STRUCTURE AND EXPRESSION OF THE MOUSE DNA LIGASE I GENE

Joanne Kirsty Jessop

A Thesis Presented for the Degree of Ph.D.
Institute of Cell and Molecular Biology
The University of Edinburgh
February 1995
DECLARATION

The composition of this thesis and the work presented in it are my own, unless otherwise stated. The experiments were designed in collaboration with my supervisor Dr. David W. Melton.

Joanne Kirsty Jessop
February 1995
# Contents

- **Abstract**
- **Acknowledgements**
- **Abbreviations**

## Chapter 1 Introduction

1.1 Foreword

1.2 DNA Repair

- 1.2.1 Direct Reversal of DNA Damage
- 1.2.2 Base Excision Repair
- 1.2.3 Nucleotide Excision Repair
- 1.2.4 Post-Replication Repair

1.3 Human Diseases Arising From Defective DNA Repair

- 1.3.1 Xeroderma Pigmentosum
- 1.3.2 Cockaynes Syndrome
- 1.3.3 Ataxia Telangiectasia
- 1.3.4 Fanconi Anaemia
- 1.3.5 Blooms Syndrome
- 1.3.6 The DNA Ligase I-Deficient Cell Line 46BR

1.4 Mammalian DNA Ligases

- 1.4.1 DNA Ligase I
  - 1.4.1.1 Substrate Specificity of DNA Ligase I
  - 1.4.1.2 The Human DNA Ligase I Gene
  - 1.4.1.3 DNA Ligase I Structure
  - 1.4.1.4 Regulation of DNA Ligase I Activity
  - 1.4.1.5 Intracellular Location and Function of DNA Ligase I
- 1.4.2 DNA Ligase II
- 1.4.3 DNA Ligase III

1.5 Gene Targeting and the Creation of Mouse Models for Human Genetic Disease

1.6 Project Aim

## Chapter 2 Materials and Methods

2A Materials

- 2A.1 Suppliers of Laboratory Reagents
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A.1.1 Restriction Endonucleases and other DNA/RNA Modifying Enzymes</td>
<td>48</td>
</tr>
<tr>
<td>2A.1.2 Standard Laboratory Reagents</td>
<td>48</td>
</tr>
<tr>
<td>2A.1.3 Bacterial Media Reagents</td>
<td>48</td>
</tr>
<tr>
<td>2A.1.4 Mammalian Cell Culture Reagents</td>
<td>48</td>
</tr>
<tr>
<td>2A.1.5 Radioactive Reagents</td>
<td>48</td>
</tr>
<tr>
<td>2A.1.6 Antibiotics</td>
<td>48</td>
</tr>
<tr>
<td>2A.1.7 Antisera</td>
<td>49</td>
</tr>
<tr>
<td>2A.1.8 Human DNA Ligase I cDNA</td>
<td>49</td>
</tr>
<tr>
<td>2A.1.9 Oligonucleotides</td>
<td>49</td>
</tr>
<tr>
<td>2A.2 Media</td>
<td>50</td>
</tr>
<tr>
<td>2A.2.1 Bacterial Media</td>
<td>50</td>
</tr>
<tr>
<td>2A.2.2 Mammalian Tissue Culture Media</td>
<td>51</td>
</tr>
<tr>
<td>2A.3 Bacterial Strains</td>
<td>52</td>
</tr>
<tr>
<td>2A.4 Mammalian Cell Lines</td>
<td>53</td>
</tr>
<tr>
<td>2A.5 Plasmids and Cloning Vectors</td>
<td>53</td>
</tr>
<tr>
<td>2A.6 Buffers</td>
<td>54</td>
</tr>
<tr>
<td>2B Methods</td>
<td>55</td>
</tr>
<tr>
<td>2B.1 Bacterial Culture</td>
<td>55</td>
</tr>
<tr>
<td>2B.1.1 Growth of <em>Escherichia coli</em></td>
<td>55</td>
</tr>
<tr>
<td>2B.1.2 Storage of <em>E. coli</em></td>
<td>55</td>
</tr>
<tr>
<td>2B.1.3 Transformation of <em>E. coli</em></td>
<td>55</td>
</tr>
<tr>
<td>2B.2 Nucleic Acid Isolation</td>
<td>56</td>
</tr>
<tr>
<td>2B.2.1 Small Scale Preparation of Plasmid DNA</td>
<td>56</td>
</tr>
<tr>
<td>2B.2.2 Large Scale Preparation of Plasmid DNA</td>
<td>56</td>
</tr>
<tr>
<td>2B.2.3 Preparation of Genomic DNA from Mammalian Cells</td>
<td>57</td>
</tr>
<tr>
<td>2B.2.4 Preparation of RNA from Mammalian Cells</td>
<td>57</td>
</tr>
<tr>
<td>2B.2.5 Small-Scale Preparation of Bacteriophage M13 Replicative Form DNA</td>
<td>58</td>
</tr>
<tr>
<td>2B.2.6 Preparation of Single-Stranded Bacteriophage M13 DNA</td>
<td>59</td>
</tr>
<tr>
<td>2B.2.7 Preparation of Bacteriophage λ DNA</td>
<td>59</td>
</tr>
<tr>
<td>2B.3 Quantification of Nucleic Acids</td>
<td>60</td>
</tr>
<tr>
<td>2B.3.1 Estimation of DNA Concentration</td>
<td>60</td>
</tr>
<tr>
<td>2B.3.2 Estimation of RNA Concentration</td>
<td>60</td>
</tr>
</tbody>
</table>
2B.4 DNA Manipulation
2B.4.1 Purification
2B.4.2 Restriction of DNA with Endonucleases
2B.4.3 Dephosphorylation
2B.4.4 Filling-In Recessed 3’ DNA Termini
2B.4.5 Ligation
2B.4.6 Addition of Linkers

2B.5 Electrophoresis of Nucleic Acids
2B.5.1 Electrophoresis of DNA in Agarose Gels
2B.5.2 Electrophoresis of RNA in Agarose Gels
2B.5.3 Recovery of DNA from Agarose Gels

2B.6 Transfer of Nucleic Acids from Agarose Gels to Membranes
2B.6.1 DNA Transfer
2B.6.2 RNA Transfer

2B.7 Nucleic Acid Hybridisation
2B.7.1 Labelling DNA by Random Priming with Hexadeoxy-ribonucleotide Primers
2B.7.2 Separation of Unincorporated Nucleotides from Labelled DNA
2B.7.3 Hybridisation
2B.7.4 Autoradiography
2B.7.5 Stripping Probes from Filters

2B.8 cDNA and Genomic Libraries
2B.8.1 cDNA Synthesis
2B.8.2 Preparation of Packaging Extracts for in vitro Packaging of Bacteriophage λ
2B.8.3 Library Construction
2B.8.4 Plating out Bacteriophage λ to generate Plaques
2B.8.5 Probing Libraries

2B.9 DNA Sequencing
2B.9.1 Templates for Sequencing
2B.9.2 Annealing Primer to Template DNA
2B.9.3 Sequencing
2B.9.4 Electrophoresis of Sequenced DNA

2B.10 Amplification of DNA by the Polymerase Chain Reaction
2B.10.1 Polymerase Chain Reaction Experiments 70
2B.10.2 Screening Colonies Using the Polymerase Chain Reaction 71

2B.11 Cell Culture 71
2B.11.1 Growth of Mammalian Cells 71
2B.11.2 Electroporation 71
2B.11.3 Calcium Phosphate/DNA Precipitation and Colony-Forming Assay 71
2B.11.4 Ethyl Methane Sulphonate Survival Assay 72
2B.11.5 3-Aminobenzamide Survival Assay 72

2B.12 Protein Methods 72
2B.12.1 Expression of Proteins in E. coli 72
2B.12.2 Preparation of Protein Extracts from Mammalian Cells 73
2B.12.3 Electrophoresis of Proteins 73
2B.12.4 Transfer of Proteins From Polyacrylamide Gels to Membranes 74
2B.12.5 Probing Protein Blots with Antibodies 74

Chapter 3 Phenotypic Correction of the Human Cell Line 46BR 75
3.1 Construction of a Vector Expressing Human DNA Ligase I cDNA 76
3.2 Transfection 79
3.3 Detection of DNA Ligase I Transfectants 79
3.4 Expression of Wild-Type Human DNA Ligase I cDNA 82
3.5 Correction of Ethyl Methane Sulphonate Sensitivity in 46BR 85
3.6 Correction of 3-Aminobenzamide Sensitivity in 46BR 88

Chapter 4 Mouse DNA Ligase I cDNA 91
4.1 Cloning of Mouse DNA Ligase I cDNA 92
4.2 The Ligase Clones Encode Mouse DNA Ligase I 92
4.3 Restriction Analysis of Mouse DNA Ligase I cDNA 97
4.4 Sequencing of Mouse DNA Ligase I cDNA 97

Chapter 5 Expression of Mouse DNA Ligase I cDNA 113
5.1 Introduction of a Marker into Mouse DNA Ligase I 114
5.2 Expression of Flag-Ligase in E. coli 119
5.3 Expression of Flag-Ligase in 46BR Under Control of the
Chapter 5 Phosphoglycerate Kinase Promoter
5.4 Expression of Flag-Ligase in 46BR Under Control of the Human β-Actin Promoter 124

Chapter 6 Mouse DNA Ligase I Gene Structure 138
6.1 Southern Hybridisation Analysis 139
6.2 Mouse DNA Ligase I Gene Map 155

Chapter 7 Cloning of a Genomic Fragment Containing the Arginine to Tryptophan Mutation Site 163
7.1 Cloning of Mouse DNA Ligase I Genomic Fragments 164
7.2 Restriction Mapping and Intron/Exon Structure of the 13.5kb EcoRI Genomic Fragment 167
7.3 Identification of the Arginine to Tryptophan Mutation Site in the 13.5kb Mouse EcoRI Genomic Fragment 177

Chapter 8 Gene Targeting at the Mouse DNA Ligase I Locus 182
8.1 Targeting Vector Construction 184
8.2 Targeting the Mouse DNA Ligase I Gene 188
8.3 Mouse Models for DNA Ligase I Deficiency 199

Chapter 9 Discussion 201

References 211
ABSTRACT

46BR is a cell line derived from a patient suffering from immunodeficiency, growth retardation and sunlight sensitivity. DNA replication and repair is affected in 46BR cells which are sensitive to the alkylating agent ethylmethane sulphonate and to 3-aminobenzamide. Aberrant DNA ligase I activity has been demonstrated in these cells, and mutations have been identified in both alleles of the DNA ligase I gene. The first mutation specifies a change of glutamate 566 to lysine, a charge alteration which renders the enzyme completely inactive and which could be lethal in a homozygote. The second mutation is a change of arginine 771 to tryptophan which leaves the enzyme with some residual activity. The transformed derivative of 46BR is either hemizygous or homozygous for this second mutation, so cells expressing only this mutant form of DNA ligase I are still viable. This mutation is therefore a good candidate for introduction into the mouse gene in order to create a mouse model for DNA ligase I deficiency.

Prior to making a mouse model it was necessary to show that the mutations at the DNA ligase I locus in 46BR give rise to the aberrant cellular phenotype. To do this, wild-type human DNA ligase I cDNA was introduced into 46BR in an expression vector. Selection for transformants was provided by a bacterial neomycin phosphotransferase gene. Transformants were assayed for survival in ethylmethane sulphonate and 3-aminobenzamide, which demonstrated a correlation between the presence of wild-type DNA ligase I sequences within the cells and rescue of sensitivity to these agents. Wild-type human DNA ligase I cDNA therefore complements the defect in 46BR.

A 3.2kb mouse DNA ligase I cDNA clone was obtained by probing a mouse embryonic stem cell cDNA library with full-length human DNA ligase I cDNA. The cDNA clone is almost complete, but lacks sequence from the 3' end, including the polyadenylation signal. At the amino acid level, homology between human and mouse DNA ligase I is extremely high, with most sequence variation in the N-terminal regulatory domain. In the C-terminal catalytic domain the amino acid sequences are virtually the same, with identical amino acids in the same positions at which the two mutations were identified in 46BR.

The mouse cDNA clone was used to analyse the mouse DNA ligase I gene structure by hybridisation to mouse genomic DNA. A discrete number of hybridising fragments were observed, with their sizes showing that the mouse gene which occupies 39-42kb of DNA is smaller than the 53kb human gene. A minimum of
eleven exons were identified, compared to 28 in the human gene. However no results were obtained incompatible with both genes having the same exon structure.

A 13.5kb EcoRI genomic fragment of the mouse gene containing the last five exons was identified as including sequence equivalent to the site of the Arg771 to Trp mutation in 46BR. This fragment was cloned from a mouse genomic library by probing with the mouse cDNA. Primers designed from the ends of human exon 23 which encodes Arg771 were used to show that the eighth mouse exon contains corresponding sequence. The same primers were used to sequence this exon, revealing an arginine residue at the same position as Arg771 in the human sequence.

The 13.5kb EcoRI genomic fragment has been used in gene targeting experiments to remove the last five exons from the mouse DNA ligase I gene in mouse embryonic stem cells. The next step is to reintroduce these exons, but with a mutation specifying an arginine to tryptophan alteration in the eighth exon at a position equivalent to human Arg771, and eventually to use the resulting cells to create a mouse model for DNA ligase I deficiency.
ACKNOWLEDGEMENTS

During the course of my PhD a great number of people have given me endless help and encouragement. I would particularly like to thank my supervisor Dr. David Melton for his helpfulness, support and advice, and Nik Somia for his invaluable assistance at the start of this project. I am also extremely grateful to all the other members of the group at the cutting edge; Thomas, Jim, Patrick, Wu, Carolanne, Ann-Marie, Angela, Sabine, Rebecca, Richard, Tracy and Niki for their patience, wit, charm, sense of humour, and of course, for their readiness to share their wealth of knowledge and technical ability. Thanks are also due to Darren for agreeing to continue this work. Other people also deserve recognition; Jean and Joan for the provision of clean glassware, and Frank and Graham for their excellent photographic work without which I would have been sunk.

Finally, thanks must go to my family and friends who have supported and encouraged me throughout the course of my work. I am especially grateful to my parents whose complete ignorance of matters scientific has always provided me with a refreshing change in times of stress, to Helen for suffering the whole experience of 'writing up' along with me, and last but certainly not least, to my husband Mike for providing me with a cast-iron deadline for finishing!
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adenosine</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>AMP</td>
<td>adenosine monophosphate</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>AP</td>
<td>apurinic/apyrimidinic</td>
</tr>
<tr>
<td>Arg</td>
<td>arginine</td>
</tr>
<tr>
<td>AT</td>
<td>ataxia telangiectasia</td>
</tr>
<tr>
<td>BAP</td>
<td>bacterial alkaline phosphatase</td>
</tr>
<tr>
<td>bp</td>
<td>base pair(s)</td>
</tr>
<tr>
<td>BS</td>
<td>Blooms syndrome</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>C</td>
<td>cytidine</td>
</tr>
<tr>
<td>C-</td>
<td>carboxyl-</td>
</tr>
<tr>
<td>cDNA</td>
<td>DNA complementary to RNA</td>
</tr>
<tr>
<td>CS</td>
<td>Cockaynes syndrome</td>
</tr>
<tr>
<td>d</td>
<td>deoxyribo</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>EMS</td>
<td>ethylmethane sulphonate</td>
</tr>
<tr>
<td>ENU</td>
<td>N-ethyl-N-nitrosoare</td>
</tr>
<tr>
<td>ERCC</td>
<td>excision repair cross-complementing</td>
</tr>
<tr>
<td>ES cells</td>
<td>embryonic stem cells</td>
</tr>
<tr>
<td>FA</td>
<td>Fanconi anaemia</td>
</tr>
<tr>
<td>FACC</td>
<td>Fanconi anaemia cross-complementing</td>
</tr>
<tr>
<td>G</td>
<td>guanosine</td>
</tr>
<tr>
<td>Glu</td>
<td>glutamate</td>
</tr>
<tr>
<td>G418</td>
<td>Geneticin</td>
</tr>
<tr>
<td>HPRT</td>
<td>hypoxanthine -guanine phosphoribosyl transferase</td>
</tr>
<tr>
<td>HSV</td>
<td>Herpes simplex virus</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>IVS</td>
<td>intervening sequence</td>
</tr>
<tr>
<td>LIG</td>
<td>ligase</td>
</tr>
<tr>
<td>Lys</td>
<td>lysine</td>
</tr>
<tr>
<td>MLTV</td>
<td>mouse ligase targeting vector</td>
</tr>
<tr>
<td>MNNG</td>
<td>methylnitro-nitrosoguanidine</td>
</tr>
<tr>
<td>MNU</td>
<td>N-methyl-N-nitrosoare</td>
</tr>
</tbody>
</table>
N any nucleoside
N- amino-
NAD nicotinamide adenine dinucleotide
NADH reduced form of NAD
nt nucleotide(s) (number of sequence)
OD optical density
p plasmid or chromosome short arm
SCE sister chromatid exchange
SDS-PAGE sodium dodecyl sulphate polyacrylamide gel electrophoresis
Ser serine
SV40 Simian virus 40
PCR polymerase chain reaction
PEG polyethylene glycol
pfu plaque-forming unit(s)
P_i inorganic phosphate
PMSF phenylmethylsulphonyl fluoride
PP_i inorganic pyrophosphate
q chromosome long arm
r ribo
r (superscript) resistance/resistant
RF replicative form
RNase ribonuclease
S sedimentation constant
ss single-stranded
T Thymidine
tk thymidine kinase
Trp tryptophan
UTR untranslated region
UV ultraviolet
X-Gal 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
XP xeroderma pigmentosum
3AB 3-aminobenzamide
'
(prime) denotes a gene truncated at the indicated side

single letter amino acid abbreviations:
A alanine
C cysteine
D  aspartate
E  glutamate
F  phenylalanine
G  glycine
H  histidine
I  isoleucine
K  lysine
L  leucine
M  methionine
N  asparagine
P  proline
Q  glutamine
R  arginine
S  serine
T  threonine
W  tryptophan
Y  tyrosine
1.1 Foreword

Our genetic material is under constant assault from the environment and from processes occurring within the cells themselves. Damage to DNA may arise during DNA replication and recombination, as a result of spontaneous DNA hydrolysis under physiological conditions and from exposure to radiation and other DNA damaging agents. It is essential that the genetic material be maintained in an informationally active form and so cells have evolved a number of mechanisms by which to repair DNA damage. DNA repair pathways appear to be conserved between bacterial and mammalian cells, therefore the better understood prokaryotic systems have provided a model for investigations in mammalian cells. Work is ongoing to elucidate the various steps involved in mammalian DNA repair mechanisms and this work has been greatly facilitated by the use of cell lines mutant in various aspects of DNA repair. Some of these cell lines are derived from humans with hereditary diseases arising from specific DNA repair defects. In most cases, different cell lines isolated from patients suffering from the same disease fall into a number of complementation groups. This genetic heterogeneity reflects the complex nature of DNA repair mechanisms and results from the involvement of a number of enzymes in each pathway. Eventual identification of the defective loci in the disease cell lines should aid understanding of the many steps involved in mammalian DNA repair pathways. It would then be possible to correlate the genetic defects with the observed clinical and cellular abnormalities of the disease. Once a particular mutation at a specific locus has been characterised, it can then be introduced into the equivalent mouse gene by gene targeting and used to create a mouse model for the disease. This would allow further investigations into the effects of that mutation.

1.2: DNA Repair

DNA can become altered in a number of ways which affect the fidelity of replication and introduce mutations into genes, as well as affecting normal cellular processes of DNA metabolism (Friedberg 1985, and references therein). Figure 1.1 shows some examples of DNA damage. Base mismatches may be introduced during DNA synthesis as a result of DNA polymerase replicative infidelity, or because base-pairing properties may have been altered by a number of mechanisms. For example, bases very occasionally undergo tautomeric shifts which allow anomalous base-pairing, resulting in base misincorporation into the daughter DNA strand during DNA synthesis. Abnormal bases may also be incorporated or may arise from spontaneous deamination, which most frequently results in the production of uracil from cytosine. In addition, bases may be lost completely by direct cleavage of the
Abasic sites result from depurination and depyrimidination which may occur spontaneously or from the action of DNA glycosylases. Pyrimidine dimers, both cyclobutane dimers and 6-4 photoproducts, are the major types of damage produced by ultraviolet irradiation. Monofunctional agents can react with DNA to produce bulky adducts. Cross-links, both intra- and inter-strand, arise from bifunctional agents which can also cross-link proteins to DNA (not shown). DNA strand breaks may be single or double (not shown). Abnormal or mispaired bases may also be present, either as a result of misincorporation or from deamination or other DNA damage.
c) depurination/depurination
depyrimidination pyrimidine
dimer bulky adduct interstrand
crosslink

<table>
<thead>
<tr>
<th>T</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>T</th>
<th>A</th>
<th>G</th>
<th>T</th>
<th>G</th>
<th>T</th>
<th>C</th>
<th>C</th>
<th>G</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>U</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>C</td>
</tr>
</tbody>
</table>

abnormal strand break 6-4 photoproduct intrastrand crosslink
glycosyl bond between the base and the deoxyribose moiety, a process known as depurination or depyrimidination. Damage arising from alkylation can affect the base-pairing properties of DNA and may also affect DNA metabolism by formation of bulky adducts and intra- or inter-strand cross-links, thus blocking the path of enzymes and preventing DNA strand separation. Ultra-violet light is another cross-linking agent, causing the formation of pyrimidine dimers (cyclobutane dimers and 6-4 photoproducts). Additionally, there are clastogenic factors, such as ionising radiation, which produce DNA strand breaks.

Cells have evolved a number of mechanisms by which such damage to the genetic material can be repaired (Grossman 1981, Lindahl 1982, Friedberg 1985, Lindahl and Wood 1989, and references therein). These mechanisms are described briefly in this chapter.

1.2.1: Direct Reversal of DNA Damage

Mechanisms of direct repair involve a single enzyme which catalyses a single reaction to return the DNA to its original state. The simplest example of direct reversal is ligation of DNA strand breaks which may be caused by agents such as ionising radiation. Other examples of direct reversal are photolysis, transmethylation, and direct insertion of purines.

Irradiation with ultra-violet light at wavelengths close to the absorption maximum of DNA causes cross-linking, mainly intrastrand, by formation of pyrimidine dimers. Such intrastrand cross-links are repaired by enzymatic photoreactivation. The enzyme photolyase binds to pyrimidine dimers and, using visible light at wavelengths between 300 and 600nm, breaks the covalent bond between bases and thus promotes monomerisation.

Transmethylation is the removal of alkyl groups transferred to DNA by alkylating agents such as methyl nitro-nitrosoguanidine (MNNG) and ethyl methanesulphonate (EMS). A major reaction product of alkylating agents, O⁶-methylguanine, is thought to be highly mutagenic and is removed by O⁶-methylguanine methyl transferase. A methyl group is transferred from the DNA to a cysteine residue on the enzyme, forming S-methyl cysteine and resulting in permanent inactivation of the enzyme. The enzyme therefore catalyses a reaction with suicide kinetics. Other alkyl groups such as ethyl, propyl and butyl moieties are also removed from DNA by this enzyme.

Depurination is the loss of purine bases from DNA, resulting in the formation of apurinic sites. These sites are repaired by a number of mechanisms, of which one
is an example of direct reversal of DNA damage. An enzyme called DNA purine insertase catalyses reinsertion of appropriate purine bases into apurinic sites.

1.2.2: Base Excision Repair

Base damage is recognised by a DNA glycosylase which hydrolyses the glycosylic bond, releasing free modified base and generating an apurinic or apyrimidinic (AP) site. DNA glycosylases are highly specific, for example, 3-methyladenine-DNA glycosylase removes only 3-methyladenine. AP sites are recognised by AP endonucleases which hydrolyse the phosphodiester bond 5' of the AP site and then remove a short patch of DNA. The gap is subsequently filled in by a DNA polymerase and ligated (figure 1.2).

1.2.3: Nucleotide Excision Repair

In this case DNA damage is recognised by local perturbation of the DNA double helix. An incision is then made on either side of the lesion and the gap is filled in by a DNA polymerase and ligated (figure 1.3). This repair pathway can correct several types of lesion, including alkylation, cross-links and dimers. Mammalian cells show preferential excision repair of transcribed DNA, with damage in untranscribed regions being repaired much more slowly.

1.2.4: Post-Replication Repair

In culture, cells halt DNA synthesis following exposure to DNA damaging agents. This is presumably because, if damage persists into DNA replication, it could potentially be highly mutagenic. For example some types of damage which alter the base-pairing properties of DNA, would result in base misincorporations. Also, if unrepaired, some types of damage can interfere with the process of DNA replication itself. Lesions such as pyrimidine dimers and DNA adducts block the path of DNA polymerases, thus interfering with DNA synthesis. There are several possible methods by which cells can tolerate such damage and allow DNA replication to continue.

In *Escherichia coli* there is evidence for two types of damage tolerance mechanism. In the first case DNA synthesis halts at a lesion and then continues on the other side leaving a gap which is then repaired by recombination. Alternatively, translesion synthesis may occur, where a randomly selected base is inserted opposite the lesion allowing DNA synthesis to continue uninterrupted. There is a 75% probability that an incorrect base will be inserted, therefore this type of repair is referred to as 'error-prone' and is mutagenic.
Figure 1.2: Base excision repair (adapted from Thompson 1989). A damaged base is recognised by a specific DNA glycosylase which hydrolyses the glycosyl bond and generates an abasic (AP) site. The AP site is recognised by an AP endonuclease which makes a strand incision by hydrolysing an adjacent phosphodiester bond. A short patch of DNA is then excised and the gap is filled in by a DNA polymerase (unscheduled DNA synthesis). Finally, a DNA ligase rejoins the two strands.
1. **Base Defect**

2. **Base Removal** (glycosylase)

3. **Incision** (AP endonuclease)

4. **Excision**

5. **Polymerisation**

6. **Ligation**
Figure 1.3: Nucleotide excision repair (adapted from Thompson 1989). DNA damage causes distortion of the DNA double helix which is recognised by a complex of proteins. The complex makes a strand incision about 22 bases 5' and 5 bases 3' of the damage and about 29 bases of DNA are excised. A polymerase then fills in the gap as represented by two closed circles (unscheduled DNA synthesis) and a DNA ligase rejoins the strand break.
helix distorting defect

damage recognition and incision

polymerisation

ligation
Evidence suggests at least three tolerance mechanisms in mammalian cells. Damage may temporarily halt DNA synthesis which, when restarted, continues across the lesion by an unknown mechanism (Sarasin and Hanawalt 1980). As in bacteria, damage may be bypassed by recombination with the sister chromatid (Fornace 1983). Finally, a 'copy-choice' model proposes that the second daughter strand is used as a template to allow DNA synthesis to continue past the lesion (Higgins et al 1976).

1.3: Human Diseases Arising from Defective DNA Repair

If DNA repair mechanisms are deficient, cells will accumulate DNA damage. This leads to increased chromosome instability with consequent effects on development and predisposition to cancer. Defective DNA repair processing mechanisms have been implicated in five rare autosomal recessive diseases since cells from patients suffering from these diseases display defective repair and/or increased chromosome instability. These diseases are; xeroderma pigmentosum, ataxia telangiectasia, Cockaynes syndrome, Fanconi anaemia and Blooms syndrome (reviewed by Friedberg in 1985, Timme and Moses in 1988 and Cleaver in 1990). These diseases are summarised in table 1.1.

1.3.1: Xeroderma Pigmentosum (XP)

This autosomal recessive disease is characterised by extreme sensitivity to sunlight and a very high rate of skin cancers (2000-fold increased) as well as an elevated rate of internal cancers (20-fold increased). Certain patients also suffer from neurological disorders (deSanctis-Cacchione syndrome). The incidence of XP varies from 1/200,000 in Europe to 1/40,000 in Japan. Classical XP is caused by a defect in the excision repair mechanism, as shown by a substantially reduced rate of DNA synthesis outside S-phase (unscheduled DNA synthesis) in response to irradiation with ultraviolet light (Cleaver 1968).

There are seven complementation groups designated A to G, plus variant (typical clinical symptoms, while exhibiting normal unscheduled DNA synthesis, but an abnormal pattern of DNA replication following ultra-violet irradiation). These groups were assigned according to acquisition by cultured cells of the ability to perform ultra-violet induced unscheduled DNA synthesis when fused to cells of another complementation group. Evidence suggests that in XP groups A to G it is the incision step that is defective, while XP variant cells are deficient in post-replication repair (Lehmann et al 1975). The existence of these groups indicates that there are at least seven gene products involved in excision repair in humans. Patients from
<table>
<thead>
<tr>
<th>disease</th>
<th>defect</th>
<th>complementation groups</th>
<th>cloned genes</th>
</tr>
</thead>
</table>
| xeroderma pigmentosum | defective excision repair in classical XP, defective post-
replication repair in XP variant | 8                      | ERCC2 (XP-D), ERCC3 (XP-B), ERCC5 (XP-G), XPA, XPC |
| Cockaynes syndrome  | defective preferential repair of transcribed DNA                       | 3                      | ERCC6 (CS-B)  |
| ataxia telangiectasia | no blockage of DNA synthesis following DNA damage?                   | 5                      |               |
| Fanconi anaemia     | unknown                                                                | 4                      | FACC          |
| Blooms syndrome     | a general defect affecting several enzymes?                            | 1                      |               |
| 46BR                | DNA ligase I deficiency                                                |                        | LIG1          |
complementation groups B, D and G may also suffer from Cockaynes syndrome (see below).

A number of artificially produced rodent excision repair mutant cell lines exist. These have been used to clone excision repair genes according to their ability to complement the repair deficiency of the mutant rodent cell lines. Such genes are therefore named excision repair cross-complementing (ERCC). Some of these genes have been found to be involved in XP (for review see Bootsma and Hoeijmakers 1994). The genes ERCC2, ERCC3 and ERCC5 are involved in XP-D, XP-B and XP-G respectively, while the gene ERCC4 may be involved in XP-F (Biggerstaff et al 1993, van Vuuren et al 1993). Additionally, the genes XPA and XPC, which are involved in XP-A and XP-C have also been cloned.

1.3.2: Cockaynes Syndrome (CS)

Patients suffering from the autosomal recessive disorder CS characteristically show arrested growth and development, while suffering from sun-sensitivity and deafness, optic atrophy, mental retardation and several other developmental aberrations. There does not appear to be any increased risk of developing tumours. CS may be, but is not always, associated with xeroderma pigmentosum.

In culture, CS cells are extremely sensitive to killing by ultra-violet irradiation. Wild-type cells exhibit preferential repair of transcribed regions of the genome in response to irradiation with ultra-violet, but this is defective in CS cells (Mullenders et al 1988, Venema et al 1990). In CS cells, unlike wild-type cells, there is no selection of the transcribed DNA strand over the non-transcribed strand of active genes, both are repaired at the same rate (van Hoffen et al 1993). Further, although the rate of repair of transcribed DNA is higher than that of transcriptionally inactive DNA in CS cells, it is still slower than repair of the nontranscribed strand of active genes in wild-type cells (van Hoffen et al 1993). Inhibition of RNA synthesis also follows ultra-violet irradiation of some CS cells (Mayne and Lehmann 1982). This is consistent with defective preferential repair of transcribed DNA, since pyrimidine dimers would block the path of RNA polymerases. There would only be a transient delay in RNA synthesis following ultra-violet irradiation in wild-type cells because transcriptionally active DNA is repaired first, thus removing the block to the RNA polymerase.

At least three CS complementation groups, A to C, have been identified on the basis of recovery of normal rates of RNA synthesis following ultra-violet irradiation (Lehmann 1982). Patients in groups A and B suffer from CS alone, while complementation group C is represented by patients who may also be members of
xeroderma pigmentosum groups B, D and G. The excision repair gene ERCC6 (Troelstra et al 1990) complements the ultra-violet sensitivity of CS complementation group B cells (Troelstra et al 1992). ERCC6 encodes a protein of 1493 amino acids with homology to the helicase family of proteins and may therefore be involved in searching for, or recognition of, damage on the transcribed strand of active genes.

1.3.3: Ataxia Telangiectasia (AT)

AT affects up to 1/40,000 people and in these patients the nervous and immune systems are particularly affected. They are exceptionally susceptible to infections and to neoplasia of the lymphoreticular system. Cerebellar ataxia results in a staggering walk with a severe lack of muscle coordination and progressive mental retardation. In addition there is prominent dilation of small blood vessels in the skin and eyes (telangiectasia). Like XP and CS, this disease displays an autosomal recessive mode of inheritance.

The first indication of a DNA repair defect came with the observation that X-rays, which cause DNA strand breaks, are fatal to many AT cancer patients when used for radiotherapy. AT cells are sensitive to X-irradiation, which elevates the already high number of chromosomal breaks. AT cell lines do not exhibit the blockage of DNA synthesis which is seen in normal cells in response to ionising radiation. Because of this, it has been suggested that the disease is not due to a DNA repair defect per se, but rather to the lack of inhibition of DNA synthesis allowing DNA replication prior to repair (Painter and Young 1980). This property has been used to define AT complementation groups, of which there are at least five, designated AB, C, D, E and V1 (Jaspers et al 1988).

Although no AT genes have yet been cloned, linkage analysis using 31 ataxia telangiectasia families has localised at least one AT gene to chromosome 11q22-23 (Gatti et al 1988). It remains to be seen whether this region of chromosome 11 contains a cluster of AT genes, or just one.

1.3.4: Fanconi Anaemia (FA)

Pancytopenia (depression of all types of blood cells), skeletal defects, stunted growth, bone marrow failure and predisposition to develop acute myeloid leukaemia distinguish this autosomal recessive disease. The incidence of this disease may be up to 1/40 000.

FA cells show high levels of spontaneous chromosome aberrations which are increased still further by DNA cross-linking agents. The molecular cause of the
disease is yet to be identified, but removal of intra-strand cross-links appears to be abnormally slow.

Cell fusion experiments have indicated that there are at least four complementation groups, designated A to D (Strathdee et al. 1992a). A cDNA for complementation group C (FACC) has been cloned by complementation of the sensitivity of FA group C cells to cross-linking agents (Strathdee et al. 1992b). The cDNA encodes a protein of 557 amino acids. There is no significant homology with any other known protein sequences and the function of the FACC gene product remains unidentified.

1.3.5: Blooms Syndrome (BS)

This disease, although rare in the general population, is relatively common among Ashkenazi Jews and the Japanese. There is some clinical variation in BS, but typical symptoms include retarded growth, sunlight sensitivity and sun-sensitive facial telangiectasia in a characteristic butterfly pattern. Immunodeficiency results in repeated infections during infancy and early childhood. This usually improves with age, however patients often then go on to develop severe chronic lung disease, probably arising from respiratory tract infections. Slight to moderate mental retardation is sometimes present. One of the most notable clinical features is a predisposition to develop cancer, of no particular type, at an early age, 24.8 years on average (German and Passarge 1989).

In culture, the most notable feature of BS cells is the elevated level of sister chromatid exchanges (SCEs) compared to normal cells (Chaganti et al. 1974), and this property has been useful in the diagnosis of BS. Interestingly however, some BS patients have two populations of circulating lymphocytes, exhibiting high and low SCE levels (German et al. 1977).

It appears that BS is represented by a single complementation group (Weksberg et al. 1988). Cell fusion experiments were carried out with BS cell lines isolated from patients with dissimilar ethnic backgrounds (Ashkenazi Jewish, French-Canadian, Mennonite and Japanese). The cell lines used were all of the high-SCE type and failed to complement this characteristic in each other. The mutation responsible for BS is therefore located within the same gene in all these cell lines, despite the fact that BS patients show substantial clinical variation. It may be that this reflects allelic differences or variable interaction with other gene products rather than several defective loci.

BS cells exhibit increased rates of spontaneous mutation and somatic recombination. The glycophorin A assay has been used to investigate this in vivo
Glycophorin A is a protein found on the surface of erythrocytes which has two codominantly expressed alleles, M and N. Using fluorescently labelled monoclonal antibodies where the colour of fluorescence is allele-specific, heterozygotes (MN), homozygotes (MM or NN) and hemizygotes (MØ or NØ) can be distinguished. Cells from heterozygous BS blood donors were analysed by this means. A high level of somatic recombination was indicated by the high frequency of homozygous variants, while the high frequency of hemizygous variants obtained is consistent with increased genomic instability and chromosome breakage.

Sensitivity to a number of different DNA-damaging agents is another characteristic of BS cells. For example, they are hypersensitive to mitomycin C (an alkylating agent which produces bulky adducts and interstrand crosslinks) as measured by cell killing (Hook et al 1984). They are also sensitive to the alkylating agents ethylmethane sulphonate (Arlett and Harcourt 1978) and N-ethyl-N-nitrosourea (Kurihara et al 1987). Sensitivity to these agents implies that BS cells are defective in DNA repair. The high levels of SCEs in these cells are consistent with this, since SCEs are thought to occur as a result of unrepaired DNA damage persisting into S phase.

DNA replication is abnormal in BS cells. The rate of DNA chain growth is retarded (Hand and German 1975). Furthermore, short-term pulse labelling experiments using $[^3H]$tritium allowed the detection in BS cells of aberrant 20kb DNA replication intermediates which are not seen in normal cells (Lönn et al 1990) indicating a deficiency in DNA ligation. A defect in ligation is further suggested by the work of Rünger and Kraemer (1989). A linearised, replicating, shuttle vector plasmid was passaged through BS cells and used to transform bacteria to give a measure of ligation. The in vivo joining efficiency was found to be three-fold reduced in BS cells. In addition, sequencing revealed a number of aberrant joining events. Deletions, point mutations, plasmid DNA reverse insertions, non-plasmid DNA insertions and filling in of cohesive ends were all observed.

Consistent with work suggesting that DNA ligation is defective in BS cells, reduced DNA ligase I activity has been demonstrated in lymphoid cell lines from BS patients (Willis and Lindahl 1987, Chan et al 1987). Two types of abnormal DNA ligase I have since been identified (Willis et al 1987). Five independently derived cell lines exhibited a thermolabile enzyme with reduced levels of activity, designated type I-1. However in one cell line the enzyme showed normal activity and heat-stability, but was apparently present in a dimeric form within the cell. This enzyme has been designated type I-2. It has been shown that similar amounts of DNA ligase
I protein are present in normal and BS cell lines (Chan and Becker 1988), therefore the observed reduction in DNA ligase I activity in type I-1 BS cells is due to impaired activity rather than abnormally low levels of the enzyme. This was confirmed by demonstration that formation of DNA ligase I reaction intermediates is affected in the cell lines examined. These results strongly suggested that the primary defect in BS is in DNA ligase I, which would explain a number of the abnormalities observed in BS cells, such as the elevated rate of sister chromatid exchanges, increased somatic recombination rate, delayed maturation of DNA replication intermediates and reduced and error-prone ligation of plasmid DNA. Further, the presence of two separate alterations in the same enzyme is consistent with a single BS complementation group.

In contrast to these findings, other laboratories have shown, under different conditions, that levels of DNA ligase I activity are not reduced. Kenne and Ljunquist (1988) demonstrated that DNA ligase I activity is two to three-fold higher than normal in the BS line GM1492 which expresses a DNA ligase-stimulatory factor. Mezzina et al (1989) found that DNA ligase I activity in BS cells was equal to or greater than that in normal cell lines.

Although BS cells do apparently contain aberrant DNA ligase I, recent work has demonstrated that the primary defect is not at this locus. Human DNA ligase I cDNA has been cloned and sequenced (Barnes et al 1990) and cDNA sequence comparisons have revealed that the amino acid sequence of DNA ligase I from BS cells contains no alterations (Petrini et al 1991, Barnes et al 1992). It has also been shown that DNA ligase I cDNAs from BS cell lines encode a functional enzyme, since they can complement S. cerevisiae DNA ligase cdc9 mutants (Petrini et al 1991). In addition, the DNA ligase I gene has been mapped to human chromosome 19 (Barnes et al 1990, Barnes et al 1992a), yet the high sister-chromatid exchange phenotype characteristic of BS cells is complemented by human chromosome 15 (McDaniel and Schultz 1992). The DNA ligase locus itself is therefore not mutated since, as well as these observations, it has been shown that expression of the gene is normal (Chan and Becker 1988, Petrini et al 1991). It is possible that some other factor is at fault, such as defective post-translational modification or some other factor mediating activity of the enzyme.

Reports have been made concerning a number of other biochemical abnormalities in BS cells (for review see German and Passarge 1989). These include an increased rate of synthesis of 'a set of proteins' (Mallick et al 1982), increased activity of a DNA recombining protein (Kenne and Ljunquist 1984) and aberrant regulation of the DNA repair enzymes uracil-DNA glycosylase (Gupta and Sirover
1984) and hypoxanthine-DNA glycosylase (Dehazya and Sirover 1986). Abnormal regulation of \(0^6\)-methylguanine methyltransferase has also been observed (Kim et al 1986). It has been reported that uracil-DNA glycosylase from BS cells is immunologically altered (Seal et al 1988) and has different properties to the enzyme from normal cells, including a changed isoelectric point, \(K_m\) and \(V_{\text{max}}\), and increased thermolability (Seal et al 1991) Another study has shown, however, that the activity of the enzyme is normal (Vilpo and Vilpo 1989). It has further been shown that more DNA breakage occurs in BS cells in response to 5-fluorouracil treatment, an effect which depends upon uracil-DNA glycosylase levels (Lönn et al 1990).

Elevated levels of the enzyme superoxide dismutase have been detected (Nicotera et al 1989) and it has also been shown that BS cells sustain abnormally high levels of DNA damage arising from free radicals (Poot et al 1990). This is consistent with the significantly raised amount of active oxygen present within the cells (Nicotera et al 1993). It is therefore likely that BS cells are defective in the detoxification of oxygen free radicals.

In normal cells the tumour suppressor gene product p53 accumulates in the nuclei following damage to DNA and is thought to play a role in arresting the cell cycle in G1, thus allowing DNA repair to take place prior to replication. In BS cells there is no p53 build-up (Lu and Lane 1993), and in one BS line (GM1492) p53 is undetectable, although cell cycle arrest occurs as normal (van Laar et al 1994).

The significance of such a wide range of abnormalities is unclear, but when these are considered together with the findings on the DNA ligase I abnormality, it may be that there is a general defect in BS cells which affects several enzymes.

1.3.6: The DNA Ligase I-Deficient Cell Line 46BR

A fibroblast cell line (46BR) has been derived from a patient, now deceased, who exhibited some of the characteristic clinical symptoms of BS - retarded growth, sunlight sensitivity and immunodeficiency. She was diagnosed as having a malignant lymphoma and died of pneumonia, aged 19 (Webster et al 1992).

46BR cells exhibit a deficiency of DNA ligase I. Defective ligation was first suggested by initial characterisation of the cell line. Teo et al (1982, 1983a) found that 46BR cells are sensitive to ionising (\(\gamma\)) radiation, slightly sensitive to the DNA cross-linking agent mitomycin C, and sensitive to alkylating agents such as \(N\)-methyl-\(N\)-nitrosourea (MNU), MNNG and EMS, but not to \(N\)-ethyl-\(N\)-nitrosourea (ENU). This sensitivity to a wide range of DNA damaging agents indicated a defect in a common step of DNA repair, either DNA polymerisation to fill in gaps created
by damage excision, or ligation. This is consistent with the lack of sensitivity to ENU, since its major reaction products, such as O6-ethylguanine, are removed directly rather than by a repair process involving polymerisation or ligation. Dimethyl sulphate reaction products are removed by the excision repair pathway and DNA strand breaks are produced by the first repair step - excision of the damaged nucleotides. In 46BR these strand breaks accumulate (Teo et al 1983b), indicating a defect at a later stage of excision repair - DNA polymerisation or ligation. No reduction in repair DNA synthesis could be detected therefore the defect was identified as being at the ligation step. These workers also demonstrated that 46BR cells are hypersensitive to 3-aminobenzamide, which inhibits polyADP-ribose polymerase, an enzyme required as well as ligase for efficient ligation, further showing that there is a ligation defect in 46BR. A similar study by Squires and Johnson (1983) showed that single-strand breaks accumulate in the DNA of 46BR cells following ultraviolet irradiation, even in the presence of inhibitors of DNA polymerisation. This indicated that, during repair of ultraviolet-induced damage via the excision repair pathway, DNA strand breaks were not accumulating following the incision step because of defective gap-filling, but due to a defect in the later step of DNA ligation.

Defective ligation in 46BR was demonstrated more directly by measuring accumulation of replication intermediates (Henderson et al 1985). Transient accumulation of low molecular weight DNA was observed, indicating abnormally slow joining of Okazaki fragments. However doubt has recently been cast over the validity of this conclusion. Prigent et al (1994) measured the rate of Okazaki fragment joining and found that the initial rate, at least, was normal.

When 46BR cell extracts were size-fractionated and assayed for DNA ligase I activity it was found that the enzyme present appears to have the type I-2 defect (Lehmann et al 1988). Subsequent to the demonstration that ligation is indeed abnormal in 46BR, due to aberrant DNA ligase I activity, mutations were identified on each allele of the DNA ligase I gene (Barnes et al 1992b). One mutation is at the active site of the enzyme, and specifies a change of Glu566 to Lys. This Glu residue is highly conserved between human DNA ligase I and DNA ligases of E. coli, Vaccinia, T7 and the yeasts S. cerevisiae and S. pombe (Tomkinson et al 1991b). ATP binds covalently to the active site of the enzyme, forming an enzyme-adenylate intermediate (figure 1.4). Glu566 is separated from the site of ATP binding to the enzyme by only one amino acid and the charge alteration from acidic to basic significantly impairs the enzyme's ability to form a DNA ligase I-adenylate intermediate, and completely removes its ability to complement a temperature
Figure 1.4: Reaction mechanism of ATP-dependent DNA ligases (adapted from Lehman 1974). A: An enzyme-adenylate complex is generated by formation of a phosphoamide bond between AMP and a lysine residue at the active site of the enzyme. Pyrophosphate is released in this step of the reaction. B: The enzyme-adenylate complex reacts with nicked DNA and AMP is transferred from the enzyme to the 5' phosphoryl group, creating a covalent DNA-adenylate intermediate. C: Enzyme-catalysed nucleophilic attack by the 3' hydroxyl group on the 5' phosphoryl group generates a phosphodiester bond, sealing the nick and releasing AMP.
A

\[
E\text{-}(\text{Lys})\text{-NH}_3^+ + \text{adenosine-O-P-O-P-O-P-O} \rightarrow E\text{-}(\text{Lys})\text{-N}_2\text{P-O-adenosine} + H_2O + PP_i
\]

ENZYME  ATP  ENZYME-ADENYLATE COMPLEX

B

\[
E\text{-}(\text{Lys})\text{-N}_2\text{P-O-adenosine} + \text{NICKED DNA} \rightarrow \text{DNA-ADENYLATE COMPLEX} + \text{ENZYME}
\]

C

\[
\text{ENZYME} + \text{O-P-O-adenosine} \rightarrow \text{SEALED NICK} + \text{AMP}
\]
sensitive *E. coli* DNA ligase mutant (Kodama *et al* 1991). DNA ligase I carrying the Glu566 to Lys alteration is therefore probably enzymatically inactive. The second mutation, which was maternally inherited, is an alteration of Arg771 to Trp. Again, this has occurred in a region of the enzyme which is extremely highly conserved. The allele carrying this mutation is the only one expressed in the SV40-transformed derivative of 46BR, called 46BR.1G1. Extracts from these cells were found to show a decreased ability of DNA ligase I to form an enzyme-adenylate intermediate (Barnes *et al* 1992b), indicating that the Arg771 to Trp mutation leaves the enzyme with reduced activity.

The activity of DNA ligase I from 46BR.1G1 is reduced about 20-fold (Prigent *et al* 1994). The enzyme is almost entirely present in an adenylated form *in vivo* indicating that adenylation occurs as normal and the defect is therefore at a later stage of the reaction mechanism. This is consistent with earlier reports of a decreased ability to form an enzyme-adenylate intermediate since, if most of the enzyme is already adenylated, only a low level of adenylation with labelled ATP will be measurable. Following adenylation, the AMP moiety is transferred from enzyme to DNA, forming a covalent DNA-adenylate intermediate (figure 1.4). During the ligation reaction accumulation of DNA-adenylate intermediates can be detected, indicating that formation of this reaction intermediate is normal but that the ability of the enzyme to perform phosphodiester bond formation is reduced.

46BR is so far the sole human representative of DNA ligase I deficiency and the cell line will be useful in investigating the role of DNA ligase I in cellular metabolism.

1.4: Mammalian DNA Ligases

Following the demonstration of defective DNA ligase I in Blooms syndrome and in the human cell line 46BR, much interest has been focused on mammalian DNA ligases, of which three kinds have been identified (table 1.2).

DNA ligases are essential for replication and recombination of DNA. In addition, all mechanisms of DNA repair involving strand breakage and rejoicing depend ultimately on DNA ligases for success. Mammalian DNA ligases mediate joining of single strand breaks in double stranded DNA in the presence of ATP and Mg2+ ions.

The reaction mechanism is apparently identical to that elucidated for bacterial and bacteriophage enzymes, although bacterial enzymes require NAD+. Overall, a phosphodiester bond is formed between a 3'-hydroxyl group and a 5'
Table 1.2: Properties of Mammalian DNA Ligases

<table>
<thead>
<tr>
<th></th>
<th>DNA ligase I</th>
<th>DNA ligase II</th>
<th>DNA ligase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>ligation of:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oligo(dT).poly(dA)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>oligo(dT).poly(rA)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>oligo(rA).poly(dT)</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>size estimated by SDS-PAGE</td>
<td>125-130kDa</td>
<td>67-72kDa</td>
<td>100kDa (plus 46kDa)</td>
</tr>
<tr>
<td>Km for ATP</td>
<td>low</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>recognition by DNA ligase I-specific antisera</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>apparently involved in DNA replication</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>apparently involved in DNA repair</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
phosphoryl group (reviewed by Lehman, 1974). There are three steps involved in this reaction (figure 1.4):

(1) The enzyme reacts with ATP, forming a covalent ligase-adenylate intermediate, with release of pyrophosphate. Adenylation occurs by means of a phosphoamide bond to a lysine residue on the protein.

(2) The adenylyl group is transferred from the enzyme to DNA, forming a pyrophosphate linkage between AMP and the 5' phosphoryl group.

(3) A phosphodiester bond forms between the 3'-hydroxyl and the activated 5'-phosphoryl group with elimination of AMP.

This reaction mechanism was confirmed for mammalian enzymes by isolation of the two reaction intermediates. First, incubation of mammalian DNA ligase with [14C]ATP produced a labelled enzyme. This enzyme could ligate DNA single strand breaks in the absence of added ATP in a reaction releasing AMP. Also, ATP could be released from the complex by addition of pyrophosphate (Söderhäll and Lindahl 1973a). Therefore the adenylated ligase is an intermediate allowing the reaction to proceed in either direction. Second, incubation of adenylated ligase with nicked DNA at 0°C produced covalently labelled DNA still containing single strand breaks (Söderhäll 1975).

DNA ligases also catalyse the reverse reaction. In the presence of AMP and Mg²⁺ they act as topoisomerases, mediating the progressive loss of superhelical turns from covalently closed double stranded DNA circles. This activity of a human DNA ligase is inhibited by pyrophosphate and ATP (Montecucco et al 1988). Ligation activity is inhibited by histones, but appears to be stimulated by polyADP-ribose (Creissen et al 1982, Chashi et al 1983). The enzyme which synthesises polyADP-ribose, called polyADP-ribose polymerase, appears to be involved in the ligation reaction. This enzyme utilises NAD⁺ to produce (ADP-ribose)n. Its activity increases in response to DNA damage and is specifically inhibited by 3-aminobenzamide (Durkacz et al 1980). Inhibition of polyADP-ribose polymerase results in the accumulation of dimethyl sulphate-induced DNA strand breaks as well as increased cytotoxicity of the alkylating agent MNU and of γ-radiation (Durkacz et al 1980, Nduka et al 1980), which initially suggested that it is involved in DNA repair in some way. A direct link between this enzyme and DNA ligases was indicated by the work of Creissen and Shall in 1982. They showed that DNA ligase II is induced in response to treatment with dimethyl sulphate and that this increase in activity is blocked by 3-aminobenzamide, thus demonstrating that
polyADP-ribose polymerase is required for ligation. It is now thought that the enzyme binds tightly to DNA strand breaks and is released by automodification with polyADP-ribose, which then allows ligation to take place (Satoh and Lindahl 1992). If release is inhibited by 3-aminobenzamide, the polyADP-ribose polymerase will remain bound and interfere with ligation. The role of polyADP-ribose polymerase in ligation remains unknown.

1.4.1: DNA Ligase I

Of the three DNA ligases which have been identified in mammalian cells, DNA ligase I is thought to be the major activity. The three ligases can be distinguished in a number of ways, including their substrate specificities, size and immunological reactivity.

1.4.1.1: Substrate Specificity of DNA Ligase I

As well as joining single and double strand breaks in DNA, including blunt ends, DNA ligase I is also active on the synthetic substrates oligo (dT) bonded to poly (dA), and oligo (rA) bonded to poly (dT) (Arrand et al 1986, Tomkinson et al 1991a). It has a high affinity for ATP during the ligation reaction (Tomkinson et al 1991a).

1.4.1.2: The Human DNA Ligase I Gene

Human DNA ligase I cDNA has been cloned by two separate methods (Barnes et al 1990). A cDNA library was screened with oligonucleotide probes designed from the sequences of tryptic peptides of bovine DNA ligase I, and in parallel, a human cDNA library in a yeast expression vector was tested for complementation of a temperature sensitive DNA ligase mutation (cdc9) in S. cerevisiae. Identical clones were obtained by both methods. The sequence of an apparently full-length 3.2kb clone revealed that it encodes a 102kDa protein of 919 amino acids. The gene encoding human DNA ligase I is present on the long arm of chromosome 19 at position 13.2-13.3 (Barnes et al 1990, Barnes et al 1992a). It occupies 53kb of DNA and consists of 28 exons, ranging in size from 54 bases to 173 bases, which are transcribed to produce a single 3.2kb mRNA (Noguiez et al 1992). Like many other 'housekeeping' genes, the promoter region is highly GC-rich and lacks a TATA box.
1.4.1.3: DNA Ligase I Structure

The N-terminus of DNA ligase I is blocked to sequencing (Tomkinson et al 1990). The enzyme is highly asymmetric with a Stokes radius of 52-53Å (Teraoka & Tsukada 1982, Tomkinson et al 1990) and hydrodynamic measurements suggest a molecular weight for the native enzyme of 98kDa (Tomkinson et al 1990) which is close to the value of 102kDa determined by sequencing (Barnes et al 1990). However DNA ligase I has an apparent molecular weight of 130kDa (Teraoka and Tsukada 1982) or 125kDa (Tomkinson et al 1990, Lasko et al 1990) as determined by SDS-PAGE. Additionally, some workers have detected a 200kDa protein with a half-life of just 30 minutes by immunoprecipitation with DNA ligase I-specific antibodies (Teraoka & Tsukada 1985), which rapidly breaks down to an active 130kDa form (Teraoka & Tsukada 1986). The discrepancy between the calculated molecular weight and that apparent from SDS-PAGE may be because sequencing has revealed that DNA ligase I has an unusually high proline content and as a result runs anomalously slowly on polyacrylamide gels.

Human DNA ligase I cDNA is 40% homologous to the DNA ligases of S. cerevisiae and Schizosaccharomyces pombe, with homology being restricted to the C-terminal region of the enzymes (Barnes et al 1990). It has been shown that this region contains the catalytic domain, since limited proteolysis of DNA ligase I generates a relatively protease-resistant 85kDa fragment from the C-terminus which retains catalytic activity (Tomkinson et al 1990). Additionally, in experiments where the human DNA ligase I cDNA was truncated from the 5' end, followed by functional expression in a conditional lethal ligase mutant of E. coli to look for complementation at the non-permissive temperature, it was shown that an 80kDa C-terminal fragment is sufficient to provide ligation activity. However loss of only 22 amino acids from the C-terminus abolished complementation of the mutation (Kodama et al 1991).

Inhibition of DNA ligase I activity by pyridoxal phosphate indicates the presence of a lysine residue at the active site of the enzyme, as would be expected (Tomkinson et al 1991b). The position of this lysine residue within the enzyme was determined by labelling with [3H]ATP, followed by trypsin digestion and chromatographic purification of labelled peptides (Tomkinson et al 1991b). Sequencing of adenylated peptides revealed a stretch of amino acids also present in the predicted amino acid sequence of human DNA ligase I. Sequences present in DNA ligases from S. cerevisiae, S. pombe, vaccinia virus, E. coli, and bacteriophages T3, T7 and T4 also show some homology with the peptide sequence. Contained within this sequence is an active site motif, Lys-Tyr/Ala-Asp-Gly-X-Arg,
present in all ATP-dependent DNA ligases studied so far. The lysine residue (Lys568 in the human enzyme) in this motif is presumed to be the site of adenylation and is consistently $332 \pm 20$ amino acid residues from the C-terminus of the enzyme. The functional significance of some of the amino acid residues around the active site lysine of human DNA ligase I has been tested by in vitro mutagenesis (Kodama et al 1991). Alteration of the lysine residue at position 568 to histidine or arginine completely abolished the ability of the active catalytic domain of human DNA ligase I to form a covalent enzyme-adenylate intermediate or to complement E. coli conditional lethal ligase mutants. Further, alteration of Glu566 to Lys (one of the mutations identified in 46BR) had a similar effect, as did alteration of the residues at positions 569, 570, 571 and 573, although enzyme-adenylate formation was not impaired in every case.

1.4.1.4: Regulation of DNA Ligase I Activity

The function of the N-terminal domain of DNA ligase I appears to be regulatory. When DNA ligase I purified from calf thymus is dephosphorylated, it is also inactivated (Prigent et al 1992). In contrast, the C-terminal catalytic domain retains activity despite phosphatase treatment. It therefore appears that the N-terminus regulates activity via phosphorylation, while the catalytic domain on its own is unregulated. When full-length human DNA ligase I cDNA is expressed in E. coli, the resulting protein is inactive until phosphorylated in vitro, whereas the C-terminal catalytic domain, lacking the N-terminus, is active independently of phosphorylation (Prigent et al 1992). This is consistent with other reports that, when expressed in E. coli, the C-terminal catalytic domain, but not full-length DNA ligase I, is active and can complement ligase-deficient E. coli cells (Kodama et al 1991). When expressed in S. cerevisiae, on the other hand, both full-length human DNA ligase I and the catalytic domain complemented ligase deficient cdc9 mutants (Barnes et al 1990). It is possible, therefore, that active DNA ligase I can only be expressed by cells which possess the specific kinase necessary for phosphorylation and this kinase may only be present in eukaryotic cells. Of the kinases tested, only casein kinase II (ubiquitous throughout eukaryotes) appeared able to activate the enzyme, by phosphorylation at serine residues. Seven S_ _E motifs, which are possible casein kinase II phosphorylation sites (Kemp and Pearson 1990), are present in the N-terminal domain of human DNA ligase I. Of these putative phosphorylation sites, two are potentially 'strong' substrates for the enzyme, and it was suggested that one of these is the site of phosphorylation (Prigent et al 1992).
In contrast to these observations, separate studies where human DNA ligase I cDNA was expressed in *E. coli*, demonstrated that expression of a full-length cDNA produced an active protein, even though it was not phosphorylated (Teraoka *et al* 1993). The major difference between these experiments is that Kodama *et al* (1991) and Prigent *et al* (1992) expressed DNA ligase I as a fusion protein with 23 amino acids of β-galactosidase, while Teraoka *et al* (1993) had altered the protein only by addition of a threonine residue immediately after the initiator methionine.

There have been reports of other factors which affect the activity of DNA ligase I. A heat-resistant protein stimulatory factor, which also acts upon DNA ligase II, was observed in the Blooms syndrome fibroblast cell line GM1492 (Kenne and Ljungquist 1988), while a heat-sensitive DNA ligase I stimulatory factor has been isolated from human embryo kidney cells (Fairman *et al* 1992). An inhibitor of DNA ligase I has also been identified (Yang *et al* 1992, Yang *et al* 1993). The inhibitor is heat-labile, with a molecular weight of 55-75kDa and appears to block ligation activity by a reversible physical interaction with the catalytic domain of the enzyme. Formation of the enzyme-adenylate intermediate is unaffected, but subsequent transfer of the adenylate moiety to DNA cannot take place when DNA ligase I is complexed with the inhibitor.

1.4.1.5: Intracellular Location and Function of DNA Ligase I

Immunocytochemical and indirect immunofluorescence studies on bovine kidney cells localised DNA ligase I to the nucleus, but not nucleoli. However, in mitotic cells it is present in the cytoplasm following breakdown of the nuclear membrane (Lasko *et al* 1990).

The observation that DNA ligase I levels increase in rapidly dividing cells such as malignant cells or those found in regenerating liver (Söderhäll 1976, Chan and Becker 1985) suggests that the major function of this enzyme is in DNA replication. Moreover, it has been demonstrated that levels of DNA ligase I mRNA increase during proliferation in mouse and human fibroblasts and in human lymphocytes (Montecucco *et al* 1992). Also, the observed pattern of DNA ligase I expression in response to proliferation resembles that of DNA polymerase α (Wahl *et al* 1988), which is involved in DNA replication. Further, when cells are induced to differentiate, DNA ligase I mRNA levels fall (Montecucco *et al* 1992). Waga *et al* (1994) have demonstrated that DNA ligase I is an essential component for replication of Simian virus 40 replication *in vitro*. DNA ligase I has also been shown to be a component of mammalian DNA replication complexes. A 21S complex which can replicate Simian virus 40 DNA *in vitro* has been isolated from HeLa cells.
The complex contains DNA polymerase α, DNA primase, a 3'-5' exonuclease, RNase H, topoisomerase I and DNA ligase I (Li et al 1993, Li et al 1994). Wu et al (1994) isolated a 17S complex from a mouse mammary carcinoma cell line, capable of replicating polyomavirus DNA in vitro. The components of this complex were identified as DNA polymerases α and δ, DNA primase, proliferating cell nuclear antigen, DNA helicase, DNA topoisomerases I and II, and DNA ligase I. These observations support the hypothesis that DNA ligase I is involved in DNA replication.

However, DNA ligase I mRNA levels also increase in response to ultraviolet irradiation (Montecucco et al 1992) and the DNA ligase I-deficient human cell line 46BR is sensitive to a wide range of DNA damaging agents (Teo et al 1982, 1983a). This suggests a role for DNA ligase I in DNA repair.

1.4.2: DNA Ligase II

A second DNA ligase was identified when it was discovered that two DNA ligase activities from calf thymus are separable by hydroxyapatite chromatography and gel filtration (Söderhäll and Lindahl 1973b). Size-fractionation of the enzymes by gel filtration, followed by ligation activity assays, revealed a large peak of activity due to the high molecular weight DNA ligase I and a minor peak of activity resulting from the smaller DNA ligase II. The two ligases are also distinguishable immunologically. Antiserum which neutralises the activity of DNA ligase I does not affect DNA ligase II (Söderhäll and Lindahl 1975).

SDS-PAGE followed by immunoblotting with an antibody specific for calf thymus DNA ligase II indicates that this enzyme consists of a single 68kDa polypeptide (Teraoka and Tsukada 1986), while other workers have estimated its size as 67kDa (Yang et al 1990) or 72kDa (Tomkinson et al 1991a). Like DNA ligase I, the N-terminus of DNA ligase II is blocked to sequencing (Tomkinson et al 1991a). Also like DNA ligase I, DNA ligase II is a nuclear enzyme (Söderhäll and Lindahl 1975).

DNA ligases I and II have a number of distinct properties. DNA ligase II is more thermolabile than DNA ligase I and has more specific pH requirements for activity (Söderhäll and Lindahl 1973b). Unlike DNA ligase I, DNA ligase II cannot catalyse blunt-end DNA joining or ligation of oligo(rA) molecules bonded to a poly (dT) template, while it can join oligo (dT) molecules bonded to a poly (dA) template and oligo (dT) molecules bonded to a poly (rA) template (Arrand et al 1986, Tomkinson et al 1991a). DNA ligase II has a high Km for ATP (a low affinity for ATP), while DNA ligase I has a low Km (Tomkinson et al 1991a).
DNA ligases can be labelled at the active site with $\alpha$$^32$P-ATP since one of the ligation reaction intermediates is a covalent enzyme-AMP complex. When DNA ligases I and II were labelled in this way, followed by partial proteolysis and SDS-PAGE, a labelled fragment common to both enzymes was observed (Yang et al 1990). This indicates that the active sites of both enzymes are closely related. However other workers have been unable to identify any proteolytic catalytic fragments shared between DNA ligases I and II (Roberts et al 1994). This is consistent with previous reports that an antibody which inhibits the activity of DNA ligase I does not affect DNA ligase II (Söderhäll and Lindahl 1975), while, in contrast, an antibody which neutralises DNA ligase II activity has no effect on DNA ligase I activity (Teraoka et al 1986).

DNA ligase II is apparently not induced in response to cell proliferation (Söderhäll and Lindahl 1975). Thus there is no elevation of DNA ligase II levels present in rapidly dividing cells (mouse ascites tumour cells and calf thymus) over other cells (calf liver, calf spleen, human placenta and rabbit spleen). It has also been demonstrated that DNA ligase II levels are similar in normal and regenerating rat liver (Söderhäll 1976, Chan and Becker 1985) while DNA ligase I levels increase 15-fold during liver regrowth. This suggests that DNA ligase II is not extensively involved in DNA replication.

1.4.3: DNA Ligase III

During purification of DNA ligase II from calf thymus, using for identification its ability, unlike DNA ligase I, to ligate an oligo (dT).poly (rA) substrate, an additional ligase activity able to act on this substrate was detected (Tomkinson et al 1991a). This ligase activity has different purification properties to both DNA ligases I and II and is detectable as a polypeptide of 100kDa in size as estimated by SDS-PAGE. It is apparently associated with a second polypeptide of 46kDa. It has different substrate specificities to both DNA ligases I and II since it can ligate the oligo (dT).poly(rA) substrate and also oligo (dT).poly(dA) and oligo (rA).poly (dT) substrates. DNA ligase I can ligate only the latter two substrates, while DNA ligase II can only ligate the first two. The newly discovered ligase activity has a similar low Km for ATP, and a similarly sized catalytic domain (87kDa, compared to 85kDa), to DNA ligase I, but is immunologically distinct from it. This ligase activity therefore represents a separate enzyme from DNA ligases I and II and has been designated DNA ligase III. It is possible that other workers had already isolated DNA ligase III, although if so, it was erroneously identified as DNA ligase II. A ligase activity sharing characteristics of DNA ligase III, most notably a
size of 100kDa, a low Km for ATP and the ability to ligate an oligo(dT).poly (rA) substrate, was isolated from rat liver (Elder and Rossignol 1990).

Peptide fingerprinting of DNA ligases has revealed several small fragments indistinguishable between DNA ligases II and III, indicating that their active site regions around the ATP binding sites are very closely related (Roberts et al 1994).

DNA ligase III has a Stokes radius of 58Å (Tomkinson et al 1991a) and comigrates with DNA ligase I (which has a Stokes radius of 52-53Å, Teraoka and Tsukada 1982, Tomkinson et al 1990) during gel filtration, resulting in a single peak of high molecular weight ligase activity (Tomkinson et al 1993). In SV40-transformed human fibroblasts the major ligation activity in this peak is due to DNA ligase III. However, in calf thymus DNA ligase III contributes only 5-10% of DNA ligation activity.

DNA ligase III is not induced in proliferating rat liver (Elder and Rossignol 1990) which suggests that it may not be heavily involved in the replication of DNA. It may therefore have a role in DNA repair, a possibility which is supported by the isolation of a protein complex that repairs double stranded breaks and deletions by recombination (Jessberger et al 1993). The complex, called RC-1, contains at least five proteins and so far a DNA polymerase (ε), a 5'-3' exonuclease and a DNA ligase have been identified. The ligase activity is probably DNA ligase III since it can act on an oligo(dT).poly (rA) substrate and is about 100kDa in size. This implicates DNA ligase III in DNA repair and recombination. It has further been shown that DNA ligase III interacts with the DNA repair protein XRCC1 (Caldecott et al 1994) since it copurifies with XRCC1 during affinity chromatography. This raises the interesting possibility that XRCC1 is another component of RC-1.

1.5: Gene Targeting and the Creation of Mouse Models for Human Genetic Disease

Gene targeting permits the precise alteration of specific genes in cultured cells which can then be used to generate animals in which every cell carries the desired alteration (reviewed by Capecchi in 1989 and Melton in 1990). As well as complete gene inactivation, examples of alterations which can be made include domain removal, active site mutations, point mutations and elimination of protein-protein recognition sites. This ability to modify genes at will means that gene targeting is an excellent technique for investigating the role of gene products, both at the cellular and multicellular levels. It is also potentially of great use in gene therapy and the study of human genetic diseases. In this respect, creation of a mouse model for a human disease by introduction of mutations associated with that disease, looks
promising for studies into disease progression and even evaluation of possible treatments.

Some of the work presented in this thesis involves gene targeting, therefore it seems appropriate to briefly describe the main aspects of this technique. The procedure is to design a targeting vector which will cause the desired alteration in mouse embryonic stem (ES) cells. The vector should contain DNA sequence homologous to the gene of interest, but carrying the modification. When introduced into the cells, homologous recombination between vector DNA and the endogenous gene introduces the desired alteration at the chosen locus. There are two classes of targeting vector, called insertion or replacement, according to whether the whole vector is inserted or part of the genome is replaced with exogenous DNA containing the alteration (figure 1.5). Insertion vectors, by their very nature, are extremely disruptive, while replacement vectors have the advantage that they can be used to make more subtle alterations.

When DNA is introduced into the ES cells most becomes randomly integrated into the genome by joining of the exogenous DNA ends with the chromosomal DNA (Folger et al 1982). Targeting events occur much more rarely and all targeting strategies therefore include one or more methods of enrichment. Due to the extreme rarity of targeting events, the success or failure of a targeting experiment may rely on the method of enrichment used. The simplest example of enrichment is to target a directly selectable gene, such as HPRT. HPRT is involved in the purine salvage pathway and it is possible to select for, or against, cells expressing the wild type gene. Cells in which the enzyme is active can grow in HAT medium which contains hypoxanthine, aminopterin (an inhibitor of de novo synthesis of purines and pyrimidines) and thymidine. Cells can bypass the metabolic block resulting from the presence of aminopterin in the medium by utilising hypoxanthine and thymidine in the purine salvage pathway, unless HPRT is mutated. Selection against active HPRT is achieved using the cytotoxic analogue 6-thioguanine. In addition, because it is present on the X chromosome, only one copy of the HPRT gene needs to be altered in male ES cells for the cellular phenotype to reflect the change. Because of these useful properties, HPRT has been extensively used in gene targeting experiments. However, unlike HPRT, most other genes are not directly selectable so alternative methods of selecting for targeting events have been developed. One such method is known as positive-negative selection, so-called because cells which have integrated input DNA are selected for integration of a marker (positive selection), while negative selection is used against cells in which DNA has integrated randomly (figure 1.6). This can be achieved using the Herpes
Figure 1.5: Insertion and replacement targeting vectors. Targeting of the HPRT gene is illustrated. The cells being targeted have a wild-type HPRT gene, consisting of 9 exons (indicated by the filled-in boxes), and can therefore be selected against in 6-thioguanine. The targeting vector contains a genomic fragment extending from exon 6 to exon 9, but with exon 8 replaced by a neomycin phosphotransferase gene (neo, hatched box). A: Insertion - The vector is linearised within the region of homology, i.e. between exons 6 and 7. Recombination between the vector and the target gene therefore results in the insertion of the whole vector, with consequent duplication of the region of homology. Thus the gene ends up with two copies of the region spanning exons 6 to 9. However, in the first copy, exon 8 has been replaced with a neomycin phosphotransferase gene, conferring resistance to the synthetic antibiotic G418. The genotype of the targeted cells is therefore now hpr\(^r\) G418\(^r\) and they will grow in medium supplemented with 6-thioguanine and G418. B: Replacement - The vector is linearised outwith the region of homology so that recombination results in replacement of the region between exons 6 and 9 with the equivalent region from the vector. In this case there is no duplication of genomic sequences, but exon 8 has again been replaced with a neomycin phosphotransferase gene rendering the cells hpr\(^r\) G418\(^r\).
Figure 1.6: Positive-negative selection. Targeting of the HPRT gene is illustrated. The cells being targeted have a wild-type HPRT gene, consisting of 9 exons (indicated by the filled-in boxes), and can therefore be selected against in 6-thioguanine. The targeting vector contains a genomic fragment extending from exon 6 to exon 9, but with exon 8 replaced by a neomycin phosphotransferase gene (*neo*, hatched box). A viral thymidine kinase gene has been placed adjacent to exon 9 (*tk*, shaded box). A: Targeting - Homologous recombination occurs between the two regions of homology flanking the neomycin phosphotransferase gene, which then integrates into the gene, replacing exon 8. No other part of the vector inserts into the gene, therefore the thymidine kinase gene is lost. Targeted cells can therefore be selected for growth in 6-thioguanine and gancyclovir. B: Random Integration - The linearised vector inserts into the genome via its ends. The whole plasmid, including the thymidine kinase gene, integrates at a random position. The HPRT gene is therefore not disrupted and the cells will die in the presence of 6-thioguanine and gancyclovir.
TARGETING

RANDOM INTEGRATION
simplex virus thymidine kinase gene (HSV-\textit{tk}). Cells expressing HSV-\textit{tk} can be selectively killed using the nucleoside analogue gancyclovir, which is specific for the viral enzyme. Overall, the effect is selection for cells in which a targeting event has occurred. A replacement vector is used, which contains homologous DNA sequences with the positive marker gene inserted, and the HSV-\textit{tk} gene adjacent to the region of homology. This design makes it probable that homologous recombination will result in loss of the HSV-\textit{tk} gene thus rendering cells resistant to gancyclovir, while also becoming selectable for the inserted marker and mutated in the targeted gene. However, should random integration occur, the HSV-\textit{tk} gene will be retained since most random integrations of linear DNA molecules occur via the ends and therefore the cells can be selectively killed. Other methods of enrichment, such as insertional activation of marker genes, have also been used.

Following selection, cells are screened for targeted alterations by procedures such as Southern blotting or PCR, where one primer is specific for the alteration and the other for the target locus. A PCR product will therefore only be seen if the gene has been correctly modified.

Once the alteration has been made, targeted ES cells are injected into a recipient blastocyst which is surgically implanted into a foster mother (figure 1.7). ES cells are isolated from the inner cell mass of a blastocyst and are capable of differentiating to form all different cell types. Thus, in the resulting animal targeted ES cells, because they are pluripotent, should have contributed to all the body tissues. The mouse is therefore called a chimaera since its tissues are composed of donor and recipient cells. If the injected cells have also contributed to the germ line some gametes will carry the altered gene and it is possible, by breeding, to obtain an animal homozygous for the targeted alteration.

1.6: Project Aim

At the start of this project it was thought that the human disease Blooms syndrome results from a defect at the DNA ligase I locus and initially the aim was to generate a mouse model for Blooms syndrome by introducing a mutation into the mouse DNA ligase I gene. It soon became evident that the DNA ligase I gene itself is not mutated in BS cells and therefore the aim of the project shifted from generating a mouse model for Blooms syndrome, to generating a model for DNA ligase I deficiency.

The initial purpose of the project is to provide information enabling the introduction of a mutation into the mouse DNA ligase I gene in embryonic stem
Figure 1.7: Generation of mice homozygous for a targeted alteration. Once a gene has been targeted in embryonic stem cells, targeted cells are injected into mouse embryos. These are then surgically implanted into a pseudopregnant foster mother. The resulting offspring should be chimaeras since embryonic stem cells are pluripotent and can therefore potentially contribute to all tissues in the body. In this example the quality of the chimaeras is assessed by coat colour, using embryonic stem cells derived from mice with a dark coat, and recipient embryos which will develop a light coat. A chimaera is therefore recognizable by patches of light and dark coat colour. Chimaeras are then bred and, if the donor ES cells have contributed to the germ line of the chimaera, the targeted alteration should be inherited by some of the offspring, which will have a dark coat.
Isolate Blastocyst

Isolate Inner Cell Mass

Culture ES Cells

Inject Targeted ES Cells Into Host Embryo Gene Targeting

Reimplant Into Foster Mother

Chimaera

Targeted ES Cell Derived Progeny

Host Embryo Derived Progeny
cells, by gene targeting, which will partially impair the activity of the enzyme. The eventual intention is to use the targeted ES cells to create a mouse model for DNA ligase I deficiency.

Mutations have been identified in both alleles of the DNA ligase I gene of a human fibroblast cell line (46BR) displaying a number of abnormalities. Prior to the introduction of one of these mutations into the mouse gene in order to generate DNA ligase I-deficient mice, the aim is to demonstrate that the observed abnormalities of this cell line arise directly from the mutated DNA ligase I. Following this, the structure of the mouse DNA ligase I gene will be investigated in order to identify parts of the gene which will be of use in gene targeting experiments and such experiments will be then be carried out.
CHAPTER 2

MATERIALS AND METHODS
2A Materials

2A.1 Suppliers of Laboratory Reagents

2A.1.1 Restriction Endonucleases and other DNA/RNA Modifying Enzymes
  Boehringer Mannheim plc.
  GIBCO BRL Life Technologies
  New England Biolabs inc.
  Pharmacia LKB Biotechnology

2A.1.2 Standard Laboratory Reagents
  BDH Chemicals
  Calbiochem-Novabiochem Co.
  Fisons Chemicals
  GIBCO BRL Life Technologies
  ICN Flow Ltd.
  Sigma Chemical Co.

2A.1.3 Bacterial Media Reagents
  Becton-Dickinson UK Ltd.
  Difco Laboratories
  Baltimore Bio. Labs

2A.1.4 Mammalian Cell Culture Reagents
  GIBCO BRL Life Technologies
  ICN Flow Ltd.
  Sera-Lab
  Sigma Chemical Co.

2A.1.5 Radioactive Reagents
  Amersham International plc.
  $\alpha^{32}$P-dCTP ~3000 Curies/mmole
  $\alpha^{35}$S-dATP ~600 Curies/mmole
  $[3H]$-hypoxanthine ~7.9 Curies/mmole

2A.1.6 Antibiotics
  ampicillin: Beecham Research Laboratories
  G418: GIBCO BRL Life Technologies
penicillin G: Sigma Chemical Co.
Streptomycin: Sigma Chemical Co.

2A.1.7 Antisera
IBI limited:
mouse anti-FLAG M1 monoclonal antibody (calcium dependent), 3.1µg/µl
and N-terminal FLAG.BAP control protein, 0.33µg/µl
Promega:
alkaline phosphatase-conjugated anti-mouse IgG 1mg/ml

2A.1.8 Human DNA Ligase I cDNA
The cDNA was kindly provided by Dr. Tomas Lindahl (Barnes et al 1990).

2A.1.9 Oligonucleotides
Except for the universal primer (Pharmacia LKB Biotechnology) all
oligonucleotides were synthesised by the Oswel DNA Service at the University of
Edinburgh (Department of Chemistry).

<table>
<thead>
<tr>
<th>NAME</th>
<th>SEQUENCE 5' TO 3'</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>572M</td>
<td>TTTTCTCAGGAGATGTGGC</td>
<td>human DNA ligase I cDNA, 5' end (←)</td>
</tr>
<tr>
<td>573M</td>
<td>CACAGAGCCTCGCTTTGCC</td>
<td>human β-actin promoter (→)</td>
</tr>
<tr>
<td>027N</td>
<td>GATCCGCCGCCCGTCCACAC</td>
<td>human β-actin promoter (→)</td>
</tr>
<tr>
<td>028N</td>
<td>CCGCCAGGATGACGAGGC</td>
<td>cloning junction in pBAneo HL between promoter and cDNA (→)</td>
</tr>
<tr>
<td>482W</td>
<td>ATGGGATGATATTATTTATTCTT</td>
<td>mouse DNA ligase I cDNA, 3' of BamHI site (←)</td>
</tr>
<tr>
<td>483W</td>
<td>GTCCGACTGCAGCGCCAGCATGG ACTACAAGGACGACGATGATA AGCGAGAAGTATTATTATGTCAT</td>
<td>mouse DNA ligase I cDNA, introducing FLAG sequence (→)</td>
</tr>
<tr>
<td>887Y</td>
<td>CTGAAAGAAGACTACCTTGA</td>
<td>human DNA ligase I, exon 23, 5' end (→)</td>
</tr>
<tr>
<td>888Y</td>
<td>CTTGCATATGGCCTGCAGCT</td>
<td>human DNA ligase I, exon 23, 3' end (←)</td>
</tr>
<tr>
<td>G4607</td>
<td>CTCAGAAGCCAGGTACATTG</td>
<td>mouse DNA ligase I, for sequencing</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>G4608</td>
<td>CTGCCAACATGCGAGAAAGT</td>
<td>mouse DNA ligase I, for sequencing</td>
</tr>
<tr>
<td>G8320</td>
<td>TCCTTCACGTGACTGCTCAAG</td>
<td>mouse DNA ligase I, for sequencing</td>
</tr>
<tr>
<td>G8321</td>
<td>CCCTATGTGAGGATTGATGG</td>
<td>mouse DNA ligase I, for sequencing</td>
</tr>
<tr>
<td>G8322</td>
<td>CAGGCACTTTCAATGGGATG</td>
<td>mouse DNA ligase I, for sequencing</td>
</tr>
<tr>
<td>H2066</td>
<td>GTTCTCTGTCTTTGTCAGGG</td>
<td>mouse DNA ligase I, for sequencing</td>
</tr>
<tr>
<td>H3123</td>
<td>CAGTTGACGCTTTGGGAAT</td>
<td>mouse DNA ligase I, for sequencing</td>
</tr>
<tr>
<td>M13 reverse</td>
<td>AACAGCTATGACCATG</td>
<td>for sequencing</td>
</tr>
<tr>
<td>universal</td>
<td>GTAAAAACGACGCGACGT</td>
<td>for sequencing</td>
</tr>
</tbody>
</table>

2A.2 Media

2A.2.1 Bacterial Media

Luria Broth (LB):

- Difco Bacto-tryptone, 10g
- Difco bacto-yeast extract, 5g
- NaCl, 5g
- per litre, adjusted to pH 7.2 with NaOH

Luria Agar:

- as LB with 15g per litre Difco agar

Terrific Broth (TB):

- Difco bacto-tryptone, 12g
- Difco bacto-yeast extract, 24g
- glycerol, 4ml
- per 900ml, plus 100ml sterile 0.17M KH₂PO₄/0.72M K₂HPO₄ after autoclaving
M9 Minimal Agar:
- Difco Bacto Agar, 15g
- 10X M9 salts, 100ml
- 20% glucose, 20ml
- 0.1M MgSO₄, 10ml
- 1mg/ml vitamin B1, 10ml
- 0.1M CaCl₂, 10ml

per litre

10X M9 salts:
- Na₂HPO₄, 60g
- KH₂PO₄, 30g
- NaCl, 5g
- NH₄Cl, 10g

BBL Agar:
- Baltimore BioLabs Trypticase, 10g
- Difco Bacto Agar, 10g
- NaCl, 5g

per litre, adjusted to pH 7.0 with NaOH

BBL Top Agar:
- Baltimore BioLabs Trypticase, 10g
- Difco Bacto Agar, 6.5g
- NaCl, 5g
- 10mM MgSO₄

per litre, adjusted to pH 7.0 with NaOH

Antibiotics
When required, ampicillin to a final concentration of 100μg/ml was added immediately prior to use.

2A.2.2 Mammalian Tissue Culture Media
Glasgow Modified Eagles (BHK21) Medium (McPherson, 1962, with modifications by W. House, Medical Research Council Institute of Virology, University of Glasgow, 1964) was supplied by GIBCO BRL Life Technologies as 10X concentrate. 1X medium was supplemented with 1X non-essential amino acids,
1mM sodium pyruvate, 2mM L-Glutamine, 10% foetal calf serum (GIBCO BRL Life Technologies), 50U/ml penicillin and 50µg/ml streptomycin.

2A.3 Bacterial Strains

<table>
<thead>
<tr>
<th>NAME</th>
<th>GENOTYPE</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TG1</td>
<td>F',traD36, lacI&lt;sup&gt;q&lt;/sup&gt;, Δ(lacZ)M15, proA&lt;sup&gt;+&lt;/sup&gt;B&lt;sup&gt;+&lt;/sup&gt;/supE, Δ(hsdM-crB)&lt;sub&gt;5&lt;/sub&gt; (rk&lt;sup&gt;−&lt;/sup&gt;mK&lt;sup&gt;−&lt;/sup&gt;McrB&lt;sup&gt;−&lt;/sup&gt;), thi, Δ(lac-proAB)</td>
<td>Maniatis et al 1982</td>
</tr>
<tr>
<td>XL1-Blue</td>
<td>F', Tn10, proA&lt;sup&gt;+&lt;/sup&gt;B&lt;sup&gt;+&lt;/sup&gt;, lacI&lt;sup&gt;q&lt;/sup&gt;, Δ(lacZ)M15/recA1, endA1, gyrA96, (Nal&lt;sup&gt;+&lt;/sup&gt;), thi, hsdR17, (rk&lt;sup&gt;−&lt;/sup&gt;mK&lt;sup&gt;−&lt;/sup&gt;), supE44, relA1, lac</td>
<td>Bullock et al 1987</td>
</tr>
<tr>
<td>DH5α</td>
<td>supE44, ΔlacU169 (φ80lacZΔM15), hsdR17, RecA1, endA1, gyrA96, thi&lt;sup&gt;−&lt;/sup&gt;-1</td>
<td>Hanahan 1983</td>
</tr>
<tr>
<td>BHB2688</td>
<td>N205, recA, [limm43, 4, clts, b2, red, Eam, Sam/I]</td>
<td>Hohn + Murray 1977</td>
</tr>
<tr>
<td>BHB2690</td>
<td>N205, recA, [limm43, 4, clts, b2, red, Dam, Sam/I]</td>
<td>Hohn + Murray 1977</td>
</tr>
<tr>
<td>K802</td>
<td>F&lt;sup&gt;−&lt;/sup&gt;e14&lt;sup&gt;−&lt;/sup&gt;, (McrA&lt;sup&gt;−&lt;/sup&gt;) lacYI or Δ(lac)6, supE44, galK2, galT22, rfbD1, metB1, mcrB1, hsdR2 (rk&lt;sup&gt;−&lt;/sup&gt;mK&lt;sup&gt;−&lt;/sup&gt;)</td>
<td>Maniatis et al 1982</td>
</tr>
<tr>
<td>NM514</td>
<td>hsdR&lt;sup&gt;−&lt;/sup&gt;, (rk&lt;sup&gt;−&lt;/sup&gt;mK&lt;sup&gt;−&lt;/sup&gt;), Lyc7 λimm434 cl&lt;sup&gt;−&lt;/sup&gt;</td>
<td>obtained from N. Murray, Edinburgh</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>F&lt;sup&gt;−&lt;/sup&gt;, hsdS, gal (DE3)</td>
<td>Studier + Moffatt 1986</td>
</tr>
</tbody>
</table>
### 2A.4 Mammalian Cell Lines

<table>
<thead>
<tr>
<th>NAME</th>
<th>DESCRIPTION</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>E14</td>
<td>wild-type male mouse embryonic stem cell line</td>
<td>Handyside 1989</td>
</tr>
<tr>
<td></td>
<td>isolated from 129/Ola blastocysts</td>
<td></td>
</tr>
<tr>
<td>E14TG2a</td>
<td>HPRT-deficient derivative of E14 obtained by selection in 6-thioguanine</td>
<td>Hooper 1987</td>
</tr>
<tr>
<td>HeLa</td>
<td>human cervical carcinoma cell line</td>
<td>Puck + Marcus 1955</td>
</tr>
<tr>
<td>RJK88</td>
<td>HPRT-deficient Chinese hamster lung fibroblast cell line</td>
<td>Fuscoe 1983</td>
</tr>
<tr>
<td>46BR.1G1</td>
<td>transformed human fibroblast cell line with aberrant DNA ligase I activity</td>
<td>Webster et al 1982</td>
</tr>
<tr>
<td>MRC5V1</td>
<td>transformed wild-type human fibroblast cell line</td>
<td>Huschtscha + Holliday 1983</td>
</tr>
</tbody>
</table>

### 2A.5 Plasmids and Cloning Vectors

<table>
<thead>
<tr>
<th>NAME</th>
<th>DESCRIPTION</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC8</td>
<td>general purpose cloning vector</td>
<td>Norrander 1983</td>
</tr>
<tr>
<td>pBluescriptII KS+</td>
<td>general purpose cloning vector</td>
<td>Thummel 1988</td>
</tr>
<tr>
<td>λEMBL4</td>
<td>replacement vector used to construct genomic library</td>
<td>Frischau et al 1983</td>
</tr>
<tr>
<td>λNM1149</td>
<td>insertion vector used to construct cDNA library</td>
<td>obtained from N. Murray, Edinburgh</td>
</tr>
<tr>
<td>M13mp18 and M13mp19</td>
<td>vectors used to generate single-stranded DNA for sequencing, with polylinkers in opposite orientations to allow both DNA strands to be sequenced</td>
<td>Yanisch-Perron et al (1985)</td>
</tr>
<tr>
<td>pLK444</td>
<td>expression vector utilising the human β-actin promoter, with \textit{neo} driven by the SV40 early promoter</td>
<td>Gunning \textit{et al} 1987</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>pHL</td>
<td>human DNA ligase I cDNA cloned in pBluescript SK-, under the control of a yeast promoter</td>
<td>Barnes \textit{et al} 1990</td>
</tr>
<tr>
<td>pDWM111</td>
<td>HPRT minigene cloned into pUC8</td>
<td>Selfridge \textit{et al} 1992, Magin \textit{et al} 1992</td>
</tr>
<tr>
<td>pHPT5</td>
<td>construct of full-length mouse HPRT cDNA with poly A sequences</td>
<td>Konecki \textit{et al}, 1982</td>
</tr>
</tbody>
</table>

### 2A.6 Buffers

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA:</td>
<td>24 parts chloroform, 1 part isoamylalcohol</td>
</tr>
<tr>
<td>denaturing buffer:</td>
<td>0.5M NaOH, 1.5M NaCl</td>
</tr>
<tr>
<td>HBS:</td>
<td>8% NaCl, 0.37% KCl, 0.126% Na$_2$HPO$_4$ dihydrate, 1% D-glucose, 5% HEPES, pH to 7.2 with NaOH</td>
</tr>
<tr>
<td>10X MOPS:</td>
<td>200mM 3-(N-morpholino) propane-sulphonic acid (MOPS), 50mM sodium acetate, 10mM EDTA, to pH 7.0 with NaOH</td>
</tr>
<tr>
<td>neutralising buffer:</td>
<td>3M NaCl, 0.5M Tris-HCl pH 7.0</td>
</tr>
<tr>
<td>PBS:</td>
<td>8mM K$_2$HPO$_4$, 1.5mM KH$_2$PO$_4$, 150mM NaCl</td>
</tr>
<tr>
<td>PCA:</td>
<td>25 parts redistilled phenol, 24 parts chloroform, 1 part isoamylalcohol</td>
</tr>
<tr>
<td>phage buffer:</td>
<td>50mM Tris-HCl pH 7.5, 100mM NaCl, 10mM MgSO$_4$, 2% gelatine</td>
</tr>
</tbody>
</table>
sample buffer: 20% glycerol, 100mM EDTA, 0.1% bromophenol blue

20X SSC: 3M NaCl, 0.3M tri-sodium citrate pH 7.0

STE: 10mM Tris-HCl pH 8.0, 1mM EDTA, 150mM NaCl

10X TBE: 0.9M Tris-HCl, 0.9M boric acid, 20mM EDTA, pH 8.0

TE: 10mM Tris-HCl pH 8.0, 1mM EDTA

TM: 50mM Tris-HCl pH 8.0, 50mM Tris-HCl pH 8.9, 50mM MgCl₂

2B Methods

2B.1 Bacterial Culture

2B.1.1 Growth of Escherichia coli

Cells were grown at 37°C in LB or TB. Cells grown on L plates supplemented with ampicillin were incubated at 30°C to prevent growth of satellite colonies.

2B.1.2 Storage of E. coli

For short term storage (4-6 weeks) bacteria were streaked onto agar plates and stored at 4°C. Strains TG1 and XL1-blue were kept on minimal agar to maintain the F' plasmid.

For longer term storage, 900μl fresh overnight culture was mixed with 100μl dimethyl sulphoxide (DMSO) and frozen at -70°C. Cells were recovered by scraping the surface of the frozen culture with a sterile inoculating loop and streaking out on an agar plate.

2B.1.3 Transformation of E. coli

Bacteria were transformed with DNA by the method of Mandel and Higa (1970) with modifications by Dagert and Ehrlich (1974). 0.5ml of an overnight culture of the E. coli strain to be transformed was added to 50ml of LB supplemented with 1ml of 1M MgCl₂. Cells were grown with vigorous shaking at 37°C to OD₆₀₀nm=0.2. The cells were chilled on ice for 5min then spun down at 15,000rpm for 15min at 4°C. Cells were resuspended in 20ml ice-cold
transformation buffer (50mM CaCl₂, 10mM Tris-HCl pH 7.5) and incubated for 30min on ice. Cells were re-pelleted at 15,000rpm for 15min at 4°C, resuspended in 2ml ice-cold transformation buffer and incubated overnight on ice (or for a minimum of 2 hours) prior to use.

About 10ng of DNA in a volume of 5-10μl was added to 100μl of competent cells which were then incubated on ice for 30min, followed by heat shocking at 37°C for 5min. 400μl warm LB was added and the cells incubated at 37°C for 1 hour (to allow expression of antibiotic resistance) before plating to dryness on LB agar with ampicillin selection. Plates were then incubated overnight at 30°C.

If appropriate, 30μl of 2% X-Gal (in dimethylformamide) and 20μl of 100mM IPTG were also added per plate to aid identification of recombinants, with recombinant colonies appearing white and non-recombinants appearing blue.

2B.2 Nucleic Acid Isolation
2B.2.1 Small-Scale Preparation of Plasmid DNA

The method used is a modification of the method described by Ish-Horowicz and Burke (1981). Bacteria were grown at 37°C in 5ml LB or TB supplemented with ampicillin. 1.5ml of the culture was transferred to an Eppendorf tube and the cells pelleted by centrifugation for 5min in a microcentrifuge. Cells were resuspended in 300μl of solution P1 (50mM Tris-HCl pH 8.0, 10mM EDTA, 100μg/ml RNase A), then lysed by addition of 300μl of solution P2 (20mM NaOH, 1% SDS) and gentle mixing by inversion. 300μl of solution P3 (2.55M potassium acetate pH 4.8) was added followed by gentle mixing. Precipitated chromosomal DNA, SDS and protein were sedimented by centrifugation for 10min in a microcentrifuge. The supernatant was phenol extracted to remove contaminating proteins, then nucleic acids precipitated with 600μl of isopropanol and spun down for 10min in a microcentrifuge. The pellet was washed twice with 70% ethanol, dried and resuspended in 50μl sterile distilled water and stored at -20°C.

2B.2.2 Large-Scale Preparation of Plasmid DNA

Bacteria were grown for 24 hours by vigorous shaking at 37°C in 80ml of TB supplemented with ampicillin. Cells were pelleted by spinning at 4,000rpm for 20min at 4°C. Cells were then lysed and chromosomal DNA and proteins precipitated as with the small scale plasmid preparation method, but using 10ml of solutions P1, 2 and 3. Following addition of solution P3, the lysed culture was incubated for 15min on ice and the precipitated chromosomal DNA/SDS/protein complex sedimented by centrifugation at 20,000rpm for 30min at 4°C. The
supernatant was removed promptly, filtered and applied to a Qiagen-tip 500 column, previously equilibrated with 10ml of buffer QBT (750mM NaCl pH 7.0, 50mM MOPS, 15% ethanol, 0.15% Triton X-100). After washing twice with 30ml buffer QC (1M NaCl pH 7.0, 50mM MOPS, 15% ethanol), the DNA was eluted with 15ml buffer QF (1.25M NaCl pH 8.2, 50mM MOPS, 15% ethanol). 12ml of isopropanol was added to precipitate the DNA which was then spun down at 4,000rpm for 20min at 4°C. The DNA was washed with 70% ethanol, dried, resuspended in sterile distilled water and stored at -20°C.

2B.2.3 Preparation of Genomic DNA from Mammalian Cells

Genomic DNA was prepared by the method of Pellicer et al (1978). Cultured cells were harvested by scraping with a plastic policeman (Costar) and collected in 10ml PBS. Two to four confluent 64 cm² dishes were used, yielding 10^7-10^8 cells depending on the cell line. Cells were pelleted by centrifugation at 1,300rpm for 5min and then rinsed three times in PBS by pelting and resuspending in 10ml of PBS. Cells were then gently resuspended in hypotonic solution (10mM Tris-HCl pH 8.0, 10mM NaCl, 3mM MgCl₂). The expanded cells were pelleted and resuspended in 10ml hypotonic solution with 0.2% Triton X-100. This caused cell lysis and intact nuclei were pelleted and resuspended in 7.6 ml of 10mM Tris-HCl pH 8.0, 400mM NaCl, 10mM EDTA. 0.4ml 10% SDS and 100μl 30mg/ml proteinase K were added to the lysate which was incubated at 37°C overnight. RNA was degraded by addition of 200μl 10mg/ml RNase A and incubation at 37°C for a further hour. Residual proteins and RNase A were hydrolysed by addition of 100μl 30mg/ml proteinase K and incubation at 37°C for another hour. This was followed by two phenol extractions, by addition of an equal volume of PCA and shaking for 15min. The phases were separated by spinning at 4,000rpm for 15min at 4°C. After the second extraction, the aqueous phase was extracted with an equal volume of CA. DNA was then precipitated by addition of 0.25 volumes of 5M NaCl and 2 volumes of cold 100% ethanol. and recovered by spooling onto a sealed pasteur pipette. The DNA was rinsed by immersion in 70% ethanol, dissolved in 1ml sterile distilled water and stored at 4°C. This method yielded 1-2mg of high molecular weight DNA.

2B.2.4 Preparation of RNA from Mammalian Cells

RNA was prepared using a modification of the methods described by Strohman et al (1977) and MacDonald et al (1987). Cultured cells were harvested and washed in PBS as for preparation of genomic DNA. After the third PBS rinse cells were resuspended in 8ml of 6M guanidine hydrochloride, 10mM DTT, 25mM
EDTA, pH 7.0. The cell suspension was transferred to a glass homogeniser and the cells macerated with 30 strokes of the homogeniser and transferred to a centrifuge tube. The homogeniser was rinsed out to recover residual homogenate with 2ml of the guanidine hydrochloride solution. 1/30 volume of potassium acetate, pH 5.0 and 1/2 volume of cold 100% ethanol were added to the homogenate, mixed, and the solution left at -20°C for at least 4 hours to precipitate the RNA. The RNA was pelleted by centrifugation at 12,000rpm for 20min at 4°C. The supernatant was discarded and the RNA re-precipitated twice more. The resulting pellet was suspended in 3ml of 0.1M Tris-HCl pH 8.9, 0.1M NaCl, 1mM EDTA, 1% SDS. An equal volume of PCA was added and the sample extracted by vigorous shaking for 10min, followed by centrifugation at 4,000rpm for 10min. RNA was precipitated from the aqueous layer by addition of 1/5 volume of 3M sodium acetate, pH 5.0 and 2 volumes of cold 100% ethanol. The RNA was left to precipitate for 4 hours at -20°C and then pelleted by spinning at 15,000rpm for 25min at 4°C. The RNA pellet was resuspended in 0.5ml sterile distilled water and precipitated by addition of 1/10 volume 3M sodium acetate, pH 5.0, 2 volumes cold 100% ethanol and incubation at -70°C for 15min. RNA was pelleted by centrifugation for 10min in a microcentrifuge, washed twice with 70% ethanol, dried, resuspended in 50-100μl of sterile distilled water and stored at -70°C. This method yielded 50-200μg of RNA depending on the cell line used.

Polyadenylated RNA was prepared by incubating total RNA at 65°C for 5min, followed by cooling on ice and application to an oligo (dT)-cellulose column (Pharmacia LKB Biotechnology) according to the manufacturers instructions. Polyadenylated RNA was eluted from the column, reincubated at 65°C for 5min and applied to a second identical column. Following elution from the second column, RNA was precipitated by addition of 0.2 volumes of 2M NaCl and 3 volumes cold 100% ethanol followed by incubation at -20°C for at least two hours. Precipitated RNA was pelleted by centrifugation for 10min at 4°C in a microcentrifuge, the supernatant was discarded, and the polyadenylated RNA was redissolved in 20μl of TE buffer and stored at -70°C.

2B.2.5 Small-Scale Preparation of Bacteriophage M13 Replicative Form DNA

50μl of an overnight culture of E. coli strain TG1 was added to 2ml of LB. To this was added one M13 plaque and the infected culture was grown for 4-4.5 hours at 37°C. 1ml of culture was removed to an Eppendorf tube and bacterial cells pelleted for 5min in a microcentrifuge. The supernatant was tipped off and any remaining removed carefully with a pipette. The pellet was resuspended in 100μl of
solution 1 (50mM glucose, 10mM EDTA, 25mM Tris-HCl pH 8.0), followed by addition of 200μl of solution 2 (0.2M NaOH, 1% SDS) and the tube contents mixed by inversion. 150μl of solution 3 (3M potassium acetate, 11.5% glacial acetic acid) was added, mixed by inversion and cell debris spun down for 5min in a microcentrifuge. The supernatant was phenol extracted and precipitated by addition of 1/10 volume of 3M sodium acetate and 2.5 volumes of cold 100% ethanol. DNA was spun down for 10min in a microcentrifuge, washed in 70% ethanol, resuspended in 25μl TE buffer and stored at -20°C. A typical yield from this method was about 500 ng of M13 DNA.

2B.2.6 Preparation of Single-Stranded Bacteriophage M13 DNA

50μl of an overnight culture of *E. coli* strain TG1 was added to 2ml of LB. One M13 plaque was added and the culture grown for 4-4.5 hours at 37°C. 1.5ml of the culture was transferred to an Eppendorf tube and cells pelleted for 5min in a microcentrifuge. This was repeated with the supernatant. 1.2ml of supernatant was transferred to a fresh tube and phage precipitated by addition of 300μl of 20% PEG 6000, 2.5M NaCl. This was mixed by inversion and left to stand at room temperature for 15min, followed by spinning for 20min in a microcentrifuge. The supernatant was removed and the tube spun again briefly and any remaining solution removed carefully with a pipette. The pellet was resuspended in 300μl TE buffer and phenol extracted. Single-stranded DNA was precipitated from the upper aqueous layer by addition of 1/10 volume 3M sodium acetate and 2.5 volumes of cold 100% ethanol. The DNA was pelleted for 20min in a microcentrifuge, washed with 70% ethanol, resuspended in 12μl of sterile distilled water and stored at -20°C. This yielded 1-2μg of DNA, sufficient for two sets of sequencing reactions.

2B.2.7 Preparation of Bacteriophage λ DNA

An overnight culture of *E. coli* was grown up in 5ml LB supplemented with 50μl of 20% maltose and 50μl of 1M MgSO4. 50μl of the overnight culture and 3-5 bacteriophage plaques were added to 22ml of LB supplemented with 250μl of 20% maltose, 250μl of 1M MgCl2 and 2.5ml of 100mM CaCl2. The culture was grown at 37°C in a conical flask with vigorous shaking for about 9 hours until cell lysis occurred. A few drops of chloroform were added and the lysed culture shaken for a further 10min. Cell debris was spun down at 15,000rpm for 15min at 4°C and the supernatant stored at 4°C. 12ml of supernatant was transferred to an ultracentrifuge tube and 5μl each of DNase I and RNase A (both 2mg/ml) added. The lysate was incubated at 37°C for one hour. Phage were then pelleted by ultracentrifugation at
27,000rpm for 90min at 10°C. The phage pellet was resuspended in 200μl of 0.5M Tris-HCl pH 8.0, transferred to an Eppendorf tube and 200μl of proteinase K (2mg/ml) in 0.5M Tris-HCl pH 8.0 added and left to digest for one hour at 37°C. Phage DNA was purified by addition of an equal volume of PCA, mixing for 5min and spinning for 5min in a microcentrifuge. This was repeated with an equal volume of CA. DNA was precipitated from the upper aqueous layer by addition of 1/10 volume of 3M sodium acetate and 2.5 volumes of cold 100% ethanol. The DNA was washed with 70% ethanol, dried, resuspended in 50-100μl of sterile distilled water and stored at -20°C.

2B.3 Quantification of Nucleic Acids

2B.3.1 Estimation of DNA Concentration

DNA samples were diluted in 1ml of distilled water and the absorbance at 260 and 280nm measured in a spectrophotometer (Perkin-Elmer, Lambda 15, UV/VIS Spectrophotometer). OD\textsubscript{260}=1.0 represents a concentration of 50μg/ml for DNA. The ratio OD\textsubscript{260}/OD\textsubscript{280} gives an estimate of nucleic acid purity. For DNA, a value of around 1.8 indicates a pure preparation.

2B.3.2 Estimation of RNA Concentration

RNA concentrations were measured as for DNA. OD\textsubscript{260}=1.0 represents an RNA concentration of 40μg/ml. OD\textsubscript{260}/OD\textsubscript{280}=2.0 indicates a pure preparation of RNA.

2B.4 DNA Manipulation

2B.4.1 Purification

One volume of chloroform and one volume of redistilled phenol equilibrated in STE buffer were added to a solution of DNA to be purified. The mixture was vortexed for 1min and spun for 5min in a microcentrifuge to separate the phases. An equal volume of chloroform was added to the upper aqueous layer, which was then vortexed and centrifuged for 2min. DNA was precipitated from the upper aqueous layer by addition of 1/10 volume 3M sodium acetate pH 5.0 and 2.5 volumes cold 100% ethanol. DNA was pelleted by spinning for 10min in a microcentrifuge, washed with 70% ethanol, dried and resuspended in an appropriate volume of sterile distilled water.
2B.4.2 Restriction of DNA with Endonucleases

DNA was digested with about 5 units of endonuclease per µg of DNA, using buffer and temperature conditions recommended by the manufacturer. For double digests involving enzymes with different buffer requirements, reactions were either carried out under buffer conditions intermediate between the requirements of the two enzymes, or DNA was digested in two steps. Reactions were halted by heating at 65°C for 10min, by phenol extraction or by addition of sample buffer.

2B.4.3 Dephosphorylation

Bacterial alkaline phosphatase (BAP) hydrolyses 3' and 5' phosphates from DNA and RNA. It can be used for dephosphorylating vectors prior to insert ligation to prevent vector recircularisation and increase the frequency of recombinants. About 1ng of DNA in 90µl of sterile distilled water was mixed with 10µl of 10X BAP buffer (0.65M Tris-HCl pH 7.5, 95mM MgCl2, 50mM DTT) and 66 units of BAP. The reaction mix was incubated for one hour at 65°C, followed by phenol extraction and two sodium acetate/ethanol precipitations to remove all traces of the enzyme.

2B.4.4 Filling-In Recessed 3' DNA Termini

5' overhanging DNA ends generated by restriction endonucleases were converted to blunt ends by "filling-in" with the Klenow fragment of E. coli DNA polymerase I. About 1µg of DNA in 39µl of sterile distilled water was mixed with 5µl of nick translation buffer (500mM Tris-HCl pH 7.5, 100mM MgCl2, 10mM DTT, 500µg/ml BSA), 5µl 5mM dNTPs and 6 units of Klenow. The reaction mix was incubated at room temperature for one hour and the reaction halted by phenol extraction. Klenow was also used to remove 3' overhanging ends by means of a lesser exonuclease activity, under the same reaction conditions.

2B.4.5 Ligation

Ligations were carried out in a 10µl reaction volume containing vector and insert DNA, 50mM Tris-HCl pH 7.6, 10mM MgCl2, 1mM DTT, 1mM ATP, 5% PEG 8000 and 1 unit of T4 DNA ligase. Reactions were incubated for 1-3 hours at room temperature or overnight at 15°C. Ligation products were then used to transform E. coli.
2B.4.6 Addition of Linkers

500µg of linkers were dried and resuspended in 3µl of TM buffer in a microcentrifuge tube. The tube was heated in a waterbath for 15min at 80°C then transferred to a 37°C waterbath for 15min to allow the linkers to anneal.

The linkers were then kinased for 1.5 hours at 37°C by mixing 2µl of annealed linkers with 1µl of linker kinase buffer (2M Tris-HCl pH 7.5, 1M MgCl₂, 1M DTT, 50mM ATP, 10 mg/ml BSA), 10 units of T4 kinase and 6µl of sterile distilled water.

10µl of kinased linkers were then ligated to 7µl of blunt-ended fragment in the presence of 1µl of linker kinase buffer and 2 units of T4 DNA ligase.

2B.5 Electrophoresis of Nucleic Acids

2B.5.1 Electrophoresis of DNA in Agarose Gels

DNA fragments were separated in 0.8-2% BRL electrophoresis grade agarose with 0.5µg/ml ethidium bromide in 1X TBE buffer. DNA samples were mixed with 1/10 volume sample buffer. Electrophoresis was carried out horizontally at 70-100V. Bacteriophage λ DNA cut with HindIII and φX174 replicative form DNA cut with HaeIII or 100bp ladder (GIBCO BRL Life Technologies) were used as size markers. DNA was visualised by ultra-violet illumination.

2B.5.2 Electrophoresis of RNA in Agarose Gels

RNA samples were run out on denaturing agarose gels. 1.4% agarose gels were made up in 1X MOPS, 0.66M formaldehyde and 0.5µg/ml ethidium bromide. 20µg of total RNA in 20µl of sterile distilled water was added to an equal volume of formamide sample buffer (2.3X MOPS, 12mM sodium acetate pH7.0, 50% deionised formamide, 11% formaldehyde) and 1/4 volume of sample buffer. Samples were heated for 5min at 65°C and snap chilled on ice. Samples were then loaded onto the gel and run in 1X MOPS at 100V for 3-4 hours.

2B.5.3 Recovery of DNA from Agarose Gels

If the DNA was to be recovered by phenol extraction it was electrophoresed through low melting point agarose and cut out of the gel. An equal volume of 1X TBE, 0.2M NaCl was added and the gel melted at 65°C. 3/4 volume redistilled phenol equilibrated in 1X TBE, 0.1M NaCl was added, the mixture vortexed and then spun for 5min in a microcentrifuge. The phenol extraction was repeated with the upper aqueous layer. An equal volume of butanol was added, the solution vortexed and spun for 5min in a microcentrifuge. The lower aqueous layer was
precipitated with 1/10 volume of 3M sodium acetate pH 5.0 and 2 volumes of cold 100% ethanol, spun for 10min in a microcentrifuge and the pellet washed twice with 70% ethanol, dried and resuspended in a suitable volume of sterile distilled water.

For DNA fragments larger than 500bp, DNA was run through a regular agarose gel and purified using Geneclean (Bio101 Inc.). 0.5 volume of "TBE modifier" (a concentrated salt solution) and 4.5 volumes of 6M NaI were added to the gel slice. The agarose was dissolved by heating at 55°C, then cooled on ice for 5min. An appropriate amount (5μl/5μg DNA and 2μl/μg DNA thereafter) of "glassmilk" (a silica matrix suspended in water) was added and incubated on ice for 5min to allow the DNA to bind to the matrix. The "glassmilk" was pelleted by centrifugation for 5s in a microcentrifuge and the supernatant discarded. The pellet was washed three times in "NEW WASH" (NaCl, ethanol, water mix), by resuspending in 250μl of the "NEW WASH" and pelleting. The DNA was eluted from the "glassmilk" by two cycles of resuspension in sterile distilled water (10μl/5μl "glassmilk") and heating at 55°C, followed by spinning for 30s and removal of the supernatant. The DNA-containing supernatant was stored at -20°C.

2B.6 Transfer of Nucleic Acids from Agarose Gels to Membranes

2B.6.1 DNA Transfer

This procedure was originally developed by Southern (1975) and modified by Smith and Summers (1980). 1-10μg of genomic DNA was cut with an appropriate restriction enzyme for at least 6 hours to ensure complete digestion, then the resulting fragments were separated on a 0.8% agarose gel. DNA was electrophoresed overnight at 45V, then photographed. The gel was soaked in denaturation buffer with gentle agitation for 30min. The DNA was transferred to Genescreen Plus nylon membrane (Du Pont) by capillary action using denaturation buffer as the transfer medium (Reed and Mann 1985). A wick made of wet blotting papers was placed on a platform with both ends of the blotting paper immersed in denaturing buffer in a reservoir underneath the platform. The gel was placed on top of the wick and a sheet of membrane (previously equilibrated in denaturing buffer) cut to the same size as the gel was placed on top of the gel. The edges were sealed with plastic films. Three sheets of blotting paper soaked in distilled water were laid on top of the membrane, followed by a stack of dry paper towels. A glass plate was placed on top of the paper towels and a weight laid on top. Transfer was allowed to continue for 12-48 hours after which the membrane was neutralised for 30min in neutralising buffer and air dried before use.
2B.6.2 RNA Transfer

Following electrophoresis the gel was photographed then soaked for two 20min periods in 10X SSC with gentle agitation. Transfer was as for DNA, except that 10X SSC was used as the transfer medium. When transfer was complete, the membrane was washed in 2X SSC and baked for 2 hours at 80°C.

2B.7 Nucleic Acid Hybridisation
2B.7.1 Labelling DNA by Random Priming with Hexadeoxy-ribonucleotide Primers

DNA labelled to high activities was obtained using the randomly primed DNA labelling method (Feinberg and Vogelstein 1983). This method allows DNA available only in small amounts to be labelled. It is based on the hybridisation of a mixture of hexanucleotides to the DNA to be labelled. Many sequence combinations are represented in the hexanucleotide primer mixture, which leads to binding of primer to template in a statistical manner. The complementary strand is synthesised from the 3’ hydroxyl termini of the hexanucleotide primers using Klenow enzyme, during which radiolabelled dNTP is incorporated into the newly synthesised DNA strand.

OLB buffer containing nucleotides and random primers is required. It is prepared by mixing 50µl solution A (1.25M Tris-HCl pH 8.0, 0.125M MgCl₂, 25mM β-mercaptoethanol, 0.5mM each of dGTP, dATP and dTTP), 125µl of solution B (2M HEPES buffer adjusted to pH 6.6 with NaOH) and 75µl of solution C (random hexanucleotides OD₂₆₀nm=90 in TE buffer, Pharmacia LKB Biotechnology).

DNA probes were purified by electrophoresing through agarose gels twice prior to labelling. 100ng of DNA probe in 32µl sterile distilled water was denatured by boiling for 5min, followed by incubation at 37°C for 10min. 12µl of OLB buffer, 2µl of 10mg/ml BSA, 5µl of α-³²P-dCTP (50 µCi) and 2 units of Klenow were added to the denatured DNA at room temperature. Incubation was continued for 4 hours at room temperature, then unincorporated nucleotides were separated from labelled DNA by chromatography on a Sephadex G-50 column.

2B.7.2 Separation of Unincorporated Nucleotides from labelled DNA

A Sephadex G-50 column was used to remove unincorporated α-³²P-dCTP from labelled DNA probe (Sambrook 1989). The column was prepared by packing a 1ml blue tip with Sephadex G-50 suspended in sterile distilled water. The Sephadex was washed with 10µl of denatured herring sperm DNA (10mg/ml) by centrifugation.
for 5 min at 900 rpm. The labelling mix was then applied to the column and eluted into an Eppendorf tube by centrifugation. 450 µl of herring sperm DNA (10 mg/ml) was added to the purified probe which was then boiled for 5 min to denature the DNA prior to hybridisation.

**2B.7.3 Hybridisation**

Filters were prehybridised in 30-50 ml of 6X SSC, 1% SDS, 10% dextran sulphate and 100 µg/ml of denatured herring sperm DNA (10 mg/ml) in a hybridisation oven (Hybaid Limited) for two hours at 65°C.

Hybridisation was performed by adding the denatured probe to the prehybridisation mixture and incubating in the hybridisation oven overnight at 65°C for Southern filters or at 60°C for Northern filters.

Non-specifically bound DNA molecules were removed by washing the filter, twice for 5 min in 2X SSC at room temperature, twice for 30 min in 2X SSC, 1% SDS at 65°C or 60°C and twice for 10 min in 0.1X SSC at room temperature. Radioactive DNA molecules bound to the filter were then visualised by autoradiography.

**2B.7.4 Autoradiography**

Following the washing procedure filters were sealed in a plastic bag to prevent drying out. Autoradiography was performed using X-OMAT AR (Kodak) or Cronex (Du Pont) X-ray film in a cassette with intensifying screens (Cronex Lightning Plus, Du Pont). Cassettes were stored at -70°C during exposure to slow the reversal of activated bromide crystals to their stable form, to give a sharper signal.

**2B.7.5 Stripping Probes from Filters**

Radiolabelled DNA probes hybridised to single-stranded nucleic acid immobilised on nylon were removed by the method described in the manual supplied with Genescreen Plus membranes. Southern filters were boiled three times for 10 min in 0.1X SSC, 1% SDS. Northern filters were washed five times for 3 min in hot 0.1X SSC, 0.01% SDS. Stripped filters were autoradiographed to confirm that deprobing was complete. Filters were then incubated in prehybridisation solution and hybridised as usual.
2B.8 cDNA and Genomic Libraries

2B.8.1 cDNA Synthesis

Synthesis of cDNA was performed using a You-Prime cDNA Synthesis Kit (Pharmacia LKB Biotechnology). Mouse (E14) polyadenylated RNA was denatured for 10 min at 65°C, then chilled on ice. Murine reverse transcriptase catalysed oligo dT-primed first strand synthesis for one hour at 37°C in the presence of RNAguard, RNase/DNase-free BSA and dNTPs. The second DNA strand was synthesised by incubation in the presence of RNase H, Klenow and dNTPs for one hour at 12°C, followed by one hour at 22°C. RNase H nicks the RNA strand of the RNA/DNA duplex, allowing the Klenow to replace the RNA with DNA by nick translation. Pure Klenow was then added for 30 min at 37°C to fill in any overhangs and ensure all cDNA ends were blunt. The cDNA was then phenol extracted and purified on a Sephacryl S-300 spun column. EcoRI/NotI adaptors were ligated on to the cDNA ends overnight at 15°C and the EcoRI-ended cDNA phosphorylated with T4 Polynucleotide Kinase for one hour at 37°C. The cDNA was again phenol extracted and purified on a spun column to remove excess adaptors. This method produced cDNA up to about 7 kb, with most about 4 kb in size.

2B.8.2 Preparation of Packaging Extracts for in vitro Packaging of Bacteriophage λ

Packaging extracts were prepared by the method of Hohn and Murray (1977), with modifications by Scalenghe et al (1981). There are two components to the packaging extracts, a freeze thaw lysate and a sonicated extract, both prepared from induced phage lysogens. The sonicated extract was prepared from E. coli BHB2690, a strain unable to synthesise D protein which is required during maturation of the phage capsid. The freeze thaw lysate was prepared from E. coli BHB2688 which cannot synthesise phage proheads, but accumulates D protein. Neither extract can package phage DNA in isolation, but the mixed extracts can.

The freeze thaw lysate was prepared by inoculating 10 ml of an overnight culture of BHB2688 into 3X two litre flasks containing 500 ml of LB and growing to OD₆₀₀nm = 0.3 at 30°C while shaking vigorously (about 2 hours). Cells were then induced by standing the flasks in a 45°C waterbath for 15 min followed by further propagation at 37°C for 2 hr with vigorous shaking. Cells were spun down at 7,500 rpm for 10 min at 4°C and the supernatant discarded. Cells were resuspended in 3 ml of ice cold freeze thaw buffer (10% sucrose, 50 mM Tris-HCl pH 7.5) and 150 µl of lysozyme (2 mg/ml in 250 mM Tris-HCl pH 7.5) added, then mixed gently by inversion. The solution was quick frozen by immersion in liquid nitrogen. The lysate
was then thawed at room temperature and transferred to ice immediately on thawing. 150μl of buffer M1 (6μl 500mM Tris-HCl pH7.5, 300μl 500mM spermidine/100mM putrescine pH 7.0, 9μl 1M MgCl₂, 75μl 10mM ATP, 1μl 14M β-mercaptoethanol) was added. Cell proteins were partially purified by ultracentrifugation at 40,000rpm for 1 hour at 4°C. 30μl aliquots of the supernatant were dispensed into pre-cooled Eppendorf tubes and snap frozen by immersion of the tube into liquid nitrogen. Aliquots of freeze thaw lysate were stored at -70°C.

To prepare the sonicated extract, E. coli BHB2690 was grown, induced and pelleted as for the freeze thaw lysate. The supernatant was discarded and the pellet resuspended in 5.6ml buffer A (20mM Tris-HCl pH 8.0, 3mM MgCl₂, 0.5% β-mercaptoethanol, 1mM EDTA). The suspension was sonicated in 3 second bursts, with 60 second cooling on ice in between to prevent the temperature rising above 4°C. Sonication was continued until the solution was clear, then the cell debris was pelleted by centrifugation at 10,000rpm for 10min at 4°C. 1/6 volume buffer M1 was added to the supernatant and 50μl aliquots were distributed into pre-chilled Eppendorfs, snap-frozen in liquid nitrogen and stored at -70°C.

2B.8.3 Library Construction

5μg of λ DNA (λNM1149 and λEMBL4 for cDNA and genomic libraries respectively) and 10μg of insert DNA (cDNA or EcoRI-digested genomic DNA) were digested to completion, mixed and ligated in a total volume of 20μl. The ligation mix was packaged in vitro by mixing 7μl buffer A, 1-5μl ligation mix, 1μl buffer M1, 10μl sonicated extract and 12μl freeze thaw lysate (see section 2B.8.2). Packaging was allowed to continue at room temperature for one hour, then 500μl phage buffer was added and the phage plated out on a suitable strain of E. coli (strain NM514 was used for the cDNA library, K802 was used for the genomic library).

2B.8.4 Plating out Bacteriophage λ to Generate Plaques

50ml LB supplemented with 1ml of 20% maltose and 0.5ml 1M MgSO₄ was inoculated with a single colony of E. coli and the cells allowed to grow to log phase (OD₆₀₀nm=0.3-0.5). Cells were spun down for 15min at 1,500 rpm and resuspended in 5ml 10mM MgSO₄. Phage were added to 100μl of cells and left at room temperature for 15min to allow phage adsorption. 3ml molten BBL top agar (47°C) was added to the cells and phage, then poured over dried BBL plates, followed by incubation at 37°C until plaques were visible (12-16 hours).
2B.8.5 Probing Libraries

Phage were plated out and the plates chilled to 4°C. Nylon filters (Hybond-N, Amersham International plc.) were placed on the surface of the agar and the phage allowed to adsorb for 5min. Filters were oriented on the plates by piercing through the membrane and agar with a needle in an asymmetric pattern. Filters were then lifted off without removing the top agar and placed, DNA side up, on filter paper soaked in denaturing buffer for 5min. They were then transferred to neutralising buffer for 5min, followed by a 5min wash in 2X SSC. Filters were then air dried and baked for 10-20min at 120°C.

Filters were prehybridised for one hour at 65°C in 7% SDS, 0.5M NaHPO₄ (0.5M Na₂HPO₄ to pH 7.0 with 0.3M NaH₂PO₄) in a hybridisation oven (Hybaid Limited). Radiolabelled probe was then added and allowed to hybridise overnight at 65°C. Filters were then rinsed twice in 5% SDS, 0.04M NaHPO₄ at room temperature and twice in 1% SDS, 0.04M NaHPO₄ for 5min at 65°C and autoradiographed.

Aided by the orientation marks, plaques corresponding to positive signals were picked and the phage allowed to elute overnight at 4°C in 1ml phage buffer with 50μl chloroform added.

2B.9 DNA Sequencing

2B.9.1 Templates for Sequencing

Single-stranded templates were generated by cloning DNA fragments to be sequenced into M13mp18 or M13mp19. Phage were then plated out and recombinants selected with X-Gal/IPTG. Phage from recombinant white plaques were grown up for DNA preparation (2B.2.6).

Double-stranded templates were DNA inserts cloned in pBluescriptII KS+. DNA was propagated in E. coli strains DH5α or XL1-Blue and prepared by the large-scale method (2B.2.2).

2B.9.2 Annealing Primer to Template DNA

For single-stranded templates, 10μl of template (1-2μg of DNA) was mixed with 2μl of primer (4μg/ml) and 2μl of annealing buffer (1M Tris-HCl pH 7.6, 100mM MgCl₂, 160mM DTT) and incubated at 70°C for 3min. Following a 5min incubation at room temperature the annealed template and primer was ready for sequencing.

4-8μg of double-stranded plasmid DNA was mixed with 16μl of 1M NaOH and sterile distilled water up to a total volume of 40μl. This was incubated at room
temperature for 10 min, then 3μl of sterile distilled water, 7μl of 3M sodium acetate pH 5.0 and 120μl of cold 100% ethanol were added to precipitate the denatured DNA. DNA was pelleted for 10 min in a microcentrifuge, washed with 100μl of 70% ethanol and dried briefly. The DNA was redissolved in 10μl of sterile distilled water, and 2μl of primer (2μg/ml) and 2μl of annealing buffer added. The mixture was incubated at 65°C for 5 min, 37°C for 10 min and room temperature for 5 min, after which it was ready to proceed with sequencing.

2B.9.3 Sequencing

Sequencing by the dideoxy chain termination method was carried out using the T7 Sequencing Kit (Pharmacia LKB Biotechnology). Reactions were the same for both single and double-stranded templates.

14 μl of annealed template/primer was mixed with 3μl of dATP labelling mix (1.375μM each of dGTP, dTTP, dCTP and 333.5mM NaCl), 1μl of α-35S-dATP and 2μl of T7 DNA polymerase (1U/μl diluted in 20mM Tris-HCl pH 7.5, 5mM DTT, 100μg/ml BSA, 5% glycerol). The mixture was incubated for 5 min at room temperature to allow the synthesis of short labelled DNA molecules.

Meanwhile, 2.5μl of termination mixes were dispensed into Eppendorf tubes and prewarmed to 37°C. Termination mixes were chosen to allow the reading of sequence from the primer to position 500 ("short") or from position 50-800 ("long").

| TERMINATION MIXES | 
|-------------------|---------------|
| 'A' Mix - Short    | 840μM each of dCTP, dGTP and dTTP, 93.5μM dATP, 14μM ddATP, 40mM Tris-HCl pH 7.6, 50mM NaCl |
| 'C' Mix - Short    | 840μM each of dATP, dGTP and dTTP, 93.5μM dCTP, 17μM ddCTP, 40mM Tris-HCl pH 7.6, 50mM NaCl |
| 'G' Mix - Short    | 840μM each of dATP, dCTP and dTTP, 93.5μM dGTP, 14μM ddGTP, 40mM Tris-HCl pH 7.6, 50mM NaCl |
| 'T' Mix - Short    | 840μM each of dATP, dCTP and dGTP, 93.5μM dTTP, 14μM ddTTP, 40mM Tris-HCl pH 7.6, 50mM NaCl |
| 'A' Mix - Long     | 840μM each of dCTP, dGTP and dTTP, 93.5μM dATP, 2.1μM ddATP, 40mM Tris-HCl pH 7.6, 50mM NaCl |
Following the labelling reaction, 4.5μl of reaction mix was added to each of the four pre-warmed termination mixes and incubated for 5min at 37°C. Reactions were stopped by addition of 5μl of sample buffer (0.3% Bromophenol Blue, 0.3% Xylene Cyanol FF, 10mM EDTA pH 7.5, 97.5% deionised formamide). Prior to loading on a gel, samples were heated for 2min at 80°C.

2B.9.4 Electrophoresis of Sequenced DNA

DNA was electrophoresed through a 0.2mm thick 0.8% polyacrylamide gel using a Bio-Rad Sequi-Gen Sequencing Cell. Acrylamide solution was made up with 50g of urea, 30ml of distilled water, 10ml of 10X TBE buffer and 20ml of 40% acrylamide/2.105% bisacrylamide (Easigel, Scotlab). Gels were polymerised by addition of 640μl of 1% ammonium persulphate and 30μl of TEMED. Gels were typically run at 50-60W to maintain the gel at a temperature of 55°C during the run. Gels were transferred to filter paper, covered with cling film and dried for two hours under vacuum at 80°C. Ladders of bands were visualised by autoradiography.

2B.10 Amplification of DNA by the Polymerase Chain Reaction

2B.10.1 Polymerase Chain Reaction Experiments

A typical PCR reaction consisted of 1-10ng of plasmid DNA or 100ng of genomic DNA, with 1μl of 5mM dNTPs, 5μl of 10X PCR buffer (500mM KCl, 15mM MgCl₂, 0.1% gelatin, 0.1M Tris-HCl pH 8.5, 4.5% Triton X-100, 4.5% Tween 20), 0.5μl of Taq DNA polymerase, 100ng of each primer, and sterile distilled water up to a final volume of 50μl. Conditions varied for different primers, but denaturation was generally carried out at 94°C for 1min, followed by appropriate annealing, and synthesis at 72°C for 30s to 1min depending on the size of the desired product. 35 cycles of these conditions were used in each experiment. All
experiments were carried out using a Dri-Block cycler (Techne PHC-2). 10-20μl of each reaction was then electrophoresed through an agarose gel to visualise products.

2B.10.2 Screening Colonies using the Polymerase Chain Reaction

PCR was performed on genomic DNA from cultured mammalian cells without the need for a lengthy DNA preparation to be carried out. This was done by the method of McMahon and Bradley (1990). Half a colony (about 1,000 cells) was spun down in a microcentrifuge for 30s and the medium removed using a pipette. The cell pellet was resuspended in 1X PCR buffer with proteinase K (0.1mg/ml) and incubated at 65°C for two hours. The proteinase K was inactivated by heating at 90°C for 15min after which the samples were ready to be used in PCR reactions. 10-20μl of sample was used per reaction.

2B.11 Cell Culture

2B.11.1 Growth of Mammalian Cells

Mammalian cells were maintained in Glasgows modified Eagles medium supplemented with 1X non-essential amino acids, 1mM sodium pyruvate, 2mM L-glutamine, 10% foetal calf serum, 50U/ml penicillin and 50μg/ml streptomycin. Cultures were incubated at 37°C in 5% CO2 in a humidified incubator.

2B.11.2 Electroporation

Prior to electroporation plasmid DNA was phenol extracted, precipitated and resuspended in sterile distilled water. 30μg of DNA in 100μl of sterile distilled water was added to 5X10^5 cells in 0.8ml of serum-free 1X medium. Cells were pulsed at 250V, 500μF (Gene Pulse apparatus, Biorad), incubated at room temperature for 10min and plated out between two 100mm dishes. G418 selection (500μg/ml) was applied to the cells two days later.

2B.11.3 Calcium Phosphate/DNA Precipitation and Colony-Forming Assay

RJK88 cells were plated out in 60mm dishes at a density of 1.2 X10^5 cells/dish on the day prior to transfection. To 10-20μg of DNA was added 1ml of HBS and 62μl of 2M CaCl2. The CaCl2 was added slowly, dropwise, and the solution left to precipitate for 45min at room temperature. The precipitate was then added to the cells and left at room temperature for 20min, with frequent agitation. 10ml of medium was added to the dishes and the cells incubated overnight. The medium was then removed and replaced with medium containing HAT (1mM hypoxanthine, 10μM aminopterin, 100μM thymidine). Two days later the medium was replaced.
with fresh HAT medium and the cells left to form colonies. After one week colonies were fixed (Carnoy's Fixative, 3 parts methanol, 1 part glacial acetic acid) and stained (10% crystal violet).

2B.11.4 Ethyl Methane Sulphonate Survival Assay

Cells were plated out in 30mm dishes using 5X10⁴ cells per dish for MRC5V1 and 2.5X10⁵ cells for 46BR.1G1 and its transfected derivatives. Cells were incubated overnight, then exposed to 0-15mM EMS in complete medium for two hours. Each dose point was carried out in duplicate. Cells were then rinsed twice in warm PBS and incubated in complete medium for five days. Surviving cells were counted using a [³H]-hypoxanthine incorporation assay (Cleaver 1989). Cells were washed twice in serum-free medium and incubated in serum-free medium with 0.5μCi/ml [³H]-hypoxanthine for 4-5 hours. Cells were washed twice in PBS and lysed with 1% SDS. The lysate was collected and RNA and DNA precipitated by addition of 0.8 volumes of 20% trichloroacetic acid. The precipitate was collected onto Whatman GF/C filters and the incorporation measured by scintillation counting. The percentage survival was calculated as (mean growth in EMS/mean growth in control medium) X100.

2B.11.5 3-Aminobenzamide Survival Assay

Cells were plated out in 60mm dishes using 200 cells per dish for MRC5V1 and 1000 cells per dish for 46BR.1G1 and its transfected derivatives. Cells were incubated overnight then exposed to 0-8mM 3AB in complete medium. Each dose point was done in duplicate. Dishes were fixed and stained after 10-15 days. Cell survival was determined by eluting the stain with 5ml 70% ethanol and measuring the absorbance at 595nm. The percentage survival was calculated as (mean plating efficiency in 3AB/mean plating efficiency in control medium) X100.

2B.12 Protein Methods

2B.12.1 Expression of Proteins in E. coli

Sequences cloned in pBluescriptII KS+ under control of the T7 promoter were expressed in E. coli strain BL21(DE3) which is a λ DE3 lysogen. DE3 is a λ derivative, carrying a DNA fragment containing a single copy of the gene encoding T7 RNA polymerase under control of the lacUV5 promoter. Addition of IPTG induces the lacUV5 promoter to produce T7 RNA polymerase. Following induction, any genes present within the cell under control of a T7 promoter will be transcribed.
Cells were grown and induced by a modification of the method of Studier and Moffatt (1986). Cells were grown overnight in 5ml of LB supplemented with ampicillin (where appropriate) and 1/2ml of fresh overnight culture was diluted into 5ml of LB supplemented with ampicillin (as appropriate). After growth for 30min, IPTG was added to a final concentration of 2mM. 1ml aliquots of cells were taken at 0, 1, 2 and 3 hours following induction. Cells were spun down for 5min in a microcentrifuge and resuspended in 100-200µl of protein sample buffer (25% stacking gel buffer [see later], 15% glycerol, 3% SDS, 20mM DTT, 0.1% bromophenol blue). The samples were denatured by boiling for 10min, then the cell extracts were electrophoresed through a polyacrylamide gel.

2B.12.2 Preparation of Protein Extracts from Mammalian Cells

6×10⁶ cells were pelleted by centrifugation for 5min at 1300rpm, washed in PBS and repelleted. The cell pellet was then resuspended in 200µl of protein sample buffer (see above) and denatured by boiling for 10min. 50µl of protein extract was loaded onto each gel.

2B.12.3 Electrophoresis of Proteins

Molecular weight markers were obtained from Sigma Chemical Co. Markers were either low molecular weight, from 14.2kDa to 66kDa, or high molecular weight, from 30kDa to 200kDa.

Gels consisted of a stacking gel upon a 7.5% separating gel. The separating gel was poured first and was made up with 10ml of lower Tris (1.5M Tris-HCl, 0.4% SDS, pH 8.8), 7.5ml of 40% acrylamide/2.105% bisacrylamide (Easigel, Scotlab) and 22.5ml of distilled water, then polymerised by addition of 120µl of 10% ammonium persulphate and 60µl of TEMED. The stacking gel was made up with 5ml of upper Tris (0.5M Tris-HCl, 0.4% SDS, pH 6.8), 2.2ml of acrylamide solution and 12.8ml of distilled water, then polymerised by addition of 40µl of 10% ammonium persulphate and 80µl of TEMED.

Gels were run in 1X electrode buffer (10X electrode buffer consists of 28g of Tris-HCl, 20g of SDS and 143g of glycine per litre) at 30W until the dye front reached the end of the gel. Gels were then either blotted onto nylon membranes or fixed (6g of trichloroacetic acid, 200ml of methanol and 70ml of ethanol per litre) and stained (1g of Coomassie blue, 50ml of methanol and 50ml of distilled water, diluted 12ml into 500ml of fix). Following staining for 30min, gels were destained (10% glacial acetic acid, 30% isopropanol) to visualise the protein bands. Destained gels were sealed in plastic bags for storage.

73
2B.12.4 Transfer of Proteins from Polyacrylamide Gels to Membranes

Proteins were transferred from gels to Genescreen Plus nylon membrane (Du Pont) by electroblotting using a Trans-Blot SD semi-dry transfer cell (Bio Rad) according to the manufacturer's instructions. Essentially, the gel and membrane cut to the same size were equilibrated in Bjerrum and Schafer-Nielsen transfer buffer (48mM Tris-HCl pH 9.2, 39mM glycine, 20% methanol) for 30min. The unit was then assembled for blotting by placing in the unit, in order, a piece of blotting paper soaked in transfer buffer, the nylon membrane, the gel and another piece of blotting paper soaked in transfer buffer. Proteins were transferred from large gels for 30-60min at 25V with a limiting current of 3mA/cm² and from mini-gels at 15V with a limiting current of 5.5mA/cm².

2B.12.5 Probing Protein Blots with Antibodies

Western blots were blocked for 1 hour in 1% Boehringer blocking agent (Boehringer Mannheim plc), then incubated with 100μl of the primary antibody (M1 mouse anti-FLAG monoclonal antibody, 10μg/ml in PBS) in 50ml of incubation solution (1% blocking agent, 0.3% Tween-20, 1mM CaCl₂ in PBS) for 1 hour. The membrane was washed three times for 5min in washing solution (0.3% Tween-20, 1mM CaCl₂ in PBS) then incubated with 6μl of the secondary antibody (alkaline phosphatase-conjugated anti-mouse IgG, 1mg/ml, Promega) per 50ml of incubation solution. The membrane was washed five times for 5min then blotted dry and sealed in a plastic bag with colour substrate solution. This consisted of 10ml of 100mM Tris-HCl pH 9.5, 100mM NaCl, 50mM MgCl₂ with 45μl of NBT solution and 35μl of X-phosphate solution (both from the nucleic acid detection kit, Boehringer Mannheim plc). Colour was allowed to develop overnight in the dark without agitation.
The human cell line 46BR carries a mutation at each allele of the DNA ligase I gene (Barnes et al 1992b). The first mutation, a change of Glu566 to Lys renders the enzyme completely inactive. The second mutation, Arg771 to Trp leaves the enzyme with some residual activity. Therefore, 46BR cells lack a completely functional DNA ligase I enzyme. However, a direct link between the DNA ligase I mutations and the phenotype of the cells still remains to be demonstrated. This chapter describes demonstration of this link by complementation of the DNA ligase I mutations with wild-type DNA ligase I cDNA. This work was carried out in association with Dr. Nik Somaia and published in *Mutation Research* (294: 51-58). A reprint of the paper is at the back of this thesis.

3.1 Construction of a Vector Expressing Human DNA Ligase I cDNA

An expression vector was constructed for the purpose of expressing wild type human DNA ligase I cDNA in the cell line 46BR. A 4.3kb EcoRI/HindIII fragment was cut out of plasmid pLK444 (Gunning et al 1987) and cloned into pBluescriptII KS+. This fragment contains the human \( \beta \)-actin promoter (3kb), a 78bp 5' untranslated region and the first intron of the human \( \beta \)-actin gene (832bp). The intron is included because it contains an enhancer sequence. A 2.4kb EcoRI fragment containing the bacterial neomycin phosphotransferase gene under control of the mouse metallothionein promoter (Selfridge et al 1992) was cloned into the EcoRI site, upstream of the \( \beta \)-actin promoter. This was introduced to provide selection for transfected cells, by acquisition of resistance to the synthetic antibiotic G418. Because both ends of this fragment had EcoRI overhangs it could ligate into the vector facing either way, but an orientation was selected so that transcription would proceed in the same direction from both promoters. This completed construction of the expression vector which was designated p\( \beta \)Aneo (figure 3.1). There is no polyadenylation signal built into the vector which therefore relies on cloned sequences to provide one.

Wild type human DNA ligase I cDNA was isolated from pHL (Barnes et al 1990) as a 3.2kb KpnI fragment. This fragment contains a virtually intact cDNA, complete with poly A tail and lacking only about 90 bases from the 5' untranslated region. It was isolated from a cDNA library by complementation of *S. cerevisiae* DNA ligase-deficient cdc9 mutants and therefore encodes a functional protein. XhoI linkers were attached to the ends of this fragment which was then ligated into the SalI site of p\( \beta \)Aneo. Again the insert was clonable in either direction and the correct orientation was selected so that the 5' end of the cDNA was positioned adjacent to the \( \beta \)-actin promoter. This plasmid was designated p\( \beta \)AHLneo.
Figure 3.1: Expression vector pβAHLneo. Human DNA ligase I cDNA (open box) was placed under the control of the human β-actin promoter (heavy shading) which includes the 5' untranslated region (UTR, closed box) and first intron (IVS1, light shading). The bacterial neomycin phosphotransferase gene driven by the mouse metallothionein promoter (neo, open box) lies upstream. Arrows indicate the direction of transcription. Key restriction sites are shown. EcoRI sites generating the diagnostic 1.5kb DNA ligase I cDNA fragment are also shown.
pβAHLneo

EcoRI  EcoRI  SalI/XhoI  EcoRI  EcoRI  XhoI/SalI

neo  human B-actin promoter UTR IVS I  human DNA ligase I cDNA

78bp  832bp  1.5kb

2.4kb  4.3kb  3.2kb
3.2 Transfection

Prior to transfection into cultured cells, plasmids pBAneo (a negative control) and pBAHLneo were linearised by restriction with Scal which has a recognition site within the ampicillin resistance gene of pBluescript. Random integration into the genome is likely to occur via DNA ends therefore linearisation at this position maximises the probability that the neoT gene and the cDNA will be cotransfected.

Both plasmids were transfected into the human fibroblast cell lines MRC5V1 and 46BR.1G1 by electroporation. MRC5V1 is an SV40-transformed wild-type cell line, while 46BR.1G1 is derived from the DNA ligase I-deficient cell line 46BR by transformation with pSV3gpt. This 46BR derivative is homozygous or hemizygous for the DNA ligase I mutation at Arg771, since the other allele with the Glu566 to Lys mutation was lost during transformation (Barnes et al 1992b). 20 to 30 G418-resistant colonies were obtained from 5X10^5 transfected cells and were grown either as pools or separately as clones.

3.3 Detection of DNA Ligase I Transfectants

Genomic DNA was prepared from 46BR.1G1 and pools of 46BR.1G1 transfected with pBAneo and pBAHLneo, as well as six 46BRpBAHLneo clones. The DNA was digested with EcoRI, electrophoresed and blotted. The Southern blot was then probed with human DNA ligase I cDNA (figure 3.2). This revealed seven bands from all cell types, specific to the DNA ligase I gene, of sizes 10.7kb, 7.5kb, 5.3kb, 4.4kb, 3.6kb, 3.3kb and 2.5kb. Three bands arising from the presence of the plasmid in the cells were also expected (figure 3.1). A band of about 4.8kb would result from the cDNA probe hybridising to an EcoRI fragment containing the β-actin promoter and 5' end of the cDNA. A band of this size is clearly present in the 46BRpBAHLneo pool as well as in 46BRpBAHLneo clones two and six. A band of unpredictable size will be derived from the 3' end of the cDNA. Its size depends on the extent of DNA between the plasmid integration site and the next EcoRI site, and there are a number of bands present which could result from this. A 1.5kb band arising from an EcoRI fragment present within the cDNA was used as a marker for the presence of the cDNA within the cells. It is visible in the 46BRpBAHLneo pool and in all 46BRpBAHLneo clones except clone four, which therefore all contain wild type human DNA ligase I cDNA. Presence of the cDNA was undetectable in clone four cells although they must have integrated some plasmid sequence since they are resistant to G418. The 1.5kb band was not visible when the cells were untransfected or when they were transfected with the control plasmid, confirming that it is specific
Figure 3.2: Southern blot of genomic DNA from 46BR.1G1 and transfectants digested with EcoRI and probed with human DNA ligase I cDNA. The diagnostic 1.5kb band is indicated by an arrow. Sizes are given in kb.
for the cDNA. When the intensity of the 1.5kb band is compared to that of the genespecific bands it is apparent that the clones have integrated differing numbers of copies of the cDNA. For example, clones one and six carry more cDNA inserts than clone three.

3.4 Expression of Wild Type Human DNA Ligase I cDNA

Five of the 46BRpβAHLneo clones have wild-type DNA ligase I cDNA sequences integrated into the genome. However it does not follow that the cDNA is being expressed in the cells. Demonstration of expression at the protein level was unlikely to be successful since the endogenous mutant and exogenously introduced wild-type enzymes differ by only one amino acid. Attempts to demonstrate expression were consequently carried out at the messenger RNA level. Plasmid derived messages should be slightly longer than the endogenous 3.2kb mRNA due to the extra 78bp 5' untranslated region from the β-actin promoter. This small size difference was undetectable when RNA was prepared, blotted and probed with full-length human DNA ligase I cDNA (figure 3.3). Only the endogenous 3.2 kb band specific for DNA ligase I was visible. There was also no obvious difference in levels of the 3.2kb DNA ligase I message, therefore cDNA transcription from the plasmid was not significantly increasing the total amount of the 3.2kb mRNA within the cells. However, high molecular weight transcripts in the range 6-9kb were seen, corresponding to the presence of the cDNA within the cells (figure 3.3). The size of these novel transcripts suggests faulty initiation or termination of transcription. It is possible that transcription could be initiating from the mouse metallothionein promoter and continuing right through to the end of the cDNA, and the Northern blot was probed with the neor gene to investigate this possibility. If the high molecular weight DNA ligase I-specific transcripts were initiating from the metallothionein promoter they would include neo sequences and would be detectable with the neo probe. However, only messages arising from transcription of the neor gene itself were detectable. The neo probe did not visibly hybridise to any high molecular weight transcripts (data not shown). The lack of hybridisation between the transcripts and neo therefore indicates that ligase transcript initiation is not occurring from the metallothionein promoter.

A different approach was tried, using the 78bp untranslated region as a probe, but this probe showed only non-specific binding to the ribosomal RNA subunits (data not shown).

The 78bp 5' untranslated region was used to try to demonstrate expression by other means. Twenty base PCR primers specific for this region were designed
Figure 3.3: Northern blot of RNA from MRC5V1, 46BR.1G1 and transfectants probed with human DNA ligase I cDNA. The arrow 'a' marks the position of the endogenous 3.2kb DNA ligase I transcript. Region 'b' marks vector-derived transcripts. The asterisk indicates the position of the 28S ribosomal RNA subunit. Loadings are equal as visualised by ethidium bromide staining of the gel (data not shown).
MRC5V1
46BR
46BR pβAneo
46BR pβAHLneo
pool
clone 1
clone 2
clone 3
clone 4
clone 5
clone 6
(573M, 027N, 028N), to be used in conjunction with a primer specific for the 5' end of the cDNA (572M). Any DNA amplified using these primers could only arise from the plasmid, giving a product of about 1.2kb in size, or from an mRNA arising from transcription from the plasmid, giving a PCR product of about 500bp due to intron splicing. However, due to their high GC content, the β-actin primers showed cross-hybridisation and the PCR experiments failed (data not shown). The presence of the high molecular weight transcripts remained the only demonstration of expression of wild-type human DNA ligase I in the transfected cells.

There appears to be a correlation between the number of cDNA inserts in a cell line and the extent of transcription of the cDNA. For example figure 3.2 shows that 46BRpβAHLneo clone 2 has integrated more copies of the plasmid than the other clones. There is a corresponding increase in the level of plasmid-derived DNA ligase I transcripts (figure 3.3).

3.5 Correction of Ethyl Methane Sulphonate Sensitivity in 46BR

EMS causes damage to DNA by alkylation, which can occur at a number of different reaction sites, both within bases and along the DNA backbone. O6-alkylguanine is one of the major mutagenic reaction products and is removed directly from DNA in mammalian cells by O6-methylguanine methyl transferase (Friedberg 1985 and references therein). Other alkylation damage is removed by the nucleotide excision repair system. 46BR cells are sensitive to EMS, probably due to a defective ligation step during nucleotide excision repair, whereas wild-type cells are relatively insensitive because their DNA repair mechanisms are functional and the alkylation damage can be removed.

Along with MRC5V1, 46BR.1G1, two control clones of 46BRpβAneo and pooled 46BRpβAHLneo cells, 46BRpβAHLneo clones 1, 2 and 6 were selected for use in the EMS survival assay since all three clones express the wild-type cDNA at higher levels than the other clones.

Cells were exposed to EMS at a range of concentrations from 0 to 15mM and the number of cells surviving was determined (figure 3.4). As anticipated, MRC5V1 cells were relatively insensitive, with 30% of cells surviving at the highest dose. 46BR.1G1 and the 46BRpβAneo clones were extremely sensitive with a maximum of 0.4% of cells surviving the highest dose of EMS. The three 46BRpβAHLneo clones expressing wild-type DNA ligase I showed an ability to survive EMS intermediate between that of wild-type MRC5V1 and 46BR, with between 2% and 4% of cells surviving at the highest dose. All three clones were rescued to the same extent and there was no correlation with the number of integrated cDNA copies.
Figure 3.4: Cell survival following EMS treatment of MRC5V1, 46BR.1G1 and transfectants. a, b and c are points used to calculate the extent of rescue. For each point standard errors of the mean were less than 7% of the survival value plotted.

+ MRC5V1
x 46BR.1G1
O 46BRpβAneo clone 1
  46BRpβAneo clone 2
□ 46BRpβAHLneo clone 1
Δ 46BRpβAHLneo clone 2
▲ 46BRpβAHLneo clone 6
The extent of correction was calculated by dose modification at constant survival (Cleaver 1990). This was done by expressing the ratio of the differences in EMS concentration between 46BR and the correctants, and between 46BR and MRC5V1, where 50% of cells survive \((b-a)/(c-a)\) in figure 3.4. By this calculation, transfection with wild-type DNA ligase I sequences has corrected the EMS sensitivity of 46BR by about 60%.

### 3.6 Correction of 3-Aminobenzamide Sensitivity in 46BR

3AB is an inhibitor of the enzyme polyADP-ribose polymerase which is required along with ligase for efficient ligation. Sensitivity to 3AB is therefore a direct measure of ligation. All cells are sensitive to 3AB at high doses, but 46BR cells are hypersensitive to 3AB since they are killed even at much lower doses. Survival in 3AB of MRC5V1, 46BR.1G1, 46BRpβAneo clone 1 and 46BRpβAHLneo clone 2 was investigated.

Cells were exposed to 3AB at a range of doses from 0 to 8mM and surviving cell numbers determined (figure 3.5). MRC5V1 cells showed 100% cell survival up to 8mM, where survival dropped to 50%. 46BR.1G1 and 46BRpβAneo clone 1 were extremely sensitive even at low doses. However, 46BRpβAHLneo clone 2 cells exhibited 100% cell survival at all doses, displaying complete correction of the 3AB hypersensitivity phenotype of 46BR upon introduction of wild-type DNA ligase I.

Survival in EMS is a measure of DNA repair, while survival in 3AB is a more direct measure of ligation activity, which could account for the discrepancy between correction of sensitivity to EMS and 3AB.

The correlation between the presence of wild-type DNA ligase I sequences in cells, the ligase-specific high molecular weight transcripts and rescue of the mutant phenotype demonstrates that the mutations identified in the DNA ligase I gene of 46BR are responsible for the mutant phenotype of these cells.
Figure 3.5: Cell survival following 3AB treatment of MRC5V1, 46BR.1G1 and transfectants. Each point is the mean from two separate determinations.

+ MRC5V1  
× 46BR.1G1  
O 46BRpβAneo clone 1  
Δ 46BRpβAHLneo clone 2
CHAPTER 4

MOUSE DNA LIGASE I cDNA
Following the demonstration that the abnormalities in the human cell line 46BR, and the clinical symptoms of the patient from whom the cell line was derived, arise from a mutation at the DNA ligase I locus, work towards creating a mouse model for DNA ligase I deficiency began. Before any gene targeting experiments could be carried out, information on the structure of the mouse DNA ligase I gene was required. The first step in obtaining this information was the cloning and sequencing of mouse DNA ligase I cDNA.

4.1 Cloning of Mouse DNA Ligase I cDNA

A mouse embryonic stem cell cDNA library (titre=10^6 pfu) constructed in λNM1149 (with the help of Dr. Nik Somia) was screened using full-length human DNA ligase I cDNA, isolated as a KpnI fragment from pHL, as a probe. Ligase-specific signals were obtained at a frequency of about 1/10,000. Seven clones were obtained and the ligase inserts subcloned into the EcoRI site of pBluescriptII KS+. The inserts were all different, of sizes 0.9kb, 1.7kb, 2.4kb, 2.6kb, 2.7kb, 2.9kb and 3.2kb (figure 4.1a), but share some common sequence, since the largest, 3.2kb, cDNA insert hybridised to them all (figure 4.1b).

All of the ligase clones were expected to be complete right up to the 3' end and foreshortened from the 5' end, if at all. This is due to the nature of cDNA synthesis. When cDNAs are synthesised, reverse transcriptase initiates at the 3' end of the messenger RNA, using an oligo dT primer which anneals to its poly A tail. That all the clones were apparently complete at the 3' end and truncated from the 5' end was confirmed by restriction mapping (data not shown). The five longest clones all have a BamHI recognition site 2kb from the 3' end, and a variable distance from the 5' end depending on the size of the insert. The two smaller clones are shorter than 2kb and therefore lack the BamHI site.

4.2: The Ligase Clones Encode Mouse DNA Ligase I

The largest mouse ligase insert is approximately 3.2kb in size. This is equal in size to the human DNA ligase I messenger RNA, and slightly longer than the human DNA ligase I cDNA used as a probe, which lacks about 90bp from the 5' end. The 3.2kb mouse ligase insert hybridises to a single 3.2kb message from both mouse and human RNA (figure 4.2) and these same transcripts are detected by reprobing with the human DNA ligase I cDNA (figure 4.2). The mouse ligase clone therefore encodes mouse DNA ligase I and is apparently full-length.
Figure 4.1: Mouse DNA ligase I cDNA clones. a) λNM1149 DNA was digested with NotI and electrophoresed through a 1% agarose gel. The positions of the 19.9kb and 8.8kb λ arms are indicated. b) The DNA fragments were transferred to a nylon membrane and probed with the 3.2kb cDNA insert. Molecular weight marker sizes are in kb.
Figure 4.2: Northern blot of human (HeLa) and mouse (E14) total RNA probed with the 3.2kb mouse cDNA (ML) and human DNA ligase I cDNA (HL). Molecular weight marker sizes are in kb. The intensity differences arise because ML hybridises more strongly to the mouse DNA ligase I mRNA, while HL hybridises more strongly to the human DNA ligase I mRNA.
4.3 Restriction Analysis of Mouse DNA Ligase I cDNA

Recognition sites within the 3.2kb mouse DNA ligase I clone were determined for a number of enzymes by restriction mapping (figure 4.3). When this map is compared to the human DNA ligase I cDNA restriction map a number of sites appear to be conserved. The two PstI sites, the SstI, BamHI and BglII sites, and two of the PvuII sites seem to correspond to sites within the mouse sequence. This further confirms that the mouse clone is related to the human DNA ligase I cDNA.

4.4 Sequencing of Mouse DNA Ligase I cDNA

Mouse DNA ligase I cDNA was sequenced by the dideoxy method using a combination of single and double stranded sequencing. About 300 bases of the sequence was obtained by Carolanne McEwan. The nucleotide sequence of the 3.2kb mouse DNA ligase I cDNA clone is given in figure 4.4. This clone consists of exactly 3172 bases. The sequencing confirmed that the above-mentioned BamHI, BglIII and PvuII sites are equivalent to sites within the human sequence, while the PstI and SstI sites do not correspond to human sites.

The cDNA has a 5' untranslated region of 256 bases up to the first ATG. Three ATGs within 66 bases follow, all in frame, and sequence homology with the human cDNA begins after the second ATG. The second ATG is also part of an initiation sequence (Kozak 1984), CCAACATGC, with only two base changes from the consensus, an A at position -2 instead of a C, and a C at position +4 instead of a G. The human cDNA has two ATGs. The first is part of the same Kozak sequence as in mouse, and is the translation start codon. It seems likely, therefore, that the second ATG in the mouse cDNA is the start codon. Assuming that this is correct, the mouse DNA ligase I cDNA encodes a protein of 916 amino acids. This is very similar in size to the 919 amino acid human DNA ligase I (Barnes et al 1990).

At the amino acid level, sequence conservation between mouse and human DNA ligase I cDNAs is extremely high, at 84% (figure 4.5). There are 147 amino acid differences between the two species. The C-terminal domains (amino acids 250-919 in the human enzyme, Kodama et al 1991) are virtually identical, while there is some sequence divergence towards the N-terminus. In particular, there are two regions of nonhomology with the human enzyme at positions 102-108 and 167-212 in the mouse enzyme (figure 4.5). This is consistent with the C-terminal domain of DNA ligase I being required for catalysis, while the N-terminus is regulatory, via interactions with other cellular factors. The C-terminal portion of the enzyme would
Figure 4.3: Restriction map of mouse DNA ligase I cDNA. For comparison a restriction map of the human DNA ligase I cDNA is also shown. Sites which are apparently conserved between the two cDNAs are shown in bold type.
Figure 4.4: Nucleotide sequence of mouse DNA ligase I cDNA. ML: Sequence of the 3.2kb clone obtained by probing an embryonic stem cell cDNA library with full-length human DNA ligase I cDNA. MLP: Sequence of the clone obtained by Savini et al (1994). The 5' and 3' untranslated regions are given in lower case, and the coding region is given in upper case. Vertical lines between nucleotides indicate homology, while spaces indicate a lack of conservation. Dots indicate gaps in the sequence. The upstream ATG, the initiation and termination codons, and the polyadenylation signal, AATAAA are given in bold print with double underlining. The primers listed were used for sequencing: 887Y, 888Y, G4607, G4608, G8320, G8321, G8322, H2066, H3123, and M13 universal primer. These primers were used to obtain sequence as follows:

- bases 1-337: M13 universal primer, by single-stranded sequencing in M13
- bases 338-490: G4608, by double-stranded sequencing in pBluescriptII KS+
- bases 491-628: H3123, by double-stranded sequencing in pBluescriptII KS+
- bases 629-859: H2066, by double-stranded sequencing in pBluescriptII KS+
- bases 860-1087: G8322, by double-stranded sequencing in pBluescriptII KS+
- bases 1088-1396: M13 universal primer, by single-stranded sequencing in M13
- bases 1397-1654: M13 universal primer, by single-stranded sequencing in M13
- bases 1655-1984: G4607, by double-stranded sequencing in pBluescriptII KS+
- bases 1985-2215: M13 universal primer, by single-stranded sequencing in M13
- bases 2216-2428: G8320, by double-stranded sequencing in pBluescriptII KS+
- bases 2429-2588: 888Y, by double-stranded sequencing in pBluescriptII KS+
- bases 2589-2793: 887Y, by double-stranded sequencing in pBluescriptII KS+
- bases 2794-2992: G8321, by double-stranded sequencing in pBluescriptII KS+
- bases 2993-3172: M13 universal primer, by double-stranded sequencing in pBluescriptII KS+
Figure 4.5: Amino acid sequence of DNA ligase I. MLP: Sequence of the mouse clone obtained by Savini et al (1994). ML: Sequence of the 3.2kb clone obtained by probing an embryonic stem cell cDNA library with full-length human DNA ligase I cDNA. HL: Sequence of human DNA ligase I (Barnes et al 1990). •: Amino acids conserved between all three sequences. ¡: Amino acids conserved between mouse sequences, but not human. *: Amino acids conserved between the mouse sequence presented in this thesis and the human sequence, but not the published mouse sequence. -: Amino acids that lack an equivalent in either the human or mouse enzyme. Putative phosphorylation sites (S_ _E motifs and the two 'strong' human sites S_66EGEEEDE and S_141EDEDRE), the ATP binding site (KYDG) and the amino acids equivalent to the mutations in 46BR at glutamate 566 and arginine 771 are indicated by underlining.
therefore be expected to be highly conserved, while the N-terminal portion could have evolved to interact with species-specific cellular factors.

Human DNA ligase I has been shown to require activation by casein kinase II phosphorylation in its N-terminal domain, and several putative sites for phosphorylation at serine residues have been identified within the enzyme (Prigent et al 1992). These sites consist of S_ _E motifs and are all possible substrates for casein kinase II (Kemp and Pearson 1990). However, two appear to be 'stronger' phosphorylation sites encompassing an S_ _E motif and it was suggested that one or other is the substrate for casein kinase II. The sequences of these two 'strong' sites are S66EGEEDE and S141EDEDRE. It is notable that the mouse enzyme lacks both of these two 'strong' casein kinase II phosphorylation sites (figure 4.5). It is therefore unlikely that Ser66 or Ser141 is the site of phosphorylation in the human enzyme. A number of other S _E motifs are present in the mouse enzyme and many of these are homologous to sites within the human enzyme. In particular there are two motifs present in the N-terminal domains of both human and mouse enzymes at positions 27 and 163 (human) or 161 (mouse). It is possible that one or other of these is the site of phosphorylation by casein kinase II. Further work on the phosphorylation patterns of these enzymes is required to clarify this point.

Amino acid sequence around the ATP binding site is highly homologous to the human cDNA. A stretch of 77 amino acids has only two conservative alterations. This region includes the Lys-Tyr-Asp-Gly motif observed in all eukaryotic DNA ligases studied so far (Tomkinson et al 1991b). DNA ligases from the viruses Vaccinia, T3 and T7 also have this motif. The Lys residue in the motif is the site of ATP binding, via a phosphoamide bond. The Glu residue two amino acids upstream from the ATP-binding site motif (position 564) is equivalent to the site at which the Glu566 to Lys mutation occurred in 46BR (Barnes et al 1992b).

Sequence in the region equivalent to that in which the Arg771 to Trp mutation occurred in 46BR is also highly conserved. At the amino acid level, the sequences are virtually identical and there is an arginine residue in mouse (position 769) corresponding to human Arg771.

The 3.2kb mouse DNA ligase I cDNA clone is not quite full-length. There is a 3' untranslated region of 117 bases, shortened by the loss of several bases from the 3' end. This includes the polyadenylation signal (Proudfoot and Brownlee 1976) and poly A tail. Two of the shorter cDNA clones, of sizes 2.7kb and 2.9kb, were sequenced from the 3' end and sequence obtained from the 2.7kb clone is given in figure 4.6. This clone includes a further 33 bases, containing a polyadenylation signal, AATAAAA, but still lacking a poly A tail. The 2.9kb clone has a shorter 3'
Figure 4.6: Nucleotide sequence of the 3' untranslated region of the 2.7kb mouse DNA ligase I cDNA clone. The polyadenylation signal, AATAAA, is underlined. The end of the 3' untranslated region of the 3.2kb mouse DNA ligase I cDNA clone is indicated by an arrow.
CCTCTGCTCTCCTGGGGCCTGGGGTGGAGGTTGCGAGAACATGGGGCAC

TGTTGGATTGGTGGTTGATTGGTGTGTGAGTCACAGGAGTGAAT

CCTGGGGCTGGAATGCTTTTTTCTCATAAATGACTTTTAGATGCC
untranslated region than the 2.7kb clone, but also has a polyadenylation signal and no poly A tail.

Mouse DNA ligase I cDNA has also been cloned and sequenced very recently by Savini et al (1994). Their clone (figure 4.4, figure 4.5) is virtually identical to the cDNA clone described in this chapter, showing 99% identity at both the nucleotide and amino acid levels. There are 11 base changes within the coding region, but most of the nucleotide alterations are silent, resulting in only six amino acid differences between the two clones, differences which may have arisen simply because the RNA used was obtained from separate mouse strains. Savini et al used RNA from NIH 3T3 Swiss mice (Todaro and Green 1963) while the RNA used in this work was taken from cells derived from 129/Ola mice (Olac Ltd 1976). However, it is notable that in every case, the mouse sequence presented here agrees with the human sequence, while the published sequence is the one that differs. Many of the alterations are conservative, such as the exchange of serine for alanine and isoleucine for valine at positions 734 and 814. On the other hand, some of the substitutions are not conservative, such as the exchange of cystine for serine at position 443. None of these differences are located in parts of the enzyme which have yet been identified as vital, although all are positioned in the highly conserved C-terminal domain of the enzyme.

There are other differences between the two mouse clones. The clone obtained by Savini et al has a shorter 5' untranslated region consisting of 59 nucleotides, compared to the 304 nucleotide 5' UTR of the clone presented in this thesis (figure 4.4). Within the 5' untranslated regions there are two nucleotide alterations at positions -39 and -40, where the sequence is GG in the published sequence, and AA in the sequence given here. The 3' untranslated region of the clone obtained by Savini et al is 151 nucleotides long while that of the clone given in this thesis consists of 117 bases (figure 4.4). There are two minor differences here as well, with there being four Gs present 14 and 44 bases after the stop codon in my sequence, but only three Gs at the same positions in the published sequence. Additionally, when the sequence of the longer 3' UTR from the 2.7kb mouse cDNA clone (figure 4.6) is compared, there is a further difference. An A has been exchanged in the published sequence for a G which is present in the 2.7kb clone 14 nucleotides down from the poly A signal. The clone obtained by Savini et al extends three bases further downstream than the 2.7kb clone but still lacks a poly A tail. Finally, Savini et al identified a recognition site for the restriction enzyme Smal within their clone. There is no recognition site for this enzyme within the clone given in this thesis. The absence of the site has been confirmed both by restriction
mapping and by sequencing. There are, then, several differences between the two mouse DNA ligase I cDNAs and, in every case where the amino acid sequence is affected, the published sequence differs from human DNA ligase I while the sequence presented here is the same as that of the human enzyme. However, in all important aspects, the clones appear to be identical. This is of interest since it confirms the absence of the two 'strong' casein kinase II phosphorylation sites identified in the human sequence.

In conclusion, mouse DNA ligase I cDNA has been cloned and sequenced. The cDNA is incomplete since it lacks a number of bases from the 3' end, but encodes the entire enzyme. It is highly homologous to the human enzyme and, with the exception of putative phosphorylation sites, retains all major features identified so far.
CHAPTER 5

EXPRESSION OF MOUSE DNA LIGASE I cDNA
Following the cloning of mouse DNA ligase I cDNA, attempts were made to express the cDNA in the cell line 46BR.1G1 for two reasons: (1) To demonstrate that the cDNA encodes a full-length active DNA ligase I enzyme by rescue of the DNA ligase I deficiency of 46BR1G1. (2) To improve upon the previous expression studies utilising the human DNA ligase I cDNA in the vector p8AHLneo (chapter 3) by modification of the mouse enzyme to make it detectable in the presence of the endogenous human enzyme and thus enable proper demonstration of expression of a wild-type enzyme within the cells.

5.1 Introduction of a Marker into Mouse DNA Ligase I

The C-terminal domain of DNA ligase I is involved in catalysis (Tomkinson et al 1990), therefore a marker peptide could interfere with ligation activity if placed within this region of the enzyme. The N-terminal domain is thought to be involved in regulation of DNA ligase I activity and in interaction with other cellular factors (Prigent et al 1992). However, it is less likely that a marker peptide would affect enzyme function if located in the regulatory domain of the enzyme rather than in the catalytic domain. Sequence encoding a hydrophilic marker peptide called 'FLAG' (Hopp et al 1988), consisting of eight amino acids, was therefore placed at the 5' end of the mouse DNA ligase I cDNA coding region. This peptide has been successfully used by other workers to tag proteins without affecting biological activity (for example see Kunz et al 1992).

A 64 base oligonucleotide (483W, figure 5.1) was designed for use in conjunction with a 20 base primer (482W) specific for a region of the mouse DNA ligase I cDNA immediately 3' of the BamHI site (figure 5.2). The 64 base primer consists of a NotI site preceded by eight bases, which is the minimum distance from DNA ends required for efficient NotI cutting. This is combined with a consensus Kozak translation initiation sequence (CCGCCATGG, Kozak 1984), which is followed by the flag sequence (Asp-Tyr-Lys-Asp-Asp-Asp-Lys) specified by codons chosen according to mouse codon usage tables. Finally, there are the first 20 bases from the mouse DNA ligase I cDNA coding region. There are three ATG codons at the 3'end of the cDNA and homology with the human cDNA begins after the second ATG. Additionally, there is a Kozak signal immediately upstream of the second ATG. It has therefore been assumed that the second ATG in the mouse cDNA is the true translation start site. Consequently the 20 bases of cDNA coding region are the first 20 bases following the second ATG.

The primers were used to amplify a 1kb section of mouse DNA ligase I cDNA, from the start of the coding region to 18 bases 3' of the BamHI site, by PCR
5' GTCCGACTTCGGCCGCTAGTGAACACTACAAGGACGACGATGATAAG

Asp Tyr Lys Asp Asp Asp Asp Lys

FLAG

cDNA
Figure 5.2: Introduction of the FLAG sequence into mouse DNA ligase I cDNA by PCR. A fragment of the mouse cDNA is represented, extending from the 5' end to downstream of the BamHI site. The hatched box represents the 5' untranslated region of the cDNA. Binding sites for the two primers used are indicated. Primer 482W binds the non-coding strand, while primer 483W binds the coding strand. The resulting 1kb PCR product is shown, with the 5' untranslated region replaced with the FLAG sequence (Closed box). The modified cDNA 5' end was isolated from this fragment as a 0.3kb NotI/SstI fragment. PCR conditions were as follows: denaturing for 1 min at 94°C, annealing for 1 min at 58°C and synthesis for 1 min at 72°C.
(figure 5.2). In this 1kb fragment all of the cDNA 5' sequence up to the second ATG has been lost, and replaced with a NotI site and 24 bases encoding the flag peptide.

The PCR fragment was digested with NotI and BamHI and cloned into NotI/BamHI-digested pBluescriptII KS+ for propagation. The modified 5' end of the cDNA was isolated from this plasmid as a 0.3kb SstI fragment, using one SstI site from pBluescript and the other from within the PCR fragment. The 5' end of the 3.2kb mouse DNA ligase I cDNA clone was removed as a 0.6kb SstI fragment, using one SstI site from pBluescript and the other from within the cDNA, and replaced with the modified 5' end. This produced a modified mouse DNA ligase I cDNA with the entire 303 bases of 5' untranslated region removed, including the first ATG, and replaced with a NotI site and the flag sequence. This plasmid was designated pfML.

The entire modified 5' end of the cDNA was sequenced, which verified that the desired changes had been made, and that no mutations had been introduced during the PCR amplification (figure 5.3).

5.2 Expression of Flag-Ligase in E. coli

The plasmid, pfML was introduced into E. coli strain BL21(DE3) (Studier and Moffatt 1986), which expresses T7 RNA polymerase from a chromosomal copy of the gene when induced. In pfML, the modified cDNA is under the control of the T7 promoter of pBluescriptII KS+ and will therefore be expressed when the cells are induced to make T7 RNA polymerase.

Cultures of BL21(DE3) cells transformed with pfML were induced to produce T7 RNA polymerase, and cell extracts electrophoresed through a polyacrylamide gel. Coomassie blue staining of the gel (figure 5.4a) revealed a ladder of protein bands in all tracks. However, at least three novel bands are visible in all tracks containing extracts from one of the cultures of cells transformed with pfML (clone 2). The largest of these novel bands is slightly larger than the 115kDa marker, and the others are approximately 30kDa or less, with the smaller proteins possibly arising from degradation of the largest polypeptide.

Human DNA ligase I is 125-130kDa in size as determined by SDS-PAGE (Teraoka and Tsukada 1982, Tomkinson et al 1990) and mouse DNA ligase I is likely to be of a similar size. Consequently the protein visible slightly above the 115kDa size marker could be mouse DNA ligase I with the marker peptide attached. The appearance of this band even when the cells were uninduced is probably due to the basal level of transcription of the T7 RNA polymerase gene as observed by other workers (Studier and Moffatt 1986).
Figure 5.3: Sequence of the 0.3kb NotI/SstI fragment containing the modified 5' end of the cDNA. The FLAG sequence was introduced as required and no base alterations were introduced into the cDNA sequence during amplification. The positions of the NotI site, Kozak sequence, flag and the start of the mouse DNA ligase I cDNA are indicated.
NotI

<table>
<thead>
<tr>
<th>Kozak</th>
<th>FLAG</th>
</tr>
</thead>
</table>

GCGGCCGCGCATGGACTCAAGGACGACGATGATAAGCGAGAAAGTTAT --- mouse DNA ligase I

GTCATTTTTTTCAACTACAAAAAGAAGGTAAAGCGAAGAAGCCAGAGAAGG

AGACACCCAGCAGCATCAGAGAGAAGGAACCCCCTCCAAGGTTGGCGCTG

AAGGAGAGGAATCAAGTGCTGCGCAGAGATCTCTCCAGTGAAAGGAGAC

AGGAAGGAAGGTAGCCCGAGTTCTGAGCTGAGGAGGAGGACGAAGATG

AAGCCCTGACCCCCAAAGTCCAGAAGCCTGTGTAACAGACTCTGAACAG

AGCTC
Figure 5.4: Expression of pfML in *E. coli*. a) Extracts from untransformed BL21(DE3), BL21(DE3) transformed with pBluescriptII KS+ and BL21(DE3) transformed with pfML (2 separate clones) were prepared at 0, 1, 2 and 3 hours following induction and electrophoresed through a polyacrylamide gel. The gel was stained with Coomassie blue to visualise the protein bands. The positions of novel bands, faintly visible only in all clone 2 tracks, are marked with an arrow. Protein marker sizes are given in kDa. b) An identical gel to that in figure 5.4a was electroblotted and probed with the anti-FLAG monoclonal antibody MI. The single band clearly visible in clone 2 tracks and also visible in clone 1 tracks corresponds to the largest of the novel bands seen on the Coomassie-stained gel.
An identical gel was electroblotted onto a nylon membrane and probed with an anti-FLAG monoclonal antibody (figure 5.4b). This revealed a single band in all clone 2 tracks at a position equivalent to the largest novel protein band observed on the Coomassie-stained gel. Additionally, a much fainter band was visible in all clone 1 tracks at the same position. None of the presumed degradation products were visualised and therefore are either unrelated to the flag-DNA ligase I protein or have lost the flag sequence during breakdown.

These experiments demonstrate that a single protein of approximately the same size as DNA ligase I and carrying the flag marker peptide is transcribed from pfML. Therefore the modified cDNA, when expressed, produces a protein which is apparently full-length and which is detectable using the anti-flag monoclonal antibody. This protein was not isolated and assayed for activity in order to show that the cDNA encodes an active protein, since expression of full-length DNA ligase I in E. coli produces an inactive enzyme which requires further post-translational modification before it can catalyse ligation reactions (Kodama et al 1991).

5.3 Expression of Flag-Ligase in 46BR Under Control of the Phosphoglycerate Kinase Promoter

A new vector was constructed to express the modified mouse DNA ligase I cDNA in 46BR. The vector was designed to eliminate the problems with faulty initiation or termination of transcription experienced with the expression vector pBAneo (chapter 3).

The expression vector was constructed in pBluescriptII KS+ by Carolanne McEwan. First, the mouse phosphoglycerate kinase promoter was cloned into the plasmid as a 550bp EcoRI/PstI fragment. The phosphoglycerate kinase promoter was chosen because it has been used successfully to express HPRT minigenes (Magin et al 1992). This was followed by a 2.4kb EcoRI fragment containing the bacterial neomycin resistance gene under control of the mouse metallothionein promoter to provide selection for transformants in G418. This is the same neo' gene as that used in pBAneo and it was cloned into the vector with the metallothionein promoter in both orientations. Finally, a 1.3kb fragment containing the last two exons and 3' untranslated region of the HPRT gene was cloned into the SstII site of the vector by addition of SstII linkers. This fragment was included to provide transcription termination and polyadenylation signals, thereby eliminating the need for cloned sequences to provide their own polyadenylation signals and also promoting proper termination of transcription. This completed construction of the new expression vectors (figure 5.5) which were designated pDWM201 (neo' in the same orientation.
Figure 5.5: Expression vectors pDWM201 and pDWM202. The mouse phosphoglycerate kinase promoter (shaded box) drives expression, with transcription termination sequences provided by the 3' end of the HPRT gene. This includes exons 8 (hatched box) and 9 (closed box) as well as the eighth intron (open box). The neomycin phosphotransferase gene is driven by the mouse metallothionein promoter which is present in the same orientation as the phosphoglycerate kinase promoter in pDWM201 and in the opposite orientation in pDWM202. Key restriction sites are given.
as the phosphoglycerate kinase promoter) and pDWM202 (neo' and the phosphoglycerate kinase promoter in opposite orientations).

A cDNA encoding mouse HPRT (isolated from pHPT5 as a 1.3kb PstI fragment) was cloned into both vectors so that they could be tested for their ability to drive expression from the phosphoglycerate kinase promoter of cDNAs inserted in the appropriate orientation. These plasmids were designated pDWM203 (pDWM201 + HPRT) and pDWM204 (pDWM202 + HPRT). They were then introduced into HPRT-deficient RJK88 chinese hamster cells by calcium phosphate precipitation by Angela Pow. Equal quantities of each plasmid were introduced into equal numbers of cells. At the same time, pDWM111 (Magin et al 1992) was also transfected into RJK88 cells. This plasmid contains an HPRT mini-gene and cells transfected with it can express active HPRT enzyme. It was included in the experiment as a positive control for HPRT expression in the RJK88 fibroblasts. Transformed cells were selected for growth in the synthetic antibiotic G418.

Total RNA was prepared from untransformed RJK88 cells, from RJK88 cells transformed with pDWM203 and pDWM204, as well as from the wild-type embryonic stem cell line E14. The RNA was size-fractionated, blotted and probed with HPRT cDNA (figure 5.6) This revealed a 1.5kb HPRT message, present only in RNA prepared from the wild-type cells or those transformed with the vectors expressing HPRT cDNA. No HPRT-specific message was detectable in RJK88 RNA in the absence of the expression vectors. Therefore the vector was successfully expressing HPRT cDNA. However an actin reprobing of the filter revealed that the RNA loadings were equal (data not shown) and therefore that HPRT was being expressed from the vectors at a level substantially lower than that in wild type cells. An additional 5.1kb band of unknown identity was visible in all the RNA taken from RJK88 cells.

Cells expressing active HPRT can be selected by growth in HAT medium which contains hypoxanthine, aminopterin and thymidine. Aminopterin is a toxic base analogue which blocks de novo synthesis of purines and pyrimidines. However, cells can bypass the metabolic block using hypoxanthine and thymidine, if they utilise the purine salvage pathway. HPRT is one of the enzymes involved in this pathway, therefore cells can only grow in HAT medium if they are HPRT+. RJK88 fibroblasts are HPRT−, so transformants expressing HPRT from an integrated plasmid can be selected.

As an additional test for HPRT expression from the vectors, HPRT+ cells were selected in HAT medium by Angela Pow and the extent of HPRT expression determined by counting colonies (figure 5.7). RJK88 cells transformed with

127
Figure 5.6: Northern blot of RNA prepared from the HPRT-deficient cell line RJK88, RJK88 transformed with plasmids pDWM203 and pDWM204, and the wild-type cell line E14. The blot was probed with HPRT cDNA. The positions of the ribosomal sub-units are indicated and their sizes given in kb. The position of the 1.5kb HPRT messenger RNA is marked by an arrow.
Figure 5.7: Bar chart showing the extent of HPRT expression in RJK88 cells transformed with pDWM111, pDWM203 and pDWM204, as measured by counting HAT-resistant colonies. Each experiment is the mean of two dishes.
pDWM111 showed the highest level of HPRT expression, but cells transformed with both pDWM203 and pDWM204 were also expressing HPRT. There was no significant difference in HPRT expression between pDWM203 and pDWM204. These plasmids differ only in the orientation of the neo\textsuperscript{r} gene, which therefore apparently does not affect expression.

The discrepancy between the higher levels of HPRT expression in the wild-type cell line E14, and from pDWM111 and the much lower expression levels from vectors pDWM203 and 204 could be due to the new expression vectors not working as efficiently, or because HPRT expression is enhanced by intron sequences present in the endogenous gene and in the minigene in pDWM111, but not in pDWM203 and 204.

Mouse DNA ligase I cDNA and the cDNA with the flag marker were cloned into the NotI site of pDWM201, placing them under control of the phosphoglycerate kinase promoter. The new plasmids were designated pPGKMLneo and pPGKfMLneo. The vectors were linearised with XhoI since the mouse DNA ligase I cDNA contains a Scal site, and introduced into the cell line 46BR.1G1 by electroporation. Following transformation, G418-resistant colonies were grown as pools and assayed for sensitivity to 3-aminobenzamide. The hypersensitivity of 46BR to 3-aminobenzamide is rescued 100% by wild-type human DNA ligase I cDNA (chapter 3, and repeated in parallel with these experiments - data not shown), but no correction was detectable by introduction of the mouse cDNA, with or without the marker (data not shown). This raised a number of possibilities. The mouse cDNA clone could be mutated and not encode a functional enzyme, or perhaps the mouse enzyme cannot complement the human defect. However, sequencing had so far indicated an extremely high degree of homology between mouse and human cDNAs with no obviously significant amino acid alterations between the two. Since both tagged and untagged mouse cDNAs had given a negative result, there was no evidence for the removal of the 5′ untranslated region and addition of the marker having affected expression. A more likely explanation for the failure of the mouse cDNA to rescue the 3-aminobenzamide hypersensitivity is that the mouse enzyme was simply not being expressed at a high enough level. Using the anti-FLAG monoclonal antibody to probe protein extracts from cells transformed with the marked cDNA confirmed this. No tagged protein was detectable (data not shown).
5.4 Expression of Flag-Ligase in 46BR Under Control of the Human β-Actin Promoter

Given the failure of the expression vector pDWM201, the tagged and untagged mouse DNA ligase I cDNAs were cloned into the original expression vector pBAHLneo, under control of the human β-actin promoter. This meant that the mouse cDNA should be transcribed since the vector had already been used successfully in other experiments. However, this vector relies on cloned sequences to provide their own polyadenylation signal and the presence of this signal at the 3' end of the mouse DNA ligase I clone had not yet been confirmed by sequencing. Since cDNA synthesis commences from an oligo dT primer which hybridises to the poly A tail of messenger RNA, it was assumed that the cDNA clone would turn out to be complete right up to the poly A tail.

No suitable sites could be identified for linearisation of the plasmids prior to introduction into mammalian cells. Consequently, the plasmids were introduced as closed circular molecules. This meant that there was no preference for the vectors to integrate into the genome with the neo' module, the human β-actin promoter and the cDNA maintained as one intact unit. Instead, the plasmid could break anywhere, including within one of these elements, with the possible result that either transformed cells would not be resistant to G418, or that they would be unable to transcribe the cDNA, or both. The plasmids were introduced into 46BR.1G1 by electroporation and G418-resistant colonies were grown up as clones or as pools.

Initially, pools of transformed cells were assayed for sensitivity to 3-aminobenzamide. Their pattern of survival was indistinguishable from that of untransformed 46BR.1G1 cells, although cells transfected with human DNA ligase I cDNA showed wild-type levels of survival (data not shown). No rescue of the hypersensitivity to 3-aminobenzamide was detectable with the cDNA tagged or untagged (data not shown). This could have been because only a very small proportion of the cells resistant to G418 were also expressing the mouse DNA ligase I cDNA since the plasmid was not linearised. Twelve G418-resistant colonies were tested for presence of the tagged mouse DNA ligase I cDNA by PCR. The primers originally used to introduce the FLAG sequence (482W, 483W), were used in the PCR assay. These primers produce a 1kb fragment from the 5' end of the mouse cDNA with the marker. This 1kb product was seen with eight out of the twelve colonies screened (figure 5.8). This was an unexpectedly high proportion, but still does not adequately demonstrate that these colonies all carry an intact cDNA under control of the human β-actin promoter. However, at this stage it was discovered, by sequencing, that the mouse DNA ligase I cDNA clone used in these experiments is
Figure 5.8: Screening G418-resistant colonies transfected with pBAfMLneo for the presence of mouse DNA ligase I cDNA sequences by PCR, using primers 482W and 483W. Denaturing was carried out at 94°C for 1min, annealing at 58°C for 1min and DNA synthesis at 72°C for 1min. STD: molecular weight markers, 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb and 1.6-0.2kb at 100bp intervals, C: negative controls, Reactions carried out in the absence of DNA substrate or in the absence of primers, fML: reaction carried out using pfML as substrate, 1-12: reactions carried out on extracts prepared from colonies of cells transfected with pBAfMLneo. The position of the 1kb PCR product is indicated by an arrow.
not complete. It is truncated from the 3' end and lacks a polyadenylation signal and poly A tail. This fact could account for the failure of the mouse cDNA to rescue the 3-aminobenzamide hypersensitivity of 46BR.1G1, since any messenger RNAs transcribed from the integrated plasmid would probably be too unstable to be translated and produce mouse DNA ligase I protein in significant amounts (Jackson and Standardt 1990). This is not inconsistent with the demonstrated expression of the cDNA in *E. coli* since bacterial messenger RNAs do not require a poly A tail for stability.

The absence of any mouse DNA ligase I enzyme within the cells was confirmed by preparing Western blots of protein extracts from each of the positive clones and probing with the anti-FLAG monoclonal antibody. No protein was detectable (data not shown).

A further complication to these experiments was discovered with the publishing of the mouse DNA ligase I cDNA sequence (Savini *et al* 1994). Comparison of their sequence with the sequence obtained for the mouse DNA ligase I cDNA clone described in chapter 4 revealed a reading error. This had resulted in the first two codons following the initiator methionine being read as CGAGAA when in fact they should have been CAGAGA. At the amino acid level this translated to a change from glutamine and arginine, to arginine and glutamate. Introduction of the FLAG sequence onto the N-terminus of mouse DNA ligase I by PCR (figure 5.2) resulted in the incorporation of this sequencing error into the cDNA and therefore into the protein. However, since these two amino acid alterations are at the extreme N-terminus of the protein which has already been altered by addition of the eight amino acid FLAG sequence, it is unlikely that they would affect the function of the enzyme.

Subsequent to these experiments, work carried out by Teraoka *et al* in 1993 suggested that N-terminal modification of DNA ligase I can affect its activity. It had previously been shown that DNA ligase I expressed in *E. coli* as a fusion protein with β-galactosidase at the N-terminus is inactive (Prigent *et al* 1992). However Teraoka *et al* found that active DNA ligase I could be expressed in *E. coli* when the only alteration present was a threonine residue inserted immediately 3' of the initiator methionine. Obviously, small changes at the N-terminus of this enzyme can affect its activity. However, these experiments were carried out in *E. coli*, and Prigent *et al* also showed that the inactive enzyme could be activated by casein kinase II. This kinase is present in all mammalian cells and would probably also phosphorylate the FLAG DNA ligase I. A further difficulty is that the N-terminus of DNA ligase I is blocked, presumably due to post-translational modification.
(Tomkinson et al. 1990). If the presence of FLAG affects this modification then it is possible that it could also affect enzyme activity. In view of the fact that no detectable FLAG DNA ligase I was produced in the mammalian cells, while unmodified DNA ligase I is readily detectable it seems unlikely that any of these factors led to the failure of the experiment.

In retrospect it would have been preferable to have confirmed the presence of a polyadenylation signal within the cDNA prior to trying to express it from a vector which does not provide this signal. However the failure of the experiments using the expression vector pDWM201, which provided a signal, cannot be due to the lack of a poly A sequence in the cDNA. This vector also failed to produce detectable quantities of mouse DNA ligase I protein. It seems that the separate experiments with the two different expression vectors failed for different reasons. The first experiments, using the phosphoglycerate kinase vector failed because transcription from the plasmid was low (figure 5.6) and no detectable protein was produced. The second experiments probably failed because mouse DNA ligase I mRNAs without a poly A tail are probably too unstable to be translated (Jackson and Standardt 1990).
CHAPTER 6

MOUSE DNA LIGASE I GENE STRUCTURE
Two mutations have been identified in the DNA ligase I gene of the human cell line 46BR (Barnes et al 1992b), both of which have a detrimental effect on DNA ligase I activity. Understanding of how these mutations give rise to all the abnormalities which were observed in the patient from whom the cell line was derived could potentially be gained from a mouse model for DNA ligase I deficiency. In order to create a mouse model, one of the 46BR mutations must first be introduced into the mouse DNA ligase I gene in embryonic stem cells by gene targeting. The first mutation, Glu566 to Lys, is at the active site of the enzyme, rendering it completely inactive, and could therefore be lethal in a homozygote. However the enzyme with the second mutation, Arg771 to Trp, retains residual activity and cells expressing only this mutant form are viable (Barnes et al 1992b). Consequently the Arg771 to Trp mutation was selected as the best candidate for gene targeting experiments.

In order to target the mouse gene, some structural information was first required and this chapter describes investigation of the mouse DNA ligase I gene structure. This was done at a preliminary level to provide information enabling isolation of a mouse genomic fragment containing sequence equivalent to the site of the Arg771 to Trp mutation identified in 46BR.

6.1 Southern Hybridisation Analysis

Mouse DNA ligase I cDNA was used to investigate the structure of the mouse DNA ligase I gene by hybridisation to mouse genomic DNA. DNA was prepared from the embryonic stem cell line E14TG2a and cut with BamHI, EcoRI or HindIII. A Southern blot of this restricted DNA was probed with full-length mouse DNA ligase I cDNA which revealed a complex pattern of bands (figure 6.1, table 6.1). Eight BamHI fragments of sizes 9.4kb, 7.2kb, 6.4kb, 3.9kb, 3.6kb, 1.9kb, 1.2kb and 1.1kb, five EcoRI fragments of sizes 13.5kb, 10.0kb, 5.9kb, 5.2kb and 4.5kb, and seven HindIII fragments of sizes 10.7kb, 6.7kb, 4.2kb, 3.9kb, 2.1kb, 1.8kb and 1.7kb. The bands are of different intensities, not because of partial digestion, but due to varying amounts of coding region in each fragment allowing differing amounts of cDNA probe to bind. Totalling the sizes of the bands in each digest suggests the mouse gene covers 31-39kb of DNA.

For comparison, genomic DNA was prepared from the human cell line HeLa and also digested with BamHI, EcoRI and HindIII. This DNA was blotted and probed with full-length human DNA ligase I cDNA (figure 6.1), disclosing a different, but equally complex pattern of bands. Three BamHI fragments of sizes
Figure 6.1: Southern blot of mouse (E14TG2a) and human (HeLa) genomic DNA digested with BamHI, EcoRI and HindIII and probed with mouse and human DNA ligase I cDNAs respectively. Molecular weight marker sizes are given in kb.
Table 6.1: Genomic Fragments Hybridising to Mouse DNA Ligase I cDNA and cDNA Subfragments.

<table>
<thead>
<tr>
<th>Probe</th>
<th>Hybridising Fragments (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BamHI</td>
</tr>
<tr>
<td>mouse DNA ligase I cDNA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
</tr>
<tr>
<td>0.6kb EcoRI/PstI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>6.4</td>
</tr>
<tr>
<td>0.5kb PstI/BamHI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td>0.35kb BamHI/PvuII</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6kb PvuII/SalI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>1.9</td>
</tr>
<tr>
<td>0.35kb SalI/PvuII</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>5.2</td>
</tr>
<tr>
<td>0.4kb PvuII/PstI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5kb PstI/EcoRI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>1.1</td>
</tr>
</tbody>
</table>
10.7kb (probably an unresolved multiple band), 6.7kb and 3.7kb, seven EcoRI fragments of sizes 10.7kb, 7.5kb, 5.3kb, 4.4kb, 3.6kb, 3.3kb and 2.5kb (the same pattern as seen with 46BR genomic DNA, chapter 3, figure 3.2), and two HindIII fragments of about 40kb and 18.5kb in size. The size of the largest HindIII fragment is approximate since the largest size marker used was 23.13kb. Other workers have observed a different pattern of BamHI, EcoRI and HindIII fragments within the human DNA ligase I gene (Barnes et al 1992a, Noguiez et al 1992). They have found that the human gene consists of 28 exons spanned by four BamHI fragments, seven EcoRI fragments and five HindIII fragments (figure 6.2). Apart from the different numbers of human gene fragments, very few of the fragments are related in size. For example, figure 6.1 shows three BamHI fragments of 10.7kb, 6.7kb and 3.7kb, while the published human gene structure has four BamHI fragments, all of which are more than 10kb in size. The discrepancies may have arisen because the human DNA came from different sources and partly also because of inaccuracies in measuring fragment sizes from Southern blots. Overall, when the bands are added up, it seems from these results that the human gene occupies 31.8 to about 60kb of DNA, which is consistent with the accurately determined value of 53kb (Noguiez et al 1992).

In order to determine how the mouse genomic fragments relate to the mouse cDNA it was necessary to work out how they are organised within the gene. To this end the mouse cDNA was cut into sub-fragments (figure 6.3) and each used in turn to probe the genomic DNA. The results from this analysis are summarised in table 6.1.

Initially, the 5' 0.6kb EcoRI/PstI fragment of the cDNA was used as a probe (figure 6.4a). It hybridised to the 7.2kb and 6.4kb BamHI fragments, to the 5.9kb and 4.5kb EcoRI fragments and to the 10.7kb and 6.7kb HindIII fragments, which must therefore be at the 5' end of the gene. This shows that the 0.6kb of 5' cDNA sequence spans two genomic fragments. It follows that there must be an intron interrupting this segment of sequence since the cDNA itself does not contain any BamHI, EcoRI or HindIII sites in this region.

Next, the DNA was probed with the 0.5kb PstI/BamHI fragment (figure 6.4a), which bound to the 7.2kb and 3.9kb BamHI fragments, to the 10.0kb and 4.5kb EcoRI fragments and to the 10.7kb and 3.9kb HindIII fragments. This allowed the first few fragments of the gene to be arranged in order. Both probes used so far bound to the 7.2kb BamHI fragment, but only the first probe bound to the 6.4kb BamHI fragment, therefore these fragments are ordered 5'-6.4kb-7.2kb-3.9kb-3'.
Figure 6.3: Restriction map of mouse DNA ligase I cDNA. Each fragment shown was used as a probe. The EcoRI sites are not within the cDNA itself, but are present within the adaptors ligated onto the cDNA ends during cDNA synthesis. Drawn to scale. Sizes are given in kb.
Figure 6.4a: Mouse genomic DNA probed with 0.6kb EcoRI/PstI and 0.5kb PstI/BamHI fragments of mouse DNA ligase I cDNA. The Southern blot of mouse genomic DNA digested with BamHI, EcoRI and HindIII (figure 6.1) was stripped of probe and reprobed with the 0.6kb EcoRI/PstI fragment. A similar Southern filter was probed with the 0.5kb PstI/BamHI fragment. Molecular weight marker sizes are given in kb. For reference, the mouse DNA ligase I cDNA restriction map (from figure 6.3) is reproduced.
Eco RI  PstI  BamHI  PvulI  Sall  PvulI  PstI  EcoRl

5'  0.6  0.5  0.35  0.6  0.35  0.4  0.5  3'

0.6kb EcoRI/PstI
BamHI  EcoRI  HindIII

0.5kb PstI/BamHI
BamHI  EcoRI  HindIII

23.130-
9.419-
6.559-
4.371-
2.322-
2.028-
1.357-
1.078-
The HindIII fragments are ordered 5'-6.7kb-10.7kb-3.9kb-3' because the two probes overlapped on the 10.7kb fragment.

The 0.35kb BamHI/PvuII cDNA fragment hybridised to the 3.6kb BamHI fragment, to the 10.0kb EcoRI fragment and to the 10.7kb and 3.9kb HindIII fragments (figure 6.4b). Therefore the fourth BamHI fragment is the 3.6kb fragment, while there was no new information on the order of the EcoRI fragments. The reappearance of the 10.7kb HindIII band was somewhat confusing but re-examination of the original pattern (figure 6.1) showed that this band is very intense and is possibly a doublet. This extended the order of the HindIII fragments to 5'-6.7kb-10.7kb-3.9kb-10.7kb-3'.

Next, the 0.6kb PvuII/Sall fragment was used as a probe (figure 6.4b). It bound to the 9.4kb, 3.6kb and 1.9kb BamHI fragments, to the 10.0kb and 5.2kb EcoRI fragments and to the 10.7kb HindIII fragment. The order of the BamHI bands became 5'-6.4kb-7.2kb-3.9kb-3.6kb-9.4kb-1.9kb-3' and the order of the EcoRI fragments became 5'-5.9kb-4.5kb-10.0kb-5.2kb-3'. This probing also confirmed that there are indeed two 10.7kb HindIII fragments.

Then the Southern was probed with the 0.35kb SalI/PvuII cDNA fragment (figure 6.4c). It hybridised to the 9.4kb BamHI fragment, to the 13.5kb and 5.2kb EcoRI fragments and to the 10.7kb and 1.8kb HindIII fragments. Therefore the order of the BamHI fragments became 5'-6.7kb-7.2kb-3.9kb-3.6kb-1.9kb-9.4kb-3', the complete order of the EcoRI fragments is 5'-5.9kb-4.5kb-10.0kb-5.2kb-13.5kb-3', and the HindIII pattern was extended to 5'-6.7kb-10.7kb-3.9kb-10.7kb-1.8kb-3'.

When the DNA was probed with the 0.4kb PvuII/PstI fragment (figure 6.4c), it picked up the 9.4kb and 1.2kb BamHI fragments, the 13.5kb EcoRI fragment and the 2.1kb, 1.8kb and 1.7kb HindIII fragments. The BamHI band pattern became 5'-6.4kb-7.2kb-3.9kb-3.6kb-1.9kb-9.4kb-1.2kb-3' and the HindIII pattern was lengthened to 5'-6.7kb-10.7kb-3.9kb-10.7kb-1.8kb-2.1kb/1.7kb-3'.

Finally, the 0.5kb PstI/EcoRI cDNA fragment was used (figure 6.4c), revealing BamHI 3.9kb and 1.1kb fragments, the EcoRI 13.5kb fragment and HindIII 4.2kb and 1.7kb fragments. This shows that there is apparently a second 3.9kb BamHI fragment. Unexpectedly, this probe did not pick up the 1.2kb BamHI fragment which hybridised to the 0.4kb PvuII/PstI cDNA fragment. It is therefore probable that the PstI site is very close to the 3' end of the exon within the 1.2kb BamHI fragment or else very close to the 5' end of the exon contained within the downstream BamHI fragment. This would mean that there is not enough homology on one side of the PstI site for hybridisation to occur and the cDNA probes used would therefore not detect a common fragment. The final pattern of BamHI bands is
Figure 6.4b: Southern blot of mouse genomic DNA digested with BamHI, EcoRI and HindIII (figure 6.1) stripped and reprobed with 0.35kb BamHI/PvuII and 0.6kb PvuII/SalI fragments of mouse DNA ligase I cDNA. Molecular weight marker sizes are given in kb. The 3.6kb BamHI fragment to which the 0.6kb PvuII/SalI probe hybridised is not visible in this figure, although it could be seen on the original autoradiogram. For reference, the mouse DNA ligase I cDNA restriction map (from figure 6.3) is reproduced.
Figure 6.4c: Southern blot of mouse genomic DNA digested with BamHI, EcoRI and HindIII (figure 6.1) stripped and reprobed with 0.35kb Sall/PvuII, 0.4kb PvuII/PstI and 0.5kb PstI/EcoRI fragments of mouse DNA ligase I cDNA. Molecular weight marker sizes are given in kb. The 9.4kb BamHI fragment to which the 0.4kb PvuII/PstI probe hybridised is barely visible in this figure, although it could be seen on the original autoradiogram. For reference, the mouse DNA ligase I cDNA restriction map (from figure 6.3) is reproduced.
EcoRI  PstI  BamHI  Pvull  Sall  Pvull  PstI  EcoRI

5'  0.6  0.5  0.35  0.6  0.35  0.4  0.5  3'

0.35kb BamHI/Pvull
BamHI EcoRI HindIII

0.6kb Pvull/Sall
BamHI EcoRI HindIII

<table>
<thead>
<tr>
<th>Size (kb)</th>
<th>23.130</th>
<th>9.419</th>
<th>6.559</th>
<th>4.371</th>
<th>2.322</th>
<th>2.028</th>
<th>1.357</th>
<th>1.078</th>
</tr>
</thead>
</table>

rT 1.357— 1.078—
EcoRI  PstI  BamHI  PvuII  Sall  PvuII  PstI  EcoRI

0.35kb Sall/PvuII  0.4kb PvuII/PstI  0.5kb PstI/EcoRI

BamHI  EcoRI  HindIII

23.130  9.419  6.559  4.371  2.322  2.028  1.357  1.078
The order of the 3.9kb and 1.1kb fragments could not be resolved by these probings. The complete HindIII fragment pattern is 5'-6.7kb-10.7kb-3.9kb-10.7kb-1.8kb-2.1kb-1.7kb-4.2kb-3'.

All three complete individual fragment patterns are shown in figure 6.5. In this figure, a convention of one exon per hybridising fragment has been used. Exons have been placed so that the BamHI, EcoRI and HindIII sites lie within introns since, with the exception of one BamHI site, the cDNA itself contains no sites for these enzymes. Therefore any sites within the gene cannot be in the coding region and must lie in introns. The exons are not necessarily single since they may contain other introns which, lacking a recognition site for one of the three restriction enzymes used, would not have been detected. The PstI site at the 3' end has been placed within the 1.2kb BamHI fragment, although it could equally be in the next fragment downstream.

6.2 Mouse DNA Ligase I Gene Map

A map of the mouse DNA ligase I gene was obtained by fitting together the three restriction patterns (figure 6.6). In this final map the gene has eleven exons. This is a minimum number for reasons already explained. Exact positions of some of the restriction sites with respect to the exons have yet to be determined, for example, the 3' end of the 9.4kb BamHI fragment could lie between the seventh and eighth exons, or between the eighth and ninth exons.

Throughout this analysis there was no evidence for a mouse DNA ligase I pseudogene which would have been detectable as a single fragment hybridising to all (or most) of the probes used. Therefore all fragments identified are authentic segments of the gene. The sum of the BamHI fragments suggests that the gene is 38.6kb in size, the sum of the EcoRI fragments suggests a size of 39.1kb and the sum of the HindIII fragments suggests a size of 41.8kb. These values are all in close agreement, which indicates that there are probably no large intron fragments missing from the gene map which is therefore probably complete. Overall then, the gene occupies 39-42kb of DNA. This figure has increased from the initial estimate of 31-39kb due to the presence of doublet bands.

When what is known of the mouse gene structure is compared with the human gene structure (Noguiez et al 1992, figure 6.6), it appears that the mouse gene is somewhat smaller than the human gene which occupies 53kb of DNA in contrast to 39-42kb. The human gene has 28 exons, while only eleven were identified in the mouse gene. However, the nature of the analysis meant that mouse
Figure 6.5: Order of mouse DNA ligase I genomic fragments. Individual patterns are shown for BamHI, EcoRI and HindIII digested DNA. No attempt has been made to reconcile numbers of exons, positions of restriction sites from the cDNA etc. Fragments are drawn to scale. Hatched boxes indicate exons.
BamHl

5' - Pstl  BamHl  Pvull  SalI  Pvull  Pstl - 3'
6.4  7.2  3.9  3.6  1.9  9.4  1.2  3.9/1.1

EcoRI

5' - Pstl  BamHl  Pvull  SalI - 3'
5.9  4.5  10.0  5.2  13.5

HindIII

5' - Pstl  BamHl  Pvull  SalI  Pvull  Pstl - 3'
6.7  10.7  3.9  10.7  1.8  2.1  1.7  4.2
Figure 6.6: Mouse DNA ligase I gene map. The arrangement of BamHI, EcoRI and HindIII restriction fragments within the gene is shown. Alternative positions for some restriction sites are given, based entirely on the Southern analysis results, without taking fragment sizes into account. Hatched boxes indicate exons. For comparison, the structure of the human DNA ligase I gene (from figure 6.2) is reproduced (not to the same scale).
introns were only detectable by the presence of a BamHI, EcoRI or HindIII site and it is probable that many introns went undiscovered because they lack one of these sites. Given the very high degree of sequence conservation between human and mouse DNA ligase I cDNAs, it is likely that the exon structures are identical, with both genes having 28 exons, and therefore size differences between the two genes must be due to varying intron dimensions. This has been observed in other genes. For example both the human and mouse HPRT genes have nine exons, while the introns vary in size, so that the human gene occupies 42kb of DNA (Kim et al 1986) and the mouse gene occupies 33kb (Melton et al 1984).

By knowing which cDNA features lie in which of the cDNA fragments used to probe the gene structure, it is possible to determine where these features are located in the gene. For example, the ATP binding site and site equivalent to the first 46BR mutation (Glu566 to Lys) are positioned immediately 3' of the SalII site in the cDNA and therefore lie within the 9.4kb BamHI, 5.2kb EcoRI and 10.7kb HindIII genomic fragments.

At the time this analysis was carried out no sequence information was available for the region of mouse DNA ligase I cDNA around the site of the second 46BR mutation (Arg771 to Trp). However a number of the restriction sites within the mouse cDNA are apparently conserved between mouse and human cDNAs (figure 6.7). In particular, there is a PvuII site 0.9kb from the 3' end of both mouse and human cDNAs. The site of the Arg771 to Trp mutation lies downstream from this PvuII site in the human DNA ligase I sequence. If this PvuII site corresponds to the mouse PvuII site at the same position then the site of the Arg771 to Trp mutation is contained within the 13.5kb mouse EcoRI genomic fragment. It was subsequently confirmed by sequencing that the PvuII site is conserved between human and mouse DNA ligase I cDNAs, and therefore that the 13.5kb EcoRI genomic fragment is indeed the fragment of interest.
Figure 6.7: Comparison of restriction sites between human and mouse DNA ligase I cDNAs. Both cDNAs are drawn to scale. Sizes are given in kb. The PvuII site 0.9kb from the 3' end of both human and mouse cDNAs is shown in bold type.
CHAPTER 7

CLONING OF A GENOMIC FRAGMENT CONTAINING THE ARGinine TO TRYPTOPHAN MUTATION SITE
Sequence corresponding to the site of the Arg771 to Trp mutation identified in the human cell line 46BR should be present within a 13.5kb EcoRI fragment of the mouse gene. This chapter concerns cloning of this 13.5kb genomic fragment and identification of an arginine codon equivalent to human Arg771.

7.1 Cloning of Mouse DNA Ligase I Genomic Fragments

There are five mouse DNA ligase I EcoRI genomic fragments, of sizes 13.5kb, 10.0kb, 5.9kb, 5.2kb and 4.5kb. For construction of a library, the two largest fragments are suitable for cloning into a bacteriophage λ replacement vector which accepts inserts in the size range 7-20kb. However the smaller fragments are more likely to be clonable in an insertion vector. The range in fragment sizes consequently prevents the entire gene being cloned from a single EcoRI library. Since the 13.5kb fragment was the one of interest, a genomic library of mouse EcoRI fragments was therefore constructed in the replacement vector λ EMBL4 (Frischauf et al 1983). This library was expected to allow cloning of the two largest DNA ligase fragments, with the possibility of also obtaining some of the smaller fragments as multiple inserts with other fragments.

The library was screened using full-length mouse DNA ligase I cDNA as a probe. The full-length cDNA was used in order to be able to detect any of the DNA ligase I genomic fragments which might be present. Ligase signals were detected at a frequency of 1/100,000 and seven clones giving a positive signal were plaque purified. Of these seven, five contained a 13.5kb fragment and two contained a 10.0kb fragment (figure 7.1a). Five of the clones also carried a second fragment, ranging in size from about 1.5kb to 4kb. None of the clones appeared to have any of the three smaller EcoRI DNA ligase I genomic fragments as inserts.

The identity of the clones was confirmed by probing Southern blots of the genomic clones with probes specific for the 13.5kb and 10.0kb ligase genomic fragments. A 0.4kb PvuII/PstI fragment from the 3' end of the mouse DNA ligase I cDNA is specific for the 13.5kb fragment (chapter 6, figure 6.4c) and hybridised to this fragment in five of the clones (figure 7.1b). When a 0.35kb BamHI/PvuII cDNA fragment specific for the 10.0kb genomic fragment (chapter 6, figure 6.4b) was used as a probe it hybridised to a 10.0kb fragment in the remaining two clones (figure 7.1c). No additional ligase-specific fragments were detectable by probing with the full-length mouse DNA ligase I cDNA (figure 7.1d). The 13.5kb fragment was subcloned into the EcoRI site of pUC8 and, since this fragment was the one of interest, nothing further has yet been done with the 10.0kb fragment.
Figure 7.1: DNA ligase I genomic clones. a) Inserts cut out of \( \lambda \) EMBL4 cloning vector with EcoRI and electrophoresed through a 0.8% agarose gel. Positions of the 19.9kb and 8.8kb \( \lambda \) arms, and 10.0kb and 13.5kb inserts are indicated. Tracks 3 and 4 show the 10.0kb fragment. All other tracks show the 13.5kb fragment. Multiple inserts are visible in tracks 2, 3, 4, 5 and 6. Equivalent fragments in different lanes have not migrated evenly, possibly due to unequal loading of the gel. b) Southern blot of the DNA gel probed with the 0.4kb PvuII/PstI fragment of mouse DNA ligase I cDNA specific for the 13.5kb genomic fragment. This fragment is visible in tracks 1, 2, 5, 6 and 7. A second band visible in track 6 above the 13.5kb band is probably due to incomplete cutting. c) Southern blot of the DNA gel stripped and reprobed with the 0.35kb BamHI/PvuII fragment of mouse DNA ligase I cDNA specific for the 10.0kb genomic fragment. This fragment is visible in tracks 3 and 4. Again, the two extra larger bands are probably due to incomplete cutting. d) Southern blot of the DNA gel stripped and reprobed with full-length mouse DNA ligase I cDNA. Extra bands are probably due to partial digestion. STD: molecular weight markers 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb, 1.4kb, 1.1kb, 0.9kb.
7.2 Restriction Mapping and Intron/Exon Structure of the 13.5kb EcoRI Genomic Fragment

More detailed information was required on the structure of the 13.5kb genomic clone before targeting of the mouse DNA ligase I gene could commence. Initially, several restriction enzymes with rare recognition sites were tested for cutting the 13.5kb genomic fragment and, in addition to those sites already identified, sites were found for the enzymes BstEII, EcoRV, KpnI and SstI (figure 7.2).

The Southern blotting analysis carried out previously (chapter 6) had shown that the 13.5kb fragment contains the last five of the exons which were detectable in the mouse DNA ligase I gene (chapter 6, figure 6.6). For the purposes of this analysis, these exons are numbered from 7 to 11, although there are almost certainly many more than eleven exons present within the gene (chapter 6). Three of these exons still remained to be positioned unambiguously within the fragment with respect to restriction sites. Figure 7.3 shows that exon 8 has three possible positions, exon ten has two, and exon eleven has four possible locations. To identify the exact locations of these exons, the 13.5kb genomic clone was cut with each enzyme for which a recognition site has been identified (except BstEII and EcoRV, because these sites are located very close to an SstI site), singly, and also in double digests. The resulting fragments were probed with a 1.25kb Sall/EcoRI (3') fragment of mouse DNA ligase I cDNA which hybridises to exons 7 to 11 of the mouse DNA ligase I gene. (figure 7.4). The reasoning behind this experiment was that only those restriction fragments containing exons would hybridise to the probe. Figure 7.3 shows that there are three possible locations for exon number 8. It could lie within a 0.6kb HindIII/SstI fragment, in a 0.8kb SstI/BamHI fragment or within a 0.7kb BamHI/HindIII fragment. These possibilities were resolved by digestion with BamHI and HindIII which produced 9 fragments (figure 7.4a). Two of these fragments are of interest, a 1.3kb HindIII/BamHI fragment which contains the first two possible locations for exon 8, and the 0.7kb BamHI/HindIII fragment which is the third possible location for exon 8. The 1.3kb band did not hybridise to the probe, while the 0.7kb fragment did hybridise (figure 7.4b). Exon number 8 therefore lies within the 0.7kb BamHI/HindIII fragment (Figure 7.5). There are two KpnI sites within the 13.5kb fragment, producing KpnI fragments of 14.7kb and 1.4kb from the whole plasmid (13.5kb fragment in pUC8). Figure 7.3 shows that exon 10 could be located in two positions, either within the 1.4kb KpnI fragment or upstream of it. When the KpnI fragments were blotted and probed with the mouse DNA ligase I
Figure 7.2: Map of the 13.5kb EcoRI genomic fragment of mouse DNA ligase I showing restriction sites. For reference, the sizes of all BamHI and HindIII restriction fragments are given. The sizes of some of these fragments are slightly different to those given in chapter 6 because cloning of the 13.5kb EcoRI genomic fragment allowed more accurate restriction mapping to be carried out.

B: BamHI
BE: BstEII
H: HindIII
K: KpnI
RI: EcoRI
RV: EcoRV
S: SstI
Figure 7.3: Map of the 13.5kb EcoRI genomic fragment of mouse DNA ligase I showing restriction sites. This map gives information about the clone prior to any Southern blotting analysis to precisely locate exons. Exons for which the position with respect to restriction sites is known are filled in, others are hatched and the alternative positions given. For reference, the sizes of all BamHI and HindIII restriction fragments are given.

B: BamHI
BE: BstEII
H: HindIII
K: KpnI
RI: EcoRI
RV: EcoRV
S: SstI
Figure 7.4: Mapping positions of the exons within the 13.5kb genomic fragment. a) The whole plasmid (13.5kb EcoRI fragment cloned into the EcoRI site of pUC8) was cut with BamHI, HindIII, KpnI, SstI, BamHI/HindIII, BamHI/KpnI, BamHI/SstI, HindIII/KpnI, HindIII/SstI and KpnI/SstI and electrophoresed through a 0.8% agarose gel. b) Southern blot of the DNA gel probed with a 1.25kb SalI/EcoRI (3') fragment of mouse DNA ligase I cDNA. Sizes are given in kb. The bands in the HindIII/KpnI digest are due to only partial cutting.
Figure 7.5: Final map of the 13.5kb EcoRI genomic fragment of mouse DNA ligase I. This fragment is shown cloned into pUC8 (not to scale) and the positions of BamHI and HindIII sites within the polylinker are indicated. Restriction fragment sizes are given for reference to figure 7.4. Hatched boxes indicate exons. Subfragments used for subsequent PCR analysis are indicated by 1, 2, 3 and 4. The position of an arginine residue equivalent to the site of the Arg771 to Trp mutation is marked by R.

B: BamHI
BE: BstEII
H: HindIII
K: KpnI
RI: EcoRI
RV: EcoRV
S: SstI
cDNA, only the 14.7kb fragment hybridised to the probe (figure 7.4b). The 1.4kb KpnI fragment therefore does not contain any exons and exon 10 must consequently be located in the position upstream of this fragment (figure 7.5). Similarly, there are four possible positions for exon 11 (figure 7.3). However two of these locations can be eliminated because they lie within the 1.4kb KpnI fragment which has already been shown not to contain any exons. The remaining two possibilities are (1) within a 1.0kb SstI fragment which overlaps the 1.4kb KpnI fragment, and therefore to the 3' side of the KpnI site contained therein, and (2) within a 2.4kb SstI/HindIII fragment. Digestion of the plasmid with SstI and HindIII produced 9 fragments (figure 7.4a), including the 1.0kb and 2.4kb fragments of interest. Only the 1.0kb fragment hybridised to the probe, while the 2.4kb fragment did not. This therefore positions exon 11 within the 1.0kb SstI fragment and, since it must lie 3' of the KpnI site, the exon is contained in the 0.6kb KpnI/SstI fragment. The locations of these three exons were confirmed by all the other digests. There were however two problems with this analysis. First, the HindIII/KpnI digest was only a partial and was therefore not used. Secondly, hybridisation of exon number 7 was difficult to detect in many cases. This exon is present in a 1.6kb fragment from the 5' end of the 13.5kb EcoRI genomic fragment (figure 7.3). Throughout the analysis in chapter 6 this fragment hybridised very faintly, but always visibly to mouse DNA ligase I cDNA probes. However, during this analysis, hybridisation of the 1.6kb HindIII fragment was never detected. Also, in other digests which produced a fragment containing only this exon and no others, the fragment containing exon 7 only hybridised to the probe very faintly. For example a BamHI digest produced a 6.9kb fragment containing exon number 7, which bound to the probe much less than the other hybridising fragments (figure 7.4b). That this fragment hybridises at all indicates that it must contain an exon, and a BamHI/HindIII double digest which cuts this fragment down to 3.9kb HindIII, 1.6kb HindIII and 1.3kb HindIII/BamHI fragments demonstrates that it is not either of the fragments flanking the 1.6kb HindIII fragment that carries the exon, since these do not hybridise either (figure 7.4b). BamHI/KpnI, BamHI/SstI, KpnI/SstI and SstI digests produce the same result. Therefore the results from chapter 6 indicate that exon 7 must be located within the 1.6kb fragment, while the results presented in this chapter are consistent with this location, but do not actually confirm it.
7.3 Identification of the Arginine to Tryptophan Mutation Site in the 13.5kb Mouse EcoRI Genomic Fragment

Having obtained further structural information for the 3' end of the gene, the next step was to locate the actual site within the mouse gene equivalent to Arg771 in human DNA ligase I. Since no relevant mouse DNA ligase I sequence was available, twenty base primers (887Y and 888Y) were designed from the human cDNA sequence (Barnes et al 1990, Noguiez et al 1992) specific for each end of the 153bp human exon 23 which carries the Arg771 to Trp mutation. PCR was performed using these primers on the 13.5kb genomic fragment itself, as well as upon four sub-fragments from the region containing the eighth, ninth and tenth mouse exons (fragments 1, 2, 3 and 4, figure 7.5). A PCR product of about 150 bp was seen in the presence of the whole 13.5kb genomic fragment as well as sub-fragments 2 and 3 (figure 7.6). The only exon common to these two fragments is the eighth exon, which therefore contains sequence equivalent to human exon number 23. There must be a very high degree of homology between mouse and human sequences in this region, as shown by the fact that the human-specific primers hybridised to the mouse DNA, and produced an amplified DNA product which co-migrates on a gel with the PCR product from human exon 23 (data not shown).

The same primers were used to sequence the mouse DNA corresponding to human exon 23 (figure 7.7). The sequence obtained was identical to that obtained by sequencing of the mouse DNA ligase I cDNA in this region. Of 153 bases there are 17 changes between mouse and human resulting in only two alterations at the amino acid level. Both alterations are conservative, a change from alanine to serine and from aspartate to glutamate. The other nucleotide changes are almost all silent third base alterations, therefore this region of DNA ligase I is extremely highly conserved and is probably crucial for proper enzyme function. It is therefore unsurprising that the Arg771 to Trp mutation identified here in 46BR adversely affects human DNA ligase I activity. There is an arginine codon in the mouse sequence in exactly the same position as that at which the mutation occurred in 46BR and it is highly probable that an equivalent alteration in the mouse gene would have the same detrimental effect.

Sequencing also indicated that, for this exon at least, the intron/exon boundaries are identical to those in the human gene (Noguiez et al 1992), with the flanking nucleotides also being homologous. This supports the theory that the mouse and human DNA ligase I genes both have the same exon structure, with 28 exons in each. The sequenced exon conforms to the GT-AG splicing consensus (Padgett et al 1986) with a GT splice donor dinucleotide and an AG acceptor.
Figure 7.6: PCR analysis of the 13.5kb EcoRI genomic fragment of mouse DNA ligase I. DNA fragments plus primers were denatured for 1min at 94°C, annealed for 30s at 69°C and synthesis was carried out at 72°C for 30s. Control reactions were carried out with no DNA or with no primers. C: control, ML: PCR on mouse DNA ligase I cDNA, 13.5: PCR on the 13.5kb EcoRI genomic fragment of mouse DNA ligase I, 1, 2, 3 and 4: PCR on genomic sub-fragments (see figure 7.3). The exon 23-specific product is indicated by an arrow. Sizes are in kb.
Figure 7.7: Nucleotide and amino acid sequence of human exon 23 and the equivalent mouse exon. Human Arg771 is circled. Base and amino acid alterations are underlined. Flanking bases from the introns are given in lower case.
<table>
<thead>
<tr>
<th>Protein Sequence</th>
<th>Human DNA Sequence</th>
<th>Mouse DNA Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>L K K D Y L D G V G D T L D L V V</td>
<td>ccagCTGAAGAAGGACTACCTTGATGGCGTGAGGAGCCCTGGACCTGGTGT</td>
<td>ccagCTGAAGAAGGACTACCTTGATGGCGTGAGGAGCCCTGGACCTGGTGT</td>
</tr>
<tr>
<td>I G A Y L G R G K K A G R Y G G F</td>
<td>ATCGGCCTACCTGGGCGGGGAAGGCGGCCTGGGCTTACGCGGGGC</td>
<td>ATGGTGCCCTACCTGGGCGGGGAAGGCGGCCTGGGCTTACGCGGGGC</td>
</tr>
<tr>
<td>L L A S Y D E D S E E L Q A I C K</td>
<td>CTGCTGGCATTCCTACGAGGACAGTGAAGGATGCAGGCCATATGCAAGG</td>
<td>CTGTTGCTACATGAGGAGAAGATTACGGCTCAGGTATATGCAGG</td>
</tr>
<tr>
<td>L L A A Y D E E S E E L Q A I C K</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 8

GENE TARGETING AT THE MOUSE DNA LIGASE I LOCUS
A deleterious mutation is present on each allele of the DNA ligase I gene in the patient from whom the cell line called 46BR was derived, and these cells are therefore deficient in DNA ligase I. This deficiency has given rise to a number of abnormalities, both clinical and cytological, some of which overlap with the symptoms of Blooms syndrome, also shown to involve DNA ligase I deficiency. A mouse model for BS could prove invaluable in studying the disease, however no mutation has yet been identified in cell lines established from these patients. It is therefore not yet possible to carry out gene targeting experiments to generate a model of classical BS. 46BR may represent a variant form of Blooms syndrome, but whether this is true, or not, awaits identification of the underlying defect in BS. If this is the case, introduction of one of the mutations identified in 46BR into the mouse DNA ligase I gene could yield a model for a form of Blooms syndrome. However, even if the defects in the cell line 46BR are unrelated to Blooms syndrome and represent a unique case, it is of interest to generate a mouse model for DNA ligase I deficiency.

A 13.5kb EcoRI fragment of the mouse DNA ligase I gene containing an arginine codon equivalent to human Arg771 has been cloned and the site of this codon within the fragment has been identified. In the human cell line 46BR this codon has mutated on one allele of the gene, and now specifies tryptophan, an amino acid alteration which significantly reduces the activity of the enzyme. This EcoRI fragment is to be used to introduce an arginine to tryptophan mutation in the equivalent position to that in 46BR by gene targeting. This is to be done in two steps. First a null mutation will be introduced, knocking out one allele of the gene completely, by removal of several exons, followed by a second targeting step in which the exons will be restored, but specifying an Arg to Trp mutation.

It is necessary to carry out this targeting in two steps, rather than one, because it would be impossible to screen large numbers of cells efficiently for a single base alteration. Screening for the loss of several exons and then for their replacement along with the mutation, is much simpler.

Following my work on the structure of the mouse DNA ligase I gene, it was originally proposed that I would carry out the first targeting step. However this proved not to be possible, due to lack of time. This work was instead carried out by others in the laboratory and I would therefore like to thank Carolanne McEwan for construction of the vector and PCR, Anne-Marie Ketchen for cell culture work and Jim Selfridge for PCR, Southern analysis, embryo injection and mouse work. Since I did not actually do the gene targeting work myself, this chapter is intended to be an overview rather than a detailed description of the experiments.
8.1: Targeting Vector Construction

A targeting vector was to be constructed which would remove several exons, including the exon carrying sequence equivalent to the human Arg771 to Trp mutation, from the 3' end of one allele of the DNA ligase I gene in mouse embryonic stem cells, thus creating a null mutation. Since targeting events are extremely rare, a method of enrichment for this happening was also required. This was to be provided by positive-negative selection (Mansour et al 1988) based on the method used by Stacey et al in 1994. The cells in which the gene was to be targeted are male and HPRT-deficient, therefore a targeting vector which introduces a functional copy of the HPRT gene will allow cells containing the plasmid to be selected by means of HPRT activity. Cells which have integrated the plasmid will express functional HPRT and will consequently be able to grow in HAT medium, while non-expressing cells will die (positive selection). However, the plasmid DNA will not always integrate by homologous recombination, but will also integrate randomly to generate non-targeted HAT-resistant cells. A method of selection against random integrants is therefore also necessary (negative selection). Random integration generally occurs via DNA ends in which case the whole plasmid is inserted. Therefore if a plasmid is designed so that when linearised it will have a negatively selectable marker at one end, it should be possible to select against random integrants. In this case the marker to be used was the viral thymidine kinase gene. Cells expressing this gene can be selected against using gancyclovir which is a nucleoside analogue specific for the viral enzyme.

With this in mind, the targeting vector was designed so that homologous recombination between two regions of homology with the mouse DNA ligase I gene would remove the four 3' exons and replace them with a functional HPRT minigene. A viral thymidine kinase gene was placed adjacent to one region of homology, providing selection against random integration of the plasmid.

Targeting vector construction was as follows (figure 8.1). First, the 13.5kb EcoRI mouse DNA ligase I genomic fragment cloned in pUC8 was digested with SstI and religated. This removed two SstI fragments which carry the last four exons of the gene. Following SstI digestion, all but one of the BamHI sites contained within the 13.5kb genomic fragment had been removed. This left one BamHI site in the 3' untranslated region, and one in the pUC8 polylinker. The vector was next digested with BamHI and religated, which removed all but 1.8kb of the 3' flanking region. Therefore the 13.5kb genomic fragment had been modified by removal of the
last four exons and of the 3' end. This left 3.3kb from the 5' end of the fragment, containing the single remaining exon, and 1.8kb of the 3' flanking region.
Figure 8.1: Targeting vector construction. The starting material was the 13.5kb EcoRI genomic fragment (open box) cloned in pUC8. This fragment contains the last five exons (hatched boxes) of the mouse gene. First, the plasmid was digested with the enzyme SstI and religated. This removed a region of DNA containing the last four exons. Next, the plasmid was digested with BamHI and religated, thus removing the 3' end of the genomic fragment. A 2.7kb fragment containing an HPRT minigene under control of a phosphoglycerate kinase promoter (shaded boxes) was then cloned into the now unique SstI site. This was followed by the cloning of a 2.8kb fragment containing a thymidine kinase gene driven by a phosphoglycerate kinase promoter (filled-in box) into the now unique BamHI site. It was cloned in as a BamHI/BclII fusion at the upstream end of the fragment to preserve a unique BamHI site at the downstream end. The resulting vector contains two regions of homology with the mouse DNA ligase I gene, of 3.3kb and 1.8kb. These are interrupted by the HPRT minigene, and a thymidine kinase gene lies 3' of the 1.8kb region of homology. This vector is called MLTV (mouse ligase targeting vector). Key restriction sites are shown. R marks the site of the codon equivalent to that specifying Arg771 in the human gene. This codon is not present in MLTV.

B: BamHI
Bcl:BclII
H: HindIII
K: KpnI
Rl: EcoRI
S: SstI
3.3kb  2.7kb  1.8kb  2.8kb

MLTV
An HPRT minigene, with HPRT controlled by the mouse phosphoglycerate kinase promoter (Magin et al 1992) was cloned into the SstI site of the modified genomic fragment, in place of the missing four exons. A viral thymidine kinase gene, also driven by the mouse phosphoglycerate kinase promoter, was cloned into the BamHI site by a BamHI/BclI fusion at the 5' end of the fragment and as a simple BamHI ligation at the 3' end, thus preserving a unique BamHI site at the 3' end of this marker for linearisation purposes.

The final targeting vector, constructed in pUC8, consists of two blocks of homology with the endogenous mouse DNA ligase I gene, of 3.3kb and 1.8kb, interrupted by an active HPRT minigene, and with a viral thymidine kinase gene placed adjacent to the 1.8kb block of homology (figure 8.1). There are two unique linearisation sites within the vector, an EcoRI site at the 5' end of the 3.3kb block of homology, and a BamHI site at the 3' end of the thymidine kinase gene. This vector was designated MLTV (mouse ligase targeting vector).

8.2: Targeting the Mouse DNA Ligase I Gene

Figure 8.2 shows the predicted alteration of the mouse DNA ligase I gene following targeting with MLTV. Recombination occurs between the two regions of homology flanking the HPRT minigene, and a region of the endogenous DNA ligase I gene containing the last four exons is replaced with HPRT. In correctly targeted cells, no other part of the vector should be inserted into the genome and therefore the thymidine kinase gene is lost along with other vector sequences. The targeted gene consequently lacks four exons from the 3' end, but retains 3' flanking sequences.

MLTV was linearised by EcoRI digestion and introduced into mouse embryonic stem cells by electroporation. Transfected cells were enriched for targeting events by growth in HAT medium supplemented with gancyclovir. Cells surviving this selection were then screened for targeting of the DNA ligase I gene by PCR. Screening for targeted alleles of the gene was achieved using one primer specific for the 3' end of the HPRT minigene, and one specific for the 3' flanking region of the gene, 1.8kb from the predicted site of HPRT insertion (figure 8.3). These primers should therefore give a PCR product of 1.8kb in size. The second primer is specific for sequence outwith that included in the targeting vector - its sequence is present within the BamHI fragment deleted from the vector during its construction. Consequently, only correctly targeted cells can give rise to this PCR product since one primer is specific for vector-derived sequences, and the other is specific for endogenous DNA ligase I sequences. 115 clones resistant to HAT and
Figure 8.2: Targeting the mouse DNA ligase I gene in HPRT− embryonic stem cells. The linearised vector, MLTV, is shown lying adjacent to the homologous region of the mouse DNA ligase I gene. A section of the gene containing exons 8 to 11 is depicted as looping out since it has no homology with MLTV. Homologous recombination should occur between the two regions of homology, around exon 7 and in the 3' flanking region. The result of this recombination event is shown. Exons 8 to 11 have been lost from the gene and replaced with the HPRT minigene (shaded boxes). No other part of MLTV has integrated into the gene, therefore the thymidine kinase gene (closed box) has been lost. The resulting cells will be HPRT+ due to integration of the HPRT minigene and will therefore be able to grow in HAT medium. They will also be resistant to gancyclovir since the thymidine kinase gene is not present.
Figure 8.3: Screening for targeting events at the mouse DNA ligase I locus. A map of the 13.5kb EcoRI genomic fragment used in targeting vector construction is shown. The sizes and positions of EcoRI, BamHI and HindIII restriction sites detectable by hybridisation to mouse DNA ligase I cDNA are indicated. Arrows pointing upstream indicate that the 10.7kb HindII and 9.4kb BamHI fragments overlap with the 13.5kb EcoRI fragment and the next EcoRI fragment upstream. The structure of part of the modified gene, following targeting, is shown below. This section of the gene is all that is left of the 13.5kb fragment (see figure 8.2 for details). The remaining hybridising fragments are indicated below the diagram. Above, the two arrows mark the positions at which the PCR primers used for screening will bind. One is specific for the 3' end of the HPRT minigene, while the other is specific for the 3' flanking region of the gene, within a BamHI/HindIII fragment. The 1.8kb PCR product, spanning the junction between HPRT and the continuing gene, is indicated. Open boxes: mouse DNA ligase I gene, hatched boxes: exons, shaded boxes: HPRT minigene. Key restriction fragments are given.

B: BamHI
H: HindIII
S: SstI
RI: EcoRI
gancyclovir were screened by this method and three gave rise to the 1.8kb PCR product (figure 8.4). These positives were named numbers 12, 53 and 106.

The three positive clones were further analysed by Southern blotting. Figure 8.3 shows the BamHI and HindIII fragments which are contained within the 13.5kb EcoRI genomic clone, as well as those fragments which overlap with it. Also shown are the fragments which are predicted to be present following a targeting event. It can be seen from this figure that mouse DNA ligase I cDNA should hybridise to the 13.5kb EcoRI fragment; 10.7kb, 1.8kb, 2.1kb, 1.7kb and 4.2kb HindIII fragments and 9.4kb, 1.2kb, 1.1kb and 3.9kb BamHI fragments from the untargeted gene. However, in the targeted gene, exchange of a 4.9kb internal fragment for the 2.7kb HPRT minigene should alter these fragment patterns. It would therefore be predicted that the 13.5kb EcoRI fragment would be reduced to a 11.3kb fragment, while the 2.1kb, 1.7kb and 4.2kb HindIII fragments would disappear. Similarly, the 1.2kb, 1.1kb and 3.9kb BamHI fragments would be lost and the 9.4kb fragment would be extended to a 13.1kb fragment.

However this analysis is complicated by the fact that there are two copies of the gene since DNA ligase I is autosomal. Therefore all of the fragments from the wild-type gene will be seen in addition to those derived from the targeted gene. A Southern blot should consequently show a wild-type pattern of bands, but with additional 11.3kb EcoRI and 13.1kb BamHI fragments. The other alterations should be detectable as differences in band intensities. For example, in targeted cells the intensities of the 2.1kb, 1.7kb and 4.2kb HindIII bands should be half that of those in untargeted cells.

Figure 8.5 shows the results of probing a Southern blot of genomic DNA prepared from each of the three clones with a 2.2kb BamHI/EcoRI fragment of mouse DNA ligase I cDNA (the 3' two thirds of the cDNA). Table 8.1 summarises the predicted and actual results from this experiment. When the DNA was digested with EcoRI only three bands were detectable from clones 53 and 106 as well as control wild-type DNA. These are 13.3kb, 10.0kb and 5.2kb in size, with the two smaller bands arising from parts of the gene not affected by the targeting, but detectable with the probe used (see chapter 6). No extra 11.3kb fragment arising from the targeted gene was visible. However, in the DNA from the targeted cells, the intensity of the 13.5kb EcoRI fragment was reduced compared to that in the DNA from the non-targeted cells. It therefore appears that the 13.5kb fragment was lost following targeting, as predicted, and it is likely that the expected 11.3kb fragment is present after all, probably as an unresolved band due to comigration with the 10.0kb fragment (since if it comigrated with the 13.5kb fragment there would be no
Figure 8.4: Screening for targeting events using PCR. A sample of the clones able to grow in HAT medium supplemented with gancyclovir is shown (numbers 50-60). Genomic DNA prepared from each clone was used for PCR with the primers described in figure 8.3. The 1.8kb product indicated by an arrow in the clone 53 track is specific for a targeting event. Its absence in all the other clones indicates that they have not been correctly targeted. HM-1: genomic DNA was prepared from the wild type embryonic stem cell line in which the targeting experiments were carried out. STD: Molecular weight markers of sizes 23.1kb, 9.4kb, 6.6kb, 4.4kb, 2.3kb, 2.0kb and 1.6kb down to 0.7kb at 100bp intervals.
Figure 8.5: Screening for targeting events by Southern blotting. Genomic DNA was prepared from each of the targeted clones identified by PCR (clones 12, 53 and 106). DNA was then digested with either EcoRI, BamHI or HindIII, electrophoresed through an agarose gel, blotted and probed with mouse DNA ligase I cDNA. HM-1: genomic DNA was prepared from the wild type embryonic stem cell line in which the targeting experiments were carried out.
Table 8.1: Genomic fragments arising from the mouse DNA Ligase I gene following targeting. Fragment sizes in brackets are those fragments arising from parts of the gene outwith the 13.5kb EcoRI fragment used in the targeting experiments. Fragments marked with an asterisk are those which were present at a reduced intensity in the DNA from targeted cells compared to non-targeted cells.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Fragments Predicted from Wild-Type Gene</th>
<th>Fragments Predicted from Targeted Gene</th>
<th>Fragments Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI</td>
<td>13.5kb</td>
<td>11.3kb</td>
<td>13.5kb*</td>
</tr>
<tr>
<td></td>
<td>(10.0kb)</td>
<td>(10.0kb)</td>
<td>(10.0kb)</td>
</tr>
<tr>
<td></td>
<td>(5.2kb)</td>
<td>(5.2kb)</td>
<td>(5.2kb)</td>
</tr>
<tr>
<td>BamHI</td>
<td>9.4kb</td>
<td>13.1kb</td>
<td>13.1kb</td>
</tr>
<tr>
<td></td>
<td>3.9kb</td>
<td>(3.6kb)</td>
<td>9.4kb*</td>
</tr>
<tr>
<td></td>
<td>(3.6kb)</td>
<td>(1.9kb)</td>
<td>3.9kb*</td>
</tr>
<tr>
<td></td>
<td>(1.9kb)</td>
<td></td>
<td>(3.6kb)</td>
</tr>
<tr>
<td></td>
<td>1.2kb</td>
<td></td>
<td>(1.9kb)</td>
</tr>
<tr>
<td></td>
<td>1.1kb</td>
<td></td>
<td>1.2kb*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.1kb*</td>
</tr>
<tr>
<td>HindIII</td>
<td>10.7kb</td>
<td>10.7kb</td>
<td>10.7kb</td>
</tr>
<tr>
<td></td>
<td>4.2kb</td>
<td>(3.9kb)</td>
<td>4.2kb*</td>
</tr>
<tr>
<td></td>
<td>(3.9kb)</td>
<td>1.8kb</td>
<td>(3.9kb)</td>
</tr>
<tr>
<td></td>
<td>2.1kb</td>
<td></td>
<td>2.1kb*</td>
</tr>
<tr>
<td></td>
<td>1.8kb</td>
<td></td>
<td>1.8kb</td>
</tr>
<tr>
<td></td>
<td>1.7kb</td>
<td></td>
<td>1.7kb*</td>
</tr>
</tbody>
</table>
reduction in intensity). The same pattern is visible in the clone 12 track, but there is an additional unidentified larger fragment present.

When the genomic DNA was digested with BamHI, seven fragments hybridised to the probe from clones 53 and 106. These are 13.1kb, 9.4kb, 3.9kb, 3.6kb, 1.9kb, 1.2kb and 1.1 kb in size, with the 3.6kb and 1.9kb bands arising from other parts of the gene not involved in the targeting (see chapter 6). The 13.1kb fragment is the product expected to arise from the correctly targeted gene and is not present in the DNA from non-targeted cells. The other predictions are also borne out, since the 9.4kb, 3.9kb, 1.2kb and 1.1kb bands are all present at a reduced intensity, as would be expected if they were lost following targeting. The same pattern of bands is present in the clone 12 track, except that the predicted 13.1kb fragment is absent and has been replaced with an unidentified larger fragment.

Digestion with HindIII produced the six bands expected to be seen, of sizes 10.7kb, 4.2kb, 3.9kb, 2.1kb, 1.8kb and 1.7kb, with the 3.9kb fragment arising from a part of the gene upstream from the site of targeting. The correctly targeted gene was expected to be detectable only by a reduction of certain band intensities, since no new fragments should be generated which could hybridise to the probe used. Unequal loadings between the non-targeted and targeted DNA meant that comparisons were difficult to make. However, by comparing the 10.7kb and 1.8kb bands it can be seen that there is very much less wild-type DNA present on the blot than DNA from clones 12 and 53, yet the intensities of the 4.2kb, 2.1kb and 1.7kb bands are very similar in all three tracks. The 10.7kb and 1.8kb bands should be unaffected by targeting, but the 4.2kb, 2.1kb and 1.7kb bands were predicted to be lost following targeting. For these three bands to be present at the same intensity in wild-type and targeted DNA indicates that there must have been a reduction in their intensity in the clone 12 and 53 tracks which contain more DNA overall than the wild-type tracks. Therefore both clones 12 and 53 are shown by HindIII digestion to have been correctly targeted.

On the basis of the results obtained, together with the PCR analysis, clones 53 and 106 are correctly targeted, while the situation with clone 12 is uncertain. It has probably been targeted with a tandem plasmid integration. Clones 53 and 106 were therefore used for further work.

8.3: Mouse Models for DNA Ligase I Deficiency

The gene targeting experiments were carried out in mouse embryonic stem cells so that mice carrying the altered gene could then be generated. ES cells are pluripotent, with the result that, if they are injected into embryos they can contribute
to all the tissues of the resulting animal, which is then called a chimaera. If they have also contributed to the germ line, the animal will be able to pass on the genes from the injected cells to its offspring. This means that it is possible, by breeding, to generate animals which are homozygous for a gene which was targeted at only one allele in the original embryonic stem cells.

Mouse embryos were injected with the targeted clones 53 and 106. Chimaeras were generated from both clones, although the best chimaera was judged to have arisen from clone 53. This chimaera was bred to produce mice heterozygous for the DNA ligase I null mutation. These mice are apparently phenotypically normal, indicating that the single active allele of DNA ligase I is sufficient for the requirements of the cell. These mice are now being mated in attempts to produce mice homozygous for the mutation. Such mice would be completely lacking in DNA ligase I activity and it is uncertain whether or not they would be viable. To date 10 homozygous wild-type mice and 12 heterozygotes have been obtained from these crosses. No homozygous mutants have been observed, so they are apparently unable to survive even as far as birth. The findings so far therefore suggest that mice homozygous for the DNA ligase I null mutation are not viable. This is not unexpected, since DNA ligase I is thought to have a crucial role in both replication and repair of DNA.

Work is currently underway to perform the second stage of the targeting. The deleted exons will be reintroduced into the gene, carrying the mutation specifying an arginine to tryptophan alteration. Such targeted cells will then be used to make a '46BR' mouse.
46BR is a human cell line carrying a damaging mutation on each allele of the DNA ligase I gene and it is therefore defective in the activity of this enzyme (Barnes et al 1992b, Prigent et al 1994). It has the same aberrant phenotype as its transformed derivative, called 46BR.1G1 (Lehmann et al 1988). 46BR.1G1 is sensitive to ethylmethane sulphonate and 3-aminobenzamide, and this sensitivity is rescued by transfection with wild-type human DNA ligase I sequences. A direct link therefore exists between the DNA ligase I deficiency and the abnormal cellular phenotype of 46BR. Cell lines derived from Blooms syndrome patients are also defective in DNA ligase I activity (Willis and Lindahl 1987, Chan et al 1987, Willis et al 1987) and it is interesting that 46BR also displays other abnormalities seen in BS cell lines, including sensitivity to certain DNA damaging agents (Teo et al 1982, 1983a, Hook et al 1984, Arlett and Harcourt 1978). Blooms syndrome cell lines and 46BR also exhibit some significant differences. Most notable of these dissimilarities is the fact that 46BR displays only a slightly elevated rate of sister chromatid exchanges (Teo et al 1983a). On the other hand, all BS fibroblast cell lines studied to date have a significantly elevated sister chromatid exchange level (Chaganti et al 1974), a property which is used in Blooms syndrome diagnosis. Additionally, the individual from whom the cell line 46BR was originally derived suffered from many of the symptoms of Blooms syndrome (Webster et al 1992, German and Passarge 1989). Like Blooms syndrome patients, she was sensitive to sunlight, severely immunodeficient and her growth was retarded. Her death at 19 years of age followed a series of lung infections, which BS patients also typically suffer from, and she had been diagnosed as having a malignant lymphoma although this was not confirmed. This is also characteristic of Blooms syndrome since development of cancer while still young is one of its most notable clinical features.

Such clinical similarities and the discovery that both 46BR and BS cell lines have abnormal DNA ligase I activity has led to the suggestion that 46BR may represent a form of Blooms syndrome. This is despite the fact that as yet, all BS cell lines studied appear to represent a single complementation group (Weksberg et al 1988), which 46BR is not a member of. The disease site in BS cells is apparently located on chromosome 15 (McDaniel and Schultz 1992), while the DNA ligase I gene is present on chromosome 19 (Barnes et al 1990, Barnes et al 1992a). 46BR does not simply denote another Blooms syndrome complementation group, however, since the cells display many properties not shared with BS cells. It may therefore represent a variant form of the disease.

In diseases such as xeroderma pigmentosum, a specific enzyme pathway is involved, in this case the excision repair pathway, (Cleaver 1968, Cleaver 1990 and
and the many complementation groups identified arise from the many gene products participating in that pathway. However, there is a variant form of XP, displaying typical clinical features but different cytological properties to the other seven XP complementation groups. It is thought that the incision steps of excision repair are defective in XP groups A to G, while in the variant form, it is post-replication repair that is affected. In this disease therefore, there are seven complementation groups displaying the same characteristics because they are defective at the same step of excision repair, and a variant form, showing slightly different cytological features because a different repair pathway is affected, but the same clinical features because the overall effect on repair is the same. A variant form of ataxia telangiectasia has also been identified (Jaspers et al. 1988), but the circumstances in this case are less well understood. The situation in Blooms syndrome appears to be more complex since DNA ligase I is only one of a number of enzyme abnormalities that have been identified and which apparently all arise from a single defect. In this case it seems that it is not a single enzymic pathway that is affected, but rather that there is a single defect affecting many functions. Further, DNA ligase I, itself, is not an enzyme involved in a single specific pathway, but instead appears to take part in many cellular processes. It seems likely that the overlap between BS and 46BR may have occurred simply because of a wide-ranging defect in BS cells which affects DNA ligase I among several other enzymes. Clarification of this point awaits identification of the defective locus in Blooms syndrome cells. 46BR could be regarded as a variant form of Blooms syndrome since, as with XP variant, it displays many typical clinical symptoms of the disease, together with different cellular features from classical disease cell lines. Further, the same enzyme is affected in both 46BR and BS cell lines. This classification probably also depends to some extent on the definition of a variant disease form.

A mouse model for Blooms syndrome would be extremely useful in the study of the disease, and possibly in the evaluation of potential therapies. However, since the primary defect in this disease remains unknown it is not yet feasible to create such a model. If 46BR is regarded as a variant form of Blooms syndrome then a mouse carrying the 46BR mutation would be a model for a type of BS. Unfortunately, even if this is the case, the model would not be particularly useful for purposes of studying BS since the underlying defect is different and, as far as is known, 46BR is a unique case. It could, nevertheless, still be invaluable in the study of DNA ligase I, its function during mammalian growth and development, and the consequences of its deficiency.
Mammalian DNA ligase I is very highly conserved. Mouse and human enzymes are 84% homologous at the amino acid level and there are very few sequence differences within the C-terminal domains. More variation occurs in the N-terminal domains, with two regions of nonhomology in particular. The C-terminal domain is involved in catalysis (Tomkinson et al 1990, Kodama et al 1991) and the N-terminal domain is thought to be involved in regulation (Prigent et al 1992). Given that DNA ligase I apparently has such a vital role within the cell, it is not surprising that the structure of the catalytic domain should be so highly conserved. If the role of the N-terminus is purely regulatory, and if this occurs via interaction with other proteins, then some sequence divergence would be expected. DNA ligase I is known to interact with a number of proteins. Heat-resistant and heat-labile stimulatory factors have been identified, as well as a heat-labile inhibitor (Kenne and Ljunquist 1988, Fairman et al 1992, Yang et al 1992, Yang et al 1993). It seems probable that these proteins interact with regions of the N-terminal domain and it is interesting to speculate that the two stretches of nonhomology in this domain may be binding sites for one or more of them. If the regulatory factors have not been as well conserved as DNA ligase I during evolution, their recognition sequences may also have diverged with the result that regions of nonhomology would arise within an otherwise highly conserved enzyme. As well as these unidentified protein factors, mammalian DNA ligase I can be phosphorylated by casein kinase II and this phosphorylation is a requirement for enzyme activation (Prigent et al 1992). Casein kinase II is known to interact with the N-terminal domain of DNA ligase I since the C-terminal domain is active independently of phosphorylation. Casein kinase II is a ubiquitous enzyme and therefore its recognition sites would be expected to be conserved. Indeed, a number of concensus casein kinase II phosphorylation sites are present throughout the sequences of both mouse and human enzymes. The human enzyme also contains two particularly strong concensus sequences and it has been suggested that one of these is the site of phosphorylation (Prigent et al 1992). However, the mouse amino acid sequence lacks both of these sites indicating that phosphorylation occurs elsewhere. Apart from these differences there are apparently no other major changes between mouse and human DNA ligase I and the sites of the two mutations identified in the DNA ligase I sequence of 46BR are identical in the mouse enzyme.

Although the mouse and human enzymes are very similar, there has been a lot of divergence in their gene structures. The human gene is significantly larger, occupying 53kb of DNA (Noguiez et al 1992), while the mouse gene occupies only 39-42kb. The restriction fragment patterns of both genes are completely different,
but this is unsurprising since most of the sites investigated lie within introns and would therefore not be expected to be conserved. Superficially, the exon structures of the two genes are also unrelated, with the human gene consisting of 28 exons (Prigent et al 1992), while throughout this thesis the mouse gene has been referred to as having only 11. However, as was explained in the relevant chapter (chapter 6), this is a minimum number due to the nature of the analysis, and there is no evidence to suggest that the mouse gene does not also consist of 28 exons. It is noteworthy that the only exon which was looked at in detail (human exon 23) was found to have an exact parallel in the mouse gene. The equivalent mouse exon consists of an identical number of nucleotides with very few alterations. The splice sites are identical and even the first few bases of flanking intron sequence are conserved. It seems likely from this finding that the other exons are similarly conserved, and that the mouse gene is also made up of 28 exons.

The identity between human and mouse DNA ligase I sequences means that it is possible to introduce the same mutations into the mouse enzyme as are present in the human enzyme in 46BR by gene targeting, and thus go on to create a mouse model for the DNA ligase I deficiency of 46BR. The mutations in 46BR both arose as a result of a single base change (Barnes et al 1992b) which means that the targeting will introduce an extremely subtle alteration into the gene. Unfortunately, the DNA ligase I gene is not directly selectable and therefore some other method of screening for targeting events is necessary. Single base changes can sometimes be detected as the introduction or loss of a restriction enzyme recognition site. However this still requires the preparation, digestion, electrophoresis, blotting and probing of genomic DNA from every clone to be tested for targeting events. Such a method of screening is therefore prohibitively tedious for large numbers of clones. A better method, involving two targeting events has been devised for targeting the mouse DNA ligase I gene in male HPRT-deficient embryonic stem cells. The first step is to create a null mutation by replacement of the last few exons of the mouse DNA ligase I gene with an HPRT minigene which can be selected for. Next, the gene is to be targeted again, to replace the HPRT minigene with the missing exons, containing the equivalent of the Arg771 to Trp mutation of 46BR. The second targeting event can be selected for by loss of HPRT activity.

The mutation to be introduced into the mouse DNA ligase I gene specifies an alteration of an arginine residue (Arg768 in mouse) to tryptophan. The equivalent mutation in the human gene reduces the activity of the enzyme by an unknown mechanism (Barnes et al 1992b, Prigent et al 1994). It retains the ability to form a covalent enzyme-adenylate intermediate, indeed most is present in the adenylated
form *in vivo* (Prigent *et al* 1994). The mutation must therefore affect a step in the reaction mechanism that follows adenylation. Subsequent to adenylation, the AMP moiety is transferred to a 3’ phosphoryl group in a nicked DNA strand, which is then sealed accompanied by release of AMP (Lehman 1974). Transfer of AMP to DNA by human DNA ligase I carrying the Arg771 to Trp mutation is also apparently normal since accumulation of the DNA-AMP reaction intermediate can be measured (Prigent *et al* 1994). Therefore it must be the final step of ligation that is abnormal. How the mutation could have this effect is unclear. It is possible that the mutation itself directly affects ligation activity. Alternatively, inability to catalyse phosphodiester bond formation may be related to the altered aggregation properties of the mutant enzyme which seems to bind tightly to another similarly sized protein. (Lehmann *et al* 1988).

Replacement of the last few exons of the mouse DNA ligase I gene with an HPRT minigene has already been carried out, and the targeted cells used to generate a chimaera. No animals homozygous for the DNA ligase I null mutation have been bred from this chimaera, although homozygous wild-type and heterozygous animals have been obtained. The numbers of progeny which have been born suggest that some homozygous mutants should have been obtained if they are viable. It therefore appears that development is severely compromised in animals lacking DNA ligase I activity. Investigations are currently underway into how long DNA ligase I homozygous mutant embryos are able to survive before development is terminated.

A demonstration that DNA ligase I is essential for viability is not unexpected for several reasons. Consider the DNA ligase I-deficient cell line 46BR and its transformed derivative 46BR.1G1. During the transformation of 46BR one of the DNA ligase I mutations was lost and 46BR.1G1 consequently retains only the allele of the gene carrying the Arg771 to Trp mutation (Barnes *et al* 1992b). This leaves the enzyme with a reduced, but measurable amount of ligation activity and the cell line is therefore not totally DNA ligase I-deficient, maintaining about 1/20 of normal activity (Prigent *et al* 1994). The other mutation, Glu566 to Lys, renders the enzyme inactive by modifying the active site. That this mutation, which inactivates DNA ligase I, was the one to be lost during transformation suggested originally that it could be lethal in a homozygote. This is the reason why the Arg771 to Trp mutation was selected for gene targeting experiments.

A further reason for supposing that a homozygous DNA ligase I null mutation might be lethal is implied by the growth of 46BR. As a consequence of the reduced DNA ligase I activity, 46BR cells are extremely sickly in culture, growing at best, at only 1/4 the rate at which healthy wild-type cells grow. They are also
much smaller and grow in clumps, instead of spreading out as wild-type cells would. The cells are obviously struggling to grow, yet they retain some DNA ligase I activity. Similarly, the patient from whom 46BR was established was severely developmentally retarded, although she did have one copy of the DNA ligase I gene specifying an enzyme with residual activity. If the partial removal of DNA ligase I activity is so detrimental, then its complete absence would be expected to have very serious repercussions for the cell.

DNA ligase I is apparently involved in two vital aspects of cellular metabolism. The sensitivity of 46BR to a wide range of DNA damaging agents demonstrates that DNA repair is faulty in these cells, and therefore DNA ligase I must be required for DNA repair to take place (Teo et al 1982, 1983a). The induction of DNA ligase I messenger RNA in response to ultraviolet irradiation also suggests involvement in DNA repair (Montecucco et al 1992). However, DNA ligase I is also induced in response to cell proliferation, implying a role in DNA replication (Chan and Becker 1985, Söderhäll 1976). This is confirmed by the presence of DNA ligase I within DNA replication complexes which have been isolated from mammalian cells (Malkas et al 1990, Wu et al 1994), and also by the observation that joining of replication intermediates is delayed in 46BR cells (Henderson et al 1985). Such an enzyme, which is required for the processes of DNA repair and replication, would be expected to be essential to the cell, and therefore to the whole animal.

How does a deficiency of DNA ligase I activity make embryos inviable? Homozygous DNA ligase I null mutants would be expected to be defective in both DNA repair and DNA replication. In the '46BR' patient (Webster et al 1992), aberrant DNA repair was manifest in her sensitivity to sunlight and suspected cancer at death, but was not actually the cause of her death. She lived to the age of 19, and died of other causes before accumulation of DNA damage could be fatal. Although she was probably not as deficient in DNA repair as a homozygous DNA ligase I null mutant might be, it is likely that the null mutant would survive for some time before the build-up of damaged DNA became overwhelming. With regard to this, it is interesting to note that mice have been generated which are homozygous for a null mutation in the ERCC-1 gene, which is involved in the incision step of the excision repair pathway (McWhir et al 1993). ERCC-1 homozygous mutants were found to survive pregnancy to term, but then died before weaning, due to liver failure, arising from oxidative stress. Likewise, DNA ligase I mutants would probably also survive pregnancy if they were purely defective in DNA repair. It seems more likely then, that the homozygous DNA ligase I mutants are not viable due to defective DNA
replication. When cells are growing and dividing DNA ligase I levels increase, presumably because the enzyme is required for ligation of DNA replication intermediates. An absence of ligation would interfere with DNA replication, which would, in turn, affect the division of cells during embryo development. It therefore seems reasonable to assume that homozygous DNA ligase I null mutants are inviable because a block on DNA replication creates a block on cell division, which then halts embryo development.

Given the sequence conservation between mouse and human DNA ligase I enzymes it is reasonable to assume that, once it has been correctly targeted, the activity of the mouse enzyme would be affected in the same way as that of the human enzyme by introduction of one of the mutations identified in 46BR. It is however by no means certain that this would have the same overall effect on the mouse carrying the mutation. A number of mouse models for human diseases have been generated, and the results have not always been as expected.

Consider the human disease Lesch-Nyhan syndrome (Lesch and Nyhan 1964). The symptoms of this disease include delayed mental retardation and compulsive self-mutilation, which arise as a result of HPRT deficiency (Seegmiller et al 1967). Two independent mouse models for Lesch-Nyhan syndrome were generated in 1987 using HPRT-deficient embryonic stem cells (Hooper et al 1987, Kuehn et al 1987). However the typical Lesch-Nyhan symptoms were not present spontaneously in either case. It was subsequently shown that mice rely more on the enzyme APRT than on HPRT for their purine salvage pathway, whereas the situation is reversed in humans. Therefore loss of HPRT activity does not have such serious consequences in mice as in humans. However, when APRT activity was blocked in HPRT-deficient mice, they started to exhibit self-injurious behaviour much like that observed in Lesch-Nyhan syndrome patients (Wu and Melton 1993). Two models for cystic fibrosis have been generated independently, and it is interesting that the phenotypes of both differed significantly. Snouwaert et al (1992) obtained mice which were very seriously affected and died soon after birth, while the symptoms of the mice generated by Dorin et al (1992) were milder. However in both cases the mice suffered from disorders associated with cystic fibrosis and therefore apparently represent valid models for the disease. A further human disease for which a mouse model has been generated is Gauchers disease, of which there are two forms, with either mild or very severe symptoms (Tybulewicz et al 1992). The mutant mice exhibited extreme features of the disease, dying within 24 hours of birth, and thus are similar to the severe form of Gauchers disease. A recent interesting example is the generation of mice deficient in the enzyme urate oxidase (Wu et al 1994).
Humans naturally lack this enzyme which converts the end product of purine metabolism, uric acid, to more soluble allantoin. As a result they are predisposed to hyperuricemia, leading to gouty arthritis and renal stones. Other mammals have retained the enzyme, which protects them from developing hyperuricemia, and they cannot therefore be used to study this disorder. Inactivation of urate oxidase in mice was shown to confer obligate hyperuricemia and its associated disorders, rather than just a predisposition to the disease, thus demonstrating that the murine enzyme is indispensable. These examples show that, while mice can be genetically engineered to carry mutations specifying human diseases, they will not necessarily develop the same form of the disease.

In the light of the above examples, the possibility that mice carrying the arginine to tryptophan DNA ligase I mutation may not exactly reflect the 46BR phenotype should be considered. Murine cells may rely on their different DNA ligases to a different extent than human cells, as was found to be the case with HPRT. With regard to this, the patterns of induction of DNA ligase I are identical between human and mouse cells, suggesting that its requirement by the cells is the same (Söderhäll 1976, Chan and Becker 1985).

It is conceivable that DNA ligase II or III could compensate for a lack of DNA ligase I activity, thus diminishing the effects of the mutation, but there is no evidence to support this to date. If the multiple forms of DNA ligase act in a similar manner to the multiple forms of DNA polymerases within mammalian cells then this could happen. There are five different DNA polymerases designated α, β, γ, δ and ε (Wang 1991, and references therein). Each has a specific function: Polymerase α carries out lagging strand synthesis during DNA replication; Polymerase β is thought to carry out short patch DNA repair synthesis; Polymerase γ synthesises mitochondrial DNA; Polymerase δ carries out leading strand synthesis during DNA replication, and polymerase ε may be involved in DNA replication and ultraviolet induced repair synthesis. Thus it would appear that each enzyme is adapted for a specific purpose and that they should not be interchangeable. However, in vitro at least, polymerase β can take over the role of polymerase α and carry out both leading and lagging strand synthesis. The specific cellular functions of mammalian DNA ligases have not yet been as extensively studied. DNA ligase I is thought to be involved in repair and replication of DNA, while all that is known about DNA ligase II is that it does not appear to take part in DNA replication (Söderhäll and Lindahl 1975, Söderhäll 1976, Chan and Becker 1985). DNA ligase III also apparently plays no part in DNA replication (Elder and Rossignol 1990), but may be involved in DNA repair and recombination (Jessberger et al 1993). As with mammalian DNA
polymerases it seems probable that each ligase has its own specific function, although the analogy may end there. It seems unlikely that either DNA ligase II or III could take over the function of DNA ligase I because neither seems to function during DNA replication. Both DNA polymerases α and β, on the other hand are thought to carry out single specific tasks which are closely related. The demonstration that mouse embryos lacking any DNA ligase I activity cannot survive even until birth also tends to suggest that the other ligases cannot substitute for DNA ligase I.

Conclusion

The aim of this project was to investigate the structure of the mouse DNA ligase I gene with a view to introducing a deleterious mutation by gene targeting. The purpose of this was to create a mouse model for DNA ligase I deficiency. So far, the gene targeting work has proceeded satisfactorily, with the knock-out of one allele of the DNA ligase I gene in mouse embryonic stem cells. These have been injected into recipient blastocysts, and the mutated gene transmitted successfully by the resulting chimaera. No animals have been obtained which are homozygous for the DNA ligase I mutation and it seems probable that such a genotype is lethal. The subsequent stage of gene targeting is underway, to restore the mutated DNA ligase I gene, but with a point mutation that affects DNA ligation activity. As with the DNA ligase I null mutation, mouse embryonic stem cells carrying the point mutation will be used to generate chimaeras which will be bred in the hope of obtaining animals homozygous for the mutation. These animals should suffer from the same DNA ligase I deficiency as the '46BR' patient and it will be of great interest to study its effects in the mouse.


fibroblast strain 46BR and cell lines representative of Bloom's syndrome. *Cancer Res.* **48**: 6343-6347


221


223


224


225


Todaro, G.J. and Green, H. (1963) Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* 17: 299-313


agents. In Ataxia-telangiectasia - A Cellular and Molecular Link between Cancer, Neuropathology and Immune Deficiency, B.A. Bridges and D.G. Harnden, eds. (New York: John Wiley and Sons Ltd.) , pp 379-386


Phenotypic correction of a human cell line (46BR) with aberrant DNA ligase I activity

Nikunj V. Somia *, J. Kirsty Jessop * and David W. Melton

Institute of Cell and Molecular Biology, The University of Edinburgh, Darwin Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland, UK

(Received 22 June 1992)
(Revision received 20 November 1992)
(Accepted 11 December 1992)

Keywords: 46BR; Bloom's syndrome; DNA ligase I

Summary

Ligation of DNA after replication and repair is a prerequisite for the preservation of DNA and chromosome structure and function. Biochemical studies with Bloom's syndrome cells have revealed an abnormal DNA ligase I activity. However, genetic analysis has not revealed any differences in transcript levels or in the cDNA sequences of DNA ligase I between Bloom's syndrome and normal cells. Another human cell line, 46BR, derived from an immunodeficient patient, also has an abnormal DNA ligase I. This cell line has recently been demonstrated to harbour two different missense mutations, one at each allele of DNA ligase I. These mutations resulted in a decreased ability of partially purified cell extracts to form an enzyme—adenylate reaction intermediate. We show that 46BR hypersensitivity to an alkylating agent, ethyl methanesulphonate, and to the polyADP-ribose polymerase inhibitor 3-aminobenzamide, is rescued by transfection of wild-type DNA ligase I sequences. This provides additional genetic evidence that the defect in 46BR is at the DNA ligase I locus.

DNA repair is a central process in the maintenance of life, and its deficiency results in several human diseases (Friedberg, 1985). Most studies with cells derived from patients with DNA-repair deficiency and with those derived from in vitro mutagenesis programmes have yielded mutants deficient in the early stages of DNA excision repair. Nucleotide excision repair involves the recognition of damage, with subsequent incision to each side leaving a gapped duplex which is filled by a DNA polymerase activity utilizing the complementary strand as a template. A defect in the incision step results in a disorder termed xeroderma pigmentosum that is characterized by the development of skin cancers and neurological lesions (Friedberg, 1985, and refs. therein). However, some mutants have been identified that have lesions in other steps of DNA repair and these will prove informative in elucidating aspects of repair and chromosome dynamics. For exam-
pie, Cockayne’s syndrome cells have normal levels of general excision repair capacity, but are defective in preferential excision repair of active genes (Venema et al., 1990).

Recently Bloom’s syndrome (BS) cells were demonstrated to have aberrant DNA ligase I activity (Chan et al., 1987, 1988; Willis and Lindahl, 1987; Willis et al., 1987). BS is an autosomal recessive disorder characterized by both pre- and post-natal growth retardation, sun-sensitive telangiectatic facial lesions and a high incidence of cancer (German, 1974). DNA ligase I alterations have also been reported for cell types other than BS cells (Rusquet, 1988). The most extensively studied of these cell lines, 46BR, was derived from an immunodeficient patient also suffering from growth retardation and sunlight sensitivity. A lesion in DNA ligase I for 46BR was initially suggested by the persistence of strand breaks after treatment of cells with the alkylating agent dimethyl sulphate (DMS), although unscheduled DNA synthesis (a measure of DNA synthesis during excision repair) was normal (Teo et al., 1983a). Hence the initial stages of repair are normal, but ligation seems to be affected.

The cellular phenotypes of 46BR and cells derived from BS patients show some overlap, but also some significant differences. For example, while both cell types are sensitive to ethyl-methane-sulphonate (EMS), 46BR cells are hypersensitive to 3-aminobenzamide (3AB), but BS cells are only sensitive to 3AB at high doses (Teo et al., 1983b and refs. therein; Lehmann et al., 1988). In addition, the most remarkable feature of BS cells is the 10-20-fold increased spontaneous sister-chromatid exchange frequency which is not apparent in 46BR cells. Isolation of human DNA ligase I cDNA clones has made it possible to demonstrate that the DNA ligase I gene is present and transcribed normally in BS cells (Petrini et al., 1991; Barnes et al., 1992a), and DNA ligase I cDNAs derived from these cells have a normal nucleotide sequence and are able to complement Saccharomyces cerevisiae DNA ligase cdc 9 mutants. Furthermore, the DNA ligase I gene is present on chromosome 19 (Barnes et al., 1992b), yet the high sister-chromatid exchange phenotype of BS cells is complemented by human chromosome 15 (McDaniel and Schultz, 1992). However, missense mutations have been identified in both ligase I genes in 46BR cells (Barnes et al., 1992a). The mutation at one allele changes the charged amino acid Glu566 to Lys. This residue is found in a highly conserved segment of DNA ligase I known to bind ATP, and subsequently form a DNA ligase I—adenylate reaction intermediate. Furthermore, it has been demonstrated (Kodama et al., 1991) that this mutation results in loss of complementing activity of an Escherichia coli lig and a greatly impaired ability to form a ligase—AMP intermediate. The mutation at the second allele (Arg771 to Trp) is also in a conserved residue. It has been demonstrated (Barnes et al., 1992a) that this is the only allele expressed in a transformed 46BR cell line, and that in vitro assays of partially purified 46BR cell extracts show a decreased ability to bind ATP. These in vitro results provide a strong basis to conclude that the mutations may be responsible for the phenotype of 46BR cells.

These observations may be interpreted in a number of ways. In BS the DNA ligase I gene is not mutated, but either post-translational modification or some other factor mediating DNA ligase I activity is abnormal. Alternatively, BS cells may have a more general defect resulting in a broad spectrum of proteins behaving aberrantly, whereas 46BR possesses mutations at the DNA ligase I loci and hence exhibits a phenotype truly reflective of a ligase mutant, as opposed to BS cells where only a subset of the total phenotype is the result of aberrant ligase action.

Before establishing an animal model for diseases associated with DNA ligase I deficiency, we report genetic complementation of the EMS and 3AB hypersensitivity of 46BR cells by introducing a vector expressing human ligase I cDNA. These results extend the work of Barnes et al. (1992a) by providing additional evidence that the alterations at the 46BR DNA ligase I loci are causal to the phenotype, in contrast to the situation in BS cells.

Materials and methods

Cell lines and culture

The pSV3gpt transformed fibroblast cell line 46BR.1G1 was a gift from Dr. A. Lehmann (Leh-
Fig. 1. Expression vector pβAHLneo. The human ligase I cDNA (open box) was put under the control of the human β-actin promoter (heavy shading). The 5' untranslated region (UTR, closed box) and first intron (IVS 1, light shading) of β-actin are also present upstream of the ligase cDNA. The bacterial neomycin phosphotransferase gene was driven by the mouse metallothionein promoter (neo, open box). The arrows illustrate the direction of transcription. Key restriction sites are also illustrated to highlight the 3.2-kb XhoI fragment used in hybridization experiments. EcoRI sites which generate the diagnostic 1.5-kb ligase cDNA fragment are also marked.

Southern hybridization

10 μg of restricted DNA was size fractionated on a 0.8% agarose gel and blotted onto GeneScreen Plus (DuPont) according to a modification (Reed and Mann, 1985) of the Southern (1975) procedure. Prehybridization and hybridization was in a solution of 6x SSC, 1% SDS, 10% dextran sulphate and 100 μg/ml sonicated herring sperm DNA at 65°C. Filters were hybridized with random primed 32P-labelled ligase I cDNA probes (Feinberg and Vogelstein, 1984). The washing procedure was that recommended for GeneScreen Plus membranes.

Northern blot analysis

Total cellular RNA was prepared as described (MacDonald et al., 1987; Stroman et al., 1977). 20 μg of total RNA was size fractionated in a 1.4% agarose - 0.66 M formaldehyde gel and blotted onto GeneScreen Plus Membrane using 10x SSC. Prehybridization, hybridization and washing procedures were carried out as for Southern blot analysis, but at 60°C.

Assay to measure EMS sensitivity

Cells were plated into 30-mm dishes (5 x 10^4 cells for MRC5V1 and 2.5 x 10^5 cells for 46BR and transfectants), incubated overnight and then exposed to EMS in complete medium (dose range 0-15 mM) for 2 h. Each dose point was done in duplicate. The cells were rinsed twice with warm PBS and incubated in complete medium for a further 5 days. Surviving cell number was determined utilizing a [3H]hypoxanthine incorporation assay (Cleaver, 1989). Essentially cells were

mann et al., 1988), as was the SV40 transformed control cell line MRC5V1 (Huschtscha and Holliday, 1983). Cells were maintained in Glasgow modified Eagle’s medium (GMEM), supplemented with 10% foetal calf serum, 50 U/ml penicillin, 50 μg/ml streptomycin sulphate, 1 mM sodium pyruvate, 2 mM L-glutamine and 1 x non-essential amino acids. Media for propagating transfectants was additionally supplemented with 500 μg/ml G418 (Geneticin, Gibco).

Plasmids and DNA preparations

Plasmid pβAHLneo is illustrated in Fig. 1. It was constructed by cloning the EcoRI/HindIII fragment from pLK444 (Gunning et al., 1987) into Bluescript (Strategene). This fragment contained the 3-kb human β-actin promoter sequence, a 78-bp 5'-untranslated region and a 832-bp intron which included an enhancer sequence. Into the EcoRI site of this plasmid, upstream of the β-actin gene fragment, was cloned a 2.4-kb EcoRI fragment containing the bacterial neomycin phosphotransferase gene regulated by the mouse metallothionein-I gene promoter (Selfridge et al., 1992). This plasmid was designated pβAneo. Human DNA ligase I cDNA was cloned into pβAneo, downstream of the β-actin gene fragment, by isolation of a 3.2-kb KpnI fragment from pHL (Barnes et al., 1990), attachment of XhoI linkers and ligation into the SalI site of pβAneo. This recombinant plasmid was designated pβAHLneo. All plasmids were propagated in E. coli TG1 cells. Plasmid DNA was isolated and purified using Qiagen columns according to the manufacturer's instructions. Plasmids were linearized by incubation with the restriction enzyme Scal, which has a recognition sequence in the ampicillinR gene of the Bluescript vector, prior to transfection. Genomic DNA was isolated according to the method of Pellicer et al. (1978).

Southern hybridization

10 μg of restricted DNA was size fractionated on a 0.8% agarose gel and blotted onto GeneScreen Plus (DuPont) according to a modification (Reed and Mann, 1985) of the Southern (1975) procedure. Prehybridization and hybridization was in a solution of 6 x SSC, 1% SDS, 10% dextran sulphate and 100 μg/ml sonicated herring sperm DNA at 65°C. Filters were hybridized with random primed 32P-labelled ligase I cDNA probes (Feinberg and Vogelstein, 1984). The washing procedure was that recommended for GeneScreen Plus membranes.

Northern blot analysis

Total cellular RNA was prepared as described (MacDonald et al., 1987; Stroman et al., 1977). 20 μg of total RNA was size fractionated in a 1.4% agarose - 0.66 M formaldehyde gel and blotted onto GeneScreen Plus Membrane using 10 x SSC. Prehybridization, hybridization and washing procedures were carried out as for Southern blot analysis, but at 60°C.

Assay to measure EMS sensitivity

Cells were plated into 30-mm dishes (5 x 10^4 cells for MRC5V1 and 2.5 x 10^5 cells for 46BR and transfectants), incubated overnight and then exposed to EMS in complete medium (dose range 0-15 mM) for 2 h. Each dose point was done in duplicate. The cells were rinsed twice with warm PBS and incubated in complete medium for a further 5 days. Surviving cell number was determined utilizing a [3H]hypoxanthine incorporation assay (Cleaver, 1989). Essentially cells were
washed twice in serum-free GMEM and incubated in serum-free GMEM with 0.5 μCi/ml [3H]hypoxanthine for 4–5 h. Cells were washed twice with PBS and lysed with 1% SDS. The lysate was collected and high molecular weight nucleic acids precipitated by addition of 0.8 vol. 20% TCA. The precipitate was collected onto Whatman GF/C filters and the incorporation measured by scintillation counting.

**Assay to measure 3AB sensitivity**

Cells were plated into 60-mm dishes (200 cells for MRC5V1 and 1000 cells for 46BR and transfectants) in complete medium containing different concentrations (0–8 mM) of 3AB. Each dose point was done in duplicate. Dishes were fixed and stained after 10–15 days and surviving colonies were counted. The % survival in 3AB for each cell line was calculated as (mean plating efficiency in 3AB/mean plating efficiency in control medium) × 100.

**Electroporation of cells**

5 × 10^5 cells were incubated with 30 μg of the appropriate linearized plasmid in 0.8 ml GMEM. The plasmid was electroporated into the cells using a Biorad electroporation apparatus (500 μF, 350 V). Cells were left for 10 min and then plated into 60-mm dishes. G418 selection (500 μg/ml) was applied 2 days after electroporation and colonies were either picked when visible, or grown as pools.

**Results**

**Construction and transfection of a vector expressing the human DNA ligase I cDNA**

Fig. 1 illustrates the expression vector pβAHLneo, designed to express the human DNA ligase I cDNA under the control of the human β-actin promoter. An intron is also included since some reports suggest that splicing of a transcript can increase its expression (Palmiter et al., 1991). This intron also contains an enhancer element (Gunning et al., 1987). Plasmid pβAneo is identical except that it lacks the ligase I cDNA sequences. This plasmid was used as a negative control in all experiments. In our hands transfecting these plasmids into 46BR cells was not possible utilizing CaPO₄-coprecipitation, but electroporation was successful. 20–30 G418 resistant colonies were generated from an initial number of 5 × 10^5 cells electroporated. These were either allowed to grow together as pools or primary colonies were picked and clonal transfectants were established. Cells transfected with pβAHLneo were selected for G418 resistance only, i.e. for the expression of the neo gene. There was no selection for DNA ligase I expression, although this should, in most cases, have cotransfected with the neo sequences.

**Detection of ligase I transfectants**

Fig. 2 illustrates a Southern blot analysis of genomic DNA digested with EcoRI and probed with human ligase cDNA (the 3.2-kb XhoI fragment illustrated in Fig. 1). DNA was derived from 46BR or 46BR transfectants. Although detailed analysis has not been completed, most of these hybridising fragments represent parts of the...
endogenous DNA ligase I gene. Integration and retention of vector derived ligase sequences in clones can be seen as novel DNA fragments appearing in cell lines transfected with pβAHLneo. A novel 1.5-kb band is seen in 5 out of 6 of the 46BRpβAHLneo clones examined and in the pool of 46BRpβAHLneo transfectants. The exception, clone 4 must contain neo sequences (since it is G418 resistant) but has lost the ligase sequences on integration. This 1.5-kb band is derived from an internal fragment of ligase cDNA (see Fig. 1). The sizes of some other novel bands visible in the transfectants are determined by the nearest cellular EcoRI recognition sequence to the site of integration. These ligase cDNA sequences are not apparent in transfectants harbouring the control plasmid pβAneo. There is also a difference in the number of copies of cDNA sequences integrated in the cell lines, as determined by the intensity of the novel cDNA-derived 1.5-kb band, compared to the signals due to the endogenous ligase I gene. For example, clones 2 and 6 contain a greater number of inserts than clone 3.

Expression of ligase cDNA sequences

Fig. 3 illustrates a Northern blot analysis of total RNA derived from 46BR and transfectant cells, probed with human ligase cDNA. A transcript of 3.2 kb (Barnes et al., 1990) is apparent in all cell lines with no obvious structural alteration revealed for 46BR compared to MRC5V1 transcripts. A slightly larger transcript, with an additional 78 bp (derived from the β-actin 5’ untranslated region; see Fig. 1) was expected from the pβAHLneo vector. This should have been revealed as a broader more intense ligase I signal in RNA derived from pβAHLneo transfected cells. No such transcripts were detected in cell lines transfected with pβAHLneo, however other transcripts were revealed in the size range 6–9 kb. These are not seen in 46BR RNA, in the pool of 46BRpβAneo transfectants, nor in the transfectant 46BRpβAHLneo clone 4 which has lost the ligase sequences. Thus, these are due to the transcription of the vector-derived ligase sequences, but initiation is probably occurring upstream of the β-actin promoter. From size considerations and the similarity of the novel large transcripts between independent clones, this initiation is probably occurring within plasmid sequences, possibly from the metallothionein promoter which controls neo expression (see Fig. 1). There is a correlation between the extent of ligase transcription from pβAHLneo and the number of integrations of this plasmid in a cell line. This is revealed on comparison of 46BRpβAHLneo clone 2 with other clones. Clone 2 has more insertions of the ligase sequence (see Fig. 2) than the other clones. This is paralleled by an increase in the amount of the vector-derived ligase transcripts apparent in clone 2 compared with other clones (Fig. 3). This increase is not due to unequal loading as visualized by ethidium bromide staining of the gel (data not shown), comparing clones 1, 2, 5 and 6.

Correction of EMS hypersensitivity in 46BR

Fig. 4 illustrates the survival of MRC5V1, 46BR and 46BR transfectants exposed to EMS. 46BR and its derivatives transfected with pβAneo exhibit hypersensitivity to EMS, whereas MRC5V1 reveals a pattern of survival typical of
normal cells (Teo et al., 1983b). 46BR derivatives transfected with pβAHLneo show a sensitivity intermediate between MRC5V1 and 46BR. The three pβAHLneo transfectants assayed (clones 1, 2, 6) were rescued to the same extent and there was no correlation with the copy number of integrated ligase sequences. The extent of correction can be calculated by dose modification at constant survival (Cleaver, 1990). That is by expressing the ratio of the differences in EMS concentration between 46BR and the correctants and between 46BR and MRC5V1 where 50% of cells survive [(b−a)/(c−a), in Fig. 4]. 46BR has been corrected by approximately 60% on transfection of pβAHLneo sequences. MRC5V1 transfected with pβAHLneo showed no difference in survival compared to MRC5V1 (data not shown).

Correction of 3AB hypersensitivity in 46BR

Survival of MRC5V1, 46BR and 46BR transfectants exposed to 3AB is illustrated in Fig. 5. 46BR and a derivative transfected with pβAneo (clone 1) are hypersensitive to 3AB, while the survival of MRC5V1 is typical of normal cells (Lehmann et al., 1988). The 46BR derivative transfected with pβAHLneo (clone 2) shows a pattern of survival similar to that of MRC5V1, indicating complete correction of the hypersensitivity to 3AB phenotype of 46BR by the introduction of functional ligase I sequences.

Discussion

The expression vector utilized in this study has been problematic due to the failure of correct initiation from the β-actin promoter. No novel transcripts were visualized at the expected size (78 bp greater than the 3.2-kb endogenous ligase I transcript) in pβAHLneo transfectants. Attempts to distinguish between the endogenous
and vector ligase transcripts by mapping the initiation sites using primer extension analysis, or PCR strategies were unsuccessful. Primer extension products were halted by secondary structure in the 5' untranslated region of the ligase I mRNA. PCR strategies utilizing amplimers based on sequences in the ligase coding region and the untranslated region of β-actin contributed by the vector were unsuccessful due to cross-hybridization of the necessarily GC-rich β-actin amplimer. The GC-richness of the 78pb β-actin UTR also made it unsuitable as a probe for Northern blot analysis. Nevertheless expression from this vector of high molecular weight transcripts was revealed by probing pβAHneo transfectants with ligase I cDNA. More significantly, there is a strong correlation between the presence of introduced wild-type ligase I sequences, the presence of novel ligase transcripts and the rescue of the EMS and 3AB hypersensitivity phenotypes. Correction of the EMS hypersensitivity was only partial (60%), while correction of 3AB hypersensitivity was complete. This discrepancy could arise because sensitivity to EMS is a much less direct measure of ligase function than survival in 3AB, which inhibits an enzyme (poly ADP-ribose polymerase) required with ligase for efficient ligation.

It is apparent that both 46BR and BS have a biochemical abnormality of DNA ligase I. However, there are clear differences between the phenotypes of 46BR and BS cells. No changes were found in the nucleotide sequence of BS ligase I cDNA which was shown to be functional by rescuing S. cerevisiae cdc9 ligase mutants (Petrini et al., 1991). This suggests that, in vivo, DNA ligase I acts in a complex with (an)other protein(s). It is tempting to speculate that in the case of 46BR, the mutation (Arg771 to Trp) in the DNA ligase I gene has resulted in an inappropriate interaction of this protein in the complex. Note that this mutation still retains some ligase activity, and may have an effect on protein/protein interactions that ligase I takes part in. A consequence of this is that rescue of the mutant phenotype is possible by transfection of wild-type ligase I sequences whereas this would not be possible for BS cells. In BS cells, another protein in this complex could have undergone a change which affects its interactions with DNA ligase I. This could result in ligase I being incorrectly modified and interacting aberrantly in the complex. Furthermore the protein altered in BS cells may modulate or modify other proteins and hence give rise to additional phenotypes such as increased spontaneous sister chromatid exchange.

In conclusion, this paper presents genetic evidence that the mutant phenotype in a mammalian cell line with altered DNA ligase I activity can be rescued on transfection and expression of wild-type DNA ligase I cDNA. These results also suggest that the phenotypes that overlap between BS cells and 46BR are due to a shared ligase aberration, although this may be due to mutations in different proteins. With reference to our aim to produce animal models for ligase deficiency it should be possible to generate these by homologous recombination in embryonic stem cells to create single nucleotide alterations to the mouse DNA ligase I gene. Cells from the resulting animals could then be checked for concordance with the 46BR phenotype.

**Acknowledgement**

We are grateful to Drs. D.E. Barnes and T. Lindahl for informing us at the inception of this project of their preliminary evidence for the presence of a DNA ligase I mutation in the 46BR cell line. We would like to thank Thomas Magin for critical reading of the manuscript and Angela Pow for assistance with the cell culture work. N.S. was supported by a postdoctoral fellowship from the Medical Research Council. J.K.J. was supported by a training award from the MRC. This work was supported by the Cancer Research Campaign.

**References**


