

The Type I Restriction Systems of *Escherichia coli*

by

Victoria Anne Barcus

Thesis presented for the degree of Doctor of Philosophy

Institute of Cell and Molecular Biology

University of Edinburgh

1993



To my parents,
and to my sister,
for their love and encouragement

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Noreen Murray, for her guidance, kindness, and for many interesting discussions, scientific and otherwise.

I would also like to thank Ken Murray.

My thanks go to my second supervisor, Chris Boyd, for providing strains and many useful suggestions (and great Birthday parties!).

Thanks also to Annette Campbell, Diane Ternent, and Anne Daniel, for their invaluable and good-humoured technical advice and assistance, and for their friendship.

I would like to acknowledge the rest of the people in Noreen's group, past and present, all of whom contributed to my work in one way or another, including Laurie Cooper, Gill Cowan, Dave Dryden, Alex Gann, P. Kannan, Julia Kelleher, Gareth King, Lynn Powell, Peter Thorpe, Damion Willcock, and Markus Winter (for guiding me through the complexities of MacDraw with windows).

My thanks also go to everyone in the media and washing up rooms, particularly Helen Hibbert and Lorraine McWilliams; to Graham and Frank for photographs; and to everyone in the stores.

Thanks also to Howard Ochman, Thomas Whittam, and David Thaler for providing strains.

I am particularly grateful to Lindsay Riddick for typing the majority of this thesis, efficiently and cheerfully, on countless Sunday afternoons.

I wish to thank Ann, Morag, Carol, Caroline, David, the gang at Tower Farm, everyone on the sixth floor, and the Friday night regulars at the Union Bar, for their friendship over the years.

Finally, special thanks to Stuart for patience, understanding, and a shoulder to lean on, and to his parents for providing a second home.

Most of all, I want to thank my family for their support, not only financial (Barcus Research Council?), but also moral, and for their encouragement and love. Special thanks to my mother for cooking and cleaning while I wrote, typing references, and helping with the proofreading.

ABSTRACT

Of the three types of restriction and modification (r-m) systems in *Enterobacteriaceae*, type I are the most complex both in structure and in function. Three genes, *hsdR*, *M*, and *S* encode the multisubunit enzyme. While the products of the *hsdM* and *S* genes alone can form a methylase, the products of all three genes together form an enzyme which functions as both an endonuclease and a methylase. The product of the *hsdS* gene confers sequence specificity to both the restriction and modification activities.

Eleven known type I r-m systems have been separated into three families on the basis of complementation tests, immunological cross-reactivity, and DNA hybridization of their coding sequences. The genes for two of the families reside at the same location on the chromosome, those of the third may be plasmid-encoded. Each member of a family recognizes a unique DNA sequence. Sequence homologies within families allow the use of probes for detection of related genes in other bacteria.

While research on type I systems has focused on their mechanism of action and their genetic determinants, extensive surveys for their presence among *E.coli* strains have not been documented. An understanding of the prevalence and diversity of type I r-m systems in natural populations is basic to any consideration of either their relevance to a bacterial population or of the evolutionary forces acting at the *hsd* locus. This thesis describes the use of biological and molecular screens to determine the distribution and diversity of type I r-m systems in natural isolates of *E.coli*.

A screen was developed using phage P1 and a Dar^- mutant derivative to detect restriction by type I r-m systems in natural isolates of *E.coli*. Unfortunately, a minority of the strains tested were sensitive to productive infection by P1. While several of the strains appeared to have a type I r-m system, as evidenced by a reduced plating efficiency of the Dar^- mutant, this suggestion could not be proven by demonstration of concomitant modification of the surviving phage.

Daniel *et al.* (1988) used DNA hybridization to screen a sample of enteric bacteria for sequences homologous to type I systems from each of two families. Although some related systems were detected in more than one genus, the majority lacked DNA with similarity to either probe. Whether the bacteria encoded alternative systems or lacked them altogether could not be concluded.

The hybridization study of Daniel *et al.* (1988) has now been extended to the ECOR collection of wild type *E.coli* strains, believed to be representative of the species as a whole (Ochman and Selander, 1984a). Among thirty-five strains tested, the DNA of nine hybridized to one family-specific probe and three to the other. This suggests that type I systems are more common than previously believed.

By cloning the relevant genes from a number of these strains and analysing them in a restriction deficient derivative of *E.coli* K-12, some of the systems were found to have novel specificities, and others the same as the previously documented *EcoK* and *EcoB* systems.

A phylogenetic tree based on the above results does not correspond to one constructed on the basis of protein polymorphisms, as determined by multilocus enzyme electrophoresis (MLEE), a method commonly used among population geneticists to determine the overall relatedness of strains (Selander *et al.*, 1986). Type I r-m systems of the same family were found in distantly-related strains, and systems of different families were found in strains deemed to be closely-related on the basis of MLEE.

These results are consistent with selective pressure at the *hsd* locus for variation in specificity of type I systems, perhaps due to frequency-dependent selection by bacteriophages on their hosts (Levin, 1986; Sharp *et al.*, 1992). Acquisition of novel *hsd* genes by *E.coli* strains may have occurred by intra- and interspecific horizontal transfer (Sharp *et al.*, 1992).

ABBREVIATIONS

AdoMet	S-adenosylmethionine
ATP	adenosine triphosphate
bp	base pair(s)
c.f.u.	colony forming units
DTT	dithiothreitol
EDTA	diaminoethanetetra acetic acid
e.o.p.	efficiency of plating
g	standard acceleration due to gravity
<i>hsd</i>	host specificity for DNA
kb	kilobase(s)
kD	kilodalton(s)
m.o.i.	multiplicity of infection
PCR	polymerase chain reaction
PEG	polyethylene glycol
p.f.u.	plaque forming units
r-m	restriction and modification
SDS	sodium dodecyl sulphate
SSC	standard saline citrate
ts	temperature sensitive
UWGCG	University of Wisconsin Genetics Computer Group
Δ	deletion

CONTENTS

DECLARATION	i
ACKNOWLEDGEMENTS	ii
ABSTRACT	iii
ABBREVIATIONS	v
<u>CHAPTER ONE</u> INTRODUCTION	1
1.1 HISTORICAL BACKGROUND	2
1.2 TYPE II R-M SYSTEMS	5
1.3 Type III R-M SYSTEMS	7
1.4 TYPE I R-M SYSTEMS	10
A Genetic Determinants	10
B Enzyme Families	14
C Reaction Mechanisms of Type I R-M Systems	19
D Sequence-Specific DNA	21

E Phage Encoded DNA Modifications	42
a) T-even phages	42
b) Phage Mu	43
c) Phage T1	44
F The Dar Antirestriction Function of Phage P1	44
1.7 DNA TRANSFER	47
A Transformation	47
B Conjugation	48
C Transduction	51
<u>CHAPTER TWO</u> MATERIALS AND METHODS	53
2.1 STRAINS	54
A Bacterial Strains	54
B Phage Strains	54
C Plasmids	54
2.2 ENZYMES AND CHEMICALS	54
2.3 MEDIA	65
2.4 STANDARD SOLUTIONS	66
2.5 MICROBIAL TECHNIQUES	67
A Long Term Storage of Bacterial Cells	67
B Preparation of Plating Cells	67
C Preparation of λ Plate Lysates	67
D Phage Titration	69
E Spot Tests	69

F	Constuction of λ cI857att Lysogens	69
G	Curing of λ cI857 Lysogens	70
H	Preparation of P1Cm1c1ts Lysogens	70
I	Preparation of P1Cm1c1ts Lysates	71
J	Preparation of P1kc Phage: Plate Lysates	71
K	Preparation of P1kc Phage: Liquid Lysates	71
L	Cotransduction Using P1 Lysates	72
M	Bacterial Conjugation	72
 2.6 DNA TECHNIQUES		 73
A	Ethanol Precipitation of DNA	73
B	Rapid Large-Scale Preparation of Plasmid DNA	73
C	Alkaline Lysis Plasmid "Miniprep"	74
D	Single Colony Gel Analysis of Plasmid DNA	74
E	Bacterial DNA "Miniprep"	75
F	Restriction Endonuclease Digestion of DNA	75
G	Ligation of DNA	76
H	Agarose Gel Electrophoresis	76
I	Purification of DNA Fragments from Low Melting Point Agarose	76
J	Recovery of DNA from Low Melting Point Agarose Using GELase™	77
K	Transformation of Competent Cells	77
L	Electrotransformation	78
M	<i>In vitro</i> Packaging of λ DNA	79
N	Transfer of DNA From Plaques to Nylon	79
O	Transfer of DNA from Agarose Gels to Nylon Membranes	80

P Labelling of DNA Probes and Hybridization	81
a) DNA hybridization using the Amersham ECL Gene Detection System	81
b) Nonradioactive DNA labelling and detection using Digoxigenin	81
<u>CHAPTER THREE</u> A BIOLOGICAL SCREEN FOR TYPE I R-M SYSTEMS IN WILD TYPE <i>E. COLI</i> ISOLATES	83
3.1 INTRODUCTION	84
3.2 DEVELOPMENT OF THE P1 SCREEN	88
3.3 RESULTS	89
A Application of the P1 Screen to Laboratory Strains of Known Specificities	89
B Extension of the P1 Screen to Natural Isolates of <i>E. coli</i>	94
C P1 Cotransduction of Putative <i>hsd</i> Genes from Wild Type <i>E. coli</i> Strains	98
3.4 DISCUSSION	100
<u>CHAPTER FOUR</u> THE DISTRIBUTION OF TYPE IA AND IB <i>hsd</i> GENES AMONG SOME MEMBERS OF THE ECOR COLLECTION	106
4.1 INTRODUCTION	107
4.2 RESULTS	111
A Hybridization Using <i>hsd</i>-Specific Sequences as Probes	111
B Hybridization Using DNA Flanking the <i>hsd</i> Region as Probes	124
4.3 DISCUSSION	131

CHAPTER FIVE	ANALYSIS OF SPECIFICITIES OF SOME TYPE IA R-M SYSTEMS IN MEMBERS OF THE ECOR COLLECTION	142
5.1	INTRODUCTION	143
5.2	RESULTS	145
	A Construction of λ Libraries	145
	B Complementation analysis of the λhsd phages	148
	C Transfer of the Phage-encoded <i>hsd</i> Genes to the <i>E.coli</i> Chromosome	150
5.3	DISCUSSION	157
	A Functional Analysis of ECOR Type Ia R-M Systems	157
	B Evolution of Type I R-M Systems in <i>E.coli</i>	168
	REFERENCES	178

CHAPTER ONE

INTRODUCTION

1.1 HISTORICAL BACKGROUND

Host-controlled restriction and modification was first described in the 1950's (Luria and Human, 1952; Bertani and Weigle, 1953; Luria, 1953; Lederberg, 1957), when it was observed that the genotype of the host bacterium in which a bacteriophage was propagated affected the phenotype of the progeny phage. For example, phage lambda (λ) propagated on *E. coli* C (λ .C) showed a lower efficiency of plating (e.o.p.) on *E. coli* K-12 than on *E. coli* C. However, λ .K plated with equal efficiency on either strain (Bertani and Weigle, 1953).

This apparent difference was non-hereditary, gained as a result of one cycle of growth in a particular host, and lost following a cycle of growth in a different host. As the adaptation was a function of the host strain on which the phage was last propagated, the term "host-specificity" was employed to describe the phenomenon.

Lederberg (1957) demonstrated that the restriction of the phage host range was associated with degradation of the phage DNA. Phage T1 labelled with ^{32}P were grown on *E. coli* B and the progeny (T1.B) were used to infect *E. coli* B and *E. coli* B(P1) (*E. coli* B lysogenic for phage P1). Detection of the ^{32}P label in the surrounding medium without recovery of active phage particles suggested that the labelled DNA was degraded upon infection of the *E. coli* B(P1) cells.

It was not until the 1960's that the molecular basis for the restriction phenomenon was determined by Arber and his colleagues (Arber and Dussoix, 1962; Dussoix and Arber, 1962; Arber *et al.*, 1963).

Arber and Dussoix (1962) provided evidence that the host specificity is carried on the phage DNA. They followed the fate of the host specificity of λ previously grown on *E. coli* K-12 lysogenic for bacteriophage P1 (λ .K12(P1)) after a single cycle of infection on K-12. The majority of the progeny phage had lost the ability to grow on *E. coli* K-12(P1) and were now λ .K-12, carrying only the host specificity of the

new strain. By labelling the λ .K-12(P1) parental DNA with the heavy isotope deuterium, they were able to show that those phage particles which retained the ability to grow on K-12(P1) after a single cycle of growth on K-12 retained at least one strand of parental (heavy) DNA as a result of semi-conserved replication or no replication. Those particles containing only newly-synthesized DNA were now adapted for growth on the new host, K-12, but not the original host, K-12(P1). These observations left little doubt that host specificity was imparted to the DNA.

Dussoix and Arber (1962) showed in marker rescue experiments that genetic markers from λ .K are rescued by λ .K(P1) when both phages coinfect K-12(P1) host cells. The rescue frequency was highest when infection by the appropriately-modified λ .K(P1) preceded infection by λ .K. This reflected competition between degradation of the λ .K genome and rescue of markers by recombination. Infection with λ .K first, resulted in degradation of the λ .K genome so that fewer markers were available upon superinfection by λ .K(P1). By infecting K-12(P1) host cells with ^{32}P -labelled λ .K phage, Dussoix and Arber (1962) additionally showed that the restriction of the phage was due to nucleolytic degradation of the infecting DNA inside the cell.

Arber and Morse (1965) demonstrated that host-specific restriction also applies to the transfer of chromosomal and episomal DNA from a donor to a recipient cell. Recombination frequencies were found to be 100-fold lower in crosses between HfrK-12 donors and F^- K-12(P1) recipients than between crosses of "like" donors and recipients or between K-12(P1) donors and K-12 recipients. This also applied to the acceptance of the fertility factor, F, and its F' derivatives *Fgal* and *Flac*, as well as resistance transfer factors (Glover *et al.*, 1963; Arber and Morse, 1965). These results led to the generalization that restriction provides a defense against any foreign DNA entering the host cell, whether it be viral, plasmid, episomal, or chromosomal.

These studies led to the proposal of a two-enzyme system controlling host-specific restriction and modification (Arber, 1965a). The absence of host-specificity in a DNA molecule would be recognized by both a restriction enzyme and a modification enzyme. The latter would modify the DNA, imparting on it the host

specificity that would render it resistant to the action of the restriction enzyme. Restriction would occur only if this host specificity was absent, destroying the DNA by endonucleolytic cleavage. Since no genetic or functional changes were observed during host-controlled modification, it appeared that host specificity does not alter the content, the copying, or the reading of the genetic message.

A further hint to the biochemical nature of host-imparted modification was provided by the observation that infection by λ of *met*⁻ *E. coli* K-12 or B strains starved for methionine resulted in progeny phage which were poorly modified (Arber, 1965b). Thus, methionine must be required for the modification function. Arber (1965b) hypothesized that the host-specific modification is conferred by alkylation of specific sites on the phage DNA, with methionine acting as a donor of methyl groups.

Evidence that an enzyme was involved in the restriction and modification process was first provided by Takano and colleagues (1966). Unmodified λ DNA was shown to be specifically inactivated by cell-free sonicates of *endoI*⁻ strains of K-12 carrying an *fi*⁻ R factor. Infectious λ DNA was mixed with sonicated cell extract. The residual infectivity of the DNA was then determined. Unmodified infectious λ DNA was inactivated to a greater extent than modified λ DNA.

Subsequent identification and purification of restriction enzymes from *E. coli* K-12 (*EcoK*) and *E. coli* B (*EcoB*) (Meselson and Yuan, 1968; Linn and Arber, 1968; Roulland-Dussoix and Boyer, 1969) confirmed the enzyme hypothesis.

Isolation of a restriction endonuclease from *Haemophilus influenzae* Rd which had a lower molecular weight and less stringent cofactor requirements than the *E. coli* enzymes (Smith and Wilcox, 1970) led to the proposal of division of the enzymes into two classes (Boyer, 1971). Type I would consist of the two *E. coli* restriction enzymes, *EcoK* and *EcoB*, which required ATP, S-adenosylmethionine (AdoMet) and magnesium ions (Mg²⁺) for restriction (Meselson and Yuan, 1968; Roulland-Dussoix and Boyer, 1969). Type II would include the enzyme from *H. influenzae* Rd and the enzyme encoded on the *fi*⁻ R-factor, both of which required only Mg²⁺ for

endonuclease activity (Smith and Wilcox, 1970; Hedgpeth *et al.*, 1972).

In more recent years, the characterization of a restriction enzyme in *H. influenzae* Rf (Kauc and Piekarowicz, 1978) and one encoded by the *E. coli* plasmid P15 as well as that of phage P1 (Arber and Wauters-Willems, 1970; Reiser and Yuan, 1977; Hadi *et al.*, 1983; Iida *et al.*, 1983) has led to the proposal of a third group of restriction enzymes, designated type III. These were separated from the other two groups on the basis of differences in cofactor requirements and enzyme activities.

1.2 TYPE II R-M SYSTEMS

Type II (reviewed by Wilson and Murray, 1991; Bennett and Halford, 1989; Wilson, 1988; Bickle, 1987) are the simplest of the three types of restriction and modification systems. They comprise two enzymes, an endonuclease encoded by the *res* gene, and a separate methylase, the product of the *mod* gene. The *res* and *mod* genes are often plasmid-borne, and are usually, if not always, adjacent to one another. The genes may be in either order relative to one another on the plasmid, transcribed convergently or divergently, or in the same direction by two promoters or one.

Only Mg^{2+} is required for type II restriction, and AdoMet as a methyl donor for modification. The endonuclease usually recognizes a palindromic sequence of four to eight base pairs (bp) which may be continuous or interrupted, and cleaves the DNA symmetrically within this sequence. While the endonuclease generally functions as a homodimer, the methyltransferase apparently acts as a monomer, methylating a specific residue within the recognition sequence, one strand of DNA at a time. Type II methyltransferases vary in the residue that they methylate, which may be either adenine or cytosine.

Approximately 150 different specificities have been identified (Kessler and Manta, 1990). Many type II r-m systems from different bacteria recognize identical

sequences and are referred to as isoschizomers.

Type II endonucleases of different specificity are entirely dissimilar in sequence. Those that are isoschizomers vary in the degree of identity. Similarities among methyltransferases also depend on specificity, though methylases that act on the same residue (i.e. adenine or cytosine) show additional sequence conservation.

Some type II r-m systems are so similar that they are clearly descendants of the same ancestral system. For example, *Cfr91* from *Citrobacter freundii* and *XmaI* from *Xanthomonas* have methyltransferases that share 81% identity and endonucleases with 75% identity (see Wilson and Murray, 1991).

The type II endonuclease and methyltransferase have the capacity to recognize and interact with the same DNA sequence independently of one another. Therefore, the scope for evolution to form new specificities ought to be limited. Evolution of one component enzyme to recognize a new specificity should necessitate compensatory evolution in the other component in order to avoid suicide of the cell.

There is in general no sequence similarity between a given methyltransferase and its cognate endonuclease. It would be anticipated that the two enzymes would at least share sequences involved in recognition of the specific DNA sequence. However, this appears not to be the case, arguing against a common ancestry for endonucleases and their corresponding methyltransferases.

Some type II r-m systems appear to have formed as partnerships between independently evolving genes that became linked. *EcoRI* and the chromosomally-encoded *RsrI* of *Rhodobacter sphaeroides* catalyse identical reactions and have the same gene organization. While the endonucleases of the two systems share 49% amino acid sequence identity (Stephenson *et al.*, 1989), the methylases are much less similar, differing in length and base composition, with only 16% amino acid identity (Kaszubska *et al.*, 1989). This suggests that the methyltransferases evolved from different ancestors and were joined by their partner endonucleases in independent

events, resulting in the formation of distinct, but catalytically identical r-m systems.

The apparently independent assembly of many of these r-m systems only adds to the puzzle of how endonucleases have come to be associated with particular methylases.

A subclass of type II restriction enzymes, type IIs (shifted cleavage) differ from type II in that they cleave DNA at a position outside of their recognition sequences (reviewed by Szybalski *et al.*, 1991). Like type II enzymes, the endonuclease and methylase act independently and have similar cofactor requirements. However, the type IIs enzymes recognize generally asymmetric DNA sequences and cleave the DNA at a precise distance from the recognition sequence, usually 1 to 20bp away. As the recognition sites are not symmetric, the type IIs modification enzyme comprises two methylases, one for methylating each strand of the recognition site.

1.3 Type III R-M SYSTEMS

The type III enzymes (such as those of phage P1 and the plasmid P15) are complex, multifunctional enzymes able to methylate or cleave unmodified DNA (for reviews see Wilson and Murray, 1991; Bickle, 1987).

At present, only four type III r-m systems have been identified. These include *EcoP1*, encoded by the P1 prophage (Iida *et al.*, 1983; Hadi *et al.*, 1983), the plasmid-encoded *EcoP15* found in *E.coli* 15T⁻ (Reiser and Yuan, 1977; Arber and Wauters-Willems, 1970; Iida *et al.*, 1983; Hadi *et al.*, 1983), the chromosomally-located *StyLT1* from *S.enterica* serovar typhimurium (DeBacker and Colson, 1991a), and *HinfII* from *Haemophilus influenzae* (Kauc and Piekarowicz, 1978).

Type III r-m systems are encoded by two genes, *res* and *mod* (reviewed in Wilson and Murray, 1991; Bickle 1987). The *mod* gene product provides sequence specificity for both DNA methylation and restriction. On its own, the *mod* gene

product acts as a methyltransferase, while as a complex with the *res* gene product it forms a two-subunit methyltransferase and endonuclease. Both genes are transcribed from a promoter located before *mod*, and there is evidence for a second promoter between the genes for transcription of the *res* gene alone (Iida *et al.*, 1983). The methylation reaction requires AdoMet as a methyl donor and Mg^{2+} . Cleavage by the restriction enzyme requires ATP and is stimulated by AdoMet. In the presence of both cofactors, the methylation and cleavage reactions compete with each other. The recognition sequences of type III enzymes are asymmetric, and five to six nucleotides in length. Methylation of the target sequence occurs on one strand only; the complementary strand lacks an adenine for methylation.

The endonuclease requires the presence of two unmodified target sites, in inverse orientation on the DNA (Meisel *et al.*, 1992). This is tantamount to defining the recognition sequence as a palindromic sequence with a non-specific spacer of variable length. The question arises as to how P1-modified DNA avoids restriction following replication, which should result in the production of unmodified sites. However, all unmodified sites generated by replication of modified (i.e. hemimethylated) DNA will be in the same orientation and thus refractive to cleavage (figure 1.1), providing a control against restriction of "self" DNA (Meisel *et al.*, 1992). Methylation is independent of the orientation of the sites.

The restriction enzymes cleave DNA, with unmethylated target sequences in the appropriate orientation, 25 to 27bp in the 3' direction from the target site.

Comparison of the EcoP1 and EcoP15 *res* genes has shown that they are almost totally conserved. Consequently, the subunits of the two systems are interchangeable (Bickle, 1987). The *mod* genes have highly-conserved amino and carboxyl ends, but apparently unrelated sequences in the middle of the molecules (Hümbelin *et al.*, 1988). The implication is that this region is involved in DNA sequence recognition. Hümbelin *et al.* (1988) hypothesized that a change in specificity could be brought about by changing these "cassettes" of non-homology.

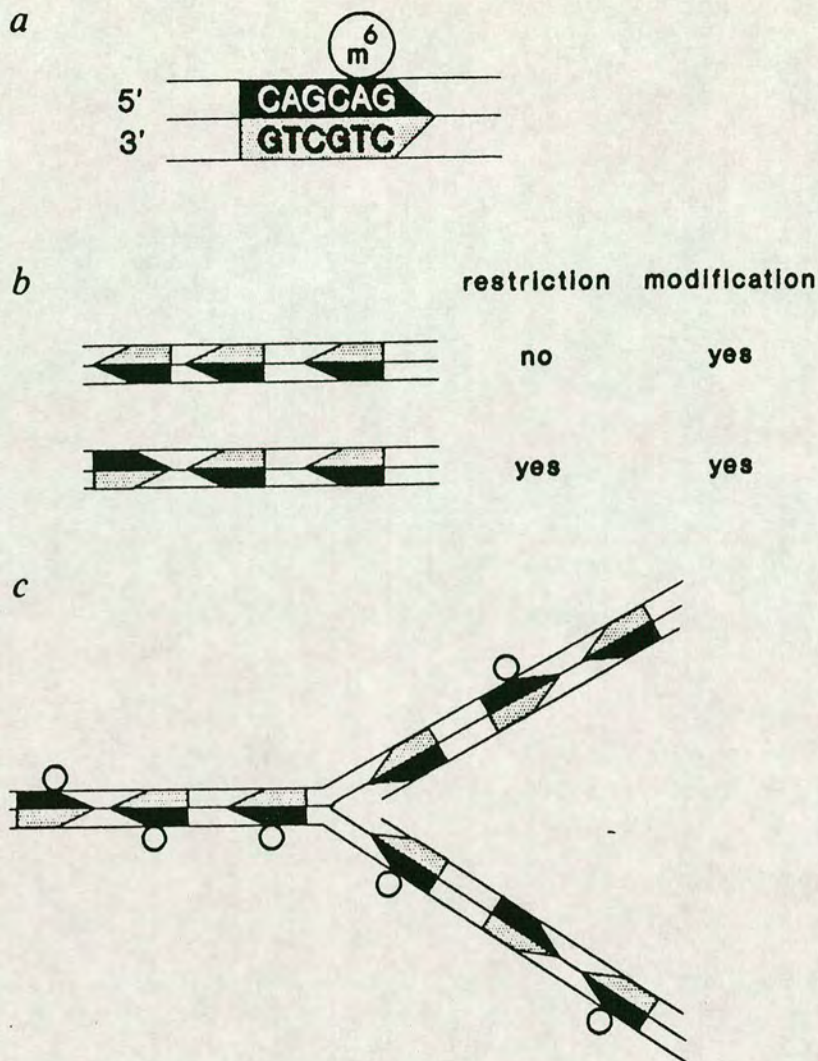


Figure 1.1: Strand-bias model for cleavage of DNA by the type III restriction enzyme, *EcoP15*.

a. The recognition sequence of *EcoP15*; the arrowhead indicates the direction of the cleavage site, 25-27bp downstream of the target sequence; m^6 shows the position of the methylated adenine.

b. Restriction requires two unmethylated recognition sequences in inverse orientation; methylation is independent of orientation.

c. All unmodified recognition sites generated during replication of modified (i.e. hemimethylated) DNA will be strand-biased; the daughter DNA molecules will have modified target sequences where the methylated adenine residues (indicated by circles) are in the parental DNA strand, and unmodified sites where thymine residues are in the parental strand. As a result, all unmethylated sites in the daughter DNA will have the same orientation. (From Meisel *et al.*, 1992)

Interestingly, unlike most r-m systems, the genes for the *Sty*LT1 system cannot be transferred to an unmodified host without a lethal effect (DeBacker and Colson, 1991b). However, the genes may be transferred to a strain expressing the modification function (Mod⁺), and the *res* gene alone may be expressed in the absence of a *mod* gene (DeBacker and Colson, 1991b). This has implications for the transfer of these genes among natural populations. Indeed, the transfer of a conjugative plasmid carrying the *Sty*LT1 genes induces the death of unmodified recipient cells by host DNA degradation (DeBacker and Colson, 1991b). Additionally, any mutation resulting in a new specificity will kill the host cell, ensuring the invariability of the LT1 specificity.

1.4 TYPE I R-M SYSTEMS

Type I systems are by far the most complex of the three types. They are composed of three subunits, encoded by three genes, *hsdR*, *hsdM* and *hsdS* (*hsd* for host specificity of DNA). The *hsdM* and *S* gene products form a methylase, while the *hsdR* product is additionally required for the restriction endonuclease. The *S* (specificity) subunit dictates the sequence recognized by the enzyme. The enzymes recognize asymmetrical, hyphenated sequences consisting of a trimer and a tetra- or pentamer of constant sequence separated by a nonspecific spacer of fixed length, generally six to eight bases in length. For example, the type I r-m system of *E. coli* K-12, *EcoK*, recognizes the sequence 5'AAC (N₆) GTGC. While the recognition sequence is also the site of methylation (one adenine is methylated in each strand), cleavage of unmodified DNA occurs at apparently random sites up to several kilobases (kb) away from the recognition sequence. The restriction enzyme, which also functions as a methylase and ATPase, requires ATP, Mg²⁺ and AdoMet for cleavage. The methyltransferase requires AdoMet only, though it is stimulated by ATP and Mg²⁺.

A Genetic Determinants

The genetic determinants of type I restriction and modification were defined

largely by complementation analysis between restriction proficient or deficient (r^+ or r^-) mutants and modification proficient or deficient (m^+ or m^-) mutants of the r - m systems of *E. coli* K-12 (*EcoK*) and B (*EcoB*).

Using Hfr mapping, Boyer (1964) first showed that the genetic loci determining host specificity in *E. coli* K-12 and B were probably allelic and that this region controlled both restriction and modification properties. These results were later confirmed by Glover and Colson (1969), who used Hfr crosses and P1 transduction to show that both the restriction and modification loci mapped near to *serB*, at approximately 98.5 minutes on the *E. coli* genetic map (Bachmann and Low, 1980).

In 1966, Wood isolated restriction-deficient mutants of *E. coli* K-12 and B. Half of these mutants retained normal modification activity (r^-m^+), while the remainder were deficient in both restriction and modification (r^-m^-). As the r^-m^- and r^-m^+ mutant phenotypes arose with nearly the same frequency, it seemed unlikely that the r^-m^- phenotypes resulted from two mutations. A third gene, concerned with both restriction and modification functions was postulated. A mutation in this gene would give rise to mutants deficient in both processes.

This theory was supported by experiments of Boyer and Roulland-Dussoix (1969) in which F' plasmids encoding restriction and modification systems were used to construct a number of partial permanent diploids with different arrangements of mutant and wild type *EcoB* alleles. These strains were then tested for their ability to restrict and modify λ DNA. A mutation conferring the r^-m^+ phenotype was able to complement one conferring an r^-m^- phenotype to produce an r^+m^+ diploid. However, independently isolated r^-m^- mutants were unable to complement each other, as were independently isolated r^-m^+ mutants. This indicated that the mutations resulting in a deficiency in restriction in the two strains must be located in different genes.

Interstrain complementation tests between the host specificity genes of *E. coli*

K-12 and B confirmed these results and defined the genetic determinant for specificity (Boyer and Roulland-Dussoix, 1969; Glover, 1970). The host specificity genes of K-12 and B were so similar that their subunits were interchangeable. An F' episome expressing an $r_K^- m_K^+$ enzyme could complement an $r_B^- m_B^-$ strain to give rise to a diploid with K-specificity ($r_K^+ m_K^+$). The reverse experiment gave similar results, with an $r_B^- m_B^+ / r_K^- m_K^-$ merodiploid encoding a functional enzyme with B-specificity. Thus, a specificity determinant required for both restriction and modification was supplied by the F' episome.

A mutation resulting in a third possible phenotype, $r^+ m^-$, had not been sought, as it was believed that this mutation would be lethal. However, m^- mutants were isolated by mutagenizing $r_B^- m_B^+$ strains to give rise to second-step $r_B^- m_B^-$ mutants. These second-step mutants could be complemented by a wild type $r_K^+ m_K^+$ strain resulting in a diploid expressing both specificities, that is, $r_{KB}^+ m_{KB}^+$. This indicated that both K and B specificity determinants were functional, and that the mutations must lie elsewhere. An experiment demonstrating complementation of the second-step $r^- m^-$ mutant by a first-step $r^- m^-$ mutant implied that three genes are involved in type I restriction and modification (Boyer and Roulland-Dussoix, 1969; Glover, 1970).

A general model was proposed, in which three genes encode three polypeptides involved in the restriction and modification of DNA. One gene (now referred to as *hsdS*) encodes a recognition, or specificity, polypeptide, a second (*hsdR*) encodes a restriction polypeptide, and the third (*hsdM*) a modification polypeptide. The experiments did not determine whether a single oligomeric enzyme was responsible for both restriction and modification activities, or whether two enzymes, both comprising a specificity subunit and either a restriction or modification subunit was involved. However, the requirement of AdoMet for restriction as well as modification activity was considered more compatible with the first possibility (Boyer and Roulland-Dussoix, 1969; Glover, 1970).

Although able to assign mutations resulting in various restriction and

modification phenotypes to the *hsdR* and *hsdS* genes, the complementation data did not establish the effect of a single step mutation in the *hsdM* gene in the absence of mutations in the other genes.

Hubacek and Glover (1970) isolated temperature-sensitive mutants ($r_K^{ts}m_K^+$) in order to select second-step mutants deficient in modification ($r_K^{ts}m_K^-$) at the restrictive temperature of 42°C without creating a lethal phenotype. In all second-step mutants encountered, it was found that the deficiency in restriction was greater than the deficiency in modification, so that any modification-deficient or temperature-sensitive mutant was invariably also restriction-deficient.

Complementation tests between the first-step or second-step mutants and an $r_B^-m_B^+$ (i.e. *hsdR*⁻) F' resulted in diploids fully-functional in both modification and restriction (that is, not temperature sensitive). Thus, the defects seen in restriction and modification in the mutants were due to mutations in the *hsdM* gene in the presence of functional *hsdR* and *hsdS* genes. This demonstrated that the *hsdM* gene product was required for restriction as well as modification.

As the earlier experiments concluded that both the *hsdS* and *hsdR* genes are required for restriction (Boyer and Roulland-Dussoix, 1969; Glover, 1970), all three subunits must be required for the formation of the restriction endonuclease. Because an *hsdR* mutant was impaired in restriction and not methylation, it was concluded that only the *hsdM* and *hsdS* gene products were required for the formation of the active modification methyltransferase.

In vitro analysis of purified enzymatic extracts from various *E. coli* K-12 and B mutants (Kühnlein *et al.*, 1969; Hadi and Yuan, 1974) supported the results of the *in vivo* experiments. While extracts from *hsdR*⁻ mutants were proficient in modification, *hsdS*⁻ mutant enzymes were deficient. However, the two mutant extracts were able to complement each other to give modification proficient enzyme activity.

Sain and Murray (1980) cloned the *hsd* genes of *E. coli* K-12 in λ using a combination of *in vitro* and *in vivo* techniques. Complementation tests confirmed the presence of three *hsd* genes, and analysis of deletion derivatives of the λ *hsd* phages and plasmids carrying the *hsd* region established the order of the genes as *hsdRMS*. Genetic evidence identified two promoters, one upstream of *hsdM* effecting transcription of both *hsdM* and *hsdS* (p_{mod}), and a second upstream of *hsdR* (p_{res}). The direction of transcription from both promoters was the same (figure 1.2). The fact that *hsdM* and *S* have a separate promoter from *hsdR* could allow differential control in the production of modification and restriction activities, so that modification activity could occur in the absence of restriction. Such differential control has been reported for *EcoP1* (Arber, 1974); upon entry of P1 phage into a host cell, the modification activity encoded by the P1 *mod* gene is expressed before the restriction activity. Similarly, Glover and Colson (1969) found evidence for sequential control during conjugation experiments involving the transfer of K-specific *hsd* genes. Such differential control would only be advantageous in such a situation where the genes are being transferred to a new bacterial strain, as the preferred activity of the *EcoK* enzyme complex is dictated by the DNA substrate. Hemimethylated DNA, such as would be found following semi-conserved replication, is a good substrate for methylation but not restriction by *EcoK* (Vovis *et al.*, 1974; Suri *et al.*, 1984a; Kelleher *et al.*, 1991).

Loenen *et al.* (1987) found no evidence that transcription from p_{res} is influenced by the products of the *hsdM* and *S* genes. Nevertheless, when a possibility of coordinated control is prevented, "self destruction" may occur. For example, when a λ *hsdMS* phage enters a cell in which the R polypeptide is expressed by a multicopy plasmid, the cell's own DNA is apparently cleaved (Fuller-Pace *et al.*, 1985; Loenen *et al.*, 1987; Kelleher *et al.*, 1991).

B Enzyme Families

The eleven known type I r-m systems of the *Enterobacteriaceae* have been divided into three families. The Ia (or K) family comprises three systems from

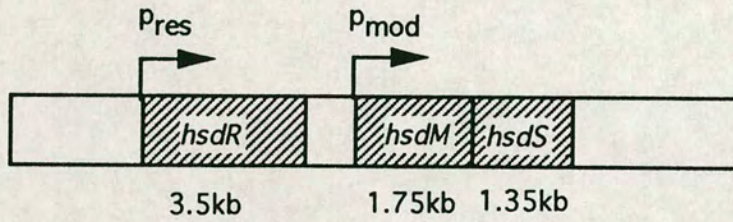


Figure 1.2: The *hsd* region of *E. coli* K-12. The sizes of the genes are given in kilobases; the direction of transcription from the two promoters is indicated by arrows.

E. coli, *EcoK*, *EcoB*, and *EcoD* (from *E. coli* E166), as well as *StySB* and *StySP* from *Salmonella enterica* serovars typhimurium LT2 and potsdam, respectively. The Ib (or A) family has two *E. coli* members, *EcoA* encoded by strain 15T⁻, and *EcoE* from a natural isolate, A58 (Fuller-Pace *et al.*, 1985; Suri *et al.*, 1985b). Additionally, a third member, *CfrA*, has been identified in *Citrobacter freundii* (Daniel *et al.*, 1988; Kannan *et al.*, 1989). While the Ia and Ib family systems are chromosomally-encoded and allelic, in the sense that the *hsd* genes map to the same location on the chromosome (Fuller-Pace *et al.*, 1985), the first members of a third family of systems (Ic or R124) are plasmid-borne. The IncFIV plasmid R124 encodes a type I system, as does the plasmid pDXXI of *E. coli* ET7 (Price *et al.*, 1987a; Piekarowicz *et al.*, 1985).

Within a family, there is strong sequence conservation between the *hsd* genes. The interchangeability of enzyme subunits, described earlier for *EcoK* and *EcoB* (Boyer and Roulland-Dussoix, 1969; Glover, 1970), is also documented for either *EcoK* and *EcoB* and *StySB* (Van Pel and Colson, 1974), and between *StySB* and *StySP* (Bullas and Colson, 1975).

Molecular approaches, using DNA hybridization, immunological tests, and DNA sequence comparisons provide an easy way of subdividing type I systems into families. Cross-reactivity of antibodies raised against the R and M subunits of *EcoK* with other members of the K family as well as DNA hybridization using probes made from the *hsd* region of the *E. coli* K-12 chromosome have demonstrated extensive sequence similarity between the *EcoK*, B and D systems (Murray *et al.*, 1982; Daniel *et al.*, 1988). A weaker, but nonetheless marked similarity is also seen between *EcoK* and the SB and SP systems of *S. enterica* (Murray *et al.*, 1982; Daniel *et al.*, 1988).

Immunological cross-reactivity is not demonstrable between members of the Ia, Ib and Ic families, and DNA hybridization has failed to detect sequence similarity between families (Murray *et al.*, 1982; Fuller-Pace *et al.*, 1985; Price *et al.*, 1987a; Daniel *et al.*, 1988).

Inter- and intraspecific sequence comparisons at the nucleotide and predicted amino acid level support the family classifications. While the *EcoK* M polypeptide has 96.6% amino acid identity with *EcoB* and approximately 95% identity with *StySP* or *StySB*, the level of identity with M polypeptides of *E. coli* type I systems in other families is much lower (Sharp *et al.*, 1992). The M subunit of *EcoK* shares only 32% identity with that of *EcoA*, and 26% identity with *EcoR124*. These values are similar to those seen when the *Salmonella* type Ia family M subunits are compared with *EcoA* or *EcoR124*. Sequence divergence between families is even greater when the R polypeptides are compared (Murray *et al.*, submitted).

Each member of a family has a unique recognition sequence, dictated by the *hsdS* gene product (Table 1.1).

Unlike type II r-m systems (see section 1.2), the multisubunit nature of type I and type III systems means that they have an S subunit common to both restriction and modification. Consequently, there will be more opportunity for evolution of new specificities.

An additional member of the type Ic family is chromosomally-located. The *prf* locus encodes an anticodon nuclease which cleaves host tRNA^{lys} upon infection of phage T4. The nuclease is normally latent; a phage-encoded factor is required for activation (Amitsur *et al.*, 1989; 1992). The locus comprises four genes, three of which, *prfA*, *prfD*, and *prfB* share similarity to the *EcoR124 hsdM*, *R*, and *S* genes, respectively (Linder *et al.*, 1990). *prfC*, which has no similarity to any of the *hsd* genes is inserted into the 120bp that normally separates *hsdS* and *hsdR*. The *prf* genes reportedly show type Ic r-m activity (C. Tyndall and T. Bickle, cited in Amitsur *et al.*, 1992). It has been postulated that this anticodon nuclease could serve as a second defence against phage infection, by preventing expression of phage genomes that have escaped type I restriction. Its existence as a latent protein complex would ensure a rapid response, even when host gene expression has been turned off by the T4 phage (Amitsur *et al.*, 1992).

Table 1.1: Recognition Sequences of Type I R-M Systems

<i>EcoK</i>	5'	AACNNNNNNGTGC TTGNNNNNNNCACG <div style="text-align: center;"> * * </div>	Kan <i>et al.</i> , 1979
<i>EcoB</i>	5'	TGANNNNNNNTGCT ACTNNNNNNNACGA <div style="text-align: center;"> * * </div>	Ravetch <i>et al.</i> , 1978 Lautenberger <i>et al.</i> , 1978
<i>EcoD</i>	5'	TTANNNNNNNGTCY AATNNNNNNNCAGR <div style="text-align: center;"> * * </div>	Nagaraja <i>et al.</i> , 1985c
<i>SrySB</i>	5'	GAGNNNNNNRRTAYG CTCNNNNNNNYATRC <div style="text-align: center;"> * * </div>	Nagaraja <i>et al.</i> , 1985b
<i>SrySP</i>	5'	AACNNNNNNGTRC TTGNNNNNNNCAYG <div style="text-align: center;"> * * </div>	Nagaraja <i>et al.</i> , 1985b
<i>SrySQ</i>	5'	AACNNNNNNRRTAYG TTGNNNNNNNYATRC <div style="text-align: center;"> * * </div>	Nagaraja <i>et al.</i> , 1985a
<i>EcoA</i>	5'	GAGNNNNNNNGTCA CTCNNNNNNNCAGT <div style="text-align: center;"> * * </div>	Suri <i>et al.</i> , 1984b
<i>EcoE</i>	5'	GAGNNNNNNNATGC CTCNNNNNNNTACG <div style="text-align: center;"> * * </div>	Cowan <i>et al.</i> , 1989
<i>CfrA</i>	5'	CAGNNNNNNNGTGG GACNNNNNNNCACC <div style="text-align: center;"> * * </div>	Kannan <i>et al.</i> , 1989
<i>EcoR124</i>	5'	GAANNNNNNRTCG CTTNNNNNNYAGC <div style="text-align: center;"> * * </div>	Price <i>et al.</i> , 1987b
<i>EcoR124/3</i>	5'	GAANNNNNNNRTCG CTTNNNNNNNYAGC <div style="text-align: center;"> * * </div>	Price <i>et al.</i> , 1987b
<i>EcoDXXI</i>	5'	TCANNNNNNNATTC AGTNNNNNNNTAAG <div style="text-align: center;"> * * </div>	Piekarowicz and Goguen, 1986

* indicates methylated adenine residues.

Y indicates either purine may be present.

R indicates either pyrimidine may be present.

N indicates any base may be present.

C Reaction Mechanisms of Type I R-M Systems

Most of what is understood about the mechanisms of DNA recognition, methylation, and cleavage is derived from studies using purified *EcoK* and *EcoB* (for reviews see Bickle, 1982; 1987; Wilson and Murray, 1991).

Although the enzyme alone has been reported to be able to bind DNA, it is believed that binding of the cofactor AdoMet induces a conformational change to an activated form of the enzyme, with greater affinity for DNA (Yuan *et al.*, 1975). This initial specific binding occurs regardless of the methylation state of the target DNA (Yuan *et al.*, 1975; Bickle *et al.*, 1978). A change in conformation has been postulated to be required before further steps in the reaction can occur (Bühler and Yuan, 1978).

After this initial activation, AdoMet is no longer required in the steps leading to endonuclease cleavage of unmodified DNA (Hadi *et al.*, 1975; Yuan *et al.*, 1975). However, free AdoMet is required in later steps of the modification reaction as a methyl donor (Burckhardt *et al.*, 1981a).

The methylation state of the DNA dictates the course the reaction will follow (Burckhardt *et al.*, 1981a,b). In the presence of ATP, the enzyme dissociates from fully-modified recognition sites, enabling it to interact with other DNA molecules (Bickle *et al.*, 1978). If the DNA is hemimethylated, the enzyme methylates the other strand to give a fully-modified target sequence. This methylation is stimulated by ATP and Mg^{2+} (Vovis *et al.*, 1974; Suri *et al.*, 1984a), although the reaction is able to proceed very slowly in the absence of either, but not both, cofactors (Haberman *et al.*, 1972; Burckhardt *et al.*, 1981a; see Bickle, 1982).

Restriction of unmodified DNA follows the formation of a tightly-bound complex in the presence of ATP which may be detected by its retention on nitrocellulose filters (Yuan and Meselson, 1970; Bickle *et al.*, 1978). This filter-binding complex commits the enzyme to the restriction mode. The bound

endonuclease does not turn over following cleavage (Eskin and Linn, 1972a).

Although type I restriction endonucleases remain bound to the recognition site throughout the cleavage reaction (Bickle *et al.*, 1978; Rosamond *et al.*, 1979; Endlich and Linn, 1985a), DNA cleavage occurs at a considerable distance from this site, often several kb away (Horiuchi and Zinder, 1972; Adler and Nathans, 1973; Murray *et al.*, 1973; Bickle *et al.*, 1978; Rosamond *et al.*, 1979; Yuan *et al.*, 1980; see Bickle, 1982).

Studier and Bandyopadhyay (1988) have proposed a comprehensive model for translocation and cleavage of DNA. The enzyme is believed to translocate the DNA towards itself in an ATP-powered reaction (Endlich and Linn, 1985a; Yuan *et al.*, 1980). The choice of cleavage sites, though seemingly random, appears to be where two enzymes, translocating DNA bidirectionally, collide (Studier and Bandyopadhyay, 1988; Brammar *et al.*, 1974). When translocation is synchronized, double-stranded cleavage occurs roughly at the midpoint between two recognition sites (Studier and Bandyopadhyay, 1988).

Following cleavage, ATP hydrolysis continues for long periods (Horiuchi *et al.*, 1974; Eskin and Linn, 1972b; Yuan *et al.*, 1972). This may reflect continued translocation after cleavage by enzymes bound to their recognition sites, or hydrolysis by stalled enzyme molecules (Studier and Bandyopadhyay, 1988).

In the absence of the R polypeptide, the M and S polypeptides are able to form a methyltransferase (Lautenberger and Linn, 1972; Suri and Bickle, 1985; Dryden *et al.*, in press). The methyltransferase functions independently of ATP and Mg^{2+} , requiring only AdoMet (Lautenberger and Linn, 1972). The common requirement of AdoMet for both restriction and modification may provide a control ensuring that restriction occurs only if the cell has the ability to modify (see Wilson and Murray, 1991). The enzyme transfers a methyl group from the cofactor to an adenine residue on either strand in the recognition sequence, resulting in the formation of N⁶-methyladenine (Vovis and Zinder, 1975). Hemimethylated DNA is

the preferred substrate for the methyltransferase (Suri and Bickle, 1984). Unmodified DNA is a poor substrate for the *EcoK* methylase (Suri and Bickle, 1985). The type Ia methyltransferase is unusual among prokaryotic methylases in this substrate preference; the type Ib enzymes are able to methylate hemimethylated and unmethylated DNA equally well (Suri and Bickle, 1985). In this case, modification competes with restriction of unmethylated DNA.

As hemimethylated DNA is the product of semi-conserved replication of fully-modified DNA, the ability to discriminate between unmethylated and hemimethylated DNA should allow the cell to ensure its own DNA is modified, but invading DNA is cleaved. *In vivo*, the type Ia methylation of unmodified DNA may be enhanced by the λ ral gene product while the restriction activity is suppressed (Zabeau *et al.*, 1980; Loenen and Murray, 1986). This implies that Ral alters the way the enzyme responds to the methylation state of the DNA, and in fact Ral-independent mutant *EcoK* methyltransferases have been isolated which appear not to differentiate between hemimethylated and unmethylated DNA (Kelleher *et al.*, 1991).

It may be that *EcoK* and *EcoA* have evolved different strategies to optimize the barrier to infection by foreign DNA. If an infecting λ phage escapes *EcoK* restriction, it should be a poor substrate for modification. As *EcoA* methylates unmodified and hemimethylated DNA equally well, it may have evolved a more effective endonuclease to compensate (see Kelleher *et al.*, 1991). A single unmodified target in λ is restricted five times more effectively by *EcoA* than *EcoK* (Arber and Wauters-Willems, 1970; Murray *et al.*, 1973).

D Sequence-Specific DNA Recognition

The DNA sequences recognized by all known type I r-m systems are asymmetric and bipartite, comprising two short defined regions separated by a non-specific spacer of fixed length. Table 1.1 lists the recognition sequences of the type I systems, with asterisks denoting the methylated adenines. The spacers may be 6-8bp in length, such that the methylated adenines are invariably 10-11bp apart.

Consequently, the methylated adenines lie in successive major grooves, along one face of the double helix (Nagaraja *et al.*, 1985b).

Many recognition sequences are degenerate, so that an enzyme is able to recognize more than one sequence. For example, *Sty*SP recognizes 5' AAC (N₆) GTRC, accommodating either purine at one site in the tetranucleotide component of the recognition sequence. This is a degenerate form of the *Eco*K recognition sequence, 5' AAC (N₆) GTGC. As a result, DNA sequences that are modified by *Sty*SP will be protected from cleavage both by *Sty*SP and *Eco*K (Bullas *et al.*, 1980).

The availability of structurally and functionally similar enzymes within a family, which bind to different sequences, is useful for the study of protein:nucleic acid interactions, and the evolution of new specificities (Gough and Murray, 1983).

Gough and Murray (1983) determined the *hsdS* gene sequences of the related *E. coli* systems *Eco*K, *Eco*B, and *Eco*D for comparative sequence analysis to identify regions involved in recognizing the target DNA. As S polypeptides bind to M polypeptides which may be interchanged between species (Boyer and Roulland-Dussoix, 1969; Glover and Colson, 1969), it was anticipated that there would be strong conservation between different *hsdS* genes. It was anticipated that sequence recognition would be dictated by conserved regions with minor variations accounting for recognition of different sequences. Unexpectedly, the *hsdS* genes showed very little similarity. Sequence conservation was confined to two short regions, one of ~100bp in the middle of the gene, and a second of ~250 bp in the distal portion. These regions were so highly conserved that a common ancestry was indicated. Flanking the conserved region were two stretches of apparently non-homologous DNA, of approximately 600 and 700bp. The *hsdS* genes of the *Salmonella* type Ia systems, *Sty*SP and *Sty*SB, also showed this pattern of homology and non-homology (Fuller-Pace and Murray, 1986; Gann *et al.*, 1987). It was difficult to imagine that such extensive variation between different S polypeptides merely reflected differences in sequence specificity, but this now appears to be the case (Cowan *et al.*, 1989).

A diagrammatic representation of a typical type Ia S polypeptide is given in figure 1.3. The amino terminal 150 amino acids, encoded by the 5' variable region, make up the amino recognition domain. The central conserved region of 35 amino acids separates the amino recognition domain from the carboxyl recognition domain of 180 residues in length, encoded by the 3' variable region. The carboxyl conserved region of 80 amino acids is encoded by the 3' conserved region.

Gough and Murray (1983) hypothesized that the two apparently non-homologous (or variable) regions of the *hsdS* gene could be due to diversification as a result of the independent insertion of modules of DNA. Alternatively, these regions could have a common ancestry made less recognizable over time due to many changes in the sequence. Support for this second hypothesis came from the fact that short stretches of DNA conserved in both sequence and position were identified in *hsdS_K* and *hsdS_D*.

Two possible models were proposed for the roles of the conserved and variable regions in DNA recognition. In the first, the highly-conserved carboxyl region would be involved in a common function such as subunit interaction. The less-stringently conserved central region, while maintaining a common configuration expected of an active site, would have enough variability to encode the different specificities. Differences in the sequences flanking this region would be acceptable as long as they maintained the necessary tertiary structure.

In the second model, favoured by Gough and Murray (1983), the two variable regions would encode domains of the S polypeptide imparting the specificity of recognition. These two domains could correlate with recognition of the two specific regions of the recognition sequence, and sequence variation would account for the different specificities.

Comparison of the predicted amino acid sequences of the *EcoK*, *EcoB* and *EcoD* S polypeptides identified a repeated sequence (Argos, 1985). These domains overlapped the central and ^{carboxyl} conserved regions, but were not confined to them (figure

1.3). Secondary structure predictions indicated that these repeats would be α -helical in nature, compelling Argos to propose that the S polypeptide acts as a "pseudo-dimer" in DNA recognition, with each putative α -helical region recognizing one of the two defined components of the recognition sequence. As the repeats extended into the variable regions, this could allow for greater sequence diversity than would the model of Gough and Murray (1983) in which sequence specificity was altered by changes in the central conserved region.

The "Argos repeats" have since been identified in all known type I S polypeptides for which sequence data are available, and coincide with the only sequence similarity common to all type I S polypeptides (Kannan *et al.*, 1989). Within these repeats there is the same degree of similarity between families as that detected within a single polypeptide. It has been suggested that these repeats are the only visible remnants of a gene duplication in the Ia family (Argos, 1985; Gann *et al.*, 1987).

P1-mediated cotransduction of the *hsd_{SP}* genes with *serB* into an *E. coli* strain carrying the *hsd_{SB}* genes resulted in the isolation of a recombinant strain encoding a system with a novel specificity, designated *StySQ* (Bullas *et al.*, 1976). Heteroduplex analysis of the *StySP*, *StySB*, and *StySQ* *hsd* genes revealed that the *hsdS* gene conferring SQ specificity derived its 5' variable region from the *hsdS_{SP}* gene and its 3' variable region from *hsdS_{SB}* (Fuller-Pace *et al.*, 1984). Nucleotide sequence analysis localized the crossover to the longest region of identity within the central conserved region (Fuller-Pace and Murray, 1986). Accordingly, the *StySQ* recognition sequence was shown by Nagaraja *et al.* (1985a,b) to comprise the trimeric component of the *StySP* recognition sequence and the pentameric component of the *StySB* recognition sequence (figure 1.4). These results confirmed that the recombination event that generated the new specificity had reassorted two DNA recognition domains each specifying recognition of one of the two defined sequence components. Comparison of the amino acid sequences of the conserved regions of the *StySP* and *StySB* S polypeptides placed constraints on the localization of the two domains. Fuller-Pace and Murray (1986) suggested that the amount of variation in

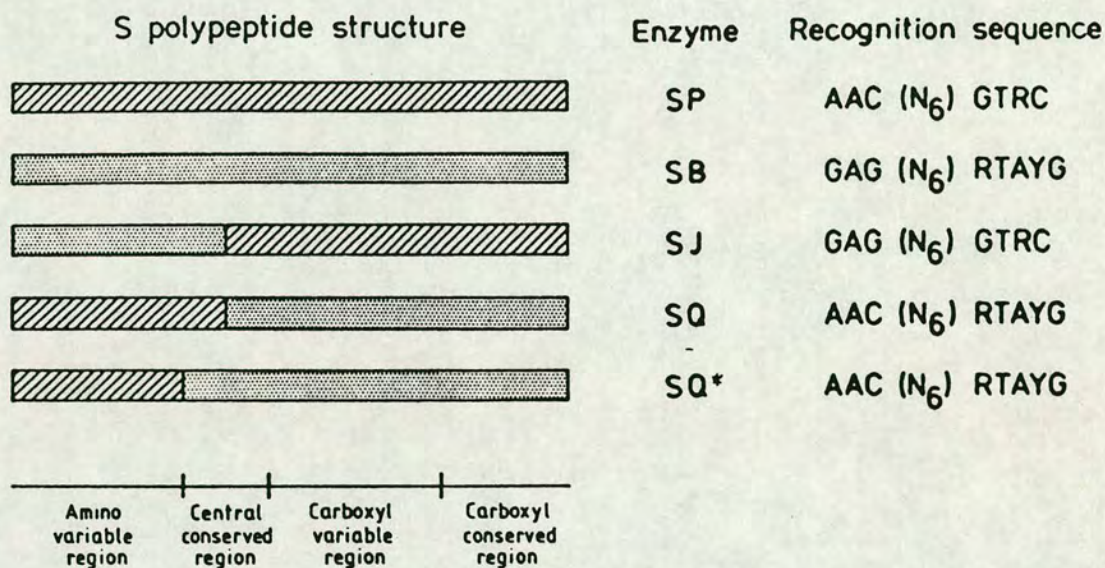


Figure 1.4: Schematic diagram of the hybrid S polypeptides produced by homologous recombination of the *StySB* and *StySP* *hsdS* genes, and their resulting recognition sequences. Regions originating from *StySP* are hatched, those from *StySB* are stippled. The regions of conservation and variation in the S polypeptides are indicated at the bottom of the figure; the amino and carboxyl variable regions are referred to in the text as the amino and carboxyl recognition domains. Recombination of the S polypeptides, and thus reassortment of the recognition domains, results in an enzyme that recognizes a hybrid target sequence. (From Cowan *et al.*, 1989)

this region was too little to allow recognition of different target sequences, and speculated that at least part of the variable regions were involved in the determination of sequence specificity.

Circumstantial support for the correlation of the variable regions with the recognition domains was provided by sequence comparison of the *hsdS* genes of *EcoK* and *StySP*, both of which recognize the trimeric sequence 5' AAC. In this case, the 5' variable regions are conserved in a region which is normally highly variable between different *hsdS* genes (Fuller-Pace and Murray, 1986). This is consistent with their sharing one recognition domain which, during the course of evolution, has become associated with different second domains.

To test the prediction that the S polypeptides of *StySP* and *StySB* contain two separate, independent recognition domains, Gann *et al.* (1987) made the recombinant gene of reciprocal structure, namely with the 5' half of the *StySB hsdS* gene and the 3' half of the *StySP hsdS* gene, with the crossover located in the same portion of the central conserved region that generated *StySQ*. This new specificity gene, designated *hsdS_{SJ}*, was shown to encode a polypeptide recognizing the predicted hybrid sequence (figure 1.4).

These results suggested that DNA recognition is specified by two independent domains encoded by the two variable regions of the *hsdS* gene. However, the recombination events generating *StySJ* and *StySQ* also reassorted minor differences in the central conserved region (Fuller-Pace and Murray, 1986), so that the model was not conclusively proven. In fact, the crossover point lay within one of the repeats which Argos (1985) had implicated in DNA sequence recognition. Between the 5' end of the central conserved region and the point of exchange that resulted in the formation of *StySQ* and *StySJ*, the parental genes differ in four codons. In order to demonstrate that sequence specificity is defined solely by the variable regions of the S polypeptide, Cowan *et al.* (1989) used site-directed mutagenesis of the *hsdS_{SQ}* gene to change all four codons to those of *StySB*. This gene, *hsdS_{SQ*}*, has only the 5' variable region of the *hsdS_{SP}* gene, the remainder being identical to that of *hsdS_{SB}*

(figure 1.5). The specificity of *StySQ** was shown *in vivo* to be identical to that of *StySQ*, conclusive evidence that the 5' variable region encodes a functional domain which confers the specificity for the trinucleotide component of its target sequence (Cowan *et al.*, 1989). By analogy, it is expected that the 3' variable region of the *hsdS* gene confers recognition of the tetra- or pentanucleotide component of the target sequence.

Heteroduplex analysis of two members of the type Ic family, *EcoR124* and *EcoDXXI* revealed an *hsdS* gene structure similar to that of Ia family members (Gubler *et al.*, 1992). Two variable regions of 420 and 540bp are separated by a constant region of 180bp, with a second constant region of 170bp at the 3' end. *In vitro* construction of hybrid *hsdS* genes from *EcoR124* and *EcoDXXI* resulted in a molecule recognizing a hybrid target (Gubler *et al.*, 1992), providing further support that the variable regions of the *hsdS* gene dictate specific sequence recognition.

Predicted amino acid sequences of the S polypeptides encoded by members of the Ib family of type I enzymes has revealed an organization analogous to that of the Ia and Ic families, with two extensive variable regions separated by conserved regions (Cowan *et al.*, 1989; Kannan *et al.*, 1989). Unlike the Ia enzymes, the Ib specificity polypeptides have an amino terminal conserved region of ~100 amino acids, upstream of the variable region that corresponds to the amino recognition domain (encoded by the 5' variable region) (Kannan *et al.*, 1989). The Ib family S polypeptides all possess a repeated sequence in addition to the Argos repeats that is not present in type Ia or Ic S polypeptides. This repeat, detectable even at the nucleotide level, is found in the amino and central conserved regions (Kannan *et al.*, 1989). These may be examples of a further duplication event during the course of evolution of the Ib enzymes.

EcoA and *EcoE* both recognize the trinucleotide 5'GAG as the trimeric component of their target sequences (Cowan *et al.*, 1989). Like *EcoK* and *StySP*, the 5' halves of the *EcoA* and *EcoE* *hsdS* genes are remarkably similar, while the 3' variable regions are easily identified. Comparison of the *EcoA* and *EcoE* S

polypeptide sequences with members of the Ia family detected similarity in only one enzyme, *StySB* (Cowan *et al.*, 1989). Residues 101-247 of either *EcoA* or *EcoE*, and the amino recognition domain of *StySB* showed 44% identity. These amino acids correspond to the amino recognition domains of these S polypeptides, and are the only regions of similarity in otherwise divergent enzymes from different families. Thus, unrelated enzymes share a domain that is capable of recognizing the trinucleotide 5'GAG (Cowan *et al.*, 1989).

There is no evidence to implicate particular residues in a recognition domain in defining specificity, although when enzymes recognize different sequences, no similarity is detectable. Enzymes recognizing the same trinucleotide component share a high degree of similarity, though less so when enzymes are members of different families. This may reflect the need to interact with their respective M subunits. It is possible that most residues in a recognition domain are important in defining sequence recognition, some being directly involved in DNA binding, while many others influence by presentation (Cowan *et al.*, 1989).

The identification of similar domains dictating recognition of identical sequence is in agreement with the concept of modules of DNA that have been acquired by a number of enzymes independently (Gough and Murray, 1983). Another mechanism for alteration of specificity is illustrated by members of the type Ic family. *EcoR124* and *EcoR124/3* recognize sequences that differ only in the length of their non-specific spacers (Price *et al.*, 1989). The *EcoR124/3* specificity arose spontaneously from *EcoR124*, and the expression of the two specificities is able to switch reversibly *in vivo* (Glover *et al.*, 1983). The *hsdS* genes were found to be identical except for the fact that a 12bp sequence in the centre of the genes is present twice in the *EcoR124 hsdS* gene and three times in the *EcoR124/3 hsdS* gene (Price *et al.*, 1989). Heteroduplex analysis of either the *EcoR124* or *EcoR124/3 hsdS* genes with the *EcoDXXI hsdS* gene, which specifies recognition of a sequence sharing neither component of the recognition sequence with either *EcoR124* or *EcoR124/3*, localizes these repeats to the conserved region between the two variable regions (Gubler *et al.*, 1992). Indeed, sequence analysis revealed that the central conserved region of

EcoDXXI contains this 12bp sequence repeated three times, corresponding to a recognition sequence with a non-specific spacer of seven nucleotides. Gubler and Bickle (1991) determined the effects of altering the length and sequence of the conserved region. It was found that a number of amino acid substitutions were tolerated with no effect on enzyme activity, so long as the number of repetitive elements was maintained. Reduction in the number of repetitive elements to less than two, or greater than three, led to a decrease in restriction activity. Gubler and Bickle (1991) concluded that the repetitive amino acid motif forms part of a flexible interdomain linker, perhaps responsible for correct positioning of the recognition domains on the DNA. The change in spacer length between *EcoR124* and *EcoR124/3* is likely to have occurred either by unequal crossing over during recombination or by slippage during replication resulting in an extra repeat of four amino acids in the S polypeptide of *EcoR124/3* (Price *et al.*, 1989). This crossing-over could provide a further natural mechanism for alteration of sequence specificity.

1.5 RESTRICTION OF MODIFIED DNA

Some restriction systems in *E. coli* are specific for methylated DNA sequences. The *mcrA*, *mcrBC*, and *mrr* loci encode such systems (see Wilson and Murray, 1991 for a review; Raleigh, 1992 for a review of McrBC).

The McrA and McrBC systems were originally detected by their ability to restrict T-even phages which have 5-hydroxymethylcytosine (HMC) replacing the cytosine residues in their DNA. The functions were named RglA and RglB (restricts glucoseless phages) because the sensitivity to restriction was dependent on the failure of the phage to glucosylate its DNA. Normally, T-even phages glucosylate their HMC residues, rendering the DNA resistant to most restriction systems (see section 1.6 D). As a failure to modify the HMC residues is rare in the T-even phages, the Rgl systems appeared to be redundant. However, a much wider role for the Rgl systems later became apparent. Difficulties encountered in cloning the genes for certain modification methylases in *E. coli* were shown to be modification-dependent, as well as sequence-specific (Noyer-Weidner *et al.*, 1986; Raleigh and Wilson,

1986). λ or plasmid DNA containing 5-methylcytosine or N⁴-methylcytosine was restricted while DNA containing 6-methyladenine was not (Noyer-Weidner *et al.*, 1986; Raleigh and Wilson, 1986; Blumenthal *et al.*, 1985). The name of these restriction systems was changed to Mcr to reflect their *in vivo* functions more accurately. The restriction phenotypes of McrA and McrB are mutually independent, with different sequence specificities (Raleigh and Wilson, 1986).

It was later verified that *rglA* and *rglB* did correspond to *mcrA* and *mcrB*, respectively, when the genes were cloned and mapped (Raleigh *et al.*, 1989). McrA is encoded by a prophage-like element, e14, at a site on the chromosome near *purB* at 25 minutes on the *E.coli* map. The *mcrB* gene is found at approximately 99 minutes on the *E.coli* map, in the "immigration control region", so-called because it also contains the *hsd* genes and *mrr*, which encodes another restriction system (Raleigh *et al.*, 1989). *mcrB* was mapped to a position downstream of *hsdS*.

The *mcrA* and *mcrB* genes have been cloned (Raleigh *et al.*, 1989; Krüger *et al.*, 1992; Ross *et al.*, 1987; Sozhamannan and Dharmalingam, 1988) and *mcrB* sequenced (Dila *et al.*, 1990; Ross *et al.*, 1989).

Genetic and sequence analysis shows that the *mcrB* locus actually comprises two genes, *mcrB* and *mcrC* (Dila *et al.*, 1990; Ross *et al.*, 1989). The two genes direct the synthesis of three proteins; a 51kD protein (McrB_L) and a 33kD protein (McrB_S) are translated in the same frame from *mcrB*, while a single protein of 38kD is encoded by *mcrC* (see Raleigh, 1992). The function of McrB_S is not known. McrC expands the range of modified sequences restricted by McrB_L; some sequences are restricted by McrB_L alone, while others (such as the T-even phages) are only restricted by McrB_L⁺ McrC⁺ hosts (Dila *et al.*, 1990). Unlike the other *E.coli* restriction systems, McrBC has been shown *in vitro* to have an absolute requirement for Mg²⁺ and GTP, and is in fact inhibited by ATP (Sutherland *et al.*, 1992). A potential GTP-binding site has been located in the McrB polypeptide sequence (Dila *et al.*, 1990).

McrBC restriction *in vivo* and *in vitro* is sequence-specific. A consensus sequence that is sensitive to McrBC, RmC, has been identified (Raleigh and Wilson, 1986; Sutherland et al., 1992). Purified McrBC has been shown to cleave both strands of the DNA at multiple positions in a small region. This cleavage occurs between modified cytosine residues in the sequence RmC (N₄₀₋₈₀) RmC (Sutherland et al., 1992). The modified residues must be appropriately spaced, though they can be on different strands of DNA. In this respect, McrBC restriction appears to resemble type I restriction, in which it is believed that two bound enzymes translocate DNA until they collide and cleave at this point, between two unmodified recognition sequences (see section 1.6 D). However, there is no evidence for a DNA translocation mechanism for McrBC (Sutherland et al., 1992).

The biological role of Mcr restriction is unclear. Raleigh and Wilson (1986) have suggested that it provides a back-up to the restriction and modification systems of *E. coli*, by providing protection against phages which have escaped restriction due to modified DNA. That these systems pose a threat to bacteriophages is suggested by the fact that phages have developed an antirestriction nuclease (Arn) function specific against Mcr restriction (Dharmalingam and Goldberg, 1976; Dharmalingam et al., 1982; see section 1.6 D).

1.6 ANTIRESTRICTION SYSTEMS

Double-stranded DNA, whether phage or plasmid, will be subject to host-encoded restriction upon entering a bacterial cell. A number of phages and plasmids have developed mechanisms for avoiding restriction. Antirestriction mechanisms have been found in virtually every phage that has been examined (reviewed by Krüger and Bickle, 1983) and in certain conjugative plasmids (Delver et al., 1991; Belogurov et al., 1992; Read et al., 1992).

Some of these antirestriction functions involve inhibition of restriction by acting either on the enzyme itself or by destroying cofactors, by modification of infecting DNA to make it refractory to cleavage, or by stimulation of the modification

function of the r-m system. Others act by as yet unsolved mechanisms. This section provides a review of some of these mechanisms, that have apparently evolved to protect infecting DNA from restriction by type I enzymes.

A Avoidance of Recognition Sequences

Perhaps the simplest way to evade restriction is to avoid inclusion of recognition sequences for r-m systems in the DNA. Natural selection for the loss of restriction sites might be expected to act in phages that lack any antirestriction functions. That phages can evolve to lose recognition sequences has been demonstrated in the laboratory. For example, by cycles of infection on *EcoRI* restricting and non-restricting strains, Murray and Murray (1974) isolated λ phages lacking one or more of the five *EcoRI* recognition sites.

Sharp (1986) compared the observed frequencies of restriction sites to the expected frequencies in the genomes of a number of DNA phages. Calculations of expected frequencies took into account that frequencies of the 16 possible dinucleotides vary between different phylogenetic groups (Nussinov, 1984; Sharp, *et al.*, 1985). Attention was focused on 6bp palindromes, as this is the length of typical recognition sites for type II restriction enzymes. The expected frequency of type I recognition sites was also calculated by multiplying the frequencies of the two recognition components.

No clear pattern of avoidance of type I and type III recognition sites was evident (Sharp, 1986). This might not be surprising, as a number of the phages used in his analysis, for example T7, encode antirestriction functions active against type I enzymes. However, in a number of phages, including T7, certain 6bp palindromes were underrepresented in the genome. In particular, these sequences corresponded to those that are recognized by type II restriction enzymes encoded by bacteria which are hosts to T7, namely members of the *Enterobacteriaceae*. A similar result has been reported by Schroeder *et al.* (1986).

Interestingly, the *Bacillus* phage $\phi 29$, though not showing an avoidance of enterobacterial r-m recognition sites, has fewer sequences recognized by *Bacillus* type II enzymes.

Two phages, G4 and λ , did not show such an obvious lack of sequences. In λ , though 6bp palindromes are significantly underrepresented, the effect is not more pronounced for known recognition sequences than other 6bp sequences. Sharp (1986) suggested that λ might not conform to the pattern because it is a temperate rather than lytic phage, and while integrated in the host chromosome could avoid restriction.

Recognition sequences for the type III enzyme *EcoP1* occur more frequently in the phage T7 genome than expected by chance. This may be explained by the fact that the sequence overlaps with that for T7 primase, an essential enzyme which should be subject to positive selection (Schroeder *et al.*, 1986).

Thus, though antirestriction functions for type II r-m systems may not exist in a number of phages, the avoidance of the recognition sequence for these enzyme may provide a passive defense (Sharp, 1986).

B Inhibition of Enzyme Function

a) T3 and T7

T7 phages are resistant to both restriction and modification by the type I enzymes *EcoK* and *EcoB*, *SlySB* and *EcoR124*, though not type II r-m systems such as *EcoRV* (Mark and Studier, 1981; Miyazaki *et al.*, 1982) or the type III system, *EcoP1* (Moffatt and Studier, 1988). This resistance to restriction is due to the early gene, *0.3*, which encodes an antirestriction function sometimes referred to as *Ocr* (overcomes classical restriction; see Krüger and Bickle, 1983, for a review). T3 phage also carry a *0.3* gene with a similar function, though sequence analysis reveals no obvious similarity between the *0.3* genes of T3 and T7 (Hughes *et al.*, 1987).

Much of the analysis of the 0.3 protein function has been carried out on the T7 0.3 gene by Studier and his colleagues (Mark and Studier, 1981; Dunn and Studier, 1981; Bandyopadhyay *et al.*, 1985; Moffatt and Studier, 1988). Purified T7 0.3 protein has been shown to inhibit the cleavage and ATPase activities of *EcoB* and *EcoK* *in vitro* (Dunn and Studier, 1981; Mark and Studier, 1981; Bandyopadhyay *et al.*, 1988).

To determine the mechanism of action of the 0.3 gene product, Bandyopadhyay *et al.* (1985) examined the effect of adding purified ³⁵S-labelled protein to a mixture of *EcoK* or *EcoB* and DNA, under a variety of conditions. It was shown that the 0.3 protein does not bind to DNA. Rather, it binds tightly to the type I holoenzyme, and to one or other of the small enzyme subunits (that is, either M or S), effectively preventing the enzyme from binding DNA. Addition of the 0.3 protein to a mixture of *EcoK* and fully-modified DNA resulted in the inactivation of the enzyme, probably as it dissociated from the recognition site. If the 0.3 protein was added to *EcoK* and unmodified DNA before the cofactor ATP, restriction was effectively blocked. If, however, addition of ATP preceded the 0.3 protein, formation of the filter-binding complex (see section 1.4 C) and ATP hydrolysis were not blocked. Nevertheless, although degradation of the DNA was evident, the degree of restriction was less in the presence of the antirestriction protein. Thus, the sensitive step appeared to be after ATP-powered DNA translocation, but prior to cleavage.

Methylation of hemimethylated DNA was also found to be inhibited by the 0.3 protein when added at any stage of the reaction (Bandyopadhyay *et al.*, 1985).

Such experiments suggest that the binding site on the type I enzyme for the 0.3 protein is accessible until the formation of the filter-binding complex, and remains protected through the ATP-driven DNA translocation. When translocation is finished, the binding site once again becomes accessible before cleavage occurs (Bandyopadhyay *et al.*, 1985). The fact that modification is also affected supports the theory that the 0.3 protein affects a common function, namely binding of the type

I enzyme to the DNA. Protection of T3 and T7 DNA requires *de novo* synthesis of the 0.3 protein (see Krüger and Bickle, 1983).

The double-stranded, linear T7 genome enters a host cell upon infection left end first (Pao and Speyer, 1973). Entry of the T7 DNA is coupled to transcription (Chamberlin *et al.*, 1970; Zavriev and Shemyakin, 1982), initially by the host's RNA polymerase. Transcription proceeds in a wave from left to right along the DNA, perhaps providing a mechanical process for drawing the phage DNA into the cell (Zavriev and Shemyakin, 1982). Though the early DNA before the 0.3 gene is accessible to RNA polymerase, it is not vulnerable to restriction by *EcoK* or *EcoB* *in vivo* (Eskridge *et al.*, 1967; Eskin *et al.*, 1973; Krüger *et al.*, 1977). By labelling T7 DNA with ³²P, Moffatt and Studier (1988) demonstrated that wild type T7 DNA was not susceptible to restriction by *EcoK*, B, or P1 until 6 to 7 minutes after infection, even though there are 24 *EcoP1* recognition sites and one *EcoB* site in the DNA entering the host cell over this period. Mutant T7 DNA with *EcoK* recognition sequences introduced into this region were also refractive to restriction. As the 0.3 gene enters the host cell approximately 3 to 4 minutes after infection begins, this block to restriction would mean that the 0.3 gene would be transcribed and translated in time to protect the T7 DNA. After this time, it was found that T7 was still protected from *EcoK* and B restriction, though it became susceptible to cleavage by *EcoP1*, no doubt because 0.3 protein does not afford protection against this restriction enzyme (Moffatt and Studier, 1988). The authors suggested that, upon entry into the host, the T7 DNA is segregated to a compartment of the cell, accessible to enzymes involved in transcription and translation, but not restriction.

In addition to the *Ocr* phenomenon discussed above, the T3 phage 0.3 gene also encodes an enzyme which hydrolyses intracellular AdoMet, thus destroying a cofactor both for restriction and modification (Spoerel *et al.*, 1979). *In vitro* studies with purified T3 0.3 protein have distinguished between the *Ocr* and AdoMet-ase functions (Spoerel *et al.*, 1979). Although the T7 0.3 protein lacks this function, it is nevertheless as effective as the T3 0.3 protein at protecting against type I

restriction. Additionally, mutants have been isolated which do not hydrolyse AdoMet yet still express the 0.3 function (Spoerel *et al.*, 1979; see Krüger and Bickle, 1983).

The protection afforded by the 0.3 protein is also effective for any other DNA introduced into cells along with T3 or T7. Plasmid DNA entering by conjugation as well as by transformation has been shown to be protected (reviewed in Krüger and Schroeder, 1981).

b) IncI and IncN plasmids

Conjugative plasmids also may encode mechanisms for the evasion of host restriction upon entry into a cell. A number of the large, self-transmissible plasmids of the IncI and IncN incompatibility groups, including ColIb-P9 and pKM101, possess homologous antirestriction functions (Boulnois and Wilkins, 1978; Belogurov *et al.*, 1985; Delver *et al.*, 1991; Kotova *et al.*, 1988, cited in Delver *et al.*, 1991; Wilkins *et al.*, 1991; Belogurov *et al.*, 1992; Read *et al.*, 1992).

The non-essential *ard* gene (*ard* for antirestriction determinant) of the IncI1 plasmid ColIb-P9 has been shown to alleviate restriction of the plasmid by *EcoK* when entering cells via conjugation (Delver *et al.*, 1991; Belogurov *et al.*, 1992; Read *et al.*, 1992). The *ard* gene maps in the ColIb leading region (Delver *et al.*, 1991; Belogurov *et al.*, 1992; Read *et al.*, 1992), and thus is transferred early during conjugation.

Restriction of λ DNA by members of all known type I r-m systems is alleviated 26-fold if the infected cell expresses the Ard function and over 100-fold if expression of Ard is from a ColIb-P9 plasmid derepressed for conjugal activity (Delver *et al.*, 1991; Read *et al.*, 1992). This suggests that Ard plays a role in establishment of conjugated plasmids. However, the Ard function has no effect versus restriction by the type II enzyme, *EcoRI*, nor the type III system, *EcoP1* (Delver *et al.*, 1991; Read *et al.*, 1992).

There have been conflicting reports as to whether the Ard function also affects type I modification. This depended on the level of expression of *ard*. A single round of infection of λ V.K in a $r_K^+ m_K^+$ host carrying a derepressed ColIb plasmid produces fully-modified phage (Read *et al.*, 1992). However, overproduction of Ard by a multicopy plasmid interferes with both *EcoK* restriction and modification of λ DNA (Delver *et al.*, 1991; Belogurov *et al.*, 1992; Read *et al.*, 1992). This suggests that Ard is inhibiting a feature common to both restriction and modification, though at low levels Ard preferentially blocks cleavage.

A homologous antirestriction function, designated ArdA, has been detected in cells carrying the IncN plasmid PKM101 (Belogurov *et al.*, 1985; 1992). The ArdA protein is of a similar size and has approximately 60% amino acid identity to Ard although little sequence similarity is generally seen between members of these two plasmid groups (Falkow *et al.*, 1974). The *ardA* gene, like *ard*, provides protection against restriction by type I r-m systems, but has additionally been shown to have some effect against type II restriction, though not type III (Belogurov *et al.*, 1992). Two genes, *ardR* and *ardK*, were also detected in conjunction with *ardA* (Belogurov *et al.*, 1992). Attempts to clone the *ardA* gene on a multicopy plasmid had failed, suggesting a potentially lethal phenotype. However, expression of the *ardK* gene product in *cis* or in *trans* inhibited this lethality (Belogurov *et al.*, 1992). The *ardR* gene product apparently controls ArdA activity, as transposon mutagenesis of this gene increases ArdA activity 200-fold (Belogurov *et al.*, 1992). Such regulatory genes have not been identified in conjunction with *ard* (Belogurov *et al.*, 1992), though a "derepression" effect similar to that of inactivation of *ardR* was noticed upon deletion of sequence upstream of *ard* (Belogurov *et al.*, 1992).

The proposed amino acid sequence for both Ard and ArdA implies an acidic nature (Delver *et al.*, 1991; Belogurov *et al.*, 1992; Read *et al.*, 1992). The 0.3 protein of phage T7 is also very acidic. It may be that Ard and ArdA, like the 0.3 protein, interact with the DNA binding site of type I enzymes (Delver *et al.*, 1991; Belogurov *et al.*, 1992).

It has been demonstrated that *Ard* must be produced in the recipient cell in order that transfer of plasmid DNA by conjugation may evade type I restriction (Read *et al.*, 1992). For this to occur, presumably synthesis of the complementary strand to the single-stranded DNA entering the cell is required. How this double-stranded DNA could evade type I restriction is a paradox. The antirestriction function does not protect plasmid DNA entering an *E.coli* cell by transformation (Read *et al.*, 1992), presumably because transforming DNA is double stranded and therefore a substrate for restriction endonuclease cleavage. However, Read and his colleagues (1992) have proposed a model similar to the T7 model of Moffatt and Studier (1988), in which the incoming conjugative DNA is targeted to a compartment of the cell, safe from type I restriction endonucleases, but accessible to host polymerases involved in complementary strand synthesis and transcription of the *ard* gene.

A second process for evasion by conjugative plasmids of both type I and type II restriction may also exist (Read *et al.*, 1992). This mechanism involves transfer of multiple copies of the plasmid, known to occur when certain plasmid genes are prevented from being expressed during transfer to the recipient cell (Boulnois and Wilkins, 1978). A delay of expression of these genes could allow transfer of further copies of the plasmid DNA if the first was restricted, resulting in substrate saturation and overcoming the restriction barrier.

C Stimulation of Host Modification Function

Unmodified DNA is only poorly modified by the *EcoK* methyltransferase (Suri and Bickle, 1985). However, when unmodified λ^+ phage is propagated in an $r_K^- m_K^+$ derivative of *E.coli* K-12, the resulting phage are fully modified even after a single round of infection. This is due to the action of the product of the non-essential *ral* gene of λ (Zabeau *et al.*, 1980; Loenen and Murray, 1986; Kelleher *et al.*, 1991). *Ral* expressed by modified $\lambda^+.K$ phage which has infected an $r_K^+ m_K^+$ host will protect superinfecting $\lambda^+.0$ phage (Zabeau *et al.*, 1980). Indeed, superinfecting T7.0 phage is protected from restriction in *E.coli* K-12 pre-infected by $\lambda^+.K$. Unmodified $\lambda^+.0$ is restricted as efficiently as $\lambda ral^- .0$ upon infecting a restricting host, implying that

Ral is not produced in time to protect its own DNA (Loenen and Murray, 1986). While Ral greatly reduces restriction, it also enhances modification (Zabeau *et al.*, 1980; Loenen and Murray, 1986). The *ral*-encoded protection is confined to the type Ia family of r-m systems; no protection is afforded against *EcoA* or *EcoR124* (Loenen and Murray, 1986), nor the type II restriction enzymes *EcoRI* and *EcoRII* or the type III r-m system *EcoP1* (Zabeau *et al.*, 1980).

Isolation of *ral*-deficient λ mutants allowed detailed mapping of the *ral* gene to a region of the early leftward operon near *N* (Debrouwere *et al.*, 1980). An open reading frame located between genes *N* and *Ea10* has been assigned to *ral*, and this open reading frame (orf) has been cloned in a plasmid expression vector (Loenen and Murray, 1986), enabling further analysis of its action.

Unlike the 0.3 protein of phages T3 and T7, Ral does not prevent DNA methylation and hence must not interfere with binding of the enzyme to DNA. Rather than directly blocking the restriction activity of *EcoK* or B, Ral appears to alter the kinetics of methylation (Zabeau *et al.*, 1980; Loenen and Murray, 1986; Loenen *et al.*, 1987). Type Ia r-m systems differ from other families of type I enzymes in that the methyltransferase modifies unmethylated DNA only very poorly, the preferred substrate for methylation being hemimethylated DNA (Vovis *et al.*, 1974; Suri and Bickle, 1985). It is possible that Ral might modulate both restriction and modification activity of *EcoK* by altering the substrate specificity of the methylase.

Kelleher *et al.* (1991) compared the modification of λral^- phages following a single round of infection of an $r_K^- m_K^+$ host when the incoming phages were either modified or unmodified. It was demonstrated that the progeny phage resulting from infection by a K-modified λral^- phage were fully-modified, presumably because the daughter DNA molecules after a single cycle of replication are hemimethylated and consequently good substrates for methylation. The progeny from infection by unmodified λral^- phage were incompletely modified even after multiple rounds of infection on an $r_K^- m_K^+$ host.

It was proposed that the presence of Ral changes the *EcoK* enzyme from functioning as a maintenance methylase, requiring methylation of one DNA strand for efficient methylation of the other strand, to a *de novo* methylase, able to modify DNA lacking any imprint of methyl groups (Kelleher *et al.*, 1991).

Ral-like activity has also been detected in the hybrid phage λ_{rev} (Toothman, 1981), and the lambdoid phage P22 (Semerjian *et al.*, 1989).

D Incorporation of Unusual Bases in Phage DNA

The T-even phages (T2, T4 and T6) are naturally resistant to the majority of restriction enzymes they encounter during infection of their hosts (reviewed by Revel, 1983; Krüger and Bickle, 1983; Mosig and Eiserling, 1988). The DNA of these phages is modified at several levels. Cytosine is completely replaced by the unusual base, hydroxymethylcytosine (HMC; Wyatt and Cohen, 1952). Additionally, the HMC residues are glucosylated to varying degrees in the different phages (Lehman and Pratt, 1960).

T4 DNA containing non-glucosylated HMC is resistant to the action of a number of restriction enzymes, including *EcoB* (see Krüger and Bickle, 1983), but not *EcoP1* (Revel and Georgopoulos, 1969), although methylation is blocked by HMC (Hattman, 1983). In addition, the *mcrA* and *mcrBC* genes encode restriction enzymes which cleave DNA sequences containing methylated cytosines, including HMC (Revel, 1967; 1983; see section 1.5). Originally designated *rglA* and *rglB* (restriction of glucose-less phages), these proteins will inactivate HMC-containing T-even phage DNA, though only if it lacks glucosylation (Revel, 1967; 1983; Raleigh and Wilson, 1986). McrA has been shown to cleave sequences containing non-glucosylated HMC residues in all T-even phages, while McrBC is active against T2 and T4, but not T6, DNA (Revel, 1967).

Glucosylation is catalysed by α - and β -glucosyltransferases, using host uridine

diphosphoglucose as a donor of glucose residues (Kornberg *et al.*, 1959; 1961). The extent of glucosylation of HMC residues varies between the T-even phages. In phage T4, all HMC residues are modified, 70% by α -glucosyltransferases, and 30% by β -glucosyltransferases, while in T2 and T6, 25% of HMC residues are not glucosylated (Mosig and Eiserling, 1988). However, T6 DNA has been shown to contain glucose residues which have been further glucosylated (Mosig and Eiserling, 1988). Fully-glucosylated HMC-containing DNA is resistant to most restriction systems in *E. coli*, including Mcr, suggesting that such an antirestriction mechanism evolved from pressure exerted by the Mcr (or Rgl) systems (see Krüger and Bickle, 1983). In addition, the T-even phages encode a gene, *arn* (for anti-restriction endonuclease), which produces an inhibitor of McrBC (RglB) (Dharmalingam and Goldberg, 1976; Dharmalingam *et al.*, 1982). Although expressed early during phage infection, Arn is not effective in preventing Mcr from cleaving the phage's non-glucosylated DNA. However, superinfecting phage is protected by the *arn* gene product (Dharmalingam and Goldberg, 1976; Dharmalingam *et al.*, 1982).

E Phage Encoded DNA Modifications

a) T-even phages

Phages T2 and T4, but not T6, encode a DNA adenine methylase (Dam) which modifies approximately 0.5% to 1.0% of the adenines to 6-methyladenine, mainly in the sequence 5'GATC. This methylation is nonessential, as unmethylated phage are viable (Hattman, 1970; 1983). Little DNA sequence similarity is seen between the T4 and *E. coli dam* genes, though considerable similarity at the protein level is evident. However, the phage DNA is a poor substrate for the host Dam methylase as the HMC residues interfere with DNA recognition (Hattman, 1983). The only proven role of this adenine methylase is prevention of restriction by *EcoP1* in T2 and T4 mutants lacking glucosylation. Non-glucosylated T2 and T4 phages, normally sensitive to *EcoP1* restriction, may mutate to a new form which is resistant (Revel and Georgopoulos, 1969; Hattman, 1970). This mutation has been shown to be correlated with an increase in phage DNA methylation (Hattman, 1970).

The variety of antirestriction mechanisms evolved by the T-even phages suggests a long coexistence between host and phage.

b) Phage Mu

Bacteriophage Mu expresses a novel modification function which provides a degree of protection against type I r-m systems. The *mom* gene (*modification of Mu*), which is nonessential for phage growth (Toussaint, 1976), specifically converts adenine to N⁶-(1-acetamido)-adenine (Hattman, 1979; Swinton *et al.*, 1983) within the sequence 5'(C/G)-A-(C/G)-N-Py 3' (Kahmann, 1984). The source of the acetamido group has not yet been identified, nor have the enzymatic requirements and steps been determined (Kahmann and Hattman, 1987). Approximately 15% of the adenines in the phage genome are modified by Mom (Hattman, 1979). As a result, Mu which is Mom⁺ is relatively insensitive to several type I and type III r-m systems *in vivo* including *EcoK*, *EcoB*, *EcoA* and *EcoP1* (Toussaint, 1976). Notably, all of these enzymes have recognition sequences which overlap the Mom target sequence (see Kahmann, 1984; Kahmann and Hattman, 1987). *EcoRI*, which is not sensitive to modification by Mom, does not have such an overlapping target site. However, Mom-specific modification protects Mu at least partially against a number of other type II systems *in vitro* (see Kahmann, 1984).

Mom not only modifies Mu DNA, but also acts in *trans* on plasmids, superinfecting λ and induced prophages, and the host chromosome (Toussaint, 1976).

Expression of the *mom* gene occurs late in the Mu lytic cycle, after replication of the phage DNA. Expression is deleterious to the host (Kahmann *et al.*, 1985) and consequently, expression is under tight control requiring both phage- and host-encoded transactivators (reviewed by Kahmann and Hattman, 1987).

c) Phage T1

Though it is restricted by EcoP1 (Drexler and Christensen, 1961), phage T1 is unaffected by *EcoK* and *EcoB* restriction (Lederberg, 1957). It has been proposed that T1 specifies a methyltransferase that modifies some residues in its DNA which may include recognition sequences for some restriction enzymes (Wagner *et al.*, 1979). A demonstration that during T1 infection of *E. coli dam*⁻ strains the 5'GATC sites present in the phage DNA were almost completely methylated lends support to such a theory (Auer and Schwieger, 1984).

F The Dar Antirestriction Function of Phage P1

Although phage P1 is sensitive to type II and type III restriction, it is only weakly restricted by type I r-m systems. This is due to the products of two nonessential phage genes, *darA* and *darB* (*dar* for *defense against restriction*) (Iida *et al.*, 1987a; reviewed by Krüger and Bickle, 1983; Yarmolinsky and Sternberg, 1988).

Phages that are *darA*⁻ have lowered efficiencies of plating on members of the type Ia and Ib r-m systems (including *EcoK*, B, A, *StySB*, SP and SQ) as well as decreased efficiency in generalized transduction (Iida *et al.*, 1987a). P1*darB*⁻ phages, while sensitive to type Ia restriction, are no longer restricted by the type Ib system *EcoA* (Iida *et al.*, 1987a). This implies that the *darA* gene product provides protection against type Ib restriction. It follows that the *darB* protein protects P1 from type Ia restriction, though *darB* appears not to function in the absence of *darA* (Iida *et al.*, 1987a).

The *darA* and *darB* genes have been mapped by transposon mutagenesis (Iida *et al.*, 1982; 1987a,b). The *darA* gene forms part of an operon between the resident *IsI* element and the invertible C-segment of P1. The *darB* gene is clearly separate from the *darA* operon, between the *res* and *mod* genes and the *IsI* element (Iida *et*

al., 1987a).

One *darA*⁻ phage, P1CmTc1, was obtained as a plaque-forming recombinant of P1Cm1 and P1Tc1 (Mise and Arber, 1976; Iida *et al.*, 1982; 1987a,b). P1Cm1 carries a 2.8kb *IsI*-flanked *Cm*^R transposon at the resident *IsI* element (Arber *et al.*, 1978; Iida *et al.*, 1987b), while P1Tc1 has a *Tn10* (*Tc*^R) transposon of 9.3kb inserted between the *IsI* element of P1 and the C-segment (Iida *et al.*, 1987b). Recombination between P1Cm1 and P1Tc1 would result in a *Tc*^R*Cm*^R P1 phage with an oversized genome (Iida *et al.*, 1987b). This may have been an intermediate to the formation of P1CmTc1, in which an *Is*-mediated inversion, and a deletion in the *dar*-C-segment region reduce the genome to a packageable size (Iida *et al.*, 1987b).

SDS-polyacrylamide gel electrophoresis of caesium chloride gradient-purified wild type P1 particles and mutant derivatives have demonstrated that the Dar proteins are associated with the phage particle. The *darA* gene product is a protein of 68 kilodaltons (kD), while the *darB* gene product is 200kD (Iida *et al.*, 1987a; Streiff *et al.*, 1987). Interestingly, phages which are genotypically *darA*⁻ lack both the DarA and DarB proteins, while *darB*⁻ phages lack DarB only (Iida *et al.*, 1987a). It appears that the *darB* gene product is not packaged in the absence of *darA*, explaining the finding that *darA*⁻ phages are phenotypically DarA⁻ and B⁻ (Iida *et al.*, 1987a).

Evidence reported by Iida and coworkers (1987a) suggests that the *dar* genes encode phage head internal proteins, injected into the host in association with the phage DNA. Firstly, a Dar⁺ phage is able to protect a Dar⁻ phage following coinfection of an r⁻m⁻ host. All of the phage recovered after this coinfection are phenotypically Dar⁺, even when genotypically *dar*⁻. This suggests that the Dar⁺ phage is producing a factor that protects the *dar*⁻ DNA. However, the Dar function does not act in *trans*; coinfection of a restricting host (r⁺m⁺) by Dar⁺ and Dar⁻ phages results in only the Dar⁺ phage escaping restriction. Therefore, it appears that the Dar proteins only protect the DNA they are associated with during infection, though DNA packaged with the proteins will be protected upon further infection. This is supported by evidence that λ DNA and even host DNA, packaged in a P1

head, is protected from restriction though λ packaged in λ heads is not.

The mechanism of action of the Dar proteins is not clear. It has been shown that Dar does not inactivate the restriction enzyme, as infection of an $r_K^+ m_K^+$ by P1.K does not prevent superinfecting λ V.P1 from restriction (Iida *et al.*, 1987a).

Although *dar* and *mom* map to regions of their respective phage genomes that have structural and functional similarities, they prevent restriction by unrelated mechanisms. Demonstration of *EcoK* activity *in vitro*, assayed by ATPase activity, shows that while Mom chemically modifies the Mu genome so that purified Mu DNA is refractive to *EcoK* restriction, purified DNA from a Dar^+ P1 phage is sensitive (Iida *et al.*, 1987a). Therefore, Dar does not confer a chemical modification on the DNA.

Interestingly, it has been shown that modification of P1 DNA is stimulated by Dar during lytic growth (Iida *et al.*, 1987a). In a single round of infection on an $r^- m^+$ host, Dar^+ phages are modified 100-fold more effectively than Dar^- phages.

The *darA* gene has been cloned on a multicopy plasmid (Streiff *et al.*, 1987). As a result, it was possible to show that the *darA* gene is not expressed in cells which have not been infected by P1 or in cells which are lysogenized by $P1dar^-$ phages. However, after induction, the protein is produced, and phenotypically Dar^+ phages are formed. This implies that expression of the *darA* gene must be controlled by a phage-encoded factor (Streiff *et al.*, 1987).

The related phage P7 has also been shown to be phenotypically Dar^+ , and to produce proteins which cross-react with Dar-specific antibodies (Streiff *et al.*, 1987), suggesting they encode homologous antirestriction functions.

The wide range of antirestriction mechanisms expressed by coliphages and plasmids, particularly against type I restriction and modification systems, implies that these systems provide a major barrier to infection and establishment of foreign DNA

elements in enterobacterial species. The diversity of the antirestriction functions is apparently the result of convergent evolution of elements addressing the same problem. Other restriction systems, for example Mcr and type II r-m enzymes, may have evolved as a backup to the type I r-m systems.

1.7 DNA TRANSFER

Bacterial DNA transfer, the stable acquisition of genetic information from another organism, occurs by three principal mechanisms, conjugation, transduction, and transformation.

A Transformation

Transformation is likely to be the most general and least species-specific type of DNA transfer system. It involves the uptake of naked DNA fragments from the surrounding environment by bacteria. It requires a specialized capacity of the bacterium to uptake DNA. Described for both Gram-positive (e.g. *Streptococcus* and *Bacillus*) and Gram-negative (e.g. *Haemophilus* and *Neisseria*) organisms, this attainment of "competence" is an active and complex process involving changes in cell chemistry and physiology.

In Gram-negative organisms such as *H. influenzae*, DNA in the form of linear fragments or covalently-closed circular plasmids is bound to receptors on the cell surface and taken up as single-stranded linear molecules (see, e.g., Levin, 1988b). These may be integrated into the chromosome by homologous recombination, or recircularized as plasmids.

E. coli is not believed to be naturally transformable. A growth-dependent physiological state of competence has not been identified in this species (see Mazodier and Davies, 1991), though treatment with divalent cations can render it competent for DNA uptake in the laboratory. Whether *E. coli* encounters the conditions for competence in its natural environment is not known. There have been no reports of

transformation of *E. coli* occurring in nature, suggesting that this mechanism does not play a major role in gene exchange in this species (Levin, 1981). Nonetheless, it is possible that plasmid DNA acquired by transformable species could then be transferred to *E. coli* by conjugation (Amábile-Cuevas and Chicural, 1992).

B Conjugation

Many plasmids specify conjugation systems that enable transfer of DNA between bacteria (see Willetts and Wilkins, 1984; Willetts, 1985; Hardy, 1986, for reviews). This ability relies on the expression of the *tra* (transfer) operon, which in the widely-studied F plasmid comprises approximately 24 genes. The *tra* gene products specify the synthesis of sex pili in the donor and the formation of cell-cell contact, initiate and mediate transfer of DNA, and in some cases initiate replication of the complementary strand of the single-stranded DNA molecule in the recipient cell. Initiation of transfer requires site-specific cleavage of a DNA strand at *oriT*. Non-conjugative (i.e. lacking the *tra* operon) plasmids may be mobilized by the *tra* gene products provided in *trans*, if they carry *oriT* and a 1.8kb *mob* (mobilization) region.

Conjugative plasmids are also able to mediate transfer of chromosomal DNA, following integration of the plasmid via recombination of *Is* elements common to the plasmid and the chromosome. The result is an Hfr (*High frequency of recombination*) strain. Expression of the *tra* genes by the integrated plasmid can result in the transfer of both itself and the genomic DNA to a recipient cell, providing the potential for recombination between donor and recipient DNA. The Hfr has the potential to transfer the entire chromosome. However, mating complexes usually break before this can occur.

Integration of conjugative plasmids is reversible. Occasionally the plasmid may excise aberrantly, carrying some chromosomal sequence with it. These F' (or R', etc.) plasmids may transfer DNA to a recipient cell, but transfer of a specific marker will occur at a lower frequency than transfer mediated by Hfr strains.

F plasmids and some ColV plasmids differ from most conjugative plasmids in that their *tra* genes are expressed constitutively. Almost all F-like and I-like sex pili, for example, are expressed only in a small proportion of their host cells (0.1%; see Willetts, 1985). The *traJ* gene product is believed to be an inducible repressor. Derepressed plasmids are expected to place a burden on their host cells, and thus are at a selective disadvantage (Hardy, 1986). Consequently, conjugation of chromosomal DNA will occur at a much lower frequency than that seen with F.

Until recently, it was expected that conjugation as a means of DNA transfer would have a narrow host range, restricted to closely-related species or genera. Indeed, the sex pili of different groups of bacteria differ markedly in structure, and the interaction between the pilus and the receptor on the recipient cell surface was expected to be highly specific (see Mazodier and Davies, 1991). *Streptococcus* species, for example, encode a conjugation system that apparently lacks pili (Clewell and Weaver, 1989). However, conjugation systems have now been shown to transfer DNA between a wide variety of genera (see Mazodier and Davies, 1991), including from *Streptococcus* to *E.coli* (Trieu-Cuot *et al.*, 1988), and from *E.coli* to the Gram-positive genera *Streptomyces* and *Mycobacterium* (Gormley and Davies, 1991). Within the *Enterobacteriaceae*, the recipient cells can form mating pairs with all possible donor groups. The association between more distantly-related bacteria is mechanically weaker, and consequently genetic transfer would be expected to be less efficient.

Thus, conjugation could provide a route for the dissemination of bacterial genes in nature. This would provide a mechanism for achieving rapid adaptation to environmental change, such as the introduction of antibiotics. However, stable acquisition of chromosomal markers will be limited by the requirement for sequence similarity for recombination. Recombination is abolished by 10-20% sequence divergence, due to the action of the mismatch repair proteins of the recipient (Rayssiguier *et al.*, 1989).

An obvious barrier to the transmission of DNA is the presence of restriction

and modification systems in the recipient cell. *Streptococcus pneumoniae*, a naturally-transformable bacterial species, has apparently evolved a "policing" system whereby DNA injected by bacteriophages may be differentiated from transforming DNA and cleaved before a productive infection can ensue (Cerritelli *et al.*, 1989). The *DpnII* r-m system of *S.pneumoniae* recognizes and restricts DNA lacking the appropriate modification. DNA escaping cleavage is modified by the corresponding methyltransferase, encoded by the *dpmM* gene. Additionally, the system includes a methylase, DpnA, specific for single-stranded DNA. Therefore, incoming single-stranded DNA, such as transforming DNA, will not be a substrate for restriction and will instead be methylated. Consequently, if the DNA recircularizes and replicates to become a double-stranded plasmid, it will be hemimethylated and thus not a substrate for *Dpn* restriction. Similarly, if the DNA is incorporated into the chromosome, it will become hemimethylated. DNA introduced by phage will be double-stranded and thus not a substrate for the DpnA methylase. Reports that conjugative plasmid transfer was susceptible to *DpnII* restriction (Guild *et al.*, 1982) were explained by the fact that though conjugation may involve transfer of a single strand of DNA, concurrent synthesis of the complementary strand might take place before DpnA methylation (Cerritelli *et al.*, 1989).

E.coli is not naturally transformable and there is no r-m system in *E.coli* which would correspond to the Streptococcal *DpnII* system. However, some conjugative plasmids have evolved systems whereby they can avoid cleavage following transfer to a restricting strain (see section 1.4 B). If, as Read *et al.* (1992) proposed, incoming conjugative DNA is targeted to a compartment of the cell, it will not be exposed to restriction endonucleases until after it has replicated and expressed the Ard antirestriction protein. The fact that the Ard function is enhanced when the *tra* operon is derepressed (Read *et al.*, 1992) supports the importance of Ard during conjugation. Ard was described for plasmids of the IncI and IncN incompatibility groups. Whether or not conjugative plasmids of other incompatibility groups, including F, also express the Ard function is not currently known.

Additionally, conjugative plasmids may overcome the restriction barrier by

substrate saturation (see section 1.4 B). This process allows the plasmid to overcome both type I and type II restriction (Read *et al.*, 1992). By transferring multiple copies of the plasmid DNA, the endonucleases may become overwhelmed, allowing a copy of the plasmid to evade restriction and express *Ard*.

C Transduction

Generalized transduction may take place when a phage, such as P1, inadvertently packages host rather than phage DNA during lytic growth and transfers it to another cell (for reviews see Masters, 1985; Yarmolinsky and Sternberg, 1988). P1 packages DNA by a headful mechanism, and so has the capacity to transfer as much as 100kb of donor DNA, the length of the P1 genome. It has been estimated that approximately 0.3% of the total phage particles released during lysis are transducing phages (Ikeda and Tomizawa, 1965).

Recombination and thus stable acquisition of chromosomal markers in the recipient cell occurs within an hour of being injected, and requires the presence of a functional *recA* gene (Sandri and Berger, 1980). Only 1-2% of injected DNA is ever incorporated into the recipient chromosome as an intact fragment, and the size of the incorporated DNA was reported to be generally 10kb in size (Sandri and Berger, 1980). Again, stable acquisition of markers requires a degree of similarity.

Specialized transducing phages, for example λ , contain bacterial DNA linked to phage DNA (reviewed by Weisberg, 1987). These transducing phages result from aberrant excision of prophage DNA following induction, so that chromosomal DNA adjacent to the phage integration site is packaged. When this DNA is injected into a recipient cell, acquisition of DNA markers results from lysogeny of the phage, usually independent of the recombination systems of the recipient cell, or by RecA-mediated recombination with existing chromosomal sequences. Specialized transduction is a more limited form of DNA transfer than generalized transduction, both in the amount of chromosomal DNA transferrable, and the specific genes packaged.



Unmodified transduced DNA would be expected to experience a similar fate to infective phage upon injection into a restricting host. Phage-encoded antirestriction functions, however, may serve to protect the packaged DNA. As described in section 1.4 F, the Dar proteins of phage P1 protect from type I restriction any DNA packaged in a P1 phage head, including host DNA (Iida *et al.*, 1987a). Transduction efficiencies of host markers packaged in Dar⁻ phage heads was found to be over 10³-fold less than in Dar⁺ phage heads when injected into K- or B-restricting strains.

The majority of documented antirestriction functions appear to have evolved primarily to protect DNA from type I restriction. An exception could be the *ardA* gene product of the conjugative plasmid pKM101, which has been shown to provide a degree of protection against the type II r-m system *EcoRI* (Belogurov *et al.*, 1992). Additionally, some generalized mechanisms such as substrate saturation would be expected to act against different types of restriction. The array of antirestriction functions evolved to perform the same role implies that type I systems pose a serious threat to foreign DNA entering a host.

CHAPTER TWO

MATERIALS AND METHODS

2.1 STRAINS

A Bacterial Strains

See Table 2.1.

B Phage Strains

See Table 2.2.

C Plasmids

a) *hsd_K* plasmids

See figure 2.1 A.

b) *hsd_A* plasmids

See figure 2.1 B.

c) Other plasmids

See Table 2.3.

2.2 ENZYMES AND CHEMICALS

DNA polymerase (Klenow fragment) and T4 DNA ligase were purchased from Boehringer Mannheim (UK) Ltd; restriction enzymes were from Boehringer Mannheim (UK) Ltd, Bethesda Research Laboratories (UK) Ltd or Northumbria Biologicals Ltd; DNase I, RNase A and lysozyme from Sigma Chemical Company.

Standard agarose was supplied by Miles Laboratory Ltd.; low melting point agarose, ethidium bromide and bromophenol blue were all from Sigma Chemical Company; GEL-aseTM was from Cambio Ltd.

TABLE 2.1 BACTERIAL STRAINS^a

A Laboratory Strains Representing Different Type I R-M Specificities

Strain	Relevant Features	Reference/Source
C600	<i>hsd_K</i>	Appleyard (1954)
NM515	<i>hsd_D</i>	Daniel <i>et al.</i> (1988)
L4001	<i>hsd_{SB}</i>	Bullas and Colson (1975)
L4002	<i>hsd_{SP}</i>	Bullas and Colson (1975)
NM551	<i>hsd_{SQ}</i>	Fuller-Pace <i>et al.</i> (1984)
NM627	<i>hsd_{SJ}</i>	Gann <i>et al.</i> (1987)
WA2899	<i>hsd_A</i>	Arber and Wauters-Willems (1970)
NM649	<i>hsd_{CfrA}</i>	Daniel <i>et al.</i> (1988)
NM144	<i>hsd_{R124}</i>	Lab. Collection
C-1a	<i>E. coli</i> C (no <i>hsd</i> genes)	Bertani (1968)

^aUnless otherwise stated, all the bacteria are derivatives of *E. coli* K-12.

B Strains Used in P1 Screen for Type I R-M Systems in Wild Type *E.coli*

a) Wild Type *E.coli* Strains

Strain	Relevant Features	Reference/Source
629	<i>hsd_B</i> +	Daniel <i>et al.</i> (1988)
653 ^b		Daniel <i>et al.</i> (1988)
BLD4		Daniel <i>et al.</i> (1988)
BRLET2		Daniel <i>et al.</i> (1988)
BRLET13		Daniel <i>et al.</i> (1988)
RM01A		Daniel <i>et al.</i> (1988)
RM66A		Daniel <i>et al.</i> (1988)
RM74A		Daniel <i>et al.</i> (1988)
E161 ^b		K. Kaiser
VB1		V. Barcus

^bThese strains were isolated by K. Cartwright at the Western General Hospital, Edinburgh, and classified by him as *E.coli* using the Analytical Profile Index 20E system.

B Strains Used in P1 Screen for Type I R-M Systems in Wild Type *E.coli*
(cont'd)

b) Laboratory Strains

Strain	Relevant Features	Reference/Source
ED8739 (NM1)	<i>supF hsdS</i>	Borck <i>et al.</i> (1976)
NM555	<i>hsdMS_AΔ2</i>	Fuller-Pace <i>et al.</i> (1985)
A58	<i>hsd_E⁺</i>	Duguid <i>et al.</i> (1955)
CB303	BMH71-18 carrying pDS1108	A.C.Boyd
NM550	<i>hsdSBΔ9</i> derivative of L4001; <i>str^R</i> , <i>Lac⁺</i>	Fuller-Pace <i>et al.</i> (1984)
VB2	<i>hsdR124</i> genes in W3110	V. Barcus
NM498	<i>dnaCts hsdRΔ4</i>	Lab. Collection
NM667	<i>(deoD-thr)Δ hsdR⁺ dnaC⁺</i> derivative of NM498	Lab. Collection
SM346	<i>hsd⁺ mutL::Tn5</i>	D. Thaler
WA3012	<i>hsd⁺ mutS::Tn5</i>	D. Thaler
NM718	<i>mutS::Tn5</i> derivative of NM667	V. Barcus
NM719	<i>mutL::Tn5</i> derivative of NM667	V. Barcus

C Strains Used in Construction of λ Libraries

Strain	Relevant Features	Reference/Source
W3110	<i>hsd_K</i>	Yanofsky and Ito (1966)
NM679	(<i>mcrC-mrr</i>) Δ^a derivative of W3110	Lab. Collection
NM715	<i>recA13</i> derivative of W3110	Lab. Collection
VB3	NM679(P2 cox)	V.Barcus
NM496	<i>hsdS</i> Δ derivative of C600	Lab. Collection
K803	<i>hsdS3</i>	Wood (1966)
NM522	<i>hsdMS</i> $\Delta 5$	Gough and Murray (1983)
ED8654	<i>hsdR</i>	Murray <i>et al.</i> (1977)
NM621	<i>hsdR recD</i>	Whittaker <i>et al.</i> (1989)
N3098	<i>ligts7 hsdK⁺</i>	Pauling and Hamm (1968)

^aThis deletion spans the *hsd* genes.

D *E.coli* Strains From the ECOR Collection^a Used in DNA Hybridization

a) Group A

ECOR No.	Previous Designation
1	RM74A
2	STM1
3	WIR1
4	RM39A
5	RM60A
6	RM66C
7	RM73C
8	RM77C
9	FN98
10	ANI
11	C97
12	FN59
13	FN10
14	P62
15	FN3
16	RM191F
17	RM200Q
18	RM210F
19	RM210J
20	RM213I
21	RM213K
22	RM215C
23	RM183E
24	FN33
25	MSI

D *E.coli* Strains From the ECOR Collection Used in DNA Hybridization
(cont'd)

b) Non-Group A Strains

ECOR No.	Previous Designation
31	RM12
37	RM44B
42	DAR1
45	RM201C
46	RM202F
48	C90
58	RM185S
65	RM202I
67	RM217T
68	RM224H
69	RM45EM
70	RM70B

* The ECOR collection, made up of 72 natural isolates of *E.coli* from a variety of hosts and geographical locations, was compiled by Ochman and Selander (1984a).

TABLE 2.2 PHAGE STRAINS

Strain	Relevant Features	Reference/Source
NM5	λ^+	N.E. Murray
NM63	$\lambda cI26$	N.E. Murray
NM123	$\lambda cI857 Sam7$	N.E. Murray
NM144	$\lambda h^{82} b522 cI^-$	N.E. Murray
NM220	$\lambda h^{80} cI26$	N.E. Murray
NM243	λvir	N.E. Murray
NM574	$ral^+ cI857$ λ replacement vector	Borck <i>et al.</i> (1976)
NM1048	$hsdM_K S_K$ genes in $\lambda NM574$	Sain and Murray (1980)
NM1049	$hsdM_K$ in $\lambda NM574$	Sain and Murray (1980)
EMBL3	$ral^- cI^-$ λ replacement vector	Frischauf <i>et al.</i> (1983)

TABLE 2.2 PHAGE STRAINS (cont'd)

Strain	Relevant Features	Reference/Source
AD12	5.3kb <i>EcoRI</i> fragment of <i>E.coli</i> C DNA spanning a region that in <i>E.coli</i> K-12 would contain <i>hsd_K</i> .	Daniel <i>et al.</i> (1988)
P1Cm1	<i>darA</i> ⁺ <i>darB</i> ⁺ <i>c1ts225</i>	Arber <i>et al.</i> (1978)
P1CmTc1Δ <i>darA</i>	Dar ⁻ <i>c1ts225</i>	Iida <i>et al.</i> (1982)
<i>P1-15darB</i> :: IS30:: <i>Tn9c1ts</i>	<i>darB</i>	Caspers <i>et al.</i> (1984)
P1kc		Lennox (1955)

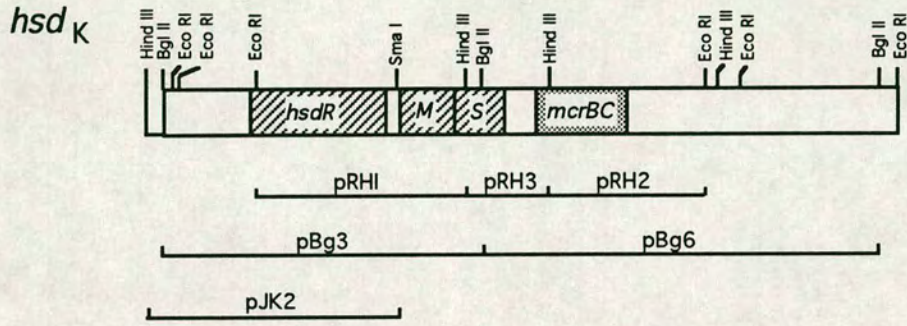
Figure 2.1: Plasmid Strains.

A. *hsd_K*-specific plasmids. The cloned DNA in the plasmids is indicated beneath a map of the *hsd* region of *E. coli* K-12. pRH1, pRH2, pRH3, pBg3, and pBg6 are all derivatives of pBR322 (Sain and Murray, 1980). The extremities of the *E. coli* K-12 DNA are flanked by *Bgl* II targets. The *Hind* III site to the left of the map is in pBR322 and was used to subclone the *hsdR* gene in pAT153 to give pJK2 (Kelleher *et al.*, 1991).

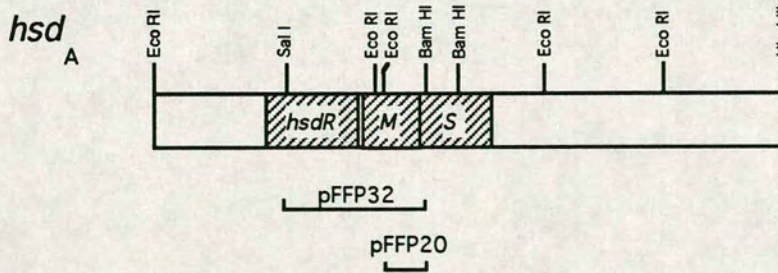
B. *hsd_A*-specific plasmids. The cloned DNA is indicated beneath a map of the *hsd* region of *E. coli* 15T. Both pFFP32 and pFFP20 are derivatives of pBR322 (Daniel *et al.*, 1988).

Figure 2.1

A. *hsdK* Plasmids



B. *hsd_A* Plasmids



1 kb

TABLE 2.3 OTHER PLASMIDS

Plasmid	Relevant Features	Source/Reference
pBR322	ColE1-like replicon carrying Amp ^r , Tet ^r , and multiple cloning sites.	Bolivar <i>et al.</i> , 1977
pDS1108	ColE1::Tn3 (<i>mob</i> ⁺ , Amp ^r)	A.C.Boyd

Ampicillin (Penbritin) was purchased from Beecham Pharmaceuticals; chloramphenicol, kanamycin (sulphate), rifampicin (rifamycin AMP) and tetracycline (hydrochloride) were all from Sigma Chemical Company; streptomycin (sulphate) from Northumbria Biologicals Ltd;

Vitamin B₁, Dithiothreitol (DTT), 2-mercaptoethanol and Isopropyl- β -D-Thiogalactoside (IPTG) were supplied by Sigma Chemical Company; 5-Bromo-4-Chloro-3-Indoyl- β -D-Galactoside (X-Gal) by Boehringer Mannheim (UK) Ltd.

Spermidine trihydrochloride and putrescine (tetramethyldiamine) were supplied by Sigma Chemical Company.

The ECL Gene Detection System and the λ *In Vitro* Packaging Kit were purchased from Amersham International plc; the DIG Luminescent Detection System was from Boehringer Mannheim (UK) Ltd. AMPPD was obtained from Tropix Inc.

Hybond-N filters were supplied by Amersham International plc; x-ray film and film cassettes by DuPont (UK) Ltd or Amersham International plc.

Standard laboratory chemicals were purchased from Sigma Chemical Company, BDH Chemicals Ltd or Fisons Scientific Equipment.

2.3 MEDIA

L-Broth: Difco Bacto tryptone (10g), Difco Bacto yeast extract (5g), NaCl (5g), distilled H₂O to 1 litre; adjusted to pH 7.2 with NaOH before autoclaving.

L-Agar: Difco Bacto tryptone (10g), Difco Bacto yeast extract (5g), NaCl (5g), Difco agar (15g), distilled H₂O to 1 litre; adjusted to pH 7.2 with NaOH before autoclaving.

BBL-Agar: Baltimore Biological Labs. trypticase (10g), NaCl (5g), Difco agar(10g), distilled H₂O to 1 litre.

BBL Top Agar: As for BBL agar but only 6.5g Difco agar added per litre.

Minimal Agar: Difco agar (4g), distilled H₂O to 300ml. After autoclaving, the following sterile solutions were added: 5x Spizizen salts (80ml), glucose (4ml, 20% w/v), vitamin B₁ (0.1ml, 2mg/ml), amino acids as required.

H₂O Top Agar: Oxoid agar No.3 (0.6g), distilled H₂O to 100ml.

Lactose MacConkey Agar: Bacto MacConkey agar base (20g), distilled H₂O to 500ml. After autoclaving, a sterile solution of lactose was added (25ml, 20% w/v), and antibiotic if required.

2x TY Broth: NaCl (10g), Difco Bacto yeast extract (10g), Difco Bacto tryptone (16g), distilled H₂O to 1 litre.

5x Spizizen Salts: (NH₄)₂SO₄ (10g), K₂HPO₄ (70g), KH₂PO₄ (30g), tri-sodium citrate dihydrate (30g), MgSO₄ · 7H₂O (1g), distilled H₂O to 1 litre.

Phage Buffer: KH₂PO₄ (3g), Na₂HPO₄ (7g), NaCl (5g), MgSO₄ · 7H₂O (1ml, 0.1M), CaCl₂ (10ml, 0.1M), gelatin (1ml, 1% w/v), distilled H₂O to 1 litre.

Other specialized buffers are described in the text.

All media were sterilized by autoclaving (15lb/in², 15 minutes).

2.4 STANDARD SOLUTIONS

TE Buffer: Tris (10mM), EDTA (1mM), adjusted to appropriate pH with HCl.

20x SSC: NaCl (3M), tri-sodium citrate (0.3M).

20 x TBE Buffer: Tris (1.78M), boric acid (1.78M), EDTA (0.05M).

10 x TAE Buffer: Tris (0.4M), sodium acetate trihydrate (0.2M), EDTA (0.02M), NaCl (0.18M), adjusted to pH 8.2 with acetic acid.

Ethidium Bromide: 10mg/ml in distilled H₂O. Stored at 4°C, protected from light.

Antibiotic Solutions: See Table 2.4.

2.5 MICROBIAL TECHNIQUES

A Long Term Storage of Bacterial Cells

A single colony was picked into L-broth (5ml) and incubated overnight with aeration at the appropriate temperature. A flamed metal wire was then dipped into the culture and used to inoculate an L-agar stab. This was incubated overnight with a loose cap, which was subsequently tightened and sealed with parafilm. The stab was stored at room temperature.

B Preparation of Plating Cells

A fresh overnight was diluted 20-fold into L-broth and grown at 37°C to mid-log phase ($O.D._{650}=0.5$, approximately 2×10^8 cells/ml). The cells were pelleted in a bench centrifuge (3,000g, 5 min.), resuspended in half the volume of MgSO₄ (10mM), and stored at 4°C.

C Preparation of λ Plate Lysates

A single plaque was picked into 1ml of phage buffer containing a few drops of chloroform, and vortexed briefly. The phage suspension (0.1-0.2ml) was added to 0.2ml plating cells and allowed to adsorb at room temperature for 15 minutes. Molten top agar (3ml) containing MgCl₂ (10mM) was then added and the mixture poured onto a fresh L-agar plate. The plate was incubated without inversion until confluent cell lysis was observed (usually 6-8 hours). The agar was then overlaid

TABLE 2.4 ANTIBIOTICS

Antibiotic	Solvent	Stock Concentration	Final Concentration
Ampicillin	H ₂ O ^a	100mg/ml	100µg/ml
Chloramphenicol	Ethanol (100% v/v)	20mg/ml	15µg/ml
Kanamycin	H ₂ O	25mg/ml	25µg/ml
Rifampicin	DMF	100mg/ml	as required
Streptomycin	H ₂ O	20mg/ml	10µg/ml
Tetracycline	Ethanol (50% v/v)	15mg/ml	10µg/ml

^aAntibiotics dissolved in H₂O were sterilized by filtration.

All antibiotics were stored at -20°C.

When used in agar plates, the antibiotic was added immediately before pouring.

with L-broth (4ml) and stored at 4°C overnight. The L-broth was decanted and a few drops of chloroform added. Cell debris was pelleted by centrifugation (3,000g, 5 min.), the phage suspension transferred to a fresh bottle and stored at 4°C.

D Phage Titration

Serial dilutions of the phage stock were made in phage buffer. A 0.1ml volume of each dilution was mixed with 0.2 ml plating cells and allowed to adsorb for 20 minutes at room temperature. The mixture was plated in top agar (3ml) containing MgSO₄ (10mM) on dry BBL agar plates and incubated overnight with inversion at 37°C.

E Spot Tests

Spot tests may be used to obtain a rough estimate when titring lysates, or determining the efficiency of plating of phage on a particular strain. It allows a number of dilutions to be tested on the same agar plate.

A lawn of cells was prepared by adding 0.2ml plating cells to top agar (3ml) and pouring onto dry BBL plates. Serial dilutions of the phage stock were made, and aliquots (10μl) were spotted onto the lawn. The spots were allowed to dry before overnight incubation.

F Construction of λ*cI857att* Lysogens

λ phages which are *att* are unable to integrate by site-specific recombination, and therefore rely on homologous recombination to integrate into the bacterial chromosome. λ phages with a temperature-sensitive repressor gene (*cI857*), will form such lysogens when grown at the permissive temperature of 30°C.

To construct the lysogens, a lawn of freshly-prepared plating cells was spotted with 10μl of serially diluted λ*cI857* and incubated overnight at 30°C. Lysogens were

selected from the cells arising within the area of lysis by streaking samples onto L-agar plates seeded with approximately 10^9 p.f.u. each of two homoimmune λcI^- phages of different host ranges. Cells that grew were purified and tested for lysogeny by their ability to grow at 30°C but not 42°C , and their sensitivity to λvir , but not homoimmune λcI^- , phage.

G Curing of $\lambda cI857$ Lysogens

At 42°C , the temperature-sensitive λ repressor gene ($cI857$) is inactivated, allowing the lysogenic phage to enter the lytic mode. However, at 32°C , spontaneous recombination may result in excision of the phage genome without derepression and concomitant cell lysis. In such cases, the phage lysis genes remain inert and the λ DNA is segregated, resulting in the eventual loss of the phage genome from the dividing cells. Such cured lysogens may be selected by virtue of their ability to grow at 42°C without being lysed.

An overnight culture of the lysogenic strain was streaked onto the surface of an L-agar plate prespread with the chelating agent tri-sodium citrate ($200\mu\text{l}$, 1M) and incubated overnight at 42°C . Surviving colonies were purified on L-agar plates at 37°C , and tested for their sensitivity to infecting λcI^- .

H Preparation of P1Cm1c1ts Lysogens

This method was essentially as described by Miller (1974). A fresh overnight of cells was diluted 50-fold in L-broth supplemented with CaCl_2 (5mM) and grown with aeration to mid-log phase. 0.2ml cells were mixed with 0.1ml P1Cm1 phage and left to preadsorb for 20 minutes at room temperature. Serial dilutions of the mixture were spread with a sterile glass spreader onto L-agar containing chloramphenicol ($15\mu\text{g/ml}$) and incubated for 24 hours at 30°C .

I Preparation of P1Cm1c1ts Lysates

The protocol was as described by Miller (1974). Cells lysogenic for P1Cm1c1ts were grown at 30°C with aeration in L-broth supplemented with MgSO₄ (10mM) and chloramphenicol (15μg/ml). When the cells reached a density of 2x10⁸/ml, the cultures were shifted to 40°C for 35 minutes to induce the phage, and then to 37°C with good aeration for an additional 60 minutes to allow cell lysis to take place. After this time, a few drops of chloroform were added and the lysate vortexed briefly. Cell debris was pelleted by centrifugation (3,000g, 5 min.), the phage supernatant decanted into a glass bottle and stored at 4°C.

J Preparation of P1kc Phage: Plate Lysates

Preparation of P1kc plate lysates was as described by Miller (1974). A fresh overnight was diluted 25-fold in L-broth supplemented with CaCl₂ (5mM) and incubated with aeration at 37°C to mid-log phase. The cells were harvested (3,000g, 5 min.) and resuspended in one-tenth of the original volume of L-broth containing CaCl₂. Aliquots (0.1ml) of P1kc diluted in phage buffer (neat, 10⁻¹, 10⁻²) were mixed with the cells (0.1ml) and preadsorbed for 20 minutes at 37°C. Top agar (2.5ml) supplemented with CaCl₂ (5mM) was then added and the mixture poured over fresh L-agar plates. Incubation was at 37°C without inversion for 8 hours or overnight. At the end of this time, L-broth (4ml) was added to the plate before storage at 4°C for a few hours or overnight. The L-broth was decanted, a few drops of chloroform added and the mixture vortexed. Cell debris was pelleted by centrifugation (3,000g, 5 min.). The supernatant containing the P1 lysate was transferred to a fresh bottle and stored at 4°C.

K Preparation of P1kc Phage: Liquid Lysates

A fresh overnight culture was diluted 20-fold in L-broth supplemented with CaCl₂ (5mM), and grown at 37°C with good aeration to mid-log phase. P1 phage were added to the culture at an m.o.i. of 0.1, and incubation at 37°C was continued

until lysis occurred. Chloroform (30 μ l) was added to the lysate and the flask was agitated for a further 15 minutes. If lysis was not evident after 4 hours, chloroform was nevertheless added. The lysate was clarified by centrifugation in a Sorvall (10,000g, 10 min.), or in a bench centrifuge (3,000g, 15 min.), and the supernatant was titred.

L Cotransduction Using P1 Lysates

A fresh overnight culture (5ml) of the strain to be transduced was pelleted by centrifugation (3,000g, 5 min.) and resuspended in MC buffer (5ml). The cells were incubated with aeration for 15 minutes. Serial dilutions of the P1 lysate were prepared using phage buffer, and 0.1ml of each dilution was added to 0.1ml of the cell suspension in small glass test tubes. The mixture was incubated at 37°C for 20 minutes to allow phage adsorption. Sodium citrate (0.2ml, 1M) was then added to prevent re-adsorption of the P1 phage. The contents of each tube were mixed with H₂O top agar (2.5ml) and poured onto selective agar. The plates were incubated at the required temperature for 36 hours. Transductants were purified and tested for cotransduction of the *hsd* genes by assaying for restriction and modification activity.

M.C. Buffer: MgSO₄ (10mM); CaCl₂ (5mM)

M Bacterial Conjugation

Cultures of the donor and recipient strains were grown overnight at 37°C in the presence of selective antibiotics. The donor cells were then diluted 20-fold in L-broth and grown to mid-log phase. Recipient cells were also diluted in L-broth, but incubated with aeration to stationary phase. Equal volumes (400 μ l) of donor and recipient were gently mixed in an Eppendorf tube. The mixture was then poured over the surface of a dry, prewarmed L-agar plate and incubated for 2 hours at 37°C. After this time, L-broth (1ml) was added to resuspend the cells, which were pipetted into an Eppendorf tube. The cells were washed three times with L-broth and streaked onto selective agar using a sterile loop. Exconjugants were screened for the presence

of plasmids using single colony gel analysis (section 2.6D).

2.6 DNA TECHNIQUES

A Ethanol Precipitation of DNA

DNA in solution was precipitated by the addition of a one-tenth volume of sodium acetate (3M, pH 5.0) and two volumes of absolute ethanol. Following incubation of the solution at -20°C for 1 hour or -70°C for 20 minutes, the precipitated DNA was pelleted by centrifugation (11,600g, 10 min.) and washed with ethanol (70% v/v). The DNA was dried under vacuum, and resuspended in the appropriate volume of TE buffer (pH 7.5).

B Rapid Large-Scale Preparation of Plasmid DNA

A fresh overnight of the plasmid-carrying strain was diluted 100-fold in L-broth containing the appropriate antibiotic and incubated with aeration overnight at 37°C . The cells were pelleted by centrifugation (3,000g, 10 min.) and resuspended in lysis buffer (3.5ml). Lysozyme (8mg, dissolved in 0.5ml lysis buffer) was added, and the mixture left on ice for 20-40 minutes. Freshly prepared alkaline SDS (8ml) was then added and the incubation on ice continued for a further 10 minutes. Sodium acetate (5ml, 3M, pH 5.0) was added and the suspension gently mixed by inversion before incubation on ice for 10 minutes. The resulting chromosomal DNA and protein precipitate was sedimented by centrifugation (13,000g, 15 min.) and the supernatant transferred to a fresh tube. Remaining protein was extracted using phenol:chloroform:isoamyl alcohol (25:24:1). The plasmid DNA in solution was precipitated with absolute ethanol, and resuspended in sterile H_2O (0.5ml). RNase ($5\mu\text{l}$, 10mg/ml) was added to the suspension, which was then incubated at 37°C for 20 minutes. Protein was removed by repeated extraction with phenol:chloroform:isoamyl alcohol (25:24:1) until the interphase between the organic and aqueous phases was clear. The plasmid DNA was then precipitated by the addition of sodium acetate and ethanol, and resuspended in TE buffer ($500\mu\text{l}$, pH

7.5).

Lysis Buffer: Tris-HCl (25mM, pH 8.0); EDTA (10mM, pH 8.0), glucose (1% w/v).

Alkaline SDS: NaOH (0.2M); SDS (1% w/v).

C Alkaline Lysis Plasmid "Miniprep"

Plasmid DNA for restriction analysis was isolated using a modified version of the method of Birnboim and Doly (1979). A 1ml volume of cells from an overnight culture were harvested by centrifugation (11,600g, 5 min.). The cell pellet was resuspended in lysis solution (100 μ l) and incubated at room temperature for 5 minutes. Alkaline SDS (200 μ l) was added, the solution mixed gently by inversion and placed on ice for 5 minutes. Chromosomal DNA and cell debris were precipitated by the addition of potassium acetate (150 μ l, 3M, pH 4.8) and a further 5 minute incubation on ice, followed by centrifugation (11,600g, 5 min.). The supernatant was transferred to a fresh tube and two volumes of absolute ethanol were added to precipitate the plasmid DNA. The precipitate was pelleted by centrifugation (11,600g, 10 min.), washed with ethanol (70% v/v), dried under vacuum, and resuspended in TE buffer (30 μ l, pH 7.5). Generally, a 5 μ l volume was used for restriction digestion.

D Single Colony Gel Analysis of Plasmid DNA

This method, essentially as described by Boyd and Sherratt (1986), allows rapid screening of plasmids within single clones. Cells cultured from a single colony were spread on the appropriate agar using a sterile loop to give a patch of 1cm². After incubation for 24 hours, the cells were toothpicked into an Eppendorf tube containing sample buffer (200 μ l) and incubated at room temperature to allow lysis to occur. The lysate was cleared by centrifugation (11,600g, 20 min., 4°C); the cell debris was removed with a toothpick. The supernatant was loaded directly onto an agarose gel (0.7% w/v) partially submerged in TBE electrophoresis buffer.

Electrophoresis was carried out in this way until the DNA samples had run into the gel. TBE buffer was then added to cover the gel, and electrophoresis was continued.

Sample Buffer: Ficoll (2.5% w/v); SDS (1% w/v); bromophenol blue (5% w/v); RNase (40 μ g/ml), in 1xTBE.

E Bacterial DNA "Miniprep"

This method is based upon that of Redfield and Campbell (1987). A 1ml volume of a fresh overnight culture was harvested by centrifugation (11,600g, 5 min.) and resuspended in a 0.5ml volume of Tris (40mM, pH 8), EDTA (20mM, pH 8), and lysozyme (10mg/ml). Following incubation at 37°C for 30 minutes, the cells were lysed by the addition of SDS (0.1ml, 5% w/v) and RNase (100 μ g/ml) was added at this point. The cell debris was repeatedly extracted with an equal volume of phenol (pre-equilibrated with TE). The phenol and aqueous layers were mixed by gentle inversion, usually on a rotating wheel, and then separated by centrifugation (11,600g, 7 min.). This was repeated until the interface between the phenol and aqueous phases appeared clear. A final extraction with chloroform removed any residual phenol. To precipitate the DNA, NaCl (15 μ l, 5M) and ethanol (1ml) were added to the supernatant, which was stored at -20°C overnight. The DNA was then pelleted by centrifugation (11,600g, 10 min.), washed, dried, and resuspended in TE buffer (100-200 μ l). A typical yield was 20 μ g of DNA.

F Restriction Endonuclease Digestion of DNA

DNA was digested with restriction enzymes in a volume of up to 150 μ l containing 0.5-10 μ g DNA, under conditions recommended by the suppliers. Reactions were incubated at the appropriate temperature for 2-3 hours. When heat-labile enzymes were used, the reaction was stopped by heating at 70°C for 5 minutes. Digests were stored at -20°C.

G Ligation of DNA

DNA was ligated using T4 DNA ligase, in a volume of 10-20 μ l containing Tris-HCl (20mM, pH7.5), MgCl₂ (10mM), DTT (10mM), ATP (0.6mM), DNA (250-500ng), and T4 DNA ligase (0.1-0.2 Weiss units). The ligation reaction was incubated at 16°C overnight, after which time the ligase was inactivated by heating at 70°C for 5 minutes. More recently, the ligation buffer supplied with the ligase was used in the reaction, namely Tris-HCl (66mM, pH 7.5), MgCl₂ (5mM), DTE (1mM), and ATP (1mM).

H Agarose Gel Electrophoresis

DNA fragments were separated by electrophoresis through horizontal agarose gels submerged in 1xTBE buffer. The concentration of agarose in the gel (0.6-1.2% w/v) varied according to the desired range of separation of the DNA molecules.

DNA samples (0.1-2 μ g) were mixed with loading buffer (5-10 μ l) before loading into the wells of the submerged gel. Electrophoresis was carried out at 5 volts cm². The DNA was stained by soaking the gel in ethidium bromide (1mM) for 20 minutes, then destained in H₂O for 10 minutes. The stained DNA was visualized by illumination with long-wave UV light.

Loading buffer: Ficoll 400 (40% w/v); EDTA (100mM, pH7.5); bromophenol blue (0.05%) in 1xTBE.

I Purification of DNA Fragments from Low Melting Point Agarose

The DNA fragment was separated by electrophoresis through low melting point agarose and removed in as small a slice of gel as possible. An equal volume of 1xTBE and NaCl (0.2M) was added. The agarose was melted by heating at 65°C in a water bath. Two extractions with redistilled phenol equilibrated with 1xTBE, NaCl (0.1M) were carried out. The aqueous layer was extracted repeatedly with an

equal volume of butanol until the final volume was approximately 400 μ l. The DNA was precipitated with sodium acetate and absolute ethanol. The pellet was resuspended in sodium acetate (100 μ l, 0.3M) and precipitated a second time by the addition of ethanol (250 μ l). The pellet was washed twice with ethanol (70% v/v), dried under vacuum and resuspended in distilled, sterile H₂O. The DNA solution was stored at -20°C.

J Recovery of DNA from Low Melting Point Agarose Gels Using GELase™

GELase™ is an enzyme preparation that digests the carbohydrate backbone of molten agarose into small, ethanol-soluble oligosaccharides. Thus, DNA may be precipitated directly from the digested agarose and used for cloning or labelling. The "high activity" protocol accompanying the enzyme was used in purifying DNA fragments. Basically, the DNA band was cut from a TAE-buffered low melting point agarose gel and soaked for one hour in GELase™ buffer. The buffer was then removed and the agarose slice melted at 65°C. Following equilibration of the temperature of the molten agarose to 40°C, the appropriate volume of GELase™ was added, and the mixture incubated for at least one hour. After this time, the DNA was precipitated from the solution with sodium acetate and ethanol, resuspended in H₂O and stored at -20°C.

K Transformation of Competent Cells

Cells were made competent for the uptake of DNA using a modified version of the method described by Mandel and Higa (1970). A fresh overnight culture was diluted 50-fold in L-broth and grown at 37°C to mid-log phase. The cells were harvested by centrifugation (2,000g, 10 min., 4°C), then resuspended in an equal volume of cold MgCl₂ (100mM). The cells were immediately centrifuged again, and resuspended in half the original volume of cold MgCl₂ (100mM). The cells were then pelleted once more and resuspended in cold CaCl₂ (1ml, 100mM). DNA was added to the competent cells (200 μ l) in a 5 ml tube and left on ice for 30 minutes. The mixture was then heat shocked by incubation at 42°C for 90 seconds. Transformation

of cells with plasmid DNA required the addition of L-broth (1ml) and incubation at 37°C for one hour to allow expression of the plasmid-encoded antibiotic resistance before spreading serial dilutions onto agar containing the appropriate antibiotic and incubating overnight. A typical transformation efficiency was 3×10^5 transformants/ μ g DNA.

L Electrotransformation

a) Preparation of Cells

Cells were prepared as described by Heery and Dunican (1989). Cells were cultured overnight in 2xTY broth (10ml). The culture was chilled on ice, then pelleted by centrifugation (2,000g, 10 min., 4°C). The pellet was washed three times in cold, distilled, sterile H₂O, then resuspended in an equal volume of H₂O and kept on ice. Cells for future use were resuspended in glycerol (10%), fast frozen in aliquots in liquid nitrogen or on dry ice, and stored at -70°C.

b) Electrotransformation

Electroporation was carried out according to the protocol supplied with the apparatus. After thawing on ice, 40 μ l cells were dispensed into cold polypropylene tubes. DNA, generally 10ng in low ionic strength buffer, was added to the cells. The mixture was then transferred to a cold, 0.2cm electroporation cuvette, and shaken to the bottom. The Gene Pulser Apparatus (Biorad) was set at 25 μ F and 2.5Kv, the Pulse Controller at 200 Ω . The cuvette was placed in the chamber slide and passed between the contacts in the base of the chamber. The cells were pulsed once, producing a pulse of 4.5-5msec and a field strength of 12.5kV/cm. The cuvette was then removed from the chamber and the cells immediately suspended in SOC buffer (1ml). The suspension was transferred to a Bijou bottle and incubated with aeration at 37°C for 30 minutes. The cells were diluted and spread over selective agar with a sterile glass spreader and incubated overnight at 37°C. A typical transformation efficiency was 2×10^7 transformants/ μ g DNA.

SOC Buffer: Bacto tryptone (2%), Bacto yeast extract (0.5%), NaCl (10mM), MgCl₂ (10mM), MgSO₄ (10mM), glucose (20mM).

M *In vitro* Packaging of λ DNA

In vitro packaging mixes, namely Freeze Thaw Lysate (FTL) and Sonicated Extract (SE) were kindly donated by Annette Campbell. The packaging reaction was carried out by mixing in this order:

7 μ l Buffer A
0.2-0.5 μ g DNA (up to 10 μ l)
1 μ l Buffer M1
6 μ l SE
8 μ l FTL (added immediately after SE)
1 μ l terminase (Chow et al., 1987)

The reaction was incubated for 2.5 hours at 23°C. Phage buffer (0.5ml) was then added and the phage titrated on the appropriate indicator strain. When a higher packaging efficiency was required, the Amersham *In Vitro* Packaging Kit was used.

Buffer A: Tris-HCl (20 μ l, 1M, pH 8.0); MgCl₂ (3 μ l, 1M);
 β -mercaptoethanol (0.5 μ l); EDTA (2 μ l, 0.5M, pH 7.5);
H₂O (975 μ l).

Buffer M1: Tris-HCl (6 μ l, 0.5M, pH 7.5); spermidine/putrescine (300 μ l, 50mM/0.1M, neutralised with Tris base); MgCl₂ (9 μ l, 1M); ATP (75 μ l, 0.1M); β -mercaptoethanol (1 μ l); H₂O (110 μ l).

N Transfer of DNA From Plaques to Nylon

The procedure used to transfer phage DNA from plaques to nylon filters was modified from that described by Benton and Davies (1977). Cells, preadsorbed with phage (3-5x10³ p.f.u./ml) were plated in BBL top agar containing MgSO₄ (10mM) on dry BBL agar plates. After overnight incubation at 37°C, the plates were placed at 4°C for 2 hours to harden the agar. After this time, dry nylon filter discs were

then placed on the surface of the plates for 2 minutes to allow transfer of phage and DNA. When duplicate filters were desired, the second transfer time was extended to 5 minutes. The filters were carefully peeled from the plates and placed, DNA side up, onto a pad of blotting paper soaked with NaOH (0.5M) for 5 minutes. The filters were next immersed in denaturation buffer for 20 seconds, and then dipped twice in neutralization buffer and once in 2xSSC for 20 seconds each. The filters were briefly blotted, and DNA was cross-linked by UV irradiation in a Stratagene Stratalinker (12,000joules/cm²).

Denaturation buffer: NaOH (0.1M); NaCl (1.5M).

Neutralization buffer: Tris-HCl (0.5M, pH 7.5); NaCl (1.5M).

O Transfer of DNA from Agarose Gels to Nylon Membranes

An adaptation of the method of Smith and Summers (1980) was used, which itself was a modification of the original procedure described by Southern (1975). The agarose gel containing the DNA fragment was placed in HCl (0.25M, 15 min.) to partially hydrolyse the DNA by depurination. After rinsing in distilled H₂O (10 min.), the gel was immersed in denaturing solution (1.5M NaCl/0.5M NaOH, 30 min.). The gel was rinsed again in distilled H₂O (10 min.), and then neutralized in ammonium acetate (1M, pH 8.0, 1 hour). For transfer, the gel was placed on a glass plate. A nylon membrane, cut to the same dimensions as the gel, was briefly soaked in ammonium acetate and then placed on top of the gel. Three pieces of blotting paper, again cut to the same size as the gel, were immersed ammonium acetate and placed on top of the nylon membrane. The blotting paper sheets were in turn covered by a 6mm stack of dry paper towels. A glass plate, weighed down by a beaker of H₂O, ensured even transfer. Transfer was effected for 5 hours or overnight. After this time, the DNA was cross-linked to the filter using a Stratagene Stratalinker (12,000joules/cm²). The gel was restained to verify that the DNA had transferred.

P Labelling of DNA Probes and Hybridization

a) DNA hybridization using the Amersham ECL Gene Detection System

The Amersham ECL Gene Detection System was used to label probes nonradioactively for use in Southern hybridization. This system is based on enhanced chemiluminescence, the emission of light as a byproduct of a chemical reaction. Positively-charged horseradish peroxidase (hrp), when added to a solution of denatured linearised DNA, will form weak ionic bonds which may then be crosslinked by the addition of glutaraldehyde. Hybridization of the labelled probe to membrane-bound DNA is carried out overnight. To detect the bound probe, the filter is covered with a solution containing a substrate of hrp and luminol, a chemiluminescent substrate. Reduction of the hrp substrate is coupled to the oxidation of luminol, which emits light as it decays to the ground state. This light may be captured on blue-sensitive x-ray film as a permanent record.

DNA labelling, hybridization, and detection were carried out as described in the protocol accompanying the kit.

b) Nonradioactive DNA labelling and detection using Digoxigenin

The Boehringer nonradioactive DNA labelling and detection kit was used to detect specific DNA sequences in Southern hybridizations as well as in plaque lifts.

The kit is based upon the incorporation of dUTP coupled to digoxigenin during random-primed DNA synthesis by Klenow enzyme, using a single-stranded DNA template. The labelled DNA, generally a purified fragment or a linearized plasmid, is hybridized overnight to DNA immobilized on nylon filters. Target sequences are detected by enzyme immunoassay; the filter is incubated with anti-digoxigenin antibody conjugated to alkaline phosphatase. The resulting antibody-hapten complexes are visualized by the addition of a chemiluminescent substrate for alkaline phosphatase, AMPPD (3-(2'-Spiroadamantane)-4-methoxy-5-(3"-

phosphoryloxy)-phenyl-1,2-dioxetane). As the dephosphorylated substrate decomposes via an unstable intermediate, light is emitted which may be captured on x-ray film.

Labeling, hybridization, and detection were carried out as described in the protocol accompanying the kit. To allow reprobing, bound probe was removed from the filter by washing twice in alkaline SDS (0.2M NaOH, 0.1% SDS) at 37°C for 15 minutes, or latterly in a solution of formamide (50%), Tris (50mM, pH 8.0), and SDS (1%) at 68°C.

When using the Amersham ECL Gene Detection System, prehybridization and hybridization were carried out at 37°C in the supplied buffer, with NaCl added to a final concentration of 0.5M, as suggested in the protocol. Washes were done in the recommended buffers, namely 6M urea, 0.5xSSC, and 0.4% SDS (twice for 20min. at 37°C), followed by two washes in 2xSSC (5min. each at room temperature).

The conditions chosen for hybridization and washes when using the Boehringer DIG kit were for relatively low stringency, essentially as described by Daniel *et al.* (1989). Hybridization was carried out at 37°C in a buffer consisting of 5xSSC (0.75M NaCl) and 50% formamide, but with 0.1% lauryl sarcosine, 0.02% SDS, and 2% "block" (provided by the manufacturer). Washes were done at room temperature in 2xSSC and 0.1% SDS (twice for 5 min.) and subsequently at room temperature in 1xSSC and 0.1% SDS (twice for 15min.).

CHAPTER THREE

A BIOLOGICAL SCREEN FOR TYPE I

R-M SYSTEMS

IN WILD TYPE *E. COLI* ISOLATES

3.1 INTRODUCTION

It is generally accepted that restriction enzymes function to prevent the acquisition and expression of foreign DNA (Arber, 1965a). A consequence of this could be protection against phage infection, which would be advantageous for the initial establishment of bacterial populations (Levin, 1986, 1988a). In addition, the flow of genetic information between populations of bacteria could be reduced. This hypothesis is supported by the finding that the division of a number of *Salmonella* strains into groups based on their host specificity systems correlates well with their taxonomical classification into serotypes (Bullas *et al.*, 1980). Alternatively, restriction and modification systems could stimulate the exchange of genetic information by generating DNA fragments and thus stimulating homologous recombination (S.Lederberg, cited in Radding, 1973; Roberts, 1978; Endlich and Linn, 1985b; Price and Bickle, 1986). Type I enzymes, which cleave DNA in an apparently random fashion, would seem particularly suited for such a role. However, there is little evidence in support of these alternative functions.

If type I restriction and modification systems do confer an advantage to the host, they might be expected to be widespread in nature. A knowledge of the prevalence of type I r-m systems is basic to any consideration of their relevance. We know already that they are not essential to cell survival; *E.coli* C lacks any r-m system, and laboratory strains with *hsd* deletions remain viable (Murray *et al.*, 1982). However, no extensive survey of the distribution and diversity of type I systems has yet been carried out.

As a result of the intensive search for restriction enzymes with novel specificities for use in recombinant DNA technology, restriction and modification systems have been detected in almost every taxonomic group of bacteria. Most of these enzymes are, of course, type II, which cleave the DNA within a defined sequence and are therefore particularly useful as tools for molecular biologists. The diversity of the sequences recognized by type II r-m systems is immense; approximately 180 different DNA sequence specificities have been identified so far

(Roberts and Macelis, 1992). Many more type II systems have been found to be isoschizomers, demonstrating the wide distribution of these enzymes among bacteria.

Detection of type II restriction enzymes usually involves assaying for endonuclease activity in cell extracts (Wilson, 1988), for example by electrophoresis of DNA digests through agarose gels (Sharp *et al.*, 1973). The near-random nature of endonucleolytic cleavage by type I restriction enzymes has precluded the use of this method to detect these systems. Rather, detection of type I r-m systems has relied on restriction of bacteriophages *in vivo* (Bertani and Weigle, 1953; Colson and Colson, 1971; Bullas and Colson, 1975; Bullas *et al.*, 1980), or detection of related systems by Southern hybridization using probes from *hsd* genes of known families (Murray *et al.*, 1982; Fuller-Pace *et al.*, 1985; Daniel *et al.* 1988). Therefore, the eleven known type I systems may not be representative of the diversity of these systems in nature.

A screen to determine the distribution and diversity of type I r-m systems could take two forms. Firstly, genes with similarity to known systems could be detected by hybridization to DNA probes constructed from *hsd* genes of known type I r-m systems. Such a search has been carried out for DNA sequences with similarity to the *hsd* genes of *EcoK* and *EcoA* (Daniel *et al.*, 1988). DNA from 20 laboratory and wild type *E.coli* strains (mostly hospital isolates), as well as samples of DNA from other related enteric bacteria were screened for similarity to either a K-specific or an A-specific probe. Although *S.enterica* serovar typhimurium encodes three r-m systems, its DNA hybridized only to the K-probe, due to the presence of the related *hsd_{SB}* genes. The DNA of *C.freundii* showed homology to the A-probe only, identifying an A-related system now known as *CfrA*. No hybridization was seen either with *Shigella boydii* or with *Klebsiella pneumoniae*, which is believed to possess an r-m system (Streicher *et al.*, 1974). Surprisingly, only six of the twenty *E.coli* isolates hybridized to either probe, four to the *hsd_K* probe and two to the *hsd_A* probe. Thus these results indicate that while related specificity systems were detected in more than one genus, the majority of the *E.coli* strains were judged to lack either K- or A- related *hsd* genes.

These results highlight the limitations of screening with DNA probes. Only *hsd* genes with DNA sequences similar to the probes will be detected and other systems may be overlooked. In order to determine whether most bacterial strains have alternative type I systems or lack them altogether, we would need probes to identify all families.

Alternatively, a biological screen for type I r-m systems may be developed. This has been carried out in *Salmonella* strains by looking for evidence of restriction of a number of bacteriophages (Bullas and Colson, 1975; Bullas *et al.*, 1980; Ryu *et al.*, 1988). Bullas and Colson (1975) screened 43 *Salmonella* strains for evidence of restriction of P3, a temperate bacteriophage with a wide host range including many *Salmonella* and *E.coli* strains (Nutter *et al.*, 1970). Eleven of the *Salmonella* strains were sensitive to P3 infection. The chromosomally-encoded LT restriction and modification system (Colson *et al.*, 1970) was detected in almost all these strains; in addition, five novel systems with unique recognition sequences in five different serotypes of *Salmonella* were identified. One of these systems, *StySP* in *S.enterica* serovar potsdam, was demonstrated to be allelic to the *S.enterica* serovar typhimurium *StySB* system using P1 cotransduction with the nearby *serB*. The *StySP* and *StySB* systems were later shown to be related to the K system of *E.coli* both by complementation analysis (Van Pel and Colson, 1974), and by immunological cross-reactivity and DNA hybridization (Murray *et al.*, 1982). The other four systems identified in the survey were not analysed further. A more extensive investigation of the occurrence of restriction and modification systems in different *Salmonella* serotypes was later carried out (Bullas *et al.*, 1980). *Salmonella* strains representing 85 serotypes were screened for restriction activity. Four bacteriophages, P3, P9a, P22 and L, were used in the survey in the hope of avoiding problems associated with resistance of the bacteria to infection or refractiveness of the phages to restriction. Twenty-eight strains were found to be sensitive to at least one of the tester phages. Twelve of these bacterial strains were shown to have restriction systems with new specificities, by assaying with phages modified against the known *Salmonella* restriction enzymes, i.e. LT, SA and SB. In P1 cotransduction experiments, eight of the novel systems were demonstrated to be closely linked to *serB*, and in fact four

were allelic to *SlySB* suggesting that they were type I. In addition, two of the systems (SM and ST) were shown to behave as alleles of SB in complementation tests.

The results of Bullas and his coworkers (Ryu *et al.*, 1988) suggest that there are alternative, unrelated type I r-m systems which could form another enzyme family or families. Although apparently allelic, they share little similarity with members of the Ia and Ib families. However, some limitations of using phages as indicators of type I restriction systems were evident. Often, wild type bacteria are resistant to phage infection. Conversely, many phages are resistant to certain r-m systems, whether due to phage-encoded antirestriction mechanisms (Krüger and Bickle, 1983), or to a lack of recognition sequences in their genomes (Krüger and Bickle, 1983; Schroeder *et al.*, 1986; Sharp, 1986). For these reasons, a biological screen has not been possible for natural populations of *E. coli*. However, a study has suggested that as many as 70% of wild-type *E. coli* strains are sensitive to infection by P1 (Robeson *et al.*, 1980), although few (approximately 2.1%) were sensitive to λ , the phage commonly used to detect restriction.

Bacteriophage P1 is only weakly restricted by type I enzymes. This is due to an antirestriction mechanism mediated by the products of two nonessential phage genes, *darA* and *darB* (Iida *et al.*, 1987a). Iida and his colleagues (1987a) demonstrated that the Dar function (Dar for *defense against restriction*) specifically protects DNA packaged in P1 phage heads from restriction by type I systems. Little protective effect against type II systems such as *EcoRI* or *EcoRII* was noted, nor against the type III systems *EcoP1* and *EcoP15*.

Phages with mutations in either of the two *dar* genes have been isolated (Iida *et al.*, 1987a) enabling analysis of the antirestriction functions (for a more detailed discussion see Chapter 1, section 1.6 F). Interestingly, the products of the two *dar* genes appear to provide protection against restriction by separate subsets of the type I systems. Although wild type P1 phage is insensitive to restriction by a number of members of both the Ia and Ib families of type I restriction enzymes, P1 with a

mutation in the *darB* gene demonstrates a reduced efficiency of plating when infecting strains encoding members of the Ia family including *EcoK*, *EcoB*, *StySB*, *StySP* and *StySQ*, while still remaining refractive to restriction by the Ib family enzyme, *EcoA*. This suggests that the *darB* gene product provides protection specifically against restriction by members of the type Ia family enzymes. P1 phages which are *darA*⁻ are sensitive to restriction by members of both the Ia and Ib families of type I enzymes. SDS polyacrylamide gel electrophoresis indicates that while the *darA* gene product is present in phage particles which are *darB*⁻, neither *darA* nor *darB* gene products are found in *darA*⁻ phages. Although it appears that the *darB* gene product protects against restriction by Ia systems and the *darA* gene product against Ib systems, this cannot be stated conclusively as no P1 mutants have been isolated that include only the *darB* gene product within the phage head. The DarB protein may require DarA in order to protect against restriction enzymes of the Ia family, thus conferring a DarB⁻ phenotype on a *darA*⁻ phage.

P1 and its *dar* mutant derivatives therefore would seem ideal candidates for a screen for type I restriction systems in natural populations of *E.coli*. Comparison of the efficiencies of plating of wild type P1 and a P1 *darA* mutant on an *E.coli* strain may elucidate the presence of a type I r-m system if the *darA* mutant is restricted while the wild type P1 is not.

3.2 DEVELOPMENT OF THE P1 SCREEN

The phages chosen for use in the screen were the Dar⁺ phage P1Cm1 (Arber *et al.*, 1978) and a *darA* mutant derivative, P1CmTc1 (Iida *et al.*, 1982). P1Cm1 carries the chloramphenicol resistance gene from the R factor NR1 probably as a result of an illegitimate recombination event (Mise and Arber, 1976). The *darA*⁻ phage P1CmTc1, was produced through a recombination event between P1Cm1 and P1Tc1, carrying tetracycline resistance (Mise and Arber, 1976). P1CmTc1 carries deletions and substitutions between the resident *IsI* element of P1 and the invertible C segment, resulting in the loss of approximately 10kb of the phage genome, and a corresponding *darA*⁻ phenotype (Iida *et al.*, 1982; Iida *et al.*, 1987a,b).

The chloramphenicol resistance gene carried by both tester phages provides a useful marker for screening; infecting phages which have escaped restriction may be detected as chloramphenicol-resistant lysogens when spread on antibiotic containing agar. Colonies are easier to count than the tiny P1 plaques, and the even more minute plaques seen with P1CmTc1 (Mise and Arber, 1976). Both tester phages also carry the *c1ts225* allele (Iida and Arber, 1977), allowing the formation of lysogens at 30°C, and production of lysates upon temperature induction (Chapter 2, sections 2.5 H; I).

Unmodified lysates of the P1Cm1 and P1CmTc1 phages were prepared on an *r⁻m⁻* strain, NM1. To assay for restriction, the unmodified P1Cm1 and P1CmTc1 phages were added to exponentially-growing test bacteria and a non-restricting control. After a twenty-minute incubation at room temperature to allow phage adsorption, the cells were serially diluted and spread onto L-agar supplemented with 15 µg/ml chloramphenicol. The plates were incubated for 18 hours at 30°C. A comparison of the counts of chloramphenicol-resistant lysogens for the test strain and the control determined the efficiency of plating (or more correctly, the efficiency of colony formation) of the two phages (figure 3.1); fewer colonies on the test strain were indicative of restriction. A lowered e.o.p. for the *darA* mutant phage than the wild type P1 suggested the presence of a type I *r-m* system. To confirm that a low e.o.p. was due to restriction, lysogens were picked from the test plates and induced to give lysates. The resulting phage should be modified against the restriction system and thus able to form colonies efficiently when plated on the test strain.

3.3 RESULTS

A Application of the P1 Screen to Laboratory Strains of Known Specificities

Initially, the screen was tested on a sample of laboratory strains encoding type I systems of known specificities, in order to repeat and extend the results reported by Iida and his colleagues (1987a). Table 3.1 shows the efficiency of plating of the two P1 tester phages on members of the Ia, Ib and Ic families of type I systems. P1Cm1

Figure 3.1: The Screen for Restriction of P1 Phage.

To facilitate screening, derivatives of wild type P1 and a *darA* mutant (P1Cm1 and P1CmTc1) were used and phage escaping restriction were detected as chloramphenicol-resistant lysogens. Unmodified P1Cm1 and P1CmTc1 were added to exponentially-growing test bacteria and a non-restricting control. After phage adsorption, the cells were serially diluted and spread onto L-agar supplemented with 15 μ g/ml chloramphenicol. A comparison of the counts of chloramphenicol-resistant lysogens for the test strain and the control determined the efficiency of plating of the phage; fewer colonies on the test strain were indicative of restriction. A lower efficiency of plating (e.o.p.)* for the *darA* mutant than for the wild type phage suggests the presence of a type I restriction and modification system.

A typical result obtained when testing an r^+m^+ strain with the two phages is shown here. The wild type P1Cm1 phage forms lysogens with almost equal efficiency on the test strain and the non-restricting control. The *darA* mutant phage, however, shows a reduced e.o.p. on the restricting strain as shown by the formation of fewer lysogens. Comparing the e.o.p.'s of the two phages indicates that the test strain encodes a type I restriction and modification system.

Phage induced from the P1 *darA* lysogens are tested for their efficiency of plating on the test strain; the e.o.p. of phage modified by the type I system should return to 1.

* e.o.p. in the context of the screen refers to the efficiency of chloramphenicol-resistant colony formation.

Figure 3.1

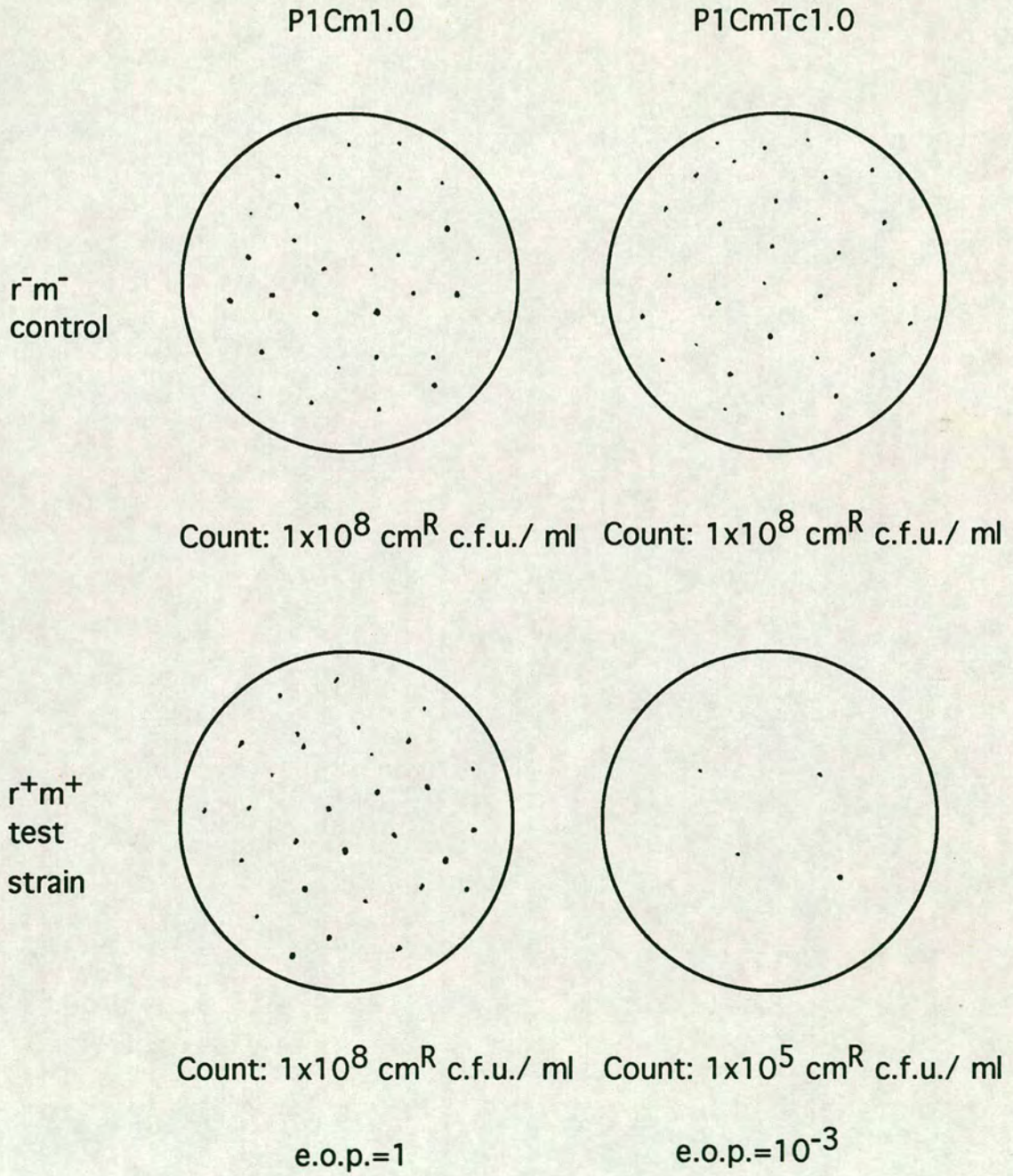


Table 3.1: Restriction and Modification of the P1 Tester Phages by Laboratory Strains Representing Three Families of Type I R-M Systems.

Strain	Restriction System	Efficiency of Plating		Ratio P1Cm1/P1CmTc1	Efficiency of Plating	
		P1Cm1.0 ^a	P1CmTc1.0		P1Cm1.X ^d	P1CmTc1.X
C600	<i>EcoK</i>	9x10 ^{-2b}	8x10 ⁻⁴	115.4	1.4	0.4
		5x10 ^{-2c}	7x10 ⁻⁴	77.8		
L4002	<i>StySP</i>	0.8 ^b	2x10 ⁻³	453	0.6	1.1
		9x10 ^{-2c}	2x10 ⁻³	51.8		
		4x10 ^{-2b}	2x10 ⁻³	22.2		
WA2899	<i>EcoA</i>	1 ^c	8x10 ⁻³	125	2.8	2.4
		0.3 ^c	2x10 ⁻³	168.4		
NM144	<i>EcoR124</i>	0.1 ^b	1x10 ⁻³	1000	0.3	0.1
		0.3 ^b	1x10 ⁻³	300		

- ^a Unmodified phage (P1.0) were prepared by induction of lysogens arising from infection of an r^m host, NM1.
- ^b Phage used were P1Cm1.0 from N.E.Murray.
- ^c Phage used were P1Cm1.0 from W.Arber.
- ^d Modified phage were prepared by induction of lysogens arising on the test strain.

phage from the two sources were tested on these strains; although giving slightly variable results, the efficiency of plating was only slightly reduced for both phages, mirroring the findings of Iida and colleagues (1987a). However, the *darA*⁻ P1CmTc1 phage repeatedly demonstrated a reduced efficiency of plating on Ia, Ib, and Ic family members when compared to the wild type P1. This lowered plating efficiency was shown to be due to restriction in all cases. Induction of the lysogens resulted lysates of phage fully modified against restriction and able to grow on the test strain with a plating efficiency of 1.

B Extension of the P1 Screen to Natural Isolates of *E. coli*

The screen for type I restriction systems was then extended to a selection of ten natural isolates of *E. coli*. Many of these strains were amongst those judged to lack homology to *hsd_K* and *hsd_A* in the hybridization study described earlier (Daniel *et al.*, 1988). One strain, 629, was already known to encode a type I r-m system with B-specificity, and therefore provided an internal control.

Results obtained from this experiment were not as clear-cut as they had been with the laboratory *E. coli* strains. Although Robeson *et al.* (1980) suggested that 70% of wild type *E. coli* strains are sensitive to P1 infection as judged by the formation of chloramphenicol-resistant P1cm1 lysogens, the situation appears to be far more complex. Eight out of the ten strains tested did allow the formation of lysogens, but few could actually be considered truly P1-sensitive.

A number of stages in the P1 infection seemed to be affected; these are outlined below, and in Table 3.2.

a) Formation of Lysogens

Two strains (BLD4 and BRL-ET2) did not support the growth of P1 at all, as evidenced by a complete lack of lysogens. These were taken to be P1 resistant, though restriction could not be ruled out. Strain E161 was included in this

Table 3.2: Results of Extending the P1 Screen to Natural Isolates of *E.coli*.

Strain	Efficiency of Transfer of Chloramphenicol Resistance				Titre of Lysate Following Induction ^b (cfu ml ⁻¹)		Modification of Tester Phages ^c	
	<u>P1Cm1.0</u>		<u>P1CmTc1.0</u>		<u>P1Cm1.X</u>	<u>P1CmTc1.X</u>	<u>P1Cm1.X</u>	<u>P1CmTc1.X</u>
	formation of lysogens ^a	eop	formation of lysogens	eop				
629	+	3x10 ⁻³	+/-	10 ⁻⁴ 3x10 ⁻⁴	<10 ^{5 d} 3x10 ⁻⁶	0	nd ^e	nd
653	+	2x10 ⁻⁴ 2x10 ⁻⁵	+/-	10 ⁻⁶	<10 ⁵	<10 ⁵	nd	nd
VB1	++	2x10 ⁻² 2x10 ⁻³	+	2x10 ⁻⁴ 7x10 ⁻⁶	1x10 ⁷	2x10 ⁵	no	yes
BLD4	-	<10 ⁻⁶	-	<10 ⁻⁶	nd	nd	nd	nd
BRL-ET2	-	<10 ⁻⁶	-	<10 ⁻⁶	nd	nd	nd	nd

BRL-ET13	+	2×10^{-4}	+/-	1×10^{-6}	7×10^8	4×10^7	yes	yes
RM01A	+	1×10^{-3} 2×10^{-4}	+	4×10^{-3} 1×10^{-4}	$< 10^5$	$< 10^5$	nd	nd
RM66A	+/-	3×10^{-5} 5×10^{-6}	+/-	7×10^{-6} $< 10^{-8}$	1×10^7	$< 10^5$	partial	nd
RM74A	++	0.5 1×10^{-2}	++	3×10^{-3} 3×10^{-4}	5×10^6	1×10^6	yes	yes
E161	+/-	3×10^{-6}	+/-	3×10^{-6}	0	0	nd	nd

^a The number of lysogens on chloramphenicol-containing plates was scored from - (no lysogens) to ++ (many lysogens even at high dilutions).

^b Lysogens forming on test *E. coli* strains were heat induced and the lysates titred on NM555 (r^m).

^c Phage were tested for modification by infecting test strain and looking for improved efficiency of plating.

^d A titre of less than 10^5 cfu ml⁻¹ was generally too low to give significant results in modification tests.

^e nd= not determinable.

group; it gave very few colonies and none yielded phage.

The wild type P1Cm1 was able to form lysogens on the other strains, though often at a vastly reduced efficiency of plating, for example, 629, 653 and RM66A. Strain RM74A, and perhaps also VB1, were exceptions to this rule; P1Cm1 had a plating efficiency of 10^{-1} to 10^{-2} for RM74A, and 10^{-2} to 10^{-3} for VB1.

The *Pldar* mutant had difficulty in lysogenizing any of the strains tested, with the exception of RM74A and VB1. On the other strains few, or no lysogens appeared. This might imply the presence of a type I restriction system, perhaps in conjunction with another r-m system of type II or type III. If this were the case, the wild type P1 would suffer restriction by the type II or III system, while in addition the *Pldar* phage would be cut back further by the type I system, to the extent that restriction was complete and no lysogens arose. Low plating efficiency due to restriction should disappear once the phage are modified.

b) Induction of Lysogens

Phage lysates were prepared by heat induction of the lysogens (see Chapter 2, section 2.5 I). Titres were calculated from the number of lysogens produced by infection of a non-restricting laboratory strain. Here again, problems arose as a number of strains yielded poor titre lysates, particularly of the P1CmTc1 phage. Indeed, in one case (RM01A), the titre of the *Pldar* lysate was only 1×10^2 colony forming units (c.f.u./ml). These lysates were considered to be of little use for detecting modification. Low titres correlated with a degree of resistance of the wild type strain to phage infection. Therefore, infection of the strain, even by modified phage, generally yielded fewer lysogens than did infection of the control. Titres of 10^5 c.f.u./ml or greater on the non-restricting control were considered to be sufficiently high to give reliable results. These were obtained from both P1Cm1 and P1CmTc1 lysogens of VB1, BRL-ET13 and RM74A, and from P1Cm1 only for RM66A. The phage in these lysates were tested for modification by infecting the *E. coli* strains and looking for improved plating efficiency. Both P1 phages plated on

RM74A showed modification, with e.o.p.'s approaching 1. A similar effect was seen with BRL-ET13 on one occasion. The e.o.p. of P1Cm1.RM66A was slightly improved but induction of P1CmTc1.RM66A lysogens yielded lysates such low titre as to make interpretation of modification results impossible. Although P1Cm1.VB1 was apparently poorly modified, P1CmTc1.VB1 exhibited modification.

Thus, an improved efficiency of plating indicating modification of the phage surviving restriction by an r-m system was demonstrable in only two strains, BRL-ET13 and RM74A. One could postulate that these strains might encode type I r-m systems as judged by the difference in the e.o.p.'s of the two tester phage, though this result was not repeatable in BRL-ET13. It was not possible, therefore, to conclude which of the ten *E.coli* strains encoded a type I r-m system and which did not. The fact that 629, which is known to encode a system with B-specificity, gave no indication of type I restriction when tested with the P1 screen suggests that a number of type I r-m systems would be overlooked due to the apparent resistance of the wild type *E.coli* strains. One can conclude that as a screen for the distribution of type I r-m systems detection of restriction of P1Cm1 and P1CmTc1 is of little value. However, in some cases, such as RM74A and BRL-ET13, a suggestion of type I restriction by the tester phage might be worth pursuing with further experiments.

C P1 Cotransduction of Putative *hsd* Genes from Wild Type *E.coli* Strains

Bullas and his colleagues were able to transfer the *hsd* genes of a number of *Salmonella enterica* serotypes by cotransduction with the linked marker *serB* to make *E.coli/Salmonella* hybrids (Colson and Colson, 1971; Bullas and Colson, 1975; Bullas *et al.*, 1980). In this way, it was possible to analyse further the *hsd* genes in a laboratory bacterial strain that was amenable to genetic manipulation. Similar cotransduction experiments were undertaken with the wild type *E.coli* strains BRL-ET13 and RM74A, both putative carriers of type I r-m systems. The *dnaC* gene was chosen as a marker for transduction; this gene is slightly closer to the *hsd* region on the *E.coli* linkage map, located at 99 min., than *serB* at 99.5 min. (Bachmann, 1987),

and therefore should give a higher frequency of cotransduction. As a recipient, NM498 was chosen. This strain has a temperature-sensitive *dnaC* gene, prohibiting growth at 42°C. It is also restriction deficient but modification proficient ($r_K^- m_K^+$). Transductants selected for *dnaC*⁺ by growing at 42°C were tested for replacement of the resident *hsd_K* genes by *hsd* genes of a novel specificity. Unmodified λVir (λV.0) was spotted on a lawn of the transductant to test for restriction; surviving phage were picked to test for modification with K-specificity, or a new, non-K specificity, or no modification, indicating that the DNA cotransduced with *dnaC* does not contain *hsd* genes.

P1kc phage was used for the cotransduction experiments as it grows well on *E. coli* K-12 (Lennox, 1955). Again, titres of P1 lysates made on the wild type strains were often low, and still grew poorly when tested for modification on the wild type strain. Both plate lysates and liquid lysates were attempted for all strains in which type I restriction was suggested. Only RM74A and BRL-ET13 yielded lysates of high titre (approximately 10¹⁰ p.f.u./ml when tested on NM555). These lysates were used to transduce NM498 to *dnaC*⁺.

Transduction with P1.RM74A gave few *dnaC*⁺ transductants. A total of 30 colonies grew on eight plates; all 30 were purified at 42°C. The specificities of the transductants were tested by spotting λV.0 and λ modified by growth on RM74A (λV.RM74A) on lawns of RM74A. If genes encoding a new system had been cotransduced with the *dnaC* gene, then one would expect λV.0 to be restricted and λV.RM74A to be protected against restriction. If the transductants retained K-modification, then both indicator phages would be modified against K-restriction if subsequently tested on C600 ($r_K^+ m_K^+$). It was found that all 30 transductants retained their K-modifying phenotypes, and thus, cotransduction of the *hsd* genes of RM74A was absent in all cases. The absence of cotransduction may suggest a lack of homology in this region of the genome between RM74A and NM498, reducing the efficiency of generalized transduction. However, it is also possible that many or all of these colonies were simply from cells which had reverted to *dnaC*⁺.

Transduction of NM498 to *dnaC*⁺ with P1.BRL-ET13 yielded 131 colonies on eight plates. Sixty of these colonies were purified and tested for cotransduction of the BRL-ET13 *hsd* genes. None of these transductants was restricting, indicating that the *hsd*_K genes were not replaced by *hsd* genes conferring an r⁺m⁺ phenotype of a different specificity. In fact, 14 of the transductants were found to have lost the K-modifying phenotype without concomitant gain of BRL-ET13 *hsd* genes. This implies that there are no *hsd* genes in this region of the BRL-ET13 genome. If a type I r-m system is encoded by BRL-ET13, it must be non-allelic to the other chromosomally-encoded *hsd* genes.

3.4 DISCUSSION

Screening with phage P1 proved to be of little value in determining the distribution and diversity of type I restriction and modification systems among natural isolates of *E.coli*.

Results of testing the laboratory strains with the P1 screen had been promising. In keeping with the results documented by Iida *et al.* (1987a), members of type Ia and Ib families were detected. In addition, restriction and modification of the P1 phages by the type Ic system, *EcoR*124, was evident.

Unfortunately, resistance of wild type *E.coli* to both P1 lysogenization and lytic growth made it impossible to say unequivocally which of the ten natural isolates encoded type I systems. In a number of cases, restriction was suggested by a lower e.o.p. on the wild type strain as compared to the non-restricting control. For example, a reduced e.o.p. of P1Cm1, seen with all the strains, though less so for RM74A, might be indicative of a type II or type III system. RM74A has been shown to have a type II system, as well as another, unidentified restriction and modification system (N.E.Murray, unpublished observation). In addition, fewer lysogens upon P1CmTc1 infection than P1Cm1 infection, as seen with 629, 653, VB1, BRL-ET13, RM66A, and RM74A, could be interpreted as evidence of a type I restriction system encoded by these strains. The fact that strain 629 encodes a system with B-specificity

(Gough and Murray, 1983) gives credence to such an interpretation. However, suggestion of restriction must be confirmed by demonstrating concomitant modification of the surviving phage. In most cases this was not possible. Titres of lysates made on many strains, such as 629, 653, RM01A, and RM66A, were too low to give significant numbers of lysogens upon testing for modification. Therefore, in these cases, it was not possible to differentiate between P1 restriction and P1 resistance. Similar results were observed when testing for restriction by A58, a λ -resistant natural isolate (Duguid *et al.*, 1955) encoding the type Ib system, *EcoE* (Fuller-Pace *et al.*, 1985). Wild type P1Cm1 plated on A58 with an efficiency of 0.1, while the *dar*⁻ phage P1CmTc1 repeatedly showed a plating efficiency of 3×10^{-3} . This 30-fold difference in plating efficiency attests to restriction by *EcoE*. However, modification was not demonstrable as no phage were recovered after induction of P1Cm1.A58 and P1CmTc1.A58 lysogens.

RM74A and BRL-ET13 were the only strains in which a degree of modification was evident for both P1 phages. The differential plating ability of the two P1 phages also suggested the presence of type I r-m systems, with perhaps a type II or type III system additionally in BRL-ET13. VB1 also appeared to encode a type I r-m system, though modification was only demonstrated for P1CmTc1. The lack of modification of P1Cm1 by VB1 was puzzling. It may indicate that the reduced efficiency of plating of P1Cm1.NM1 on VB1 was due to a degree of resistance of the strain to infection by this phage though not to P1CmTc1.

It was apparent that the results of Robeson *et al.* (1980) suggesting that 70% of *E. coli* strains are sensitive to P1 were misleading. They screened for the ability of P1Cm to lysogenize the bacteria but made no mention of making lysates. Resistance of the wild type *E. coli* strains was obviously not limited to one step of P1 infection. For the formation of chloramphenicol-resistant lysogens, the bacteria must be permissive for P1 adsorption, DNA injection and cyclization, and maintenance of the P1 prophage during successive rounds of chromosomal replication. Production of modified phage lysates requires phage induction and vegetative growth before the eventual lysis of the cell.

The first stage of infection of *E. coli* by P1 is adsorption. The P1 receptor in *E. coli* is the terminal glucose of the lipopolysaccharide (LPS) core of the outer membrane (Sandulache *et al.*, 1984). More than one core type has been identified in *E. coli* (Stocker and Mäkela, 1978). Wild type ("smooth") *Salmonella enterica* serovar typhimurium has an O-antigenic side chain of repeating oligosaccharide units of mannose, abequose, rhamnose, and galactose attached to the terminal glucose of its LPS core. This side chain renders the *Salmonella* strain resistant to P1, presumably because it interferes with the ability of P1 to reach the glucose of the core (Ornellas and Stocker, 1974). P1-sensitive *Salmonella* mutants can be isolated. These "rough" forms have lost, for example, their side chains and part of the LPS core due to a mutation in the *galE* gene which results in a galactose-deficient LPS (Ornellas and Stocker, 1974). Bullas and coworkers (1980) found that only two strains of all the *Salmonella* in their collection were sensitive to P1 infection. It was necessary, therefore, to select for *galE* derivatives of their strains before conducting P1 cotransduction experiments.

A large number of O-side chains exist in *Salmonella*, accounting for the variety of O-serotypes in these bacteria (LeMinor, 1984). *E. coli* are believed to have over 170 different O-antigens (Ørskov, 1984). It may well be that a number of wild type *E. coli* strains express O-antigens which have a similar effect as those in *Salmonella*, in causing resistance to P1 by interfering with adsorption. In support of this is the fact that *E. coli* K-12, B, and C, which are P1 sensitive, have no O-side chains. Selection for *galE* mutants might overcome P1 resistance in strains with O-antigens containing galactose; however there could well be other host factors which would still prevent productive P1 infection.

At any point in a phage life cycle that one of its gene products interacts with a host-derived factor, a productive infection may be halted by a mutation that prevents this interaction. For example, wild type (N-dependent) λ phage requires a host factor, the product of the *nusA* gene, for maximal function of the λN gene product during *E. coli* infection (Baron *et al.*, 1970; Friedman and Baron, 1974). The *N* gene product acts with NusA through *nut* sites in the λ genome to render subsequent

downstream transcription resistant to a number of termination signals (Olson *et al.*, 1982). NusA is thought to recognize an 8bp sequence known as *boxA* (Olson *et al.*, 1982). N-dependent λ fail to grow in *Salmonella*, and in *E. coli* encoding *Salmonella* NusA (Friedman and Baron, 1974). Comparison of the *boxA* sequence of λ and of the *Salmonella* bacteriophages 21 and P22 highlight a base pair difference in a highly-conserved portion of the *boxA* sequence (Olson *et al.*, 1982). Friedman and Olson (1983) isolated a mutant λ phage able to utilize the *Salmonella* NusA. The phage carried two types of mutations, one in its *N* gene, and a second in the *nutR*-associated *boxA* sequence, an A:T to T:A transversion resulting in a *boxA* sequence identical to the conserved portion of the *boxA* sequences of phages 21 and P22. Thus, a single point mutation in *boxA* and a corresponding change in *N* to allow the *N* gene product to interact with the mutant *boxA* sequence, is sufficient to change the host range of λ from *E. coli* to *Salmonella*. If such a dramatic effect on the ability of λ to infect *E. coli* may be caused by a point mutation, it seems likely that similar effects might be seen with P1.

Attempts to transfer the putative *hsd* genes from the natural isolates of *E. coli* to a λ -sensitive laboratory *E. coli* strain failed. P1 cotransduction, confined to strains sensitive to P1kc did not transfer the *hsd* genes from either BRL-ET13 or RM74A. The results with RM74A suggested that either cotransduction of the *hsd* region with *dnaC* did not take place, or the colonies growing on the chloramphenicol plates at 42°C were merely revertants to *dnaC*⁺. The cotransduction experiment suggested BRL-ET13 does not have *hsd* genes which are allelic to the *hsd* genes of the chromosomally-encoded type I r-m systems. In this respect, it resembles *E. coli* C, which has no *hsd* genes at this location (Daniel *et al.*, 1988), although it appears to differ in that evidence suggests that restriction and modification genes must be located elsewhere in BRL-ET13. A lack of homology with the recipient chromosome in the region of *dnaC* might have contributed to the failure of cotransduction of the *hsd* genes of RM74A. The experiments of Rayssiguier *et al.* (1989) showed that such problems due to poor homology could be overcome by using recipient bacteria in which the methyl-directed mismatch system has been disrupted. Inactivation of either the *mutS*, *mutL*, or *mutH* genes by transposon mutagenesis was shown to increase the

frequency of transduction between *Salmonella enterica* serovar typhimurium and *E.coli*, which are approximately 20% divergent in DNA sequence, up to 100-fold. It was possible that such a relaxation in the requirement for homology for recombination might improve the transduction frequencies between the wild type *E.coli* strains and the K-12 derived recipient. *E.coli* strains with a *tn5* transposon inactivating either the *mutS* or *mutL* genes were kindly provided by David Thaler. P1 transduction of these mutant genes created recipients for cotransduction experiments. As the reversion frequency to *dnaC*⁺ was high in NM498 possessing the mutator phenotype (N.E.Murray, pers. comm.), *serB*-deleted recipients with the transposon in *mutS* or *mutL* were constructed instead. However, the mutator phenotype had little effect on the transduction frequency, perhaps because of the greater distance of the *serB* marker from prospective *hsd* genes.

As only one of the wild type strains, RM74A, is λ -sensitive, the possibility of conferring λ -sensitivity to some of the other strains was considered. The λ receptor, the product of the *lamB* gene, also plays a role in maltose uptake and utilization. Transformation of λ -resistant bacteria with a plasmid carrying the *lamB* gene (DeVries *et al.*, 1986) may confer λ -sensitivity, and the transformants may be selected by their simultaneous ability to utilize maltose, by plating on maltose MacConkey agar plates. However, all the λ -resistant wild type strains were able to utilize maltose, suggesting that they already encode the λ receptor and that resistance must affect some other stage in the λ life cycle.

The possibility that a host specificity system might be carried on a plasmid was also examined. Most of the wild type bacteria (though not RM74A) used in the screen harbour large, presumably conjugative, plasmids. The genes for the *EcoRI* r-m system are carried on a large R plasmid with conjugal transfer ability (Roulland-Dussoix *et al.*, 1974; Hedgpeth *et al.*, 1972). Indeed, the *hsd* genes of *EcoR124* are carried on a large plasmid (Bannister and Glover, 1968; Hedges and Datta, 1972). Conjugal transfer of the plasmids present in the natural isolates into a $r_K^- m_K^-$ *E.coli* strain was attempted in the hope, however unlikely, that cotransfer of restriction and

modification genes also occurred. As none of the plasmids were found to carry antibiotic resistance genes, it was necessary to transform the strains with a mobilizable plasmid carrying ampicillin resistance as a selectable marker in conjugation. Cotransfer of two plasmids was successful only with VB1, and the control strain harbouring the R124 plasmid. However, although conjugation transferred an $r_{R124}^+m_{R124}^+$ plasmid to the recipient, conjugation of the plasmid from VB1 did not result in an $r_{VB1}^+m_{VB1}^+$ phenotype.

Another tempting method of transferring *hsd* genes to a laboratory *E.coli* strain is the P1 cloning system of Sternberg (1990, 1992). Briefly, this system will package the P1 vector DNA with insert DNA as large as 100kb into P1 phage particles for injection into an *E.coli* host. Once inside, the DNA cyclizes via *lox-cre* recombination, and is maintained stably as a plasmid. In this way, expression of a novel type I r-m system by the prophage-encoded *hsd* genes could be demonstrated, assuming that introduction of extra restriction systems would not be harmful to the cell. However, in the absence of experience in making P1 packaging extracts and given the expense of commercially produced ones, this approach was not attempted.

It is concluded that the P1 screen can do little more than suggest the presence of type I restriction in P1-sensitive bacteria. Bullas *et al.* (1980) found it necessary to include four *Salmonella*-specific bacteriophages in their screen for restriction among *Salmonella* strains representing 85 different serotypes. Even then, only 23 strains were sensitive to any one of the tester phages. Possibly, a successful screen for type I restriction and modification systems in *E.coli* should have involved screening with more than one bacteriophage. Results from testing phage T7 and its mutant derivative, T7ocr, for their ability to infect the ten *E.coli* strains indicated that including these phages would not have expanded the scope of the screen. Phage P3, with a host range including *E.coli* and *Salmonella* (Nutter *et al.*, 1970) was also considered, but a shortage of time precluded testing this as an indicator phage.

CHAPTER FOUR

THE DISTRIBUTION OF TYPE IA AND IB *hsd* GENES AMONG SOME MEMBERS OF THE ECOR COLLECTION

4.1 INTRODUCTION

Screening bacteria with phage as indicators of restriction was found to be of limited use (Chapter 3) since it is confined to a minority of natural isolates, those which are sensitive to every aspect of infection by the phage or phages used. Screening by DNA hybridization, on the other hand, is not confined to phage-sensitive bacteria, though it is restricted by the availability of probes.

Using probes constructed from *hsd_K* and *hsd_A*, Daniel *et al.* (1988) screened a number of standard laboratory strains as well as wild type *E. coli* isolates for evidence of related type I r-m systems. Their results, based on a small number of strains, suggested that the majority of the wild type strains screened did not possess *hsd* genes of the Ia or Ib families of type I systems. If strains commonly lack type I r-m systems, this would cast doubts upon their general importance to the bacterial cell.

Results consistent with the occurrence of alternative, though homologous, *hsd* genes conferring different specificities have been reported in *S. enterica* strains of different serotypes (Bullas *et al.*, 1980). Eight *Salmonella* strains with *hsd* genes apparently allelic to *hsd_K* and *hsd_A* were tested with A-specific and SB-specific probes; only one, ST from *S. enterica* serovar thompson, appeared to hybridize weakly to the SB-specific probe (Ryu *et al.*, 1988). It is possible that alternative specificity systems also occur in *E. coli* strains.

This chapter describes an extension of the hybridization survey to the ECOR collection of wild type *E. coli* strains (Ochman and Selander, 1984a). The ECOR collection is composed of 72 wild type *E. coli* strains isolated from a number of sources and geographical locations, comprising clinical and non-clinical isolates from man, as well as domestic and zoo animals. The collection is considered to be representative of the genotypic diversity of the species as a whole on the basis of multilocus enzyme electrophoresis (MLEE; Ochman *et al.*, 1983; Whittam *et al.*, 1983). MLEE detects mobility variation of enzymes in starch-gel electrophoresis caused by differences in electrostatic charge (Selander *et al.*, 1986). As these

differences are due to amino acid substitutions, mobility variants may be directly equated with alleles at the corresponding gene locus. A number of mobility variants, or electromorphs, may be detected for a particular enzyme. Isolates may be characterized by their combinations of electromorphs over a number of loci, giving electromorph profiles. In developing the ECOR collection, electromorph profiles over eleven loci were examined (Ochman and Selander, 1984a). Estimates of nucleotide sequence divergence as determined by DNA hybridization agree with classification by MLEE (Ochman *et al.*, 1983), as does biotyping based on, among other things, the ability of the strains to utilise seven sugars (Miller and Hartl, 1986).

The ECOR collection, constructed on the basis of enzyme polymorphism, provides a phylogenetic framework within which the distribution of other traits may be determined and analysed in evolutionary terms. A phylogenetic tree (figure 4.1) shows the relationship of these strains as determined by MLEE (Selander *et al.*, 1987; Herzer *et al.*, 1990). The ECOR collection has been used in a number of studies, and as a result a picture is arising of mechanisms involved in the evolution of *E.coli*.

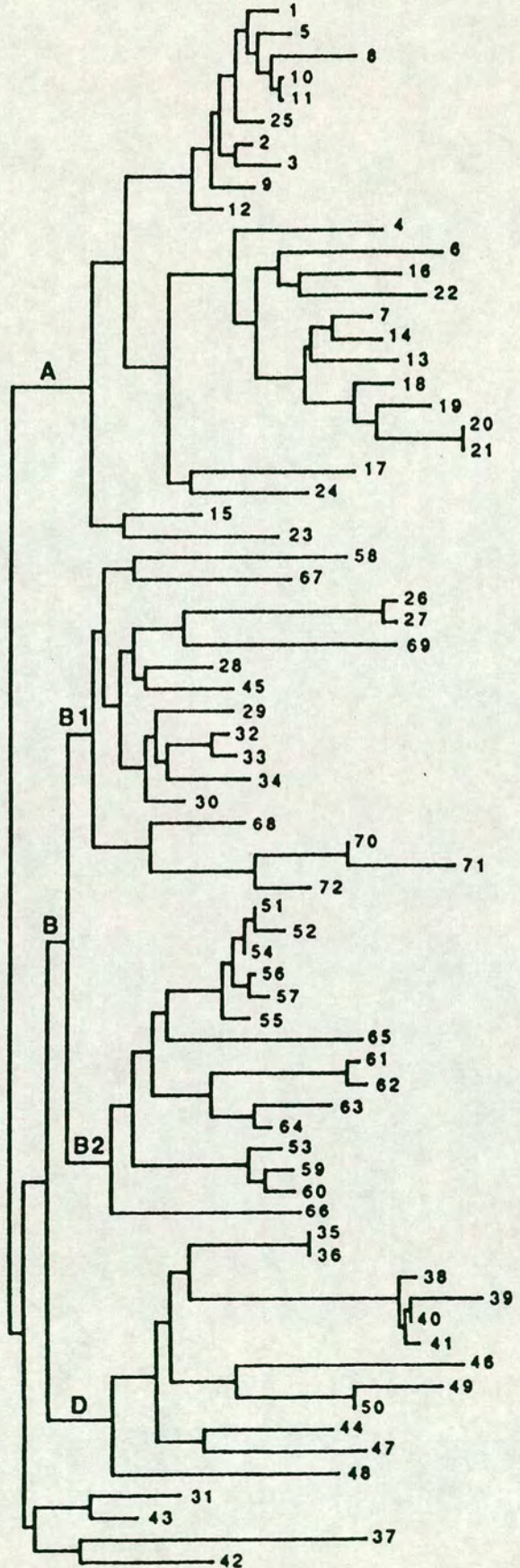
Screening members of the ECOR collection for the presence of K- or A-related *hsd* genes should give an indication of the distribution and frequency of type Ia and Ib enzymes in *E.coli*.

The distribution of these systems among the strains of the ECOR collection should provide information of an evolutionary nature. One question to be asked is whether distribution corresponds to the grouping of the strains based on the results of MLEE. Perhaps type I enzymes, or particular families of enzymes, are confined to groups of closely-related strains. If, however, similar systems are found in bacterial strains deemed to be generally dissimilar on the basis of MLEE, this would have important implications. It might suggest that variation in enzyme mobilities is not a reliable method of classification of *E.coli*. Alternatively, and more likely, it might indicate horizontal transfer of *hsd* genes between distantly-related strains, even though recombination on a large scale is believed to be uncommon in *E.coli* (Selander and Levin, 1980; Caugant *et al.*, 1981; Milkman and Crawford, 1983; Ochman and

Figure 4.1: Phylogenetic distribution of *E.coli* strains of the ECOR collection.
The tree shows the genetic distances based on polymorphisms of enzymes encoded by 38 loci. (From Herzer *et al.*, 1990)

Figure 4.1

Distance
0 0.10



Selander, 1984b, Whittam *et al.*, 1984).

4.2 RESULTS

A Hybridization Using *hsd*-Specific Sequences as Probes.

EcoRI digests of bacterial DNA were transferred to nylon filters following separation by electrophoresis through agarose gels. The filter was hybridized with probes labelled non-radioactively, using initially the Amersham ECL gene detection system, but more recently the Boehringer DIG kit. The probes were generally linearized plasmid consisting of a DNA fragment cloned in pBR322. Linearized pBR322 was also used as a probe in most cases, to differentiate hybridization to sequences within pBR322, as opposed to hybridization to the insert DNA. Occasionally, purified DNA either from a plasmid or a λ phage was used for probing. Typical results are shown in figures 4.2 and 4.3.

Initially, DNA from strains throughout the ECOR collection was probed with a K-specific probe (linearized pRH1; see Chapter 2, figure 2.1 A) and an A-specific probe (linearized pFFP32; figure 2.1 B). The sixteen strains probed included a number that had been used in other surveys of the genetic diversity in *E.coli* (Milkman and Crawford, 1983; Dubose *et al.*, 1988; Dykhuizen and Green, 1991), as well as a sample chosen to represent different sections of the phylogenetic tree.

Three of the strains hybridized to either the K or the A probe (figures 4.2 and 4.3; Table 4.1). One of the strains, ECOR70, had a single *EcoRI* fragment of approximately 5.5kb which hybridized to the K-specific probe (figure 4.2 A). Two, ECOR17 and 42, were positive with the A-specific probe (figure 4.3 A). The hybridization pattern of ECOR42 was similar to 15T⁻ (*hsd_A*; see figure 4.2 B), with two fragments of approximately 4kb and 5kb hybridizing to the probe. ECOR17 may have shown a similar pattern, though a pBR322-specific band would be masking the 5kb band (figure 4.3 D). Both strains were rehybridized with a purified fragment from a plasmid carrying DNA from *hsd_A* (pFFP20; Chapter 2, figure 2.1 B). Only

Figure 4.2: Results of screening a sample of *E.coli* strains from the ECOR collection for the presence of K- and A-related *hsd* genes.

The DNA of the bacterial strains was digested with *EcoRI* and, after separation by electrophoresis through an agarose gel, was transferred to a nylon filter for hybridization.

The filter was probed with:

- a) Linearized pRH1 (K-specific)
- b) Linearized pFFP32 (A-specific)

Sequences which hybridized to both pRH1 and pFFP32 were taken to be specific to the vector, pBR322.

The order of strains are as follows:

Lane 1: C600 (*hsd_K*)

2: 15T (*hsd_A*)

3: ECOR4

4: ECOR16

5: ECOR45

6: ECOR67

7: ECOR69

8: ECOR70

9: ECOR68

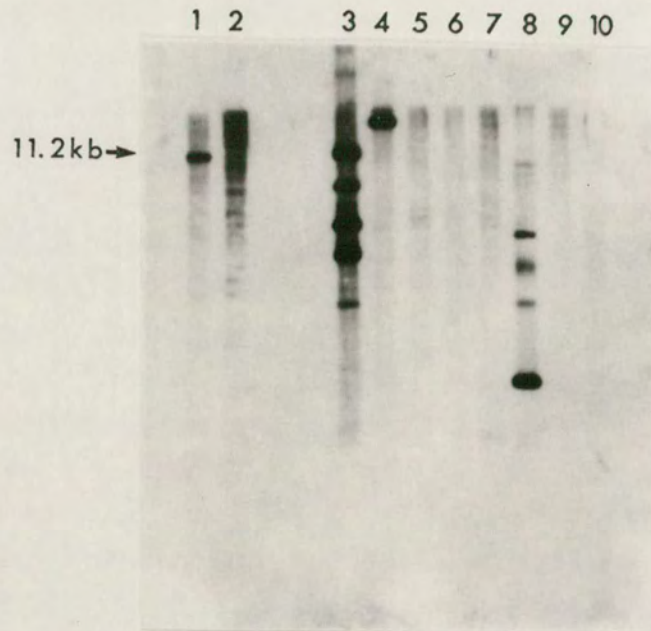
10: ECOR65

Size markers are indicated to the left of the photograph.

A summary of the data from the filters is given in Tables 4.1 and 4.2.

Figure 4.2

A) pRHI



B) pFFP32

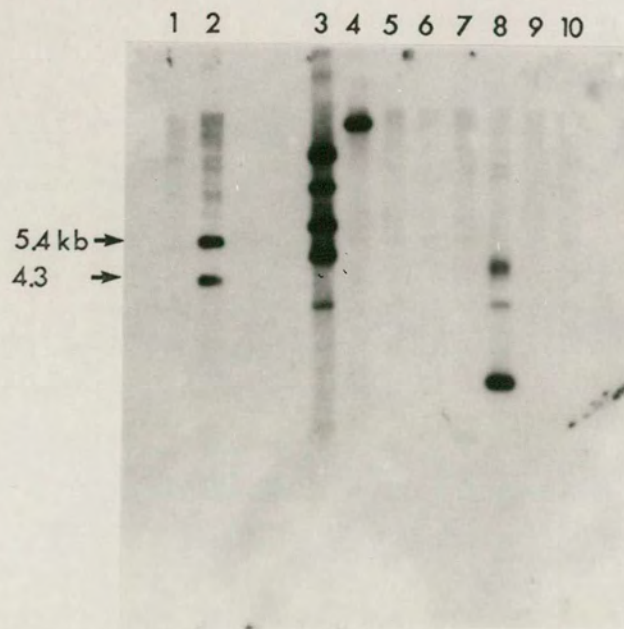


Figure 4.3: Results of screening a number of wild-type *E.coli* strains from the ECOR collection using *hsd_K*- and *hsd_A*-specific probes.

The DNA of the isolates was digested with *EcoRI* and separated by electrophoresis through an agarose gel prior to transfer to a nylon filter for hybridization.

The filter was probed with:

- A) Linearized pFFP32 (*hsd_A*-specific)
- B) Purified insert DNA from pFFP20 (*hsd_A*-specific)
- C) Linearized pRH1 (*hsd_K*-specific)
- D) Linearized pBR322.

Sequences which hybridized to both pRH1 and pFFP32 were taken to be specific to the vector, pBR322. This was verified by probing with linearized pBR322.

The order of strains is as follows:

- Lane 1: C600 (*hsd_K*)
- 2: ECOR17
- 3: ECOR19
- 4: ECOR31
- 5: ECOR37
- 6: ECOR42
- 7: ECOR46
- 8: ECOR48
- 9: ECOR58
- 10: ECOR65
- 11: ECOR68

A size marker of 11.2kb is indicated.

A summary of the data from the filters is given in Table 4.1 and 4.3.

Figure 4.3

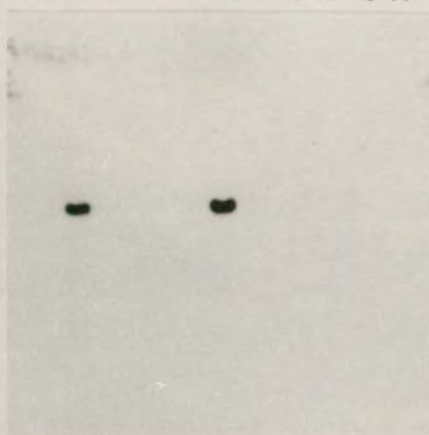
A) pFFP32

1 2 3 4 5 6 7 8 9 10 11



B) pFFP20

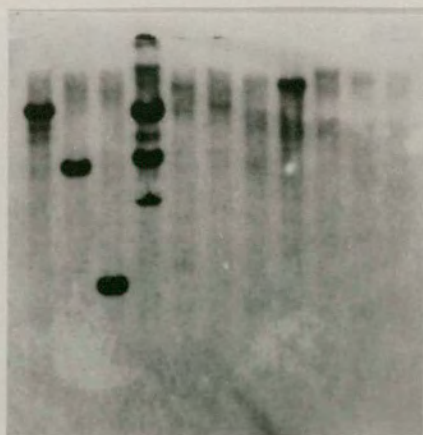
1 2 3 4 5 6 7 8 9 10 11



C) pRHI

1 2 3 4 5 6 7 8 9 10 11

11.2kb →



D) pBR322

1 2 3 4 5 6 7 8 9 10 11

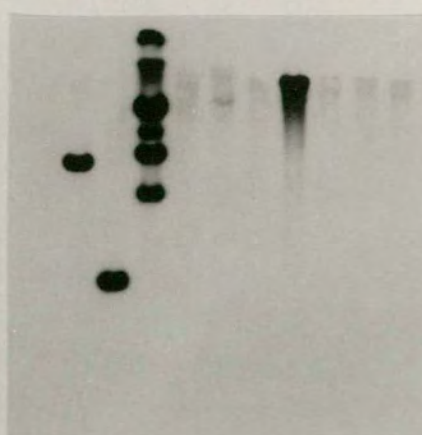


TABLE 4.1: Results of Screening a Sample of ECOR Strains with Type Ia, Ib, and McrBC-Specific Probes.

<u>ECOR No.</u>	<u>HYBRIDIZATION TO^a</u>		
	<i>hsd_K</i>	<i>hsd_A</i>	<i>mcrBC</i>
4	-	-	-
16	-	-	-
17	-	5.4kb 4.1kb	-
19	-	-	-
31	-	-	-
37	-	-	-
42	-	5.3kb 4.1kb	-
45	-	-	-
46	-	-	-
48	-	-	-
58	-	-	-
65	-	-	-
67	-	-	-
68	-	-	-
69	-	-	-
70	5.5kb	-	8.3kb

^a The approximate size of the *EcoRI* fragment hybridizing to the probe is indicated; "-" indicates no hybridization.

the 4kb fragments hybridized to this probe (figure 4.3 B), which is consistent with an *EcoRI* site being located in a similar position as that of *hsd_A* (see Chapter 2, figure 2.1). Unfortunately, this gave no further insight as to whether two fragments carry the A-like *hsd* region of ECOR17. The possibility of a positive result being overlooked due to hybridization to the vector background must therefore be considered when interpreting results.

The other thirteen strains tested, though often hybridizing to pBR322, did not display additional bands with either the K- or A-specific probes (Table 4.1).

As it was not feasible to include all 72 ECOR strains in the study, and because of the low frequency of positives in the strains so far examined, it was thought to be most informative to look at one particular section of the collection. Therefore, the study focused on the strains belonging to Group A of the ECOR collection (see figure 4.1). Based on results of MLEE, if *E.coli* K-12 and B were included in the collection, they would be assigned to this group (Herzer *et al.*, 1990). It seemed reasonable to suggest that strains closely related to K-12 and B on the basis of MLEE, might also resemble these two strains in encoding type Ia r-m systems. Obviously, the correlation is not perfect, as ECOR17 from Group A had already been shown to have similarity to *hsd_A*.

All of the 25 Group A *E.coli* natural isolates were tested. Probes specific for *hsd_K*, *hsd_A*, and *mcrBC* genes were used and each filter included *E.coli* K-12 (*hsd_K*) and *E.coli* 15T⁻ (*hsd_A*). Autoradiographs for one set of DNA samples hybridizing with each probe are presented (figure 4.4), and the data for all filters are summarized in Table 4.2.

Nine isolates hybridized to the K-specific probe, pRH1 (Table 4.2). Most, like the C600 control, had a single large *EcoRI* fragment hybridizing to the pRH1 probe. The *EcoRI* fragment containing the entire *hsd_K* region in C600 is 11.2kb. ECOR5, 12, 13, 23 and 24 each had one K-specific fragment ranging in size from ~10kb to 12.5kb (Table 4.2). ECOR25, like ECOR70, appeared to have a smaller

Figure 4.4: An example of the screening of strains of the ECOR collection for the presence of K- and A- related *hsd* genes, *mcrBC*, and sequences flanking the *hsd* region. The DNA of the bacterial strains was digested with *EcoRI*, separated by electrophoresis through an agarose gel, and then transferred to a nylon filter for hybridization.

The filter was probed with:

A) Linearized pFFP32 (*hsd_A*-specific)

B) Linearized pRH1 (*hsd_K*-specific)

C) Linearized pRH2 (*mcrBC*-specific)

D) Linearized pBR322

E) A 5.3kb region of *E.coli* C spanning a region which, in *E.coli* K-12, includes the *hsd* region (see figure 4.6).

The order of the strains is as follows:

Lane 1: ECOR17

2: ECOR42

3: ECOR15

4: 15T (*hsd_A*)

5: C600 (*hsd_K*)

6: ECOR70

7: ECOR12

8: ECOR13

9: ECOR24

Size markers are indicated on the sides of the photographs.

Probing with pBR322 confirms that sequences which hybridized to all three probes, pRH1, pRH2, and pFFP32, were pBR322-specific. The photograph also highlights problems encountered when rehybridizing using the Boehringer DIG gene detection system. The 5.3kb C-probe (filter E), used prior to hybridization with pBR322, was not removed completely before application of the second probe (filter D). This is evidenced by faint traces of C-specific bands (e.g. ECOR17, Lane 1) which had given particularly strong signals. A summary of the data from filters A, B, C is given in Table 4.1, those from filter E in Table 4.3.

Figure 4.4

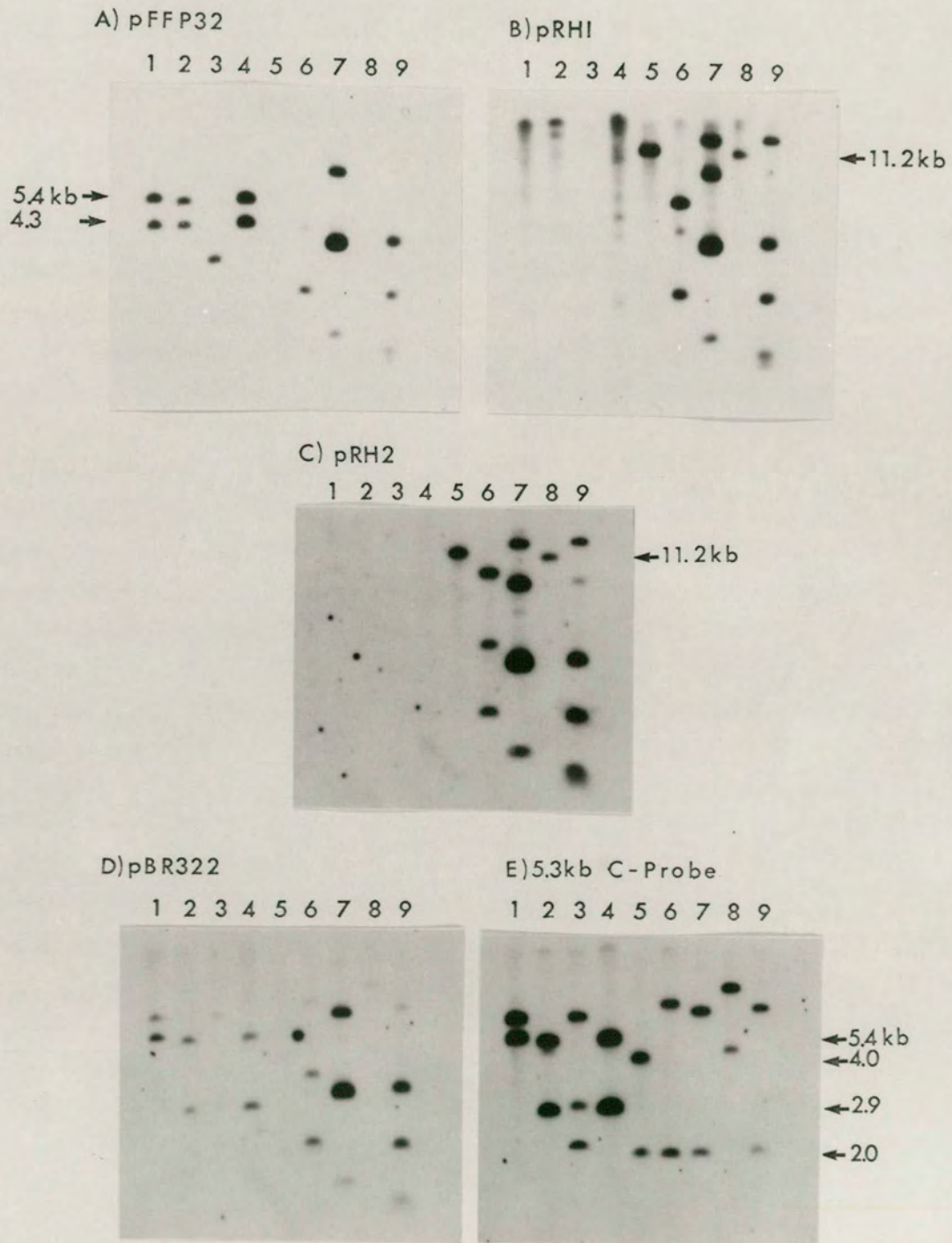


TABLE 4.2: Results of Screening ECOR Strains of Group A with Type Ia,Ib, and McrBC-Specific Probes.

<u>ECOR No.</u>	<u>HYBRIDIZATION TO^a</u>		
	<i>hsd_K</i>	<i>hsd_A</i>	<i>mcrBC</i>
1	-	-	-
2	-	-	-
3	-	-	-
4	-	-	-
5	11.2kb	-	-
6	-	-	-
7	-	-	-
8	-	-	-
9	-	-	-
10	-	-	-
11	12kb, 5kb	-	12kb
12	12.3kb	-	12.3kb
13	10.2kb	-	10.2kb
14	5.8kb	-	11.7kb
15	-	3.1kb	-
16	-	-	-
17	-	5.4kb, 4.1kb	-
18	-	-	-
19	-	-	-
20	-	-	-
21	-	-	-
22	-	-	-
23	11.2kb	-	11.2kb
24	12.3kb	-	12.3kb
25	5.4kb	-	-

^a The approximate size of the *EcoRI* fragment hybridizing to the probe is indicated.

single *EcoRI* fragment of approximately 5-6kb. Both also had other pBR322-specific fragments which could conceivably be masking further K-specific fragments. ECOR11 appeared to have two fragments of 12kb and 5kb hybridizing to the *hsd_K* probe. ECOR14 had a single fragment of 6kb which hybridized to pRH1, and a larger pRH2-specific fragment of approximately 12kb. Therefore, both ECOR14 and ECOR11 seem likely to have extra *EcoRI* sites in this region of the chromosome which are absent in K-12.

One strain in Group A, in addition to ECOR17, was found to have DNA with similarity to *hsd_A*. ECOR15 has a single *EcoRI* fragment of roughly 3kb which hybridizes to the A-specific probe. There are no pBR322-like *EcoRI* fragments in this strain which might conceal a further A-specific fragment (figures 4.4 A and D).

Figure 4.5 shows the hybridization results superimposed on the phylogenetic tree of the ECOR strains constructed on the basis of multilocus enzyme electrophoresis. Similarity based on *hsd* genes and similarity based on enzyme polymorphisms did not always correlate. Type Ia r-m systems were not confined to strains which were closely related to K-12 or B; *hsd_K*-like genes were also identified in a strain (ECOR70) belonging to a different group.

The division of type I families according to strain similarity was also not possible. Members of two families were identified both within Group A and also in other groups. In fact, some pairs of strains deemed most closely related on the basis of MLEE showed homology with *hsd* genes of alternative families (for example, ECOR17 and 24; ECOR15 and 23). In other cases, for example ECOR10 and 11, *hsd* genes were detected in one strain, but not the other.

The DNA digests of the ECOR isolates used in this survey were also hybridized with a probe (pRH2; Chapter 2, figure 2.1 A) consisting of most of the *mcrBC* region and some downstream sequence from *E.coli* K-12. Daniel *et al.* (1988), when screening a number of enteric bacteria with pRH2, found that *mcrBC* was only detected in strains encoding a type Ia restriction system, as if the *hsd_K*-like


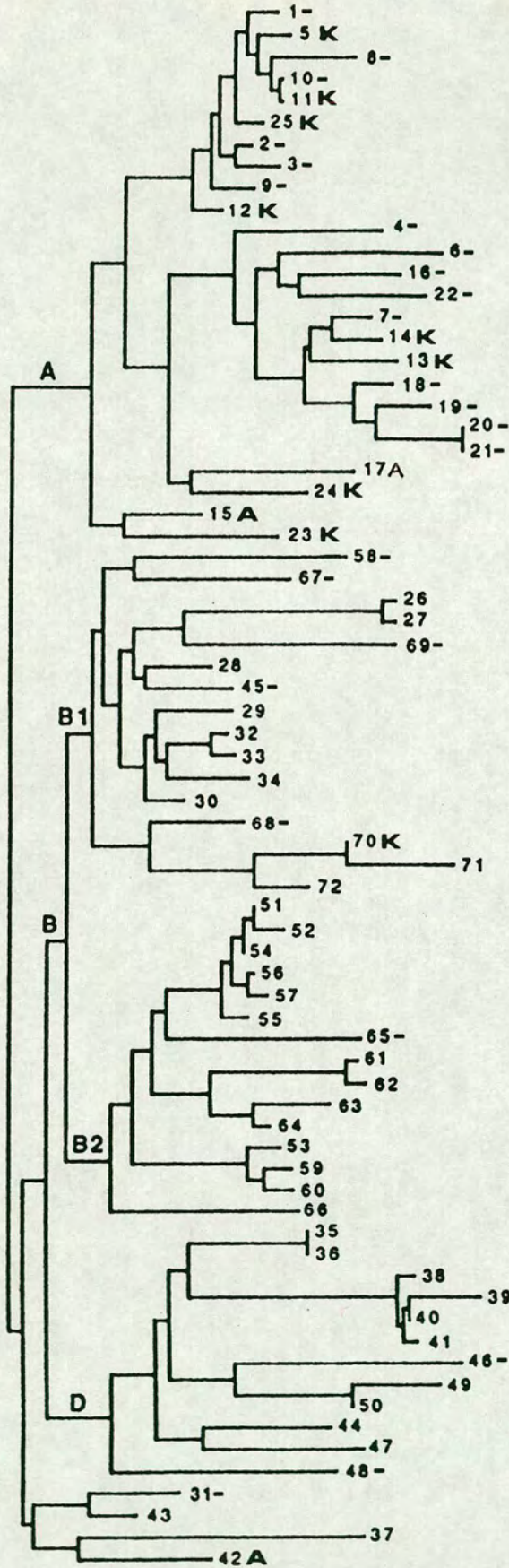
Figure 4.5: Distribution of *hsd_K*- and *hsd_A*-related sequences among some members of the ECOR collection. Results obtained from Southern hybridization analysis were applied to a phylogenetic tree of the ECOR strains showing relatedness based on multilocus enzyme electrophoresis (from Herzer *et al.*, 1990). Strains marked with a 'K' hybridize to the K-specific probe, pRH1; those with an 'A' to the A-specific probe, pFFP32. Strains not hybridizing to either probe are designated '-'.


Figure 4.5

Distance
0 0.10



genes and *mcrBC* genes were acquired concomitantly on a mobile genetic element. However, homology was not detected in all Ia-family strains; both E166 (encoding *EcoD*) and *S. enterica* serovar typhimurium lacked these genes.

Tables 4.1 and 4.2 reinforce the conclusions of Daniel *et al.* (1988). Evidence of *mcrBC* genes was found only in isolates determined to have DNA with homology to *hsd_K*; no A-like systems among these strains were found to be associated with *mcrBC*. Strains lacking K- or A-like systems were also negative. Two K-like strains, ECOR5 and 25, did not hybridize to the *mcrBC* probe. Thus, *mcrBC* appears only to be associated with a subset of strains possessing K-like *hsd* genes.

B Hybridization Using DNA Flanking the *hsd* Region as Probes

E. coli C lacks DNA with homology to any known restriction genes. Probes specific for *hsd_K*, *hsd_A* and *mcrBC* do not hybridize to DNA from *E. coli* C (Daniel *et al.*, 1988). However, DNA flanking the *hsd_K* region does hybridize to DNA sequences in *E. coli* C. In fact, flanking probes from either side of the *hsd_K* region both hybridize to the same DNA fragment in an *EcoRI* digest of *E. coli* C DNA (Daniel *et al.*, 1988). The size of the fragment, 5.3kb, precludes the presence of genes encoding a type I r-m system in this region of the genome.

It was not possible to conclude whether ECOR strains which do not hybridize to K- or A-specific probes lack *hsd* genes, or have *hsd* genes which are sufficiently dissimilar to be undetectable by hybridization with these probes. As probes from other families become available, such a question will be more easily addressed. However, results from probing *E. coli* C with flanking DNA suggest that such an approach might be extended to some of the probe-negative ECOR isolates.

The 5.3kb *EcoRI* fragment from *E. coli* C has been cloned in λ (Daniel *et al.*, 1988). From this, it was possible to purify the fragment to make a probe. This probe therefore consists of DNA from *E. coli* C which is similar to that flanking the *hsd* and *mcrBC* genes in K-12 and in fact spans a region in K-12 of approximately

18kb.

The size of the fragment in a genomic *EcoRI* digest that hybridized to the C probe may suggest, in a manner similar to that seen when probing *E.coli* C with flanking DNA, the absence or possible presence of *hsd* genes. This, of course, relies largely on the location of *EcoRI* sites within the DNA of the strain to be tested. If the probe hybridized to a single *EcoRI* fragment judged to be too small to include *hsd* genes, this would suggest a lack of *hsd* genes in this region of the chromosome. If it hybridized to a large fragment, it might imply that *hsd* genes are present, although this could be misleading as the similarity to the probe could be confined to a small portion of the fragment. Hybridization to two or more fragments might suggest *hsd* genes lying within two flanking regions, but would require further analysis.

The C probe was initially applied to a filter containing *EcoRI* digests of a sample of ECOR strains which had hybridized to either pRH1 or pFFP32 (figure 4.4 E). For both K (C600) and A (15T⁻) DNA digests, the C probe hybridized to two fragments. In K, these were judged to be a 2kb fragment upstream of *hsdR*, and a 4kb fragment downstream of *mcrBC* (see figure 4.6 for a map). In A, the fragments were most likely the 5.4kb fragment which hybridized to pFFP32, and a 2.9kb fragment downstream of *hsdS* (figure 4.6).

The results of probing all of the ECOR strains previously screened for the presence of *hsd_K*- and *hsd_A*- related genes are summarized in Table 4.3.

Two of the three A-positive strains, ECOR17 and 42, had two fragments which hybridized to the C-probe, and in fact, ECOR42 showed a similar hybridization pattern to that seen with A (Table 4.3). The third A-positive strain, ECOR15, had three fragments.

Most of the K-like strains, including ECOR70 (figure 4.4 E), demonstrated a pattern of hybridization similar to that seen with K, that is, at least two fragments hybridized to the C probe, the smaller of which was approximately 2kb (Table 4.3).

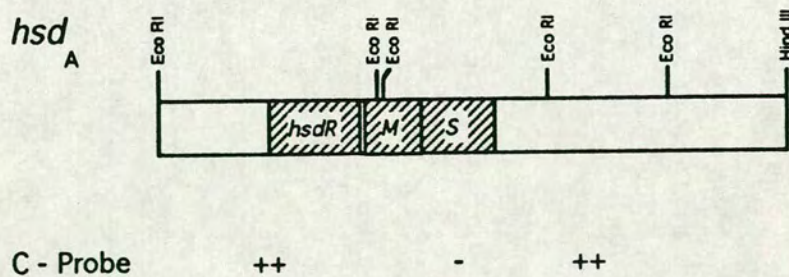
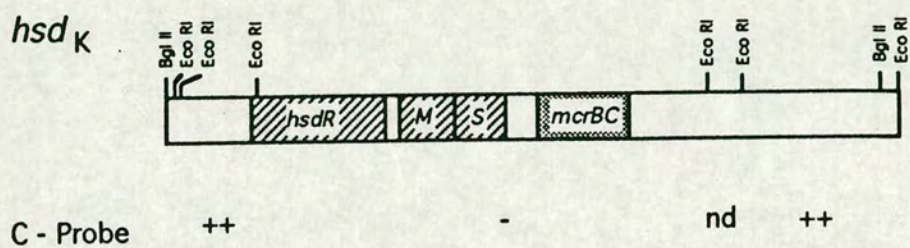
Figure 4.6: Hybridization of the 5.3kb C-probe and probes derived from this fragment to regions flanking the *hsd* genes of C600 (K) and 15T⁻ (A).

A nylon filter containing *EcoRI* digests of DNA including that of C600 and 15T⁻ was probed with the purified 5.3kb C-specific *EcoRI* fragment from λ AD12.

Results of probing C600 and 15T⁻ with the C-probe are summarized here beneath maps of the *hsd_K* and *hsd_A* regions ("-", "+", and "++" indicate degree of hybridization; "nd" means not determinable). The probe hybridizes to a 2kb fragment upstream of the C600 *hsdR* gene and a 4kb fragment downstream of *mcrBC*. With 15T⁻, the probe hybridized to a 5.4kb fragment which includes *hsdR* and some upstream, and a 2.9kb fragment downstream of *hsdS*.

An example of the results of probing some members of the ECOR collection with the 5.3kb C-probe may be seen in figure 4.4.

Figure 4.6



1kb

TABLE 4.3: Results of Hybridization of the 5.3kb C-probe to Members of the ECOR Collection.

A) Group A Strains

<u>ECOR No.</u>	<u>Hybridization to</u> C-probe ^a
1	5kb, 4kb
2	20kb, 12.5kb, 4kb
3	4kb
4	5kb
5	9.5kb, 2.1kb
6	9.5kb
7	4.5kb
8	4kb
9	6kb, 2.5kb
10	6.5kb, 2.0kb
11	9.5kb, 2.0kb
12	7kb, 2.1kb
13	10kb, 4.5kb
14	2.1kb
15	6.5kb, 3.5kb, 2.2kb
16	4.5kb
17	6.5kb, 5.5kb
18	-
19	-
20	3kb
21	2.9kb
22	4.5kb
23	7kb, 4.5kb, 2.1kb
24	8kb, 2.2kb
25	11kb, 7kb

TABLE 4.3 (cont'd)

B) Non-Group A Strains

<u>ECOR No.</u>	<u>Hybridization to C-probe^a</u>
31	-
37	4kb
42	5kb, 2.7kb
45	10kb
46	19.5kb, 5.5kb
48	6kb
58	5.5kb
65	7.5kb
67	7.5kb
68	3kb
69	7.5kb

^a Approximate sizes of bands are given; sizes greater than 3.0kb were rounded off to the nearest 0.5kb. "-" indicates no hybridization.

In some cases, one of the fragments hybridizing to the C probe also hybridized to pRH1, for example ECOR13. In contrast, probing ECOR14 with the C probe only detected a 2kb *EcoRI* fragment (Table 4.3). This suggests that in this case, DNA sequence homology with K-12 is restricted to one of the two regions flanking the *hsd* genes.

Although a number of ECOR strains which were negative with the K-specific and A-specific probes had a single fragment of approximately 5kb or less hybridizing to the C probe (Table 4.3), the result obtained with ECOR14 precludes any interpretation of such a finding. In order to conclude that no *hsd* genes are found in this region of the chromosome, it is necessary to demonstrate that both the "upstream" and "downstream" portions of the chromosome are hybridizing to the one fragment. To achieve this, one must reprobe the strains using peripheral portions of the C probe.

Few of the Group A strains had a single *EcoRI* fragment hybridizing to the C probe large enough to carry *hsd* genes. The DNA of some strains, for example ECOR19, in fact, did not hybridize at all, suggesting a lack of homology in this region. One strain, ECOR6, however, may have a C-specific fragment large enough to include *hsd* genes (Table 4.3), warranting further investigation. Additionally, when the C probe was applied to those strains outside of Group A that had not hybridized to either the K- or A-specific probe, a single fragment was detected in all except ECOR46. Some of these fragments appeared large enough to include *hsd* genes and it might be worth cloning them. The large fragments that hybridize to both flanking sequences are candidates for future analysis.

A problem of probes not completely washing off filters before rehybridizing, associated specifically with the Boehringer DIG gene detection system, was encountered, and is illustrated in figure 4.4 (D). In this case, the filter was probed with linearized pBR322 subsequent to hybridization with the C-probe. Faint traces of C-probe specific bands (e.g. ECOR17, Lane 1) which gave particularly strong signals on filter E (figure 4.4), are still visible after washing and rehybridizing with

pBR322. Similar problems have also been encountered by others (A.J.B.Campbell and G.King, pers. comm.). As a long series of detection steps are necessary after each hybridization in order to confirm that a probe has been successfully removed by washing, the Boehringer DIG system is obviously not ideal for experiments involving successive hybridizations.

The problems associated with incomplete removal of probes did not affect the conclusions presented. Re-examination of the results obtained when carrying out the screening with pRH1, pRH2, and pFFP32 confirmed that the problems had not arisen until after this time.

4.3 DISCUSSION

Screening thirty-six members of the ECOR collection with K-specific and A-specific probes identified *hsd_K*-like sequences in nine strains, and *hsd_A*-like sequences in three. The *hsd* genes were detected more commonly, but not exclusively, in Group A of the ECOR collection; ten out of twenty-five (40%) of Group A strains hybridized to either one or the other of the two probes. In the remaining groups of the ECOR collection, two out of eleven (19%) were positive.

These results suggest that type Ia and Ib r-m systems are more common among natural isolates than previously believed. Daniel *et al.* (1988) tested sixteen wild type *E.coli* strains of which five (31%) showed similarity to either the *hsd_K* or *hsd_A* probe. Four of these strains had previously been shown to encode type I systems (Gough and Murray, 1983; Fuller-Pace *et al.*, 1985), consequently only one of the remaining twelve strains whose restriction genotypes are unknown hybridized to either of the probes.

The results of multilocus enzyme electrophoresis indicate that *E.coli* K-12 and B are closely related to Group A strains (Herzer *et al.*, 1990). In fact, K-12 has an electromorph type identical to that of ECOR2 (R.K.Selander, in Biserčić *et al.*, 1991). Milkman and Bridges (1990) compared PCR fragments from fifteen genes

among a number of ECOR strains by restriction analysis. They reported that some strains in Group A (ECOR4, 16, 19, and 21) have identical, or very similar, sequences to K-12 for most genes examined. They referred to this group of strains as a level II clone, and suggested that this clone spans all of the Group A strains. Therefore, the apparent predominance of K-like *hsd* genes among the Group A strains could correlate with their proximity on the phylogenetic tree.

Two strains in Group A, namely ECOR15 and 17, shared homology with an *hsd_A*-specific probe. In both cases, the strains most closely-related to them were found to have K-like *hsd* genes. Additionally, K-like *hsd* genes were detected in ECOR70, a strain not included in Group A. Thus, a phylogenetic tree based on the results of the screen for *hsd* genes would not correspond to the phylogenetic tree based on MLEE. Daniel *et al.* (1988) reported similar findings in their screen of a number of enteric bacteria for genes with homology to *hsd_K* and *hsd_A*. Although BLXA, a wild type strain identical to *E.coli* B in nineteen out of twenty enzymes examined, was found to encode a type I system with B-specificity, such correlation was not generally the case. BLD4, BRL-ET2 and BRL-ET13, almost identical to K-12 in the proteins analysed, gave no evidence of specificity systems in the hybridization screen, though evidence of type I restriction has been demonstrated for BRL-ET13 (see Chapter 3). In the present screen, ECOR2, though identical to K-12 on the basis of MLEE (R.K.Selander, in Biserčić *et al.*, 1991), was found to lack a type Ia r-m system.

These results might suggest that classification of *E.coli* strains on the basis of multilocus enzyme electrophoresis is unreliable. Biserčić *et al.* (1991) expressed such doubts when they carried out sequence comparisons of *gnd* genes among selected members of the ECOR collection. Their results suggested much more variation in amino acid sequence in the enzyme encoded by *gnd*, 6-phosphogluconate dehydrogenase, than indicated using MLEE. However, *gnd* is notable for its high level of synonymous substitutions (base pair changes that do not alter the amino acid sequence), a finding that may correlate with its proximity to the *rfb* gene cluster, encoding the highly variable O-antigen of the bacterial outer membrane. As the *rfb*

locus is probably subject to frequency-dependent selection for rare O-antigen types, it may be that variants of the linked *gnd* gene are coselected with the *rfb* genes (Selander *et al.*, 1987; Biserčić *et al.*, 1991; Dykhuizen and Green, 1991).

If, as generally believed, MLEE provides a reliable classification of strains, the presence of *hsd_A*-like genes in strains of Group A is hard to explain without invoking lateral transfer. Sharp *et al.* (1992) reported that intraspecific divergence between members of the Ia and Ib families is as high as the divergence seen between *Bacillus* and *E.coli*. They invoked horizontal transfer of the *hsd* region from a distantly-related species to explain the presence of Ib family systems in *E.coli*. Transfer of the *hsd_A*-like genes to members of Group A could have occurred either directly from another species, or indirectly through another *E.coli* strain.

Though believed to be rare, horizontal transfer and recombination of genes have been invoked to explain polymorphisms detected in genes in a number of surveys of genetic variation in *E.coli* and *Salmonella* (Herzer *et al.*, 1990; Brown *et al.*, 1991; Raleigh, 1992; Groisman *et al.*, 1992; Lee *et al.*, 1992; Marklund *et al.*, 1992; Wang *et al.*, 1992). Often, the genes in question display unusual codon usage, or a G+C content different than that of the rest of the DNA of the species in which they are found, suggesting recent transfer.

Salmonella enterica, like *E.coli*, is highly polymorphic for the O-antigen, the product of the *rfb* gene cluster. Extensive homology in the flanking region of the *rfb* clusters of some antigenic groups borders *rfb* regions with limited homology. These alternative *rfb* regions may be due to one or more recombination events replacing genes with other, unrelated sets (Wang *et al.*, 1992). Because the G+C content in all of the *rfb* clusters is lower than is normal for *S.enterica*, and the codon usage is atypical, they are believed to have been acquired by lateral transfer (Brown *et al.*, 1991; Jiang *et al.*, 1991; Lee *et al.*, 1992; Wang *et al.*, 1992). Most of the divergence, in the form of recombination followed by genetic drift, is thought to have taken place in the donor strain followed by recent transfer into *S.enterica* strains, as the low G+C content is still evident (Verma and Reeves, 1989; Jiang *et al.*, 1991;

Lee *et al.*, 1992; Wang *et al.*, 1992).

Herzer *et al.* (1990) screened for the ability to produce multicopy single-stranded DNA (msDNA), a branched DNA-RNA molecule, in *E. coli* strains of the ECOR collection as well as a number of clinical isolates. The synthesis of msDNA relies upon a chromosomally-encoded reverse transcriptase, and strains were screened for the presence of this enzyme. Evidence of msDNA was found in 13% of strains in the ECOR collection. Although msDNA was distributed throughout the collection, it would appear that the ability to produce msDNA was stably maintained in a clonal manner, associated with particular electromorph types. In support of this, clinical strains found to produce msDNA often had the same MLEE types as msDNA-producing members of the ECOR collection, though these strains were isolated at different times from unassociated hosts in different geographical locations. Herzer and his colleagues suggested that msDNA-encoding genes underwent divergence in other species, and were acquired by *E. coli* lineages independently by horizontal transfer. This transfer was recent enough that closely related strains, such as ECOR35 and 36, differed in the ability to produce msDNA. This theory was supported by the observation that the reverse transcriptase genes involved in the production of msDNA have different codon usage from that normally seen in protein-encoding genes in *E. coli*, arguing for the acquisition of these genes from a distantly-related species. Once acquired, the genes remained stably in association with a particular electromorph type.

It may be that ECOR15 and 17, like their closest relatives in the ECOR collection, at one time had K-like *hsd* genes, but a fairly recent recombination event replaced these with the allelic *hsd_A* genes. Perhaps other strains which failed to hybridize to the K-probe have also acquired alternative, unrelated *hsd* genes.

Marklund *et al.* (1992) also found evidence of horizontal transfer in *E. coli* when screening members of the ECOR collection for the presence of the *pap* and *prs* operons that encode alternative forms of P pili. *E. coli* isolates associated with urinary tract infections generally possess genes encoding P pili, which bind to Gal α 1-4Gal-

containing glycolipids via their G-adhesins (Lund *et al.*, 1987). This binding property is believed to be an important virulence factor (Marklund *et al.*, 1992). P pili formation is encoded by a cluster of eleven genes. The G-adhesin, which mediates glycolipid binding, is the product of the *papG*, or homologous *prsG*, gene. Three classes of G-adhesins exist, each with a unique receptor specificity (Strömberg *et al.*, 1990; 1991); different classes are encoded by different, though homologous, gene clusters. Members of one class are highly related, though substantial sequence divergence is found between different classes. The different classes are believed to confer different tissue tropisms. The ECOR collection was screened using class-specific probes (Marklund *et al.*, 1992). Although none hybridized to a Class I-specific probe, a number of ECOR strains hybridized to either the Class II or Class III probe. These strains belonged primarily to the B2 and D groups of the collection. This was taken to mean that the genes were acquired after separation of *E. coli* to form the four lineages. In a manner analogous to the results of the *hsd* probing, and the msDNA probing (Herzer *et al.*, 1990) pairs of closely-related strains were found to encode different classes of P pili. This was postulated to be due to introduction of the entire P pili gene clusters separately into related strains.

The *pap* genes that encode components for the adhesive tip of the P pili, namely *papE*, *F*, and *G*, have a lower G+C content than do the *pap* genes required for the structure and assembly of the pilus, and in fact lower than the G+C ratio of *E. coli* in general (Marklund *et al.*, 1992). This suggests recruitment of these genes from a bacterial species with a low G+C content.

Although mainly associated with Groups B2 and D, P pili genes were occasionally found in other groups, including Group A (Marklund *et al.*, 1992). This may be a further example of horizontal transfer.

Horizontal transfer has also been invoked in the evolution of the *mcrBC* restriction system, which cleaves DNA with methylated cytosines (see Chapter 1, 1.5; Raleigh, 1992). The *mcrBC* genes have so far only been found in association with members of the Ia family of type I r-m enzymes. Not all strains of *E. coli* with *hsd*

genes that are members of this family possess *mcrBC* genes; hybridization results indicate that both E166, encoding the *EcoD* system, and *S. enterica* serovar typhimurium lack homology with *mcrBC* (Daniel *et al.* 1988). A probe constructed from a plasmid including most of *mcrBC* hybridized only to ECOR strains which also shared similarity to the *hsd_K*-specific probe. However, as found by Daniel *et al.* (1988), *mcrBC* was not always associated with K-like *hsd* genes. DNA from both ECOR5 and 25 both hybridized to the K-probe, pRH1, but nevertheless did not hybridize to the *mcrBC* probe (Table 4.1).

Raleigh (1992) hypothesized that the *mcrBC* genes form a cassette which was acquired by an ancestral *E. coli* strain independently of an *hsd* cassette. Again, evidence to support this was the fact that the G+C content of the *mcrBC* region in *E. coli* K-12 is significantly lower than the G+C content of the *hsd_K* genes, and of K-12 DNA in general. In fact, DNA with such a G+C content (40% G+C) is only found within *Enterobacteriaceae* in *Proteus* and *Providentia*, genera distantly related to *E. coli* (Raleigh, 1992). Inverted repeats on either side of this GC-rich region in K-12 suggest acquisition of this region by a transposition event (Dila *et al.*, 1990). The lack of homology with *mcrBC* seen in ECOR5 and 25 is consistent with this theory of independent acquisition of the *hsd* and *mcrBC* regions. Determination of the specificities of the type I systems encoded by these strains is necessary to check whether some strains with the same specificity as *EcoK* lack *mcrBC*.

It is not known whether strains which did not hybridize to the K-probe or the A-probe possess unrelated *hsd* genes or lack them altogether. Before we can satisfactorily explain the pattern of distribution of *hsd* genes among strains of the ECOR collection, the *hsd* genotypes for all or most of the strains must be determined. Additional probes for unrelated families, such as the plasmid-encoded type Ic family, may detect further *hsd* genes in the ECOR collection. *hsd* genes that map to the same region of the chromosome as those of *E. coli* K-12 have been identified in various serovars of *S. enterica* (Bullas *et al.*, 1980). One such strain, *S. enterica* serovar blegdam, encodes a system that may be the first member of a fourth family of type I systems (A.J.B. Campbell, pers. comm.). Other strains may encode systems that

define further families. Cloning of the relevant *hsd* genes is necessary to provide further probes for analysis of the ECOR collection.

A probe made from the DNA of *E. coli* C, known to hybridize to DNA both upstream and downstream of the *hsd* genes of *E. coli* K-12, was thought to be suitable for detecting regions that might flank alternative *hsd* genes in the DNA of strains of the ECOR collection. DNA fragments flanking the *hsd_K* genes in *E. coli* K-12 were also considered as potential probes. These fragments, cloned in pBR322 (pBg3 and pBg6; Sain and Murray, 1980), include the *hsd* and *mcr* regions of K-12, and extend approximately 2kb upstream of *hsdR* (pBg3) and 6.5kb downstream of *mcrBC* (pBg6; see Chapter 2, figure 2.1). The 5.3kb C-fragment was identified by probing *E. coli* C DNA with these two plasmids (Daniel *et al.*, 1988), even though it is devoid of *hsd* genes. The C-fragment was considered preferable to pBg3 and pBg6, in part because of this lack of sequences specific to *hsd_K* and *mcrBC*. In addition, it was hoped that the C-fragment would encompass DNA extending further into flanking regions than pBg6, and especially pBg3. Thus there would be greater chance of this single probe including sequences common to all *E. coli* strains. It was hoped that by screening with the C probe, it might be possible to gain some information as to whether strains might have DNA that could encode alternative *hsd* genes. Homologous flanking sequences were detected in most strains. The results obtained, however, indicated that in order to determine whether *hsd* genes could be located in this region of the chromosome, it would be necessary to divide the C probe into upstream and downstream components.

Initial hybridization experiments (data not shown) using putative upstream and downstream probes were ambiguous. This may have been the result of technical problems associated with the removal of one probe before application of the second. Nevertheless, the use of flanking probes provides a simple way of cloning intervening sequences.

Using sequences that flank the chromosomally-located *hsd* regions as probes can facilitate the detection of unrelated *hsd* genes. Isolation of a flanking probe from

S. enterica serovar typhimurium which hybridizes to DNA in *S. enterica* serovar blegdam enabled cloning of *hsd* genes believed to be the first representative of a fourth family of type I r-m systems (A.J.B.Campbell, pers. comm.).

The polymerase chain reaction (PCR) was another route considered for the detection of unrelated specificity systems. PCR has been used to amplify portions of the *gap*, *putP*, and *ompA* genes from a number of enteric bacteria for sequence comparisons (Lawrence *et al.*, 1991a, 1991b; Nelson *et al.*, 1991, Nelson and Selander, 1992). Milkman and Bridges (1990) amplified DNA from fifteen regions of the chromosomes of *E. coli* K-12 and some ECOR strains using PCR. By selecting oligonucleotide primers from portions of the *hsd* region, it might be possible to amplify the DNA in between. In order to detect all type I systems, it would be necessary to use primers derived from regions conserved in all members of all families of type I systems. The use of sequences flanking the *hsd* region as primers, such as those in the 5.3kb fragment of C, was rejected, as this would involve amplification of long segments of DNA. Instead, conserved sequences were sought within the *hsdR* and *hsdM* genes. A number of regions in the R and M polypeptides are conserved between members of different families. Using these as primers would require amplification of approximately 2.5-3.5kb of DNA.

Because identity at the amino acid level disguises mismatches at the nucleotide level due to silent substitutions, the nucleotide sequences of the *EcoK* and *EcoA* *hsdR* and *M* genes were first compared. The nucleotide sequence of *EcoE* was then included, as appreciable divergence is found even within the Ib family of enzymes, members of the plasmid-encoded type Ic family were excluded from the comparisons, as they would introduce even more variation, making identification of suitable primers more difficult.

The nucleotide sequences of conserved regions in the *hsdR* and *hsdM* genes of *EcoA*, *EcoE*, and *EcoK* were aligned and compared using the computer programs PILEUP and PRETTY (Devereux *et al.*, 1984). The aligned sequences were examined by eye for possible PCR primers, avoiding sequences with mismatches or

Figure 4.7: Sequence comparisons of conserved regions of the *hsdR* and *hsdM* genes of *EcoA*, *EcoE*, and *EcoK* highlight sequences that may be used as primers for the polymerase chain reaction (PCR). The nucleotide sequences were aligned using the University of Wisconsin Genetics Computer Group computer programs PILEUP and PRETTY (Devereux *et al.*, 1984). Only the sequences most suitable as PCR primers are shown. The amino acid sequence for each conserved region is given; where alternative amino acids are encoded, both are displayed. The nucleotide sequences of the coding strands are shown beneath the amino acids, with mismatches indicated by lower case letters. Conserved nucleotides in flanking codons were also included. A consensus primer is given beneath each region, with an 'X' placed at positions where all three sequences have a different nucleotide. At these degeneracies, a mixture of all four nucleotides could be used (Compton, 1990). An arrow indicates the direction of extension of the primer by the DNA polymerase. For the third *hsdM* primer, two arrows denote primers of different lengths, a 16mer with two mismatches or a 27mer with 5 mismatches.

runs of G's and C's at the 3' end (Innis and Gelfand, 1990). Examples of sequences which are potential primers are given in figure 4.7. The DNA sequence comparisons are shown, with mismatches indicated. A consensus primer is given under the sequences, indicating the direction of primer extension by the DNA polymerase. Three possible sequences were located within *hsdM*, and two within *hsdR*. All contained a number of mismatches, reflecting the high level of both inter- and intrafamilial divergence even within conserved regions of the R and M polypeptides. Such mismatches will reduce the specificity of the primer. Degeneracy of a nucleotide in all three sequences may be incorporated into the primer by including a mixture of all four nucleotides (Compton, 1990).

Published reports of the use of consensus primers to amplify DNA regions in diverse species have generally focused on genes for metabolic enzymes (Milkman and Bridges, 1990; Lawrence *et al.*, 1991a; Nelson *et al.*, 1991; Nelson and Selander, 1992), which may not be under evolutionary pressure for sequence diversity. Judging from the sequence divergence even between members of the same family of type I enzymes, it seems unlikely that primers may be designed that will amplify all *hsd* regions. However, although PCR will not provide conclusive results particularly concerning the absence of type I enzymes in natural isolates of *E. coli*, it may enable detection of some novel systems among probe-negative strains.

As it was not feasible to verify that the strains tested in the hybridization experiments corresponded to the original ECOR strain numbers, the possibility of mixing of strains could not be dismissed. However, for the majority of strains analysed, bacteria from two sources (usually Howard Ochman and Thomas Whittam) were tested. As the results were identical for strains from both sources, this was considered to be sufficient proof that no mix-ups had occurred.

Differences in recognition sequences between strains, associated with divergence in the sequences of the *hsdS* genes, will provide further clues to the evolution of type I r-m systems. Classification on the basis of specificity of some of the K-like systems detected in the ECOR strains is the subject of the next chapter.

CHAPTER FIVE

ANALYSIS OF SPECIFICITIES

OF SOME TYPE IA R-M SYSTEMS

IN MEMBERS OF THE ECOR COLLECTION

5.1 INTRODUCTION

Although all type I r-m systems perform analogous functions and have a similar genetic organisation, they have been separated to date into three discrete families on the basis of a number of criteria. Interchangeability of enzyme subunits, originally used to demonstrate the relatedness of *EcoK* to other family members (Boyer and Roulland-Dussoix, 1969; Glover and Colson, 1969; Bullas and Colson, 1975), is confined to members of the same family, as is cross-reactivity of antibodies raised against the R or M subunits (Murray *et al.*, 1982; Suri and Bickle, 1985; Fuller-Pace *et al.*, 1985). Hybridization studies reinforce this division into families, demonstrating the close similarity of *hsdR* and *hsdM* within, but not between, families (Murray *et al.*, 1982; Daniel *et al.*, 1988).

Sequence comparisons within and between families of such diverse enzymes may provide information of both a functional and evolutionary nature. Interfamilial comparisons may identify conserved motifs within enzyme subunits which are involved in common functions. Analysis of the predicted amino acid sequence of *EcoK* identified a motif common to all adenine methylases (Loenen *et al.*, 1987). Indeed, this short sequence is also found in the M subunits of members of the Ib and Ic families (Sharp *et al.*, 1992; Price *et al.*, 1989). Examination of the predicted amino acid sequence of the R polypeptides of type I systems highlighted the so-called "DEAD box" sequences, including an ATP-binding site, found in enzymes with helicase activity (Gorbalenya and Koonin, 1991). This may account for the ability of type I restriction enzymes to translocate DNA from the recognition sequence to a cleavage site up to several kilobases away (Murray *et al.*, submitted).

Comparison of the predicted amino acid sequence of the S polypeptide of *EcoK* with its relatives identified two extensive variable regions separated by sequences of near identity (Gough and Murray, 1983). Analogous regions, of approximately 150-180 amino acids, also occur in members of the Ib family (Kannan *et al.*, 1989), and in *EcoR124* (Gubler *et al.*, 1992). Each of these regions have been correlated with one of the two components of the DNA recognition sequence (Fuller-

Pace *et al.*, 1984; Nagaraja *et al.*, 1985b; Fuller-Pace and Murray, 1986; Gann *et al.*, 1987; Gubler *et al.*, 1992). It has been demonstrated that the specificity for the trinucleotide component is confined to the amino variable region (Cowan *et al.*, 1989).

When two S polypeptides, even within the same family, specify different recognition sequences, no similarity has been detected in their variable regions. In contrast, *EcoK* and *StySP*, which both recognize the trinucleotide component of the target sequence 5'AAC, show 90% identity throughout their amino variable regions (Fuller-Pace and Murray, 1986). The amino recognition domain of the *StySB* S polypeptide, which specifies recognition of the trinucleotide, 5'GAG, has 44% identity with the amino recognition domains of *EcoA* and *EcoE*, both of which also recognize 5'GAG (Cowan *et al.*, 1989), although the remainder of the polypeptide is dissimilar. Therefore, enzymes which have generally dissimilar amino acid sequence share a domain that specifies recognition of 5'GAG. The amino acids conserved between members of different families which recognize the same sequence should include those which interact specifically with the nucleotides of the recognition sequence. Variation within the target recognition domains may reflect the need of the S polypeptides to interact with their corresponding M polypeptides which are highly divergent between families.

Sequence comparisons between the S subunits of all type I families reveal repeats, evident only at the amino acid level, which represent the only sequence common to all type I S polypeptides (Kannan *et al.*, 1989). The same level of similarity is seen between S polypeptides of different families as within the repeats of a single polypeptide (Kannan *et al.*, 1989). If these repeats have functional significance, they should identify regions involved in functions common to all type I systems.

Regions of the S polypeptides which are conserved between members of a family may be involved in family-specific functions, such as subunit interaction. The central conserved region has been postulated to maintain the recognition domains the

appropriate distance apart for target recognition (Price *et al.*, 1989; Gubler and Bickle, 1991).

Inter- and intrafamilial comparisons also provide information concerning evolutionary relationships of *hsd* genes. Sharp *et al.* (1992) compared the nucleotide and predicted amino acid sequences of the *hsdM* genes between members of the Ia family, and between families. They concluded that all three families of enzymes are homologous and speculated that a number of evolutionary processes resulted in the levels of divergence seen between bacterial species within the Ia family, and between members of different families.

Identification of further members of known families will add to our database for comparative sequence analysis within and between enzyme families. Constraints to sequence variation within a family may be revealed. Although it has been postulated that frequency-dependent selection should favour divergence (Levin, 1986; 1988a; Sharp *et al.*, 1992), a level of similarity must be retained in order that family relationships are maintained.

This chapter describes the construction of λ libraries of the DNA of some of the ECOR strains which showed homology to an *hsd_K*-specific probe (Chapter 4). By cloning *hsdM* and *S* or at least *hsdS*, it was hoped that classification of the type I systems encoded by these ECOR strains on the basis of specificity would be possible. In this way, alternative systems with novel specificities could be identified. As a result, more clues to the evolution and diversity of type I r-m systems and their distribution throughout the ECOR collection would be revealed.

5.2 RESULTS

A Construction of λ Libraries

The K-specific probe, pRH1, hybridizes to a single *EcoRI* fragment of 11.2kb in *E.coli* K-12, which encompasses *hsdM* and *S*, and most of *hsdR* (Sain and Murray,

1980; see Chapter 2, figure 2.1 A). Additionally, the *mcrBC* genes are located in this *EcoRI* fragment (Ross *et al.*, 1987; Raleigh *et al.*, 1989). Efforts to clone the *hsd* genes of the K-like ECOR strains were concentrated on those strains, namely ECOR5, 12, 13, 23, 24, and 25, in which similarity to pRH1 and, in most cases, pRH2 was also confined to a single *EcoRI* fragment (see Chapter 4). ECOR14 was also included for further analysis. In this case, pRH1 hybridized to an *EcoRI* fragment of approximately 6kb, while pRH2 hybridized to a larger fragment of 11.7kb (Table 4.2), indicating the presence of an *EcoRI* somewhere in the intervening DNA.

For each of the strains, construction of a λ library of *EcoRI* fragments allowed cloning of the pRH1-specific fragment in the λ att⁻ *cI857 EcoRI* replacement vector, λ NM574 (Borck *et al.*, 1976). The phage libraries were packaged *in vitro* and recovered on the r_k⁻m_k⁺ host ED8654. Generally, 5x10³-1x10⁴ plaques were screened. Recombinant phages carrying the K-specific *EcoRI* fragment were detected by plaque hybridization using pRH1 as a probe.

To confirm that the DNA fragment cloned in λ NM574 was the K-specific fragment from the *EcoRI* genomic digest of each of the strains, the recombinant λ DNA was hybridized with pRH1 following Southern transfer of *EcoRI* digests. The DNA digests were also probed with pRH2 and pRH3 (see Chapter 2, figure 2.1 A) to determine the extent of the region in the cloned DNA fragment. The pRH3 plasmid consists of the majority of the *hsdS* gene of K-12, plus intergenic DNA extending 900bp downstream of *hsdS*, and into the beginning of *mcrB*.

The results of the probing are summarized in Table 5.1. All but one of the recombinant phages contained DNA that hybridized to the *hsd_k*-specific probe, pRH1. The sizes of the K-specific fragments corresponded to the sizes of the K-specific bands in Southern hybridization of the DNA of the ECOR strains. λ 574-14(b) was identified in the λ 574-14 library by both pRH2 and pRH3, although it showed no homology to pRH1.

Table 5.1: Determination of the Extent of the Similarity of the Cloned DNA to *hsd_K* and *mcrBC*.

Phage ^a	Hybridization to the Following Probe ^b		
	pRH1	pRH3	pRH2
λNM1048 ^c	11.2kb	11.2kb	11.2kb
λNM574 ^d	-	-	-
λNM574-5 ^e	11kb	11kb	-
λNM574-12	12kb	12kb	12kb
λNM574-13	10kb	10kb	10kb
λNM574-14(a)	6kb	-	-
(b)	-	11kb	11kb
λNM574-23	11kb	11kb	11kb
λNM574-24	12kb	12kb	12kb
λNM574-25	5.4kb	-	-

^a Preparation of phage DNA and Southern hybridizations were kindly carried out by Diane Ternent. Phage DNA was purified and digested with *EcoRI*, followed by separation through an agarose gel and transfer to a nylon filter.

^b The size of DNA fragments hybridizing are shown. "-" indicates no hybridization.

^c λNM1048 is *hsdM_K⁺S_K⁺ mcrBC⁺*.

^d λNM574 is the vector DNA.

^e Recombinant λ*hsd* phages are named according to the ECOR strain number of the bacterial DNA they contain.

^f Two different λ clones were isolated from the ECOR14 library by screening with pRH1, pRH2, and pRH3.

Phages λ 574-12, λ 574-13, λ 574-23, and λ 574-24 all carried DNA that hybridized to pRH1, pRH2 and pRH3 (Table 5.1). This suggested that *hsdM* and *S*, and perhaps all or most of *hsdR*, as well as *mcrBC* were located in the insert DNA cloned in the phages. The cloned DNA of λ 574-5, though lacking homology with pRH2, did hybridize to pRH1 and pRH3. This implies that at least some of the *hsdS* gene of ECOR5 is located in λ 574-5, as well as *hsdM*, and perhaps *hsdR*. The lack of hybridization to pRH2 is consistent with the absence of *mcrBC*; this correlates with the result obtained when probing the *EcoRI* digest of ECOR5. λ 574-25 failed to hybridize to either pRH2 or pRH3. This casts doubts on the presence of *hsdS* in the cloned fragment in λ 574-25 and implies the absence of *mcrBC*, suggesting that only *hsdM* and possibly *hsdR* were included. This would correlate with the small size of the *EcoRI* fragment in a genomic digest of ECOR25 DNA, which hybridized only to pRH1 and not pRH2 in Southern hybridizations (see Chapter 4, Table 4.2). Alternatively, it could indicate that the *hsdS* gene of ECOR25 is so dissimilar to the K-12 *hsdS* gene that no cross-hybridization was detected.

The DNA carried by the two λ 574-14 phages corresponded to what was anticipated. λ 574-14(a) carries the smaller, pRH1-specific fragment, while the larger, pRH2 and pRH3-specific fragment appears to be the DNA cloned in λ 574-14(b). It may be that λ 574-14(b) includes the *hsdS* gene of ECOR14 as well as *mcrBC*, as the DNA in pRH3 consists of the majority of the K-12 *hsdS* gene.

B Complementation analysis of the λ hsd phages.

The extent of the *hsd* region carried by the recombinant λ phages was determined by complementation analysis.

Phage-encoded M and S polypeptides can interact with a host-encoded R polypeptide of the same family to form an active restriction endonuclease *in vivo*. If the host lacks the genes to encode a functional methylase, its DNA will be unmodified and presumably vulnerable to restriction by the endonuclease. The result is a reduced plating efficiency of the phage on the host strain, most likely due to cell death before

a productive phage infection. This effect requires excess production of the R subunit in the host cell, encoded by a multicopy plasmid (Fuller-Pace *et al.*, 1985).

Such a complementation test was used to analyse mutants resulting in modification-deficient phenotypes of λ *hsd*_A and λ *hsd*_E phages (Fuller-Pace *et al.*, 1985). It was possible to localize the lesions in *hsdM* or *hsdS* by determining the efficiency of plating of these phages on an *hsd*-deleted host strain harbouring either an *hsdR*-containing plasmid, or an *hsdR*⁺*M*⁺ plasmid. A phage will show a reduced efficiency of plating if it provides the functional gene(s) missing from the plasmid. Therefore, a λ *hsdM*⁻*S*⁺ phage infecting a host harbouring a plasmid carrying *hsdR* and *hsdM* would show a reduced efficiency of plating when compared to a control strain because the phage-encoded S polypeptide would be complemented by the plasmid-encoded M polypeptide. However, λ *hsdM*⁺*S*⁻, or λ *hsdM*⁻*S*⁻ would not show a reduced efficiency of plating, due to the absence of a functional S polypeptide (Fuller-Pace *et al.*, 1985). Similar experiments have been used to localize mutations within the *EcoK* methylase genes (Kelleher *et al.*, 1991).

This complementation test is referred to as a "killing test" because of the effect on the host cell when an active restriction endonuclease is formed. Killing tests were used to assess the capability of phages believed to be carrying *hsd* genes from ECOR5, 12, 13, 14, 23, 24 and 25 to form active restriction complexes with plasmid-encoded *hsd*_K gene products.

The plasmid pBg3 encodes the *hsdR* and *hsdM* genes of *E. coli* K-12 (Sain and Murray, 1980). When an *r*_K⁻*m*_K⁻ strain (NM522) is transformed with pBg3, it may be used to assay the λ *hsd* phages isolated from the libraries of the K-like ECOR strains for the presence of an active *hsdS* gene. A phage encoding an *hsdS* gene whose product is able to interact with the plasmid-encoded R and M polypeptides, will be phenotypically Kill⁺, manifest as a lowered efficiency of plating when compared to a control strain. However, if an inactive *hsdS* gene product is produced, no complementation will occur and a normal infection will ensue. An *m*_K⁻ strain transformed with the plasmid pJK2, extends the assay to phage-encoded *hsdM* genes.

This plasmid carries only the *hsdR* gene of K-12 (Kelleher *et al.*, 1991). Thus, a phage showing a reduced e.o.p. on this transformant must encode functional *hsdM* and *hsdS* gene products.

The results of killing tests using 522(pBg3) and 522(pJK2) are summarized in Table 5.2. All of the phages, with the exception of λ 574-14 and λ 574-25, gave a Kill⁺ phenotype when infecting 522(pBg3). This indicates that these phages carry a functional *hsdS* gene. Three of these phages, λ 574-5, λ 574-12, and λ 574-24, also encode a functional M polypeptide, as evidenced by a Kill⁺ phenotype when infecting 522(pJK2). The Kill⁻ phenotypes of λ 574-13 and λ 574-23 on 522(pJK2) could indicate either that these phages did not encode active *hsdM* genes, or that the M polypeptide was not interacting with the plasmid-encoded *hsdR_K* gene product.

Both λ clones carrying *hsd_K*-like DNA of ECOR14 were phenotypically Kill⁻ on both 522(pBg3) and 522(pJK2). Hybridization of the DNA carried by the two phages using K-specific probes implied the presence of an *EcoRI* cleavage site in the region of *hsdS*. A cleavage site within *hsdS* would account for a Kill⁻ phenotype in both killing test strains, regardless of the activity of the phage-encoded *hsdM* gene.

Paradoxically, though infection of 522(pJK2) by λ 574-25 resulted in cell death, an active restriction complex was not produced upon infection of 522(pBg3). Although the phage must encode the ECOR25 *hsdM* and *hsdS* gene products, this *hsdS* gene product was unable to complement the plasmid-encoded R and M polypeptides. One possible explanation of this result is that while the M and S polypeptides of ECOR25 are able to form an active restriction complex with the R subunit of *EcoK*, the S polypeptide of ECOR25 is unable to interact with the M subunit of *EcoK* and cannot, in the context of the killing test where the M subunit of *EcoK* is in great excess over the ECOR25 M subunit, interact with the latter.

C Transfer of the Phage-encoded *hsd* Genes to the *E.coli* Chromosome.

Having established that the *hsdS* genes of ECOR5, 12, 13, 23 and 24 all encode a functional S polypeptide able to form an active restriction enzyme in the

TABLE 5.2: Determination of the Ability of the ECOR *hsd* Gene Products to Complement *hsd_K* Gene Products Using Killing Tests.

<u>PHAGE</u>	<u>KILL PHENOTYPE^b</u>	
	NM522/pJK2 (<i>hsdR</i> ⁺)	NM522/pBG3 (<i>hsdR</i> ⁺ <i>M</i> ⁺)
λNM1048 (<i>hsdM_K</i> ⁺ <i>S_K</i> ⁺)	+	+
λNM1049 (<i>hsdM_K</i> ⁺ <i>S_K</i> ⁻)	-	-
λ574-5	+	+
λ574-12	+	+
λ574-13	-	+
λ574-14(a)	-	-
(b)	-	-
λ574-23	-	+
λ574-24	+	+
λ574-25	+	-

^a The phages tested were those isolated from libraries of the ECOR strains.

^b The "Kill" phenotype is determined by infecting an *r_K*⁻*m_K*⁻ host (NM522) expressing the *EcoK hsdR* gene or *hsdR* and *M* on a plasmid (pBg3 or pJK2, respectively) with λ*hsd* phages. If the *hsd* gene products expressed by the phage complement the plasmid-encoded *hsd_K* gene products, the unmodified bacterial DNA will be cleaved by the resulting restriction enzyme before productive infection by the phage. This "Kill⁺" phenotype is manifest by a reduced e.o.p. of the phage, usually on the order of 10⁻³. "Kill⁻" indicates no complementation.

^c λNM1048 is a positive control.

^d λNM1049 is a negative control.

presence of the *EcoK* R and M subunits, the next step was to classify the type I systems on the basis of the specificities dictated by these S subunits. Simple genetic tests should allow the differentiation between systems that have a previously-identified specificity, and those that encode novel specificities. Further analysis would be required to determine the recognition sequences of the latter.

Integration of the λ *hsd* phages into the genome of an *hsdR⁺M⁺S⁻* *E.coli* strain by homologous recombination could result in the formation of an *hsd* region encoding an active restriction enzyme with the sequence specificity dictated by the phage-encoded *hsdS* gene. Two *E.coli* strains, NM496 and K803, were chosen as hosts. NM496 has most of the *hsdS* gene deleted, while K803 has an *hsdS* gene inactivated by a single point mutation. As λ 574 is *att⁻*, integration of the phage DNA may only occur by recombination within regions of homology. It was hoped that the *hsdM* gene could provide such a region for a crossover event. This relies on the assumption that the phage carries an *hsdM* gene of sufficient similarity that recombination may occur. In this way, transcription of the *hsdS* gene from the phage would be under the control of its natural promoter.

As the lysogens could be unstable, spontaneously cured cells were selected as survivors of incubation at 42°C. Although a proportion of phages will recircularize and leave the bacterial chromosome by the same recombination event in which they integrated, some may recombine in such a way as to leave behind the *hsdS* gene introduced by the phage. These bacteria will therefore have a stable *hsd* region with the specificity of the ECOR strain from whence the *hsdS* gene came.

Restriction-proficient cured cells were only isolated from K803 cells lysogenic for λ 574-12, and λ 574-24. All cells harbouring λ 574-5 or λ 574-23 prophages lost their restriction ability upon curing. Therefore, analysis of the other ECOR *hsd* genes were carried out at 30°C, using lysogens.

Initially, the lysogens or cured cells were infected with λ V.0 as a test for restriction. The plating efficiency of λ V.0 on both the ECOR12 and 24 cured

TABLE 5.3: Spot Tests Using Various Laboratory Phages to Detect Restriction by the λ hsd Lysogens.

<u>LYSOGEN^a</u>	<u>E.O.P. OF THE FOLLOWING PHAGE</u>				
	λ V.0	P1CmTc1.0	P3	ϕ 80	ϕ 82
K803(5) ^b	2×10^{-4}	nd ^c	nd	nd	nd
K803(12)	7×10^{-4}	nd	nd	nd	nd
K803(13)	1	1	0.5	1	1
NM496(13)	1	1	0.6	nd	0.16- 0.5
K803(23)	1	1×10^{-2}	4×10^{-2}	nd	nd
K803(24)	1×10^{-3}	nd	nd	nd	nd
NM469(24)	1×10^{-3}	nd	nd	nd	nd

^a Either K803 or NM496 were lysogenized by λ phages carrying the *hsdS* genes of the ECOR strains.

^b The number in brackets indicates the ECOR strain from which the *hsd* region carried by the phage originated.

^c nd=not determined.

lysogens was approximately 10^{-3} , and 10^{-4} on the ECOR5 lysogen (Table 5.3).

Neither the ECOR13 nor the ECOR23 lysogens restricted λ V.0 (Table 5.3). This may suggest that an active restriction complex is not being formed, though interaction of both the ECOR13 and ECOR23 S polypeptides with R and M from *EcoK* were evident in killing tests. Alternatively, λ may lack target sequences for the ECOR13 and 23 endonucleases, and thus escape restriction. To test the latter possibility, a number of other laboratory phages were tested for sensitivity to restriction by either of the lysogens.

Initially, P3 and the Dar^- P1 phage, P1CmTc1 were tested as they have larger genomes than λ and therefore are more likely to contain recognition sequences. The ECOR23 lysogen restricted both unmodified P1CmTc1 and P3 DNA (Table 5.3). Therefore, the *hsdS* gene from ECOR23 encodes an S polypeptide that recognizes sequences in the DNA of P1CmTc1 and P3, but not λ . This result is consistent with the S polypeptide of ECOR23 conferring a novel specificity, as all known type I r-m systems recognize sequences within the DNA of λ (Kessler and Manta, 1990)

Four λ 574-13 lysogens of NM496 (NM496(13)) and six of K803 (K803(13)) were tested for the ability to restrict P1CmTc1, P3, ϕ 80, and ϕ 82. All of these phages plated with an efficiency approaching one (Table 5.3).

The e.o.p. of ϕ 82 was slightly lowered on NM496(13); plating efficiencies of ϕ 82.0 varied from 0.16 to 0.5 among the four lysogens tested (Table 5.3). This could imply that ϕ 82 has a single target sequence in its genome for the restriction system of ECOR13; a single target in λ for a type I r-m system will result in an e.o.p. of 10^{-1} for λ V.0 upon infection of this strain (Gann *et al.*, 1987). However, the fact that ϕ 82 plated with an efficiency of one on all K803(13) lysogens tested argues against such a possibility.

λ phages modified by propagation on *E.coli* strains encoding type I r-m systems of different specificities were used to determine whether the specificities of

the ECOR5, 12 and 24 systems were the same as any of those previously identified. The e.o.p. of various appropriately modified λ phage was checked on the restriction-proficient lysogens (or cured cells).

Cells carrying restriction systems with the specificity of ECOR12 or ECOR24 restricted λ V.0, λ V.SB, and λ V.SQ. However, λ carrying either *EcoK* or *StySP* modification was able to plate on both strains with an efficiency of one (Table 5.4 A and B). The *StySP* r-m system recognizes a degenerate form of the *EcoK* recognition sequence; while the target sequence of *EcoK* is 5' AAC(N₆)GTGC, that of *StySP* is 5' AAC(N₆)GTRC, where R is either purine (Nagaraja *et al.*, 1985b). Thus, *StySP* will cleave DNA with unmodified *EcoK* target sequences, as well as DNA with the target including the tetrameric sequence 5'GTAC. Because both λ V.K and λ V.SP are protected from restriction by either the ECOR12 or ECOR24 systems, this suggests that both encode K-specific restriction enzymes (Table 5.4 A and B). To confirm this, λ phages recovered after propagation on K803(12) and K803(24) were tested for modification against *EcoK* and *StySP* restriction. Protection only against K-specific restriction was imparted by the *hsd*_{ECOR12} and *hsd*_{ECOR24}-specific methylases. It was concluded that both ECOR strains encode type I r-m systems with K-specificity.

K803(5) lysogens restricted λ V modified by propagation on strains encoding methylases with specificities of all known type I r-m systems other than *EcoR124/3* and *EcoDXXI* approximately 10³-fold (data not shown). This implies that ECOR5, like ECOR23, encodes a type I r-m system with a novel specificity for the Ia family.

The DNA of ECOR70, a non-Group A strain, was found to hybridize to pRH1 (Chapter 4, Table 4.1). Rather than cloning the *hsd* genes of this strain by making a library, the specificity of the type I r-m system was determined using the P1*darA* mutant, P1CmTc1, as an indicator of restriction. Like λ , P1CmTc1 evading restriction by the type I r-m system should be modified with the specificity of the strain. These modified phages may then be tested on laboratory strains encoding the spectrum of type I r-m systems.

TABLE 5.4A: ECOR12 Encodes a Type I R-M System with K-Specificity.

<u>STRAIN</u>	<u>E.O.P. OF THE FOLLOWING PHAGE</u>			
	$\lambda V.0$	$\lambda V.K^b$	$\lambda V.SP^b$	$\lambda V.K803(12)^b$
K803 ($r_K^- m_K^-$)	1	1	1	1
C600 ($r_K^+ m_K^+$)	7×10^{-4}	1	1	1.2
L4002 ($r_{SP}^+ m_{SP}^+$)	7×10^{-5}	3×10^{-3}	1.1	6×10^{-3}
K803(12) ^a	9×10^{-4}	1	1.1	1.4

TABLE 5.4B: ECOR24 Encodes a Type I R-M System with K- Specificity.

<u>STRAIN</u>	<u>E.O.P. OF THE FOLLOWING PHAGE</u>			
	$\lambda V.0$	$\lambda V.K^b$	$\lambda V.SP^b$	$\lambda V.K803(24)^b$
K803	1	1	1	1
C600	1×10^{-3}	1.1	1	1
L4002	1×10^{-4}	3×10^{-3}	1.4	1×10^{-2}
K803(24) ^a	1×10^{-3}	1.2	0.8	1

^a K803(12) and K803(24) are strains which have been cured of the λ prophage, though the ECOR *hsdS* gene remains on the chromosome. An inability of λcI to productively infect these strains indicates that they are no longer lysogens (data not shown).

^b Modified λV was obtained by making plate lysates on the relevant strains: C600 for K-modification, L4002 for SP, and K803 carrying the *hsdS* gene of ECOR12 or ECOR24.

Experiments were conducted using the same methods as described in Chapter 3, that is, using formation of lysogens as an indication of plating efficiency.

Table 5.5 (A) gives the results of infecting ECOR70 with unmodified P1Cm1 and P1CmTc1. The wild type P1 has an efficiency of plating of approximately 10^{-3} to 10^{-4} , while the e.o.p. of the P1*darA* phage is less than 10^{-6} . This difference in plating efficiency could be indicative of type I restriction. Heat induction of the P1CmTc1.ECOR70 lysogens gave lysates of modified phage. These phage were then tested for the ability to infect *E.coli* laboratory strains encoding type Ia r-m systems (Table 5.5 B). The plating efficiency of P1CmTc1.ECOR70 was very poor on strains encoding *EcoK* and *StySP*, while somewhat less so on *StySB*. The e.o.p. of the phage on a B-restricting strain (NM270), however, approached one. This implied that ECOR70 has a type I r-m system with B-specificity.

To confirm B-specificity, unmodified P1CmTc1, P1CmTc1.B and P1CmTc1.ECOR70 were plated on a laboratory *E.coli* strain encoding *EcoB* (NM270) and ECOR70 (Table 5.5 C). As was often the case when using phage P1 to screen wild type *E.coli* for type I r-m systems (Chapter 3), even phage modified by growth on ECOR70 had difficulty infecting this strain. Unmodified P1CmTc1 formed no lysogens when infecting ECOR70. P1CmTc1.B had an e.o.p. of 10^{-4} on ECOR70 and P1CmTc1.ECOR70 showed a slightly improved plating efficiency of 4×10^{-3} . Such results could not definitively prove *hsd*_{ECOR70} has B-specificity. However, infection of NM270 by P1CmTc1.ECOR70 gave results consistent with ECOR70 encoding a B-specificity system; the e.o.p. of P1CmTc1.ECOR70 was greater than one on NM270.

5.3 DISCUSSION

A Functional Analysis of ECOR Type Ia R-M Systems.

Among the nine ECOR strains whose DNA hybridized to an *hsd*_K- specific probe, two strains were shown to encode systems with K-specificity, one with B-

TABLE 5.5: Determination of the Specificity of the ECOR70 Type I R-M System Using the *darA*⁻ Phage, P1CmTc1.

A) Efficiency of Plating of Unmodified Wild Type P1 (P1Cm1.0) and P1CmTc1 (P1CmTc1.0).

<u>STRAIN</u>	<u>E.O.P. OF THE FOLLOWING PHAGES</u>	
	P1Cm1.0	P1CmTc1.0
NM555 ^a	1	1
ECOR70	5x10 ⁻³	< 10 ⁻⁶

B) Efficiency of Plating of P1CmTc1.ECOR70 on Laboratory *E. coli* Strains Encoding Type Ia R-M Systems.

<u>STRAIN</u>	<u>SPECIFICITY</u>	<u>E.O.P. of P1CmTc1.ECOR70^b</u>
NM555 ^a	0	1
C600	K	1.4x10 ⁻³
NM270	B	0.6
NM515	D	3x10 ⁻⁵
L4001	SB	7x10 ⁻²
L4002	SP	4x10 ⁻⁴

TABLE 5.5 (cont'd)

C) Confirmation of B-Specificity of the Type I R-M System of ECOR70.

<u>STRAIN</u>	<u>E.O.P. OF THE FOLLOWING PHAGES</u>		
	P1CmTc1.0	P1CmTc1.ECOR70	P1CmTc1.B ^c
NM555 ^a	1	1	1
ECOR70	$< 10^{-6}$	4×10^{-3}	1×10^{-4}
NM270 (r _B ⁺ m _B ⁺)	4×10^{-2}	4	5×10^{-1}

^a NM555 is a r⁻m⁻ control.

^b Lysogens of P1CmTc1 on ECOR70 were heat-induced to obtain a lysate of modified phage.

^c Induction of P1CmTc1 lysogens of NM270 produced *EcoB*-modified phage.

Table 5.6: Summary of the Analysis of the *λ*hsd Phages.

Phage	Hybridization to the Following Probe			Killing Test Results		Lysogens: e.o.p. of:		Specificity
	pRH1	pRH3	pRH2	pJK2 (<i>hsdR</i> ⁺)	pBg3 (<i>hsdR</i> ⁺ <i>M</i> ⁺)	<i>λ</i> V.0	Other Phages	
<i>λ</i> 574-5	+	+	-	+	+	2x10 ⁻⁴	nd ^a	novel
<i>λ</i> 574-12	+	+	+	+	+	7x10 ⁻⁴	e.o.p. = 1 for <i>λ</i> V.K, <i>λ</i> V.SP	K
<i>λ</i> 574-13	+	+	+	-	+	1	e.o.p. = 1 for P1CmTc1, P3, <i>φ</i> 80, <i>φ</i> 82	nc ^b
<i>λ</i> 574-14 (a)	+	-	-	-	-	nl ^c	nl	nc
(b)	-	+	+	-	-	nl	nl	nc

λ 574-23	+	+	+	-	+	1	e.o.p. = 1×10^{-2} for P1CmTc1; 4×10^{-2} for P3	novel
λ 574-24	+	+	+	+	+	1×10^{-3}	e.o.p. = 1 for λ V.K, λ V.SP	
λ 574-25	+	-	-	+	-	nl	nl	K nc

^a "nd" indicates not determined. λ modified against restriction by almost all type I r-m systems showed e.o.p.'s of 10^{-3} - 10^{-4} .

^b "nc" means the specificity has not been concluded.

^c "nl" indicates that no lysogens were isolated.

specificity, and two with novel specificities. The analysis of the others is incomplete. A summary of the results of hybridization analysis of the λ *hsd* phages, killing tests, and attempts to classify the specificities of some of the ECOR type I systems is given in Table 5.6.

Both ECOR12 and 24 were found to encode systems indistinguishable from *EcoK*. Unmodified λ V was restricted by lysogens expressing type I systems with either ECOR12 or 24 specificity, while λ V.K and λ V.SP were protected. Phage recovered from infection of lysogens expressing either the ECOR12 or 24 type I system were protected only against *EcoK* restriction. It is interesting to note that these were the only two of five systems tested in which cells cured of the lysogenic λ *hsd* phage retained the ECOR *hsdS* gene. In order for this to occur, recombination to recircularize the DNA of the λ phage must take place downstream of the ECOR *hsdS* gene. This method was successfully used to transfer *hsd_B* and *hsd_D* genes to the bacterial chromosome (Gough and Murray, unpublished observation). Downstream homology between the K-12 chromosome and the bacterial DNA within the phage, such as the intergenic region between *hsdS* and *mcrBC*, *mcrBC* itself, or more distant flanking DNA common to both K-12 and ECOR12 or ECOR24 could provide sites for homologous recombination and excision of the phage DNA. The fact that it was not possible to cure any of the other λ *hsd* phages without losing the corresponding ECOR *hsd* genes suggests poor homology in the DNA flanking the ECOR *hsd* regions.

Both ECOR5 and ECOR23 were found to encode type Ia r-m systems with novel specificities. Although the DNA of λ V contains recognition sequences for the ECOR5 system, λ V modified by any of the members of the Ia and Ib families, as well as *EcoR124*, was sensitive to restriction by the ECOR5 endonuclease. There were no target sequences for the ECOR23 type I system in λ V, as evidenced by an e.o.p. of one on the lysogen expressing a system with ECOR23 specificity. However, both P1CmTc1 and P3 phages were restricted by this r-m system. This is the first example of a type I system which does not restrict λ . DNA of the ECOR strains carrying *hsd* genes with K or B specificity hybridized to an *mcrBC*-specific

probe (see Chapter 4), in concordance with previously published findings. In addition, the *mcrBC* probe also hybridized to the DNA of ECOR23, a strain encoding a type I system of novel specificity. Therefore, association with *mcrBC* is not limited to *hsd_K* and *hsd_B* in *E.coli*. As discussed in Chapter 4, the *mcrBC* genes have not been detected in strains specifying *EcoD*, *StySP* and *SB*, and type Ib systems. Further analysis of this region in ECOR23 may aid the understanding of how *mcrBC* came to be associated with certain *hsd* regions.

The specificities of type I r-m systems may be deduced *in vivo*, obviating the need for enzyme purification. This method, first described by Gann and his colleagues (1987), was used to determine the specificities of the recombinant r-m system, *StySJ* (Gann *et al.*, 1987) as well as that of *EcoE* (Cowan *et al.*, 1989) and *CfrA* (Kannan *et al.*, 1989). The method made use of the fact that M13 phage plates with an efficiency of one on an F' derivative carrying the *hsd* genes of the above systems, indicating that there were no target sequences in M13. M13 phages with cloned inserts of known sequence were then screened; the introduction of a target site in the fragment would result in a reduced efficiency of plating. A computer-aided search comparing M13 clones carrying targets was then used to determine the recognition sequence, assuming that it would have the typical trimeric and tetra- or pentameric components, separated by a non-specific spacer.

Such a method could be used to determine the recognition sequences of ECOR5 and 23, if transferred to the chromosome of an F' *E.coli* strain. As this method relies on the M13 lacking sites for restriction enzyme, a similar procedure using λ carrying inserts of known sequence could be used to determine the specificity of the ECOR23 system, as λ lacks target sites for this system. Stable transfer of the *hsd* genes of ECOR23 to the host chromosome would obviate the need to consider immunity problems.

It was not possible to draw any conclusions about the specificities of the type I r-m systems of ECOR13, 14 and 25. Because the *hsdS* gene product of ECOR13 was shown to complement the *Eco_K* R and M polypeptides in killing tests (see Table

5.6), λ lysogens carrying the ECOR13 *hsdS* gene should produce a functional restriction enzyme. However, a number of lysogens of both K803 and NM496 did not restrict unmodified λ , P1CmTc1, P3, ϕ 80, or ϕ 82 (Table 5.6). It seems unlikely that all these phages would lack the recognition sequence for the ECOR13 r-m system. A more plausible explanation may be that the integration of the λ DNA into the chromosome resulted in a recombinant *hsd* region unable to produce a functional restriction enzyme. The *hsdM* gene product of ECOR13 did not complement the plasmid-encoded *EcoK* R polypeptide in a killing test (Table 5.6). Additionally, results from testing the e.o.p. of wild type P1 and the Dar^- P1CmTc1 on ECOR13 gave no evidence of an active restriction enzyme (data not shown). If integration of the λ *hsd* phage occurred within the *hsdM* gene, it might result in the formation of an *hsd* region with an inactive *hsdM* gene.

Results of probing λ 574-25 indicated that the cloned DNA may not include an *hsdS* gene, as pRH1 but not pRH3 showed hybridization. However, in killing tests the ECOR25 M and S polypeptides together were able to form an active restriction complex with a plasmid-encoded R polypeptide, demonstrating that the ECOR25 *hsdS* and *M* genes carried by the phage are active. The ECOR25 S subunit alone was not able to interact with plasmid-encoded *EcoK* R and M subunits. Taken together, these results suggest that the *hsdS* gene of ECOR25 is sufficiently dissimilar to the *EcoK* *hsdS* gene that the S polypeptide of ECOR25 is unable to interact with M from *EcoK*. The M subunit of ECOR25, though divergent enough to interact with its corresponding S polypeptide, has retained the ability to interact with the *EcoK* R subunit. It may be that the ECOR25 *hsd* genes approach the lower limit to the degree of similarity necessary to participate in the family-specific interaction of subunits.

Two different λ clones were recovered from an *EcoRI* library of ECOR14, each with a different pattern of hybridization when probed with pRH1, pRH2 and pRH3. The hybridization results (Table 5.6) suggested the presence of an *EcoRI* site within the ECOR14 *hsdS* gene; pRH3, which in K-12 hybridizes mostly to *hsdS* and DNA between *hsdS* and *mcrB*, and pRH1 hybridized to the DNA of different clones. Consistent with this, neither of the λ clones was able to complement the *hsdR* and *M*

genes of pBg3 in killing tests (Table 5.6). Therefore, in order to clone the ECOR14 *hsd* region, it will be necessary to make a library from *Sau3A* or *EcoRI* partial digests. The *hsd* region of ECOR11 must also include an *EcoRI* site as pRH1 hybridized to two fragments in an *EcoRI* digest of the ECOR11 chromosome (Chapter 4, Table 4.2). Thus, cloning of the *hsd* region of ECOR11 will also rely on libraries of partial digests.

Construction of an *EcoRI* library of ECOR70 DNA was not considered to be high priority, as the specificity of its type I r-m system had been established previously. Although P1CmTc1 was able to lysogenize ECOR70 only very poorly, phage recovered after heat induction of these lysogens had *EcoB* modification. However, confirmation of this B-specificity will require cloning of the ECOR70 *hsd* region.

Sequence comparisons of these new systems with currently known ones may provide information concerning structure and function of type I r-m systems. Including the novel *hsd* genes in comparisons will highlight regions involved in common functions, such as subunit interactions. The ECOR25 *hsd* sequence may prove particularly informative in this respect, as its S polypeptide appeared not to interact with the M polypeptide of *EcoK* in killing tests. This may narrow down the sequence identity required for this interaction.

Comparison of a number of S polypeptides recognizing common target sequence components may identify amino acids that are directly involved in DNA binding. Candidate amino acids could then be tested by site-directed mutagenesis. Comparison of the amino recognition domains of *EcoA* and *EcoE* with *StySB*, all of which recognise the trinucleotide component 5'GAG, has already been documented (Cowan *et al.*, 1989). Comparison of systems with the same specificity should also disclose any amino acid changes that have occurred during evolution. A change may imply that a certain amount of flexibility is allowed at this residue without loss of function.

Figure 5.1: Classification according to specificity of the *EcoK*-like type I r-m systems among members of the ECOR collection. The specificities are superimposed on the phylogenetic tree of the ECOR strains showing the results of the DNA hybridization experiments of Chapter 4 (figure 4.5).

Key to symbols:

* = K-specificity

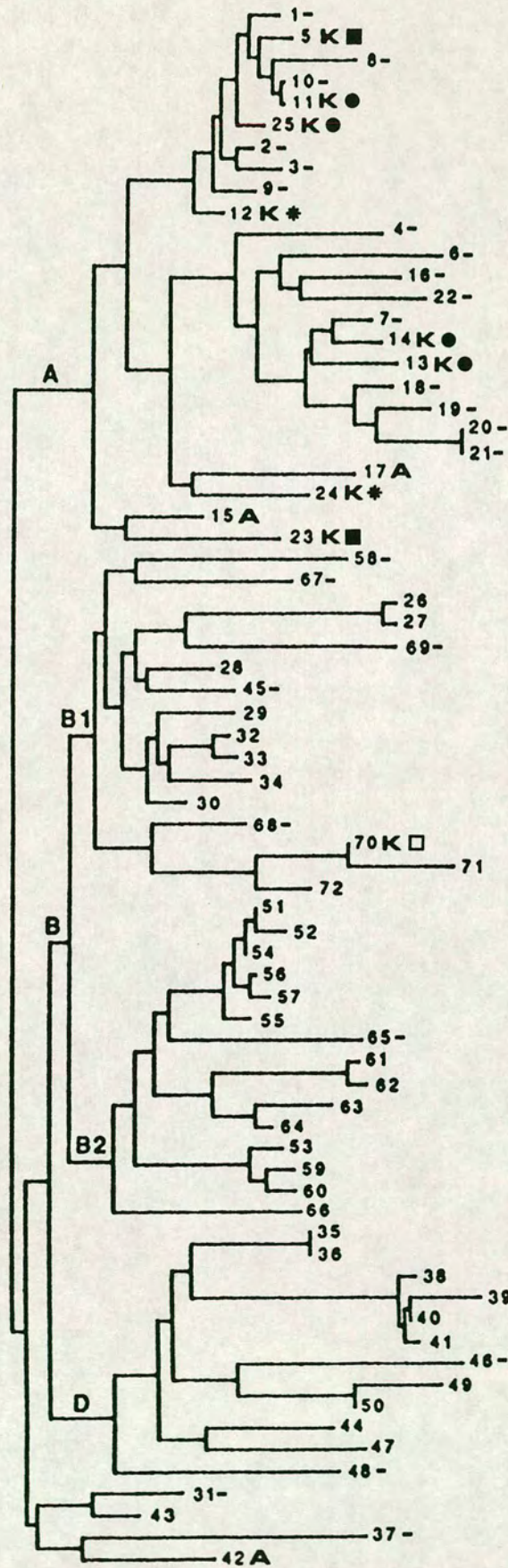
□ = B-specificity

■ = specificity novel to the type Ia family

● = classification of specificity not yet determined.

Figure 5.1

Distance
0 0.10



B Evolution of Type I R-M Systems in *E.coli*.

The distribution of specificities within the ECOR collection (shown in figure 5.1) supports the finding that classification based on type I r-m systems does not agree with that based on polymorphisms in metabolic enzymes (Chapter 4). A cluster of closely-related strains in Group A of the ECOR collection that would include K-12 (R.K.Selander, in Biserčić *et al.*, 1990) contains another strain with K-specificity, ECOR12, as well as a strain with a novel specificity, ECOR5. Two further strains within this cluster, ECOR11 and 25, encode systems with as yet undetermined specificities. It may be the case that further variation in specificity exists.

Obviously, determination of the specificities of the K-like systems must precede a definitive understanding. However, the pattern already evident, involving the occurrence of type I systems of alternative specificities and even alternative families (Chapter 4) among strains deemed to be closely-related on the basis of enzyme electrophoresis, suggests evolutionary pressure for variation of specificity. This is consistent with the theory of Levin (1986; 1988a), that bacteriophages will exert frequency-dependent selection on their hosts, so that type I r-m systems with rare specificities will be favoured. As a particular specificity becomes more common, it will confer less of an advantage to the host as more phages will acquire modification with this specificity. The bacteria in a population encoding this specificity system will decrease in frequency until they approach the point of extinction, at which point this, now rare, specificity will once again provide an advantage. This should ensure that a particular specificity is never lost from the population, and should also favour divergence resulting in the formation of new specificities.

Frequency-dependent selection is expected to act primarily on the *hsdS* gene, as S dictates the specificity of the enzyme. The capacity of the *hsdS* gene to evolve to recognize new sequences has been demonstrated in the laboratory. Recombination

within conserved regions of *hsdS* genes may reassort the recognition domains of the S subunit, producing systems with recombinant and hence novel specificities (Fuller-Pace *et al.*, 1984; Nagaraja *et al.*, 1985b; Fuller-Pace and Murray, 1986; Gann *et al.*, 1987; Gubler *et al.*, 1992). New domains could be acquired by horizontal transfer. Sequence specificity may also be changed by altering the length of the conserved region between two recognition domains (Price *et al.*, 1989; Gubler and Bickle, 1991). This could provide a further mechanism for the generation of diversity. No information exists on the effects of changing the amino acid sequence within a domain.

Within the type Ia family, the number of synonymous (nucleotide substitutions that do not alter amino acid sequence) and non-synonymous (substitutions altering amino acid sequence) substitutions were calculated for the *hsdM* genes of *EcoB*, *EcoK*, *StySB* and *StySP* (Sharp *et al.*, 1992). The extent of interspecific divergence of *hsdM* suggested that these genes have diverged from a common ancestor of *E. coli* and *Salmonella*. In contrast, intraspecific divergence of the *hsdM* genes of *EcoK* and *EcoB* was found to be exceptionally high, both at synonymous and non-synonymous sites when compared to intraspecific divergence in 28 other genes for which sequence data is available for *E. coli* K-12 and B (Sharp *et al.*, 1992). This was attributed to Levin's frequency-dependent selection. The high level of non-synonymous changes suggests that selection directly affects *hsdM*, rather than being an indirect effect from selection for divergence in *hsdS*.

High intraspecific diversity has been reported in other systems in which differentiation between "self" and "non-self" is the central role. The major histocompatibility complex (MHC) of mammals binds processed proteins and presents them to T-lymphocytes; the MHC proteins provide self-context to enable the T-cells to recognize antigens as foreign. Comparison of allelic sequences within a species shows that the ratio of replacement (non-synonymous) to silent (synonymous) substitutions is high, suggesting strong selection for divergence (Lundberg and McDevitt, 1992). However, comparisons between species, for example humans and chimpanzees, and rats and mice has shown that a large part of the MHC

polymorphism predates speciation and has been maintained by natural selection (Figuroa *et al.*, 1988; Lawlor, *et al.*, 1988; Lundberg and McDevitt, 1992). A similar mechanism for distinguishing self and non-self exists in many families of plants. The self-incompatibility locus allows recognition and rejection of self-pollen. Within a species of the *Solanaceae* family, amino acid identity may be as low as 40% (Ioerger *et al.*, 1990). Additionally, some alleles appear to be more similar to alleles in other species than those in the same species. Ioerger and his colleagues (1990) hypothesized that some of the polymorphism at the self-incompatibility locus predated divergence of these species and has been maintained by natural selection. Strictly neutral evolution would not have maintained these polymorphisms over such long periods of time, and Additionally, new mutations should be favoured as rare alleles will have an advantage. There is less chance of pollen bearing the same allele landing on a stigma with the same allele (S.Wright, in Ioerger *et al.*, 1990).

Sharp *et al.* (1992) pointed out the parallels between these systems and the type Ia family of r-m systems, and hypothesized that the K-like systems were subject to similar frequency-dependent selection, maintaining ancient alleles and favouring new, rare specificities arising by recombination.

Interfamilial comparisons of *E.coli hsdM* genes revealed a far higher divergence between *EcoK* and *EcoA* than seen within the type Ia family, even in comparisons between *E.coli* and *Salmonella* (Sharp *et al.*, 1992). Horizontal transfer was invoked to explain the presence of these "pseudodoalleles" in *E.coli*.

Horizontal transfer may also be responsible for the high level of divergence seen in comparisons of the *hsdR* and *M* genes of *EcoA* and *EcoE*, members of the type Ib family (Murray *et al.*, submitted). The estimated number of nucleotide substitutions per synonymous site again exceeded the divergence of *E.coli* and *Salmonella*. The *EcoA* and *EcoE* gene sequences are thus dissimilar enough to have originated from different species.

It may be that the type Ia family within *E.coli* also includes systems that have

been acquired by horizontal transfer. Sequence comparisons of the K-like *hsd* genes detected among members of the ECOR collection with known Ia-family *hsd* genes should reveal any evidence of such a phenomenon; ECOR25 could be a candidate. The fact that the ECOR25 *hsd* genes were detected using a K-specific probe classifies the system as type Ia. However, the S polypeptide of ECOR25 appeared unable to interact with the *EcoK* M polypeptide, suggesting a high level of divergence. Although sequence comparisons of *hsd_A* and *hsd_E* indicate extensive divergence, their subunits are still able to interact in complementation tests (Fuller-Pace *et al.*, 1985).

Raleigh and her colleagues (Dila *et al.*, 1990; Raleigh, 1992) noticed that the variable regions in the *hsdS* genes of *EcoK*, have a lower G+C content than the adjacent conserved regions and in fact lower than the G+C composition of the *E. coli* genome in general. These low G+C regions (in many cases less than 40% G+C) are evident in the *hsdS* genes of all known type Ia and Ib r-m systems, and perhaps the type Ic systems additionally (P. Thorpe, pers. comm.). Such a base composition is more typical of distantly-related enteric species such as *Proteus* and *Providentia* (Dila *et al.*, 1990; Raleigh, 1992), implying that the sequences encoding the recognition domains of the type I systems have been acquired by lateral transfer from other species. It has been hypothesized that insertion of such "cassettes" into *hsdS* genes could alter the DNA sequences which the enzyme recognizes (Dila *et al.*, 1990; Raleigh, 1992).

There are examples in various bacterial species of recombination of short regions of DNA resulting in the acquisition of advantageous genes, or in the variation of existing genes (e.g. Spratt *et al.*, 1989; Smith *et al.*, 1991; Smith, 1992; Spratt *et al.*, 1992) The *mcrBC* genes of *E. coli* have so far only been found in association with certain type Ia *hsd* regions. Raleigh and her colleagues have proposed that the *mcr* genes, like *hsdS*, compose a cassette which has been inserted into the genomes of certain *E. coli* strains. The low G+C content, in conjunction with DNA sequences resembling insertion elements are consistent with lateral transfer of this region from a distantly-related species (Dila *et al.*, 1990; Raleigh, 1992). Variation in *mcrBC* may exist; Raleigh *et al.* (1989) reported that the *mcrBC* gene products of *E. coli* B

appear to have a different activity level or sequence specificity from that found in K-12.

A number of Gram-negative pathogenic bacteria, including *Haemophilus influenzae*, *Neisseria meningitidis*, and certain strains of *E.coli* produce polysaccharide capsules. These capsules are highly variable; *E.coli*, for example, produces more than 70 chemically and serologically distinct capsules, or K antigens, which are divided into group I and group II based on a number of criteria (Jann and Jann, 1989). The capsules are important virulence factors, permitting evasion of the host's immune response. The group II capsule of *E.coli* are very similar to those of *H.influenzae*, whether due to a common origin (Boulnois and Jann, 1989), or convergence to a common method of altering surface antigenicity (Kroll and Moxon, 1990).

Conversion of capsular types resulting in a change in serotype may occur by a cassette mechanism. In both *E.coli* and *H.influenzae*, the chromosomal loci involved in the production of different capsular polysaccharides have a common organization independent of serotype; a central serotype-specific DNA segment is flanked by DNA sequences involved in conserved aspects of capsular biosynthesis (Boulnois and Jann, 1989; Kroll *et al.*, 1989), though the boundaries between the central region (region 2) of the *E.coli* group II K antigens have not been unequivocally defined. For *H.influenzae*, the conversion of one capsular type to another has been demonstrated in the laboratory (Kroll *et al.*, 1989). A strain with serotype b may be transformed to the other five serotypes, producing recombinant capsule (*cap*) loci in which the only change is within the central region. This presumably is also the case for the *E.coli* group II (*kps*) locus. Though conversion of one antigen type to another has not been observed, only one form of antigen is produced by a strain at any time (Boulnois and Jann, 1989).

Evidence of horizontal transfer of these central regions, resulting in a change in serotype, has been reported for *H.influenzae* (Kroll *et al.*, 1989; Kroll and Moxon, 1990). Multilocus enzyme electrophoresis separates *H.influenzae* into two divisions,

and hybridization studies of the *cap* region supports this (Musser *et al.*, 1988). Strains of different serotypes remain clonally distinct (Musser *et al.*, 1988). Ninety-nine percent of strains with b-serotype are found in division I, and the remainder in division II (Musser *et al.*, 1988). While division I and II genes in the *cap* loci not involved in specifying serotype b-specific functions are believed to have diverged by genetic drift from a common ancestor, genes conferring the type-specificity of the capsule show no such divergence. Kroll and Moxon (1990) reported that in a partial amino acid sequence of the *cap* loci of one homologous pair, there were 95/795 differences (12%) in the non-serotype specific regions, but no differences (0/250 amino acids) in the central region. This suggests recent horizontal transfer between the two divisions, in spite of the fact that the chromosomal location of *cap* in the two divisions appears to differ (Kroll and Moxon, 1990). Thus, the central region of the *cap* locus may be thought of as a cassette, enabling *H.influenzae* to alter its serotype.

A similar mechanism of altering capsular antigenicity has been hypothesized for *E.coli* (Boulnois and Jann, 1989), even though *E.coli* is not naturally transformable whereas *H.influenzae* is. In their model, Boulnois and Jann suggest that the whole *kps* region could be acquired, or just region 2 as a cassette. Recombination within region 2 could also generate diversity.

Lateral transfer of serotype-specific capsular biosynthesis genes has also been reported in Gram-positive bacterial species. More than 80 different polysaccharide types have been found in *Streptococcus pneumoniae*. Coffey *et al.* (1991) found strong evidence that genes determining one capsule type can replace those determining another capsule type. A strain was encountered which expressed a serotype 19 capsule but was otherwise indistinguishable from a serotype 23F clone on the basis of MLEE analysis of 14 enzyme-encoding loci, antimicrobial resistance pattern, and fingerprint patterns of the genes encoding PBP1A, 2B, and 2X. Coffey and colleagues (1991) hypothesized that horizontal transfer and recombination of the genes determining the expression of serotype 19 capsule had occurred.

Further examples of putative horizontal transfer events of cassettes of DNA

to generate diversity of surface antigens were discussed in Chapter 4. These included the *pap/prs* gene clusters in *E.coli*, encoding the P-pilus, and the *rfb* cluster in *Salmonella enterica*, which encodes the highly variable O-antigen.

Lateral transfer and recombination has also been reported in the genes encoding protein subunits of flagella in *Salmonella* (Smith *et al.*, 1990). In *Salmonella* strains, antigenic diversity of these subunits (flagellins) is seen at two levels. Firstly, two genes encode structurally-distinct flagellins, phase 1 (*fliC*) and phase 2 (*fliB*) (Iino *et al.*, 1988). Secondly, the central, antigen-determining part of the flagellin protein is highly polymorphic; for *fliC*, this polymorphism is believed to be due to recombination as the phylogeny of *fliC* does not correspond to the phylogeny of the chromosome in general as determined by MLEE (Smith *et al.*, 1990). The sequences of the central portion of *fliC* for several serovars of *Salmonella* have been compared (Smith *et al.*, 1990). Within a serovar, the central region of the *fliC* gene is identical (Smith *et al.*, 1990; Smith and Selander, 1990). *S.enterica* serovars heidelberg and typhimurium LT2 are closely related on the basis of MLEE (Beltran *et al.*, 1988), yet differ in the central regions of their *fliC* genes by 19%. The central region of the *fliC* allele of *S.enterica* serovar muenchen proved to differ from those of heidelberg and typhimurium LT2 by 50%, though this strain is also closely related to both strains. Yet the *fliC* allele of serovar muenchen was found to differ from that of *S.enterica* serovar typhi by only 1.3%, though the chromosomal genotypes of these strains are very dissimilar. Thus, lateral transfer and recombination were invoked to explain the variation within the central region of the phase 1 flagellins of these four strains (Smith *et al.*, 1990). Pressure for the expression of novel combinations of surface antigens could result in the selection of strains in which lateral transfer of the central region of the *fliC* gene as a cassette from distantly-related strains has occurred.

Alternative restriction systems in *Streptococcus pneumoniae* recognize the sequence 5'GATC, but react differently according to the methylation state of the sequence. *DpnI* cleaves DNA containing the methylated sequence, while *DpnII* cleaves the unmodified sequence and it also comprises a methylase which will modify

the DNA of the cell. The genes encoding these systems are unrelated, and only one of the two systems is found in a particular cell. Hybridization studies indicate that the genes responsible for one restriction phenotype are not detected in cells of the opposite phenotype (Lacks *et al.*, 1986). Lacks and his colleagues (1986) hypothesized that the genes encoding the alternative restriction systems (including a methylase gene in the case of *DpnII*) are genetic cassettes which insert into identical sites in the Pneumococcal chromosome, making use of the homology of adjacent regions for recombination. Upon infection of a susceptible cell by a virus, lysis would result in release of the gene cassette which could be acquired by another cell via transformation. Interestingly, these alternative restriction systems are only active on viral DNA; sequences acquired by transformation are not affected (Lacks and Greenberg, 1975). It would seem that such a mechanism is designed to protect a population of bacteria against phage infection, ensuring that some bacteria survive an onslaught of virus carrying a particular restriction phenotype (Lacks *et al.*, 1986). An analogy may be drawn between the *Dpn* restriction systems of *Pneumococcus* and the type I r-m systems of *Enterobacteriaceae*. The presence of type I systems with alternative specificities within an *E.coli* population should ensure the bacteriophage with DNA carrying a particular modification are not protected against cleavage upon infection of all hosts.

Although sequence comparisons have provided convincing evidence of recombination of DNA segments on the order of a few hundreds or thousands of base pairs between *E.coli* strains (Dykhuizen and Green, 1986; DuBose *et al.*, 1988; Stoltzfus *et al.*, 1988; Milkman and Bridges, 1991; Biserčić *et al.*, 1991; Marklund *et al.*, 1992), evidence of interspecific recombination is limited. Nelson and colleagues (Nelson *et al.*, 1991; Nelson and Selander, 1992) failed to identify any recombination events between *Salmonella* and *E.coli* in the *gapA* and *putP* loci. However, they did find evidence that the highly divergent *E.coli gnd* locus includes sequences recruited from other species of bacteria, such as *Klebsiella* (Nelson and Selander, 1992). Sharp *et al.* (1992) found evidence of horizontal transfer of a short sequence of the *hsdM* gene between *S.enterica* serovar potsdam and *E.coli*. Recombination between the *gnd* genes of *E.coli* and *S.enterica* serovar typhimurium

LT2 has also been reported, based on sequences with a high degree of similarity shared between LT2 and certain *E.coli* clones (Barcak and Wolf, 1988; Biserčić *et al.*, 1991).

The paucity of evidence of interspecific recombination may simply reflect the fact that few studies have included members of both *E.coli* and *Salmonella*. However, the clonal nature of *E.coli* argues against extensive recombination, as does the fact that enteric species form discrete taxonomic units, distinguishable both by genotyping and by biotyping (Ochman and Wilson, 1987). A potential mechanism for limiting stable transfer of DNA between *E.coli* and *S.enterica* serovar typhimurium has been described (Rayssiguier *et al.*, 1989). The requirement for sequence homology for genetic recombination is relaxed in *mutH*, *L* or *S* mutants, suggesting that the methyl-directed mismatch repair system plays a role in preventing extensive interspecific recombination.

The hypothesis that the recognition domains of the S polypeptides of type I r-m systems have been acquired from distantly-related species, or that the presence of *hsd_A* and *hsd_E* genes in *E.coli* is attributable to horizontal transfer, may thus seem difficult to reconcile. It could be that the predecessors of the type Ib family were encoded by the common ancestor of *E.coli* and *Salmonella*; the recent finding of an A-like type I system in *S.enterica* serovar kaduna (D.Ternent, pers. comm.) would be consistent with this theory. Whether the high level of divergence seen in members of different families, or even between *EcoA* and *EcoE* could be maintained intraspecifically over such a long divergence time is not known, although it is believed to be unlikely (Sharp *et al.*, 1992).

The picture emerging, supported by the results of screening the ECOR collection for type I r-m systems of two families, is one of an array of evolutionary forces acting on the *hsd* locus to promote diversity. These include frequency-dependent selection on more than one *hsd* gene, as well as intra- and interspecific recombination. As a result, specificities would be less likely to be lost from the population, and different specificities might be found among closely-related strains.

Results of MLEE analysis suggest that while the chromosomal genes of *E. coli* are highly polymorphic, linkage disequilibrium is high and the number of distinct genotypes is limited (Selander and Levin, 1980; Whittam *et al.*, 1983; Selander *et al.*, 1987). This may be due to the action of periodic selection for clones of higher fitness coupled with random extinction of cell lines, and low rates of large-scale recombination (Levin, 1981). The result is a clonal population of geographically widespread lineages, in which recombination is too rare or not over sufficiently large enough regions to obscure phylogenetic relationships or prevent particular electromorph types from being maintained over long periods of time. Within this framework, however, some loci may be under substantial pressure for diversity. These loci are generally associated with surface antigens and capsules involved in evasion of host immune systems, in which variation would provide a selective advantage. Type I r-m systems would be included in this category, as recognition of incoming foreign DNA should be an advantage.

REFERENCES

- Adler, S.P., and Nathans, D. (1973). *Biochem. Biophys. Acta* **299**: 177-188.
- Amábile-Cuevas, C.F., and Chicurel, M.E. (1992). *Cell* **76**: 189-199.
- Amitsur, M., Morad, I., Chapman-Shimshoni, D., and Kaufmann, G. (1992). *EMBO J.* **11**: 3129-3134.
- Amitsur, M., Morad, I., and Kaufmann, G. (1989). *EMBO J.* **8**: 2411-2415.
- Appleyard, R.K. (1954). *Genetics* **39**: 440-452.
- Arber, W. (1965a). *Ann. Rev. Microbiol.* **19**: 365-378.
- Arber, W. (1965b). *J. Mol. Biol.* **11**: 247-265.
- Arber, W., and Dussoix, D. (1962). *J. Mol. Biol.* **5**: 18-36.
- Arber, W., Hattman, S., and Dussoix, D. (1963). *Virology* **21**: 30-35.
- Arber, W., and Morse, M.L. (1965). *Genetics* **51**: 137-148.
- Arber, W., and Wauters-Willems, D. (1970). *Mol. Gen. Genet.* **108**: 203-217.
- Arber, W. (1974). *In: Prog. Nucl. Acid Res. Mol. Biol.* **14**. Ed. Cohn, W.E., Academic Press, New York, pp. 1-37.
- Arber, W., Iida, S., Jütte, H., Caspers, P., Meyer, J., and Hänni, C. (1978). *Cold Spring Harbor Symp. Quant. Biol.* **43**: 1197-1208.
- Argos, P. (1985). *EMBO J.* **4**: 1351-1355.
- Auer, B., and Schweiger, M. (1984). *J. Virol.* **49**: 588-590.
- Bachmann, B.J., and Low, K.B. (1980). *Microbiol. Rev.* **44**: 1-56.
- Bachmann, B.J. (1987). *In: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* Eds. Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umberger, H.E., Am. Soc. Microbiol., Washington, D.C., pp. 807-876.
- Bandyopadhyay, P.K., Studier, F.W., Hamilton, D.L., and Yuan, R. (1985). *J. Mol. Biol.* **182**: 567-578.
- Bannister, D., and Glover, S.W. (1968). *Biochem. Biophys. Res. Comm.* **30**: 735-

- Barcak, G.J., and Wolf Jr., R.E. (1988). *J. Bacteriol.* **170**: 372-379.
- Baron, L.S., Penido, E., Ryman, I.R., and Falkow, S. (1970). *J. Bacteriol.* **102**: 221-233.
- Belogurov, A.A., Yussinov, T.N., Kotova, V.U., and Zavilgelsky, G.B. (1985). *Mol. Gen. Genet.* **198**: 509-513.
- Belogurov, A.A., Delver, E.P., and Rodzevich, O.V. (1992). *J. Bacteriol.* **174**: 5079-5085.
- Beltran, P., Musser, J.M., Helmuth, R., Farmer III, J.J., Friedrichs, W.M., Wachsmuth, I.K., Ferris, K., McWhorter, A.C., Wells, J.G., Cravioto, A., and Selander, R.K. (1988). *Proc. Natl. Acad. Sci. USA* **85**: 7753-7757.
- Bennett, S.P., and Halford, S.E. (1989). *In: Current Topics in Cellular Regulation*, Ed. Horecker, B.L., Academic Press, New York, pp. 57-104.
- Benton, W.D., and Davis, R.W. (1977). *Science* **196**: 180-182.
- Bertani, G., and Weigle, J.J. (1953). *J. Bacteriol.* **65**: 113-121.
- Bertani, L.E. (1968). *Virology* **36**: 87-103.
- Bickle, T.A. (1982). *In: Nucleases* Ed. Linn, S.M., and Roberts, R.J. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 85-108.
- Bickle, T.A., Brack, C., and Yuan, R. (1978). *Proc. Natl. Acad. Sci. USA* **75**: 3099-3103.
- Bickle, T.A. (1987). *In: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. Eds. Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umberger, H.E., Am. Soc. Microbiol., Washington, pp. 692-696.
- Birnboim, H.C., and Doly, J. (1979). *Nucleic Acids Res.* **7**: 1513-1523.
- Bisereić, M., Feutrier, J.Y., and Reeves, P.R. (1991) *J. Bacteriol.* **173**: 3894-3900.
- Blumenthal, R.M., Gregory, S.A., and Cooperider, J.S. (1985). *J. Bacteriol.* **164**: 501-509.
- Bolivar, F., Rodriguez, R.L., Greene, P.J., Bétlach, M.C., Heyneker, H.L., and Boyer, H.W. (1977). *Gene* **2**: 95-113.
- Borck, K., Beggs, J.D., Brammar, W.J., Hopkins, A.S., and Murray, N.E. (1976).

- Mol. Gen. Genet. **146**: 199-207.
- Boulnois, G.J., and Wilkins, B.M. (1978). *J. Bacteriol.* **133**: 1-9.
- Boulnois, G.J., and Jann, K. (1989). *Mol. Microbiol.* **3**: 1819-1823.
- Boyd, A.C., and Sherratt, D.J. (1986). *Mol. Gen. Genet.* **203**: 496-504.
- Boyer, H. (1964). *J. Bacteriol.* **88**: 1652-1660.
- Boyer, H.W. (1971). *Ann. Rev. Microbiol.* **25**: 153-176.
- Boyer, H.W., and Roulland-Dussoix, D. (1969). *J. Mol. Biol.* **41**: 459-472.
- Brammar, W.J., Murray, N.E., and Winton, S. (1974). *J. Mol. Biol.* **90**: 633-647.
- Brown, P.K., Romana, L.K., and Reeves, P.R. (1991). *Mol. Microbiol.* **5**: 1873-1881.
- Bühler, R., and Yuan, R. (1978). *J. Biol. Chem.* **253**: 6756-6760.
- Bullas, L.R., and Colson, C. (1975). *Mol. Gen. Genet.* **139**: 177-188.
- Bullas, L.R., Colson, C., and Van Pel, A. (1976). *J. Gen. Microbiol.* **95**: 166-172.
- Bullas, L.R., Colson, C., and Neufield, B. (1980). *J. Bacteriol.* **141**: 275-292.
- Burckhardt, J., Weisemann, J., and Yuan, R. (1981a). *J. Biol. Chem.* **256**: 4024-4032.
- Burckhardt, J., Weisemann, J., Hamilton, D.L., and Yuan, R. (1981b). *J. Mol. Biol.* **153**: 425-440.
- Caspers, P., Dalrymple, B., Iida, S., and Arber, W. (1984). *Mol. Gen. Genet.* **196**: 68-73.
- Caugant, D.A., Levin, B.R., and Selander, R.K. (1981). *Genetics* **98**: 467-490.
- Cerritelli, S., Springhorn, S.S., and Lacks, S.A. (1989). *Proc. Natl. Acad. Sci. USA* **86**: 9223-9227.
- Chamberlin, M., McGrath, J., and Waskell, L. (1970). *Nature* **228**: 227-231.
- Chow, S., Daub, E., and Murialdo, H. (1987). *Gene* **60**: 277-289.
- Clewell, D.B., and Weaver, K.E. (1989). *Plasmid* **21**: 175-184.
- Coffey, T.J., Dowson, C.G., Daniels, M., Zhou, J., Martin, C., Spratt, B.G., and

- Musser, J.M. (1991). *Mol. Microbiol.* **5**: 2255-2260.
- Colson, C., Colson, A.M., and Van Pel, A. (1970). *J. Gen. Microbiol.* **60**: 265-271.
- Colson, C., and Colson, A.M. (1971). *J. Gen. Microbiol.* **69**: 345-351.
- Compton, T. (1990). *In: PCR Protocols: A Guide to Methods and Applications.* Eds. Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., Academic Press. Inc., London, pp. 39-46.
- Cowan, G.M., Gann, A.A.F., and Murray, N.E. (1989). *Cell* **56**: 103-109.
- Daniel, A.S., Fuller-Pace, F.V., Legge, D.M., and Murray, N.E. (1988). *J. Bacteriol.* **170**: 1775-1782.
- DeBacker, D., and Colson, C. (1991a). *Gene* **97**: 103-107.
- Debrouwere, L., Zabeau, M., van Montague, M., and Schell, J. (1980). *Mol. Gen. Genet.* **179**: 75-80.
- Delver, E.P., Kotova, V.U., Zavilgelsky, G.B., and Belogurov, A.A. (1991). *J. Bacteriol.* **173**: 5887-5892.
- Devereux, J., Haeberli, P., and Smithies, O. (1984). *Nucleic Acids Res.* **12**: 387-395.
- DeVries, G.E., Raymond, C.K., and Ludwig, R.A. (1986). *Proc. Natl. Acad. Sci. USA* **81**: 6080-6084.
- Dharmalingam, K., and Goldberg, E.B. (1976). *Nature* **260**: 454-456.
- Dharmalingam, K., Revel, H.R., and Goldberg, E.B. (1982). *J. Bacteriol.* **149**: 694-699.
- Dila, D., Sutherland, E., Moran, L., Slatko, B., and Raleigh, E.A. (1990). *J. Bacteriol.* **172**: 4888-4900.
- Drexler, H., and Christensen, J.R. (1961). *Virology* **13**: 31-39.
- Dubose, R.F., Dykhuizen, D.E., and Hartl, D.L. (1988). *Proc. Natl. Acad. Sci. USA* **85**: 7036-7040.
- Duguid, J.P., Smith, I.W., Dempster, G., and Edmunds, P.N. (1955). *J. Path.* **70**: 335-348.
- Dunn, J.J., and Studier, F.W. (1981). *J. Mol. Biol.* **148**: 303-330.
- Dussoix, D., and Arber, W. (1962). *J. Mol. Biol.* **5**: 37-49.

- Dykhuizen, D.E., and Green, L. (1991). *J. Bacteriol.* **173**: 7257-7268.
- Endlich, B., and Linn, S. (1981). In: *The Enzymes*, (3rd Edition) Ed. Boyer, P.B., Academic Press, New York, pp. 137-156.
- Endlich, B., and Linn, S. (1985a). *J. Biol. Chem.* **260**: 5720-5728.
- Endlich, B., and Linn, S. (1985b). *J. Biol. Chem.* **260**: 5729-5738.
- Eskin, B., and Linn, S. (1972a). *J. Biol. Chem.* **247**: 6183-6191.
- Eskin, B., and Linn, S. (1972b). *J. Biol. Chem.* **247**: 6192-6196.
- Eskin, B., Lautenberger, J.A., and Linn, S. (1973). *J. Virol.* **11**: 1020-1023.
- Eskridge, R.W., Weinfeld, H., and Paigen, K. (1967). *J. Bacteriol.* **93**: 835-844.
- Falkow, S., Guerry, P., Hedges, R.W., and Datta, N. (1974). *J. Gen. Microbiol.* **85**: 65-76.
- Figueroa, F., Günther, E., and Klein, J. (1988). *Nature* **335**: 265-267.
- Friedman, D.I., and Baron, L.S. (1974). *Virology* **58**: 141-148.
- Friedman, D.I., and Olson, E.R. (1983). *Cell* **34**: 143-149.
- Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N.E. (1983). *J. Mol. Biol.* **170**: 827-842.
- Fuller-Pace, F.V., Bullas, L.R., Delius, H., and Murray, N.E. (1984). *Proc. Natl. Acad. Sci. USA* **81**: 6095-6099.
- Fuller-Pace, F.V., Cowan, G.M., and Murray, N.E. (1985). *J. Mol. Biol.* **186**: 65-75.
- Fuller-Pace, F.V., and Murray, N.E. (1986). *Proc. Natl. Acad. Sci. USA* **83**: 9368-9372.
- Gann, A.A.F., Campbell, A.J.B., Collins, J.F., Coulson, A.F.W., and Murray, N.E. (1987). *Mol. Microbiol.* **1**: 13-22.
- Glover, S.W. (1970). *Genet. Res. Camb.* **15**: 237-250.
- Glover, S.W., Schell, J., Symonds, N., and Stacey, K.A. (1963). *Genet. Res.* **4**: 480-482.
- Glover, S.W., and Colson, C. (1969). *Genet. Res. Camb.* **13**: 227-240.

- Glover, S.W., Firman, K., Watson, G., Price, C., and Donaldson, S. (1983). *Mol. Gen. Genet.* **190**: 65-69.
- Gorbalenya, A.E., and Koonin, E.V. (1991). *FEBS Letts.* **291**: 271-281.
- Gormley, E., and Davies, J. (1991). *J. Bacteriol.* **173**: 6705-6708.
- Gough, J.A., and Murray, N.E. (1983). *J. Mol. Biol.* **166**: 1-19.
- Groisman, E.A., Saier, Jr., M.H., and Ochman, H. (1992). *EMBO J.* **11**: 1309-1316.
- Gromkova, R., and Goodgal, S.H. (1976). *J. Bacteriol.* **127**: 848-854.
- Gubler, M., Braguglia, D., Meyer, J., Piekarowicz, A., and Bickle, T.A. (1992). *EMBO J.* **11**: 233-240.
- Gubler, M., and Bickle, T.A. (1991). *EMBO J.* **10**: 951-957.
- Guild, W.R., Smith, M.D., and Shoemaker, N.B. (1982). *In: Microbiology 1982.* Ed. Schlessinger, D., American Society for Microbiology, Washington, D.C., pp. 82-92.
- Haberman, A., Heywood, J., and Meselson, M. (1972). *Proc. Natl. Acad. Sci. USA* **69**: 3138-3141.
- Hadi, S.M., and Yuan, R. (1974). *J. Biol. Chem.* **249**: 4580-4586.
- Hadi, S.M., Bickle, T.A., and Yuan, R., (1975). *J. Biol. Chem.* **250**: 4159-4164.
- Hadi, S.M., Bächli, B., Iida, S., and Bickle, T.A. (1983). *J. Mol. Biol.* **165**: 19-34.
- Hardy, K. (1984). *Bacterial Plasmids* (2nd edit.). Van Nostrand Reinhold Co. Ltd, Wokingham, pp. 23-54.
- Hattman, S. (1970) *Virology* **42**: 359-367.
- Hattman, S. (1979). *J. Virol.* **32**: 468-475.
- Hattmann, S. (1983). *In: Bacteriophage T4.* Eds. Mathews, C.K., Kutter, E.M., Mosig, G., and Berget, P.B., American Society for Microbiology, Washington D.C., pp. 152-155.
- Hedges. R.W., and Datta, N. (1972). *J. Gen. Microbiol.* **71**: 403-405.
- Hedgpeth, J., Goodman, H.M., and Boyer, H.W. (1972). *Proc. Natl. Acad. Sci. USA* **69**: 3448-3452.

- Heery, D.M., and Dunican, L.K. (1989). *Nucleic Acids Res.* **17**: 8006.
- Herzer, P.J., Inouye, S., Inouye, M., and Whittam, T.S. (1990). *J. Bacteriol.* **172**: 6175-6181.
- Horiuchi, K., and Zinder, N.D. (1972). *Proc. Natl. Acad. Sci. USA* **69**: 3220-3224.
- Horiuchi, K., Vovis, G.F., and Zinder, N.D. (1974). *J. Biol. Chem.* **249**: 543-552.
- Hubacek, J., and Glover, S.W. (1970). *J. Mol. Biol.* **50**: 111-127.
- Hughes, J.A., Brown, L.K., and Ferro, A.J. (1987). *Nucleic Acids Res.* **15**: 717-730.
- Hümbelin, M., Suri, B., Rao, D.N., Hornby, D.P., Eberle, H., Pripfl, T., Kenel, S., and Bickle, T.A. (1988). *J.Mol.Biol.* **200**: 23-29.
- Iida, S., and Arber, W. (1977). *Mol. Gen. Genet.* **153**: 259-269.
- Iida, S., Meyer, J., Kennedy, K.E., and Arber, W. (1982). *EMBO J.* **1**: 1445-1453.
- Iida, S., Meyer, J., Bächli, B., Stålhammar-Carlemalm, M., Schrickel, S., Bickle, T.A., and Arber, W. (1983). *J. Mol. Biol.* **165**: 1-18.
- Iida, S., Streiff, M.B., Bickle, T.A., and Arber, W. (1987a). *Virology* **157**: 156-166.
- Iida, S., Meyer, J., and Arber, W. (1987b). *FEMS Microbiol. Lett.* **43**: 117-120.
- Iino, T., Komeda, Y., Kutsukake, K., Macnab, R.M., Matsumura, P., Parkinson, J.S., Simon, M.I., and Yamaguchi, S. (1988). *Microbiol. Rev.* **52**: 533-535.
- Ikeda, H., and Tomizawa, J. (1965). *J. Mol. Biol.* **14**: 85-109.
- Innis, M.A., and Gelfand, D.H. (1990). In: *PCR Protocols: A Guide to Methods and Applications*. Eds. Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., Academic Press, Inc., London, pp. 3-12.
- Ioerger, T.R., Clark, A.G., and Kao, T.-H. (1990). *Proc. Natl. Acad. Sci. USA* **87**: 9732-9735.
- Jann, B., and Jann, K. (1990). *Curr. Topics Microbiol. Immunol.* **150**: 19-42.
- Jansson, P.-E., Lindberg, A.A., Lindberg, B., and Wollin, R. (1981). *Eur. J. Biochem.* **115**: 571-577.
- Jiang, X.M., Neal, B., Santiago, F., Lee, S.J., Romana, L.K., and Reeves, P.R. (1991). *Mol. Microbiol.* **5**: 695-713.

- Kahmann, R. (1984). *Curr. Topics Microbiol. Immunol.* **108**: 29-47.
- Kahmann, R., and Hattman, S. (1987). In: *Phage Mu*. Eds. Symonds, N., Toussaint, A., Van De Putte, P., and Howe, M.M. Cold Spring Harbor Laboratory, N.Y., pp. 93-109.
- Kahmann, R., Seiler, A., Wulczyn, F.G., and Pfaff, E. (1985). *Gene* **39**: 61-70.
- Kan, N.C., Lautenberger, J.A., Edgell, M.H., and Hutchison III, C.A. (1979). *J. Mol. Biol.* **130**: 191-209.
- Kannan, P., Cowan, G.M., Daniel, A.S., Gann, A.A.F., and Murray, N.E. (1989). *J. Mol. Biol.* **209**: 335-344.
- Kaszubska, W., Aiken, C., O'Connor, C.D., and Gumpert, R.I. (1989). *Nucleic Acids Res.* **17**: 10403-10425.
- Kauc, L., and Piekarowicz, A. (1978). *Eur. J. Biochem.* **92**: 417-426.
- Kelleher, J.E., Daniel, A.S., and Murray, N.E. (1991). *J. Mol. Biol.* **221**: 431-440.
- Kessler, C., and Manta, V. (1990). *Gene* **92**: 1-248.
- Kornberg, A., Zimmerman, S.B., Kornberg, S.R., and Josse, J. (1959). *Proc. Natl. Acad. Sci. USA* **45**: 772-785.
- Kornberg, S.R., Zimmerman, S.B., and Kornberg, A. (1961). *J. Biol. Chem.* **236**: 1487-1493.
- Kroll, J.S., and Moxon, E.R. (1990). *J. Bacteriol.* **172**: 1374-1379.
- Kroll, J.S., Zamze, S., Loynds, B., and Moxon, E.R. (1989). *J. Bacteriol.* **171**: 3343-3347.
- Krüger, D.H., Schroeder, C., Hansen, S., and Rosenthal, H.A. (1977). *Mol. Gen. Gen.* **153**: 99-106.
- Krüger, D.H., and Schroeder, C. (1981). *Microbiol. Rev.* **45**: 9-51.
- Krüger, D.H., and Bickle, T.A. (1983). *Microbiol. Rev.* **47**: 345-360.
- Krüger, T., Grund, C., Wild, C., and Noyer-Weidner, M. (1992). *Gene* **114**: 1-12.
- Kühnlein, U., Linn, S., and Arber, W. (1969). *Proc. Natl. Acad. Sci. USA* **63**: 556-562.
- Lacks, S., and Greenberg, B. (1975). *J. Biol. Chem.* **250**: 4060-4066.

- Lacks, S.A., Mannarelli, B.M., Springhorn, S.S., and Greenberg, B. (1986). *Cell* **46**: 993-1000.
- Lautenberger, J.A., and Linn, S. (1972). *J. Biol. Chem.* **247**: 6176-6182.
- Lautenberger, J.A., Kan, N.C., Lackey, D., Linn, S., Edgell, M.H., and Hutchison III, C.A. (1978). *Proc. Natl. Acad. Sci. USA* **75**: 2271-2275.
- Lawlor, D.A., Ward, F.E., Ennis, P.D., Jackson, A.P., and Parham, P. (1988). *Nature* **335**: 268-271.
- Lawrence, J.G., Ochman, H., and Hartl, D.L. (1991a). *J. Gen. Microbiol.* **137**: 1911-1921.
- Lawrence, J.G., Hartl, D.L., and Ochman, H. (1991b). *J. Mol. Evol.* **33**: 241-250.
- Lederberg, S. (1957). *Virology* **3**: 496-513.
- Lee, S.J., Romana, L.K., and Reeves, P.R. (1992). *J. Gen. Microbiol.* **138**: 1843-1855.
- Lehman, I.R., and Pratt, E.A. (1960). *J. Biol. Chem.* **235**: 3254-3258.
- Le Minor, L. (1984). *In: Bergey's Manual of Systematic Bacteriology*. Eds. Krieg, N.R., and Holt, J.G., Williams and Wilkins, London, pp. 427-458.
- Lennox, E.S. (1955). *Virology* **1**: 190-206.
- Levin, B.R. (1981). *Genetics* **99**: 1-23.
- Levin, B.R. (1986). *In: Evolutionary Processes and Theory*. Eds. Karlin, S., and Nero, E., Academic Press, Inc., New York, pp. 669-688.
- Levin, B.R. (1988a). *Philos. Trans. R. Soc. London, Ser. B* **319**: 459-472.
- Levin, B.R. (1988b). *In: The Evolution of Sex in Bacteria*. Eds. Michod, R.E., and Levin, B.R., Sinauer Assoc. Inc., Sunderland, Mass., pp. 194-211.
- Linn, S., and Arber, W. (1968). *Proc. Natl. Acad. Sci. USA* **59**: 1300-130.
- Linder, P., Doelz, R., Gubler, M., and Bickle, T.A. (1990). *Nucleic Acids Res.* **18**: 7170.
- Loenen, W.A.M., Daniel, A.S., Braymer, H.D., and Murray, N.E. (1987). *J. Mol. Biol.* **198**: 159-170.
- Loenen, W.A.M., and Murray, N.E. (1986). *J. Mol. Biol.* **190**: 11-22.

- Lund, B., Lindberg, F., Marklund, B.-I., and Normark, S. (1987). *Proc. Natl. Acad. Sci. USA* **84**: 5898-5902.
- Lundberg, A.S., and McDevitt, H.O. (1992). *Proc. Natl. Acad. Sci. USA* **89**: 6545-6549.
- Luria, S.E. (1953). *Cold Spring Harbor Symp. Quant. Biol.* **18**: 237-244.
- Luria, S.E., and Human, M.L. (1952). *J. Bacteriol.* **64**: 557-569.
- Mandel, M., and Higa, A. (1970). *J. Mol. Biol.* **53**: 159-162.
- Mark, K.K., and Studier, F.W. (1981). *J. Biol. Chem.* **256**: 2573-2578.
- Marklund, B.-I., Tennent, J.M., Garcia, E., Hamers, A., Baga, M., Lindberg, F., Gaastra, W., and Normark, S. (1992). *Mol. Microbiol.* **6**: 2225-2242.
- Masters, M. (1985). *In: Genetics of Bacteria.* Eds. Scaife, J., Leach, D., and Galizzi, A., Academic Press. Inc., London, pp. 197-215.
- Mazodier, P., and Davies, J. (1991). *Ann. Rev. Genet.* **25**: 147-171.
- Meisel, A., Bickle, T.A., Krüger, D.H., and Schroeder, C. (1992). *Nature* **355**: 467-469.
- Meselson, M., and Yuan, R. (1968). *Nature* **217**: 1110-1114.
- Milkman, R., and Crawford, I.P. (1983). *Science* **221**: 378-380.
- Milkman, R., and Bridges, M.M. (1990). *Genetics* **126**: 505-517.
- Miller, R.D., and Hartl, D.L. (1986). *Evolution* **40**: 1-12.
- Miller, J. (1974). *Experiments in Molecular Genetics* (2nd edit.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. 201-205.
- Mise, K., and Arber, W. (1976). *Virology* **69**: 191-205.
- Miyazaki, J., Fujisawa, H., Ryo, Y., and Minagawa, T. (1982). *Virology* **122**: 1-7.
- Moffatt, B.A., and Studier, F.W. (1988). *J. Bacteriol.* **170**: 2095-2105.
- Mosig, G., and Eiserling, F. (1988). *In: The Bacteriophages.* Ed. Calendar, R., Plenum Press, London, pp. 521-606.
- Murray, N.E., and Murray, K. (1974). *Nature* **251**: 476-481.
- Murray, N.E., Manduca de Ritis, P., and Foster, L.A. (1973). *Mol. Gen. Genet.*

120: 261-281.

Murray, N.E., Gough, J.A., Suri, B., and Bickle, T.A. (1982). *EMBO J.* 1: 535-539.

Musser, J.M., Kroll, J.S., Moxon, E.R., and Selander, R.K. (1988). *Proc. Natl. Acad. Sci. USA* 85: 7758-7762.

Nagaraja, V., Shepherd, J.C.W., and Bickle, T.A. (1985a). *Nature* 316: 371-372.

Nagaraja, V., Shepherd, J.C.W., Pripfl, T., and Bickle, T.A. (1985b). *J. Mol. Biol.* 182: 579-587.

Nagaraja, V., Steiger, M., Nager, C., Hadi, S.M., and Bickle, T.A. (1985c). *Nucleic Acids Res.* 13: 389-399.

Nelson, K., and Selander, R.K. (1992). *J. Bacteriol.* 174: 6886-6895.

Nelson, K., Whittam, T.S., and Selander, R.K. (1991). *Proc. Natl. Acad. Sci. USA* 88: 6667-6671.

Noyer-Weidner, M., Diaz, R., and Reiners, L. (1986). *Mol. Gen. Genet.* 205: 469-475.

Nussinov, R. (1984). *Nucleic Acids Res.* 12: 1749-1763.

Nutter, R.L., Bullas, L.R., and Schmetz, R.L. (1970). *J. Virol.* 5: 754-764.

Ochman, H., Whittam, T.S., Caugant, D.A., and Selander, R.K. (1983). *J. Gen. Microbiol.* 129: 2715-2726.

Ochman, H., and Selander, R.K., (1984a). *J. Bacteriol.* 157: 690-693.

Ochman, H., and Selander, R.K. (1984b). *Proc. Natl. Acad. Sci. USA* 81: 198-201.

Ochman, H., and Wilson, A.C. (1987). *In: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology.* Eds. Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umberger, H.E., American Society for Microbiology, Washington, D.C., pp. 1649-1654.

Olson, E.R., Flamm, E.L., and Friedman, D.I. (1982). *Cell* 31: 61-70.

Ornellas, E.P., and Stocker, B.A.D. (1974). *Virology* 60: 491-502.

Ørskov, F. (1984). *In: Bergey's Manual of Systematic Bacteriology.* Eds. Krieg, N.R., and Holt, J.G., Williams and Wilkins, London, pp. 420-423.

Pao, C.-C., and Speyer, J.F. (1973). *J. Virol.* 11: 1024-1026.

- Pauling, C., and Hamm, L. (1968). *Proc. Natl. Acad. Sci. USA* **60**: 1495-1502.
- Piekarowicz, A., Goguen, J.D., and Skrzypek, E. (1985). *Eur. J. Biochem.* **157**: 387-393.
- Piekarowicz, A., and Goguen, J.D. (1986). *Eur. J. Biochem.* **167**: 111-115.
- Price, C., Shepherd, J.C.W., and Bickle, T.A. (1976). *EMBO J.* **6**: 1493-1497.
- Price, C., and Bickle, T.A. (1986). *Microbiol. Sci.* **3**: 296-299.
- Price, C., Pripfl, T., and Bickle, T.A. (1987a). *Eur. J. Biochem.* **167**: 111-115.
- Price, C., Shepherd, J.C.W., and Bickle, T.A. (1987b). *EMBO J.* **6**: 1493-1497.
- Price, C., Lingner, J., and Bickle, T.A. (1989). *J. Mol. Biol.* **205**: 115-125.
- Price., C., Lingner, J., Bickle, T.A., Firman, K., and Glover, S.W. (1989). *J. Mol. Biol.* **205**: 115-125.
- Radding, C.M. (1973). *Annu. Rev. Genet.* **7**: 87-111.
- Raleigh, E.A. (1992). *Mol. Microbiol.* **6**: 1079-1086.
- Raleigh, E.A., and Wilson, G. (1986). *Proc. Natl. Acad. Sci. USA* **83**: 9070-9074.
- Raleigh, E.A., Trimarchi, R., and Revel, H. (1989). *Genetics* **122**: 279-296.
- Ravetch, J.V., Horiuchi, K., and Zinder, N.D. (1978). *Proc. Natl. Acad. Sci. USA* **75**: 2266-2270.
- Rayssiguier, C., Thaler, D.S., and Radman, M. (1989). *Nature* **342**: 396-401.
- Read, T.D., Thomas, A.T., and Wilkins, B.M. (1992). *Mol. Microbiol.* **6**: 1933-1941.
- Redfield, R.J., and Campbell, A.M. (1987). *J. Mol. Biol.* **198**: 393-404.
- Reiser, J., and Yuan, R. (1977). *J. Biol. Chem.* **252**: 451-456.
- Revel, H.R. (1983). *In: Bacteriophage T4*. Eds. Mathews, C.K., Kutter, E.M., Mosig, G., and Berget, P.B., American Society for Microbiology, Washington D.C., pp.156-165.
- Revel, H.R. (1967) *Virology* **31**: 688-701.
- Revel, H.R., and Georgopoulos, C.P. (1969). *Virology* **39**: 1-17.

- Roberts, R.J. (1978). *Nature* **271**: 502.
- Roberts, R.J., and Macelis, D. (1992). *Nucleic Acids Res.* **20**: 2167-2180.
- Robeson, J.P., Goldschmidt, R.M., and Curtiss III, R. (1980). *Nature* **283**: 104-106.
- Rosamond, J., Endlich, B., and Linn, S. (1979). *J. Mol. Biol.* **129**: 619-635.
- Ross, T.K., Achberger, E.C., and Braymer, H.D. (1987). *Gene* **61**: 277-289.
- Ross, T.K., Achberger, E.C., and Braymer, H.D., (1989). *J. Bacteriol.* **171**: 1974-1981.
- Roulland-Dussoix, D., and Boyer, H. (1969). *Biochem. Biophys. Acta* **195**: 219-229.
- Roulland-Dussoix, D., Yoshimori, R., Greene, P., Betlach, M., Goodman, H.M., and Boyer, H.W. (1974). *In: Microbiology 1974*. Ed. Schlessinger, D., American Society for Microbiology Press, Washington D.C., pp. 187-198.
- Ryu, J.-I., Rajadas, P.T., and Bullas, L.R. (1988). *J. Bacteriol.* **170**: 5785-5788.
- Sain, B., and Murray, N.E. (1980). *Mol. Gen. Genet.* **180**: 35-46.
- Sandri, R.M., and Berger, H. (1980). *Virology* **106**: 14-29.
- Sandulache, R., Prehm, P., and Kamp, D. (1984). *J. Bacteriol.* **160**: 299-303.
- Schroeder, C., Jurkschat, H., Meisel, A., Reich, J.-G., and Krüger, D. (1986). *Gene* **45**: 77-86.
- Selander, R.K., and Levin, B.R. (1980). *Science* **210**: 545-547.
- Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N., and Whittam, T.S. (1986). *Appl. Environ. Microbiol.* **51**: 873-884.
- Selander, R.K., Caugant, D.A., and Whittam, T.S. (1987). *In: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. Eds. Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umberger, H.E., American Society for Microbiology, Washington, D.C., pp.1625-1648.
- Semerjian, A.V., Malloy, D.C., and Poteete, A.R. (1989). *J. Mol. Biol.* **207**: 1-13.
- Sharp, P.A., Sugden B., and Sambrook, J. (1973). *Biochemistry* **12**: 3055-3063.
- Sharp, P.M. (1986). *Mol. Biol. Evol.* **3**: 75-83.
- Sharp, P.M., Rogers, M.S., and McConnell, D.J. (1985). *J. Mol. Evol.* **21**: 150-160.

- Sharp, P.M., Kelleher, J.E., Daniel, A.S., Cowan, G.M., and Murray, N.E. (1992). *Proc. Natl. Acad. Sci. USA* **89**: 9836-9840.
- Smith, G.E., and Summers, M.D. (1980). *Anal. Biochem.* **109**: 123-129.
- Smith, H.O., and Wilcox, K.W. (1970). *J. Mol. Biol.* **51**: 379-391.
- Smith, J.M., Dowson, C.G., and Spratt, B.G. (1991). *Nature* **349**: 29-31.
- Smith, J.M. (1992). *J. Mol. Evol.* **34**: 126-129.
- Smith, N.H., and Selander, R.K. (1990). *J. Bacteriol.* **172**: 603-609.
- Smith, N.H., Beltran, P., and Selander, R.K. (1990). *J. Bacteriol.* **172**: 2209-2216.
- Southern, E.M. (1975). *J. Mol. Biol.* **98**: 503-517.
- Sozhamannan, S., and Dharmalingam, K. (1988). *Gene* **74**: 51-52.
- Spoerel, N., Herrlich, P., and Bickle, T.A. (1979). *Nature* **278**: 30-34.
- Spratt, B.G., Zhang, Q.-Y., Jones, D.M., Hutchison, A., Brannigan, J.A., and Dowson, C.G. (1989). *Proc. Natl. Acad. Sci. USA* **86**: 8988-8992.
- Spratt, B.G., Bowler, L.D., Zhang, Q.-Y., Zhou, J., and Smith, J.M. (1992). *J. Mol. Evol.* **34**: 115-125.
- Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M., and Greene, P.J. (1989). *Gene* **85**: 1-13.
- Sternberg, N. (1990). *Proc. Natl. Acad. Sci. USA* **87**: 103-107.
- Sternberg, N.L. (1992). *Trends Genet.* **8**: 11-16.
- Stoltzfus, A., Leslie, J.F., and Milkman, R. (1988). *Genetics* **120**: 345-358.
- Streicher, S.L., Shanmugan, K.T., Ausubel, F., Morandi, C., and Goldberg, R.B. (1974). *J. Bacteriol.* **120**: 815-821.
- Streiff, M.B., Iida, S., and Bickle, T.A. (1987). *Virology* **157**: 167-171.
- Stocker, B.A.D., and Mäkelä, P.H., (1978). *Proc. Roy. Soc. Lond. Ser. B.* **202**: 5-30.
- Strömberg, N., Marklund, B.-I., Lund, B., Ilver, D., Hamers, A., Gaastra, W., Karlsson, K.-A., and Normark, S. (1990). *EMBO J.* **9**: 2001-2010.
- Strömberg, N., Nyholm, P.-G., Pascher, I., and Normark, S. (1991). *Proc. Natl.*

- Acad. Sci. USA **88**: 9340-9344.
- Studier, F.W., and Bandyopadhyay, P.K. (1988). Proc. Natl. Acad. Sci. USA **85**: 4677-4681.
- Suri, B., Nagaraja, V., and Bickle, T.A. (1984a). Curr. Top. Microbiol. Immunol. **108**: 1-9.
- Suri, B., Shepherd, J.C.W., and Bickle, T.A. (1984b). EMBO J. **3**: 575-579.
- Suri, B., and Bickle, T.A. (1985). J. Mol. Biol. **186**: 77-85.
- Sutherland, E., Coe, L., and Raleigh, E.A. (1992). J. Mol. Biol. **225**: 327-348.
- Swinton, D., Hattman, S., Crain, P.F., Cheng, C.-S., Smith, D.L., and McCloskey, J.A. (1983). Proc. Natl. Acad. Sci. USA **80**: 7400-7404.
- Szybalski, W., Kim, S.C., Hasan, N., and Podhajski, A.J. (1991). Gene **100**: 13-26.
- Takano, T., Watanabe, T., and Fukasawa, T. (1966). Biochem. Biophys. Res. Comm. **25**: 192-198.
- Toothman, P. (1981). J. Virol. **38**: 621-631.
- Toussaint, A. (1976). Virology **70**: 17-27.
- Trieu-Cuot, P., Carlier, C., and Courvalin, P. (1988). J. Bacteriol. **170**: 4388-4391.
- Van Pel, A., and Colson, C. (1974). Mol. Gen. Genet. **135**: 51-60.
- Verma, N., and Reeves, P.R. (1989). J. Bacteriol. **171**: 5694-5701.
- Vovis, G.F., Horiuchi, K., and Zinder, N.D. (1974). Proc. Natl. Acad. Sci. USA **71**: 3810-3813.
- Vovis, G.F., and Zinder, N.D. (1975). J. Mol. Biol. **95**: 557-568.
- Wagner, E.F., Auer, B., and Schweiger, M. (1979). J. Virol. **29**: 1229-1231.
- Wang, L., Romana, L.K., and Reeves, P.R. (1992). Genetics **130**: 429-443.
- Weisberg, R.A. (1987). *In: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. Eds. Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umberger, H.E., American Society for Microbiology, Washington, D.C., pp. 1169-1176.
- Whittaker, P.A., Campbell, A.J.B., Southern, E.M., and Murray, N.E. (1988). Nucleic Acids Res. **16**: 6725-6736.

- Whittam, T.S., Ochman, H., and Selander, R.K. (1983). *Proc. Natl. Acad. Sci. USA* **80**: 1751-1755.
- Whittam, T.S., Ochman, H., and Selander, R.K. (1984). *Mol. Biol. Evol.* **1**: 67-83.
- Wilkins, B.M., Rees, C.E., Thomas, A.T., and Read, T.D. (1991). *Plasmid* **25**: 227.
- Willetts, N. (1985). *In: Genetics of Bacteria*. Eds. Scaife, J., Leach, D., and Galizzi, A., Academic Press, London, pp. 197-215.
- Willetts, N., and Wilkins, B. (1984). *Microbiol. Rev.* **48**: 24-41.
- Wilson, G.G. (1988). *Trends Genet.* **4**: 314-318.
- Wilson, G.G. and Murray, N.E. (1991). *Ann. Rev. Genet.* **25**: 585-627.
- Wood, W.B. (1966). *J. Mol. Biol.* **16**: 118-133.
- Wyatt, G.R., and Cohen, S.S. (1952). *Nature* **170**: 1072-1073.
- Yanofsky, C., and Ito, J. (1966). *J. Mol. Biol.* **21**: 313-344.
- Yarmolinsky, M.B., and Sternberg, N. (1988). *In: The Bacteriophages (Vol.1)*. Ed. Calendar, R., Plenum Press, London, pp. 291-438.
- Yuan, R., and Meselson, M. (1970). *Proc. Natl. Acad. Sci. USA* **65**: 357-362.
- Yuan, R., Heywood, J., and Meselson, M., (1972). *Nature New Biology* **240**: 42-43.
- Yuan, R., Bickle, T.A., Ebbers, W., and Brack, C. (1975). *Nature* **256**: 556-560.
- Yuan, R., Hamilton, D.L., and Burckhardt, J. (1980). *Cell* **20**: 237-244.
- Zabeau, M., Friedman, S., van Montague, M., and Schell, J. (1980). *Mol. Gen. Genet.* **179**: 63-73.
- Zavriev, S.K., and Shemyakin, M.F. (1982). *Nucleic Acids Res.* **10**: 1635-1652.

BCN: 0 680,845 Owner:LOADER Clvl: 0 Table: MON
 Record Status: Acc/Hol
 Original operator: IRENE 29-10-93 Update operator: IRENE 29-10-93
 Proof List : y 00-00-00 (last print) MARC Request : n
 MARC to CIRC : n Remove from CIRC : n
 Non-Circulating : n LSP Request : n

001 Geac006808451
 004 00fae162895860
 008 931029s1993 st L 0 eng
 Date <931029>Natur <s>Date1 <1993>Date2 < >Cntry <st >Illus < >Intel < >Matde <L>Form < >
 Gov < >Conf < >Fest < >Index <0>Tithd < >Litfm < >Biog < >Lang <eng>Per < >BL < .
 049 00faE
 100 10faBarcusfhVictoria Anne
 245 14faThe type I restriction systems of Escherichia colifdby Victoria A. Barcus
 260 00fc1993
 502 00faThesis (Ph.D.)--University of Edinburgh, 1993

No errors found

HOLDINGS INFORMATION

001 Geac006808451
 966 00fc1f1SP.COLfmREFfsTheses Sect.1
 966 2 00fb30150011569560fc2f1DARWINfmNORMALfsTheses Section

Error Summary of record

Tag 966 Occ 01 Error no. 00044
 Missing mandatory subfield b

650 00 \$a Bacteriophages

LC

\$x Genetics

\$x Theses

650 00 \$a Escherichia coli

LC

\$x Genetics

\$x Theses

650 00 \$a Restriction enzymes, DNA

LC

\$x Theses