THE CHINESE HAMSTER PHOSPHOGLYCERATE KINASE GENE FAMILY

by

Stephen J. Rawson

A thesis submitted for the degree of Doctor of Philosophy

of the University of Leicester

Department of Biochemistry,

University of Leicester.

October, 1986.
ACKNOWLEDGEMENTS

I would like to thank the following people for their invaluable help: My supervisor, Dr. Michael J. Morgan, and Dr. Pelin Faik for their guidance and practical assistance during four years of laboratory work and writing; Jamie Walker for much helpful practical assistance, and also Judy Calvert, Kathy Bowness and Jan Smith for their practical support during my first two or three years in Leicester; Pam McTurk for all her electron microscopy work for heteroduplex analysis; John Keyte for his work in synthesizing numerous oligonucleotide primers.

Thanks are also extended to members of other laboratories in this department for their practical help and advice.

I also acknowledge the receipt of a Medical Research Council studentship for three years from October 1982, and a Wellcome Trust grant for the period of October to December of 1985.
## CONTENTS

### Chapter 1: Introduction

1.1 Introduction ............................................. 1  
1.2 Regulation of Gene Expression ............................. 3  
   1.2.1 Gene expression in prokaryotes ................... 4  
   1.2.2 Gene expression in eukaryotes ...................... 7  
   1.2.2.1 Transcriptional control ........................ 9  
   1.2.2.2 Translational control .......................... 12  
1.3 Phosphoglycerate kinase ................................. 15  
   1.3.1 PGK: General characteristics ..................... 15  
   1.3.2 Isozymes of PGK ................................. 16  
   1.3.3 PGK: A highly conserved enzyme ................... 17  
   1.3.4 The human PGK multigene family .................. 19  
   1.3.5 Genetic variants ............................... 22  
1.4 The Subject of this Thesis .............................. 23

### Chapter 2: Materials and Methods

2.1 Materials .................................................... 26  
   2.1.1 Cell lines and bacterial strains .................. 26  
   2.1.2 Enzymes and nucleic acids ....................... 26  
   2.1.3 Chemicals .......................................... 27  
   2.1.4 Standard solutions ................................. 28  
   2.1.5 Standard bacterial culture media ................. 29  
2.2 Methods ..................................................... 30  
   2.2.1 Cell culture ........................................ 30  
   2.2.2 Preparation of macroserum ......................... 30  
   2.2.3 Cloning of CHO cells ............................... 31  
   2.2.4 Single-cell plating ............................... 31
2.2.5 Glycerol cultures ........................................ 31
2.2.6 Plasmid preparation (with amplification) .............. 32
2.2.7 Plasmid preparation (no amplification) .................. 33
2.2.8 Preparation of DNA from cultured animal cells ......... 35
2.2.9 Isolation of DNA from low melting point gel ............ 36
2.2.10 "Freeze squeeze" extraction of DNA from agarose gel .... 37
2.2.11 DNA-mediated gene transfer .................................. 37
2.2.12 Transfer of DNA from agarose gel to nitrocellulose filter ...... 39
2.2.13 Hybridisation of nitrocellulose-bound DNA with radiolabelled DNA probe .................. 40
2.2.14 Plaque transfer and in situ hybridisation of plaques ....... 41
2.2.15 Phage titrations ........................................ 42
2.2.16 Phage stock (plate lysate) .................................. 43
2.2.17 Lambda phage preparation - liquid lysate ............... 43
2.2.18 Lambda phage DNA preparation .............................. 44
2.2.19 Preparation of alkaline-denatured salmon sperm DNA .... 46
2.2.20 M13 template DNA preparation (mini-prep) ............... 47
2.2.21 Determination of insert orientation in M13 recombinants (C-test) ............ 48
2.2.22 Large scale M13 DNA template preparation ............... 48
2.2.23 Preparation of M13 replicative form DNA .................. 49
2.2.24 Radioactive labelling of DNA by 'nick-translation ......... 50
2.2.25 5'-end labelling of oligonucleotides ...................... 50
2.2.26 Determination of counts incorporated into labelled DNA .... 51
2.2.27 Gel electrophoresis of DNA .......................... 52
2.2.28 Restriction endonuclease digestion of DNA  ...... 52
2.2.29 Alkaline phosphatase treatment of DNA .......... 53
2.2.30 Ligation of linear DNA fragments ............... 53
2.2.31 In vitro packaging of ligated DNA ......... 54
2.2.32 Transformation of bacteria .................... 55
2.2.33 Polyacrylamide gel electrophoresis ............ 56
2.2.34 DNA sequencing (dideoxy chain termination) . 57
2.2.35 Polyacrylamide gel purification ............... 60
2.2.36 Spun-column purification ........................ 61

Chapter 3: Gene Transfer in Chinese Hamster Ovary Cells

3.1 Introduction ............................................ 62
3.1.1 Carbohydrate metabolism in Chinese hamster
        ovary cells ....................................... 62
3.1.2 Gene transfer in cultured animal cells .......... 66

3.2 Results and Discussion .............................. 68
3.2.1 Sensitivity of R1.1.7 cells to G418 .......... 68
3.2.2 DNA mediated gene transfer ..................... 69
3.2.3 Stability of G418 resistance .................. 71
3.2.4 Level of G418 resistance ....................... 71
3.2.5 Southern transfer and hybridisation analysis 72
3.2.6 Modification of transfection system .......... 74
3.2.7 Co-transformation and selection for GPI
        positive transformants .......................... 76
3.2.8 Transformation of R1.1.7 cells using the human
        PGK cDNA ....................................... 78

3.3 Summary ................................................. 78
Chapter 4: Analysis of the Chinese hamster PGK gene family

4.1 Introduction ............................................. 81

4.2 Strategy ................................................... 83

4.2.1 Southern transfer analysis of Chinese hamster PGK gene family ......................... 83

4.2.2 Construction of Chinese hamster genomic library ........................................ 84

4.2.3 Screening of the library and isolation of recombinant phage ................................. 86

4.3 Results and Discussion ..................................... 87

4.3.1 Southern transfer analysis of Chinese hamster PGK gene family ......................... 87

4.3.2 Analysis of recombinant phage containing hamster PGK sequences ........................ 92

4.3.3 Heteroduplex analysis .................................... 97

4.3.4 Further discussion ........................................ 99

4.4 Summary .................................................. 101

Chapter 5: Sequence analysis of the Chinese hamster PGK sequences.

5.1 Introduction ................................................ 103

5.2 Strategy ................................................... 105

5.2.1 M13 cloning .............................................. 105

5.2.2 Screening M13 plaques ................................ 106

5.2.3 Determination of M13 insert orientation .................................................. 106

5.2.4 DNA sequencing ........................................... 107

5.3 Results and Discussion ..................................... 108

5.3.1 Analysis of the putative PGK exon .................................................. 108

5.3.1.1 Coding sequence ........................................ 108

5.3.1.2 Potential splice junctions ................................ 111

5.3.1.3 Potential 3' splice signal ................................ 112
5.3.1.4 Comparison with the exon/intron structure of the human PGK gene ................................ 113
5.3.2 Structure of Chinese hamster PGK pseudogenes ......................................................... 115
5.3.2.1 PGK coding region ....................................................... 115
5.3.2.2 5′ Non-coding region ................................................. 118
5.3.2.3 3′ Non-coding region ................................................. 120
5.3.3 Generation of hamster PGK pseudogenes ......................................................... 121
5.3.4 Origin of the pseudogenes ........................................................................ 125
5.3.5 Evolutionary history of the hamster PGK genes ..................................................... 126
5.3.6 Pseudogene function ....................................................................................... 130

5.4 Summary .......................................................................................................... 131

Final Summary .................................................................................................... 134

References ......................................................................................................... 138
ABBREVIATIONS

PGK  3-phosphoglycerate kinase
GPI  glucose phosphate isomerase
HK   hexokinase
HPRT hypoxanthine phosphoribosyltransferase
DHFR dihydrofolate reductase
HMGCoA 3-hydroxy,3-methylglutaryl coenzyme A
DPG  1,3-diphosphoglycerate
3-PG 3-phospho-D-glycerate
DTT  dithiothreitol
EDTA ethylene diamine tetraacetic acid
SDS  sodium dodecyl sulphate
BSA  bovine serum albumin
PVP  polyvinylpyrrolidone
PEG  polyethylene glycol 6000
TCA trichloroacetic acid
DMSO dimethyl sulphoxide
IMS industrial methylated spirits
ATP adenosine 5′-triphosphate
GTP guanosine 5′-triphosphate
DNA deoxyribonucleic acid
RNA ribonucleic acid
mRNA messenger RNA
tRNA transfer RNA
dCTP 2′-deoxycytidine 5′-triphosphate
dGTP 2′-deoxyguanosine 5′-triphosphate
dTTP thymidine 5′-triphosphate
dATP 2′-deoxyadenosine 5′-triphosphate
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddCTP</td>
<td>2',3'-dideoxycytidine 5'-triphosphate</td>
</tr>
<tr>
<td>ddGTP</td>
<td>2',3'-dideoxyguanosine 5'-triphosphate</td>
</tr>
<tr>
<td>ddTTP</td>
<td>2',3'-dideoxythymidine 5'-triphosphate</td>
</tr>
<tr>
<td>ddATP</td>
<td>2',3'-dideoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>AGPT</td>
<td>aminoglycosyl 3'-phosphotransferase</td>
</tr>
<tr>
<td>DMGT</td>
<td>DNA-mediated gene transfer</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-[N-Morpholino]propanesulphonic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-thiogalactoside</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-Bromo, 4-chloro, 3-indolyl β-D-galactopyranoside</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>HEC</td>
<td>Human embryo cell</td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs(s)</td>
</tr>
<tr>
<td>μm</td>
<td>micrometre(s)</td>
</tr>
<tr>
<td>nm</td>
<td>nanometre(s)</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION

1.1 Introduction

The study of carbohydrate metabolism in prokaryotes and eukaryotes has led to the discovery of numerous metabolic pathways. The investigation of the regulation of these pathways has been greatly facilitated by the isolation of mutant cell lines which have been used to examine the effects of enzyme deficiencies on carbohydrate metabolism and also to isolate the genes coding for such enzymes by complementation of these deficiencies.

Using animal cells in culture, the utilisation of carbohydrates by both "wild-type" and variant cells has been studied by a number of investigators. Growth rates, conditions for optimum growth, and lactic acid production, whilst growing on different carbohydrates, are some of the features investigated in human (Eagle et al., 1958; Melnykovych & Bishop, 1972), Chinese hamster (Faik & Morgan, 1976; Burns et al., 1976; Dahl et al., 1976) and mouse cell lines (Burns et al., 1976).

Work in this laboratory has, for a number of years, concentrated on the investigation of carbohydrate utilisation in cultured Chinese hamster ovary (CHO) cells. Variants with altered abilities to utilise carbohydrates have been isolated (Faik & Morgan, 1976, 1977a, 1977b) and have been of use in gaining insights into the regulation of mammalian carbohydrate
metabolism. Work with these variants, using techniques of genetic manipulation as a more direct approach to the investigation of the regulation of gene expression, should lead to a greater understanding of the expression of mammalian genes - both in general, and in particular those coding for enzymes of carbohydrate metabolism and other "housekeeping" enzymes. A great deal can be learned from studying isolated gene sequences and their RNA transcripts, and from investigating the production, processing, and translation of the transcripts.

The parental, or "wild-type", cell line which has been used in this laboratory in the investigation of carbohydrate metabolism, and from which a number of variant cell lines have been derived, is the Chinese hamster ovary cell line, CHO-K1 (Tjio & Puck, 1958). A variant cell line, R1.1.7, isolated as a result of having altered carbohydrate utilising abilities, was found to be deficient in both phosphoglycerate kinase (PGK) and glucose phosphate isomerase (GPI) and to have a reduced ability, in vivo and in vitro, to convert glucose to lactic acid, but it was still able grow at almost the same rate as the parental CHO-K1 cells (Morgan & Faik, 1980). The characteristics and requirements of CHO-K1 and R1.1.7 are discussed in detail in chapter 3.

The deficiency of PGK in R1.1.7 cells makes this cell line an ideal candidate for the investigation of the expression of this enzyme in vivo in a system with little background enzyme activity. Chapter 3 describes the setting up of such an in vivo expression system for PGK sequences introduced into R1.1.7 cells by DNA-mediated transformation. Chapters 4 and 5 describe the isolation and characterisation of a number of Chinese hamster PGK sequences from CHO-K1 cells. It was unfortunate that those
sequences isolated were not capable of expression, being incomplete PGK sequences and pseudogenes, and for this reason the project developed into an investigation of the structures of these Chinese hamster PGK sequences.

1.2 Regulation of Gene Expression

Carbohydrate metabolism, like other metabolic pathways, is regulated at two principal levels: "fine" control of the activity of individual enzymes in the pathway and "coarse" control of their synthesis. In "fine" control, the catalytic activity of an enzyme is subject to the concentration of its substrates and products, and also to effector molecules which enhance the enzyme's activity and inhibitors which bring about a reduction in the catalytic activity. Enzymes may also be inter-converted between active and inactive forms by covalent modification, for example phosphorylation of an amino acid side chain. These effects are usually reversible. "Coarse" control of the activity of an enzyme determines the amount of functional enzyme present by regulating the expression of the genes coding for the enzymes. Here, regulation may take place at the level of transcription of an RNA message from the DNA sequence, processing of the RNA, translation of the message into a polypeptide, or post-translational modification resulting in the final, active, form of the enzyme.

A great deal of work has been carried out by the isolation of individual genes and examination of the factors that influence their expression. The following sections summarise some of the features of the regulation of gene expression in prokaryotes and eukaryotes.
1.2.1 Gene expression in prokaryotes.

Much of the work on prokaryote gene expression has been carried out using the bacterium *Escherichia coli*. Genes coding for proteins which are involved in a particular metabolic process are usually organised in single transcriptional units called operons. The *lac* operon codes for β-galactosidase, galactoside permease and thiogalactoside transacetylase, and the arabinose operon codes for the enzymes that convert arabinose to xylulose5-phosphate. Other operons that have also been studied include those that encode enzymes involved in the biosynthesis of tryptophan (*trp*), and histidine (*his*). An operon which consists of a number of structural genes (nine in the case of the *his* operon) is transcribed from a single promoter region into a polycistronic messenger RNA molecule, and this is translated into a number of polypeptides.

Transcription is initiated in *E. coli* by the binding of RNA polymerase to a promoter sequence upstream of the structural genes, and it is at this stage that the major controls of gene expression operate (for review see McClure, 1985). Promoter sequences contain two consensus regions of DNA sequence which are involved in RNA polymerase recognition and binding. Approximately 35 bp upstream from the site of transcript initiation there is a consensus sequence, TTGACA, and 10 bp upstream of the initiation site is found the "Pribnow" box, TATAAT (Rosenberg & Court, 1979; Siebenlist et al., 1980) (Figure 1.1). The termination of transcription is signalled by a GC-rich region containing a palindromic repeat, and an AT-rich region at the actual termination site (Rosenberg & Court, 1979). These sequences are recognised by RNA polymerase and by a
Figure 1.1

DNA sequence elements controlling gene expression in prokaryotes and eukaryotes.

Promoter sequences are illustrated in prokaryotes (A), and eukaryotes (B), upstream of the transcriptional start points. The Shine-Dalgarno sequence (S-D) is found upstream of the translational start point in prokaryotes. The enhancer elements of eukaryotic genes may be located either upstream or downstream of the transcriptional start point.
A. 30-40 bp

- TTGACA
- TATAAT
- S-D
- ATG...

B. <100 bp

- GGGCGG
- CCAAT
- TATA

Enhancer
termination factor, rho \((\rho)\), which interacts with RNA polymerase. Translation of the messenger RNA initially involves the binding of the transcript to a ribosome. A semi-conserved sequence in the transcript, known as the Shine-Dalgarno sequence, is complementary to the 3' end of the 16S ribosomal RNA (Shine & Dalgarno, 1975). This sequence is found between 3 and 12 bp upstream of the initiation codon, AUG, which specifies the first amino acid, methionine, in the initial polypeptide product. Translation continues with the mRNA passing through the active site of the ribosome, and is terminated when one of the three termination codons (TGA, TAG, or TAA) is reached at the end of the coding sequence.

The efficiency of both transcription and translation are probably subject to the precise arrangement of the promoter region consensus sequences and the Shine-Dalgarno sequence, and their positions with respect to the RNA polymerase binding site and the translation initiation codon respectively.

Post-translational modification of synthesized polypeptides is often necessary before the protein can function properly, for example, proteins which are secreted through membranes are initially synthesized with hydrophobic signal sequences at their amino termini, and these are subsequently removed by a proteolytic enzyme.

Gene expression in bacteria is, however, primarily regulated at the level of transcription (McClure, 1985). The efficiency of transcript production is dependent on the ability of RNA polymerase to initiate this process. Operons have been shown to contain regulatory genes which code for protein factors that interact with the DNA in the region of the RNA polymerase
binding site. These proteins may be repressors, as in the case of the lac operon, where such a protein binds to an operator site downstream from the promoter region thus blocking RNA polymerase from initiating transcription. The regulatory protein may, as in the case of the arabinose (ara) operon, act as both a repressor and an activator of transcription initiation depending on whether a secondary regulatory molecule is present. These molecules which interact with the regulatory proteins are usually substrates or products of the enzymes encoded by the operons, for example, β-galactosides are inducers which act by binding to the regulatory protein of the lac operon and altering its conformation such that it can no longer bind to the operator region, and so transcription can occur. Tryptophan, however, inhibits transcription of the trp operon by binding to the trp repressor thereby allowing it to bind to the operator region of the operon.

Thus, the substrates and products of an enzyme's action can regulate its activity on the level of "coarse" control of enzyme synthesis as well as the "fine" control of the enzyme's specific activity. In addition to the effects of substrates and products on transcription, cyclic AMP promotes the initiation of transcription in many inducible operons. In the case of the lac operon, cyclic AMP binds to a receptor protein, CAP, which can then bind to the promoter region of the operon, thus stimulating the initiation of transcription by RNA polymerase. It is thought that cyclic AMP acts in a similar way in the induction of other operons.

The efficiency of translation in prokaryotes varies for different mRNA species and as a result of the presence of
regulatory factors in the cellular environment. Ribosome binding is thought to be dependent on the secondary structure of the 5' end of the mRNA which would influence the accessibility of the Shine-Dalgarno sequence and the initiation codon (Iserentant & Fiers, 1980). A more direct regulatory influence on mRNA translation in prokaryotes is evident in a number of bacteria and bacteriophages where protein synthesis has been shown to be subject to translational repression by protein factors. A T4 phage protein involved in DNA replication and repair regulates its own expression by binding to its mRNA, thus blocking translation (Lemaire et al., 1978). A number of ribosomal proteins also repress translation of their mRNAs, in certain cases, by binding to them at the same sites at which they interact with the ribosomal RNA (Olins & Nomura, 1981).

1.2.2 Gene expression in eukaryotes.

Transcription and translation of eukaryotic genes are highly complex processes about which a great deal remains to be understood. The mechanisms involved are more complex than those observed in prokaryotes, so there are a greater variety of processes that could be involved in the regulation of gene expression.

Unlike prokaryotic genes, those of eukaryotes are generally transcribed as separate mRNAs. Whether a gene is transcribed or not can depend on the presence of initiating factors or inducing molecules, or whether the DNA is 'masked' in some way from the transcriptional apparatus of the cell. Methylation of certain regions of a gene, for example, has been shown to influence transcription: there is a correlation between undermethylation
of a gene and its potential to be transcribed (for reviews see Razin & Riggs, 1980; Doefler, 1981 & 1983).

Following transcription in the nucleus, the primary DNA transcripts undergo a number of modifications which could also be the objects of regulatory processes (for review see Revel & Groner, 1978). Primary transcripts are spliced to remove the intervening sequences of non-coding DNA (introns) which break up the coding sequence in most eukaryotic genes. A tail of adenylic acid residues ("poly A" tail) is added to the transcript, and this is thought to play a part in mRNA stability. Eukaryotic mRNAs also undergo methylation and generally contain a 'cap' structure at their 5' ends consisting of a terminal 7-methyl guanosine residue linked to an adjacent methylated nucleotide (Shatkin, 1976). This structure is recognised during the initiation of translation and also protects the 5' end of the mRNA from exonuclease digestion.

Translation of the mature mRNA involves a complex initiation mechanism requiring the ribosome and at least ten initiation factors. Modification of some of these factors, and of the mRNA itself, as a response to stimulation or inhibition by different cellular factors and inducing molecules such as hormones, has been shown to be involved in the regulation of the initiation of translation (for review see Austin & Kay, 1982).

Thus, in eukaryotes, transcription, processing, translation, and their regulation, are subject to a number of environmental and developmental signals mediated by a number of different protein factors which bind to specific regions of DNA, RNA, or other components of the transcriptional or translational machinery. Some of the features of eukaryotic gene expression
and some of the regulatory processes which are involved in overall and specific control of the efficiency of transcription and translation are described below.

1.2.2.1 Transcriptional control.

Control of the rate and accuracy of the initiation of transcription is dependent on a number of promoter elements upstream of a gene's coding sequence, usually within 100 bp of the transcriptional start point (Figure 1.1). The site of initiation of the transcript appears to be fixed by the closest promoter element, a conserved sequence, TATAA, known as the "TATA box" which is found 25 to 30 bp from this point (Breathnach & Chambon, 1981). There is a specific protein factor required for initiation that binds to this site (Parker & Topol, 1984; Sawadogo & Roder, 1985). Further upstream, conserved sequences are found which have been shown to be required for the efficient initiation of transcription by RNA polymerase II (for review see Kadonaga et al., 1986) and to which protein transcription factors have been shown to bind (Jones et al., 1985; Dynan & Tjian, 1985). These elements, the sequence CCAAT (the "CAAT box") and GC-rich sequences containing the consensus GGGCGG, have been found both separately and in combination in different eukaryotic promoter regions: the mammalian β-globin promoters each contain a "CAAT box", the Simian Virus 40 (SV40) early promoter region contains six GC-rich promoter elements and the herpes simplex virus thymidine kinase (HSV-TK) promoter contains a "CAAT box" flanked by two GC-rich elements. The proteins which bind to these two consensus sequences are transcription factors known as Sp1 (which binds to the GC-rich promoter element) and CTF
(CAAT-binding transcription factor), although the mechanisms by which transcription is stimulated by these factors are not known.

It is notable that promoter regions containing the GC-rich Sp1 binding sequence generally lack the "TATA box" and are found as the upstream control elements of genes coding for so-called "housekeeping" enzymes which catalyse essential and universal metabolic functions (for review see Dynan, 1986). The binding of Sp1 to the GC-rich promoter element and the consequent activation of transcription have been demonstrated for the mouse dihydrofolate reductase (DHFR) gene (Dynan et al., 1986). Other examples of such promoters include those of the genes coding for mouse and human hypoxanthine phosphoribosyl transferase (HPRT), the human and hamster DHFR, human adenosine deaminase and human phosphoglycerate kinase. This element is not, however, confined to housekeeping gene promoters as such sequences are found in the upstream regions of some growth control genes, for example, the Ha-ras gene (Ishii et al., 1985). Nor do these promoters necessarily lack the "TATA box", for example, the 5' flanking region of the human triose phosphate isomerase gene contains both these promoter elements (Brown et al., 1985). Thus, the promoter regions of housekeeping genes, which maintain relatively low levels of messenger RNA synthesis, differ from those containing "TATA boxes" which control the transcription of genes such as the globins, immunoglobulins and histones, which are transcribed during particular periods of the cell cycle and stages of cell development, or in specialised tissues.

The transcription of a number of eukaryotic genes is also regulated by 'enhancer' sequences (Figure 1.1) which are
generally 100 to 200 bp long and may be positioned in either orientation, both upstream or downstream of the transcription initiation site. The most notable characteristic of enhancer elements is that they are usually located some distance (at least 1 kb) from the promoter region of the gene which they influence. A consensus sequence (NTGTGG(A/T)(A/T)AG) has been determined for elements found in a number of enhancers (Weiher et al., 1983), although this sequence is not universal. Like promoter elements, enhancers appear to exert their effects by interacting with cellular factors (for review see Sassone-Corsi & Borrelli, 1986). The mechanism by which enhancers affect transcriptional activity is unlikely to be similar to those by which the promoter elements modulate transcription, due to the great distances between them and their targets. It is possible that transcription may be modulated by interactions between enhancer and promoter elements or their binding factors, resulting from a looping out of the intervening DNA.

Upstream promoter elements and enhancers are, in some cases, thought to be involved in mediating tissue specific or cell specific stimulation of transcription, and have been shown to be influenced by a number of inducers, for example steroid hormones and heavy metals in the cases of the human metallothionein-II and mouse metallothionein promoters respectively (Sassone-Corsi & Borrelli, 1986). As well as sequences which are involved in the stimulation of transcription, there are, as in prokaryotes, negative regulatory sequences and their binding factors which reduce mRNA synthesis, for example the large T antigen of SV40, the 'blocker' sequence of the ovalbumin gene promoter, and the suppression of hydroxymethylglutaryl Coenzyme A (HMGCoA)
reductase transcription by sterols (for references see Sassone-Corsi & Borrelli, 1986).

1.2.2.2 Translational control.

Although the expression of specific genes is generally regulated through transcriptional control, the overall activity of protein synthesis in a cell and the synthesis of specific proteins are also subject to the availability or modification of the numerous protein and nucleic acid factors involved in translation, and the environmental conditions in the cell. These include the presence of activating or inhibiting molecules, hormonal stimulation or repression, energy provision, and pH (for review see Austin & Kay, 1982).

Cellular mRNA concentration depends on the efficiency of processing of the initial transcript and the stability of the mature RNA as well as the rate of gene transcription. It is likely that, for different mRNA species, post-transcriptional modifications such as splicing and capping will take place with different efficiencies. The regulation of mRNA stability has also been shown to have a great effect on the levels of mRNA in the cell (Harpold et al., 1979; Tobin, 1979). The half-lives of certain mRNAs have been shown to be under the influence of hormonal action, for example, the mRNA half-lives of conalbumin and ovalbumin in the chick oviduct, and of casein in mammary gland organ cultures, are increased in the presence of steroids and prolactin respectively (McKnight & Palmiter, 1979; Guyette, 1979). mRNA degradation can also be induced as a mechanism for decreasing its cellular concentration and consequently its translation rate (Austin & Kay, 1982).

Translational controls are generally more important for the
regulation of overall protein synthesis, and allow sensitive responses of protein synthesis to controlling factors where mRNA concentrations cannot be influenced so rapidly because of their relatively long lifetimes (varying, for different mRNAs, from a few minutes to at least 24 hours). Indeed, in eukaryotic cells lacking nuclei, such as reticulocytes, protein synthesis can only be regulated at the level of translation.

Translational regulation more usually involves control of the rate of initiation (Fan & Penman, 1970; Austin & Clemens, 1981) and much of the work carried out in this area has involved the use of reticulocytes. Initiation of translation begins with the formation of a complex between methionyl tRNA, GTP and an initiation factor, eIF-2. This complex then binds to the 40S ribosomal subunit in a process involving a number of other initiation factors (see Austin & Kay, 1982). In the reticulocyte, globin synthesis is dependent on the presence of haem, the prosthetic group of haemoglobin. In the absence of haem, a haem controlled inhibitor (HCI) is activated and phosphorylates the α subunit of eIF-2, thus blocking initiation (Ochoa, 1983). This general mechanism of control of initiation, by the prevention of methionyl tRNA binding to the 40S ribosomal subunit, is thought also to be applicable to other mammalian cell systems (see Voorma et al., 1983).

There are numerous other mechanisms by which translation could be influenced (Austin & Kay, 1982). The availability of ribosomes may be controlled by reversible inactivation by inhibitory proteins or phosphorylation. Changes in intracellular pH and the concentration of certain other ions have been correlated with changes in translation rates, different mRNAs sometimes requiring different ionic conditions.
for the stimulation of protein synthesis. Overall protein synthesis, and the synthesis of specific proteins, could be limited by the availability of charged tRNA molecules, particularly the initiating methionyl tRNA, and may be inhibited competitively by increased concentrations of uncharged tRNAs.

The rate of translation of different individual mRNA species can be regulated by a number of more specific mechanisms. Certain mRNAs may be inherently more efficient than others at initiation, as has been suggested for the translation of α- and β-globin in reticulocytes (Lodish, 1974), where β-globin competes more effectively in ribosome binding. The secondary structure of the mRNA may play an important part in the efficiency of initiation by influencing its binding to ribosomal constituents and other initiation factors. The efficiency of mRNA binding and initiation can also be influenced by modification of the ribosomes or initiation factors, as is thought to occur in certain viral infections, for example, in the late phase of reovirus infection modification of the initiation factor eIF-4E, or mRNA cap-binding factors, results in a stimulation of translation of the uncapped viral mRNAs (Skup et al., 1981). Protein factors have also been shown to bind to mRNA, preventing the message in the resulting ribonucleoprotein (mRNP) from being translated. This is a mechanism that allows, for example, a rapid increase in protein synthesis when fertilisation or maturation of the eggs of many animals stimulates the release of 'stored' mRNA from the mRNP complexes. As well as regulating the overall rate of protein synthesis, there are thought to be separate mechanisms for making different 'stored' mRNAs available for translation in this way, thus allowing selective regulation of protein
synthesis (Flynn & Woodland, 1980). Translation may also be inhibited or stimulated by a class of translational control RNA (tcRNA) molecules which have been reported to influence the translation of mRNA, in most cases in a non-specific way.

Thus, there exist a variety of mechanisms which may act independently or cooperatively to regulate protein synthesis at a translational level in eukaryotic cells. However, for the most part, the regulation of gene expression occurs, more economically, at the level of transcription.

1.3 Phosphoglycerate Kinase

1.3.1 Phosphoglycerate kinase (PGK): general characteristics.

PGK is an essential "housekeeping" enzyme of glycolysis, being the first energy-yielding step in the pathway. It catalyses the reversible conversion of 1,3-diphosphoglycerate (1,3-DPG) to 3-phospho-D-glycerate (3-PG), resulting in the production of ATP from ADP in the forward direction (Figure 1.2).

PGK has been isolated from tissues of a number of mammalian species, and has been shown to be made up of a single polypeptide chain with a molecular weight of at least 45,000. The number of amino acids found in the PGK proteins thus far isolated varies from 416 in the human (Michelson et al., 1983) and the horse (Merret, 1981) to 424 in the mouse (Lee et al., 1980). The polypeptide chain is folded into two domains, one of which binds the ADP or ATP, and the other which, it is thought, might bind the 1,3-DPG or 3-PG (Banks et al., 1979). Watson et al. (1982), however, have provided evidence that the phosphoglycerate binding site is located in the C-terminal
Figure 1.2

Metabolic scheme illustrating the steps of glycolysis pertinent to this study.

Phosphoglycerate kinase (PGK, E.C.2.7.2.3) catalyses the conversion of 1,3-diphosphoglycerate (1,3-DPG) to 3-phospho-D-glycerate (3-PG). The phosphoryl group from the acyl phosphate of 1,3-DPG is transferred to ADP to yield ATP. Glucose phosphate isomerase (GPI, E.C.5.3.1.9) catalyses the interconversion of glucose-6-phosphate and fructose-6-phosphate.
domain, as is the nucleotide binding site. The substrates are thought to be brought together into an active site through the hinge bending action of the double chain link between the domains (Banks et al., 1979; Merret, 1981).

1.3.2 Isozymes of PGK.

Two isozymes of PGK are found in mammals. One of these, PGK-1, is expressed in all somatic cells and is coded for by an X-linked gene (Chen et al., 1971; Cooper et al., 1971; VandeBerg et al., 1973a; Kozak et al., 1974). PGK-2 has an autosomal locus (Cooper et al., 1971) and is generally detectable in most mammals only in sperm and in testicular tissue during spermatogenesis (VandeBerg et al., 1983). There are reported exceptions, however, where it is claimed that this isozyme is expressed weakly in somatic cells: for example in the dog, fox and kangaroo (VandeBerg et al., 1973b). The two PGK isozymes each consist of a single major electrophoretic band of activity and can be resolved from each other on starch gels. The two are also immunologically distinct: although there is a weak reaction of antiserum to PGK-1 with PGK-2 in the mouse, this is detectable only by enzyme inactivation and not by immunoprecipitation, and there is no reaction of antiserum to PGK-2 with PGK-1 (Pegoraro et al., 1978). In human cell extracts, antiserum to PGK-1 does not have any inactivating effect on the PGK activity of PGK-2 (Chen et al., 1976).

Comparison of the biochemical properties of the two mouse isozymes (Pegoraro & Lee, 1978) revealed that PGK-2 has a lower thermal stability than PGK-1 and is more rapidly inactivated by acidic conditions. However, the conservation of their properties is demonstrated by their similar molecular weights,
pH dependence and temperature dependence of activity, substrate specificity, and also their kinetic properties: similar Km values have been determined for PGK-1 and PGK-2 (Chen et al., 1976; Pegoraro & Lee, 1978). The amino acid compositions of the two isozymes are also remarkably similar (VandeBerg, 1985).

PGK immunological activity (Kramer, 1981), enzymatic activity, and translation in post-meiotic cells have been demonstrated (Kramer & Erickson, 1981). In addition, in vitro translational studies using mRNA isolated from testicular germ cells showed that PGK-2 mRNA was present only in the mature testis and late spermatid cell fractions (Erickson et al., 1980), being synthesized in the haploid spermatids (Erickson et al., 1985). PGK-2 appears in the mouse testis at the same time as spermatids are first seen. A possible functional significance of this is that, being autosomally encoded, its message can be synthesized in post-meiotic sperm cells bearing the Y-chromosome and during X-chromosome inactivation throughout spermatogenesis following early meiotic prophase (VandeBerg et al., 1983). In this way PGK activity can be maintained in sperm cells to allow glycolysis to take place. It is also likely that the difference in properties of PGK-1 and PGK-2 reflects the different environment in spermatozoa (VandeBerg, 1985).

1.3.3 PGK: A highly conserved enzyme.

PGK has been isolated from a number of species, and the amino acid compositions and polypeptide sequences of some of these enzymes have been determined (for review see VandeBerg, 1985). In mammalian species, the isozyme studied has generally been the universally abundant PGK-1.

X-ray analysis has shown that PGK from yeast (Bryant et al.,
1974) and horse muscle (Blake & Evans, 1974) have very similar tertiary conformational structures, consisting of two separate domains of approximately equal size. Watson et al. (1982) have confirmed from crystallographic studies that the conformation of the enzyme has been highly conserved during its evolution. Banks et al. (1979) have suggested that these domains move towards each other to bring the phosphoglycerate substrate into close proximity to the nucleotide substrate as part of the catalytic mechanism.

That homology also exists at the level of the primary structure of PGK in yeast and mammalian species was realised by the determination of their amino acid compositions and sequences. The primary structure of horse PGK was determined and related to its tertiary structure by Banks et al. (1979). Human PGK was the next PGK enzyme for which the complete amino acid sequence was determined (Huang et al., 1980a & 1980b), followed by the yeast PGK sequence (Watson et al., 1982) which was also deduced from the yeast PGK gene sequence (Hitzeman et al., 1982). Comparison of the three enzymes (Hitzeman et al., 1982) demonstrated considerable homology. Horse and human PGK appeared to contain only 14 amino acid differences in the total of 416 making up the polypeptides. One apparent difference was the insertion of an extra lysine residue in the human PGK (Huang et al., 1980a), although the cDNA sequence of human PGK (Michelson et al., 1983) later revealed that this amino acid was not represented in the DNA sequence. The yeast PGK protein sequence containing 415 amino acids has approximately 65% homology with the human and horse sequences, although there are certain regions which are highly conserved, perhaps indicating
sections which are important for the binding or catalytic functions of the PGK enzyme.

The yeast PGK gene has been isolated by immunological screening of recombinant plasmid clones (Hitzeman et al., 1980), and sequenced (Hitzeman et al., 1982). Later, cDNA clones for human X-linked PGK were isolated using mixtures of synthetic oligodeoxynucleotide probes (Singer-Sam et al., 1983; Michelson et al., 1983), and the DNA sequence was determined (Michelson et al., 1983). If the amino acid sequence predicted by the human cDNA is correct, then there are only 12 differences between human and horse PGK-1 (representing 97% protein sequence homology) each of which could be explained by single base substitutions in the DNA sequence.

The amino acid compositions of a larger number of mammalian PGK-1s have been determined, and those of the mouse (Lee et al., 1980), sheep (Stewart & Scopes, 1978) and rabbit (Lee et al., 1980) are very similar to those of the horse and human (VandeBerg, 1985). The amino acid composition of PGK-2 from the mouse (Lee et al., 1980) and sheep (Stewart and Scopes, 1978) are also very similar to those of the PGK-1 enzymes (VandeBerg, 1985). The similarity of these PGK-2 enzymes to PGK-1 suggests that the two isozyme loci may have arisen from a gene duplication event before the divergence of the mammalian orders, and the amino acid compositions and sequences in those species examined show that both isozymes have been highly conserved during their evolution.

1.3.4 The human PGK multigene family.

Human PGK cDNA sequences (Michelson et al., 1983; Singer-Sam et al., 1983) have been used as probes to study the organisation
of the human PGK gene family and to map PGK sequences to specific chromosomal regions. Initial analysis of human PGK sequences revealed a complex array of PGK-related sequences, both X-linked and autosomal (Michelson et al., 1983). A single functional PGK gene has been mapped on the human X chromosome to band Xq13 (Miller et al., 1984). This PGK gene has been cloned and its sequence determined (Michelson et al., 1985b). It was found to cover 23 kb of DNA, the coding sequence being interrupted by 10 introns which range in size from 0.2 to 5.7 kb. Earlier, the promoter region and first exon of this gene had been isolated and sequenced (Singer-Sam et al., 1984): Three start points for transcription were determined, by primer extension and S1 nuclease mapping, at positions 86, 96 and 97 bp upstream of the initiation codon. The promoter region was found to lack the "TATA box" (Corden et al., 1980) and "CAAT box" (Benoist et al., 1980) often found upstream of eukaryotic transcriptional start points, but it is GC-rich and contains an 8 bp direct repeat, and a possible binding site for the transcription factor Sp1 similar to the promoter sequences of other housekeeping genes.

It has also been established that there is a second, independent, X-linked PGK gene, and this has been sequenced and identified as an intronless pseudogene located proximally to the functional gene on the long arm of the X chromosome (Michelson et al., 1985a). Despite having a high degree of homology with the PGK cDNA, this gene contains a number of deletions resulting in shifts in the reading frame, and base changes producing premature termination codons and sequences of inappropriate amino acid codons, thus rendering it unable to code for a normal PGK enzyme. It also completely lacks the intervening sequences
found in the functional gene. The pseudogene exhibits some of the properties of a processed pseudogene originally generated via an mRNA intermediate: the gene shows homology with the functional gene up to one of the transcriptional start points, but not beyond, and in the 3' untranslated region up to the probable polyadenylation site. However, unlike many processed pseudogenes, a 'poly A' tract representing the remains of the 'poly A' tail of the message is not observed. The gene is flanked by a direct repeat which may be taken as evidence of integration of the pseudogene into the chromosome (see chapter 5 for a discussion of pseudogenes).

Two autosomal PGK sequences have been described, one of which was isolated and mapped to human autosome 6 in the p23-q12 region (Szabo et al., 1984). This is also the region containing the major histocompatibility locus (HLA). Since the gene coding for PGK-2 in the mouse is closely linked to the H2 histocompatibility complex (Eicher et al., 1978) it was thought that the isolated PGK sequence might represent part of the gene for PGK-2. More recently, Michelson et al. (1985a) mapped autosomal PGK sequences to 6p12-21.1 and concluded that the four human PGK genes which they estimated to be present in the genome are dispersed to only two chromosomal loci, one on the long arm of the X chromosome and the other on the short arm of chromosome 6. An autosomal intronless pseudogene for human PGK has been isolated and sequenced (Tani et al., 1985a). The clone containing this sequence overlaps with the clone isolated by Szabo et al. (1984) which was originally thought to code for PGK-2. This pseudogene contains deletions which result in frame shifts, base changes, and an in-phase termination codon. Direct repeats also flank the gene, which contains homology with the
human PGK cDNA in the 3' untranslated region and an adenine-rich region downstream of the polyadenylation signal. These characteristics, like those of the X-linked PGK pseudogene, identify this PGK sequence as a processed pseudogene.

Interestingly, in contrast to the conclusion of Michelson et al. (1985a), Willard et al. (1985) have used a panel of somatic cell hybrids to assign an autosomal PGK-related DNA sequence to human chromosome 19. This does not, however, refute Michelson’s evidence. Since a PGK sequence on chromosome 6 has been identified as a pseudogene, perhaps the sequence mapped to chromosome 19 represents the testis-specific PGK-2 gene. Further cloning and molecular analysis of this PGK gene sequence (or perhaps a fifth PGK sequence) will be required to finally determine the nature and location of the PGK-2 gene.

1.3.5 Genetic variants.

Genetic variants of PGK have been discovered in humans and a number of other species (for review see VandeBerg, 1985). Of the seventeen human PGK-1 variants identified, five have normal enzyme activity and are fairly common in some of the populations in which they have been studied. Twelve rare variants are associated with reduced PGK activity and each of these has been observed in single, unrelated families (for refs see: Huang et al., 1980c; DiMauro et al., 1983): Nine of these deficiencies are associated with haemolytic anaemia, and six of the nine with mental disorders; two variants are associated with rhabdomyolysis (a muscular disease) and renal failure, although not with haemolytic anaemia and mental disorders; one variant, PGK-München (Fujii et al., 1980), has no clinical deficiencies. PGK enzymes isolated from some of the variants exhibit different
electrophoretic mobilities from the "wild-type" protein, so the mutations responsible for the PGK deficiencies probably occur in the coding sequences of these genes. Single amino acid substitutions have in fact been determined for three of these PGK variant enzymes: PGK-München, PGK-Uppsala (Fujii & Yoshida, 1980) and PGK-Tokyo (Fujii et al., 1981).

Studies of the genomic aspects of the PGK-Matsue variant (Yoshida & Miwa, 1974) is in progress (Yoshida & Tani, 1983) and, although a mutation responsible for this deficiency has not been identified, it has been determined that this variant has a normal PGK mRNA content, and results suggest that the major cause of enzyme deficiency is a large increase in mutant enzyme degradation (Tani et al., 1985b).

Molecular analysis of further variant PGK genes and study of their transcription and translation should eventually clarify the precise molecular abnormalities which bring about their reduced activities.

1.4 The Subject of this Thesis

The project described in this thesis involves the use of the "wild-type" Chinese hamster ovary cell line, CHO-K1, to investigate the Chinese hamster PGK gene family. Restriction digests of genomic DNA were probed with a human PGK cDNA and, in addition, this probe was used to isolate PGK sequences from a CHO-K1 DNA library. It was intended to study the expression of such PGK sequences using the PGK-deficient variant of CHO-K1, R1.1.7 and, with this plan in mind, a 'transfection system' was developed for the efficient introduction of exogenous DNA into R1.1.7 cells. The development of this system, which was shown
to take up and express exogenous PGK sequences, is described in chapter 3. Using this system it would be possible to introduce different PGK sequences into R1.1.7, including any isolated from a CHO-K1 DNA library, and investigate their expression in a background free from endogenous PGK activity.

R1.1.7 is also of interest in its own right, since there are a number of possible reasons why PGK activity is not seen in this variant: mutations may exist in the coding sequences, altering the substrate binding or conformation of the protein to such an extent that catalysis is not possible. Another possibility is that there may be mutations in the non-coding regions of the PGK gene sequence which prevent efficient or accurate transcription, processing or translation of the message. Although this cell line was only investigated by comparing the hybridisation patterns of its DNA with a PGK cDNA probe to those observed in CHO-K1 DNA, it would have been interesting to determine whether any messenger RNA was produced in R1.1.7 cells and, if so, whether the level is comparable to that seen in CHO-K1. Ultimately, the precise nature of the cause of this deficiency could only be determined following the isolation of the deficient PGK gene in R1.1.7.

The direction in which this project turned after the development of the above-mentioned transfection system, however, was towards the general organisation of the Chinese hamster PGK gene family, and the genes which make up this group. The use of the transfection system in studying the isolated PGK sequences was not possible, as the sequences isolated were not appropriate for such research as they were not capable of expression.

The first stages in the investigation, described in chapter 4, involved probing digests of Chinese hamster DNA with
a human PGK cDNA probe. Hybridisation of this probe to male and female hamster DNA was performed to determine whether any of the sequences were X-linked and might therefore correspond to the Chinese hamster PGK-1 gene. Hybridisation to DNA from different hamster cell lines and hamster tissue was also compared to determine whether there had been any DNA rearrangements while these cells were in culture. The construction of a genomic library and the isolation of Chinese hamster PGK sequences from the library are described. These isolated PGK sequences were analysed, firstly by hybridisation to different regions of the PGK cDNA probe and heteroduplex mapping to determine their orientation and general organisation in the recombinant clones. Secondly, as described in chapter 5, the DNA sequences of the isolated PGK gene sequences were examined in detail. This analysis included the determination of their potential to code for all or part of a functional PGK enzyme, their alignment and comparison with the sequence of the human PGK cDNA to determine the level of inter-species conservation of the sequences, and evaluation of their evolutionary history.
Chapter 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Cell lines and Bacterial strains.

The major cell type used was the Chinese hamster ovary cell line, CHO-K1 (Tjio & Puck, 1958). R1.1.7 (Faik & Morgan, 1977b) is a phosphoglycerate kinase-deficient cell line derived from CHO-K1. A3 is a male Chinese hamster cell line (kindly supplied by Professor P. Pearson, Dept. of Human Genetics, Leiden, Netherlands).

Two human cell lines were also used, these being a human embryonic cell line, HEC (isolated in this laboratory), and GM0743 (Yoshida & Miwa, 1974) which is deficient in phosphoglycerate kinase.

E.Coli bacterial strains (kindly provided by Professor W.J.Brammar.) were used for the preparation of plasmid DNA, creating genomic libraries, and cloning.

2.1.2 Enzymes and Nucleic acids.

Restriction enzymes were from Bethesda Research Laboratories (UK) Ltd. (Cambridge, England), and Pharmacia Ltd. (Milton Keynes, England). Klenow polymerase was from Pharmacia Ltd., or prepared by Marion Sweeney (this department). DNA polymerase I (Kornberg polymerase), Deoxyribonuclease I, Proteinase K, T4 DNA ligase and calf intestinal alkaline phosphatase were from Boehringer Mannheim GmbH (West Germany), pancreatic ribonuclease
and lysozyme from Sigma Chemical Co. Ltd. (Poole, England) and Pronase from Calbiochem-Behring Corp. (La Jolla, U.S.A.).

Salmon sperm DNA was from Sigma Chemical Co, Ltd. and yeast transfer RNA from Boehringer Mannheim GmbH.

17-mer oligonucleotides, used as sequencing primers, were synthesized by John Keyte (this department).

2.1.3 Chemicals.

Ham's F12 medium (Ham, 1965) and F12D medium (Kao & Puck, 1974) were from Flow Laborotaries (Rickmansworth, England), and foetal calf serum and Genticin (G418) were from Gibco Europe Ltd. (Paisley, Scotland). Tryptone, yeast extract and agar were from Difco Laborotaries (Detroit, U.S.A.) and trypcticase from BBL Microbiology Systems, Becton Dickinson & Co. (U.S.A.). Agarose was from Bethesda Research Laborotaries (UK) Ltd. Ficoll 400 and sephadex were from Pharmacia Fine Chemicals AB (Uppsalal, Sweden). Acrylamide, urea and SDS were from Serva, Feinbiochemica (Heidelberg, West Germany); ammonium persulphate, TEMED, 2-mercaptoethanol, TCA, PVP and Amberlite monobed resin (MB-1) were from BDH Chemicals Ltd. (Poole, England). Radiochemicals were from Amersham International plc. (Amersham, England). Solvents were from BDH Chemicals Ltd. or Fisons plc. (Loughborough, England). Carbon sources for cell culture were also from Fisons plc, and other fine chemicals were from Sigma Chemical Co. Ltd. and Boehringer Mannheim GmbH.
2.1.4 Standard solutions

saline D  
$\text{NaCl}(8\text{g}.\text{l}^{-1}), \text{KCl}(0.4\text{g}.\text{l}^{-1})$, 
$\text{MgSO}_4\cdot7\text{H}_2\text{O}(0.154\text{g}.\text{l}^{-1}), \text{CaCl}_2\cdot6\text{H}_2\text{O}(0.024\text{g}.\text{l}^{-1})$, 
$\text{Na}_2\text{HPO}_4(0.18\text{g}.\text{l}^{-1}), \text{KH}_2\text{PO}_4(0.15\text{g}.\text{l}^{-1})$.

saline G  
$\text{NaCl}(8\text{g}.\text{l}^{-1}), \text{KCl}(0.4\text{g}.\text{l}^{-1})$, 
$\text{Na}_2\text{HPO}_4\cdot7\text{H}_2\text{O}(0.045\text{g}.\text{l}^{-1}), \text{KH}_2\text{PO}_4(0.03\text{g}.\text{l}^{-1})$.

phage buffer  
$\text{KH}_2\text{PO}_4(3\text{g}.\text{l}^{-1}), \text{Na}_2\text{HPO}_4(7\text{g}.\text{l}^{-1}), \text{NaCl}(5\text{g}.\text{l}^{-1})$, 
$1\text{M MgSO}_4(1\text{ml}.\text{l}^{-1}), 0.1\text{M CaCl}_2(1\text{ml}.\text{l}^{-1})$, 
$1\%$ gelatin$(1\text{ml}.\text{l}^{-1})$.

TE buffer  
$10\text{mM Tris-HCl}, \text{pH} 7.5$, $1\text{mM EDTA}$.

STE buffer  
$10\text{mM Tris-HCl}, \text{pH} 8.0$, $100\text{mM NaCl}, 1\text{mM EDTA}$.

SSC  
$0.15\text{M NaCl}, 15\text{mM Na-citrate}, \text{pH} 7.0$.

Tris-acetate  
$40\text{mM Tris.}, 20\text{mM Na-acetate}, 1\text{mM EDTA}, \text{pH} 8.2$ 
(with glacial acetic acid).

TBE buffer  
Tris$(10.9\text{g}.\text{l}^{-1})$, Boric acid$(5.5\text{g}.\text{l}^{-1})$, 
EDTA$(0.93\text{g}.\text{l}^{-1})$.

Denhardt's  
$0.2\%$ Ficoll $400,000$, $0.2\%$ BSA, $0.2\%$ PVP.

DDS  
Denhardt's solution plus $0.1\%$ SDS, 
$50\mu\text{g}.\text{ml}^{-1}$ salmon sperm DNA (alkaline denatured).

phenol/CHCl$_3$  
$250\text{ml phenol}, 250\text{ml chloroform}, 10\text{ml isoamyl}$ 
alcohol, equilibrated with $10\text{mM Tris-HCl}, \text{pH} 7.5$.

Restriction 
buffers:

Low salt  
$10\text{mM Tris-HCl}, \text{pH} 7.5$, $10\text{mM MgCl}_2$.

Medium salt  
$50\text{mM NaCl}, 10\text{mM Tris-HCl}, \text{pH} 7.5$, $10\text{mM MgCl}_2$.

High salt  
$100\text{mM NaCl}, 50\text{mM Tris-HCl}, \text{pH} 7.5$, $10\text{mM MgCl}_2$.

Sma I buffer  
$15\text{mM Tris-HCl}, \text{pH} 7.5$, $60\text{mM MgCl}_2$, $15\text{mM KCl}$.
2.1.5 Standard bacterial culture media.

L-broth  
D/B Tryptone(10g.1\textsuperscript{-1}), D/B yeast extract (5g.1\textsuperscript{-1}), NaCl(5g.1\textsuperscript{-1}), glucose(1g.1\textsuperscript{-1}).

L-agar  
D/B Tryptone(10g.1\textsuperscript{-1}), D/B yeast extract (5g.1\textsuperscript{-1}), NaCl(5g.1\textsuperscript{-1}), D/B agar(17g.1\textsuperscript{-1}).

BBL-agar  
BBL Trypticase(10g.1\textsuperscript{-1}), NaCl(5g.1\textsuperscript{-1}), D/B agar (10g.1\textsuperscript{-1}).

BBL-top agar  
BBL Trypticase(10g.1\textsuperscript{-1}), NaCl(5g.1\textsuperscript{-1}), D/B agar (or agarose)(6.5g.1\textsuperscript{-1}).

M9 minimal agar  
D/B agar(15g.1\textsuperscript{-1}), \(\text{Na}_2\text{HPO}_4\)(6g.1\textsuperscript{-1}), \(\text{KH}_2\text{PO}_4\)(8g.1\textsuperscript{-1}), NaCl(0.5g.1\textsuperscript{-1}), \(\text{NH}_4\text{Cl}\)(1g.1\textsuperscript{-1}), 1M MgSO\textsubscript{4}(2ml.1\textsuperscript{-1}), 1M CaCl\textsubscript{2}(0.05ml.1\textsuperscript{-1}).
2.2 Methods.

2.2.1 Cell culture.

Chinese hamster ovary (CHO) cells were grown as monolayers in Ham's F12 medium (Ham, 1965) supplemented with 4% foetal bovine serum or 6% foetal bovine macroserum (Faik & Morgan, 1977a) and carbon sources added to the required concentration (usually 10mM). Culture dishes were incubated at 37 °C in an humidified atmosphere of 5% CO₂ in air.

Cells were sub-cultured by aspirating the medium, washing the cells twice with saline D, and incubating them at 37 °C, in 0.025% trypsin until the cells became rounded (approx. 5 min.). The cells were then washed off the surface of the plate and suspended in culture medium before aliquots were dispensed into new plates containing fresh medium.

2.2.2 Preparation of macroserum.

The macromolecular components of foetal bovine serum were prepared by passing 400 ml of serum through a sephadex G-50 column (150 X 450 mm) and eluting with 4 litres of saline solution (Kao & Puck, 1974). The diluted eluant (approx. 800 ml) was concentrated back to 400-600 ml using Sephadex G-25, sterilised by filtration, and stored at -20 °C. Each batch of macroserum was tested, in a single cell plating experiment, for the absence of glucose by its inability to support the growth of CHO-K1 cells except in the presence of added glucose.
2.2.3 Cloning CHO cells.

Large colonies containing at least fifty cells were located and their positions marked. The medium was removed, the cells washed twice with saline D, and sterile cloning cylinders placed over well separated colonies.

The cells were then trypsinised as above and transferred to 35mm plates containing fresh medium.

2.2.4 Single-cell plating.

This method was used to determine the cytotoxicity of various substances or conditions.

A cell suspension was diluted to 5,000 cells per ml, and 0.1 ml aliquots were dispensed into a number of culture dishes containing culture medium. The cells were then allowed to plate out overnight before the medium was removed and replaced by fresh medium containing the substance under investigation at a number of different concentrations. After 7 to 10 days of incubation, during which time the plates were not moved, the medium was poured off and the cells were fixed with 10% formalin and stained with 0.01% crystal violet. Colonies (of 50 cells or more) were then counted.

2.2.5 Glycerol cultures.

Bacteria were grown up in 10ml L-broth (containing appropriate antibiotics), with shaking, at 37 °C, for at least 5 to 6 hr. The cells were spun down at 3,000 r.p.m. for 5-10 min. in a bench centrifuge, the supernatant was removed, and the cells were suspended in 1 ml of 10mM MgSO$_4$ by vortexing. The bacteria were then transferred to a small glass bottle to which 2 ml of sterile glycerol was added. This was mixed by
vortexing, and stored at -20 °C.

2.2.6 **Plasmid preparation** (chloramphenicol amplification)

A single colony of *E. coli* was used to inoculate 25ml of L-broth (containing appropriate antibiotics). This "starter" culture was grown up, at 37 °C, and used to inoculate 800ml of L-broth. When the absorbance of the broth, at 650nm, reached 0.8, the bacteria were in the late log-phase of growth, and chloramphenicol was added to a final concentration of 200 μg.ml⁻¹. The incubation was continued overnight.

The cells were pelleted, the supernatant poured off and the pellet resuspended in 25% sucrose, 0.05M Tris-HCl pH 8.0 (8 ml), on ice. The cells were lysed by mixing with 1.5 ml lysozyme (10 mg.ml⁻¹) for 3 min., 3 ml of 0.4M EDTA, pH 8.0 were then added, and after a further 3 min., 12 ml of 2% Triton X100, 50mM Tris-HCl pH 8.0, 0.0625M EDTA were mixed in. The lysate was centrifuged to pellet the cell debris, and the supernatant was collected. Caesium chloride was dissolved in the supernatant (19.25g per 20 ml) and ethidium bromide (1.25 ml of 10 mg.ml⁻¹ solution per 20 ml supernatant) was mixed in, before centrifuging at 39,000 r.p.m. for 60 hr. at 20 °C.

The centrifuge tubes were observed under U.V. light and the lower fluorescent DNA band (plasmid DNA) was drawn out through a needle.

The ethidium bromide was extracted by shaking with CsCl-sat. isopropanol, the extraction being repeated two or three times. The DNA was transferred to dialysis tubing (previously boiled for 20 min. in 0.2M EDTA) and dialysed against two litres of 10mM Tris-HCl, pH 8.0, 1mM EDTA for 5 hr. with four changes of
buffer.

The preparation was phenol-extracted and, after discarding the phenol, was extracted three times with ether. Finally, the DNA was ethanol-precipitated, washed with 70% ethanol, and dissolved in 1mM Tris-HCl, pH 8.0, 0.1 mM EDTA. The concentration of DNA was determined by measuring the absorbance (A) at 258 nm. (DNA concentration = A_{258nm} \times 50 (\mu g \cdot ml^{-1}).)

2.2.7 Plasmid preparation (no amplification).

10 ml of L-broth (containing appropriate antibiotics) were inoculated with a colony of bacteria and incubated, with shaking, at 37 °C, for 6-7 hr. or overnight. A few ml. of starter culture were added to 200 ml. volumes of L-broth (with appropriate antibiotics) in 21 baffled flasks and incubated, with shaking, overnight at 37 °C.

The cells were pelleted by centrifuging at 6,000 r.p.m. for 8 min. at 4 °C. The supernatant was poured off and the pellets kept on ice. The cells were lysed (Birnboim & Doly, 1979) by resuspending them in 4 ml of lysis buffer (25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 50 mM glucose) containing 2 mg ml\(^{-1}\) lysozyme. After standing on ice for 5 min., the cells were transferred to a 40 ml plastic centrifuge tube to which 12 ml of alkaline SDS (0.2 M NaOH, 1% SDS) was added. This was carefully but thoroughly mixed and left for 5 min. on ice. 16 ml of ice-cold 3 M potassium acetate solution\(^1\) was then added, mixed by

\(^1\)To make up 3 M KOAc, 11.5 ml of glacial acetic acid and 28.5 ml of H\(_2\)O were added to 60 ml of 5 M KOAc. The resulting solution was 3 M with respect to potassium and 5 M with respect to acetate.
inverting sharply several times, and left on ice for a further 10 min.

The lysate was cleared by centrifugation at 8,000 r.p.m. for 10 min. at 4 °C, with the centrifuge brake off. The supernatant was carefully poured through a funnel containing glass wool, and collected in a larger plastic centrifuge tube. 0.6 volumes of isopropanol was added, mixed well, and centrifuged at 8,000 r.p.m. for 10 min. at 20 °C. The supernatant was carefully poured off and the pellet was gently rinsed in 70% ethanol, then allowed to drain and dry for 1-2 hr. The pellet was dissolved in 2 ml of sterile TE buffer and transferred to a 10 ml plastic tube. Any pellet remaining was rinsed out with TE buffer and also transferred to the tube. The volume was made up to 6 ml, and to this was added exactly 6g of caesium chloride and 0.2 ml of 10 mg.ml⁻¹ ethidium bromide. The caesium chloride was dissolved by gently shaking and leaving in a warm water-bath.

Before centrifuging, it was important to check that the refraction of this solution was less than 1.396. If not, the solution was diluted with TE buffer until this figure was reached.

The solution was transferred to 10 ml polycarbonate ultracentrifuge tubes which were then topped up with liquid paraffin and balanced. The tubes were loaded into a 10 X 10 Titanium ultracentrifuge rotor and centrifuged at 40,000 r.p.m. for 40 hr. at 20 °C.

The upper, and then the lower, DNA bands were removed under U.V. light using a syringe and needle, and the lower band was extracted at least three times with an equal volume of caesium chloride-saturated isopropanol to remove the ethidium bromide.
The preparation was then dialysed against two litres 0.1x TE buffer for at least 5 hr., changing the buffer three times. (N.B. an alternative to dialysis was to dilute the caesium chloride down at least two-fold and ethanol-precipitate.)

The dialysed solution was concentrated by ethanol-precipitation.

2.2.8 Preparation of DNA from cultured animal cells. (Jeffreys & Flavell, 1977a.)

Cells were scraped from the surface of the dishes in saline D, spun down, and resuspended in 150mM NaCl, 100mM EDTA pH 8.0 (approximately 1.5 ml per 10^7 cells). The cells were lysed by adding an equal volume of SDS, and swirling for 1 min. An equal volume of phenol/chloroform mix was added and carefully mixed until an emulsion was formed. The lysates were centrifuged in siliconised corex tubes at 10,000 r.p.m. and 20 °C (using swing-out rotor) for 8 min., and the aqueous layers pipetted off. The phenol layers were re-extracted by mixing with an equal volume of TE buffer, and re-centrifuged. The aqueous layers were pooled and mixed with 0.1 volumes of 2M sodium acetate pH 5.6 and 2.5 volumes of absolute ethanol. The solution was swirled gently until the DNA clotted. The DNA was washed with 70% ethanol, then dissolved in 5mM Tris-HCl, pH 7.5, 10mM NaCl, 0.5mM EDTA. The solution was incubated at 37 °C with pancreatic ribonuclease (50 μg.ml⁻¹) for 15 min., and then with proteinase K or pronase (100 μg.ml⁻¹) for another 15 min., before being ethanol-precipitated and re-dissolved as before. The ethanol-precipitation was repeated four times and the DNA was finally washed in 70% ethanol, vacuum-dried, and dissolved in 0.1x TE buffer.
2.2.9 Isolation of DNA from low melting point agarose gel.
(Weislander, 1979.)

The DNA to be isolated was loaded onto a 0.8% low melting point agarose gel and run slowly overnight. (Resolution was better if the gel was run in the cold at 4 °C). The gel was stained in ethidium bromide (0.2 ml of 10 mg.ml⁻¹ in 21 of distilled H₂O) and the DNA of interest was cut out of the gel under U.V. light. The gel slice was placed in a 1.5 ml microfuge tube and an equal volume of 0.1M Tris-acetate pH 7.5, 5mM EDTA, 0.5M NaCl, was added.

The gel was then gently macerated using a glass rod or sealed Pasteur pipette, and heated to between 65 and 70 °C, with occasional shaking, until the agarose melted. An equal volume of buffered phenol was added, and the mixture was shaken gently for 20 min. until a white emulsion formed. This was then centrifuged at high speed (13,000 X g) in a microfuge for 2 min. A dense white layer of agarose formed at the interface. The upper aqueous layer was removed and re-extracted with phenol, then any residual phenol in the aqueous phase was removed by extracting twice with 2 volumes of ether. Excess ether was blown off using a Pasteur pipette.

The DNA was ethanol-precipitated, centrifuged at high speed for 10 min., washed with 70% ethanol, vacuum-dried, and dissolved in TE buffer, 10mM Tris-HCl, pH 7.5, or sterile distilled water.
2.2.10 "Freeze-squeeze" extraction of DNA from agarose gel. (Tautz & Renz, 1983.)

The DNA was electrophoresed slowly in the cold through a thin 0.6% agarose gel. The gel was stained with ethidium bromide and the DNA cut from the gel under U.V. light.

The gel pieces were equilibrated in at least ten times their volume of 0.3M sodium acetate, 1mM EDTA, pH 7.0, for 30-40 min. Each gel piece was then placed in a small (0.5 ml) microfuge tube, previously punctured at the bottom and plugged with sterile, siliconised glass wool. This was frozen in liquid nitrogen for at least 10 min., then placed in a larger capless microfuge tube and centrifuged immediately at high speed for 10 min. at room temperature.

The solution remaining in the large tubes was pooled in clean tubes (up to 0.3 ml solution), centrifuged for 1 min., and the supernatant removed, carefully leaving any agarose behind. 0.01 volumes of 1M MgCl₂, 10% acetic acid, and 2.5 volumes of cold absolute ethanol were added to the solution, mixed, and frozen for 20 min. in dry ice / IMS. The DNA was centrifuged for 10 min. at high speed, vacuum-dried briefly, and dissolved in Tris-HCl buffer or H₂O.

2.2.11 DNA-mediated gene transfer. (Graham & Van der Eb, 1973; Graham et al., 1979.)

Preparation of DNA. The DNA to be used in the experiment was mixed with carrier DNA (if required) and then ethanol-precipitated and pelleted by centrifugation. After washing with 70% ethanol, and another short centrifugation, the DNA was dissolved in 0.1x TE buffer to a concentration of 80 μg.ml⁻¹.

Transfection. 100 mm petri dishes were seeded (5-7 × 10⁵
cells/plate) the day before DNA was to be added to the plates. The next morning the medium was changed, and three hours later the calcium phosphate-DNA co-precipitate was added to the plates.

The co-precipitate was formed as follows:

For each plate to be used, the appropriate amount of DNA was made up to 0.44 ml with distilled H₂O.

'Solution B' was formed by adding 0.06 ml of 2M CaCl₂ for every 0.44 ml of DNA solution, and mixing rapidly. 'Solution A' was formed by mixing 0.5 ml of Hepes buffered saline, pH 7.1, to 0.01 ml of 100x phosphate solution (1 : 1 of 70mM Na₂HPO₄: NaH₂PO₄) for each plate to be used. 'Solution B' was added dropwise to 'solution A' whilst bubbling air through to mix them. The precipitate thus formed was left to stand at room temperature for 35 min.

Approximately three hours after changing medium, the precipitate was added to the plates (1ml per plate). Three and a half hours later, 1.5 ml of dimethylsulphoxide (DMSO) was added dropwise to each plate (to give a final concentration of 10% DMSO) whilst gently swirling the medium. After 25 min. exposure to DMSO, the medium was removed and replaced with fresh medium containing 0.1 mg.ml⁻¹ kanamycin to prevent bacterial contamination.

Selection. The next day the cells were sub-cultured: cells from each 100mm culture dish were divided between two 150 mm dishes containing non-selective medium. Selection was applied 45 hr. after the DNA was added. The medium was replaced with fresh selective medium every four days, taking care to move the culture dishes as little as possible to minimise the formation
of satellite colonies. Colonies usually become evident about two weeks after transfection.

2.2.12 **Transfer of DNA from agarose gel to nitrocellulose filter.** (Southern, 1975.)

Digested DNA was electrophoresed through a 0.8% agarose gel, and then stained with ethidium bromide. The gel was observed using a strong U.V. light source, and photographed alongside a scale. The area of the gel to be blotted was marked out and then carefully cut out.

The gel was gently washed in 0.25M HCl for 7.5 min., drained, and washed again in HCl for the same length of time. It was then quickly rinsed twice in distilled water to remove excess acid, and washed in 0.5M NaOH, 1.5M NaCl, 1mM EDTA for two 30 min. periods to denature the DNA. After a brief wash in H$_2$O, the gel was neutralized by washing in 0.5M Tris-HCl pH 7.5, 3M NaCl for 4 x 15 min., followed by a final wash in 20x SSC for 20 min.

The transfer apparatus was set up as follows: A piece of "Scotchbrite" (or in earlier blots, a sheet of glass supported by rubber bungs) was placed in a large tray, and 20x SSC was poured in. A sheet of Whatmann 3MM paper was soaked in 20x SSC and placed over the "Scotchbrite" (or over-hanging the glass plate to act as a wick). The paper and tray were covered with cling-film, with a space cut out, over which the gel was placed such that it slightly overlapped the cling-film around its edges. Air bubbles were always avoided. A piece of Sartorius nitrocellulose filter (0.45 μm pore size, cut slightly larger than the gel) was soaked in 3x SSC and carefully placed over the gel, and over this was placed a sheet of 3MM paper soaked in 3x
SSC, and 5 dry sheets. A wad of paper towels was placed over this, and a weight (a flask containing 100 ml. of water) resting on a glass plate was placed on top. The DNA is transferred to the nitrocellulose filter by the action of 20x SSC being drawn up through the gel and filter.

The transfer was continued overnight, after which the nitrocellulose filter was washed in 3x SSC for 20 min., blotted dry, and baked for 2-4 hr. at 80 °C.

2.2.13 Hybridisation of nitrocellulose-bound DNA with radiolabelled probe. (Jeffreys & Flavell, 1977b.)

The nitrocellulose filter to which the DNA of interest was bound was cut up into strips to fit the hybridisation chambers. The strips were first wetted in 3x SSC, then incubated in degassed 3x SSC (usually at 65 °C) for 30 min., and the hybridisation chamber gently rocked in a 65 °C water-bath. The filters were transferred into pre-warmed, degassed, Denhardt’s solution (Denhardt, 1966) for 3 hr., then into 1x DDS for 1 hr.

The radioactive probe was boiled for 5-10 min., and added to another hybridisation chamber containing degassed, pre-warmed 1x DDS, to which the filters were then transferred. This was incubated in a water bath overnight, with gentle rocking, to allow the probe to bind to the DNA on the filters.

Next day the filters were transferred into pre-warmed 1x DDS and washed for 10-15 min. This wash was repeated six times. The filters were then washed in pre-warmed 0.3x SSC containing 0.1% SDS and 50 μg.ml⁻¹ alkaline denatured salmon sperm DNA, for two 10 min. periods (stringency washes). Finally, the filters were washed in 3x SSC for 20 min. before being blotted onto 3MM paper to remove excess solution, and dried for 30-45 min. at
37 °C.

The filter was reconstructed as in the original blot and stuck down onto filter paper. This was covered with cling-film and placed in a photographic cassette case with a sheet of X-ray film and an intensifying screen, and left at -70 °C.

The film was developed after an appropriate period of time (anything from a few hours to over a month).

2.2.14 Plaque transfer and in situ hybridisation of plaques.

(Benton & Davies, 1977.)

Phage were plated out on L-agar and incubated at 37 °C overnight. The plates were then left for at least 30 min. at 4 °C to harden the top-layer agar. Meanwhile 85mm Sartorius nitrocellulose filter circles (0.45 μm pore size) were labelled and marked with asymmetrically placed dots so that the plates could later be orientated relative to the filters. A filter was placed on a plate for 15 min., and the position of the dots marked on the bottom of the plate. The filter was then carefully inverted (phage uppermost) onto a damp pad of 3MM papers soaked in 0.5M NaOH, left for 10 min. to denature the phage, then transferred to a pad of 1M Tris-HCl, pH 7.5 for 2 min., then to another of the same, and finally to a pad of 1M Tris-HCl, pH 7.4, 1.5M NaCl for 15 min., before blotting, drying briefly at room temperature, and baking at 80 °C for 2 hr.

Prior to incubation with the probe, the filters were washed in degassed hybridisation solution (6x SSC, 0.1% Ficoll 400,000, 0.1% BSA, 0.1% PVP), at the hybridisation temperature, for at least 4 hr. This reduced non-specific binding of the radioactive probe.

The probe was denatured by boiling for 5-10 min., and
immediately added to a hybridisation chamber containing degassed, pre-warmed, hybridisation solution. The filters were then added to the chamber and incubated overnight.

The filters were washed in pre-warmed 3x SSC, 0.1% SDS for 45 min. This wash was repeated six times until a decrease in radioactivity of the filters could no longer be detected using a Geiger counter. This was then followed by 2 x 30 min. stringency washes. The filters were blotted, arranged on a sheet of 3MM paper, and at least three positions were marked on the paper with radioactive ink. The sheet was covered with cling-film and placed in a cassette case with X-ray film and an intensifying screen, and left at -70 °C. The film could usually be developed the next morning. The positions of the dots on the filters were then marked on the film, the plates positioned correctly on the film, and positive plaques were picked.

2.2.15 Phage titrations.

The bacteria were grown in L-broth to an absorbance of 0.6 at 650nm, spun down in a bench-top centrifuge (3,000 r.p.m. for 5 min.), and suspended in half their original volume of 10mM MgSO$_4$.

100μl of various dilutions of phage suspension (for example $10^{-2}$, $10^{-4}$, $10^{-6}$ and $10^{-7}$) were added to 100μl of bacterial cells, mixed, and left to stand at room temperature for 10-15 min.

2-3 ml. of melted BBL top-layer agar (at 55 °C) were added to this, mixed, and then poured onto fairly thick BBL plates. The plates were left at room temperature for 10 min. to allow the agar to set, before being incubated overnight at 37 °C.

Plaques were counted the next morning to calculate the
number of phages, or plaque-forming units (pfu), in the original phage suspension.

2.2.16 **Phage stock (plate lysate).**

A single plaque was picked using a toothpick or Pasteur pipette, and vortexed in 1 ml of phage buffer. 100 μl of this suspension was mixed with 100 μl of cells in the logarithmic phase of growth, and left at room temperature for 10-15 min. (A drop of chloroform was added to the remaining phages to prevent bacterial growth). 2-3 ml of molten BBL top-layer agar was added, mixed, poured onto an L-agar plate, and incubated overnight at 37 °C.

2-3 ml of phage buffer were pipetted onto the plate, and the top-layer agar was scraped off into a sterile plastic tube. This was centrifuged at 3,000 r.p.m. for 10 min. to pellet the agar debris, the supernatant was poured off, and a drop of chloroform was added. This phage stock was stored at 4 °C. Phage titrations were performed to determine the pfu.ml⁻¹ of all phage stocks.

2.2.17 **Lambda phage preparation - liquid lysate.** (Blattner et al., 1977.)

A small volume of L-broth was inoculated with a bacterial colony, and when the cells were in the logarithmic phase of growth, a few ml of the cells were added to 200 ml of L-broth, containing 10mM MgSO₄, in a 21 flask. The bacteria were incubated, with vigorous shaking, at 37 °C to an absorbance of 0.3 at 650nm (equivalent to 1-2 X 10⁷ cells.ml⁻¹), when 2-4 X 10⁹ phage were added (i.e. 1 phage : 1 bacteria). After one hour, 2 X 20 ml of culture were diluted 10-fold into 2 X 21
flasks of L-broth (+ 10mM MgSO$_4$), and incubated at 37 °C overnight with vigorous shaking.

The next morning, the remaining cells were lysed by adding 3 ml of chloroform to each flask, and shaking was continued for a further 20 min. Cell debris was removed by centrifugation at 10,000 r.p.m. for 10 min. at 4 °C. An aliquot of the supernatant was taken for a phage titration.

2.2.18 Lambda phage DNA preparation. (Yamamoto et al., 1970.)

4g of NaCl and 10g of PEG were added to a 200ml liquid lysate of at least 2 X 10$^9$ pfu.ml$^{-1}$, and dissolved by gentle swirling. This was left at 4 °C overnight (or on ice for 4hr.), for the phages to precipitate, and was then centrifuged at 10,000 r.p.m. for 10 min. at 4 °C. The supernatant was poured off and the pellet was covered with 2-3ml of phage buffer. The pellet was resuspend by shaking gently on ice.

A saturated solution of caesium chloride (CsCl) was made up by dissolving 65g of CsCl in 35g of H$_2$O, and autoclaving. This was used to make up three solutions of different densities, with phage buffer, in order to prepare a block gradient as follows:

<table>
<thead>
<tr>
<th>CsCl</th>
<th>Phage buffer</th>
<th>Density (g.ml$^{-1}$)</th>
<th>Refractive index</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.7ml</td>
<td>12.3ml</td>
<td>1.3</td>
<td>1.37</td>
</tr>
<tr>
<td>11.1ml</td>
<td>8.9ml</td>
<td>1.5</td>
<td>1.384</td>
</tr>
<tr>
<td>14.0ml</td>
<td>6.0ml</td>
<td>1.7</td>
<td>1.395</td>
</tr>
</tbody>
</table>

The gradient was made up in a 15ml polypropylene centrifuge tube. Firstly, 2.5ml of the 1.7g.ml$^{-1}$ CsCl solution was pipetted into the tube. The other two solutions (1.5g.ml$^{-1}$, then 1.3g.ml$^{-1}$) were carefully layered on top, and the
resuspended phages were layered on top of the gradient. The volume was made up with phage buffer to just below the top of the tube. The gradient was centrifuged at 24,000 r.p.m. for 2 hr. at 20 °C in a SW 6 X 16.5 rotor. The phage band was observed just below the 1.3g.ml⁻¹ and 1.5g.ml⁻¹ interface, and was withdrawn through a syringe needle inserted through the side of the tube.

The suspension of phage was dialysed against TE buffer for an hour at 4 °C. Heat-treated ribonuclease (RNase) was added to the dialysis bag to a concentration of 20μg.ml⁻¹, and the dialysis continued against fresh TE buffer at room temperature for 1 hr. Pre-digested pronase² was added to the phage suspension to a concentration of 1mg.ml⁻¹, and dialysis continued against pronase buffer (20mM Tris-HCl, pH 8.0, 1mM EDTA, 100mM NaCl, 0.01% Triton X100) at 37 °C, for a further hour.

The dialysed solution was phenol-extracted twice, and centrifuged at 10,000 r.p.m. for 8 min. in corex tubes. The aqueous layer was ether extracted three times, excess ether blown off, and was ethanol-precipitated. The DNA normally precipitated at room temperature, but could be frozen at -20 °C for 20 min. to ensure better precipitation. The DNA was pelleted (10,000 r.p.m., for 20 min. at 0-4 °C), the supernatant poured off, and the DNA rinsed with cold 70% ethanol. The DNA pellet was then drained, vacuum-dried, and

²Pronase was made up at 20mg.ml⁻¹ in 20mM Tris-HCl, pH 8.0, 1mM EDTA, 100mM NaCl, and pre-digested by incubating at 37 °C for an hour. It was stored at -20 °C.
dissolved in TE buffer, 10mM Tris-HCl pH 7.5 or distilled H₂O. The concentration was determined by measuring its absorbance at 258nm (see section 2.2.6).

2.2.19 Preparation of alkaline-denatured salmon sperm DNA.
(W.J. Brammar, pers. comm.)

1 g of Sigma type III salmon sperm DNA was dissolved in 500ml of distilled H₂O by heating in a boiling water bath. To this, 20ml of 0.5M EDTA, pH 8.0, was added. When the solution was clear, 15ml of 10M NaOH was added, with stirring, and the pH checked with pH paper (pH should be 14). After a further 20 min. in the boiling water bath, the DNA solution was cooled on ice and 20ml of 1M Tris-HCl, pH 7.5, was added. Concentrated HCl was added slowly, with stirring, until the pH was between 7 and 8.

The solution was extracted with 0.5 volumes of phenol/chloroform, and ethanol-precipitated by adding 2 volumes of cold absolute ethanol. The DNA was centrifuged out at 10,000 r.p.m. for 30 min. at 4 °C in a 6 X 250ml rotor. The precipitate was washed with 70% ethanol, vacuum-dried, and dissolved in 40ml of distilled H₂O. The DNA was made up to 1x SSC by the addition of 10ml of 5x SSC. The DNA concentration was determined by measuring the absorbance at 258nm. (For double-stranded DNA, an absorbance of 1.0 is equivalent to 50µg.ml⁻¹, but for singlestranded DNA it is equivalent to approximately 40µg.ml⁻¹.)
2.2.20 M13 template DNA preparation (mini-prep). (Schreir & Cortese, 1979.)

M13 plaques were picked into 1.5 ml aliquots of L-broth in 10 ml sterile plastic tubes, and shaken vigorously at 37 °C for 5-6 hr. Each culture was then poured into a 1.5 ml microfuge tube, and centrifuged for 5 min. at high speed. The supernatant was transferred into a clean tube and centrifuged again for 5 min. The supernatant was removed (taking only 0.8 ml so as not to disturb any pelleted bacteria) and 0.2 ml of 20% PEG, 2.5M NaCl was added. This was left at room temperature for 45 min. with occasional mixing, and then centrifuged at high speed for 5 min. The supernatant was flicked off, and the tube centrifuged for a few seconds to bring the remaining liquid to the bottom. All traces of the PEG supernatant were removed using a drawn out capillary tube. The PEG pellet was clearly visible.

The pellet was dissolved in 100 μl of 10mM Tris-HCl, pH 7.5, 0.1mM EDTA, and 50 μl of buffered phenol/chloroform was added. This was vortexed for 10 sec., left to stand for 5 min., vortexed again, and centrifuged for 1 min. at high speed. The aqueous layer was removed, the phenol layer re-extracted with 100 μl of 10mM Tris-HCl, pH 7.5, 0.1mM EDTA, and the pooled aqueous layers phenol extracted 2 or 3 times again. After three chloroform extractions and three ether extractions, the solution was ethanol-precipitated by the addition of 20 μl of 3M Na-acetate, pH 5.5, and 0.5 ml absolute ethanol, and frozen in a dry ice/IMS bath. The tube was centrifuged for 10 min. in a microfuge, all the supernatant removed, and the DNA pellet dried, and dissolved in 25μl of 10mM Tris-HCl, pH 7.5, 0.1mM EDTA.
The yield of the template was estimated by running a small aliquot (usually 5 µl) through a 0.7% agarose gel.

2.2.21 Determination of insert orientation in M13 recombinants (C-test). (Messing, 1983.)

Single stranded M13 phage DNA cannot hybridise to itself. However, if DNA fragments are ligated into the double-stranded DNA of two phages (the replicative form) in opposite orientations, the single-stranded form of one will contain the strand of the insert which is the complement of the insert strand of the other. In this case, the single stranded phages will hybridise to each other, and can be distinguished from non-hybridising phage DNA by the slower mobility of their DNA through an agarose gel.

To determine whether identical inserted DNA fragments in two M13 recombinant phages were in the same orientation, they were mixed and allowed to hybridise as follows: 10 µl of supernatant from liquid cultures of both M13 phages were added to 10 µl of 10x medium salt 'restriction buffer' and 3 µl of 1% SDS, carefully but thoroughly mixed, and overlayed with 20 µl of paraffin oil. After incubating at 65 °C for one hour, the reaction mixes were electrophoresed through an agarose gel.

2.2.22 Large scale M13 DNA template preparation. (Schrier & Cortese, 1979.)

100 µl of an overnight culture of JM101 and 50 µl of M13 phage supernatant from a mini template preparation were added to 20 ml of L-broth. This culture was incubated for 5-6 hr., at 37 °C, with vigorous shaking.

The bacteria were pelleted by centrifugation at 10,000
r.p.m. for 8 min. at 20 °C. The supernatant was collected (care
being taken not to disturb the bacteria), and centrifuged as
before. 1 ml of supernatant was removed and stored at -20 °C,
and most of the remaining solution was carefully removed and
mixed with 4 ml of 20% PEG, 2.5 M NaCl. This was kept at room
temperature for 45 min., with occasional mixing (during which
time the solution became cloudy), then centrifuged at 10,000
r.p.m. for 8 min. at 20 °C. All the PEG solution was removed
and the pellet dissolved in 200 μl of TE buffer, and transferred
to a microfuge tube.

This solution was phenol extracted, extracted with
chloroform, ether extracted, then ethanol-precipitated, and
frozen in dry ice/IMS for at least 15 min. The DNA was
pelleted, dried, and dissolved in 100 μl of 10mM Tris-HCl,
ph 7.5, 0.1mM EDTA.

2.2.23 Preparation of M13 replicative form (RF) DNA.

JM101 were transformed with M13 DNA (see section 2.2.32) and
plated out on BBL agar plates. Next morning, a well-separated
blue plaque was picked into 10 ml of dilute JM101 (50 μl of an
overnight culture in 10 ml of L-broth). After overnight
incubation at 37 °C, the bacteria were pelleted by
centrifugation at 10,000 r.p.m. for 10 min., and the supernatant
containing the M13 phage was collected.

500 ml of L-broth were inoculated with a few ml of an
overnight culture of JM101, and when the absorbance at 650nm
reached 0.5, the 10 ml of supernatant were added. After shaking
vigorously at 37 °C for one and a half hours, chloramphenicol
was added to a concentration of 250 μg.ml⁻¹. After a further
90 min. incubation, the cells were harvested by centrifuging at
8,000 r.p.m. for 10 min. at 4 °C.

The cells were lysed and the DNA preparation continued as in a standard plasmid preparation (see section 2.2.7).

2.2.24 Radioactive labelling of DNA by 'nick-translation'.

(Jeffreys & Flavell, 1977a.)

0.1 μg of double stranded DNA was incubated with 20 μCi of [α-32p]dCTP in the presence of 4 μM dATP, 4 μM dGTP, 4 μM dTTP, 50mM Tris-HCl, pH 7.8, 5mM MgCl₂, 10mM β-mercaptoethanol, 8 pg deoxyribonuclease I (DNase I) and 6 units of DNA polymerase (Kornberg polymerase), in a total volume of 25 μl, at 15 °C for 1 hr. To stop the reaction, 25 μl of nick quench mix (10mM Tris-HCl, pH 7.5, 10mM EDTA, 0.5% SDS) and 25 μl of phenol/chloroform mix were added, the tube was flicked to mix the contents, and then centrifuged for 1 min. The aqueous layer was collected, and the phenol layer was re-extracted with TE buffer. The pooled aqueous layers were ethanol-precipitated with 20 μl of 5mg.ml⁻¹ salmon sperm DNA. The ethanol-precipitation was repeated, and the labelled DNA was dissolved in 0.5 ml of TE buffer.

2.2.25 5'-end labelling of oligonucleotides. (Maxam & Gilbert, 1980.)

10 picomoles of DNA (oligonucleotide) were mixed in a microfuge tube with 5 μl of 10x T4 kinase buffer (0.5M Tris-HCl, pH 7.5, 1mM EDTA, 0.1M MgCl₂, 10mM dithiothreitol, 1mM spermidine), 3 μl of [γ-32p]ATP (10μCi.μl⁻¹), and made up to 49 μl with distilled H₂O. Finally, 1 μl (10 units) of T4 polynucleotide kinase (ATP : 5'-dephosphopolynucleotide 5'-phosphotransferase) was added. Following a 30 min.
incubation at 37 °C, the reaction was stopped by the addition of 1 μl of 10% SDS and 1 μl of 0.5M EDTA, pH 7.0. 50 μl of phenol (saturated with 10mM Tris-HCl, pH 7.5) was then added, the tube was vortexed for 10 sec., then centrifuged for 2 min. The phenol layer was re-extracted with 50 μl of TE buffer and the pooled aqueous phases were ether extracted. This was ethanol-precipitated by adding 1 μl of yeast tRNA (10mg.ml⁻¹), 0.1 volumes of 2M Na-acetate pH 5.0, and 2-3 volumes of absolute ethanol. After freezing in dry ice/IMS for 15 min., the tube was centrifuged for 15 min., and the pellet rinsed with 70% ethanol, vacuum-dried for 5-10 min., and dissolved in 0.5 ml H₂O. (5 μl was removed, and the amount of label incorporated was determined using a scintillation counter.) The probe was stored at -20 °C.

2.2.26 Determination of counts incorporated into labelled DNA.

The counts incorporated into the DNA were determined by precipitating the DNA with TCA as follows:

100 μg of salmon sperm DNA was added to a 5 μl aliquot of the labelled DNA in a sterile plastic tube. After the addition of 5 ml of ice-cold 10% TCA, this was left on ice for 30 min. The solution was filtered through a Whatmann glass-fibre filter (GF/C, 2.5 cm circle), the tube rinsed twice again with 5 ml of 10% TCA, and the washings filtered. The filter was then washed through with 5 ml of absolute ethanol, dried well under an infra-red lamp, and transferred to a scintillation vial with 5 ml of scintillation fluid (Fisofluor 3). The counts incorporated were determined using a scintillation counter.
2.2.27 Gel electrophoresis of DNA. (Sharp et al., 1973.)

For a 0.7% gel on a large gel plate (approximately 20cm X 20cm), 1.4 g of agarose was added to 200 ml of 1x Tris-acetate buffer and heated on a hot plate/stirrer. When the agarose had boiled, it was allowed to cool to 50 °C, then poured onto a gel plate which had previously been sealed at the ends with tape, and had a comb positioned about 2 mm above the surface. The gel was then allowed to set for at least 30 min. (If low melting-point agarose was used, the gel was allowed to set at 4 °C.) Before loading DNA samples into the wells formed by the comb, a few µl of loading buffer were mixed with the DNA samples in microfuge tubes.

Loading buffer was made by dissolving 20 mg of agarose at 100 °C in 10 ml of 10mM Tris-HCl, pH 7.5, 20mM EDTA, 10% glycerol, and 0.025% bromophenol blue. This was then cooled at 4 °C, and when set, was forced several times through a syringe needle to make a fine slurry.

2.2.28 Restriction endonuclease digestion of DNA.

DNA was digested as recommended by the manufacturers of the restriction enzymes used. The digestions were carried out in the minimum appropriate volume of solution. The DNA was added to a reaction mix containing a standard restriction buffer, DTT at the required concentration (usually 1mM), and nuclease-free BSA at a concentration of 100 µg.ml⁻¹. Finally, the restriction enzyme was carefully mixed in and incubated, generally at 37 °C, for an appropriate length of time (half an hour to several hours). To ensure complete digestion of DNA, the amount of enzyme used and the duration of digestion were increased to give, in combination, a 5-fold 'excess' over the recommended
conditions. The enzymes were usually inactivated by heating at 65 °C for 10 min.

2.2.29 Alkaline phosphatase treatment of DNA. (Seeburg et al., 1977; Ullrich et al., 1977.)

The DNA was made up to 18 μl with distilled H₂O, and 2 μl of 10x phosphatase buffer (0.1M Tris-HCl, pH 9.2, 1mM EDTA) was added. 0.75 μl of calfintestinal alkaline phosphatase (1 unit.μl-1) was then mixed in, and the reaction incubated at 37 °C for 30 min. Another 0.75 μl of alkaline phosphatase was then added, and the incubation continued for a further 20 min. To inactivate the phosphatase, 20 μl of H₂O, 5 μl of 10x STE buffer and 2.5 μl of 10% SDS were added, and this was incubated at 68 °C for 15 min. The volume was then made up to 100 μl with H₂O and extracted twice with phenol, and four times with ether, then ethanol-precipitated, and dissolved in 20 μl of sterile distilled H₂O.

2.2.30 Ligation of linear DNA fragments. (Sugino et al., 1977.)

10-20 ng of linearised vector was mixed with a three-fold molar excess of insert DNA fragments. If the concentration of the DNA was not known, a number of ratios of insert to vector was used to determine the optimal value.

1 μl of TM buffer, 1 μl of 10mM DTT, 0.5 μl of nuclease-free BSA (2μg.μl-1), and 1 μl of 10mM ATP, was added to the DNA, and this was made up to 10 μl with distilled H₂O. One unit of T4 DNA ligase per μg of DNA was then added, mixed in using the end of the pipette tip, and incubated at room temperature for 4 hr., or at 12 °C overnight.
2.2.31 *In vitro* packaging of ligated DNA. (Enquist & Sternberg, 1979.)

Packaging reactions were carried out by mixing the ligated DNA with two preparations of phage extracts, each preparation lacking an essential component required for *in vitro* phage assembly, but together containing all the proteins necessary for the formation of viable phage. In this protocol the two preparations were a Freeze-Thaw Lysate (FTL) and a Sonic Extract (SE), both kindly provided by Ian Charles of this department.

The components of the reaction mix were mixed, in the following order, in a 1.5 ml microfuge tube as follows: 7 μl of buffer A (20 μl of 1M Tris-HCl, pH 8.0, 3 μl of 1M MgCl₂, 0.5 μl of β-mercaptoethanol, 10 μl of 0.1M EDTA, pH 7.0, 966 μl of H₂O), DNA (in 1-2 μl), 1 μl of buffer B (6 μl of 1M Tris-HCl, pH 7.5; 300 μl of 0.1M spermidine-3HCl, pH 7.0, and 300 μl of 0.2M putrescine, pH 7.0, each neutralised with Tris base; 18 μl of 1M MgCl₂, 15 μl of 0.1M ATP, pH 7.0 neutralised with NH₄OH, 2 μl of β-mercaptoethanol, 224 μl of H₂O), 3.5 μl of sonic extract, and 5 μl of freeze-thaw lysate. After centrifuging for a few seconds to spin the contents to the bottom of the tube, the reaction was incubated at room temperature for 90 min. The reaction was stopped by adding 230 μl of phage buffer followed by 10 μl of chloroform. A phage titration was then carried out to determine the number of viable phage recovered.
2.2.32 Transformation of bacteria. (Mandel & Higa, 1970.)

0.1 ml of an overnight culture of bacteria\(^3\) was diluted into 30 ml of L-broth, and grown at 37°C, with shaking, until the absorbance at 650nm was approximately 0.5. 1.5 ml aliquots of the culture were dispensed into microfuge tubes and centrifuged at high speed for 10-12 sec. The supernatant from each tube was discarded, and the pellet resuspended in 0.5 ml of buffer, pH 7.0 (0.1M MOPS pH 7.0, 10mM RbCl), by vortexing. This was centrifuged for 10-12 sec., the supernatant discarded, and the pellet resuspended by vortexing in 0.5 ml buffer, pH 6.5 (0.1M MOPS, pH 6.5, 10mM RbCl, 50mM CaCl\(_2\)). This was then left on ice for 1.5 - 2 hr. The cells (now very fragile) were pelleted by centrifugation for 10 sec., the supernatant discarded, and the cells resuspended in 150 µl of buffer, pH 6.5. 3 µl of DMSO, followed by the DNA which was being used to transform the bacteria, were then added. After leaving the cells on ice for 30 min., they were heat-shocked (by incubating at 55°C for 30 sec.) and then immediately placed on ice for a minimum of 2 min.

250 µl of bacterial cells in L-broth (the remainder of the cells which were not made competent), 30 µl of IPTG (at 30 mg.ml\(^{-1}\) in H\(_2\)O), and 30 µl of X-gal (at 30 mg.ml\(^{-1}\) in dimethylformamide) were added to each tube. The contents of the tube

\(^3\)When using JM101 in transformations, a single colony was picked from an M9 minimal plate supplemented with 100 µl of 1M glucose and 100 µl of 0.1% vitamin B1 (in H\(_2\)O), and used to inoculate an overnight culture.
microfuge tubes were then added to 4.5 ml of molten top layer agar in a sterile 10 ml tube, and this was poured onto a BBL agar plate, allowed to set at room temperature, then incubated overnight at 37 °C.

2.2.33 Polyacrylamide gel electrophoresis. (Maxam & Gilbert, 1977.)

Two glass plates (one of them notched at one end) were taped firmly together around the sides and bottom, with plastic-card strips along the sides to separate the plates.

For a gel of 20cm X 37cm, 40 ml of urea/acrylamide was made up in electrophoresis buffer as follows:

16.8 g of urea was dissolved in 4 ml of 10x TBE, pH 8.3, the required amount of 40% acrylamide solution⁴ (for example 6 ml for a 6% gel) and water to make the total volume up to 40 ml. When the urea had dissolved, 320 µl of 10% ammonium persulphate was added and the solution degassed briefly. To catalyse the polymerisation of the acrylamide, 25 µl of TEMED was added. The acrylamide solution was taken up immediately in a 25 ml pipette and carefully poured between the glass plates, avoiding air bubbles. The well former was then placed between the

⁴40% acrylamide was made by dissolving 190 g of acrylamide and 10 g of bis-acrylamide in 500 ml of H₂O. This was deionised by adding 25 g of Amberlite MB-1 (mixed bed resin) and stirring slowly for two hours. The resin was removed by filtration and the acrylamide was stored at 4 °C. If the solution was cloudy, it was degassed on a water pump.
plates at the top of the gel and the plates were clamped with bull-dog clips down the sides. The gel was allowed to set for at least 1 hr.

When set and ready to use, the well former was removed, the tape removed from the bottom of the gel, and the gel was clamped onto the gel tanks. After filling the tanks with TBE electrophoresis buffer, the DNA samples (mixed with a few microlitres of formamide-dyes mix⁵ and heated at 80 °C for 5-10 min.) were loaded into the wells using fine drawn-out capillary tubes. Immediately prior to loading the samples, urea, which was produced from the top of the gel, was forced out of the wells using a Pasteur pipette.

2.2.34 DNA sequencing (dideoxy chain termination). (Sanger et al., 1977.)

The first stage of the sequencing reactions was the annealing of the primer to the template DNA. 0.5μg (4μl) of template DNA was mixed with 0.1 picomoles (0.6 ng) of oligonucleotide primer, 2.4 μl of 10x sequencing buffer (0.1M Tris-HCl, pH7.5, 0.05M MgCl₂), and H₂O, to give 10 μl of template/primer mix. This was incubated at 65 °C for 5 min., centrifuged for a few seconds, and left for at least 15 min. at

---

⁵The formamide-dyes mix was made up by stirring 100 ml of formamide for 30 min. with 5 g of Amberlite MB-1, then removing the resin by filtration. EDTA was added to a concentration of 20mM, followed by 30 mg of xylene cyanol FF and 30 mg of bromophenol blue. This was stored at room temperature.
room temperature. Meanwhile, four uncapped microfuge tubes were
labelled, placed on ice, and into these were pipetted 1 μl each of
0.5 mM ddCTP, 0.4 mM ddGTP, 0.1 mM ddATP, and 0.5 mM ddTTP, with
1 μl of the corresponding mix of deoxynucleoside triphosphates
(C°, G°, A° or T°). The latter were made up according to the
following table (volumes in μl):

<table>
<thead>
<tr>
<th></th>
<th>C°</th>
<th>G°</th>
<th>T°</th>
<th>A°</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 mM dCTP</td>
<td>5</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>0.5 mM dGTP</td>
<td>100</td>
<td>5</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>0.5 mM dTTP</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>50 mM Tris•HCl,</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>pH8.0, 1 mM EDTA.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N.B. ddNTPs and stock solutions of 0.5 mM triphosphates were
made up in H2O and stored at -20 °C.

The nucleotide mixtures were allowed to warm to room
temperature, and 2.5 μl of template/primer mix added to each
tube. 0.5 μl (5 μCi) of [α-32P]dATP (or 1 μl (8 μCi) of
[α-35S]dATP) was mixed with 0.5 μl of cold 10 mM dATP, and made up
to 8 μl with H2O. 0.4 μl of the Klenow subfragment of DNA
polymerase I (5 units μl-1) was then quickly mixed in and
dispensed in 2 μl volumes into each of the reaction mixtures to
start the reactions. The reactions were left at room
temperature for 15 min. (or 35 min. if using [α-35S]dATP) before
the addition of 2 μl of chase mix (0.5 mM dATP in 5 mM Tris•HCl,
pH8.0, 0.1 mM EDTA). After a further 15 min., 4 μl of
formamide-dyes mix was added to stop the reactions.

The gel was pre-run for approximately 10 min., the slots
were flushed out to remove the urea, and the samples (incubated
immediately beforehand in an 80 °C water bath for 10 min. to denature the DNA) were layered immediately into the wells. The samples were run into the gel rapidly (at 1500 volts) then the power was reduced to 1200-1300 volts (at 30 milliamps limiting current) until the bromophenol blue dye was approximately 2 cm from the bottom of the gel.

For most sequencing gels, in order to read up to 400 bases from the primer, sequencing reactions were carried out using one and a half times the volumes stated above, and the reaction products were split into three: the second and third sets of samples were loaded when the bromophenol blue from the previous samples reached the bottom of the gel.

When the bromophenol blue dye was approximately 2 cm from the bottom of the gel, the buffer was aspirated off, the gel plates were removed and carefully separated, and the plate to which the gel was attached was placed in fixing solution (10% acetic acid, 10% ethanol) for at least 15 min. Excess fixer was carefully drained off and a sheet of Whatmann 3MM paper placed onto the gel. The glass plate was then inverted and the paper, with the gel still attached, was pulled away from the plate. The gel was covered with 'Saranwrap', avoiding air bubbles and large creases, and dried in a gel drier at 80 °C. The gel was then exposed to X-ray film.

N.B. Using $^{32}$P, the gel could be exposed without fixing and drying. Also, using $^{32}$P, an image was obtained on film after an overnight exposure at room temperature without an intensifying screen. With $^{35}$S, although a two day exposure at room temperature was usually required, the bands were more clearly defined. Completed sequencing reactions using $^{35}$S were sometimes stored at -20 °C overnight.
2.2.35 Polyacrylamide gel purification.

Oligonucleotides, synthesized in this department using solid phase synthesis on Whatmann 3MM paper discs (Matthes et al., 1984; Brenner & Shaw, 1985), were purified as follows:

Samples were resuspended in 100μl of H_2O, and ethanol-precipitated by adding Na-acetate to 0.3M, and 3 volumes of cold absolute ethanol, then freezing overnight at -70°C. After centrifuging at high speed for 15 min., the pellet was washed in cold 90% ethanol, then resuspended in 10μl of H_2O. 3μl of this was mixed with 2μl of formamide-dyes mix, and boiled for 2 min., before loading into a 5mm slot in a 20% acrylamide/urea gel (pre-run at 45 watts, with the voltage limited to 1800 volts).

The bromophenol blue dye was run to about 5 cm from the bottom of the gel, and the gel plates separated leaving the gel attached to one plate. Cling film was stretched over the gel avoiding creases and bubbles. The gel and film were pulled away from the plate and placed, gel upwards, back on the plate. A second sheet of cling film was then stretched over this.

An indicator thin layer chromatography plate was placed between the gel and the plate, and the oligonucleotide bands observed and marked under a U.V. lamp (short wavelength). The band was cut out, and the gel slice placed in a microfuge tube with 100μl of H_2O and left overnight at room temperature for the oligonucleotide to elute from the gel. The solution was then desalted using a 1ml spun column of fine G-25 sephadex equilibrated in H_2O. After making the volume up to 1ml with H_2O, the concentration of oligonucleotide was estimated by
measuring the absorbance at 260nm (concentration of DNA \( \mu g.ml^{-1} \) = \( A_{260} \times 37 \)). For use as primers, samples were diluted to 0.3ng.\( \mu l^{-1} \) with \( H_2O \).

2.2.36 Spun-column purification. (Maniatis et al., 1982.)

A 1ml disposable syringe was plugged with a small quantity of sterile glass wool. The syringe was filled with swollen Sephadex (G-25, fine, for oligonucleotide purification) which had been autoclaved in \( H_2O \) (or appropriate buffer) and then cooled. The syringe was inserted into a plastic centrifuge tube and centrifuged at 1600 X g for 4 min. in a bench centrifuge. More Sephadex was added, and the process was repeated until a packed column of 0.9ml volume was obtained. 0.1ml of \( H_2O \) was then added, and the column centrifuged as before. This was repeated several times until the volume of \( H_2O \) recovered from the column was also 0.1ml. The DNA was then applied to the column in a volume of 0.1ml and the column centrifuged as before, but with a small microfuge tube placed under the syringe to collect the eluant.
Chapter 3

GENE TRANSFER IN CHINESE HAMSTER OVARY CELLS

3.1 Introduction

3.1.1 Carbohydrate metabolism in Chinese hamster ovary cells.

The Chinese hamster ovary cell line, CHO-K1 (Tjio & Puck, 1958), grows on only a limited number of carbon sources, these being glucose, galactose, fructose and mannose (Faik & Morgan, 1977a). This contrasts with the metabolic diversity found in the whole animal where other carbohydrates such as lactose, starch and maltose can be utilised by various specialized tissues. Provision of suitable conditions and the application of selection pressures have led to the isolation of variants of CHO-K1 cells which are now able to utilise starch, maltose (Scannell & Morgan, 1982), lactose, sucrose and ribose (Faik & Morgan, 1977b), presumably through the expression of appropriate genes which are not expressed in CHO-K1.

The mutant cell line, R1.1.7, is a variant of CHO-K1 which has some unusual characteristics of carbohydrate utilisation. These features make R1.1.7 a particularly useful cell line, with the potential for use as a system in which two particular genes involved in carbohydrate metabolism can be isolated, and their expression studied. These are the genes coding for phosphoglycerate kinase (PGK, E.C.2.7.2.3) and glucose phosphate isomerase (GPI, E.C.5.3.1.9). Table 3.1 illustrates some of the properties of R1.1.7 cells and compares them to those of the "wild-type" CHO-K1.
Table 3.1

Comparison of properties of CHO-K1 cells and R1.1.7 cells.
<table>
<thead>
<tr>
<th></th>
<th>CHO-K1</th>
<th>R1.1.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth on: ribose</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>glucose</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>fructose</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>galactose</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>mannose</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Mannose toxic to cells</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Respiration dependent</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Lactate production</td>
<td>HIGH</td>
<td>LOW</td>
</tr>
<tr>
<td>Inositol requirement</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Phosphoglycerate kinase</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Glucose phosphate isomerase</td>
<td>YES</td>
<td>NO</td>
</tr>
</tbody>
</table>
R1.1.7 was first isolated as a variant which was able to grow on ribose (Faik & Morgan, 1977b). Like CHO-K1 cells, R1.1.7 cells grow on glucose, galactose and fructose, but although CHO-K1 cells grow on mannose, R1.1.7 cells cannot. The latter are, in fact, killed by mannose, even in the presence of other utilisable carbon sources (Faik & Morgan, 1977b). A second notable characteristic of R1.1.7 cells is that, unlike CHO-K1, during growth on glucose they secrete very little lactic acid into the medium (Faik & Morgan, 1977b), suggesting some deficiency in the glycolytic process.

The differences in mannose and glucose metabolism were further investigated by determining the activities of a number of enzymes involved in carbohydrate metabolism (Morgan & Faik, 1980). In R1.1.7 cell extracts, GPI activity was less than 5% of the level found in CHO-K1 cell extracts, and PGK activity was 0.2% of the "wild-type" activity. In measurements of the overall rate of glycolysis in cell-free homogenates, the rate of production of lactate in R1.1.7 was approximately 30% of the rate in CHO-K1, and full stimulation to the "wild-type" activity was only possible on the addition of small amounts of both purified PGK and GPI. Therefore, the decreased ability of R1.1.7 cells and cell-free extracts of R1.1.7 to convert glucose to lactic acid can be attributed to the reduced PGK and GPI activities in this cell line. A consequence of this deficiency is that R1.1.7 cells are able to derive very little energy from glycolysis.

It has been shown that R1.1.7 must obtain its energy entirely from respiration (Morgan & Faik, 1981): R1.1.7 cells are rapidly killed by exposure to respiratory inhibitors, such as oligomycin, at concentrations that do not affect CHO-K1.
cells over the same period of time, and it has been shown that glutamine and pyruvate are utilised as respiratory substrates.

Mannose toxicity is a phenomenon that has been observed in the honey-bee, and it has been shown to be due to the absence of mannose phosphate isomerase (MPI), resulting in the lethal accumulation of mannose-6-phosphate (Sols et al., 1960). To determine whether a defect in a mannose utilising enzyme was responsible for the mannose toxicity in R1.1.7, mannokinase (MK) and MPI were assayed (Morgan & Faik, 1980), but it was found that the activities of these enzymes were similar to those present in CHO-K1. Thus, the mannose toxicity in R1.1.7 cells was not due to a build up of mannose-6-phosphate consequent to a deficiency of MK or MPI.

Investigations into the levels of glycolytic intermediates in R1.1.7 cells revealed that, when growing on mannose, there is an accumulation of fructose-6-phosphate (F-6-P) and fructose-1,6-bisphosphate (FBP) (M.J. Morgan, unpublished results) which is probably responsible for the inability of R1.1.7 cells to grow on mannose, and its toxicity. In R1.1.7 cells, the lack of PGK results in an accumulation of FBP, since it cannot be metabolised via glycolysis to pyruvate, and the F-6-P formed from mannose cannot be isomerised to glucose-6-phosphate (G-6-P) because of the deficiency in GPI (Figure 3.1). Further evidence that mannose toxicity is due to the deficiency of both enzymes is that mutant cell lines lacking solely GPI, for example the hamster cell line DS7 (Pouysségur et al., 1976), or just PGK, for example the human cell line GM0743 (derived from the variant PGK Matsue ; Yoshida & Miwa, 1974), do not demonstrate the sensitivity to mannose that can be seen in R1.1.7 cells, presumably because in each case a channel exists for the further
Figure 3.1

Metabolic scheme illustrating some aspects of carbohydrate metabolism in R1.1.7 cells.
metabolism of FBP.

Another consequence of the enzyme deficiencies in R1.1.7 was the discovery that in order to grow on ribose, which is metabolised via the pentose phosphate pathway, R1.1.7 requires inositol to be provided in the culture medium.

Inositol is required for the biosynthesis of phosphatidyl inositol, an important constituent of cell membranes. Its immediate precursor, inositol-1-phosphate, is formed from glucose-6-phosphate (G-6-P) (Figure 3.1). In cell-free extracts of the GPI-deficient R1.1.7 cells, F-6-P, but not G-6-P, accumulates with ribose-5-phosphate as a substrate (Morgan, 1981). Thus, G-6-P cannot be synthesised by the pentose phosphate pathway, and consequently, this explains R1.1.7's requirement for inositol.

The major feature which lends the R1.1.7 cell line to the study of introduced PGK or GPI genes is the very low background of PGK and GPI activities in the cells. The toxicity of mannose to R1.1.7 cells provides the basis of a selection system for cells which are transformed to a GPI or PGK positive (GPI⁺ or PGK⁺) phenotype, and as a result would not be subject to a lethal accumulation of sugar phosphates.

A further refinement of this selection can be made by making use of the inositol requirement of R1.1.7 resulting from the deficiency of GPI. Cells acquiring GPI activity after gene transfer would no longer exhibit this requirement, so selection for transformants of R1.1.7 to a GPI⁺ phenotype, after gene transfer, could be made by subjecting the cells to culture medium containing mannose as the sole carbon source, and lacking inositol.

In a similar way, the dependence of R1.1.7 on respiratory
substrates can be used as a means to refine the selection system to allow for the survival of PGK\(^+\) transformants. This could be carried out by including mannose in the selection system, but reducing the concentrations of pyruvate and glutamine, normally required for respiration in R1.1.7, to levels where only transformant cells able to glycolyse mannose for energy production could survive.

3.1.2 Gene transfer in cultured animal cells.

The introduction of exogenous genes into mammalian cells, resulting in genetic transformation, has been successfully carried out using DNA-mediated gene transfer (DMGT). In most instances the recipient cells used have been mouse L-cells or 3T3 cells because they show particularly high frequencies of transformation (Wigler et al., 1977; Pellicer et al., 1978 & 1980; Weinberg, 1981). Although it is usually found that CHO cells are less easily transformed, transformation frequencies being 50-100 fold lower than with mouse L-cells (Graf et al., 1979; Lewis et al., 1980; Abraham et al., 1982), in certain sublines, frequencies only 5-10 times lower (\(>3 \times 10^{-4}\)) have been observed (Nairn et al., 1982). However, CHO cells (Kao & Puck, 1974) have a number of advantages as recipient cells in gene transfer experiments, including a stable and close to diploid karyotype, which is useful in the analysis of somatic cell hybrids, and the existence of a large number of mutant phenotypes (Siminovitch, 1979; Adair et al., 1979; Siciliano et al., 1983) including the PGK\(^-\), GPI\(^-\) cell line, R1.1.7.

The principles for the identification of transformant clones of cells which have acquired and are expressing the GPI\(^+\) or PGK\(^+\) gene after DMGT has been described above. The identification
and isolation of an introduced gene from the transformants can be achieved following co-transformation of the R1.1.7 cells with genomic DNA isolated from a PGK<sup>+</sup>/GPI<sup>+</sup> source, along with DNA for which a molecular probe is available, and which would confer a selectable phenotype on the recipient cells. The plasmid pTCF (Grosveld et al. 1982) fulfills these requirements, as it contains an aminoglycosyl-3′-phosphotransferase (AGPT) gene, whose gene product confers resistance to a number of antibiotics. In the plasmid pTCF, the AGPT gene has been inserted downstream of the Herpes Simplex virus thymidine kinase (HSV-TK) eukaryotic promoter region to allow expression in eukaryotic cells. AGPT inactivates the antibiotic G418 which, unlike most other antibiotics, inhibits protein synthesis in eukaryotes as well as in prokaryotes. Cells transformed using pTCF DNA and expressing the AGPT gene should be resistant to G418 and would therefore survive and form colonies in selection medium containing the antibiotic.

It has been shown that selected genes become linked to other exogenously added DNA during, or at some period subsequent to, gene transfer, and remain stable in the genome of the recipient cells (Ruddle, 1979 & 1980; Scangos et al., 1981). Therefore, colonies which appear after co-transformation and selection in mannose medium containing G418 may contain pTCF sequences covalently linked to DNA containing the donor PGK or GPI gene. In this situation, a pTCF probe could be used to isolate these genes for which a direct DNA probe is not available.

The DNA isolated from co-transformant clones would be used to generate a DNA library in a lambda-based vector. The recombinant phage containing pTCF sequences would be identified by plaque hybridisation, and DNA isolated from these phages
would be used to transform R1.1.7 cells to mannose resistance (GPI\(^+\) or PGK\(^+\)), selecting separately for PGK\(^+\) and GPI\(^+\) as described. Thus, recombinant phage containing GPI or PGK genes would be identified. However, this second round of transformation may be limited in usefulness as it depends on an individual recombinant phage containing the full coding region and all the elements required for the expression of the gene. The alternative to this would be to analyse each recombinant isolated and identify those containing PGK or GPI sequences by, for example, searching for overlapping regions of DNA.

Once isolated, the same transformation system that was used to isolate these genes could be used in the study of the expression of the genes.

3.2 Results and Discussion

The aim was to set up a transformation system which could be used initially for the isolation of the genes of interest and, secondly, to study their expression in the CHO cell system.

A process of DNA-mediated gene transfer using the technique of calcium phosphate-DNA precipitation (Graham & Van der Eb, 1973; Graham et al., 1980) was used to transform cells of the Chinese hamster ovary cell line, R1.1.7. The plasmid pTCF, containing the AGPT gene, was used to confer resistance to G418 on the recipient cells.

3.2.1 Sensitivity of R1.1.7 cells to G418.

Before carrying out a transformation experiment, it was necessary to determine the precise conditions which would allow the survival of cells expressing the AGPT gene of pTCF, while
killing all untransformed cells. The concentration of G418 to be used in this selection was determined by carrying out single cell platings over a range of G418 concentrations from 0 to 500 μg per ml of medium. Figure 3.2 shows, firstly, that R1.1.7 cells appear to be more sensitive to the antibiotic G418 than CHO-K1 cells and, secondly, that the lowest concentration of G418 which resulted in total cell killing was 250 μg.ml\(^{-1}\). As a result it was decided to use a G418 concentration of 300 μg.ml\(^{-1}\) to select for G418 resistance in R1.1.7 cells exposed to pTCF DNA in DMGT experiments.

3.2.2 DNA mediated gene transfer.

Twenty-one 100 mm culture dishes were seeded with R1.1.7 cells (at 6.5 X 10\(^5\) cells per dish) in modified F12 medium (i.e. F12 lacking glucose, pyruvate and glutamine) containing 5% foetal calf serum and 10mM glucose. Seven dishes (S1-S7) were treated with a co-precipitate of calcium phosphate containing 1.6μg of pTCF DNA along with 16 μg of salmon sperm carrier DNA (in 1 ml). Seven other dishes (C1-C7) were treated with a calcium phosphate co-precipitate containing 1.6 μg of pTCF DNA and 16 μg of chick embryo DNA. Seven control dishes were exposed to a calcium phosphate precipitate without DNA. In each group of seven dishes, five were exposed to 10% DMSO for 25 min. as described under "Materials and Methods", and the remaining two were not exposed to DMSO.

The following day, the cells were trypsinised and counted, and each 100mm culture dish was split between two 150mm dishes. 42-45 hr. after exposure to the DNA, the medium was replaced with selective medium (G418 at 300 μg.ml\(^{-1}\)). Twelve days after exposing the cells to selective medium, large colonies were
Figure 3.2

Sensitivity of R1.1.7 cells to the antibiotic G418.

R1.1.7 (o-o) cells were exposed to a range of concentrations of G418. After one week, the number of surviving colonies at each concentration was counted and expressed as a percentage of the number of colonies appearing in medium lacking G418 (% survival). For each concentration of G418, percentages were calculated from duplicate dishes at plating densities of 500 cells per 35 mm dish. The results of a previous experiment with CHO-K1 cells (•-•) are included as a comparison.
AMOUNT OF G418 ADDED TO EACH 35mm DISH (µg)
Table 3.2

Transformation of R1.1.7 cells to G418 resistance.

Transformation frequencies are calculated as the number of G418 resistant colonies appearing as a proportion of the number of calcium phosphate precipitate-treated cells plated out.
<table>
<thead>
<tr>
<th>Frequency (transformation)</th>
<th>Colonies resistant (No. of cells)</th>
<th>Plating efficiency (No. of cells used)</th>
<th>Plates exposed to caps 4 ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7 X 10^6</td>
<td>0</td>
<td>1.4 X 10^6</td>
<td>ctch + PTCF + salmon + PTCF</td>
</tr>
<tr>
<td>3.3 X 10^6</td>
<td>0</td>
<td>2.16 X 10^6</td>
<td>ctch + PTCF + salmon + PTCF</td>
</tr>
<tr>
<td>9.3 X 10^5</td>
<td>10</td>
<td>1.08 X 10^7</td>
<td>ctch + PTCF + salmon + PTCF</td>
</tr>
<tr>
<td>0.9 X 10^5</td>
<td>40</td>
<td>8.0 X 10^6</td>
<td>ctch + PTCF + salmon + PTCF</td>
</tr>
<tr>
<td>1.4 X 10^5</td>
<td>11</td>
<td>1.79 X 10^7</td>
<td>ctch + PTCF + salmon + PTCF</td>
</tr>
<tr>
<td>2.5 X 10^5</td>
<td>44</td>
<td>1.79 X 10^7</td>
<td>ctch + PTCF + salmon + PTCF</td>
</tr>
</tbody>
</table>

Controls

6.7 (−DMSO)

1-5 (−DMSO)

C6, C7 (−DMSO)

C4, C5 (−DMSO)

SG, S7 (−DMSO)

S1-55 (−DMSO)
observed in a number of culture dishes. A number of these colonies were cloned and cultured in selective medium. All the dishes were then fixed and stained, the colonies counted, and transformation frequencies calculated (Table 3.2).

Transformation frequencies in all cases were found to be very low - between 0.9 and $2.5 \times 10^{-6}$ (equivalent to 0.7-2.0 colonies per μg of pTCF DNA). These frequencies are at least 20 times lower than those found by other investigators with other CHO cell lines (Nairn et al., 1982). This may simply be a reflection of clonal variation among CHO sublines with respect to transformation efficiency with DNA-mediated gene transfer. Alternatively, some factor in the formation of the calcium phosphate-DNA co-precipitate, or the technique of transformation, may have contributed to the low frequencies of transformation.

Both salmon sperm and chick DNA were used as carrier to investigate whether or not the efficiency of DNA-mediated gene transfer was affected by the type of carrier DNA used. Transformation frequencies were a little higher with salmon sperm DNA as carrier, but due to the small number of colonies appearing, such a small difference between these two sets of experiments cannot be regarded as significant.

Transformation frequencies for culture dishes treated with DMSO were about double those for dishes not treated with DMSO suggesting that DMSO had a positive effect on some part of the process of DNA-mediated gene transfer. It has, however, been reported by other investigators that DMSO does not significantly enhance the efficiency of transformation by DMGT in CHO cell lines (Srinivasan & Lewis, 1980).
3.2.3 Stability of G418 resistance.

Twenty-three colonies of G418 resistant cells were cloned and cultured in selective medium. The first stage in the characterisation of these clones was to determine how stable the G418 resistance was under non-selective conditions. Cells of each G418 resistant clone were cultured in non-selective medium for three weeks. After this period, cells were plated into both selective and non-selective medium at a cell density of 500 cells per 35 mm culture dish and, after seven days, the cells were fixed and stained, and the colonies were counted. Comparison of the number of cells growing up to form colonies in selective medium with the number appearing in non-selective medium indicates how stable the G418 resistance is in individual transformant clones (Table 3.3).

The results show that over half of the resistant clones tested remained resistant over a period of three weeks in non-selective medium (stability >90%). The remaining clones exhibited a range of stabilities, the lowest being around 6% (C34). It was also observed that a number of clones grew more slowly in selective medium than in non-selective medium, but there was no direct correlation between this and the stability of G418 resistance.

3.2.4 Level of G418 resistance.

Eight G418 resistant clones were chosen for further characterisation and exposed to G418 at 0, 0.5, 1, 2 and 4 mg.ml⁻¹. Table 3.4 shows that three of the eight clones tested grew at 2 mg.ml⁻¹ G418, although more slowly than at lower concentrations. There was no correlation between the stability of G418 resistance in individual clones and the level
Table 3.3

Stability of G418 resistance in transformant clones.

Cells from each clone were cultured in non-selective medium for three weeks, then plated into selective and non-selective medium. % stability is the number of colonies arising in selective medium as a proportion of the number appearing in non-selective medium.
<table>
<thead>
<tr>
<th>Clone</th>
<th>No. colonies from cells plated into:</th>
<th>% stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Selective medium (+500µg.ml⁻¹G418)</td>
<td>Non-selective medium (-G418)</td>
</tr>
<tr>
<td>S2</td>
<td>329</td>
<td>364</td>
</tr>
<tr>
<td>S3</td>
<td>292</td>
<td>332</td>
</tr>
<tr>
<td>S7</td>
<td>282</td>
<td>313</td>
</tr>
<tr>
<td>S8.1</td>
<td>92</td>
<td>393</td>
</tr>
<tr>
<td>S8.2</td>
<td>381</td>
<td>392</td>
</tr>
<tr>
<td>S25.1</td>
<td>300</td>
<td>378</td>
</tr>
<tr>
<td>S35</td>
<td>294</td>
<td>315</td>
</tr>
<tr>
<td>C6</td>
<td>125</td>
<td>213</td>
</tr>
<tr>
<td>C10</td>
<td>385</td>
<td>370</td>
</tr>
<tr>
<td>C13</td>
<td>23</td>
<td>294</td>
</tr>
<tr>
<td>C14</td>
<td>112</td>
<td>316</td>
</tr>
<tr>
<td>C15</td>
<td>200</td>
<td>208</td>
</tr>
<tr>
<td>C18</td>
<td>437</td>
<td>400</td>
</tr>
<tr>
<td>C20</td>
<td>402</td>
<td>396</td>
</tr>
<tr>
<td>C30</td>
<td>133</td>
<td>155</td>
</tr>
<tr>
<td>C34</td>
<td>14</td>
<td>227</td>
</tr>
</tbody>
</table>
Cells from eight G418 resistant clones were plated out into non-selective medium, at 500 cells per well, in 25-well culture dishes. The following day, the cells were exposed to fresh medium containing G418 at 0, 0.5, 1, 2 and 4 mg.ml$^{-1}$. One week later, cells were fixed and stained, and the number of colonies surviving at different concentrations of G418 were compared. 

++++ denotes maximum growth of colonies.
- denotes no growth.
<table>
<thead>
<tr>
<th></th>
<th>G418 concentration (mg/ml⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td><strong>R1.1.7 (control)</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>S7</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>S22</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>S25.1</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>S35</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>C6</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>C10</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>C18</strong></td>
<td>+++</td>
</tr>
<tr>
<td><strong>C20</strong></td>
<td>+++</td>
</tr>
</tbody>
</table>
of G418 resistance: clone C6 grew at 1 mg.ml⁻¹ G418 but had less than 60% stability after three weeks in non-selective medium; clone S25.1, which grew at 2 mg.ml⁻¹ G418, has less than 80% stability, yet clone S35, although at least 90% stable, did not survive even at 500 µg.ml⁻¹ G418.

3.2.5 Southern transfer and hybridisation analysis.

DNA was isolated from R1.1.7 cells and from each of the eight G418 resistant cell lines chosen for further characterisation. 15 µg of each DNA preparation and 0.2-0.4 ng of pTCF DNA were digested with the restriction enzyme EcoRI and electrophoresed through a 1% agarose gel. The DNA was transferred to a sheet of nitrocellulose filter and probed with ³²P-labelled pTCF DNA. The filter was then left in contact with X-ray film for a month, with an intensifying screen, at -70°C.

The autoradiogram in figure 3.3 shows that DNA isolated from each of the eight G418 resistant clones contains EcoRI restriction fragments that hybridise strongly to the probe. There are also two faint bands in the R1.1.7 DNA track, but these appear in all the other hamster DNA tracks and probably correspond either to slight homologies in the Chinese hamster genome or to fragments derived from repeated sequences in the genome. The presence of the AGPT sequence in all the transformant clones provides evidence that the G418 resistance is due to the expression of this gene.

In six of the eight transformants, the hybridising fragments are of higher molecular weight than the EcoRI restriction fragment of pTCF corresponding to the AGPT gene, indicating that there has been some recombination within this fragment and that the AGPT gene has therefore probably been integrated into the
**Figure 3.3**

**Southern transfer analysis of G418 resistant clones.**

pTCF DNA (positive control, '+'), R1.1.7 DNA (negative control), and DNA from eight G418 resistant transformant clones, was digested with EcoR1, electrophoresed, and transferred to a sheet of nitrocellulose filter. The DNA was probed with the $^{32}$P-labelled 2 kb EcoR1 fragment from pTCF containing the aminoglycosyl-3' phosphotransferase gene (Grosveld *et al.*, 1982).
genome of the resistant clones. In the two tracks containing hybridising fragments of the same size as the AGPT EcoRI fragment (S35 and C18), the pTCF DNA may also have been integrated but through recombination outside that restriction fragment of the plasmid.

If the level of G418 resistance in the transformant clones was dependent on the number of copies of the AGPT gene in the cells, DNA from those clones which exhibit particularly high levels of resistance (S25.1, C10, C20) should show more intense hybridisation to the probe, or hybridisation of a number of fragments of DNA. Similarly, DNA from S35 which has a very low level of resistance to G418 would be expected to show fainter hybridisation to the probe. This experiment did not demonstrate either of these phenomena: there was only one hybridising band in each case and all the bands were of similar intensity. Thus in the case of these transformants, the level of G418 resistance did not appear to be associated with copy number of the introduced AGPT gene.

An alternative explanation for the differences in the levels of G418 resistance may be that the position of integration of the gene in the genome of the recipient cell may have some influence on the degree of expression of the gene.

Further analysis and investigation of the transformants was not continued any further. Instead, efforts were concentrated on improving and developing this mammalian transfection system in order to increase transformation frequencies to a level that would make the isolation of PGK* and GPI* transformants possible from recipient cells exposed to genomic donor DNA.
3.2.6 Modification of transfection system.

In experiments to improve the transformation efficiencies in gene transfer experiments, DNA-mediated gene transfer by the DNA-calcium phosphate co-precipitation method was carried out as before, using G418 resistance to select for the expression of AGPT in recipient cells, but with a number of modifications.

In all experiments, the DNA added to the dishes consisted of 2 μg of pTCF and 18 μg of carrier DNA isolated from human embryo cells (HEC), except for two control dishes which were exposed to 20 μg of HEC DNA each, and two further controls in which DNA was not added to the calcium phosphate precipitate.

The standard conditions of transfection involved exposure of the cells to 10% DMSO for 25 min. after a 4 hr. exposure to the calcium phosphate precipitate (adsorption time). This was followed by a 45 hr. expression time (the time maintained in non-selective medium before being exposed to the selective medium). The variations to this protocol are shown in table 3.5.

The results show that the highest transformation frequencies (1.9 x 10^{-5}) were obtained by increasing the adsorption time to 16 hr., and 'shocking' the cells with DMSO at the end of this period. The cells which were not treated with DMSO after their 16 hr. adsorption time also gave a higher transformation frequency than the corresponding dish with a 4 hr. adsorption time. Exposure of the cells to 0.5% glycerol during the time which they were exposed to the DNA increased further the transformation efficiency (8.5 x 10^{-6}), but not to the level resulting from DMSO treatment. Figure 3.4 shows that increasing the adsorption time from 4 to 16 hr. increases the frequency of
Table 3.5

Transformation frequencies of R1.1.7 to G418 resistance using various conditions of DMGT.

All dishes except the controls were exposed to 2 μg of pTCF and 18 μg of carrier HEC DNA.

The variations to the protocol were as follows:

1) A 'standard' dish was treated as previously described.
2) As 1) but without DMSO treatment.
3) As 1) but with an increased adsorption time of 16 hours.
4) As 3) but without DMSO treatment.
5) As 3), without DMSO treatment, but exposed to 0.5% glycerol for the 16 hours adsorption time.
6) As 1) but with 27 hours adsorption time.
7) As 1) but with 65 hours adsorption time.
8) As 1) but with cells split directly into selective medium to determine whether transformants would plate out in the presence of G418.
9) As 1) but the DNA-calcium phosphate precipitate was added immediately on formation i.e. it was not left for 35 min. before adding to the dish.

Negative controls:

10) As 1) but using 20 μg of HEC DNA only.
11) As 1) but using no DNA in transfection.

Large colonies, observed approximately two weeks after transfection, were fixed, stained and counted.
### Conditions of DMGT

<table>
<thead>
<tr>
<th>Adsorption time (hr.)</th>
<th>± DMSO etc.</th>
<th>Expression time (hr.)</th>
<th>Transformation frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) 4</td>
<td>+ DMSO</td>
<td>45</td>
<td>3.6 x 10^{-6}</td>
</tr>
<tr>
<td>2) 4</td>
<td>- DMSO</td>
<td>45</td>
<td>2.0 x 10^{-6}</td>
</tr>
<tr>
<td>3) 16</td>
<td>+ DMSO</td>
<td>45</td>
<td>1.9 x 10^{-5}</td>
</tr>
<tr>
<td>4) 16</td>
<td>- DMSO</td>
<td>45</td>
<td>5.1 x 10^{-6}</td>
</tr>
<tr>
<td>5) 16</td>
<td>+ glycerol(a)</td>
<td>45</td>
<td>8.5 x 10^{-6}</td>
</tr>
<tr>
<td>6) 4</td>
<td>+ DMSO</td>
<td>27</td>
<td>1.1 x 10^{-6}</td>
</tr>
<tr>
<td>7) 4</td>
<td>+ DMSO</td>
<td>65</td>
<td>1.0 x 10^{-5}</td>
</tr>
<tr>
<td>8) 4</td>
<td>+ DMSO</td>
<td>45</td>
<td>1.6 x 10^{-6} (b)</td>
</tr>
<tr>
<td>9) 4</td>
<td>+ DMSO</td>
<td>45</td>
<td>1.3 x 10^{-6} (c)</td>
</tr>
<tr>
<td>10) 4 (HEC DNA)</td>
<td>+ DMSO</td>
<td>45</td>
<td>7.9 x 10^{-7}</td>
</tr>
<tr>
<td>11) 4 (no DNA)</td>
<td>+ DMSO</td>
<td>45</td>
<td>2.3 x 10^{-7}</td>
</tr>
</tbody>
</table>

(a) Glycerol at 0.5% for 16 hours.

(b) Cells plated directly into selective medium.

(c) Precipitate added immediately after forming.
Figure 3.4
Effect of adsorption time and 'cell shocking' agents on the transformation efficiency of R1.1.7.

Histogram of selected results from table 3.5 showing the effect of increasing the time of adsorption of the calcium phosphate-DNA precipitate, and the effects of DMSO or glycerol, on the frequency of transformation of R1.1.7 cells to G418 resistance.
16 HRS ADSORPTION TIME
4 HRS ADSORPTION TIME

TRANSFORMATION FREQUENCY (x 10^-5)

- DMSO  + 0.5% GLYCEROL  + DMSO

0  0.2  0.5  1  1.5  2
transformation by between 2.5 times (-DMSO) and 5 times (+DMSO): Treatment with cell 'shocking' agents also results in higher frequencies of transformation, 10% DMSO being more effective than 0.5% glycerol.

Decreasing the expression time to 27 hr. resulted in a reduced transformation frequency and, similarly, increasing this period to 65 hr. produced an increase in the frequency of transformation. (N.B. the expression time is limited to about 70 hr. as, at this stage, the cells approach confluence in the dishes and enter stationary phase).

Plating the cells directly into selective medium decreased the transformation frequency by over 50%, indicating that G418 probably has some effect on the ability of the recipient cells to plate out successfully. The period during which the calcium phosphate precipitate is left at room temperature also seems to be an important part of the process of precipitate formation, as where it was added to the cells immediately on formation, transformation frequencies were reduced by over 60%.

Increasing the adsorption time to 16 hr. and the expression time to 65 hr. each resulted in frequencies of transformation of over $1 \times 10^{-5}$.

To determine any increase in transformation frequency as a result of increasing both the adsorption and expression times, a further set of transfections was carried out. The adsorption and expression times were increased independently and in combination using, as a 'standard', an adsorption time of 4 hr. and an expression time of 45 hr. In each case, the cells were treated with DMSO. The results of these experiments were ambiguous as in most of the culture dishes satellite colonies were formed as a result of cells becoming detached from
previously existing colonies and plating out again. It was cautiously estimated that the number of transformants arising from the standard conditions used and from increasing adsorption time and expression time alone, corresponded roughly to the results observed in previous experiments. In plates where these conditions were combined, it was estimated that the frequency of 'original' colonies exceeded that observed under any other set of conditions. The total number of colonies, including satellite colonies, was very much greater than in the other experiments (at least 100 per dish). Had this result represented 'original' colonies only, the transformation frequency would have been greater than $5 \times 10^{-5}$. However, the only conclusion that can be drawn from this experiment is that by combining a 16 hr. adsorption time with a 65 hr. expression time, the transformation frequency is increased to a level greater than the frequency observed using each of these two conditions separately.

3.2.7 Co-transformation and selection for GPI positive transformants.

Using the conditions of DNA-mediated gene transfer which had resulted in the highest transformation frequencies (treatment with DMSO for 25 min. after 16 hr. exposure to the DNA-calcium phosphate precipitate followed by 65 hr. expression time), 104 x 100mm culture dishes seeded at $5 \times 10^5$ cells per dish were each exposed to a calcium phosphate precipitate containing 2 µg of pTCF and 18 µg of high molecular weight whole cellular DNA from a human source. The human DNA would act as a carrier and, more importantly, as a potential source of human genomic GPI$^+$ sequences. In half of the dishes the human DNA used was
isolated from HEC cells in culture, and that used in the remaining dishes was isolated from human placental tissue. Two days after transfection, the cells \((1 \times 10^8)\) were trypsinised and transferred to roller bottles \((110\text{mm diameter} \times 285\text{mm})\) : cells from \(3 \times 100\text{mm} \) dishes were transferred into each roller bottle. The selection medium used was F12D (which lacks inositol ; Kao & Puck, 1974) containing 8% macroserum, 5mM mannose, 2mM pyruvate, and 2mM glutamine. This medium would select for cells expressing glucose phosphate isomerase activity (see section 3.1.1). Medium was changed every three days, and the roller bottles were checked for colonies. Unfortunately, most of the bottles became contaminated with bacteria, and at no time were any colonies of cells observed.

The fact that only seven roller bottles (representing less than 30% of the cells used in the experiment) remained uncontaminated reduced the probability of observing transformants. Other investigators have found that CHO thymidine kinase negative \((tk^-)\) cells can be transformed at a frequency of \(10^{-7}\) using whole cellular DNA isolated from CHO thymidine kinase positive \((tk^+)\) cells (Abraham et al., 1982), so that had the transformation efficiency of R1.1.7 been much lower, the isolation of transformants would have been statistically unlikely.

It was also possible that the selection conditions may have been too stringent : if any of the cells were transformed to \(GPI^+\), the GPI activity might not have been at a sufficiently high level to enable the cells to grow in the selection medium.

Our intention at this point was to repeat the experiment on a larger scale. However, this was not followed through, as a cDNA probe for the human phosphoglycerate kinase (PGK) gene,
contained on the plasmid PGK-5 (Figure 4.1) became available, which meant that it would be possible to probe for PGK gene sequences directly in a DNA library. In addition, it thus became possible to investigate R1.1.7 as a system for the study of the expression of introduced PGK genes.

3.2.8 Transformation of R1.1.7 cells using the human PGK cDNA.

Using the plasmid PGK-5 in DMGT experiments, work was carried out to demonstrate the usefulness of R1.1.7 as a suitable 'test' system in which to study PGK gene expression. Transfection of R1.1.7 was carried out using a calcium phosphate precipitate of both pTCF (containing the AGPT gene) and PGK-5 (containing the human PGK cDNA downstream of an SV40 early promoter). After selection in G418, a number of clones were isolated and cell extracts of each were assayed for PGK activity. Approximately half of the G418 resistant clones proved to be co-transformants which expressed the human PGK gene (demonstrated by starch gel electrophoresis, since the human enzyme has a different migration rate to that of the Chinese hamster). All of these clones grew in the selection medium devised for PGK+ transformants of R1.1.7 (Faik et al., 1985).

3.3 Summary

R1.1.7 is a mutant cell line derived from the Chinese hamster cell line CHO-K1. Analysis of its glycolytic functions revealed that R1.1.7 was deficient in two enzymes of glycolysis: phosphoglycerate kinase (PGK) and glucose phosphate isomerase (GPI). One consequence of this is that mannose is toxic to R1.1.7 cells.
A selection system has been devised, based on this mannose toxicity such that, in the selective medium, CHO-K1 cells (PGK⁺, GPI⁺), grow but R1.1.7 cells (PGK⁻, GPI⁻) are rapidly killed. The selection system can be adapted such that if the gene for either of these enzymes is reintroduced into R1.1.7 cells, and expressed, the cells will grow and form colonies.

DNA-mediated gene transfer, by exposure of the recipient cells to a co-precipitate of DNA and calcium phosphate, was used to determine transfection conditions which could be used to introduce exogenous DNA into R1.1.7 cells. As a test system, the AGPT gene, carried on the plasmid pTCF, was introduced into R1.1.7 cells, and selection of transformants was carried out by exposing DNA-treated cells to the antibiotic G418. Using this system, the transformation of R1.1.7 cells to a G418⁺ phenotype resulted in the production of stable transformants, some of which could survive at concentrations of 2mg.ml⁻¹ G418. Furthermore, the presence of pTCF sequences in the DNA of these G418⁺ transformants was demonstrated by Southern transfer and hybridisation analysis.

A number of conditions of DMGT were altered to increase the transformation frequencies. Treatment of the cells with DMSO for 25 min., after exposure to the co-precipitate, resulted in transformation frequencies higher than those obtained by omitting this step or by exposure to 0.5% glycerol for the duration of exposure to the DNA. In addition, the transformation frequencies were increased when the adsorption time was increased from 4 hr. to 16 hr. and when the expression time was increased from 45 hr. to 65 hr. Using the optimal conditions we had defined, an attempt was made to introduce pTCF and human DNA into R1.1.7 cells in the same...
precipitate in the hope of co-transforming the cells to G418 resistance and a GPI+ phenotype. Unfortunately, using the selection system devised to select for GPI+ transformants, we were unable to isolate any cells expressing GPI. However, we had developed an expression system which could be useful for the study of the expression of introduced functional GPI genes or PGK genes.

The acquisition of a plasmid containing the human PGK cDNA, and the successful transformation of R1.1.7 cells to a PGK+ phenotype, enables this transformation system to be used to investigate the expression of the human PGK cDNA or any other introduced PGK gene.
Chapter 4

ANALYSIS OF THE CHINESE HAMSTER PGK GENE FAMILY

4.1 Introduction

Two isozymes of phosphoglycerate kinase (PGK) have been shown to exist in all mammalian species examined: PGK-1 has been identified genetically as having a locus on the X-chromosome (Chen et al., 1971) and PGK-2 is encoded by an autosomal locus.

Human X-linked PGK cDNAs have been isolated and cloned into plasmid vectors (Michelson et al., 1983; Singer-Sam et al., 1983), and these have been used by a number of investigators to probe genomic digests of mammalian DNA. These studies have revealed complex patterns of hybridisation to the PGK cDNA, suggesting that there may be a greater array of PGK sequences than just the two loci previously identified. Experiments performed using DNA isolated from human cell lines containing four X-chromosomes have shown hybridisation dosage of DNA fragments containing X-linked human PGK sequences, and other fragments that did not exhibit increased hybridisation in the 4X DNA and thus were derived from autosomal DNA (Singer-Sam et al., 1983 & 1984; Hutz et al., 1984; Michelson et al., 1983 & 1985a).

Using subfragments of the PGK cDNA, investigators were able to estimate the existence of four independent phosphoglycerate kinase genes, two located on the X-chromosome and two with autosomal loci. Using a panel of human-mouse somatic cell hybrids they showed that, in the human, both the X-linked genes
mapped to the same chromosomal band (Xq13), and that the
autosomal PGK genes mapped to chromosome 6 (p12-21.1) (Michelson
et al., 1985a).

The X-linked human PGK sequences have been cloned and
sequenced: the functional gene has been shown to span 23 kb of
DNA and is interrupted by 10 introns (Michelson et al., 1985b),
and the second gene has been identified as an intronless
pseudogene which is found proximally to the active X-linked gene
(Michelson et al., 1985a).

In the present study, the Chinese hamster PGK gene family
has been analysed by probing a number of digests of genomic DNA
with a human PGK cDNA, and experiments have been carried out to
identify which restriction fragments were X-linked by looking
for an increased hybridisation signal in DNA from female cells
compared to the signal from DNA from male cells.

In order to obtain more detailed information about the PGK
sequences of the Chinese hamster, a number of recombinant phages
were isolated from a library of genomic DNA obtained from the
"wild-type" hamster cell line CHO-K1. These were analysed by
hybridisation of restriction digests to the complete human PGK
cDNA and to subfragments of the cDNA, and also by heteroduplex
mapping. Of the PGK sequences isolated, one was shown to
represent a single 0.55 kb region homologous to the cDNA,
derived from the X-chromosome, which could be an exon of the
X-linked functional PGK gene. Three other sequences were
isolated and found to be intronless PGK sequences with homology
to the human PGK cDNA along the full length of the coding
sequence of the cDNA, and to varying extents into its 3' untranscribed
region. One of these intronless genes was identified as being X-linked.

82
4.2 Strategy

4.2.1 Southern transfer analysis of Chinese hamster PGK gene family.

Initial experiments were carried out using Southern transfer to analyse the Chinese hamster PGK gene family. The X-linked human PGK cDNA which had been cloned in the plasmid PGK-5, shown in figure 4.1 (S.M. Gartler, personal communication), was used to probe restriction digests of human and Chinese hamster DNA. (The cDNA was isolated from PGK-5 by restricting the plasmid with BamHI and extracting the smaller 1.8 kb fragment, containing the cDNA, from a low melting point agarose gel.)

Genomic DNA was extracted from the following sources: HEC (human embryo cells; female), human placental tissue (male), GM0743 (a cell line isolated from a human male with a severe PGK deficiency; Yoshida & Miwa, 1974), CHO-K1 ("wild-type" Chinese hamster ovary cells) and R1.1.7 (the PGK deficient variant of CHO-K1). DNA from each source was digested with EcoRI, BamHI, HindIII and XbaI and, after electrophoresis, transferred to nitrocellulose filters before hybridisation with the probe. The $^{32}$P-labelled human PGK cDNA was hybridised with the filters at 65 °C in 3x SSC and washed, with final stringency conditions of 2 x 15 min. washes in 0.3x SSC at 65 °C. After autoradiography, the hybridisation patterns of the fragments from the human and hamster DNA were compared. In addition, the fragment sizes of the variants GM0743 and R1.1.7 were examined with reference to the patterns observed with DNA from the "wild-type" human and hamster cells.

DNA isolated from CHO-K1, R1.1.7, a male Chinese hamster...
The diagram shows the restriction sites used to digest the plasmid for the purposes of a number of experiments referred to in this work, for example, *HindIII* digestion produced fragments which were used to probe for three different regions of the PGK gene (5', middle (M), and 3'). The shaded area represents the location of the human phosphoglycerate kinase cDNA.
Hind SV40 ori
Bam HI
/P s t I
Amp
/Pst I
Hind III
cDNA
Hind III
Bam HI
cell line (A3), and from male and female Chinese hamster tissue, was digested with EcoRI. DNA from male and female Chinese hamster tissue was also digested with XbaI and HindIII. After southern transfer analysis, by comparing the intensity of hybridisation of the PGK cDNA to different restriction fragments in male (XY) and female (XX) hamster DNA, it was possible to determine that some of the DNA fragments were derived from the Chinese hamster X-chromosome. (X-linked fragments showed approximately twice the hybridisation signal in DNA derived from female cells than from male cells). The sizes of the fragments observed in DNA isolated from animal tissue were compared to those seen in DNA extracted from the Chinese hamster cell lines.

4.2.2 Construction of Chinese hamster genomic library.

To create a genomic library of hamster DNA, DNA produced by Sau3A partial digestion was inserted into the 41 kb lambda replacement vector EMBL4 (Frischauf et al., 1983).

EMBL4 (figure 4.2) has a large capacity for inserted DNA (15 to 20 kb), and contains polylinker sequences, with multiple restriction sites, flanking the central replaceable fragment of the phage for simplicity of excision and cloning. In addition, it is possible to select for recombinant phages genetically by virtue of their spi- phenotype (Zissler et al., 1971).

To prepare the vector DNA, EMBL4 was first digested with BamH1, cutting in the polylinkers to create the 'sticky ends' for ligation to the hamster DNA Sau3A fragments. Further digestion with SalI removed short fragments of polylinker DNA from the ends of the middle fragment (Figure 4.2). Digestion was followed by removing samples from the reaction mixes and electrophoresing them through a 0.7% agarose gel. (Restriction
Figure 4.2

The lambda derived bacteriophage EMBL4: 
Vector for the CHO-K1 genomic library.

The arrangement of restriction sites around the replaceable 
13 kb central 'stuffer' fragment is shown. E = EcoRI, 
B = BamHI, S = SalI.
by SauI could be followed this way as this enzyme also cuts at two sites within the middle fragment). Isopropanol precipitation of the vector DNA removed the short fragments of linker DNA (which remained in the supernatant) and thus eliminated the possibility of the middle fragments competing with the hamster DNA partial digestion products during ligation to the vector arms.

The Sau3A partial digest of hamster DNA was prepared as follows: In a pilot experiment, the DNA was digested with Sau3A (1 unit per 20μg of DNA) at 37 °C. Samples were removed after 10 min. and at two and a half minute intervals thereafter for 20 min., heated at 65 °C for 10 min. and placed on ice. The samples were then electrophoresed through a 0.7% agarose gel to determine the incubation time at which the greatest proportion of fragments was in the 10 to 20 kb range. This experiment was repeated, but samples were removed at 1 min. intervals around this optimum time period (20 to 30 min. of digestion). These optimum conditions were scaled up and 200μg of DNA were digested using the same time, temperature, and DNA concentration, but half the enzyme concentration (Seed et al., 1982). The DNA was then extracted with phenol/chloroform and ether, ethanol precipitated, washed with 70% ethanol and dissolved in 0.5 ml of TE buffer. The digested DNA was then centrifuged through a 10% to 40% sucrose gradient (in 10mM Tris-HCl, 10mM EDTA, 0.2M Na-acetate, pH 8.0) at 26,000 r.p.m. for 24 hr. at 20 °C in an MSE 6 X 14ml swing-out rotor. Fractions of the gradient were taken and run on an agarose gel to determine which part of the gradient contained DNA fragments of the required size.

Using 0.5μg of vector DNA per ligation reaction, a number of ligations of EMBL4 arms and hamster DNA were carried out using
different molar ratios of insert to vector DNA. To show whether ligation was successful, samples of reaction mixes were removed before and after ligation and electrophoresed through a 0.7% agarose gel. One fifth of the ligation products from each reaction was packaged into phage in vitro (Enquist & Sternberg, 1979) and the titre of recombinants (plaque forming units, pfu) determined by plating out dilutions of the packaging products with the P2 lysogen Q359 (Karn et al., 1980) (table 4.1 and figure 4.3). The highest yield of recombinant phages per test experiment (1.04 X 10^4 recombinants) was obtained using a 2:1 molar ratio of insert DNA to vector DNA.

Using the optimal ratio determined by the test ligations and in vitro packagings, a number of scaled up ligations, using 4μg of EMBL4 arms and 8mg of Sau3A partial digest of CHO-K1 DNA, were carried out. The ligation products were packaged, using an Amersham in vitro packaging kit, to yield a library of 6.2 X 10^5 recombinant phage in total.

4.2.3 Screening of the library and isolation of recombinant phage.

The phage library was plated out with Q359 and screened using duplicate sets of nitrocellulose filters (Benton & Davis, 1977). The human PGK cDNA, isolated from the plasmid PGK-5, was used as a probe. EMBL4 was used as a negative control and was plated with the *E.coli* strain C600 (Appleyard, 1954). After five rounds of purification, eleven phage were isolated that hybridised strongly to the probe, and DNA was prepared from each.
Table 4.1

Titrations of recombinant phage with O359
produced by packaging ligation products.

Ratios of insert DNA to vector DNA used in the ligation reactions varied from 1:2 to 8:1, with the ratios of 1:1 and 2:1 in duplicate. (In addition, 0.4μg of undigested EMBL4 DNA was packaged and the products titrated using the E.coli strain C600; a control was included where no DNA was added to the packaging reaction and the products plated with C600).

*total pfu = pfu in total (250μl) of in vitro packaging products.
<table>
<thead>
<tr>
<th>DNA packaged (insert:vector ratio)</th>
<th>Dilution</th>
<th>pfu/50μl</th>
<th>*total pfu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:2</td>
<td>$10^{-1}$</td>
<td>76</td>
<td>$3.8 \times 10^3$</td>
</tr>
<tr>
<td>1:1</td>
<td>$10^{-1}$</td>
<td>143</td>
<td>$7.15 \times 10^3$</td>
</tr>
<tr>
<td>1:1</td>
<td>$10^{-1}$</td>
<td>184</td>
<td>$9.2 \times 10^3$</td>
</tr>
<tr>
<td>2:1</td>
<td>$10^{-1}$</td>
<td>160</td>
<td>$8.0 \times 10^3$</td>
</tr>
<tr>
<td>2:1</td>
<td>$10^{-1}$</td>
<td>207</td>
<td>$1.04 \times 10^4$</td>
</tr>
<tr>
<td>3:1</td>
<td>$10^{-1}$</td>
<td>108</td>
<td>$5.4 \times 10^3$</td>
</tr>
<tr>
<td>4:1</td>
<td>$10^{-1}$</td>
<td>84</td>
<td>$4.2 \times 10^3$</td>
</tr>
<tr>
<td>5:1</td>
<td>$10^{-1}$</td>
<td>46</td>
<td>$2.3 \times 10^3$</td>
</tr>
<tr>
<td>6:1</td>
<td>$10^{-1}$</td>
<td>45</td>
<td>$2.25 \times 10^3$</td>
</tr>
<tr>
<td>7:1</td>
<td>$10^{-1}$</td>
<td>50</td>
<td>$2.5 \times 10^3$</td>
</tr>
<tr>
<td>8:1</td>
<td>$10^{-1}$</td>
<td>54</td>
<td>$2.7 \times 10^3$</td>
</tr>
<tr>
<td>EMBL4</td>
<td>$10^{-4}$</td>
<td>170</td>
<td>$8.5 \times 10^6$</td>
</tr>
<tr>
<td>-DNA</td>
<td>$10^0$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 4.3

Titrations of recombinant phage from in vitro packaging reactions.

Packagings were carried out using the equivalent of 0.1μg of vector DNA ligated to different amounts of insert DNA (horizontal axis). The vertical axis represents the total number of recombinants obtained from each packaging reaction (taken from table 4.1).
4.3 Results and Discussion

4.3.1 Southern transfer analysis of Chinese hamster PGK gene family.

The results of the hybridisation of the PGK cDNA to digests of human and Chinese hamster DNA are shown in figures 4.4 and 4.5. Figure 4.4a shows the hybridisation of the human PGK cDNA with EcoRI digests of human and hamster DNA. Nine fragments of human DNA hybridised with the probe, their sizes being 16, 11.6, 9.6, 8.1, 6.5, 5.0, 4.0, 3.6 and 2.8 kb. This pattern of hybridisation confirms that observed by Michelson et al. (1983 and 1985a) except that Michelson does not make reference to the weakly hybridising 5.0 kb fragment.

Seven of these EcoRI fragments showed a significantly greater signal with the female HEC DNA than with the male DNA, as compared with the other two (9.6 and 8.1 kb) bands where the signals are similar with male and female DNA. This demonstrates the existence of both X-linked and autosomal PGK sequences in human genomic DNA.

It has been shown that the six X-linked DNA fragments observed by Michelson et al. represent two independent PGK genes. The 11.6, 6.5, 4.0 and 2.8 kb fragments constitute the X-linked PGK gene which is expressed in all somatic cells (Michelson et al., 1985a & 1985b) and the 3.6 kb and 16 kb fragments contain the PGK sequence of an X-linked processed pseudogene (Michelson et al., 1985a).

The 5.0 kb X-linked EcoRI fragment seen in Figure 4.4a might represent a PGK-related sequence which has sufficient homology with the probe to produce a weak signal, yet had not been detected previously, or it could be a fragment peculiar to the
Figures 4.4 and 4.5

Southern transfer analysis of human and
Chinese hamster PGK gene sequences.

Human and hamster DNA, digested with EcoRI (4.4,A), BamHI (4.4,B), HindIII (4.5,A) and XbaI (4.5B), were electrophoresed, transferred to nitrocellulose filter, and probed with the human PGK cDNA from PGK-5 (figure 4.1). The sources of DNA used were: HEC, human embryo cells in culture; HP, human placental tissue; GM0743, human cultured cells from a PGK deficient male; CHO-K1, the "wild-type" hamster cell line; R1.1.7, the hamster PGK deficient variant. The sizes (kb) of the hybridising fragments are indicated. (See also figure 4.6 for EcoRI digests of Chinese hamster DNA.)
cell line which may have arisen by mutation.

Figures 4.4b and 4.5 show two hybridization of the HSR cDNA probe to human and hamster DNA digested with Sall, Hinfl, and EcoRI. Similar patterns of hybridization have been observed with human digested using BglII (Singer-Bax et al., 1984c and 1986a).

In the human, four small fragments hybridize with the

![Image with bands and numbers]

shows six fragments of homologous DNA (7.3, 5.4, 3.8, 2.9, 2.3, and 1.8 kb), and there are 5.4 kb and 4.6 kb fragments present in the human and 7.3 kb and 2.9 kb fragments to be autosome and the 3.8 kb and 2.9 kb fragments to be linked, but it is not clear whether the 3.8 kb and 2.9 kb fragments are derived from the X-chromosome or from autosome.
HEC cell line which may have arisen by mutation.

Figures 4.4b and 4.5 show the hybridisation of the PGK cDNA to human and hamster DNA digested with BamHI, HindIII, and XbaI. The same patterns of hybridisation have been observed with human DNA digested using BamHI (Singer-Sam et al. 1984) and with HindIII (Michelson et al., 1985a).

In the human, four BamHI fragments hybridise with the probe: the 4.9 kb and 6.7 kb fragments hybridised more strongly to the female HEC DNA than to the male human DNA digests, suggesting that the PGK sequences on these fragments are X-linked. The signals from the 2.1 kb and 9.3 kb bands are similar in all three digests indicating that these fragments are not derived from the X-chromosome. However, Singer-Sam et al. (1984) have shown, using 4X and XY human DNA, that the 9.3 kb fragment is in fact X-linked.

The HindIII digests of human DNA show eight bands of hybridising DNA fragments (19, 7.6, 5.4, 4.8, 2.7, 0.8, 0.6 and 0.5 kb). Two of these (4.8 kb and 0.8 kb) are equally intense in the male and female DNA tracks and represent autosomally derived sequences. Michelson et al. (1985a) have shown that the 7.6 kb band consists of two co-migrating DNA fragments with PGK sequence homology, one being autosomal, and the other being X-linked. The remaining five fragments appear to be derived from the X-chromosome.

Hybridisation of XbaI digested human DNA with the PGK cDNA shows six fragments of homologous DNA (7.3, 5.4, 3.8, 3.3, 2.9 and 1.8 kb). Of these, the 5.4 kb and 1.8 kb fragments appear to be autosomal and the 7.3 kb and 3.3 kb fragments to be X-linked, but it is not clear whether the 3.8 kb and 2.9 kb fragments are derived from the X-chromosome or from autosomal
DNA.

Comparison of the hybridisation patterns of GM0743 DNA with those of HEC and placental DNA does not reveal any difference in the number or size of DNA fragments hybridising to the PGK cDNA. This result was confirmed by the work of Yoshida & Tani (1983). It has since been shown that the PGK deficiency in GM0743 is not due to a reduced transcriptional or post-transcriptional processing rate, but that the most likely causes of the enzyme deficiency are an increased rate of mutant PGK degradation combined with a reduced specific activity due, perhaps, to a point mutation in the coding sequence (Tani, et al. 1985b).

Michelson et al. (1985a) have carried out a more detailed analysis of the human PGK multigene family. This generally confirms the hybridisation patterns observed and the assignment of particular DNA fragments to the X-chromosome. Their analysis has revealed the existence of an X-linked PGK gene (PGK-1) closely linked to an intronless pseudogene which is proximal to the functional gene. They have cloned and sequenced this X-linked pseudogene and have also cloned and determined the structure of the functional PGK gene. Their results, along with those of Szabo et al. (1984), suggest the presence of two PGK genes on chromosome 6, possibly a complete functional gene (PGK-2) and an associated pseudogene.

An EcoRI digestion of Chinese hamster DNA produces eight DNA fragments which hybridise to the human PGK cDNA (13.5, 11.0, 9.8, 8.6, 7.8, 4.2, 2.6 and 1.9 kb : see Figures 4.4a and 4.6). This pattern also confirms that observed by other researchers (Michelson et al., 1983). BamHI digestion (Figure 4.4b) produces four strongly hybridising fragments (16, 10, 7.8 and 6.0 kb) and three weakly hybridising fragments (4.5, 2.5 and
Figure 4.6

Southern transfer analysis of EcoRI digested male and female Chinese hamster DNA.

The DNA was digested to completion with EcoRI, electrophoresed, transferred to a nitrocellulose filter, and probed with the human PGK cDNA from PGK-5. The sources of DNA were: CHO-K1, the "wild-type" cell line originally derived from a female Chinese hamster; R1.1.7, the GPI, PGK deficient variant derived from CHO-K1; A3, a cell line derived from a male Chinese hamster; the brain tissue from a male (MALE) and a female (FEMALE) Chinese hamster.

The sizes (kb) of the fragments are indicated.
<table>
<thead>
<tr>
<th>CHO-K1</th>
<th>RI.1.7</th>
<th>A3</th>
<th>MALE</th>
<th>FEMALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(XX)</td>
<td>(XY)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 13.5
- 11.8
- 9.8
- 8.6
- 7.8
- 4.2
- 2.6
- 1.9
At least eleven fragments are produced by digestion with HindIII (6.2, 4.6, 3.5, 2.3, 1.8, 1.3, 1.0, 0.85, 0.6, 0.55, and at least one of 0.4 kb or less). At least eight fragments are produced by digestion with XbaI (7.4, 6.0, 4.7, 4.3, 3.3, 3.0, 2.8 and 2.5 kb – see figure 4.5).

The fact that the human PGK cDNA hybridises strongly with non X-linked sequences and also with hamster sequences under conditions of high stringency (0.3x SSC at 65 °C) demonstrates a high degree of homology between X-linked and autosomal PGK sequences and interspecific conservation of sequence homology.

The human PGK cDNA hybridised to at least as many Chinese hamster DNA fragments as to those generated by the digestion of human DNA. This suggests that the Chinese hamster PGK gene family is at least as complex as that in the human where the existence of a minimum of four genes has been demonstrated.

Comparing the hybridisation patterns of CHO-K1 DNA with those of DNA from the PGK deficient variant of CHO-K1, R1.1.7, there are no observable differences in the number or sizes of DNA fragments. Thus, in R1.1.7, there appear to be no major structural changes in the nature of deletions or insertions of DNA in the coding regions of any of the PGK sequences which might explain the deficiency of PGK in this variant cell line. The change in R1.1.7 responsible for this lack of PGK activity might be a base change, or base changes, in the coding sequence or a small deletion or insertion of one or a few bases not detected in the Southern transfer analysis. These events would render the protein produced inactive, either by a resulting shift in the reading frame, by the incorporation of inappropriate amino acids, or by the appearance in the coding
sequence of a termination codon. Alternatively, it is possible that there might be a mutation in a non-coding region of the PGK gene resulting in incorrect transcription or processing of the messenger RNA.

Since it is the X-linked gene that is generally expressed in somatic cells, it is reasonable to assume that it is this gene that is affected in R1.1.7. The precise identification of the defect could be made by isolating the X-linked PGK gene from R1.1.7 and sequencing both the coding sequences and the control regions of the gene.

The present studies have shown that for most of the DNA fragments in the digests of human DNA it was possible, by comparing the strengths of the hybridisation signals in DNA from male and female cells, to identify which were derived from the X-chromosome and which from autosomal DNA. Therefore, in order to investigate the Chinese hamster PGK gene family further, digests of male and female DNA were probed with the PGK cDNA to identify any X-linked DNA fragments.

The resulting autoradiograms of HindIII and XbaI digests of DNA extracted from male and female animals are shown in Figure 4.7. Those obtained from EcoRI and BamHI digests were very unclear, and the only conclusion that could be drawn was from the EcoRI digests (Figure 4.6), where the 9.8 kb fragment gives a relatively greater signal than the remaining fragments in the female DNA, but not with the male DNA. Figure 4.7 shows that the fragments which appear to be X-linked are the 3.5, 1.8 and 1.0 kb fragments of the HindIII digests and the 4.7, 3.0 and 2.5 kb fragments of the XbaI digests. It is also possible that the 6.0 kb XbaI fragment might be X-linked.

If these few fragments identified as being X-linked are the
Figure 4.7

Southern transfer analysis of male and female Chinese hamster DNA digested with HindIII and XbaI.

The DNA was digested to completion with HindIII (A) and XbaI (B), electrophoresed, transferred to nitrocellulose filters, and probed with the human PGK cDNA.

The sources of DNA were brain tissue a from male and a female Chinese hamster.
only X-linked PGK sequences in the Chinese hamster genome, then a greater proportion of PGK sequences are autosomal than is the case in the human PGK gene family, where in the EcoRI digest, 6 of the 8 fragments containing PGK sequences have been identified as being X-linked, and 6 of the 9 HindIII fragments are also X-linked (Michelson et al., 1985a). In the XbaI digested hamster DNA, 3 or 4 of the eight fragments containing PGK sequences appear to give a greater hybridisation signal in the female DNA track (Figure 4.7), but only three HindIII fragments can be identified as being X-linked.

Much clearer results would have been obtained by using hamster DNA from cells containing more than two X-chromosomes, but such cell lines were not available.

4.3.2 Analysis of recombinant phage containing hamster PGK sequences.

DNA isolated from the recombinant phages which hybridised to the human PGK cDNA was digested using the following hexanucleotide recognising restriction enzymes in single and double digests: EcoRI, BamHI, HindIII, XbaI, SalI and XhoI. The digestion products were electrophoresed through 0.7% agarose gels, the sizes of the resulting fragments obtained, and restriction maps were constructed (Figure 4.8). These maps fell into four distinct non-overlapping groups (I, II, III and IV), suggesting that there are at least four gene sequences present in the Chinese hamster PGK gene family.

Southern transfer analysis of DNA isolated from four of the recombinant phages (8.1 from group I, 2.2 from group II, 10.1 from group III, and 13.1 from group IV) was carried out. Single and double digests of the DNA, using some of the above-mentioned
Restriction maps of the recombinants isolated, in all four groups of overlapping clones, are shown. The areas covered by each clone are illustrated by lines immediately below the restriction maps. E = EcoRI, X = XbaI, H = HindIII, B = BamHI, O = XhoI, S = SalI.

(The broken line above the restriction map of group I recombinants indicates a 2.3 kb BamHI fragment used as a probe in a later experiment - see figure 4.15).
enzymes, were probed with the full length PGK cDNA (Figures 4.9 and 4.10), and then with three probes produced by radiolabelling the three HindIII fragments of PGK-5 (Figures 4.11, 4.12 and 4.13). One PGK-5 HindIII fragment (1 kb) contained the 5' region of the cDNA (445 base pairs of the coding sequence and 44 base pairs of 5' non-coding DNA), the second consisted of 594 base pairs of the middle section of the cDNA, and the third (3.3 kb) contained the 3' region of the cDNA (212 base pairs of coding sequence, 434 base pairs of 3' untranslated DNA and a poly A tail). The fragments which hybridised to these probes are listed in table 4.2 and are shown on the restriction maps in figure 4.14.

Group I was represented by two recombinant clones: 8.1 and 13.2. In 8.1, the human PGK cDNA hybridised to a 0.75 kb XbaI/HindIII fragment of insert DNA, and a 1.1 kb HindIII/XhoI fragment. Hybridisation with the different regions of the PGK cDNA showed that the 0.75 kb XbaI/HindIII fragment contained homology with the 5' and middle regions of the cDNA, and that the HindIII/XhoI fragment only contained homology with the middle region of the cDNA. These restriction fragments were part of larger 3.5 kb and 1.8 kb HindIII fragments and a 3.0 kb XbaI fragment (figures 4.9 and 4.14). Southern transfer experiments previously described (see figure 4.7) suggested that restriction fragments of these sizes were derived from the Chinese hamster X-chromosome. It seems, therefore, that Group I represents PGK gene sequences that are X-linked.

These PGK gene sequences are part of an EcoRI fragment that is not fully represented in the insert of 8.1 or 13.2 so, in order to verify the X-linkage of this sequence, the size of the genomic EcoRI fragment on which it was contained was determined.
Restriction enzyme digests of recombinant 8.1 were hybridised with the human PGK cDNA from PGK-5.

E = EcoRI, B = BamHI, X = XbaI, H = HindIII,  
O = XhoI, S = SalI.

The sizes of the marker fragments from the HindIII digestion of lambda DNA are shown.
Figure 4.10

Southern transfer analysis of PGK sequences in recombinants 2.2, 10.1, and 13.1.

Recombinants 2.2 (group II), 10.1 (group III) and 13.1 (group IV), digested with EcoRI (E), XbaI (X), and HindIII (H), were probed with the human PGK cDNA. The fragments shown here are represented on the restriction maps in figure 4.15.

The sizes of the lambda marker fragments (m) are indicated.
Figures 4.11, 4.12, and 4.13

Southern transfer analysis of CHO-K1 PGK sequences in the recombinants: Hybridisation with different regions of the human PGK cDNA.

One clone from each group of recombinants was digested to produce the smallest restriction fragments which were seen to hybridise with the PGK cDNA in figures 4.9 and 4.10, and probed with the three HindIII fragments of PGK-5 containing the 5' (4.11), middle (4.12), and 3' (4.13) regions of the PGK cDNA (see figure 4.1).

The sizes of the lambda marker fragments (m) are indicated. ' + ' = the positive control, PGK-5 digested with HindIII (sizes of fragments shown).
Figure 4.11

5' PROBE
Figure 4.12

2.2 10.1 8.1 13.1
m + X, E, X, X, O, X,
H H H H H H H H H

23.1 9.4 6.6 4.4
2.3 2.0
(+) 0.6 0.56

MIDDLE PROBE
Figure 4.13

3' PROBE
Table 4.2

Recombinant restriction fragments containing PGK sequences.

The restriction fragments which hybridised to different regions of the human PGK cDNA were determined by comparing the sizes of the hybridising bands in figures 4.12, 4.13, and 4.14 with the restriction maps of the recombinants (Figure 4.8).

X = XbaI, H = HindIII.
<table>
<thead>
<tr>
<th>Group</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recomp.</td>
<td>8.1</td>
<td>2.2</td>
<td>10.1</td>
<td>13.1</td>
<td>1.3 Kbp H</td>
<td>1.1 Kbp H</td>
<td>0.5 Kbp H</td>
<td>0.2 Kbp X/H</td>
</tr>
<tr>
<td>5' + M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and compared to the sizes of EcoRI fragments in the Southern transfer analysis experiments (Figure 4.6).

CHO-K1 DNA was digested with EcoRI and electrophoresed on two adjacent lanes through an agarose gel, then transferred to a sheet of nitrocellulose filter. The DNA from one lane was probed with the human PGK cDNA, and that from the other was probed under the same conditions with a 2.3 kb BamHI fragment from 8.1 (Figure 4.8) which was isolated by 'freeze squeeze' extraction from an agarose gel. This BamHI fragment was chosen to detect that specific EcoRI fragment which was partly represented in 8.1.

Of the EcoRI fragments detected by the PGK cDNA, the 2.3 kb BamHI fragment from 8.1 hybridised to the 9.8 kb fragment (Figure 4.15). This was shown in an earlier Southern transfer experiment to be X-linked, confirming that the PGK sequences in 8.1 were derived from the X chromosome. (N.B. The BamHI fragment also produces a dark background of hybridisation with CHO-K1 DNA, probably demonstrating that it contains some high copy number repeated DNA sequence.)

Group II includes recombinants 2.2, 3.1 and 5.1. 2.2 contains the following restriction fragments which hybridise to the human PGK cDNA (Figure 4.10): two HindIII fragments of approximately 0.6 kb and 0.5 kb, and two HindIII/XbaI fragments approximately 0.5 kb and 0.2 kb in length. When probed with the 5', middle, and 3' regions of the cDNA, the HindIII/XbaI fragments hybridised with the 3' probe (Figure 4.13), the 0.5 kb HindIII fragment with the middle region (Figure 4.12), and the 0.6 kb HindIII fragment with the 5' region of the cDNA (Figure 4.11). Figure 4.8 shows that in 2.2 and 5.1, the HindIII/XbaI fragments are part of a larger 0.7 kb HindIII fragment of insert
Figure 4.15

Southern transfer analysis of CHO-K1 DNA PGK sequences using two different probes.

CHO-K1 DNA was digested with EcoRI and electrophoresed through an agarose gel in two adjacent tracks. After transfer to nitrocellulose filter, one track was hybridised with a 2.3 kb BamHI fragment from recombinant 8.1 (A) (see figure 4.8), and the second with the human PGK cDNA (B).
DNA, and that the total hybridising region is contained within a 2 kb EcoRI fragment and within two XbaI fragments of 7.8 kb and 2 kb. These single digest fragments were identified on the autoradiographs from the Southern transfer experiments with Chinese hamster DNA (Figures 4.4 and 4.5).

Recombinants 2.1, 4.1, and 10.1 comprise group III. The total hybridising region is covered by all three and consists of three adjacent restriction fragments: a HindIII fragment (1 kb), a HindIII/XbaI fragment (approximately 0.5 kb) and an XbaI fragment (approximately 0.2 kb). The 1 kb HindIII fragment hybridised to the 5' region of the cDNA (Figure 4.11) and the middle probe (Figure 4.12), the 0.5 kb HindIII/XbaI fragment to the middle section of the cDNA (Figure 4.12), and the 0.2 kb XbaI fragment to the 3' region of the cDNA (Figure 4.13). In 10.1 the HindIII/XbaI and XbaI fragments are contained within a 1.5 kb HindIII fragment, and this hybridising DNA is part of a 6.0 kb XbaI fragment (as well as the above-mentioned 0.2 kb fragment). In addition, it is contained within a 7 kb EcoRI fragment and a 10 kb BamHI fragment. From figure 4.7, it was determined that the 1.0 kb HindIII fragment, and possibly the 6.0 kb fragment, were X-linked. Thus, the Chinese hamster PGK sequence represented in the group III recombinants is probably derived from the X chromosome.

Group IV consists of recombinants 3.2, 12.1 and 13.1. Figure 4.10 (track X + H) shows that, in recombinant 13.1, the hybridisation to the PGK cDNA is restricted to two HindIII restriction fragments (approximately 0.4 and 1.3 kb) and an XbaI fragment (less than 0.2 kb). The 1.3 kb HindIII fragment hybridises strongly to the 5' region of the cDNA (Figure 4.11) and weakly to the middle region (Figure 4.12), the 0.4 kb
HindIII fragment to the middle region (Figure 4.12), and the <0.2 kb XbaI fragment to the 3' probe (Figure 4.13). Single-enzyme digests (figure 4.10) showed that the PGK cDNA hybridises to three HindIII fragments (2.3 kb, 1.3 kb and 0.4 kb), and two XbaI fragments (2.8 kb and <0.2 kb). Combining the restriction maps of the inserts of recombinants 12.1 and 13.1 (Figure 4.8) shows that this region is located on an EcoRI fragment approximately 13.5 kb long. There is only one internal BamHI site in 12.1, and there are none in 13.1, so the BamHI fragment containing this PGK sequence in the Chinese hamster must be at least 18 kb in length.

Group I recombinants differ from the others in that the restriction fragments which hybridise to the PGK cDNA contain hamster DNA sequences which are only homologous to the 5' and middle regions of the cDNA probe (figure 4.14). This suggests that these sequences might represent one or more exons of a PGK gene.

If this is the case, then the intron downstream of the hybridising region must be at least 3.5 kb long. As there appears to be no homology to the probe for at least 11 kb upstream of the hybridising restriction fragments (see Figure 4.8), it would be reasonable to assume that in an active gene the point of initiation of translation might be found within the 0.65 kb XbaI/HindIII fragment. The alternative, that there is an intron of at least 11 kb upstream from this fragment, seems less likely.

In groups II, III, and IV, the 5', middle, and 3' probes together hybridise to the hamster DNA over relatively short lengths of DNA. The individual fragments which hybridise to the three probes are shown in figure 4.14. The total hybridising
Figure 4.14

Restriction maps showing fragments containing Chinese hamster PGK sequences.

The DNA fragments which hybridise to the human PGK cDNA (see figures 4.9 and 4.10) are denoted by wide black lines. 5', M, and 3' indicate the regions of the PGK cDNA to which each fragment hybridised (see table 4.2).

E = EcoRI, X = XbaI, H = HindIII, B = BamHI, O = XhoI, S = SalI.
region in group II is restricted to one 2 kb EcoRI fragment. In group III it is restricted to two adjacent fragments totaling approximately 1.7 kb in length, and in group IV the hybridising region is contained on restriction fragments with a combined length of less than 2 kb. As the coding region of the cDNA is 1.25 kb long, very little space would remain for intron material in the fragments containing PGK sequences. It seems unlikely, therefore, that these recombinants contain PGK genes with intervening sequences.

4.3.3 Heteroduplex analysis.

Heteroduplex analysis of a number of recombinant phage was carried out using a 3.4 kb PstI fragment from PGK-5 as a probe (Figure 4.1). This fragment contained the 5' untranslated region of the PGK cDNA, 1,369 base pairs of coding sequence, 434 base pairs of 3' untranslated sequence, a poly A tail of 50 base pairs and approximately 1.5 kb of plasmid DNA 3' to the cDNA. The presence of non-homologous plasmid DNA 3' to the cDNA enabled the orientation of the PGK sequences in the recombinants to be determined. The results of the heteroduplex mapping are shown in figures 4.16, 4.17, 4.18 and 4.19.

Recombinant 2.2 (group II) hybridised with the PstI fragment of PGK-5 along one continuous length of DNA, estimated to be 1,320 bp long (Figure 4.17). There was a short, 50 bp, single stranded section at the 5' end of the cDNA which did not hybridise to the insert, and approximately 2 kb downstream of the coding region of the cDNA corresponding to non-homologous plasmid DNA. As the PGK coding region is 1369 base pairs in length, the hamster DNA appears not to hybridise with the cDNA along the total length of its coding region.
Figures 4.16, 4.17, 4.18, and 4.19.

Heteroduplex analysis of recombinant DNA.

Heteroduplex molecules were formed between the large PstI fragment of PGK-5 containing the human PGK cDNA (see figure 4.1) and recombinants 8.1 (group I; fig. 4.16), 2.2 (group II; fig. 4.17), 4.1 (group III; fig. 4.18) and 13.1 (group IV; fig. 4.19). The photographs are represented in line drawings, where arrows indicate the extent of annealing. In the case of 8.1, 'SB' indicates a "snap-back" caused by the presence of an inverted repeat sequence upstream of the PGK sequence. The PGK DNA sequences in the recombinants are indicated by shaded boxes in the linear diagrams.

Figures were calculated from standard measurements of M13mp9 DNA (double stranded) and pBR322 DNA, and are given to the nearest 10 base pairs.
Figure 4.16

(sb) PGK-5 / / 250 / inverted................................... repeat PGK-5

8.1 C

250 960 250 560 570
Figure 4.19

There is no detectable signal above background in the segment of the DNA representing the PGK-5 region. This would represent an approximately 3.2 kb DNA fragment containing the coding sequence. In addition, there are approximately 1.5 kb of non-coding DNA 3' to the coding sequence. In situ hybridisation with an exon which would be at least 1.5 kb would not give rise to a detectable signal above background.
In the case of recombinants 4.1 (group III - figure 4.18) and 13.1 (group IV - figure 4.19), the cDNA hybridises to the insert DNA along a greater length: approximately 1,610 bp in 4.1, and 1,522 bp in 13.1. In addition, there appears not to be a short tail of single stranded DNA at the 5' end of the cDNA. In both groups of recombinants, the double stranded regions corresponding to sequence homology between the cDNA and the insert DNA are considerably larger than the coding region of the cDNA, indicating some homology to untranslated regions of the cDNA. As the 5' untranslated region of the cDNA is only 44 bp in length, homology between 13.1 and the cDNA must continue into the 3' untranslated sequence for at least 140 bp, and in the case of 4.1, for at least 200 bp.

Recombinant 8.1 (group I - figure 4.16), contains an uninterrupted region of homology with the human PGK cDNA, approximately 570 base pairs in length. This region of homology begins at a position in the coding sequence some 250 base pairs downstream from the PstI site of PGK-5, and as such cannot include 150 to 200 bp of the 5' region of the coding sequence. There is no detectable homology anywhere in the 9.5 kb of hamster DNA contained in 8.1 upstream of this region. This would represent part of an unusually large intron if the PGK sequence represented an exon which was part of an active gene. The DNA in this region contained a 'snap-back', indicating the presence of an inverted repeat sequence (approximately 250 bp long) in the DNA upstream of the PGK sequence. In 8.1, there are approximately 3.5 kb of non-homologous hamster DNA 3' to the hybridising DNA. In an active gene, this would represent part of an intron which would be at least 3.5 kb long.

Thus, the hamster PGK sequence of the group I recombinant
phages appears to represent a single exon from an X-linked PGK gene, approximately 570 base pairs long, with an unusually large intron upstream and an intron of at least 3.5 kb downstream. Alternatively, this PGK sequence might represent a section of vestigial DNA derived from another PGK gene sequence. DNA sequencing to further investigate these two possibilities would reveal whether potential splice sites exist at the extremes of the coding sequence and whether there are any base changes which would exclude the possibility that the sequence could code for part of an active enzyme.

PGK DNA sequences represented in groups II, III and IV are all very similar in arrangement. Heteroduplex analysis showed that, in each case, the region of homology extended in one continuous section along the full length of the cDNA. In groups III and IV, at least 150 base pairs of 3' untranslated DNA hybridised to the cDNA. No single stranded loops of hamster DNA were observed, demonstrating the absence of introns. It is unlikely that any of these PGK sequences represent active genes in the Chinese hamster as intronless eukaryotic genes are exceptionally rare (Schaffner et al., 1978; Nagata et al., 1980; Houghton et al., 1981; Stein et al., 1983). These sequences may therefore be inactive intronless pseudogenes.

4.3.4 Further discussion.

Single restriction digest fragments on which all of the above PGK sequences are found were identified with restriction fragments found in Southern transfer analyses of Chinese hamster DNA (Figures 4.4 and 4.5). Of the fragments in different digests representing the PGK gene family, approximately half have been isolated in recombinant phage from the CHO-K1 DNA
library. Of the nine EcoRI fragments detected by the human PGK cDNA, four are represented in the recombinant phage, one in each group. Seven BamHI fragments hybridise to the PGK cDNA, and the four largest are represented, at least in part, one in each group of recombinants. Five of the eight XbaI fragments in CHO-K1 which contain sequences homologous to the human PGK cDNA are represented, there being one in each of groups I, III, and IV, and two in group II (in addition to those 0.2 kb fragments in groups III and IV which were not detected in the Southern transfer experiments with genomic DNA). There are at least 14 HindIII fragments which are detected by the human PGK cDNA, and 10 of these correspond to fragments found in the recombinant phage.

The fragments which are present in the recombinants isolated are generally those which hybridised most strongly to the probe in the CHO-K1 genomic blots. This is particularly noticeable in the BamHI digests. They all have continuous homology with the cDNA for at least 0.55 kb. Other, less well hybridising bands on these autoradiograms presumably represent fragments containing shorter lengths of PGK coding DNA (i.e. shorter exons) which were not as easily detected under the conditions of hybridisation used, both in the Southern transfer experiments and in probing the library. These sequences might still be represented in the CHO-K1 library (now amplified) and could be isolated by probing with the PGK cDNA at lower stringency than the original screening.

In the human PGK gene family, it has been shown that there exists an intronless pseudogene on the X-chromosome closely linked to the expressed PGK gene (Michelson et al., 1985a). The presence of a putative intronless pseudogene on the X-chromosome
of the Chinese hamster (represented in 10.1), as well as a functional X-linked PGK gene, would represent another similarity with the situation in the human PGK gene family besides the close sequence homology. The X-linked human PGK genes have been shown to be located on the same chromosomal band (Xq13). There is, however, no evidence for close direct linkage, as there are examples of translocations which separate these genes, and the two PGK sequences have not been isolated within a single clone of genomic DNA (Michelson et al., 1985a). The existence of such a linkage between a Chinese hamster X-linked intronless pseudogene and its functional counterpart could be resolved by mapping of the genes using a panel of somatic cell hybrids containing X-chromosome/autosome translocations or by the isolation of genomic clones linking both sequences.

4.4 Summary

Digests of human and hamster DNA, using a number of restriction enzymes, were probed with the X-linked human PGK cDNA and revealed, in each case, several DNA fragments containing homologous sequences. The hybridisation patterns of the human DNA were confirmed by the work of other investigators who have also identified X-linked restriction fragments. In both human and hamster, the complex hybridisation patterns suggest the existence of a number of independent phosphoglycerate kinase genes. Work carried out recently (Michelson et al., 1985a) has resulted in the identification of at least four PGK genes in the human genome, two of which have closely linked X-chromosomal loci (the X-linked functional PGK gene and an intronless pseudogene). At least one of the
remaining genes, a pseudogene, is located on chromosome 6p. A PGK-related sequence has also been mapped independently to chromosome 19 (Willard et al., 1985).

A number of fragments from restriction digests of Chinese hamster DNA were assigned to the X-chromosome by their increased hybridisation dosage in DNA from female as compared to male animals.

Recombinant phage were isolated from a genomic CHO-K1 DNA library by screening with the human PGK cDNA. The inserts of eleven recombinants were mapped, and these fell into four distinct groups of overlapping DNA fragments. Hybridisation of restriction digests of the recombinants to the complete PGK cDNA and to three subfragments of the cDNA (5', middle, and 3') showed that the PGK sequences of one group were homologous to parts of the 5' and middle regions and that those in the remaining three groups had homology with all three sections of the cDNA. Heteroduplex analysis further revealed that the first group had homology with the cDNA over one continuous section of DNA for approximately 0.57 kb, and that the other three each hybridised to the PGK cDNA along the full length of the coding sequence, and in two groups for at least 150 bp into the 3' untranslated region of the cDNA.

The 0.57 kb of PGK sequence in the first group of recombinants was identified as being X-linked, and as such might represent an exon belonging to the functional X-linked PGK gene. The PGK sequences found in the inserts of the remaining recombinants, one of which was assigned to the X-chromosome, probably represent intronless pseudogenes.
In this chapter, the isolation and characterisation of the four PGK-related Chinese hamster DNA sequences contained in the EMBL4 recombinant clones are described. DNA sequencing has revealed that these sequences represent three PGK pseudogenes, and a large exon from the Chinese hamster X-linked gene, PGK-1.

Pseudogenes are DNA sequences found in gene families, which exhibit considerable similarity to the functional genes, but also contain mutations which prevent the synthesis of functional protein products (for review see Wilde, 1986). Many contain mutations that interfere with the processing of a message or prevent transcription altogether. If the control regions are present and could potentially initiate efficient transcription and translation, mutations in the coding regions could still preclude accurate translation. Premature termination codons can arise through base changes and through deletion/insertion mutations which would produce shifts in the reading frame. Frame shifts would also completely alter the sequence of codons that would be translated in a potential polypeptide.

Initially, pseudogenes were found linked to their functional counterparts, for example globin pseudogenes (Proudfoot & Maniatis, 1980; Lacy & Maniatis, 1980, etc.) and often showed similar intron/exon structures. Later, many more were
discovered that were dispersed on different chromosomes in the genome, for example the pseudogenes related to mouse α-globin (Vanin et al., 1980), human metallothionein (Karin & Richards, 1982) and dihydrofolate reductase (Chen et al., 1982). Pseudogenes in this category were generally found to lack intervening sequences and contain other features which are found in messenger RNA molecules. These features include 5' homology with the corresponding functional gene up to the point where transcription is initiated and homology at the 3' end of the gene up to a point just beyond the polyadenylation signal, where a sequence of adenine nucleotides is often observed. In addition, these pseudogenes are generally flanked by direct repeat sequences suggesting that they were originally integrated into staggered breaks in the chromosomal DNA (Wilde, 1986). The Chinese hamster PGK genes exhibit a number of characteristics found in the latter class of intronless genes.

Two human PGK pseudogenes have been isolated and sequenced. One of these is X-linked and maps to the same chromosomal band as the human functional X-linked PGK gene (Michelson et al., 1985a). The second is located on chromosome 6 (Tani et al., 1985). Both human pseudogenes, like those found in the hamster, lack intervening sequences and exhibit the characteristics described above for the pseudogenes resembling messenger RNA sequences.

In this study, the Chinese hamster PGK genes have each been sequenced for their entire coding regions and to varying extents into their 5' and 3' non-coding regions. Mutations are present in all three which would prevent the production of a functional enzyme. Sequence divergence calculations carried out in order
to estimate the approximate time at which the pseudogenes diverged from the functional PGK gene suggested that each arose independently between 20 and 50 million years ago.

The putative exon from the Chinese hamster X-linked PGK gene corresponds to amino acids 54 to 244 of the human X-linked PGK gene, and contains no obvious critical mutations in its DNA sequence. Other features of its sequence, such as the splice sites and the 3’ splicing signal, are consistent with it being part of a functional gene.

5.2 Strategy

5.2.1 M13 Cloning.

One clone from each group of recombinant phage was chosen for detailed analysis of the phosphoglycerate kinase sequences. Recombinant 8.1 (group I) was digested to completion with HindIII and XbaI, and ligated separately into the double-stranded replicative form of the vectors M13mp18 and M13mp19 (Norrander et al., 1983) which had also been digested with these two enzymes but had, in addition, been treated with alkaline phosphatase to remove their 5’ phosphate groups. DNA fragments were inserted into the two vectors in opposite orientations. Similarly, recombinants 2.2 (group II), 10.1 (group III), and 13.1 (group IV) were digested with HindIII and ligated to HindIII-cut and alkaline phosphatase-treated M13mp18.

The ligation products were used to transform the E.coli strain JM101, which was then plated onto BBL agar plates in the presence of IPTG and X-gal to distinguish phage containing vector DNA (blue plaques) from those containing recombinant DNA.
(white plaques) (Messing et al., 1977).

Plates containing up to 50 white plaques were used in screening for recombinant phages containing inserts that hybridised to the human PGK cDNA.

5.2.2 Screening M13 Plagues.

Plaques resulting from transformations of JM101 with M13mp18 ligated to 2.2, 10.1, or 13.1 HindIII fragments, were transferred onto nitrocellulose filter circles and hybridised with the PGK cDNA radioactively labelled by 'nick-translation'. Twenty-four white plaques which hybridised to the cDNA were picked from each set of transformations (except in the case of the M13/10.1 recombinants, where only ten white plaques hybridised, and all of these were picked). Twenty-four white plaques were also picked from each of the M13mp18/8.1 and M13mp19/8.1 transformations.

The M13 plaques were each incubated in 1.5 ml of L-broth at 37 °C, with shaking, for 6 hours, the bacteria pelleted, and 10 µl samples of each supernatant containing the M13 phages were spotted onto three nitrocellulose filter circles. The filters were hybridised with radio-labelled subfragments (5′, middle and 3′) of the PGK cDNA. The results, shown in figure 5.1, illustrate the arrangement of supernatant samples on the filters (a), the hybridisation of these supernatants to the probes (b), and list the recombinant M13 phages which hybridise to each of the probes (c).

5.2.3 Determination of M13 Insert Orientation.

The HindIII/XbaI fragments from 8.1 were inserted into the M13 vector DNA in pre-determined orientations : the
**Figure 5.1**

Identification of CHO-K1 DNA fragments contained in M13 recombinant clones.

XbaI/HindIII fragments from 8.1 were ligated into M13mp18 and M13mp19, and HindIII fragments from 2.2, 10.1, and 13.1 were ligated into M13mp18.

Figure 5.1a. shows the arrangement of M13 clones, containing insert fragments from the CHO-K1/EMBL4 recombinant phage, on the nitrocellulose filters. Figure 5.1b. shows the results of hybridising the filters to the 5', middle, and 3' regions of the human PGK cDNA. Figure 5.1c. lists the M13 clones which contain XbaI/HindIII fragments (from 8.1) and HindIII fragments (from 2.2, 10.1 and 13.1), showing whether they hybridise to the 5', middle, or 3' regions of the human PGK cDNA.

* In 10.1, the middle and 3' regions are contained on the same HindIII fragment (see figure 4.8).

+ In clone 24 from 13.1, two different fragments appear to be ligated together in the M13 vector.
multi-restriction site polylinker into which they were ligated in M13mp18 was in the opposite orientation to that in M13mp19, and since there were different restriction sites at each end of the fragments, they could only ligate into each vector in a particular orientation.

The HindIII fragments of 2.2, 10.1, and 13.1 were inserted into M13mp18 in both orientations. The relative orientations of the inserts in those M13 clones containing the Chinese hamster PGK gene sequences were determined as described in section 2.2.21 of "Materials and Methods".

Single-stranded DNA from these phages was then prepared and used as templates for sequencing reactions.

5.2.4 DNA Sequencing.

DNA sequencing was carried out using the dideoxy chain termination technique (Sanger et al., 1977). Initially, a universal primer was used to prime the sequencing reactions. M13 template DNA was prepared as described in section 2.2.20 of "Materials and Methods". The primer was annealed to the M13 DNA adjacent to the polylinker and thus primed directly into the insert. The products of the sequencing reactions were split into three, and the second and third samples loaded onto the 6% polyacrylamide/urea sequencing gel when the bromophenol blue dye from the previous sample had reached the bottom of the gel. In this way, it was possible to read at least 300 bases, and often over 400 bases, of DNA sequence from each sequencing gel. Having determined the DNA sequence thus far into the inserts, 17 mer oligonucleotides were synthesized as described by Matthes et al., 1984 and Brenner & Shaw, 1985. These oligonucleotides
were complementary to the DNA, usually approximately 250 bases from the original priming point. The complements of these oligonucleotides were also synthesized. By using these oligonucleotides as primers for the next round of sequencing on the same M13 templates, the DNA sequences further into the inserts were determined, as well as those of the DNA strands complementary to the sequences already determined (Figure 5.2).

In this way, the sequences of both DNA strands of the Chinese hamster DNA inserts were determined. An example of a typical autoradiogram from a DNA sequencing gel is shown in figure 5.3.

5.3 Results and Discussion

5.3.1 Analysis of the putative PGK gene exon.

5.3.1.1 Coding sequence.

The two XbaI/HindIII restriction fragments of recombinant 8.1 (2.2 kb and 0.75 kb) containing the PGK nucleotide sequence were each sequenced from their common HindIII sites (in M13mp18), and the 0.75 kb fragment was also sequenced in the opposite direction, from its XbaI terminus (in M13mp19) (Figure 5.2). The nucleotide sequence is shown in Figure 5.4 aligned with the human PGK cDNA sequence.

The region of homology between the PGK sequence of 8.1 and the X-linked human PGK cDNA begins at the second nucleotide of the codon for Gly 53, and continues for a length of 575 bases up to and including the codon for Leu 244. Comparison of the
Figure 5.2

Strategy for sequencing CHO-K1 PGK DNA sequences.

'UP' denotes the points where sequencing was primed using the M13 universal primer. Primers complementary to 17-nucleotide stretches of DNA within the inserts were used as internal primers and are denoted by '→' and a number, e.g. A1, and its reverse complement, A1-RC, for priming on the opposite DNA strand.
Figure 5.3
DNA sequencing gel.

The autoradiogram shows the arrangement of bands in a DNA sequencing gel. The sequence is read from bottom to top using the bases indicated at the top of the tracks. Two sequences are shown (A and B). Samples in (I) and (II) are from the same reaction mixes, those in (II) having been loaded onto the gel later in the run. A_ shows the sequence of the CHO-K1 DNA from recombinant 8.1, upstream of the PGK coding sequence, using primer B1-RC (see figure 5.2). B_ shows the sequence in 8.1 from primer B1 (see figure 5.2). In B_, the sequence of the 3' splice signal and that of the 3' splice junction are indicated.

'+' denotes the points at which the sequences continue in (I), at the bottom of the gel, from the points indicated in (II).
A  B
CGAT  CGAT
^  «
5  8
n
. - g
a a m
I «

A  B
CGAT  CGAT
Ü j *
-É 3
1
j =
- - 4
2
-%. TTC*
«J** 3 splice signal
'S - Z
I splice ICC A'
The partial amino acid sequence of the Chinese hamster PGK, deduced from the DNA sequence, is shown above the hamster DNA sequence and the positions of the amino acids in the full length PGK polypeptide are indicated. '*' indicates a base difference between the human and hamster PGK coding sequence. Where base differences result in amino acid differences, the amino acid from the human PGK sequence is shown below the codon. The sequences underlined are, from top to bottom, a.) the putative 3' splice signal, b.) the 3' splice junction (AG<), c.) the internal HindIII site in the hamster coding sequence, and, d.) the 5' splice junction (>GT).
CHPGK-1: GCAAGGTAATCTTCACACCACCTTTTCTTGTTTGCCATGAGATCTTGAC

CHPGK-1: CATCTTGCCCCCTTCTTCACTACAGAGAAGTGCCAATCCTGTTTGAG

CHPGK-1: 

Hu.cDNA: CC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT

Hu.cDNA: GGT GTT CCC ATG CCT GAC AAG TAC TCC TTA GAG CCA GTT GCT

Hu.cDNA: GGT GCC ATG CCT GAC AAG TAC TCC TTA GAG CCA GTT GCT

Hu.cDNA: CC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT

Hu.cDNA: GGT GTT CCC ATG CCT GAC AAG TAC TCC TTA GAG CCA GTT GCT

Hu.cDNA: GGT GCC ATG CCT GAC AAG TAC TCC TTA GAG CCA GTT GCT

Hu.cDNA: GGT GTT CCC ATG CCT GAC AAG TAC TCC TTA GAG CCA GTT GCT

Hu.cDNA: GGT GCC ATG CCT GAC AAG TAC TCC TTA GAG CCA GTT GCT

Hu.cDNA: GGT GTT CCC ATG CCT GAC AAG TAC TCC TTA GAG CCA GTT GCT
Ser Leu Ser Lys Leu Gly Asp Val Tyr Val Asn Asp Ala Phe
CHPGK-1: TCA CTG TCC AAA CTT GGG GAT GTC TAT GTC AAT GAT GCT TTT
Hu.cDNA: TCA CTT TCC AAG CTA GGG GAT GTC TAT GTC AAT GAT GCT TTT
* * *

Gly Thr Ala His Arg Ala His Ser Ser Ser Met Val Gly Val Asn
CHPGK-1: GGA ACT GCA CAC CGA GCC CAC AGC TCC ATG GTG GGT GTG AAT
Hu.cDNA: GGC ACT GCT CAC AGA GCC CAC AGC TCC ATG GTA GGA GTC AAT
* * * *

Leu Pro Gln Lys Ala Gly Gly Phe Leu Met Lys Lys Glu Leu
CHPGK-1: CTG CCA CAG AAG GCT GGT GGA TTT TTG ATG AAG AAG GAG CTG
Hu.cDNA: CTG CCA CAG AAG GCT GGT GGG
* ** *

Asn Tyr Phe Ala Lys Ala Leu Glu Ser Pro Glu Ser His Phe
CHPGK-1: AAC TAC TTT GCC AAA GCT ATC TTG GGA GGA GCT AAA GTT GCA GAC AAG ATC CAG
Hu.cDNA: AAC TAC TTT GCA AAG GCC TTG GAG AGC CCA GAG CGA CCC TTC
* * * * * * *

Leu Ala Ile Leu Gly Gly Ala Lys Val Ala Asp Lys Ile Gln
CHPGK-1: CTG GCT ATC TTG GGA GGA GCT AAA GTT GCA GAC AAG ATC CAG
Hu.cDNA: CTG GCC ATC TTG GGA GGA GCT AAA GTT GCA GAC AAG ATC CAG
* * * *

Leu Ile Asn Asn Met Leu Asp Lys Val Asn Glu Met Ile Ile
CHPGK-1: CTG ATC AAT AAT ATG CTG GAC AAA GTC AAT GAG ATG ATT ATT
Hu.cDNA: CTC ATC AAT AAT ATG CTG GAC AAA GTC AAT GAG ATG ATT ATT
* * *

Gly Gly Gly Met Ala Phe Thr Arg Leu
CHPGK-1: GGT GGT GGA ATG GCT TTC ACC AGG CTG＞GTAAACCCACAGAATCG
Hu.cDNA: GGT GGT GGA ATG GCT TTC ACC TTC CTG d.
* *** *

Phe

CHPGK-1: CTGGCCTGAACAGGGGGAGCACTCGACACCGGATATTTCTGTTAAGTG.
nucleotide sequences reveals 45 mismatches representing a sequence divergence of 7.8%. (This difference represents the sum of the base changes in each of the DNA sequences since diverging from their common ancestral gene.) None of the mismatches in the Chinese hamster PGK coding sequence result in the appearance of termination codons, so the sequence could be transcribed without premature termination. As there are no insertion or deletion mutations, the sequence can be read 'in frame' throughout its entire length.

Of the base changes, 14 are changes at 'replacement sites', that is, they have resulted in changes which would bring about a change of amino acid. The remaining 29 are 'silent' base changes which do not produce codons for different amino acids. The consequent difference in amino acid sequence (14 out of 191 amino acids represented) represents a 7.3% divergence of protein primary structure. Of particular note are differences in 4 out of 6 amino acids at positions 104 to 109 (Gly-Asn-Ala-Phe-Ala-Ser in the hamster, compared to Glu-Lys-Ala-Cys-Ala-Asn in the human) and residues 205 and 206 (Ser-His in the hamster, contrasting with Arg-Pro in the human protein). The remaining differences between the PGK enzymes of the two species are the following isolated amino acid changes (human → hamster): Ser → Asn (residue 61), Val → Ala (82), Ser → Thr (114), Ala → Asp (134 and 144), Val → Ile (139), Glu → Asp (147) and Phe → Arg (243).

Comparison of the amino acid sequence of the human PGK with that of the horse (Banks et al., 1979; Hitzeman et al., 1982; Michelson et al., 1983) shows only four differences over the region represented in the hamster PGK sequence, so the human PGK
DNA sequence over this region is more closely related to the horse sequence than to the hamster sequence. However, there are 18 amino acid differences between the hamster and horse coding sequence, so the primary structure of the hamster protein represented is more closely related to human PGK than to the horse enzyme.

Although there are 18 amino acid differences between the horse and the deduced Chinese hamster PGK sequence between residues 55 and 244, there is still considerable sequence homology (90.5%) between the two proteins.

The region of the hamster PGK gene covered codes for a section of polypeptide which, in the horse, is represented approximately equally in each of the two domains of the protein, beginning in the N-terminal ATP/ADP binding domain of the protein and extending into the C-terminal domain of the protein. A number of amino acids in the horse PGK protein have been implicated as being involved in the binding of ADP and ATP (Banks et al., 1979) due to their close proximity to the proposed binding site. These include arginine residues 65, 122 and 170, histidines 62, 169 and 172, and glutamates 127 and 128. Glycine residues 185 and 186 are thought to be important in allowing freedom of movement of the two domains with respect to each other as they form the main link between the domains. All of these amino acids are conserved in the Chinese hamster sequence as well as in the human PGK sequence. Thus, none of those amino acids identified as, perhaps, having a direct effect on the ATP/ADP binding properties of the enzyme or the movement of the domains in catalysis, have been replaced.
5.3.1.2 Potential splice junctions.

Examination of the DNA sequences immediately flanking the exons of eukaryotic genes has revealed the presence of consensus sequences (Breathnach et al., 1978; Breathnach & Chambon, 1981; Mount, 1982). The consensus sequence determined by Mount (1982) at the 5' end of the exon (the 3' splice junction of the intron) is 5'..(T/C)\(_N\)(C/T)AG<G..3', and at the 3' end of the exon (5' splice junction) is 5'..(C/A)AG>GTAAGT..3'. Breathnach et al. (1978) noted the requirement for an AG immediately adjacent to the first nucleotide of an exon and GT adjacent to the most 3' coding nucleotide.

The DNA sequences at the 5' and 3' ends of the putative Chinese hamster PGK coding sequence in 8.1 are shown in Figure 5.4. Although these sequences do not conform exactly to the consensus sequences, they do conform to the GT/AG rule which, with very few exceptions, is conserved absolutely in all eukaryotic genes. As the 3' terminus consensus sequence of an intron ends in AG, the putative splice junction at the 5' end of the Chinese hamster PGK coding sequence would be between the guanosine and cytosine residues in the 1st and 2nd positions of the Ala 54 codon. Thus, although homology of the hamster DNA with the human PGK cDNA begins three bases upstream of this position, the potential exon would only contribute two nucleotides of the Ala 54 codon, and none of the Gly 53 codon, to the final processed transcript. At the 3' end of the hamster PGK sequence, the GT required at the 5' terminus of the downstream intron places the splice junctions precisely where the effective homology with the human PGK sequence ends (between the codons for Leu 244 and Lys 245), and although the final G of the putative exon differs from the equivalent nucleotide in the
human sequence, the amino acid remains unchanged.

5.3.1.3 Potential 3' splice signal.

Using computer searches, Keller & Noon (1984) have identified potential 3' signals which, they propose, are required as part of the general mechanism for the splicing of precursor messenger RNAs in eukaryotes. They have determined a number of consensus sequences for the 3' splice signals of a number of different organisms. By examining base positions -10 to -60 from the 3' splice junction they determined a consensus sequence of C-T-G-A-C for the intron sequences of rodents (rats and mice). Examination of the Chinese hamster DNA sequence upstream of the PGK coding sequence reveals a sequence of five nucleotides (T-T-G-A-C) which matches well with the consensus sequence, at base positions -54 to -58 with respect to the 3' splice junction (Figure 5.4).

Investigations of intron splicing (Grabowski et al., 1984; Padgett et al., 1984; Ruskin et al., 1984) have led to the isolation of a splicing intermediate in the form of a lariat, in which the 5' terminal G of an intron is covalently joined to a point near the 3' splice site of the intron. The branch points are always at A residues (Padgett et al., 1984), and, in those cases examined, these A residues were found to be part of conserved sequences (Ruskin et al., 1984) which matched well with the 3' splice signals. In all the 3' splice signals determined, the A in position 4 is conserved absolutely (Keller & Noon, 1984), suggesting a role for these potential splice signals in the splicing mechanism.
5.3.1.4 Comparison with the exon/intron structure of the human PGK gene.

The potential Chinese hamster PGK exon is compared to the exon/intron arrangement of the human X-linked PGK gene (Michelson et al., 1985b) in Figure 5.5. It is immediately obvious that the structures are very different. The 572 base pair PGK sequence in the hamster is represented in the human PGK gene by most of exon 3 (109 bp), all of exons 4 (145 bp), 5 (104 bp) and 6 (120 bp), and most of exon 7 (94 bp). Including introns, this is equivalent to nearly 9.5 kb of DNA. Thus, if the hamster PGK sequence represents an exon which is part of a functional gene, there is no conservation of exon/intron arrangement at all in this complete section of the coding region. As the splice sites in the hamster sequence occur at positions which are in the middle of exons 3 and 7 of the human PGK gene, the difference cannot be explained simply by the precise loss of introns 3, 4, 5 and 6 in the evolution of the Chinese hamster PGK-1.

The lack of intron 6 is particularly interesting, as this is within that part of the gene that codes for the nucleotide-binding domain of the human PGK enzyme. Michelson et al. (1985b) have suggested an intron-mediated pathway for the evolution of the nucleotide-binding domain of a number of proteins. They have compared the human PGK DNA sequence to the secondary structure of the horse PGK enzyme and determined that the exons in the human gene (separated by introns 6, 7, 8 and 9) corresponded to the subdomains of the nucleotide binding region of the protein. Comparing this to the structure of the chicken glyceraldehyde-3-phosphate dehydrogenase and the maize alcohol dehydrogenase, they found that this arrangement of exons and
Figure 5.5

Comparison of the putative Chinese hamster PGK-1 exon with the intron/exon structure of the human PGK-1 gene.

The putative hamster PGK exon is aligned with the corresponding region of the human PGK-1 gene to show the relative complexity of the intron/exon structure of the human gene.

In the human PGK-1 gene, E3, E4, E5, E6, and E7 = exons 3, 4, 5, 6, and 7; I3, I4, I5, I6, and I7 = introns 3, 4, 5, 6, and 7. Shaded regions of the exons represent regions of the human PGK gene coding for part of the nucleotide binding domain.
introns was conserved in these genes - including intron 6 between two β-strands thought to be involved in binding the nucleotide substrate in human PGK. With such inter-species (including inter-kingdom) and inter-enzyme conservation of intron/exon structure in these nucleotide-binding domains, the Chinese hamster PGK gene might also be expected to contain introns in the same positions. However, it is possible that, in the evolutionary history of the Chinese hamster PGK gene, intron 6 was lost without the loss of any adjacent coding sequence. It should also be noted that the position of the splice site at the 3' end of the hamster PGK sequence is consistent with the proposed intron-mediated evolution of the nucleotide-binding domain: It is only seven amino acid residues upstream of the position of intron 7 in the human gene, and still, therefore (assuming a similar secondary structure for human, horse and hamster PGK enzymes), between two β-strands of the nucleotide binding domain which are represented in separate exons in the above diverse enzymes (Michelson et al., 1985b).

There are a number of features of the Chinese hamster PGK DNA sequence isolated in recombinant 8.1 which suggest that it represents an exon which is part of the functional X-linked PGK gene (PGK-1). This Chinese hamster PGK sequence, denoted by 'CHPGK-1', has the following properties: The full sequence represents an open reading frame of 572 bases with 92.2% homology with the human PGK cDNA. This reading frame is not interrupted by deletions or insertions with respect to the human DNA sequence. The amino acid changes which have been deduced from the hamster nucleotide sequence do not involve any of the
amino acids which have so far been directly implicated in the binding or catalytic activities of the PGK enzymes studied. The hamster PGK coding region has putative splice sites at its 5' and 3' termini and, in addition, a potential 3' splice signal has been identified 54 to 58 bp upstream of this sequence. The major difference between the hamster PGK sequence and the human X-linked PGK gene is that this putative exon differs greatly in exon/intron arrangement from the situation in the human.

5.3.2 Structure of Chinese hamster PGK pseudogenes.

The DNA sequences of all three intronless PGK sequences (PGK-II, PGK-III, and PGK-IV) were determined over those regions which are homologous to the coding region of the human PGK cDNA (Figure 5.6). The nucleotide sequence upstream of the coding region was determined, in each case, beyond the 5' terminus of the cDNA (the transcriptional start of the human PGK gene) and the point at which homology between the three genes ends. The PGK sequences of PGK-II and PGK-III have been sequenced for the entire region corresponding to the 3' untranslated sequence of the human cDNA. The DNA sequence of PGK-IV has been determined up to a position 175 nucleotides downstream of the translation termination codon. In all cases, the sequences have been aligned with the human PGK cDNA to give maximum homology between the sequences. (It is more appropriate to use the human cDNA for comparison as only a limited amount of Chinese hamster PGK coding sequence data has been obtained.)

5.3.2.1 PGK coding region.

It can be seen in Figure 5.6 that, when compared to the human PGK cDNA, all three intronless PGK sequences contain
Figure 5.6

Coding regions of the Chinese hamster pseudogenes.

The coding regions of the Chinese hamster pseudogenes are shown aligned with the coding sequence of the human PGK cDNA. Termination codons, both in-phase with the correct PGK coding sequence and in-phase with the pseudogene sequences altered by frame shifts, are underlined. Also underlined are the HindIII sites within these genes. Regions of DNA which have been deleted in each of the three pseudogenes, (____), are flanked by direct repeats (underlined in the human cDNA sequence).
<table>
<thead>
<tr>
<th>Human cDNA</th>
<th>PGK-ψII</th>
<th>PGK-ψIII</th>
<th>PGK-ψIV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Ser Leu Ser Asn Lys Leu Thr Leu Asp Lys Leu</td>
<td>ATG TCG CTT TCT AAC AAG CTG ACG CTG GAC AAG CTG</td>
<td>ATG TCG ATT TCT AAA AAG CTG ACT TTG GAC CAG CTG</td>
<td>ATG TCG CTT TCT AAC AAG CTG ACT TTG GAC AAG CTG</td>
</tr>
<tr>
<td>GTC</td>
<td>GTC</td>
<td>GTC</td>
<td>GTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cDNA</td>
<td>GAC GTT AAA GGG AAG CGG GTC GTT AGA GTC GAC TTC AAT</td>
<td>GAC ATG AAG GGG AAG AGG GTG GTG ATG AGA GTG GAC TTT AAT</td>
<td>GAC ATG AAG GGG AAG AGG GTG GTG ATG AGA GTG GAC TTT AAT</td>
</tr>
<tr>
<td>ψII</td>
<td>GAC ATG AAG GGG AAG AGG GTG GTG ATG AGA GTG GAC TTT AAT</td>
<td>GAC ATG AAG GGG AAG AGG GTG GTG ATG AGA GTG GAC TTT AAT</td>
<td>GAC ATG AAG GGG AAG AGG GTG GTG ATG AGA GTG GAC TTT AAT</td>
</tr>
<tr>
<td>ψIII</td>
<td>GGC ATG AAG GGG AAA CAG GTC GTG TTG AGG GTC GAC TTC AAG</td>
<td>GAG GTG AAG AAG CAG GTC GTG ATG AGT ATG GAC TTC AAT</td>
<td>GAG GTG AAG AAG CAG GTC GTG ATG AGT ATG GAC TTC AAT</td>
</tr>
<tr>
<td>ψIV</td>
<td>GAA GTG AAG AAG CAG GTC GTG ATG AGT ATG GAC TTC AAT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
</tr>
<tr>
<td>26 27 28 29 30 31 32 33 34 35 36 37 38 39</td>
<td>Val Pro Met Lys Asn Asn Glu Ile Thr Asn Asn Glu Arg Ile</td>
<td>Val Pro Met Lys Asn Asn Glu Ile Thr Asn Asn Glu Arg Ile</td>
<td>Val Pro Met Lys Asn Asn Glu Ile Thr Asn Asn Glu Arg Ile</td>
</tr>
<tr>
<td>cDNA</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
</tr>
<tr>
<td>ψII</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
</tr>
<tr>
<td>ψIII</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
</tr>
<tr>
<td>ψIV</td>
<td>GTC CCT ATG AAG AAC AAT CAG ATG ACA AAC AAC CAG AGG ATT</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
</tr>
<tr>
<td>40 41 42 43 44 45 46 47 48 49 50 51 52 53</td>
<td>Lys Ala Ala Val Pro Ser Ile Lys Phe Cys Leu Asp Asn Gly</td>
<td>Lys Ala Ala Val Pro Ser Ile Lys Phe Cys Leu Asp Asn Gly</td>
<td>Lys Ala Ala Val Pro Ser Ile Lys Phe Cys Leu Asp Asn Gly</td>
</tr>
<tr>
<td>cDNA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
</tr>
<tr>
<td>ψII</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
</tr>
<tr>
<td>ψIII</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
</tr>
<tr>
<td>ψIV</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
<td>AAG GCT GCT GTC CCA AGC ATC AAA TTC TGC CTG GAC AAT GGA</td>
</tr>
<tr>
<td>54 55 56 57 58 59 60 61 62 63 64 65 66 67</td>
<td>Ala Lys Ser Val Val Leu Met Ser His Leu Gly Arg Pro Asp</td>
<td>Ala Lys Ser Val Val Leu Met Ser His Leu Gly Arg Pro Asp</td>
<td>Ala Lys Ser Val Val Leu Met Ser His Leu Gly Arg Pro Asp</td>
</tr>
<tr>
<td>cDNA</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
</tr>
<tr>
<td>ψII</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
</tr>
<tr>
<td>ψIII</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
</tr>
<tr>
<td>ψIV</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
<td>GCC AAG TCG GTA GTC CTT ATG AGC CAC CTA GGC CGG CCT GAT</td>
</tr>
<tr>
<td>68 69 70 71 72 73 74 75 76 77 78 79 80 81</td>
<td>Gly Val Pro Met Pro Asp Lys Tyr Ser Leu Glu Pro Val Ala</td>
<td>Gly Val Pro Met Pro Asp Lys Tyr Ser Leu Glu Pro Val Ala</td>
<td>Gly Val Pro Met Pro Asp Lys Tyr Ser Leu Glu Pro Val Ala</td>
</tr>
<tr>
<td>cDNA</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
</tr>
<tr>
<td>ψII</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
</tr>
<tr>
<td>ψIII</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
</tr>
<tr>
<td>ψIV</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
<td>GGT GTG CCC ATG CCT GAC AAG TAG TCC TTA GAG CCA GTT GCT</td>
</tr>
</tbody>
</table>
**cDNA**: GAG GAA CTC AAA TCT CTG CGC AAG GAT GTT CTG TTC TTG

**ψII**: GAG GAA CTC AAA TCT CTG CGC AAG GAT GTT CTG TTC TTG

**ψIII**: GCA GAA CTC AAA TTT CTG CGC AAG GAT GTT CTG TTC TTG

**ψIV**: GAA GCA CTC AAA TCT CTG ATT GGT AAG GAT GTT CTG TTC TTA

**cDNA**: CCA GCT GGT GCC TCT GTC ATC CTG GAG AAC CTC CGC TTC

**ψII**: AAG GAC TGT (__________)TC ATC CTG CTG GAG AAC CTC TGC TTC

**ψIII**: CCA GCA GCT GCC ACT GTC ATC CTG CTG GAG AAC ACT CTC TTC

**ψIV**: CCA GCA GCT GCC ACT GTC AGA CTT CTG GAG AAC CTC TGC TTC

**cDNA**: CAT GTG GAG GAA GAA GGG AAG GGA AAA GAT GCT TCT GGG AAC

**ψII**: CAT GTA GAG GAA GAA GGG AAG GGA AAA GAT GCT TCT GGG AAC

**ψIII**: CAT GTA GAG GAA GAA GGG AAG GGA AAA GAT GCT TCT GGG AAC

**ψIV**: CAT GTA GAG GAA GAA GGG AAG GGA AAA TAT GCT TCT GGG AAC

**cDNA**: AAG GTT AAA GCC GAG CCA GCC AAA ATA GAA GCT TCT CGA GCT

**HindIII**

**ψII**: AAG AGT AAA GCT GAA CCA GCC AAA ATT GGT GCT GCC CAA ACT

**ψIII**: AAG ATT AAA GCT GAG CCA GCC AAA ATT GAT GCT TCT AGA GCC

**ψIV**: AAG ATT AAA GCT GAA CCA GCC AAA ATT GAT GCT TCT CAG GTC

**cDNA**: TCA CTT TCC AAG CTA GGG GAT GTT ATC ATC AAT GAT GCT TTT

**ψII**: TCA CTT CCC AAA CTT GGG GAT GTT TAC ATC AAT GAT GCT TTT

**ψIII**: TCA CTT CCC AAA CTT GGG GAT GTT TAC ATC AAT GAT GCT TTT

**ψIV**: TCA GTG TCC AAA CTT GGG TAT GTT TAT GTC AAT GAT GCT TTT
\textbf{cDNA} : GGC ACT GCT CAC AGA GCC CAC AGC TCC ATG GTA GGA GTC AAT
\textbf{cDNA} : CTG CCA CAG AAG GCT GGT GGG TTG ATG AAG AGG GAG CTG
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : CTG GCC ATC CTG GGC GGA GCT GGT GCA GAC AAG ATC CAG
\textbf{cDNA} : CTC ATC AAT ATG CTG GAC AAA GAT GAG ATG ATT ATT
\textbf{cDNA} : TTC ATC AAT AAT ATG CTG GTC AAA GTC AAT GAG ATG ATC ATT
\textbf{cDNA} : GGT GGT GGG ATG CTT CTG GCA AAT ATG GGT GCA GAC AAG ATC AAG
\textbf{cDNA} : GAT GGT GCA ATG GCT TTT ACC TTC CTT AAG GTC GTC CAC AAC AAC
\textbf{cDNA} : GAT GGT GCA ATG GCT TTT ACC TTC CTT AAG GTC GTC CAC AAC AAC

\textbf{cDNA} : GGT GGT GCA CAC CCA ACC AAG CAC TCC ATG AGT GGT GTG GTG AGT
\textbf{cDNA} : GGT GGT GCA CAC CCA ACC AAG CAC TCC ATG AGT GGT GTG GTG AGT
\textbf{cDNA} : GGT GGT GCA CAC CCA ACC AAG CAC TCC ATG AGT GGT GTG GTG AGT
\textbf{cDNA} : GGT GGT GCA CAC CCA ACC AAG CAC TCC ATG AGT GGT GTG GTG AGT
\textbf{cDNA} : GGT GGT GCA CAC CCA ACC AAG CAC TCC ATG AGT GGT GTG GTG AGT
\textbf{cDNA} : GGT GGT GCA CAC CCA ACC AAG CAC TCC ATG AGT GGT GTG GTG AGT

166 167 168 169 170 171 172 173 174 175 176 177 178 179
Gly Thr Ala His Arg Ala His Ser Ser Met Val Gly Val Asn

\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC

180 181 182 183 184 185 186 187 188 189 190 191 192 193
Leu Pro Glu Lys Ala Gly Gly Phe Leu Met Lys Lys Glu Leu

\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC

194 195 196 197 198 199 200 201 202 203 204 205 206 207
Asn Tyr Phe Ala Lys Ala Leu Glu Ser Pro Glu Arg Pro Phe

\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC

208 209 210 211 212 213 214 215 216 217 218 219 220 221
Leu Ala Ile Leu Gly Gly Ala Lys Val Ala Asp Lys Ile Glu

\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC

222 223 224 225 226 227 228 229 230 231 232 233 234 235
Leu Ile Asn Asn Met Leu Asp Lys Val Asn Glu Met Ile Ile

\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC

236 237 238 239 240 241 242 243 244 245 246 247 248 249
Gly Gly Gly Met Ala Phe Thr Phe Leu Lys Val Leu Asn Asn

\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
\textbf{cDNA} : AAC TAC TTT CCC AAA GCT TTG GAG AGT CCT GAA TGC CCC TTC
cDNA : ATG GAG ATT GGC ACT TCT CTG TTT GAT GAA GAG GGA GCC AAG

ΨII : GTG GAG ATT GGC ACT TTT CTG CTA GAT GAA GAG AAG GGC TAC AAG
ΨIII : ATG GAG ATT GGC ACT TCT CTG TAT GAT GAA GAG GGC AAG
ΨIV : ATG GAG ATT GGC ACC TCT CTA TAT GAT AAG GAG GGG GCC AAG

264 265 266 267 268 269 270 271 272 273 274 275 276 277
Ile Val Lys Asp Leu Pro Val Asp Leu Thr Ala Asp Ala Thr Lys Phe

ΨII : ATG GTC AAA GAC CTA ATG TCC AAA GCT GAG AAT GTT GTG
ΨIII : ATG GTC AAA AAT CTC ATG GCC TAA GCA AAG AAA ATG GTT CTG
ΨIV : ATG GTC AAA GAT CTC ATG GCC AAA GCT GAG AAA ATG GTT CTG

278 279 280 281 282 283 284 285 286 287 288 289 290 291
Lys Ile Thr Leu Pro Val Asp Phe Val Thr Ala Asp Lys Phe

ΨII : AAG ATT ACC TTG CCT GTT GAC TTT GTC ACT GCT GAC AAA TTT
ΨIII : AAG ATT ACC TTG CCT GTT GAC TTT GTC ACT GCT GAC AAA TTT
ΨIV : AAG ATT ACC TTG CCT GTT GAC (______________) CAA TTT

292 293 294 295 296 297 298 299 300 301 302 303 304 305
Asp Glu Asn Ala Lys Thr Gly Glu Ala Thr Val Ala Ser Gly

ΨII : AAT GAG ATT GCC AAG ACT GCC CAA GCT ACT GTG CCT TCT GCC
ΨIII : GAT GAG ATT GCC AAA ACT GAC CAA GCT ACT GTG GCC TCT GTT
ΨIV : GAT GAG ATT GCC AAA ACT GAC CAA GCT ACT GTG GCC TCT TGT

306 307 308 309 310 311 312 313 314 315 316 317 318 319
Ile Pro Ala Gly Trp Met Gly Leu Asp Cys Gly Pro Glu Ser

ΨII : ATA CCT GCT GCC TGG ATG GGC TGG ACT GTC TCT GAA AGC
ΨIII : GTA CCT GCT GCC TGG ATG GGC TGG ACT GTC TCT GAA AGC
ΨIV : ATA CCT GCT GCC TGG ATG GGC TGG ACT GTC TCT GAA AGC

320 321 322 323 324 325 326 327 328 329 330 331 332 333
Ser Lys Lys Tyr Ala Glu Ala Val Thr Arg Ala Lys Gln Ile

ΨII : AGC AAG AAA TAT GCT GAG GCT GGG GCC CCA GCA AAG CAG ATT
ΨIII : AGC AAG AAA TAT GCT GAG GCT GTG GCT CAA GCT AAG CAG ATC
ΨIV : AGC AAG AAA TAT GCT AAG GTG ACC CGA GCT AAG CAG ATT
<table>
<thead>
<tr>
<th>cDNA</th>
<th>GGT TGG AAT CCT GTG GGG GTA TTT GAA TGG GAA GCT TTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψII</td>
<td>GTT TGG AAT GGA CCT GCT GGG GTT TTT GAT TGG GAA GCT TTT</td>
</tr>
<tr>
<td>ψIII</td>
<td>GTT TAG AAT CCT GTG GGG GTT TTT GAA GAT TTT</td>
</tr>
<tr>
<td>ψIV</td>
<td>GTT TGG AAT GGA CCT GTT GGG GTT TTT GAA TGG GAA GCT TTT</td>
</tr>
</tbody>
</table>

HindIII

<table>
<thead>
<tr>
<th>cDNA</th>
<th>GCC CGG GGA ACC AAA GCT CTC ATG GAT GAG GTG GAAGCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψII</td>
<td>GCA AGG GGA ACC AAG CTC CTC AAG GAT GAG GTG GTA AAA GCC</td>
</tr>
<tr>
<td>ψIII</td>
<td>GCA AGG GGA ACC AAG CTC CTC ATG GAT GAG GTG GTA AAG GCC</td>
</tr>
<tr>
<td>ψIV</td>
<td>GCA AGG GCC AAG CTC CTC GAT GAT GAG GTG GTA AAA GCC</td>
</tr>
</tbody>
</table>

Thr Ser Arg Gly Cys Ile Thr Ile Ile Gly Gly Gly Asp Thr

<table>
<thead>
<tr>
<th>cDNA</th>
<th>ACT TCT AGG GCC TGC ATC ACC ACT ATA GGT GGT GGA GAC ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψII</td>
<td>ACT TCT AGG GTT TGC ATC ATT ACT ATA GGT GGT GGA GAC ACT</td>
</tr>
<tr>
<td>ψIII</td>
<td>ACA TCT AGG TGC ATC ACT ATC ATA GGT GGT GGA GAC ACT</td>
</tr>
<tr>
<td>ψIV</td>
<td>ACT TCT AGA GGT TGC ATG ACT ATC ATA GGT GGT GGA GAC ACT</td>
</tr>
</tbody>
</table>

Ala Thr Cys Cys Ala Lys Trp Thr Glu Asp Lys Val Ser

<table>
<thead>
<tr>
<th>cDNA</th>
<th>GCC ACT TGC TGT GCC AAA TGG AAC ACG GAT AAA GTC AGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψII</td>
<td>GCC ACT TGC TGT GCC AAA GGG AAT ACA GAG GAT AAA GTC AGC</td>
</tr>
<tr>
<td>ψIII</td>
<td>GCC ACT TGC TAT GCC AAA TGG AAC AAT GAG GAT AAA GTC AGC</td>
</tr>
<tr>
<td>ψIV</td>
<td>GCC ACT TGC TAT GCC AAA TGG AAC AAT GAG GAT AAA GTC AGC</td>
</tr>
</tbody>
</table>

His Val Ser Thr Gly Gly Gly Ala Ser Leu Glu Leu Leu Glu

<table>
<thead>
<tr>
<th>cDNA</th>
<th>CAT GTG AGC ACT GGG GTT GCC AGC TTG GAG CTC CTG GAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψII</td>
<td>CAT GTC AGC ACT GGG GTT GAT ACC ACT CTA GAA CTC CTG GAA</td>
</tr>
<tr>
<td>ψIII</td>
<td>CAT GTC AGC ACT GGG GTT GAT GCC AGC CTA GAA TTC CTA GAA</td>
</tr>
<tr>
<td>ψIV</td>
<td>CAT GTC AGC ACT GGG G T GTT GCC AGC CTA GAG CTC CTG GAA</td>
</tr>
</tbody>
</table>

Gly Lys Val Leu Pro Gly Val Asp Ala Leu Ser Asn Ile TER

<table>
<thead>
<tr>
<th>cDNA</th>
<th>GGT AAA GTC CTT CCT GGG GTG GAT GCT CTC AGC AAT ATT TAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ψII</td>
<td>GGT AAA GTC CTT CCT GGG GTG GAT GCT CTC AAG AAT GTT TAG</td>
</tr>
<tr>
<td>ψIII</td>
<td>GGT AAA GTC CTT CTC GGG GTG GAT GCT CTC AGC AAT GTT TAG</td>
</tr>
<tr>
<td>ψIV</td>
<td>GGT AAA GTC CTT CCT GGG GTG GAT GCT CTC AGC AAT GTT TAG</td>
</tr>
</tbody>
</table>
numerous base differences, a deletion of at least fifteen bases, and premature 'in-phase' termination codons. These factors all preclude the transcription and translation of the Chinese hamster sequences into functional PGK enzymes.

PGK-II contains an in-phase termination codon (TAA) in place of Gln 37 (CAG) of the human cDNA. Consequently, if transcription of this sequence were initiated, the resulting message would code for a polypeptide that was only 36 amino acids long. There is also a deletion of 49 nucleotides, with respect to the cDNA, including codons Val 99 to Ser 114 and the first nucleotide of Val 115. In addition to removing 16 amino acids from a potential protein, this deletion also introduces a frame shift into the coding sequence. Deletion mutations have also resulted in the loss of the third nucleotide of Lys 183 and the first nucleotide of Pro 317. The frame shifts resulting from these deletions and the large deletion of 49 bases have resulted in the generation of 13 other termination codons 'in-phase' with the sequence of PGK-II before the final termination codon denoting the end of the coding sequence. There is also a second termination codon (TAA) 'in-phase' with the human PGK sequence in place of Lys 271 (AAA).

PGK-III contains an in-phase termination codon (TAG) instead of Glu 78 (GAG), so a polypeptide only 77 amino acids in length would be produced by translation of a messenger RNA product from this sequence. The next mutation likely to have severe consequences for a potential protein product is a deletion resulting in the loss of the second nucleotide of Cys 107. The consequent frame shift results in the appearance of 10 termination codons 'in-phase' with the new reading frame. There
is, in addition, another termination codon (TAG) 'in-phase' with the human cDNA in place of Trp 335 (TGG). PGK-III also contains a deletion of 60 nucleotides over the region coding for amino acids Lys 229 to Asn 249. This deletion removes 20 amino acids from the potential coding sequence, but does not result in a further shift in the reading frame.

In PGK-IV, the codon for Ser 1 (TCG) in the human cDNA has been replaced by a termination codon (TAG). Thus, with a termination codon immediately following the initiation codon, translation of a message could not occur beyond initiation. There is a second termination codon (TGA) in place of Arg 205 (CGA) of the cDNA. The first deletion mutation observed in the sequence of PGK-IV consists of a deletion of the three bases of the codon for Leu 180. This is followed closely by the insertion of a cytosine residue between the third nucleotide of Pro 181 and the first nucleotide of Gln 182. This results in a frame shift and thus a 'meaningless' reading frame in which 10 termination codons are observed. A deletion mutation has resulted in the loss of the first nucleotide of Ser 270, the reading frame thus returning to that of the human cDNA. Deletion of the third nucleotide of Thr 351 and a single base in position 2 of the Gly 395 codon create further frame shifts. Six more premature termination codons are generated in these altered reading frames. In addition to the above mutations, PGK-IV also contains a deletion of 15 nucleotides coding for Phe 285 to Asp 289 inclusive.

Examination of the human cDNA sequence at the positions equivalent to the end points of the 49 bp deletion in PGK-II reveals a direct repeat of GG(A/T)CTGT. One of these repeats remains in PGK-II. Similarly, in the human PGK gene in
positions equivalent to the ends of the 60 bp deletion in PGK-III and the 15 bp deletion in PGK-IV, there are also direct repeats and one of each is retained in their respective pseudogenes. These repeats are TGCNNACAA in PGK-III, and CTGNTGAC in PGK-IV. These features suggest that the deletions arose by homologous recombination between the repeats either in the hamster cells or during the cloning of the hamster sequences in their lambda vectors. The repeated sequences are underlined in Figure 5.6 in the human gene sequence.

From the evidence above, insertions, deletions, and nucleotide changes resulting in frame shifts, mis-sense mutations, and termination codons within the coding sequence, render the DNA sequences of all the intronless Chinese hamster PGK genes unable to code for a functional PGK protein. Thus, PGK-II, -III and -IV will henceforth be refered to as PGK-\(\psi\)II, -\(\psi\)III and -\(\psi\)IV respectively, to indicate that they are PGK pseudogenes.

5.3.2.2 5' Non-coding region.

Figure 5.7 shows the sequence of PGK-\(\psi\)II, -\(\psi\)III and -\(\psi\)IV upstream of their initiation codons, aligned with the relevant regions of the 5' non-coding sequences of the human X-linked PGK gene (Singer-Sam et al., 1984) to show maximum homology.

The Chinese hamster pseudogenes have much lower homology with the human gene than with each other, in particular over the first 48 bp, 5' to the coding region. Beyond this, PGK-\(\psi\)III and PGK-\(\psi\)IV contain deletions of 26 bp with respect to the human sequence, while the sequence of PGK-\(\psi\)II follows more closely that of the human PGK gene. Upstream of these deletions, the hamster pseudogenes contain between 34 and 38 bp of sequence
Figure 5.7

5' non-coding regions of the Chinese hamster pseudogenes.

The hamster pseudogenes are aligned with the 5' non-coding and flanking regions of the human PGK-1 gene. The proposed Sp1 binding site of the human PGK gene, consisting of an 8 bp direct repeat, is underlined. The asterisks indicate the three transcriptional start points identified for the human PGK gene.

All 'ATG' sequences appearing in the DNA sequence of the hamster pseudogenes, downstream of the proposed Sp1 binding site in the human PGK gene, are also underlined.

Base positions upstream of the translational start are indicated for the human PGK cDNA.
that is absent from the human PGK gene. (The sequence of PGK-$\psi$II has not been determined beyond this point.) After a short sequence of significant homology (6 bp), PGK-$\psi$III contains an extra 9 bp, and PGK-$\psi$IV an additional 33 bp, with respect to the human PGK sequence before the final region of homology between these sequences. This homology continues up to position 126 bp from the initiation codon of the human gene, which is 27 bp upstream of the most 5' transcriptional start point. Beyond the homologous regions, only PGK-$\psi$III of the pseudogenes has been sequenced.

In PGK-$\psi$III, the most 5' regions sequenced lack any evidence of the usual features of eukaryotic promoters that are found upstream of the transcriptional start point (see section 1.2.2.1 of "Introduction"). These features are the "TATA box", usually found between 20 and 40 bp upstream of the start of transcription, and the "CAAT box" or the GC-rich Sp1 binding site, found, in most cases, between 70 and 100 bp from this point. A possible promoter in the human PGK sequence was noted by Singer-Sam et al. (1984), who compared the PGK promoter region to that of the hypoxanthine phosphoribosyl transferase (hprt) gene (Melton et al., 1984). Both genes contain direct repeats of at least 8 bp in a GC-rich region of DNA, which form the boundaries of a highly conserved region of 30 bp to which RNA polymerase may bind. Each GC-rich repeat in the PGK gene represents potential binding sites for the transcription factor Sp1. It is interesting to note that the homology between the human PGK gene, PGK-$\psi$III and PGK-$\psi$IV ends just inside the most 3' of these direct repeats (underlined in Figure 5.7). This suggests that these two hamster pseudogenes might represent the sequence of a messenger RNA transcribed immediately from the
region of the RNA polymerase binding.

It is also worthy of note that the homology between the human PGK gene and the human X-linked pseudogene (Michelson et al., 1985a) ends at a point 3 bp downstream of two adjacent nucleotides which have been identified, by primer extension and S1 nuclease protection experiments, as transcription start points (Singer-Sam et al., 1984). A third transcription start point was also identified 5 bp downstream of this point.

As has been discussed previously, these three hamster PGK sequences each contain insertion/deletion mutations, base changes and larger deletions which preclude the synthesis of a functional PGK enzyme. However, there may also be mutations within the 5' non-coding regions of the genes which could affect the efficiency and accuracy of transcript production if transcription were to be initiated. The hamster PGK sequences each contain at least one ATG sequence between the point where homology with the human PGK gene begins and the ATG at the correct site of initiation of translation (underlined in Figure 5.7). If translation is normally initiated at the AUG sequence closest to the 5' terminus of the messenger RNA (Kozak, 1978), then a message from any of these genes would bring about premature initiation of translation resulting in an inappropriate amino acid sequence.

5.3.2.3 3' Non-coding region.

All three pseudogenes contain the same termination codon, TGA, at the end of the PGK coding sequence. The sequences downstream of this point are shown in Figure 5.8. Apart from an additional length of 13 nucleotides, the human 3' untranslated
Figure 5.8

3' non-coding regions of the Chinese hamster pseudogenes.

The hamster pseudogene sequences are aligned with the
3' non-coding region of the human PGK cDNA. The
polyadenylation signals in the human cDNA and PGK-ψIII
sequences are underlined, as is the 'poly A' tract of PGK-ψIII.

Base positions downstream of the termination codon are
indicated for the human PGK cDNA.
cDNA : GTTA AAAAGAAAG....TGAGCAGT ...................................

ψII : GGTAGAAATATAAGGGTAGCTAT.CAATTCTGCTAGTTTTGGGATAATTGTTGCT
ψIII : GT...AAAAGAAAG...TGAGCTAT ...................................

290 300 310 320

| |
| |

cDNA : .GTTAGCTTAGTTCTCTTTTCTATTTATGATGGTT.ATTATGATTAGC
ψII : AATTGTCCCCAATAGT.CAGTAAGGATATTTGCTCAATA..TTTATTAT.......C
ψIII : .AAAAGTTCAGT .... TCTCTTT.GA.GTAGCCCT....TGGTAGC

330 340 350 360 370 380

| |
| |

cDNA : TTTGTCACTGTTTCTACACTCAGCATGGAAACAAGATGAAATTCCATTTGTAGGTAG
ψII : TTTCACA ..........................................................AG
ψIII : TTTGTCACTCATT.CATGACACAGCATATATATAGATGAATTCGCTTAGGT...

390 400 410 420 430

| |
| |

cDNA : TCAGACA.AAATT.CATGATCCATTAAGTAAACAAATACAAATATGGTGCCATTGA
ψII : .CAGG.A.AAATTTCTT...CATT.A.GTAA
ψIII : TT.GG.AGAAAGT.GATGCTCTATT........AATAAAGATGTCCACTGACAAAA
sequence shares significant homology with all three pseudogenes up to a point 162 bp from the termination codon. Beyond this position, limited sequence homology is maintained along the full length of the 3' untranslated region up to the 'poly A' tail of the human cDNA. This homology is restricted to a number of 'blocks' into which similar sequences can be aligned by introducing a number of large deletions and insertions. The general sequence structure of PGK-ΨIII remains similar to that of the human cDNA, but PGK-ΨII contains a number of deletions and insertions, the largest deletions being 18, 29 and 48 bp long, and the insertions, 9 and 40 bp long. The sequence of PGK-ΨIV was only determined up to a point equivalent to 201 bp from the termination codon in the cDNA. PGK-ΨIII has been sequenced through the hexanucleotide polyadenylation signal, AATAAA, which is conserved in the pseudogene, and precedes the site of addition of the 'poly A' tail in the functional gene's message (Proudfoot & Brownlee, 1976). At the extreme of the sequenced region of PGK-ΨIII, starting 12 bp downstream of the polyadenylation signal, there appears to be an adenine-rich tract, 7 bp in length, in a position equivalent to the 'poly A' tail in the human cDNA. This, therefore, probably represents the remains of a 'poly A' tail.

5.3.3 Generation of hamster PGK pseudogenes.

The gross structure of the three hamster pseudogenes differs from that of the functional human PGK gene by the complete absence of intervening sequences. This loss is precise, and as such is unlikely to have come about as a result of random deletions in the DNA of a pre-existing PGK gene. If such
intronless pseudogenes have descended from an ancestral gene which contained introns, then it is more likely that intron loss would have occurred via a molecular mechanism comparable with RNA splicing which removed the intervening sequences exactly. The existence of such a mechanism has been proposed for a rat preproinsulin gene which lacks an intron that has been shown to exist in its ancestral preproinsulin gene (Perler et al., 1980).

A different mechanism that could result in the removal of a number of introns from a gene might involve an RNA intermediate. The PGK pseudogenes could have arisen through gene conversion or translocation between a PGK messenger RNA and a pre-existing PGK gene. This mechanism has been suggested for the loss of the intervening sequence of the mouse α-Ψ3 globin pseudogene (Vanin et al., 1980). The observation that the intervening sequences of a mouse α-globin gene are lost after one round of productive infection through RNA intermediates in a retrovirus vector (Shimotohno & Temin, 1982) suggests that a retrovirus could be involved in the formation of pseudogenes.

Another potential explanation for the existence of intronless pseudogenes is that they could conceivably be descendants of a primitive gene which had never contained intervening sequences. However, if divergence from the ancestral gene were so far back in evolutionary history, a functionless gene might now be expected to have diverged to such an extent that its sequence was not recognisably related to the functional gene. It is not, however, certain that pseudogenes are without function — possible functions for pseudogenes are discussed later in this chapter.

Another proposal for a mechanism which could generate intronless genes, and one which is widely thought to be the most
likely explanation, is that gene sequences found in RNA are incorporated into breaks in chromosomal DNA. These sequences could be represented on an mRNA molecule itself or on a DNA reverse transcript of the mRNA. Reverse transcriptase activity required for such a mechanism could have come from an endogenous retrovirus or from transient retroviral infection of germ line cells (Bernstein et al., 1983). Another possibility is that a normal DNA polymerase has some secondary activity enabling it to use RNA as a template: this activity has been demonstrated in vitro using DNA polymerase β and synthetic RNA templates (Weissbach, 1977).

Many intronless pseudogenes exhibit a number of characteristics which suggest that they may be derived from processed mRNAs. For this reason, these pseudogenes are referred to as "processed" pseudogenes (Wilde, 1986). Most processed pseudogenes are colinear with the normal cellular messenger RNAs, starting at the 'cap' site at the 5' end of the mRNA and containing 'poly A' tracts of various lengths correctly positioned relative to the polyadenylation signal at their 3' ends. Another characteristic of processed pseudogenes is the presence of direct repeats of up to 25 bp, flanking the sequences homologous to the mRNA. These repeats seem to be generated by the integration of the processed gene sequence into a staggered break in the chromosomal DNA, and the subsequent repair of the single stranded regions. The sites of integration often have A-T rich sequences, such regions being more prone to local melting, and thus breakage, of the DNA duplex. Examples of such pseudogenes include a human β-tubulin pseudogene (Wilde et al., 1982), pseudogenes of human metallothionein (Karin & Richards, 1982), human glyceraldehyde -3- phosphate
dehydrogenase (Benham et al., 1984), dihydrofolate reductase (Chen et al., 1982; Masters et al., 1983) and human triose phosphate isomerase (Brown et al., 1985). The Chinese hamster PGK pseudogenes also conform, to some extent, to this pattern of structure. The position where homology of PGK-ψIII and PGK-ψIV with the human X-linked PGK gene sequence begins is not equivalent to any of the capping sites for the human mRNA transcripts (Singer-Sam et al., 1984) (it is 29 bp upstream from the most 5' capping site – figure 5.7), but it is still possible that the pseudogenes were generated by reverse transcription of an mRNA transcript initiated at the 3' end of a possible RNA polymerase binding site. It may be that when the pseudogenes were generated, transcription of the functional PGK gene was normally initiated at this point, either immediately from the RNA polymerase binding site or with RNA polymerase binding more distantly from the transcription site, and that during the evolution of the human PGK gene the sites of initiation of transcription shifted downstream to the present positions. Alternatively, the source of the pseudogenes might be an aberrant transcript initiated upstream of the usual transcriptional start point. In this case, PGK-ψIII and PGK-ψIV would probably be derived by duplication of an original processed pseudogene, as it seems unlikely that both would be derived independently from aberrant transcripts with similar incorrect capping sites.

The presence of a polyadenylation signal and a short 'poly A' tract in PGK-ψIII in positions equivalent to the same features in the human cDNA also supports the theory that this pseudogene was generated via a messenger RNA intermediate and presumably integrated into the chromosomal DNA at the point
where the 'poly A' sequence finishes.

Further sequencing of the hamster pseudogenes would reveal whether there are 'poly A' tracts at the 3' ends of the other two pseudogenes. Similarly, it is not known if direct repeats exist at the termini of the pseudogenes, representing the points of integration of a PGK mRNA or putative processed retrotranscript into the chromosomal DNA. However, the sequence immediately upstream of the point in PGK-3III where homology with the human PGK gene finishes is A-T rich (A and T nucleotides make up 65% of the first 99 bases) - a common feature of the integration sites of processed pseudogenes.

5.3.4 Origin of the pseudogenes.

Processed pseudogenes are found throughout most, if not all, individuals of a species and are inherited with the rest of the genome. As such, they must have originally appeared in the germ line of ancestral individuals. With a mechanism of generation relying on a messenger RNA template, processed pseudogenes would be expected to be formed only from genes that are expressed in germ line cells. Such genes would include those coding for "housekeeping" enzymes, of which PGK is an example, or genes which are expressed preferentially in the germ cells. This raises an interesting possibility regarding the origin of PGK pseudogenes, in that during spermatogenesis the autosomal PGK gene (PGK-2) is transcribed. It has been shown that, in mammals, the isozyme of PGK which predominates in the testis and sperm is encoded by an autosomal locus (VandeBerg et al., 1973). More recently, Erickson et al. (1985) detected post-meiotic synthesis of PGK-2 mRNA in mouse spermatids. In fact, the rate of synthesis of PGK-2 (and thus, presumably, the rate of PGK-2
mRNA synthesis) is nearly two orders of magnitude higher than the rate of synthesis of PGK-1 in the spermatids (Kramer & Erickson, 1981). If this also occurs in the Chinese hamster, then if any of the PGK pseudogenes arose during the spermatid stage of spermatogenesis they are likely to be derived from, and related more closely to, the autosomal PGK gene. Unfortunately there is no sequence data available as yet for the autosomal PGK gene of any species, so it is not possible to determine whether the DNA sequences of any of the three hamster PGK pseudogenes are more closely related to the autosomal gene or to the X-linked PGK gene.

5.3.5 Evolutionary history of the hamster PGK genes.

The evolutionary relationships between the three PGK pseudogenes and the functional PGK gene of the Chinese hamster have been estimated by calculating the sequence divergence of that part of the coding sequence represented in the PGK exon of CHPGK-1 and the equivalent regions of each of the pseudogenes, according to the method of Perler et al. (1980). Sequence divergence is calculated in terms of nucleotide substitutions that lead to changes of amino acid (replacement site substitutions) or those that do not lead to amino acid replacements (silent site substitutions). Evolutionary divergence times are calculated from the sequence divergence by multiplying by a factor which represents the estimated rate of base substitutions. Such rates of divergence have been established for globin genes (Estratiadis et al., 1980).

As the replacement sites of processed pseudogenes are unlikely to have been under any selective pressure since their generation, whereas those of the functional gene would have been
under selective pressure due to the functional constraints placed on the enzyme, pseudogene divergence times would be estimated more accurately by calculating the percentage of silent site substitutions (Lacy & Maniatis, 1980). Apart from any requirements of the mRNA for a stable secondary structure, or specific codon requirements, silent sites in genes should not be under any selective pressure (Perler et al., 1980). Therefore, the rate of silent site substitutions calculated for the globin genes can be applied to the PGK genes.

Table 5.1 shows the calculated sequence divergence of the pseudogenes at replacement sites and at silent sites, and the predicted divergence times of the pseudogenes from the functional PGK gene (assuming that they arose from the X-linked gene (CHPGK-1) and not the autosomal PGK gene). The rate of silent site substitutions calculated for globin genes is 0.008 substitutions per one million years for two initially identical sequences (Efstratiadis et al., 1980; Perler et al., 1980). In addition, the divergence between the human PGK gene and the Chinese hamster PGK gene is shown (over the region represented in the exon of CHPGK-1). Silent site substitutions predict that PGK-ψII diverged from the functional PGK gene approximately 49 million years ago, and that PGK-ψIV diverged approximately 39 million years ago. PGK-ψIII is predicted to have diverged much more recently, approximately 20 million years ago. It must be noted that these calculations assume that the silent sites in functional genes accumulate mutations at the same rate in the equivalent sites in the pseudogenes. However, it is likely that there is selective pressure against some such changes in functional genes to account for bias in codon usage, so the
Table 5.1

Corrected percentage divergence (±100) between the coding sequences of the Chinese hamster PGK-1 gene and A) the three hamster PGK pseudogenes and, B) the human PGK-1 gene.

The values for the nucleotide changes per site over the regions of the PGK sequences equivalent to the putative Chinese hamster PGK-1 exon (nucleotides 163 to 732 of the human PGK cDNA) were calculated, and these figures corrected for multiple events, using the method of Perler et al. (1980). Divergence times were calculated assuming 0.008 substitutions per silent site per one million years (Efstratiadis et al., 1980; Perler et al., 1980).
<table>
<thead>
<tr>
<th>Gene pair</th>
<th>Replacement site divergence</th>
<th>Silent site divergence</th>
<th>Estimated divergence time (million yrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHPGK-1/PGK-ΨII</td>
<td>0.099</td>
<td>0.390</td>
<td>49</td>
</tr>
<tr>
<td>CHPGK-1/PGK-ΨIII</td>
<td>0.115</td>
<td>0.164</td>
<td>20</td>
</tr>
<tr>
<td>CHPGK-1/PGK-ΨIV</td>
<td>0.147</td>
<td>0.314</td>
<td>39</td>
</tr>
<tr>
<td>B.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHPGK-1/Hu.PGK-1</td>
<td>0.095</td>
<td>0.705</td>
<td>88</td>
</tr>
</tbody>
</table>
figures calculated for the pseudogenes should be taken as maximum divergence times. Calculations of the divergence time of the human and hamster PGK genes predicts that the event took place approximately 88 million years ago, although silent site substitutions accumulating over this time scale would be approaching saturation point and for this reason the nucleotide substitution rate would not be as reliable as over a shorter evolutionary time period (Perler et al., 1980). This divergence time correlates well with the estimated time of the adaptive radiation of the mammals 85 million years ago (Romera-Herrera et al., 1973; McKenna, 1975). Although replacement site substitutions, rather than silent site substitutions, more accurately reflect the divergence times of functional genes, calculations based on, for example, globin gene evolutionary rates (Efstratiadis et al., 1980) would not be appropriate, as genes coding for proteins with different structural constraints have different rates of replacement site nucleotide substitutions (Wilson et al., 1977; Perler et al., 1980). The results in table 5.1 also illustrate the increased relative rate of replacement site substitutions to silent site substitutions in the PGK pseudogenes compared to the functional genes.

The divergence times estimated for the pseudogenes are based on an assumption that there has been no exchange or transfer of any DNA between any of the pseudogenes or any other functional or non-functional PGK-related sequences in the genome during their evolutionary histories. If this is not the case, and there have been occurrences of gene conversion or translocation, these events would reduce the apparent sequence divergence of the genes involved (Ohta, 1984).

Such a mechanism might explain how introns can be lost from
genes without disruption of the coding sequence. For example, homologous recombination and a consequent transfer of DNA from one of the hamster pseudogenes to the functional PGK gene might have resulted in the present structure of the exon of CHPGK-1 isolated from the Chinese hamster genomic library (if the ancestral gene in fact had a structure containing introns as found in the human gene). This would only have been possible, however, if the gene conversion event had occurred before the pseudogene had accumulated any critical mutations in the region transferred to the functional gene. Such an event would be much more likely to occur if the genes involved were physically linked. Of the PGK pseudogenes isolated, only PGK-ψIII has been identified as possibly being located on the X-chromosome, and therefore potentially closely linked to the functional PGK gene. In addition, its sequence is more closely related to that of the exon of CHPGK-1 than are the sequences of PGK-ψII and PGK-ψIV. If this increased sequence homology is a result, at least in part, of an event such as gene conversion, then the resulting apparent decrease of the sequence divergence would mean that the estimated divergence time for PGK-ψIII and CHPGK-1 would be an underestimation of the actual divergence time. If comparisons of the sequence of PGK-ψIII with that of CHPGK-1 outside the region that has been isolated were to demonstrate a greater sequence divergence than that within the exon, this evidence would support the idea of gene conversion between these two genes, possibly resulting in the loss of pre-existing introns.

Alternatively, a recombination event involving an RNA intermediate or an intronless DNA sequence which has either not been isolated or has since been deleted from the genome might equally be responsible for such intron loss.
5.3.6 Pseudogene function.

It has been assumed that processed pseudogenes are transcriptionally inactive, as in general they do not contain complete promoter regions. In any case, accumulation of mutations in the genes would eventually destroy the functional capacity of any translation products. However, although expression of active enzymes is not possible in the hamster PGK pseudogenes, they could conceivably play a part in the control of the expression of a functional PGK gene. Such functions have been suggested which might result in a higher degree of conservation of pseudogene sequence (Wilde, 1986).

A possible mechanism of modifying gene expression might be through transcription of the DNA strand of a pseudogene complementary to that used in the functional gene. The resulting 'anti-transcript' could potentially form an RNA-RNA duplex with the transcript of the functional gene, preventing it from binding to the ribosome. This mechanism of transcript 'pairing' has been suggested as a model for setting thresholds where molecular gradients are responsible for the formation of discrete structures in development (McCarrey & Riggs, 1986). In these situations, pseudogenes would act as a source of intracellular inhibitors.

Evidence for this, and other possibilities, might be found by probing messenger RNA with probes specific for 'sense' and 'anti-sense' transcripts of the pseudogenes, or investigating the ability of cellular RNA to protect such probes from S1 nuclease.
5.4 Summary

The four Chinese hamster PGK DNA sequences analysed by restriction mapping and heteroduplex mapping in the previous chapter have now been further analysed by DNA sequencing. One of these sequences represents a region of 572 nucleotides of the PGK coding region with 92.2% homology with the human PGK cDNA. The DNA sequences at the 5' and 3' ends of this region are consistent with their being putative splice sites, and there exists a putative 3' splice signal upstream of the 3' splice site. Although the intron/exon arrangement in the hamster PGK sequence is very different from the human X-linked gene where four intervening sequences break up the coding region represented, the hamster PGK sequence itself does not contain any obviously critical mutations. Thus, there is every reason to believe that this hamster PGK sequence represents part of a functional PGK gene (CHPGK-1) and is not a vestigial region of PGK coding sequence or a truncated pseudogene.

The remaining three PGK sequences isolated have been shown to represent the entire coding sequence of the PGK gene. The coding regions cover the same length as that of the human cDNA showing that, at least at the time the pseudogenes were generated, the hamster PGK gene, like that of the human, coded for a protein 416 amino acids long. The coding regions of all three pseudogenes contain deletion and transition mutations, one contains a single base insertion, and all contain a large deletion of at least 15 nucleotides (although the latter may have arisen during the cloning and manipulation of the isolated hamster sequences). These mutations would render any transcript
produced unable to code for a functional enzyme, due to the presence of premature termination codons and frame shifts resulting in 'meaningless' amino acid coding sequences. The 5' regions of two pseudogenes (PGK-ψIII and PGK-ψIV) have been sequenced, and have been shown to be homologous to the human X-linked PGK gene up to a point upstream of the present-day transcriptional start points of the human PGK gene but downstream of the region postulated to be involved in the binding of RNA polymerase. In PGK-ψII and PGK-ψIII the 3' regions contain homology with the human cDNA up to the putative polyadenylation signal, and the presence of a short 'poly A' tract has been demonstrated in PGK-ψIII. It is not known whether any 'poly A' tracts exist in PGK-ψII and PGK-ψIV.

It is probable that the pseudogenes arose independently in the germ line cells of an ancestral individual by the reverse transcription of a processed messenger RNA molecule. The estimated times of these events, assuming that they arose from the X-linked PGK gene are, at most, 20, 39, and 49 million years ago (for PGK-ψIII, -ψIV, and -ψII, respectively). Comparison of the sequence of the hamster PGK exon with the human cDNA, to determine sequence divergence, suggested that they diverged from their common ancestral PGK gene approximately 88 million years ago.

It is possible that the difference in intron/exon structure between the X-linked PGK genes of the human and the Chinese hamster is due to recombination events between the functional hamster gene and a pseudogene, resulting in removal of introns from the hamster sequence (assuming that the ancestral gene contained these introns). However, further evidence for such an event, and for the presence of other features of the pseudogenes
which might confirm them as processed pseudogenes derived from a PGK mRNA, would be revealed by the isolation and analysis of the remainder of the Chinese hamster functional X-linked PGK gene, and further sequencing of the PGK pseudogenes.
This project has involved an analysis of the Chinese hamster PGK gene family: some of the PGK sequences that constitute this group have been isolated and their DNA sequences determined. A transfection system for the study of PGK gene expression in vivo has also been developed.

PGK deficient R1.1.7 cells were used as recipients for CaPO$_4$-DNA co-precipitate mediated gene transfer. Successful transformation to G418 resistance, by the introduction of an aminoglycosyl-3′-phosphotransferase gene into the cells, was consistently achieved and this system was used to optimise transfection conditions. Subsequently, a human PGK cDNA was successfully used to transform R1.1.7 cells to a PGK-positive phenotype. It should now be possible to use this in vivo expression system to study the expression of any introduced PGK gene or GPI gene. It would prove particularly useful for the investigation of transcription and processing (by mRNA assay), and translation (by enzyme or immunological assay) of cloned genes modified by in vitro mutagenesis, where the effects of specific nucleotide sequence changes on the accuracy and efficiency of gene expression could be determined. In addition, cloned PGK genes from existing human PGK variants could be examined, using this system, to learn more about the nature of their deficiencies.

R1.1.7, itself, has been shown not to contain any rearrangement of DNA involving the PGK sequences (observable by Southern transfer analysis), and it is hoped that the precise
cause of its PGK deficiency will, at some point, be determined.

Four different PGK sequences have been isolated from a Chinese hamster DNA library. One of these has been identified as a putative exon belonging to a functional PGK gene, and was shown to be X-linked (CHPGK-1). This exon contains no known critical mutations, its terminal sequences are consistent with being splice junctions, and a putative 3' splice signal has been identified upstream of the 3' splice site. Although the intron/exon structure of the human X-linked PGK gene is very different from the hamster putative exon over this section of coding region, introns, perhaps originally present in an ancestral gene, could conceivably have been removed from the hamster PGK gene by homologous recombination with a pseudogene or another mRNA related sequence.

The three remaining PGK sequences have been shown, by the presence of critical mutations in their coding regions, to represent pseudogenes (PGK-ψII, -ψIII, and -ψIV). One of these, PGK-ψIII, appears to be X-linked. All three were shown, by heteroduplex mapping and DNA sequencing, to be intronless pseudogenes. The lack of intervening sequences, combined with (in the cases of PGK-ψIII and -ψIV) the presence of homology with the human PGK gene only as far as a possible RNA polymerase binding region, suggests that these may be "processed" pseudogenes which are derived from processed messenger RNA molecules in the germ cells of ancestral individuals. PGK-ψII and -ψIII have been sequenced as far as their putative polyadenylation sites, and PGK-ψIII has been shown to contain a 'poly A' tract downstream from this point. This again points to processed messenger RNA molecules as intermediates in the formation of these pseudogenes. Evolutionary divergence times
for the pseudogenes, from an ancestral gene common to the X-linked Chinese hamster PGK gene, have been estimated by calculating sequence divergence in terms of 'silent' site substitutions. These estimates are put at 20, 39 and 49 million years ago and probably represent maximum values.

Bearing in mind that other mammals examined thus far have been shown to express two different PGK genes, one (X-linked PGK-1) in all somatic cells and the second (autosomal PGK-2) only in post-meiotic sperm cells, it is reasonable to assume that this is likely to be the case in the Chinese hamster. This being so, the Chinese hamster PGK gene family must consist of at least five independent PGK gene sequences - two functional genes and three pseudogenes. In addition, this work suggests that there are probably no more than this number of PGK sequences in the gene family: Of the fragments observed to hybridise to the human PGK cDNA probe in digests of Chinese hamster DNA, at least half have been accounted for in the recombinant clones that have been isolated. Functional genes containing intervening sequences are likely to be represented by more than one of these hybridising bands, particularly where the fragments are of lower molecular weight. However, the situation may well not be as simple as this: It is possible that, if the genomic digests and DNA library were probed at lower stringency, hybridisation to other fragments would be seen. Such additional bands may represent smaller exons of the functional genes. Other PGK-related sequences which might be observed may represent pseudogenes which have diverged further from the sequences of the functional genes. If pseudogene sequences are not under any selective pressure, they should evolve at a rate much greater than that at which functional genes evolve, and it is possible
that there exists a range of PGK pseudogenes in the Chinese hamster genome which have diverged considerably from the functional PGK gene, eventually fading into the background of genomic DNA. The isolation of further sequences from a hamster genomic library would contribute to the answering of these questions.

This work has produced some detailed information regarding the Chinese hamster PGK gene family: Most of the data has been generated by the sequencing of three intronless pseudogenes. As more pseudogenes are isolated from a range of organisms, questions as to their role, if any, in cell development, gene expression, or other cellular functions, will be addressed, and more will be learned about the mechanisms involved in their generation and evolution.

It has been of most interest, however, to relate this work with that carried out, by a number of investigators, on the human PGK gene family, and comparing the putative Chinese hamster exon with the same region of the human PGK-1 gene. In doing so, it has been demonstrated clearly that whilst the DNA of an organism undergoes substantial rearrangement or mutation during its evolution (as shown by the completely unrelated patterns of DNA restriction fragments which contain PGK sequences in each species, and also by the differences in exon arrangement between the human and hamster PGK genes), the coding sequences, particularly in the case of PGK, are very strictly maintained by the cells replicative machinery over millions of generations of the organism and over many millions of years of evolution.
REFERENCES


Fertility and Sterility 27, 699-701.


Nature 316, 774-778.

Trends in Genetics 2, 196-197.


J. Biol. Chem. 233, 551-558.


J. Biol. Chem. 255, 6421-6423.


Virology 52, 456-467.


Matthes, H.W.D., Zenke, M., Gundström, T., Staub, A.,
synthesis of over one hundred oligonucleotides on a microscale.
EMBO J. 3, 801-805.


with base specific chemical cleavages.
Methods enzymol. 65, 499-560.

as a mechanism for threshold setting in development: A possible
function for pseudogenes.

McClure, M.R. (1985) Mechanism and control of transcription in

McKenna, M.C. (1975) In: Luckett, W.P. & Szalay, F.S. (Eds)

McKnight, G.S. & Palmiter, R.D. (1979) Transcriptional
regulation of the ovalbumin and conalbumin genes by steroid

Melnykovych, G. & Bishop, C.F. (1972) Utilisation of hexose and
synthesis of glycogen in two strains of HeLa cells.
In vitro. 7, 397-405.


The Chinese hamster phosphoglycerate kinase gene family

by Stephen J. Rawson.

Abstract.

I. A phosphoglycerate kinase (PGK) deficient variant cell line, R1.1.7, derived from the Chinese hamster ovary cell line, CH0-K1, is used as a recipient in the development of a DNA-mediated transfection system designed for the study of in vivo expression of exogenous PGK gene sequences.

II. Several PGK DNA sequences are detected in the Chinese hamster genome by hybridisation of genomic DNA digests with a human PGK cDNA probe. A number of these sequences are shown to be X-linked. Four of the PGK sequences observed in the blot hybridisations are isolated from a CH0-K1 DNA genomic library and analysed by probing with different regions of the PGK cDNA, heteroduplex mapping and DNA sequencing. One of these sequences is a 572 bp exon from the functional X-linked PGK gene (PGK-1) which is expressed in all somatic cells. The three remaining PGK sequences are intronless pseudogenes which were apparently derived independently, from an ancestral Chinese hamster PGK gene, within the last 50 million years. They exhibit features typical of 'processed' pseudogenes which are generated via an mRNA intermediate and integrated into breaks in the chromosomal DNA.