TYPE I RESTRICTION AND MODIFICATION SYSTEMS

by

FRANCES VICTORIA FULLER-PACE

A thesis presented for the Degree of

Doctor of Philosophy

at the University of Edinburgh

Department of Molecular Biology
University of Edinburgh

October, 1985
To David

and

To Pamela and Douglas
FOREWORD

This thesis and the results contained within it are, unless otherwise stated, my own work. The original ideas for these projects were suggested by my supervisor, Noreen Murray, and many of the approaches were devised in discussion with her.

Frances V. Fuller-Pace
Department of Molecular Biology
University of Edinburgh.
ACKNOWLEDGEMENTS

It is an impossible task to acknowledge all those who have contributed to making my stay in Edinburgh so enjoyable.

I am very grateful to my supervisor, Noreen Murray, for the advice and encouragement she has given me throughout the course of this work - and for lending a shoulder when things went wrong.

My thanks go to the following persons in the Molecular Biology Department: my advisor, Richard Hayward, and Ken Murray for their encouragement and guidance on many occasions; Andrew Coulson for advice on computing and sequence analysis; Anne Daniel for invaluable advice on DNA sequencing and for spending many hours helping me with computing; Gill Cowan for collaboration in the hsd A and hsd E work; Annette Campbell for providing in vitro packaging extracts; Annie Wilson for help with figures; Graham Brown for photographs, Kathie Harris for typing the manuscript; and everybody in the media and washing up rooms, particularly Marion Payter and Helen Hibbert for brightening up my sequencing days. Thanks also to Eric Kawashima of Biogen, Geneva, for gifts of synthetic oligonucleotide primers.

Finally, special thanks to David for his advice and help with figures, protein gels, proof reading of the manuscript, for being chief cook and housekeeper for the last few months, and for his encouragement and understanding over the last three years.
The hsd A genes encoding EcoA, a restriction and modification system first identified in Escherichia coli 15T\(^{-}\) (Arber and Wauters-Willems, 1970) and which appears to be unrelated to those of hsd K (Murray et al., 1982), were cloned in \(\lambda\). On the basis of DNA homology, related genes conferring a new specificity (hsd E) were identified in a natural isolate of E. coli and also cloned in \(\lambda\).

Analyses of the hsd A and hsd E systems showed that the overall organisation of the genes encoding EcoA and EcoE closely parallels that for EcoK. Each enzyme is encoded by three genes of which only one, hsdS, confers the specificity of DNA recognition. The genes are in the same order as those encoding EcoK, i.e. hsdR, hsdM, hsdS. The evidence indicates that EcoA and EcoE are Type I restriction endonucleases, but they appear to identify an alternative family to EcoK. For both families the hsdR polypeptide is by far the largest, but the sizes of the other two polypeptides are reversed, with the smallest polypeptide of EcoK being the product of hsdS, and the smallest for the EcoA family the product of hsdM.

A survey of enterobacterial strains for the presence of hsd K- or hsd A-related genes, using DNA hybridisation, detected homology in only a small minority; strains which shared homology with hsd K failed to hybridise with hsd A and vice-versa.
The DNA sequences recognised by the *E. coli* hsd K and the *Salmonella* hsd SP systems differ by only one base-pair (Nagaraja et al., 1985). The DNA sequence of the hsd SP specificity gene was determined and the amino acid sequence of the polypeptide it encodes inferred and compared with the sequence reported for hsd K (Gough and Murray, 1983). The DNA homology between the hsd SP and hsd K specificity genes is considerably greater than that between those of hsd K, hsd B and hsd D, in which two non-homologous regions of about 500 base-pairs flank a central conserved region of about 100 base-pairs, with a second conserved region of about 250 base-pairs at the end encoding the carboxy terminus of the polypeptide (Gough and Murray, 1983). Comparisons of the DNA sequences of the SP and K hsdS genes and of the amino acid sequences of the polypeptides they encode, showed that the central and distal regions which share homology in hsd K, hsd B and hsd D are conserved but, in addition, the proximal region, which is variable in the *E. coli* hsdS genes, is also conserved when hsd K and hsd SP are considered. This conserved "variable" region correlates with the conserved trimeric component of the recognition sequence, AAC, and supports the notion that the variable domains of the specificity polypeptide impart the specificity of recognition. Computer analysis of the DNA sequences of the distal variable regions of the SP and K hsdS genes failed to detect homology. If the variable domains do indeed
impart specificity of recognition, then the non-homologous distal variable regions of the SP and K hsdS genes, despite their very similar tetrameric recognition sequence, imply either an independent origin or extensive changes in related genes of phylogenetically ancient origin.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdoMet</td>
<td>S-adenosyl methionine</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base-pair/s</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>CsCl</td>
<td>caesium chloride</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymidine triphosphate</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleoside triphosphate</td>
</tr>
<tr>
<td>ddATP</td>
<td>dideoxyadenosine triphosphate</td>
</tr>
<tr>
<td>ddCTP</td>
<td>dideoxycytidine triphosphate</td>
</tr>
<tr>
<td>ddGTP</td>
<td>dideoxyguanosine triphosphate</td>
</tr>
<tr>
<td>ddTTP</td>
<td>dideoxythymidine triphosphate</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleoside triphosphate</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>e.o.p.</td>
<td>efficiency of plating</td>
</tr>
<tr>
<td>FTL</td>
<td>freeze-thaw lysate</td>
</tr>
<tr>
<td>HCl</td>
<td>hydrochloric acid</td>
</tr>
<tr>
<td>hsd</td>
<td>host specificity DNA/determinant</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase-pairs</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Mg$^{2+}$</td>
<td>magnesium ions</td>
</tr>
<tr>
<td>MgSO$_4$</td>
<td>magnesium sulphate</td>
</tr>
<tr>
<td>m.o.i.</td>
<td>multiplicity of infection</td>
</tr>
<tr>
<td>NaAc</td>
<td>sodium acetate</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>NH$_4$AC</td>
<td>ammonium acetate</td>
</tr>
<tr>
<td>O.D.$_{650}$</td>
<td>optical density at 650 nm</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>p.f.u.</td>
<td>plaque forming units</td>
</tr>
<tr>
<td>phage</td>
<td>bacteriophage</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>RF</td>
<td>replicative form</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SE</td>
<td>sonicated extract</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate buffer</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>UV</td>
<td>ultra violet light</td>
</tr>
<tr>
<td>UWGCG</td>
<td>University of Wisconsin Genetics Computer Group</td>
</tr>
<tr>
<td>Xgal</td>
<td>5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-Thiogalactoside</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>C. freundii</td>
<td>Citrobacter freundii</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>H. influenzae</td>
<td>Haemophilus influenzae</td>
</tr>
<tr>
<td>K. pneumoniae</td>
<td>Klebsiella pneumoniae</td>
</tr>
<tr>
<td>S.</td>
<td>Salmonella</td>
</tr>
<tr>
<td>Sh.</td>
<td>Shigella</td>
</tr>
<tr>
<td>λ</td>
<td>Lambda</td>
</tr>
<tr>
<td>imm$^\lambda$</td>
<td>immunity of phage $\lambda$</td>
</tr>
<tr>
<td>imm$^{21}$</td>
<td>immunity of phage 21</td>
</tr>
<tr>
<td>Δ</td>
<td>deletion</td>
</tr>
<tr>
<td>μCi</td>
<td>micro Curie</td>
</tr>
<tr>
<td>ts</td>
<td>temperature sensitive</td>
</tr>
</tbody>
</table>
CONVENTIONS

1. Genotypes are indicated by underlining gene symbols. For \textit{hsd} genes, individual genes are underlined, e.g. \textit{hsdR}; where an \textit{hsd} system is referred to, only \textit{hsd} is underlined, e.g. \textit{hsd A}.

2. Phenotypes are indicated as follows:
   
e.g. \textit{r}^{-}\textit{m}^{+} \quad \text{restriction minus, modification plus}
   \textit{r}^{-}\textit{m}_{K}^{+} \quad \text{restriction minus, modification plus, with K specificity.}

3. Restriction enzyme nomenclature is according to Smith and Nathans (1973).

4. Lysogens are designated by the bacterial host followed by the phage in brackets, e.g. \textit{477 (λ)}-NM477 lysogenic for \textit{λ}.

5. Where \textit{λ} has been grown in a particular host this is indicated as follows:
   
   \textit{λ}.477 - \textit{λ} grown in NM477
   \textit{λ}.A - \textit{λ} grown in an A-modifying host
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOREWORD</td>
<td>i</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>vi</td>
</tr>
<tr>
<td>CONVENTIONS</td>
<td>ix</td>
</tr>
<tr>
<td>CONTENTS</td>
<td>x</td>
</tr>
</tbody>
</table>

### CHAPTER 1  GENERAL INTRODUCTION

1.1 Historical background  
1.2 The genetic determinants for Type I restriction and modification systems  
1.3 Restriction and modification systems in Salmonella  
1.4 Type I restriction endonucleases  
   a. Purification and structure  
   b. Recognition sequences  
   c. Sites of methylation and cleavage  
   d. Reaction mechanisms  
1.5 Anti-restriction mechanisms  
1.6 The distribution of Type I restriction and modification systems among bacteria

### CHAPTER 2  MATERIALS AND METHODS

2.1 Materials  
   a. Bacterial strains  
   b. Phage strains
2.1 Materials (contd.)

2.1 c. Plasmids

- M13 vectors

2.1 d. M13 vectors

2.1 e. Chemicals

2.1 f. Bacteriological media

2.1 g. Buffers and stock solutions

2.1 h. Special buffers and solutions for particular techniques

2.2 Methods

a. Phage and bacteriological techniques

   (i) Long term storage of bacterial cells

   (ii) Preparation of plating cells

   (iii) Preparation of phage stocks

   (iv) Phage titrations

   (v) Preparation of phage liquid lysates

   (vi) Selection of λ deletion derivatives using EDTA

   (vii) Phage crosses

   (viii) Phage complementation tests

   (ix) Construction of lysogens/dilysogens

b. DNA techniques

   (i) Phenol extraction

   (ii) Butanol extraction

   (iii) Ethanol precipitation
### 2.2 Methods (contd.)

#### b. (iv) Extraction of DNA from phage .. 86

(v) Extraction of chromosomal DNA from bacteria .. .. .. 86

(vi) Extraction of plasmid DNA .. 87

(vii) Preparation of M13 replicative form (RF) DNA .. .. .. 89

(viii) Plasmid "mini-prep" .. .. .. 90

(ix) DNA restriction .. .. .. 91

(x) Agarose gel electrophoresis .. 92

(xi) DNA ligation .. .. .. 93

(xii) Transfection/transformation .. 94

(xiii) Transfection using "Hanahan" competent cells .. .. .. 96

(xiv) *In vitro* packaging .. .. .. 98

(xv) Elution of DNA from agarose gels 99

(xvi) Southern transfer .. .. .. 101

(xvii) Plaque hybridisation .. .. .. 102

(xviii) Nick translation .. .. .. 103

(xix) Hybridisation .. .. .. 104

#### c. Analysis of polypeptides

(i) Preparation of minicells .. 106

(ii) Labelling of minicells .. .. 108

(iii) Phage infection and labelling of UV-irradiated cells .. 109

(iv) SDS-polyacrylamide gel electrophoresis .. .. 111
2.2 Methods (contd.)

2.2 d. DNA sequencing

(i) Sonication and cloning in M13 .. 113
(ii) DNA preparation from M13 phage lysates .. .. .. .. .. 116
(iii) M13 strand-specific hybridisation 117
(iv) Sequencing reactions .. .. 119
(v) Sequencing gels .. .. .. .. 121
(vi) Sequence data analysis .. .. .. 122

CHAPTER 3 CLONING AND ANALYSIS OF hsd A AND hsd E

3.1 Introduction .. .. .. .. .. .. .. .. 123

3.2 Results

a. Cloning of hsd A genes in λ .. .. .. 131
b. Analysis of λ hsd A; isolation and analysis of deletion derivatives .. 133
c. Identification and cloning of A-related hsd genes .. .. .. .. .. .. 135
d. Identification of polypeptides encoded by the hsd A genes .. .. .. .. .. 138
e. Identification of hsdM and hsdS .. .. 141
f. Direction of transcription of hsd A genes 146
g. Effect of Ral on modification by EcoA 148

3.3 Discussion .. .. .. .. .. .. .. .. 149
CHAPTER 4 THE DISTRIBUTION OF TYPE I
RESTRICTION AND MODIFICATION
SYSTEMS AMONG ENTEROBACTERIA

4.1 Introduction 154
4.2 Results - Hybridisation of bacterial DNAs
with hsd K and hsd A probes 158
4.3 Discussion 162

CHAPTER 5 DNA SEQUENCE OF THE hsdS GENE
OF S.POTSDAM AND COMPARISON
WITH THAT OF THE E.coli K12 hsdS
GENE

5.1 Introduction 170
5.2 Results
   a. Orientation of the hsdS gene in λhsd SP 174
   b. Cloning in M13 176
   c. Sequencing strategy 178
5.3 Discussion 186

REFERENCES 196
APPENDIX I 208
APPENDIX II 214
CHAPTER 1

GENERAL INTRODUCTION
1.1 HISTORICAL BACKGROUND

Host controlled restriction and modification is responsible for the host-specific barriers to inter-strain and inter-species transfer of DNA observed in a variety of bacterial species. The phenomenon of host specificity was first described in the early 1950's by Luria and Human (1952), who observed that bacteriophage T2, which had been propagated in Escherichia coli B, plated with a low efficiency on E. coli K12. Similar results were obtained by Bertani and Weigle (1953) for phage lambda (λ) which had been grown in E. coli C and then transferred to E. coli K12. Bertani and Weigle also showed that this low efficiency of plating (e.o.p.) was not due to lack of adsorption. Phage which survived the transfer to the new host were adapted to grow normally in this host but they lost this adaptation if they were subsequently propagated in their original host. Therefore, the adaptation was not inherited and could not be explained by a mutation in the phage genome; it was rather a function of the bacterial strain in which the phage had been last propagated.

The first indication that the low e.o.p. (restriction) of infecting phage was due to degradation of the phage DNA came from an observation by Lederberg (1957) that when 32P-labelled phage T1 grown in E. coli B (T1.B) were transferred to E. coli B lysogenic for phage P1 [E. coli B (P1)], the phage were restricted and the label subsequently diffused out of the cells. (Phage P1 carries its own
restriction and modification system, different from those of E.coli; Lederberg, 1957; Arber and Dussoix, 1962). Little was known about the molecular basis of this phenomenon until Arber and his colleagues, using phage λ grown in different strains of E.coli, showed that a given host specificity imparts a specific modification to the viral DNA and that DNA lacking the appropriate modification underwent rapid nucleolytic hydrolysis upon entry into the host cell (Arber and Dussoix, 1962; Dussoix and Arber, 1962; Arber et al., 1963). From 32P-labelling experiments, using λ modified to one specificity and grown for one cycle in a host with a new specificity, they showed that, while the DNA of most phages was degraded on entering the new host, a small proportion of the DNA survived and viable phage were produced. This was deduced from the observation that 32P label was present in both progeny phage and in the growth medium as acid-soluble phosphate, having diffused out of the bacterial cells (Dussoix and Arber, 1962). By deuterium-labelling in similar single cycle infection experiments, Arber and Dussoix (1962) showed that the majority of the phage progeny had acquired the new specificity and contained only newly synthesised DNA, but some had retained the specificity of the original infecting phage. Of those with the original specificity, most had semi-conserved the infecting phage DNA while a very few had fully conserved this DNA. The semi-conserved DNA (in which only one strand was modified to the specificity of the new host,
the other being derived from the original infecting phage) was shown to be resistant to restriction by the "new" host (Arber et al., 1963), indicating that DNA modification was a "dominant" property. Dussoix and Arber (1962) also carried out marker rescue coinfection experiments using *E. coli* K12 lysogenic for Pl [**E. coli** K12 (Pl)] and genetically marked λ.K and λ.K (Pl)* phages. [**E. coli** K (Pl)] restricts λ.K but not λ.K (Pl) phages due to the restriction and modification system possessed by Pl.] These experiments showed that markers from λ.K phages could be rescued in λ.K (Pl) phages if both phages coinfected the *E. coli* K12 (Pl) host. As marker rescue was competing with DNA degradation, the probability of rescue of markers from λ.K phages was much higher if the λ.K (Pl) phages infected first. These results gave further evidence that the DNA of the phage was successfully injected into the host cell and then degraded. The above and infection experiments using phenol extracted λDNA** (Dussoix and Arber, 1965) confirmed that host specificity is a property of the phage DNA. Arber and Morse (1965) then showed that zygotic induction of λ and the formation of recombinants for genetic markers via conjugation were "restricted" in crosses between *E. coli* Hfr K12 donors and F− K12 (Pl)

*λ.K phages were grown on *E. coli* K12 and were therefore modified against restriction by the K system. Similarly λ.K (Pl) phages, which were propagated on *E. coli* K12 (Pl) were modified against restriction by both the K system and that of the Pl prophage.

**The infection experiments were carried out using 434 unmodified λDNA in conjunction with modified imm helper phage, as described by Kaiser and Hogness (1960).
recipients. These studies led to the generalisation that not only viral, but plasmid, episomal, and bacterial DNA was subject to restriction and modification.

From these observations, a two-enzyme system was proposed, consisting of a restriction endonuclease recognising a unique sequence and introducing double-stranded scissions and a modification enzyme recognising the same sequence and modifying it, rendering it resistant to hydrolysis by the endonuclease (see Arber, 1965a). In addition, biological experiments (Arber, 1965b) showed that methionine was essential for modification, suggesting that this process involved methylation of the DNA. Thus the bacterial host's own DNA would be resistant to cleavage by its restriction endonuclease by being appropriately methylated, while foreign DNA, lacking the relevant methylation, would be specifically recognised and hydrolysed by the restriction endonuclease. Such restriction and modification enzymes could account for the observed barriers to transfer of DNA between different procaryotic cell types.

Evidence that an enzyme was indeed involved first came from the finding that unmodified \( \lambda \) DNA was specifically inactivated by a crude extract of \( \text{E.coli} \) carrying a \( \text{fi}^- \) R-factor (Takanö et al., 1966). The endonuclease hypothesis was validated in 1968 with the identification and isolation of restriction endonucleases from \( \text{E.coli} \) K, \( \text{E.coli} \) B and \( \text{E.coli} \) K lysogenic for \( \text{P1} \) (Meselson and Yuan, 1968; Linn and Arber, 1968; Roulland-Dussoix and Boyer, 1969). In 1970, Smith and Wilcox described the isolation
and characterisation of a novel endonuclease, from a strain of *Haemophilus influenzae*, which was smaller than the enzymes isolated from *E.coli* and did not have the same specific cofactor requirements. Studies on the complexity of the structure and cofactor requirements of the *E.coli*, *Haemophilus*, and R-factor endonucleases led to the initial sub-division of restriction and modification systems into two classes (see Boyer, 1971). The first class consisted of the *E.coli* systems while the second included those of the *Haemophilus* strains studied and R factors. However, further investigation and the isolation of several new restriction endonucleases from a variety of bacterial species resulted in the reorganisation of the classes to give the present three classes or types (Reiser and Yuan, 1977; Kauc and Piekarowicz, 1978).

The simplest enzymes are those of Type II, e.g. *EcoRI*. In these, the restriction endonuclease contains a single subunit and the modification methylase is a separate enzyme. These enzymes require no cofactor other than magnesium ions (Mg\(^{2+}\)) and they cleave DNA within or close to the sequences that they recognise. Their recognition sequences are generally tetrameric or hexameric and symmetrical. (For reviews, see Modrich, 1979; Modrich and Roberts, 1982). In contrast, Type I and Type III enzymes are multifunctional in that the same enzyme can catalyse both restriction and modification. Type III enzymes (e.g. that encoded by phage P1) require
adenosine triphosphate (ATP) and Mg$^{2+}$ for activity and are stimulated by S-adenosyl methionine (AdoMet). They require ATP as a co-factor but are not ATPases. Two contiguous genes encode the two subunits of the enzyme; a "modification" gene encodes the modification subunit and also imparts the specificity of DNA recognition while a "restriction" gene encodes the additional subunit essential for endonucleolytic activity. Type III enzymes recognise relatively simple asymmetric DNA sequences and cleave DNA some 24-27 bases away from the recognition sequence. (Recent reviews include Yuan, 1981; Bickle, 1982.) The Type I enzymes, such as those encoded by the *E.coli* strains K12 and B are the most complex. These multifunctional enzymes require ATP, AdoMet and Mg$^{2+}$ for restriction and, in addition to their activities as restriction endonucleases and modification methylases, are also DNA-dependent ATPases. They are encoded by three chromosomal genes: *hsdR* (*hsd* for host specificity DNA), *hsdM* and *hsdS*. The product of the *hsdS* gene is responsible for recognition of the DNA sequence imparting specificity; that of *hsdM* together with that of *hsdS*, is required for modification, while the products of all three genes are necessary for restriction. Type I restriction endonucleases recognise asymmetric, hyphenated DNA sequences and cleave at relatively random locations which may be several thousand base-pairs (bp) from the recognition sequence. (For recent reviews see Modrich, 1979; Yuan, 1981; Bickle, 1982).
Type I restriction and modification systems and the enzymes they encode are considered in more detail below and in the introductions to relevant chapters. Particular reference is given to the genetic experiments leading to the identification of the genetic determinants of hsd K and hsd B, the enzymes encoded by these and other systems and their mechanism of activity, using EcoK as a model and, finally, the distribution of the various known Type I hsd systems among bacterial species and strains.

1.2 THE GENETIC DETERMINANTS FOR TYPE I RESTRICTION AND MODIFICATION SYSTEMS

The most extensive genetic analyses of Type I restriction and modification systems have been carried out on those of E.coli K12 and E.coli B. Most of these analyses relied on complementation tests using a variety of restriction-deficient and modification-deficient mutants of the E.coli K12 and B strains. The hsd K and hsd B systems are allelic and map at the same region of the chromosome, at 98.5 mins on the E.coli map (Boyer, 1964; Glover and Colson, 1969; Bachmann and Low, 1980). The close relationship of these systems and the finding that the subunits encoded by the respective hsd K and hsd B genes were interchangeable, allowed inter-strain as well as intra-strain complementation tests (Boyer and Roulland-Dussoix, 1969; Glover and Colson, 1969).
Restriction-deficient mutants, first isolated by Wood (1966) were found to be of two different phenotypic classes: $r^-m^+$, with normal modification, and $r^-m^-$, which were also deficient in modification. These two classes of mutants complemented each other to give functional restriction and modification (Boyer and Roulland-Dussoix, 1969). Another group of $r^-m^-$ mutants, derived from the original $r^-m^+$ ones by a second-step mutation, were subsequently shown to complement first-step $r^-m^-$ mutants giving a $r^+m^+$ phenotype, suggesting that the mutations in the first- and second-step $r^-m^-$ mutants had occurred in different cistrons. In interstrain complementation tests, using $r^-B^-m^-B^-$ first- and second-step mutants and $r^+K^+m^+K^+$ strains, diploids from first-step mutants were shown to lack B specificity, while second-step mutants complemented the K-specific strain, giving $r^+m^+$ diploids with both K and B specificity (Boyer and Roulland-Dussoix, 1969; Glover, 1970). This indicated that second-step $r^-m^-$ mutants retained, while first-step mutants lacked, a functional specificity gene. These findings thus showed that mutations in two cistrons lead to the same $r^-m^-$ phenotype, implying that two genes are required for modification. They also defined three genes responsible for host-specific restriction and modification, later named $\text{hsdR}$, $\text{hsdM}$ and $\text{hsdS}$ (as described in the previous section), with $\text{hsdS}$ conferring host specificity. Other complementation tests, using temperature sensitive derivatives of $r^-m^+$ mutants
(Hubacek and Glover, 1970), which were subsequently found to have functional \textit{hsdS} genes and mutations in the \textit{hsdM} gene, characterised the function of the \textit{hsdM} gene; these mutants were deficient in both restriction and modification, indicating that the product of the \textit{hsdM} gene is required for both of these activities.

The \textit{hsd K} and \textit{hsd B} complementation tests are described more specifically in the introduction to Chapter 3.

Further analysis of \textit{hsd} genes relied on the cloning of the \textit{hsd K} genes in \textit{\lambda} (Sain and Murray, 1980). Complementation tests using deletion derivatives of the \textit{hsd K} region in \textit{\lambda} showed that the order of the genes in \textit{hsd K} is \textit{hsdR}, \textit{hsdM} and \textit{hsdS}, and that they are arranged in two transcriptional units, with two promoters, one for \textit{hsdR} and one for \textit{hsdM} and \textit{hsdS}. This could allow for expression of \textit{hsdM} and \textit{hsdS}, i.e. modification, without expression of \textit{hsdR}, i.e. restriction, although the biological significance of this has not been demonstrated. The approximate sizes of the genes, calculated from the molecular weights of the polypeptides they encode (described in the next section), taking an amino acid to have a molecular weight of 110 (Sain and Murray, 1980), would be as shown in Figure 1.1.

Sain and Murray (1980) also provided molecular evidence for the relatedness between the \textit{hsd} systems of \textit{E.coli} K12 and \textit{E.coli} B. They used a probe comprising most of the \textit{hsdR} gene and all of the \textit{hsdM} and \textit{hsdS} genes.
of E. coli K12 and showed, by DNA hybridisation studies, that it shared extensive homology with the DNA of E. coli B, but none with that of E. coli C, which has no known restriction and modification system. This not only corroborated the findings from the earlier genetic experiments, but also indicated that, in E. coli C, the structural genes conferring host specificity are absent. Therefore, the lack of restriction and modification activity observed in this strain is not due to a loss of function but due to the absence of the relevant DNA.

1.3 RESTRICTION AND MODIFICATION SYSTEMS IN SALMONELLA

Many Salmonella serotypes are now known to possess a variety of Type I restriction and modification systems and, in several cases, a particular strain encodes more than one specificity (Bullas, et al., 1980). With
Salmonella most of the interest has focused on the identification of the various systems found in different serotypes and comparisons of these systems with other known ones, particularly those of E.coli K12 and E.coli B.

The first indication of the presence of host-controlled restriction and modification in Salmonella was given by the detection of the existence of a barrier to the acceptance of DNA carrying E.coli specificity by S.typhimurium (Zinder, 1960; Miyake, 1962). It was later shown that the fertility mutation in S.typhimurium resulted in the loss of previously undetected host restriction towards E.coli DNA (Colson and Colson, 1967; Okada et al., 1968). The key observation in these studies was that the restriction affecting E.coli DNA also affected bacteriophage P22 DNA. Colson et al. (1969) isolated and characterised forty eight restriction-deficient mutants of S.typhimurium LT2, using phage P22, and showed that these fell into two groups, $r^- m^+$, with normal modification, and $r^- m^-$, which were deficient in modification. They suggested that the genetic control of host specificity in Salmonella was therefore similar to that of E.coli i.e. that the S.typhimurium LT2 system was of Type I. Subsequently, using F-mediated conjugation, Colson et al. (1970) mapped the chromosomal location of this system, which they called LT, at 8.5 minutes on the Salmonella map (Sanderson and Hartman, 1978), a location which is different from that of the E.coli hsd genes.
In 1971, Colson and Colson, using bacteriophage L, identified a new *S. typhimurium* host specificity system, SA, in addition to the previously discovered LT system. The genes for SA were mapped at 98 minutes, i.e. in approximately the same region of the chromosome as the *hsd K* genes, considering the similarity of the *E.coli* and *S. typhimurium* linkage maps (Sanderson and Demerc, 1965). In addition, on transferring the genes to *E.coli*, they were found to be fully functional. Colson and Van Pel (1974), using haploid hybrids between *S. typhimurium* Hfr and *E.coli* F\(^{-}\), described yet another host restriction and modification system, SB, active on phage \(\lambda\). The genes of the SB system mapped between those of SA and *serB*, presumably at the same location as those of *hsd K*. The SB system had not previously been detected as, unlike SA, it is not active on the *Salmonella* phages P22 and L, while active on phages \(\lambda\) and P3. Pl- and P22-mediated transduction in *S. typhimurium* hybrids established that the genes of the SA and SB systems are non-allelic, even though they are linked (Colson and Van Pel, 1974). SB was subsequently found to complement both *hsd K* and *hsd B* suggesting that these systems are related, but there was no evidence for the relatedness of SA as this complemented neither *hsd K* nor *hsd B* (Van Pel and Colson, 1974).

Bullas and Colson (1975a) screened forty two *Salmonella* strains with bacteriophage P3, which is resistant to restriction by the SA and LT systems but
sensitive to restriction by SB. They identified five new DNA restriction and modification systems in five different serotypes. Genetic analysis of one of these (SP) from *S. potsdam*, involving P1-transduction of SP genes into *S. typhimurium* and *E. coli/S. typhimurium* hybrids showed that the SP genes are allelic and functionally homologous to those of SB, and therefore to those of *hsd K* and *hsd B*. (The other four specificities have not been investigated.) A similar experiment employing P1-mediated cotransduction with *serB* from *S. potsdam* to an *E. coli/S. typhimurium* hybrid led to the isolation of a new specificity SQ (Bullas and Colson, 1975b; Bullas et al., 1976). Many of the *serB*\(^{+}\) recombinants obtained had lost specificity but one had acquired a new specificity (SQ) with genes allelic to those of SB and functionally related to the *E. coli hsd K* genes. It was suggested that SQ had arisen from recombination between the specificity (*hsdS*) genes of SB and SP. This explanation was subsequently confirmed by heteroduplex analysis of the *hsdS* genes of all three systems (Fuller-Pace et al., 1984). SQ is the first example of a new specificity resulting from recombination between existing specificity genes and it has considerable implications concerning the evolution of such systems in bacteria.
1.4 TYPE I RESTRICTION ENDONUCLEASES

1.4.a. Purification and Structure

Type I restriction and modification systems were the first systems identified as having endonuclease capable of specific recognition of DNA sequences and much of the work in the 1960's and early 1970's was based on the belief that they would be valuable as reagents for analysis of DNA molecules and as systems for the study of DNA sequence recognition by proteins. Although the former belief proved unfounded, as the extraordinary complexity of the reactions mediated by Type I endonucleases became apparent, the use of such enzymes as models for the study of DNA-protein interactions has provoked considerable research.

The restriction endonucleases encoded by all documented Type I systems have been purified. The first, and most extensively studied, are those encoded by E.coli K12 (Meselson and Yuan, 1968) and E.coli B (Roulland-Dussoix and Boyer, 1969; Eskin and Linn, 1972 a, b), EcoK and EcoB respectively. (The nomenclature is as in Smith and Nathans, 1973). As expected from the allelic nature of the \textit{hsd K} and \textit{hsd B} genes, EcoK and EcoB have similar structures. Both have three subunits with molecular weights of about 135,000, 62,000 and 50,000 (\(\alpha\), \(\beta\), and \(\gamma\) respectively), but the proportions of each of the subunits in the holo-enzyme appear to be different. EcoK was reported to have the structure \(\alpha_2\beta_2\gamma\) (see Meselson et al., 1972) from molecular weight
estimation by gel filtration and glycerol-gradient centrifugation of the holo-enzyme and sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis of the subunits. With EcoB, at least three oligomeric structures have been identified from sucrose gradient purification and molecular weight estimation by polyacrylamide gel electrophoresis, the major form being $a_2b_4\gamma_2$ (Eskin and Linn, 1972a). These differences in structure may, however, be due to differences in the growth conditions of the bacteria or in the purification procedures.

Although EcoK and EcoB restriction endonucleases also show modification activity (see Yuan, 1981; Bickle, 1982), modification enzymes lacking restriction activity have been purified from wild-type strains of E. coli B (Kühnlein et al., 1969; Lautenberger and Linn, 1972) and more recently from a strain of E. coli K12 lysogenic for a λ derivative carrying the hsd K region (see Bickle, 1982; Suri et al., 1984a). Both these contained the two smaller subunits (β and γ) indicating that the large (α) subunit is the product of the hsdR gene. The EcoK methylase was found to contain the β and γ subunits in equimolar amounts (Suri et al., 1984a); in the case of the EcoB methylase, again several forms were obtained but freshly isolated enzyme had the structure $2\beta_1\gamma_1$ (Lautenberger and Linn, 1972). Analysis of deletions in the hsd K region cloned in λ (Sain and Murray, 1980) correlated the β subunit with the hsdM gene and the γ subunit with hsdS.
The restriction endonucleases encoded by the K-related systems hsd D (from E.coli El66), hsd SB (from S.typhimurium), hsd SP (from S.potsdam) and hsd SQ (from recombination between hsd SB and hsd SP), have since been purified. All these resemble EcoK and EcoB in their structure, function and cofactor requirements. The subunits of the SB, SP and SQ enzymes are present in similar ratios as those of EcoK as judged from SDS polyacrylamide gel electrophoresis (Nagaraja et al., 1985 b, c); the ratios of the subunits of EcoD have not been documented (Nagaraja et al., 1985a).

The EcoA restriction endonuclease has also been purified (Suri et al., 1984b). Two protein species were obtained during purification; one contained two different subunits of molecular weights 64,000 and 55,000 and was a functional methylase; the other consisted of a single polypeptide of molecular weight 98,000 which had no enzymatic activity. However, on mixing the two species, restriction activity was obtained which was similar to that of other Type I enzymes, indicating that the largest subunit is encoded by hsdR. In the case of the two smaller subunits, however, it was not possible to assign functions to particular subunits or to correlate subunits with respective genes.

1.4.b. Recognition sequences

The DNA sequences recognised by all the known EcoK-related restriction endonucleases and by EcoA have been determined. They all have the same asymmetrical and hyphenated structure, with two constant domains
**Figure 1.2** DNA sequences recognised by Type I restriction endonucleases

- **Eco K**
  
  \[
  5'-A\, A^*\, C\, N\, N\, N\, N\, N\, N\, G\, T\, G\, C-3'
  \]
  
  \[
  3'-T\, T\, G\, N\, N\, N\, N\, N\, N\, C\, A^*\, C\, G-5'
  \]
  
  Kan et al., 1979

- **EcoB**
  
  \[
  5'-T\, G\, A^*\, N\, N\, N\, N\, N\, N\, N\, N\, T\, G\, C\, T-3'
  \]
  
  Lautenberger et al., 1978; Ravetch et al., 1978; Sommer and Schaller, 1979; Lautenberger et al., 1979

- **EcoD**
  
  \[
  5'-T\, T\, A^*\, N\, N\, N\, N\, N\, N\, N\, N\, G\, T\, C\, Y-3'
  \]
  
  + Nagaraja et al., 1985a

- **SB**
  
  \[
  5'-G\, A^*\, G\, N\, N\, N\, N\, N\, N\, R\, T\, A\, Y\, G-3'
  \]
  
  Nagaraja et al., 1985b

- **SP**
  
  \[
  5'-A\, A^*\, C\, N\, N\, N\, N\, N\, N\, G\, T\, R\, C-3'
  \]
  
  Nagaraja et al., 1985b

- **SQ**
  
  \[
  5'-A\, A^*\, C\, N\, N\, N\, N\, N\, N\, R\, T\, A\, Y\, G-3'
  \]
  
  Nagaraja et al., 1985c

- **EcoA**
  
  \[
  5'-G\, A^*\, G\, N\, N\, N\, N\, N\, N\, G\, T\, C\, A-3'
  \]
  
  Suri et al., 1984b

  \[
  3'-C\, T\, C\, N\, N\, N\, N\, N\, N\, N\, C\, A^*\, G\, T-5'
  \]

* indicates methylated adenine residues, see section c.

+ Y refers to either pyrimidine, R to either purine

Boxed in bases indicated conserved positions.
consisting of a trimer and a tetramer (a pentamer for SB and SQ) of specific sequence, separated by a spacer of non-specific sequence, of a fixed, but different length (6-8 bases) for the various enzymes. The recognition sequences are as shown in Figure 1.2.

An interesting feature emerging from these sequences is the degeneracy found in those recognised by EcoD and by the Salmonella enzymes (see Fig. 1.2), particularly in the case of the sequences recognised by SB and SQ where a pentanucleotide with two degenerate positions replaces the tetramer found in all the other recognition sequences. From the sequences recognised by the Salmonella restriction endonucleases SB, SP and SQ, it is seen that one part of the sequence for SQ is derived from SP, the other from SB (Nagaraja et al., 1985c); this was a crucial prediction from the finding that the hsdS gene of SQ resulted from recombination between those of SB and SP (Fuller-Pace et al., 1984; see Chapter 5). The similarity of the EcoK and SP recognition sequences is very striking, which corroborates the previous finding that SP modification protects DNA from restriction by EcoK (Bullas et al., 1980; Nagaraja et al., 1985b). This is of particular interest in terms of the evolution of different specificities within families of restriction and modification systems.
1.4.c. Sites of methylation and cleavage

One of the most interesting aspects of Type I restriction endonucleases is that while they modify (methylate) DNA within the sequences they recognise (Smith et al., 1972; Kühnlein and Arber, 1972), they cleave DNA at another site which may be a considerable distance away from the recognition site (Horiuchi and Zinder, 1972; Murray et al., 1973b; Brammar et al., 1974).

(i) Methylation

Arber (1965b) was the first to show that methionine was necessary for modification. Subsequent studies using enzymes purified from various modification-proficient strains of E.coli B indicated that AdoMet was an essential cofactor and also acted as a methyl donor in modification in vitro (Kühnlein et al., 1969; Kühnlein and Arber, 1972). Smith et al. (1972) and Kühnlein and Arber (1972) investigated nucleotide methylation in B-specific modification and showed that two methyl groups were incorporated within each EcoB recognition site, with methylation occurring at adenine residues. Vovis and Zinder (1975) subsequently showed that one adenine residue on either strand is methylated and, furthermore, that DNA, in which only one strand was methylated, was resistant to restriction but underwent further
methylation during modification to give DNA with both strands methylated.

Kan et al., (1979) considering the EcoK and EcoB recognition sequences (the only sequences available at the time) suggested that despite the differences in length of the non-specific spacers, the EcoK and EcoB sequences could be lined up so that four out of the seven specific bases were conserved in both sequences; these would include the methylated residues. However, this homology is no longer apparent when all the sequences in Figure 1.2 are examined. One A.T base-pair is conserved in each specific domain for all the sequences carried by the EcoK-related enzymes (see Fig.1.2). This is not the case, however, with the sequence recognised by EcoA (encoded by hsd A), which, as discussed later, does not appear to be related to hsd K (Suri et al., 1984b). These conserved base-pairs are the most likely targets for methylation, one methyl group being added to an adenine residue on each strand, as shown for EcoB by Vovis and Zinder (1975). In the lower strand, analysis of methylated oligonucleotides suggested that the first adenine residue in the tetranucleotide domain of the EcoB recognition sequence was methylated (Van Ormondt et al., 1973). This suggestion gained credence from the finding that this residue is conserved in all the EcoK-
related recognition sequences; in fact in that of EcoK, there are no other adenine residues in the lower strand (see Fig.1.2). In the case of the upper strand, the second adenine residue in the specific trinucleotide was shown to be the methylation site. When this site overlaps with a HindII site, EcoK modification protects against restriction by HindII, indicating that the adenine residue common to both recognition sites is methylated (Roy and Smith, 1973; see Bickle, 1982). This residue is again conserved in all the sequences recognised by enzymes belonging to the K family and in that recognised by EcoB, it is the only adenine residue in the upper strand (see Fig.1.2).

In all the Type I recognition sequences determined to date (Kan et al., 1979; Suri et al., 1984b; Nagaraja et al., 1985a,b,c) the methyl groups are placed in the specific domains such that one adenine residue in the upper strand of the trimer and one in the lower strand of the tetramer (or pentamer) are methylated; these are depicted as A* in Figure 1.2. The methylated residues are ten bases apart for K-related enzymes (eleven for EcoA); this together with the constraints on the length of the non-specific spacer suggests that the enzyme interacts with one face of the DNA helix as the specific bases in the recognition sequences, including the methylated residues, would
be in two successive major grooves with most of the non-specific sequence in the minor groove separating them (Ravetch et al., 1978; Nagaraja et al., 1985b).

(ii). Cleavage

Early studies on restriction by Type I endonucleases were based on the belief that these enzymes cleaved within the sequences they recognised and many of the initial investigations involved the recognition sites in DNA molecules that were normally restricted by such enzymes.

Rühnlein and Arber (1972) identified two recognition sites for EcoB in bacteriophage fd DNA and showed that the enzyme interacts with these sites for both restriction and modification. Franklin and Dove (1969) showed that there are at least three targets for EcoK in λ, giving a restriction coefficient of 1000. (This is in contrast to the situation with EcoA, which has only one target in λ and shows a restriction coefficient of 100; Arber et al., 1972). By genetic manipulation of phi 80/λ hybrids, Murray et al. (1973a) mapped two of the EcoK recognition sequences in λ but, contrary to earlier expectations, restriction of λ DNA molecules by EcoK did not produce discrete fragments (Murray et al., 1973b). The first indication that cleavage occurs at a site distinct from the recognition sequence had previously come from in vitro experiments with.
purified EcoB and fl replicative form DNA molecules (Horiuchi and Zinder, 1972). This was subsequently confirmed by in vivo experiments using the same system (Hartmann and Zinder, 1974). Other in vivo experiments with λ trp phages in exo V-deficient cells (Brammar et al., 1974) confirmed that EcoK cleaves outside the recognition sequence and also showed that cleavage does not preferentially occur near the recognition sequence. In addition, Murray et al. (1973b) and Brammar et al. (1974) showed that, when two EcoK sites were present in a λ DNA molecule, cleavage preferentially occurred between them and that certain regions of the molecule (in this case the trp operon) were relatively insensitive to restriction by EcoK. This finding suggests that, although EcoK cleavage is not sequence-specific, it is not totally random and may be affected by other features of the DNA molecule.

1.4.d. Reaction mechanisms

Type I restriction endonucleases catalyse an extraordinary variety of different reactions, acting as restriction endonucleases, DNA site-specific methylases and DNA-dependent ATPases. They even show a DNA gyrase-like action which in recent years has generated new interest in the mechanism of action of these enzymes and their interaction with DNA. All K-related restriction
endonucleases appear to be very similar in their action (see Bickle, 1982; Nagaraja et al., 1985a,b,c) although some differences have been reported between the mechanisms of EcoK and EcoB (Rosamond et al., 1979; Yuan et al., 1980).

The first evidence concerning the action of Type I restriction endonucleases as multifunctional enzymes catalysing both methylation and endonucleolytic cleavage came from a report indicating that highly purified EcoK, apart from cleaving DNA efficiently, transferred methyl groups from AdoMet to adenine residues of unmodified DNA (Haberman et al., 1972). Subsequently, Vovis et al., (1974) showed that both modification and restriction by EcoB required, or were stimulated by the same cofactors (i.e. Mg$^{2+}$, ATP, and AdoMet) and that the type of reaction exhibited by EcoB (i.e. modification or restriction) depended upon the nature of the recognition sequence on the DNA. Unmodified recognition sequences resulted in rapid cleavage, or slow methylation, while sequences in which one DNA strand was modified were a substrate for rapid methylation. Heteroduplex recognition sequences (with only one DNA strand modified) had previously been shown to be a substrate for methylation but not restriction, which requires unmodified DNA (Yuan et al., 1972; Vovis et al., 1973).

Yuan and Meselson (1970) gave an indication of the process involved in restriction and modification reactions when they showed that, in the presence of Mg$^{2+}$, ATP and
Fig. 3 The reaction mechanism of the restriction endonuclease EcoK.
AdoMet, \textit{EcoK} forms a specific complex with the DNA substrate, which can be detected by its retention on nitrocellulose membranes. The formation of this complex is dependent on the activation of \textit{EcoK} by AdoMet, which acts both as an allostERIC effector (involved in \textit{EcoK} activation) and as a methyl group donor in methylation (Hadi \textit{et al.}, 1975). Three enzyme-DNA complexes can be differentiated experimentally in the course of a cleavage reaction by \textit{EcoK} (Yuan \textit{et al.}, 1975). These comprise (i) an initial complex at a non-specific site on the DNA, (ii) a recognition complex formed at a specific \textit{EcoK} recognition site (sK) on the DNA molecule and (iii) a cleavage complex, dependent on ATP and requiring an unmodified recognition sequence. The reaction mechanism of restriction and modification, including the various complexes formed between \textit{EcoK} and the DNA substrate, have been described by Burckhardt \textit{et al.} (1981 a, b). These are outlined in Figure 1.3. \textit{EcoK} is used as a model, but the general features of the mechanism described below are believed to be valid for the other K-related enzymes (see Bickle, 1982; Nagaraja \textit{et al.}, 1985 a, b,c). The reactions are sequential and involve several steps. The early steps (1-4) are common for both restriction and modification; the nature of the recognition sequence (sK) then determines which reaction follows (steps 5-9).

\underline{Steps 1-2}

\textit{EcoK} binds to AdoMet non-covalently and is activated, by a conformational change, to \textit{EcoK*}.  

\underline{Figure 1.3: from Burckhardt \textit{et al.} (1981a)}
The activated form of the enzyme is then able to recognise and bind to the specific recognition (sK) sites on the DNA molecule. AdoMet acts as an allosteric effector at this stage and it has been shown that a mutation in the \textit{hsdM} gene results in the loss of AdoMet binding and consequently in the inability to modify or restrict DNA (Buhler and Yuan, 1978). Once activated to EcoK*, the enzyme does not require any further AdoMet for restriction although it still requires AdoMet for modification. A mutation in the \textit{hsdS} gene was found to prevent enzyme-DNA binding; this is thought to be due to the lack of activation of the enzyme to EcoK* (Hadi \textit{et al.}, 1975). (This explains the r" m" phenotype obtained with mutations in either the \textit{hsdM} or \textit{hsdS} genes, as seen in Section 1.2).

**Step 3**

EcoK* then binds to DNA non-specifically, regardless of the presence or absence of sK recognition sequences, forming an initial-complex. Electron microscopic studies (Brack \textit{et al.}, 1976) have shown that the enzyme will bind to mutant DNA lacking recognition sites.

**Step 4**

The bound EcoK* will then seek an sK site, in an as yet unexplained fashion (Yuan \textit{et al.}, 1980), and bind tightly to this site to form a recognition
complex. This binding occurs regardless of whether the sK sequence is modified or not (Bickle et al., 1978). ATP then induces a conformational change in EcoK* which allows it to recognise whether the sK sequence is unmodified, heteroduplex (hemi-methylated) or modified. This conformational change does not require ATP hydrolysis as the same effect was obtained with the $\beta,\gamma$-imido analog of ATP (Bickle et al., 1978). The modification state of the sK site determines whether restriction or modification will follow (Hadi et al., 1975; Yuan et al., 1980).

**Step 5**

If the sK site is modified, ATP causes release of the enzyme from the DNA without ATP hydrolysis (Bickle et al., 1978).

**Step 6**

In the case of an unmodified sK sequence, ATP induces a conformational change of EcoK* to a new form, EcoK+, as seen from electron microscopic studies, and AdoMet is released (Bickle et al., 1978). This new form is then able to cleave DNA; a mutation in the hsdR gene allows the formation of a recognition complex and normal methylation but appears to block the ATP-induced conformational change from EcoK* to EcoK+ (Bühler and Yuan, 1978). This explains the $r^{-}m^{+}$ phenotype resulting from such
mutations. ATP again acts as an allosteric effector at this point and no ATP hydrolysis is involved. The \textit{EcoK}^+ -DNA complex is retained on nitrocellulose filters, hence the term \textit{filter binding complex}.

**Steps 7-8**

Various mechanisms have been suggested for the way in which \textit{EcoK}^+, bound to the DNA molecule at the recognition site, can cleave DNA at a site which may be a few kilobase-pairs away. The following models have been proposed.

(i) \textit{EcoK} dissociates from the sK site and interacts at random sites on the same or a different DNA molecule. This is improbable as Meselson and Yuan (1968) showed that the enzyme will only cleave unmodified DNA and will not cleave modified DNA in trans when incubated with a mixture of modified and unmodified DNA. \textit{EcoB} behaves in the same way (Vovis \textit{et al.}, 1973).

(ii) ATP triggers movement of the enzyme along the DNA molecule until it reaches a possible cleavage site. However, evidence to date suggests that the enzyme remains bound to the recognition site, even after endonucleolytic action (Bickle \textit{et al.}, 1978), which also contradicts the previous model.
Figure 4 Model of the DNA Translocation Mechanism by Eco K
Eco K+ binds at an unmodified sK site and forms a filter binding complex in the presence of ATP due to a conformational alteration of the enzyme, shown as Eco K+ on the second line of the Figure. This altered complex can now bind at a second site on the DNA substrate, presumably in any of the four configurations depicted on the first and third lines. (A and B are used to indicate relative ends of the DNA molecule; the arrows depict the direction in which translocation will occur.) Concurrent with ATP hydrolysis, the Eco K winds the DNA past itself, and, depending upon the orientation of the second binding site, it will either form a twisted loop which will increase in size (left side of fourth line) or decrease in size (right side of fourth line).
(iii) EcoK remains stably bound to the sK site and cleaves following a random second contact with the same DNA molecule.

(iv) EcoK stays bound to the sK site and winds the DNA molecule past it.

Yuan et al. (1980) proposed a mechanism involving elements from the latter two models, but in which the second contact with the DNA molecule, resulting in cleavage, is not completely random. They suggest that once EcoK* is transformed to EcoK', a second DNA-binding site becomes accessible on the enzyme, which is then able to interact with DNA randomly on either side of the sK sequence. EcoK' then begins to wind the DNA past itself in a process coupled to ATP hydrolysis until a possible cleavage site is reached, where a single-stranded break occurs, followed by cleavage in the other strand close to the first cut (Meselson and Yuan, 1968). This mechanism of translocation forms twisted loops in the DNA molecule as seen in Figure 1.4. Such loops have been observed in electron micrographs and this translocation activity has been likened to that of DNA gyrase (Yuan et al., 1980). A conflicting account was given for the mechanism of EcoB cleavage by Rosamond et al. (1979) who produced electron microscopic evidence that the enzyme "tracks" along the DNA in one direction only.
away from the trimeric part of the recognition sequence, thus cleaving only on the 5' side of the sequence. Although it had been suggested that the differences were due to the different reaction conditions used for the two enzymes (see Yuan, 1981), more recent investigations into DNA translocation by EcoB, using other electron microscope techniques still detected DNA translocation only in one direction (Endlich and Linn, 1985a), implying that EcoK and EcoB have slightly different mechanisms of action.

Exactly what constitutes a "cleavage" site is not known, but it is possible that the enzyme translocates the DNA at a constant rate with a given probability of cleavage at each phosphodiester bond. Any structural features of the DNA causing a slowing down of translocation would increase the probability of cleavage. Thus certain regions of a particular DNA molecule would be preferentially cleaved.

Cleavage is a two step reaction; the enzyme cuts one DNA strand and creates a gap, possibly by exonucleolytic digestion. This gap would stop translocation and allow a cut in the opposite strand. Evidence that DNA cleavage involves two steps and requires two enzyme molecules came from the finding that under conditions of excess DNA only single-stranded breaks were obtained (Meselson and Yuan, 1968;
Roulland-Dussoix and Boyer, 1969; Adler and Nathans, 1973). The suggestion that cleavage is accompanied by exonucleolytic degradation, creating a gap, is based on the observation that EcoB digestion releases about 75 bases in the form of acid soluble oligonucleotides per cleavage event (Kimball and Linn, 1976); there is no evidence of this in the case of EcoK. A curious feature of the DNA ends produced by both EcoK and EcoB is that they are resistant to the action of polynucleotide kinase, even after phosphatase treatment (Eskin and Linn, 1972a; Murray et al., 1973b). This may be due to the presence of short 3' single-stranded extensions produced by the exonuclease action.

Following, or possibly concomitant with, DNA cleavage, Type I restriction endonucleases are irreversibly transformed into potent ATPases (Eskin and Linn, 1972b; Yuan et al., 1972). This transformation is accompanied by a loss of endonuclease activity; thus each enzyme molecule cuts only one DNA strand. The endonuclease remains bound to the recognition site and does not appear to turn over as a nuclease, but it is capable of rapid turnover as an ATPase (Eskin and Linn, 1972a, b); this may explain the observation that the ATPase activity continues long after DNA cleavage is complete (Yuan et al., 1972). Endlich and Linn (1985a), from studies on the mechanism of cleavage.
by EcoB, have suggested that DNA translocation, mediated by ATP hydrolysis, allows the enzyme to scan the DNA molecule, prior to cleavage, for the presence of modified recognition sites and to check that the DNA has not been already cleaved. The observed post cleavage ATPase activity may therefore be a consequence of the fact that while, after restriction, the enzyme has lost its endonucleolytic activity, it retains the ability to scan the DNA. The ATPase activity was found to be dependent on the loop structure formed by the binding of the enzyme to the DNA; if this structure was cleaved or disrupted the ATPase activity stopped (Endlich and Linn, 1985a). Therefore, although the ability of the enzyme to continue scanning the DNA molecule after cleavage would appear to be wasteful in terms of energy, in vivo ATP hydrolysis is probably stopped by the degradation of restricted DNA by cellular nucleases.

Such a scanning model has additional implications. Endlich and Linn (1985b) suggested that the structure at the 5' termini of restricted DNA, which blocks T4 polynucleotide kinase action (Eskin and Linn, 1972a; Murray et al., 1973b), may also be responsible for the finding that, in the case of EcoB, only one double-stranded cleavage is introduced per recognition site (Eskin and Linn, 1972a). When such 5' termini were treated with λ
exonuclease, which removes about 300 bases, the
"new" termini could be phosphorylated and the DNA
was susceptible to restriction. This implied that
the normal lack of cleavage of DNA which has already
been restricted is not due to a change in the
recognition site, or to the enzyme/DNA complex at
this site, but to an as yet unknown structure at
the 5' termini. According to the "scanning" model
(Endlich and Linn, 1985a,b), a fresh EcoB molecule
would "track" along restricted DNA (or wind the
DNA past it) to the 5' end and the structure there
would induce release of the enzyme before the DNA
is cleaved. This means that the enzyme would normally
scan past the cleavage site and then track back to
it, a procedure which also implies that cleavage
sites are not entirely random. Such a system,
although very complex, would avoid wasteful cleavage
of DNA which had already been restricted.

The above model however if taken at face
value does not explain how double-stranded cleavage
of the DNA is achieved by two enzyme molecules.
The second cleavage, by a second enzyme molecule,
would not be possible if exactly the same scanning
procedure applied, as the second enzyme molecule
would encounter the 5' terminus (if unidirectional
translocation is assumed) created by the first
enzyme molecule and would be released. However,
Endlich and Linn (1985b) propose a mechanism in
which the first enzyme molecule remains anchored to both the recognition site and the site of single-stranded cleavage and may direct the second enzyme molecule to verify the unmodified nature of the other strand. Support for this idea is provided by the finding that if the enzyme/DNA complex is disrupted, single-stranded gapped intermediates are not a substrate for the second EcoB scission.

It is not known whether this complex scanning procedure is unique to EcoB or whether it is a property of other Type I restriction endonucleases. In any case, it highlights the complexity of these enzymes and indicates that they are perhaps more efficient in their mode of action and their energy requirements than previously thought.

Step 9.

When EcoK* finds an sK site which is semi-modified (hemi-methylated), the recognition complex formed causes EcoK* to methylate the unmodified strand (Burckhardt et al., 1981 a, b). The enzyme has at least two binding sites for AdoMet: an effector site which is involved in the transition to EcoK* and a methyl transfer site. Once the enzyme has bound to semi-modified DNA, the nature of the complex formed allows the methyl transfer site to become accessible and this catalyses the transfer of a methyl group from AdoMet to the adenine residue in the unmodified strand. ATP stimulates methylation
but it is not an essential requirement (Vovis et al., 1974; Burckhardt et al., 1981a). EcoK appears to turn over during methylation, thus behaving catalytically; this is in contrast to its behaviour during endonucleolytic cleavage, as seen above. This was suggested from the finding that the addition of more heteroduplex DNA, AdoMet and ATP, to an EcoK methylation reaction resulted in increased methylation (Burckhardt et al., 1981a).

Note: Reaction Mechanisms of EcoA

The enzymatic activities of EcoA (Suri and Bickle, 1985) are basically similar to those of classical Type I enzymes. In all known cases, modification methylation is catalysed by both a two-subunit modification enzyme and the three-subunit restriction endonuclease, which is capable of both restriction and modification. However, one significant difference has been observed in the modification reaction carried out by EcoA; while hemi-methylated DNA is methylated by both EcoA and EcoK at comparable rates, unmodified DNA is methylated efficiently by EcoA (Suri and Bickle, 1985), but inefficiently by enzymes of the K family (Vovis et al., 1974; Burckhardt et al., 1981a). This may reflect a difference between the reaction mechanisms of EcoA and those of EcoK-related enzymes.
1.5 ANTI-RESTRICTION MECHANISMS

Bacteriophages encounter a range of different hosts in their natural environment and thus potentially a variety of restriction and modification systems. Several phages have evolved mechanisms to counteract restriction, which allows them to infect restricting hosts of different specificities with a higher probability of survival. Anti-restriction mechanisms have been detected in many coliphages, including T3, T5, T7, λ, P1 and Mu (see Bickle, 1982). One of the most extensively studied mechanisms is that mediated by the early gene ral (restriction alleviation) in λ, located between cIII and N, which both alleviates restriction and enhances DNA modification (Zabeau et al., 1980). The action of Ral is of particular interest with respect to the study of Type I restriction endonucleases, as it has been found to counteract restriction by the Type I enzymes EcoK and EcoB but has no effect on the Type II enzymes EcoRI and EcoRII or the Type III enzyme EcoPl. Debrouwere et al. (1980b) suggested that Ral also antagonises other E.coli enzymes which, like Type I enzymes, are ATPases, e.g. RecBC nuclease and Rho. However, as regards its anti-restriction activity it is not known whether its effect on EcoK and EcoB reflects a general effect on all Type I restriction endonucleases.

Unmodified phages can also escape restriction when infecting restricting hosts at high multiplicity. Heip et al. (1974) showed that when E.coli K12 or E.coli B
cells were infected with a non-modified $\lambda$ phage (helper phage) at high multiplicity and subsequently with a non-modified $\lambda$ vir phage (test phage), the efficiency of plating of the $\lambda$ vir phage was increased. They suggested that the helper phage was saturating the restriction system due to the high multiplicity of infection. This would correlate with the finding that the restriction enzyme does not turn over in its activity as an endonuclease, as discussed in section 1.4d (Eskin and Linn, 1972a). Zabeau et al. (1980) subsequently found that the reduction in restriction of $\lambda$ vir phage by pre-infection with $\lambda$ helper phage observed by Heip et al. (1974) was due to the expression of ral in the helper phage. Although Ral did not alleviate restriction of the helper phage, it relieved restriction of the $\lambda$ vir phage when provided in trans. By using $\lambda$ral$^+$ and $\lambda$ral$^-$ phages in similar experiments Zabeau et al. (1980) showed that while $\lambda$ cannot use the product of its own ral gene to escape restriction, it can provide restriction alleviation in trans. The mechanism of this is, however, as yet unknown.

1.6 THE DISTRIBUTION OF TYPE I RESTRICTION AND MODIFICATION SYSTEMS AMONG BACTERIA

Since the initial discovery of the phenomenon of restriction and modification, there has been considerable interest in the occurrence of host specificity in different strains of bacteria. The examination of different bacterial
strains and species for the presence of Type I restriction and modification systems has shown that, at least among enterobacteria, several strains possess these chromosomally-encoded systems. Nevertheless, many strains have not been examined and some which have been extensively studied, e.g. E.coli C, do not appear to have such systems. Among the strains which have been found to express host-specific restriction and modification, the number and groups of systems present, even if only Type I systems are considered, show considerable variation.

Extensive studies of the distribution of these systems among E.coli strains have not been documented. However, the reported E.coli hsd systems suggest the presence of two families of Type I restriction and modification systems among strains of E.coli: hsd K, from E.coli K12, and the related hsd B and hsd D, from E.coli B and E.coli El66 respectively, constitute the K-family while hsd A, from E.coli 15T (Arber and Wauters-Willems, 1970), which does not appear to be related to hsd K but physiologically behaves as a Type I system (Lark and Arber, 1970), could constitute an alternative family. The first evidence of the close relationship between hsd K and hsd B came from genetic complementation tests (Boyer and Roulland-Dussoix, 1969). Similar experiments involving hsd A failed to show any relatedness between this system and hsd K (W. Arber, pers. comm.). All the hsd systems mentioned above are allelic and map at the same region of the E.coli chromosome (Boyer, 1964; Glover and Colson, 1969;
Arber and Wauters-Willems, 1970). Molecular evidence of the relationship between hsd K, hsd B, and hsd D was obtained from DNA hybridisation and immunological cross-reaction studies (Murray et al., 1982); these failed to detect any homology between hsd A and the K-related systems. The genetic and molecular experiments therefore support the idea of a K family of restriction and modification systems with the possibility of an alternative A family. Proof that an A family exists requires the identification and isolation of A-related systems and analysis of the members of this new family with respect to the genetic organisation of the systems and the restriction endonucleases they encode. Comparisons of the A-related systems would indicate whether the A family, if found, is truly an alternative to the K family. This is the approach employed in this thesis. Screening of a variety of bacterial strains, using DNA hybridisation techniques, for the presence of systems sharing DNA homology with hsd A resulted in the isolation of hsd E. Analyses of the A and E systems, including complementation tests comparable to those carried out between hsd K and hsd B (Boyer and Roulland-Dussoix, 1969; Glover, 1970; Hubacek and Glover, 1970) indicate that hsd A and hsd E constitute a family of restriction and modification systems alternative to that of hsd K. The subunits of the endonucleases encoded by hsd A and hsd E are found to be interchangeable; this highlights their close relationship. On the other hand, when hsd A is compared with hsd K, there is no evidence of
DNA homology (Murray et al., 1982) and the subunits of their respective endonucleases are not interchangeable (W. Arber, pers. comm.). On the basis of these criteria the $\text{hsd K-related}$ and $\text{hsd A-related}$ systems can be divided into two families. The variation of Type I restriction and modification systems among $E.\text{coli}$ strains and perhaps even the occurrence may be somewhat limited, as so far only the strains mentioned above have been shown to possess these systems. However, a true estimate of the overall distribution, and frequency of occurrence, would require the screening of many more natural isolates of $E.\text{coli}$.

Studies on the restriction and modification properties of $\text{Salmonella}$ have concentrated on the characterisation of the various Type I systems found in several different serotypes and comparison of these systems with those of $E.\text{coli} \ K12$ and $E.\text{coli} \ B$. The result of this approach has been that studies on the distribution of such systems among different $\text{Salmonella}$ serotypes have been considerably more comprehensive than those involving $E.\text{coli}$ strains. These investigations have shown that most known $\text{Salmonella}$ serotypes possess at least one Type I restriction and modification system and in many cases more than one system are present (Bullas et al., 1980). The analysis of $\text{Salmonella hsd}$ systems has however been complicated by the fact that some are not active on the $\text{Salmonella}$ phages P22 and L and therefore went unnoticed in the early experiments. The construction of
Salmonella/E. coli hybrids, the transfer of Salmonella hsd genes to E. coli, and the use of phages P3 and \lambda in addition to P22 and L, allowed the identification and analysis of several restriction and modification systems in a wide variety of Salmonella serotypes.

Three different chromosomally located classes of systems have been described in Salmonella: the LT system, which Colson et al. (1970) mapped at 8.5 minutes; the SA system which was mapped close to serB at 98 minutes (Colson and Colson, 1971; Colson and Van Pel, 1974); the hsd K related systems, SB, SP and SQ, located between serB and SA (Colson and Van Pel, 1974). (These systems were described in detail in section 1.3). Although only the K-related systems have been extensively studied and confirmed as Type I systems (Van Pel and Colson, 1974; Bullas and Colson, 1975a; Bullas et al., 1976), the other systems are also thought to be of type I from genetic and DNA methylation studies (Colson et al., 1969; Hattman, 1971; Colson and Van Pel, 1974; Hattman et al. 1976). Bullas et al. (1980) screened Salmonella strains representing eighty five different serotypes for the presence of restriction and modification systems and showed that they fell into three groups on the basis of their restriction-modification phenotype:

(i) those lacking restriction and modification systems;
(ii) those possessing the LT system;
(iii) those which, in addition to LT possess other serB-linked systems such as SA or SB.
In group (iii) some strains only carry SB- or SP-like (K-related) systems, while others have the SA system in conjunction with SB- or SP-like ones.

The relationship between SB, SP, SQ and the \textit{E.coli} hsd K and hsd B systems was shown by genetic complementation tests (Van Pel and Colson, 1974), and by DNA hybridisation and immunological cross-reaction experiments (Murray et al., 1982). Both these approaches failed to detect any relatedness between SA and the \textit{E.coli} hsd K family. Following the finding that the SB, SP and SQ systems were allelic and functionally homologous to those of \textit{E.coli} K12 and \textit{E.coli} B, interest has tended to focus on these systems.

Bullas et al. (1980) also observed, from their extensive survey, that the specificity of the serB-linked systems was unique for each serotype, with different strains of the same serotype expressing the same specificity. Thus, the classification of serotypes according to their restriction-modification systems paralleled taxonomic classification (Borman et al., 1944; Kauffmann and Edwards, 1952), suggesting that subdivision according to serB-linked systems may be significant in defining a "biological grouping" of the different serotypes, which may be ultimately useful in describing the \textit{Salmonella} species. This observation may also have implications concerning the evolution and distribution of such systems among different species of bacteria.
Another restriction and modification system, also thought to be of Type I, has been identified in *Haemophilus influenzae* (Gromkova and Goodgal, 1976). The presence of such a system in *H. influenzae* is of particular interest as all the Type I systems reported to date have been found in members of the family *Enterobacteriaceae*. If it is indeed Type I, this would pose further questions concerning distribution of these systems. However, it has not been extensively studied, and any conclusions must await further analysis.

Surveys of the genetic diversity in natural *E. coli* populations, using electrophoretic techniques to study variation in chromosomal genes encoding enzymes, and in the distribution of cryptic plasmids, identified a large number of electrophoretic types, and in most cases the plasmid composition of each type was unique (Selander and Levin, 1980; Caugant *et al.*, 1981). However, despite the extensive variability, the number of distinctive "genotypes" appeared to be rather limited, with identical clones, as regards electrophoretic type, being often obtained from unassociated hosts (Selander and Levin, 1980).

A survey of the distribution of Type I restriction and modification systems among a range of *E. coli* natural isolates and investigation of their relatedness to the known *E. coli* systems may provide some other parameters for classification. Such a survey would be of particular interest if the *E. coli* strains were also classified by some other means as it would highlight any correlation between
the two classifications. A survey of this kind, especially if a range of other enterobacteria is included, may also give an insight into the ways in which these restriction and modification systems have evolved and diversified and may even provide some information concerning their physiological role in the natural environment.

The Type I restriction endonucleases studied so far all recognise asymmetric, hyphenated DNA sequences, with two specific domains flanking a spacer of definite length but unspecified sequence (see section 1.4.b). The finding that the hsdS genes of hsd K and hsd B share a conserved region in the middle of the gene flanked by two non-homologous regions has led to suggestions that the variable regions may correlate with the specific domains of the recognition sequence (Gough and Murray, 1983). Heteroduplex analysis of the hsdS genes of the Salmonella SB, SP and SQ systems indicated that the SB and SP genes, like those of hsd K and hsd B, share a central conserved region flanked by two variable regions and showed that in the case of SQ, the hsdS gene has one variable region from SB, the other from SP (Fuller-Pace et al., 1984). This confirmed earlier suggestions that the origin of the SQ specificity was recombination between the hsdS genes of SB and SP (Bullas et al., 1976). The finding that the sequence recognised by SQ derives one domain from SB, the other from SP (Nagaraja et al., 1985 b,c) is consistent with the correlation of the variable regions of the hsdS gene with the specific domains of the recognition
sequence. The recombination event that generated SQ therefore reassorted two domains of the recognition polypeptide each recognising a specific domain of the recognition sequence. It is tempting to believe that reassortment of the variable regions of the hsdS gene is important in the diversification of sequence specificity, in a way reminiscent of the diversification of immunoglobulin variability (Gough, 1981).
CHAPTER 2

MATERIALS AND METHODS
2.1 MATERIALS

(a) BACTERIAL STRAINS (i)

<table>
<thead>
<tr>
<th>Strain No.*</th>
<th>Specificity</th>
<th>Relevant mutations/ features</th>
<th>Source +</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA2899</td>
<td>A</td>
<td>hsd A genes in E.coli K12</td>
<td>W.A.</td>
<td>W. Arber, pers. comm.</td>
</tr>
<tr>
<td>WA2552</td>
<td>A</td>
<td>hsdR</td>
<td>W.A.</td>
<td>Arber &amp; Wauters-Willems (1970)</td>
</tr>
<tr>
<td>NM555</td>
<td>A</td>
<td>hsdMS₂ derivative of WA2899</td>
<td>-</td>
<td>This thesis</td>
</tr>
<tr>
<td>5K</td>
<td>K</td>
<td>hsdR</td>
<td>N.E.M.</td>
<td>Hubacek &amp; Glover (1970)</td>
</tr>
<tr>
<td>NM477</td>
<td>K</td>
<td>hsdMSΔ(Δ5)</td>
<td>N.E.M.</td>
<td>Gough &amp; Murray (1983)</td>
</tr>
<tr>
<td>477 (λ)</td>
<td>K</td>
<td>imm² lysogen of NM477</td>
<td>-</td>
<td>This thesis</td>
</tr>
<tr>
<td>477 (21)</td>
<td>K</td>
<td>imm²¹ lysogen of NM477</td>
<td>-</td>
<td>This thesis</td>
</tr>
<tr>
<td>NM531</td>
<td>K</td>
<td>hsdR recA 13</td>
<td>N.E.M.</td>
<td>-</td>
</tr>
<tr>
<td>M159</td>
<td>K</td>
<td>uvrA</td>
<td>N.E.M.</td>
<td>Jaskunas et al. (1975)</td>
</tr>
<tr>
<td>WL542</td>
<td>K</td>
<td>minA minB</td>
<td>W.L.</td>
<td>Loenen, pers. comm.</td>
</tr>
<tr>
<td>HB101</td>
<td>B</td>
<td>hsdS recA</td>
<td>N.E.M.</td>
<td>Boyer &amp; Roulland-Dussoix (1969)</td>
</tr>
<tr>
<td>NM522</td>
<td>K</td>
<td>lac-proΔhsdMSΔ</td>
<td>N.E.M.</td>
<td>Gough &amp; Murray (1983)</td>
</tr>
</tbody>
</table>
(a) Bacterial Strains (i)

Footnotes:

*Derivatives of *E.coli* K12 unless otherwise stated

+W.A. Werner Arber
N.E.M. Noreen E. Murray
G.C. Gianni Cesareni
W.L. Wil Loenen
Natural isolates used in screening for hsd K- and hsd A-related sequences

<table>
<thead>
<tr>
<th>Strain No.</th>
<th>Relevant features</th>
<th>Source**</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A58</td>
<td>E.coli (API)*</td>
<td>K.K./K.C.</td>
<td>Duguid et al. (1955)</td>
</tr>
<tr>
<td>A101</td>
<td>E.coli (API)</td>
<td>K.K./K.</td>
<td>Duguid et al. (1955)</td>
</tr>
<tr>
<td>A204</td>
<td>Citrobacter (API)</td>
<td>K.K./K.C.</td>
<td>Duguid et al. (1955)</td>
</tr>
<tr>
<td>E136</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>E147</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>E148</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>E161</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>E163</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>E164</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>E171</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>581</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
<tr>
<td>653</td>
<td>E.coli (API)</td>
<td>K.K./K.C.</td>
<td>-</td>
</tr>
</tbody>
</table>
### (a) Bacterial Strains (ii) (continued)

<table>
<thead>
<tr>
<th>Strain No.</th>
<th>Relevant features</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRLET-13</td>
<td><em>E. coli</em> K-like</td>
<td>B.R.L.</td>
<td>Caugant et al. (1981)</td>
</tr>
<tr>
<td><em>E. coli</em> W</td>
<td>lysogenic for phage WØ</td>
<td>N.E.M.</td>
<td>Glover and Kerszman (1967)</td>
</tr>
<tr>
<td><em>S. typhimurium</em> 4247</td>
<td>hsd SA:LT:SB</td>
<td>N.E.M.</td>
<td>Bullas et al. (1980)</td>
</tr>
<tr>
<td><em>Shigella boydii</em></td>
<td>wild type+</td>
<td>K.K./K.C.</td>
<td></td>
</tr>
<tr>
<td><em>Shigella dysenteriae</em></td>
<td>wild type</td>
<td>K.K./K.C.</td>
<td></td>
</tr>
<tr>
<td><em>Shigella flexneri</em></td>
<td>wild type</td>
<td>K.K./K.C.</td>
<td></td>
</tr>
<tr>
<td><em>Shigella sonnei</em></td>
<td>wild type</td>
<td>K.K./K.C.</td>
<td></td>
</tr>
<tr>
<td><em>Citrobacter freundii</em></td>
<td>wild type</td>
<td>K.K./K.C.</td>
<td></td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em></td>
<td>Wild type (M5a1)</td>
<td>Dixon et al. (1977)</td>
<td></td>
</tr>
</tbody>
</table>
Control *E. coli* strains used in the screening programme

<table>
<thead>
<tr>
<th>Strain No.</th>
<th>Specificity system</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>C600</td>
<td>hsd K‡</td>
<td>N.E.M.</td>
<td>Appleyard (1954)</td>
</tr>
<tr>
<td><em>E. coli</em> B</td>
<td>hsd B‡</td>
<td>N.E.M.</td>
<td>Studier (1969)</td>
</tr>
<tr>
<td>15T⁻</td>
<td>hsd A; P15</td>
<td>W.A.</td>
<td>Arber &amp; Wauters-Willems (1970)</td>
</tr>
<tr>
<td><em>E. coli</em> C-1a</td>
<td>no known system</td>
<td>N.E.M.</td>
<td>Bertani (1968)</td>
</tr>
<tr>
<td>CR63</td>
<td>hsd K (λ resistant)</td>
<td>N.E.M.</td>
<td>Appleyard <em>et al.</em> (1956)</td>
</tr>
</tbody>
</table>

* API refers to Analytical Profile Index 20E system, Analytab Products, Division of Ayerst Laboratories Inc., 200 Express Street, Plainview, New York 11803. 20 biochemical tests are performed on each bacterial strain and the profile is matched with that of a large number of coliform strains to give a "best fit". All strains were classified by Keith Cartwright at the Western General Hospital, Edinburgh.

** J.P.D.  J. P. Duguid  
K.K.  Kim Kaiser  
K.C.  Keith Cartwright  
B.R.L.  Bruce R. Levin  
C.K.  Christina Kennedy  
N.E.M.  Noreen E. Murray  
W.A.  Werner Arber  

*** Strains were compared with *E. coli* K and *E. coli* B according to electrophoretic mobility of several enzymes produced by the strains; see Selander & Levin (1980), Caugant *et al.* (1981) All strains had previously been classified as *E. coli* by A.P.I.
Footnotes (continued)

† Shigella and Citrobacter strains were the standard wild type strains kept at the Western General Hospital, Edinburgh

‡ The host specificity systems K, B, and D belong to the K-family; see Gough and Murray (1983).
(b) PHAGE STRAINS (λ)

<table>
<thead>
<tr>
<th>Strain No.*</th>
<th>Relevant features</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM761</td>
<td>att-redΔ, imm^{21} c^{+} nin 5</td>
<td>Murray et al (1977)</td>
</tr>
<tr>
<td>NM762</td>
<td>att-redΔ cIΔ nin 5</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>FFP 1**</td>
<td>hsd A genes in NM762</td>
<td>This thesis</td>
</tr>
<tr>
<td>NM1185</td>
<td>hsd SP genes in NM762</td>
<td>Fuller-Pace et al. (1984)</td>
</tr>
<tr>
<td>NM5</td>
<td>λ^{+}</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>NM63</td>
<td>λ cI 26</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>NM144</td>
<td>h^{82} b522 immλ cI</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>NM243</td>
<td>λ vir</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>NM507</td>
<td>λ imm^{21} cI</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>NM675</td>
<td>h^{80} att^{80} cI857 nin 5</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>NM848</td>
<td>h^{82} b522 imm^{21} cI</td>
<td>&quot;&quot;</td>
</tr>
<tr>
<td>NM1090</td>
<td>λral 18 cI60</td>
<td>Debrouwere et al. (1980a)</td>
</tr>
<tr>
<td>NM1094</td>
<td>λral^{+} cI60</td>
<td>&quot;&quot;</td>
</tr>
</tbody>
</table>

* All phages, except for FFP1 were obtained from Noreen E. Murray

** Deletion, mutant, and other derivatives of FFP1 (λhsd A), which were isolated or constructed for this thesis, are listed in the relevant chapters.
(c) PLASMIDS

(i) Vectors

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Confers resistance to*</th>
<th>Source**</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBR322</td>
<td>Ap, Tc</td>
<td>N.E.M.</td>
<td>Bolivar et al. (1977)</td>
</tr>
</tbody>
</table>

(ii) Derivatives of pBR322:

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relevant features</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRH1</td>
<td>hsd K genes†</td>
<td>N.E.M.</td>
<td>Sain &amp; Murray (1980)</td>
</tr>
<tr>
<td>pGC1</td>
<td>hsd E genes†</td>
<td>G.M.C.</td>
<td>Fuller-Pace et al. (1984)</td>
</tr>
<tr>
<td>pBgH</td>
<td>hsdS (E.coli K12)**</td>
<td>N.E.M.</td>
<td>Murray et al. (1982)</td>
</tr>
</tbody>
</table>

* Ap Ampicillin
  Tc Tetracycline

** N.E.M. Noreen E. Murray
  G.M.C. Gill M. Cowan
  K.K. Kim Kaiser

† pRH1 contains DNA from within the hsd K region and does not encode functional polypeptides. pGC1 contains complete, functional hsd E genes.

‡ pK13 is a derivative of pBR322 which forms monomers and oligomers with a 2.15 kb unit size

++ pBgH contains a 400 bp HindIII-BglIII fragment from within the hsdS gene of E.coli K12.
<table>
<thead>
<tr>
<th>Vector</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mp10</td>
<td>Messing (1983)</td>
</tr>
<tr>
<td>Mp11</td>
<td>&quot;</td>
</tr>
<tr>
<td>Mp18</td>
<td>Yanisch-Perron et al. (1985)</td>
</tr>
<tr>
<td>Mp19</td>
<td>&quot;</td>
</tr>
</tbody>
</table>
(e) CHEMICALS

(i) Enzymes

Restriction endonucleases were purchased from Amersham International plc, New England Biolabs and NBL Enzymes Ltd.

T4 DNA ligase was obtained from New England Biolabs or prepared in this laboratory.

E.coli DNA polymerase was purchased from NBL Enzymes Ltd., while E.coli DNA polymerase (Klenow fragment) was obtained from Pharmacia Ltd.

DNase I, RNase A, Proteinase K and lysozyme were all obtained from Sigma Chemical Company Ltd.

(ii) Primers

M13 sequencing primer (17-mer) and M13 hybridisation probe-primer were obtained from New England Biolabs.

Synthetic oligonucleotide primers were a gift from Eric Kawashima, Biogen SA, Geneva.

(iii) Nucleotide Triphosphates

Adenosine triphosphate was supplied by Sigma Chemical Company Ltd.

Deoxynucleotide triphosphates and dideoxy-nucleotide triphosphates were purchased from Pharmacia Ltd.
(iv) **Gel Components**

Standard and Seakem LE pure agarose were purchased from Miles Laboratories Ltd.

Acrylamide and bis-Acrylamide were supplied by BDH Chemicals Ltd.

Ammonium persulphate and TEMED (N,N,N',N'-tetramethyl ethylene diamine) were obtained from Sigma Chemical Company Ltd.

(v) **Radiochemicals**

All radiochemicals, namely $\alpha^{32}$P deoxycytidine triphosphate, $\alpha^{35}$S deoxyadenosine triphosphate and $35$S-methionine were supplied by the Radiochemical Centre, Amersham International, PLC.

(vi) **Scintillation Components**

Soluene-350, 2,5-Diphenyloxazole (PPO), 1,4-bis[2-Methyl-5-Phenloxazolyl]-Benzene (di-Methyl POPOP) and 2-4'-tert-butylphenyl[-5-4''-biphenyl]-1,3-4-oxadiazole (Butyl PBD) were purchased from Packard Instruments Ltd.

(vii) **Antibiotics**

Ampicillin (Beecham Pharmaceuticals-Penbritin) was obtained from The Royal Infirmary, Edinburgh.

Tetracycline hydrochloride, D-cycloserine, and Chloramphenicol were purchased from Sigma Chemical Company Ltd.

(viii) **Bacteriological Media**

Baltimore Biological Laboratories trypticase was purchased from Becton-Dickinson (UK) Ltd. Other media were supplied by Difco Laboratories.
(ix) **Miscellaneous Compounds**

Vitamin B1, Threonine, Leucine, L-methionine, Dithiothreitol, Dimethyl sulphoxide, β-mercaptoethanol, Spermidine trihydrochloride and Putrescine (Tetramethyl diamine), and Isopropyl-β-D-Thiogalactoside (IPTG) were obtained from Sigma Chemical Company Ltd.

5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside (Xgal) was supplied by Boehringer Corporation (London) Ltd.

Bromophenol Blue, Coomassie Brilliant Blue G250, Ethidium bromide and Xylene cyanol FF were obtained from BDH Chemicals Ltd.

Difco Methionine Assay Medium was purchased from Difco Laboratories.

Protein gel standard markers (Electrophoresis Calibration Kit) were purchased from Pharmacia Ltd.

All other reagents were of the highest grade commercially available.

**Addresses of Suppliers**

Amersham International plc, P.O. Box 16, Amersham,
Buckinghamshire, HP7 9LL

Becton -Dickinson (UK) Ltd., Between Towns Road,
Cowley, Oxford, OX4 3LY

BDH Chemicals Ltd., Broom Road, Poole, BH12 4NN

Bethesda Research Laboratories (UK) Ltd., P.O. Box 145,
Science Park, Cambridge, CB4 4BE

Boehringer Corporation (London) Ltd., Boehringer
Mannheim House, Bell Lane, Lewes, East Sussex BN7 1LG
Difco Laboratories, P.O. Box 14B, Central Avenue, East Molesey, Surrey, KT8 OSE.
Miles Laboratories, Ltd., P.O. Box 37, Stoke Poges, Slough, SL2 4LY
NBL Enzymes Ltd., 3 Laburnum Terrace, Ashington, Northumberland, NE63 0XX.
New England Biolabs., CP Laboratories Ltd. (Distributors), P.O. Box 22, Bishops Stortford, Herts.
Packard Instruments Ltd., 13-17 Church Road, Caversham, Berks., RG4 7AA.
Pharmacia Ltd., Pharmacia House, 351 Midsummer Boulevard, Central Milton Keynes, MK9 3YY.
Sigma Chemical Company Ltd., Fancy Road, Poole, Dorset, BN17 7NH.

(f) BACTERIOLOGICAL MEDIA
All media were sterilised by autoclaving at 15 lb/in² for 15 minutes before use.

L-Broth
Difco Bacto Tryptone 10 g
Difco Bacto Yeast Extract 5 g
Sodium Chloride 10 g
Distilled water to 1 litre
Adjusted to pH 7.2 with sodium hydroxide before autoclaving.
L-Agar

- Difco Bacto Tryptone: 10 g
- Difco Bacto Yeast Extract: 5 g
- Sodium chloride: 10 g
- Difco Bacto Agar: 15 g
- Distilled water: to 1 litre

Adjusted to pH 7.2 with sodium hydroxide before autoclaving.

BBL Agar

- Trypticase (Baltimore Biological Labs.): 10 g
- Sodium chloride: 10 g
- Difco Bacto Agar: 10 g
- Distilled water: to 1 litre

pH unaltered.

BBL Top Layer

As for BBL agar, but Difco Bacto Agar added to 0.65%. Top layer was supplemented with 1mM magnesium sulphate for \( \lambda \) phages, except when EDTA was added for the selection of deletion derivatives.

Minimal Agar

- Difco Bacto Agar: 15 g
- Distilled water: 790 ml

pH unaltered

The following were added as sterile solutions to the molten agar after autoclaving:
Spizizen Salts (5x) * 200 ml
Glucose (20%) 10 ml
Vitamin B1 (2 mg ml⁻¹) 0.25 ml

*Spizizen Salts (5x)
Ammonium sulphate 10 g
di-Potassium hydrogen phosphate 70 g
Potassium dihydrogen phosphate 30 g
tri-Sodium citrate .2H₂O 5 g
Magnesium sulphate 7H₂O 1 g
Distilled water to 1 litre

M9-Maltose Minimal Medium
M9 Salts (4x) ** 250 ml
Magnesium sulphate .7H₂O (1M) 1 ml
Maltose (20%) 15 ml
Distilled water to 1 litre
All compounds mixed as pre-sterilised solutions.

**M9 Salts (4x)
di-Sodium hydrogen phosphate (anhydrous) 28 g
Potassium dihydrogen phosphate 12 g
Sodium chloride 2 g
Ammonium chloride 4 g
Distilled water to 1 litre

Phage Buffer
di-Sodium hydrogen phosphate (anhydrous) 7 g
Potassium hydrogen phosphate 3 g
Sodium chloride 5 g
Magnesium sulphate $\cdot 7H_2O (0.1M) \quad 10 \text{ ml}$
Calcium chloride $(0.01M) \quad 10 \text{ ml}$
Gelatin $(1\% \text{ w/v}) \quad 1 \text{ ml}$
Distilled water \quad to 1 litre

**Xgal-Indicator Plates**

5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside (Xgal) and the inducer Isopropyl-β-D-Thiogalactoside (IPTG) were added to the BBL top layer before adding cells/phage, for indicator plates.

- **Xgal** $(20 \text{ mg ml}^{-1}) \quad 30 \mu l$
- **IPTG** $(20 \text{ mg ml}^{-1}) \quad 20 \mu l$
- **BBL Top Layer** \quad 2.5 ml

**Antibiotics**

- **Ampicillin** \quad 100 $\mu g \text{ ml}^{-1}$
- **Tetracycline** \quad 10 $\mu g \text{ ml}^{-1}$
- **Chloramphenicol** $^+$ \quad 150 $\mu g \text{ ml}^{-1}$

Antibiotics were used at the above concentrations and, when used in Agar, were added to the molten medium immediately before pouring.

$^+$Chloramphenicol was used at this concentration for plasmid amplification when growing cells carrying pBR322 or derivatives of this plasmid.
BUFFERS AND STOCK SOLUTIONS

Tris-HCl (1M) Tris Base in distilled water, pH adjusted as required with concentrated hydrochloric acid before autoclaving.

EDTA (0.5M) Ethylenediaminetetraacetate (disodium salt) in distilled water, adjusted to pH 7.5 and pH 8.0 as required with 10M sodium hydroxide before autoclaving.

- Ammonium sulphate (1M) in distilled water; sterilised by autoclaving
- Magnesium sulphate (1M)
- Calcium chloride (1M)
- Magnesium chloride (1M)
- Potassium chloride (1M)
- Sodium chloride (1M and 5M)
- Ammonium acetate (5M) in distilled water; sterilised by filtration
- Potassium acetate (5M)
- Sodium acetate (3M)
- Sodium hydroxide (10M) - in distilled water
- SDS (sodium dodecyl sulphate) 10% in distilled water
- Triton X-100 20% in distilled water
- Glucose (20%) all in distilled water; sterilised
- Maltose (20%)
- Gelatin (2 mg ml⁻¹; and 10 mg ml⁻¹) by autoclaving
- DTT (Dithiothreitol) (100 mM and 1M) in distilled water; stored at -20°C.
DNase I (10 mg ml$^{-1}$) in distilled water; stored at -20°C.

RNase A (10 mg ml$^{-1}$) in distilled water; boiled for 10 minutes to remove endogenous DNase activity and stored at -20°C.

Proteinase K (10 mg ml$^{-1}$) in distilled water; prepared immediately before use.

**TE**

- Tris-HCl (1M) pH 8.0 10 ml
- EDTA (0.5M) pH 8.0 2 ml
- Distilled water 988 ml

**TBE (10x)**

- Tris Base 108 g
- Boric acid 55 g
- EDTA 9.3 g
- Distilled water to 1 litre

Buffer was filtered before storage.

**SSC (20x)***

- Sodium chloride 175.3 g
- tri-Sodium citrate 88.2 g
- Distilled water to 1 litre

*Subsequently diluted as required in distilled water.

**Equilibrated Phenol** - Phenol was distilled, saturated with TE and stored at -20°C, protected from light.
DNA Sample Buffer (5x)

Ficoll 400 2 g
Bromophenol Blue 20 mg
Distilled water to 10 ml

Ethidium Bromide - 10 mg ml$^{-1}$ in distilled water; stored at 4°C, protected from light.

Special Buffers and Solutions for Particular Techniques

(i) DNA Restriction

Restriction Buffers (10x)

<table>
<thead>
<tr>
<th></th>
<th>Low Salt</th>
<th>Medium Salt</th>
<th>High Salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl pH 8.0</td>
<td>10mM</td>
<td>10mM</td>
<td>50mM</td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td>10mM</td>
<td>10mM</td>
<td>10mM</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>-</td>
<td>50mM</td>
<td>100mM</td>
</tr>
</tbody>
</table>

Buffers were prepared from sterile 1M stock solutions. They were then used according to the requirements of each enzyme.

Sma I Restriction Buffer (10x)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl pH 8.0</td>
<td>150mM</td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td>50mM</td>
</tr>
<tr>
<td>Potassium chloride</td>
<td>150mM</td>
</tr>
</tbody>
</table>

Sterile 1M solutions of the components were used. (Sma I required Potassium chloride.)
DNA Ligation

Ligase Dilution Buffer
- Tris-HCl (1M) pH 7.5: 0.2 ml
- Potassium chloride (1M): 1 ml
- Glycerol (90%): 5.74 ml
- Gelatin (10 mg ml⁻¹): 50 μl
- Triton X-100 (20%): 50 μl
- β-mercaptoethanol: 10 μl
- Distilled water: 3.95 ml

Stored at -20°C.

Ligase Buffer (10x)
- Magnesium chloride (1M): 1 ml
- Tris-HCl (1M) pH 7.5: 6.6 ml
- EDTA (0.5M) pH 8.0: 0.2 ml
- Distilled water: 2.2 ml

Stored at -20°C.

ATP (Adenosine-5'-triphosphate, di-Sodium salt)
10mM solution in sterile distilled water, stored at -20°C.

Transfection/Transformation

SSC/Calcium chloride
- SSC (1X): 6 ml
- Calcium chloride (0.1M): 8 ml

Components mixed as pre-sterilised solutions.
TFB

MES* (1M) pH 6.2 2 ml
Potassium chloride 1.49 g
Manganese chloride 1.75 g
Calcium chloride 0.29 g
Hexamine cobalt chloride 0.16 g
Distilled water to 200 ml

Adjusted to pH 6.15 ± 0.1 before sterilisation by filtration.

*MES (2-[N-morpholino]ethanesulphonic acid); 1M solution stored at -20°C.

DMSO (Dimethyl sulphoxide)
Spectroscopic grade DMSO purged with nitrogen for 15 minutes and 200 μl aliquots stored at -70°C. Aliquots were thawed slowly and never re-frozen.

DTT
2.25 M solution in sterile, 40mM potassium acetate; stored at -20°C.

(iv) **In vitro Packaging**

**Buffer A**

Tris-HCl (1M) pH 8.0 20 μl
EDTA (0.5 M) pH 7.5 2 μl
Magnesium chloride (1M) 3 μl
Distilled water 975 μl

1 μl of β-mercaptoethanol added after autoclaving; buffer stored at -20°C.
**Buffer Ml**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl (0.5M) pH 7.5</td>
<td>6 µl</td>
</tr>
<tr>
<td>Spermidine (50mM)/Putrescine (100 mM)*</td>
<td>300 µl</td>
</tr>
<tr>
<td>Magnesium chloride (1M)</td>
<td>9 µl</td>
</tr>
<tr>
<td>ATP (0.1M)**</td>
<td>75 µl</td>
</tr>
<tr>
<td>β-mercaptoethanol</td>
<td>1 µl</td>
</tr>
<tr>
<td>Distilled water</td>
<td>110 µl</td>
</tr>
</tbody>
</table>

* Spermidine trihydrochloride/ Putrescine (Tetramethylenediamine); solution neutralised with Tris Base and stored at -20°C.

** Neutralised with ammonium hydroxide and stored at -20°C.

Buffer Ml prepared immediately before use.

---

**(v)** Extraction of Bacterial and Plasmid DNA

**Sucrose/Tris-HCl/EDTA**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl (1M) pH 8.0</td>
<td>5 ml</td>
</tr>
<tr>
<td>EDTA (0.5M) pH 8.0</td>
<td>8 ml</td>
</tr>
<tr>
<td>Sucrose</td>
<td>25 g</td>
</tr>
<tr>
<td>Distilled water</td>
<td>to 100 ml</td>
</tr>
</tbody>
</table>

**Lysozyme Solution**

10 mg ml⁻¹ dissolved in above sucrose/Tris-HCl/EDTA solution. The lysozyme solution was prepared just prior to use.
**Triton Lysis Buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triton X-100 (10%)*</td>
<td>1 ml</td>
</tr>
<tr>
<td>Tris-HCl (1M) pH 8.0</td>
<td>12.5 ml</td>
</tr>
<tr>
<td>EDTA (0.5M) pH 8.0</td>
<td>5 ml</td>
</tr>
<tr>
<td>Distilled water</td>
<td>81.5 ml</td>
</tr>
</tbody>
</table>

* In distilled water

**Equilibrated Phenol**

Distilled phenol was equilibrated as follows and stored at -20°C, protected from light.

(i) **Plasmid DNA** — saturated with TE

(ii) **Bacterial DNA** — saturated with:

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl pH 8.0</td>
<td>0.5 M</td>
</tr>
<tr>
<td>EDTA pH 8.0</td>
<td>10 mM</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>10 mM</td>
</tr>
<tr>
<td>SDS</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

(vi) **Plasmid "Mini-Prep"**

**Solution I**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl (1M) pH 8.0</td>
<td>2.5 ml</td>
</tr>
<tr>
<td>EDTA (0.5M) pH 8.0</td>
<td>2 ml</td>
</tr>
<tr>
<td>Glucose (20%)</td>
<td>4.5 ml</td>
</tr>
<tr>
<td>Distilled water</td>
<td>91 ml</td>
</tr>
</tbody>
</table>

**Solution II**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium hydroxide (10M)</td>
<td>1 ml</td>
</tr>
<tr>
<td>SDS (10%)</td>
<td>5 ml</td>
</tr>
<tr>
<td>Distilled water</td>
<td>44 ml</td>
</tr>
</tbody>
</table>

Solution II was prepared just before use.
Solution III

Potassium acetate (5M) 60 ml
Acetic acid (glacial) 11.5 ml
Distilled water 28.5 ml

Southern Transfer and Plague Hybridisation

Denaturation Buffer

Sodium hydroxide (10M) 50 ml
Sodium chloride 87.7 g
Distilled water to 1 litre

This gives 0.5M NaOH/1.5M NaCl.

Neutralisation Buffer

Tris Base 60.5 g
Sodium chloride 87.7 g
Hydrochloric acid (concentrated) 39 ml
Distilled water to 1 litre

This gives 0.5M Tris-HCl/1.5M NaCl, pH 7.5.

Ammonium Acetate Solution

1M ammonium acetate in distilled water, adjusted to pH 8.0 with sodium hydroxide.

Nick Translation/Hybridisation

Nick Translation Buffer

Tris-HCl (1M) pH 7.5 210 µl
Magnesium chloride (1M) 21 µl
Gelatin (2 mg ml⁻¹) 10 µl
Distilled water 759 µl
Stored in 250 µl aliquots at -20°C.
**Deoxynucleotide Triphosphate (dNTP) Buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nick Translation Buffer*</td>
<td>250 μl</td>
</tr>
<tr>
<td>dNTP mix**</td>
<td>10 μl</td>
</tr>
<tr>
<td>β-mercaptoethanol</td>
<td>2.5 μl</td>
</tr>
<tr>
<td>Distilled water</td>
<td>737.5 μl</td>
</tr>
</tbody>
</table>

* See above

**dNTP mix contains:** deoxy-adenosine triphosphate (dTTP), deoxy-adenosine triphosphate (dATP), deoxy-guanosine triphosphate (dGTP), deoxy-cytidine triphosphate (dCTP), deoxy-thymidine triphosphate (dTTP); all at 2 mM in sterile distilled water and stored at -20°C.

dNTP Buffer was stored in 100 μl aliquots at -20°C.

**DNA Polymerase I Dilution Buffer (2x)**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl (1M) pH 7.5</td>
<td>100 μl</td>
</tr>
<tr>
<td>Ammonium sulphate (2M)</td>
<td>100 μl</td>
</tr>
<tr>
<td>Gelatin (2 mg ml⁻¹)</td>
<td>1 ml</td>
</tr>
<tr>
<td>β-mercaptoethanol</td>
<td>2 μl</td>
</tr>
<tr>
<td>Distilled water</td>
<td>1 ml</td>
</tr>
</tbody>
</table>

Buffer stored at -20°C.

**DNase I Dilution Buffer**

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNase I (20 μg ml⁻¹ in TE)</td>
<td>1 μl</td>
</tr>
<tr>
<td>Polymerase I dilution buffer (2 x)*</td>
<td>0.5 ml</td>
</tr>
<tr>
<td>Glycerol</td>
<td>0.5 ml</td>
</tr>
</tbody>
</table>

* See above

Final concentration of DNase I = 2 x 10⁻⁵ mg ml⁻¹.

Solution stored at -20°C.
Denhardt's Solution (20x)

- Bovine Serum Albumin: 0.4 g
- Ficoll 400: 0.4 g
- Polyvinylpyrrolidone: 0.4 g
- Distilled water: to 100 ml

Stored in 5 ml aliquots at -20 °C.

Hybridisation Solution

- Formamide: 50 ml
- SSC (20x): 5 ml
- Denhardt's solution*: 5 ml
- Distilled water: 25 ml

* See above

Scintillation Fluid

- Butyl-PBD: 4 g
- Toluene: to 1 litre

(ix) M12 Template Hybridisation

Hin Buffer (10x)

- Tris-HCl (1M) pH 8.0: 1 ml
- Sodium chloride (1M): 6 ml
- Magnesium chloride (1M): 0.66 ml
- Distilled water: 2.34 ml

dNTP mix

- dATP: all at 500 μM - Nick Translation
- dTTP: dNTP (2mM) mix used, but diluted
- dGTP: 1/4 (see above Section viii)
Hybridisation Buffer

- SSC (20x) 12.5 ml
- Denhardt's solution (20x)* 12.5 ml
- SDS 10% 0.5 ml
- Distilled water 24.5 ml

* See Nick Translation/Hybridisation (above, section viii).

Analysis of Polypeptides

Minicells M9 Incubation Solution

- M9 salts (4x) 2.5 ml
- Glucose (20%) 250 μl
- Vitamin B1 (1 mg ml⁻¹) 10 μl
- Magnesium sulphate (1M) 10 μl
- Threonine (1 mg ml⁻¹) 100 μl
- Leucine (1 mg ml⁻¹) 100 μl
- Cycloserine (1 mg ml⁻¹) 200 μl
- Distilled water 6.83 ml

Solutions of the components were prepared in distilled water. M9 incubation solution was prepared immediately prior to use.

Solubilising Scintillation Fluid

- Soluene-350 50 ml
- PPO 1.8 g
- di-Methyl POPOP 50 ml
- SDS (10%) 0.45 ml
- Toluene 450 ml
Sample Buffer

- Tris-HCl (0.5M) pH 6.8 1 ml
- SDS (10%) 1 ml
- Glycerol 1 ml
- Bromophenol Blue 4 mg
- Distilled water 7 ml

Protein Size Markers

(Electrophoresis calibration kit, supplied by Pharmacia Fine Chemicals Ltd.). The kit contained:

- Phosphorylase B 94,000*
- Bovine Serum Albumin 67,000
- Ovalbumin 43,000
- Carbonic Anhydrase 30,000
- Soybean Trypsin Inhibitor 20,000
- α-Lactalbumin 14,400

* Molecular weight

3 μl were loaded for each marker track, after boiling for 3 minutes in the presence of 5 μl TE and 5 μl sample buffer.

Stock Solutions for SDS-Polyacrylamide Gels

(i) 30% Acrylamide/0.8% bis-Acrylamide
(ii) Tris-HCl (1.5M) pH 8.8
(iii) Tris-HCl (0.5M) pH 6.8

all in distilled water and stored at 4°C
**Electrophoresis Buffer**

- Tris Base: 3 g
- Glycine: 14.4 g
- SDS (10%) : 10 ml
- Distilled water: to 1 litre

**Coomassie Blue Stain**

- Coomassie Brilliant Blue G250 : 1 g
- Acetic acid (glacial): 100 ml
- Methanol: 500 ml
- Distilled water: 400 ml

**Destain/Fix**

- Acetic acid (glacial): 100 ml
- Methanol: 100 ml
- Distilled water: 800 ml

**DNA Sequencing**

**(xi)**

**LTB (Long-term Buffer for M13 phage)**

- Tris-HCl (1M) pH 7.5: 50 ml
- Magnesium chloride (1M): 10 ml
- DTT (1M): 10 ml
- Distilled water: 30 ml

LTB stored at 4°C.

**TM Buffer**

- Tris-HCl (1M) pH 8.5: 10 ml
- Magnesium chloride (1M): 5 ml
- Distilled water: 85 ml
Polyethylene glycol/Sodium chloride (PEG/NaCl)

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyethylene glycol 6000</td>
<td>20 g</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>14.6 g</td>
</tr>
<tr>
<td>Distilled water</td>
<td>to 100 ml</td>
</tr>
</tbody>
</table>

40% Acrylamide

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrylamide</td>
<td>38 g</td>
</tr>
<tr>
<td>bis-Acrylamide</td>
<td>2 g</td>
</tr>
<tr>
<td>Distilled water</td>
<td>to 100 ml</td>
</tr>
</tbody>
</table>

Stirred with 20 g of Amberlite MB-1 resin and filtered. Stored at 4°C, protected from light.

0.5 x TBE Gel Mix

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrylamide solution (40%)</td>
<td>15 ml</td>
</tr>
<tr>
<td>TBE (10x)</td>
<td>5 ml</td>
</tr>
<tr>
<td>Urea</td>
<td>46 g</td>
</tr>
<tr>
<td>Distilled water</td>
<td>to 100 ml</td>
</tr>
</tbody>
</table>

5 x TBE Gel Mix

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrylamide solution (40%)</td>
<td>15 ml</td>
</tr>
<tr>
<td>TBE (10x)</td>
<td>50 ml</td>
</tr>
<tr>
<td>Urea</td>
<td>46 g</td>
</tr>
<tr>
<td>Sucrose</td>
<td>5 g</td>
</tr>
<tr>
<td>Bromophenol Blue</td>
<td>5 mg</td>
</tr>
<tr>
<td>Distilled water</td>
<td>to 100 ml</td>
</tr>
</tbody>
</table>

Both gel mixes were prepared weekly and stored at 4°C.
Ammonium persulphate
A 25% solution in distilled water was made monthly and stored at 4°C, protected from light.

Formamide Dyes
- De-ionised Formamide* 10 ml
- Xylene cyanol FF 10 mg
- Bromophenol Blue 10 mg
- EDTA (0.5M) 0.2 ml

* Stirred with 0.2 g of Amberlite MB-1 resin and filtered.

Deoxynucleotide/Di-deoxynucleotide tri-phosphates
(dNTP/ddNTP)

A = Adenosine
T = Thymidine
G = Guanosine
C = Cytidine

dNTP - stored as 50mM solutions at -20°C
ddNTP - stored as 10mM solutions at -20°C
### ddNTP Chain Termination Mixes

<table>
<thead>
<tr>
<th>Chain terminated at *</th>
<th>T°</th>
<th>C°</th>
<th>G°</th>
<th>A°</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM dTTP</td>
<td>2.2</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>50mM dCTP</td>
<td>2.5**</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>50mM dGTP***</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>10mM ddTTP</td>
<td>22.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10mM ddCTP</td>
<td>7.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10mM ddGTP</td>
<td>15.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1mM ddATP+</td>
<td>7.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5mM dTTP+</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5mM dCTP</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5mM dGTP***</td>
<td>12.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Distilled water</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
</tbody>
</table>

* ° denotes respective chain termination mix; e.g. T° mix terminates the chain at a T

** All amounts in μl

*** In the case of dITP reactions (see Methods, Section d, iv) inosine was substituted for guanosine in dGTP.

+ A low concentration of ddATP was used. No dATP was added as the radioactive label used was α-\(^{35}\)SdATP.

‡ A small amount of the relevant dNTP was added to each chain termination mix to ensure that the chain was not terminated too early.
Mixes were prepared from above stock solutions and stored in 50 µl aliquots at -20°C. Amounts are approximate as each new batch of ddNTP's had to be checked and the amounts adjusted to give a satisfactory range in the sizes of terminated chains.

**dNTP Chase**

<table>
<thead>
<tr>
<th>dNTP</th>
<th>Concentration</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>dTTP</td>
<td>50 mM</td>
<td>5 µl</td>
</tr>
<tr>
<td>dCTP</td>
<td>50 mM</td>
<td>5 µl</td>
</tr>
<tr>
<td>dGTP</td>
<td>50 mM</td>
<td>5 µl</td>
</tr>
<tr>
<td>dATP</td>
<td>50 mM</td>
<td>5 µl</td>
</tr>
<tr>
<td>Distilled water</td>
<td>980 µl</td>
<td></td>
</tr>
</tbody>
</table>

Stored in 100 µl aliquots at -20°C.
2.2 METHODS

N.B. All buffers, solutions and bacteriological media mentioned in Methods, are described in Materials.

Centrifugations, unless otherwise stated, were carried out in a Sorvall centrifuge. Microfuge centrifugations were all carried out at 12,000 rpm (13,000 g).

2.2.a PHAGE AND BACTERIOLOGICAL TECHNIQUES

(i) Long Term Storage of Bacterial Cells:

A single colony of the appropriate strain was picked into 5 ml of L-broth and shaken overnight at the required temperature. 0.8 ml of the overnight culture was added to 0.15 ml of sterile glycerol in a screw top polypropylene tube and stored at -20°C.

(ii) Preparation of Plating Cells:

An overnight bacterial culture was diluted 50-fold and grown to mid-logarithmic phase (O.D. 650 0.5, i.e. 2-3 x 10^8 cells ml^-1). The cells were harvested by centrifugation and resuspended in half the volume of 10 mM magnesium sulphate (MgSO_4) and stored at 4°C. Fresh plating cells were prepared weekly.

(iii) Preparation of Phage Stocks:

A single plaque was picked into 1 ml of phage buffer, 2-3 drops of chloroform were added and the mixture briefly "whirlimixed". 0.1-0.2 ml of this suspension
was adsorbed to 0.2 ml of a fresh overnight bacterial culture containing 10 mM MgSO$_4$ for 15 minutes at room temperature. 3 ml of BBL top layer were added and the mixture poured on to a fresh L-agar plate. The plate was incubated at 37°C until confluent lysis was observed, usually after 6-8 hours. 3-4 ml of L-broth were added and, after storing the plates overnight at 4°C, the L-broth was decanted. Chloroform was added and mixed thoroughly. Alternatively, after adding L-broth and leaving at room temperature for 2-3 hours, the whole top layer was harvested and chloroform added as above. In both cases, agar, chloroform and cell debris were removed by centrifugation (in a bench centrifuge) at 4000 rpm for 10 minutes. The supernatant was decanted, titred, and stored at 4°C. Titres varied from $10^9$ to $10^{11}$ p.f.u.·ml$^{-1}$, depending on the phage and the bacterial host.

(iv) Phage Titrations:

Phage stocks were serially diluted in phage buffer. 0.1 ml of the appropriate dilutions were adsorbed to 0.2 ml of plating cells for 15 minutes at room temperature; in each case, 2.5 ml of molten BBL top layer, containing 1 mM MgSO$_4$, were added and the mixture was poured on to a BBL agar plate. Plates were then incubated overnight at 37°C. For spot tests, a lawn of cells was prepared by adding 0.2 ml of plating cells to 2.5 ml of molten BBL top layer and pouring on to a "dry" BBL agar plate. Drops of the appropriate

* p.f.u. plaque-forming units
phage dilutions were spotted using a 0.1 ml pipette. Alternatively 10 µl were spotted using a micropipette.

(v) **Preparation of Phage Liquid Lysates:**

(A 100 ml culture was found to be a suitable volume, yielding about $10^{12}$ p.f.u.)

A fresh overnight culture of the host bacterium was diluted 50-100 fold in to 100 ml of L-broth containing 0.5 ml of 1 M MgSO$_4$ and grown at 37°C, with good aeration, to O.D.$_{650}$ of 0.5 (i.e. $2-3 \times 10^8$ cells ml$^{-1}$). Phage, from a phage stock as described above, was added to a multiplicity of infection (m.o.i.) of 0.2 for imm$^\lambda$ phages and 1 for imm$^{21}$ phages. The turbidity of the culture was followed at 650 nm until lysis occurred, usually 2-4 hours later. 0.2 ml of chloroform was added and, after 15 minutes, the lysate was clarified by centrifugation at 4000 rpm for 10 minutes in Nalgene polypropylene bottles, using a Sorvall centrifuge and GSA rotor. The supernatant was then titred.

**Polyethylene Glycol Precipitation:**

The lysate was concentrated by Polyethylene glycol (PEG) 6000 essentially as described by Yamamoto et al. (1970). 4 g Sodium chloride, 10 µl DNase I (10 mg ml$^{-1}$), and 10 µl RNase A (10 mg ml$^{-1}$) were added to the clarified lysate which was then left at room temperature for 1 hour. 10 g of PEG 6000 were added, allowed to dissolve and the lysate was then left at 4°C overnight. The phage were pelleted by centrifugation at 10,000 rpm for 10 minutes
using a Sorvall GSA rotor and resuspended gently (over 2-3 hours on ice) in 5 ml of phage buffer. Remaining debris was removed by centrifugation at 4000 rpm for 5 minutes in a Sorvall SS34 rotor. Yields of phage were typically 60-100%. The phage were then further purified by a caesium chloride (CsCl) step gradient.

CsCl Step Gradients:

CsCl step gradients (Thomas and Abelson, 1966) were prepared in 14 ml MSE polycarbonate centrifuge tubes using three densities of CsCl, namely 31% w/w, 45% w/w, 56% w/w in phage buffer. The CsCl solutions were clarified by centrifugation at 10,000 rpm for 20 minutes and 1.5 ml "steps" of each were used for the gradients. These were prepared by putting the lightest CsCl solution in the centrifuge tube first and underlaying the other solutions in order of density using a long pasteur pipette. The phage suspension was then carefully overlaid using a 10 ml pipette. Centrifugation was carried out at 33,000 rpm and 15°C for 2 hours using an MSE 6 x 14 ml Titanium rotor in an MSE preparative centrifuge. This typically gave three bands, the middle one containing the phage. The band was collected by puncturing the side of the tube using a 21 gauge needle. For viable phage the CsCl was removed by dialysis against phage buffer.

(vi) Selection of λ Deletion Derivatives using EDTA:

This method of selecting λ deletion derivatives is based on that of Parkinson and Huskey, 1971.
BBL agar plates containing a range of EDTA concentrations of 2-10 mM were prepared by adding sterile 0.5M EDTA pH 8 to molten BBL agar before pouring. Plates were used within 2-3 days as evaporation would alter the EDTA concentration. The range of EDTA concentrations used depended on the phage and on the size of the desired deletions.

The plates were first checked by spot tests as described in phage titrations, but without MgSO₄ in the plating cells or the top layer. Standard phages including those with full-size genomes and phages with known deletions were used as controls. The phage from which deletions were to be isolated was also included. These spot tests not only checked whether the EDTA selection was working but also estimated the range of EDTA concentrations within which deletions were easily isolated from the relevant phage. Once this was confirmed, phages were plated as for titrations to obtain single plaques. Individual plaques were picked into 1 ml of phage buffer containing 2 drops of chloroform and purified on BBL/EDTA plates. Individual plaques were again picked and phage stocks prepared as described earlier. The location and extent of the deletions were estimated by comparison of endonuclease restriction patterns of the original phage and deletion derivatives after DNA was prepared from a liquid lysate.
(vii) **Phage Crosses:**

Freshly prepared plating cells were co-infected at a m.o.i. of 5 of each of the two parent phages. After 15 minutes adsorption at room temperature the infected cells were diluted 100-fold into pre-warmed L-broth and grown at 37°C, with aeration, for 1½ hours. Chloroform was added and, after centrifugation to remove cell debris and chloroform, the supernatant was titred on a permissive host for total progeny and on a selective host for the required recombinant. Recombinant single plaques were picked and tested as appropriate. After purification by single plaque isolation, phage stocks were prepared as described earlier.

(viii) **Phage Complementation Tests:**

Freshly prepared plating cells were co-infected at a m.o.i. of 2 of each of the two phages in the complementation tests, in a microfuge tube. After 15 minutes adsorption at room temperature, the infected cells were harvested by centrifugation at 6000 rpm for 5 minutes. The supernatant, containing unadsorbed phage, was discarded and the cells were resuspended in 0.4 ml of $10^{-2}$ M MgSO$_4$. They were subsequently diluted 100-fold into pre-warmed L-broth and shaken at 37°C for 45 minutes (1 cycle). Chloroform was added and, after centrifugation to remove cell debris and chloroform, the phages were titred on suitable hosts to estimate the degree of complementation.

**N.B.** When phage/plasmid complementation tests were
carried out, a lysate (section iii, above) of the phage to be tested was prepared using cells carrying the relevant plasmid and the phage were subsequently tested for complementation by titration on suitable hosts. Multiple cycles of infection were found to be required for phage/plasmid complementation tests, hence the preparation of a lysate rather than the single cycle infections described above for phage/phage complementation tests.

(ix) **Construction of Lysogens/Di-lysogens:**

Fresh plating cells of the bacteria to be lysogenised were prepared. These were infected with the appropriate phage (or phage and helper phage) at a m.o.i. of 5 and allowed to adsorb for 15 minutes at room temperature. They were then diluted 50-fold into L-broth and grown for 3-4 hours at the appropriate temperature. The resultant culture was serially diluted in L-broth and plated on L-agar plates which had been spread with $10^9$ p.f.u. each of two homimmune phages of different host-ranges and carrying a clear mutation in the repressor gene - "killer phages". (Under these conditions only lysogens or cells resistant to both phages should survive and typically 70-80% of surviving colonies were lysogens.) Survivors were picked and tested for lysogeny. These were purified by single colony isolation without killer phage selection and frozen stocks were prepared.
2.2.b DNA TECHNIQUES

(i) Phenol Extraction:

DNA solutions were normally phenol extracted by mixing with equal quantities of TE-equilibrated phenol. In the case of bacterial DNA solutions, the phenol was equilibrated with Tris-HCl, EDTA, sodium chloride (NaCl) and sodium dodecyl sulphate (SDS) (see Materials, section v). The phenol and aqueous phases were separated by brief centrifugation.

(ii) Butanol Extraction:

Ethidium bromide was removed from DNA solutions by extraction with TE-saturated butan-1-ol until the DNA solution appeared colourless. Generally 2 or 3 extractions with an equal volume of butanol were sufficient. In each case the butanol and aqueous layers were separated by brief centrifugation.

(iii) Ethanol Precipitation:

3M sodium acetate-(NaAc) was added to DNA solutions to give a final concentration of 0.3M NaAc. After mixing, 3 times the volume of ethanol was added. (Plastic microfuge tubes or siliconised glass "corex" tubes were used to prevent the DNA from sticking to the tubes.) The tubes were placed at -70°C for 2 hours and then the DNA was pelleted by centrifugation at 10,000 rpm for 30 minutes using a Sorvall HB4 rotor in the case of corex tubes, or for 15 minutes in a microfuge in the case of microfuge tubes. The supernatant was poured off and the pellet was washed with 70% ethanol and dried.
under vacumm. It was then dissolved in an appropriate volume of TE.

(iv) Extraction of DNA from Phage:
Phage liquid lysates were prepared as described earlier and, after the phage band was collected, the CsCl was dialysed out against TE. (This causes some of the phage heads to rupture, releasing the DNA.) The DNA solution was phenol extracted 3 times and ether extracted with an equal volume of Diethyl ether, 1-3 times until it became clear. It was then dialysed for 24 hours against 3 changes of TE. If the phage was to be used as a cloning vector it was purified through a second gradient, in this case an equilibrium gradient. The phage band from the step gradient was mixed with a 41.5% w/w preclarified CsCl solution (see section (a), v, above) in a 5 ml MSE polyallomer tube and concentrated by centrifugation at 33,000 rpm and 15°C for 24 hours in a 6 x 5 ml Titanium rotor in an MSE preparative centrifuge. The phage band was collected by puncturing the side of the tube and treated as described above for phage from step gradients. (100-200 μg of DNA were obtained from a 100 ml culture.)

(v) Extraction of Chromosomal DNA from Bacteria:
The procedure used for the extraction of DNA from bacteria was modified from that described by Marmur, 1961. Generally, a 250 ml bacterial culture was used, yielding 1-1.5 mg of DNA.
A single colony was picked into 5 ml of L-broth and shaken overnight at 37°C. The overnight culture was used to inoculate 250 ml of L-broth in a 1 litre conical flask and this was grown to OD$_{650}$ 0.7-1.0 at 37°C, with aeration. The cells were harvested by centrifugation at 5,000 rpm for 10 minutes in a Sorvall GSA rotor, resuspended in 20 ml of TE and harvested as before. The pellet was resuspended in 6 ml of Tris-HCl/EDTA/sucrose, transferred to 50 ml Nalgene polypropylene tubes and 0.85 ml of lysozyme (10 mg ml$^{-1}$) were added. The tubes were shaken gently in a 37°C waterbath for 10 minutes and then placed on ice for 5 minutes. 3.25 ml of 0.25M EDTA pH 8.0 were added and, after 5 minutes on ice, 6.75 ml of Triton lysis buffer were added, followed by 1 ml of 10% SDS, 0.1 ml of 10 mg ml$^{-1}$ proteinase K, and 0.2 ml of 10 mg ml$^{-1}$ RNase A. The solution was shaken gently overnight at room temperature and extracted with phenol equilibrated with Tris-HCl/EDTA/NaCl/SDS. For each extraction, after the phenol was added, the tubes were shaken gently for 30 minutes and the phenol and aqueous layers were then separated by centrifugation for 5 minutes at 5,000 rpm in a Sorvall SS34 rotor. The DNA was ether extracted 1-3 times until the aqueous layer cleared and finally dialysed overnight against 3 changes of 50 mM Tris-HCl pH 8.0/10 mM EDTA pH 8.0/10 mM NaCl.

(vi) Extraction of Plasmid DNA:

Plasmid DNA was extracted from bacterial
strains carrying plasmids using a procedure based on the methods described by Clewell and Helinski, (1969, 1972). A 500 ml culture was prepared in each case yielding 0.3-0.5 mg of DNA.

A single colony was picked into 10 ml of L-broth containing 100 μg ml\(^{-1}\) ampicillin and shaken overnight at 37°C. The overnight culture was diluted (50-fold) into 500 ml of L-broth in a 2-litre conical flask containing 0.5% glucose and 100 μg ml\(^{-1}\) ampicillin and grown to OD\(_{650}\) 0.5 at 37°C, with aeration. 75 mg of chloroamphenicol were added, (this prevents protein synthesis but not plasmid replication) and the culture was shaken overnight at 37°C. The cells were harvested by centrifugation at 5,000 rpm for 10 minutes using a Sorvall GSA rotor, resuspended in 20 ml of TE and again harvested. The pellet was resuspended in 6 ml of Tris-HCl/EDTA/sucrose and transferred to a 100 ml conical flask. 1 ml of lysozyme (10 mg ml\(^{-1}\)) was added, followed by 0.5 ml of 0.5M EDTA. The flask was shaken gently, and warmed if necessary, until the cells lysed. 12 ml of cold Triton lysis buffer were added and the flask was left on ice for 10 minutes. The mixture was transferred to a Nalgene 50 ml polypropylene tube and chromosomal DNA was sedimented by centrifugation at 15,000 rpm for 40 minutes, using a Sorvall SS34 rotor. The supernatant was carefully decanted into a clean tube and extracted with 15 ml of TE-equilibrated phenol. The phenol and aqueous phases were separated by centrifugation at
8,000 rpm for 10 minutes in a Sorvall SS34 rotor. The aqueous layer was taken off by pipette and equal volumes were placed in two siliconised glass (corex) 30 ml tubes. The DNA was ethanol precipitated, as described in section b (iii), above, but the pellets were not dried. The pellets from both tubes were each dissolved in 5 ml of TE; the solutions were pooled and incubated at 37°C for 30 minutes with 10 μl of RNase A (10 mg ml⁻¹). 0.5 ml of ethidium bromide (10 mg ml⁻¹) were added and CsCl was dissolved in the solution to give a final density of 1.55 mg ml⁻¹. The solution was transferred to a Beckman 14 ml quick-seal polyallomer tube; centrifugation was carried out at 38,000 rpm for 48-60 hours at 18°C using a Sorvall Ultracentrifuge and 50 Ti rotor. This gave two well separated bands as seen under u.v. illumination, the lower of which contained the supercoiled plasmid DNA. The tube was punctured at the top and the plasmid band was removed by a further puncture at the side of the tube. Extraction with TE-saturated butan-1-ol removed ethidium bromide and the CsCl was dialysed out against TE. The DNA was finally ethanol precipitated and resuspended in 200-300 μl of TE.

(vii) Preparation of M13 Replicative Form (RF) DNA:

M13 RF DNA was prepared using, essentially, the procedure for preparation of plasmid DNA.

Single plaques of the relevant M13 vector (or M13 clone) were obtained on NM522 as described in section a (iv). A single plaque was picked into 1 ml of L-broth and
shaken at 37°C for 3 hours. At the same time, a colony of NM522 was inoculated into 20 ml of L-broth and grown in the same way. (Colonies of NM522 were kept on minimal agar as explained below, section b, (xiii). 0.5 ml of the M13 plaque suspension and 5 ml of the NM522 culture were subsequently used to inoculate 500 ml of L-broth in a 2-litre conical flask which was shaken overnight at 37°C. The cells were harvested and RF DNA was prepared as described above for plasmid DNA extraction (section b, (vi)).

(viii) **Plasmid "Mini-Prep":**

This quick method for preparing plasmid DNA based on that described by Ish-Horowitz and Burke (1981), is particularly useful for obtaining DNA suitable for endonuclease restriction. This allows analysis of several plasmid clones without the need for large scale preparation.

A single colony was picked into 2 ml of L-broth with the appropriate antibiotic and shaken overnight at 37°C. 1 ml of this overnight culture was harvested in a microfuge tube by centrifugation for 5 minutes. The pellet was resuspended in 100 μl of Solution I and incubated at room temperature for 5 minutes. 200 μl of fresh Solution II were added, mixed gently and the tube was placed on ice for 5 minutes. 150 μl of pre-cooled Solution III were added and, after 5 minutes on ice, the precipitated protein, SDS (from Solution II), and chromosomal DNA were removed by centrifugation for
3 minutes. 1 ml of ethanol was added and, after incubation at room temperature for 5 minutes, the plasmid DNA was precipitated by centrifugation for 5 minutes. The pellet was washed with 70% ethanol, dried in a desiccator and dissolved in 40 \( \mu l \) of TE. 10 \( \mu l \) aliquots were used for endonuclease restriction. If small fragments (i.e. smaller than 600 bp) were expected boiled RNase A was added to the restriction buffer to a final concentration of 2 \( \mu g \text{ ml}^{-1} \).

(ix) **DNA Restriction:**

Restriction endonuclease digestion of DNA was generally carried out in a total volume of 20 \( \mu l \). Large scale digests were performed in 50 or 100 \( \mu l \) depending on the amount of DNA used and the constituent reagents were scaled up accordingly. A typical reaction mixture contained:

- DNA \( 1-3 \mu g \) (as required)
- Restriction buffer (10x)* \( 2 \mu l \)
- DTT (100mM) \( 1 \mu l \)
- Restriction endonuclease 1 unit per \( \mu g \) of DNA*
- Distilled water to 20 \( \mu l \)

*The type of restriction buffer used depended on the particular enzyme requirements (see Maniatis et al., 1982)

+ When *E.coli* DNA was restricted 2 units of enzyme were used per \( \mu g \) of DNA.
Reactions were carried out at 37°C for 2 hours, to ensure complete digestion, and subsequently stopped by heating at 70°C for 10 minutes, followed by rapid cooling on ice. In the case of restriction enzymes which were not heat-labile, EDTA was added to a final concentration of 10 mM to stop the reaction. When λ DNA was restricted for electrophoresis, the digest was always heat treated even if EDTA was also necessary. The heat treatment dissociated the λ cohesive ends which facilitated the interpretation of the restriction patterns. The restricted DNA was then either loaded directly on an agarose gel for electrophoresis or stored at 4°C (short term) or -20°C (long term) until required.

(x) Agarose Gel Electrophoresis:
Electrophoresis was carried out in 0.7% w/v agarose gels, in 1 x TBE and at 1 volt cm⁻¹ for 16 hours. Before the DNA was loaded on to the gel 5 μl of DNA sample buffer were added to each digest. The DNA bands were visualised by staining in distilled water containing 1 mg litre⁻¹ of ethidium bromide for 20 minutes followed by de-staining for 20 minutes in distilled water. Gels were photographed under uv light using an X4 red filter and Ilford FP4 film.

λcI857 DNA digested with EcoRI, HindIII, or a mixture of both enzymes provided size markers for gels. The sizes of the fragments obtained from λcI857 under these conditions have been estimated by Philippsen et al. (1978).
DNA Ligation:

Ligation conditions were essentially as described by Wilson et al. (1977). DNA was digested with restriction endonucleases as described above (section ix). After restriction, in the case of heat-labile enzymes, EDTA was added to give a final concentration of 10 mM and the digest was heated at 70°C for 10 minutes. For enzymes which were not heat-labile, EDTA was added as above, followed by 0.4 volumes of 5M ammonium acetate (NH₄Ac) and 2 volumes of isopropanol. After 10 minutes at room temperature the DNA was recovered by centrifugation at 12,000 rpm for 10 minutes and the pellet was resuspended in TE to give a DNA concentration of 0.05-0.1 μg μl⁻¹.

The ligation reaction mixture was prepared as follows:

10 mM ATP 1 μl
100 mM DTT 1 μl see media
Ligase buffer (10x) 1 μl
Vector DNA 0.05-0.1 μg*
Donor DNA *
Ligase 1-2 units*
Distilled water to 10 μl

*The amounts of vector and donor DNA used depended on the number of "ends" available and therefore on the size and number of fragments produced by restriction of the donor DNA. In cases where a high background of religated vector was likely an excess
of donor fragments (ends) was used. When phage libraries were being constructed a more suitable ratio of vector to donor ends was 1:1 (see Maniatis et al., 1982). When E.coli libraries were made in phage λ, empirical tests were carried out to optimise ratios. This was also the case in the cloning of bacterial DNA in M13 for sequencing.

Ligase was diluted in ligase dilution buffer to the required concentration. For blunt end ligation neat ligase was used (50-100 units).

The ligation reaction mixture was incubated at 15°C for 16 hours and stored at 4°C until required. In the case of M13 clones, long term storage was at -20°C.

(xii) Transfection/Transformation:

For routine transfection and transformation, cells were made competent for the uptake of DNA using a modification of the calcium chloride procedure of Mandel and Higa (1970).

A single colony of the host strain (usually 5K or NM477 for transfection and HB101 for transformation) was picked into 5 ml of L-broth and shaken overnight at 37°C. The overnight culture was diluted 50-fold into L-broth and grown at 37°C, with aeration, to mid-logarithmic phase, i.e. OD₆₅₀ of 0.5 (2-3 x 10⁸ cells ml⁻¹). The cells were harvested by centrifugation at 4,000 rpm, in the cold, for 10 minutes using a bench centrifuge and resuspended in half the original volume of cold 0.1M
CaCl$_2$ (calcium chloride). The cells were again harvested, resuspended in 1/20 of the original volume of cold 0.1M CaCl$_2$ and kept on ice for at least 30 minutes. They remained competent for several hours. Transfection and transformation were carried out in 5 ml glass dilution tubes.

For transfection, a range of DNA amounts from 10 ng to 50 ng of the ligated DNA was used. 100 μl of cold SSC/CaCl$_2$ were added to each DNA sample, followed by 100 μl of competent cells. Two control "transfections" were carried out, one using unrestricted vector phage to determine the competence of the host and the other containing only cells and SSC/CaCl$_2$ to check for contaminating phage. After 10 minutes on ice, the samples were "heat shocked" at 37°C for 5 minutes. Subsequently, 2.5 ml of BBL top layer agar, supplemented with 10$^{-2}$M MgSO$_4$, were added and the mixture was poured on to BBL agar plates. The plates were incubated overnight at 37°C.

In the case of transformation, approximately 50 ng of the ligated DNA were used as one sample. Controls included a transformation with 50 ng vector plasmid DNA, to determine the competence and cells and SSC/CaCl$_2$, to check for background of antibiotic resistant colonies. The transformation procedure was as described above for transfection but, after the heat shock, 1 ml of L-broth was added to each sample and the tubes were incubated at 37°C for 1 hour to allow expression of antibiotic resistance
genes by the plasmid (e.g. β-lactamase by pBR322).
5 µl, 10 µl and 50 µl aliquots were plated on L-agar plates containing 100 µg ml⁻¹ of ampicillin. (For the cells/SSC/CaCl₂ sample only a 50 µl aliquot was plated.) The plates were incubated at 37°C for 16-24 hours.

For both transfection and transformation a competence of 10⁵ - 10⁶ pfu/transformants per µg of DNA was generally observed.

(xiii) Transfection using "Hanahan" Competent Cells:
This procedure for making competent cells for transfection is based on the Hanahan transformation method (Hanahan, 1983). It was found to be particularly useful for transfecting NM522 with M13 phage, giving a competence of about 10⁷ pfu per µg of M13 DNA.

Transfection of NM522, with M13 clones, for sequencing was carried out using this method.

A single colony on NM522, grown on a minimal agar plate* was picked into 5 ml of L-broth and shaken overnight at 37°C. The overnight culture was diluted 500-fold in L-broth, grown at 37°C with aeration, to logarithmic phase (OD₆₅₀ 0.6-0.7) and then placed on ice for 30 minutes. The cells were harvested by

* The NM255 chromosome is pro⁻ (pro Δ), but this strain is F' and has a pro (proline synthesising) gene on the F' factor. Therefore strains only appear as pro⁺ if they are F'. Thus plating on minimal agar plates selects for F' cells.
centrifugation at 4,000 rpm for 10 minutes in the cold, using glass universal bottles and a bench top centrifuge, and resuspended in half the volume of TFB. After 20 minutes on ice they were harvested as before, resuspended in 1/12.5 of the original volume of TFB and again kept on ice for 20 minutes. (A 25 ml culture gave 2 ml of transfection cells.) For every 2 ml of cells, 70 µl of DMSO from a frozen stock were added, followed, after 5 minutes on ice, by 70 µl of 2.25M DTT in 40mM potassium acetate. The bottles were kept on ice for 10 minutes and, following the addition of 70 µl of DMSO, were left on ice for a further 5 minutes. 200 µl aliquots of the "competent" cells were placed in chilled 5 ml glass dilution tubes, DNA (usually less than 50 ng) was added and the tubes were left on ice for 30 minutes. They were heat shocked at 42°C for 90 seconds and again chilled on ice for 2 minutes. Control samples included transfection with intact M13 to check competence and a tube containing treated cells only to check for contaminating phage. This method was generally used to transfect cells with an M13 "ligation mix" in which DNA had been inserted in a polylinker in the M13 β-galactosidase gene** (Messing et al., 1977, 1981). Therefore 2.5 ml of molten BBL top layer supplemented with the indicator Xgal and the inducer IPTG were added to the cells and the mixture was subsequently poured on to minimal agar plates, where it is easier to distinguish white plaques from blue plaques. Plates were incubated overnight at 37°C.
** The β-galactosidase gene in NM522 is deleted and the F' factor carried by this strain has a deletion (lac ZΔM15) in the operator proximal region of its β-galactosidase gene. On infection with M13 phage, β-galactosidase production (resulting in blue plaques in the presence of Xgal and IPTG) depends on α ω intracistronic complementation, the α peptide being provided by the M13 phage. Thus parent M13 phages give blue plaques while those containing inserts in the polylinker give white plaques, due to insertional inactivation of the M13 β-galactosidase gene.

(xiv) **In vitro Packaging:**

In vitro packaging mixes, namely Freeze Thaw Lysate (FTL) and Sonicated Extract (SE) were kindly donated by Annette J. B. Campbell.

The packaging reaction mixture was prepared by adding the reagents in the following order:

Buffer A 7 μl
DNA 1-2 μl*
Buffer M1 1 μl
SE 6 μl
FTL 10 μl

*The DNA volume and concentration were critical for packaging. If the volume exceeded 2 μl the packaging efficiency was considerably lower. In addition, a minimum concentration of DNA was necessary; a lowered efficiency was observed when the amount of DNA added was less than 200 ng.
The reaction mixture was incubated at 25°C for 1 hour and subsequently diluted with 0.5 ml of phage buffer. A control reaction mixture in which no DNA was present was prepared to check for endogenous DNA or contaminating phage. Test aliquots of 10 and 50 µl were added to 0.2 ml of the relevant plating cells and 2.5 ml of molten BBL top layer agar, and were poured on to BBL agar plates. Plates were incubated overnight at 37°C. The volume required to give 200-300 plaques per plate was thus calculated and the remaining phage were plated to give this concentration of plaques. Generally, resulting plaques were probed for the desired recombinants and for this 200-300 plaques per plate were found to be a suitable number.

(xv) Elution of DNA from Agarose Gels:

The following method, based on one described by Maniatis et al., (1982) was found to be most efficient in the recovery of DNA from agarose gels, especially when the DNA was subsequently used for cloning. When DNA was to be eluted from agarose gels, Seakem LE agarose (0.7% w/v in 1 x TBE) was used for electrophoresis to reduce the impurities which, if carried over in the eluted DNA, might impair subsequent restriction or ligation of the DNA. In this case, electrophoresis was carried out at 1 volt per cm, in submerged gels (using 1 x TBE) and both gel and electrophoresis buffer contained ethidium bromide at a concentration of 1 mg litre⁻¹. This made the bands visible under U.V. light.
[A long-wave-length lamp (300-360 nm) was used to minimise damage to the DNA.] When the fragment which was to be eluted was well separated from other fragments present, the current was switched off and buffer removed until the surface of the buffer was just below the gel surface. A trough was cut in the gel directly in front of the leading edge of the DNA band, 5 mm wider than the band on either side. At the side of the trough away from the band, a slit was cut 5 mm into the gel on either side of the trough. A piece of dialysis tubing was inserted in the slits and pushed underneath the gel to form a base for the trough and a "wall" on the side furthest from the band (see figure 2.1).

![Figure 2.1](image)

The trough was filled with TBE buffer from the tank and the current was switched on again and the voltage was raised to 150 volts. At this stage it was possible to see the band eluting out of the gel under uv light. When all the DNA had been eluted (after approximately 10
minutes) the current was reversed briefly (for 1 minute) to release any DNA which had attached to the dialysis tubing. After switching off the current the buffer was collected from the trough by micropipette and transferred to a microfuge tube. Any DNA remaining in the trough was recovered using a further 300 μl of buffer. Ethidium bromide was removed by extraction with TE-saturated butan-1-ol and the DNA was subsequently phenol extracted and ethanol precipitated. It was finally dissolved in TE; the volume of TE used depended on the amount of DNA eluted and the final concentration was usually in the range 100-300 μg ml⁻¹.

(xvi) Southern Transfer:
This method (Smith and Summers, 1980) of transferring DNA from an agarose gel to a nitrocellulose filter is a modification of the original procedure (Southern, 1975).

After electrophoresis, the agarose gel was stained with ethidium bromide, photographed as described above (section (x)) and transferred to a plastic tray after the unused borders had been trimmed away. It was then soaked in 0.25M HCl for 15 minutes. (This partially hydrolysed the DNA by depurination which facilitated the transfer of large DNA fragments.) After rinsing the gel in distilled water for 10 minutes, the DNA was denatured in NaOH/NaCl denaturation buffer for 30 minutes. The gel was again rinsed and subsequently soaked in 1M NH₄Ac (pH 8.0) for 1 hour. For transfer, it was placed on a
glass plate and covered with a nitrocellulose filter, of the same dimensions, which had previously been soaked in NH$_4$Ac. Any air bubbles trapped between the gel and the filter were removed. Three sheets of blotting paper, cut to the same size and also soaked in NH$_4$Ac, were placed on top of the filter and, again, air bubbles were removed. These sheets were in turn covered by a 6 cm stack of dry blotting paper, cut just smaller than the gel and topped by a glass plate. (The weight of the glass plate maintained even contact between the gel and the filter.) Transfer was effected overnight; the filter was briefly rinsed in 2 x SSC (diluted from a 20 x SSC stock), blotted dry and finally heated at 80°C under vacuum for 2 hours. The gel was re-stained with ethidium bromide and viewed under UV light to check whether DNA transfer had been efficient.

(xvii) **Plaque Hybridisation:**

An abbreviated version of the method described by Benton and Davis (1977) was used for the transfer of DNA from plaques to nitrocellulose filters. Phage were plated in BBL top layer on dry BBL agar plates, using a suitable host, to obtain single plaques (200-300 per plate). After overnight incubation at 37°C the plates were kept at 4°C for 1 hour. DNA was transferred from the plaques by laying nitrocellulose filter discs on the surface of the agar and leaving for 1 minute. The filters were carefully peeled off the plates and placed, plaque side uppermost, on a pad of blotting paper.
saturated with NaOH/NaCl denaturation buffer for 2 minutes. (Pre-cooling the plates at 4°C allowed removal of the filters without damage to the top layer agar.) The filters were then placed in a beaker containing 200 ml of Tris-HCl/NaCl neutralisation buffer for a further 2 minutes and finally rinsed briefly in 2 x SSC, blotted dry and heated at 80°C under vacuum for 2 hours. Up to three filters could be prepared from each plate. In each case, both the plate and the filters were asymmetrically marked to facilitate alignment after hybridisation.

(xviii) **Nick Translation:**

The procedure used for nick translation was based on the methods described by Maniatis et al. (1975) and Rigby et al. (1977).

α-32P deoxycytidine triphosphate (α32P dCTP) was the labelled nucleotide triphosphate used. The Nick Translation reaction mixture was prepared as follows:

\[
\begin{align*}
\alpha-32P \text{ dCTP} & : 10 \mu Ci \\
dNTP buffer (2 \mu MdNTP)* & : 20 \mu l \\
DNA (0.2-1.0 \mu g) & : 2-4 \mu l \\
DNase (2 \times 10^{-5} \text{ mg ml}^{-1}) & : 1 \mu l \\
DNA polymerase I & : 1 \text{ unit}
\end{align*}
\]

* The dNTP buffer contained no dCTP.

The mixture was incubated at 15°C for 2 hours and the reaction was subsequently terminated by adding 100 μl of 10 mM EDTA. A sephadex G-50 column, equilibrated with
TE\(^+\) was used to remove unincorporated radioactive label. The reaction mixture was passed through the sephadex column, using TE as column buffer, and the first peak of radioactivity (localised using a mini-monitor), containing the label bound to the DNA, was collected in a 1 ml fraction. (The second peak was due to unincorporated label.) A 1 \(\mu\)l sample was spotted on a glass paper filter, dried and the activity was counted in scintillation fluid.

\(^+\)Sephadex G-50 was equilibrated as follows:

30 g of sephadex G-50 were added to 250 ml of TE in a 500 ml bottle, autoclaved at 15 lb in\(^{-2}\) for 15 minutes and allowed to cool at room temperature. The supernatant was decanted and replaced with an equal volume of TE. The sephadex was stored at 4°C. Columns were prepared in Bio-rad 8 ml borosilicate glass chromatography columns and washed several times with TE before and after use.

(xix) Hybridisation

Nitrocellulose filters prepared as described in sections (xvi) and (xvii) above, were pre-hybridised in 50-100 ml of hybridisation solution, depending on the number of filters, for 30 minutes. Hybridisation was carried out overnight at 37°C in sealed polythene bags containing 10-20 ml of hybridisation solution according to the number and size of the filters. Denatured sonicated calf thymus DNA was included for hybridisation at a concentration of 25 \(\mu\)g per ml of
hybridisation solution. The radioactive "probe" was prepared by boiling the appropriate amount of labelled DNA (see above, section (xviii)) for 5 minutes and rapidly cooling on ice. 1-3 million cpm were used for each filter containing DNA transferred from a gel or for each batch of 10 filters with DNA transferred from plaques. After hybridisation, filters were washed in 2 x SSC/0.1% SDS at 37°C and subsequently in 1 x SSC/0.1% SDS at room temperature. In each case, two 30 minute washes were carried out. Finally the filters were thoroughly rinsed in 1 x SSC, blotted dry and placed between two layers of "Saran Wrap" plastic film. For autoradiography, Cronex x-ray film and intensifying screens were used. The film was pre-flashed and exposed as required at -70°C.
2.2.c. ANALYSIS OF POLYPEPTIDES

(i) Preparation of Minicells:

This method for preparing minicells was adapted from that described by Reeve (1979). The minicell strain WL542 was used as a host for plasmids in the analysis of polypeptides. Minicells were separated from nucleated cells by centrifugation in sucrose gradients. These were prepared in 30 ml glass (Corex) centrifuge tubes by freezing 20 ml of a 22% w/v sucrose solution in 1 x M9 salts at -20°C overnight and subsequently thawing slowly (overnight) at 4°C.

For each preparation, 5 ml of L-broth containing the appropriate antibiotic (in this case ampicillin, as the plasmid vector used was pBR322) were inoculated with a colony of WL542 containing the relevant plasmid and shaken overnight at 37°C. 1 ml of the overnight culture was added to 500 ml of L-broth, containing ampicillin, in a 2 litre flask which was subsequently shaken overnight at 37°C. Nucleated cells were sedimented by centrifugation* of the overnight culture at 2,000 rpm in a GSA rotor. Minicells were then harvested by centrifugation of the supernatant at 7,000 rpm for 10 minutes in the same rotor. These were resuspended in 5 ml of 1 x M9 salts in 50 ml polypropylene tubes and spun at 2,000 rpm for 3 minutes in an SS34 rotor.

*This and subsequent centrifugations were carried out at 4°C using a Sorvall centrifuge and Sorvall rotors. Minicell suspensions, sucrose gradients and M9 salts solutions were kept cold at all times.
to remove more nucleated cells. The supernatant was carefully layered on top of a sucrose gradient (see above) and the minicells were concentrated by centrifugation at 5,000 rpm for 20 minutes in an HB4 rotor. This centrifugation caused most nucleated cells to form a pellet or band at the bottom of the tube and minicells to remain as a thick band about half way up the gradient. The top two-thirds of the minicell band was collected with a Pasteur pipette and transferred to a polypropylene tube. 20 ml of 1 x M9 salts were added to dilute the sucrose and, after gentle mixing, the minicells were harvested by centrifugation at 8,000 rpm for 10 minutes in an SS34 rotor. They were resuspended in 2 ml of 1 x M9 salts and layered on to another sucrose gradient. The sucrose gradient centrifugation was repeated 2-3 times until most nucleated cells (seen as a pellet at the bottom of the gradient) were removed. After the final centrifugation, the minicell band was collected and the OD$_{650}$ of an appropriate dilution was determined. The minicells were harvested after dilution with M9 salts as described above and finally resuspended in 1 x M9 salts containing 30% v/v glycerol to give an OD$_{650}$ of approximately 2 ($2 \times 10^{10}$ minicells ml$^{-1}$). 200 µl aliquots were placed in microfuge tubes, frozen in liquid nitrogen and stored at -70°C. (Minicells remain stable in this form for several months.)
(ii) **Labelling of Minicells:**

For each plasmid analysed, the frozen minicell suspension was thawed slowly on ice and then spun at 12,000 rpm for 5 minutes in a microfuge. The pelleted minicells were resuspended in 200 µl of M9 incubation mixture and incubated at 37°C for 1 hour. This "pre-incubation" caused effective translation of stable endogenous mRNA thus preventing it from interfering with translation of plasmid encoded mRNA after labelling. Cycloserine was used in the incubation mixture to kill viable (nucleated) cells. After incubation 20 µl of Difco methionine assay medium containing 20 µCi of $^{35}$S-methionine were added and the reaction mixture was again incubated at 37°C for 1 hour. (The Difco methionine assay medium improved the efficiency of $^{35}$S-methionine incorporation.) The minicells were harvested by centrifugation in a microfuge for 5 minutes; the supernatant was decanted and the minicells were resuspended in 20 µl of TE. 20 µl of sample buffer and 2 µl of β-mercaptoethanol were added and the samples (6-10 minicell strains were usually labelled in one experiment) were boiled for 5 minutes. This disrupted the cells and denatured the proteins. 10 µl from each sample were loaded on an SDS-polyacrylamide gel (see below, section (iv)) and the remainder was stored at -20°C. 5-10 µl was generally a suitable volume, the exact volume being determined from the relative band intensities observed in a trial gel. A control sample, in which the original WL542 host had
been transformed with the plasmid vector (in this case pBR322), was included to identify polypeptides encoded by the vector.

(iii) Phage Infection and Labelling of UV-Irradiated Cells:

The analysis of polypeptides encoded by DNA cloned in \( \lambda \) vectors was carried out by infection of UV-irradiated cells and subsequent labelling with \( ^{35} \mathrm{S} \)-methionine, essentially as described by Jaskunas et al. (1975). The UV-sensitive bacterial strain M159 (Jaskunas et al., 1975) was infected with phage that had been concentrated in a CsCl step gradient.

A single colony of M159 was picked into 5 ml of L-broth and shaken overnight at 37°C. 0.1 ml of the overnight culture was used to inoculate 10 ml of M9 maltose minimal medium and again shaken overnight at 37°C. This culture was further diluted in M9 maltose medium to an \( \mathrm{OD}_{650} \) of 0.1. 50 ml of the diluted culture were grown, with aeration, at 37°C to an \( \mathrm{OD}_{650} \) of 0.5 (approximately \( 2 \times 10^8 \) cells ml\(^{-1}\)). 30 ml of cells were harvested by centrifugation at 8,000 rpm and 15° for 10 minutes, using Nalgene 50 ml polypropylene tubes in a Sorvall centrifuge and SS34 rotor, and resuspended in 4.5 ml of cold M9 maltose minimal medium containing 20 mM \( \mathrm{MgSO}_4 \). This gave approximately \( 10^9 \) cells ml\(^{-1}\). 4 ml of cells were placed in a sterile petri dish on ice and UV-irradiated for 7½ minutes at 4,500 ergs mm\(^{-2}\) using a pre-calibrated UV Lamp. Irradiation was carried out in
the dark to minimise photoreactivation and the petri dish was gently swirled to ensure that all cells were irradiated. The culture was subsequently poured into a flask wrapped in aluminium foil and kept on ice for 10 minutes.

For each phage analysed, duplicate microfuge tubes were prepared (for "early" and "late" infections) containing $1.25 \times 10^8$ phages and kept on ice. 50 µl aliquots of cells ($5 \times 10^7$) were added; this gave a m.o.i. of 2.5. Control tubes with and without the vector phage were included. The tubes were left on ice for 10 minutes to allow phage adsorption and then transferred to a 37°C water bath. After 2 minutes, 200 µl of pre-warmed M9 maltose minimal medium containing L-methionine carrier were added to each tube followed by 20 µCi of $^{35}$S-methionine after a further 3 minutes, for "early" infections, and 25 minutes, for "late" infections. In each case the cells were labelled for 10 minutes and chased with "cold"L-methionine for 10 minutes. After cooling on ice they were harvested by centrifugation in a microfuge at 4°C for 5 minutes. They were resuspended in 1 ml of cold acetone, again harvested and finally resuspended in 20 µl of TE. 20 µl of sample buffer and 5 µl of β-mercaptoethanol were added to each sample. After boiling for 5 minutes, 5 µl aliquots were taken from each sample and the radioactivity was counted in solubilising scintillation fluid. Suitable amounts from each sample (usually 40,000-50,000 cpm) were loaded
on an SDS-polyacrylamide gel (section (iv) below) and remaining samples were stored at -70°C.

(iv) SDS-Polyacrylamide Gel Electrophoresis of Polypeptides:

Electrophoresis of polypeptides was carried out in 10% polyacrylamide gels in the presence of 0.1% SDS using the discontinuous system described by Laemmli (1970). The constituents for a gel of dimensions 27 cm x 16 cm were as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Separating Gel</th>
<th>Stacking Gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% acrylamide/0.8% bis-acrylamide</td>
<td>10 ml</td>
<td>1.5 ml</td>
</tr>
<tr>
<td>1.5M Tris-HCl pH 8.8</td>
<td>7.5 ml</td>
<td>-</td>
</tr>
<tr>
<td>0.5M Tris-HCl pH 6.8</td>
<td>-</td>
<td>2.25 ml</td>
</tr>
<tr>
<td>10% SDS</td>
<td>0.3 ml</td>
<td>90 µl</td>
</tr>
<tr>
<td>Distilled water</td>
<td>12.2 ml</td>
<td>5.2 ml</td>
</tr>
<tr>
<td>TEMED</td>
<td>20 µl</td>
<td>10 µl</td>
</tr>
<tr>
<td>25% ammonium persulphate*</td>
<td>20 µl</td>
<td>10 µl</td>
</tr>
</tbody>
</table>

*prepared in distilled water, immediately prior to use.

The separating gel solution was poured between glass plates (which were separated by 1.5 mm spacers and sealed with 1% w/v agarose) and overlaid with water-saturated butanol. When the separating gel had set, the butanol was rinsed off with distilled water and a comb for forming wells was inserted between the plates. The stacking gel solution was prepared, poured on to the separating gel.
and allowed to set. The comb was carefully removed and the walls of the wells were straightened prior to loading. Samples were loaded using a Terumo microsyringe (MSG-50). Electrophoresis, in Tris/glycine/SDS buffer, was carried out overnight at 100 volts, constant voltage, until the tracking dye in the samples had travelled the length of the gel. Standard protein markers were used for size estimation.

The gel was stained in Coomassie Blue for 20 minutes at 37°C and destained for 2 hours in 2-3 changes of acetic acid/methanol. Both staining and destaining were carried out at 37°C without shaking. The gel was transferred to blotting paper, covered with "Saran Wrap" plastic film and dried under vacuum at 80°C for 1 hour. The Saran Wrap was removed and autoradiography was carried out at room temperature using Cronex x-ray film.
2.2.d. DNA SEQUENCING

This was carried out by the M13 dideoxy chain termination method (Sanger et al., 1977, 1980). DNA fragments generated either by endonuclease restriction or by sonication (Dieninger, 1983) were cloned in the M13 vectors Mpl0, Mpl1, Mpl8 and Mpl9 (see Messing, 1983, Yanisch-Perron et al., 1985). \( \alpha^{35} \text{SdATP} \) was used as radioactive label in the sequencing reactions and electrophoresis was in 0.35 mm polyacrylamide buffer gradient gels (Biggin et al., 1983).

(i) Sonication and Cloning in M13

A modification of the method described by Deininger (1983) was used to generate random fragments for sequencing. A restriction fragment containing the DNA to be sequenced had previously been cloned in pBR322. Instead of purifying the relevant DNA fragment and self ligating to create a circular molecule for sonication, whole plasmid (circular) DNA was sonicated.* (Sonication of a circular molecule gives a more random breakage of the DNA.) Recombinant phages resulting from cloning of these fragments in suitable M13 vectors were subsequently screened by hybridisation and those containing fragments of pBR322 were discarded.

* In one case, however, a small purified fragment (of approximately 700 bp) was sonicated as a linear fragment.
Sonication was carried out in an Electrosonic H60-2 sonicating waterbath containing 500 ml of water. The DNA was placed in a 0.5 ml Sarstedt tube which was secured in a polystyrene float in the middle of the waterbath so that the bottom of the tube was below the surface of the water. A test sonication was carried out to determine the conditions producing fragments ranging from 300-600 bp in size. 3 μg of DNA in 30 μl of TE were sonicated for 10 minutes in 2 minute "bursts" with 30 second intervals to reduce heating of the DNA solution. 5 μl aliquots were removed after each "burst" and subsequently checked by electrophoresis to estimate the degree of breakage and the size of the fragments produced. Generally sonication for 6 or 8 minutes was sufficient. 10 μg of DNA, in 30 μl of TE were sonicated for subsequent cloning and sequencing. Sonication generated fragments with ragged ends; these were therefore repaired before cloning. End repair was carried out by adding 4 μl TM buffer, 1 μl dNTP chase and 20 units of *E. coli* DNA polymerase, and incubating at 15°C for 3-4 hours. Sonication also produced some DNA fragments of less than 100 bp. Small fragments were removed by electrophoresis separation. The end-repaired DNA was subjected to electrophoresis at 100 volts through a 0.7% w/v Seakem LE agarose gel in 1 x TBE buffer. Ethidium bromide was included in the gel and the buffer; this made the DNA visible under UV light, which allowed the progress of the DNA through the gel to be monitored. Electrophoresis for 15-20 minutes was generally sufficient; the leading part of the
DNA smear (containing the small fragments) was then cut out, forming a trough, and the remaining DNA fragments were recovered by electro-elution into the trough (see Methods, section b (xv)). (A 300 bp DNA marker track was included in the gel; the part of the smear below the position of the marker was cut out.) The small fragments were removed to avoid the problem of reading non-contiguous sequences from clones with multiple inserts.

The eluted fragments, after phenol extraction and ethanol precipitation, were cloned in the relevant M13 vector which had been restricted with Sma I to generate blunt ends. The ratio of sonicated fragment ends to vector ends in the ligation was such that there was a considerable excess of donor ends; this reduced the extent of vector self-religation.* Three different ligations, using varying amounts of donor fragments, were carried out to determine the best ratio. The ligated DNA was subsequently used to transfect the bacterial strain NM522 by the Hanahan procedure (Methods, section b (xiii)) and white (recombinant) plaques were picked and purified.

* Phosphatase treatment of the restricted vector DNA prior to ligation was found to greatly reduce vector self-ligation, but also lowered the overall ligation efficiency. Therefore the vector was not generally phosphatase-treated.
(ii) DNA Preparation from M13 Phage Lysates:

DNA was purified from M13 phage lysates (i.e. single stranded phage) for subsequent use as a template for DNA sequencing, using a method based on that of Sanger et al., (1980).

A fresh overnight culture of NM522, from a single colony, was diluted 100-fold in L-broth and 1.5 ml aliquots were dispensed into Bijou bottles. For each M13 clone, a fresh single plaque was picked into a bottle containing the diluted NM522 culture and the bottles were shaken vigorously at 37°C for 5 hours. The cultures were then transferred to microfuge tubes and clarified by spinning for 5 minutes. The supernatants were transferred to clean microfuge tubes, 150 µl of PEG/NaCl were added to each tube and after 20 minutes at room temperature the phage were pelleted by centrifugation in a microfuge for 10 minutes. The PEG/NaCl was drawn off and, after centrifugation for 1 minute, any residual PEG/NaCl was removed using tissue paper. The pellets were resuspended in 100 µl of TE and extracted with 50 µl of TE-equilibrated phenol. After the phenol extraction, the aqueous layer was carefully taken off and transferred to a clean microfuge tube. For each sample, the DNA was ethanol precipitated by adding 10 µl of 3M NaAc and 300 µl of ethanol, and leaving overnight at -20°C. The DNA was recovered by microfuge centrifugation for 10 minutes, dissolved in 100 µl of TE and again ethanol precipitated. For the second precipitation
the tubes were left at -70°C for 1 hour. After centrifugation, the pellets were washed with 1 ml of 95% ethanol, dried under vacuum and finally resuspended in 50 μl of TE. "Templates" were stored at -20°C until required. The second ethanol precipitation was carried out to ensure that all traces of phenol had been removed; any phenol carried over in the template DNA gave "background" and "laddering" effects in the sequencing reactions.

(iii) M13 Strand-Specific Hybridisation:

This hybridisation procedure, modified from that of Hu and Messing (1982), employs an M13 template (single-stranded DNA) as a probe to screen for M13 clones which would, when sequenced, give the opposite strand. The hybridisation primer used anneals to the 5' region of the vector sites and, under the radio-labelling conditions used, synthesises a labelled (-) strand complementary to all or part of the M13 vector DNA but leaves the cloned DNA single stranded. This single-stranded DNA will, therefore, specifically hybridise to single-stranded DNA of the complementary sequence.

The M13 template probe was prepared by first annealing the M13 hybridisation primer to the template DNA and then labelling with α-32PdCTP. The annealing was effected by mixing 1 μl M13 template DNA, 1 μl (10 x) Hin buffer; 1 μl hybridisation primer and 3 μl distilled water; heating in a boiling water bath for 3 minutes and subsequently cooling, slowly to room temperature. The annealed primer/template were then mixed with
10 µCi α-32PdCTP, 3 µl dNTP mix, 1 unit DNA polymerase (Klenow fragment) and 6 µl distilled water. After incubation at 15°C for 90 minutes, the reaction was terminated by the addition of 100 µl of 10 mM EDTA and the probe was stored at -20°C until required. The probe was not denatured before use; 50-60 µl were sufficient for one hybridisation.

For this hybridisation procedure phage lysates of the M13 clones in question were prepared to provide the necessary single-stranded DNA. 5 µl of each lysate were spotted on to a nitrocellulose filter, allowed to dry and then treated as for plaque hybridisation (see Methods, section b (xvii)). When template DNA was used instead of lysates, 3 µl of each template were spotted on to a nitrocellulose filter soaked in 2 x SSC; the filter was blotted dry and kept at 80°C under vaccum for 2 hours. Prior to hybridisation, the filter was shaken gently at 65°C for 6 hours in 30 ml of M13 hybridisation buffer containing 50 µg ml⁻¹ of denatured (by boiling for 5 minutes), sonicated calf thymus DNA. (Sonicated calf thymus DNA was kindly donated by Anne S. Daniel.) The M13 probe (see above) was added and the filter was hybridised overnight under the same conditions. After hybridisation, the filter was washed twice in 1 x SSC/0.1% SDS and twice in 0.5 x SSC/0.1% SDS. All washes were for 30 minutes at 65°C. Finally, the filter was blotted dry, placed between two layers of "Saran Wrap" plastic film. For autoradiography Cronex x-ray film and
intensifying screens were used; the film was pre-flashed and exposed as required at -70°C.

(iv) Sequencing Reactions:

Sequencing reactions were carried out essentially as described by Bankier and Barrell (1983).

The DNA templates which were to be sequenced were first annealed to M13 sequencing primer. In each case, 8 μl template DNA, 1 μl TM buffer and 1 μl M13 sequencing primer were mixed in a microfuge tube and incubated at 65°C for 1 hour. The tubes were subsequently cooled to 4°C and centrifuged briefly prior to their use in the sequencing reactions.

These reactions were carried out in lidless microfuge tubes set in plastic racks which fit the Eppendorf Model 5413 centrifuge. These racks allow 8 clones to be sequenced at one time and have an added advantage in that they can be both boiled and frozen. Therefore, all stages of the reactions could be carried out in these racks. A Hamilton PB600-1 fitted with a 1710 LT syringe and siliconised micropipette tips were used to dispense 2 μl aliquots of reagents as speed was essential to obtain satisfactory chain termination reactions. Droplets of individual reagents were kept separate on the side of the tubes; this ensured that all reactions were started, and stopped, at exactly the same time (by centrifugation). For each clone, 2 μl aliquots of annealed primer/template were placed in 4 reaction tubes. 2 μl pf ddNTP mixes (T°, C°, G°, A°) were added to the respective tubes for
all clones. A Klenow/label mix for 8 clones was prepared on ice, consisting of:

- 7.2 μl 100 mM Tris-HCl pH 8.5
- 7.2 μl 100 mM DDT
- 24 units α-35SdATP
- 12 units E.coli DNA polymerase, Klenow fragment
- 48 μl sterile distilled water

2 μl aliquots of the Klenow-label mix were added, without delay to all samples and the reactions were started simultaneously by brief centrifugation. After 25 minutes at room temperature*, 2 μl of dNTP chase were added. The reactions were "chased" for 25 minutes at room temperature and the samples were subsequently stored at -20°C until required. Before loading on a polyacrylamide/urea gel, 2 μl of Formamide dyes were added to each tube and the tubes were heated in a boiling water bath for 5 minutes. Usually only a third of the reaction mix was loaded; the remainder was stored at -20°C and if a second gel was required, the samples were boiled for 1 minute before loading.

* For sequences which were difficult to read, particularly G-C-rich regions, inosine was substituted for guanosine in dGTP for all ddNTP mixes (Mills and Kramer, 1979) as it disrupts secondary structures, hence reducing band compression. Inosine has a lower efficiency than guanosine in base pairing, therefore the reactions were allowed to proceed for 35 minutes at room temperature.
(v) **Sequencing Gels:**

Urea/polyacrylamide gradient gels were prepared basically as described by Biggin *et al.*, (1983). Gels were poured between glass plates of dimensions 40 cm x 20 cm with 0.4 mm spacers and sealed with "Sellotape" thermosetting tape (1607). Each gel contained 25 ml of 0.5 TBE gel mix and 6 ml of 5 x TBE gel mix supplemented with 50 μl ammonium persulphate/25 μl TEMED and 12 μl ammonium persulphate/6 μl TEMED respectively, for polymerisation. The gradient was prepared in a 25 ml pipette by drawing up 7 ml of the 0.5 x TBE mix, followed by the 5 x TBE mix. The bromophenol blue in the 5 x TBE mix showed a two-phase solution separated by a diffuse boundary; 8-10 air bubbles were then allowed to run up the pipette to create the gradient. This was then poured between the glass plates and topped by the remaining 0.5 x TBE gel mix. Sharkstooth combs, 0.4 mm thick and with 3 mm slots (Bethesda-Research-Laboratories) were used to form the loading wells; these gave finer bands and generally allowed an extra 50-100 bases to be read off each gel. The main modification of the method of Biggin *et al.*, (1983) was in the buffers used for electrophoresis. Instead of using 1 x TBE for electrophoresis, 0.5 x TBE was used in the top tank with 2 x TBE in the bottom tank. This gave improved gels with sharper bands (Heather E. Houston, pers. comm.). Electrophoresis was at 1.25 K volts and 20 m Amps and was continued until the bromophenol blue in the loading dye was within 2 cm of the bottom of the gel.
After separation of the plates the gel was fixed in 10% acetic acid/10% methanol v/v for 15 minutes, transferred to blotting paper, covered with "Saran Wrap" plastic film and dried at 80°C under vacuum for 1 hour. The Saran Wrap was removed and autoradiography was carried out overnight at room temperature using Cronex x-ray film.

(vi) **Sequence Data Analysis**

This was carried out using a VAX 11/750 computer and the Staden (1982) and University of Wisconsin Genetics Computer Group (Devereux et al., 1984) programs.
CHAPTER 3

CLONING AND ANALYSIS OF

hsd A AND hsd E
3.1 INTRODUCTION

The most extensively studied Type I restriction and modification systems are those of *E.coli* K and *E.coli* B. These are allelic and are located close to *serB* at 98.5 minutes on the *E.coli* map (Boyer, 1964; Glover and Colson, 1969; Bachmann and Low, 1980).

Extensive complementation analyses using a variety of restriction-deficient mutants of *E.coli* K and B (Boyer and Roulland-Dussoix, 1969; Glover and Colson, 1969; Glover, 1970; Hubacek and Glover, 1970) showed that three genes are involved in restriction and modification. They also led to the hypothesis that the product of only one of the three genes, *hsdS*, imparts the specificity for recognition of the DNA sequence characteristic of each system, that of a second gene, *hsdM*, together with *hsdS* is required for modification, while the product of the third gene *hsdR*, together with the other two, is essential for restriction.

Restriction-deficient (r−) mutants, originally isolated by Wood (1966), were found to be of two different classes: r−m+, with normal modification and r−m−, which were also deficient in modification. These occurred with comparable frequencies and therefore r−m− mutants were unlikely to be true double mutants. Hence it was proposed that they arose from a mutation in the *hsdS* gene, while r−m+ mutants were due to a mutation in the *hsdR* gene. Mutations in the *hsdM* gene were, at the time, presumed to be lethal as it was thought that they would...
give a $r^+m^-$ phenotype. However, Hubacek and Glover (1970) isolated temperature-sensitive mutants of *E.coli* K12 which were defective in restriction and modification and these were subsequently found to have mutations in the *hsdM* gene (see below). These mutations were not lethal at restrictive temperatures but resulted in a deficiency in restriction as well as in modification, indicating that the product of the *hsdM* gene was also required for restriction.

Complementation tests with F' merodiploids (Boyer and Roulland-Dussoix, 1969; Glover, 1970) showed that $r^-m^+$ (*hsdR^-M^+S^+*) mutants complemented $r^-m^-$ (*hsdR^-M^-S^-*) mutants, while independent $r^-m^+$ mutants did not complement, as was the case with independent $r^-m^-$ ones. The subunits of the endonucleases encoded by *E.coli* K and *E.coli* B were found to be interchangeable, thus allowing complementation tests between the $r^-m^+$ and $r^-m^-$ strains. Such inter-strain tests using mutants from *E.coli* K and *E.coli* B defined the *hsdS* gene (Boyer and Roulland-Dussoix, 1969). A $r_K^-m_K^+/r_B^-m_B^-$ diploid had a $r_K^+m_K^+$ phenotype as the *hsdS* gene of the *E.coli* K mutant was functional. In the same way, a $r_B^-m_B^+/r_K^-m_K^-$ diploid had a $r_B^+m_B^+$ phenotype. Some $r_B^-m_B^+$ mutants, mutated to $r_B^-m_B^-$ by a second-step mutation (Boyer and Roulland-Dussoix, 1969) were found to complement the wild type $r_K^+m_K^+$ to give a $r_{KB}^+m_{KB}^+$ phenotype, indicating that both specificity (*hsdS*) genes were functional. The second-step $r^-m^-$ mutants were subsequently found to complement first step $r^-m^-$ ("**hsdS**")
mutants, giving a $r^{-m^+}$ phenotype. Thus the mutations in first- and second-step $r^{-m^-}$ mutants were in different cistrons, even though they resulted in the same phenotype. In another experiment, using $r_{K}^{-m_{K}^{ts}}$ mutants, Hubacek and Glover (1970) showed that these mutants, which were deficient in both restriction and modification at high temperatures, complemented $r_{B}^{-m_{B}^+}$ mutants to give K- and B-specific restriction and modification, even at high temperatures. In both these tests, the hsdS genes were shown to be functional; therefore, these mutations, which gave rise to a $r^{-m^-}$ phenotype, must have occurred in the hsdM gene. This not only characterised the function of the hsdM gene, but highlighted the requirement for the product of this gene in both restriction and modification.

In vitro complementation tests using purified enzymes from E.coli K and E.coli B gave the same results as those obtained from genetic tests. Kühnlein et al. (1969) purified EcoB from various mutants of E.coli B and showed that the extract from a $r_{B}^{-m_{B}^+}$ mutant had modification, but no restriction, activity, while that from a $r_{B}^{-m_{B}^-}$ mutant lacked both activities. One fraction derived from a $r_{B}^{-m_{B}^+}$ strain only showed modification activity; when this was mixed with the extract from a $r_{B}^{-m_{B}^-}$ mutant, it could both modify and restrict DNA. Hadi and Yuan (1974) purified enzymes from a restriction-deficient, but modification-proficient, mutant of E.coli K ($r_{K}^{-m_{K}^+}$) and from a mutant which lacked
both activities \((r_K \cdot m_K)\), and showed that, when mixed, these complemented each other to give an enzyme which could both modify and cleave DNA.

Investigations into the restriction and modification properties exhibited by various species of Enterobacteriaceae resulted in the identification of other systems which map in the same region of the chromosome as those of E.coli K and B. These included the hsd A system from E.coli 15T\(^{-}\) (Arber and Wauters-Willems, 1970) and two from Salmonella serotypes, hsd SB from S.typhimurium and hsd SP from S.potsdam (Colson and Van Pel, 1974; Bullas and Colson, 1975a). The Salmonella SB and SP systems were found to be not only related to each other but also to the E.coli K and B systems. Complementation tests using Hfr/F\(^{-}\) hybrids indicated efficient exchange between the subunits of the enzymes (Van Pel and Colson, 1974; Bullas and Colson, 1975a). The E.coli 15T\(^{-}\) system (hsd A), however, did not appear to be related to that of E.coli K on the basis of complementation tests (W. Arber, pers. comm.), despite the finding that these genes behave as alleles at the same locus in experiments using phage P1 to mediate genetic recombination (W. Arber, N. E. Murray, pers. comm.).

More recently, molecular experiments which readily demonstrated similarity between hsd K, hsd B, hsd D (a new specificity carried by an E.coli strain El66; Murray, et al., 1982), and the hsd SB and SP systems, failed to detect any relatedness between the hsd K and hsd A systems (Murray et al.1982). These experiments relied on
immunological cross-reactivity, using antiserum prepared against the hsdR and hsdM subunits of EcoK and DNA hybridisation (using probes from within the hsd K region) as a measure of homology. Nevertheless, early physiological experiments, involving the hsd A system, indicated that AdoMet is required for both restriction and modification, implying that the hsd A genes encode a Type I restriction and modification system (Lark and Arber, 1970). Earlier DNA hybridisation experiments (Sain and Murray, 1980), using \( \lambda \) derivatives carrying the hsdMS genes of E.coli K12 as probes, showed DNA homology between E.coli K and E.coli B and, significantly, no homology with E.coli C. E.coli C does not possess any known restriction and modification system and its lack of homology with the E.coli K12 restriction and modification genes implies a lack of structural genes rather than a functional defect.

Several Type I restriction endonucleases have now been purified extensively. They all consist of three non-identical subunits but there have been varying reports concerning the relative proportions and the exact molecular weight of the subunits. All the K-related enzymes contained subunits of molecular weight about 135,000, 60,000 and 50,000 (see Bickle, 1982; Nagaraja et al., 1985a,b,c). (Details of enzyme structure and reaction mechanisms are given in the general introduction, Chapter 1.) Modification enzymes purified from a wild-type strain of E.coli B (Lautenberger and Linn, 1972) and, more
recently, from an \textit{E. coli} K12 strain lysogenic for a \(\lambda\) derivative carrying the \textit{hsd} K region (see Bickle, 1982; Suri et al., 1984a), contained the two smaller subunits, thus identifying the large subunit as that encoded by \textit{hsdR}. Genetic analysis, involving complementation tests between deletion derivatives of a \(\lambda\) phage carrying the \textit{hsd} K region and various plasmids containing sections from the same region, followed by analysis of the polypeptides produced by the various \(\lambda\textit{hsd}\) phages after infection of UV-irradiated cells (Sain and Murray, 1980), identified the smallest polypeptide as the product of the \textit{hsdS} gene and the intermediate polypeptide as that of the \textit{hsdM} gene.

The analysis of the \(\lambda\textit{hsd}\) K deletion derivatives also showed the organisation of the \textit{hsd} genes, namely that they are in the order \textit{hsdR}, \textit{hsdM} and \textit{hsdS}, and that they are arranged in two transcriptional units with a promoter for \textit{hsdR} and a promoter for \textit{hsdM} and \textit{hsdS}, all genes being transcribed in the same direction. This organisation could allow differential control in modification and restriction. Assuming that under normal, \textit{in vivo}, conditions restriction and modification are carried out by the same polypeptide complex and that the preferred activity is determined by the nature of the DNA substrate (Vovis, et al., 1974), control of this sort would only be advantageous when intact \textit{hsd} genes are transferred to a bacterial strain which lacks the same specificity. Biological experiments have shown that when phage P1 (which carries a Type III restriction
and modification system) enters *E. coli*, the modification activity determined by this phage is expressed before its restriction activity (Arber and Dussoix, 1962; Arber et al., 1975). A similar sequential control was suggested from experiments involving conjugational transfer of the *hsd* K genes (Glover and Colson, 1969). The genetic organisation of the *E. coli hsd B* and *hsd D* genes was, as expected, found to be similar to that of the *hsd K* genes (Gough and Murray, 1983) but there was no information concerning the organisation of the *hsd A* genes.

In this chapter, the cloning of the *hsd A* genes in *λ* is described. *λhsd* phages, which modify their DNA with A specificity, are easily recognised as they plate normally on an A-restricting host, while other phages plate with an efficiency of $10^{-2}$. This method of recognising *λhsd* phages was used in the isolation of *λhsd K* phages (Borck et al., 1976). Three *hsd A* genes are defined and it is shown that, as for *hsd K* (Sain and Murray, 1980), the genes are in the order *hsdR*, *hsdM* and *hsdS*. The distinction between the *hsdM* and *hsdS* genes relied on the identification and cloning of a new specificity, E, which, although related to A, recognises a different DNA sequence. The subunits of the restriction endonucleases encoded by *hsd A* and *hsd E* are interchangeable and this allowed inter-strain complementation tests which identified the *hsdS* gene. Analysis of the polypeptides encoded by the *hsd A* genes allowed the correlation of the polypeptides with respective genes.
The λhsd A phage described in this chapter has been used in the purification of the EcoA restriction and modification enzyme (Suri et al., 1984b). A dilysogen of E.coli C (a naturally restrictionless strain) carrying a \( \lambda \text{att}^- \text{red}^- \text{cI857 Q}\text{S}^- \) derivative of λhsd A and an \( h^{80} \text{cI857 Q}\text{S}^- \) helper phage allowed the overproduction of EcoA, following heat induction, without cell lysis. The enzyme was found to consist of 2 protein species, separable by column chromatography. One contained two different subunits of molecular weights 64,000 and 55,000 and was a functional methylase. The second species consisted of a single polypeptide of molecular weight 98,000 which, alone, had no enzymatic activity but, when associated with the two subunits of the first species, formed a functional restriction endonuclease. The enzyme is, therefore, similar to other Type I enzymes. The reactions involved in restriction and modification also resemble those of other Type I systems in their requirement for AdoMet and ATP (Suri et al., 1984b).

The DNA sequences recognised by the restriction endonucleases encoded by the Type I systems mentioned above have been determined. All, including that recognised by EcoA, have the same overall hyphenated structure consisting of a trimer and a tetramer (or pentamer) of specific sequence, separated by a spacer of non specific sequence, of a fixed, but different, length (6-8 bases) for the various enzymes (see Bickle, 1982; Suri et al., 1984b; Nagaraja et al., 1985 a,b,c). (Details of these recognition sequences are
given in the general introduction, Chapter 1.4.b.)
The similarity between the EcoA recognition sequence and
those recognised by enzymes of the K family again underlies
a basic similarity between these systems.

The A-specificity system, therefore, may identify a
new family of Type I restriction systems. All the evidence
to date suggests that it is indeed a Type I system, in
spite of its lack of homology with those of the K-family
(Murray et al., 1982). The findings from the early
physiological experiments (Lark and Arber, 1970) have been
corroborated by the evidence relating to the enzyme
structure and reaction requirements and to its recognition
sequence (Suri et al., 1984b).

3.2 RESULTS
3.2.a. Cloning of hsd A Genes in λ

DNA from WA2899, a derivative of E.coli K12 in
which the hsd K genes had been replaced by the hsd A genes
(W. Arber, pers. comm.) was digested with HindIII and a
library of recombinants was made in the HindIII replacement
vector λNM762. The phage library was recovered on a non-
restricting (r^−), non-modifying (m^−) E.coli K12
derivative, NM477.* The efficiency of plating (e.o.p.) of the
resulting library of phages was less than 10^{-2} on the r^+_A
strain WA2899. Phages which include the genes required

*Since no complementation has been observed between the
hsd K and hsd A systems (W. Arber, pers. comm.) and no DNA
homology has been detected between the two sets of genes
(Murray et al., 1982), the non-restricting, non-modifying
derivative of E.coli K12 (NM477) was generally used for
propagating λ_hsd A phages.
for A-modification might be expected to occur at a frequency of $10^{-3}$, but they will modify their own DNA, thereby enabling them to plate with high efficiency on the $r_A^+$ strain WA2899. Enrichment for $\lambda hsd$ A phages was therefore effected by propagating the phage library on WA2899. After subsequent growth in the non-modifying host NM477, one hundred of the resulting phages were tested for A-modification and four were found to be modified. WA2899 was used rather than E.coli 15T$, the original strain in which the $hsd$ A system was identified, as the latter strain has an additional plasmid encoded restriction system (Arber and Wauters-Willems, 1970).

The four $\lambda hsd$ A phages isolated were all found to contain a 15 kb HindIII fragment. To determine whether these phages carried the whole $hsd$ region, i.e. the genes required for both restriction and modification, dilysogens of these phages were made in the $r_K^-m_K^-$ host NM477 and these were checked for restriction and modification activity. (Dilysogens were made using a heteroimmune helper phage as the $\lambda hsd$ A phages were integration-deficient.) $cI^+$ phages are needed to make lysogens. Therefore, $cI^+$ derivatives of the $\lambda hsd$ A phages were constructed by changing the right arm of the vector $\lambda NM762$ (imm$^1 cI\Delta \lambda$) for that of $\lambda NM761$ (imm$^{21} cI^+$). This was carried out in vitro by restricting $\lambda NM761$ and $\lambda hsd$ A with HindIII and ligating. imm$^{21} cI^+$ phages, recovered by transfection of NM477, were recognised as turbid plaques and those carrying the $hsd$ insert were subsequently
### Table 3.1

Efficiency of plating of $\lambda$vir phages on lysogens of $\lambda$hsd A

<table>
<thead>
<tr>
<th>Bacterial strains</th>
<th>E.o.p. of phages:**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda$vir.O</td>
</tr>
<tr>
<td>NM477 ($r_Km_K^{-}$)</td>
<td>1</td>
</tr>
<tr>
<td>WA2899 ($r_A^+m_A^+$)</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>NM477 ( hsd A)*</td>
<td>$10^{-2}$</td>
</tr>
</tbody>
</table>

* NM477 lysogenic for $\lambda$hsd A $imm^{21} CI^+$ and $\lambda^+$ (helper)

** $\lambda$vir.O: non-modified

$\lambda$vir.K: K-modified

$\lambda$vir.A: A-modified

$\lambda$vir.477 ($\lambda$hsd A): $\lambda$vir.O phages after growth in NM477 ($\lambda$hsd A)

Notes: NM477 and WA2899 were used as non-restricting and restricting controls, respectively.

NM477 ($\lambda$hsd A) restricted $\lambda$vir.K confirming that the A specificity is different from that of E.coli K12
identified as they plated with normal efficiency on WA2899. Dilysogens of NM477 were made by mixed infection with the integration deficient $\lambda hsd A^+ cI^+$ and a heteroimmune helper, $\lambda^+$; they were selected in the presence of $\lambda imm^2 cI^+$ ($\lambda NM507$) and $h^82 imm^2 cI^+$ ($\lambda NM848$) phages to minimise the background of $\lambda$ resistant colonies. Lysogens were then checked for their restriction and modification ability $\lambda vir.O$, $\lambda vir.K$ and $\lambda vir.A$ (i.e. $\lambda vir$ modified with no, K and A specificity respectively). All were found to restrict $\lambda vir.O$ and $\lambda vir.K$ but not $\lambda vir.A$, and all surviving phages were found to be modified with A specificity (see Table 3.1). This confirmed that the lysogens, and hence the original $\lambda hsd A$ phages carried all the genes necessary for A-specific restriction and modification. This restriction and modification ability was irrespective of the orientation of the cloned DNA fragment in the $\lambda hsd A$ phages; these phages therefore included not only the $hsd A$ genes but also the promoter, or promoters, necessary for their expression.

3.2.b. Analysis of $\lambda hsd A$; Isolation and Analysis of Deletion Derivatives

Restriction analysis of $\lambda hsd A$ DNA using EcoRI, HindIII, BamHI and SalI gave the map shown in Figure 3.1. To localise more precisely the $hsd A$ genes within the 15kb insert, deletion derivatives of $\lambda hsd A$ were selected on BBL agar supplemented with 0.4 mM EDTA. Twenty such phages were isolated and tested for A-modification; nine
Table 3.2
Efficiency of plating of λvir phages on NM477 lysogenic* for λhsd A deletion derivatives

<table>
<thead>
<tr>
<th>Bacterial strains</th>
<th>λvir.O</th>
<th>λvir.A</th>
<th>λvir.NM477 (λhsd Δ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM477 ( r^K_m^K )</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WA2899 ( r^A_m^A )</td>
<td>10^-2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NM477 (λhsd Δ6)</td>
<td>10^-2</td>
<td>1</td>
<td>-***</td>
</tr>
<tr>
<td>NM477 (λhsd Δ12)</td>
<td>10^-2</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>NM477 (λhsd Δ4)</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>NM477 (λhsd Δ8)</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

* in each case dilysogens with λhsd A imm21 CI derivatives and λ' as helper.

** λvir.O: non-modified
λvir.A: A-modified
λvir.NM477 (λhsd Δ): λvir.O phages after growth in any of the dilysogens tested

*** -: not tested

Notes: NM477 and WA2899 were respective non-restricting and restricting controls.
retained the ability of modify their DNA. The location
and extent of the deletions were then determined by
restriction analysis of phage DNA and this correlated
with their modification phenotype as shown in Figure 3.2.
Phages retaining A modification could be divided into two
classes, I and II, according to the location of their
deletions (see Fig. 3.2.). Dilysogens of phages
from both of these classes (namely Δ6, Δ12, Δ4, and Δ8)
were made in NM477 to determine whether a functional
hsdR gene had been retained. These were made as
described above (section a) using imm21C1+ derivatives
of the phages. Lysogens of Class I deletion phages
were rA+; those of Class II were rA- (see Table 3.2).
All were, as expected, modification proficient. Since one
Class I deletion (Δ12) removes the EcoRI site marked Eco*
(see Fig. 3.3), the entire hsd region must lie to the
left of this EcoRI site. Class II deletions remove at
least part of the hsdR gene, thereby locating this gene
to the left-hand part of the insert. The ability of
Class II deletion derivatives to modify DNA implies the
presence of a promoter, separate from that for hsdR,
which allows expression of the modification genes. This
would be analogous to the situation with hsd K (Sain
and Murray, 1980). A third class of deletions, III,
which fail to modify their DNA, locate genes essential for
modification to the right of the hsdR gene. An additional
deletion derivative (Δ21) was constructed in vitro by
deleting the DNA between the two BamHI sites (see Fig. 3.3);
this Class III deletion (r m -), indicates that DNA to the right of the BamHI site, marked Bam+ in Figure 3.3, is required for modification.

The hsd genes were therefore localised to the left hand region of the HindIII insert as shown in Figure 3.3. However, although an analysis of the EcoA polypeptides (Suri et al., 1984b) implied three genes, two of which encode the modification methylase, the roles of the two modification polypeptides were not distinguished. By analogy with hsd K, it is tempting to call the genes hsdM and hsdS with the implication that the polypeptide encoded by hsdS determined the specificity of the restriction and modification system. This hypothesis could be checked and the hsdS gene identified if an alternative, but different specificity were available in which the polypeptides could be interchanged with those of hsd A, in complementation tests comparable to those carried out with E.coli K12 and E.coli B (Boyer and Roulland-Dussoix, 1969; Glover, 1970; Hubacek and Glover, 1970).

For this purpose, a survey of natural isolates of E.coli was carried out with the aim of identifying and isolating specificity systems related to hsd A.

3.2.c. Identification and Cloning of A-related hsd Genes

The localisation of the hsd A genes identified an EcoRI-BamHI fragment (Eco+-Bam+ in Fig. 3.7) as being entirely with the hsd A region. This fragment was subcloned in pBR322 (pFFP20 in Fig. 3.7) and used as a
probe to screen the DNAs of twenty five natural isolates of E. coli, including 15T−, for the presence of hsd genes sharing homology with this segment of hsd A. The E. coli DNAs were digested with EcoRI and with HindIII respectively to give two alternative fragment patterns and, after separation by electrophoresis through agarose gels, the DNA fragments were transferred to nitrocellulose filters. The filters were first hybridised with 32P-labelled pBR322 DNA and, three to four weeks later (to allow decay), with 32P-labelled pFFP20. This prior hybridisation identified fragments which shared homology with the pBR322 vector rather than with the hsd insert and reduced the background of non-specific hybridisation as seen in Figure 3.4.

In addition to E. coli 15T− and WA2899, the DNA of two isolates of E. coli shared homology with the hsd A probe; Figure 3.4 shows the hybridisation pattern from EcoRI digests. DNA from each of these two isolates, namely A58 and A101, was restricted with EcoRI, HindIII, and BamHI, again separated by electrophoresis through an agarose gel and transferred to a nitrocellulose filter. Duplicate filters were prepared, hybridised first with pBR322 (vector control) as described above, and subsequently with two different 32P-labelled probes, namely pFFP19 and pFFP21 (see Fig. 3.8). These DNA fragments cover the entire hsd A region as deduced in section a, above. These different hybridisations were carried out to determine which of the three enzymes (EcoRI, HindIII and BamHI) was most likely to give a single DNA
## Table 3.3

Efficiency of plating of A-related $\lambda$hsd phages on WA2899

<table>
<thead>
<tr>
<th>Phages</th>
<th>E.o.p. on bacterial strains:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NM477 ($r_K^{-}m_K^{-}$)</td>
</tr>
<tr>
<td>$\lambda^+$</td>
<td>1</td>
</tr>
<tr>
<td>(non-modified)</td>
<td></td>
</tr>
<tr>
<td>$\lambda$hsd A</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda$hsd A58</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda$hsd A101*</td>
<td>1</td>
</tr>
</tbody>
</table>

* All five $\lambda$hsd A101 phages behaved in the same way.
fragment of a suitable size, containing all the genes necessary for restriction and modification. The hybridisation patterns obtained with these probes (Fig. 3.5) indicated that the most suitable enzyme for cloning the *hsd A*-related genes, in both cases, was *HindIII*. A fragment of 16 kb for A58 and one of 11 kb for A101, hybridised to both probes. Restriction with *EcoRI* and *BamHI* resulted in either very large fragments (greater than 20 kb) or several small fragments hybridising to each of the two probes (see Fig. 3.5). The *hsd* genes were therefore cloned in the *λimm* 21 cI+ *HindIII* replacement vector NM761 and the phage libraries were recovered by *in vitro* packaging using the *rK mK* host NM477. [The use of the *imm* 21 cI + vector (XNM761) rather than λNM762 which is *imm* cIΔ, allowed dilysogens to be made as described in section 3.a.] *λhsd* phages were detected by plaque hybridisation, using α32P-labelled pFFP20 as probe. Detection by DNA homology was necessary because the *E.coli* strains A58 and A101 are resistant to phage λ. In each case two thousand plaques were hybridised; A58-derived plaques yielded one *λhsd* phage, while A101-derived ones gave five such phages. The *λhsd* phage derived from A58 was restricted by the *rA* + strain WA2899, those derived from A101 were not (see Table 3.3). This suggests that A101 has the A specificity while A58 may have a new specificity. Dilysogens of each *λhsd* phage were made in NM477, as in the case of *λhsd A* phages, and found to be restriction proficient (see Table 3.4), indicating that in all cases the *λhsd* phages carried all
Table 3.4
Efficiency of plating of \textit{\textlambda}vir\textit{\textphage} on dilysogens\textsuperscript{*} carrying A-related \textit{\textlambda}hsd\textit{\textphage}es

<table>
<thead>
<tr>
<th>Bacterial strains</th>
<th>\textit{\textlambda}vir.O</th>
<th>\textit{\textlambda}vir.K</th>
<th>\textit{\textlambda}vir.A</th>
<th>\textit{\textlambda}vir.NM477 (\textit{\textlambda}hsd A58)</th>
<th>\textit{\textlambda}vir.NM477 (\textit{\textlambda}hsd A101)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM477 (r\textsubscript{K} m\textsubscript{K})</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WA2899 (r\textsuperscript{+} m\textsubscript{A})</td>
<td>10\textsuperscript{-2}</td>
<td>10\textsuperscript{-2}</td>
<td>1</td>
<td>10\textsuperscript{-2}</td>
<td>1</td>
</tr>
<tr>
<td>NM477 (\textit{\textlambda}hsd A)\textsuperscript{*}</td>
<td>10\textsuperscript{-2}</td>
<td>10\textsuperscript{-2}</td>
<td>1</td>
<td>10\textsuperscript{-2}</td>
<td>1</td>
</tr>
<tr>
<td>NM477 (\textit{\textlambda}hsd A58)\textsuperscript{*}</td>
<td>10\textsuperscript{-3}</td>
<td>10\textsuperscript{-3}</td>
<td>10\textsuperscript{-3}</td>
<td>1</td>
<td>10\textsuperscript{-3}</td>
</tr>
<tr>
<td>NM477 (\textit{\textlambda}hsd A101)\textsuperscript{*}</td>
<td>10\textsuperscript{-2}</td>
<td>10\textsuperscript{-2}</td>
<td>1</td>
<td>10\textsuperscript{-2}</td>
<td>1</td>
</tr>
</tbody>
</table>

\textsuperscript{*} in each case dilysogens with \textit{\textlambda}hsd \textit{\textimm}\textsuperscript{21} cI\textsuperscript{+} phages and \textit{\textlambda}\textsuperscript{+} as helper

\textsuperscript{**} \textit{\textlambda}vir.O: non-modified
\textit{\textlambda}vir.K: K-modified
\textit{\textlambda}vir.A: A-modified
\textit{\textlambda}vir.NM477 (\textit{\textlambda}hsd A58)/\textit{\textlambda}vir.NM477 (\textit{\textlambda}hsd A101): \textit{\textlambda}vir.O phages after growth in the respective dilysogens
Table 3.4

Notes: Dilysogens of Λ-related phages restricted

Λvir.K indicating that they are not K-specific.

The efficiency of plating of Λ phages on the
Λhsd A58 dilysogen is 10-fold lower than that
on the Λhsd A101 or the Λhsd A dilysogens,
suggesting that there are more targets for the
A58 specificity on Λ. There is one target
for EcoA in Λ (Arber et al., 1972).

NM477 and WA2899 were respective non-restricting
and restricting controls.
the genes required for restriction and modification. The restriction properties of these lysogens, as seen from Table 3.4, confirmed that A101 had A specificity, while A58 had a different specificity. This was designated E.

Restriction analysis of DNA from the \( \lambda_{hsd} \) phages derived from A58 and A101 corroborated these results. Within the hsd region, as deduced from the analysis of deletion derivatives (section b, above), the restriction map of the A101-derived phage was identical to that of \( \lambda_{hsd} A \) (from restriction with EcoRI, HindIII and BamHI); on the other hand, that of the A58- derived phage (\( \lambda_{hsd} E \)) was different (see Figs. 3.6 and 3.7).

\( \lambda_{hsd} E \) was analysed by Gill M. Cowan, using the same strategy as that used for \( \lambda_{hsd} A \). The organisation of the hsd genes was found to be similar to that of \( \lambda_{hsd} A \) (Fuller-Pace et al., 1985). Deletion and mutant derivatives of \( \lambda_{hsd} E \) (isolated by Gill M. Cowan) were subsequently used in complementation tests with derivatives of \( \lambda_{hsd} A \). These tests are described later.

3.2.d. Identification of Polypeptides Encoded by the hsd A Genes

Segments of the hsd A region were subcloned in pBR322, using HB101 as host, and the derivatives were analysed to establish the order of the genes for the three polypeptides produced by the hsd A system (Suri et al., 1984b) (see Fig. 3.8). These plasmids were then used to transform a minicell strain WL542 and the polypeptides were
labelled with $^{35}$S-methionine and analysed after electrophoresis through SDS-polyacrylamide gels followed by autoradiography.

As expected pFFP30 (Fig. 3.8), which contained the HindIII insert from $\lambda$hsd AΔ 6 ($r_A^+m_A^+$) (see Fig. 3.3), conferred A-specific restriction and modification ability to the ($r_B^-m_B^-$) strain HB101 and encoded three polypeptides with mobilities consistent with those reported for the Eco A polypeptides (Sun, et al., 1984b) (see Fig. 3.9). The polypeptides (Fig. 3.9) encoded by plasmids (Fig. 3.8) carrying different segments of the hsd A genes established the order of the coding sequences for the three polypeptides. The analysis of $r_A^-m_A^+$ deletion derivatives of hsd A (see section b and Fig. 3.3) indicated that the hsdR gene is located in the leftmost segment of the hsd A region. This gene was shown previously, by Suri et al. (1984b), to encode the largest of the three hsd A polypeptides.

pFFP26 carried the leftmost segment and only encodes this polypeptide (see Figs. 3.8 and 3.9). Transformation of WA2552 (a $r_A^-m_A^+$ derivative of WA2899) with pFFP26 produced a strain which was restriction-proficient, confirming the presence of a functional hsdR gene in this plasmid and correlating this gene with the large polypeptide. The production of a restriction-proficient strain was shown to be due to complementation between the subunits of the enzyme rather than to recombination as, on curing the strain of pFFP26, it reverted to its original ($r_A^-m_A^+$) phenotype. (Curing was effected by replacing the pFFP26, in the WA2552 host, with pBR322. This was carried out by
transforming the tetracycline-sensitive, pFFP26-carrying strains with pBR322 and selecting for tetracycline-resistant colonies. Due to incompatibility, pBR322 could only be established in a cell if it replaced its derivative pFFP26).

pFFP31 (Fig. 3.8) conferred A-modification but not restriction, ability to HB101, indicating that this plasmid carries the genes required for modification. pFFP31 encodes the two smaller polypeptides (see Fig. 3.9), which were shown by Suri et al. (1984b) to comprise the modification enzyme. The smaller of these is encoded by the gene next to hsdR (pFFP32 in Figs. 3.8 and 3.9) and the remaining polypeptide by the rightmost gene (pFFP19 in Figs. 3.8 and 3.9). In the case of hsd K, the gene next to hsdR encodes the polypeptide of intermediate size while the rightmost gene encodes the smallest polypeptide (Sain and Murray, 1980). This could either imply that the order of the genes encoding the modification enzyme subunits is reversed in hsd A or that these subunits are not analogous, as regards size, to those of hsd K.

As shown above, pFFP19 and pFFP31 respectively encode the intermediate and small subunits of the modification enzyme; therefore, strains carrying each of these two plasmids provide an ideal alternative to deletion derivatives of λhsd phages for complementation tests. These are described in the following section.
3.2.e. Identification of hsdM and hsdS

The modification methylases of EcoK and EcoB are composed of two polypeptides encoded by the genes hsdM and hsdS, where hsdS alone is responsible for sequence specificity. Mutations in either hsdM or hsdS result in a modification-deficient phenotype. A distinction between hsdM and hsdS mutants requires complementation tests between systems conferring different specificities, in which case each functional hsdS gene imparts a distinctive phenotype, i.e. a system-specific modification (see Section 3.1). It follows that, if the two subunits of the EcoA methylase (Suri et al., 1984b) are analogous to those of EcoK, complementation between modification-deficient mutants of the A and E systems would correlate one of the complementation groups of each system with a specificity gene. This, of course, depends on each mutation inactivating only one of the genes required for modification.

Modification-deficient deletion mutants, such as those shown in Figure 3.2, have the advantage that they are readily located on a physical map. However, such deletions frequently inactivate more than one gene, in some cases simply because they remove a promoter sequence common to more than one gene. All λhsd A deletions resulting in modification-deficient phages were found to inactivate both modification genes and thus were not suitable for complementation tests. Point mutants were therefore needed in the tests since,
Table 3.5
Interactions between phage and plasmid encoded \hsd\ genes

<table>
<thead>
<tr>
<th>Phage Derivative*</th>
<th>Phenotype</th>
<th>E.o.p. on HB101 carrying plasmids:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda\hsd A)</td>
<td>(m_A^+)</td>
<td>10^{-2} 1 1 1</td>
</tr>
<tr>
<td>(\lambda\hsd A \Delta 6, \Delta 12)</td>
<td>(m_A^+)</td>
<td>10^{-3} 1 1</td>
</tr>
<tr>
<td>(\lambda\hsd A \Delta 4, \Delta 8)</td>
<td>(m_A^+)</td>
<td>10^{-3} 1 1</td>
</tr>
<tr>
<td>(\lambda\hsd A \Delta 2, \Delta 21)</td>
<td>(m_A^-)</td>
<td>1 1 1</td>
</tr>
<tr>
<td>(\lambda\hsd A 1512, 3375)</td>
<td>(m_A^-)</td>
<td>10^{-2} 1 1</td>
</tr>
<tr>
<td>(\lambda\hsd A 1042, 2648)</td>
<td>(m_A^-)</td>
<td>1 1 1</td>
</tr>
</tbody>
</table>

* Deletion mutants (see Fig. 3.3) are given the symbol \(\Delta\), other numbers refer to point mutants isolated after growth on the \mutD5\ strain RP526.

** Figure 3.8 shows pFFP21 and pFFP32.
unless polar in nature, they should inactivate only one
gene. M Odification-deficient point mutants were derived
from \( r^-m^+ \) deletion derivative \( \lambda hsd \AA 8 \) (see Fig. 3.3)
following growth in the mutD5 host RP526.* Mutants,
recognised as phages that were restricted by WA2899, were
obtained at a frequency of \( 10^{-3} \) and were readily divided
into two classes by various types of complementation tests.
(Four such mutants were isolated).

In the first test, the mutants were grown in HB101
carrying the plasmid pFFP21 (see Fig. 3.8). This plasmid,
as seen in Figure 3.9, encodes the hsdR subunit and the
smaller of the two modification subunits. The \( \lambda hsd \)
mutant phages were divided into two classes on the basis
of their response; one class grew as well as on a control
strain (HB101 carrying pBR322), the other showed a greatly
reduced efficiency of plating (see Table 3.5). These
results can be understood if the phages showing a reduced
e.o.p. encoded the component missing from the plasmid,
thereby producing a functional enzyme. The resulting
restriction endonuclease would degrade the unmodified DNA
of both host and phage genomes, leading to unproductive
infection. To test this hypothesis, representatives of

* The mutD5 gene product (or lack of it) is active in DNA
replication and gives a phenotype showing a high mutation
rate. This suggested that it acts at the replication fork
allowing the incorporation of incorrect base-pairs
(Fowler et al., 1974; see Cox, 1976). Scheuermann et al.
(1983) provided evidence that the mutD5 mutation affects
dnaQ, the gene product of which is the \( \xi \) subunit of
DNA polymerase III, hence its involvement in DNA
replication.
the three classes of deletion derivatives shown in Figure 3.3 and the original $\lambda$hsd A phage were grown in HB101 carrying pFFP21. Modification-proficient phages showed a low e.o.p. while modification-deficient ones (which had inactivated both modification genes; see section b, above) plated normally (see Table 3.5). This indicated that restriction of the host DNA was occurring rather than simply restriction of the infecting phage's DNA. Phages showing a low e.o.p. were therefore referred to as "kill". ($\lambda$hsd E deletion and mutant derivatives, isolated by G. M. Cowan, showed a similar "killing" effect; Fuller-Pace et al., 1985). Tests using the plasmid PBSllARvb (Suri and Bickle, 1985), which has the same genetic content as pFFP21 and also encodes the same polypeptides, gave the same results indicating that this effect is due to the genes carried by these plasmids and not a property peculiar to the pFFP21 construction. The dependence of the killing effect on the presence of the plasmid-encoded restriction subunit was tested by growing the same phages on HB101 carrying the plasmid pFFP32 (Fig. 3.8). This plasmid differs from pFFP21 in that it lacks the DNA from the extreme left of the hsd A region, assuming the orientation shown in Figure 3.8, and therefore does not encode the restriction subunit (see Fig. 3.9). All the phages plated normally on this strain (see Table 3.5) confirming the need for a functional hsdR gene in the plasmid.
Table 3.6

Efficiency of plating of \( \lambda \hsd \) phages on plasmid-carrying strains, when both phage and host chromosome are A-modified.

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Phenotype</th>
<th>E.o.p. on bacterial strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda \hsd A )</td>
<td>( m_A^+ )</td>
<td>0.5</td>
</tr>
<tr>
<td>( \lambda \hsd A \Delta 6, \Delta 12, )</td>
<td>( m_A^+ )</td>
<td>0.7</td>
</tr>
<tr>
<td>( \lambda \hsd A \Delta 4, \Delta 8 )</td>
<td>( m_A^+ )</td>
<td>0.65</td>
</tr>
<tr>
<td>( \Lambda \hsd A \Delta 2, \Delta 21 )</td>
<td>( m_A^- )</td>
<td>1</td>
</tr>
<tr>
<td>( \Lambda \hsd A 1512, 3375 )</td>
<td>( m_A^- )</td>
<td>1</td>
</tr>
<tr>
<td>( \lambda \hsd A 1042, 2648 )</td>
<td>( m_A^- )</td>
<td>1</td>
</tr>
<tr>
<td>( \lambda \hsd E 25, 2118^* )</td>
<td>( m_E^- )</td>
<td>( 10^{-4} )</td>
</tr>
<tr>
<td>( \lambda \hsd E 11, 1447^* )</td>
<td>( m_E^- )</td>
<td>1</td>
</tr>
</tbody>
</table>

* hsds E point mutants isolated by G. M. Cowan.
25 and 2118 were kill\(^+\), 11 and 1447 were kill\(^-\)
(from Fuller-Pace et al., 1985)

All phage lysates were assayed on the \( r_A^+ \) strain WA2899 and shown to be fully modified.
The most likely explanation is that kill\(^+\) phages retain, while the plasmid pFFP21 lacks, a functional specificity gene. To test this the r\(_{A}^{-}\)m\(_{A}^{+}\) strain WA2552 was transformed with the plasmid pFFP21. In this situation the host chromosome is already modified against the EcoA restriction endonuclease. All \(\lambda\)hsd\(_{A}\) modification-deficient deletion derivatives and point mutants to be used were propagated on WA2552 to protect their DNA against A-specific restriction. These phages were then tested on WA2552 carrying pFFP21 and were all found to plate normally (see Table 3.6). Killing of the WA2552/pFFP21 strain would however be expected if the phages entering the cell encode a specificity other than that of A and complement the plasmid-encoded functions to give a restriction endonuclease of different specificity. This was the case when A-modified \(\lambda\)hsd\(_{E}\) kill\(^+\) phages were used to infect the A-modified strain carrying pFFP21 (see Table 3.6), indicating that the killing effect was due to the incoming, different, specificity provided by the \(\lambda\)hsd\(_{E}\) kill\(^+\) phages. Complementation with the plasmid-encoded subunits produced an E-specific restriction endonuclease which degraded the host's A-modified DNA. This result, therefore, showed that kill\(^+\) phages are hsd\(_{S}^{+}\) and that pFFP21 is hsd\(_{S}^{-}\). Since the plasmid pFFP21 does not encode the subunit of intermediate size (see Fig. 3.9), this polypeptide is identified as the A specificity subunit.

This was confirmed by intra- and inter-specific complementation tests, involving both \(\lambda\)hsd\(_{A}\) and \(\lambda\)hsd\(_{E}\)
Table 3.7
Intra- and inter-specific phage/phage complementation tests

<table>
<thead>
<tr>
<th>λhsd mutants**</th>
<th>E.o.p.* on Bacterial strains:</th>
<th>NM531</th>
<th>NM522/pBR322</th>
<th>WA2899</th>
<th>NM522/pGC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A kill⁺</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>A kill⁻</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>E kill⁺</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>E kill⁻</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td>10⁻⁵</td>
<td></td>
</tr>
<tr>
<td>A kill⁺. A kill⁺</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>A kill⁻. A kill⁻</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A kill⁺. A kill⁻</td>
<td></td>
<td>1</td>
<td>6 x 10⁻²‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E kill⁺. E kill⁺</td>
<td></td>
<td>1</td>
<td>-</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>E kill⁻. E kill⁻</td>
<td></td>
<td>1</td>
<td>-</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>E kill⁺. E kill⁻</td>
<td></td>
<td>1</td>
<td>-</td>
<td>10⁻¹‡</td>
<td></td>
</tr>
<tr>
<td>A kill⁺. E kill⁺</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>A kill⁻. E kill⁻</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>A kill⁺. E kill⁻</td>
<td></td>
<td>1</td>
<td>10⁻¹‡</td>
<td>10⁻⁴</td>
<td></td>
</tr>
<tr>
<td>A kill⁻. E kill⁺</td>
<td></td>
<td>1</td>
<td>10⁻³</td>
<td>5 x 10⁻¹‡</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.7 (continued)

* E.o.p. after one cycle in the non-modifying \textit{recA} strain NM531. The E-restricting strain used was NM522 carrying the plasmid pGCl, which encodes a functional EcoE restriction endonuclease (Fuller-Pace \textit{et al.}, 1985). Therefore non-restricting control strains included both NM531 and NM522 carrying the plasmid pBR322, which was the vector used in the construction of pGCl.

** $\lambda$\textit{hsd} mutants refer to modification deficient point mutants; A kill$^+$ is $\lambda$\textit{hsd} A kill$^+$ point mutant etc. For $\lambda$\textit{hsd} A phages mutant numbers 1512 and 1375 were used as kill$^+$; numbers 1042 and 2648 were used as kill$^-$. In the case of $\lambda$\textit{hsd} E phages (isolated by G. M. Cowan) numbers 25 and 1118 were kill$^+$ mutants; numbers 11 and 1447 were kill$^-$ ones (Fuller-Pace \textit{et al.}, 1985).

*** - not tested

‡ Complementation is shown by an increased e.o.p. on the appropriate restricting strain. In inter-specific tests this indicated that for both $\lambda$\textit{hsd} A and $\lambda$\textit{hsd} E phages, the kill$^+$ mutant carried a functional specificity gene.
modification-deficient mutants, which relied on these phages becoming modified as a result of complementation. These tests were effected either by coinfection of recA− cells (NM531) with two modification-deficient λhsd phages, or by propagating a modification-deficient λhsd phage in a recA− host (HB101) carrying plasmids encoding one, or the other, of the subunits of the modification enzyme. In each case specific modification was assessed by protection against the appropriate restriction system. For inter-specific tests, protection against both the A and the E restriction systems was assayed.

Complementation tests employing co-infection showed that, in both intra- and inter-specific tests, kill+ phages complemented kill− phages but not kill+ ones. In inter-specific tests, phages modified as a result of complementation acquired the specificity of the kill+ phage, indicating that kill+ phages have a functional hsdS gene (see Table 3.7).

The genetic content of the two plasmids each of which encoded one subunit of the A methylase (see Fig. 3.9) was identified by complementation tests. One plasmid, pFFP32, encoding the small polypeptide, spans the central hsd gene; the other, pFFP19, encoding the polypeptide of intermediate size, covers the gene remote from hsdR (see Figs. 3.8 and 3.9). Mutant λhsd A phages referred to as kill−, and presumed to be hsdS−, were complemented by pFFP19, whereas those identified as kill+ and presumed to be hsdM−S+, were complemented by pFFP32 (see Table 3.8).
### Table 3.8

Intra- and inter-specific phage/plasmid complementation tests

<table>
<thead>
<tr>
<th>λhsd phages*</th>
<th>Host**</th>
<th>WA2899</th>
<th>NM522/pGCl***</th>
<th>Specificity of modification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$r_A^{+}m_A^{+}$</td>
<td>$r_E^{+}m_E^{+}$</td>
<td></td>
</tr>
<tr>
<td>A kill$^+$</td>
<td>pBR322</td>
<td>$10^{-3}$</td>
<td>N/T$^+$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pFFP32</td>
<td>$10^{-1}$</td>
<td>N/T</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>pFFP19</td>
<td>$10^{-3}$</td>
<td>N/T</td>
<td>-</td>
</tr>
<tr>
<td>A kill$^-$</td>
<td>pBR322</td>
<td>$10^{-3}$</td>
<td>N/T</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pFFP32</td>
<td>$10^{-3}$</td>
<td>N/T</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pFFP19</td>
<td>$10^{-1}$</td>
<td>N/T</td>
<td>A</td>
</tr>
<tr>
<td>E kill$^+$</td>
<td>pBR322</td>
<td>$10^{-2}$</td>
<td>$10^{-3}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pFFP32</td>
<td>$10^{-2}$</td>
<td>$10^{-1}$</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>pFFP19</td>
<td>$10^{-2}$</td>
<td>$10^{-3}$</td>
<td>-</td>
</tr>
<tr>
<td>E kill$^-$</td>
<td>pBR322</td>
<td>$10^{-2}$</td>
<td>$10^{-3}$-$10^{-4}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pFFP32</td>
<td>$10^{-2}$</td>
<td>$10^{-3}$-$10^{-4}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pFFP19</td>
<td>0.5-1</td>
<td>$10^{-3}$-$10^{-4}$</td>
<td>A</td>
</tr>
<tr>
<td>E kill$^+$ (λhsd E Δ6)</td>
<td>pBR322</td>
<td>$10^{-3}$</td>
<td>0.4</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>pFFP32</td>
<td>$10^{-3}$</td>
<td>0.5</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>pFFP19</td>
<td>$10^{-1}$</td>
<td>0.4</td>
<td>A+E</td>
</tr>
</tbody>
</table>
Table 3.8 (continued)

* For $\lambda hsd$ A mutant numbers 1512, 3375 were used as kill$^+$ (presumed $hsdM^-$) phages; numbers 1042, 2648 were used as kill$^-$ (presumed $hsdS^-$) phages. In the case of $\lambda hsd$ E phages (isolated by G. M. Cowan), numbers 25 and 1118 were kill$^+$ mutants; numbers 11 and 1447 were kill$^-$ ones. $\lambda hsd$ E$\Delta 6$ was an $r_E^+ hsdR^+ M^+ S^+$ phage (Fuller-Pace et al., 1985).

** Hosts in which $\lambda hsd$ phages were propagated prior to titre on the A- and E-restricting strains; the plasmids carried by the $m_A^- m_E^- recA$ hosts are identified. For $\lambda hsd$ A phages the host was HB101, for $\lambda hsd$ E phages it was NM531.

*** NM522 carrying the plasmid pGC1 was used as the E-restricting strain; this plasmid carries the $hsd$ E restriction and modification genes (Fuller-Pace et al., 1985).

† Complementation is shown by an increased e.o.p. on the relevant restricting strain; the specificity of the acquired modification is given. With intra-specific tests between $\lambda hsd$ A phages and plasmids, resulting phages were not titred on NM522/pGC1 (N/T). $\lambda hsd$ E$\Delta 6$ is already E-modified.
Proof that pFFP19 encodes the \textit{hsdS} polypeptide of the A system, and hence that pFFP32 includes the \textit{hsdM} gene, relied on the acquisition of A-specific modification by \textit{\lambda hsd E} phages propagated in a host carrying pFFP19 (Table 3.8). As predicted, \textit{\lambda hsd E} \textit{kill}\textsuperscript{−} phages (\textit{hsdM}\textsuperscript{+}S\textsuperscript{−}) acquired only A-specific modification, whereas a modification-proficient \textit{\lambda hsd E} phage (\textit{hsdM}\textsuperscript{+}S\textsuperscript{+}) acquired protection against both A- and E-specific restriction (see Table 3.8 and Fuller-Pace \textit{et al.}, 1985).

These results therefore not only correlated the modification genes with the polypeptides they encode, thus identifying the \textit{hsdS} gene; they also confirmed that for \textit{hsd A} the order of the genes is \textit{hsdR}, \textit{hsdM} and \textit{hsdS}, the largest polypeptide being encoded by \textit{hsdR} and the smallest by \textit{hsdM}.

3.2.f. \textbf{Direction of Transcription of \textit{hsd A} Genes}

The direction of transcription of the \textit{hsd} genes in \textit{\lambda hsd A} phages was determined by infecting UV-irradiated cells (M159) and labelling polypeptides with \textsuperscript{35}S-methionine. As shown in Figure 3.10, the \textit{hsd} genes in the HindIII insert of \textit{\lambda hsd A} may be transcribed early from the leftwards \textit{\lambda} promoter \textit{P}\textsubscript{L} and late from the rightwards \textit{\lambda} promoter \textit{P}\textsubscript{R}'\textsuperscript{−}. The transcript from \textit{P}\textsubscript{R}'\textsuperscript{−} not only requires activation by \textit{pQ}, a product of early rightwards transcription from \textit{P}\textsubscript{R}'\textsuperscript{−}, but must traverse more than 20 kb of the intervening \textit{\lambda} genome.

The polypeptides encoded by \textit{\lambda hsd A} phages were labelled either early (3 to 13 minutes) or late (25 to 35 minutes)
after infection and detected by autoradiography after electrophoresis through an SDS-polyacrylamide gel. Two λhsd A phages were used in this experiment; λhsd A with the hsd genes oriented so that the hsdR gene is at the left side of the insert (as in Fig. 3.10) and λhsd A* with the hsd genes in the alternative orientation, i.e. the hsdR gene to the right of the insert. For λhsd A, with hsdR as the leftmost gene, a lag of more than 13 minutes was required before expression of the hsd polypeptides was detected (see Fig. 3.11). This indicates that the genes are oriented so that their transcription awaits activation of the late λ promoter (pR') and moderation of the early promoter (pL); i.e. that they are transcribed in the direction from left to right as shown in Figure 3.10. Transcription from hsd promoters immediately after infection is likely to be opposed by convergent transcription (Ward and Murray, 1979). As expected, for λhsd A*, with the fragment inserted in the alternative orientation, the polypeptides were detected early after infection (Fig. 3.11). In these experiments, in contrast to the minicell experiments, the hsdR polypeptide was difficult to detect and frequently not visible, as is the case in Figure 3.11. However, when the hsdR polypeptide has been detected, the timing of its appearance indicates that all three genes are transcribed in the same direction. It is not clear why the hsdR polypeptide is difficult to detect in cells infected with phage despite the anticipated transcription from the λ promoters. This could reflect
instability of the hsdR polypeptide in UV-irradiated cells, or in the presence of some currently unidentified protein encoded by phage λ.

The hsdR gene is expressed in a λhsd A prophage, as seen from the restricting ability of λhsd A dilysogens (see sections a and b above); this indicates the presence of a promoter within the HindIII insert as λ promoters are repressed in the prophage state. If the hsdR gene is indeed transcribed in the same direction as hsdM and hsdS, a promoter for hsdR would be located to the left of the coding sequence, within the region of the insert that is missing in some Class II deletion derivatives of λhsd A (e.g. Δ4 and Δ8 in Fig. 3.3). The finding that dilysogens of λhsd Δ4 and λhsd Δ8 are modification-proficient (section b, above) requires a promoter between the hsdR and hsdM genes. Such an organisation would parallel that for hsd K (Sain and Murray, 1980). However, the absence of concrete evidence concerning the direction of transcription of the hsdR gene means that divergent transcription from promoter sequences between hsdR and hsdM remains a possibility.

3.2.g. Effect of Ral on Modification by EcoA

The product of an early non-essential gene, ral (restriction alleviation), located between cIII and N, was shown by Zabeau et al. (1980), to specifically alleviate restriction by the Type I enzymes EcoK and EcoB, but not by the Type II enzymes EcoRI and EcoRII or the Type III enzyme EcoPl. It also enhanced modification
### Table 3.9

<table>
<thead>
<tr>
<th>Phage*</th>
<th>Host**</th>
<th>E.o.p. on bacterial strains***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>5K/WA2552</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$r^{-}m^{+}$</td>
</tr>
<tr>
<td>$\lambda_{ral}^{+}$</td>
<td>5K</td>
<td>$(r_{K}^{-}m_{K}^{+})$</td>
</tr>
<tr>
<td>$\lambda_{ral}^{18}$</td>
<td>5K</td>
<td>$(r_{K}^{-}m_{K}^{+})$</td>
</tr>
<tr>
<td>$\lambda_{ral}^{+}$</td>
<td>WA2552</td>
<td>$(r_{A}^{-}m_{A}^{+})$</td>
</tr>
<tr>
<td>$\lambda_{ral}^{18}$</td>
<td>WA2552</td>
<td>$(r_{A}^{-}m_{A}^{+})$</td>
</tr>
</tbody>
</table>

* Phages were previously propagated in NM477 ($r_{K}^{-}m_{K}^{+}$) for the test on EcoK modification and on NM555 ($r_{A}^{-}m_{A}^{+}$) for that on EcoA modification. NM555 is a non-restricting, non-modifying derivative of WA2899 in which the $hsd$ A genes were replaced with $\Delta 2$ (see section b, above) by homologous recombination.

** Hosts in which phages were grown for one cycle (45 minutes) prior to titre on the restricting and non-restricting bacterial strains.

*** Appropriate restricting and non-restricting strains were used for assessing the extent of modification of the $\lambda_{ral}^{+}$ and $\lambda_{ral}^{18}$ ($\lambda_{ral}^{-}$) phages; 5K and C600 for those grown in 5K and WA2552 and WA2899 for those grown in WA2552. As seen from the figures in the table above, $\lambda_{ral}^{+}$ phages show a 50-fold increase in modification by EcoK compared with $\lambda_{ral}^{18}$ ($\lambda_{ral}^{-}$) phages. In the case of EcoA modification appeared to be complete irrespective of the presence or absence of Ral.
by EcoK and EcoB. These effects are presumed to be due to modulation of the restriction and modification activities of the Type I restriction endonucleases studied. The effect of Ral on modification by EcoA was investigated, and compared with that on EcoK modification, by growing ral+ and ral- phages in appropriate r-m+ hosts for a single cycle and subsequently plating on the respective restricting strains. With the K system, ral+ phages showed a 50-fold increase in modification when compared with ral- phages. In contrast, with the A system methylation was completely effective even in the absence of Ral (see Table 3.9). These results do not rule out an enhancement of modification by Ral in the case of the A system. They do, however, suggest that the EcoA methylase activity is much more efficient than that of EcoK.

3.3. DISCUSSION

As described in the introduction to this chapter, the hsd A genes are located in the same region of the E.coli chromosome as those of hsd K (Arber and Wauters-Willems, 1970), apparently as alternative genes or alleles; however, no complementation has been observed between the subunits of the enzymes encoded by the two systems (W. Arber and N. E. Murray, pers. comm.). There is also no apparent DNA homology between the genes of the two systems (Murray et al., 1982). Nevertheless, EcoA appears to be similar to EcoK with respect to its overall structure and its requirements for ATP and AdoMet.
(Lark and Arber, 1970; Suri et al., 1984b), suggesting that EcoA is a Type I restriction endonuclease.

The experiments described in this chapter show that the organisation of the genes encoding EcoA closely resembles that of the archetypal Type I restriction and modification system of E. coli K12 (Sain and Murray, 1980). Three genes, hsdR, hsdM and hsdS encode the restriction endonuclease and a functional methylase is encoded by two of these genes (hsdM and hsdS), one of which, hsdS, confers the specificity of DNA recognition. The presence of three genes, two of which are required for modification, was implied from the finding that purified EcoA restriction endonuclease, like EcoK, comprises three polypeptides and that two of these polypeptides form a functional methylase (Suri et al., 1984b).

Initial analysis of deletion derivatives of λhsd A suggested that, as for hsd K (Sain and Murray, 1980), there are two promoters in hsd A, one from which the restriction (and perhaps the modification genes) can be expressed and one located between the hsdR and hsdM genes, allowing the expression of modification without restriction. This could provide the means for a transcriptionally determined sequential expression of first the methylase and then the complete complex with endonucleolytic activity. However, as for hsd K, there is no evidence to indicate such sequential transcription. Analysis of polypeptides produced by λhsd A phages showed that the hsdM and hsdS genes, and possibly hsdR, are transcribed in the same direction,
although the evidence concerning the direction of transcription of $\text{hsdR}$ was inconclusive. Suri and Bickle (1985) by deletion analysis of a pBR322 derivative carrying the $\text{hsdA}$ genes, showed that the three genes are transcribed in the same direction, from left to right in the orientation shown in Figures 3.1 and 3.10. The presence of a promoter to the left of $\text{hsdR}$ was indicated by the finding that deletions extending from the $\text{hsdM}$ gene into $\text{hsdR}$ resulted in the production of a truncated $\text{hsdR}$ polypeptide.

A family of restriction and modification systems, alternative to that of $\text{hsdK}$, was identified as a result of the cloning of $\text{hsdE}$, which although related to $\text{hsdA}$, recognises a different DNA sequence. This concept of families was based upon DNA homology and the interchangeability of subunits of the restriction endonucleases encoded by the respective systems. Complementation tests between modification-deficient mutants of $\text{hsdA}$ and $\text{hsdE}$, analogous to those described for $\text{hsdK}$ and $\text{hsdB}$ (Boyer and Roulland-Dussoix, 1969) were carried out, although in the case of $\text{hsdA}$ and $\text{hsdE}$ the relevant genes were carried on $\lambda\text{hsd}$ phages. These tests confirmed that restriction and modification in $\text{hsdA}$ and $\text{hsdE}$ is encoded by three genes, two of which provide functional modification, and that the overall organisation of these systems is similar to that of $\text{hsdK}$. Other complementation tests, between modification-deficient point mutants of $\lambda\text{hsdA}$ and $\lambda\text{hsdE}$ and plasmids encoding one, or the other, of the two modification polypeptides of $\text{EcoA}$ identified the specificity
gene, hsdS, and showed that the order of the genes in both hsd A and hsd E is hsdR, hsdM and hsdS.

The size of the three polypeptides encoded by EcoA (see Fig. 3.9) correlate well with those given by Suri et al. (1984b). The EcoA and EcoE polypeptides show some heterogeneity, as shown in Figure 3.9, but this is more pronounced when they are compared with those of the K family (see Bickle, 1982). For both families, the hsdR polypeptide is by far the largest, but that of EcoA (98,000) is considerably smaller than that of EcoK (135,000). In addition, the sizes of the other two polypeptides are reversed, with the smallest polypeptide of EcoK being the product of hsdS (Sain and Murray, 1980), and the smallest for EcoA and EcoE being the product of hsdM.

The A modification system differs from those of E.coli K12 and E.coli B in the efficiency with which it methylates unmodified DNA in vivo. While modification of phage in r_k^-m_k^+ hosts was very much enhanced by the presence of Ral, complete modification, even in the absence of Ral, was detected for A-specific modification by an r_a^-m_a^+ host. This corroborates the in vitro findings for EcoA, which indicated that, unlike EcoK, EcoA methylates unmodified DNA efficiently (Suri and Bickle, 1985).

The fundamental similarities and differences between the K and A families pose several questions concerning the origins of these systems. Although the
organisation of the genes and the overall structure and cofactor requirements of the enzymes encoded by these systems are similar, there are some basic differences in the reaction mechanisms, as shown by the differences in the modification efficiency. In addition, the lack of complementation between the subunits encoded by the K and A systems and the absence of DNA homology and immunological cross-reactivity separates them from the K-related systems (W. Arber, pers. comm.; Murray et al., 1982). Therefore the origin of the two families remains unclear. The differences between the systems and the lack of DNA homology would tend to suggest independent origins for the two families, or divergent evolution with the implication that these systems are phylogenetically ancient, but as yet no DNA sequence data are available to support this concept. If the K and A families are of independent origin, their apparently common location in the _E.coli_ chromosome is unexplained.
Figure 3.1

(a) 

<table>
<thead>
<tr>
<th>kb</th>
<th>λcI857</th>
<th>λhsd A</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) 

<table>
<thead>
<tr>
<th>λhsd A</th>
<th>Hin</th>
<th>Sal</th>
<th>Eco</th>
<th>Eco</th>
<th>Bam</th>
<th>Bam</th>
<th>Eco</th>
<th>Eco</th>
<th>Sal</th>
<th>Sal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>2.3</td>
<td>0.1</td>
<td>1.5</td>
<td>0.8</td>
<td>2.1</td>
<td>3.0</td>
<td>1.9</td>
<td>1.405</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.1

Restriction analysis of $\lambda\text{hsd A}$

(a) $\text{EcoRI, HindIII, BamHI and SalI}$ digests of $\lambda\text{hsd A}$. $\lambda c I857$ was used as a marker; the sizes of fragments produced by digestion of $\lambda c I857$ with $\text{EcoRI, HindIII and BamHI}$ have been estimated (Haggerty and Schleif, 1976; Philippsen et al., 1978) and some are indicated.

(b) Restriction map of the 15kb $\text{HindIII}$ insert in $\lambda\text{hsd A}$; the sizes of various restriction fragments are indicated.

(Hin, HindIII; Eco, EcoRI; Bam, BamHI; Sal, SalI.)
Figure 3.2

Analysis of λhsd A deletion derivatives

The upper line shows a restriction map of the 15 kb HindIII fragment cloned in the original λhsd A phages. Deletion derivatives of λhsd A, Δ1 to Δ20, and their phenotypes are shown; gaps indicate the DNA deleted. (Deletion end points have not been accurately determined but in each case are between the flanking restriction sites.) Eco, EcoRI; Hin, HindIII; Bam, BamHI; Sal, SalI.
Figure 3.3
Phenotypic classes of $\lambda_{hsd}$ A deletion derivatives

The upper line shows a restriction map of the 15 kb HindIII fragment cloned in the original $\lambda_{hsd}$ A phages. Representatives of the three phenotypic classes of deletion derivatives are shown; gaps indicate the DNA deleted. (Deletion end points have not been accurately determined but in each case are between the flanking restriction sites.) Eco, EcoRI, Hin, HindIII; Bam, BamHI; Sal, SalI. See text for relevance of Bam$^+$ and Eco*. 
Figure 3.3

Class I  \( r^+ m^+ \)
- \( \Delta 6 \)
- \( \Delta 12 \)

Class II \( r^- m^+ \)
- \( \Delta 4 \)
- \( \Delta 8 \)

Class III \( r^- m^- \)
- \( \Delta 2 \)
- \( \Delta 21 \)

---

1 kb
Figure 3.4

(a) pBR322

(b) pFFP20
Survey of *E. coli* strains for the presence of *hsd A*-related genes

The results obtained from hybridisation of an *hsd A* probe with EcoRI digests of *E. coli* strains are shown. Consistent results (data not shown) were obtained with HindIII digests of DNA from these strains. The probes used were as follows: (a) pBR322; (b) pFFP20-*hsd A*. All isolates, except for 4247* have been classified as *E. coli* by K. Cartwright or B. R. Levin (see bacterial strains; Chapter 2.1). After digestion and electrophoresis through an agarose gel, the DNAs were transferred to a nitrocellulose filter prior to hybridisation. Only half of the strains tested are indicated in this figure; the remainder, none of which hybridised with the *hsd A* probe, are shown in Figure 4.2. pK13, a derivative of pBR322 which forms monomers and oligomers of 2.15 kb unit size (Bouché et al., 1982) was used as a marker; marker sizes are given in kb.

The controls were as follows:

- **C600** - *hsd K* (*E. coli* K12)
- **15T** - *hsd A* (natural isolate carrying A specificity)
- **WA2899** - *hsd A* (strain used in cloning of *hsd A* genes)
- **E166** - *hsd D* (K-related system)
- **C** - no known system (*E. coli* C-la)

(contd.)
Figure 3.5

(a) 

<table>
<thead>
<tr>
<th></th>
<th>Eco</th>
<th></th>
<th></th>
<th>Eco</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WA2899</td>
<td>A58</td>
<td>A101</td>
<td>WA2899</td>
<td>A58</td>
<td>A101</td>
<td>WA2899</td>
</tr>
</tbody>
</table>

kb

6.45

4.3

2.15

(b) 

\( \lambda \ hsd \ A \)

hsdR (hsdM hsdS)

pFFP21

pFFP19

--- 1 kb
The identification of restriction fragments including the \textit{hsd} regions from A58 and A101.

(a) The patterns obtained from hybridisation of EcoRI, HindIII and BamHI digests of DNA from the bacterial strains WA2899 (\textit{hsd A}), A58 and A101 with pFFP21, which covers most of the \textit{hsd A} region, are shown. Similar patterns were obtained with the overlapping probe pFFP19 (see b) (data not shown), except that in the WA2899 EcoRI digest the doublet band of about 4kb was resolved into a single band and, in the A101 EcoRI digest, the smaller (2kb) fragment does not hybridise. The positions of markers corresponding to the sizes of pK13 monomers and oligomers are indicated (see Fig. 3.4).

WA2899 was used as the \textit{hsd A} control, instead of 15T\textsuperscript{−}, to avoid complication by the pBR322-homologous bands which normally appear with 15T\textsuperscript{−} (see Fig. 3.4). However, some faint bands appear in the EcoRI and BamHI digests; these are unexplained. The hybridisation of the \textit{hsd A} probe with A101 is considerably stronger than that with A58, suggesting more homology between \textit{hsd A} and A101. In the EcoRI digests, three bands appear with A58 and two with A101, implying the presence of EcoRI sites within the \textit{hsd} regions of these strains. In the BamHI digests, two bands of equal intensity appear with A58, corresponding to fragments of about 3kb and 6kb, while, with A101,
Figure 3.5 (contd.)

a large fragment hybridises strongly, but there is also a weak band, corresponding to a very small fragment; this again suggests the presence of BamHI sites in the hsd regions of A58 and A101.

In the case of HindIII, no faint bands appear and only one fragment hybridises with hsd A in WA2899, A58 and A101. In all three strains, the fragments were of comparable sizes; therefore HindIII was used for making libraries of A58 and A101 in λ. Subsequent digests of λhsd phages derived from A58 and A101 indicated that some weak bands were not due to DNA from within the respective hsd regions (see Figs. 3.6 and 3.7).

(b) The location of the DNA in the plasmids pFFP19 and pFFP21, corresponding to the hsd region in the 15kb insert of λhsd A is indicated. These plasmids cover the entire hsd A region and include some flanking sequences.
Figure 3.6

1  \(\lambda cI857\) HindIII
2  \(\lambda hsd\ A\)
3  \(\lambda hsd\ E\) EcoRI
4  \(\lambda hsd101\)
5  \(\lambda hsd\ A\)
6  \(\lambda hsd\ E\) HindIII
7  \(\lambda hsd101\)
8  \(\lambda hsd\ A\)
9  \(\lambda hsd\ E\) BamHI
10 \(\lambda hsd101\)
11 \(\lambda hsd\ A\)
12 \(\lambda hsd\ E\) EcoRI/HindIII
13 \(\lambda hsd101\)
14 \(\lambda hsd\ A\)
15 \(\lambda hsd\ E\) EcoRI/BamHI
16 \(\lambda hsd101\)
17 \(\lambda hsd\ A\)
18 \(\lambda hsd\ E\) HindIII/BamHI
19 \(\lambda hsd101\)
20 \(\lambda cI857\) HindIII
Figure 3.6

Restriction analysis of $\lambda$hsd A and phages carrying A-related hsd genes

Restriction digests of $\lambda$hsd A, $\lambda$hsd 101 and $\lambda$hsd E, using EcoRI (Eco), HindIII (Hin) and BamHI (Bam) are shown. HindIII digests of $\lambda cI857$ are used as markers; the sizes of these have been estimated (Philippsen et al., 1978) and are indicated.

There are similarities in the restriction patterns of $\lambda$hsd A101 (derived from bacterial strain A101) and $\lambda$hsd A; these are highlighted in the restriction maps (Fig. 3.7). $\lambda$hsd E is derived from strain A58 and has a different specificity.
Restriction maps of \( \lambda hsd \) A and phages carrying A-related \( hsd \) genes

The upper line shows a restriction map of the 15 kb \( \text{HindIII} \) fragment in \( \lambda hsd \) A. pFFP20 refers to the plasmid used as probe to identify A-related \( hsd \) systems by homology. The \( hsd \) DNA contained in this plasmid is defined by a continuous line. A restriction map of the 11 kb \( \text{HindIII} \) insert in \( \lambda hsd \) A101* indicates the similarities between this region and that in \( \lambda hsd \) A; this phage has A specificity. For \( \lambda hsd \) E (derived from the bacterial strain A58), the map of the 16 kb \( \text{HindIII} \) insert is, as expected, markedly different from that of \( \lambda hsd \) A. pGC1 refers to the plasmid which carries the entire \( hsd \) E region, the DNA contained in pGC1 being defined by a continuous line (Fuller-Pace et al., 1985).

(Eco, EcoRI; Hin, HindIII; Bam, BamHI.)

* \( \lambda hsd \) A101 refers to the \( \lambda hsd \) phages derived from the bacterial strain A101.
Figure 3.8

Subclones of the hsd A region in pBR322

The upper line shows a restriction map of the 15 kb fragment cloned in the original λhsd A phages. (Eco, EcoRI; Hin, HindIII; Bam, BamHI; Sal, SalI.) Segments from the HindIII fragment of λhsd A subcloned in pBR322 are defined by the respective continuous lines. The hsd genes within each plasmid are listed, but these were deduced from complementation tests described later. pFFP30 contains the HindIII fragment from Δ6 (Fig. 3.3). The orientation of the EcoRI fragment relative to pBR322 in pFFP19 is such that the gene in this fragment can be transcribed anticlockwise from the promoter P1, identified by Stüber and Bujard (1981). (See text for relevance of Eco⁺ and Bam⁺.)

All plasmids were used to transform WL542 to identify the polypeptides produced by the various fragments.
Figure 3.9

<table>
<thead>
<tr>
<th>Mr</th>
<th>Markers</th>
<th>pBR322</th>
<th>pFPF30</th>
<th>pFPF21</th>
<th>pFPF31</th>
<th>pFPF26</th>
<th>pFPF32</th>
<th>pFPF19</th>
<th>pFPF30</th>
<th>pGC1</th>
</tr>
</thead>
<tbody>
<tr>
<td>94k</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>67k</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>43k</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>30k</td>
<td></td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.9

Autoradiographic analysis of $^{35}$S-labelled polypeptides produced by subclones of the hsd A region

All hsd fragments were subcloned in pBR322 (see Fig. 3.8) and transferred to the minicell strain WL542. Newly synthesised proteins were labelled for 1 hour and separated by electrophoresis through a 10% SDS polyacrylamide gel. The leftmost track shows standard proteins which were stained with Coomassie blue and their positions marked with radioactive ink:

- Phosphorylase b, $M_r$ 94,000 (94k); Bovine Serum Albumin, $M_r$ 67,000; Ovalbumin, $M_r$ 43,000; Carbonic Anhydrase, $M_r$ 30,000. pFFP30 has the entire hsd A region, which specifies 3 polypeptides with mobilities corresponding to molecular weights of 98,000, 64,000 and 55,000 (Suri et al., 1984b). All the hsd A derivatives which include the hsdR gene encode the largest polypeptide ($M_r$ 98,000). pFFP21 also includes the next gene and encodes the smallest polypeptide ($M_r$ 55,000). This polypeptide, together with the third ($M_r$ 64,000), is seen in pFFP31 which codes for the methylase. The extra band seen for pFFP31 is encoded by DNA downstream of the hsd region (data not shown).

In various complementation tests described in the text pFFP26, pFFP32 and pFFP19 are shown to include hsdR, hsdM, and hsdS respectively, thus correlating polypeptides with genes and establishing the gene order (see Fig. 3.8).

/contd.
Figure 3.9 (contd.)

pGCl, which carries the entire hsd E region is included for comparison. hsd E also encodes three polypeptides but with mobilities readily distinguishable from those of hsd A.
Figure 3.10

λ promoters in the λhsd A phages

A map of the linear genome of λhsd A indicates the position of the insert relative to the λ promoters, Pr', Pr, and Pp; the orientation of the 15 kb insert is as shown in Figure 3.1. Transcription occurs from circular molecules; open arrows indicate leftwards transcription, solid arrows rightwards transcription. RMS relate to hsd genes in the HindIII (Hin) insert.
Figure 3.11
Autoradiographic analysis of polypeptides specified by \( \lambda \)hsd A phages labelled early (3 - 13 minutes) and late (25-35 minutes) after infection

Polypeptides were separated by electrophoresis through a 10% SDS polyacrylamide gel. Arrows identify the hsd polypeptides. pFFP30 tracks provide markers for the three hsd A polypeptides. \( \lambda \)hsd A has the HindIII insert oriented as in Figure 3.1. \( \lambda \)hsd A* has the insert in the opposite orientation. (Both are in the vector NM761.) For \( \lambda \)hsd A, a lag in excess of 13 minutes was required before hsd A polypeptides were detected. In contrast, for \( \lambda \)hsd A*, synthesis of these polypeptides was detected before 13 minutes. In \( \lambda \)hsd A*, therefore, the hsd genes would be transcribed from \( \delta_L \) with hsdR the first to be transcribed. Nevertheless, in the phage experiment shown, the hsdR polypeptide is not visible. On those occasions when the hsdR polypeptide has been seen, it is a weak band but its appearance is consistent with transcription of the hsdR gene being in the same direction as hsdM and hsdS. The vector \( \lambda \)NM761 itself does not encode any polypeptides with the same mobilities as those of EcoA. An extra polypeptide encoded by the \( \lambda \)hsd A phage, transcribed in the opposite direction from the hsd genes, has a mobility consistent with the extra one seen for pFFP31 (see Fig. 3.9).
CHAPTER 4

THE DISTRIBUTION OF TYPE I RESTRICTION
AND MODIFICATION SYSTEMS AMONG ENTEROBACTERIA
4.1 INTRODUCTION

As a result of the intense search for restriction endonucleases as tools in recombinant DNA research, almost every major group of gram-negative and gram-positive bacteria has been found to have at least one genus from which these enzymes can be isolated. Three different types of restriction endonucleases have been identified and, although Type I were the first to be studied, most of the enzymes purified to date are those of Type II that cut within specific DNA sequences. The diversity of these enzymes, particularly in terms of the DNA sequences they recognise is immense.

It is generally accepted that restriction endonucleases function to protect the cell from the integration or expression of foreign DNA (Arber, 1965a). Alternatively, or additionally, they may control gene flow between different populations of bacteria, by acting as a barrier to the transfer of DNA between different strains; this suggestion is supported by the finding that the classification of many Salmonella serotypes according to the hsd systems they carry correlates well with that obtained by standard taxonomical methods (Bullas et al., 1980).

Since 1970, the main interest in restriction endonucleases has been their use as specific DNA cleavage reagents and consequently, although most of the enzymes isolated have been of Type II, the systems themselves
have not been subjected to detailed analysis. In many cases, it is not known whether these endonucleases are involved in host-controlled restriction and modification \textit{in vivo}. In contrast, with Type I enzymes, interest has focused on their mechanism of action and on their genetic determinants. The standard laboratory strain \textit{E.coli} K12 has been extensively studied and has a chromosomally-determined Type I system; analysis of this and the system encoded by \textit{E.coli} B have contributed most of the present information about Type I systems and the enzymes they encode.

The \textit{E.coli} K12 and B strains have been found to carry one chromosomally-encoded Type I system, but some strains, e.g. \textit{E.coli} 15T\textsuperscript{-}, have, in addition, a plasmid-encoded system which is usually Type II or Type III (e.g. \textit{hsd} A and P15 in \textit{E.coli} 15T\textsuperscript{-}; Arber and Wauters-Willems, 1970; Reiser and Yuan, 1977) and others have only plasmid-encoded systems. Only a few \textit{E.coli} strains have been shown to carry Type I restriction and modification systems, but it is not known whether these systems are prevalent among members of the species. The situation in \textit{Salmonella} is already known to be considerably more complicated; several serotypes have two or three chromosomally-encoded \textit{hsd} systems, active on various phages. \textit{S.typhimurium} carries the LT and SA systems which so far have only been found in \textit{Salmonella} and also the K-related SB system which is allelic to \textit{hsd} K (Bullas et al., 1980). Therefore, a
particular kind of Type I system is not necessarily confined to one bacterial species. In *Salmonella*, K-related systems have always been found to map in the same region of the chromosome as those in *E. coli*, whereas the unrelated systems map in other regions. However, in *E. coli* 15T, the chromosomally-determined *hsd* A system, although mapping in the same region as that of *E. coli* K12, has not shown any relatedness to *hsd* K by genetic and molecular criteria (W. Arber, pers. comm.; Murray et al., 1982). No restriction and modification system has ever been detected in the commonly used isolate of *E. coli*, *E. coli* C, suggesting that these enzymes do not fulfil an essential role in the natural environment. In addition, NM477, a derivative of *E. coli* K12 in which the *hsdM* and *hsdS* genes were deleted is quite viable in the laboratory (Murray et al., 1982), implying that the loss of host specificity is not detrimental to the cell.

The *E. coli* systems, *hsd* K, *hsd* B and *hsd* D, and those of *S. typhimurium* (*hsd* SB) and *S. potsdam* (*hsd* SP) represent a closely related family as seen from genetic experiments (Boyer and Roulland-Dussoix, 1969; Hubacek and Glover, 1970; Van Pel and Colson, 1974; Bullas and Colson, 1975a) and from DNA cross-hybridisation and cross-reaction experiments (Murray et al., 1982), suggesting that they share a common ancestor. The *hsd* A system, on the basis of present evidence, does not appear to be related to the K family, although it maps in the
same region of the chromosome and, physiologically, appears to be of Type I. This suggested that hsd A represents a new family of Type I restriction and modification systems on the basis of experimentally defined tests for DNA homology and interchangeability of subunits; this was confirmed by the isolation of hsd E.

To date, with one exception (HindI from Haemophilus influenzae; Gromkova and Goodgal, 1976), all known Type I restriction and modification systems have been found in Enterobacteriaceae. This may have an, as yet, unclear physiological significance, or may simply reflect the fact that in the last decade there has been a bias towards Type II enzymes. The Type I systems reported in bacteria have been mainly confined to strains of E.coli and Salmonella; this is probably due to the fact that the most extensive genetic studies have been carried out on these species. Klebsiella pneumoniae, which is thought to be closely related to E.coli, also carries a restriction and modification system (Streicher et al., 1974) although it is not known if this system is of Type I or whether it is chromosomally- or plasmid-encoded. There have been no reports of hsd systems in members of the species Shigella and Citrobacter, which are, taxonomically closely related to E.coli (Borman et al., 1944; Kauffmann and Edwards, 1952).

While extensive studies of the distribution of restriction and modification systems among many Salmonella serotypes have been carried out (Bullas et al.,
1980), such a comprehensive survey of other enterobacteria
has not been reported. This chapter initiates a survey
of various species of enterobacteria for the presence
of hsd systems sharing DNA homology with hsd K or hsd A.
This includes several natural isolates of E.coli,
strains of Shigella, Citrobacter and Klebsiella
pneumoniae. Such an investigation should provide an
insight into the distribution and prevalence of hsd K-
and hsd A- related systems in enterobacteria. The
isolation and analysis of new specificities should also
eventually aid the understanding of the basis of
recognition of DNA sequences by Type I restriction
endonucleases.

4.2 RESULTS

Hybridisation of bacterial DNAs with hsd K and
hsd A probes

DNAs from natural isolates of E.coli, Citrobacter,
Klebsiella, Salmonella and Shigella, together with DNA
from E.coli K (C600) and E.coli 15T as controls, were
digested with EcoRI and, after separation by electro-
phoresis through agarose gels, the DNA fragments were
transferred to nitrocellulose filters. Duplicate filters
were prepared and hybridised with $^{32}$P-labelled pBR322 and
subsequently with $^{32}$P-labelled probes from within the hsd K
(pRH1; Sain and Murray, 1980) and the hsd A (pFFP20;
see Chapter 3, Section 2.c) regions. The prior hybridisation
with pBR322 identified fragments sharing homology with
## Table 4.1
Isolates hybridising with hsd K and hsd A probes

<table>
<thead>
<tr>
<th>Isolate*</th>
<th>Hybridisation with:</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hsd K</td>
<td>hsd A</td>
</tr>
<tr>
<td></td>
<td>(pRH1)</td>
<td>(pFPP20)</td>
</tr>
<tr>
<td><strong>E.coli</strong> K12: C600/CR63</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>E.coli B</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>El66</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>BLXA</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>629</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Salmonella 4247</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Salmonella LT2</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>E.coli 1ST⁻</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>WA2899</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>A58</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>A101</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Citrobacter freundii</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>E.coli</strong> C-la</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* Only the isolates shown shared homology with the hsd K or hsd A probes (see Figs. 4.1, 4.2 and 4.3). E.coli C was included as a control; it shared no homology with either of the probes. Unless otherwise stated, all isolates were E.coli strains. These had been classed as E.coli by Keith Cartwright or Bruce R. Levin (see Bacterial Strains, Chapter 2.)
pBR322 rather than the \textit{hsd} insert. (Both pRH1 and pFFP20 contained \textit{hsd} fragments cloned in pBR322).

The minority of isolates hybridised either with the \textit{hsd} K or \textit{hsd} A probes (see Figs. 4.1, 4.2 and 4.3). These are shown in Table 4.1. Out of the twenty nine natural isolates examined, twenty two hybridised with neither of the probes; these included many \textit{E.coli} and all the \textit{Shigella} strains. As expected, isolates which shared homology with \textit{hsd} K did not hybridise with the \textit{hsd} A probe, confirming the observed lack of relatedness between the two systems (Murray et al., 1982), despite their common chromosomal location and allelism.

The various isolates sharing homology with \textit{hsd} K or \textit{hsd} A showed different patterns of hybridisation. With the \textit{hsd} A probe (pFFP20) (Fig. 4.1c), a fragment of approximately the same size as the 4.3kb pK13 marker hybridised for both WA2899 and \textit{E.coli} 15T\textsuperscript{−}; both these carry \textit{hsd} A (WA2899 is a derivative of \textit{E.coli} K carrying \textit{hsd} A genes). The two extra bands in \textit{E.coli} 15T\textsuperscript{−} were due to hybridisation with pBR322 (see Fig. 4.1). In the \textit{E.coli} strains A58 and A101 and in \textit{Citrobacter freundii} fragments of sizes different from those in WA2899 and \textit{E.coli} 15T\textsuperscript{−} hybridised with pFFP20. A58 carries a different specificity, \textit{hsd} E, but A101 was subsequently shown to have A specificity (see Chapter 3, section 2.c). A101, therefore, might be expected to show the same hybridisation pattern as \textit{E.coli} 15T\textsuperscript{−}, or at least have the \textit{hsd} A \textit{EcoRI} fragment present in both \textit{E.coli} 15T\textsuperscript{−} and WA2899.
However, the difference in hybridisation pattern was found to be due to the absence of an EcoRI site outside the hsd region; within the hsd region itself the restriction map is the same for both λhsd A and λhsd A101 (see Figs. 3.6 and 3.7). C.freundii, which also appears to share homology with hsd A, may carry yet another specificity; confirmation of this, however, awaits cloning of the relevant DNA and studies comparable to those carried out with hsd A and hsd E (see Chapter 3). As C.freundii is resistant to λ, simple biological tests for restriction of phage by this strain are not possible.

The Salmonella strain 4247, which carries LT, SA and SB specificities, did not show any hybridisation to the hsd A probe, indicating that the SA system present in this strain shares no homology with hsd A (see Fig. 4.1).

With the hsd K probe (pRH1), the situation was more complicated. There are several known members of the K family, each showing a different hybridisation pattern; therefore the patterns obtained with other natural isolates were compared with the known ones. The strains sharing homology with hsd K included E.coli B, El66 (hsd D; Murray et al., 1982), BLXA, 629 and the Salmonella strains 4247 and LT2. (C600 and CR63 were used as the E.coli K12 controls; CR63 is a λ-resistant derivative of E.coli K12 and has K specificity.) Two independent natural isolates of E.coli showed the same hybridisation pattern as E.coli B, namely 629 and BLXA. The hsd region from 629 has been cloned in λ and shown to provide protection
against B-specific restriction (Gough and Murray, 1983). BLXA was infected with \( \lambda^\text{bla} \) (\( \beta \)-lactamase) phages; ampicillin-resistant, lysogenic, colonies were isolated and, on induction of these lysogens, the released phages were found to be modified with B specificity (N. E. Murray, pers. comm.). This finding corroborated results from tests involving comparison of the electrophoretic mobility of proteins produced by BLXA with those produced by E. coli B, which showed that BLXA was identical to E. coli B for nineteen of the twenty proteins examined (Selander and Levin, 1980; B. R. Levin, pers. comm.).

The Salmonella strains 4247 and LT2 are different isolates of S. typhimurium and both should carry the hsd K-related SB system. In both cases, a large fragment hybridised with pRH1; this correlates with that obtained for the SB system (Murray et al., 1982). However, an extra small fragment hybridised in LT2; this was unexpected as the evidence to date indicates that these two isolates, being both S. typhimurium should be identical as regards K-related hsd systems (Bullas et al., 1980; Murray et al., 1982). The other systems present in S. typhimurium, namely LT and SA, do not share homology with hsd K (Murray et al., 1982). The difference in the hybridisation patterns of the 4247 and LT2 strains may reflect the presence of an EcoRI site in the SB region of the LT2 strain as no other K-related systems have been identified in this strain in spite of extensive tests (Bullas et al., 1980). Further hybridisations with
pRH1, using DNA of both 4247 and LT2 digested with a variety of restriction endonucleases should determine the extent of any differences between the SB hsd regions of the two strains.

4.3 DISCUSSION

Among the various enterobacterial strains screened, relatively few appear to have hsd K- or hsd A-related genes according to the hybridisation results. This implies either that they lack hsd systems, or that they possess systems which are apparently related to neither hsd K nor hsd A on the basis of the tests available. A survey of several Salmonella serotypes showed that although most serotypes possess Type I restriction and modification systems, a considerable proportion carried the LT and SA systems but none related to hsd K (Bullas et al., 1980). Salmonella 4247 which carries the LT and SA systems, in conjunction with the K-related SB, shares no homology with hsd A, as seen from the hybridisation results (see Fig. 4.1). This indicates that neither LT nor SA are related to hsd A on this basis. By analogy with Salmonella, it is possible that E.coli and other enterobacterial strains possess restriction and modification systems which are unrelated to the known ones. It is generally difficult to test for the presence of these systems biologically as most natural bacterial isolates are resistant to λ. Therefore, the identification of Type I restriction and modification
systems has relied either on the transfer of these genes to a phage-sensitive bacterial strain, as in the case of the *Salmonella* SB system (Colson and Van Pel, 1974), or on the detection of DNA homology with known *hsd* genes.

The idea that restriction systems, unrelated to any of the known ones, may be present in the bacterial strains examined is supported by the observation that *K. pneumoniae* and the *E.coli* strain RM74A, which are reported to possess, as yet unclassified, restriction and modification systems (Streicher *et al.*, 1974; N. E. Murray, pers. comm.), do not share DNA homology with *hsd* K or *hsd* A (see Figs. 4.1 and 4.3). RM74A is one of the few natural isolates found to plate λ and, in biological tests was shown to carry two restriction and modification specificities; one is the same as *EcoRII*, the other, as yet unidentified, also restricts λ (N. E. Murray, pers. comm.). Attempts to clone the RM74A unknown system "X" in λ, using *EcoRI, HindIII* and *BamHI* libraries have so far failed (J. A. Gough and F. V. Fuller-Pace, unpublished observations). As seen in Figure 4.1, RM74A does not share any homology with *hsd* K or *hsd* A; therefore the relevant recombinants in a phage library cannot be identified by hybridisation. The alternative method of identifying these recombinant phages, i.e. by testing for X-specific modification, was employed, but, in each case, while phages carrying the *EcoRII*-like system were readily isolated, those with X specificity were not.
Among the Salmonella serotypes investigated by Bullas et al. (1980), restriction and modification phenotypes correlated well with previous taxonomic classification and a biological role for restriction and modification systems, in the preservation of the individual genetic identity of different serotypes was suggested. However, in the case of the E.coli isolates, the situation is less convincing. The two isolates showing the same hybridisation pattern as E.coli B, namely 629 and BLXA, carry B specificity (Gough and Murray, 1983; N. E. Murray, pers. comm.). BLXA has been found to be closely related to E.coli B with respect to the electrophoretic mobility of proteins it produces (19 our of 20 are identical; Selander and Levin, 1980; B. R. Levin, pers. comm.). This would support a correlation between restriction and modification phenotype and other classifications. However, when other E.coli strains are considered, it is seen that this is not a general feature. As regards electrophoretic mobility, the isolates BLD4, BRLET-2 and BRLET-13 were found to be identical to E.coli K12 for most proteins tested; BLD4, 20/20; BRLET-2, 15/15; BRLET-13, 14/15 (Caugant et al., 1981; B. R. Levin, pers. comm.), but none of these shared homology with hsd K (see Figs. 4.1 and 4.2). Therefore no general correlation can be seen in these E.coli strains.

The occurrence of apparently identical bacterial strains in geographically remote locations has been reported
by Selander and Levin (1980) in a survey of the electrophoretic variation of proteins from a wide range of natural isolates of \textit{E.coli}. In some cases, identical clones were isolated from hosts separated by large geographical distances and often recovered at different times. In one instance, the enzymes for a clone isolated from an infant in a hospital nursery were electrophoretically indistinguishable from the standard \textit{E.coli} K12 laboratory strain. A similar situation was found in the distribution of restriction and modification systems in the isolates surveyed in this chapter. The \textit{hsd A} system was found in both \textit{E.coli} 15T\textsuperscript{−} and A101, while the isolates 629 and BLXA were shown to carry B specificity. As the sources of these isolates were different (see bacterial strains, Chapter 2), this finding is of interest in terms of the distribution and evolution of different host specificities.

With respect to \textit{hsd A}-related genes, the presence of homology in \textit{C.freundii} is of interest. Detection of a functional \textit{hsd} system in this strain requires the cloning of the relevant genes and analyses similar to those carried out for \textit{hsd A} and \textit{hsd E} (see Chapter 3). Such a system belonging to the A family would indicate another example of the occurrence of related systems in different species of bacteria. This, of course, is already documented for the K family, of which members have been found both in different strains of \textit{E.coli} and in some serotypes of \textit{Salmonella} (Bullas \textit{et al.}, 1980;
The failure of *K. pneumoniae* and RM74A to hybridise with the *hsd K* and *hsd A* probes provides no further insight into the nature of their restriction and modification systems. The majority of the isolates included in the hybridisation experiments did not share any homology with *hsd K* or *hsd A*, but this does not necessarily mean that they lack host specificity. These isolates may have systems related to LT or SA, which are present in the majority of *Salmonella* serotypes and are in fact considerably more common than those related to *hsd K*. Another possibility is that these natural isolates have other systems unrelated to any known ones. If restriction systems act as barriers to DNA transfer, whether for protection or as a means of speciation, it is perhaps surprising to find that the majority of the strains surveyed lack such systems.

DNA hybridisation and immunological cross-reactivity experiments failed to detect any homology between *hsd K* and *hsd A* (Murray *et al.*, 1982), while complementation tests showed that the subunits of the two systems are not interchangeable (W. Arber and N. E. Murray, pers. comm.). However, without sequence data, it may be premature to claim that they are unrelated. Failure to detect DNA homology by hybridisation may not necessarily indicate a complete lack of homology and, although the concept of families of complex enzymes, the members of which show complementation in the exchange of subunits, is still valid,
it is not possible, without sequence data, to comment on the relationships between different families and their possible ancestry. Even in the presence of sequence information, the consequences of convergent evolution may be difficult to distinguish from the remnants of homology remaining after phylogenetically ancient, divergent evolution.

The apparent common location of \texttt{hsd K} and \texttt{hsd A} encourages caution in concluding whether or not they are related. If they are, then genes encoding related restriction and modification systems in bacterial strains, which so far do not appear to carry host specificity, may reside at the same location on the chromosome. It is not known whether, in strains which apparently have no restriction systems, alternative genes are present in this region. Cloning of the DNA from this region of the chromosome of bacterial strains, which apparently lack host specificity, may answer this question.

More general ways of detecting Type I restriction and modification systems are required. The method used in the survey described in this chapter involved the detection of related systems by hybridisation with probes from the \texttt{hsd} regions of known systems. This, however, is limited as only genes sharing homology with the probes used will be detected. An additional problem with this method is that many bacterial strains have sequences that share homology with pBR322 and these hybridise very strongly
to probes which are derivatives of pBR322 (see Figs. 4.2, 4.3 and 4.4). The same applies to probes containing DNA cloned in \( \lambda \) vectors as bacteria often share homology with \( \lambda \) (J. A. Gough and N. E. Murray, pers. comm.), but it is not known if this is also the case with M13 derivatives. One way of avoiding these problems is by purifying the relevant DNA fragment prior to using it as a probe.

One of the simplest ways of testing for host specificity is by infection of the bacterial cells with phage and estimation of the efficiency of plating, with subsequent testing for modification in such phages. Very few isolates are sensitive to \( \lambda \), but other phages have been used successfully (Bullas et al., 1980). This method however depends on the sensitivity of the cell to the various phages used and also on the presence of targets for the respective restriction endonucleases, in the phage genomes. An alternative biological approach involves transformation of the relevant bacterial strain with the recently isolated plasmid carrying the \( \text{lamB} \) gene (De Vries et al., 1984). Transformation with this plasmid confers a functional phage \( \lambda \) receptor to the bacterial cell, allowing phage adsorption. This however will not necessarily result in a productive infection and therefore does not guarantee that the presence of a restriction and modification system will be detected even if it has targets on the \( \lambda \) genome.

Immunological cross-reactivity tests with antisera prepared against polypeptides encoded by \( \text{hsd} \) systems
have been used, in conjunction with DNA hybridisation experiments, to confirm the relatedness between members of the K family on a molecular basis (Murray et al., 1982) but these have the same shortcomings as DNA hybridisation, in that cross-reaction is only expected to occur between closely related systems.

The detection of endonuclease activity in bacterial extracts has led to the identification of a large number of sequence-specific endonucleases including most of those purified to date. The fundamental property of a restriction endonuclease is its ability to degrade foreign (unmodified) DNA while having no effect on the DNA of the cell of its origin. Various methods have been used to detect such activity in bacterial extracts and purified enzyme preparations; these include viscometry assays (Smith and Wilcox, 1970), measurements of the inactivation of infectious λ DNA (Takano et al., 1966) and electrophoresis of DNA digests through agarose gels (Sharp et al., 1973; see Fig. 4.4). Such approaches could be used for the detection of Type I enzymes in cell-free extracts from a range of natural isolates of bacteria.
Figure 4.1

(a) pBR322

(b) pRH1

(c) pFFP20
The probes used in the hybridisations were as follows: (a) pBR322; (b) pRH1 - hsd K; (c) pFFP 20 - hsd A. All isolates, except for 4247*, have been classified as \textit{E. coli} by K. Cartwright or B. R. Levin (see Bacterial Strains, Chapter 2.1). The bacterial DNAs were digested with \textit{EcoRI} and, after electrophoresis through an agarose gel, transferred to a nitrocellulose filter prior to hybridisation. pK13, a derivative of pBR322 which forms monomers and oligomers of 2.15 kb unit size (Bouché \textit{et al.}, 1982) was used as a marker; marker sizes are given in kb.

The controls were as follows:

- \textit{C600} - hsd K (\textit{E. coli} KL2)
- 15T - hsd A (natural isolate carrying A specificity)
- WA2899 - hsd A (strain used in cloning of hsd A genes)
- \textit{E166} - hsd D (K-related system)
- \textit{C} - no known hsd system (\textit{E. coli} C-1a)

\textit{E. coli} B, BLXA, E166 and 4247* share homology with hsd K; A58 and A101 share homology with hsd A. A58 is hsd E (see Chapter 3.2c).

* 4247 is a \textit{Salmonella typhimurium} derivative carrying the specificity systems LT, SA and SB (Bullas \textit{et al.}, 1980). SB shares homology with hsd K (Murray \textit{et al.}, 1982).
Screening of enterobacterial strains for the presence of hsd K- and hsd A- related genes

The probes used in the hybridisations were as follows: (a) pBR322; (b) pRH1 - hsd K; (c) pFFP20 - hsd A. All isolates except A204* have been classified as E.coli by K. Cartwright and B. R. Levin (see Bacterial Strains, Chapter 2.1). The bacterial DNAs were digested with EcoRI and, after electrophoresis through an agarose gel, transferred to a nitrocellulose filter prior to hybridisation. pK13, a derivative of pBR322 which forms monomers and oligomers of 2.15 kb unit size (Bouché et al., 1982) was used as a marker; marker sizes are given in kb.

The controls were:
15T- - hsd A (natural isolate carrying A specificity)
CR63 - hsd K (λ-resistant derivative of E.coli K12).

629 shares homology with hsd K; no isolates share homology with hsd A. Several strains hybridise strongly with pBR322; this probably reflects the presence of pBR322-related multicopy plasmids (see Fig. 4.4).

* A204 was classified as Citrobacter by K. Cartwright (see Bacterial Strains, Chapter 2.1).

Note: The bands seen in BRLET-2 are due to homology with pBR322 and are evident in (b) and (c) and not in (a) due to the longer exposures used for (b) and (c).
The probes used in the hybridisations were as follows: (a) pBR322; (b) pRHl - hsd K; (c) pFFP20 - hsd A. The strains screened were standard enterobacterial strains. The bacterial DNAs were digested with EcoRI and, after electrophoresis through an agarose gel, transferred to a nitrocellulose filter prior to hybridisation. No pKL13 DNA was available for markers when these experiments were carried out; however the purpose of this experiment was the detection of homology and not necessarily the estimation of the sizes of any fragments hybridising to the respective probes.

Control strains used in this experiment were:

- C600 - hsd K (E.coli K12)
- 15T - hsd A (natural isolate carrying A specificity)
- C - no known hsd system (E.coli C-la)

Salmonella LT2 shares homology with hsd K while Citrobacter freundii appears to share homology with hsd A. For Salmonella LT2, two fragments share homology with pRHl (b), one large fragment of a size comparable to that seen with 4247 (Fig. 4.1) and a smaller one giving faint hybridisation with pRHl. Although the exposure in (b) was considerably longer than that for the pBR322 hybridisation (a), it is unlikely that the faint hybridisation seen in (b) is due to homology with pBR322.
as the exposure used for the hsd A probe (c) was longer, as seen by the intensity of the bands for Shigella and in (c) no such hybridisation is present in the LT2 track. 

The Shigella strains show very strong hybridisation with pBR322. With the pFFP20 (hsd A) probe (c), extra bands appear which are not evident in the shorter exposure for the pBR322 probe (a); however, these may also be due to homology with pBR322 and may be simply a result of the longer exposure. It is therefore difficult to draw any conclusions concerning these strains as any homology they might share with hsd K or hsd A could be easily masked by their strong hybridisation with pBR322.
Figure 4.4

EcoRI restriction digests of DNAs from various strains of E.coli

(a) Examples of EcoRI digests of E.coli DNA are given; C600* indicates undigested DNA. The digested DNAs include those of standard strains: C600 (E.coli K12), B (E.coli B) and various natural isolates.

(b) EcoRI digests of various natural isolates of E.coli are shown. Some very distinct bands are visible in these digests, particularly in those of E161 and E164; these correspond to multicopy plasmids and often hybridise strongly to pBR322 (see Fig. 4.2).
Figure 4.4

(a) C600* RM66A RM74A C600 B 629

(b) 147B E161 E164 E166 E171 581
CHAPTER 5

DNA SEQUENCE OF THE hsdS GENE OF S. POTSDAM AND COMPARISON WITH THAT OF THE E. coli KL2 hsdS GENE
5.1 INTRODUCTION

The concept of a family of closely related enzymes, the members of which carry out essentially the same reactions but recognise different DNA sequences, has generated considerable interest in the evolution and diversification of the different members within particular families of restriction and modification systems.

The family of hsd systems related to that of E.coli K12 has been most widely studied and includes systems found in various strains of E.coli and Salmonella. Genetic experiments involving the K family have shown that the three genes (hsdR, hsdM and hsdS) encoding the restriction endonuclease can be interchanged between different members of the family and that only hsdS confers the specificity of recognition (Boyer and Roulland-Dussoix, 1969; Hubacek and Glover, 1970; Van Pel and Colson, 1974). More recent DNA hybridisation and immunological cross-reaction experiments underlined the close relationship between these systems (Murray et al., 1982). The product of the hsdS gene of a particular system is capable of interacting with the product of the hsdM gene of any other related system to produce a functional methylase and, directly or indirectly, with the hsdR subunit in the formation of an effective restriction endonuclease. There must therefore be a basic similarity between the subunits encoded by different hsdS genes to allow this interaction and, at the
same time, a difference which causes the enzyme to recognise a different DNA sequence.

Arber and Linn (1969) suggested that mutations within the hsdS genes may result in enzymes which carry out essentially the same reaction but on a different DNA sequence. To date there has been no evidence of this occurring naturally or experimentally. Bullas et al. (1976) isolated a new restriction specificity, SQ, from P1 co-transduction experiments between S.potsdam (SP) and E.coli/S.typhimurium (SB) hybrids. SQ had a specificity different from that of SB and SP and Bullas and his colleagues suggested that it had arisen as a result of recombination between the hsdS genes of the SB and SP systems. This was the first and, to date, the only report of a new specificity being generated experimentally. However, no evidence was given at the time to confirm this suggestion.

The hsd systems SB and SP belong to the K family (Van Pel and Colson, 1974; Bullas and Colson, 1975a; Murray et al., 1982). Analysis of the DNA sequences of the hsdS genes of the related E.coli systems hsd K, hsd B and hsd D (Gough and Murray, 1983) showed that the different hsdS genes share little homology. The genes vary in length from 1335 to 1425 base-pairs and the only regions showing obvious homology were about 100 base-pairs in the middle of the gene and about 250 base-pairs at the end encoding the carboxy terminus.
It was suggested that the homologous region at the carboxy terminus of the hsdS polypeptide, which was highly conserved for all three specificities, may be involved in interaction with the hsdM subunit while the remainder of the polypeptide may be involved in DNA sequence recognition. This could correlate the two variable regions of the gene (flanking the central conserved region) with the two specific domains of the recognition sequence. Such a correlation would predict that hsdS genes, in which the variable regions are derived from different genes, would confer different specificities. If the SB and SP hsdS genes had a structure similar to that of the hsdS genes of E.coli K and B, with two variable domains flanking a central homologous region, then SQ could have arisen from recombination between the hsdS genes of SB and SP, at the central conserved region. This was indeed found to be the case from heteroduplex analysis of the hsdS genes of SB, SP and SQ when it was shown that SQ contains one variable domain from SB and the other from SP (Fuller-Pace et al., 1984; see Appendix II, this thesis). The crucial prediction (from the correlation of the hsdS gene domains and those of the recognition sequence of the enzyme) that one region of the DNA recognition sequence for SQ derives from the SB sequence, the other from that of SP, was recently confirmed by Nagaraja et al. (1985c).

The sequence recognised by SP, 5'-AAC(N)_6 GTRC-3'
(Nagaraja et al., 1985b), where N signifies unspecified base-pairs and R either purine, is remarkably similar to that recognised by EcoK: 5′-AAC(N)₆GTGC-3′ (Kan et al., 1979). The only difference between the two sequences is in the specific tetranucleotide domain; the third base is G in the EcoK sequence whereas it can be either purine in the SP sequence. The SP sequence is degenerate, as are the SB and SQ sequences (Nagaraja et al., 1985 b,c). This degeneracy was previously suggested by Bullas et al. (1980) when it was found that modification by SP gave complete protection against restriction by EcoK but not vice versa.

It would be reasonable to expect that the similarity of the recognition sequences of EcoK and SP is a reflection of the similarity between the hsdS genes of the two systems. However, DNA hybridisation studies did not detect any greater homology between these two systems than, for example, between hsd K and hsd B (Murray et al., 1982). Comparison of the DNA sequence of the hsdS gene of hsd SP with that of hsd K (Gough and Murray, 1983) should show the degree of homology shared by these genes and consequently by the polypeptides they encode.

In this chapter, the determination of the DNA sequence of the S.potsdam (SP) hsdS gene is described. The DNA sequences and the corresponding amino acid sequences of the K and SP hsdS genes are compared. The DNA sequence
data should provide information concerning the divergence of the two specificities and it may be possible to localise the base change in the recognition sequence to changes in a specific part of the hsdS gene. Amino acid sequence comparisons may also give an insight into how the two specificity polypeptides recognise differences in the DNA recognition sequence.

5.2 RESULTS
5.2.a. Orientation of the hsdS gene in λhsd SP

Digestion of λhsd phages carrying SB, SP and SQ specificity, with EcoRI, HindIII and BamHI gave the restriction maps shown in Figure 5.1a (Fuller-Pace et al., 1984). The HindIII site marked Hin* is present in λhsd SP and λhsd SQ but is replaced by an EcoRI site (Eco*) in λhsd SB. Results from hybridisation with various hsd K probes and the finding that a 400 bp probe spanning the central conserved region of the hsdS gene of E.coli K12 hybridised to a region to the right of the targets marked * (Fig. 5.1a), suggested that the variable HindIII/EcoRI sites in SB, SP and SQ identified the distal region of non-homology within the hsdS gene (Fuller-Pace et al., 1984). However, subsequent hybridisation of λhsd SB and λhsd SP DNA, digested with HindIII and BamHI, with a previously unavailable probe from within the hsdM gene of E.coli K12 (Fig. 5.1b), localised the hsdM gene of the SB and SP systems to the left of the
BamHI site marked Bam in Figure 5.1a. (The E.coli K12 hsdM probe was kindly provided by Anne S. Daniel.) This indicated that the orientation of the hsdM and hsdS genes in the SB, SP and SQ systems was not as previously thought and that the variable HindIII/EcoRI targets presumably identified the proximal variable region of the hsdS gene. Therefore SQ derives the proximal variable region from SP. The presence of a HindIII site in the variable region, proximal to the hsdM gene, is reminiscent of the hsd region of E.coli K12 which not only has a HindIII site in this region of the hsdS gene, but also has a BamHI site upstream of this gene, the distance between these two targets being approximately the same (Gough and Murray, 1983). The hsdS gene in λhsd SP was therefore presumed to be in this region of the 11kb insert and possibly completely within the 5.1kb BamHI fragment (Fig. 5.1a). This fragment was cloned in pBR322, giving the plasmid pFFP24 (see Fig. 5.2) which was subsequently used as a source of SP DNA for sequencing. The 2.7kb HindIII fragment was also cloned in pBR322 (plasmid pFFP28 in Figure 5.2); this provided a probe covering the 0.7 kb BamHI-HindIII fragment, which was later found to be useful in identifying M13 clones containing DNA from this region of the SP hsdS gene.
5.2.b. Cloning in M13

DNA fragments from the 5.1kb BamHI insert of pFFP24 were obtained in a variety of ways and cloned in Mpl0 and Mpl1; these included sonication, restriction with AluI and Sau3a and direct cloning of BamHI-HindIII fragments in these M13 vectors.

(i) Sonication

The plasmid pFFP24 was sonicated as described in Methods, section d.i and the resulting fragments were cloned in Mpl0 which had been restricted with SmaI. White plaques, the majority of which should be recombinants, were hybridised with pBR322 and pFFP24 probes and those which hybridised with both probes, i.e. which contained DNA from the pBR322 vector rather than the SP insert, were discarded. Initially 700 white plaques were examined; of those 80 hybridised with both pBR322 and pFFP24 while 70 hybridised with pFFP24 only. The remainder shared no homology with either of the plasmids, implying an 80% background of non-recombinant white plaques. This was much higher than expected; a 10% background, generally due to deletions in the M13 vector, is more usual (Heather E. Houston, pers. comm.). Subsequent cloning of fragments derived from the same SP DNA and generated by restriction with AluI or Sau3a and also from another sonication procedure, gave a similar background of white plaques (see below), suggesting that
some feature of this DNA is tending to cause deletions in the M13 vector upon insertion.
(No white plaques were obtained when the vector alone was cut and religated.) When 10 of these M13 clones were sequenced they were found to be deletion derivatives of the M13 vector. The phages which hybridised only to pFFP24 were plated to give single plaques and again hybridised with the above probes; this ensured that only recombinants containing SP DNA were selected. Out of the 70 recombinants originally selected only 47 hybridised to pFFP24 in the second hybridisation.

(ii) AluI and Sau3a fragments

For this procedure pFFP24 was restricted with BamHI and after electrophoresis through an agarose gel the BamHI insert, containing SP DNA, was separated by electroelution. This DNA was subsequently digested with AluI and Sau3a respectively and the fragments were cloned in Mpl0 which had been restricted with SmaI in the case of AluI fragments and BamHI in the case of Sau3a fragments. Separation of the insert from the pBR322 DNA, prior to restriction with AluI and Sau3a, ensured that the majority of the recombinant M13 phages obtained contained SP DNA. White plaques were purified and hybridised with pBR322 and pFFP24 as described above for clones containing sonicated DNA fragments.
Although there was again a high background, most of the recombinants were found to contain SP DNA.

(iii) Direct cloning of BamHI-HindIII fragments in Mpl0 and Mpl1

The orientation of the BamHI and HindIII cloning sites in Mpl0 and Mpl1 allow the insert to be sequenced in the direction HindIII to BamHI and BamHI to HindIII respectively (Messing, 1983). Thus BamHI-HindIII fragments cloned in these vectors provided DNA sequence data for positive and negative strands as shown in Figure 5.2: SP-27, SP-28 and SP-29. These data were a useful starting point for the computer sequence data base with which sequences originating from "random" fragments could be overlapped.

5.2.c. Sequencing Strategy

The Mpl0 recombinants containing DNA fragments generated by sonication and by AluI or Sau3a restriction were localised to particular regions within the 5.1kb BamHI fragment by hybridisation with pFFP28 and pBgH probes (see Fig. 5.2). pBgH covers the central conserved region of the E.coli K12 hsdS gene and hybridises to an area to the right of, and presumably near to, the HindIII site marked Hin* (Fuller-Pace et al., 1984). Thus the DNA fragments initially sequenced could be limited to those from the region thought to contain the hsdS gene. All recombinants containing sonicated DNA
fragments and 100 each of those with AluI and Sau3a inserts were hybridised with pFFP28 and pBglII and divided into groups according to whether they hybridised with pFFP28, pBglII or with both probes. 80 recombinants, including the three "hybridisation groups" were then sequenced and the data were compiled and compared, by computer, with the sequence data of the _E. coli_ K12 hsdS gene (Gough and Murray, 1983). This comparison was carried out on the assumption that the similarity of the recognition sequences of the SP and _hsd K_ systems (Nagaraja et al., 1985b) reflects certain similarities between the DNA sequences of the respective _hsdS_ genes. This should further aid the localisation of individual stretches of SP DNA sequences, especially in cases where the fragments sequenced did not overlap. This initial analysis gave a contiguous sequence of approximately 1000 bp, which by analogy with the _hsd K_ sequence should cover the proximal variable region and the central conserved region of the SP hsdS gene (see Fig. 5.3.a,b). This section of the SP sequence showed a remarkable similarity to that of _hsd K_; however, the only other homology between the two sets of sequence data was at the distal conserved region, i.e. at the carboxy terminus of the gene (see Fig. 5.3.a,b). The finding that all the sequences covering the proximal section of the gene gave the positive strand (except for SP-28 which resulted from the cloning of the HindIII-BamHI fragment in Mpl1; see Figure 5.2) and those covering the distal region gave only the negative strand (see
Fig. 5.3.b) suggested that there was a bias towards particular strands in the recovery of DNA fragments. M13 recombinants containing segments of these regions were therefore used as probes to select other M13 clones which, when sequenced, would give the complementary strand. The probes used were SP-9 and SP-25 as depicted in Figure 5.3.c.

A further 50 Mpl0 recombinants, which had previously been localised to this region of the hsdS gene by hybridisation with pFFP28 and pBgh (see above) were hybridised with SP-9 and SP-25 respectively, using the M13 strand-specific hybridisation method described in Methods, section d(iii). Several recombinants hybridised to both probes and subsequent analysis of the DNA of the inserts showed that they were co-clones containing non-contiguous DNA fragments. This problem was particularly common in the recombinants containing AluI fragments. This hybridisation procedure, failed to obtain many sequences for complementary strands. With respect to the proximal region of the gene, 48 M13 clones, which had specifically only hybridised with pFFP28 and were therefore confined to the 0.7kb BamHI-HindIII fragment (see Fig. 5.2), were sequenced; none gave the negative strand (except SP-28). The negative strand of this section of the gene was therefore sequenced using synthetic oligonucleotide primers, P1* and P2* (prepared by Eric Kawashima of Biogen, Geneva) and SP-28, which carries the complete 0.7kb BamHI-HindIII fragment in Mpl0 (Fig. 5.2), as template.
The DNA sequences of P1* and P2* are indicated in Appendix I.

The DNA sequence of some M13 recombinants which contained SP DNA and hybridised with the hsd K probe, pBgh (see Fig. 5.3.a), shared no significant homology with that of the hsdS gene of E.coli K12 and did not overlap with other SP sequences. There are two possibilities concerning the origin of these recombinants; they could contain very large inserts so that the section for which the DNA sequence was determined was downstream of the hsdS gene, or they could indicate a lack of homology in the distal variable regions of the hsdS genes of hsd SP and hsd K.

The M13 probe SP-9 (Fig. 5.3.c) was found to hybridise with single-stranded M13 recombinants, which had previously been shown to contain only DNA from within the 0.7kb BamHI-HindIII fragment (Figs. 5.2 and 5.3), suggesting that SP-9 extends beyond the HindIII site. This was confirmed by subsequent restriction analysis of SP-9 RF DNA. Very little sequence data was available for the distal variable region of the SP hsdS gene and even, with pBgh and SP-9 as probes (Fig. 5.3) no other recombinants giving this region of sequence were obtained. Therefore, SP-9 (RF) was used as a source of DNA covering specifically this segment of the gene, with the aim of subcloning fragments in M13 for sequencing. SP-9 was generated by the insertion of a Sau3a fragment in the BamHI site of Mpl0. There is only one HindIII target in the SP insert (see Figs. 5.2 and 5.3); this and the orientation of the
compiled as before. Although there was a high frequency of recombinants containing inserts from the ends of the sonicated DNA fragments, they were still sufficiently distributed to cover the region of interest. Computer analysis of the collective sequence data in the database showed that, with the addition of the sequences of DNA fragments derived from SP-9, some other stretches of sequence which had previously not matched with hsd K, or overlapped with any SP sequences, now fitted in a contiguous sequence covering the entire SP hsdS gene. However, there was still a strong strand bias; in the case of the distal variable region this was towards the negative strand.

Although a contiguous DNA sequence covering the hsdS gene had been obtained, there were still two regions where only data for one strand were available. These were between bp 1070 and 1120 where only data for the negative strand were available and between bp 1510 and the end of the gene where only the sequence of the positive strand was known. The gap in the positive strand between bp 1070 and 1120 was filled using a synthetic oligonucleotide primer (P3*) and templates SP-3 and SP-10 (Fig. 5.4). The recombinant SP-27, in which the 4.4 kb HindIII-BamHI fragment had been cloned (Figs. 5.1.a and 5.4) was initially used, but the primer, P3*, did not bind, perhaps due to a deletion in the relevant region of the SP-27 insert. It was not possible to obtain sequence data for the distal end of the gene from SP-9 subclones, as it was beyond the
segment covered by SP-9. The 4.4 kb HindIII-BamHI fragment (Fig. 5.1a) was therefore cloned in Mpl9 (SP-251; Fig. 5.4) so that the DNA sequence would be read in the direction BamHI to HindIII, as depicted in Figure 5.3.d. A further synthetic oligonucleotide primer (P4*) was then used to cover the last remaining gap in the negative strand at the distal end of the gene. The DNA sequences of the primers P3* and P4* are shown in Appendix I.

As the sonicated fragments from SP-9 extended to the HindIII site and SP-28 started at the same site, there was no overlap at this point in the negative strand. Analysis of the sequence data, however, indicated the presence of a PstI site approximately 160 bp to the right of the HindIII site (Fig. 5.4). This PstI site was used to make PstI deletions in SP-9, removing the DNA between this site and the PstI target in the Mpl0 vector (see Fig. 5.3.d). This gave the recombinant SP-250 (Fig. 5.4) which allowed the negative strand sequence overlapping the HindIII site to be determined.

The DNA sequence of certain G-C-rich regions was difficult to read and there were discrepancies between the sequences obtained for the positive strand and those for the negative strand. These were resolved by carrying out dITP reactions in which inosine was substituted for guanosine in dGTP (see Methods, section d(iv)). These regions included those from bp 140 to 160, bp 745 to 760 and bp 1575 to 1590.
Once the complete, double-stranded, DNA sequence was confirmed it was compiled and analysed by computer. A contiguous sequence of 1731 bp was obtained, extending from the BamHI site, past the HindIII site, to an AluI target downstream of the hsdS gene (Appendix I). An open reading frame was found extending from bp 226 to bp 1617 including the stop codon; this gave a polypeptide of 463 amino acids (Fig. 5.5), one amino acid smaller than that of the hsdS gene of E.coli K12 (Gough and Murray, 1983). The sequence data for SP therefore included about 200 bp upstream of hsdS, covering the distal end of hsdM by analogy with hsd K (A. S. Daniel and W. Loenen, pers. comm.), the complete hsdS gene and about 100 bp downstream of hsdS. Comparison of this sequence data with the respective data for hsd K showed that, within the hsdS gene, there is a striking homology in the central and distal regions which are conserved when the hsd K, hsd B and hsd D genes are compared (Gough and Murray, 1983), but, in addition, the proximal region, which is variable in the E.coli hsdS genes, is conserved when those of hsd SP and hsd K are considered. The sequences corresponding to regions upstream and downstream of the SP and K hsdS genes also shared extensive homology (see Fig. 5.6). Comparison of the amino acid sequences of the polypeptides encoded by the hsdS genes of SP and K, as expected, gave the same overall result, with very similar sequences covering the proximal variable domain and the central and distal conserved regions, but very little homology between the
sequences corresponding to the distal variable domain (see Fig. 5.7). Dot matrix comparisons of the nucleotide and amino acid sequence comparisons are given in Figures 5.8 and 5.9.

5.3 Discussion

The DNA sequences of the E.coli K-related \textit{hsdS} genes (Gough and Murray, 1983) are relatively non-homologous, but their overall organisation is uniform and the sequences upstream and downstream of the \textit{hsdS} genes are conserved.

The DNA sequence data available for \textit{hsd} SP (Fig. 5.5 and Appendix I) include sequences upstream (within the \textit{hsdM} gene) and downstream of the \textit{hsdS} gene. Comparison of these data with those for \textit{hsd} K (Gough and Murray, 1983) indicated that these sequences share significant homology (see Fig. 5.6). The DNA sequence of the \textit{E.coli} K12 \textit{hsdM} gene has recently been determined (A. S. Daniel and W. Loenen, pers. comm.) and it has been shown that the stop codon for the \textit{hsdM} gene overlaps with the start codon for the \textit{hsdS} gene as follows:

\[
\begin{array}{c}
\text{hsd K} \\
\quad \text{GAATGA} \\
\quad \quad \quad \text{hsdS (start)} \\
\quad \quad \quad \quad \text{hsdM (stop)}
\end{array}
\]

The sequence at the start of the SP \textit{hsdS} gene allows the same overlap:

\[
\begin{array}{c}
\text{hsd SP} \\
\quad \text{GCATGA} \\
\quad \quad \quad \text{hsdS (start)} \\
\quad \quad \quad \quad \text{hsdM (stop)}
\end{array}
\]
This suggests a similar organisation of the hsdM and hsdS genes.

Comparisons of the DNA sequences of the E. coli hsd K, hsd B and hsd D specificity genes (Gough and Murray, 1983) showed that in these genes, which vary in length from 1335-1425 bp, there are only two regions of obvious homology, a central one of about 100 bp and a second, of about 250 bp, at the end encoding the carboxy terminus of the specificity polypeptide. These regions are highly conserved in the three hsdS genes studied, but the remainder of each gene shares little, or no, homology with either of the other genes. Heteroduplex analysis of the hsdS genes of the Salmonella hsd systems, SB and SP, indicated the same overall arrangement with variable regions of about 500 bp flanking a central conserved region of about 100 bp (Fuller-Pace et al., 1984).

The DNA sequences recognised by the Salmonella SB, SP and SQ systems have recently been determined (Nagaraja et al., 1985 b, c). One striking feature of these sequences is the similarity between that recognised by SP, i.e. 5'-AAC(N)₆GTRC-3' and that previously determined for EcoK, i.e. 5'-AAC(N)₆GTGC-3' (Kan et al., 1979). These sequences differ by only one bp in the tetranucleotide region, which in the SP sequence is degenerate (R is either purine). Degeneracy was observed in the three Salmonella sequences (Nagaraja et al., 1985). A biological prediction of the similarity between the recognition sequences of hsd SP
and $\text{hsd K}$ comes from the earlier finding that SP modification provides protection against K restriction (Bullas et al., 1980), when a degeneracy in the SP recognition sequence was first suggested. The similarity between the recognition sequences and the overlap in specificities imply a basic similarity between the polypeptides encoded by the K and SP $\text{hsdS}$ genes and, therefore, between the DNA sequences of the $\text{hsdS}$ genes.

The homology between the DNA sequences of the SP and K $\text{hsdS}$ genes is considerably more extensive than that between the E.coli $\text{hsdS}$ genes (Gough and Murray, 1983). The central and distal regions which share homology in $\text{hsd K}$, $\text{hsd B}$ and $\text{hsd D}$ (Gough and Murray, 1983) are conserved but, in addition, the proximal region, which is variable in the E.coli $\text{hsdS}$ genes, is also conserved when $\text{hsd K}$ and $\text{hsd SP}$ are considered (see Fig. 5.6). This suggests a closer relationship between $\text{hsd K}$ and $\text{hsd SP}$ than that between $\text{hsd K}$, $\text{hsd B}$ and $\text{hsd D}$.

Computer analysis of the DNA sequences of the distal variable regions of the SP and K $\text{hsdS}$ genes, however, failed to detect any significant homology; therefore the differences in this region cannot be correlated with simple mutations.

The polypeptides encoded by the SP and K specificity genes are of approximately the same length, that of SP being one amino acid smaller than that of K (see Fig. 5.7). As in the case of the DNA sequences, there is a striking homology, not only in the central and distal
conserved regions (as seen previously for hsd K, hsd B and hsd D; Gough and Murray, 1983), but also in the proximal domain which is variable when hsd K, hsd B and hsd D are considered. A comparison of the amino acid sequences of the hsdS polypeptides, encoded by hsd SP, hsd K, hsd B and hsd D, indicated that the central and distal conserved regions are common for the four polypeptides (see Fig. 5.10).

One interpretation of the finding that the Salmonella specificity, SQ, results from recombination between the SB and SP specificities and derives one variable region from SB, the other from SP (Fuller-Pace et al., 1984), would correlate the variable domains of the hsdS polypeptide with the specific domains of the recognition sequence. This is consistent with the finding that SQ, which derives the proximal variable region from SP (see Section 5.2.a), shares the same trimeric segment of the recognition sequence as seen below:

\[
\begin{align*}
\text{hsd SB} & \quad 5'\text{-G A G N N N N N R T A Y G-3'} \\
\text{hsd SP} & \quad 5'\text{-A A C N N N N N N G T R C -3'} \\
\text{hsd SQ} & \quad 5'\text{-A A C N N N N N N R T A Y G-3'}
\end{align*}
\]

(R is either purine; Y, either pyrimidine; Nagaraja et al., 1985 b,c).

This could correlate the proximal variable region of the hsdS gene with the specific trimeric segment in the recognition sequence, in this case AAC. Alternatively, reassortment of the specific components of the recognition sequence may be due to reassortment of minor differences.
in the central conserved region.

The recognition sequence for hsd K is as follows:

\[ \text{hsd K} \ 5'-A\ A\ C\ N\ N\ N\ N\ N\ G\ T\ G\ C-3' \]

and, by analogy with SP and SQ and presuming a correlation between the variable regions of the genes and the specific segments of the recognition sequences, the trimeric component of the recognition sequence would correspond to the proximal variable region of the polypeptide. This region is well conserved in hsd SP and hsd K (Fig. 5.6) and would support such a correlation. However, the distal variable regions might be expected to share a higher degree of homology, considering that the only change in the tetrameric segment of the recognition sequences is G in hsd K to R (either purine) in hsd SP (see above). Although there is little homology in the distal variable domains of the SP and K hsdS polypeptides, there is a similarity between positions 290 and 306 where the amino acids are as follows (see Fig. 5.7):

<table>
<thead>
<tr>
<th>Position</th>
<th>SP</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>291</td>
<td>D I L V I R V N G S A D L A G</td>
<td></td>
</tr>
<tr>
<td>292</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>293</td>
<td></td>
<td></td>
</tr>
<tr>
<td>294</td>
<td></td>
<td></td>
</tr>
<tr>
<td>295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>297</td>
<td></td>
<td></td>
</tr>
<tr>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>299</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>302</td>
<td></td>
<td></td>
</tr>
<tr>
<td>303</td>
<td></td>
<td></td>
</tr>
<tr>
<td>304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>305</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Vertical lines identify identical amino acids, * shows similar amino acids; see Fig. 5.7 for key to amino acids.)

All the "similar" amino acids, except aspartic acid and glutamic acid (D and E) are hydrophobic. Aspartic acid and glutamic acid are both hydrophilic and negatively
charged. Since this similarity is not found when either the *hsd* SP or *hsd* K specificity polypeptides are compared with those of *hsd* B and *hsd* D (see Fig. 5.10), it could correlate with the close similarity of the recognition sequences for SP and K, but it is, currently, not possible to support this idea. This similarity could be the result of convergent evolution or the remnants of divergent evolution; therefore, it does not provide any information regarding the origin of the two specificities or a possible correlation between the specific components of the recognition sequence and the variable domains of the *hsdS* polypeptide.

An alternative hypothesis has recently been presented by Argos (1985) in which the specific components of the DNA sequences recognised by *hsdS* polypeptides are correlated with the central and distal conserved regions. Argos compared the amino acid sequences of the *hsd* K, *hsd* B and *hsd* D specificity genes and found that, in each of these genes, there was a domain which was repeated in the central and distal conserved regions. When these regions were aligned accordingly, the repeated domains were found to be well conserved in the three polypeptides. Predictions of the secondary structure of these polypeptides delineated two helical regions corresponding to the repeated segments in the central and distal conserved regions. From comparisons of these secondary structures with those of well characterised DNA repressor and activator proteins,
Argos (1985) suggested that hsdS polypeptides act as pseudo-dimers in DNA sequence recognition, with each putative helical region interacting with a specific component of the recognition sequence. (The absence of an exact dimer was correlated with the lack of rotational symmetry in the recognition sequence.) Thus differences in the central and distal conserved regions would result in different recognition sequences.

Argos (1985) also suggested that SQ derives the complete central conserved region from SP, in which case this region would specify the trimeric segment of the recognition sequence, i.e. AAC (see above), which is common for both specificities. This segment is also present in the hsd K recognition sequence and, by analogy with SP and SQ and using Argos's model, the amino acid sequences of the SP and K hsdS polypeptides in the central conserved region might be very similar, or even identical. The SP and K hsdS amino acid sequences can be aligned in the same way as those of hsd K, hsd B and hsd D (Argos, 1985) with the repeated segments corresponding to positions 158 to 221 and 386 to 442 (Fig. 5.10), but the homology between the SP and K sequences does not appear to be greater than that between those of the E.coli hsdS polypeptides. This and the finding that the SP and K hsdS polypeptides share a high degree of homology in the proximal variable region, suggest that the variable regions might also be involved in determining specificity; therefore, both
models are plausible. In the case of SB and SP, it is not known where, in the central conserved region, the recombination event, which generated SQ, occurred. Determination of the DNA sequence of the central conserved regions of the SB, SP and SQ **hsdS** genes may identify the cross-over point. Mutational studies of these regions and of that of **hsd K**, should test Argos's hypothesis as, according to his model, they could result in changes in the trimeric segment of the recognition sequence or alternatively, in the loss of specificity.

The recognition sequences of the six known K-related restriction and modification systems and of **hsd A** have been determined. **hsd SB** and **hsd A** recognise hyphenated DNA sequences which have the same trimeric, but different pentameric and tetrameric segments, respectively, as follows:

\[
\text{hsd SB} \quad 5'\text{-G A G N N N N N R T A Y G-3}
\]
\[
\text{hsd A} \quad 5'\text{-G A G N N N N N N N G T C A-3}
\]

The DNA sequences of their **hsdS** genes are not known. The **hsd A** system appears to be unrelated to the K family on the basis of DNA hybridisation, immunological cross-reactivity and interchangeability of subunits (Murray et al., 1982; W. Arber, pers. comm.). However, if the **hsd A** specificity gene has the same overall structure as those of the K family, with variable and conserved regions, comparison of the DNA sequences of the **hsdS** genes of **hsd A** and **hsd SB** and of the amino acid sequences
of the polypeptides they encode may provide some insight into whether the conserved or the variable regions are involved in DNA sequence recognition.

Comparisons of the predicted secondary structure of the polypeptides encoded by the various hsdS genes, for which the DNA sequence is known, may aid the correlation of polypeptide domains with specific segments of the recognition sequence. It is, however, more likely that the elucidation of the tertiary structure of the polypeptides, dependent on X-ray crystallography and preferably correlated with mutational studies, will be required before the regions involved in DNA sequence recognition can be identified unequivocally. Predictions of the structure of the hsdS polypeptides alone may not be sufficient and the ease of obtaining such critical information cannot be predicted. It will, in the first instance, be limited by the availability of crystals and it is not known, in this case, whether the crystals would require both the hsdM and hsdS subunits. In vivo, the hsdM/hsdS subunit complex is required for AdoMet activation and DNA sequence recognition (Hadi et al., 1975; Bühler and Yuan, 1978; Burckhardt et al., 1981b). The hsdS subunit may change conformation on binding to that encoded by hsdM and it is not yet known whether the hsdS subunit alone is capable of recognising a specific DNA sequence.

The origin and diversification of the various specificities known to date is still unclear. The lack
of apparent DNA sequence homology in the distal "variable" region of the SP and \( \text{K hsdS} \) genes, compared with the conservation of the proximal "variable" region, implies that the diversification of these specificities is not a result of simple mutation, but more likely, due to recombination of one of these genes with an as yet unidentified, but related, specificity gene. The generation of the SQ specificity (Bullas et al., 1976) sets a precedent for this. Such a process of recombination between existing specificity genes provides a limited means of increasing diversification.

An alternative means of diversification could involve the insertion of modules of DNA sequence; if this were the case, some evidence for related DNA sequences in various \( \text{E.coli} \) strains might be expected. Surveys of enterobacterial isolates for the presence of \( \text{hsd K} \)-related sequences (Gough and Murray, 1983) detected \( \text{hsd} \) systems but no other related sequences. Therefore, from present evidence, recombination between existing specificity genes seems a more likely mechanism of specificity diversity. Nevertheless, the evolutionary origin of the diverse variable regions remains unexplained. If restriction and modification systems are phylogenetically ancient, the origin of the various different specificities identified to date may have been obscured by many changes, which could even mask a possible common ancestry of the apparently non-homologous regions in the specificity genes (Gough and Murray, 1983).
Figure 5.1

(a) \( \lambda_{hsd} \text{ SB} \)

\[
\begin{array}{cccccc}
& \text{Hin} & \text{Bam} & \text{Eco} & \text{Hin} \\
\text{2.0} & 0.6 & 4.5 & 3.4 & 0.5 \\
\end{array}
\]

\( \lambda_{hsd} \text{ SP} \)

\[
\begin{array}{cccccc}
& \text{Hin} & \text{Bam} & \text{Eco} & \text{Hin} \\
\text{2.0} & 0.7 & 4.4 & 3.4 & 0.5 \\
\end{array}
\]

\( \lambda_{hsd} \text{ SQ} \)

\[
\begin{array}{cccccc}
& \text{Hin} & \text{Bam} & \text{Eco} & \text{Hin} \\
\text{2.0} & 0.7 & 4.4 & 3.4 & 0.5 \\
\end{array}
\]

(b) i SB SP SB/pUC

\[
\begin{array}{c}
\text{SB} \\
\text{SP} \\
\text{SB/pUC} \\
\end{array}
\]

ii SB SP SB/pUC

\[
\begin{array}{c}
\text{2.7kb} \\
\text{2.0kb} \\
\end{array}
\]
Figure 5.1

Orientation of the hsdS gene in λhsd SB and λhsd SP

(a) Restriction maps of the hsd regions of λhsd SB, λhsd SP and λhsd SQ (from Fuller-Pace et al., 1984). The positions of the EcoRI (Eco), HindIII (Hin) and BamHI (Bam) targets in the 11 kb inserts are indicated. A 400 bp probe spanning the central conserved region of the hsdS gene of E.coli K12 was shown to hybridise to a region to the right of the targets marked *, these targets were presumed to identify the distal region of non-homology (Fuller-Pace et al., 1984).

(b) [i] HindIII/BamHI double digests of λhsd SB (SB) and λhsd SP (SP) and a BamHI digest of a pUC9 derivative containing the 5kb BamHI fragment from the hsd insert in λhsd SB (SB/pUC). (The pUC9 derivative was provided by Annette J. B. Campbell.) pUC9 was used to give a better separation between the vector and the insert. The arrows indicate the 2kb HindIII-BamHI fragments in λhsd SB and λhsd SP and the pUC vector 2.7kb fragment.

[ii] Hybridisations of the DNA digests in (i), after transfer to nitrocellulose filters, with an E.coli K12, hsdM-specific probe in pBR322 (provided by Anne S. Daniel). The 2kb HindIII-BamHI fragments in λhsd SB and λhsd SP hybridise /contd.
Figure 5.1 (contd.)
to the hsdM probe, indicating that the hsdM gene is to the left of hsdS for the orientation shown in (a). In the SB/pUC track, the vector fragment hybridises, due to homology between pUC and the pBR322 vector in the probe. The 5 kb BamHI fragment in SB/pUC and the 4.5 kb/4.4kb fragments in λhsd SB and λhsd SP do not share homology with the hsdM probe. Therefore the EcoRI and HindIII targets, marked *, identify the proximal variable regions of the hsdS genes of SB and SP respectively. The same applies to SQ in which this variable region is derived from SP (Fuller-Pace et al., 1984).
Figure 5.2

\[
\begin{array}{c|c|c|c|c}
\text{Hin} & \text{Bam}^+ & \text{Hin} & \text{Bam} & \text{Eco} \\
2.0 & 0.7 & 4.5 & 3.4 & 0.5
\end{array}
\]

- \( \lambda hsd \ SP \)
- pFFP24
- pFFP28
- Mp10:SP-27
- Mp10:SP-28
- Mp11:SP-29
- pBgH*

---

1 kb
Subclones of the hsd region from λhsd SP

The upper line shows a restriction map of the 11kb insert in λhsd SP. The positions of the EcoRI (Eco), HindIII (Hin) and BamHI (Bam) targets are indicated; the sizes of the fragments are in kb.

Segments from the 11kb insert in λhsd SP were subcloned in pBR322, Mpl0 and Mpl1; these are defined by the respective continuous lines. pFFP24 and pFFP28 are derivatives of pBR322 containing the 5.1kb BamHI and the 2.7kb HindIII fragments respectively. pFFP24 was subsequently used as a source of DNA for subcloning in M13, while pFFP28 was used as a probe to identify M13 recombinants containing DNA derived from the 0.7kb BamHI-HindIII fragment. Mpl0: SP-27, Mpl0: SP-28 and Mpl1: SP-29 identify M13 clones containing respective BamHI-HindIII fragments. The arrows indicate the direction of sequence data obtained from these clones, relative to the cloning sites.

* pBgH is a 400 bp probe spanning the central conserved region of the hsdS gene of E.coli K12 (Murray et al., 1982; Gough and Murray, 1983) and hybridises to a region to the right of the HindIII site marked Hin+. The approximate location of the equivalent region in the SP hsdS gene is indicated by a dotted line. pBgH was used as a probe to identify M13 recombinants containing DNA from this region.
Figure 5.3

Initial sequence data and subsequent strategy for sequencing the SP hsdS gene

(a) The upper line shows a map of the E.coli K12 hsdS gene; certain restriction targets and the location of the start and stop codons are indicated. The shaded areas identify the central and distal regions which are conserved in hsd K, hsd B, and hsd D (Gough and Murray, 1983). The HindIII target marked * is also present in the SP hsdS gene; as indicated in Figure 5.2, the insert in the plasmid pBgH covers the region immediately to the right of this site and has homology with the equivalent region in hsd SP. Due to the homology between the K and SP hsdS genes, the DNA sequence data for the E.coli K12 hsdS gene (Gough and Murray, 1983) was useful in aligning stretches of sequence obtained for SP, especially in the initial stages.

(b) The dotted lines indicate the sequence data obtained after the initial sequencing strategy; M13 recombinants sequenced included those containing AluI, Sau3a and sonicated fragments, and those obtained by direct cloning of HindIII-BamHI fragments in Mpl0 and Mpl1 (see Fig. 5.2).

(c) As there was a very pronounced strand bias in the sequence data (see b), two of the M13 recombinants were used as probes in M13 strand-specific hybridisation /contd.
Figure 5.3

(a) E. coli K12 hsdS

(b) SP hsdS sequence

(c) SP-25

(d) SP-9

sequencing direction

--- 100 bp
Figure 5.3 (contd.)

experiments in an attempt to isolate clones, which would give the sequence of the complementary strand. These probes, namely SP-9 and SP-25 are indicated; the strand for which each provided sequence data is indicated by an arrow, i.e. SP-9 gave the negative strand and SP-25, the positive strand.

(d) The orientation of the insert (shaded) in SP-9 relative to the cloning sites in Mpl0 is indicated. (Mpl0 cloning sites are marked *). The sequencing direction in SP-9 is indicated by an arrow. SP-9 was used to make PstI deletions which allowed the sequence overlapping the HindIII site in the insert to be determined and was also a source of DNA from which a HindIII fragment containing mainly SP DNA was purified and subsequently sonicated (see text).
Figure 5.4
DNA sequence determination of the SP hsdS gene

The upper line shows a map of the 5.1kb BamHI fragment in a pFFP24; this fragment includes the entire SP hsdS gene for which the start and stop codons are indicated. The shaded areas indicate the regions in which the SP DNA sequence is homologous to that of hsd K. These include the central and distal regions which are conserved in hsd K, hsd B and hsd D (Gough and Murray, 1983; see Fig. 5.3).

The M13 recombinants used in determining the DNA sequence of the SP hsdS gene are shown; where several clones were obtained covering the same region of the gene, only representatives are given. The extent of the DNA sequence obtained for each M13 recombinant is indicated by a continuous line; (a) identifies the positive strand, (b) the negative strand. The recombinants used included those containing AluI (Alu), Sau3a (Sau) and sonicated (Son) fragments derived from the 5.1kb BamHI fragment. Recombinants generated by the cloning of HindIII-BamHI from the hsd SP region in M13 are indicated as broken lines and, in each case the segment sequenced is identified by a continuous line; SP-27, SP-28, SP-29 and SP-251. (Only the numbers of the clones are given in this figure.)

* P1, P2, P3 and P4 were the regions sequenced using the synthetic oligonucleotide primers. For P3, SP-3 and

/contd.
Figure 5.4

pFFP24  SP:hsdS

(a) Alu
   Sau
   Son.

(b) Sau
   Son.

--- 100 bp
Figure 5.4 (contd.)

SP-10 were used as templates as SP-27 was found to be unsuitable (see text).

+ SP-250 was obtained by deleting the DNA between the PstI site shown and that in the linker in Mp10 (see text and Fig. 5.3d) and allowed the sequence for the negative strand overlapping the HindIII site to be determined.
Figure 5.5
DNA sequence of the SP hsdS gene

The contiguous DNA sequence of 1731 bp extending from the BamHI site to the AluI site, indicated in Figure 5.4, is shown. This includes the entire hsdS gene for which the amino acid sequence is given. The location of the BamHI, HindIII and PstI targets is shown (see the text for the relevance of these sites). The sequence is numbered consecutively from the 5' to the 3' end.
Figure 5.6

Comparison of the DNA sequences of the \textit{hsdS} genes of \textit{E.coli} K12 and \textit{S.potsdam} (SP)

The DNA sequence obtained for \textit{hsd SP} is compared with the respective sequence for \textit{hsd K} (Gough and Murray, 1983); the upper line refers to \textit{hsd SP}, the lower to \textit{hsd K}. The DNA sequences include the entire \textit{hsdS} genes and extend upstream and downstream and are numbered consecutively from the 5' to the 3' ends. Vertical lines indicate matches; the start and stop codons for the \textit{hsdS} genes are underlined. (Gaps are introduced for alignment.) The sequences are very well conserved between bp 1 and 850 and between bp 1330 and 1731 (the extent of the sequence available for SP); i.e. in the proximal variable region, the central and distal conserved regions of the \textit{E.coli} K12 \textit{hsdS} gene, and regions upstream and downstream of the gene (Gough and Murray, 1983).
Figure 5.7

Comparison of the amino acid sequences of the polypeptides encoded by the hsdS genes of hsd SP and hsd K

The upper line refers to SP, the lower to K; vertical lines indicate matches. The amino acid sequences are very well conserved between positions 1 and 200, and 370 and the end of the gene; these include those of the proximal variable region and the central and distal conserved regions of the E.coli K12 hsdS gene (Gough and Murray, 1983). A gap of one amino acid is introduced between positions 215 and 216 in the SP sequence for alignment.

Key to amino acids:

A = alanine  
B = aspartic acid/asparagine  
C = cysteine  
D = aspartic acid  
E = glutamic acid  
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  
V = valine  
W = tryptophan  
Y = tyrosine  
Z = glutamic acid/glutamine
Figure 5.8
Dot matrix comparisons of the DNA sequences of hsd SP and hsd K

Continuous lines indicate matches; arrows locate the start and stop codons. (See Fig. 5.6 for details.)
Figure 5.9

Dot matrix comparisons of the amino acid sequences of the polypeptides encoded by the \texttt{hsdS} genes of \texttt{hsd SP} and \texttt{hsd K}

Continuous lines indicate matches. (See Fig. 5.7 for details.)
Figure 5.9

Window: 5  Stringency: 3.0  Points: 868  Density: 20.22
1-OCT-85  12:58

hsdsp.pep ck: 180, 1 to 463
Figure 5.10
Comparison of the amino acid sequences of the polypeptides encoded by the hsdS genes of hsd SP, hsd K, hsd B and hsd D

The amino acid sequences of the K, B and D hsdS genes (Gough and Murray, 1983) are compared with that of SP. Capital letters indicate matches; gaps are introduced in the sequences for alignment. As seen in Figure 5.7, the sequences for SP and K are well conserved in the proximal variable domain and the central and distal conserved regions. When these sequences are compared with those for hsd B and hsd D, however, homology is only seen in the central and distal conserved regions identified by Gough and Murray (1983), and no significant homology is seen in the proximal variable domain.

Key to amino acids:

A = alanine
B = aspartic acid/asparagine
C = cysteine
D = aspartic acid
E = glutamic acid
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
V = valine
W = tryptophan
Y = tyrosine
Z = glutamic acid/glutamine
REFERENCES

Bankier, A. T. and Barrell, B. G. (1983), Shotgun DNA Sequencing, Laboratory of Molecular Biology, MRC Centre, Cambridge.
Bullas, L. R. and Colson, C. (1975b), Genetics, 80, S17.


Lautenberger, J. A., Kan, N. C., Lackey, D., Linn, S.,
Edgell, M. H. and Hutchison, C. A. III (1978),
247, 6176-6182.
59, 1300-1306.
Luria, S. E. and Human, M. L. (1952), J.Bacteriol., 64,
557-569.
Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982),
Molecular Cloning, A Laboratory Manual (Cold Spring
Harbor Laboratory, Cold Spring Harbor, N.Y.).
Natl.Acad.Sci. USA, 72, 1184-1188.
Biochem., 41, 447-466.
Messing, J., Gronenborn, B., Müller-Hill, B. and
Hofschneider, P. H. (1977), Proc.Natl.Acad.Sci. USA,
74, 3642-3646.
USA, 76, 2232-2235.
Miyake, T. (1962), Genetics, 47, 1043-1052.
Sanderson, K. E. and Demerc, M. (1965), Genetics, 51, 897-913.
Zinder, N. D. (1960), Science, 131, 813-815.
APPENDIX I

The DNA sequence of the hsd SP region

The contiguous double-stranded DNA sequence of 1731 bp and the translation frame which includes that of the hsdS gene are shown. Targets for all the restriction enzymes compiled in the UWGCG programs available are indicated. The sequence is numbered consecutively from the 5' to the 3' end and the start and stop codons for the hsdS genes are underlined. The sequences used for the synthetic oligonucleotide primers, P1*, P2*, P3* and P4* are indicated.
BS  H  E  BS
AANX  H  I  C  SC
MULH  A  N  O  TR
H3AO  E  F  R  NF
1A42  3  1  V  11

CCTAGGCGTGCCGCTTTAGGCTAAAGGACCTATGAGGACCGACTTCTATTTCTGCTG

a: GlySerAlaArgProAsnProIleArgTrpSerGlyEndArgIleArgThrAla

---

TCGACGCCGACAGCCTGCCGGAGCCGGACGTGCTGGCGGCAGAAGCCATGGGTGAGCTGG

a: SerThrProThrAlaCysArgSerArgThrCysTrpArgGlnLysArgTrpValSerTrp

---

TGCGCAGCGTCAGTTGCTGGGAGAAGCATTTGGTGGGGTGAAGGCATGAATAGGGGGA

a: CysAlaAlaSerValAlaGlyArgSerIleTrpGlyGluGlyMetAsnArgGlyLys

---

CTGCCGGAGGGGTGGGCTACAGCTCCAGTATCTACAGTCACAACCCTAATCCGAGGAGTC

a: LeuProGluGlyTrpAlaThrAlaProValSerThrValThrThrLeuIleArgGlyVal
ACATATAGGCAAGCTCAATTATCTACGATGATTATTTGCCTATTATACGTA

301 -------------------------------------------------------

TGTATATTTTTTCGGTGGAGTTATAATAGATGTTCTACTAATAAACGGGATAATAGCA

a: ThrTyrLysLysGluGlnAlaLeuAsnTyrLeuGlnAspAspTyrLeuProIleIleArg

D
D
E

GCAAGCATAATTTCAAAATGGAAGGTAAGACTACGAGACTTAGTTTTGTGCTAAAAT

361 -------------------------------------------------------

CGTTTGGTTATAAGTTTTACCGTTCAAATGATGTCGTCACACGGATTTTTA

a: AlaAsnAsnIleGlnAsnGlyLysPheAspThrThrAspLeuValPheValProLysAsn

MM
BB
OO
22

CTTGTAAAGAAATGCAAGATATCTCCTGAGATTGTAAATTGCGATGTCCTCGGG

421 -------------------------------------------------------

GAAACATTTCTTCTAGCTTTTATAGAGGACTTCTATAACATTAACGCTACAGAAGTC

a: LeuValLysGluSerGlnLysIleSerProGluAspIleValIleAlaMetSerSerGly

H
A
C
C

AGTAAATCTGTAGTCGGTAAATCCGCACATCAACGTCTACCATTTGTGTAGTTTTCGGC

481 -------------------------------------------------------

TCATTAGACATCGCCATTTAGCGGTAGTTGCGATGTTAAACCTACATCAAGCCG

a: SerLysSerValValGlyLysSerAlaHisGlnArgLeuProPheGluCysSerPheGly

H
H
A

1
1
1

GCATTTTGCAGGCGCTTGGGACGCTTCAATTACATTGCTCATTT

541 -------------------------------------------------------

CGTAAACGCCCCGTAAACCGGGACTCTTTATGATAGAGTTTAAATGTAACGAGTAAAG

a: AlaPheCysGlyAlaLeuArgProGluLysPheIleSerProAsnTyrIleAlaHisPhe
H
B N  T I
A L  A N
N A  Q F
1 4  1 1

AATCTGGGTGCCTTAATTGTCGATTCCTGTAATGGGCTGGCTAAAAGACAAGGATTAAT

TTAGACCCCCACGGAATTAAGCTAGCTAGCACCTACCCGACGATTTTCTGTTCTTAATTTA

a: AsnLeuGlyAlaLeuIleValAspSerCysAsnGlyLeuAlaLysArgGlnGlyLeuAsn

GGTAATGAAATTTACCTATTTTGGAGTTGGCTGATTTTAAAGATGCTCAGCGTATAATTG

CCATTACTTTAATGATAAAACTCTAACCGACTAAAATTTCTACGAGTCGCATATTCCG

a: GlyAsnGluIleThrIleLeuArgLeuAlaAspPheLysAspAlaGlnArgIleIleGly

AATGAAAGAAAAATAAAATTAGACTCTAAGGAAGAAAATAAGTATTACCTAGAAAATGAT

TTACTTTCTTTTTATTTAATCTGAGATTCCTTCTTTTATTCATAAGTAAATCTTTTACTA

a: AsnGluArgLysIleLysLeuAspSerLysGluGluAsnLysTyrSerLeuGluAsnAsp

GATATTTATGTTATAAGATGGAATGCGGACTTGGCTGGCCGATTTTATGAAATAT

CTATATAAAATCAATATCTCTACTTACTTTTCACGCCTAGACCGACGCTAAATATCTTATA

a: AspIleLeuValIleArgValAsnGlySerAlaAspLeuAlaGlyArgPheIleGluTyr

H
I  D  M
N  D  B
F  E  O
1 1  2

GATATTTATGTTATAAGATGGAATGCGGACTTGGCTGGCCGATTTTATGAAATAT

CTATATAAAATCAATATCTCTACTTACTTTTCACGCCTAGACCGACGCTAAATATCTTATA

a: AspIleLeuValIleArgValAsnGlySerAlaAspLeuAlaGlyArgPheIleGluTyr

H
A  E
A  S
U  D
3  D
1  N
1  E
1  F
1  1

AAGTCAACGCCTATTGAAAGTGGGGATTTTATAGTTACGCTAGATTTCA

TTCAGTTTCGCTTTAATCTTCTCTCTCTCGTAAATATGCAAATGCGAATCTAAAGT

a: LysSerAsnGlyAspIleGluGlyPheCysAspHisPheIleArgLeuArgLeuAspSer
AATAAAATCATGTCTAGATTTTTAACCCTATTTGCAAATGAAGGCGAGGGTAGATTTTAT
TTATTTTAGTGACAGATCCTAAAATAGGAATATAACGGTTACTTTGTCCGTCCCATCTAAATA

**a:** AsnLysIleMetSerArgPheLeuThrTyrIleAlaAsnGluGlyGluGlyArgPheTyr

CTACGCAATATTTATCAACAAAGGACAGAACATACATTACCACACGAACTCAAATA
GATGCGTTATCAATAGTTGATGCTCTGTTTTATGATAATTGTTGTAGTTATTTT

**a:** LeuArgAsnSerLeuSerThrSerAlaGlyGlnAsnThrIleAsnGlnThrSerIleLys

GGACTTAGTTTTTTACTCCCTCCATTTAAAAGAAACAAGGGAATCGTCCGGCGCGTGCAA
CCTGAATCAAATAATAGGAGGTAATTTCCTTGTCTGGCTTTAGCAGGGCCGGCGCTT

**a:** GlyLeuSerPheLeuLeuProProLeuLysGluGlnAlaGluValArgArgValGlu

CAACTCTTGCCTAGGCGACACCATTGAAACAGGTTAAAACAGCCCGCCCTAGACCCGATT
GTTGAGAACCGGATGCGGCTGTGGTAACTTTTTGTCCATTGTTGCGGGACTGGGCGCAG

**a:** GinLeuPheAlaTyrAlaAspThrIleGluLysGlnValAsnAsnAlaLeuThrArgVal

AACAGCCCTACCCAGTCGATCCTGGCGCGCCTTCCGCGGAGAGCTTACCGCCCAGTGG
TTGTCGGAGTGGGTCAGCTAGGACCGCTTCCGGGGCCCCTCTCGTGGCGGGTCACC

**a:** AsnSerLeuThrGlnSerIleLeuAlaLysAlaPheArgGlyGluLeuThrAlaGlnTrp
<table>
<thead>
<tr>
<th>S</th>
<th>H</th>
<th>F</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>I</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>U</td>
<td>N</td>
<td>H</td>
<td>U</td>
</tr>
<tr>
<td>3</td>
<td>P</td>
<td>A</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>H</td>
</tr>
</tbody>
</table>

CGTGCGGAAAACCTCGATCTCATCGCGGTGAAAAACAGCCCGCCCGCCCGCTGCTGAAAAA

GCACGCCTTTTGGGACTAGAGTAGTCGCCCAATTCTGTCGCGGCGGGACGACGCTTTTT

<table>
<thead>
<tr>
<th>a:</th>
<th>Arg</th>
<th>Ala</th>
<th>Glu</th>
<th>Asn</th>
<th>Pro</th>
<th>Asp</th>
<th>Ile</th>
<th>Ser</th>
<th>Gly</th>
<th>Lys</th>
<th>Asn</th>
<th>Ser</th>
<th>Ala</th>
<th>Ala</th>
<th>Ala</th>
<th>Leu</th>
<th>Leu</th>
<th>Leu</th>
<th>Glu</th>
<th>Lys</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>H</td>
<td>N</td>
<td>I</td>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>A</td>
<td>U</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>D</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>L</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ATCAAGGCCGAACGCGCCGTCAGCGCGGGCTAAAAAACCTCGCGTTAAAAAGCCTGACGC

TAGTTCCGGCTTGCACCGGAGTCGCCCACTTCTGACGCTCTTTTTTGGAGGCGCATTTTTTCCGACTGC

<table>
<thead>
<tr>
<th>a:</th>
<th>Ile</th>
<th>Lys</th>
<th>Ala</th>
<th>Glu</th>
<th>Arg</th>
<th>Ala</th>
<th>Val</th>
<th>Ser</th>
<th>Gly</th>
<th>Lys</th>
<th>Thr</th>
<th>Ser</th>
<th>Arg</th>
<th>Lys</th>
<th>Lys</th>
<th>Ala</th>
<th>End</th>
<th>Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>I</td>
<td>M</td>
<td>H</td>
<td>N</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>N</td>
<td>S</td>
<td>H</td>
<td>P</td>
<td>C</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>P</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>I</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TTATTTTTTCGCCAGCCTCCCGCGGCTCGTTTTAAAAATTAAATCCCGACATCAAAACATCA

AATAAAAAGACGCGTGGAAGGGCGGAAGCAAAATAATAAGGGCGGAAGCATGTTTTTTG

<table>
<thead>
<tr>
<th>a:</th>
<th>Leu</th>
<th>Phe</th>
<th>Cys</th>
<th>Ala</th>
<th>Pro</th>
<th>Ser</th>
<th>Arg</th>
<th>Leu</th>
<th>Arg</th>
<th>Phe</th>
<th>Ile</th>
<th>Ser</th>
<th>Pro</th>
<th>Pro</th>
<th>Ser</th>
<th>End</th>
<th>Pro</th>
<th>Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AGAATATACCAATTCCTCGAATCCCATCCAGTCAAGCAGCATCTATACGCT

TCTTATAGGGAACGTCAATACGATGTCAGAAACATAGTCA

<table>
<thead>
<tr>
<th>a:</th>
<th>Arg</th>
<th>Ile</th>
<th>Tyr</th>
<th>Pro</th>
<th>Ile</th>
<th>Ser</th>
<th>End</th>
<th>Thr</th>
<th>Phe</th>
<th>Pro</th>
<th>Leu</th>
<th>Gln</th>
<th>Ser</th>
<th>Ile</th>
<th>Ser</th>
<th>Ala</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX II

Genetic Recombination can generate altered restriction specificity

Genetic recombination can generate altered restriction specificity (restriction enzyme evolution/protein-DNA interaction)

F. V. FULLER-PACE*, L. R. BULLAS†, H. DELIUS, AND N. E. MURRAY*

European Molecular Biology Laboratory, Postfach 10 2209, 6900 Heidelberg, Federal Republic of Germany

Communicated by Hamilton O. Smith, June 14, 1984

ABSTRACT  A recombinant strain, isolated following the transduction of an Escherichia coli recipient carrying the Salmonella typhimurium (SB) specificity genes with DNA from a donor having the Salmonella pottsdam (SP) specificity, was shown (Bullas, L. R., Colson, C. & Van Pel, A. (1976) J. Gen. Microbiol. 95, 166-172) to have neither SB nor SP specificity but to encode a novel restriction specificity, SQ. The heteroduplex analysis of the hsdS (specificity) genes of the SB and SP restriction and modification systems described here identifies a conserved sequence of around 100 base pairs flanked by two nonhomologous regions each of approximately 500 base pairs. This organization parallels that previously deduced from the DNA sequences of the hsdS genes of the related E. coli K-12, B, and D restriction systems. The present heteroduplex analyses further show that the hsdS gene conferring the SQ specificity derives one nonhomologous region from the SB gene and the other from the SP gene, as predicted from genetic exchange within the conserved sequence. This finding supports the idea that two domains of an hsdS polypeptide, which are different for each specificity, may correlate with two regions of the DNA sequence recognized. It has been shown that the recognition sequences for E. coli K-12 and B each consist of two short oligonucleotide sequences interrupted by a nonspecific sequence. A similar organization is suggested for the Salmonella specificity systems, providing the potential for evolutionary diversification of restriction specificities as a result of recombination within the conserved sequence of the hsdS gene.

The type I restriction (r) and modification (m) systems characteristic of Escherichia coli K-12 and E. coli B have been studied extensively (for reviews see refs. 1-3). The relevant enzymes EcoK and EcoB interact with a specific DNA sequence and may methylate (modify) this sequence or alternatively cleave (restrict) the DNA at relatively random locations, which may be several thousand base pairs (bp) from the interaction site.

The genes encoding the K restriction system are located counterclockwise to serB at 98.5 min on the E. coli K-12 chromosome (4, 5). Complementation tests between mutants defective in the E. coli K-12 and E. coli B restriction systems led to the hypothesis that these systems are encoded by three genes: hsdS (hsd for host specificity determinant), hsdM, and hsdR. The product of hsdS is responsible for recognition of the DNA sequence thus imparting specificity; that of hsdM, together with that of hsdS, is required for modification, while the products of all three genes are essential for restriction (6-8). Complementation tests that indicated the exchange of subunits between EcoK and EcoB (6, 9) imply that the E. coli K-12 and B restriction systems are related.

Restriction and modification systems have been demonstrated in several different Salmonella strains (10, 11), and of these, the genes encoding SB of Salmonella typhimurium (12), like those of E. coli K-12, map in the chromosome counterclockwise to serB. Complementation was observed between the products of the S. typhimurium hsdSB genes and those encoding EcoK and EcoB, thereby providing evidence that the Salmonella SB system is related to the systems of E. coli K-12 and B (13). Genetic analysis demonstrated the allelism of the SB and SP systems of Salmonella (10) and hence that the hsdK, hsdB, hsdSB, and hsdSP genes are all similar and allelic and may therefore be members of a family derived from a common ancestor.

Further experiments, involving phage P1-mediated cotransduction of hsdSP genes with serB from Salmonella pottsdam to an hsdSB strain of E. coli, led to the isolation of a strain with another specificity, SQ (14). One of the serB* recombinants in these experiments had neither SB nor SP specificity but had acquired a new system, SQ, with genes allelic to those of SB and functionally related to the K system in E. coli. It was suggested that this system arose from recombination between the hsdS genes of SB and SP. Although the possibility of creating new host specificity systems by mutation or recombination between different hsdS genes was suggested by Arber and Linn (15), SQ is the only such restriction system that has been produced experimentally.

Hybridization experiments, using DNA probes from within the hsd genes of E. coli K-12 and immunological tests using antibodies raised against either the hsdR or the hsdM polypeptide of EcoK, suggest a high degree of conservation of the structure of the hsdR and hsdM polypeptides in E. coli K-12 and B strains and, to a lesser degree, when these polypeptides are compared with those of the Salmonella SB and SP systems (16). A DNA fragment from within the hsdS gene of E. coli K-12 was shown to hybridize with the hsd genes from E. coli B, S. typhimurium, and S. pottsdam. This indicates that, even within the hsd genes, some sequence is conserved, although the extent of homology is considerably less than that between hsdR and hsdM (16).

DNA sequence analysis of the hsdS genes of E. coli K-12, E. coli B, and a system recently designated E. coli D (17) shows that within the hsdS genes of these strains, which vary in length from 1335 to 1425 bp, there are two regions that are highly conserved, one of about 100 bp and a second of about 250 bp. The remainder of each gene shares no detectable homology with the related hsdS genes. The conserved sequences are such that the 100-bp region of homology is in the middle of the gene and flanked by two nonhomologous regions, while the 250-bp conserved region is at the end of the gene encoding the carboxyl terminus of the polypeptide (17). The conserved central region and the flanking nonhomologous sequences of about 500 bp are readily identifi

Abbreviations: bp, base pairs; kb, kilobase pairs; r and m, restriction and modification.

*Present address: Department of Molecular Biology, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland.

†Present address: Department of Microbiology, Loma Linda University, Loma Linda, CA 92350.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
Materials and Methods

Bacterial and Phage Strains. The bacterial strains, all derivatives of E. coli K-12, are listed in Table 1.

The phage vector λNM762 (18) was used for cloning the hsdSB region. This vector is deleted for the ci repressor gene of λ, and where an imm<sup>21</sup> derivative was needed to provide a marker for heteroduplex analysis. im<sup>m</sup>-<sup>21</sup> Cl<sup>+</sup>. This was achieved either in vitro by using λNM761 (18) as a donor of imm<sup>21</sup> or in vivo by using an h<sup>+</sup> imm<sup>m</sup><sup>21</sup> nin donor. Temperature-inducible derivatives were isolated from crosses in which imm<sup>21</sup> was replaced by imm<sup>k</sup> cI857.

Genetic information was transferred between phage and bacterial chromosomes by homology-dependent recombination. The transfer of hsd alleles from a transducing phage to the chromosome relied on homology-dependent integration and excision. In this case, lysogens of temperature-inducible (imm<sup>k</sup> cI857) phages were selected at 32°C as colonies immune to infection by λCI<sup>-</sup> and h<sup>+</sup> imm<sup>m</sup><sup>21</sup> Cl<sup>+</sup>, and cured strains were recovered as colonies surviving at 42°C. Marker rescue from the bacterial chromosomes provided hsdSP and hsdSQ derivatives of the hsdSBΔ9 phage (see Fig. 1).

hsd phages were propagated either on NM477, a bacterium deleted for the hsdM and -S genes of E. coli K-12, or on the phage carrying the same hsd region as the transducing phage, thereby eliminating the opportunity for recombination between different hsd genes.

Plasmids for DNA Probes. Derivatives of pBR322 carrying part of the E. coli K-12 hsd genes (16, 19) were used as probes in hybridization experiments (see Fig. 1).

Media and Microbial Techniques. Media and general methods are described elsewhere (18). Deletion mutants were selected by their increased resistance to chelating agents (20).

Preparation of DNA. Bacterial DNA was purified essentially as described by Kaiser and Murray (21) except that in later preparations dialysis was used in preference to precipitation with ethanol. Plasmid DNA was purified from cleared bacterial lysates by centrifugation in CsCl/ethidium bromide (22). Phage DNA was prepared as described by Wilson et al. (23).

Table 1. Specificity genotypes of E. coli K-12 strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Specificity genotype</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM477</td>
<td>hsdKΔ5</td>
<td>Murray et al. (16)</td>
</tr>
<tr>
<td>L4001</td>
<td>hsdSB</td>
<td>Bullas et al. (14)</td>
</tr>
<tr>
<td>L4002</td>
<td>hsdSP</td>
<td>Bullas et al. (14)</td>
</tr>
<tr>
<td>L4004</td>
<td>hsdSQ; hsdSA</td>
<td>Bullas et al. (14)</td>
</tr>
<tr>
<td>NM550</td>
<td>hsdSBΔ9</td>
<td>Derivative of L4001</td>
</tr>
<tr>
<td>NM551</td>
<td>hsdSQ</td>
<td>SQ derivative of NM550</td>
</tr>
</tbody>
</table>

Restriction and Ligation of DNA. These reactions were as described by Murray et al. (16). Hybridization of DNA to Labeled Probes. The details of these experiments were as in Murray et al. (16).

Heteroduplex Analysis. Heteroduplexes were prepared from purified phage DNAs by using the cytochrome c spreading technique essentially as described by Davis et al. (24). Spreadsings were done from a mixture containing 30% (mol/mol) formamide onto a hypophase containing 10% (mol/mol) formamide. DNAs from phages P2M and M13 were used as internal length standards. Pictures were taken at a magnification of X4000. Measurements were done on 10-fold-enlarged negatives, using a digitizer connected to a Cromemco microcomputer system.

Results

Cloning the hsdSB Genes of S. typhimurium. DNA from a derivative of E. coli K-12 in which the hsdK genes have been replaced by the hsdSB genes of S. typhimurium (L4001; see Table 1) was digested with HindIII, and a library of these fragments was made, using the replacement vector λNM762. The phage library was recovered on an E. coli K-12 host with the hsdM and -S genes deleted (NM477; see Table 1); the phages formed plaques with an efficiency of less than 10<sup>-2</sup> on the rs<sup>B</sup> ms<sup>B</sup> host, L4001 (see Table 1). The phages that include the hsdM and -S genes of the SB bacterium are expected to occur at a frequency of about 10<sup>-3</sup>, but they will modify their own DNA, thereby protecting themselves from degradation on entering the rs<sup>B</sup> ms<sup>B</sup> bacterium. Enrichment for phages modified with SB specificity was readily effected by plating the library on the E. coli <sup>1</sup>ms<sup>B</sup> strain (25).

Analysis of the hsdSB Phages. The hsdSB phages include a HindIII fragment around 11 kilobase pairs (kb) in length. The positions of two EcoRI and two BamHI targets were identified (see Fig. 1).

Deletion derivatives of a hsdSB phage were selected by their ability to form plaques on Baltimore Biological Laboratories tryptase agar supplemented with 0.4 mM EDTA (20). Ten deletion derivatives were checked, and all but one failed to modify their genomes. The extent of each deletion

<table>
<thead>
<tr>
<th>a) hsd SB</th>
<th>H</th>
<th>B</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.0</td>
<td>0.6</td>
<td>4.5</td>
</tr>
<tr>
<td>b) hsd SB Δ9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) hsd K probes</td>
<td>p RH3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) hsd K genes</td>
<td>p RH1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p BH1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Analysis of hsdSB phages. (a) Restriction targets in the 11 kb HindIII fragment present in hsdSB. Restriction targets: R, EcoRI; H, HindIII; B, BamHI; numbers refer to distance, in kb, between restriction targets. (b) The extent of Δ9, a deletion shown to remove hsdS and flanking sequences. (c) Probes from the hsdK region used to detect residual homology in restriction digests of a series of deletion derivatives of hsdSB. These results (data not shown) ordered and oriented the genes in hsdSB. Both pRH3 and pBH1, which together cover hsdS and flanking sequences, fail to hybridize with the DNA remaining in hsdSBΔ9. (d) Map showing the location of the hsdK genes relative to the probes derived from them; arrows indicate the direction of transcription.
was estimated from the sizes of the fragments generated by digestion of the DNAs with HindIII, BamHI, and EcoRI. For each deletion the residual DNA homologous to each of three probes derived from different regions of the hsd genes of E. coli K-12 was identified (see Fig. 1) by hybridization (26). Since both the order of the three hsd genes of E. coli K-12 (19) and the location of the hsd DNA used as probes are known (see Fig. 1), these data oriented and ordered the hsd genes within the hsdSB phage (data not shown). Furthermore, the hybridization experiments identify hsdSB derivatives in which the entire hsdS gene is missing, since they fail to hybridize with either the pRH3 probe or the pBH1 probe, which together cover the hsdS gene and its flanking sequences (see Fig. 1). Of these deletions, hsdSBΔ9 is the most extensive, extending well into the hsdM gene and beyond the leftmost BamHI target (see Fig. 1). This phage was chosen as the means of isolating hsdSP and hsdSQ derivatives by in vivo rescue of the respective chromosomal genes.

**Generation in Vivo of λhsdSB and λhsdSQ Phages.** Derivatives of E. coli K-12 in which the hsdSP genes or the hsdSQ genes were substituted for the K specificity gene were available as λ-sensitive hosts (see Table 1). The alternative Salmonella specificity genes can therefore be transferred from the E. coli chromosome to a λ-transducing phage by homology-dependent recombination. λhsdSBΔ9 phage were propagated in the E. coli SP strain (L4002; see Table 1) and λhsdSP phages were selected after growth in the nonmodifying host NM477. λhsdSP phages grown in this host form

![Fig. 2. Restriction maps of the hsd regions of λhsdSB, -SP, and -SQ. The positions of EcoRI (R), HindIII (H), and BamHI (B) targets in the 11-kb insert in λhsdSB, -SP, and -SQ are indicated. A 400-bp probe spanning the central conserved region of the hsdS gene of E. coli K-12 was shown to hybridize to a region to the right of the targets marked *; these targets are presumed to identify the distal region of nonhomology.](image)

![Fig. 3. Heteroduplex between the DNAs from λhsdSB and λhsdSP. The empty arrow points to the central conserved sequence in the hsd region, which is flanked by two substitution loops. The filled arrow points to the substitution loop representing the different immunity regions. The circular molecules are double-stranded DNA of phage PM2 and single-stranded DNA of phage M13.](image)
The empty arrows point to substitution loops within the \( hsd \) region. The filled arrows point to the nonhomologous immunity regions.

\( \lambda hsdSQ \) was made in an analogous way by growth in the \( E. coli \) SQ host L4004. \( \lambda hsdSQ \) recombinants could not be detected by their modification phenotype, since phage \( \lambda \) has no restriction target for the SQ system (14). Enrichment for the \( \lambda hsdSQ \) phages relied on their increased buoyant density. Heavy phages were collected from a CsCl equilibrium gradient, and plaques on strain NM477 were screened by using labeled pRH3 (see Fig. 1) as probe. Plaques identified as positive were presumed to arise from phage in which the deletion had been replaced by the \( hsdS \) region from the \( hsdSQ \) genes, the only source of homologous \( hsd \) genes.

The specificity endowed by the \( hsd \) genes in the presumptive \( \lambda hsdSQ \) phage was confirmed, since the transfer of these genes to the chromosome of an \( r_{SB}^+ m_{SP}^- \) K-12 derivative (NM550; see Table 1) conferred the SQ specificity as detected by using transducing derivatives of phage \( \lambda \) that carry targets for the SQ system.

The DNAs of the \( \lambda hsdSB, -SP, \) and \( -SQ \) phages were compared and found to be very similar (Fig. 2). The inserts are of the same size but different in that the SP and SQ phages include an extra HindIII site at the expense of an EcoRI target. This difference is consistent with the homologous fragments identified in the appropriate \( E. coli \) and Salmonella DNAs by using probes from the \( hsd \) genes of \( E. coli \) K-12 (ref. 16 and unpublished data).

**Heteroduplex Analyses.** The heteroduplex molecules resulting from the three combinations of phage DNAs carrying the SB, SP, and SQ \( hsd \) regions were examined by electron microscopy. In each heteroduplex the vectors differed by an immunity marker; one phage was \( imm^2 \), and the other \( imm^1 cI \) KH54. This caused a substitution loop that could be used as a marker in the right arm of the vector. Any further nonhomologies found in the heteroduplexes were within the DNA from different \( hsd \) regions.

Heteroduplex molecules derived from \( \lambda hsdSB \) and \( \lambda hsdSP \) (Fig. 3) were like those seen for \( hsdR \) and \( hsdB \) (unpublished observations). The unpaired regions indicated nonhomologous sequences of about 500 bp flanking a conserved core of around 100 bp. The measurements obtained from these heteroduplexes are given in Fig. 5. In the case of \( E. coli \) K and B the DNA sequence (17) predicts the struct-
tasures of the heteroduplex molecules. It appears that the organization seen in *E. coli* is conserved for the *Salmonella* members of the family. Furthermore, it has been shown (16) that a probe including the central conserved sequence of the *hsdSK* gene shares homology with the *hsdS* genes of the three *Salmonella* systems. Hybridization experiments (data not shown) using this probe and DNA from *hsdSB*, -SP, and -SQ phages show that this region of homology, presumed to be the conserved central region, is upstream of the variable region identified by the EcoRI and HindIII targets (see Fig. 2).

Heteroduplexes derived from *hsdSB* and *hsdSQ* (Fig. 4A) or from *hsdSP* and *hsdSQ* (Fig. 4B) show only a single substitution loop in addition to the nonhomology in the immunity region. The loops cannot be distinguished by size, but the distances from the immunity region indicate that *hsdSQ* contains sequences homologous to *hsdSB* to the right side of the conserved central region (Fig. 5b) and sequences homologous to *hsdSP* to the left side of the conserved central region (Fig. 5c). The recombination event that generated SQ must have occurred either close to or within the conserved region, since no additional loop or knot can be detected.

### DISCUSSION

Bullas *et al.* (14) suggested that the origin of SQ, a novel DNA restriction and modification system, was recombination between the specificity genes of *S. typhimurium* and *S. pottsdam*. The DNA sequences of the related *hsd* genes of the *E. coli* K-12, B, and D systems identified a conserved region within the *hsdS* genes, flanked by nonhomologous regions. It was tempting to suggest that recombination within a similar conserved region of the *hsdS* genes of SB and SP would lead to the reassortment of flanking regions and the concomitant generation of a new specificity. The heteroduplex analysis of the *Salmonella hsd* regions confirms that the *hsdSQ* genes are indeed derived by recombination between the *hsdSB* and -SP genes.

One interpretation of this finding would correlate the two varying domains of the *hsdS* gene, and hence of the specificity polypeptide, with the two segments of the DNA recognition sequence, although reassortment of minor differences within the conserved sequences remains an alternative or additional possibility. Nevertheless, the crucial prediction that one region of the DNA recognition sequence for SQ derives from the SB sequence, the other from SP, has recently been confirmed by the data of V. Nagaraja and T. A. Bickle (personal communication).

Their data, taken with our analysis of the *hsdS* genes, permit the correlation of the distal variable region of the *hsdS* genes of SP and SQ with the recognition of the sequence A-A-C and the proximal variable region of the *hsdS* SB and SQ genes with the recognition of the sequence T-A-Y-G (Y = pyrimidine nucleoside), assuming the direction of transcription indicated in Fig. 1.

One of the two sequences recognized by the SP system is identical to that recognized by *EcoK* (V. Nagaraja and T. A. Bickle, personal communication), as expected from the finding that modification by the SP system gives complete protection against restriction by the *E. coli* K-12 system (11). Nevertheless, our earlier hybridization experiments (16) do not suggest a specially close similarity of these two specificity genes.

The potential to create new restriction specificities by recombination has significant evolutionary implications; additional genetic experiments should test the generality of our finding.

We thank V. Nagaraja and T. A. Bickle for communication of results prior to publication; R. P. Ambler, T. A. Bickle, W. Loenen, D. W. Meek, and K. Murray for constructive criticism of the manuscript; and H. Grozier and E. C. McCready for help in the preparation of the manuscript. The latter part of the work was supported in Edinburgh by a grant from the Medical Research Council.