of M13 phage stock and grown for 5 hr at 37°C with shaking. 1.2ml of culture, in an Eppendorf, was centrifuged for 5 min and 800μl of the cleared supernatant removed to a fresh tube. The original tube was topped up with the remaining culture, respun as described and 1ml of supernatant reserved as a phage stock. To the first 800μl of supernatant, 200μl of 20% PEG (Polyethyleneglycol 6000), 2.5M NaCl was added with incubation at room temperature for 15 min to precipitate the phage. After centrifugation (5 min, in a microfuge) the pellet was resuspended in 100μl of TE; the protein coats of the phage particles were removed by phenol extraction and the single-stranded DNA was recovered by ethanol precipitation, with DNA resuspended in 15μl of TE. This preparation was sufficient for two sequencing reactions (using 7μl per reaction).

### 2.2.9.2 Preparations from pBluescript

This method describes the preparation of single-stranded DNA from pBluescript phagemid.

A culture of pBluescript plasmid is grown by inoculating 5ml of L-broth with the appropriate glycerol stock and incubating overnight at 37°C. 3ml of Superbroth is inoculated with sufficient overnight culture to give OD<sub>600</sub> of 0.1 (this is approximately a 1:10 dilution) and grown at 37°C until OD<sub>600</sub>=0.3 which is approximately equivalent to 2.5×10<sup>8</sup> bacteria ml<sup>-1</sup>. Helper phage (R408, Promega) were added to a multiplicity of infection of 20 phage: 1 cell and the culture was incubated with shaking at 37°C for a further 8 hr. The culture was then heated at 65°C for 15 min, then transferred to Eppendorfs and spun for 2 min (13000g): the supernatant could then be stored at 4°C. 300μl of 20% PEG (Polyethylene glycol 6000), 2.5M NaCl was mixed with 1.2ml of supernatant and stood at room temperature for 15 min. This was then spun for 5 min in a microfuge and the supernatant discarded. The pellet was resuspended in 300μl TE and extracted once with phenol, once with phenol/chloroform and once with chloroform. The single-stranded DNA was recovered by ethanol precipitation and resuspended in 20μl TE; 7μl of single-stranded DNA was used in each sequencing reaction.
2.2.10 Southern blotting

After electrophoresis in agarose gels, DNA was denatured by submerging the gel in 1.5M NaCl, 0.5M NaOH for 30 min and neutralised by submerging in 1.5M NaCl, 0.5M Tris-HCl pH 7.5 for 60 min. The DNA was blotted (Southern, 1975) onto a nylon filter (Hybond-N, Amersham) in 20xSSC for 16-24 hr. Filters were then baked at 80°C for 2 hr to bind the DNA to the filter.

Hybridisations of Southern blots carried out in sealed plastic bags in a 65°C waterbath. Prehybridisation was carried out at 65°C for 4 hr in the following solution:

<table>
<thead>
<tr>
<th>Component</th>
<th>Final concentration</th>
<th>Volumes for 100ml solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denhardtts (100x)</td>
<td>10x</td>
<td>10ml</td>
</tr>
<tr>
<td>SSC (20x)</td>
<td>4x</td>
<td>20ml</td>
</tr>
<tr>
<td>SDS (20%)</td>
<td>0.1%</td>
<td>0.5ml</td>
</tr>
<tr>
<td>Saturated PPi (1M)</td>
<td>10mM</td>
<td>1.0ml</td>
</tr>
<tr>
<td>Denatured Herring</td>
<td>0.1mg. ml⁻¹</td>
<td>2ml</td>
</tr>
<tr>
<td>Sperm DNA (5mg. ml⁻¹)</td>
<td>-</td>
<td>66.5ml</td>
</tr>
<tr>
<td>Water</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

The denatured probe was then added and the filter hybridised overnight at either 65°C or 55°C depending on the expected level of homology between probe and blotted DNA.

After hybridisation, non-specifically bound probe was removed from the filters by washing firstly in a solution of 2xSSC, 0.1%SDS at room temperature for 30 min. Further washes at higher stringencies were then carried out: generally 0.2xSSC, 0.1%SDS at 65°C 30 min followed by 0.1xSSC, 0.1%SDS at 65°C for 10 min, unless otherwise stated. Filters were blotted of excess liquid, wrapped in clingfilm and subjected to autoradiography with Kodak X-ray film and with intensifying screens at -70°C.

2.2.11 Preparation of ³²P labelled probes

Probes were labelled by random priming (Feinberg and Vogelstein, 1984), which uses the large fragment of E. coli DNA polymerase I (Klenow fragment) to
Chapter 2

initiate second strand synthesis from random hexamer oligonucleotide-primed single-stranded DNA.

Radiolabelling of DNA probes was carried out using either an Amersham random priming oligolabelling kit (with $^{32}$P-α-dCTP supplied by Amersham or ICN) or an Amersham Rediprime kit (which utilises Amersham Redivue $^{32}$P-α-dCTP). Both kits were used exactly according to the manufacturers instructions, generally labelling between 25 and 100ng of DNA per probe. To separate labelled probe from unincorporated radionucleotides the labelling reaction was put through a Sephadex G-50 Nick Column (Pharmacia) as described by the manufacturer. Percentage incorporation was estimated by comparing the cpm given by the column with those given by the DNA fraction: incorporation was generally 60-80%. The probe was made single-stranded by boiling for 5 min, before adding it to the filter in hybridisation solution.

In order to re-use filters, probes were removed by boiling the filters for 15 min in 0.1xSSC, 1%SDS. The filters were checked for complete removal of the probe by autoradiography.

2.2.12 cDNA library screening

cDNA libraries constructed in λgt11 and Uni-ZAP XR were screened with antibody and cDNA probes. The libraries were titrated to determine the number of pfu per ml of stock by spotting serially diluted phage stock onto host cells in top-agar; incubation at 42°C for 5 hr facilitated lysis and the number of plaques were counted. The host cells were *E. coli* Y1090 for λgt11 which were plated on L-agar plates and *E. coli* XL1Blue for Uni-ZAP XR which were plated on LM-agar. For each screening, an overnight culture of the host cells was grown, in either L- or LM-broth, with 0.2% maltose. For each primary screen, 2ml of the overnight culture was centrifuged at 6000 rpm (MSE Mistral 2000) for 5 min, the supernatant drained off and $10^5$ pfu phage added, gently mixed and incubated at 37°C for 20 min. 8ml of top-agarose was mixed with the cells and phage and poured on top of the 182 mm plate. Plates were incubated at 42°C overnight or until plaques were visible (approximately 5 hr).
2.2.12.1 Screening with antibody probes. After the plaques had formed (at 42°C) the plates were removed from incubation and nitrocellulose membrane filters, pre-soaked in 10mM IPTG and air-dried, were lowered onto the plates. Registration marks were made through the filter and agar with a sterile needle to allow positive signals to be aligned later. The plate with filter was then incubated at 37°C to induce the expression of proteins. The filters were removed and washed in 1xTBS for 5 min, then in 1xTBS, 3% gelatin for 10 min. The antibody probe at 25μg. ml⁻¹ was added to the filter in 15ml 1xTBS, 3% gelatin, 0.2% sodium azide and left at room temperature overnight. After removing the first antibody, the filter was washed in 1xTBS for 5 min, 1xTBS, 0.05% NP40 for 4 min, and 1xTBS for 10 min. The filter was incubated in 5μl of peroxidase conjugated second antibody (generally rabbit anti-goat) in 12ml 1xTBS, 3% gelatin for 1.5hr and then washed as described above for removal of the first antibody. Positive recombinant plaques were visualised by staining the filter with 15 ml of the following: 60mg 5-chloro 1 napthol in 20ml ice-cold methanol mixed with 100ml 1xTBS containing 60μl of 30% hydrogen peroxide. Small purple spots indicated the position of positively staining plaques, which were taken as agar plugs into 1ml PSM with 1 drop of chloroform added. This stock was kept at 4°C overnight to allow the phage to diffuse out of the agar before plating out a 100 fold dilution for re-screening.

2.2.12.2 Screening with cDNA probes. Nitrocellulose membrane filters (2 per plate) were moistened in 2xSSC then blotted on Whatman™ paper. The first filter was placed onto the plate for 1.5 min and registration marks made with a needle: the second filter was placed onto the plate for 3 min. Both filters were then submerged in 1.5M NaCl, 0.5M NaOH for 60sec and then in 1.5M NaCl, 0.5M Tris-HCl pH 7.5 for 5 min before a brief rinse in 5xSSC. The filters were air-dried and baked for 2 hr at 80°C to fix the DNA onto the membrane. A cDNA probe was labelled with ³²P-α-dCTP, as described in 2.2.11, and hybridised to the filters in hybridisation solution (described in 2.2.10) at 65°C overnight. Filters were washed to 0.2xSSC, 0.1%SDS for 20 min at 55°C and exposed to X-ray film, with intensifying screens, at -70°C. Positive recombinants were identified by duplicated signals on autoradiographs from both duplicate filters and picked after alignment of the autoradiographs to the plates. Successive rounds of screening were completed as described in 2.2.12.1.
2.2.13 Screening chromosome 6 library

A cosmid library constructed from flow sorted human chromosome 6 material was screened with a cDNA probe. The library (constructed at Imperial Cancer Research Fund (ICRF)) is in the form of high density gridded filters, with 10000 clones per filter as described in Nizetic et al, 1994. For each screening, two filters (exact replicas) are screened together and positive clones are those with signal on both filters.

Filters were prehybridised and hybridised in the following solution (modified Church buffer):

<table>
<thead>
<tr>
<th>Component</th>
<th>Final concentration</th>
<th>Volumes for 100ml of solution.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaPi (1M), pH 7.2</td>
<td>0.5M</td>
<td>50ml</td>
</tr>
<tr>
<td>SDS (20%)</td>
<td>7%</td>
<td>35ml</td>
</tr>
<tr>
<td>EDTA (0.5M)</td>
<td>1mM</td>
<td>0.2ml</td>
</tr>
<tr>
<td>BSA</td>
<td>0.1%</td>
<td>100mg</td>
</tr>
<tr>
<td>Denatured Herring</td>
<td>50μg. ml⁻¹</td>
<td>1ml</td>
</tr>
<tr>
<td>Sperm DNA (5mg.ml⁻¹)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td>13.8ml</td>
</tr>
</tbody>
</table>

The filters were prehybridised in a sealed plastic bag at 65°C for 7hr. The cDNA probe was labelled with ³²P-α-dCTP as described in 2.2.11. Additionally, library vector DNA (Lawrist 4) was labelled with ³⁵S-α-dATP. 1μl vector DNA (at 20ng. μl⁻¹) was heated at 100°C for 5 min with 20μl ddw and 5μl primer (Amersham Megaprime Kit). 4μl of each dGTP, dTTP and dCTP and 5μl reaction buffer (also from Amersham Megaprime Kit) were added at room temperature with 5μl ³⁵S (at 1x10⁴ μCi. ml⁻¹) and 2μl Klenow enzyme. The reaction was incubated for 1hr at 37°C. The reaction was stopped by the addition of 5μl 0.2M EDTA. Unincorporated ³⁵S was removed using a Nick Column (section 2.2.11). 50ml of hybridisation solution was removed from the bagged filters and the probes were added after boiling both for 5 min. The filters were hybridised at 55°C overnight.

Non-specifically bound probe was removed from the filters by 4 washes, each for 5 min at 55°C, with 40mM NaPi pH 7.2, 1%SDS. The filters were then
arranged in a cassette with a piece of previously exposed and developed X-ray film between the filter and new X-ray film to reduce the intensity of the $^{35}$S signal. Autoradiography was overnight at -70°C. The $^{35}$S-$\alpha$-dATP labelled vector DNA allowed the positioning of all of the clones on the filter by small relatively weak signals. The clones which hybridised to the cDNA were identified by stronger signals duplicated on both filters.

The positions of positive clones were given reference numbers by aligning the autoradiographs to a numbered grid. Corresponding cosmids were collected from ICRF in the form of stabs in Eppendorfs. These were incubated at 37°C overnight. A stab from each cosmid tube was spread onto a L-agar plate (with Kanamycin at 0.1mg. ml$^{-1}$) and grown overnight at 37°C. A single colony was picked from these plates for each cosmid and grown in 3ml L-broth (with Kanamycin at 0.1mg. ml$^{-1}$) at 37°C overnight. Glycerol stocks and mini-plasmid preparations were prepared as described in 2.2.6.

2.2.14 Polymerase Chain Reaction (PCR)

Amplification reactions of genomic DNA and cDNA utilised Advanced Biotechnologies Taq enzyme with BufferV (10x BufferV is 500mM KCl, 100mM Tris-HCl pH 8.8, 15mM MgCl$_2$, 1% Triton X-100). Amplifications of 100ng-1μg of template DNA were generally carried out in 100μl reaction volumes with 1x BufferV, 1μl of dNTPs (a mix with each at 20mM) and 50pM of both forward and reverse primers. 5% formamide was added to genomic DNA amplifications to effectively raise the annealing temperature in order to overcome melting problems associated with high GC content of templates. This mix was heated at 95°C for 5min before the addition of 0.5μl Taq enzyme (at 5u. μl$^{-1}$). Reactions were topped with approximately 100μl of paraffin oil to prevent evaporation during amplification. 30-35 cycles of amplification were performed on a Hybaid thermocycler machine using the following conditions:

<table>
<thead>
<tr>
<th>Step</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturation step</td>
<td>94°C 15 sec (cDNA template), 60 sec (genomic DNA template)</td>
</tr>
<tr>
<td>Annealing step</td>
<td>Tm-5°C 30 sec</td>
</tr>
<tr>
<td>Extension step</td>
<td>72°C 30 sec</td>
</tr>
</tbody>
</table>
The annealing temperature (Tm, also described as the melting
temperature) was calculated from the sequence of the primer using the following
formula:

\[ Tm = 69.3 + 0.41(G+C\% \text{ of primer}) - \left(\frac{650}{\text{length of primer}}\right) \]

The Tm used for each amplification is the average of Tm for the forward
and reverse primers minus 5°C.

\( \lambda \text{gt11} \) and Uni-ZAP XR bacteriophage stocks were amplified directly using
commercially available primers which flank the cloning sites. The sequences of
the Uni-ZAP XR primers are (forward) 5' GTAAAATGACGGCCAGTGC 3' and
(reverse) 5' GAAACAGCTATGACCAGATTAC 3'; the sequences of the \( \lambda \text{gt11} \)
primers are (forward) 5' GGTGGCGACGACTCCTGGAGCCCG 3' and (reverse)
5' TTGACACCAGACCAATCTGGTAATG 3'. Phage stocks of raised titre
(approximately \( 10^{10} \text{pfu. ml}^{-1} \)) were produced by the procedure described in
2.2.7. 2\( \mu \)l of phage stock were diluted with 79\( \mu \)l of water and frozen at -20°C for
30 min before boiling for 5 min. The DNA released from the phage particles was
then amplified with the same mix described above with the following temperature
cycles:

- Denaturation step 94°C 15 sec
- Annealing step 52°C 30 sec
- Extension step 72°C 3 min (extended to allow
  extension of longer inserts)

2.2.15 'AT' direct cloning of unmodified PCR products

This procedure exploits the finding that Taq polymerase adds a single
adenosine nucleotide at the 3' end of amplification products. This allows the
subcloning of PCR products into blunt-cut vectors modified with a single
thymidine overhang and is based on a method described by Marchuk et al, 1991
which is generally more efficient than cloning products into blunt-cut vectors.

1\( \mu \)g of M13 mp18 vector (in a 20\( \mu \)l volume) was digested with Smal
restriction enzyme which linearised the vector to give blunt-ends. This was then
incubated with 0.5μl Taq enzyme in the presence of 2mM dTTP for 2 hr at 70°C. After phenol extraction and ethanol precipitation, the M13 vector was resuspended in 20μl of ddw. PCR products for subcloning were gel purified as described in 2.2.3 and ligated to the M13 as described in 2.2.2.3. The use of M13 vector allowed the blue/ white colour selection for clones containing insert. Vector self-ligation events were avoided because of the 3' thymidine overhang. Single-stranded DNA was generated from recombinants (2.2.9.1) and sequenced with M13 sequencing primers (2.2.16).

2.2.16 DNA sequencing

Sequencing using the dideoxy NTP termination method developed by Sanger, Nicklen and Coulson (1977) was carried out using the USB Sequenase™ Version 2 Kit. DNA fragments were synthesised using T7 DNA Polymerase, a single stranded template (the DNA to be sequenced) and a short primer to initiate the reaction. The reaction terminated when an 35S labelled dideoxy base was incorporated and the ratios of dideoxy nucleotides to deoxy nucleotides were balanced so that there was a termination representing each base in the sequence. The fragments were separated by size on a denaturing polyacrylamide gel (8% polyacrylamide/6M urea), polymerised with 1.7% TEMED and 0.34% ammonium persulphate and set in a Biorad sequencing apparatus.

2.2.16.1 Sequencing of single stranded DNA. Single-stranded templates were prepared from M13 and pBluescript as described in 2.2.9. The reactions were carried out following the USB kit instructions. Each reaction had 1pM of primer (in 1μl), 2μl of 5x reaction buffer and 7μl of single-stranded DNA template. The primers were annealed to the DNA at 65°C for 3 min and then cooled slowly to 30°C. Labelling mix was prepared: volumes for 4 reactions were 4.5μl 0.1M DTT, 9μl of 'G' mix (diluted 2:8 with ddw) and 2.5μl 35S-α-dATP. 5.5μl of the labelling mix was added to each tube of annealed template and primer and incubated at room temperature for 2 min. 3.5μl from each reaction was dispensed into each of the four termination mixes (2.5μl of A, T, G and C termination mixes, pre-warmed at 37°C) and incubated at 37°C for 5 min. The reactions were stopped with 4μl of formamide/ bromophenol blue/ xylene cyanol loading dye. Samples were heated at 80°C for 5 min just prior to loading on the gel pre-warmed to 50°C. After electrophoresis in 1xTBE buffer at 50W, the gel was
removed from the apparatus onto Whatman paper and dried in an 80°C oven for 30 min. Autoradiography was to X-ray film at room temperature for 24-48 hr.

Modifications were made to the method to extend the sequence: replacing 1.5μl of each termination mix with Extension mix (from Sequenase Kit) gave a relative extension of 2.5 fold of that obtained with the normal mixes. This was used when sequence beyond 400 bp from the primer was needed.

Another modification to the method was the substitution of inosine (dITP) in place of guanosine (dGTP) and ddITP termination mixes instead of ddGTP to remove compressions in the sequence caused by 3 or more G/C pairs in succession (dITP and ddITP were used from the Sequenase Kit). The inosine-cytosine base pair is a weaker interaction than the guanosine- cytosine base pair and the DNA is therefore more easily denatured. When DNA is not fully denatured electrophoretic mobility is greater and a band appears at an incorrect position, resulting in more than one band in one position.

2.2.16.2 Sequencing of double-stranded templates. Double-stranded DNA was also used as a template for sequencing. Gel purified PCR products (procedure detailed in 2.2.3) were rendered single-stranded for sequencing by the addition of DMSO to the annealing reaction and the termination mixes. 1μl DMSO was added to 7μl of double-stranded DNA, 1pM of sequencing primer and 2μl of 5x reaction buffer. This was boiled for 5 min, then immediately plunged onto ice for 5 min. The sequencing reactions were then carried out as described above for single-stranded templates with the additional modification of 10% DMSO in the termination mixes.

A second method for sequencing PCR products involved the use of streptavidin coated beads (Dynabeads). PCRs were carried out with one primer biotin labelled at the 5' end. The beads were used as directed by the manufacturer: beads are mixed with the PCR product and treated with 0.15M NaOH, and then collected using a magnetic particle concentrator. The supernatant (containing the non-biotin labelled strand) is removed and neutralised in 3M sodium acetate pH 6 (to a final concentration of 0.3M), ethanol precipitated and resuspended in 7μl TE. The beads are washed with TE and water and resuspended in 7μl TE. The 2 strands can now be sequenced as
described above with care that the strand on the beads remains well mixed throughout the procedure. The beads are removed after incubation at 65°C for 5 min after the sequence reaction has been stopped.

2.2.17 Isolation of RNA

RNA was extracted from small pieces of human and adult tissue using RNAzol™ B (Biogenesis), a solution containing guanidium thiocyanate and phenol similar to that described by Chomczynski and Sacchi (1987). The solution was used as described by the manufacturer. The procedure involves homogenisation of the tissue in RNAzol™ B (2ml for 100mg of tissue), extraction with chloroform (1 volume homogenate to 0.1 volume chloroform) and precipitation of the aqueous phase with isopropanol. All equipment for this procedure was rinsed in DEPC water and sterilised by autoclaving where possible. Typical concentrations of extracted RNA from muscle tissue were 4-7mg. ml⁻¹.

mRNA was isolated from total RNA using Micro-FastTrack mRNA isolation kit (from Invitrogen). This system uses oligo (dT) cellulose for adsorption of polyA⁺ RNA. The kit was used exactly as instructed by the manufacturer extracting approximately 3-15μg polyA⁺ RNA from 400μg total RNA.

2.2.18 Preparation of cDNA by Reverse Transcription and RT-PCR

First strand cDNA was prepared from RNA using BRL MMLV reverse transcriptase. 5μg of total RNA was mixed (on ice) with 7μl reverse transcriptase buffer, 3.5μl DTT (both supplied by BRL), 2μl 20mM dNTP (Pharmacia), 50pM antisense primer (or 1000pM random hexamer primers), 1μl Human Placental RNAase Inhibitor (Boehringer Mannheim) and the volume made up to 33μl with DEPC treated water. This was heated at 65°C for 10 min, placed on ice and 2μl reverse transcriptase enzyme added. The reaction was incubated at 42°C for 90 min.

5μl of first strand cDNA, prepared as described, was amplified by PCR in a 100μl volume with 1xTaq polymerase buffer (detailed in 2.2.14), 2μl 5mM dNTP,
50pM of both forward and reverse primers and 0.5µl (equivalent to 2.5units) of 
Taq enzyme. The reactions were covered in approximately 100µl paraffin oil and 
amplified under temperature conditions calculated for each primer pair.

2.2.19 RNAse Protection Assay

RNAse Protection assay was used to determine the 5' extent of cDNAs. A 
genomic fragment which extended upstream of the estimated 5' end of the cDNA 
was first subcloned into pGEM -3Zf(-), a vector which has a multiple cloning site 
flanked by the SP6 and T7 RNA polymerase promoters making possible the 
synthesis of RNA corresponding to the noncoding strand from any plasmid 
construct.

After subcloning the insert into pGEM -3Zf(-), a DNA preparation was 
made and then linearised, using a poly linker enzyme, downstream of the 
polymerase promoter and beyond the extent of the insert cDNA fragment. This 
prevents the polymerase enzyme from extending the RNA strand beyond the 
insert into the rest of the vector.

Transcription of the cDNA was carried out by mixing 1µg of linearised DNA 
with 5µl 5xtranscription buffer (200mM Tris-HCl pH 7.5, 30mM MgCl2, 50mM 
NaCl), 20units of RNAase Inhibitor (Boehringer Mannheim), 5µl 10xNTPs (5mM 
each of ATP, CTP, GTP and 250µM UTP), 3µl 32P-α-UTP (at 20µCi. µl⁻¹) and 
1.5µl either T7 or SP6 polymerase enzyme with the total volume made up to 50µl 
with DEPC treated ddw. This reaction was incubated for 1 hr at 37°C. 1µl 
RNAase-free DNAase was added to each reaction and incubated at 37°C for 10 
min to digest the DNA template. 5µl 100mM Tris-HCl pH 8.0, 100mM EDTA was 
added as a stop. The reaction was phenol/chloroform extracted and the 
phenol/chloroform phase back-extracted with 10mM Tris-HCl pH 8.0, 10mM 
EDTA, 1% SDS. The pooled aqueous phases were ethanol precipitated, and 
resuspended in 20µl hybridisation buffer (80% formamide, 40mM pipes, 10mM 
EDTA pH 6.7, 0.4M NaCl) after 15 min centrifugation at 13000g and 4°C. The 
cpm of the probe was determined by scintillation counting, and probe equivalent 
to 2x10⁵ cpm was incubated with 10µg polyA+ RNA in 30µl of hybridisation 
buffer overnight at 45°C.
Chapter 2

The sample was digested for 1 hr at 30°C with 300μl (pre-boiled) RNAase solution (40μg. ml⁻¹ RNAase A, 2μl ml⁻¹ RNAaseT, 10mM Tris, 5mM EDTA, 300mM NaCl, pH 7.5). The RNAase digests away RNA which overhangs the probe, leaving a piece of double stranded RNA made up of sample RNA and radiolabelled RNA probe. 20μl 10% SDS and 50μg Proteinase K was added for 10 min at 37°C to terminate the RNAase reaction. The samples were phenol/chloroform extracted and 10μg of tRNA (acting as 'cold carrier' for the labelled RNA) and 1ml 100% ethanol were added to precipitate the sample. The recovered pellet was suspended in 10μl loading buffer (7M urea, 1xTBE, 0.1% Bromophenol blue, 0.1% xylene cyanol) and heated at 90°C for 5 min before loading on a denaturing polyacrylamide gel (5% polyacrylamide, 6M urea). The gel was run for 2-3 hr at 50W; ³²P-α-dCTP labelled MspI cut pBR322 was run with the sample as a size marker. The gel was dried and exposed to X-ray film at -70°C.

2.2.20 Primer Extension Analysis

This method was used to determine the position of the 5' end of mRNA. An antisense primer is designed from coding sequence towards the 5' end of the RNA. The primer is end-labelled and used in a reverse transcription reaction. The resulting, labelled first-strand cDNA is sized by polyacrylamide gel electrophoresis.

40ng of antisense was end-labelled in a 20μl reaction volume with 2μl T4 polynucleotide kinase buffer (BRL), 2μl T4 polynucleotide kinase and 8μl 32p-γ-ATP for 1 hr at 37°C. 2μl 0.5M EDTA was added. 1μl of the reaction was removed for estimation of the proportion of incorporation of label. The sample was diluted with 9μl ddw and 2μl of this was spotted on to each of four small circular DE81 papers (Whatman). 2 papers were measured in scintillation fluid; the other 2 papers were first washed in 0.5M Na₂HPO₄ (3 washes), air-dried and then measured in the scintillation fluid. The proportion incorporated was estimated as (the average cpm of the washed filters)/ (average cpm of unwashed filters) and was found to be approximately 60%. The remaining 19μl of sample were ethanol precipitated, centrifuged at 13000g for 30 min (at 4°C) and resuspended in 20μl DEPC water. 1μl of this was measured by scintillation counting: specific activities were between 4x10⁶ and 4x10⁷ cpm. μg⁻¹. Between 2
and 4ng of labelled oligo, at \(7 \times 10^4\) cpm, was co-ethanol-precipitated with 20µg RNA overnight at -20°C. The recovered RNA and oligo were resuspended in 15µl DEPC ddw and 15µl 2xannealing buffer (0.5M KCl, 20mM Tris-HCl pH 8.3) and heated at 85°C for 5 min then at the annealing temperature minus 5°C (formula for calculating annealing temperature is described in 2.2.14) for 1 hr. The sample was then ethanol precipitated and the recovered sample resuspended in 12µl DEPC ddw.

The reverse transcriptase stage of the reaction used the same reagents as described in 2.2.18 with the following volumes added to the sample (in a 20µl reaction volume): 1µl 20mM dNTPs, 4µl 5xreverse transcriptase buffer, 20units RNAase Inhibitor, 200units of reverse transcriptase enzyme. After incubation at 42°C for 1 hr, the cDNA was ethanol precipitated and the recovered sample resuspended in 5µl bromophenol blue/ xylene cyanol/ glycerol loading buffer. The sample was heated at 90°C for 10 min before loading on a 6% polyacrylamide/ 6M urea gel. Electrophoresis was at 45W for 1.5 hr with \(^{32}P\) labelled \(Msp^I\) cut pBR322 vector as a size marker (as described in 2.2.19). Gels were dried and autoradiographed at -70°C overnight.

2.2.21 Isoelectric Focusing (IEF)

Protein from aqueous extracts of human tissues and from human breast milk were separated by isoelectric focusing in gels measuring 240x100x0.4mm and containing ampholines (LKB) in the ranges pH 3.5-10, pH 4-6, pH 5-8 and pH 6-8 at a ratio of 10:25:50:25 respectively. Gels were 5% acrylamide (Bio-Rad) polymerised with TEMED and ammonium persulphate. The electrode contacted the gel via a strip of 17M Whatman paper soaked in 1M NaOH positioned across the long edge of the gel at the cathode, and a similar strip soaked in 1M H₃PO₄ at the anodal end. Gels were prefocused at 300V, 8mA, 10W and 10°C for 30 min. Protein samples, generally 4µl, were loaded directly onto the gel at the cathodal end approximately 2cm from the electrode. Separation was achieved by electrophoresis for 30 min at 300V, 8mA, 10W followed by electrophoresis at 2000V, 8mA, 10W for approximately 2 hr or until approximately 4800 Vhr in total.
Gels were either passive-blotted for antibody detection or stained with an activity overlay. Passive blotting was achieved by contact with moistened nitrocellulose membrane for 2-4 hr. The membrane was blocked in 1xPBS, 0.1% Tween 20 for 20 min. The first antibody was applied at 25µg. ml⁻¹ in 1xPBS, 0.1% Tween 20 and incubated at 4°C overnight. This was removed by five 5 min washes in 1xPBS, 0.1% Tween 20. The peroxidase conjugated second antibody (rabbit anti-goat from Dakopatts) was applied as a 1:1000 dilution in 1xPBS, 0.1% Tween 20 for 2-3 hr. After a further five washes (each of 5 min) in 1xPBS, 0.1% Tween 20, the filter was stained with 500µl (5mg. ml⁻¹) diaminobenzidine (DAB), 75µl 9%H₂O₂ in 25ml 1xPBS for 5 to 20 min. The reaction was stopped by thorough rinsing in water.

The phosphoglucomutase specific activity stain overlay contained the following components: 10ml 0.5M Tris-HCl pH 8.0, 1ml 0.2M MgCl₂, 1ml NADP (5mg. ml⁻¹ solution), 0.05g glucose-1-phosphate, 100µl glucose 6 phosphate dehydrogenase (G6PD) (a 1:5 dilution of G6PD with Ammonium sulphate solution), 2ml MTT (5mg. ml⁻¹), 1ml PMS (5mg. ml⁻¹), 10ml 2% Agar Noble. The components were mixed in order, and poured onto the gel which had been sealed round the edge. Strongly stained bands were seen immediately, weaker bands were visualised after 15 min at 37°C (30 min at room temperature). The gel was photographed to keep a permanent record.
Isolation of the human PGM1 cDNA (Whitehouse et al, 1992) provided the tool required to examine the genomic structure of the PGM1 gene and questions relating to the possible correlation between protein domains revealed by three dimensional studies (Dai et al, 1992) and exon/intron boundaries, the nature of the PGM1 transcriptional promoter and the molecular basis of the common protein polymorphism.

The isolation of the 5' genomic region of PGM1, which is described in detail here, was carried out independently but as part of the group effort to characterise the entire gene structure of PGM1. Another major group effort was to identify the molecular basis of the protein polymorphism and included in the series of experiments described here is a search for genetic variation in the 5' exons.

3.1 The architecture of the 5' end of human PGM1

A human genomic DNA library constructed in the vector λ2001 was screened using the human PGM1 cDNA (Whitehouse et al, 1992) as probe. To specifically isolate a clone in the 5' region, an EcoRI/HindIII fragment spanning the first 600 base pairs of PGM1 cDNA was used as probe. The λ2001 cloning site is depicted in Fig.3.1: during the construction of the library (LeFranc et al, 1986) Sau 3A genomic fragments were inserted at the BamHI site.

A single positive recombinant, PGM1.5', was identified, purified to homogeneity and the DNA analysed by single and double restriction enzyme digests with the endonucleases XbaI, XhoI and SacI (Fig.3.2A). Sites for these enzymes flank the insert in the λ2001 cloning sites (Fig.3.1). The digested DNA was analysed by Southern blotting using the 5' 600 bp EcoRI/HindIII fragment of PGM1 cDNA (Whitehouse et al, 1992) as probe. The autoradiograph shown in Fig.3.2B revealed that a single 4.2kb XbaI fragment hybridised to this portion of the cDNA. A restriction map of the clone was constructed from the fragment size data (Fig.3.3) which showed that the 4.2kb XbaI fragment lies to one side of the
Fig. 3.1 The cloning site of λ2001. Inserts were cloned into *BamHI* (excluding *HindIII*, *EcoRI* and *XbaI* from the construct).
Fig. 3.2  **Panel A**: Single and double digests of PGM1.5' with *XbaI, SacI*, and *XhoI* restriction enzymes.  
**Panel B**: Autoradiograph of Southern blot of the same gel after hybridisation with a 5' 600bp *EcoR1/HindIII* fragment of human PGM1 cDNA as probe. The position of a 4.2kb *XbaI* fragment which hybridises to the probe is indicated.
Fig. 3.3 Map of PGM1.5' clone showing the relative positions of the 4.2kb XbaI fragment and the XbaI/HindIII 2.8kb fragment which hybridised to the 5' EcoRI/HindIII 600bp fragment of PGM1 cDNA. Restriction endonuclease sites: X, XbaI; S, Sad1; H, HindIII, X indicates an XbaI site from the vector polylinker.
λ clone. At this stage the orientation and extent of the cDNA sequence encompassed within the λ clone were not known. Digestion of the isolated 4.2kb XbaI fragment with HindIII showed that the signal from hybridising probe was confined to a single 2.8 kb XbaI/HindIII fragment (Fig.3.3). Preparative quantities of the XbaI/HindIII digest of PGM1.5' DNA were separated by agarose gel electrophoresis and the 2.8 kb fragment was isolated by elution from the gel. This fragment was subcloned into pUC19 and M13. DNA maxi-preparations of the recombinants were prepared. Restriction enzyme digests using both single enzymes (HindI, PstI and SmaI) and double combinations of enzymes (HindIII/PstI; HindIII/SmaI; PstI/SmaI) were carried out (Fig.3.4A). The digested DNA was analysed by Southern blotting using the 600 bp PGM1 probe described above: the autoradiograph is shown in Fig.3.4B. Only one fragment from each digest hybridised to the probe. A restriction enzyme map of the region, positioning HindIII, PstI and SmaI sites, was constructed (Fig.3.4C) from the restriction fragment size data. This showed that the hybridising region lies towards the centre of the 2.8kb fragment, confined to a 550bp PstI/SmaI fragment.

3.2 Sequence analysis of the 5' end of PGM1.

Simple Southern blotting and mapping procedures could not accurately determine the exact location of coding DNA within the 2.8kb XbaI/HindIII subclone or the exon/intron boundaries, therefore sequence analysis of the region was necessary.

The strategy employed is illustrated in Fig.3.5. The 2.8kb XbaI/HindIII region was subcloned into M13 mp18 and in M13 mp19, in the reverse orientation with respect to the M13 'universal' sequencing primer (denoted primer 1 in Fig.3.5A), thus allowing the two ends of the fragment to be sequenced. These sequence data were then used to design sequencing primers to 'walk' in towards the centre of the fragment (primers 2, 5, 6, 7 and 8 in Fig.3.5A). Additional primers were designed from the known cDNA sequence (Whitehouse et al, 1992): primer 3 from the 5' untranslated region positioned nt 14 to nt 37, primer 10 from nt 63 to nt 84, primer 9 from nt 119 to nt 143, primer 4 from nt 246 to nt 270 and primer 11 (not shown in Fig.3.5A) from nt 341 to nt 365. All of the
Fig. 3.4  **Panel A:** Restriction endonuclease digests of \textit{XbaI/HindIII} 2.8kb region on an ethidium bromide stained agarose gel. 
\textit{M} indicates size marker.

**Panel B:** Autoradiograph of Southern blot of the same gel after hybridisation with the 5' EcoRI/HindIII 600bp fragment of PGM1 cDNA as probe.


4 cm = 1 kb
**B**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Number of templates used</th>
<th>Number of sequence reactions with primer</th>
<th>Distance sequenced with primer (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (mp18)</td>
<td>6</td>
<td>11</td>
<td>340</td>
</tr>
<tr>
<td>1 (mp19)</td>
<td>4</td>
<td>8</td>
<td>540</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>250</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>358</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>14</td>
<td>236</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>180</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>7</td>
<td>270</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>320</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>400</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>7</td>
<td>180</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>9</td>
<td>400</td>
</tr>
</tbody>
</table>

Fig. 3.5  **A**: Sequencing strategy employed to sequence the 2.8kb *XbaI/HindIII* fragment subcloned into M13. Arrowed lines indicate the extent of sequence obtained with each primer. Exon 1 is boxed. The bold line indicates flanking sequences and is drawn to scale (1cm: 250bp).

**B**: Table summarising sequence data obtained with each primer.
primers except primer 11 were able to prime DNA synthesis from either the mp18 (primers 9 and 10) or mp19 (primers 3 and 4) subclones. The sequence corresponding to primer 11 was considered to be absent from the 2.8kb Xbal/HindIII fragment. At least two independent subclones of the fragment in M13 were sequenced with each primer. Additionally, 16% of the sequence was read in both directions and there were sequence overlaps covering 25% of the entire sequence (Fig.3.5B).

These analyses showed that the 2.8kb Xbal/HindIII fragment contained cDNA corresponding to a single exon, exon 1. Sequence analysis using primers 9 and 10 defined the 5' end of the cDNA while the 3' boundary of exon 1 was defined by sequencing using primers 3 and 4 by identifying the position where recognisable cDNA sequence was discontinued and intervening sequence began. The exon/ intron junction of exon 1 conforms to the consensus eukaryotic structure beginning with the dinucleotide GT. Exon 1 coding sequence was found to extend 246 bp in length and to end with amino acid residue 82, a glycine. The complete sequence of the 2.8kb Xbal/HindIII region is shown in Fig.3.6. It was found to contain in addition to 246bp of coding sequence in exon 1, 948bp of 5' flanking sequence and 1546bp of intron 1. Complete sequencing of the fragment showed that it did not encompass exon 2.

The sequence analyses also determined the orientation of exon 1 within the 2.8kb Xbal/ HindIII fragment and the λ PGM1.5' clone, illustrated in Fig.3.7, with the 5' end of PGM1 at one end of the phage insert. This orientation, together with the observation that λ PGM1.5' did not contain any other cDNA sequence, implied that exon 1 and exon 2 were separated by a large intron of at least 14kb. The exact location of exon 2 was eventually defined at a position 38kb downstream of exon 1. This work was carried out by Ms Wendy Putt who isolated further genomic clones (λ11, λ3, λ5 and λ4) which spanned the gap between λ PGM1.5' and exon 2. This is shown in Fig.3.8 together with clones spanning the rest of the gene.

In summary, the first exon of PGM1 has been found to encompass 246bp coding sequence and is situated 38kb upstream of exon 2.
Fig. 3.6 This is the sequence of the 2.8 kb Xbal/HindIII fragment containing 948 bp of 5' flanking sequence, 1546 bp of intron 1 sequence and 246 bp of exon 1 coding sequence (shown in bold print). The Xbal and HindIII cloning sites are underlined.
Fig. 3.7 Map of the genomic clone PGM1.5' showing position and orientation of exon 1. From this data, intron 1 is at least 14kb long. Restriction enzyme sites are: X, XbaI; H, HindIII; S, Sad.
Fig.3.8  Genomic clones which span the entire PGM1 gene. Vertical bars represent exons (numbered above). Arrowed lines represent the extent of each λ clone. Intron 1 is 38kb long and is spanned by clones λ1.5', λ11, λ3, λ4, and λ5. Scale is 1cm=7kb.
3.3 The promoter region of Human PGM1

The sequence of the 5' exon and immediate flanking regions appeared to be highly GC rich, this feature gave rise to some problems with sequence analysis in the form of GC compression. To resolve the compression, dITP mixes were used in replacement of the usual termination mixes (see Chapter 2 Methods) in the sequence reactions. As a result the sequence in the flanking region downstream of exon 1, which read (5' to 3') as CCCGGCG when compressed, could now be more accurately read as CCCCCGCGCG with the GC compression removed (Fig.3.9).

The sequence upstream of exon 1 exhibits features characteristic of housekeeping genes (Fig.3.10). There is no TATA box or CCAAT box, but the core consensus binding element for the ubiquitous transcription factor Sp1, CCCGCC, occurs six times in the 900 bp region around exon 1. Exon 1 and the region around it are highly GC rich: G+C% in the region -240 bp to +600 bp is 70% compared to 40% in bulk DNA. The incidence of CpG dinucleotides is also high in this region: the CpG/GpC ratio is 0.79 compared to 0.10 in bulk DNA. The high CpG content is reflected in the large number of Hpal restriction enzyme sites (CCGG) identified by sequence analysis: twelve sites are found in the region -240 bp to +600 bp producing fragments with an average length of 60 bp. Multiple Hpal sites are the feature of 'Hpal-tiny-fragment-islands' found in many housekeeping genes (Bird, 1986). The distributions of CpG and GpC dinucleotides and Hpal restriction enzyme sites across the entire 2.8 kb Xbal/HindIII fragment are plotted in Fig.3.11A, B and C. This illustrates the clustering of CpG dinucleotides and Hpal sites in the region spanning exon 1 and immediate flanking regions.

The human PGM1 cDNA as reported by Whitehouse et al, (1992) has a 5' untranslated region of 58bp. Since this cDNA does not necessarily contain the complete 5' untranslated region, RNAse Protection was employed to determine the transcription initiation start site and the full length of the 5' untranslated region of PGM1. The basis of the technique relies on subcloning a genomic fragment...
Fig. 3.9 Resolution of GC compression in intron 1 by sequence analysis in which dGTP is replaced by dITP, and ddGTP termination mixes are replaced with ddITP mixes. The sequence of interest is shown in bold print. 
Panel A shows compressed sequence observed using normal sequence mixes. 
Panel B shows accurate sequence observed using dITP and ddITP mixes.
Fig.3.10 Exon 1 and flanking sequence. cDNA sequence is shown in upper case. Five sites of the core consensus binding element for Sp1 are shown in bold type. The translation start site is underlined. Numbering is as given in Whitehouse et al, 1992.
Fig. 3.11  

A: *XbaI/HindIII* 2.8kb fragment drawn to scale (1cm=200bp).  
Black box denotes exon 1.  

B: Plot of GpC distribution. Each vertical line represents 5 sites.  

C: Plot of CpG distribution. Each vertical line represents 5 sites.  

D: Plot of *HpaII* restriction endonuclease sites. Each vertical line represents 1 *HpaII* site.
Fig. 3.12 **Panel A:** Illustration of a 1026bp genomic *HindII/Smal* fragment (bold line) which extends upstream of the estimated 5' UTR (dashed line) and downstream into intron 1, encompassing exon 1.

**Panel B:** Multiple cloning site of pGEM -3Z(+) plasmid with relative positions of the T7 and SP6 RNA polymerase promoters, and cloning position of genomic fragment illustrated in A: the RNA copy is made in a 5' to 3' direction.
which spans the 5' untranslated region (extending in a 5' direction such that the DNA is longer than the expected untranslated length (Fig.3.12B)) into a vector which has RNA polymerase promoter allowing an RNA copy of the DNA insert to be made with 32P radiolabel incorporated. The cRNA is then hybridised to the poly A+ RNA of interest. RNAses are used to digest any overhanging single-stranded RNA/cRNA leaving a double-stranded duplex which extends from a fixed position within the gene at the 3' end and extends as far as the full length of the 5' untranslated region at the 5' end. The vector used was pGEM 3Zf(+) (Promega) which has SP6 and T7 RNA polymerase promoters flanking a multiple cloning region (Fig.3.12A); the RNA polymerases work in a 5' to 3' direction.

Fig.3.13A shows the position relative to exon 1 of a Smal/HindII fragment used to prepare the cRNA probe. The full (undigested) length of the probe is 1026 bp which comprises 227bp of intron 1, 246bp of exon 1 coding sequence and 553bp upstream to the coding sequence. After hybridisation of this cRNA probe to 10μg of adult muscle, foetal muscle or adult gut poly A+ RNA and digestion with RNase, the size of the protected double-stranded fragment was analysed by denaturing acrylamide gel electrophoresis. MspI digested pBR322 fragments radiolabelled with 32P were used as DNA size markers together with a conventional dideoxy sequence analysis track of known DNA sequence. The largest protected fragment was 328bp long for all tissue RNAs analysed. The length of the 5' untranslated region, 82bp, was calculated by subtracting the length of the coding sequence, 246 bp from the length of the protected fragment 328 bp (Fig.3.13A). The published sequence (Whitehouse et al, 1992) has a 5' untranslated of 58bp, thus the length computed from the RNase protected fragment represents an extension of the published sequence by 24bp, and suggests that the full length of the most common PGM1 mRNA is 2342bp.

A second protected fragment of 160bp was also observed in all tissues. The 5' end of this protected fragment is positioned at nucleotide +82 into the coding sequence. The 5' end of this protected fragment has the sequence CCAGAGCAGC, curiously an identical ten base pair sequence motif is found at the 5' end of the usual transcript with no other obvious sequence homology either side. The significance of this shorter transcript is not clear. The presence of an identical 5' motif seems to imply that this sequence is associated with transcription initiation. However it may not be functional in vivo and may not be
Fig. 3.13 A: 1026bp PGM1 genomic Smal/Hincl probe comprising 227bp of intron 1, 246bp of exon 1 and 553bp of 5' sequence. B: Acrylamide gel showing RNAase protected fragments (P) alongside the undigested Smal/Hincl probe and dideoxy sequence tracks of known DNA sequence (M).
translated efficiently. The next ATG sequence downstream of this putative second transcription initiation start site is the methionine at amino acid residue 67 of the full length protein. This ATG is embedded in the sequence TTCTACATGA. In a survey of mRNA sequences, Kozak (1987) found that a purine was generally found at position -3 relative to the A of the translation initiation start site. Only 23 out of 699 mRNAs examined did not have a purine at this position, and 17 of these 'compensate' by having a G at position +4 (where A of ATG is +1) also thought to be important in strong translational efficiency. This sequence has neither of these important features. In contrast, the full length transcript ATG is embedded in the more efficient Kozak consensus sequence GCCACCATGG.

In summary, exon 1A is transcribed from a typical housekeeping promoter, with no TATA or CCAAT boxes. There appears to be a downstream transcription initiation, a feature seen in many genes with no TATA box. It is speculative whether a protein is translated from this shortened transcript.

3.4 Exon 1B, an alternative first exon in PGM1: position and sequence

In 1992, during the isolation of rabbit PGM cDNA, an unusual transcript was described by Lee et al. (1992). They found that one out of five PGM cDNAs isolated from a rabbit muscle cDNA library showed sequence differences at the 5' end when compared to rabbit PGM1 cDNA reported by Whitehouse et al. (1992), and to the other four cDNAs which they had isolated which were identical. The two types of cDNA differed significantly upstream of a position corresponding to amino acid residue 83 (Fig.3.14A). Lee et al. (1992) proposed that the cDNA may arise from a separate but closely related gene to PGM1. An alternative explanation is that the transcripts arise from the same gene but are processed differently at their 5' ends.

At the time when Lee et al. (1992) reported this cDNA sequence, the gene structure for PGM1 was not available. As it emerged careful inspection of the sequence of the novel transcript shows that the sequence homology between the novel and rare transcripts diverged precisely at the boundary of exon 1 and exon 2, with some sequence variation between the two cDNAs but with complete sequence identity in the rest of the transcript (Fig.3.14A). This lends support to the explanation that the transcripts arise from the same gene and suggests the possibility of alternative first exons. To test if
Fig. 3.14  

A: Comparison of 5' ends of rabbit PGM1 (RaPGM1) and novel rabbit sequences (nPGM). Exon 1 and part of exon 2 are shown with Exon 1 boxed (dashed lines) illustrating the sequence divergence at the exon 1/exon 2 junction. PCR primers for the amplification of the novel sequence are boxed (solid lines).

B: Agarose gel with 254bp PCR amplification product (lane 1) of rabbit genomic DNA with primers pgmppF and pgmppR. M=size marker.
this was the case, the whole gene was searched for the novel 5' sequence using a PCR procedure. PCR primers were designed from the novel rabbit cDNA sequence (Fig.3.14A). Primer pgmppF spans nucleotide 2 to 23 and pgmppR from nucleotide 233 to 255 (numbering is taken from Lee et al, 1992). These primers were demonstrated to amplify the expected product of 254 bp from rabbit genomic DNA (Fig.3.14B).

Primers pgmppF and pgmppR were then used in PCR amplifications of DNA from seven \( \lambda \) genomic clones spanning 65kb of the human PGM1 gene (Fig.3.15A). A 254bp product was amplified from two overlapping clones \( \lambda 3 \) and \( \lambda 4 \) (Fig.3.15B). Clone \( \lambda 3 \) was analysed by single and double restriction endonuclease digests with the enzymes \( XbaI \), \( XhoI \) and \( SacI \) (Fig.3.16A) and the gel Southern blotted and probed with the pgmppF/pgmppR 254bp PCR product (amplified from human genomic DNA). The autoradiograph (Fig.3.16B), revealed that the novel 5' sequence lay in a 2.4kb \( XbaI \) band towards the centre of the \( \lambda 3 \) genomic clone (Fig.3.16C). The position of the novel sequence, designated exon 1B, within this fragment was found to be 6 kb upstream of exon 2 and 32 kb downstream of exon 1 (Fig.3.17A).

In order to sequence the human genomic region corresponding to exon 1B, the 2.4 kb \( XbaI \) fragment was subcloned into M13mp18 in both orientations. Orientation was assessed making use of an assymetric \( BamHI \) site. Inserts in a 5' to 3' orientation with respect to the polylinker could be sequenced from the 3' end with the universal primer and with the PCR primer pgmppR. Those in a 3' to 5' orientation could be sequenced from the 5' end with the universal primer and with the PCR primer pgmppF (illustrated in Fig.3.17). Internal sequencing primers were designed (Fig.3.17B) : pgmppF2 (nt 234 to 255) was used to extend the sequence data into the 3' flanking region; pgmppR2 (nt 106 to 123) was used to complete the sequence at the 5' end of exon 1B (numbering in Fig.3.17 is based on the human sequence equivalent to rabbit sequence described by Lee et al (1992)). In total, 1351 bp of sequence in and around exon 1B was completed (Fig.3.17B).

Lee et al (1992) designated a methionine codon, which approximately corresponded with the terminus of the common PGM1 isoform, as the translation start site of the rabbit muscle PGM1B isoform. However, our
Fig. 3.15  
**A:** Genomic clones which span the entire PGM1 gene. Arrowed lines represent the extent of each \( \lambda \) clone. Intron 1 is 38kb long and is spanned by clones \( \lambda 1.5' \), \( \lambda 11 \), \( \lambda 3 \) and \( \lambda 4 \), Scale is 1cm=7kb.  
**B:** Agarose gel with amplification products of DNA from genomic clones \( \lambda 1.5 \), \( \lambda 11 \), \( \lambda 3 \), \( \lambda 4 \), \( \lambda 1 \), \( \lambda 6 \) and \( \lambda 7 \) (illustrated in **A**) using PCR primers pgm1pF and pgm1pR. A 254bp product was amplified from \( \lambda 3 \) and \( \lambda 4 \) DNA. M is a size marker; C is a no-DNA control amplification.
Fig. 3.16 **Panel A**: Single and double digests of PGM1 genomic clone λ3 with *XbaI*, *SacI* and *XhoI* restriction enzymes. **Panel B**: Autoradiograph of Southern blot of the same gel after hybridisation with the pgmppF/pgmppR 254bp PCR product as probe. The position of a 2.4kb *XbaI* fragment which hybridises to the probe is indicated. **Panel C**: Restriction map of clone λ3 showing *SacI* (S) and *XbaI* (X) sites. The 2.4kb *XbaI* fragment which hybridises to the probe is indicated by a bold line. Scale 1cm:1.3kb.
Fig. 3.17 A: Position of novel sequence 32kb downstream of exon 1A and 6kb upstream of 2, X=XbaI site. Exon 1B position is shown within the 2.4kb XbaI fragment, sequencing primers are indicated by arrowed lines and the extent of the region sequenced is shown below.

B: 1351bp of sequence in and around exon 1B. Sequencing primers pgmppF, pgmppF2, pgmppR and pgmppR2 are boxed. Exon 1B coding sequence is boxed (dashed line). Numbering is as Young Sup Lee et al, 1992.
further analyses of the human and rabbit sequences upstream of this proposed start site revealed an in-frame ATG codon at a more 5' position. Since it seems likely that this is the real start site it is proposed that the open reading frame continues in a 5' direction a further 42 nt, 14 amino acids (Fig.3.18). This proposal is supported by the high level of conservation shown between human and rabbit in this upstream region. Furthermore this upstream start site is embedded in the motif TGAAAGATGA in the rabbit and TGAAGGATGA in human which both conform to examples described by Kozak (1987) for initiation of translation. Given that this upstream translation start site is real, then the full length of the coding sequence of exon 1B is 300 bp (100 amino acid residues) (Fig.3.18). From these analyses it is proposed that exon 1A and exon 1B are alternative first exons for PGM1 and it seems likely that since their 5' untranslated regions are different that they are each transcribed from a separate promoter. The genomic distance between exons 1A and 1B indicates that their promoters are at least 32kb apart at the 5' end of the PGM1 gene.

Comparisons were made between the rabbit exon 1B sequence and flanking regions, and the equivalent human sequences. Exon 1B is highly conserved between human and rabbit with 91% nucleotide sequence identity in the coding region (Fig.3.18) and 93% amino acid identity. When the 5' untranslated region of the rabbit cDNA is compared with the human genomic sequence corresponding to that region, 71% homology is found across a 390nt stretch of sequence.

3.4.1 Evidence for a duplication of PGM1

Exon 1B, which is 300 nucleotides in length and encodes 100 amino acids, is 54 nucleotides longer than exon 1A and has 18 additional amino acid residues at the N-terminal end (Fig.3.19). However, there are considerable similarities between the two exons which show 58% homology at the DNA level (Fig.3.19A) and 74% overall amino acid similarity (Fig.3.19B). There are three relatively large blocks of complete amino acid identity comprising 11, 5 and 5 contiguous amino acid residues and an overall amino acid identity of 51% (Fig.3.19B). The sequence similarities between exon 1A and exon 1B show that they are related and suggest that they have arisen by a duplication event. This is an unusual observation. There are numerous examples in the literature of alternative promoters and differential splicing which generate multiple transcripts
Fig. 3.18 Comparison of rabbit (Ra) and human (Hu) exon IB and upstream sequences. Coding sequence is boxed; additional 42nt of open reading frame is in bold print; consensus motifs recognised for the initiation of translation are underlined. Vertical bars represent nt identity.
Fig. 3.19
A: Nucleotide sequence comparison between human exons 1A (hu 1A) and 1B (hu 1B). Exon 1B sequence is shown in bold type; vertical bars represent nucleotide identity. Exon 1B is 54 nt longer than exon 1A (sequence is boxed).
B: Comparison of exon 1A and 1B amino acid sequence. Vertical bars represent identity, * represents homology.
from a single transcriptional unit: aldolase A (Izzo et al., 1988), α1-antitrypsin (Perlino et al., 1987), mouse carbonic anhydrase I (Fraser et al., 1989) and the adenomatous polyposis gene (Horii et al., 1993). However, there appear to be no other examples of alternatively spliced first exons which are duplicates one of the other (see Discussion, 6.1).

The boundaries of the proposed PGM1 gene duplication event have not yet been defined and there are at least two possible models for duplication of the first exon. A simple model is that of the duplication of only a small region of the gene including exon 1 and no other coding region and translocation to a position 32kb away from the original position. A more complex model can be constructed by analysing the gene structure. A scale drawing of the gene is shown in Fig. 3.20; the distance between exon 1A and exon 1B is 32 kb, similar to the distance of 37 kb between exon 1B and exon 11. From these sizes it is possible to envisage a whole gene duplication of exons 1 to 11 placing two copies of the PGM1 gene side by side. Subsequent loss of coding sequence in exons 2 to 11 in the 5' copy may have left a total of twelve exons including a duplicated exon 1, and the unusually large first intron.

If the loss of the exons was relatively recent in evolutionary terms, then it may be possible to detect sequences, corresponding to degraded exons, in the large first intron. This possibility was investigated by Southern blotting and by PCR amplification. Restriction enzyme digests of genomic clones λ1.5', λ11 and λ3 (which span the first intron, shown in Fig. 3.8) were Southern blotted and probed with a radiolabelled fragment of the human PGM1 cDNA from exons 2 to 11. No signals were detected after autoradiography following hybridisation at 55°C and a filter wash at 0.2xSSC, 0.1%SDS at 50°C for 15 minutes. Additionally, pairs of PCR primers were designed from within exons 2, 3 and 4 and used in amplifications of clones λ1.5' and λ11 (Fig. 3.15): no products were amplified. That no remains of the missing exons could be detected can be explained by proposing that the duplication event was ancient such that many changes have accumulated in the region between 1A and 1B which prevent the detection of sequences corresponding to the lost exons. The accumulation of sequence differences between exons 1A and 1B have been fewer, but are sufficient that primers for PCR amplification can be designed from exon 1A.
Fig. 3.20 Diagram of PGM1 genomic structure drawn to scale. Vertical lines represent exons. The bold arrowed line represents the length from exon 1B to exon 11 (37kb). The dashed arrowed line represents the length from exon 1A to 1B (32kb).
sequence that do not amplify products from exon 1B, and vice versa, despite their clear relatedness.

3.4.2 The 5' extent of the PGM1.1B transcript

The rabbit PGM1 cDNA with exon 1B (PGM1.1B), described by Lee et al. (1992) has an unusually long 5' untranslated region which was determined by reverse transcription of rabbit fast skeletal muscle poly A+ RNA using a PGM specific primer followed by dATP-tailing of the product and subsequent PCR amplification. They found the 5' untranslated region to be 390bp long. A comparison of this region in rabbit and the equivalent region in human sequence showed 70% homology. This high level of homology indicates that this sequence is conserved between species and is likely to be transcribed as part of a human cDNA as well as in rabbit. In this study primer extension analysis was employed to determine the transcription initiation start site of the human PGM1 1B transcript. In this method a radiolabelled reverse orientated primer is used to prime a reverse transcription reaction from RNA of interest. The primer (nucleotide -245 to nt -267 , numbering from ATG where A=1) was end-labelled with $^{32}$P γATP, and used to prime transcription from 20μg adult skeletal muscle total RNA. The position of the primer is illustrated in Fig.3.21. The length of the first strand cDNA product was analysed by acrylamide gel electrophoresis, alongside MspI digested pBR322 fragments radiolabelled with $^{32}$P and sequence reactions of M13 single stranded DNA with the 40mer sequencing primer as size markers, followed by autoradiography.

The major primer extension product was 132 bp long (Fig.3.21) indicating that the 5'UTR of human PGM1 exon 1B is 377 bp long. This is 13 bp shorter than the rabbit equivalent (Lee et al, 1991) but extends to the position predicted by sequence comparison between rabbit 5' UTR and human sequence which indicated 70% homology to that position. These 5'UTRs of rabbit (390bp) and human (377bp) PGM1 1B are unusually long. In a survey of mRNA sequences, leader sequences of most vertebrate mRNAs were found to fall in the size range of 20-100 nt (Kozak, 1987). A common feature of long 5' UTRs is high GC content but this is not apparent in this case. Also, the long 5' UTR is unlikely
length of 5' UTR = size of primer extension product + 245bp

Fig. 3.21 A: The position and orientation of the primer (at nt position -245) for primer extension is indicated by an arrowed line: shaded area is exon 1B coding region.

B: Autoradiograph of primer extension. M is dideoxy sequencing track of known DNA sequence; pB indicates MspI cut pBR322 used as a size marker with sizes of bands indicated left of the photograph; AM indicates 2 tracks of primer extension of adult muscle RNA with 4ng (1) and 2ng (2) of primer; AG indicates negative control extension using adult gut RNA.
to be involved in favoured translation of 1B over 1A in skeletal muscle since only one out of four cDNAs cloned was of the 1B type (Lee et al., 1992).

The sequence of exon 1B and flanking regions is shown in Fig.3.22. As was found for exon 1A there is no TATA or CCAAT box. In contrast to exon 1A, the region flanking exon 1B has a relatively low GC content at 47% and the CpG/GpC ratio is 0.24 compared to 0.79 in exon 1A and 0.1 in bulk DNA. There are no Sp1 transcription factor binding site sequence motifs in or flanking exon 1B. However, there are six copies of an E-box consensus sequence CANNTG, characteristic of binding sites for the myogenic determination transcription factors, positioned at nt -276, -142, -35, +10, +22 and +287 relative to the transcription initiation start site (+1). Only one such sequence is found in the proximal promoter region of exon 1A (nt -15).

3.4.3 Expression of the PGM1.1B transcript

The rabbit PGM1 cDNA containing exon 1B was cloned from an adult fast-twitch skeletal muscle cDNA library (Lee et al., 1992). No other data regarding the expression of this transcript was available and in particular there was no information as to whether PGM1.1B was ubiquitous as is PGM1.1A, or was characteristic of muscle. In order to ascertain the pattern of expression of the 1B transcript, RNA from various human tissues were analysed by RNA PCR amplification. Three pairs of PCR primers were designed (Fig.3.23): one pair comprised a forward primer specific for human exon 1B sequence and a reverse primer taken from exon 2 sequence; a second pair comprised a forward primer specific for exon 1A and another reverse primer in exon 2; a control pair of primers comprised a forward primer taken from exon 2 and a reverse primer taken from exon 5. The exon primers in each pair are separated in genomic DNA by at least one intron, so that amplification products of cDNA can be recognised unambiguously. The amplification using the control pair amplify PGM1 RNA regardless of whether it contains exon 1A or exon 1B.

The use of random hexamer oligonucleotides as primers for reverse transcription allowed identical aliquots of the same first strand cDNA to be amplified with each of the three primer pairs described above and thus a direct comparison could be made of exon 1A and 1B products within cDNA samples. In
<table>
<thead>
<tr>
<th>Number</th>
<th>DNA Base Pair</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-281</td>
<td>ttgcatgtgt</td>
<td>ggaatcaaaa</td>
</tr>
<tr>
<td>-231</td>
<td>gtaaagaaaa</td>
<td>caggggtaaa</td>
</tr>
<tr>
<td>-181</td>
<td>acctttaact</td>
<td>tgcttagttg</td>
</tr>
<tr>
<td>-131</td>
<td>cacagatgat</td>
<td>gccaagagag</td>
</tr>
<tr>
<td>-81</td>
<td>gagcccacacc</td>
<td>agggtgtctt</td>
</tr>
<tr>
<td>-31</td>
<td>gcccccctttt</td>
<td>gcttggtgct</td>
</tr>
<tr>
<td>20</td>
<td>TACAGCTGAA</td>
<td>GGGGCTGCA</td>
</tr>
<tr>
<td>70</td>
<td>ACGGAAATGAG</td>
<td>AAAAAATAAT</td>
</tr>
<tr>
<td>120</td>
<td>GAAGTGTCCT</td>
<td>CATCAAACA</td>
</tr>
<tr>
<td>170</td>
<td>GAGTGGTTGT</td>
<td>GTGGCCCTTA</td>
</tr>
<tr>
<td>220</td>
<td>ATTAACCAAG</td>
<td>CTTTCATTT</td>
</tr>
<tr>
<td>270</td>
<td>TTAGTTACTC</td>
<td>ATAGACACCT</td>
</tr>
<tr>
<td>320</td>
<td>TGGAGTCCTCT</td>
<td>TGAGCTCCTT</td>
</tr>
<tr>
<td>370</td>
<td>AGTGAAGGATG</td>
<td>GAGTGATTC</td>
</tr>
<tr>
<td>420</td>
<td>ATGGAGAAAG</td>
<td>GTCTCTCTCC</td>
</tr>
<tr>
<td>470</td>
<td>CGATCCAAGAA</td>
<td>AAGGAAACCA</td>
</tr>
<tr>
<td>520</td>
<td>AGGAAAAAGCC</td>
<td>ATGGATACCTG</td>
</tr>
<tr>
<td>570</td>
<td>ATAGATACAT</td>
<td>AAGTGTTTCA</td>
</tr>
<tr>
<td>620</td>
<td>GCGGCTTCTTT</td>
<td>AATAAATCAG</td>
</tr>
<tr>
<td>670</td>
<td>CCAAATGGGgt</td>
<td>ggttatagta</td>
</tr>
<tr>
<td>720</td>
<td>taaccttaaat</td>
<td>ttacagagaag</td>
</tr>
</tbody>
</table>

**Fig. 3.22**

Sequence of and flanking exon 1B. The 5' untranslated and coding sequence is shown in upper case, intron regions are shown in lower case. Six E-box consensus sequences (CANNTG) are shown in bold print. The translation initiation site is underlined. Numbering is from transcription initiation start (+1).
Fig. 3.23 Position of PCR primers for the amplification of exon 1A, 1B and the exon2/exon5 control product, shown in genomic DNA and in cDNA (not to scale). Arrowed lines represent position and direction of primers, boxes represent exons, and horizontal lines represent introns.
addition identical quantities of total RNA (5µg) were used from each tissue source for cDNA preparation. These strategies were used so that a rough estimate of the relative levels of PGM1A and PGM1B could be made for each tissue. This is not entirely quantitative because the amount of messenger RNA versus ribosomal RNA in the total RNA pool used in the reverse transcription was not measured. Also, different species of RNA in each tissue pool may compete differentially for random oligo primers. Amplification products were analysed both by ethidium bromide staining and Southern blotting using the radiolabelled human 1B PCR product as probe (Fig.3.24A and B). Expression was analysed in adult skeletal muscle, heart, gut, liver, kidney and foetal muscle, heart, gut and liver. The exon 1A product was amplified from cDNA from all of the tissues examined and could be detected by simple ethidium bromide staining. However, exon 1B could only be visualised in this way in the adult skeletal muscle sample. Using autoradiography and the exon 1B product as probe, exon 1B amplification products could be detected in adult gut, adult kidney and foetal heart. The signals for adult gut and foetal heart were very faint and required a long exposure autoradiograph; the adult kidney signal was approximately 5 to 10 fold the intensity of that of adult gut. From this data it appears that PGM1.1B is most abundant in striated muscle, present at low levels in kidney and is absent or very low in cardiac and smooth muscle and other non-muscle tissues. Using the exon 1B PCR primers designed from human sequence, products of the expected size were amplified from cDNA reverse transcribed from mouse skeletal muscle mRNA, indicating that a PGM1.1B transcript is also present in this tissue (data not shown).

As PGM1.1B is strongly expressed in human skeletal muscle and rabbit PGM1.1B cDNA is present in fast twitch skeletal muscle (Lee et al, 1991), we were interested to determine if the expression was differentially distributed between different muscle fibre types. Human muscles are of approximately equal mixed Type 1, slow, and Type 2, fast, fibres but some rabbit muscles are almost exclusively of one fibre type. First strand cDNA was generated from RNA extracted from rabbit extensor digitorum longus (Type 2, fast muscle) and soleus (Type 1, slow muscle) as described above. The cDNA was amplified by the control PCR and the exon 1A and exon 1B PCRs as described above. Both exon 1A and exon 1B products were amplified from both
Fig. 3.24  

A: Exon 1A and 1B PCR products following reverse transcriptase of RNA from adult and foetal muscle, heart, gut and liver and adult kidney. Size markers are indicated by lane M, lane CB is an exon 1B no-DNA control. * indicates primer dimer band seen in some negative PCR amplifications.

B: Autoradiograph of Southern blot of the gel shown in A. Lanes are as shown in A. Exon 1B product can be detected in adult muscle, gut and kidney, and in foetal heart.
Fig. 3.25 Exon 1A and 1B PCR products following reverse transcriptase of RNA from rabbit extensor digitorum longus (EDL) (Type 2, fast muscle) and soleus (Type 1, slow muscle). Size markers are indicated by lane M. CA is the no-DNA control for the 1A amplification, CB is the no-DNA control for the 1B amplification.
extensor digitorum longus and soleus cDNA in roughly equal amounts (Fig. 3.25). This infers that exon 1B is not fibre type specific in rabbit muscle.

3.5 Search for genetic variation in the 5' exons of PGM1

A major group effort was to define the molecular basis of the common PGM1 protein polymorphism. DNAs were collected from a panel of individuals whose PGM1 phenotype was determined by isoelectric focussing analysis of a blood sample. From these individuals a working test panel was selected of DNA from individuals representing each of the PGM1 phenotypes (Table 3.1). These DNAs were then analysed by sequencing each exon in each individual to look for sequence variation. Exons 1A and 1B were sequenced in full from five and four individuals of different PGM1 phenotype, respectively, as part of this present project (Table 3.1).

For exon 1A, PCR primers were designed from sequence upstream of exon 1A coding sequence and from intron 1 to amplify the whole of 1A coding sequence. Primers were also designed from flanking sequence of exon 1B to amplify its entire coding sequence. The exon 1A forward primer was biotin labelled so that the double stranded PCR product could be split into single strands using a streptavidin-coated bead separation system. The two sets of single strands could then be sequenced using as primers those used in the PCR amplification. The sequence data obtained from exon 1A using this method was inadequate probably due to the high GC content of the region which causes secondary structures preventing the separation of the strands by the beads. An alternative strategy was devised whereby the PCR products were subcloned into M13 and then sequenced. To improve the efficiency of subcloning M13 was cut with Smal, and then TaqI polymerase used to add a T nucleotide to the blunt cut end of the vector. PCR products have an overlapping A nucleotide on their 3' ends which then ligate to the T added to M13. The inserts subcloned in this way were sequenced using the forward and reverse primers used in the initial PCR amplification as well as two internal primers designed from cDNA sequence. Exon 1A was sequenced from one individual of each of the phenotypes 2+, 2-, 1+, 1- and 2+2-. No sequence variation was detected. Using the same method,
exon 1B was sequenced from one individual of each of the phenotypes 1+, 1-, 2+ and 2+2-, and again no sequence variation was detected.

Analysis of the remaining 10 PGM1 exons identified two mutations. A C-T transition at nt 723 in the coding sequence changes the amino acid sequence from arginine to cysteine at residue 220. Another C-T transition at nt 1320 changes amino acid residue 419 from tyrosine to histidine. These two mutations show complete association with the 2/1 and +/- polymorphisms respectively (March et al, 1993).

<table>
<thead>
<tr>
<th>PGM1 phenotypes</th>
<th>No. individuals collected</th>
<th>No. individuals in test panel</th>
<th>No. individuals sequenced 1A</th>
<th>No. individuals sequenced 1B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1+</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2+</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2+1+</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2+1-</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1+1-</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2-1+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2+2-</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2-1-</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TOTALS</td>
<td>47</td>
<td>9</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3.1 Summary of the number of samples collected from individuals of the ten common PGM1 phenotypes, the number included in a working test panel and the number and types analysed by sequencing exon 1A and exon 1B.
Chapter Four
PGM3

A set of isozymes determined by the third PGM locus, designated PGM3, were identified after electrophoretic studies of tissue with high PGM activity (Hopkinson and Harris, 1968). Populations were found to be polymorphic for two common alleles at this locus. Tests for linkage using dizygotic twin data showed that PGM1 and PGM3 are not closely linked (Hopkinson and Harris, 1968). Eventually somatic cell hybridisation analysis led to the assignment of the PGM3 locus to chromosome 6, later refined to the band 6q12 (Meera Khan et al, 1984).

Similarities in biochemical properties, such as protein size, isozyme electrophoretic patterns and thermostability, suggest that PGM1, PGM2 and PGM3 have arisen by duplication and thus will show some degree of nucleotide homology. For this reason the human PGM1 cDNA was used to screen for PGM3 in the sorted chromosome 6 cosmid library (Reference Library, ICRF; Lehrach et al, 1990).

4.1 Isolation and analysis of cosmids from the sorted chromosome 6 library.
Duplicate library membranes were hybridised, as described in Chapter 2, section 2.13, with $^{32}$P-$\alpha$-dCTP labelled human PGM1 cDNA as probe at low stringency. Eight clones that hybridised to the probe were identified by grid co-ordinates following autoradiography of the filters. Cosmids corresponding to each identified clone were received from the ICRF Reference Library and were initially grown from agar stabs as described in 2.2.13. The 8 clones were ICRFc109D0446, ICRFc109L1347, ICRFc109P1721, ICRFc109I238, ICRFc109C1011, ICRFc109C1847, ICRFc109C1853, ICRFc109C1849 (in this chapter and in subsequent discussion the clones will be referred to without the prefix 'ICRFc109').

DNA mini-preparations were made of each cosmid and were digested with the restriction enzymes EcoRI, HindIII and PstI in single enzyme digests, electrophoresed and Southern blotted. The blots were probed with $^{32}$P-$\alpha$-dCTP labelled human PGM1 cDNA at 55°C overnight and were washed to a stringency
of 0.2x SSC, 0.1% SDS for 20 min at 55°C. The results are shown in Fig 4.1 (clones D0446, L1347, P1721 and I238) and Fig 4.2 (clones C1011, C1847, C1853 and C1849). All the cosmid DNAs contained sites for all three enzymes. DNA fragments hybridised to PGM1 cDNA in clones D0446, L1347, P1721, I238 and C1011 and were detected after 18 hours exposure. Signals in clones C1847, C1849 and C1853 were much weaker and were only detected following 5 day exposures.

Close examination of the 3 restriction enzyme digest patterns for clones D0446, P1721 and C1011 revealed that there were bands in common. Similar comparison of the autoradiographs showed that the smallest single hybridising fragment in each digest, a 2kb PstI fragment, was the same in each clone. This is illustrated in Fig.4.3, and indicates that clones D0446, P1721 and C1011 are overlapping cosmids. Clone P1721 was used to represent the three cosmids in further analysis.

Maxi-DNA preparations was made of cosmids L1347, P1721, I238, C1847, C1853 and C1849 and used in fluorescent in situ hybridisation (FISH) to determine their chromosome localisations. This work was carried out by Dr. Margaret Fox (UCL, London) and the hybridisations were all made to chromosome spreads from blood from a normal 46XX female. The FISH results are illustrated in Fig.4.4, Fig.4.5 and Fig.4.6. Table 4.1 summarises the localisations determined for each of the cosmids tested: P1721, C1853 and C1849 are localised to 6p21.3; C1847 to 6p23-24; I238 to 6q24 and L1347 to 9p11,9q12 (this is in the centromeric region of chromosome 9). This result also shows that chromosome 9 is a contaminant of the chromosome 6 library. Previous experience indicates that the library is likely to be around 80% chromosome 6 (D. Nizetic, personal communication).

There were no previous data to suggest a PGM-like gene on chromosome 9. Thus cosmid L1347 was of considerable interest and was further analysed. In addition, cosmids I238 (6q24) and P1721 (6p21.3) were chosen for further analysis as potential candidates for PGM3, although the chromosomal localisation of these was not the same as that given for PGM3 by analysis of somatic cell hybrids (6q12, Meera Khan et al, 1984).
<table>
<thead>
<tr>
<th></th>
<th>D0446</th>
<th>L1347</th>
<th>P1721</th>
<th>L238</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>E</td>
<td>H</td>
<td>P</td>
<td>M</td>
</tr>
<tr>
<td>kb</td>
<td></td>
<td></td>
<td></td>
<td>kb</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
</tbody>
</table>

Fig. 4.1 A: Restriction enzyme digests of cosmids D0446, L1347, P1721 and L238 with EcoRI (E), HindIII (H) and PstI (P) enzymes. M = DNA size marker.

B: Autoradiograph of Southern blot of gel shown in A after hybridisation to PGM1 cDNA.
Fig. 4.2 A: Restriction enzyme digests of cosmids C1011, C1847, C1853 and C1849 with EcoRI (E), HindIII (H) and PstI (P) enzymes. M=DNA size marker.

B: Autoradiograph of Southern blot of gel shown in A after hybridisation to PGM1 cDNA.
Fig. 4.3  *PstI* digests of cosmids D0446, P1721 and C1011 (A) paired with autoradiographs of Southern blots of the digests following hybridisation to human PGM1 cDNA (B). M=DNA size marker. *PstI* fragments in common between the three digests are indicated by arrowed lines (sizes are in kb). The 2.0 kb band was the only hybridising fragment in each digest (indicated with bold arrow).
Fig. 4.4 Fluorescent *in situ* hybridisation of cosmids L1347 (Panel A) and P1721 (Panel B) to chromosome spreads from blood of a normal 46XX female.
Fig. 4.5 Fluorescent *In Situ* Hybridisation of cosmids I238 (Panel A) and C1847 (Panel B) to chromosome spreads from blood of a normal 46XX female.
Fig. 4.6 Fluorescent *in situ* hybridisation of cosmids C1853 (Panel A) and C1849 (Panel B) to chromosome spreads from blood of a normal 46XX female.
### Table 4.1 Summary of the chromosome localisations of cosmids L1347, P1721, I238, C1847, C1853 and C1849 determined by fluorescent in situ hybridisation.

<table>
<thead>
<tr>
<th>Cosmid tested</th>
<th>Chromosome localisation determined by FISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1347</td>
<td>9p11, 9q12</td>
</tr>
<tr>
<td>P1721</td>
<td>6p21.3</td>
</tr>
<tr>
<td>I238</td>
<td>6q24</td>
</tr>
<tr>
<td>C1847</td>
<td>6p23-24</td>
</tr>
<tr>
<td>C1853</td>
<td>6p21.3</td>
</tr>
<tr>
<td>C1849</td>
<td>6p21.3</td>
</tr>
</tbody>
</table>
4.1.1 Sequence analysis of cosmid P1721 (6p23-24)

A 2kb PstI fragment of cosmid P1721 hybridised to the PGM1 cDNA probe (Fig.4.1). This band was isolated from a large scale PstI digest of P1721 DNA by gel purification and subcloned into M13mp19. A maxi-DNA preparation of the M13 recombinant was digested with polylinker restriction enzymes, SphI, HindI, XbaI and BamHI. Of these, only SphI and HindI were found to cut the insert and the smallest fragment of the PstI 2kb insert which hybridised to the probe was found to be a 500bp SphI/HindI fragment. This fragment was subcloned into M13mp18 and mp19.

Sequence data obtained from both subclones were overlapped and the fragment was determined to be 489bp in length (Fig.4.7A). A BESTFIT computer comparison of the sequence to the PGM1 cDNA sequence revealed 74% nucleotide identity to a 73bp region of PGM1 exon 5, from nt 792 to nt 864, (indicated in Fig.4.7B). However the sequence does not share the same open reading frame as PGM1 in this region, which lies upstream of the magnesium binding loop. Moreover, the nucleotide homology did not extend to the junctions of exon 5 with introns 4 and 6, and therefore it could not be determined whether the sequence had conserved PGM1 exon/intron boundaries.

The nucleotide sequence was used in a BLAST database search of protein sequences (Swissprot). The sequence, between nt 1 and 152, was found to show high homology (82%) to human Alu sequences. No other sequences with high homology to the P1721 sequence were found.

4.1.2 Sequence analysis of cosmid l238

Southern blot analysis of an EcoRI digest of cosmid l238 revealed a 3.4kb fragment which hybridised to the PGM1 cDNA probe (Fig.4.1). This 3.4kb fragment was further digested with a range of polylinker enzymes. SacI digestion gave a 1.1kb band which hybridised to the PGM1 cDNA probe and a 2.2kb band which did not. The 1.1kb SacI/EcoRI fragment was subcloned into M13mp18 and sequenced using the M13 universal primer and three internal primers; the full length of the fragment was found to be 1070 bp. An alignment to PGM1 cDNA sequence showed regions of 60% homology to exon 2 (27 bp) and
A

1  caccatgccc ggctaatattt tgtatatttt agtaaagatg gggttttact
51  atgttgccca ggctggtctc gaactcctga cctgaggttg cccaccccacc
101  ttggcctccc aaagtactga gattacaggc atgagccacc gtgcctggtct
151  actaggcagt gacttaatga actttgagge caccacccctg ataccaaccc
201  cccctcccttt agggctttgc actacotaat atccatcttc ttccttatgc
251  atctttgctat ctgtctccct cctgagaaca atgctcagc tggacaggga
301  ttttttcttc ttcacctccc tgcctgtcttc ctgtgaccca gtagggtgtc
351  tggcatgtag taggtgctca ataaacattt gtggaggaac tgaatgaatg
401  aatgaatgaa tgatagtctc ttctgcacag ccctgtgcaa ggcagcgcca
451  taacctgcct ttaggataa ggaaactgtg gcacagggt

B

144  cctggctactaggcagttagta.........cttaatgaacttttgaggcca 182
    ||||| ||| |||||| | || |||| || |||||
792  CCTGCGAACTCGGCAGTTAACTGCGTTCCTCTGGAGGACTTTGGAGGCA 841

183  cccaccctgtacatccaccccccccctt 205
    ||||| ||| ||||| || | |
842  CCACCCTGACCCCAACCTCCTCCTCACCT 864

Fig. 4.7  A Sequence of 489 bp Sphl/Hindll fragment from cosmid P1721.  
B BESTFIT sequence comparison of P1721 sequence (nt 144 to nt 205) to exon 5 of PGM1 cDNA (nt 792 to nt 864).
exon 8 (32 bp), with the region homologous to exon 2 within a 56 residue open reading frame, the longest found in the fragment sequenced (Fig.4.8). A sequence data base search showed a 45 bp region of I238 with 82% homology to a sequence tagged site on human chromosome 21 (D21s321). This region lies in an open reading frame in I238, but the significance of the homology to the STS is unclear as its function is not known.

4.1.3 Sequence analysis of cosmid L1347 (9p11, 9q12)

The Southern blot of the HindIII digest of cosmid L1347 revealed a 900bp fragment and a 4.6kb fragment which both hybridised to the PGM1 cDNA probe (Fig.4.1). The 900bp fragment was cut out and purified and subcloned into cut M13mp18. Single stranded DNA was sequenced using the M13 '40' primer. Of 5 recombinants sequenced, 3 were in one orientation (designated orientation 'a') and 2 in the other ('b') with respect to the vector, allowing a total of 316bp to be sequenced directly from the end of 'a' subclones and 366bp from the end of 'b' subclones. A 20mer sequencing primer, BF1, was designed from subclone 'a' sequence allowing a further 279bp to be sequenced: the full sequence obtained overlapped in both directions by 59bp. The full length sequence, 887 bp, is shown in Fig.4.9.

A BESTFIT computer comparison with human PGM1 cDNA revealed 74% nucleotide identity across a 191bp region. When a GAP computer comparison was used to fit all of the 887 bp to the PGM1 cDNA it revealed that high homology to the PGM1 cDNA was restricted to the region identified by BESTFIT, with low homology either side; Fig.4.10 shows part of the BESTFIT analysis. The region of high homology was equivalent to exon 5 of PGM1 cDNA and contained the same open reading frame opening and closing in the same codon phases, phase I (5') and phase 0 (3') respectively. The sequence was found to have 79% identity and 90% similarity to exon 5 at the amino acid level (Fig4.11). The 3' intron splice site was found to contain the consensus AG. The 5' splice site was found to contain AT instead of the recognised GT in positions +1 and +2 of the consensus. A survey of 3700 5' splice sites (Shapiro and Senepathy, 1987) found that 9 did not conform to the GT consensus at positions +1 and +2, but none had A in position +1. Although position +1 of this sequence does not conform to the
Fig. 4.8 450 bp of I238 sequence, with the regions of 60% homology to exons 2 and 8 of PGM1 cDNA indicated in bold type. The 56 residue open reading frame is underlined.
Fig. 4.9 Sequence of 900 bp HindIII fragment from cosmid L1347.
The internal sequencing primer, BF1, is underlined.
Fig. 4.10 GAP computer comparison of 588 bp of L1347 sequence (lower case) and human PGM1 cDNA from nt 551 to nt 1128 (upper case). Exon 5 of PGM1 (bold type) is boxed indicating the region of 74% identity between the 0.9 kb HindIII fragment of L1347 and PGM1 exon 5.
L1347 : 329 VMGPYVRKVLCDLEGPPANSAINCVPLEDFGG 360
V+GPYV+K+LC+ELG PANSA+NCVPLEDFGG
hu PGM1 : 227 VVGOPYVKILCEELGAPANSANCVPLEDFGG 258

L1347 : 361 QHPDPNLTYAMLLEAMKGEYGFAGAFAFDADG 392
HPDPNLTYA L+E MK GE+ FGAAFPDADG
hu PGM1 : 259 HHPDPNLTYAADLVEITMKSGEHDFGAAFDGDG 290

Fig. 4.11 A comparison of amino acid sequence in the region of high nucleotide homology between L1347 and exon 5 of human PGM1 (residues 227-290). Identical residues in both sequences are shown in bold type. Non-identical but homologous residues are shown by +.
consensus, 6 out of the eight nucleotide positions identified in the 5' splice site do conform to the consensus sequence.

A comparison of sequence at the 3' end of PGM1 intron 4 with the equivalent sequence upstream of L1347 coding sequence showed 21% identity over 167bp, and of sequence at the 5' end of PGM1 intron 5 with sequence downstream of the L1347 coding sequence showed 26% identity over 290bp (intron sequences are unpublished data, supplied by Edwards and Putt).

The 822bp of L1347 sequence was used in a BLAST computer database search of both cDNA (EMBL database) and protein (Swissprot) sequences. Both searches showed that phosphoglucomutase 1 was the closest match, providing further evidence that this sequence represents part of a PGM1-like cDNA and that the level of conservation of sequence is such that translation of this sequence produces a peptide highly homologous to part of the PGM1 protein.

To confirm the chromosome assignment of cosmid L1347, PCR primers were designed from sequence flanking the coding sequence (Fig.4.12A) and used in amplifications of DNAs of known chromosomal origin: cell line GM10611 (human chromosome 9 /rodent hybrid), cell line MCP6 (human chromosome 6 /rodent hybrid), a cosmid with an insert from chromosome 6q27, hamster genomic DNA and human genomic DNA. A product was amplified from DNA from cell line GM10611, which has chromosome 9 as the only human chromosome, and from human genomic DNA (Fig.4.12B). No product was amplified from chromosome 6 DNAs nor from the hamster DNA. This sequence in cosmid L1347 is the first evidence for a PGM1-like sequence localised to chromosome 9, and may represent part of only the second PGM gene family member cDNA to be cloned.

The observation that a 4.6kb HindII fragment from cosmid L1347 also hybridised to the PGM1 cDNA probe suggests that more exon sequence is contained within this fragment. As an initial step towards characterisation of this fragment, it was gel purified from a HindII digest of L1347 and subcloned into

118
A

1  ttaatacatg  aaaaagcaca  ggcgaatgaat  cattattatg  tagctgattt
51  ctttcggca  aacttttaga  aaatttcag  gtgaaaaaat  gtatgtaact
101  agatgattga  atacatcttc  tgttatcc  aacacgtaac  tggcacatag
151  tagatgcctca  ataaatagta  gttcaagcaaa  atattgttct  atattgtatc
201  tttaaaacga  ttgtaaattt  atatatttctt  tctgggttat  taatcgggtg
251  tttccgtgtc  tattcaagac  aaataactct  gggagaagg  tctctgaat
301  attaaatatt  gctgtgtatcc  tatatttagc  tatgggacct  tatgtgagaa
351  aagttctgtg  tgatgagctg  gggccccccag  ccaattctgc  aataaactgt
401  gttccttctg  aagacctttgg  agggcagcac  cctgacccca  acctgacata
451  tgcaatgact  cttcttgaag  caatgaaagg  aggagaatat  gattttggag
501  ctgcatttta  tgctgtatggg  ataagttgga  aagctctctg  tcttgctgacc
551  ctttgatcat  atatagtgacc  atgcacttga  aagcagagaa  ttaatgaaca
601  catctggagc  taacatgtat  gggcagtgtg  gccttaactg  gctctgcatc
651  tgcccttctg  ttaatgact  aagttttgca  aattaaatgc  gattacttttt
701  ctttagttccaa  gatattacatt  aagggatatg  aagttttttt  attttatat

B

Ro  Chr9  Chr6  hu  Cos6  C  M

Fig. 4.12
A: 750 bp of L1347 sequence with the PCR primers (L1347F and L1347R) for amplification of genomic samples indicated in bold type, and the region of high sequence homology to PGM1 exon 5 boxed.
B: Ethidium bromide gel showing L1347F/L1347R PCR amplifications of DNA from hamster (Ro), GM10611 cell line (Chr9), MCP6 cell line (Chr6), human (hu) and a cosmid with an insert from 6q27. C= no-DNA control. M= DNA size marker.
M13mp18. At the time of writing, restriction enzyme mapping and sequencing of this region are in progress.
Chapter Five

PGM4

The possibility of a fourth gene, distinct from PGM1, PGM2 and PGM3 and determining a phosphoglucomutase isoform, was raised by an electrophoretic study of phosphoglucomutase in human milk (Cantu and Ibarra, 1982). This study detected novel isozymes in milk and showed that while the PGM activity in non lactating mammary gland homogenates was due to PGM1 and PGM2 these isozymes were not present in milk. The authors proposed a reversible switch in gene expression from PGM1 and PGM2 to PGM4, taking place when lactation was initiated.

Another possible and alternative explanation of the novel milk PGM isozymes is that these isozymes are encoded by PGM1 and arise either by modification of PGM1 mRNA or of its protein product. Some supporting evidence for a close structural relationship between PGM1 and PGM4 isozymes comes from immunological analysis using anti-rabbit muscle PGM antiserum. The antiserum shows immunoreactivity with both PGM1 and PGM4 but does not cross react with PGM2 and PGM3 isozymes, (Drago et al, 1991). The antigenic similarity of PGM1 and PGM4 suggests that they have protein domains in common.

It seemed reasonable to make use of the antigenic similarity of PGM1 and PGM4 to search for PGM4 cDNA by screening expression cDNA libraries using the anti-human PGM1 polyclonal antibodies and the anti-rabbit PGM antibody. The human PGM1 cDNA was also used as a probe to exploit the possible nucleotide homology between PGM1 and PGM4. PGM4 cDNA was searched for in both human and mouse cDNA libraries since electrophoresis and specific activity staining of both rat and mouse milk samples show PGM bands which correspond to the human PGM4 bands (Drago, PhD Thesis, 1990)

5.1 Isolation and analysis of cDNAs from Mammary tissue cDNA libraries

A human breast cDNA library in λgt11 (Clontech), and a mouse mammary gland cDNA library (from C. Watson) in Uni-ZAP XR (Stratagene) were used in
screening for PGM4. The human library mRNA source was breast tissue, excised during mastectomy in the eighth month of pregnancy, which was well-differentiated and showed lactational competence. The mouse mammary gland library mRNA source was 11-day lactating tissue. The number of independent recombinants were $1.2 \times 10^6$ for the mouse library and $1.6 \times 10^6$ for the human library.

5.1.1 A search for PGM4 cDNA using antibody probes

Three anti-PGM antibodies were available for use as probes. The anti-rabbit muscle PGM antibodies produced by immunising a sheep with a purified preparation of rabbit skeletal muscle PGM (Whitehouse et al, 1989) and two anti-human PGM1 polyclonal antibodies, designated 6' and 10', produced by immunising a sheep with fusion protein fragments of PGM1 (Edwards, Lovegrove and Whitehouse, unpublished data). The 6' and 10' constructs extend from amino acid residue 448 to 561 (domain IV of the PGM1 protein) and from residue 156 to 561 (domains IV, III, II and 31 residues at the carboxyl end of domain I), respectively as determined from 3D protein crystallography studies by Dai et al (1992) (Fig.5.1). The anti-human PGM and anti-rabbit PGM antibodies were used to screen both the human and mouse libraries. Cross-reactivity between both the human 6' antibodies and the anti-rabbit antibodies and mouse PGM has been demonstrated by activity staining following IEF separation (Whitehouse, unpublished data).

At each primary screen of the libraries between $1.8 \times 10^5$ and $6 \times 10^5$ pfu were examined using the anti-PGM1 antibodies. Binding of the anti-PGM1 antibody was detected with a peroxidase conjugated rabbit anti-goat second antibody. Plugs containing positive plaques were picked and independently subjected to successive rounds of screening at decreasing plaque densities to purify each positive recombinant. A total of $1.08 \times 10^6$ pfu were screened from the human library (67.5% of the total number of independent clones). While $3.0 \times 10^6$ plaques were screened from the mouse library (2.5 times the total number of recombinants).
Fig. 5.1
Illustration of the relative positions of PGM1 exons and protein domains (bold print) (Dai et al., 1992). The dashed horizontal lines represent the regions of the protein expressed from the 6' and 10' pEX vector constructs which were used to raise anti-PGM1 antibodies (Edwards and Whitehouse, unpublished).
After screening the mouse mammary gland library with the anti-human PGM1 6' antibodies and the anti-rabbit PGM antibodies, 28 potential positive recombinants were isolated. However, these were all false positives as none tested positive at the secondary screen. Similarly, a number of false positives, 19, were identified after screening the human breast tissue library with the anti-human PGM1 antibodies and the anti-rabbit PGM antibodies. However, five positives selected after screening with the anti-human PGM1 10' antibodies were immunopositive after four rounds of screening and were purified to homogeneity. Table 5.1 summarises the screening results of the human and mouse libraries with the three antibodies. The failure to detect mouse PGM positive recombinants may be because by chance none of the cDNAs are in-frame in the fusion proteins (only 1 in 6 cDNAs will be in the correct reading frame) and/or the affinity of human and rabbit antibodies for mouse PGM protein may not be adequate for library screening.

PCR was used to amplify the cDNA inserts of the 5 human PGM1-like recombinant clones (MPL-1, -4, -5, -6, -7) for further analysis. Purified phage stocks were amplified to titres of between $7 \times 10^{10}$ and $1.6 \times 10^{11}$ pfu ml$^{-1}$ and used in amplifications with primers which flank the EcoRI cloning site, within the Lac Z gene in $\lambda gt11$ (Fig.5.2). The sizes of the amplification products of MPL-1, -5, -4, -6 and -7 (shown in Fig.5.3A) were 340bp, 440bp, 400bp, 450bp and 400bp respectively (the region between the $\lambda gt11$ primers is 85bp). Autoradiography following Southern blot analysis of these amplification products showed no hybridisation to the PGM1 cDNA probe (Fig.5.3B).

Clones MPL-1 and MPL-5 were chosen for sequencing. The PCR products were purified after electrophoresis in low melting point agarose gels. The products were sequenced using the $\lambda gt11$ forward and reverse primers. Sequence data using both primers was overlapped to ensure that the whole of the recombinant insert was sequenced. When sequenced, MPL-1 was found to be 252 bp long with one open reading frame of 26 amino acid residues (Fig.5.4A). In a BLAST search of the EMBL database a region of 71% identity
<table>
<thead>
<tr>
<th>Library</th>
<th>Antibody probe</th>
<th>No. pfu screened</th>
<th>No. false positives</th>
<th>No. PGM1 positive recombinants</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN</td>
<td>anti-human PGM1 6'</td>
<td>3.6x10^5</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>anti-human PGM1 10'</td>
<td>3.6x10^5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>anti-rabbit PGM</td>
<td>3.6x10^5</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>MOUSE</td>
<td>anti-human PGM1 6'</td>
<td>1.2x10^6</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>anti-human PGM1 10'</td>
<td>nt</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td></td>
<td>anti-rabbit PGM</td>
<td>1.8x10^6</td>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.1 Summary of the screening results of the human and mouse mammary tissue libraries with anti-rabbit PGM and anti-human PGM1 6' and 10' antibodies. False positives are those only detected at the primary screen. PGM1 positive recombinants are those which remained immunopositive after four rounds of screening and purification to homogeneity. nt is not tested.
Fig. 5.2
A: Position of \( \lambda g t11 \) forward and reverse primers relative to LacZ gene and \( \lambda g t11 \) left and right arms (each arm is 43.7kb). The EcoRI cloning site is indicated.
B: Sequence of \( \lambda g t11 \) across the EcoRI cloning site. Forward and reverse primers are boxed: the \( \lambda g t11 \) forward primer is 5' GGTGGCGACGACTCTGGAGCCCG 3'; the sequence of the \( \lambda g t11 \) reverse primer is 5' TTGACACCAGACCAACTGGAATG 3'. The amplification product, with no insert, is 85bp.
Fig. 5.3 Panel A: PCR amplification products of clones 1, 4, 5, 6 and 7. C, a positive control for the PGM1 probe, is a 308bp PCR product spanning nt 1909 to 2217 of PGM1 cDNA (numbering as Whitehouse et al., 1992). M is a DNA size marker (indicated in base pairs (bp)).

Panel B: Autoradiograph following Southern blot of gel in A with PGM1 cDNA as probe; lanes are as described for Panel A.
Chapter 5

Fig. 5.4  A: Sequence from recombinant clone MPL-1. The underlined region indicates the 26 amino acid residue open reading frame.  
           B: Sequence from recombinant clone MPL-5.
was found between a 44 bp region of MPL-1 and intron 3 of the human matrix Gla protein (MGP) gene, however this region did not coincide with the 26 residue open reading frame and no significant homology was identified between MPL-1 and known coding sequences. BESTFIT and GAP computer comparisons of MPL-1 and PGM1 sequences found no significant homology. Presumably MPL-1 was identified by the PGM1 antibodies because of a small shared epitope which is not obviously seen in sequence analysis.

The sequence of clone MPL-5 is 359 bp long (Fig.5.4B). An open reading frame of 119 amino acid residues was found. The sequence was used in a BLAST search of the EMBL database. Out of 25 of the closest identity matches, four matches were found to beta-galactosidase sequence and 17 matches to cloning vectors; all matches were between 97% and 100% identity and were found across the entire sequence of the clone. MPL-5 was immunopositive with the anti-PGM1 10' antibodies which, as described above, was raised using a fusion protein between the PGM1 10' region and the beta-galactosidase sequence of the cloning vector, hence the presence of anti-beta-galactosidase activity in these antibodies. It is of interest, therefore that only one recombinant was detected with the anti-beta-galactosidase activity as 1 in 6 recombinants should express Lac Z. A possible explanation is that the beta-galactosidase peptide in the fusion protein is usually unstable, and that MPL-5 is formed from a rearrangement of vector sequences which renders it more stable.

5.1.2 Isolation using Human PGM1 cDNA as probe

As an alternative approach to the search for the PGM4 cDNA, libraries were screened with the PGM1 cDNA probe. This approach seemed reasonable since the homology in protein sequence which leads to immuno-crossreactivity of PGM1 and PGM4 is likely to extend to their DNA sequences and might allow the identification of recombinant clones containing PGM4 cDNA inserts using a PGM1 cDNA probe.

The same two mammary gland libraries described above were screened. Duplicate filters containing 2-6x10^5 pfu were hybridised with ^32P radiolabelled
PGM1 cDNA as probe at 55°C then washed to a moderate stringency of 0.2xSSC at 55°C for 20 min. From the autoradiographs of the filters, positive plaques were identified and corresponding plugs picked and independently subjected to successive rounds of screening at decreasing plaque densities to purify each one or until a false positive was identified.

Table 5.2 summarises the results from these analyses. A total of \(7 \times 10^5\) pfu were screened from the human library, equivalent to 44% of the number of independent clones in the library. From the mouse library \(9 \times 10^5\) pfu were screened which is 75% of the number of cloning events in the library. Seven recombinants, five from the human library and two from the mouse library, were purified to homogeneity.

PCR was used to approximately size the inserts of the positive recombinants and to make DNA templates for sequencing. Amplified phage stock with titres between \(5 \times 10^{10}\) and \(1.5 \times 10^{11}\) pfu ml\(^{-1}\) were used in PCR amplifications with either the pBluescript primers (pBluescript is inserted in lambda sequences in Uni-ZAP XR) for mouse recombinants or \(\lambda\)gt11 PCR primers which flank the cloning site (described in section 5.1.1) for human recombinants. Table 5.3 shows the insert sizes of each of the 5 human (H1 to H5) and 2 mouse clones (M1 and M2) The sizes of the human clone inserts range from approximately 950 bp to 1750 bp. The mouse clone inserts were estimated at 2800 bp and 1400 bp.

5.1.2.1 Analysis of human recombinants

The PGM1 positive human cDNA clone PCR products were Southern blotted and probed with \(^{32}\)P labelled full length human PGM1 cDNA and washed to a relatively high stringency of 0.1xSSC, 65°C for 20 min. All 5 products hybridised to the probe indicating a high degree of sequence homology between the cDNA inserts and PGM1 cDNA. To further analyse the clones, the products were digested with \(Rsa\) and \(Hind\), restriction enzymes known to cut PGM1 cDNA. Fig.5.5A shows the digested PCR products on an ethidium bromide stained agarose gel: all 5 clones were cut by both enzymes. The gel was Southern blotted and probed with a \(^{32}\)P labelled \(SphI/PstI\) fragment which spans from nt 1060 to nt 1710 of the coding sequence of PGM1 cDNA. This probe was chosen because the cDNA inserts of the positive recombinants were shorter than
<table>
<thead>
<tr>
<th>Library</th>
<th>No. pfu screened</th>
<th>No. false positives</th>
<th>No. positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human library</td>
<td>$7 \times 10^5$</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Mouse library</td>
<td>$9 \times 10^5$</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.2 Summary of screening of human and mouse libraries with PGM1 cDNA as probe, indicating number of pfu screened, false positives identified at primary screen and positives screened to homogeneity.
<table>
<thead>
<tr>
<th>Clone</th>
<th>Size of cDNA inserts (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>950</td>
</tr>
<tr>
<td>H2</td>
<td>1750</td>
</tr>
<tr>
<td>H3</td>
<td>1500</td>
</tr>
<tr>
<td>H4</td>
<td>1600</td>
</tr>
<tr>
<td>H5</td>
<td>1050</td>
</tr>
<tr>
<td>M1</td>
<td>2800</td>
</tr>
<tr>
<td>M2</td>
<td>1400</td>
</tr>
</tbody>
</table>

Table 5.3 Summary of sizes of cDNA inserts, estimated by PCR amplification using vector primers, of the human breast tissue library recombinant clones H1, H2, H3, H4 and H5, and the mouse mammary gland library recombinant clones M1 and M2.
Fig. 5.5

Panel A: PCR products of clones H1, H2, H4, H5 and H3 cut with HindIII and Rsal on an ethidium bromide stained agarose gel. M is DNA size marker (sizes indicated in kb).

Panel B: Autoradiograph of Southern blot of agarose gel shown in Panel A, following hybridisation to the 3' PGM1 cDNA probe (SphI/PstI fragment spanning nt 1060 to nt 1710). Sizes (in kb) to the left and right of the panel correspond to hybridising HindIII fragments and Rsal fragments respectively.
PGM1 cDNA, but would include 3' sequences since cDNA was synthesised following oligo dT priming. The blot was washed to a moderate stringency of 0.2xSSC at 55°C for 20 min. The autoradiograph is shown in Fig.5.5B. All the cDNAs, except H3, hybridised to the PGM1 3' probe. The signal was confined to a single band in each digest of each cDNA: from the HindIII digests a 1.1kb band in H2 and H4, a 0.6kb band in H1 and a 0.8kb band in H5; from the Rsal digests a 1.3kb band in H2 and H4, a 0.35kb band in H1 and a 0.55kb band in H5.

HindIII and Rsal restriction maps of clones H1, H2, H4 and H5, and of PGM1 cDNA are illustrated in Fig.5.6. It can be seen that the maps of the 3' ends of the clones H1, H2, H4 and H5 appear to be alike and that the clones are all shortened versions of PGM1 cDNA synthesised from the 3' end. The restriction map of clone H3, however, could not be constructed on the basis of this data as there was no hybridisation to the PGM1 3' probe.

PCR products from positive recombinants were also sequenced with the λgt11 reverse primer to establish the extent of the 5' end of each clone. The 5' end was determined in each case by identifying the junction between vector and insert sequence: clone H1 extends to nt 1334, H2 to nt 437, H4 to nt 437 and H5 to nt 1143 (illustrated in Fig.5.7A, numbering as for human PGM1 in Whitehouse et al, 1992). The 3' ends of the PCR products from clones H1, H2, H4 and H5 were sequenced to determine whether the 3' untranslated regions of these clones were the same as that of the PGM1 cDNA. The sequencing primer used was from the 3' untranslated region of PGM1 and spans from nt 1909 to nt 1933 (illustrated in Fig.5.7B). Sequence data across a region of 200 bp from clones H1, H2, H4 and H5 revealed that all were identical to each other and to the same region in the 3' untranslated region of PGM1 (from nt 1954 to nt 2154). This sequence analysis confirmed that clones H1, H2, H4 and H5 represent shortened versions of PGM1 cDNA.

The restriction map of clone H3 did not match those of the other clones, nor did it match that of the PGM1 cDNA. Furthermore, after washes of moderate stringency, H3 did not hybridise with the SphI/PstI 3' probe. The H3 PCR product was sequenced using the λgt11 reverse primer to characterise the 5' end of the insert. The sequence obtained was identical to 32 bp of intron 1 of PGM1 (Edwards and Putt, unpublished) and the 5' end of exon 2 with the intron/exon
Fig. 5.6 Aligned restriction enzyme maps of PGM1 cDNA and clones H1, H2, H4, and H5, showing the positions of HindIII (H) and Rsal (R) restriction sites. Sizes of HindIII fragments are shown by arrowed lines and in bp. Scale is 1cm=200bp.
Fig. 5.7

A: Illustration of the extent of clones H1, H2, H4 and H5 with the most 5' nucleotide indicated in each case.

B: Position of 3' sequencing primer (arrowed line) in 3' untranslated region of PGM1 (boxed region), the sequence of the primer is shown below.
Fig. 5.8
Part of PGM1 drawn to scale (1cm=200bp) showing exons 2 to 4 with intervening sequences, the proposed structure of H3 is shown beneath. Exons are represented by boxes, arrows represent the sequence primers and the bold horizontal lines represent the regions sequenced.
junction identical to that defined by Putt et al., 1993 (illustrated in Fig. 5.8). This implied that the cDNA insert in H3 represented unspliced mRNA. To further characterise H3, the PCR product was sequenced using a primer, 2F, which spans from nt 374 to nt 396 in exon 2 of PGM1 cDNA. The sequence obtained with this primer overlapped that obtained with the λgt11 reverse primer and revealed a complete exon 2, the exon2/ intron 2 junction and 88 bp of sequence confirmed to be identical to intron 2 (Edwards and Putt, unpublished) (illustrated in Fig. 5.8).

The exon 2, intron 1 and intron 2 sequence obtained from clone H3 accounted for 284 bp. The insert was estimated at 1415 bp in length (Table 5.3), therefore to further define the nature of the remaining 1131 bp, the clone was sequenced with primers from exon 3 and exon5. Sequencing with the exon 3 primer 3F, positioned from nt 465 to 492 of the PGM1 cDNA, revealed that H3 contained at least 115 bp of exon 3 and 133 bp of intron 3 (Fig. 5.8). The exon 5 primer 5F, positioned from nt 918 to nt 939 of the PGM1 cDNA, did not prime sequence from clone H3.

From the sequence data clone H3 appeared to be a cDNA transcribed from an unspliced message during the construction of the human cDNA library. It appears to comprise part of intron 1, exon 2, intron 2, exon 3 and approximately 669 bp of intron 3. The exon 5 sequencing primer does not prime from an H3 template therefore exon 5 sequences do not seem to be present in H3. Further evidence for the exon/intron structure of clone H3 comes from restriction mapping data. Clone H3 DNA was cut with PstI, HindIII and HincII restriction enzymes in single digests and in combination in double digests; the position of the sites is illustrated in Fig. 5.9. The map of PstI, Hind III and HincII sites in clone H3 matches the map of the PGM1 genomic structure in the region between intron 1 and intron 3.

5.1.2.2 Analysis of mouse recombinants

PCR of the PGM1 positive mouse cDNA clones revealed that the insert sizes are approximately 2800 bp in M1 and 1400 bp in M2 (Table 5.3). In order to characterise the cDNAs further, the pBluescript II SK- plasmids with the inserts in the multiple cloning site, were excised from the Uni-ZAP XR phage stocks. The
Fig. 5.9
Genomic structure of human PGM1 (vertical bars represent exons) with an enlarged region showing the *Pst*I (P), *Hind*III (H) and *Hind*I (Hi) restriction sites in and around exons 2 to 4 (exons are represented by solid boxes). The *Pst*I, *Hind*III and *Hind*I sites found in Clone H3 are shown aligned to human PGM1. Scale is 1 cm = 60 bp.
single-stranded DNA templates were prepared and the 5' ends of the cDNAs sequenced using the T3 sequencing primer.

Sequence extending 335 bp was obtained from the 5' end of clone M1. The 5' sequence of the mouse homologue of human PGM1 (confusingly designated *Pgm-2*) was not available for comparison at the time of writing. However, alignment of M1 sequence to human PGM1 cDNA showed that it contained sequence equivalent to the 5' 190 bp of exon 1 coding sequence (Fig.5.10) including the translation start site. Upstream of this sequence was 145 bp of sequence which corresponds to the 5' untranslated region of this cDNA. A comparison of the 43 bp of the rat liver PGM cDNA 5' untranslated region (Rivera *et al*, 1993) and 43 bp of sequence upstream of the ATG corresponding to the start of exon 1 in M1 shows 81% identity. Alignments were made between nt 146 and nt 335 of M1 sequence and the equivalent 190bp of coding sequence of rat (Rivera *et al*, 1993), human and rabbit (Whitehouse *et al*, 1992) PGM1 cDNAs (Fig.5.11), revealing 90%, 81% and 88% identities at the nucleotide level. The high homology between M1 and PGM1 cDNAs of rat, rabbit and human implies that M1 may encode mouse PGM2. The estimated length of clone M1 at 2800 bp suggests that it may be a full length mouse *Pgm 2* (equivalent to human PGM1) cDNA and this is the first report of isolation of a complete mouse *Pgm 2*.

From the shorter clone M2, 246 bp of sequence was obtained from the 5' end of the clone insert. Prof. J.M. Friedman (Rockefeller University, New York) has derived 780 bp of sequence from a mouse *Pgm 2* cDNA (unpublished) equivalent to nt 621 to nt 1401 of human PGM1 cDNA. A comparison of the M2 sequence with this mouse *Pgm 2* sequence showed complete identity over the entire 246 bp. When aligned to human PGM1 cDNA sequence the M2 sequence is found to extend from nt 783 to nt 1029, encompassing 154 bp of exon 5 and 92 bp of exon 6 (Fig.5.12).

5.2 PCR Amplification of Human cDNA library.

An alternative approach to cloning PGM4 was to use primers designed from highly conserved regions of PGM1 to amplify cDNA from the two libraries. It was reasoned that if annealing temperatures were kept low this procedure would
**Fig. 5.10** 5' sequence of mouse clone M1 (lower case) aligned to 5' untranslated and exon 1A sequence of human PGM1 (upper case). The human PGM1 translation start site is underlined, numbering is for human PGM1 cDNA (Whitehouse *et al*, 1992).
Fig. 5.11 M1 sequence (lower case) aligned to rat and rabbit PGM and to human PGM1. The translation initiation site is underlined, numbering is for human PGM1 as given in Whitehouse et al., 1992, n is sequence not determined at that position. Sequence differences are indicated by upper case letters: M1 shows 93% identity to rat PGM, 88% identity to rabbit PGM and 81% identity to human PGM1 in this region.
Fig. 5.12 Sequence at the 5' end of mouse clone M2, aligned to human PGM1 (Hu) from nt 783 to nt 1028 encompassing 154 bp of exon 5 and 92 bp of exon 6 (numbering is for human PGM1 as given in Whitehouse et al., 1992). Vertical bars indicate identical nucleotides.
detect products of genes closely related to PGM1 and unusual products which were derived by alternative splicing of PGM1 mRNA. A preliminary investigation of this approach was made.

The human breast tissue cDNA library was used as a source of appropriate cDNA templates in amplifications with the λgt11 forward primer and primers designed from regions of the PGM1 cDNA thought likely to be conserved. Two forward and one reverse primer were designed: primer Ex2 (forward) spans from nt 334 to nt 357 of the PGM1 cDNA in a region of complete identity between rabbit and human sequence; primer Ex3/4 (forward) spans from nt 465 to nt 492 in a region of conserved sequence between rabbit and human; primer Mg (reverse) spans nt 918 to nt 939 which encompasses the crucial magnesium binding region of the enzyme. In this initial study, the following primer pair combinations were used: λgt11 forward and Ex3/4; Ex2 and Mg; Ex3/4 and Mg. The positions of the primers, relative to PGM1 cDNA, and the sizes of the expected amplification products with PGM1 cDNA as template are illustrated in Fig.5.13.

Phage stock of the Human breast tissue cDNA library, at 2x10^12 pfu ml^-1 was used in each amplification and the sizes of the amplification products estimated by agarose gel electrophoresis with standard size markers. The PCR reactions with the primer combinations of λgt11 forward with Ex3/4, and Ex3/4 with Mg did not amplify any products. However, the PCR with Ex2 and Mg primers amplified 2 products: a 600 bp product (Ex2/Mg/600) which is the size expected from amplification of PGM1 cDNA with these primers (Fig.5.13), and a 300 bp product (Ex2/Mg/300).

The Ex2/Mg PCR products were Southern blotted and then hybridised to radiolabelled PGM1 cDNA at 55°C. After washes of moderate stringency (0.2XSSC, 0.1%SDS at 55°C) and autoradiography Ex2/Mg/600 gave a strong signal and Ex2/Mg/300 gave no detectable signal. Both Ex2/Mg/600 and Ex2/Mg/300, were purified after agarose gel electrophoresis and then sequenced using the Ex2 PCR primer. 105 bp of Ex2/Mg/600 were sequenced and these matched exactly the sequence from nt 399 to nt 504 of the PGM1 cDNA confirming that this product resulted from the amplification of PGM1. 199 bp of Ex2/Mg/300 were sequenced (Fig.5.14). When this sequence was aligned, using
Fig. 5.13 Position of PCR primers Ex2, Ex3/4 and Mg in PGM1 cDNA (indicated by arrowed lines) and the expected and observed sizes of amplification products using these primers to amplify the human mammary tissue library (** = no product amplified). The sequences of the primers are:

Ex2 5' ATGGAATCCTCTCCACCCCTGCTGT 3'
Ex3/4 5' AATGGAGGGTCTGTCCAGAAGCAATAA 3'
Mg 5' GATCCCCATCTCCATCAAAGGCAGC 3'
Fig. 5.14 Sequence of Ex2/Mg/300 (lower case) aligned to human PGM1 cDNA (upper case) from nt 327 to nt 525 (numbering is for human PGM1 as given in Whitehouse et al, 1992). Region of 83% identity between Ex2/Mg/300 and PGM1 is shown in bold type. Region of 75% identity between Ex2/Mg/300 and human V beta T-cell receptor gene is underlined. Vertical bars indicate identical nucleotides.
the GAP computer comparison, to the corresponding region of PGM1 cDNA 37% identity was found, within which is an 18 bp region of 83% homology. No extensive open reading frames could be identified in Ex2/Mg/300. A BLASTN computer database search with the sequence showed a 75% identity across a 68 bp region to human V beta T-cell receptor gene which includes the region of 83% homology to PGM1 cDNA. Homology did not extend outside this region. The significance of this short stretch of homology is unclear. It lies in a region corresponding to exon 2 of PGM1 and may represent an ancient evolutionary ancestor for a domain shared between PGM1 and human V beta T-cell receptor gene.

5.3 IEF analysis of milk

In their study of PGM isozymes in milk, Cantu and Ibarra (1982) found no correlation between PGM phenotypes of milk and corresponding erythrocyte samples. On the basis of a distinct PGM4 locus they distinguished eight phenotypes encoded by four alleles. The same phenotypes were identified by Kvito and Weimer in a study of colostrum and milk samples (Kvito and Weimer, 1990) but they found that some women showed different phenotypes in colostrum compared to milk. The phenotype change was always from a "homozygote", less complex electrophoretic pattern in colostrum to a "heterozygote", more complex pattern in milk. Kvito and Weimer (1990) did not compare colostrum phenotypes with corresponding erythrocyte phenotypes.

In order to investigate the milk isozyme patterns further, samples of colostrum were collected from 6 unrelated individuals (from the Obstetrics Department, University College Hospital, London) to examine PGM phenotypes. Each sample was spun briefly to separate the less dense lipids. The aqueous proteins were then separated on an isoelectric focussing gel of pH range 4-10. Placental extracts of known PGM1 phenotype were electrophoresed with the colostrum samples for comparison. The bands with PGM activity were then visualised using a PGM enzyme activity stain in an overlay directly onto the gel. Fig.5.15A shows the IEF gel with PGM stained bands: all colostrum samples stained positively for PGM activity but varied in the intensity of the stain reflecting
Fig. 5.15

A: Colostrum samples from mothers SH, MS, SB, RK, DH and HM following isoelectric focusing and staining with a PGM activity overlay for 30 min (the PGM1 phenotype, determined by SSCP analysis, of each mother is indicated in brackets). Placental extracts of PGM1 phenotypes 2+1+, 2-1- and 1+1- were focused on the same gel for comparison and are shown after staining for 15 min. Arrowed lines indicate the primary PGM1 bands in the control samples. + indicates the anode.

B: A diagrammatic representation of the stained bands shown in A.
### SSCP typing

<table>
<thead>
<tr>
<th>Mouthwash sample</th>
<th>Exon 4</th>
<th>Exon 8</th>
<th>Haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH</td>
<td>2-1</td>
<td>+</td>
<td>2+, 1+</td>
</tr>
<tr>
<td>SB</td>
<td>2-1</td>
<td>+</td>
<td>2+, 1+</td>
</tr>
<tr>
<td>RK</td>
<td>2</td>
<td>+/-</td>
<td>2+, 2-</td>
</tr>
<tr>
<td>SH</td>
<td>1</td>
<td>+</td>
<td>1+, 1+</td>
</tr>
<tr>
<td>MS</td>
<td>2</td>
<td>+</td>
<td>2+, 2+</td>
</tr>
<tr>
<td>HM</td>
<td>1</td>
<td>+</td>
<td>1+, 1+</td>
</tr>
</tbody>
</table>

Table 5.4 Summary of the SSCP analysis of exon 4 and exon 8 from DNA extracted from mouthwash samples from mothers DH, SB, RK, SH, MS and HM, and the deduced PGM1 haplotypes of these individuals.
variation in PGM activity levels between samples. Fig.5.15B shows a diagrammatic representation of the stained bands.

To examine the PGM1 phenotypes of the mothers by a less invasive procedure than collecting blood samples, saline mouthwash samples were collected. Cell material was separated from each sample and DNA extracted. PGM1 exon 4 and exon 8 were amplified by PCR and then analysed by SSCP analysis to determine the 2/1 and +/- phenotypes in each sample. The sample processing and SSCP analyses were carried out by Dr. Ruth March (March et al, 1993). The SSCP results, and the deduced haplotypes are listed for each sample in Table 5.4: among the 6 samples there were 3 homozygotes and 3 heterozygotes. The PGM1 phenotype of each mother is indicated with the IEF analysis in Fig.5.15.

In this preliminary study of colostrum samples, complex banding patterns have been detected in all samples. However despite these complexities there are some areas of identity between the mothers' PGM1 type and the colostrum banding pattern, particularly when the cathodal (lower) half of the colostrum pattern is compared with the placental PGM1 binding patterns. So for example, a general feature is that samples from individuals with the PGM1*1 phenotype display a more cathodal (i.e. more negatively charged) band than those of a PGM1*2 phenotype. This is in keeping with the migration of the PGM1 1 isozyme which is more cathodal than the 2 isozyme. Furthermore the colostrum banding pattern is most complex in individuals with a 2-1 PGM1 phenotype. If the colostrum banding patterns are due to secondary modification of PGM1 isozymes, then this modification must, in general, add to the net negative charge. Possible mechanisms are interaction with thiol groups and/or glycosylation. Analysis of more samples would be necessary to make a definite association between IEF pattern in milk and PGM1 status of the mother, however analysis of these samples would seem to support the view that PGM4 is a modified PGM1 protein. Comparison of these results with those of Kvito and Weimer (1990) could not be made because isoelectric focussing analysis was not employed in that study.
Chapter Six
Discussion

6.1 PGM1 gene structure

The focus of this particular project was the 5' end of the PGM1 gene and the data concerning this region will be discussed in the context of the entire gene. The entire genomic structure of PGM1 was elucidated in the laboratory of Dr Y. Edwards. The gene is 75kb in length and encompasses eleven exons with ten intervening sequences (Putt et al., 1993). The length of coding regions of exons vary from 87bp (exon 11) to 300bp (exon 1B). Excluding exons 1A, 1B and 11, the average length of PGM1 exons is 150bp which falls into the most frequent exon size category of approximately 140bp as compiled for a series of eukaryotic genes (Naora and Deacon, 1982).

Gilbert (1978) proposed that intervening sequences may serve to separate portions of the coding region of genes corresponding to the structural and/or functional domains for their corresponding proteins. This feature has been recognised in the structures of a number of genes and proteins including alcohol dehydrogenase (B\"a\"ndern et al., 1984), the immunoglobulins (reviewed in Tonegawa, 1983) and the related \( \alpha \)-fetoprotein and serum albumin genes (Eiferman et al., 1981). A four domain structure of PGM1 has been described (Dai et al., 1992): the amino acid residues which flank the domains are Val-1 to Asp-189 (domain I), Ser-190 to Phe-301 (domain II), Phe-302 to Gly-420 (domain III), Arg-421 to Thr-561 (domain IV). Domains I, II and III all contain an \( \alpha_3\beta_4\alpha_1 \) motif, (Dai et al., 1992) however no correlation can be identified between these domains and the intron/ exon junctions in these regions of the gene. This is in keeping with observations on many other genes such as the carbonic anhydrase II gene (Venta et al., 1985), the \( \alpha_1 \)-antritrypsin and ovalbumin genes (Leicht et al., 1982) and the gene for amylase (MacDonald et al., 1980) where no correlation can be found between the individual exons or groups of exons and structural features of the protein. An alternative model for the positioning of introns, proposed by Fyrberg (1981), is that some introns are vestiges of transposon-like elements that have been inserted into genes and become fixed and have subsequently diverged in nucleotide sequence.
An interesting feature of the PGM1 gene structure is the unusually large first intron which is 38.5kb long. Long first introns are often associated with alternative N-terminal exons, and have been described for red cell carbonic anhydrase I, c-abelson and dystrophin (Fraser et al, 1989; Bernards et al, 1987; Boyce et al, 1991). The existence of far distal promoters, and in some cases coding sequences, may be explained by proposing that such sequences have been recruited from upstream sequences and have evolved into regulatory/coding sequences. This, presumably, is preceded by evolution of the appropriate donor and acceptor splice sites making splicing to the downstream mRNA feasible. This model was proposed to account for the structure of the mouse carbonic anhydrase I (CAI) gene which has a distal erythroid-specific promoter far upstream of the colon promoter and common coding sequences (Fraser et al, 1989). This mechanism could be generally applicable to examples where no homology is seen between the distal and proximal promoters and N-terminal exons. However, this mechanism would not account for the PGM1 alternative N-terminal exons 1A and 1B which show significant nucleotide homology and probably arose by duplication, as discussed below.

Large first introns raise the questions of whether transcription is continuous and how such introns are accurately spliced over very long distances. Bernards et al (1987) suggest that the first common c-abl exon has the ability to recognise donor splices from far upstream in a long transcript. In discussion of the mouse CAI gene Fraser et al (1989) propose that transcription of genes with very large introns may be discontinuous involving the looping of chromatin thus resulting in a much smaller transcript such that donor and acceptor splice sites in the RNA transcript would be the only possible choices for splicing (Fraser et al, 1989).

6.1.1 Alternative first exons, function and significance

The studies presented in this thesis show that two alternative first exons, 1A and 1B, are found in the human PGM1 gene. The transcript which includes exon 1A encodes the known protein sequence of the ubiquitously expressed PGM1 (this is the cDNA sequence determined by Whitehouse et al, 1992). The alternative first exon, 1B, is located 6kb upstream of exon 2 and 32 kb downstream of exon 1A. The exon 1B coding sequence is highly homologous,
91% nucleotide and 94% amino acid identity, to the 5' sequence of a cDNA (one out of four) cloned from a rabbit fast muscle library (Young Sup Lee et al., 1992). A comparison of the nucleotide sequences of human exons 1A and 1B revealed that while 1B is 54 nucleotides longer than 1A, there is surprisingly high sequence homology with 58% nucleotide identity and 74% amino acid homology. The amino acid homology included complete identity in three blocks of 11, 5 and 5 contiguous residues. These similarities strongly suggest that the exons have arisen by duplication. The boundaries of the duplication event, which have not yet been established, could have encompassed only the first exon or perhaps have involved the whole gene. Support for the latter model comes from comparing distances within the gene: the distance from exon 1B to exon 11 is approximately 35kb and from exon 1A to exon 1B is approximately 32kb. One scenario is that the whole gene was duplicated with loss of functional exons 2 to 11 in the upstream copy. No indications have been found of other duplicated exons in the large first intron but this could simply reflect the ancient nature of the event and rapid mutation rate observed in sequence which has been rendered non-coding by an initial mutation.

A possible mechanism for the duplication in PGM1 is random nonhomologous chromosomal breakage and reunion which has been recognised as an initial duplication event in the human haptoglobin gene (Hp) (Smithies et al., 1962; Maeda et al., 1984). The Hp$^2$ allele contains a partial gene duplication formed as a result of a nonhomologous crossover between Hp$^{1F}$ and Hp$^{1S}$ alleles, and as such encodes an α-polypeptide subunit almost twice the size of those coded for by the Hp$^{1F}$ and Hp$^{1S}$ alleles. There was found to be no similarity between the 5' and 3' crossover points hence the categorisation as nonhomologous. Also there were no recognised special features in the crossover regions that predispose to crossovers, so the breakage was seen as random (Maeda et al., 1984). Albumin and α-fetoprotein also show evidence of intragenic gene duplication, caused by unequal crossing over in the ancestral gene, which has resulted in triplication of a single functional domain (Kioussis et al., 1981).

The occurrence of alternative first exons which are transcribed from independent promoters is relatively common. The alternative first exon in many cases is non-coding as found for example in aldolase A (Izzo et al., 1988), α1-antitrypsin (Perlino et al., 1987), carbonic anhydrase I (Fraser et al., 1989) and the
adenomatous polyposis gene (Horii et al, 1993). Cells use specialised transcriptional systems to define tissue specific genes. Independent promoters may be needed to allow the identical protein product to be expressed in distinct tissues. Differential expression is regulated by particular tissue-specific trans-acting factors required for expression. One such example is α1-antitrypsin which is expressed in a cell type specific manner in both macrophages and hepatocytes by the use of alternate first non-coding exons and promoters (Perlino et al, 1987). Another example is the human porphobilinogen gene which has an upstream housekeeping promoter and a downstream erythroid promoter (Chretien et al, 1988). The ubiquitous mRNA contains coding exons 1 to 15, whereas the erythroid specific mRNA contains exon 2 as a 5’ non-coding region and has exons 3 to 15 as coding exons. Here, two different protein products are produced which catalyse the same enzymatic function.

The PGM1 gene is unique in having two transcripts containing alternate first exons which have their origins in a duplication event and that are both coding, thus producing two very similar protein products differing at their amino terminal. Reverse transcriptase PCR studies showed that while the exon 1A transcript is expressed ubiquitously, the exon 1B transcript has only limited tissue expression.

6.1.2 Exon 1A and the upstream promoter

The most 5’ PGM1 exon (designated exon 1A) and immediate flanking DNA was found to be highly GC rich: 70% GC compared with 40% in bulk DNA with a CpG/GpC ratio of 0.79 compared with 0.10 in bulk DNA. Both these features are characteristic of Hpall tiny-fragment (HTF) islands (Bird, 1986) found most commonly in the 5’ regions of genes that are expressed ubiquitously. The prediction, from the CpG rich nature of the 5’ PGM1 sequences, that the PGM1 transcript with exon 1A would be ubiquitously expressed was demonstrated by reverse transcriptase PCR studies and complements early PGM isozyme activity analysis (Spencer, Hopkinson and Harris, 1964) indicating that PGM1 protein was found in a wide range of tissues.

The enrichment for CpG in HTF islands is due to their unmethylated state and lack of methylation is a criterion for a true HTF island. The vast majority of methylcytosine in the vertebrate genome is found in CpGs (Gruenbaum et al,
1981) and is recognised by DNA repair mechanisms and converted to thymine, hence the low frequency of CpG in bulk DNA, about one-fifth of the expected frequency in vertebrate DNA. Twelve *HpaII* (CCGG) sites in the 900bp region including and flanking exon 1 in PGM1 were identified by sequence analysis. It would be of some interest to test the methylation state of CpGs in these *HpaII* sites by digestion with the methylation-sensitive enzyme *HpaII* and with its methylation-insensitive isoschizomer *MspI* : sites which encompass unmethylated CpGs should be cut by both enzymes. Since expression studies indicate that PGM1.A is expressed ubiquitously the firm prediction is that they will prove to be unmethylated (Bird, 1986).

It has been proposed that promoters of housekeeping genes may stay free of the effects of methylation because they are protected by permanently bound factors, thus maintaining the frequency of CpG and rendering the island a stable state. The relatively large size of islands (0.5-1.5kb) compared to binding sites for trans-acting proteins, and the clustered arrangement of CpGs may suggest that multiple factors are bound and that there may be efficient co-operative binding of the factors (Bird, 1986). One such factor may be the Sp1 transcription factor whose binding site contains the CpG dinucleotide (Kadonaga *et al*, 1986). Sp1 was originally identified as a protein from HeLa cells that binds to GC boxes (GGGCGG) in the 21 bp repeat elements of SV40 and activates *in vitro* transcription from the SV40 early promoter (Dynan and Tjian, 1983). Subsequent studies showed that a variety of cellular and viral promoters were activated by Sp1 *in vitro* and although Sp1-responsive promoters usually contain multiple GC boxes, a single binding site is sufficient for a promoter to be stimulated by Sp1. Höller *et al* (1988) suggested that Sp1 may have a special role in the protection of CpG islands from methylation and in their transcriptional regulation: binding may allow constitutive transcription from the promoter. Multiple Sp1 sites have been recognised in genes with ubiquitous expression; for example there are six sites in the HTF island of human aldolase A (the proximal promoter) (Maire *et al*, 1987), and nine sites in the HTF island of carbonic anhydrase II (Shapiro *et al*, 1987). Six Sp1 transcription binding consensus elements were identified in the 900bp proximal promoter region in and around exon 1 of PGM1. However, Sp1 sites do not occur in multiple copies in all CpG islands, for example chicken skeletal-muscle actin gene has only one Sp1 binding site (Edwards, 1990), and
furthermore the Sp1 factor has been shown to bind equally well to both methylated and unmethylated DNA (Höller et al., 1988).

HTF islands have been correlated with transcriptional control regions and are typically observed in genes that encode oncoproteins, growth factors and transcription factors as well as housekeeping proteins; all are generally recognised to be expressed at low levels (Kozak, 1991). The low abundance of such proteins is consistent with the expected difficulty of translating GC-rich, highly structured leader sequences. Elevating the translation of mRNAs with such leader sequences might require the induction of RNA helicases that could scan through the structured sequences (Kozak, 1992). RNA helicases with this particular function have not yet been identified. Evidence that such mRNAs might be translationally impaired comes from the dramatic improvement in expression seen when comparing the levels of translation with that of a shortened version of the mRNA (Kozak, 1991). Examples include human insulin-like growth factor II which produces naturally two versions of mRNA (Nielsen et al., 1990), and murine metalloproteinases inhibitor where the leader sequence was experimentally truncated (Waterhouse et al., 1990). However, in these cases it is difficult to be certain that translation is more efficient due to the stabilisation of these mRNAs and stimulated transcription (Kozak, 1992). PGM1 mRNA is relatively rare as highlighted during the cloning of the cDNA where only two cDNAs were identified from 4x10^5 recombinants in a skeletal muscle cDNA library (Whitehouse et al., 1992). With the expected difficulties of translating the GC-rich PGM1 leader sequence it may be expected that PGM1 protein would be of low abundance and indeed PGM1 isozymes cannot be detected after isoelectric focussing by staining for protein (Whitehouse et al., unpublished data).

There is no TATA box or CCAAT box in the PGM1 distal promoter, features often found in promoter regions of housekeeping genes. It has been suggested that there are two different types of promoters, TATA box-containing and TATA box-lacking promoters (Kageyama et al., 1989). The high GC content and multiple Sp1 factor binding motifs found in the 5’ region of PGM1 are features found in other TATA box-lacking promoters: for example epidermal growth factor receptor gene, (Kageyama et al., 1989) and human perlecan gene, (Cohen et al., 1993) and these features may provide clues about how gene promoters such as that of PGM1 interact with RNA polymerase II. In vitro studies
have identified at least six fractions of nuclear extracts that are required in addition to RNA polymerase II for accurate and efficient transcription initiation. Of these, fraction TFIID is known to contain a sequence-specific DNA-binding activity named the TATA box-binding protein (TBP). Binding of TBP to the TATA box is thought to be the first step in the assembly of the basal initiation complex. It has been proposed that a tethering factor functions to recruit the TFIID complex to TATA box-lacking promoters (Pugh and Tijan, 1990). TBP in TFIID fractions is present as part of a large multiprotein complex with a number of TBP-associated factors (TAFs) (Pugh and Tijan, 1991). When immunoprecipitated from the TFIID fraction this TBP-TAF complex supports transcription from a TATA-less promoter suggesting that one of the TAFs functions as the tethering factor (Pugh and Tijan, 1991). The function of the tethering factor is to replace the TATA box element of the template, and it is required by genes that are activated by Sp1 so that a productive initiation complex is assembled. Sp1 appears to be required to sequester the transcription apparatus at TATA-less promoters and is then anchored to it by the tethering factor (Pugh and Tijan, 1990).

There are thought to be factors uniquely required for transcription from promoters without TATA boxes. Nuclear factor ETF (Epidermal growth factor receptor gene Transcription Factor), which binds to stretches of cytosine and less efficiently to stretches of guanine residues, has been isolated and shown to stimulate transcription from TATA box-lacking promoters (Kageyama et al, 1989). ETF does not stimulate transcription from TATA containing-promoters although potential binding sites exist in these promoter regions. It is proposed that ETF might stabilise the binding of TFIID, in particular TBP, to promoters without a TATA box helping to form a stable transcription preinitiation complex (Kageyama et al, 1989). Stretches of cytosine residues exist in the regions flanking exon 1A (Chapter 3, Fig.3.10), providing potential binding sites for ETF. The precise nature of the stabilisation of TFIID to the PGM1 promoter is not known but is likely to involve both binding of sequence specific factors, such as Sp1 and ETF, as well as protein-protein interactions such as that of the tethering factor.

Typical of many housekeeping type genes, PGM1 appears to have more than one potential transcription initiation site. A 10bp sequence motif identical to that encompassing the major transcription initiation site, is found 82bp downstream associated with a translation start site at a position equivalent to
amino acid residue 67 of the major product. It is not clear that this start site is used in vivo since the surrounding sequence does not have features considered important for strong translational efficiency (Kozak, 1987; detailed in Chapter 3, section 3.1.3) and a corresponding truncated protein has not been described. The protein which would be produced from this shorter transcript would be 66 amino acids shorter at the N terminal than the full length PGM1 protein. The 66 terminal amino acid residues include 9 basic and 5 acidic residues and their loss would result in a relatively more negatively charged protein isoform due to the net loss of 4 basic residues. On electrophoresis such a protein would migrate further towards the anode. It is of interest to note that when protein extracts are separated by starch gel electrophoresis and activity stained for PGM, then weaker PGM1 secondary isozymes which are more negatively charged than the primary band are seen. While these have in the past been attributed to chemical modification of amino acid side chains, such as deamidation, it is of some interest to speculate whether these isoforms could represent the translation products of shorter transcripts of PGM1.

6.1.3 Exon 1B and promoter region

The limited tissue distribution of PGM1.1B is in contrast to the ubiquitous expression of PGM1.1A transcript. Reverse transcriptase PCR expression studies of a range of tissues revealed that expression of PGM1.1B is predominant in adult muscle tissue, with very low levels of expression in adult gut and kidney and foetal heart. The differences in tissue specificity of the two transcripts would suggest that PGM1.1B is transcribed from a non-housekeeping type promoter. Indeed, in contrast to exon 1A, 1B does not show the hallmarks of a CpG island, with values for the G+C% (47%) and CpG/GpC ratio (0.24) both closer to the expected values of methylated vertebrate DNA, and there is a concomitant lack of both HpaII sites and consensus binding elements for the transcription factor Sp1 (Table 6.1).

Like PGM1.1A, PGM1.1B has neither a TATA box nor a CCAAT box, elements occasionally absent in tissue-specific promoters. As discussed above, for transcription initiation from TATA-less promoters, the activity of a tethering factor has been proposed which helps stabilise the transcription initiation complex (Pugh and Tijan, 1990). However, the lack of Sp1 binding sites in the 1B promoter probably precludes involvement of the Sp1 protein in transcription from
<table>
<thead>
<tr>
<th></th>
<th>1A</th>
<th>1B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G+C%</strong></td>
<td>70%</td>
<td>47%</td>
</tr>
<tr>
<td><strong>CpG/GpC dinucleotide ratio</strong></td>
<td>0.79</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Hpall sites (CCGG)</strong></td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td><strong>TATA box</strong></td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><strong>CCAAT box</strong></td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td><strong>Consensus binding element for transcription factor Sp1 (CCCGCC)</strong></td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6.1 Comparison of features of 1A and 1B promoters.
this promoter, and thus also precludes the involvement of the tethering factor
which has been proposed to anchor Sp1 to TBP within the TFIID complex. The
nuclear factor ETF (Epidermal growth factor receptor gene Transcription Factor)
may be involved in stabilisation of the transcription complex at this promoter as
there are two potential ETF binding sites in the upstream flanking region at
positions -155 and -215 (Chapter 3, Fig.3.23). However, both sites are composed
of guanine residues on the sense strand which is the weaker orientation of the
ETF binding motif (Kageyama et al, 1989).

Interestingly, and possibly reflecting the muscle predominant expression of
exon 1B, there are six copies of the E-box consensus sequence CANNTG at
positions -276 to -271 (CATGTG), -142 to -137 (CAGATG), -35 to -30 (CATTTG),
+10 to +15 (CACCTG), +22 to +27 (CAGCTG) and +287 to +292 (CAGCTG). E-
boxes are binding sites for transcription factor proteins that contain a basic helix-
loop-helix (bHLH) motif. Four related skeletal muscle-specific regulatory factors,
MyoD, myogenin, MRF4 and Myf5 are part of this bHLH family and are myogenic
determination factors. The high sequence identity in their bHLH domains
suggests that all of these factors have similar skeletal muscle transcriptional
regulatory functions (Emerson, 1993), but as yet there are insufficient data to
predict which of the factors are most likely to occupy a given DNA sequence in vivo
(Wright, 1992). These transcription factors have been cloned, and purified
protein and antibody are available. It would be of some interest to test directly by
electrophoretic mobility shift assays (EMSA) and footprinting whether these
factors bind to the E-box sequences in the PGM1.1B promoter.

Exon 1B has a relatively long 5' untranslated region of 377bp. The
sequence appears to be conserved as there is 71% homology in this region
between rabbit and human sequence. The function of the long 5' untranslated is
unclear: it is not GC rich and therefore is probably not highly structured. Long 5'
untranslated regions may facilitate translation by allowing the queuing of several
40S rRNA molecules (Kozak, 1991)

Since many muscle specific genes show fibre type specificity, for example
myosin heavy and light chains, carbonic anhydrase III and myoglobin, a
preliminary study of exon 1B expression in relation to fibre type was made. This
study used rabbit soleus muscle which is almost entirely of type 1 (slow) fibres
and extensor digitorum longus (EDL) which is mainly of type 2a and 2b (fast) fibres. Human muscle is not useful for this study since most muscles show a chequered mixture of type 1 and type 2 fibres. These comparisons indicated that exon 1B is probably not fibre type specific. It would be of interest to extend this study to the expression of exon 1B in electrically stimulated fast type muscle tissue, which if stimulated at low frequency for prolonged periods then resembles slow type muscles in both appearance and contractile properties, to observe any concurrent changes in the level of exon 1B expression.

One intriguing possibility is that PGM1.1B may be implicated in the aetiology of myopathy associated with PGM deficiency. Three cases of PGM deficiencies have been described (Chapter 1, section 1.3.2). Thomson et al (1963) reported a case of a child with PGM deficiency and excessive glycogen storage in both gastrocnemii and shortening of these muscles which caused an abnormal gait. Slight muscle weakness was also observed. The second described case (Illingworth and Brown, 1964; Brown and Brown, 1968) had excess glycogen storage in the liver and reduced PGM activity in both liver and muscle. The third described case (Sugie et al, 1988) had secondary systemic carnitine deficiency syndrome caused by PGM deficiency, with accumulation of glycogen in the muscles. Sugie et al (1988) proposed that such deficiencies may arise due to a defect in one of the PGM isoforms: PGM1, PGM2 and PGM3 isoforms are all active in muscle. Loss of the PGM 1B transcript may lead to reduced activity in muscle leaving other tissues largely unaffected and resulting in the relatively mild phenotypes so far observed. Samples from PGM deficient tissues examined for exon 1B and 1A expression may reveal what part PGM1.1B plays in muscle PGM activity. If the deficiencies are associated with a specific loss of PGM1.1B then this would raise the possibility that its function must be discrete from that of PGM1.1A, in this case the function of the PGM1.1B amino terminal sequence would seem to be critical. Exon 1B has eighteen more amino acids at the 5' terminal than exon 1A with the sequence 5' Met-Ser-Asp-Phe-Glu-Glu-Trp-Ile-Ser-Gly-Thr-Tyr-Arg-Lys-Met-Glu-Glu-Gly 3'. Many proteins involved in transmembrane translocation or secretion have 5' leaders but no homology has been found in searching the databases between this eighteen residue sequence and known protein sequences. The sequence may represent a so far undescribed membrane spanning or membrane interactive domain.
6.1.4 Intragenic duplication and the molecular basis of polymorphism

Recent studies (March et al., 1993; Takahashi and Neel, 1993) of PGM1 sequence in human populations have confirmed the prediction that the PGM1 protein polymorphism arose by a combination of point mutation and recombination (Carter et al., 1979; Takahashi et al., 1982). The studies of the genomic structure (Putt et al., 1993) allowed the positioning of the sites of the two common mutations in exons 4 (the 2/1 mutation at nt 723) and 8 (the +/- mutation at nt 1320). The work in this project (described in Chapter 3) which contributed to the study of the polymorphism, involved amplification and sequencing of exons 1A and 1B from DNA of individuals of 2+, 2-, 1+, 1- and 2+2- phenotypes. No sequence variation was found among the individuals who were all of North European origin. A mutation has been found by Takahashi and Neel (1993) however, in exon 1 (nt 265) in DNA from individuals of Japanese origin. The identification of this third mutation in the PGM1 sequence completes the sequence data which underlie the electrophoretic differences observed between PGM1 phenotypes on which the original phylogeny was based (Carter et al., 1979; Takahashi et al., 1982). The phylogeny presumes that the PGM1 1+ allele was primal as it is the most common allele found world-wide and as this sequence is the same as that of other higher primates (Carter et al., 1979). The mutation in exon 1 may have succeeded the duplication of the recombination site as exon 1B has been detected in other species (e.g. rabbit and mouse) and yet the human population with the exon 1 mutation is small and localised. This would be refuted, however, if an equivalent mutation was found in exon 1B, which would indicate that the mutation arose before the duplication: no study has yet been made of exon 1B in other ethnic groups. The three mutations and four intragenic recombination events account for the eight alleles now observed. Two sites of recombination are proposed (between exon 1 and 4, and between exons 4 and 8) and it is highly likely that these arose by duplication.

This study introduced the notion that the entire PGM1 gene has been duplicated and we have postulated that intron 1 (lying between exon 1A and 1B) represents an ancient copy of the gene. The observation that the PGM1*3 allele arises by recombination of allele 7 with allele 2 points to the presence of a recombination site within intron 1. This would be compatible with the proposed duplicated gene structure. Given that such a duplication event occurred, then its timing might be estimated by the presence or absence of exon 1B in diversified
species. Preliminary data presented here show that exon 1B is present in human, mouse and rabbit which places the duplication event prior to the divergence of primates, rodents and lagomorphs. This study could be extended by searching for exon 1A and 1B homologous sequences in non-mammals.

Future investigations of the nature of the recombination sites will involve haplotype analysis across the recombination site in order to decrease the size of the genomic region thought to contain the site. Analysis of known recombinant individuals and their families will be particularly important in pinpointing these sites.

6.2 A search for PGM3

From comparisons of gross biochemical characteristics, including subunit structure and polypeptide size, human PGM1, PGM2 and PGM3 are thought to comprise a structurally related gene family. Many other glycolytic enzymes are encoded by families of multiple loci, with member genes encoding products closely related in size and specific activity. Studies of PGM isozymes in many species have indicated that multiple PGM loci are frequent. For example at least two PGM loci have been identified in the mouse and similarities in their phenotypic expression suggest that they may possibly be structurally, functionally and evolutionally related (Shows et al, 1969). Complete nucleotide sequences of the Saccharomyces cerevisiae PGM genes show that they are closely related to one another (Boles et al, 1994). Furthermore two closely related putative protein kinase genes, YPK1 and YKR2, occur downstream of each PGM locus suggesting that the S. cerevisiae genes arose by duplication covering an extended region of DNA (Boles et al, 1994).

There are a number of mechanisms by which multiple loci can arise. The initial event leading to duplicate gene or genes from a single gene is unpredictable (Maeda and Smithies, 1986). The initial event is difficult to analyse because in most gene families it occurred so far back in evolutionary history that subsequent mutations have altered sequences leaving them uninformative about the nature of the event. Categories of initial duplication events have been identified which include random nonhomologous chromosomal breakage and
reunion (for example in the human haptoglobin gene, Smithies et al., 1962; Maeda et al., 1984), homologous recombination between repeated elements (seen in the foetal globin gene, Shen et al., 1981), RNA-mediated DNA duplications (found in some families of repeated DNA such as the Alu family, Jagadeeswaran et al., 1981), gene amplification (demonstrated in mammalian somatic cells, Bostock, 1986) and whole chromosome duplication. Once multiple copies of a gene have been established, translocation to other chromosomes may occur giving widespread chromosomal locations to members of a gene family as has been the case in the PGM gene family. Members also diverge by accumulating deletions, insertions, duplications and base-pair substitutions. This process of gene evolution may also be affected by recombinational events between the copies of genes which can lead to unequal crossover and gene conversion. Unequal crossover leads to the expansion and contraction of gene families, and within gene families differences in the number of gene copies between individuals can lead to noticeable phenotypic differences as in the case of adult α-globin (Maeda and Smithies, 1986). Additionally, unequal crossover and gene conversion maintain homogeneity in gene family members by removing some or all of the differences that have accumulated between members. Until other members of the PGM gene family are cloned we can only speculate on their evolutionary origins.

6.2.1 Initial screening for PGM3

PGM3 was localised using somatic cell hybrids to chromosome 6q12 by Meera Khan et al., (1984). The construction of a chromosome 6 library from flow-sorted chromosome material (Nizetic et al., 1994) made it possible to consider cloning the PGM3 gene by screening with the PGM1 probe. This seemed a sensible approach since data from other multigene families has shown that a high level of homology at the nucleotide level has been conserved between members.

The screening strategy allowed the isolation of eight cosmids which hybridised to the cDNA probe at low stringency. The chromosomal localisations of six of the eight cosmids were determined, and of these three (L1347, P1721 ans L238) were further analysed. As the remaining three hybridised only weakly to the cDNA probe and were not localised to 6q12, they were not analysed.
further. The isolation of recombinants containing very short homologous sequences indicates that this was a valid method for finding homologous, but non-identical sequences: in cosmid I238 regions as short as 27bp and 32bp with 60% homology to PGM1 were detected. Cosmids P1721 (localised to 6p21.3) and I238 (localised to 6q24) may represent the remains of duplications of ancient ancestral sequences shared in common with PGM1.

Cosmid L1347, localised to the centromeric region of chromosome 9, was an unexpected find in the chromosome 6 library. Contamination of flow-sorted chromosome material with chromosomes other than the intended is not unusual, 15-20% of the chromosome 6 library is estimated to be material from other chromosomes (D. Nizetic, personal communication). The sequence data obtained from L1347 shows high homology at the nucleotide level to exon 5 of PGM1, and has high amino acid homology in the same open reading frame. Exon/intron boundaries to intron 4 and 5 are maintained, although the 5' splice site does not strictly conform to the most common consensus. To date, extensive studies of PGM activity in a wide range of tissues have not revealed isozymes that could not be accounted for by PGM1, PGM2 or PGM3. Thus the cDNA, on chromosome 9, may encode a PGM-like protein whose activity is sufficiently dissimilar to those characterised such that it has not previously been recognised. Alternatively, it may be expressed in a tissue or at a developmental stage not yet examined for phosphoglucomutase activity. Cosmid L1347 contains other fragments which hybridise to the PGM1 cDNA probe. Work is in progress to sequence these regions to piece together the cDNA.

During the course of this study, Belkin _et al._ (1994) described a cytoskeletal antigen which runs as a closely spaced doublet in SDS-gels, termed the 60/63 kDa proteins. The proteins show homology at the amino-acid level to PGM1 and were detected immunologically at the adherens junctions in muscle and in nonmuscle cells, with most abundant protein levels in smooth muscle tissues. The proteins share immunological properties with PGM1 and the peptide maps show similarities, however the 60/63 kDa proteins have no phosphoglucomutase activity (Belkin _et al._, 1994). The study also showed that the 63kDa polypeptide has a broader tissue distribution than the 60kDa polypeptide which is found only in smooth and cardiac muscle cells. The authors suggested that the proteins are isoforms which arise from post-translational modification or
alternative splicing, or might be encoded by separate genes. Subsequently, studies of the 60/63 kDa proteins, termed aciculin, showed that expression in skeletal muscle was strictly dependant on the differentiation stage of skeletal muscle cells with a major increase in protein levels during late stages of prenatal and early postnatal development (Belkin and Burridge, 1994). Although found abundantly in smooth muscle tissues and to a lesser extent in striated muscle, some types of non-muscle cells were found, by immunological methods, to have small amounts of aciculin.

Subsequent to the sequencing of part of cosmid L1347, the cDNA for aciculin has been cloned (Critchley et al, in press). Interestingly, comparison of the L1347 exon 5 like sequence with the aciculin sequence showed that they were identical. Furthermore, the very close chromosome assignments of L1347 sequence to the centromeric region of chromosome 9 (this study, determined by FISH) and aciculin to 9q12.13 (determined by somatic cell hybrid analysis) support the view that they are the same gene. However, initial expression studies by RT-PCR in our laboratory (Edwards and Putt, unpublished) show that the L1347 exon 5-like sequence is found in human adult and foetal heart, liver, gut, skeletal muscle and in adult kidney (data not shown). This distribution is different to that described by Belkin and Burridge (1994) for aciculin but is closer to that reported for the 63kDa polypeptide which is a component of the 60/63kDa proteins (recorded by Belkin et al, 1994). In addition, close inspection of the FISH assignment of cosmid LI 347 (Chapter 4, Fig.4.4A) reveals two pairs of signals on chromosome 9, one pair either side of the centromere. Thus the evidence suggests that there are two closely related genes on chromosome 9, one of which is aciculin. The protein products of these genes are of similar size, 60 and 63 kDa, and both are related to \textit{PGM1}, but have differing expression patterns. These observations which indicate two copies of a PGM gene adjacent on chromosome 9 are of interest in relation to the hypothesis that in the ancestral past two adjacent copies of PGM1 existed on chromosome 1. Despite the homology to PGM1, aciculin (to be named PGM5) has no detectable phosphoglucomutase activity which may be due to the substitution of a cysteine residue instead of the asparagine residue in the active site of PGM1 (Belkin et al, 1994).
The basic strategy of screening the chromosome 6 library in order to clone PGM3, relied on the existence of significant nucleotide homology between PGM1 and PGM3. As the assignment of PGM3 to chromosome 6 is certain (Jongsma et al., 1973) and the library contains eight genome equivalents, it must be concluded that since all hybridising sequences were selected, the homology between PGM1 and PGM3 is lower than estimated. Relatively low sequence homology (35-60%) is not uncommon amongst some members of gene families. A comparable gene family is carbonic anhydrase, where members cannot be cross-identified on a Southern blot and (as is the case for PGM1 and PGM3) an antibody raised against one member does not cross-react with the other members. In this case, homology between CA I, CA II and CA III was found to be 55-59% and between CA I and CA VI was 32% (Tashian, 1989). Other strategies for cloning PGM3 would involve PCR based methods using degenerate primers designed from regions of the PGM1 cDNA which are likely to be conserved, such as the active site region and the magnesium binding domain, and using these to amplify cDNAs from tissues and cell lines which express PGM3 but not PGM1 and PGM2. Target tissues would be the K562 cell line in which PGM3 activity is considerably increased and PGM1 is almost absent (Povey et al., 1980), and fibroblast tissue, where PGM3 levels are elevated and account for 7% of the total PGM activity (McAlpine, Hopkinson and Harris, 1970).

6.3 PGM4

The evidence for the existence of a fourth PGM locus, PGM4, comes from two electrophoretic studies of milk samples (Cantu and Ibarra, 1982; Kvito and Weimer, 1990) which revealed patterns of PGM activity not immediately recognisable as PGM1, PGM2 or PGM3. The interpretation of the isozyme patterns was complicated by finding that colostrum samples were different from the milk of the same individual (Kvito and Weimer, 1990). In general, fewer bands ("homozygous patterns") were seen in the colostrum samples than in milk ("heterozygous patterns") but there was no evidence that these were less intensely stained.

The cross-reactivity of the anti-PGM1 antibodies with the PGM4 isozymes (Drago et al., 1991(b)) indicates that the two proteins share antigenic
determinants and suggests structural homology. One possibility is that the two proteins are actually translated from the same gene and that PGM4 isozymes result from alternate splicing of the PGM1 mRNA, or from post-translational modification of PGM1 protein in a cell specific manner, or a combination of both mechanisms.

A possible mechanism of post-translational modification is interaction with thiol groups. Very recently, Dr. Neil Spencer (King's College, London) has shown that PGM1 isoforms secreted by the salivary gland are modified to give complex banding patterns and in a way which is reversible with dithiothreitol, a sulphhydryl reducing agent (Spencer, in press). Saliva PGM1 has a stronger anodal band when compared to blood PGM1 of the same phenotype; this anodal band is modified by dithiothreitol treatment so that the cathodal band becomes more intense and resembles that seen in blood, thus showing that the modification to saliva PGM1 involves interactions with sulphhydryl groups. Interactions of the sulphhydryl groups of cysteine residues in isozymes with oxidised glutathione to form a mixed disulphide, has been recognised as an *in vitro* storage change in several enzymes. It produces a characteristic stepwise increase in the anodal electrophoretic mobilities of the isozymes by increasing the net negative charge by one unit for each sulphhydryl group (Harris and Hopkinson, 1976). In view of the effects of sulphhydryl reducing agents on saliva and the additional negatively charged bands identified in the colostrum samples, it would be of interest to investigate milk and colostrum, also exocrine secretions, with sulphhydryl reagents to observe the effects on PGM isozyme patterns.

A second possible mechanism of post-translational modification is glycosylation which can be the transfer of mannose-rich oligosaccharides to arginine residues (N-glycosylation) and the addition of the monosaccharide N-acetyl-galactosamine to serine or threonine residues (O-glycosylation). N-glycosylation of milk proteins is thought to take place in the endoplasmic reticulum, while O-glycosyl linkage and other modifications take place in the Golgi membranes (Neville *et al*., 1983). Post-translational modifications could account for the changes from an apparent 'homozygote' phenotype in colostrum to a 'heterozygote' in milk as protein modification may proceed at different rates in the mammary gland cells during the production of colostrum and milk.
Further supportive evidence that milk isozymes previously identified as PGM4 may be a modified form(s) of PGM1 comes from a comparison of the electrophoretic patterns of PGM activity in milk with those of blood from the same individuals (Drago, 1991(a)). In this study the author was able to demonstrate that some bands characteristic of the common blood phenotypes were held in common by the milk phenotypes. In this present study a preliminary IEF analysis of colostrum samples (described in Chapter Five) showed some association between the PGM1 phenotype of mother and PGM banding pattern in colostrum: those individuals of PGM1*1 phenotype have a cathodal band in the colostrum sample not observed in samples from mothers of PGM1*2 phenotype. This initial study lends weight to the case that PGM4 is a post-translationally modified PGM1.

While Cantu and Ibarra (1982) proposed a switch from PGM1 and PGM2 to PGM4 activity in the mammary gland, these authors only tested milk and not lactating mammary gland samples. If a separate gene locus existed for PGM4 it would be transcribed in the lactating mammary gland and PGM4 mRNA would accumulate in the alveolar cells. Thus the mouse lactating mammary gland cDNA library was judged to be a suitable resource for the isolation of PGM4 cDNA. Furthermore the mouse library was tested by screening for β-casein which was found at a frequency within the expected range (10-20%) of the total polyA+ RNA in lactating mammary gland (communicated by C.J. Watson).

Antibody screening of the mouse cDNA library did not identify PGM cDNAs. This probably reflects the fact that only 1 in 6 cDNAs are cloned in the correct orientation and reading frame and are thus able to express the PGM1 protein. More success was achieved using the human cDNA as probe at low stringency which resulted in the isolation of cDNAs, M1 and M2, which, from the available data, appeared identical to PGM1. The mammary gland material used in the construction of the human library while not lactating showed 'lactational competence' as judged by the stage of pregnancy (third trimester) of the woman whose tissue was used. This was the most appropriate human cDNA library available for searching for PGM4. Sequence analysis showed that all the clones identified in this library were identical to PGM1.
In summary, only PGM1 cDNAs were isolated from the mammary gland libraries and in addition, from a preliminary study there appears to be an association between PGM1 phenotype of mother and PGM banding pattern in colostrum. These findings add weight to the hypothesis that PGM4 isozymes are modified forms of PGM1 isozymes.
References


Dynan, W.S., Tijan, R. 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* 32:669-680


Najar, V. A., Pullman, M. E. 1954. The occurrence of a group transfer involving enzyme (PGM) and substrate. Science 119:631-634


