ISOLATION AND CHARACTERISATION OF *E. COLI xerC*, A GENE REQUIRED FOR *cer* SITE-SPECIFIC RECOMBINATION

A thesis submitted for
the degree of
Doctor of Philosophy
at the
University of Glasgow

by

Sean David Colloms

Institute of Genetics
University of Glasgow
Church St.
Glasgow

December 1990

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

The research reported in this thesis is my own and original work except where otherwise stated and has not been submitted for any other degree.

Dedicated to my sisters and my parents,
and in memory of Lionel.

CONTENTS

ABBREVIATIONS

i) Chemicals

APS	-	ammonium persulphate
ATP	-	adenosine triphosphate
BSA	-	bovine serum albumin
DNA	-	2' deoxyribonucleic acid
dNTP	-	2' deoxy (nucleotide)
DTT	-	dithiothreitol
EDTA	-	ethylene diamine tetra-acetic acid (disodium salt)
IPTG	-	isopropyl B-D-thiogalactoside
PMSF	-	phenylmethylsulphonyl fluoride
RNA	-	ribonucleic acid
SDS	-	sodium dodecyl sulphate
TEMED	-	NNN' N' tetramethyl ethylenediamine
Tris	-	tris (hydroxymethyl) amino ethane
X-gal	-	5-bromo-4-chloro-3-indolyl-B-D-galactoside

ii) antibiotics

Ap	-	ampicillin
Cm	-	chloramphenicol
Km	-	kanamycin
Nal	-	nalidixic acid
Str	-	streptomycin
Tet	-	tetracycline

iii) Phenotype

X^r	-	resistance to X
X^s	-	sensitivity to X

iv) Measurements

bp	-	base pair
kbp	-	kilobase pair (10^3 bp)
kDa	-	kilodalton (10^3 dalton)
min	-	minute
sec	-	second

v) Miscellaneous

pfu	-	plaque forming unit
moi	-	multiplicity of infection
UV	-	ultra violet light

ACKNOWLEDGMENTS

I thank Dave, my supervisor, for his advice and enthusiasm concerning this project. I also thank Mary for keeping everything running smoothly and cleanly on the sixth floor, and the "prep room ladies" for supplying glassware, solutions and media.

Thanks to Drs. I. Bloomfield, C. Higgins, S. Kushner, C. Miller and C. Richaud for their kind gifts of plasmids and bacterial strains.

Special thanks to Marshall, who read most of this thesis while waiting for the gas man, and to Martin and Sally who each read one chapter. Gratitude is extended to Liz, Dave and the *cer* old guard - Richard and George - as well as to every inhabitant (past and present) of the sixth floor for their help and cooperation both at work and at play. Last but not least, I must thank Amy for making life in Glasgow so much more enjoyable.

SUMMARY

The formation of plasmid multimers by homologous recombination can greatly reduce the heritable stability of the natural multicopy plasmid ColE1. ColE1 carries a 250 bp site (*cer*) which, when present in direct repeat (as in a ColE1 dimer) efficiently resolves plasmid multimers to momomers. *cer*-mediated monomerisation is strongly correlated with an increase in plasmid stability. Current evidence suggests that ColE1 segregates randomly at cell division, and that *cer*-dependent monomerisation ensures stability by increasing the number of independently segregating plasmid copies (Sherratt *et al*, 1984; Summers and Sherratt, 1984).

The 250 bp *cer* site is the only ColE1 sequence required for *cer* site-specific recombination. ColE1 does not appear to encode any proteins necessary for *cer* recombination. Two *E. coli* chromosomal genes (*xer* genes) absolutely required for recombination at *cer* have been isolated and characterised (Stirling *et al*, 1988a; 1988b; 1989). One of these (*argR*) encodes the arginine repressor which binds to *cer* in the presence of L-arginine. The other (*xerB*) had not yet been characterised when this work began.

Evidence is presented here which shows that *xerB* corresponds to *pepA*, encoding aminopeptidase A. The nucleotide sequence has been extended 5' of *pepA* and potential *pepA* promoter sequences have been identified. The sequence has also been extended 3' of *pepA*, as far as *valS*, the gene encoding valyl-tRNA synthetase. Between *pepA* and *valS* is an open reading frame capable of encoding a polypeptide of molecular mass 13 kDa.

The procedure developed by Stirling (1987) was used to select further *xer* mutants. One of these carried a mutation in a previously unidentified gene, *xerC*. The *xerC* gene was isolated and mapped to 85 minutes, between the genes *dapF* (encoding diaminopimelate epimerase) and *uvrD* (encoding DNA helicase II). The nucleotide sequence of *xerC* and the flanking regions was determined. This revealed that *xerC* is capable of encoding a protein of

molecular mass 33.8 kDa. This protein has substantial amino acid sequence similarity to the lambda integrase family of site-specific recombinases. XerC is absolutely required for recombination at *cer* and also for recombination at a variant *cer* site which does not require either ArgR or PepA for recombination. This, and the similarity to the lambda integrase family, suggests that XerC is the recombinase that acts at *cer*.

The nucleotide sequence also revealed the presence of an open reading frame (*orf235*) of 235 codons between *dapF* and *xerC*, and another open reading frame (*orf238*) of 238 codons between *xerC* and *uvrD*. Tn5 and mini Mu insertions upstream of *xerC*, within either *dapF* or *orf235*, are polar on the expression of *xerC*. This suggests that *dapF*, *orf235*, *xerC* and *orf238* are all expressed on the same mRNA, from the *dapF* promoter.

Finally, a gel binding assay was developed to assay for binding of XerC to the *cer* site. This assay was used to follow the partial purification of XerC from a *xerC* over expressing *E. coli* strain. The gel binding assay was also used to show that XerC binds specifically to the crossover region of *cer* DNA.

CHAPTER ONE

INTRODUCTION

1.1 General introduction

Bacterial plasmids are extra-chromosomal hereditary elements, capable of autonomous replication and stable maintenance within their host. Most bacterial plasmids are found as covalently closed, circular DNA molecules. Plasmids encode a variety of phenotypic traits which include drug resistance, resistance to metal ions, production of enterotoxins, production of bacteriocins and altered sensitivity to mutagens (Broda, 1979). Plasmid encoded traits can confer a selective advantage, under certain environmental conditions, to their bacterial host.

Plasmids encode functions required for stable plasmid maintenance, including replication and partition functions (Nordstrom and Austin, 1989). Some plasmids also encode systems capable of killing plasmid-free segregants (Gerdes *et al*, 1985; Ogura and Hiraga, 1983b). Together, these functions ensure that naturally occurring plasmids are lost only rarely from growing cultures. Additionally, some plasmids carry conjugation or mobilisation genes which allow horizontal transfer of plasmid molecules from host to host.

Naturally occurring plasmids range in size from under 5 kbp to over 100 kbp and their copy number ranges from one to hundreds per host cell. In general, large plasmids are present at low copy number and small plasmids are present at higher copy number. Plasmid replication is tightly controlled at the level of initiation to maintain a constant copy number. The plasmid sequences responsible for this regulation are usually contained within a small region of the plasmid, consisting of a replication origin (*ori*) and replication control functions. After initiation of replication at the origin, host enzymes take over to complete plasmid replication.

1.2 Stable plasmid maintenance

The various plasmid functions that ensure stable plasmid maintenance are reviewed by Nordstrom and Austin (1989). There are two requirements that must be fulfilled

for a plasmid to be stably maintained at a constant copy number in all the cells of a population. The first of these is that plasmid replication keeps pace with host replication. Copy number control mechanisms ensure this. The second requirement is for the segregation of plasmid copies to both daughter cells at cell division. As long as both daughter cells receive at least one plasmid molecule at cell division, the copy number control mechanism can compensate for any deviation from the normal copy number.

1.2.1 Copy number control mechanisms

In all cases so far characterised, control of plasmid replication is carried out by a negative feedback loop, such that if plasmid copy number is too high, replication initiation is inhibited, and if the copy number becomes too low replication initiation is stimulated. This negative feedback is mediated by plasmid encoded repressors of initiation which provide a measure of plasmid copy number *via* gene dosage. This repressor can be a small RNA molecule (eg RNAI of ColE1 (Tomizawa *et al*, 1981) and CopA of R1 (Stougaard *et al*, 1981)), a small protein (eg Rop of ColE1 (Tomizawa and Som, 1984) and CopB of R1 (Molin *et al*, 1981)) or a series of DNA repeat sequences in the plasmid origin region which apparently titrate out plasmid initiation factors (eg the replication origins of F (Tolun and Helinski, 1981) and P1 (Chattoraj *et al*, 1984)).

1.2.2 Partition of plasmids at cell division

Plasmids can either be partitioned randomly at cell division or they can be actively partitioned. For low copy number plasmids, with only one or two copies per host chromosome, a random partition mechanism would result in the production of plasmid free cells at a very high frequency.

The low copy number plasmids F (Ogura and Hiraga, 1983a), R1 (Nordstrom *et al*, 1980) and P1 (Austin and Abeles, 1983) have active partition systems. These

systems all consist of a *cis*-acting site and two *trans*-acting Par proteins (Fig 1.1; reviewed by Austin and Nordstrom, 1990). The SopB protein of F has been purified and binds specifically to the *incD* (*sopC*) partition site of F (Mori *et al*, 1989). ParB of P1 has also been purified and binds specifically to the P1 *par* site (Davis and Austin, 1988). One or both of the R1 Par proteins probably binds to the R1 *par* site. The *par* sites of these plasmids are probably recognised by a combination of plasmid-encoded partition proteins and host proteins. Segregation "machinery" probably then acts on pairs of plasmids, to place one member of the pair into each daughter cell.

There is no evidence for the active partition of high copy number plasmids such as ColE1. The available evidence is consistent with the random partition of such plasmids at cell division (Durkatz and Sherratt, 1973; Summers and Sherratt, 1984). If this is the case, then plasmid-free segregants should be produced with a probability of 2^{1-n} per cell division, where n is the number of independently segregating plasmid units. ColE1 has an estimated copy number of 30 at cell division (Timmis, 1981). Therefore, provided that plasmid copies are free to segregate individually, ColE1 should produce plasmid-free segregants with an experimentally undetectable frequency of 1.9×10^{-9} per cell division. As predicted, ColE1 does not produce plasmid free cells at a detectable frequency.

Paradoxically, many common multicopy cloning vectors (derived from the ColE1 related plasmid pMB1) have a higher copy number than ColE1 but are less stable than ColE1. Summers and Sherratt (1984) showed that this instability is correlated with the formation of plasmid multimers by homologous recombination. Strains which produce large numbers of multimers lose multicopy plasmids at a faster rate than strains which produce few multimers (Summers and Sherratt, 1984). Copy number control mechanisms effectively count plasmid replication origins, so multimers are maintained at a lower copy number than monomers (Summers and Sherratt, 1984). This reduces the number of independently segregating units and thus

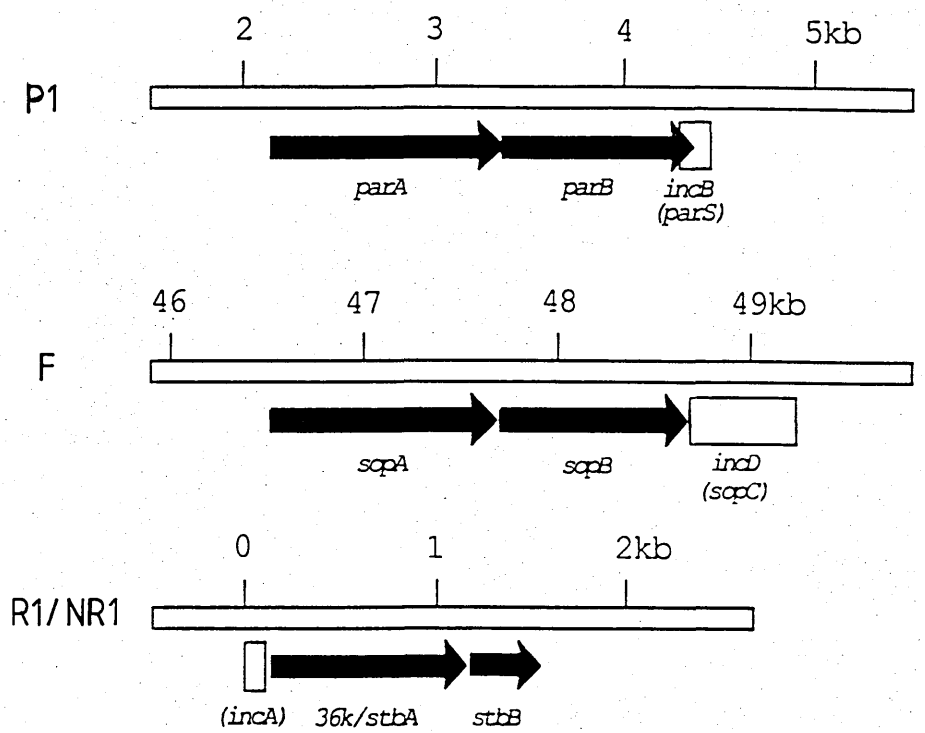


Figure 1.1 Genetic organisation of the *par* regions of the plasmids F, P1 and R1/NR1. Arrows indicate open reading frames for essential partition proteins. Open boxes show the regions containing the *cis* acting partition sites. (Taken from Nordstrom and Austin, 1989.)

increases the probability of producing plasmid-free cells.

ColE1 contains a sequence (*cer*) of approximately 250 bp, which, when present in direct repeat on a single plasmid (as in a ColE1 dimer), acts as a substrate for site-specific recombination. This site-specific recombination efficiently resolves plasmid multimers to monomers and ensures stable inheritance of ColE1 (Summers and Sherratt, 1984).

The *cer* site can stabilise the cloning vectors pUC8 and pACYC184 when present in *cis* (Summers and Sherratt, 1984). This stabilisation is correlated with an increase in the proportion of monomeric plasmid forms present (Summers and Sherratt, 1984). Site-specific recombination at *cer* is highly directional; *cer*-mediated intermolecular fusion occurs at 10^{-4} - 10^{-5} times the frequency of intramolecular deletion (D. Summers, pers. comm.).

Many other plasmids contain analogous, plasmid-stabilising, multimer resolution systems. The plasmids ColK (Summers *et al*, 1985), CloDF13 (Hakkaart *et al*, 1984), pMB1 (Greene *et al*, 1981), ColA (Morlon *et al*, 1988), pNPT16 (P. Strike, pers. comm.) and ColN (Kolot, 1990) contain recombination sites with extensive sequence similarity to ColE1 *cer* (Fig 1.2). The plasmid R1 also contains a plasmid-stabilising recombination site which is similar to *cer* only in the crossover region (Clerget, 1984; Fig 1.2). Most, if not all plasmids carry a site-specific recombination system capable of resolving plasmid multimers. Many of these systems utilise a plasmid-encoded recombinase and a recombination site with no sequence similarity to *cer*, eg Lox acts at *cre* on P1 (Austin *et al*, 1981), a resolvase acts at the R46 *per* site (Dodd and Bennett, 1986; 1987), the D-protein acts at *rfsF* on the F plasmid (Lane *et al*, 1986; O'Connor *et al*, 1986).

1.3 Site-specific recombination

Site-specific recombination reactions are catalysed by proteins known as recombinases. These proteins bind specifically to pairs of recombination sites and probably bring them together by protein-protein and protein-DNA

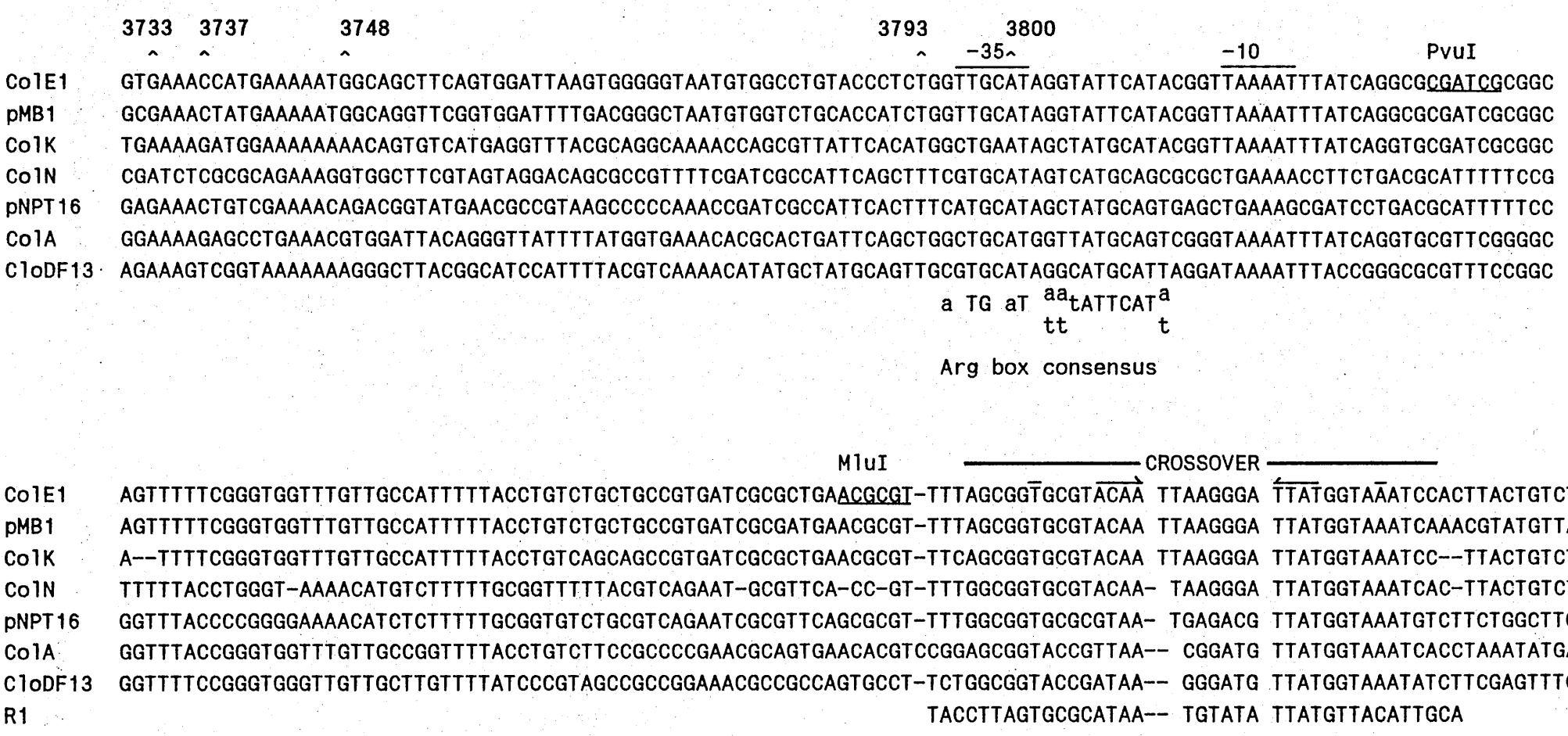


Figure 1.2 Alignment of ColE1 *cer* with the analagous sites from pMB1, ColK, CloDF13, ColA, ColN, pNPT16 and R1 (Greene *et al*, 1981; Summers *et al*, 1985; Morlon *et al*, 1988; Hakkaart *et al*, 1984; Kolot, 1990; P. Strike, pers. comm.; Clerget, 1984). The positions of the Arg box, the promoter, and the MluI and PvuI sites within *cer* are indicated. Gaps (-) have been inserted to maximise homology. The crossover region (3907-3942) as determined by sequencing the hybrid produced by recombination between ColE1 *cer* and ColK *ckr*, and a possible symmetry element within this region, are indicated. Coordinates given are those of ColE1 on the sequence of Chan (1985).

interactions. The recombinase then catalyses recombination by strand cleavage, exchange and religation. In contrast to homologous recombination, there is no requirement for extensive homology in the recombining sites. In general, the recombinase gene is found adjacent to the recombination site at which the recombinase acts.

Site-specific recombination sites are generally asymmetrical. This asymmetry is retained in the recombination reaction, ie the left hand side of one site is joined to the right hand side of the other and *vice versa*. Therefore recombination between sites in inverted repeat leads to inversion, whereas recombination between directly repeated sites leads to deletion. Recombination between sites on different DNA molecules can also occur and leads to fusion of the two molecules.

Site-specific recombination carries out a variety of different functions in bacterial systems. These include: the integration and excision of temperate bacteriophage into/from the host chromosome, the resolution of cointegrate structures formed by transposition of type II transposons, control of gene expression by inversion of small DNA segments, and plasmid-stabilising monomerisation (reviewed by Sadowski, 1986). Site-specific recombination also occurs naturally in a eukaryotic system: FLP mediated inversion in the *Saccharomyces cerevisiae* 2-micron plasmid provides a plasmid-stabilising amplification property (Volkert and Broach, 1986). Many site-specific recombination systems are being studied *in vitro*, following the lead of Nash (1975) with the lambda integrase system.

Site-specific recombination systems fall into two classes, identified both on the basis of sequence similarity amongst the recombinases and the mechanism of the recombination reaction. These two classes, the resolvase/invertase class and the Int class, are discussed separately below.

1.4 The resolvase/invertase class of recombinases

The members of this family can be further subdivided into the resolvases and the invertases. The resolvases are responsible for resolution of transposon cointegrate structures and are also involved in plasmid stabilising monomerisation (eg R46 *per*). The resolvases of Tn3 and gamma-delta are the best studied. These two enzymes and their recombination sites are interchangeable, and results obtained with one system are generally applicable to the other. The resolvases have been reviewed by Hatfull and Grindley (1988), Stark *et al* (1989b) and Sherratt (1989). The DNA invertases Gin and Cin invert segments of DNA to alter the expression of tail fibre genes in phage Mu and P1 respectively. Hin inverts a promoter containing chromosomal DNA segment to bring about antigenic phase variation in *S. typhimurium*. Pin-mediated inversion in the defective e14 viral element of the *E. coli* chromosome has an unknown function. The four invertases Gin, Hin, Cin and Pin and their recombination sites are all interchangeable (reviewed by Glasgow *et al* 1989).

The recombination site for Tn3 resolvase (*res*) is approximately 120 bp long and contains 3 dyad-symmetric sub-sites (I, II, III; see Fig 1.3) which are each thought to bind a resolvase dimer (Grindley *et al*, 1982; Kitts *et al*, 1983). Recombination occurs only at the centre of sub-site I (Reed, 1981), but the accessory sub-sites II and III are required for recombination (Grindley *et al*, 1982). The recombination sites for the DNA invertases are 26 bp long, have dyad symmetry and are probably bound by invertase dimers (Mertens *et al*, 1988; Fig 1.3). Strand cleavage occurs at the centre of these 26 bp sites (Klippel *et al*, 1988a).

Tn3 *res* sites are only recombined efficiently if they are present in direct repeat on the same, supercoiled DNA molecule (Kitts *et al*, 1983). The only protein required for recombination is resolvase. The invertase recombination sites, on the other hand, are only recombined efficiently if they are present in inverted repeat on a supercoiled DNA molecule. A copy of the

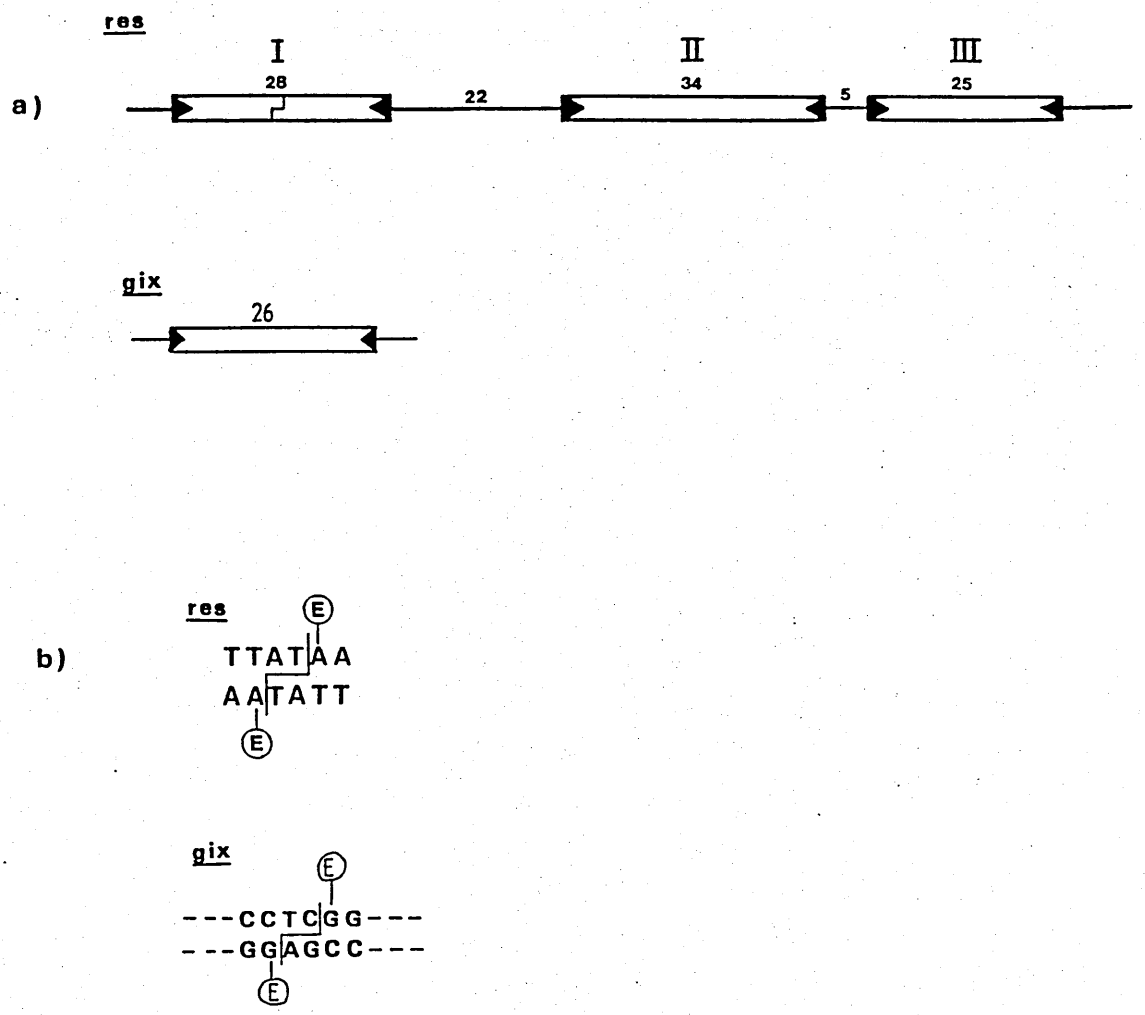


Figure 1.3 Structures of recombination sites required for Tn3/gamma-delta resolvases and the DNA invertases.
A) Resolvase and invertase binding sites are represented as open boxes with arrows indicating the imperfect dyad symmetry. Lengths of binding sites are given in arabic numbers.
B) Sequence of the centre of the invertase recombination site *gix* and sub-site I of *res*. The positions of cleavage by the recombinase proteins are shown as staggered lines. The position at which the recombinase becomes covalently attached are shown.

enhancer site (*sis*), on the same DNA molecule, is also required for efficient recombination (Kahmann *et al*, 1985). The host protein Fis, which binds to the enhancer site, is required along with the DNA invertase for efficient inversion (Kahmann *et al*, 1985).

The resolvases and invertases are small proteins (about 180 amino acids) with a high degree of sequence similarity (Fig 1.4). They contain two structural domains; a C-terminal domain of approximately 40 amino acids is largely responsible for sequence specific DNA binding activity (Abdel-Meguid *et al*, 1984), the N-terminal domain is thought to contain the catalytic site and to be responsible for protein - protein interactions.

Recombination reactions catalysed by resolvases and invertases are mechanistically very similar. Putative intermediates have been isolated which contain double strand breaks at the centre of the recombination site (Reed and Grindley, 1981; Klippel *et al* 1988a; Johnson and Bruist, 1989). These have a staggered 2 bp cut, with an overhanging 3'OH end and a recessed 5' end which is covalently attached to the recombinase protein by a phospho-serine linkage (Fig 1.3B; Reed and Moser, 1984; Klippel *et al*, 1988a). An absolutely conserved serine residue, near to the N-terminus of the resolvase/invertase proteins, appears to be the catalytic residue which becomes linked to the DNA (Hatful and Grindley, 1986; Klippel *et al*, 1988a).

The DNA invertase - and resolvase - catalysed recombination reactions take place with specific topologies (Fig 1.5). Invertase-catalysed recombination normally produces unknotted circles, with a linkage change of +4 (Kahmaan *et al*, 1987; Kanaar *et al*, 1988). Resolvase-catalysed deletion normally produces a specific (-2) catenane with a +4 linkage change (Wasserman and Cozzarelli, 1985; Boocock *et al*, 1987). It is proposed that strand exchange can take place only after the formation of specific inter-wrapped synaptic complexes. Once the synapse has formed, strand exchange probably occurs by concerted cutting, simple rotation through 180^{o} and re-ligation of the DNA (Fig 1.6).

The differences in selectivity and topology of the

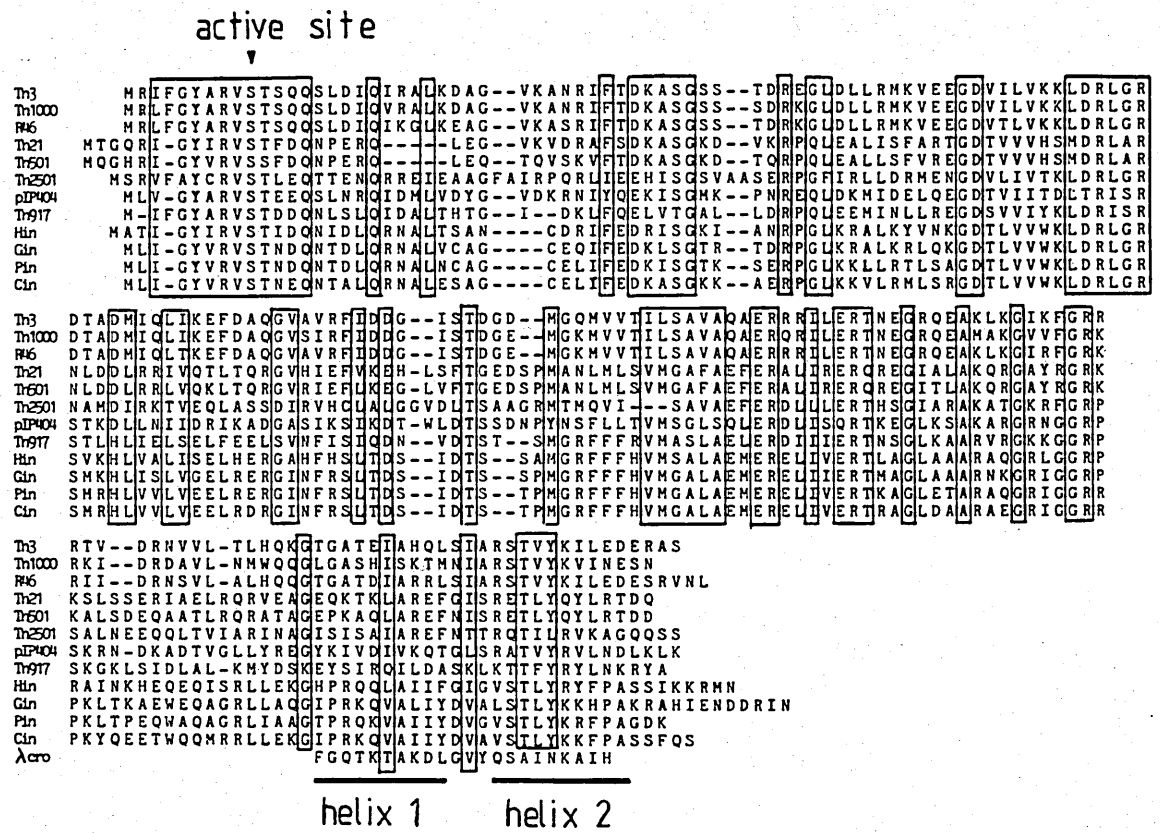


Figure 1.4 Alignment of predicted amino acid sequences of members of the resolvase and DNA invertase family of site-specific recombinases. The putative DNA binding helix-turn-helix motif is indicated, as is the presumed active site serine. (Taken from Sherratt, 1989)

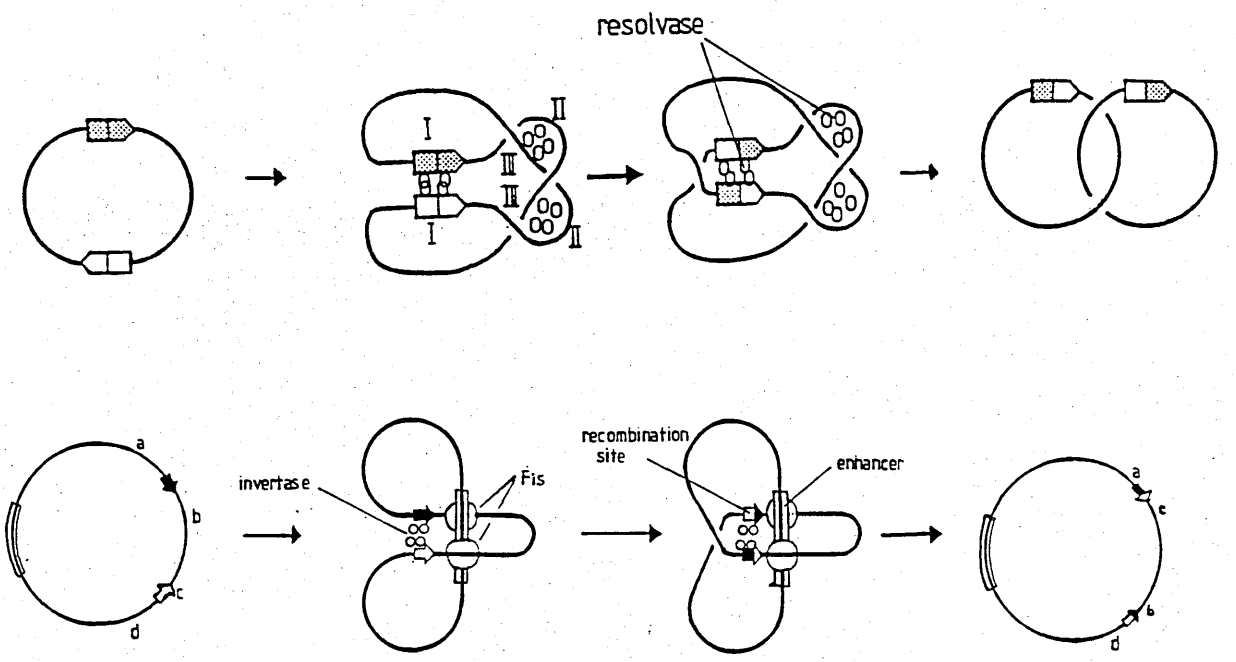


Figure 1.5 Topology of reactions catalysed by the resolvases and the DNA invertases. Resolution catalysed by resolvase goes *via* a -3 synapse and produces a simple (-2) catenane as shown. Inversion by the DNA invertases produces an unknotted circle and goes *via* a -2 synapse as shown. The -3 synapse between *res* sites may be stabilised by inter-wrapping of sub-sites II and III around resolvase. The -2 synapse in the inversion reactions may be stabilised by Fis/enhancer interactions as shown.

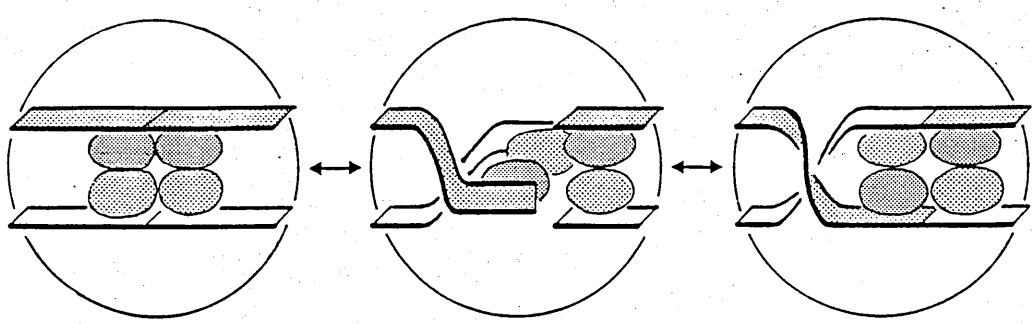


Figure 1.6 The "simple rotation" model for strand exchange. The DNA is drawn as a ribbon, and recombinase monomers are shown as stippled ellipses.

recombination reactions catalysed by resolvases and invertases can be accounted for by the different topologies required for strand exchange. Resolution reactions occur *via* a -3 synapse (Fig 1.5). This synapse is thought to be stabilised by the binding of resolvase to the inter-wrapped accessory sub-sites II and III (Fig 1.5; Stark *et al*, 1989a). The inversion reactions go *via* a -2 synapse, which is thought to be stabilised by interactions between the Fis protein and the enhancer site (Fig 1.5; Bruist *et al*, 1987; Kanaar *et al*, 1989).

1.5 The lambda integrase class of recombinases

The second category of site-specific recombinases are related to lambda integrase, and can be collectively termed the "Int" family. They have a large number of biological functions, ranging from phage integration and excision (eg lambda, Phi80, 186 Int proteins; see Thompson and Landy, 1989 for a review), regulation of *E. coli* fimbrial antigens by inverting a small promoter-containing DNA segment (FimB and FimE; Klemm, 1986; Dorman and Higgins, 1987), transposon cointegrate resolution (TnpI of Tn4430; Mahillon and Lereclus, 1988), plasmid stabilising inversion (FLP of the yeast 2 micron plasmid; Volkert and Broach, 1986), and plasmid stabilising monomerisation (Cre of P1; Austin *et al*, 1981; Abremski *et al*, 1983 and the D protein of plasmid F; Lane *et al*, 1986; O'Connor *et al*, 1986).

Recombination systems belonging to the Int class are distinguished from the resolvase/invertase systems by the following:

i) Proteins of the Int family contain two regions (domains 1 and 2) of amino acid similarity, although FLP has no match to domain 1 (Fig 5.10; Argos *et al*, 1986).

ii) The recombinase becomes covalently linked to a recessed 3' phosphate *via* a phospho-tyrosine linkage. The active site tyrosine is contained within the conserved domain 2 (Pargellis *et al*, 1988; Gronostajski and Sadowski, 1985a; Prasad *et al*, 1987; Wierzbicki *et al*, 1987).

iii) Recombination occurs *via* a Holliday junction intermediate (Pargellis *et al*, 1988; Jayaram *et al*, 1988; Hoess *et al*, 1987).

iv) The Int family appears to have somewhat less topological selectivity and specificity than the resolvases and invertases. In most systems, the recombinase can catalyse deletion, inversion and fusion reactions. (Pollock and Nash, 1983; Gronostajski and Sadowski, 1985c; Abremski *et al*, 1983). Lambda Int - and FLP - catalysed recombination appears to take place after random collision of recombination sites. Reactions catalysed by these enzymes yield multiply catenated deletion products and multiply knotted inversion products (Pollock and Nash, 1983; Beatty *et al*, 1986). Cre, on the other hand appears to use a reaction mechanism which excludes interdomainal supercoils from the synapse, and yields mainly free circles and simple catenanes as deletion products (Abremski *et al*, 1983).

The Int family of recombination systems can be divided into the minimal systems (eg FLP/FRT and Cre/*lox*), which have simple sites and require only the recombinase protein, and the complex systems (eg lambda and related phage integrases) which have complex sites and require accessory proteins as well as the recombinase.

1.5.1 Lambda integration and excision

Lambda integration and excision has been extensively studied both *in vitro* and *in vivo* (reviewed by Landy, 1989; Thompson and Landy, 1989). There are two pathways of recombination, one for phage integration and one for excision. The integrative pathway recombines *attP* (POP') with *attB* (BOB') to form *attL* (BOP') and *attR* (POB'), and requires the phage-encoded Int protein as well as the host-encoded IHF. The excisive pathway recombines *attL* with *attR* to reform *attP* and *attB*, but is not the reverse of integrative recombination. It requires Int, IHF and the phage-encoded Xis protein and is stimulated by the host-encoded Fis protein. By regulating the expression of Xis,

lambda can favour either integration or excision to suit its life cycle.

The *attP* site (240 bp) is complex, consisting of 13 different binding sites for the four proteins (Int, IHF, Xis and Fis) involved in lambda integration and excision (Fig 1.7; Bushman *et al*, 1984; Thompson *et al*, 1987). Int, IHF, Xis and Fis take part in various competitive and cooperative interactions in binding to their sites within *attP* (Bushman *et al*, 1984; Thompson *et al*, 1987).

Strand exchange occurs at two core-type Int binding sites in *attP*. These core-type sites are arranged in inverted repeat, flanking a 7 bp spacer. Int makes transient nicks at the boundaries of this spacer region, and becomes covalently linked to the 3' end of the DNA *via* a phospho-tyrosine linkage, leaving a free 5' protruding OH (Fig 1.7; Craig and Nash, 1983; Pargellis *et al*, 1988).

The *attB* site is comparatively simple, consisting of two core-type Int binding sites flanking a 7 bp spacer (Fig 1.7). The *attB* sequence is identical to the *attP* sequence for 15 bp in a region which includes the 7bp spacer. Lambda Int makes staggered nicks in *attB* in positions that correspond to the staggered nicks in *attP* (Fig 1.7; Craig and Nash, 1983).

Int binds to a second class of sequences within *attP*, the arm-type binding sites (Fig 1.7). The Int protein contains two separate domains responsible for binding to the two classes of site. Proteolytic cleavage of Int produces an N-terminal fragment of 7.5 kDa which binds specifically to the arm-type sites and a C-terminal proteolytic fragment of 32 kDa which binds to the core-type sites (Moitoso de Vargas *et al*, 1988). The C-terminal fragment retains some catalytic (topoisomerase and Holliday junction resolution) activity (Moitoso de Vargas *et al*, 1988).

Integrative recombination requires supercoiling of *attP* (Mizuuchi and Mizuuchi; 1979). Supercoiling of *attP* is thought to favour the assembly of an inter-wrapped nucleoprotein structure containing Int and IHF (Richet *et al*, 1986). This "intasome" is thought to capture an

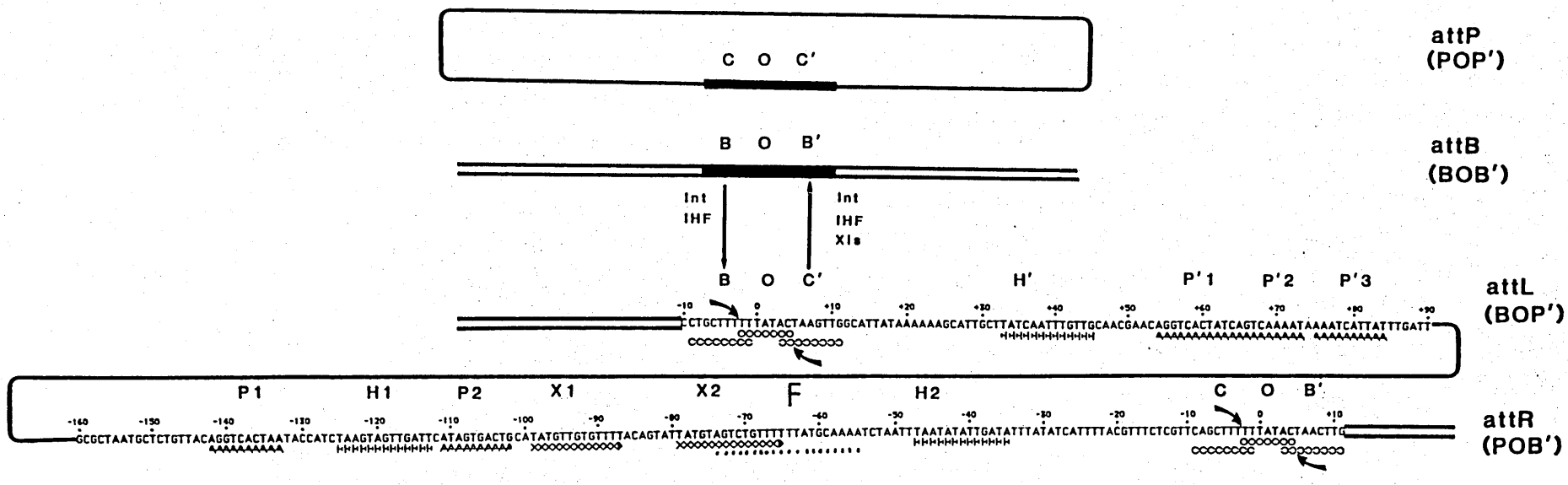


Figure 1.7 Structure of the lambda *att sites.* Recombination between *attP* and *attB* generates two hybrid sites *attL* and *attR*. The binding sites for IHF (H), Xis (X), Fis (F), arm-type Int (P) and core-type Int (C or B) are indicated. The curved arrows show the sites of strand exchange.

unbound *attB* site, after which recombination occurs (Richet *et al*, 1988). The binding of IHF to sites within *attP* bends the DNA substantially (Thompson and Landy, 1988) and is thought to help in the formation of the inter-wrapped intasome structure.

Excisive recombination does not require supercoiling and is stimulated by Fis and Xis. Xis and Fis are postulated to a play a similar role to IHF, helping to assemble specific nucleoprotein structures required for excisive recombination but inhibiting integrative recombination.

Int catalyses recombination *via* two independent strand exchange reactions (Fig 1.8). The first of these forms a Holliday structure which is then converted to complete recombinant by a second strand exchange. The first piece of evidence for this came from the observation of Hsu and Landy (1984) that lambda Int can resolve artificial Holliday junction intermediates. The presence of all four core-type Int binding sites is all that is required for this activity. Further evidence came from studying the reactions of suicide substrates with either a nick (Nunes-Duby *et al*, 1987) or a phosphorothioate (Kitts and Nash, 1988) in the 7 bp spacer region. Presumed recombination intermediates, containing Holliday junctions and covalently bound protein-DNA complexes, accumulate when these suicide substrates are incubated with Int. These experiments were used to show that strand exchange occurs in a specific order. First, the top strands (as drawn in Fig 1.8) are cut, exchanged and re-ligated. This is followed by Holliday junction branch migration, requiring homology in the spacer region. The reaction is completed by cleavage, exchange and re-ligation of the bottom strands. The order of these cleavages appears to be governed, not by the asymmetry of the core regions, but by the asymmetric arrangement of proteins bound to the P and P' arms (Kitts and Nash, 1988).

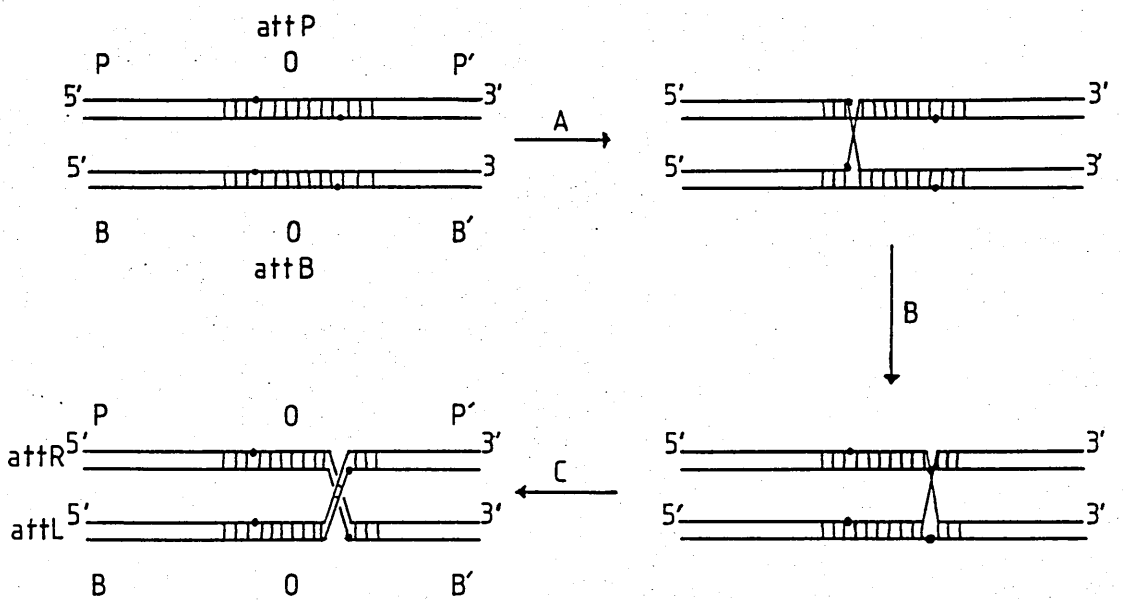


Figure 1.8 Proposed mechanism for lambda integration. Top strands are exchanged first (A), followed by Holliday junction branch migration (B) and bottom strand exchange (C) to yield recombinant.

1.5.2 FLP and Cre recombination

FLP and Cre catalyse recombination by a mechanism very similar to that used by lambda Int (see Cox, 1989 for a review of FLP and comparison to Cre). The conserved tyrosine of both these proteins appears to be involved in catalysis of recombination. An intermediate has been isolated, in which the recombination site DNA is covalently linked to the conserved (domain 2) tyrosine of FLP (Gronostajski and Sadowski, 1985a; Jayaram, pers. comm.). Mutation of the domain 2 tyrosine in Cre leads to a protein with wild type DNA binding activity but no recombinase activity (Wierzbicki *et al*, 1987).

The recombination reactions appear to go *via* Holliday junction intermediates (Jayaram *et al*, 1988; Hoess *et al*, 1987). Homology between the spacer regions of the two participating sites is required for efficient recombination (Senecoff and Cox, 1986; Hoess *et al*, 1986). This homology is probably required for branch migration of the Holliday junction across the spacer region.

The recombination sites for FLP and Cre are much simpler than the lambda *att* sites. The *lox* site of P1 contains two 13 bp Cre binding sites (corresponding to the core-type sites of lambda *att*) flanking an 8 bp spacer (Fig 1.9; Hoess and Abremski, 1985). Cre cleaves the *lox* site to leave a 6bp staggered cut, with a free protruding 5'OH and a recessed 3' phosphate covalently linked to the recombinase (Fig 1.9; Hoess and Abremski, 1985). The FRT site consists of two 13 bp FLP binding sites flanking an 8 bp spacer, but has an additional 13 bp FLP binding site located at one side (Fig 1.9; Andrews *et al*, 1985). Only the two inverted 13 bp repeats are absolutely required for FLP-mediated recombination (Gronostajski and Sadowski, 1985b). FLP cleaves the FRT site to leave an 8 bp staggered cut with a protruding 5'OH and a recessed 3' phosphate covalently linked to the recombinase (Andrews *et al*, 1987)

Neither FLP or Cre require supercoiling of their substrates, and both enzymes can carry out both

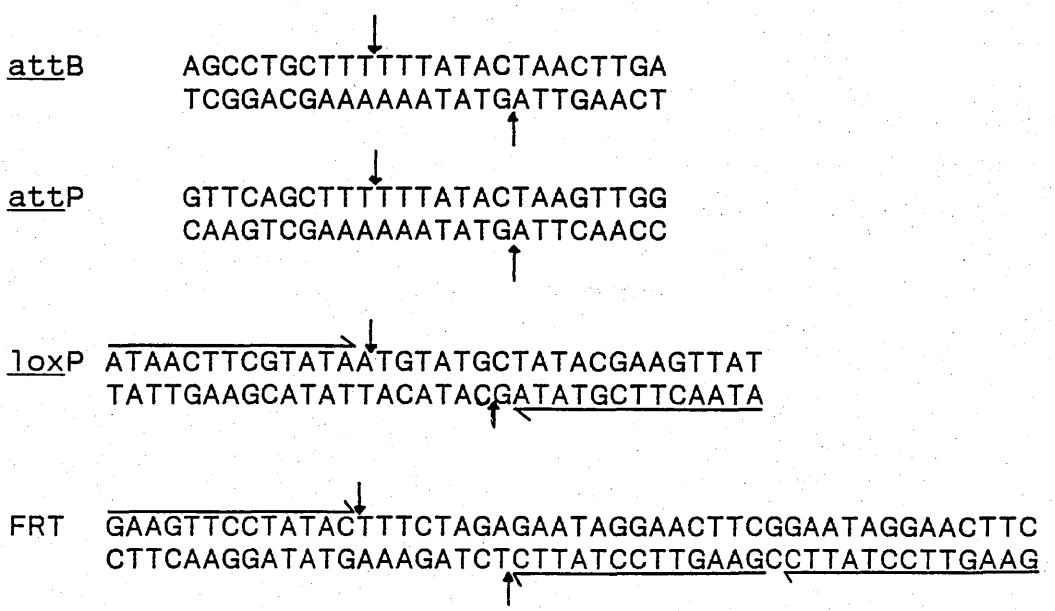


Figure 1.9 The sequences of crossover sites for lambda integrase (att sites), Cre (loxP) and FLP (FRT). Each site is cleaved on either side of a spacer region (core) by their respective recombinase. FRT and loxP contain inverted repeat sequences.

intermolecular and intramolecular (deletion and inversion) recombination events (Vetter *et al*, 1984; Abremski and Hoess, 1984).

1.6 The *xer* genes

The only ColE1 sequence required for *cer* site-specific recombination is the 250 bp *cer* site. The small size of *cer* and the absence of any coding region conserved between *cer* and related recombination sites suggested that *cer* does not encode a site-specific recombinase. This led to the proposal that host encoded functions (*xer* genes for chromosomally encoded recombinase) act at *cer* to bring about site-specific recombination (Summers *et al*, 1985).

A scheme was therefore devised to select for mutants defective in genes required for *cer* recombination (Stirling *et al*, 1988a). The selection procedure used the pseudo-dimeric plasmid pKS455, which contains two directly repeated copies of *cer* flanking a chloramphenicol resistance marker (*cat*). Site-specific recombination between the *cer* sites of pKS455 produces two circles, and effectively isolates the *cat* gene from the plasmid replication origin (see Fig 4.1). Thus, pKS455 transformants of wild-type *E. coli* strains are chloramphenicol sensitive. Mutagenised populations were transformed with pKS455 and *xer* mutants were selected by virtue of their resistance to chloramphenicol (Stirling, 1988a).

A total of 11 independent *xer* mutants were isolated from Tn5 mutagenised populations (Stewart, 1986; Stirling, 1987). This collection of mutants was used to clone the wild type *xer* genes by complementation. This revealed that the 11 mutants fell into two different complementation groups. Two mutants were defective in *xerA*, and the other nine mutants were defective in *xerB*. The *xerA* gene was mapped to 70.5 minutes (Stirling *et al*, 1988b) and the *xerB* gene was mapped to 96.5 minutes (Stirling *et al*, 1989) on the *E. coli* chromosome. Both genes have been over-expressed and their gene products purified to near homogeneity.

The nucleotide sequence of the *xerA* gene was determined and shown to be identical to the sequence of *argR*, encoding the arginine repressor (Stirling *et al*, 1988b; Lim *et al*, 1987). The arginine repressor (along with its co-repressor L-arginine) is a negative regulator of genes involved in arginine biosynthesis (see Glansdorf, 1987 for a review). ArgR binds to operator sites within the promoters of the arginine biosynthetic operons, repressing transcription. These operators each contain two copies of a loosely conserved (dyad symmetric) 18 bp Arg box. There are also two copies of the Arg box upstream of the *argR* gene, overlapping with one of the *argR* promoters. ArgR binds to these sequences *in vitro* (Lim *et al*, 1987; G. Szatmari, pers. comm.) and it seems likely that synthesis of *argR* is autoregulatory.

The *cer* site contains one copy of the 18 bp Arg box (Fig 1.2), to which ArgR has been shown to bind *in vitro* (Stirling *et al*, 1988b; G. Szatmari, pers. comm.). This binding is absolutely dependent on the presence of L-arginine. However, it is likely that intracellular levels of L-arginine are high enough under most physiological conditions for most cellular Arg boxes to be bound by ArgR. The ArgR binding site in *cer* is located approximately 100 bp from the crossover site in *cer* and overlaps with a promoter within *cer* (Fig 1.2). However, it seems unlikely that the requirement for ArgR in *cer* recombination simply reflects a need to repress transcription from this promoter. The binding of ArgR to this site in *cer* could be involved in the assembly of a higher order protein-DNA complex, necessary for normal *cer* site-specific recombination.

The predicted XerB amino acid sequence displays no sequence similarity to any known recombinase. At the time work on this thesis began, no insight had been gained into the role of XerB in *cer* recombination.

1.7 The *cer* site

The functional *cer* site was originally defined to a 282 bp HpaII - TaqI fragment from ColE1, coordinates 3687-

3969 on the sequence of Chan *et al* (1985) (Summers and Sherratt, 1984). Two directly repeated copies of this fragment are all that is required for efficient site-specific recombination. Many other plasmids contain regions with similar sequences that also serve as substrates for monomerisng site-specific recombination (Fig 1.2).

The crossover site of *cer* was initially mapped by sequencing the hybrid produced by recombination between ColE1 *cer* and the similar *ckr* site from ColK. This showed that recombination occurred within a 35 bp region close to the right hand end of *cer*, between ColE1 coordinates 3907 and 3942 (Summers *et al*, 1985). Sequence comparison of the related multimer resolution sites suggest that the right hand boundary of *cer* is probably at about coordinate 3940, where there is an abrupt sequence divergence between ColE1, pMB1 and ClODF13 (Fig 1.2).

Exonuclease III deletions into the 282 bp HpaII - TaqI fragment, from the HpaII end, were used to define the left hand boundary of the *cer* sequence. Deletion derivatives of *cer* were assayed for their ability to resolve plasmid dimers. Deletion as far as coordinate 3733 left a 233 bp site (D3733) which retained full *cer* function (Summers and Sherratt, 1988). Deletion of a further 4 bp (D3737) reduced the efficiency of dimer resolution and when a further 11 bp was removed (D3748) all dimer resolution activity was lost (Summers and Sherratt, 1988).

The *cer* sequence can be divided into three regions on the basis of sequence similarity to other *cer*-like sites (Fig 1.2). The region between coordinates 3733 and 3793 is poorly conserved among the *cer* related sites. The region between coordinates 3794 and 3906 is more highly conserved. The crossover region between 3907 and 3942 of *cer* is the most highly conserved. A recombination site from R1 shows sequence similarity only to the crossover region of *cer* (Fig 1.2; Clerget, 1984).

The least conserved region, between coordinates 3733 and 3793 is only required in one copy for *cer* recombination. Although D3748 did not recombine with

another copy of the same site, it did recombine with a full length *cer* site. Progressive deletions as far as coordinate 3793 gradually reduced the efficiency with which recombination between deleted and full length sites occurred (Summers and Sherratt, 1988). Deletion as far as coordinate 3800 produced a 169 bp site which did not recombine with a full length *cer* site (Summers and Sherratt, 1988).

Although the sequence of the left hand boundary region of *cer* is not well conserved, all of the *cer*-like sequences contain AT rich and GC rich sequences which alternate with a periodicity of approximately 10 bp. It has been shown that AT rich sequences preferentially lie with their minor grooves on the inside of bends and GC rich sequences preferentially lie with their minor grooves on the outside of bends on the nucleosome core (Drew and Travers, 1985). It has been suggested that the nucleotide sequence of the left hand boundary region of *cer* is not important *per se*, but that its ability to bend to follow a specific curved path is all that is important for *cer* recombination (Summers and Sherratt, 1988). In *cer* the function of the left hand boundary region can in part be performed by an unrelated sequence which fortuitously contains alternating AT rich and GC rich sequences in the correct position (Summers and Sherratt, 1988).

The region between coordinates 3794 and 3906 is highly conserved among the *cer*-like sites and is required in both recombining partners. This region contains a promoter which directs transcription towards the crossover region. Transcription from this promoter does not appear to be essential for *cer* recombination. A mutation which reduces transcription from this promoter by approximately 60 fold has no effect on *cer* recombination (Summers and Sherratt, 1989). The role of transcription in *cer* site-specific recombination and plasmid maintenance remains unclear.

Overlapping with the promoter is an 18 bp Arg box sequence, to which ArgR binds *in vitro*, in the presence of L-arginine (Stirling *et al*, 1988b). The D3800 deletion, which is the first deletion to abolish recombination with

a full length *cer* site, removes part of this Arg box, suggesting that the Arg box is required in both recombining partners. ArgR binding to the Arg box in *cer* may serve to bring two copies of *cer* together in a higher order synaptic complex.

In order to map the crossover site of *cer* more accurately, a plasmid was constructed that contained a copy of the ClODF13 *parB* site in direct repeat with a ColE1 *cer* site. Recombination occurred at low frequency to produce two classes of hybrid *cer-parB* site (Summers, 1989). Both classes of hybrid appeared to have been produced by a recombination event that took place within the previously defined crossover region (Fig 1.10).

The two classes of hybrid have quite different properties. The class I hybrid behaves exactly like a wild-type *cer* site, resolving multimers in a XerA and XerB dependent fashion (Summers, 1989).

The class II hybrid, though differing from the class I hybrid only by the deletion of 2 bp in the crossover region (Fig 1.10), supports both intermolecular and intramolecular recombination (Summers, 1989). Plasmids containing the type II hybrid rapidly form a mixture of multimeric forms in *E. coli* (Summers, 1989). Furthermore, recombination at the type II hybrid does not require the products of either the *xerA* or the *xerB* genes. All sequences necessary for *xerA*, *xerB* independent recombination at the type II hybrid lie within a 50 bp MluI - TaqI fragment (Summers, 1989). This fragment contains the previously defined crossover region of *cer*, but excludes the flexible left hand region and the ArgR binding site. In the presence of ArgR and XerB, the full length type II hybrid retains some bias towards intramolecular recombination. However, if either *xerA* or *xerB* is defective, or if the left hand (Arg box-containing) region of *cer* is missing, this bias towards intramolecular recombination is removed. It appears that ArgR, XerB and the upstream region of *cer* act together to impose intramolecular directionality on *cer* recombination.

Because the type II hybrid can recombine in the absence of ArgR and XerB, and because neither XerB or ArgR

Figure 1.10 Nucleotide sequences of the hybrids produced by recombination between ColE1 *cer* and ClodF13 *parB*. The top of the figure shows an alignment of the crossover regions of *cer* and *parB*. Asterisks (*) denote sequence identity. The points at which strand exchange could have occurred to produce the type I and type II hybrids are indicated. A possible inverted repeat in the crossover region is shown. The bottom of the figure shows the nucleotide sequence of the crossover regions of the type I and type II hybrids. The type I hybrid acts *in vivo* like wild-type *cer*, requiring the full length site and favouring intramolecular recombination. The type II hybrid requires at most the Sau3A - TaqI fragment, can recombine in the absence of ArgR and XerB, and carries out both intermolecular and intramolecular recombination.

show amino acid similarity to any known recombinase, it seemed likely that neither was the *cer* recombinase. It therefore seemed probable that at least one more gene required for site-specific recombination at *cer* (encoding the recombinase) remained to be discovered.

Project aims

The aims of the project were:

1) To characterise the cloned *xerB* gene further and investigate its function.

2) To isolate further *xer* mutants with mutations in new *xer* genes, in the hope that the recombinase gene could be identified.

3) To isolate and characterise any new *xer* gene(s).

CHAPTER TWO

MATERIALS AND METHODS

Table 2.1 Bacterial Strains

Strain	Genotype	Source/reference
AB1157	thr-1, leuB6, hisG4, thi-1, ara-14, △(gpt-proA)62, argE3, galK2, supE44, xyl-5, mtl-1, tsx-33 lacY1, rpsL31	Bachmann 1972
DS902 (AB2463)	AB1157, but recA13	D. Sherratt
DS903	AB1157, but recF143	D. Sherratt
DS941	DS903, but lacZ△M15, $lacI^q$	D. Sherratt
DS947	DS903, but △lacI-lacZ, sup^o	D. Sherratt
DS953	DS941, but sup^o	D. Sherratt
DS956	DS941, but xerA9 (argR::fol)	D. Sherratt
JM101	supE, thi, △(lac,proA,B), F'(traD36, pro A,B, lacZ△M15 $lacI^q$)	Yanisch-Perron et al, 1985
TG1	JM101, but hsdD5	J. Gibson
CSH50 (DS887)	ara, △(lac-pro), thi	R. Kahmann
JC7623	AB1157, but recBC, sbcBC	C. Richaud
CGSC4311 (KL226)	relA1, tonA22, pit-10, spoT1, ompF627, Hfr PO2A of Cavalli	B. Bachmann
CSX3	DS903, but xerA3 (argR::Tn5)	C. Stirling
CSX17	DS941, but xerB1 (pepA::Tn5)	C. Stirling
HOM38b	DS941, but pepA7 (spontaneous Val-Leu-NH_2 resistant mutant)	H. O'Mara
PS6 (DS981)	DS941, but xerC2 (xerC::Km^r)	P. Sykora
SC1	DS941, but xerC1	This work
SC2	CGSC4311, but xerC1	This work
SC3	JC7623, but dapF Y30	C. Richaud
SC4	JC7623, but orf235 Y13	C. Richaud
SC5	JC7623, but xerC Y17	C. Richaud
SC6	JC7623, but orf238 Y2	C. Richaud
SC7	DS941, but dapF Y30	This work
SC8	DS941, but orf235 Y13	"
SC9	DS941, but xerC Y17	"
SC10	DS941, but orf238 Y2	"
SC11	CSH50, but xerC Y17	"
SC12	CSH50, but △fimB-fimH	I. Bloomfield
SC13	SC12, but xerC Y17	This work

Table 2.2 Plasmids

Plasmid	Description	Resistance	Source/Reference
pBR322	Vector derived from pMB1	Ap,Tc	Sutcliffe 1978
pBR325	" " " "	Ap,Cm,Tet	Bolivar *et al* 1978
pUC8	Vector derived from pBR322	Ap	Vieira & Messing 1982
pUC9	" " " "	Ap	Vieira & Messing 1982
pUC18	" " " "	Ap	Yanisch-Perron *et al* 1985
pUC19	" " " "	Ap	Yanisch-Perron *et al* 1985
pUC4K	Km^r block in pBR322 derived vector	Ap,Km	Vieira & Messing 1982
pTZ18R	Vector derived from pBR322	Ap	Vieira & Messing 1987
pTZ19R	" " " "	Ap	Vieira & Messing 1987
pBAD	P_{tac} expression vector derived from pKK223-3	Ap	A. C. Boyd
pACYC184	Vector derived from p15A	Cm,Tet	Chang & Cohen 1978
pCB104	Lambda *dv*-based vector	Cm	A. C. Boyd
pCS80	pBR322::Tn5	Ap,Tet,Km	Stirling 1987
pCS100	pUC8 + rop^+ ColE1 fragment in NarI site	Ap	"
pCS110	pCS100 + *xerB* as a 3 kbp BamHI fragment	Ap	"
pCS111	pUC18 + *xerB* as a 2.8 kbp SphI fragment	Ap	"
pCS112	pUC18 + *xerB* as a 1.9 kbp HindIII fragment	Ap	"
pCS202	Lambda *dv*-based 2-*cer* reporter plasmid deletion marker = Tet^r	Cm,Tet	"
pCS210	pACYC184-based 2-*cer* reporter plasmid deletion marker = $LacZ^+$	Tet	"
pKS455	pUC9-based 2-*cer* reporter plasmid deletion marker = Cm^r	Ap,Cm	D. K. Summers
pKS492	pUC18 + 280 bp HpaII-TaqI *cer* fragment	Ap	"
pKS493	pUC18 + 280 bp HpaII-TaqI *cer* fragment	Ap	"
pGS38	pUC19 + 920 bp SphI-AccI $xerA^+$ fragment	Ap	Stewart 1986

pMAK101	pBR325 + 8.3 kbp EcoRI *xerC*$^+$ fragment	Ap	Aldea *et al* 1988
pMAK102	pBR325 + 5.9 kbp SalI *xerC*$^-$ fragment	Ap	"
pVMK42	pBR325 + partial SalI-PvuII *xerC*$^-$ fragment	Ap	"
pSDC100	pBR322 + 14.1 kbp EcoR1 *xerC1*::Tn5 fragment	Ap	This work
pSDC101	pTZ18R + 3.8 kbp HindIII-BglII *xerC*$^+$ fragment	Ap	"
pSDC102	pTZ18R + 3.8 kbp HindIII-BglII *xerC*$^+$ fragment from pMAK101 (opposite orientation from pSDC101)	Ap	"
pSDC103	pTZ19R + 1.3 kbp BglII fragment from pSDC100	Ap	"
pSDC104	pSDC102 delta 4.1	Ap	"
pSDC105	pBAD + *xerC*$^+$ HindIII-EcoRI fragment from pSDC104	Ap	"
pSDC106	pTZ18R + 1.8 kbp SalI fragment from pMAK101	Ap	"
pSDC107	pCB104 + *xerC*$^+$ HindIII-EcoRI fragment from pSDC104	Cm	"
pSDC109	pBR322 + type II hybrid at ClaI site	Ap (Tet^S)	"
pSDC110	pSDC109 + type II hybrid at PvuII site deletion substrate	Ap	"
pSDC111	pSDC109 + type II hybrid at PvuII site inversion substrate	Ap	"
pSDC112	pSDC107 + Km^r gene from pUC4K	Cm,Km	"
pSDC113	pBR322 + *cer* from pKS492 inserted at EcoRI-HindIII sites	Ap (Tet^S)	"
pSDC115	pSDC113 + *cer* at PvuII site (deletes to give pSDC116)	Ap	"
pSDC117	pSDC113 + *cer* at PvuII site (inverts to give pSDC118)	Ap	"
pSDC120	pBAD + 1.2 kbp *orf238*$^+$ NdeI fragment	Ap	"
pSDC121	pBAD + 1.1 kbp *orf235*$^+$ AccI fragment	Ap	"
pSDC123	pSDC102 delta 6.2		

Table 2.3 Bacteriophage

Phage	Description	Source/Reference
$P1_{kc}$	Generalised transducing phage	M. Masters
λ NK467	Suicide vector for Tn5 mutagenesis	N. Kleckner
M13mp18	Cloning vector derived from M13	Yanisch-Perron et al 1985
M13mp19	Cloning vector derived from M13	" " " " "
mSDC1	M13mp18 + HindIII fragment from pCS110	This work Ch. 5
mSDC2	M13mp18 + HindIII fragment from pCS110	" "
mSDC3	M13mp18 + BamHI-EcoRI fragment from pCS110	" "
mSDC4	M13mp18 + HindIII-BamHI fragment from pCS110	" "
mSDC12	M13mp19 + 1.2 kbp BglII fragment from pSDC100	" "
mSDC13	M13mp18 + HindIII-EcorI fragment from pSDC103	" "
mSDC14	M13mp19 + HindIII-EcoRI fragment from pSDC103	" "
mSDC15	M13mp19 + 389 bp BglII fragment from pVMK42	" "
mSDC16	M13mp19 + 389 bp BglII fragment from pVMK42	" "
mSDC17	M13mp19 + HindIII-BamHI fragment from pDF3 Y13	" "
mSDC18	M13mp19 + HindIII-BamHI fragment from pDF3 Y17	" "
mSDC19	M13mp19 + HindIII-BamHI fragment from pDF3 Y2	" "
mAS1	M13mp18 + 252 bp PstI fragment from pSDC102	A. Sutcliffe
mAS2	M13mp18 + 252 bp PstI fragment from pSDC102	A. Sutcliffe
mMB19.2	M13mp19 + PstI-Sau3A and Sau3A fragments from pSDC102	M.Burke
mMB19.3	M13mp19 + PstI-Sau3A and Sau3A fragments from pSDC102	M. Burke

2.1 Bacterial strains. The derivatives of *Esherichia coli* K-12 used are listed in Table 2.1. The derivatives of *Salmonella typhimurium* LT-2 used are listed in table 3.1.

2.2 Plasmids. The plasmids used and constructed in this study are listed in Table 2.2.

2.3 Bacteriophage. The bacteriophage used and constructed in this study are listed in table 2.3.

2.4 Chemicals.

CHEMICALS	SOURCE
General chemicals, biochemicals and organic solvents	BDH, May and Baker, Sigma
Media	Difco, Oxoid
Agarose	BRL
X-gal, IPTG	BRL
Radiochemicals	NEN
10 x restriction enzyme buffer	BRL, Boehringer Mannheim
Nucleotides	Boehringer Mannheim
Antibiotics	Sigma

2.5 Proteins.

Restriction and DNA modification enzymes	BRL, Boehringer Mannheim, Pharmacia

2.6 Culture media.

L Broth: 10 g tryptone, 5 g yeast extract, 5 g NaCl, made up to 1 litre in distilled water and adjusted to pH7.5 with NaOH.

L Agar: as L Broth with the addition of 15 g/l agar.

4 x Davis and Mingioli minimal salts (D&M salts): 28 g K_2HPO_4, 8 g KH_2PO_4, 1 g sodium citrate, 0.4 g $MgSO_4H_2O$, made up to 1 litre in distilled water.

Minimal agar: 25 ml D&M salts, 75 ml 2% agar in distilled water; supplemented with 0.2% glucose, 20 ug/ml thiamine (vitamin B1) and amino acids as necessary.

Phage buffer: 7 g Na_2HPO_4, 3 g KH_2PO_4, 5 g NaCl, 0.25 g $MgSO_4$, 15 mg $CaCl_2.2H_2O$ and 1 ml 1% gelatin made up to 1 litre in distilled water.

2.7 **Sterilisation.** All growth media were sterilised at 120°C for 15 minutes; supplements and buffer solutions at 108°C for 10 minutes and $CaCl_2$ at 114°C for 10 minutes.

2.8 Buffer solutions.

Electrophoresis

10 x E buffer: 242g Tris, 82g sodium acetate, 18.6g $Na_2EDTA.2H_2O$, made up to 5 litres in tap water, adjusted to pH 8.2 with glacial acetic acid.

10 x TBE buffer: 109g Tris, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$, made up to 1 litre in distilled water; pH is 8.3. (For sequencing gels, an alternative was used: 121.1g Tris, 55g boric acid, 9.3g $Na_2EDTA.2H_2O$, made up to 1 litre in distilled water)

Single colony gel loading buffer: 2% ficoll, 1% SDS, 0.01% bromophenol blue, 0.01% orange G in buffer E.

Polyacrylamide gel loading buffer: 1% ficoll, 0.1% SDS, 0.02% orange G, 0.01% bromophenol blue in distilled water.

4 x Horizontal agarose gel loading buffer: 25% sucrose, and 0.01% bromophenol blue in distilled water. SDS (0.5%) and/or protease K (0.2 mg/ml) were added when necessary.

Protein gel running buffer: 144 g Glycine, 30 g Tris base made up to 1 litre. After dilution, 10 ml of 10% SDS was added to every 1 litre of running buffer.

Protein sample buffer: 5% SDS, 50% Glycerol, 0.01% bromophenol blue and 50mM Tris-HCl (pH6.8) made up to 5% beta-mercaptoethanol immediately prior to use.

Buffers for DNA work

10 x restriction buffers: as supplied with restriction enzymes.

5 x ligation buffer: supplied by BRL.

1 x TE buffer: 10mM Tris/HCl, 1mM EDTA; pH 8.0.

20 x SSC: 3M NaCl, 300mM trisodium citrate.

EDTA: 200mM solutions of EDTA were made up and titrated with NaOH to a pH of 8.0.

10 x S1 buffer: 1.1 ml of 3M KOAc (pH 4.6), 5 ml 5M NaCl, 5 ml glycerol, 30 mg $ZnSO_4$.

10 x Nick translation buffer: 0.5 M Tris-HCl (pH7.5), 50 mM $MgCl_2$, 100 mM beta-mercaptoethanol, 250 ug/ml BSA.

50 x Denhardts solution: 5 g Ficoll, 5 g polyvinyl-pyrrolidone, 5 g BSA made up to 500 ml with H_2O.

Phenol: All phenol used in the purification of DNA contained 0.1% 8-hydroxyquinoline and was buffered against 0.5M Tris-HCl pH8.0.

2.9 Antibiotics. The antibiotic concentrations used throughout for both liquid and plate selection were as follows:

Antibiotic	Stock solution	Selective concentration
Ampicillin (Ap)	5 mg/ml (water)	50 ug/ml
Tetracycline (Tc)	1 mg/ml (10mM HCl)	10 ug/ml
Chloramphenicol (Cm)	2.5 mg/ml (ethanol)	25 ug/ml
Kanamycin (Km)	2.5 mg/ml (water)	25 ug/ml
Streptomycin (Str)	10 mg/ml (water)	100 ug/ml

All stock solutions were stored at 4^{o}C.
Antibiotics were added to molten agar which was precooled to 55^{o}C.

2.10 Indicators. X-gal (5-bromo-4-chloro-3-indolyl-B-D-galactoside) was used in conjunction with the host strains DS941 and JM101 and the pUC and M13 vectors, providing a screen for plasmids with inserts in the polylinker region. Clones containing inserts were generally white; clones lacking inserts were blue. X-gal (40mg/ml in DMF) was stored at -20^{o}C and added to L agar to a final concentration of 20ug/ml. IPTG (isopropyly-B-D-thiogalactoside) was added to a final concentration of 25 ug/ml to medium containing X-gal.

2.11 Growth conditions. Liquid culture for transformation, DNA preparations or in vivo recombination assays were routinely grown in L broth at 37^{o}C with vigorous shaking. Both L agar and minimal solid media were used. Antibiotics were used as required. Plates were generally incubated overnight at 37^{o}C.
Bacterial strains were stored in 50% L broth, 20% glycerol and 1% peptone at -20^{o}C or on L agar slopes at room temperature.

2.12 Plasmid DNA isolation.

Large scale DNA preparation (Birnboim and Doly, 1976; as modified in this laboratory).

Solutions:

I. 50mM glucose, 25mM Tris/HCl pH 8.0, 10mM EDTA.

II. 0.2M NaOH, 1% SDS; made fresh.

III. 5M potassium acetate pH 4.8; mix equal volumes of 3M CH_3COOK and 2M CH_3COOH, pH will be 4.8.

200ml cultures of stationary phase plasmid-containing cells were harvested by centrifugation (12,400g, 10 min at 4°C). The pellet was resuspended in 4 ml of solution I. 8 ml of solution II was added and the mixture was left on ice for a further 5 min. 6 ml of solution III was then added, gently mixed and the cell debris and chromosomal DNA removed by centrifugation (39,200g, 30 min at 4°C). The plasmid DNA was precipitated from the supernatant with 12 ml isopropanol for 15 minutes at room temperature. The DNA was pelleted at 27,200g for 15 min at 20°C, rinsed with 70% ethanol and then further purified by banding on a CsCl/EtBr gradient. The DNA was resuspended in 2.09 ml of TE buffer and added to 270 ul of a 15 mg/ml ethidium bromide solution. 5 g of CsCl were dissolved in 3 ml of water and added to the DNA/EtBr solution. The gradients were centrifuged in a Beckman Ti70 fixed angle rotor at 200,000g for 16 hours at 25°C. Where two bands were visible, a lower supercoiled plasmid band and an upper nicked DNA band, the lower band was removed using a 1ml syringe. The ethidium bromide was removed by repeated butanol extractions and the DNA was dialysed twice against one litre of 1 x TE buffer to remove the CsCl.

Mini preparation of DNA (modified from from Holmes and Quigley, 1981):

STET buffer: 8% sucrose, 5% Triton x-100, 50mM EDTA, 50mM Tris/HCl pH 8.0

3.0 ml of a stationary culture was harvested by centrifugation (12.100g, 30sec) and resuspended in 350 ul STET buffer in an Eppendorf tube. 25 ul of a 10 mg/ml lysozyme solution was then added and after mixing, the suspension was boiled for 45 sec and centrifuged in an Eppendorf microfuge for 15 min at 4°C. The pellet was discarded with a toothpick. The DNA was precipitated by the addition of 40 ul of 3M NaOAc and 400 ul of isopropanol. After microcentrifugation for 15 min, the pellet was washed in 70% ethanol, dried and resuspended in 50 ul TE buffer. 5-10 ul of this DNA was suitable for restriction enzyme digests. 1 ul of a 1 mg/ml RNase A solution was added to the DNA with the gel loading buffer.

2.13 **Chromosomal DNA isolation.** 5 ml of stationary phase culture was harvested by centrifugation (12,000g, 5 min, 4°C) and resuspended in 700 ul 10 x TE. 50 ul 10% SDS and 200 ul RNAase A (1 mg/ml) was added followed by incubation for 2 hours at 37°C. 100 ul of proteinase K (20 mg/ml) was added followed by a further 2 hours at 37°C. The DNA was then purified by two phenol extractions, two phenol/ chloroform extractions and two chloroform extractions. The DNA was precipitated at 4°C, by the addition of two volumes of ethanol, washed in 70% ethanol and resuspended in 200 ul of 1 x TE. This DNA was suitable for restriction enzyme digestion.

2.14 **Transformation with plasmid DNA.** Genetic transformation introduced plasmid DNA into different host strains. An overnight culture of the recipient strain was diluted 1 in 100 into 20 ml L broth and grown to a density of $2x10^8$ cells/ml (about 90 min). The cells were harvested (12,100g, 1 min, 4°C) and resuspended in 10 ml of 50mM $CaCl_2$. The cells were pelleted again, resuspended in

0.5 ml cold (4°C) 50mM $CaCl_2$ and kept on ice until use. 100 ul aliquots of competent cells were added to DNA in TE buffer and, after gentle mixing, were left on ice for 15 min. The cells were then heat shocked (2 min, 42°C or 5 min, 37°C) and returned to ice for 15 min. 100 ul of L broth was added to the cell suspension and incubated at 37°C for 90 min to allow expression of plasmid genes. Transformation to ampicillin resistance did not require this expression time. 100ul aliquots of the transformation mixture were spread onto selective plates.

2.15 Transfection with M13 DNA. This procedure was carried out exactly as above using either the bacterial strain JM101 or TG1. After the final incubation on ice, the cells were mixed with 200 ul of fresh log phase JM101 or TG1. This mixture was then added to 3 ml of soft molten agar already containing 30 ul of IPTG (15 mg/ml) and 30 ul X-gal (40 mg/ml) at 45°C. The cells were then plated out onto D&M minimal medium plates containing thiamine and glucose.

2.16 Preparation of P1 lysates. The donor strain was grown up in L broth to late log phase. The cells from 100 ul of culture were harvested by centrifugation and resuspended in 100 ul fresh L broth, 100 ul $CaCl_2$ (50mM) and 100 ul $MgCl_2$ (100mM). Sufficient P1 was added to give a m.o.i. of approximately 0.001. The mixture was then incubated at 37°C for 25 min, mixed with precooled (45°C) molten soft agar and plated onto a fresh undried L agar plate. After overnight incubation, 2.5 ml of phage buffer was added and left at room temperature for 15 min. The buffer and soft agar were placed in a centrifuge tube with a few drops of chloroform and vortexed for 30 sec. The tube was left for 30 min at room temperature, re-vortexed and then centrifuged at 12,000g for 10 min. The supernatant, containing approximately 10^9 pfu/ml, was stored at 4°C.

2.17 P1 transduction. The recipient strain was grown up to mid-log phase in L broth, harvested and resuspended in an equal volume of L broth. To 100 ul of cells was added 100

ul of $CaCl_2$ (50mM), 100 ul $MgCl_2$ (100mM) and 0.1-10 ul of P1 lysate. After 20 min incubation at 37°C, phage infection was stopped by the addition of 200 ul of 1M sodium citrate (filter sterilised). 500 ul of L broth was added followed by incubation for one hour at 37°C. The cells were then plated onto a suitable selective medium. All transductants were streaked out to single colony at least two times, to remove any contaminating P1 bacteriophage.

2.18 Single colony gel analysis. By using this technique, the plasmid content of an isolate can be observed without the need to purify the DNA. A single transformant was patched out (1cm square) on a selective plate and grown overnight. Using a toothpick, a large scrape of cells was collected and resuspended in 100 ul of single colony gel buffer. The cells were left to lyse at room temperature for 15 minutes. Cell debris and chromosomal DNA was spun down in an Eppendorf microfuge for 15 min at 4°C). 30 ul of the supernatant was loaded onto an agarose gel.

2.19 Ethanol precipitation of DNA. To the DNA solution was added one tenth volume of 3M NaOAc and 2 volumes of absolute ethanol. After mixing, the DNA was precipitated at -20°C for at least 30 mins and pelleted for 15 min at 4°C. The pellet was washed with 70% ethanol and dried.

2.20 Restriction of DNA. Restriction enzyme digests were usually performed in a total volume of 20 ul, containing 0.3-1.0 ug DNA and 2 ul of 10 x restriction buffer. 2-3 units of restriction enzyme was added per ug of DNA. The reactions were incubated at 37°C (or at the appropriate temperature) for 1-2 hours. For restrictions of larger quantities of DNA, the volume was scaled up accordingly. The enzymes were inactivated by the addition of loading buffer, heating to 70°C or by phenol extraction and ethanol precipitation if subsequent manipulations were necessary.

2.21 Calf intestinal phosphatase (CIP) treatment. Phosphatase was used to remove 5' phosphate groups from linearised vector to prevent recircularisation of the vector (thus increasing cloning efficiency). 1 unit of CIP was added directly to the digest and incubated at 37^{o}C for 15 min (or 45 min for blunt ends).

2.22 Ligation of DNA fragments. Restriction fragments were ligated in volumes of 10-20 ul, containing 1 x ligation buffer and 0.5 units of DNA T4 DNA ligase. Generally a 3:1 insert to vector ratio of fragments was used (10:1 for blunt end ligations). The reactions were incubated at 16^{o}C overnight. Aliquots of the ligation mix were used to transform competent cells.

2.23 Gel electrophoresis of DNA.

Agarose gels: 0.7-1.2% agarose gels were used.
Agarose powder was dissolved at 100^{o}C in 125 or 200ml buffer E and precooled to 55^{o}C prior to use. Horizontal gels were used to analyse restriction digests and for single colony gel analysis. Gels were usually run for 15-18 hours at 1.5V/cm in gel tanks containing 3 litres buffer E and then stained in 0.6 ug/ml ethidium bromide. The DNA was visualised on a 254 nm wavelength UV transilluminator.

Polyacrylamide gels: three types were used.
1) Polyacrylamide restriction gels.
Vertical gel kits were used with 1.5 mm spacers. The gel apparatus was sealed with 0.6% agarose in H_2O. An appropriate acrylamide gel mix was poured between the plates (with the insertion of a well former) and allowed to set for 60 min.

30ml acrylamide gel consists of:-

30% acrylamide: 0.8% bisacrylamide (w/v)	X ml (for X% gel)
10 x TBE	3 ml
H_2O	27-X ml
10% APS (w/v)	360 ul
TEMED	18 ul

The gels were run at room temperature in 1 x TBE at a constant current (25-30 mA), for 2-3 hours. DNA bands were visualised under 254nm UV illumination after staining in 0.6 ug/ml ethidium bromide for 10 min.

2) Non-denaturing polyacrylamide gels. These gels were used to separate protein:DNA complexes.
6% acrylamide 50mM Tris/Glycine pH 9.4, 0.1mM EDTA gels were usually used. Vertical gel kits were sealed with 0.6% agarose in distilled water.

30 ml of a 6% 50mM TE (pH 9.4) gel mixture contained:

30% acrylamide: 0.8% bisacrylamide (w/v)	6 ml
0.5M Tris/Glycine pH 9.4	3 ml
H_2O	20.5 ml
0.2M EDTA (pH 8.0)	15 ul

To each gel, the following reagents were added:-

10% APS (w/v)	150 ul
TEMED	15 ul

The gels were pre-run in 50mM Tris/Glycine 0.1mM EDTA at 15V/cm (30-90 min; 4^{o}C). After loading, the gels were run for 2-3 hours at 15V/cm, 4^{o}C. When the gel run was complete the gel was transferred to filter paper and dried under vacuum. Bands were visualised by autoradiography of a sheet of Kodak S1 or Fuji RX100 X-ray film for 1-3 days.

3) Polyacrylamide sequencing gels. Sequencing reactions were run on 8% high resolution polyacrylamide / urea gels as described in the 'M13 cloning / dideoxy sequencing instruction manual' published by BRL. Gels were pre-run for 45 min and run for 2-6 hours at 40W. Samples were

denatured prior to loading (80°C, 2 min). After the gel run was complete, the gel was fixed in 10% acetic acid 10% methanol for 15 min, dried under vacuum and autoradiographed.

2.24 Photography of gels. After staining in ethidium bromide, gels visualised by 254nm UV illumination were photographed using Polaroid type 67 land film or using a Pentax 35mm SRL loaded with Ilford HP5 film. Both cameras were fitted with a Kodak Wratten filter No.23A.

2.25 Electrophoresis of proteins in SDS-polyacrylamide gels. The electrophoresis of proteins followed the procedure of Laemmli (1970) using 15% running gels and 4% stacking gels. Gels were poured in vertical gel kits sealed with 0.6% agarose in H_2O. Gel mixes were made up according to the table below:

	Running Gel	Stacking Gel
30 / 0.8% acrylamide/bis	15 ml	2.0 ml
1.5M Tris-HCl pH8.8	7.5 ml	-
0.5M Tris-HCl pH6.8	-	3.75 ml
10% SDS	0.3 ml	0.15 ml
H_2O	to 30 ml	to 15 ml
TEMED	10 ul	10 ul
Ammonium persulphate (10% w/v)	150 ul	150 ul

The running gel was poured and left to set with a layer of isopropanol over it. The isopropanol was washed off and the stacking gel was poured. After the gel had set, running buffer was poured into the gel tank, and the comb was removed. Prior to loading, the protein samples were mixed with protein sample buffer and boiled for 5 minutes. The gel was run overnight at a constant current of 8 mA, leaving the dye front just above the bottom of the gel.

2.26 Nick Translation of DNA. 0.5 ug of plasmid DNA was mixed with: 5 ul 10 x nick translation buffer

20uCi alpha <^{32}P> dATP

unlabelled dCTP, dGTP and dTTP to 1mM final

1 ul 10mM $MgCl_2$

H_2O to a total volume of 48 ul.

One unit of DNA polymerase I and 3 ul of a 1 ug/ml solution of DNAase I was then added. After one hour at 16^{o}C the labelling reaction was stopped by the addition of 5 ul 200 mM EDTA. The labelled DNA was separated from unincorporated nucleotides on a Sephadex G-50 column and used to probe Southern blots.

2.27 Southern blotting and hybridisation. Agarose gels were soaked in 0.25M HCl for 15 min and then transferred to a solution of 1.5M NaCl, 0.5M NaOH for 30 min. The DNA was then transferred to a nylon membrane (Hybond N, Amersham) by a capillary blot overnight using 1.5M NaCl, 0.25 M NaOH. The blot was washed three times for 15 min in 2 x SSC and then air dried. The blot was prehybridised for 3 hrs at 65^{o}C in 10% dextran sulphate, 5 x Denhardt's solution, 1 x SSC, 0.5% SDS and 0.1 mg/ml denatured salmon sperm DNA. The blot was then hybridised overnight at 65^{o}C in 10 ml of the same solution to which the denatured <^{32}P> labelled probe had been added.

2.28 Exonuclease III deletions (adapted from Henikoff, 1987). 5-10 ug of DNA was digested with suitable enzymes to provide a recessed 3' or a blunt end for digestion by exonuclease III and a protruding 3' end to protect the remainder of the plasmid. After ethanol precipitation the DNA was dissolved in 60 ul Exo buffer (66mM Tris-HCl pH8.0, 0.66mM $MgCl_2$). The DNA was warmed to 30^{o}C and 500 units (a large excess) of exonuclease III was added and mixed rapidly. 2.5ul aliquots were removed at one minute time intervals and added to prepared tubes containing 7.5 ul of S1 mix (172 ul H_2O, 27 ul 10 x S1 buffer and 60 units of S1 nuclease). When all aliquots had been removed, they were incubated at room temperature for 30 min. 1 ul of S1 stop buffer (0.3M Tris base, 50mM EDTA) was added to

each aliquot followed by heat inactivation at 70°C for 10 min. At this point the extent of the digestion was determined by agarose gel electrophoresis. 1 ul of Klenow mix (3 ul 0.1M Tris-HCl pH8.0, 6 ul 1M $MgCl_2$, 20 ul H_2O mixed with 6 units Klenow polymerase) was then added to each reaction followed by incubation at 37°C for 5 minutes. All four dNTPs were added to a final concentration of 0.0125mM and the reactions were incubated for a further five minutes. DNA molecules were circularised by the addition of 40 ul of 1 x ligase buffer and 1 unit of DNA ligase, followed by overnight incubation. Plasmids (derivatives of pSDC102) were then transformed into DS941 *xerC1*. Derivatives of bacteriophage M13 were transfected into JM101 or TG1.

2.29 DNA sequencing. All DNA sequencing reactions were performed on single stranded M13 templates isolated as described in the BRL 'M13 cloning / dideoxy sequencing instruction manual'. All sequencing reactions were carried out using the Sequenase enzyme and the reagents supplied in the Sequenase kit (United States Biochemicals). The method used was that described in the manual supplied with the kit. Note that on warm days the initial labelling reaction carried out at room temperature was actually carried out in a 17°C water bath. This was found to reduce the incidence of sequence specific termination in all four tracks. The samples were denatured at 80°C and run on 8% polyacrylamide / urea gels as described above.

2.30 Extraction of DNA fragments for plasmid constructions from agarose gels using GENECLEAN. After staining, the gel was placed on a long wave transilluminator (300-360nm) and the band of interest was excised. The DNA was then purified using the "Glassmilk" suspension supplied with the Geneclean kit according to the manufacturers instructions.

2.31 Purification of DNA fragments for end labelling and subsequent use in Gel binding assays. The gel was stained and the band was excised on a long wave transilluminator as described above. The agarose chip was then placed in a small (500 ul) Eppendorf tube over a small amount of siliconised glass wool covering a hole in the bottom of the tube. The small Eppendorf tube was then placed in a large (1.5 ml), lidless Eppendorf tube and centrifuged at 6000 rpm for 10 min. The liquid from the bottom tube was collected and extracted twice with phenol and once with chloroform. The DNA was then precipated by the addition of NaOAc and ethanol. The DNA obtained was suitable to be end-labelled as described below

2.32 End-labelling of DNA fragments. Purified DNA fragment were end-labelled by filling in recessed 3' ends with the Klenow fragment of DNA polymerase I. The reaction contained 1-200ug/ml DNA, 50nM unlabelled nucleotides (dCTP, dGTP and dTTP), 10uCi alpha $<^{32}P>$ dATP, restriction buffer (50mM Tris/HCl pH 8.2, 10mM $MgCl_2$, 50mM NaCl) and 1 unit/ug DNA Klenow enzyme. After incubation at 16^{o}C for 1 hour, the reaction was stopped by phenol extraction, and the DNA was ethanol precipitated twice to remove unincorporated nucleotides.

2.33 Gel binding assays. Protein fractions containing XerC were diluted in a buffer that was generally 50mM Tris-HCl pH8.0, 1mM EDTA, 0.1mM DTT and from 50mM to 1M NaCl.

Binding buffer:-
10mM Tris/Glycine pH 9.4, 1mM EDTA, 10% glycerol (v/v).
If the protein was in a low salt dilution buffer NaCl was added to the binding buffer to a final concentration of approximately 100mM.

1.0 ul of each protein dilution was added to a 10 ul reaction tube containing approximately 1ng of end-labelled DNA fragment in binding buffer with 20ug/ml uncut pUC18 carrier DNA. Control (blank) reactions contained protein dilution buffer only i.e. all binding reactions in a given experiment contained the same NaCl concentration. After mixing thoroughly (vortexing) the reactions were incubated at 37^{o}C for 10 min, quenched on ice and loaded almost immediately onto a pre-run non-denaturing polyacrylamide gel at 4^{o}C.

CHAPTER THREE

FURTHER CHARACTERISATION OF *xerB*

3.1 Introduction

Mutations in the *xerB* gene were first isolated by Gillian Stewart (1986) and Colin Stirling (1987). The *xerB* gene is essential for site-specific recombination at *cer*. Plasmids containing two *cer* sites in direct repeat remain unresolved in strains with mutations in *xerB*. The *xerB* gene was isolated from a cosmid library, and genetically defined to a 1.69 kbp HindIII - AccI restriction fragment, by virtue of its complementation of the *xerB1* mutation. The slightly larger (1.92 kbp) *xerB*$^{+}$ HindIII fragment was sequenced and found to contain an open reading frame capable of encoding a protein of 503 amino acids with a predicted molecular weight of 55,300. The results of complementation experiments employing various plasmid subclones of the *xerB* region, taken together with the position of the Tn5 insertion in *xerB1* provided compelling evidence that this open reading frame represents the *xerB* gene (Stirling, 1987; Stirling *et al*, 1989).

Comparison of the restriction map of the cloned *xerB* region with the complete restriction map of the *E. coli* chromosome (Kohara *et al*, 1987) revealed that *xerB* maps between 96.5 and 97.0 minutes on the *E. coli* chromosome. Genetic data also demonstrated that *xerB* maps within 0.5 minutes (15 - 25 kbp) of *pyrB* encoding aspartate carbamoyltransferase (96.5 min) and *argI* encoding ornithine carbamoyltransferase (96.6 min) (Stirling, 1987). Other genes known to map in the 96 - 97 minute region are: *fbp* encoding fructose-bisphosphatase (96.0 min), *corB* involved in Mg^{++} transport (96.4 min), *sbaA* involved in regulation of serine and branched chain amino acid metabolism (96.7 min), *valS* encoding valyl-tRNA synthetase (96.8 min), *gntV* encoding glucokinase (96.9 min) and *leuX* encoding leucine tRNA 5 (97.0 min) (Bachmann, 1990 and Fig 3.1).

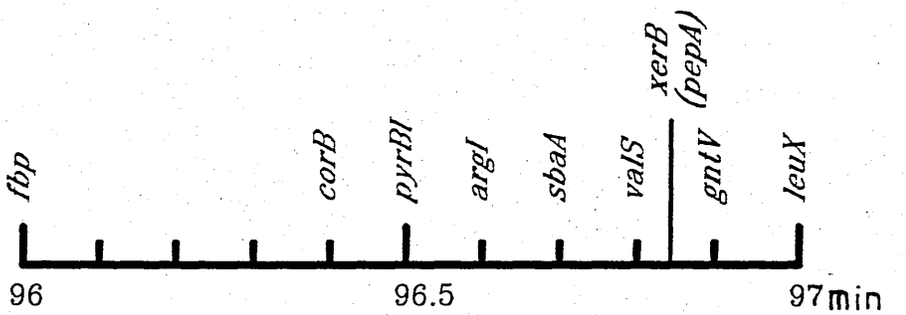


Figure 3.1 Genetic map of the 96.0 - 97.0 minute region of the E. coli **chromosome.** The approximate position of xerB (pepA) determined in this work is indicated. (Adapted from Bachmann, 1990).

3.2 The sequence upstream of *xerB*

Examination of the sequence of the 1,921 bp HindIII fragment revealed no candidates for a promoter sequence from which *xerB* could be transcribed (Stirling, 1987). However the sequenced region extended only 60 nucleotides upstream of the *xerB* start codon. The XerB$^+$ clone pCS112 contains the sequenced HindIII fragment in pUC18 in the correct orientation for *xerB* to be transcribed from P*lac*. The plasmid pCS110 contains the slightly larger BamHI fragment in pUC18 in the wrong orientation for *xerB* to be transcribed from P*lac*, and yet is also XerB$^+$ (Stirling, 1987 and Fig 3.2). This suggests that *xerB* might be transcribed from its own promoter in pCS110. If this were the case the *xerB* promoter would have to be located on the 310 bp BamHI - HindIII fragment or on the HindIII fragment itself. Because no potential promoter sequences were apparent on the HindIII fragment it was decided to extend the sequence of the *xerB* region by sequencing the BamHI - HindIII fragment.

To determine the sequence from the BamHI site towards the HindIII site the 583 bp BamHI - EcoRI fragment from pCS110 was cloned into M13mp18, producing mSD3 (Fig 3.2). To sequence the other strand the 310 bp HindIII - BamHI fragment from pCS110 was also cloned into M13mp18, producing mSD4 (Fig 3.2). The sequence of the first 349 bp of the insert in mSD3 was determine using reagents from the Sequenase kit (USB). This sequence overlapped with the first 39 bp of the sequence reported by Stirling (1987), with which it agreed completely. The sequence of the other strand (from HindIII towards BamHI) was determined from mSD4. There were no inconsistencies between the sequence determined from mSD3 and its complement determined from mSD4. The complete sequence from the BamHI site to the *xerB* start codon is shown in Fig 3.3. The sequence could have been extended a further 110 bp to the SphI site in pCS111, but this was thought not to be a priority.

The sequence upstream of *xerB* was examined for potential promoters using the program ZPROM (Chris Boyd). *E. coli* promoters have been found to contain two conserved

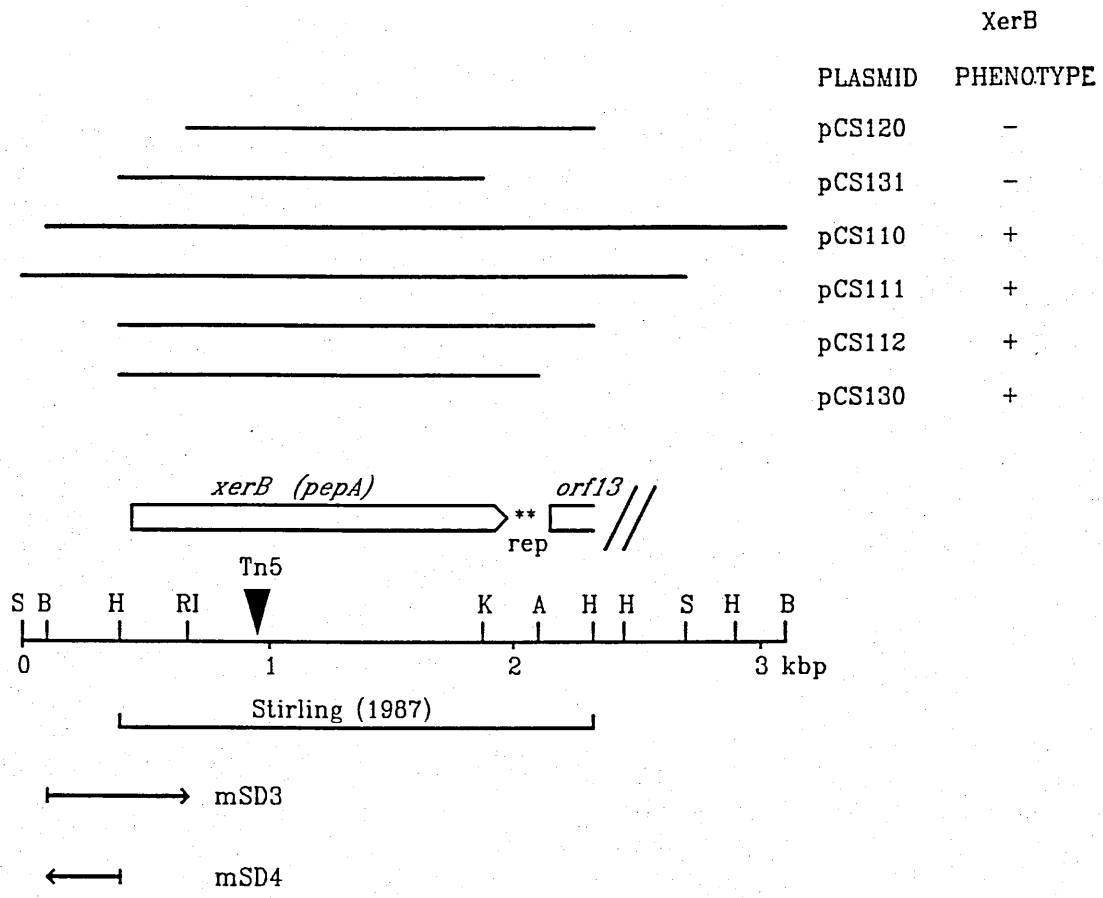


Figure 3.2 Physical map of the xerB (pepA) locus. The top of the figure shows the extent of chromosomal fragments carried by various plasmids. The XerB phenotypes of these plasmids, determined by complementation of the xerB1 mutation is also shown (taken from Stirling, 1987). Note that the insert carried by pCS112, but not that carried by pCS110, is in the correct orientation for xerB to be expressed from the plasmid lac promoter. The xerB (pepA) reading frame and the beginning of an open reading frame (orf13) identified by Stirling (1987) are shown. The position of insertion of Tn5 in xerB1::Tn5 is indicated. The fragments ligated into M13mp18 to create mSD3 and mSD4 are shown. The arrow heads indicate the direction in which these fragments were sequenced.

Restriction enzyme recognition sites are shown as follows:
A = AccI B = BamHI H = HindIII
K = KpnI RI = EcoRI S = SphI

Figure 3.3 Nucleotide sequence of the BamHI - HindIII fragment upstream of xerB (pepA). The sequence from the HindIII site at nucleotide 311 is that reported by Stirling (1987). The -35 and -10 regions of two possible xerB promoters sequences are indicated. The inferred protein sequence of the first 17 amino acids of XerB (PepA) is shown below the DNA sequence.

sequences of 6 bp just upstream from the transcription initiation site. These -35 and -10 regions, which are separated by a 16 - 18 bp spacer, are thought to be important for promoter recognition by the RNA polymerase enzyme. The consensus *E. coli* promoter sequence is

-35		-10		+1
TTGACA	16-18 bp	TATAAT	5-7 bp	START

The ZPROM program examines sequence data and finds matches to this consensus promoter sequence. Two statistics are produced to give a measure of how good the match is. Statistic I is a direct measure of how close the -35 and -10 sequences are to the consensus. Statistic II is calculated on the same basis as statistic I but also takes the spacing between the -35 and -10 regions into account. The ideal spacing between the -35 and -10 regions is thought to be 17 bp. All potential promoter sequences in the region upstream of *xerB* are given in Fig 3.4. Only those that score greater than 0.002 for statistic I or greater than 0.0002 for statistic II are shown. Most known *E. coli* promoters are found to score above 0.002 and 0.0002 for statistics I and II respectively, and these cut off points give a minimum number of false positives (Harr *et al*, 1983). The two best candidates for the *xerB* promoter are those with -35 boxes starting at nucleotides 178 and 186. However no further work was carried out to show that these sequences do indeed act as promoters *in vivo*.

3.3 The sequence downstream of *xerB*

Just downstream of *xerB*, within the 1,921 bp HindIII fragment, are two copies of the REP (repetitive extragenic palindromic) sequence (Stirling, 1987). About 500-1000 REP sequences are thought to occur dispersed throughout the *E. coli* genome (reviewed by Gilson *et al*, 1987; Higgins *et al*, 1988). They have always been found in transcribed DNA, either between coding regions in an

Plus Strand

Pos.	-35	-10	Statistic I	Statistic II	Dist
17	TTGACA GAAGAAGATCAAAAG____	CAAGAT.	0.0171314	0.0003426	15
64	TTTTGA GCGTCTCCCGCACCAGA__	TATCTT.	0.0004273	0.0004273	17
66	TTGAGC GTCTCCCGCACCAGA____	TATCTT.	0.0020727	0.0000415	15
147	TTGTTT CATGGCTTAAACGTCATT_	TATTCT.	0.0042797	0.0006420	18
178	TTGAGT CGTCGAAATCGTCGC____	TAAGAT.	0.0206821	0.0004136	15
178	TTGAGT CGTCGAAATCGTCGCTAA_	GATAAT.	0.0167675	0.0025151	18
183	TCGTCG AAATCGTCGCTAAGATAAT	TATACT.	0.0022110	0.0000442	19
186	TCGAAA TCGTCGCTAAGATAAT___	TATACT.	0.0106785	0.0016018	16
246	GTGCCA ATGCCGTAATAACGT____	TAAGAT.	0.0045553	0.0000911	15

Minus Strand

Pos.	-35	-10	I	II	Dist
388	TTTTTA CACTAAACTCCATGCAC__	TACGCT.	0.0002676	0.0002676	17
386	TTTACA CTAAACTCCATGCAC____	TACGCT.	0.0040584	0.0000812	15
348	AAGACA ACGGCGGTGGCTACAGA__	TAGAAT.	0.0005112	0.0005112	17
291	TTGCGA TGACTTCGTGTTAATCT__	TAACGT.	0.0014128	0.0014128	17
232	CTGAGA GGTGAATCCGTTGAG____	TATAAT.	0.0139481	0.0002790	15
169	ATGACG TTTAAGCCATGAAACAAGC	TAAAAT.	0.0043853	0.0000877	19
121	TTTACG GGCGTATTTAAAGTGAT__	AATCAT.	0.0002031	0.0002031	17

Figure 3.4 Potential promoter sequences upstream of xerB. The figure shows the ZPROM output of all potential promoters between the BamHI site (nucleotide 1) and the xerB (pepA) start codon (nucleotide 370). Those on the "plus" strand are in the correct orientation to drive the transcription of xerB (pepA). Most Known E. coli promoters score above 0.002 for statistic I and above 0.0002 for statistic II. The "Dist" column gives the spacing between the -35 and -10 sequences.

operon or in the 3' untranslated region of a transcript. Some occurences of REP sequences have been shown to be involved in stabilising mRNA (Newbury *et al*, 1987) and in transcription termination (Gilson *et al*, 1987) but it seems unlikely that this is the primary function of REP sequences. It has been suggested that REP sequences might be involved in nucleoid organisation and it is therefore interesting to note that several *E. coli* proteins appear to bind specifically to DNA fragments containing REP sequences. These REP binding proteins include DNA gyrase (Yang and Ames, 1988) and DNA polymerase I (Gilson *et al*, 1990).

Beyond these REP sequences is the beginning of an open reading frame (*orf13*) which extends for 63 codons to the HindIII site, after which no sequence data was available (Fig 3.2). There is evidence that this open reading frame is translated *in vivo*. A translational fusion between *orf13* at the HindIII site and the *lacZ'* coding sequence of M13mp19 (in mCS113) gave rise to blue plaques in JM101 grown on medium containing X-gal (Stirling, 1987). The arrangement of *xerB* and *orf13*, in close proximity and separated by two REP sequences, with no apparent *orf13* specific promoter, suggested that *xerB* and *orf13* are transcribed together on the same transcript. Because genes of related function are often co-transcribed, it was thought that *orf13* might have a *xer* related function. It was therefore decided to extend the sequence of the *xerB* region beyond the HindIII site further into *orf13*.

Just as this work was about to start, the sequence of *valS* (encoding valyl-tRNA synthetase) was published by two groups (Hartlein *et al*, 1987; Heck and Hatfield, 1988a; 1988b). It became apparent from the restriction map of the cloned *valS* gene that *valS* maps very close to *xerB* (Fig 3.5). Indeed it was found that the *xerB* sequence (Stirling, 1987) was separated from the published *valS* sequence only by an unsequenced HindIII fragment of 117 bp. In order to sequence this 117 bp HindIII fragment it was cloned into M13mp18 in both orientations to produce mSD1 and mSD2. Plaques formed by mSD1 were white and those

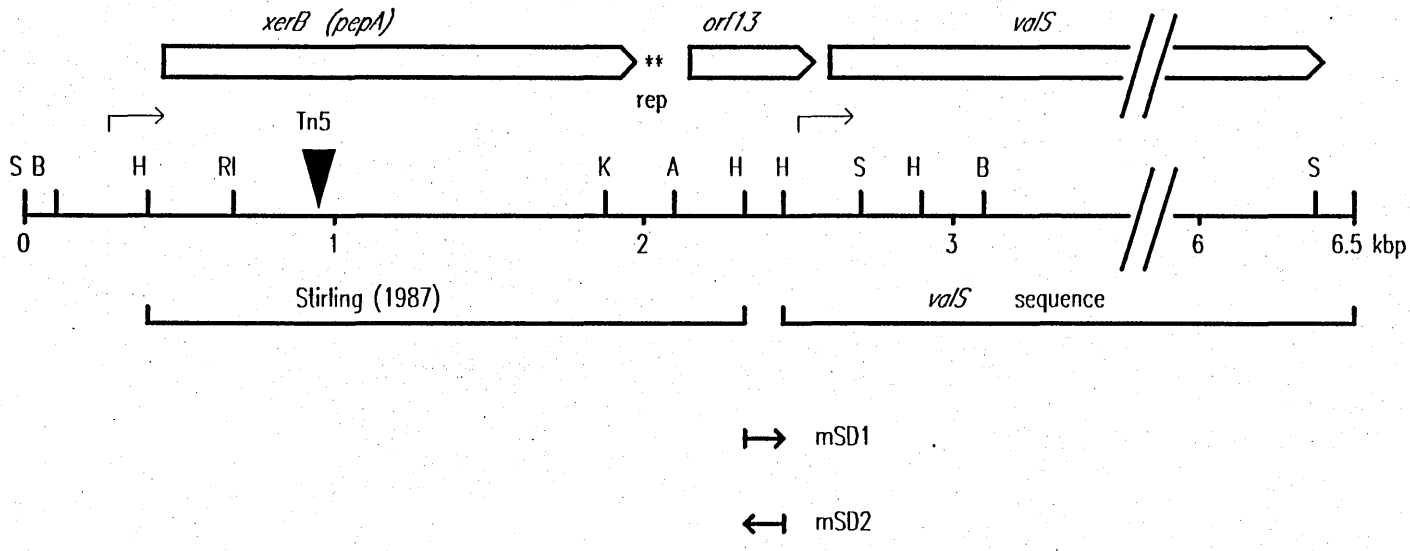


Figure 3.5 Physical map of the xerB (pepA) - valS region of the E. coli chromosome. The extent of the xerB sequence determined by Stirling (1987) is shown as is the extent of the published valS sequence (Heck and Hatfield, 1988a; 1988b; Hartlein et al, 1987). The fragments contained in mSD1 and mSD2, and the direction in which these fragments were sequenced are indicated. The approximate locations of the xerB, orf13 and valS reading frames and the two REP sequences are shown. The position of the two valS promoters identified by Heck and Hatfield (1988a) as well as the position of the possible xerB (pepA) promoters are shown (⌐⟶).

Restriction enzyme recognition sites are shown as follows:
A = AccI B = BamHI H = HindIII K = KpnI RI = EcoRI S = SphI

produced by mSD2 were blue on medium containing X-gal. The correct orientation of the sequence of the HindIII fragment was found when Heck sent us his unpublished sequence of the region upstream of *valS*. This sequence overlaps with the sequence of Stirling (1987) for 257 nucleotides, with an easily identified overlap (Fig 3.6). Since Heck was not totally confident of his sequence in this region, I tend to favour the carefully determined sequence of Stirling in the 7 areas of conflict, which are either single base-pair substitutions or single base-pair deletions. The sequence of the HindIII fragment was carefully determined from both mSD1 and mSD2, and again I favour my sequence over that of Heck in the 8 conflicts, which in this case are all single base-pair substitutions.

Examination of the complete sequence revealed that the *orf13* open reading frame extends through the 117 bp HindIII fragment and beyond the *valS* promoters (Fig 3.7). After this point the sequence of Heck and Hatfield (1988a) differs from that of Hartlein *et al* (1987). Depending on which sequence is correct, the *orf13* reading frame stops either 77 bp upstream (Heck and Hatfield, 1988a), or 14 bp downstream (Hartlein *et al*, 1987) of the *valS* start codon. If Hartlein *et al* are correct, *orf13* overlaps with the *valS* coding sequence by six codons. I find the sequence of Heck and Hatfield more plausible, but I have no data which might confirm my suspicions.

Examination of the mSD2 sequence revealed that the *orf13* open reading frame is fused "in frame" at both ends to the *lacZ'* gene of M13mp18. This explains the blue colour that was produced by mSD2 grown on X-gal containing medium, and confirms that the *orf13* reading frame carries on through the 117 bp HindIII fragment.

If the sequence of Heck and Hatfield is taken as being correct, the *orf13* encoded protein has 123 amino acids, with a calculated molecular weight of 13,700. The CODONPREFERENCE plot (see section 5.3 for more details of CODONPREFERENCE) shown in Fig 3.8 demonstrates that *orf13* has a codon composition consistent with it being a moderately expressed *E. coli* gene. The predicted amino acid sequence of the Orf13 protein bore no significant

```
Heck                                                        CGCATCAGGC
Xbcon1  ACAGCACTGA ACTCGTAGGC CTGATAAGAC ACAACAGCGT CGCATCAGGC

                                                            2035
  Heck  GCTGCGGTGT ATAGCCTGAT GCGGTAGTTA AATCCGACCA CAAGAAGCCC
Xbcon1  GCTGCGGTGT ATA.CCTGAT GCG.TATTTA AATCC.ACCA CAAGAAGCCC
                     *          * *         *

                                                            2085
  Heck  CATTTATGAA AAACGCGACG TTCTACCTTC TGGACAATGA CACCACCGTC
Xbcon1  CATTTATGAA AAACGCGACG TTCTACCTTC TGGACAATGA CACCACCGTC

                                                            2134
  Heck  GATGGCTTAA GCGCCGTTGA GCAACTGGTG TGTGAAATTG CCGCAGAAAC
Xbcon1  GATGGCTTAA GCGCCGTTGA GCAACTGGTG TGTGAAATTG CCGCAG.AAC
                                                             *
                          double stranded                   2184
  Heck  GTTGGCGCAG CGGTAAGCGC GTGCTCATCG CCTGTGAAGA TGAAAAGCAG
Xbcon1  GTTGGCGCAG CGGTAAGCGC GTGCTCATCG CCTGTGAAGA TGAAAAGCAG

                                                            2234
  Heck  GCTTACCGGC TGGATGAACC GCTGTGGGCG CGTCCGGCAG AAAGCTTTGT
Xbcon1  GCTTACCGGC TGGATGAAGC CCTGTGGGCG CGTCCGGCAG AAAGCTTTGT
                         *  *                        HindIII
                                                            2284
  Heck  TCCGCATAAT TTAGCGGGAG AAGGACCCGG CGGGCTGGCA CCGGTGGAGA
Xbcon1  TCCGCATAAT TTAGCGGGAG AAGGACCGCG CGGCGGTGCA CCGGTGGAGA
                                      **    ****
     certain                                                2334
  Heck  TCGCCTGGCC GCAAAAGCGT ACGAGCAGCC GGCGCGATAT ATTGATTAGT
Xbcon1  TCGCCTGGCC GCAAAAGCGT AGCAGCAGCC GGCGCGATAT ATTGATTAGT
                                 **
            HindIII                                         2384
Hartlein          AA GCTTTGCAGA TTTTGCCACC GCTTTCACAG AAGTGGTAGA
  Heck  CTGCGAACAA GCTTTGCAGA TTTTGCCACC GCTTTCACAG AAGTGGTAGA
Xbcon1  CTGCGAACAA GCTT

                                                            2434
Hartlein  CTTCGTTCCT TATGAAGATT C.T.CTGAAA CAACTGGCGC GCGAACGCTA
  Heck    CTTCGTTCCT TATGAAGATT CGTCCTGAAA CAACTGGCGC GCGAACGCTA
                                 * *
                                                            2484
Hartlein  TAAAGCCTAC CGCGTGGCTG GTTTCAACCT GAATACGGCA ACCTGGAAA.
  Heck    TAAAGCCTAC CGCGTGGCTG GTTTCAACCT GAATACGGCA ACCTGGAAAT
                                                              *
                                    2514
Hartlein  AATGGAAAAG ACATATAACC CACAAGATAT
  Heck    AATGGAAAAG ACATATAACC CACAAGATAT
          MetGluLys ThrTyrAsnP roGlnAsp
          valS
```

Figure 3.6 Comparison between the nucleotide sequences reported by various groups for the xerB - valS region. The sequence determined by Stirling (1987) and that reported in this work is labelled "Xbcon1". The sequence determined by Heck and Hatfield (1988a and unpublished) is labelled "Heck". The sequence reported by Hartlein et al is labelled "Hartlein". The sequence is numbered from the BamHI site at nucleotide 1 (see Fig 3.3). The sequence labelled Heck was determined on only one strand from nucleotide 1979 to 2147 (Heck, personal communications). The sequence from nucleotide 2285 is reported with high confidence (Heck, personal communications). The sequence from nucleotide 2331 is published (Hatfield and Heck, 1988a). Gaps (.) have been inserted to maximize similarity between sequences and asterisks (*) indicate all sequence conflicts. The start of the valS coding sequence is shown.

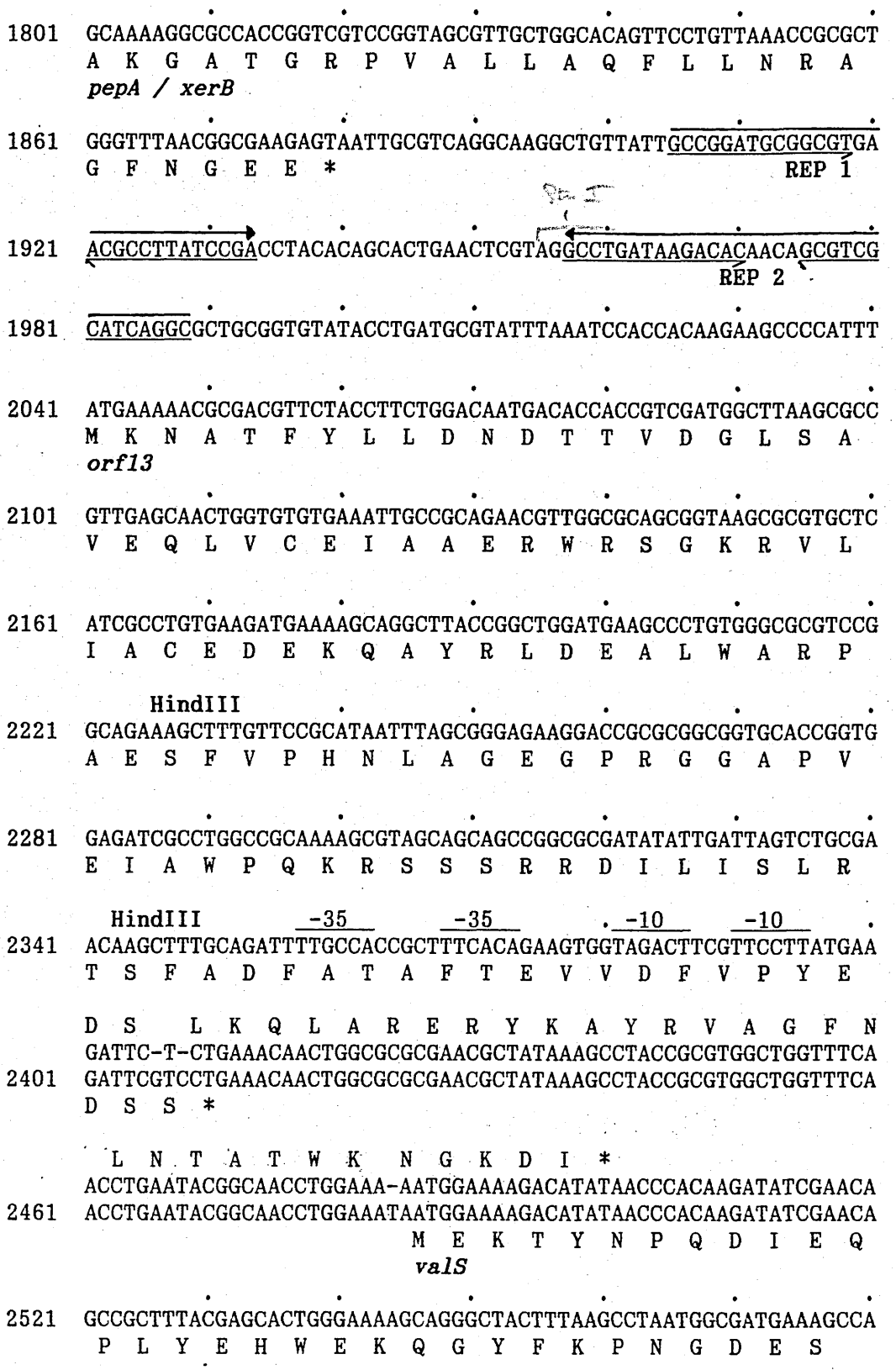


Figure 3.7 Nucleotide sequence of the region between xerB (pepA) and valS. The inferred protein sequences of Orf13, the C-terminus of XerB and the N-terminus of ValS are shown. From nucleotide 2401 to nucleotide 2520 both the sequence reported by Hartlein et al (1987) (top) and that reported by Hatfield and Heck (1988a) (bottom) are shown. The different Orf13 protein sequences inferred from these two sequences are shown. The -35 and -10 sequences of the two valS promoters identified by Hatfield and Heck (1988a) are indicated.

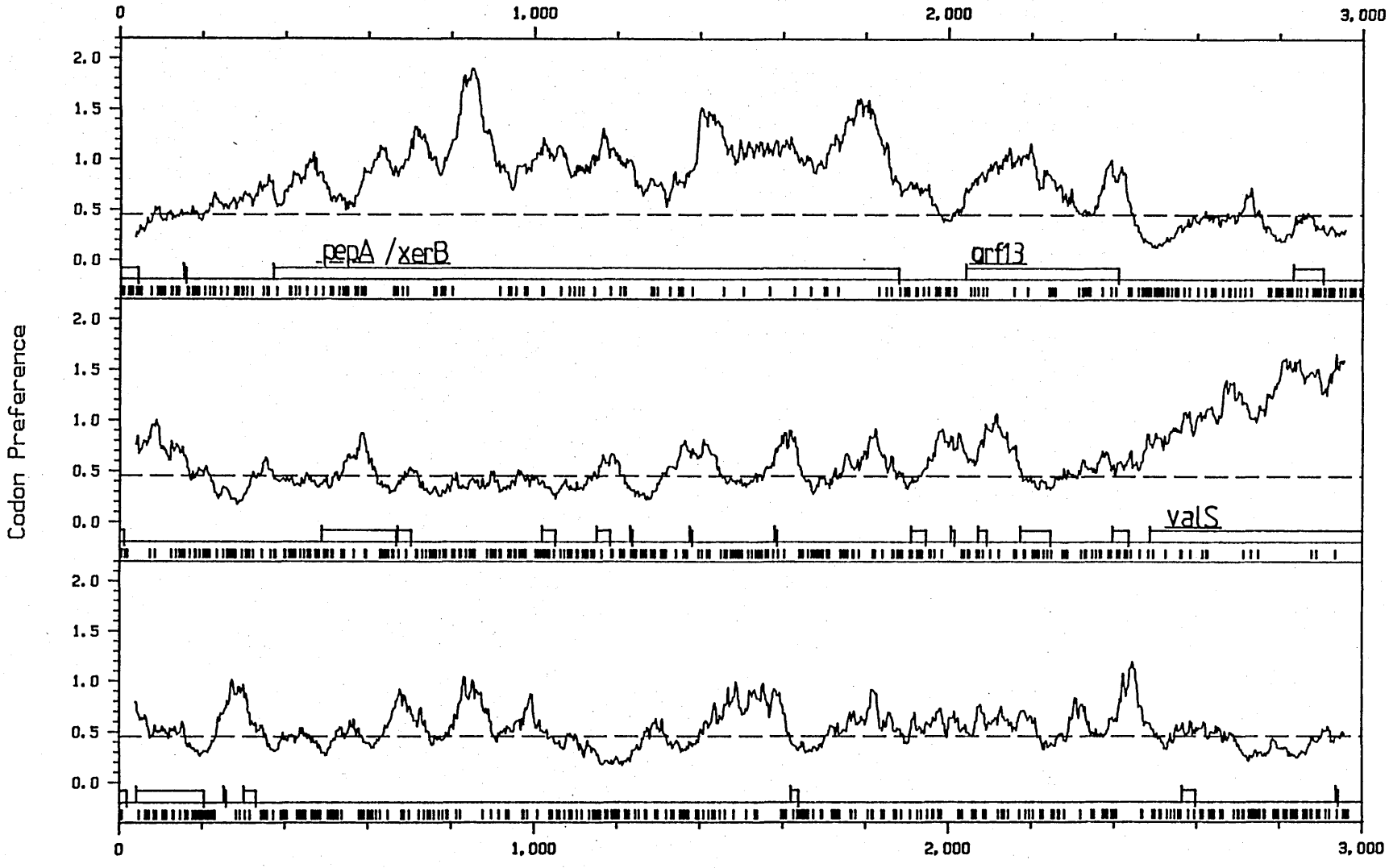


Figure 3.8 Codonpreference plot of the xerB (pepA) - valS sequence. The xerB (pepA), orf13 and valS reading frames are indicated. The plot extends only 514 bp into the 2950 bp valS coding sequence.

resemblance to any protein sequences in the databases.

A strain was constructed by Hazel O'Mara (a project student) and Mary Burke in which the chromosomal *orf13* gene was disrupted within the 117 bp HindIII fragment by the insertion of a Km^r gene. This strain was found not to be defective in *cer* site-specific recombination (unpublished results).

3.4 XerB shows amino acid sequence similarity to bovine leucine aminopeptidase

A database search for protein sequences similar to XerB was carried out with the help of John Collins (Dept. of Molecular Biology, University of Edinburgh). The search used the ICL 64 x 64 distributed array processor employing the "best local similarity" algorithm of Smith and Waterman (1981) as modified by Collins *et al* (1988). This search revealed a very striking region of similarity between the C-terminal region of XerB (residues 230-485) and the C-terminal region of bovine leucine aminopeptidase (residues 210-470). The N-terminal regions can also be aligned, but with a lower degree of similarity (Fig 3.9). The sequence of bovine leucine aminopeptidase was determined from overlapping peptides derived from leucine aminopeptidase which had been purified from bovine eye lens (Cuypers *et al*, 1982).

The molecular structure of bovine lens leucine aminopeptidase has recently been solved at 0.27 nm resolution (Burley *et al*, 1990). The protein is hexameric in structure and contains two zinc ions bound to each monomer subunit. The N-terminal domain of the protein (160 amino acids) mediates trimer - trimer interactions and is not directly involved in catalysis. The C-terminal domain (327 amino acids) contains the enzyme active site and two zinc binding regions. This correlates well with the high degree of sequence similarity between the C-termini of bovine lens leucine aminopeptidase and XerB, and the much lower degree of similarity between the N-termini of these proteins. The active site of the bovine enzyme has been identified and contains the residues Asp-255, Asp-273,

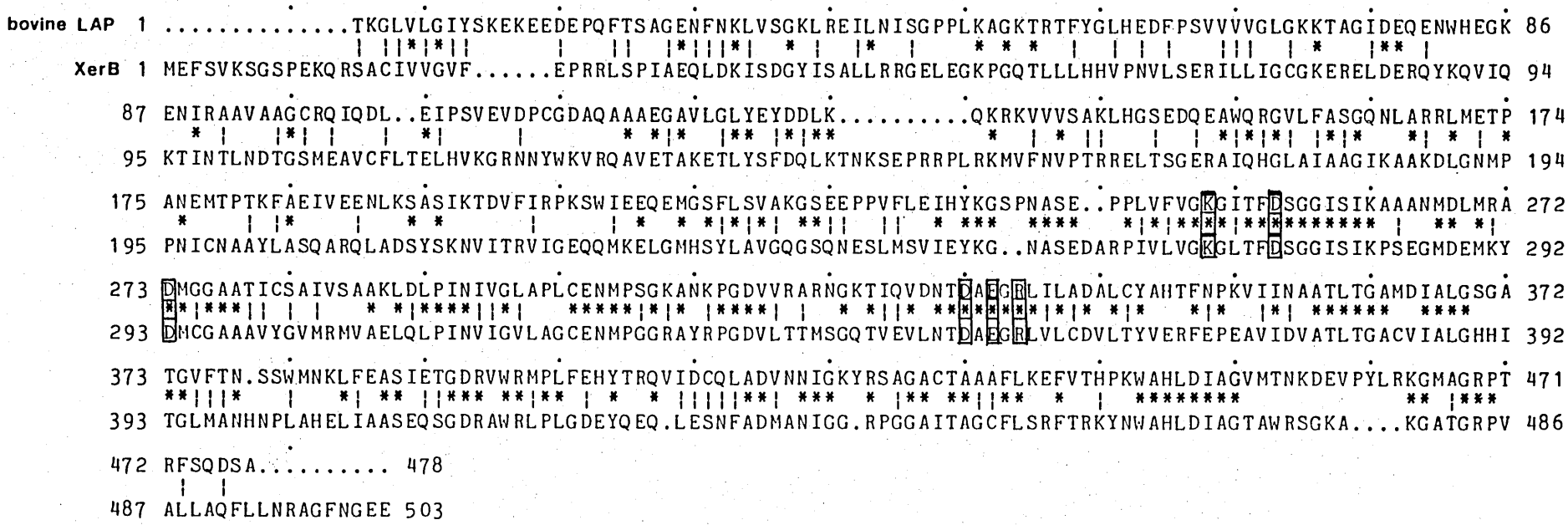


Figure 3.9 Comparison of the protein sequence of bovine lens leucine aminopeptidase with the predicted protein sequence of XerB (PepA). Gaps (.) have been introduced to maximize the similarity. Identical residues are indicated by an asterisk (*) and conservative substitutions (within the exchange groups (V, L, I, F, Y, M, W) - (A, T, G, S, C) - (H, K, R) - (D, E, Q, N)) are shown by a dash (¦). The residues identified as being in the active site of the bovine enzyme (Lys-250, Asp-255, Asp-273, Asp-332, Glu-334 and Arg-336 (Burley et al, 1990)) are shown boxed.

Asp-332, and Glu-334, which co-ordinate the two zinc ions, and also Lys-250 and Arg-336. All six amino acid residues are conserved between bovine lens leucine aminopeptidase and XerB (Fig. 3.9)

3.5 *xerB* is identical to *pepA*

The high degree of sequence similarity between XerB and leucine aminopeptidase led us to suspect that *xerB* might be the gene for one of the known *E. coli* aminopeptidases. Both *Salmonella typhimurium* and *E.coli* contain a surprising variety of enzymes that hydrolyse small peptides. These include five aminopeptidases (PepA, PepN, PepB, PepP and PepM), four dipeptidases (PepD, PepQ, PepE and PepG), a tripeptidase (PepT) and several carboxypeptidases. These peptidases are involved in the degradation of intracellular protein, especially under conditions of nutritional starvation, and also allow the utilisation of peptides supplied in the medium (See Miller, 1987 for a review).

Six peptidase activities of *S. typhimurium*, distinguishable by their electrophoretic mobility, were identified by Miller and Mackinnon (1974). *S. typhimurium* mutants deficient in four of these peptidases (PepN, PepA, PepD and PepB) were isolated (Miller and Mackinnon, 1974). The mutations causing three of these deficiencies (*pepN*, *pepA* and *pepD*) as well as two others (*pepP* and *pepQ*) were genetically mapped (Miller, 1975). It has been shown that *E. coli* contains broadly the same pattern of peptidases as *S. typhimurium*. Mutations in the genes for several of these peptidases have been isolated in *E.coli* but not mapped (Miller and Schwartz, 1978).

The *pepA* gene of *S. typhimurium* maps to 97 minutes (Miller, 1975), in the same relative position as *xerB* on the *E. coli* map. The *S. typhimurium* PepA protein was thought, because of its substrate specificities and its physical properties, to correspond to aminopeptidase I purified by Vogt (1970) (see Miller and Mackinnon, 1974). According to Vogt, aminopeptidase I has a subunit molecular weight, estimated from its mobility in SDS-

polyacrylamide gels, of 52,000. The protein was also reported to form an insoluble aggregate at low ionic strength. Both of these properties are consistent with the behaviour of purified XerB. XerB migrates in SDS polyacrylamide gels as a protein of molecular weight 55,000 and is known to precipitate in conditions of low ionic strength (Stirling, 1987). Other properties of aminopeptidase A (or aminopeptidase I) are its broad substrate specificity and its characteristic heat stability (70^{o}C for 5 minutes) (Vogt, 1970; Miller and Mackinnon 1974).

These facts taken together suggested that *xerB* might correspond to *pepA* of *E. coli*. In order to test this hypothesis, several experiments were carried out. The first of these experiments was to transform a number of *S. typhimurium* peptidase mutants with the plasmid pKS455. This plasmid is an Ap^{r} Cm^{r} 2-*cer* plasmid that loses Cm^{r} by Xer dependent site-specific recombination. *S. typhimurium* was used rather than *E. coli* because of the more extensive genetic characterisation of *pep* mutants in *S. typhimurium*. Also these strains were readily available from the lab of Dr. C. F. Higgins.

Initially only the four strains CH44, CH351, CH356 and TN1301 were transformed with pKS455. However, because of the differing restriction systems of *S. typhimurium* and *E. coli* (from which pKS455 plasmid DNA was originally isolated), this transformation was very inefficient. Only one Ap^{r} transformant was obtained. This pKS455 transformant of CH351 was found to be Ap^{r} and Cm^{r}. Plasmid DNA was prepared from this transformant by the boiling method and analysed by agarose gel electrophoresis. No band was seen in this DNA which would correspond to the deletion product of pKS455 (p456). Only bands corresponding to the various multimeric forms (monomer, dimer, trimer etc.) of pKS455 were seen (see Fig 3.10).

Plasmid DNA prepared from CH351 containing pKS455 was then used to transform CH44, CH351, CH356, TN996, TN1301, TN1302, TN1303 and TN1420. At least ten Ap^{r} pKS455 transformants of each of these strains was obtained. Transformants were then tested for resistance to

chloramphenicol. Plasmid DNA was also prepared from two transformants of each strain and examined by agarose gel electrophoresis. Strains were said to be Xer^- if pKS455 transformants were found to be resistant to chloramphenicol and contained only plasmid DNA of the correct size to be pKS455 or multimeric forms of pKS455. Strains were said to be Xer^+ if pKS455 transformants were not resistant to chloramphenicol and contained plasmid DNA of the correct size to be p456. An example of the type of gel used in this analysis is shown in Fig 3.10 and the results for all strains tested are summarised in Table 3.1.

Table 3.1 shows that all strains that carried a functional *pepA* gene were Xer^+ and all strains with a mutation in *pepA* were Xer^-. Mutations in the other *pep* genes did not affect the Xer phenotype and neither did a mutation in *tppB*, a gene which encodes a tripeptide permease.

Ideally the next experiment would have been to complement the Xer^- phenotype of the *S. typhimurium pepA* mutants with a cloned copy of the *E. coli xerB* gene. However, because of the restriction system of *S. typhimurium* it was not possible to transform the *pepA* strains with a 2-*cer* plasmid and a compatible *xerB* complementing plasmid. Instead an experiment was done that showed that a cloned *S. typhimurium pepA* gene complemented the *xerB* mutation in *E. coli*. Dr. C. Miller kindly sent us the plasmids pJG6 and pJG7. Both of these plasmids contain the *pepA* gene from *S. typhimurium* cloned as partial Sau3A fragments into the BamHI site of pBR322. DS941 *xerB1* was transformed with the Cm^r Tet^r 2-*cer* plasmid pCS202. DS941 *xerB1* containing pCS202 was then transformed with either pJG6 or pJG7 and transformants were selected on medium containing ampicillin and chloramphenicol. Plasmid DNA was examined on an agarose gel by the single colony lysis method. DS941 *xerB1* carrying pCS202 that had been transformed with either pJG6 or pJG7 had lost the band corresponding to pCS202 and instead had a band corresponding to pCS203, the deletion derivative of pCS202.

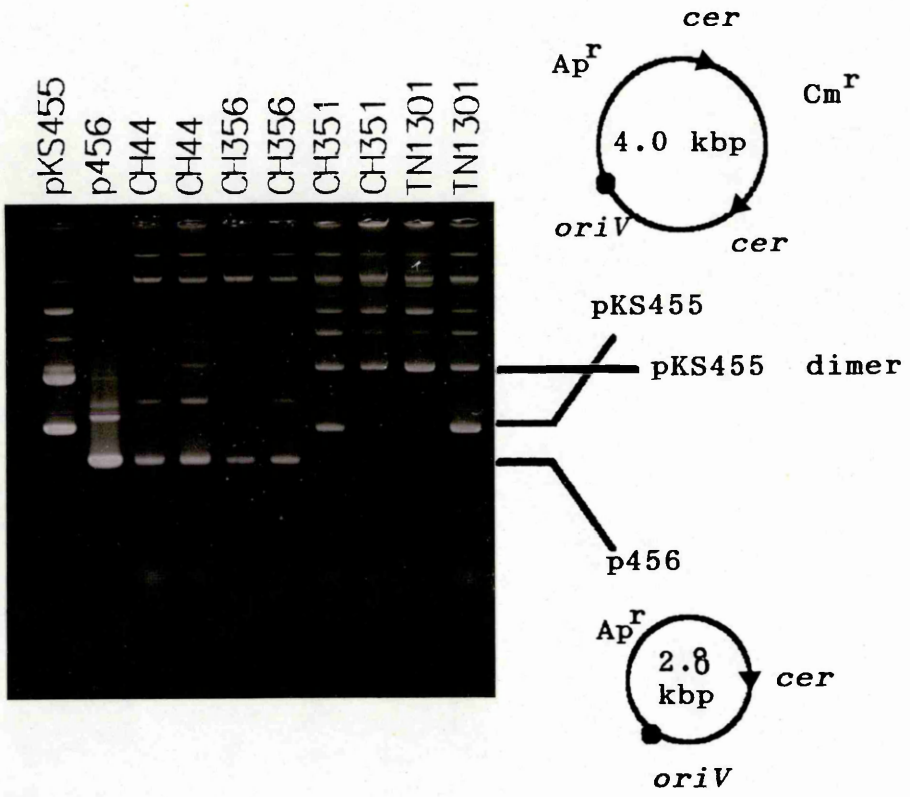


Figure 3.10 cer-mediated recombination in pepA$^+$ and pepA$^-$ strains of S. typhimurium. CH44 (pepA$^+$), CH356 (pepA$^+$), CH351 (pepA$^-$) and TN1301 (pepA$^-$) (see table 3.1) were transformed with pKS455 plasmid DNA isolated from CH351. Plasmid DNA was visualised on a 0.85% agarose single colony gel. cer mediated recombination acting on pKS455 in Xer$^+$ strains yields the plasmid p456.

Table 3.1 Xer phenotype of peptidase mutants of *S. typhimurium*

Strain	Genotype	Xer phenotype assayed by loss of Cm^r	Xer phenotype assayed by gel electrophoresis
CH44	del(*oppBC*)250	+	+
CH351	del(*oppBC*)250 *pepA*	-	-
CH356	del(*oppBC*)250 *tppB*	+	+
TN1301	del(*leuBCD*)485 *pepA16* *pepB11* del*supQ*302(*proAB*-*pepD*) *pepP1* *pepQ1*	-	-
TN1302	del(*leuBCD*)485 *pepB11* del*supQ*302(*proAB*-*pepD*) *pepP1* *pepQ1* *pepN90*	+	+
TN1303	del(*leuBCD*)485 *pepA16* *pepB11* del*supQ*302(*proAB*-*pepD*) *pepP1* *pepQ1* *pepN90*	-	-
TN996	del(*leuBCD*)485 *pepA1* *pepB1* *pepD1* *pepN10* *zjh*-829::Th1	-	-
TN1420	del(*leuBCD*)485 *pepA16* *pepB11* del*supQ*302(*proAB*-*pepD*) *pepP1* *pepQ1* *pepN80* *pepT1* *metE388* *zie*-822	-	-

The strains listed above were transformed with pKS455. The Xer phenotype of transformants was assessed by testing for sensitivity to chloramphenicol and by agarose gel electrophoresis of plasmid DNA (See Fig 3.10)

Other experiments carried out by Dr. George Szatmari and Hazel O'Mara further confirmed that *xerB* does indeed correspond to *pepA*. The results of these experiments are summarised briefly below.

1) The purified XerB protein was shown to have peptidase activity against the artificial peptide substrate leucine *p*-nitroanilide. This peptidase activity was Mn^{++} dependent and stable to heat at $70^{o}C$ (Stirling *et al*, 1989).

2) Extracts prepared from DS941 *xerB1* lacked any heat stable peptidase activity against leucine *p*-nitroanilide. Similar extracts prepared from DS941 did contain heat stable peptidase activity (O'Mara, pers. comm.).

3) Spontaneous mutants were selected which were resistant to the toxic peptide Val-Leu-NH_2 but still sensitive to valine. Val-Leu-NH_2 is hydrolysed by *pepA*$^{+}$ *pepN*$^{+}$ strains to produce valine, which is toxic to *E. coli* in high concentrations. A plasmid-borne copy of *xerB* was found to make these mutants sensitive to Val-Leu-NH_2. Several of the mutants selected in this way were found to be Xer^{-}. Those mutants selected in this way which were still Xer^{+} appeared to retain at least some PepA peptidase activity (O'Mara and Szatmari, personal communications).

3.6 Discussion

The *xerB* sequence reported by Stirling (1987) has been extended by 310 bp 5' of *xerB* and 120 bp 3' of *xerB*. The 310 bp BamHI - HindIII fragment 5' of *xerB* was found to contain two potential promoter sequences from which *xerB* could be transcribed. No further work was carried out to investigate whether *xerB* is actually transcribed from one of these promoter sequences. There was no obvious sign of any other open reading frames within the 370 bp upstream of the *xerB* coding sequence.

The sequence 3' of *xerB* was found to overlap with the unpublished sequence determined by Heck and Hatfield, 5'

of *valS* (the gene for valyl tRNA synthetase). This allowed the accurate location of *xerB* with respect to the known genes in the 96.0 to 97.0 minute region. Taken together, the *xerB* sequences reported here, those reported by Stirling (1987), and the *valS* sequence of Heck and Hatfield (1988a; 1988b) and Hartlein *et al* (1987) give 5640 bp of continuous *E. coli* chromosomal sequence.

Between *xerB* and *valS* there is a small open reading frame (*orf13*). The position of *orf13*, just downstream of *xerB* and separated from *xerB* by two REP sequences, suggests that *orf13* is co-transcribed with *xerB*. There is evidence, from a fusion of the first 63 codons of *orf13* to the *lacZ'* gene of M13mp18 (in mCS113), that *orf13* is translated *in vivo*. Another fusion between the next 39 codons of *orf13* and *lacZ'* (in mSD2) indicates that the *orf13* reading frame carries on for another 117 bp. Depending on whether the sequence data of Heck and Hatfield (1988a; 1988b) or those of Hartlein *et al* (1987) are to be believed, the *orf13* open reading frame then carries on for either a further 63 bp or a further 150 bp. This would make the protein encoded by *orf13* either 123 amino acids (with a calculated molecular weight of 13,700) or 152 amino acids (with a calculated molecular weight of 16,700). The function of *orf13* is not known, but it appears not to be essential for *cer* site-specific recombination.

The proposed *orf13* reading frame overlaps with the two *valS* promoters. If expression of *valS* is to be controlled independently of expression of the *xerB* / *orf13* operon there would have to be a terminator for the *xerB* / *orf13* transcript. This causes an interesting problem. Either the transcript could terminate before the *valS* promoters, in which case the *orf13* reading frame would be cut short, or the transcript could terminate after the promoters, beyond the end of *orf13*. In this case the terminator would have to stop only those transcripts which originated from the *xerB* promoter and not those that originated from the *valS* promoters. Alternatively there might not be any terminator for the *xerB* transcript, in which case *valS* would be transcribed by a combination of

read-through transcription from *xerB* and transcription from the *valS* promoters. These alternatives could only be distinguished by experiments to map the 3' end of the *xerB* transcript.

The predicted amino acid sequence of XerB was shown to display a high degree of sequence similarity to bovine lens leucine aminopeptidase. This similarity prompted us to investigate whether *xerB* corresponds to any of the known peptidase genes of *S. typhimurium* and *E.coli*.

All *pepA* mutant strains of *S. typhimurium* tested were found to be phenotypically Xer^-, whereas all $pepA^+$ strains were Xer^+. Furthermore, a cloned *pepA* gene from *S. typhimurium* was found to complement the Xer^- phenotype of *xerB* mutant *E.coli* strains. Other experiments showed that the *xerB* gene product has peptidase activity both *in vitro* (against the artificial peptide analogue Leu-*p*-nitro-anilide) and *in vivo* (against the dipeptide Val-Leu-NH_2).

These experiments led us to the conclusion that *xerB* is the *pepA* gene of *E. coli*. For the purposes of further discussion the *pepA* nomenclature will be used rather than *xerB* whenever possible.

The results show that, of all the peptidase genes looked at, only *pepA* is required for *cer* recombination. TN1302 is mutant in *pepB*, *pepD*, *pepP*, *pepQ* and *pepN* but is still Xer^+. None of the other peptidase genes can substitute for *pepA* as shown by the fact that CH351, which has wild type genes for all the peptidases apart from *pepA*, is Xer^-.

This specific requirement for *pepA* could be explained in two ways. One possibility is that PepA plays a structural role in the recombination machinery that acts at *cer*. PepA could interact directly with *cer* DNA and/ or with the other proteins involved in *cer* recombination. These interactions would not necessarily require the peptidase activity of PepA. It is easy to envisage how this could specifically require PepA rather than any of the other cellular peptidases, because the other peptidases are probably structurally quite different from PepA. However we have so far been unable to detect any specific interactions between PepA and either *cer* or ArgR

(Stirling, 1987; G. Szatmari, personal communications).

The other possibility is that the enzymatic (peptidase) activity of PepA plays some essential role in *cer* recombination. In this case the specific requirement for PepA must reflect some substrate specificity of PepA not possessed by any other cellular peptidases. It is indeed the case that the various peptidases have different substrate specificities. For instance, PepA, PepN and PepB are all described as broad specificity aminopeptidases (Miller, 1987), yet all three can be distinguished by their substrate specificities. PepN is the only peptidase in *E. coli* or *S. typhimurium* that shows detectable hydrolytic activity towards amino acid beta-napthylamides (Miller and Mackinnon, 1974; Miller and Schwartz, 1978). PepA appears to be the only *S. typhimurium* peptidase capable of efficiently hydrolysing the peptide analogue alafosfalin (L-alanyl-L-1-aminoethylphosphonic acid). Alafosfalin is hydrolysed *in vivo* to produce the toxic compound L-1-aminoethylphosphonic acid (Allen *et al*, 1978). However, alafosfalin is not toxic to *pepA* mutant (*pepN*$^+$, *pepB*$^+$) strains, presumably because it is not hydrolysed by these strains (Gibson *et al*, 1984).

If it is the peptidase activity of PepA that is required for recombination at *cer*, this activity could conceivably be directed against either some protein component of the recombination machinery (ArgR, XerC, or some other as yet unidentified protein) or it could be directed against a peptide. If the requirement was for PepA to act on one of the other Xer proteins, it would be expected that the removal of one or several amino acids from the N-terminus of the protein would activate it. There is at least one precedent for the activation of a protein by the removal of the N-terminal methionine. The activity of glutamine phosphoribosylpyrophosphate amidotransferase requires the removal of the N-terminal methionine to uncover a crucial cysteine residue (Tso *et al*, 1982). There is no evidence that PepA can act on the N-terminus of proteins, although there is no *a priori* reason why it could not, so long as the N-terminus is not buried in the protein tertiary structure. PepA could also

conceivably act on the N-terminus of a protein before it is fully translated. It is known, however, that PepA is not the enzyme responsible for removing the N-terminal methionine from those proteins which do not retain methionine on their N-terminus (which make up about 60% of bulk E. coli protein). PepA attacks many peptides with N-terminal amino acids other than methionine and is able to proceed beyond the N-terminal amino acid (Vogt, 1970). It now appears that PepM is the principal enzyme responsible for removing the N-terminal methionine from *E. coli* proteins (Miller, 1987).

Alternatively PepA could be required to act on a peptide rather than a protein. This peptide could be an inhibitor of Xer function, or the products (amino acids) released by the activity of PepA on this peptide could be activators of Xer function. These peptide-inhibitors or amino acid-activators could conceivably modify the DNA binding activity or recombination activity of ArgR or some other Xer protein.

Future work could possibly distinguish between a structural or an enzymatic role for PepA in *cer* recombination. Site directed mutagenesis could be used to abolish the activity of PepA without affecting the gross structure of the protein. If this mutation did not abolish the Xer activity of PepA it would show that PepA peptidase activity did not play an essential role in *cer* site-specific recombination. The crystal structure of bovine leucine aminopeptidase has recently been solved (Burley *et al*, 1990) and work is in progress to solve the structure of PepA (Isaacs, personal communications). This might help to target mutagenesis to the PepA active site. Alternatively, classical genetic techniques could be used to select for *pepA* mutations that either abolish peptidase activity without affecting Xer activity or abolish Xer activity without affecting peptidase activity.

Of course it is possible that PepA plays both a structural and an enzymatic role in *cer* recombination. If this were the case, PepA and the other Xer proteins might first assemble into a recombination complex and only then would the peptidase activity be required.

A variant *cer* site, with an altered crossover region, recombines in the absence of both ArgR and PepA (Summers, 1989). This site requires only about 50 bp around the crossover region and has lost directionality, allowing both intermolecular and intramolecular recombination. Some directionality is restored if ArgR, PepA and the full variant site (280 bp) are all present (Summers, 1989). These results suggest that ArgR, PepA and the upstream region of *cer* (which includes the ArgR binding site) somehow act together to impose directionality on *cer* recombination. This is in some ways analogous to the directionality imposed on promiscuous (mutant) Gin/*gix* recombination by Fis/enhancer interactions. Wild-type Gin does not promote recombination efficiently in the absence of the enhancer sequence and Fis (Kahmann *et al*, 1985). A mutant Gin enzyme exists that will catalyse recombination in the absence of Fis and the enhancer (Klippel *et al*, 1988b). This mutant Gin is promiscuous both in its choice of substrates (which can be supercoiled or linear) and in the reactions it will promote (inversion and deletion) (Klippel *et al*, 1988b). However, the presence of the enhancer together with Fis can still impose, to some extent, the normal substrate selectivity and reaction specificity on this mutant Gin (Klippel *et al*, 1988b).

The exact role of PepA in *cer* site-specific recombination remains mysterious. However, the results discussed in the above paragraph suggest that PepA is not the *cer* recombinase but instead plays an accessory role in *cer* recombination. This is analogous to several other systems where accessory factors are required for site-specific recombination. IHF is required for lambda integration and excision and FIS is required for the various DNA inversion systems. However, in these cases the accessory factors are known to interact specifically with sequences required for recombination (*attP*, *attL* and *attR* in the case of IHF, the enhancer site *sis* in the case of Fis). PepA, on the other hand, does not appear to interact specifically with any of the DNA sequences required for *cer* site-specific recombination.

Finally it is worth noting one more point that came

out of the experiments with *S. typhimurium*. This is the fact that *S. typhimurium* is capable of efficiently carrying out recombination at plasmid borne *cer* sites. This implies that all of the functions required for *cer* site-specific recombination are present in *S. typhimurium* as well as in *E. coli*. The *S. typhimurium argR* and *pepA* genes have already been characterised and mapped (Kelln and Zak, 1978; Sanderson and Hurley, 1987; Miller, 1987). Preliminary southern blot experiments have revealed that *S. typhimurium* also contains a homologue of the *E. coli xerC* gene (Peter Sykora, personal communication). This conservation of Xer function tends to suggest that these functions are important for bacterial fitness and probably all have some role other than just plasmid-stabilising *cer* recombination.

CHAPTER FOUR

ISOLATION AND MAPPING OF *xerC*

4.1 INTRODUCTION

Prior to this work, eleven independent *xer* mutants had been isolated (Stewart, 1986; Stirling, 1987). Two of these mutants belonged to the *argR* complementation group, the other nine belonged to the *pepA* complementation group. These eleven mutants did not make up a large enough sample to be confident that all *xer* complementaion groups were represented. Furthermore it appeared that neither ArgR or PepA was the actual recombinase that acts on *cer* because: i) neither protein displays significant amino acid similarity to any known recombinase (Stirling *et al*, 1988b; 1989), and ii) a variant *cer* site produced by David Summers (the type II hybrid between ColE1 *cer* and ClODF13 *parB*) recombines in the absence of functional *argR* and *pepA* genes (Summers, 1989). This implied that the gene for the *cer* recombinase remained to be discovered. It was therefore decided to look for further genes required for *cer* site-specific recombination.

The selection strategy used by Colin Stirling to produce nine of the eleven *xer* mutants was based on the transformation of Tn5 mutagenised populations with the plasmid pKS455 (Stirling, 1987). This plasmid contains two directly repeated copies of *cer* flanking a copy of the *cat* gene, encoding resistance to chloramphenicol. In a wild type host (Xer^+) site-specific recombination between the *cer* sites produces two circles; one (p456) contains the Ap^r gene and replication origin, the other (p457) carries only the *cat* gene (Fig 4.1). The circle carrying the *cat* gene fails to replicate and only p456 persists. The host therefore remains resistant to ampicillin but becomes sensitive to chloramphenicol. Mutants defective in genes required for *cer* site-specific recombination (*xer* genes) can be selected because they retain pKS455 in its original form and remain resistant to chloramphenicol.

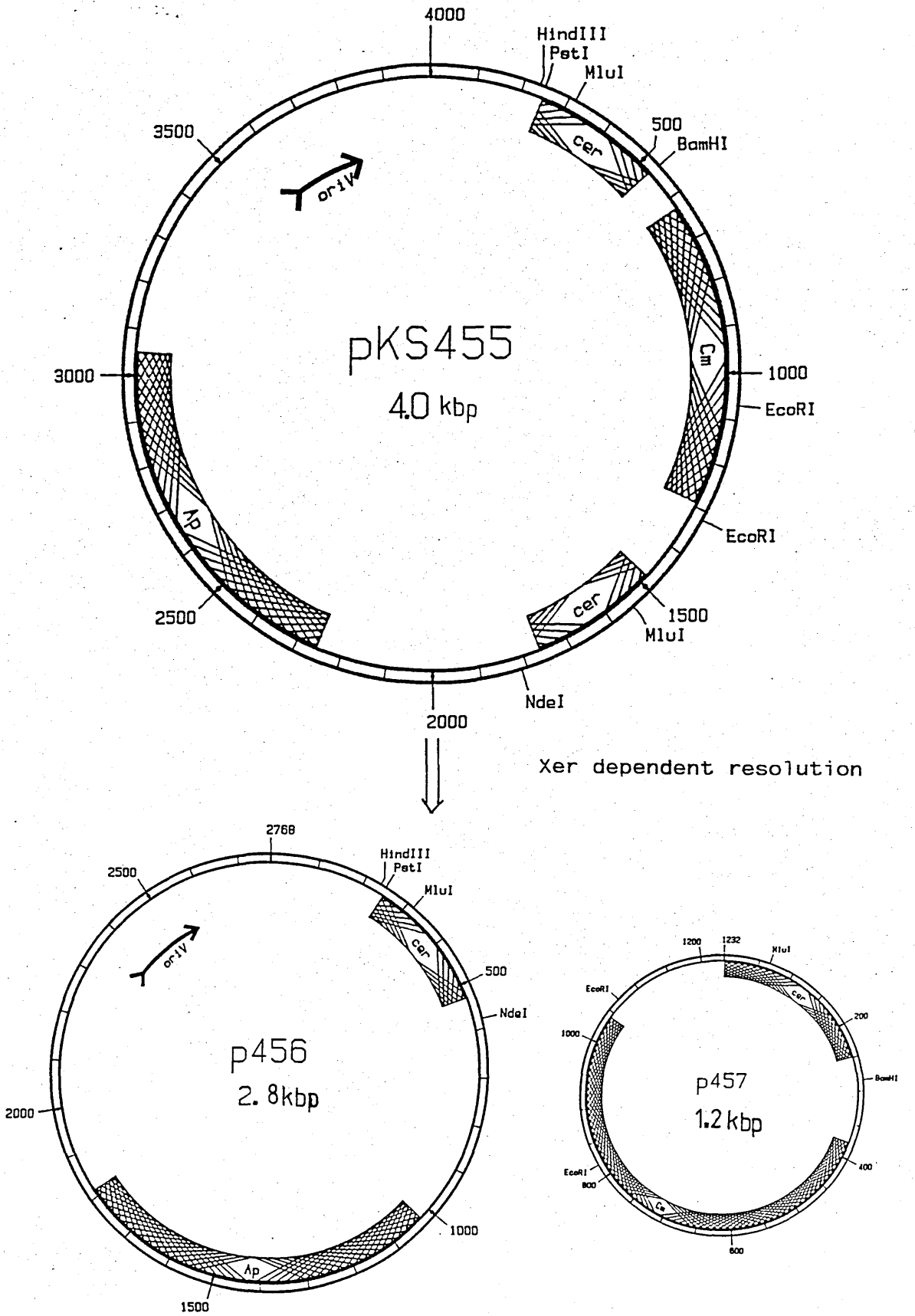


Figure 4.1 Diagrammatic representation of Xer - dependent resolution of the reporter plasmid pKS455. pKS455 is rapidly resolved in Xer$^+$ hosts. Note that, of the two resolution products, only p456 contains a replication origin.

4.2 Tn5 as a mutagen

Several treatments, such as irradiation with UV light, treatment with chemical mutagens and transposon mutagenesis can be used to increase bacterial mutation rates. Mutagenesis with the bacterial transposon Tn5 has a number of advantages over other methods of mutagenesis. Tn5 is a 5.8 kbp composite (class I) transposon made up of two copies of the 1,535 bp IS50, in inverted repeat, flanking a central region of 2.7 kbp (Fig 4.2). The central region of Tn5 carries a gene encoding neomycin phosphotransferase II, conferring resistance to neomycin and kanamycin. Two other genes in this region confer resistance to streptomycin (cryptic in *E. coli*) and CL990, a member of the bleomycin family of antibiotics (Mazodier *et al*, 1985). The properties of Tn5 have been reviewed by Berg and Berg (1983). Some of these properties, relevant to the use of Tn5 as a mutagen, are listed below:

1) Tn5 insertion mutants can be pre-selected before screening by a period of growth on kanamycin.

2) After a mutant has been obtained, the close physical and genetic linkage of Tn5 to the mutant allele can be used to map the gene, by both genetic and physical (eg restriction) means. The close association of Km^r, encoded by Tn5, with the mutation can be used to construct new strains carrying the mutation and to clone the mutated gene.

3) Tn5 transposes into the chromosome with very little sequence specificity. It inserts at many sites within a gene (Berg *et al*, 1980) and in many genes on the *E. coli* chromosome (Shaw and Berg, 1979). Very large pools of Tn5 insertion mutants (up to 10^5) can easily be generated as described below. The *E. coli* chromosome is approximately 4×10^6 bp in size, so a pool of 10^5 independent mutants contains in principle a Tn5 insertion every 40 bp or so.

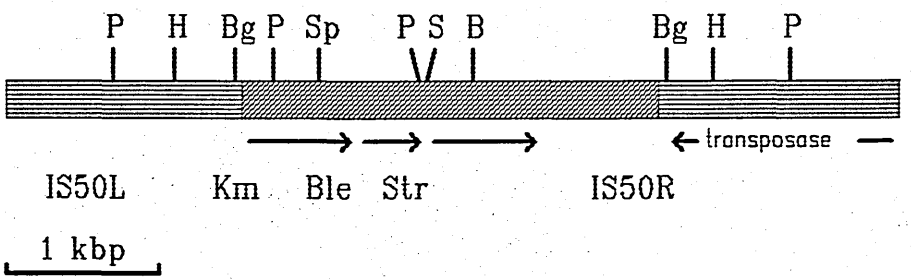


Figure 4.2 Diagrammatic representation of Tn5. Tn5 contains two copies of IS50 flanking a central region encoding resistance to kanamycin (Km), CL990 (Ble) and streptomycin (Str). Restriction enzymes recognition sites are shown as follows:

B = BamHI Bg = BglII H = HindIII
P = PstI S = SalI S = SphI

4) Insertion of Tn5 into a gene usually results in total inactivation of the target gene. Reversion of Tn5 insertion mutations by transposon excision occurs at a very low frequency, about 10^{-6} revertants/generation (Berg, 1977). Insertions within multigene operons usually exert highly polar effects on genes located downstream of the insertion. Berg *et al* (1980) noted that approximately one third of Tn5 insertions into *lacZ* resulted in a constitutive low level of expression of *lacY*. The remaining two thirds of insertions into *lacZ* totally abolished expression of *lacY*.

The uses of Tn5 as a mutagen are extensively reviewed in de Bruijn and Lupski (1984). Because of the advantages listed above it was decided to continue to use Tn5 as a mutagen.

4.3 Lambda NK467 as a Tn5 delivery vector

Lambda NK467 is a "suicide" vector, constructed by Nancy Kleckner, for the delivery of Tn5 as a mutagen. Its use for this purpose is described by de Bruijn and Lupski (1984). The genotype of lambda NK467 is delta b221, *rex*::Tn5, cI857, *O*am, *P*am. The b221 deletion removes *attP* along with *int* and *xis* so that the phage can not integrate into the *E. coli* chromosome (Davidson and Szybalski, 1971). The *O* and *P* genes are essential for phage replication, therefore lambda NK467 can be propagated in an amber suppressing strain such as DS941 but will not replicate in a Sup^o strain. The vast majority of Km^r cells received after infection of a Sup^o strains such as DS953 will result from transposition of Tn5 into the host chromosome.

4.4 The use of pKS455 in the selection of *xer* mutants

An initial period of growth on medium without choramphenicol was found to be necessary for proper expression of the *cat* gene in pKS455 transformants of Xer^- strains. When Xer^- strains were transformed with pKS455

and plated directly onto medium containing both ampicillin and chloramphenicol, no transformants were obtained. These same Xer^- pKS455 transformants were also plated onto medium containing only ampicillin and then replica plated onto medium containing ampicillin and chloramphenicol. All of these transformants were found to be resistant to chloramphenicol after replica plating. The selection of *xer* mutants from a population transformed with pKS455 was found to be a very clean selection. DS941 (a Xer^+ strain) was transformed with pKS455 and plated out onto solid medium containing ampicillin. Transformants were then replica plated onto medium containing ampicillin and chloramphenicol. Less than one in ten thousand transformants was found to be resistant to chloramphenicol.

In order to select *xer* mutants, the mutagenised population must be transformed with pKS455. This transformation is an inefficient process; less than one in a thousand viable cells put into a standard $CaCl_2$ transformation experiment is transformed at saturating concentrations of plasmid DNA. An unamplified mutant population containing 10^5 independent mutants, transformed with pKS455, will yield only about 100 ampicillin resistant colonies for the selection of *xer* mutants. *E. coli* contains of the order of 5,000 genes. Therefore in theory a population of about 15,000 mutants is needed to have a 95% chance of containing at least one mutant with a mutation in a given gene. In order to screen a population of this size for *xer* mutants, the Tn5-mutagenised population must first be amplified by a period of growth on kanamycin, prior to transformation with pKS455. This allows the generation of large numbers of pKS455 transformants which can then be screened for resistance to chloramphenicol.

Any mutations that cause a reduction in host viability will be under-represented in the amplified population. It would therefore be difficult to select *xer* mutants if they were less viable than wild type strains. The amplification step could be avoided if a more efficient method, such as electroporation or conjugation,

was used to get the test plasmid into the mutant population. Obviously if further *xer* genes are absolutely essential for host growth, mutations could not be obtained in these genes unless a screen were developed that could pick up conditional lethal *xer* mutants. Despite worries that new *xer* mutants might have a reduced viability, it was decided to use essentially the same mutagenesis and selection strategy as had been employed before.

4.5 Mutagenesis and selection of *xer* mutants

DS953 (a Sup^o strain) was grown to stationary phase in L broth, giving a density of approximately 10^9 cells/ml. To a 2.5 ml aliquot of this culture was added: $MgSO_4$ to 10mM and approximately 2.5 x 10^8 plaque forming units (pfu) of lambda NK467. The mixture was incubated at 30^oC for two hours to allow phage absorption, transposition of Tn5 and expression of Km^r. The cells were washed in 20 ml of L broth, centrifuged and resuspended in 2.5ml. 150 ul of this mutagenised population was plated onto medium containing kanamycin. After overnight growth at 37^oC, approximately 3,000 Km^r colonies were obtained. This represents a transposition frequency of 2 x 10^{-4} Km^r cells per pfu. If each Km^r colony represented an independent transposition event, then the 3,000 colonies on this plate carried a Tn5 inserted in the *E. coli* chromosome on average once every 1,500 bp. This was therefore by no means a complete saturation mutagenesis. The Km^r colonies were washed off the plate in 10 ml of phage buffer. 0.5 ml of the suspension produced was used to inoculate 20 ml of L broth for the preparation of competent cells. These were transformed with pKS455 and plated onto L agar containing ampicillin and kanamycin giving approximately 10^4 Ap^r Km^r colonies. The 3,000 original Km^r colonies were each represented on average about three times in the pKS455 transformant population. Assuming that all of the mutants were transformed with the same probability, it can be calculated (using the Poisson distribution) that 95% of the original 3,000 mutants were represented at least once in the pKS455 transformant

population. The pKS455 transformants were replica plated onto medium containing ampicillin, kanamycin and chloramphenicol. Eleven Ap^r Km^r Cm^r colonies were obtained. The entire mutagenesis and selection scheme is outlined in Fig 4.3.

Obviously the eleven colonies produced were not necessarily independent, but might have resulted from the transformation of descendants of a single Tn5 mutant. Independent mutants could be produced by repeating this procedure on separate aliquots of cells. Four aliquots of DS953 were mutagenised and transformed with pKS455 exactly as described above. Each aliquot gave approximately 2×10^3 Km^r colonies which then yielded approximately 5×10^3 Ap^r Km^r pKS455 transformants. After replica plating a total of 12 Ap^r Cm^r Km^r colonies were obtained from three of the four mutagenesis experiments (Table 4.1).

In all, a total of 23 Ap^r Cm^r Km^r colonies were isolated from four independent mutagenesis experiments. These colonies could either have represented host *xer* mutants unable to delete the *cat* gene from pKS455 or they could have represented plasmid (*cer*) mutations no longer able to lose Cm^r by recombination in a Xer^+ host. Only the eleven mutants, produced in the first mutagenesis experiment, were ever characterised in any detail.

4.6 Initial characterisation of *xer* mutants

The first step in characterising the 11 Ap^r Cm^r Km^r colonies was to examine their plasmid content by agarose gel electrophoresis. This revealed that seven colonies contained only a plasmid of the correct size to be pKS455, as expected for *xer* mutants. The remaining four colonies contained a mixture of two plasmids, one the correct size to be pKS455 (about 80% of the total plasmid DNA) and the other the correct size to be p456 (about 20% of the total plasmid DNA). This leaky Xer^- phenotype was thought to be both novel and interesting.

The next step in characterising these mutants was to see if their Xer^- phenotype was genetically linked to the

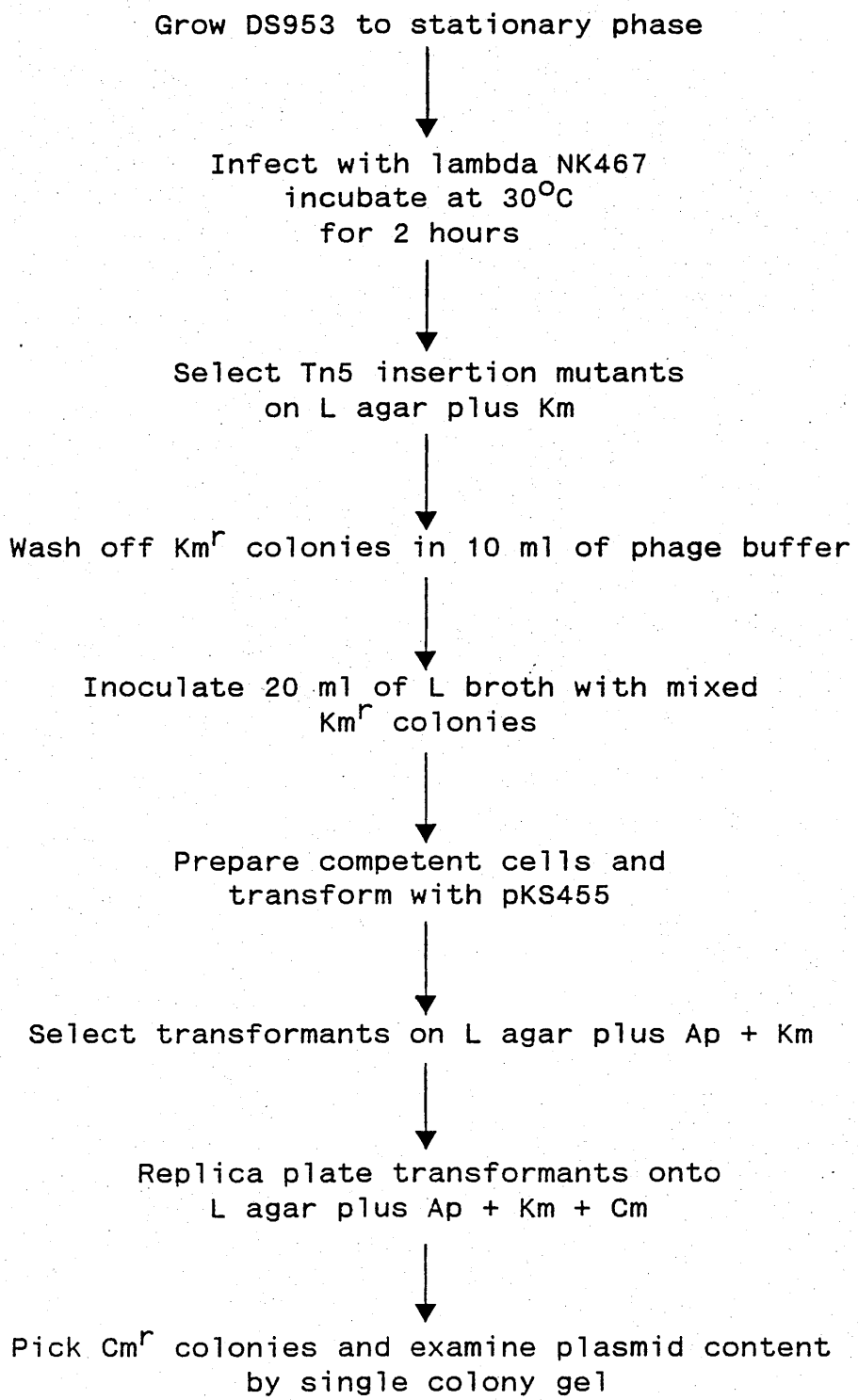


Figure 4.3 Flow chart of mutagenesis and selection of xer mutants.

Table 4.1 Number of possible *xer* mutants obtained from five independent Tn5 mutagenesis experiments.

Mutant population	No. of Tn5 mutants	No. of Ap^r Km^r pKS455 transformants	No. of Ap^r Cm^r Km^r pKS455 transformants
DS953::Tn5 #5	3000	10,000	11
DS953::Tn5 #6	2000	5,000	0
DS953::Tn5 #7	2000	5,000	8
DS953::Tn5 #8	2000	5,000	3
DS953::Tn5 #9	2000	5,000	1

Table 4.2 Co-transduction of Xer^- phenotype with Tn5

$P1_{\underline{kc}}$ lysate of strain	No. of Km^r transductants	No. of Xer^- transductants	%co-transduction	complemented by pGS38	complemented by pCS110
SCX5.1	2	2	100	-	+
SCX5.2	2	1	50	-	+
SCX5.7	2	2	100	-	+
SCX5.9	2	2	100	-	+
SCX5.11	1	0	0	n/a	n/a
SCX5.12	2	1	50	-	+
SCX5.13	2	2	100	-	+
SCX5.4*	4	4	10	-	-

The Xer phenotypes of Km^r transductants were assessed by transformation of transductants with pCS202 followed by agarose gel electrophoresis of plasmid DNA.

* SCX5.4 was the leaky *xer* mutant that retained a mixture of pKS455 and p456. All four transductants from SCX5.4 displayed a similar leaky phenotype when transformed with pCS202.

Tn5 insertion. P1 lysates were made from the seven totally Xer$^-$ strains and from only one of the leaky Xer$^-$ strains. These lysates were then used to transduce the mutations into DS941, by selecting for Tn5 encoded Km^r. Two transductants from each of the totally Xer$^-$ lysates and four from the leaky Xer$^-$ lysate were then transformed with the 2-*cer* plasmid pCS202. The Xer phenotype of each transductant was assessed by gel electrophoresis of plasmid DNA from pooled transformants.

As can be seen in Table 4.2 the Xer$^-$ phenotype cotransduces with Tn5 in all but one case, with a frequency of 50% to 100%. These results demonstrated conclusively that at least seven of the eleven original Ap^r Cm^r Km^r colonies contained genuine Tn5 linked host *xer* mutations and not *cer* or other plasmid mutations. The lack of 100% cotransduction between Tn5 and the Xer$^-$ phenotype of some of the strains could have been caused by the presence of secondary copies of Tn5, not linked to the *xer* mutation.

A mixture of pCS202 and pCS203 (the 1-*cer* deletion derivative of pCS203) was seen in transformants of all four transductants from the leaky Xer$^-$ strain. This partial resolution of pCS202 was very similar to the partial resolution of pKS455 seen in the original mutant.

An experiment was then carried out to ascertain whether the Xer$^-$ phenotype of these seven mutant strains was complemented by cloned copies of either *argR* or *pepA*. A Xer$^-$ transductant from each lysate, already containing pCS202, was transformed with either pCS110 (a *pepA*$^+$ plasmid) or pGS38 (an *argR*$^+$ plasmid). Plasmid DNA was recovered from transformants and analysed by agarose gel electrophoresis. Complementation was said to have occurred if there was no band comigrating with pCS202 and a new band had appeared which comigrated with pCS203. All six totally Xer$^-$ transductants tested were found to be complemented by pCS110 and not by pGS38 (Table 4.2). They were therefore presumed to be mutant in the *pepA* gene and were not analysed further. It is quite likely that they were not all independent mutants.

The leaky *xer* mutation, P1 transduced into DS941,

appeared not to be complemented by either *pepA* or *argR*. Transductants from the leaky *xer* mutant containing pCS202 retained a band which comigrated with pCS202, and showed no significant increase in the amount of pCS203 present when transformed with either pCS110 or pGS38. This mutant appeared to belong to a new complementation group. The mutation was therefore given the name *xerC1*::Tn5.

The phenotype of the *xerC1* mutation could be distinguished from that of *pepA* and *argR* in another way. Recombination at the type II hybrid, which does not require the function of either *pepA* or *argR*, was found to be substantially reduced in DS941 *xerC1*. DS941 *xerC1* was transformed with pSDC110, which contains two directly repeated copies of the type II (ColE1 *cer* - ClodF13 *parB*) hybrid. Intramolecular recombination (deletion) at the type II hybrid site was found to be substantially reduced in DS941 *xerC1*. Intermolecular recombination (multimerisation) appeared also to be reduced by the *xerC1* mutation. A comparison of the behaviour of the reporter plasmids pKS455, pCS202 and pSDC110 when transformed into DS941, DS941 *argR*, DS941 *pepA* and DS941 *xerC1* is shown in Fig 4.4.

4.7 Mapping and isolation of *xerC1*

The Tn5 insertion within *xerC1* was genetically mapped by interrupted mating. The *xerC1* mutation was first P1 transduced into CGSC4311. Transductants were selected for resistance to kanamycin and then transformed with pKS455 to check that they displayed the leaky Xer$^-$ phenotype. CGSC4311 is an Hfr strain, carrying the F origin of transfer at a map position of approximately 14 minutes. Chromosomal markers are transferred from this origin in an anticlockwise direction. CGSC4311 *xerC1* was mated with a Nal^r derivative of DS903, which carries the following mutations *argE* (map position 90 min), *hisG* (44 min), *leuB* (2 min) and *thr-1* (0 min). CGSC4311 is wild type at all of these loci and should therefore transfer these markers to the recipient in the order *leuB*, *thr-1*, *argE* and last of all *hisG*.

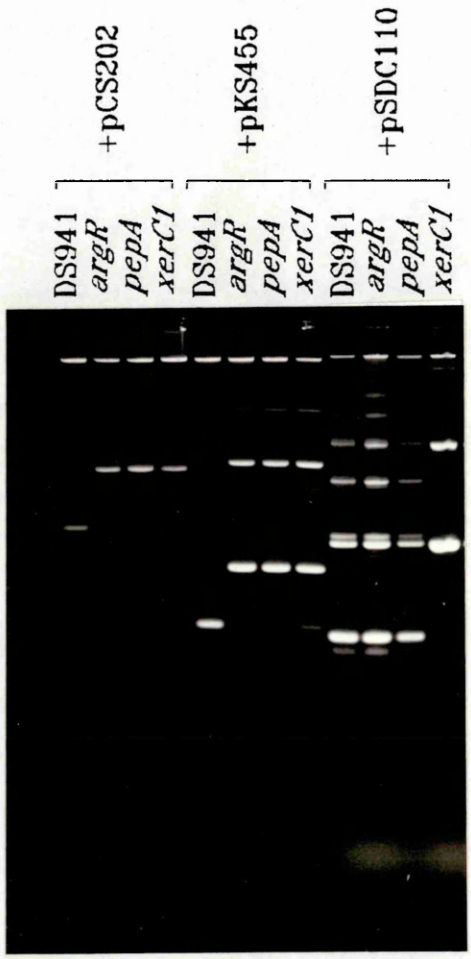


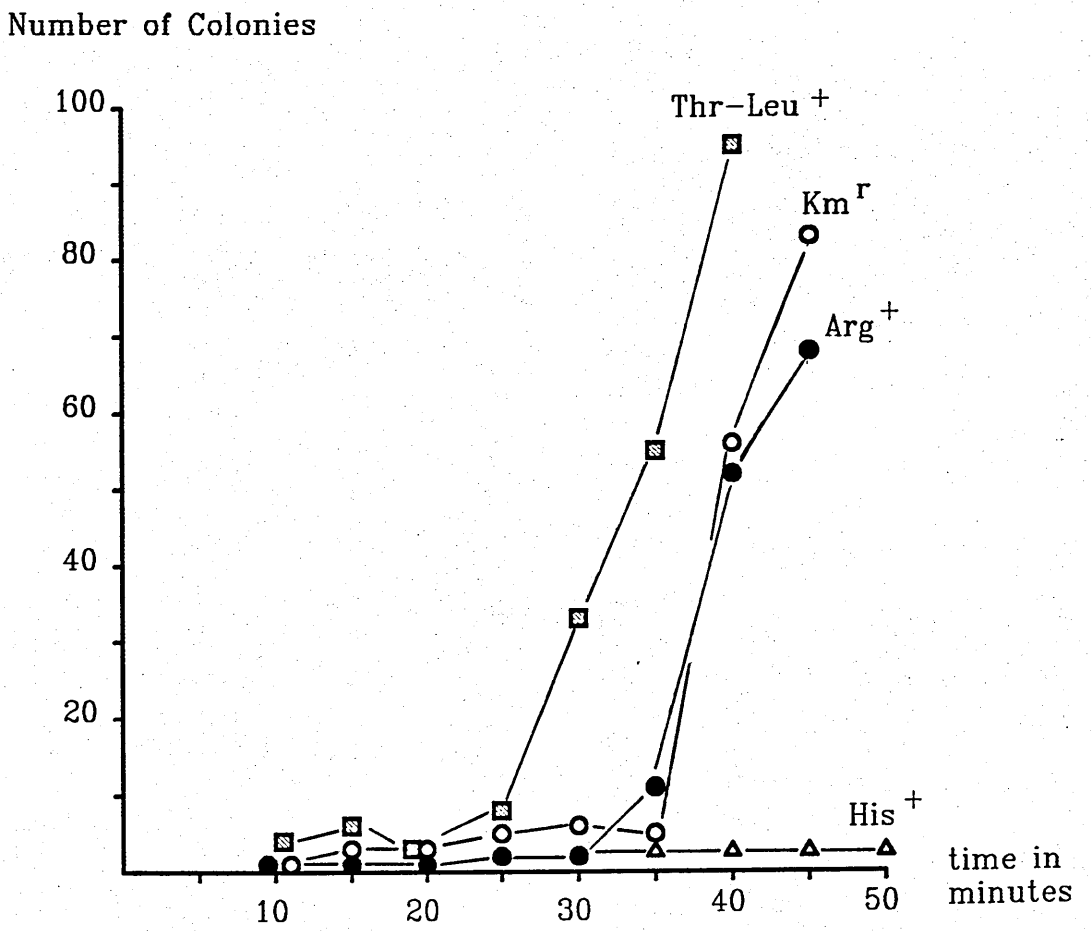
Figure 4.4 cer **mediated site specific recombination in** Xer^+ **and** Xer^- *E. coli* **strains.** DS941, DS941 *argR* (*xerA9*::fol), DS941 *pepA* (*xerB1*) and DS941 *xerC1* were transformed with the reporter plasmids pKS455, pCS202 and pSDC110. Plasmid DNA was isolated by the boiling method and run on a 1.2% agarose gel. pKS455 and pCS202 both contain two directly repeated copies of the wild-type *cer* site. pSDC110 contains two directly repeated copies of the type II ColE1 *cer* ClоDF13 *parB* hybrid (Sau3A - AluI fragment). Intramolecular site-specific recombination acting on pKS455, pCS202 and pSDC110 yield p456, pCS202 and p111 respectively.

Both donor and recipient were grown to an optical density (A_{600}) of about 0.4 and then transferred to a 37^{o} non-shaking water bath for 30 minutes. 1 ml of the donor was gently mixed with 10 ml of the recipient. 0.5 ml samples were taken at five minute intervals and interrupted by dilution in 5 ml of ice cold 1 x D&M salts supplemented with 20 ug/ml nalidixic acid, followed by vigorous vortexing. Samples were concentrated by centrifugation and resuspended in 0.5 ml 1 x D&M salts. 30 ul from each time point was plated onto medium that selected for either Thr^{+} - Leu^{+} or Arg^{+} or His^{+} or Km^{r} recipients (all plates contained nalidixic acid). The time of entry of known markers was much later than expected but the results shown in Fig 4.5 indicated that the Tn5 insertion within *xerC1* probably maps within 10 minutes of the *argE* gene at 90 minutes.

The Tn5 insertion within *xerC1* was physically mapped by Southern blot analysis. Chromosomal DNA was prepared from DS941 *xerC1*, treated with various restriction endonucleases and run on an agarose gel. DNA was transferred to a nylon membrane and probed with a mixture of ^{32}P labelled lambda DNA (to show up the lambda size markers) and pCS80 (a pBR322::Tn5 derivative) to show up chromosomal fragments containing Tn5. The results of this blot are shown in Fig 4.6 and summarised in Table 4.3. These data showed that DS941 *xerC1* carried a single chromosomal copy of Tn5.

Tn5 contains no internal EcoRI sites. It was therefore decided to clone the 14 kbp Tn5 containing EcoRI fragment seen by Southern blotting (Fig 4.6). This fragment should contain the 5.8 kbp Tn5 within a chromosomal fragment of approximately 8 kbp. It was hoped that this fragment would contain all or part of the *xerC* gene. Total DNA from DS941 *xerC1* was cut with EcoRI and ligated into pBR322 which had been treated with EcoRI and CIP to prevent ligation of the vector to itself. This ligation mix was transformed into DS941 and one Ap^{r} Km^{r} colony was obtained. The plasmid from this Ap^{r} Km^{r} colony was named pSDC100. pSDC100 plasmid DNA was isolated, cleaved by various restriction endonucleases and analysed

A)



B)

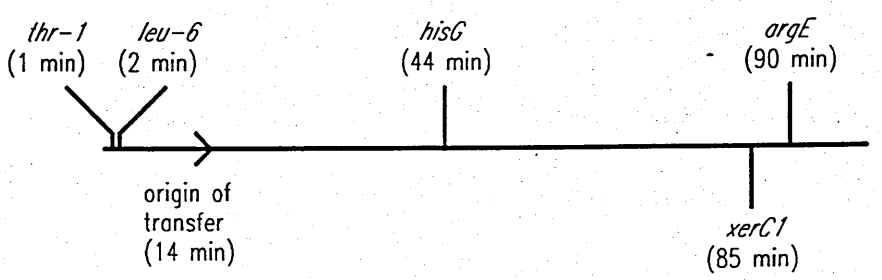


Figure 4.5 Interrupted mating between CGSC4311 xerC1 and DS903 nal^r. CGSC4311 xerC1 was mated with DS903 nal^r as described in the text. The number of (Thr^+ Leu^+ Nal^r), (Arg^+ Nal^r), (Km^r Nal^r) and (His^+ Nal^r) colonies obtained at various time points is shown in (A). The genetic map of CGSC4311, with the location of the Hfr origin of transfer, is shown in (B).

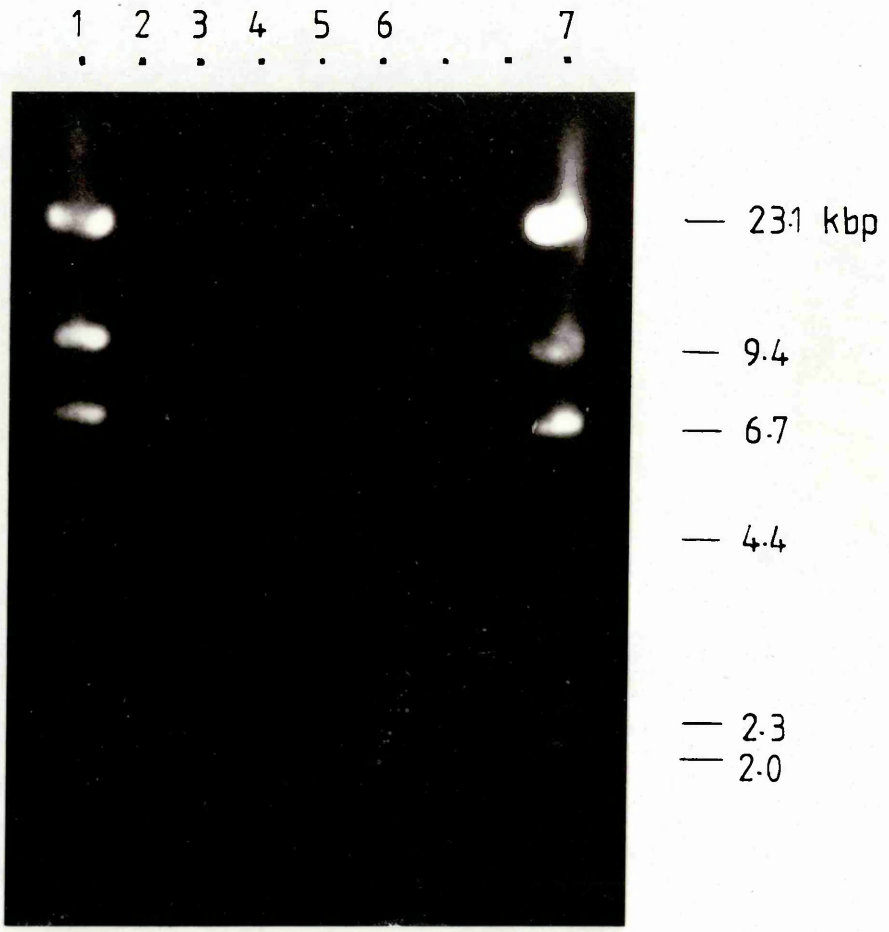


Figure 4.6 Southern blot of chromosomal DNA from DS941 xerC1. Chromosomal DNA from DS941 xerC1 was cut with various restriction enzymes, run on a 0.85% agarose gel and transferred to a nylon membrane. The blot was then probed with a mixture of nick translated pCS80 (pBR322::Tn5) and lambda DNA.

Lane	Sample
1	Lambda HindIII
2	DS941 xerC1 BamHI
3	DS941 xerC1 EcoRI
4	DS941 xerC1 HindIII
5	DS941 xerC1 KpnI
6	DS941 xerC1 SphI
7	Lambda HindIII

Table 4.3 Sizes of Tn5 containing restriction fragments from DS941 xerC1

Enzyme	No. of sites in Tn5	Fragment sizes (kbp)		
BamHI	1	7*	6	
EcoRI	0	14*		
HindIII	2	>20	3.6*	3.4*
KpnI	0	>20		
SphI	1	14	6*	

Those fragments marked with an asterisk (*) are completely contained within pSDC100. The 3.4 kbp HindIII fragment is an internal fragment from Tn5.

Table 4.4 The number of recognition sites identified in pSDC100 for various restriction enzymes.

Enzyme	No. of sites in pSDC100	Predicted No. of sites in: pBR322	Tn5	chromosomal DNA
BamHI	4	1	1	2
HindIII	4	1	2	1
EcoRI	2	1	0	2*
EcoRV	9	1	0	8
KpnI	1	0	0	1
PstI	8	1	4	3
PvuII	6	1	4	1
SalI	4	1	1	2
SphI	3	1	1	1

*The two EcoRI sites are at the boundaries between vector and chromosomal sequences.

by electrophoresis on agarose and acrylamide gels. One agarose and one acrylamide gel used in this analysis are shown in Fig 4.7. As can be seen, restriction with EcoRI produced the expected 14 kbp Tn5-containing chromosomal fragment and a 4.6 kbp vector fragment.

A breakdown of the number of various restriction sites found in pSDC100 and the number of such sites expected in both pBR322 and Tn5 is shown in Table 4.4. Using these data, and the sizes of restriction fragments estimated from plots of mobility against log [fragment length] of known markers, a restriction map of pSDC100 was deduced (Fig 4.8). The construction of this map was greatly aided by the known restriction map of two components of pSDC100, namely pBR322 and Tn5.

Restriction enzyme cleavage of pSDC100 with EcoRI yielded the 14 kbp fragment expected from the Southern blot of DS941 *xerC1* chromosomal DNA. Further comparison of the deduced restriction map of pSDC100 (Fig 4.8) with the results of the Southern blot experiment (Fig 4.6 and Table 4.3) confirmed that the desired, *xerC* linked, copy of Tn5 had been cloned in pSDC100.

A restriction map of the wild type 8.3 kbp EcoRI fragment, which was thought to contain at least part of *xerC*, was deduced by subtracting the known restriction map of Tn5 from the pSDC100 restriction map. Comparison of this deduced map of the *xerC* region with the complete restriction map of the *E. coli* chromosome (Kohara *et al*, 1987) revealed a very convincing match at a map position of 85 minutes (co-ordinates 3,716 to 3,724). Further refinement of this map position was made possible by comparison with the detailed restriction map of the *ilv - metE - udp* region of the *E. coli* chromosome (Aldea *et al*, 1988; see Fig 4.9). This allowed the position of Tn5 in pSDC100 (and therefore by implication in DS941 *xerC1*) to be accurately determined. The *xerC1* linked copy of Tn5 is located between the genes for adenylate cyclase (*cya*) and the SOS-inducible DNA helicase II (*uvrD*), in the 85 minute region of the *E. coli* chromosome. The extension of the restriction map of the *xerC* region, outside of the 8.3 kbp EcoRI fragment contained in pSDC100, confirmed the

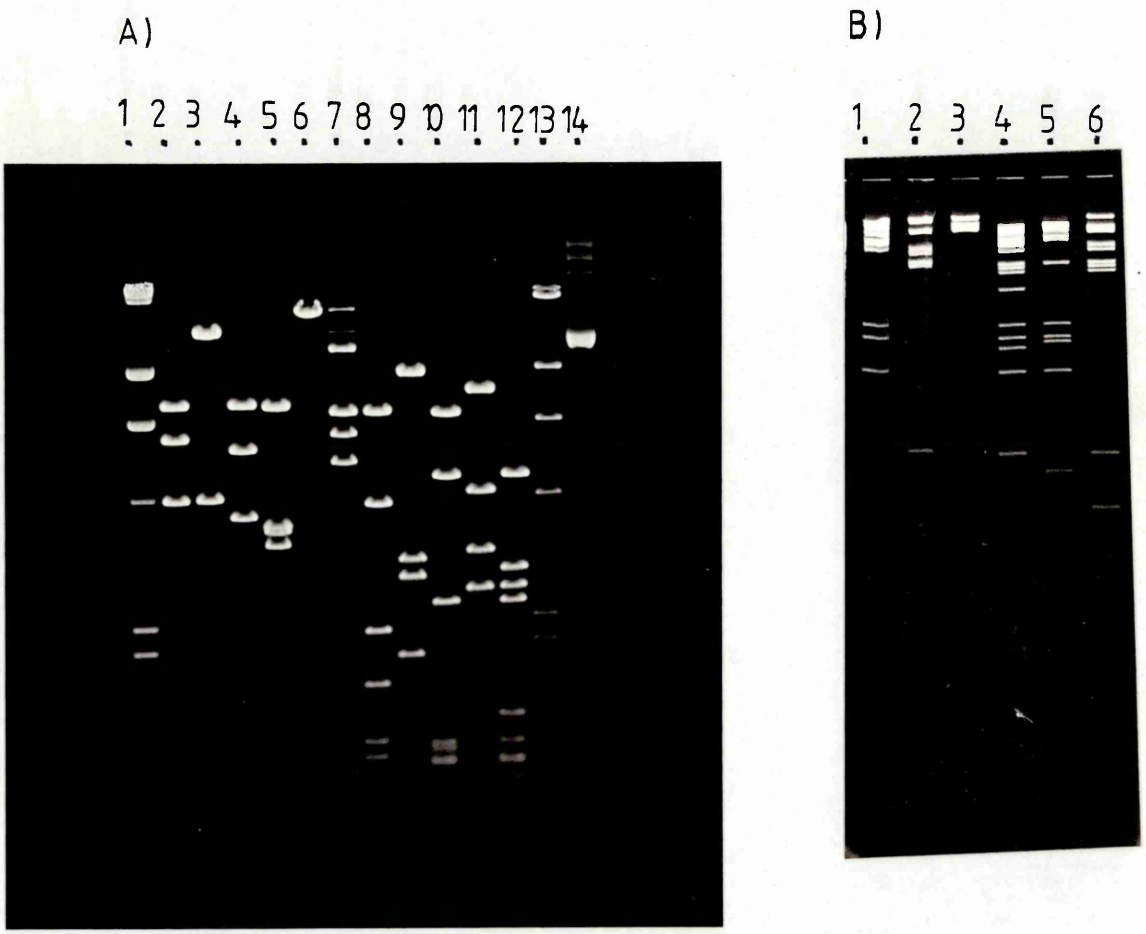


Figure 4.7 Restriction analysis of pSDC100. pSDC100 plasmid DNA was cut with various restriction enzymes and run on a 1.2% agarose gel (A) and on a 5% acrylamide gel (B).

A) Lane	Sample
1	Lambda HindIII
2	pSDC100 BamHI
3	pSDC100 EcoRI
4	pSDC100 BamHI EcoRI
5	pSDC100 HindIII
6	pSDC100 KpnI
7	pSDC100 SphI
8	pSDC100 EcoRV
9	pSDC100 PvuII
10	pSDC100 PstI
11	pSDC100 SalI
12	pSDC100 PstI SalI
13	Lambda HindIII
14	pSDC100 uncut

B) Lane	Sample
1	pSDC100 EcoRV
2	pSDC100 PstI
3	pSDC100 SalI
4	pSDC100 EcoRV PstI
5	pSDC100 EcoRV SalI
6	pSDC100 PstI SalI

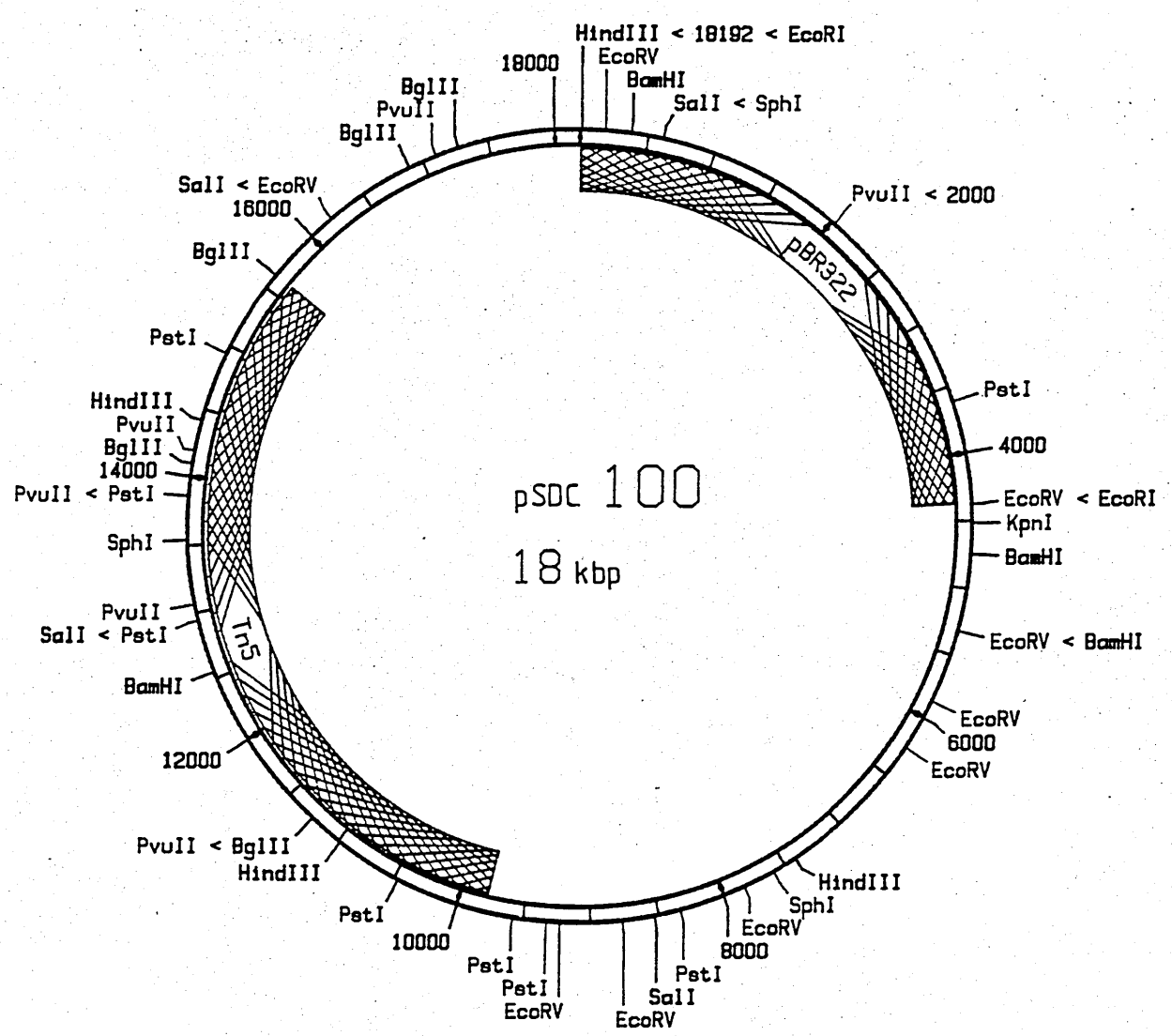


Figure 4.8 Restriction map of pSDC100 The positions of Tn5 and pBR322 within pSDC100 are shown. The presence of a "<" between restriction enzyme labels indicates that the two enzymes cut at closely linked sites (eg EcoRV < EcoRI indicates that the EcoRV cut-site is located in a clockwise direction from the EcoRI cut-site).

existence of chromosomal BamHI, HindIII and KpnI sites outwith pSDC100 in positions consistent with the Southern blot data shown in Fig 4.6 (See Fig 4.9).

4.8 Defining the wild type *xerC* gene

The entire *ilv - metE - udp* region of the *E. coli* chromosome has been restriction mapped and cloned into multicopy plasmids (Aldea *et al*, 1988). S. Kushner kindly supplied us with plasmids covering the *cya - uvrD* region. Mary Burke tested these plasmids (pMAK101, pMAK102 and pVMK42) to see if they complemented the *xerC1* mutation. The plasmids were transformed into DS941 *xerC1* already carrying pCS202. Plasmid DNA was isolated from transformants and run on an agarose gel to look for resolution of pCS202 to pCS203. Only pMAK101 was found to complement *xerC1*. This plasmid contains the same 8.3 kbp EcoRI fragment found in pSDC100 without a copy of Tn5 disrupting the *xerC* gene.

Two sub-clones were constructed which contained fragments from pMAK101 in the vector pTZ18R. pSDC106, carrying the 1.8 kbp SalI fragment (with the Tn5 insertion point located approximately in the centre), was found not to complement *xerC1*. pSDC102, carrying the 3.6 kbp HindIII - BglII fragment, was found to complement *xerC1* (Fig 4.9).

In order to further define the *xerC* complementing region carried by pSDC102, it was decided to carry out exonuclease III deletions. Chromosomal sequences carried by pSDC102 were deleted, starting from the end furthest from the Tn5 insertion point (the HindIII end). Exonuclease III deletion was carried out by the method of Henikoff (1987). pSDC102 contains unique BamHI and SstI sites in the polylinker, adjacent to the HindIII end of the insert (Fig 4.10). Treatment with SstI and BamHI produced one end (SstI) with a protruding 3' OH which blocked exonuclease III degradation of the vector, and another end (BamHI) with a recessed 3' OH which allowed exonuclease III to degrade the insert progressively on one strand in a 3' to 5' direction. By taking samples and stopping the exonuclease III reaction at various time

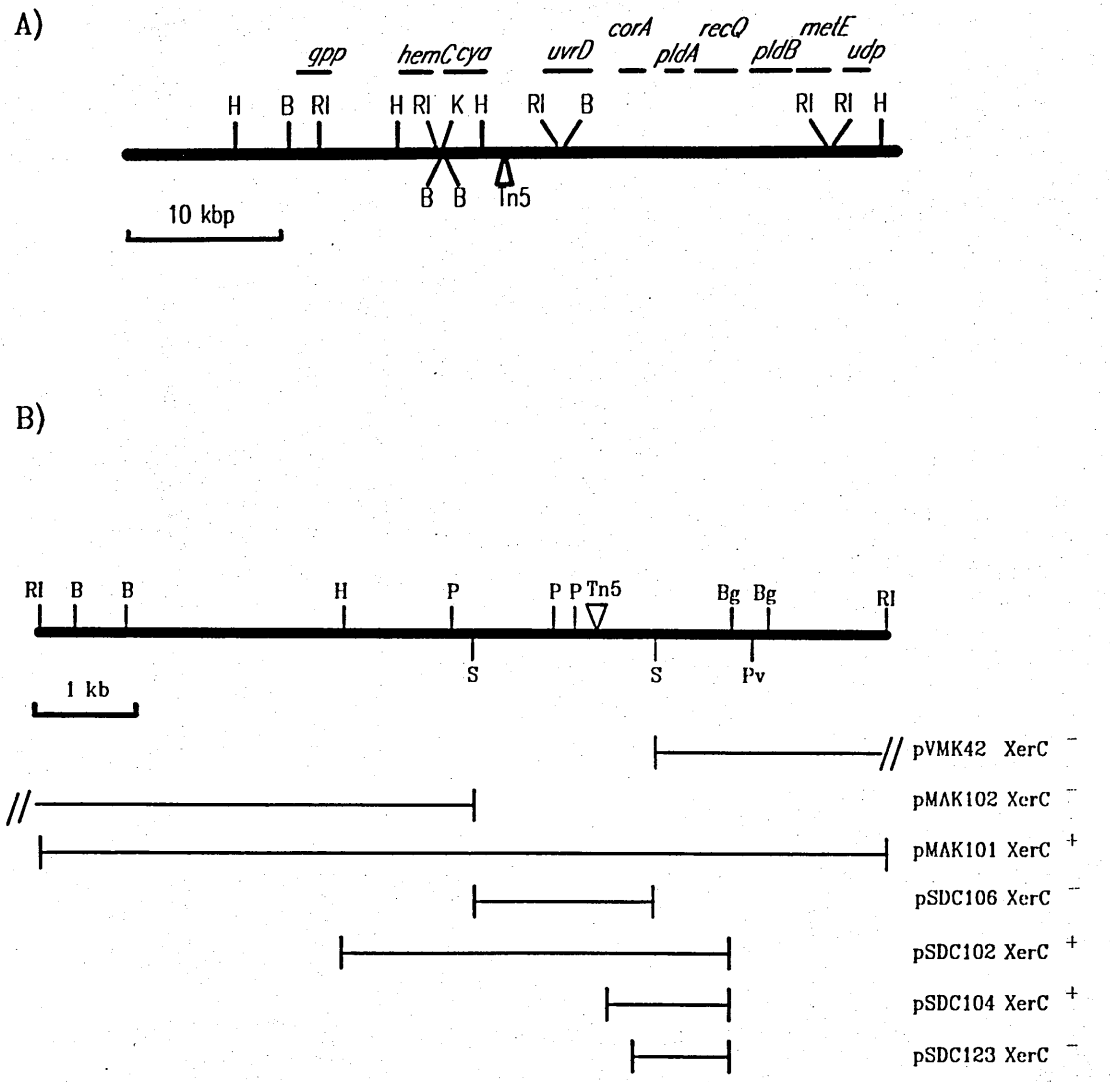


Figure 4.9 Genetic and physical map of the xerC region of the E. coli chromosome. (A) Shows the restriction map of 50 kbp of the E. coli chromosome and the location of genes known to map to this region (Adapted from Kohara et al, 1987 and Aldea et al, 1988). (B) Shows an enlarged restriction map of the 8.3 kbp EcoRI fragment into which Tn5 inserted to give xerC1::Tn5. Below this map is shown the extent of chromosomal fragments carried by various plasmids. The XerC phenotype of these plasmids was assessed by complementation of the xerC1 mutation.

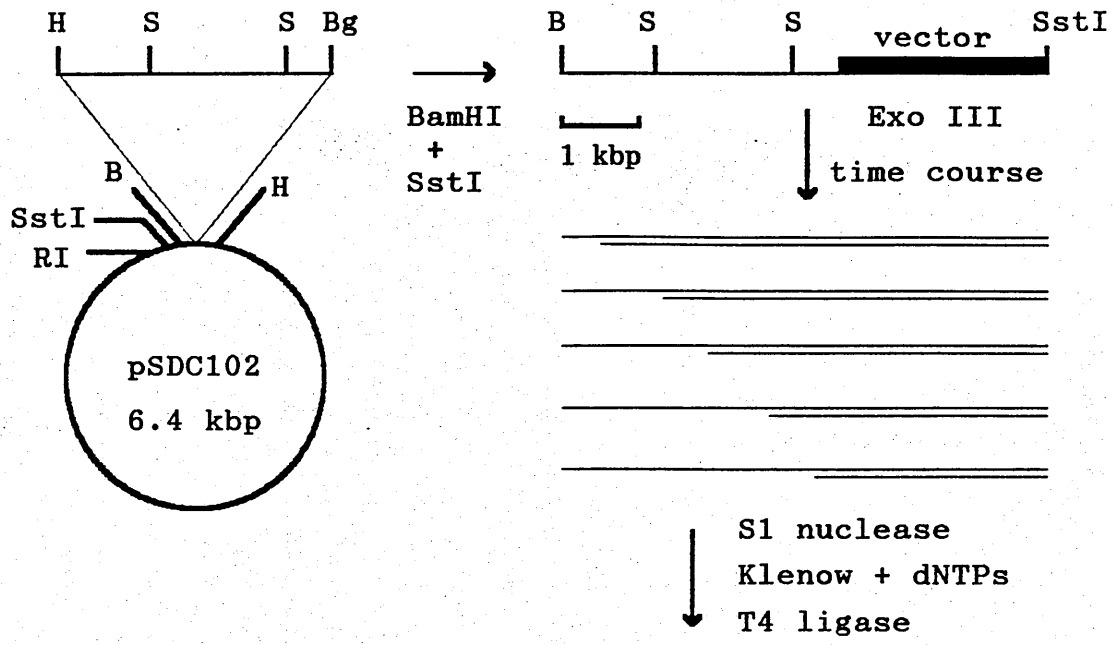
points, progressively larger deletions were obtained. Single stranded tails were removed by the action of S1 nuclease, which specifically degrades single stranded DNA. Blunt ends were then produced by the action of the Klenow fragment of DNA polymerase I in the presence of all four deoxynucleotide triphosphates. DNA molecules, containing deletions into insert sequences but not into vector sequences, were recircularised by the action of DNA ligase and transformed into DS941 *xerC1*.

After screening a number of plasmids produced in this way, a series of pSDC102 deletion derivatives was obtained. The deletion end points of this series covered the 3.8 kbp HindIII to BglII fragment in intervals of 200-300 bp (Fig 4.10). DS941 *xerC1* isolates containing these plasmids were transformed with pCS202. Transformants were patched out, DNA was obtained by the single colony lysis technique and separated by electrophoresis on a 0.85% agarose gel (Fig 4.11). Complementation of the Xer^- phenotype was assessed by the complete disappearance of pCS202 and an increase in intensity of the band corresponding to pCS203.

The smallest plasmid found to complement *xerC1*, pSDC102 $\triangle$ 4.1 (also called pSDC104), contained a chromosomal fragment that extended 1,232 bp from the BglII site towards the HindIII site. The next smallest deletion derivative of pSDC102, pSDC102 $\triangle$ 6.2 (pSDC123), contained a fragment 269 bp smaller than that in pSDC104 and did not complement *xerC1*.

4.9 Discussion

Site-specific recombination at *cer* maintains ColE1 in a monomeric state, thus ensuring its stable inheritance. ColE1 carries no protein coding sequences essential for this monomerising site-specific recombination (Summers and Sherratt, 1984). Previous work had led to the isolation of two *E. coli* chromosomal genes (*pepA* and *argR*) which are absolutely required for *cer* site-specific recombination (Stirling *et al*, 1988b; Stirling *et al*, 1989). Both of these genes have roles in *E. coli* apart



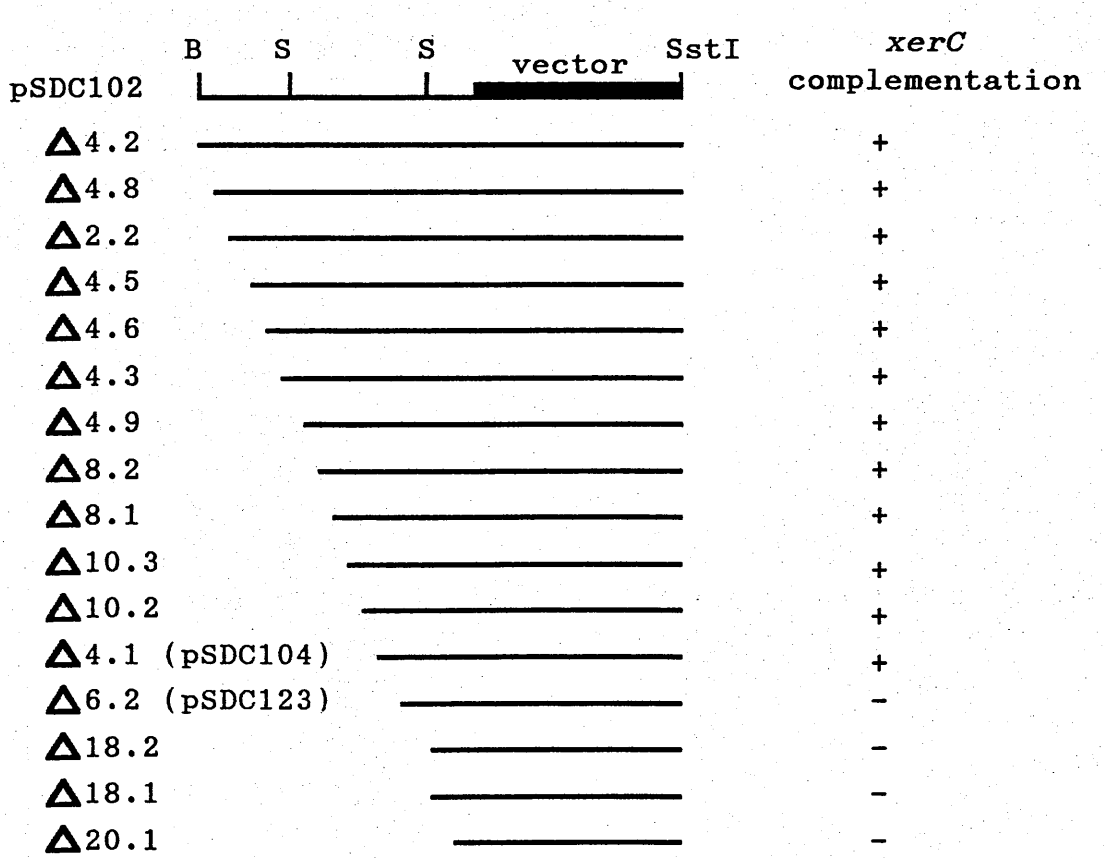


Figure 4.10 Exonuclease III deletions of pSDC102. The XerC$^+$ plasmid pSDC102 was cut with BamHI and SstI and then incubated with exonuclease III. Blunt ends were produced by the actions of S1 nuclease and the Klenow fragment of DNA polymerase I. DNA molecules were then circularised by T4 ligase. The deletion derivatives of pSDC102 produced were then tested for complementation of xerC1 (see Fig 4.11).

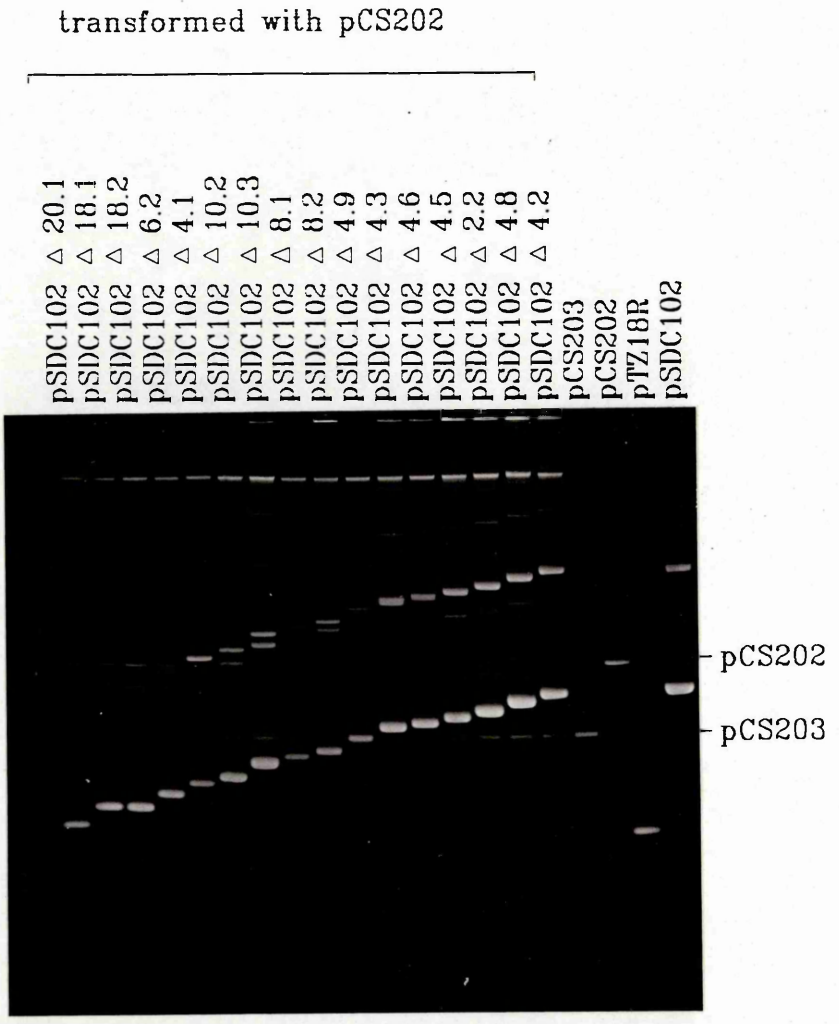


Figure 4.11 Complementation of <u>xerC1</u> with the deletion derivatives of pSDC102. DS941 <u>xerC1</u> isolates already carrying the deletion derivatives of pSDC102 were transformed with the compatible reporter plasmid pCS202. Plasmid DNA was visualised on a 0.85% agarose single colony gel.

from their action on *cer*: *argR* controls expression of the arginine biosynthesis regulon (reviewed by Glansdorff, 1987) and *pepA* encodes a peptidase which can liberate amino acids from small peptides which are either produced by protein degradation or imported from the medium (reviewed by Miller, 1987).

This chapter reports the isolation of a strain carrying a mutation in a new gene, *xerC*, involved in *cer* site-specific recombination. This *xerC* mutant strain has substantially reduced levels of site-specific recombination at *cer*. The *xerC* gene has been cloned and mapped, and is totally distinct from both *pepA* and *argR*.

The *xerC1* mutation does not completely abolish recombination at *cer*. There are two possible explanations for this: either the *xerC* gene product is not absolutely required for *cer* site-specific recombination (but substantially increases the rate of such recombination) or the *xerC* gene product is essential for recombination at *cer* but the *xerC1* mutation is not a null allele. Results discussed in the next chapter will show that the second of these possibilities is in fact true and that *xerC* is absolutely required, along with *pepA* and *argR*, for *cer* site-specific recombination. There is also some evidence to suggest that *xerC* might have a cellular role in *E. coli*.

The *xerC1* mutation substantially reduces recombination at the type II (ClоDF13 *parB* ColE1 *cer*) hybrid, which does not require either *pepA* or *argR* for recombination. This type II site is functionally much smaller than the wild type *cer* site, requiring at most 50 bp around the crossover region (Summers, 1989). This suggests that XerC acts at or near the crossover site of *cer*, making XerC a candidate for being the *cer* recombinase.

The strong possibility remains that yet more *E. coli* genes are required for *cer* site-specific recombination. The strategy described in this chapter could be used to select further *xer* mutants. In fact a further 12 possible *xer* mutants have already been isolated. It would be interesting to investigate whether these all belong to the

three known *xer* complementation groups. However any further genes required for *cer* site-specific recombination, which are also essential for *E. coli* growth, would be very difficult to identify by genetic approaches. The development of an *in vitro* recombination system would be useful in the search for further *E. coli* factors involved in *cer* site-specific recombination.

CHAPTER FIVE

THE NUCLEOTIDE SEQUENCE OF *xerC*

5.1 Introduction

In the work described in chapter 4, a minimal *xerC* complementing fragment of approximately 1.2 kbp was isolated. The chromosomal location of this fragment was accurately defined by restriction analysis to between the genes *cya* and *uvrD* in the 85 minute region. Another gene that maps to this interval is *dapF*, encoding diaminopimelate epimerase (Richaud *et al*, 1987). This enzyme catalyses the conversion of L,L-diaminopimelate to the *meso* (D,L-) form. *meso*-diaminopimelate is a direct precursor of lysine and is also an essential cell wall component. The *dapF* gene has been cloned and sequenced (Richaud *et al*, 1987; Richaud and Printz, 1988), and it was therefore possible to locate it on the restriction map of the *cya* to *uvrD* region. The gene order in this region was found to be *cya*, *dapF*, *xerC*, *uvrD* (Fig 5.1). The *uvrD* gene has also been sequenced (Finch and Emerson, 1983; 1984; Easton and Kushner, 1983). Examination of the location of these sequences on the restriction map of the *xerC* region showed that there was an unsequenced region of approximately 1.8 kbp between *dapF* and *uvrD* (Fig 5.1). Since the minimal *xerC* complementing fragment was contained within this unsequenced region, it became a priority to determine the nucleotide sequence spanning the gap between *dapF* and *uvrD*.

5.2 Sequencing of the *xerC* gene

The sequence of most of one strand (the top strand) of the *xerC* region was determined from the exonuclease III deletion derivatives of pSDC102 shown in Table 5.1a. The production of these deletion derivatives was described in Section 4.8 and is represented diagrammatically in Fig 4.10. The chromosomal fragments from these deletion derivatives were cut out as HindIII - EcoRI fragments and cloned into M13mp19 also cut with HindIII and EcoRI (Fig 5.2). Single stranded templates were prepared from these recombinant phage and sequenced using the reagents from a Sequenase kit. Overlaps between the sequence of

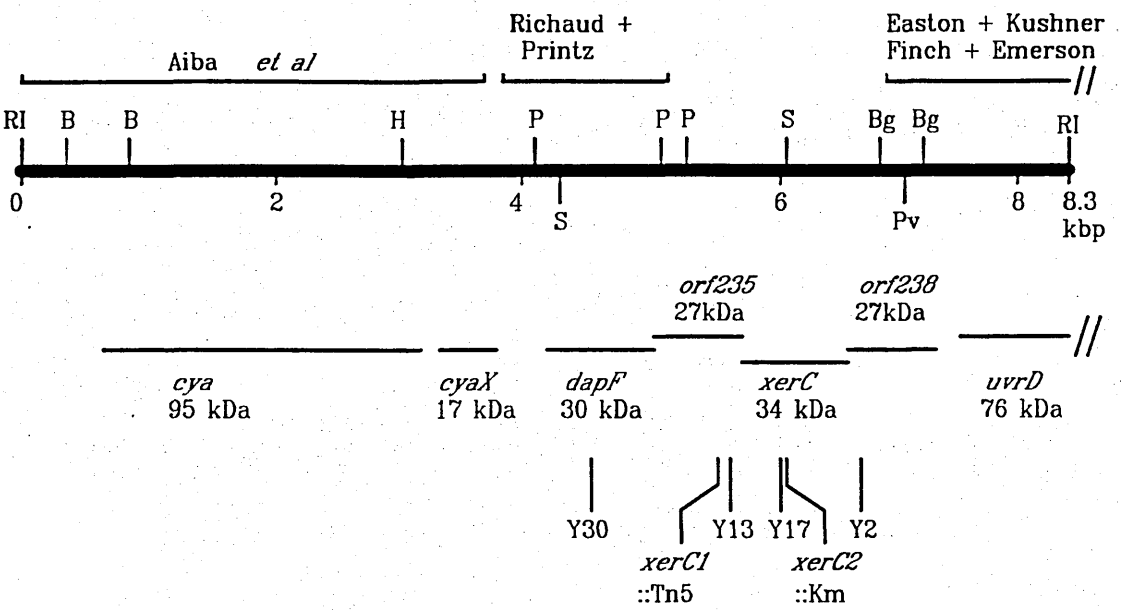


Figure 5.1 Restriction map of the xerC region of the E. coli chromosome. The top of the figure shows the extent of published sequenced in the xerC region. The cya sequence has been determined by Aiba et al (1984), the dapF sequence has been determined by Richaud and Printz (1988) and the sequence of uvrD has been determined by Easton and Kushner (1983) and Finch and Emerson (1983; 1984). The positions of the genes identified in this region are shown. The molecular weights were calculated from sequence data. The positions of insertion of Tn5 in xerC1 and of the Km^r cassette in xerC2 are shown. The sites of insertion of Mu dII PR13 (Y30, Y13, Y17 and Y2) are also indicated.

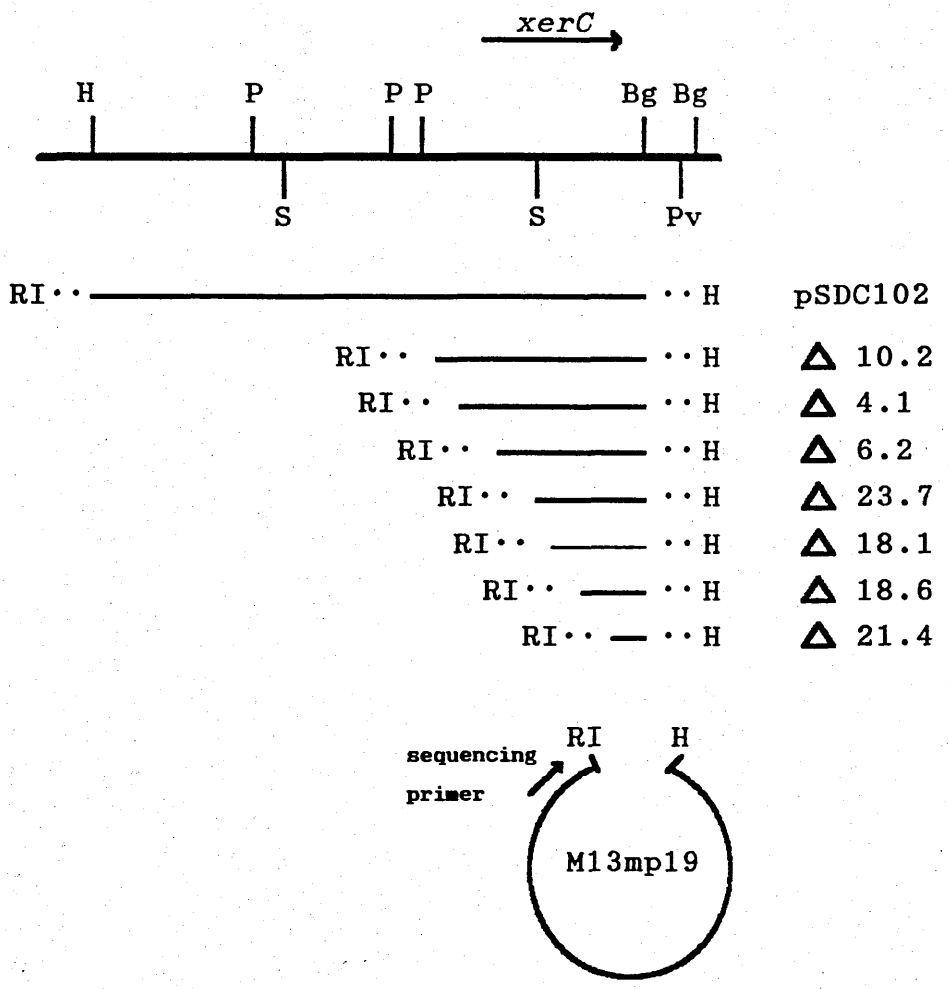


Figure 5.2 Strategy used to determine the sequence of the plus strand of xerC. Exonuclease deletion derivatives of pSDC102, produced as shown in Fig 4.10, were cloned into M13mp19 and sequenced .

successive deletions were found by eye and compared using programs from the University of Wisconsin GCG sequence analysis package running on the Glasgow University VAX (VMS3). The sequence was assembled using the SEQED program, producing 1,391 bp of continuous sequence, extending from the pSD102 Δ 10.2 endpoint at nucleotide 417 to the BglII site at nucleotide 1,807 (Table 5.1a). (All sequence coordinates quoted in this chapter refer to the numbering used in Fig 5.5)

In order to determine the sequence of the bottom strand of the *xerC* region and simultaneously to determine the exact Tn5 insertion point in *xerC1* it was decided to clone an expected 2.5 kbp HindIII - BglII fragment from pSDC100 into M13mp19 (Fig 5.3). However, restriction of pSDC100 with HindIII and BglII did not yield the expected 2.5 kbp fragment, yielding instead one fragment of 1.2 kbp and another of 1.3 kbp. This was due to the presence of an unexpected BglII site in pSDC100, located at or very near the right hand end (as drawn in Fig 5.3) of Tn5. The sequences of both Tn5 ends are 5'AG 3'. Therefore, when Tn5 inserts next to the tetra nucleotide sequence 5'ATCT3' a BglII site (AGATCT) is generated. In order to confirm that this is how the extra BglII site in pSDC100 was produced, the 1.2 kbp HindIII - BglII fragment from pSDC100 was cloned into M13mp19 cut with BamHI and HindIII, to create mSD12 (Fig 5.3). Single stranded DNA produced from mSD12 was sequenced and was found to contain a BglII site next to a Tn5 end exactly as expected.

In order to determine the exact location of the extra BglII site in pSDC100 and, by implication, the exact site of insertion of Tn5 in *xerC1*, the 1,273 bp BglII fragment from pSDC100 was sequenced from the left hand end. The 1,273 bp BglII fragment was first ligated into pTZ19R in the BamHI site to create pSDC103. The insert from this plasmid was ligated into M13mp19 as an EcoRI - HindIII fragment, to create mSD14 (Fig 5.3). Single stranded template was prepared from mSD14 and sequenced. This revealed that the BglII site had been created by Tn5 insertion next to the sequence ATCT located at nucleotide 532 (Fig 5.5). The Tn5 insertion point in *xerC1* was thus

Table 5.1 Derivation of the *xerC* sequence

A) Plus strand

Clone	Sequence co-ordinates	Gel(s)
mAS1	69 - 326	n/a
mMB19.3	390 - 432	n/a
delta 10.2	417 - 680	5 & 7
delta 4.1	576 - 878	6 & 7
delta 6.2	846 - 1141	6 & 7
delta 23.7	1108 - 1308	9 & 10
delta 18.1	1161 - 1461	6 & 7
delta 18.6	1440 - 1631	9 & 10
delta 21.4	1582 - 1807	8 & 12
mSD15	1804 - 2120	20

Gaps in plus strand 327 - 390 and 2120 - 2197.

B) minus strand

Clone	Sequence co-ordinates	Gel(s)
mSD16	2197 - 1911	20 & 21
mMB9	1986 - 1791	n/a
mSD19 (Y2)	1873 - 1792	23
mSD13	1807 - 1515	11 & 13
mSD13 delta 5a	1434 - 1355	14
mSD13 delta 2e	1413 - 1285	16
mSD13 delta 2a	1382 - 1116	14 & 15
mSD13 delta 3b	1186 - 996	14 & 15
mSD18 (Y17)	1058 - 944	23
mSD13 delta 4b	975 - 748	14 & 15
mSD13 delta 5a	775 - 678	14 & 15
mSD17 (Y13)	741 - 658	23
mSD13 delta 7b	688 - 532	14 & 15
mMB19.2	595 - 429	n/a
mMB19.3	393 - 321	n/a
mAS2	326 - 116	n/a

Gaps in minus strand 69 - 116, 393 - 429 and 1434 - 1515.

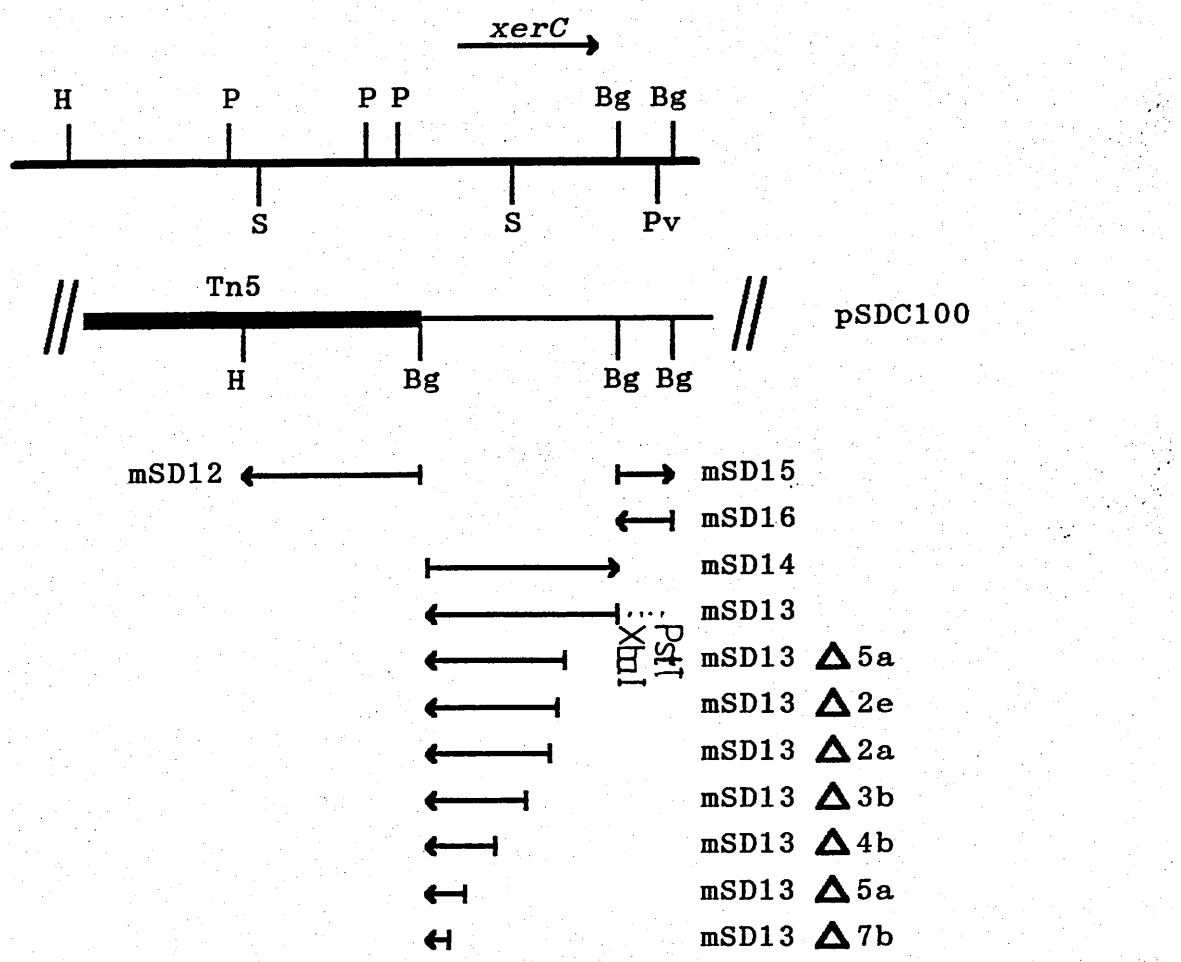


Figure 5.3 Strategy used to determine the sequence of the minus strand of xerC and also to determine the Tn5 insertion site in xerC1. The fragments shown were cloned from pSDC100 into M13mp18 and M13mp19 to produce mSD12, mSD13 and mSD14. The arrow heads indicate the direction in which these fragments were sequenced. The 389 bp BglII fragment from pVMK42 was cloned into m13mp19 to produce mSD15 and mSD16. Exonuclease III was used as described in the text to obtain the deletion derivatives of mSD13 shown.

accurately located.

The insert from pSDC103 was then cloned into M13mp18 to create mSD13 (Fig 5.3). Double stranded (RF form) DNA was made from mSD13 and cut with the restriction enzymes XbaI and PstI. A set of nested deletions was then made using exonuclease III as previously described (Fig 5.3). Sequence was obtained from mSD13 and the deletion derivatives of mSD13 shown in table 5.1b, giving the bulk of the sequence of the bottom strand between the BglII site and the Tn5 insertion point.

The 389 bp BglII fragment from pVMK42 was cloned into M13mp19 in both orientations to give mSD15 and mSD16 (Fig 5.3). Sequence from these two clones, together with sequence determined by Mary Burke (see Table 5.1b), completed the sequence between nucleotides 1807 and 2197. Finch and Emerson (1983 & 1984) and Easton and Kushner (1983) reported the *uvrD* sequence, starting from the PvuII site at 2017. There are no differences between the published *uvrD* sequences (Finch and Emerson 1983;1984; Easton and Kushner, 1983) and that determined in this work, within the 181 bp region of overlap.

The remaining sequence was determined as summarised in Table 5.1 with the help of Mary Burke, Alisdair Sutcliffe and Richard McCulloch. The sequence from the PstI site at nucleotide 61 to the pSDC102 $\triangle$ 10.2 endpoint at nucleotide 417 was determined on at least one strand from mAS1, mAS2, mMB19.2 and mMB19.3. The 252 bp PstI fragment from pSDC102 was cloned in both orientations into M13mp18 by Alisdair Sutcliffe to produce mAS1 and mAS2. Mary Burke cloned the 68 bp fragment extending rightward from the PstI site to the nearest Sau3A site, along with other random Sau3A fragments, into M13mp19 to produce mMB19.2 and mMB19.3. Sequence from these clones was assembled together by comparing it to the sequence determined by Catherine Richaud. The sequence as far as nucleotide 206 is published (Richaud and Printz, 1988) and the sequence from 207 to 741 was communicated by C. Richaud. As shown in Fig 5.4 there are several differences between our sequence and the sequence of Richaud. These inconsistencies were checked as carefully as possible and

```
                                                                      50
Richaud    TATATGACTG GCCCGGCGGT ACATGTCTAC GACGGATTTA TTCATCTATG
           Y  M  T  G   P  A  V   H  V  Y   D  G  F  I   H  L  *
           dapF
                                PstI                                 100
Richaud    AAGCAACCAG GGGAAGAACT GCAGGAAACA CTCACGGAGC TTGATGACCG
This work                     CT GCAGGAAACA CTCACGGAGC TTGATGACCG

                                                                     150
Richaud    GGCGGTTGTC GATTATCTGA TTAAAAATCC TGAGTTTTTT ATCCGTAATG
This work  GGCGGTTGTC GATTATCTGA TTAAAAATCC TGAGTTTTTT ATCCGTAATG

                                        *  *                        198
Richaud    CGCGCGCAGT AGAAGCGATA CGTGTTGCCC GCATCCGGTA CGCGGCACCG
This work  CGCGCGCAGT AGAAGCGATA CGTGT.GCC. GCATCCGGTA CGCGGCACCG

           published                                                 248
Richaud    TTTCGTTGGT CGAGTGGCAC ATGGCCCGCG CACGTAATCA TATTCATGTT
This work  TTTCGTTGGT CGAGTGGCAC ATGGCCCGCG CACGTAATCA TATTCATGTT

             *                    (stop)                            298
Richaud    CTCGAAGAGA ACATGGCGCT GTTGATGGAA CAGGCTATCG CCAACGAAGG
This work  CTGGAAGAGA ACATGGCGCT GTTGATGGAA CAGGCTATCG CCAACGAAGG

                                   PstI      *                      348
Richaud    CCTGTTTTAT CGCCTACTCT ACCTGCAGCG CATTCTCACC GCCGCCAGCA
This work  CCTGTTTTAT CGCCTACTCT ACCTGCAGCG CAGTCTCACC GCCGCCAGCA

                         *                                          *398
Richaud    GTCTCGACGA TAT.CTGATG CGCTTTCACC GCTGGGCGCG CGATCT.GGC
This work  GTCTCGACGA TATGCTGATG CGCTTTCACC GCTGGGCGCG CGATCTCGGC

                                                                     448
Richaud    CTGGCAGGTG CGAGTCTGCG CCTGTTTCCG GATCGCTGGC GCTTAGGTGC
This work  CTGGCAGGTG CGAGTCTGCG CCTGTTTCCG GATCGCTGGC GCTTAGGTGC

             *                                                       498
Richaud    GC.GTCGAAC CACACTCATC TGGCATTAAG CCGTCAGTCT TTCGAACCGC
This work  GCCGTCGAAC CACACTCATC TGGCATTAAG CCGTCAGTCT TTCGAACCGC

                                                                     548
Richaud    TGCGTATTCA GCGTTTGGGG CAGGAACAGC ACTATCTTGG GCCGCTTAAC
This work  TGCGTATTCA GCGTTTGGGG CAGGAACAGC ACTATCTTGG GCCGCTTAAC

                                                                     598
Richaud    GGACCAGAGC TGCTGGTGGT GCTACCGGAA GCGAAAGCGG TGGGATCGGT
This work  GGACCAGAGC TGCTGGTGGT GCTACCGGAA GCGAAAGCGG TGGGATCGGT

                                                                     648
Richaud    GGCGATGTCG ATGCTGGGAA GCGATGCTGA TTTGGGTGTC GTGCTGTTTA
This work  GGCGATGTCG ATGCTGGGAA GCGATGCTGA TTTGGGTGTC GTGCTGTTTA
```

```
                                                                    698
Richaud     CCAGTCGCGA TGCCAGTCAC TATCAACAAG GGCAAGGAAC GCAGTTACTT
This work   CCAGTCGCGA TGCCAGTCAC TATCAACAAG GGCAAGGAAC GCAGTTACTT

                                                                    748
Richaud     CATGAAATTG CGCTGATGTT GCCGGAGCTT CTGGAGCGTT GGA
This work   CATGAAATTG CGCTGATGTT GCCGGAGCTT CTGGAGCGTT GGATTGAACG

              xerC                                                  798
This work   CGTATGACCG ATTTACACAC CGATGTAGAA CGCTACCTAC GTTATCTGAG
              M  T  D   L  H  T   D  V  E   R  Y  L  R   Y  L  S
```

Figure 5.4 Comparison between the sequence determined by Richaud and Printz and the sequence determined in this work. The seven differences between the sequence determined by Richaud and Printz (1988 and personal communications) and the sequence determined in our lab are indicated with asterisks (*). The alignment is numbered throughout according to the sequence determined in this work. The sequence as far as nucleotide 206 is published by Richaud and Printz (1988). The inferred protein sequences of the C-terminus of DapF and the N-terminus of XerC are shown. The start (nucleotide 48) and stop (nucleotide 753) codons of orf235 are shown. Note that the two single base-pair insertions at nucleotides 175 and 178 in the sequence of Richaud and Printz introduce a stop codon into the orf235 reading frame at nucleotide 271.

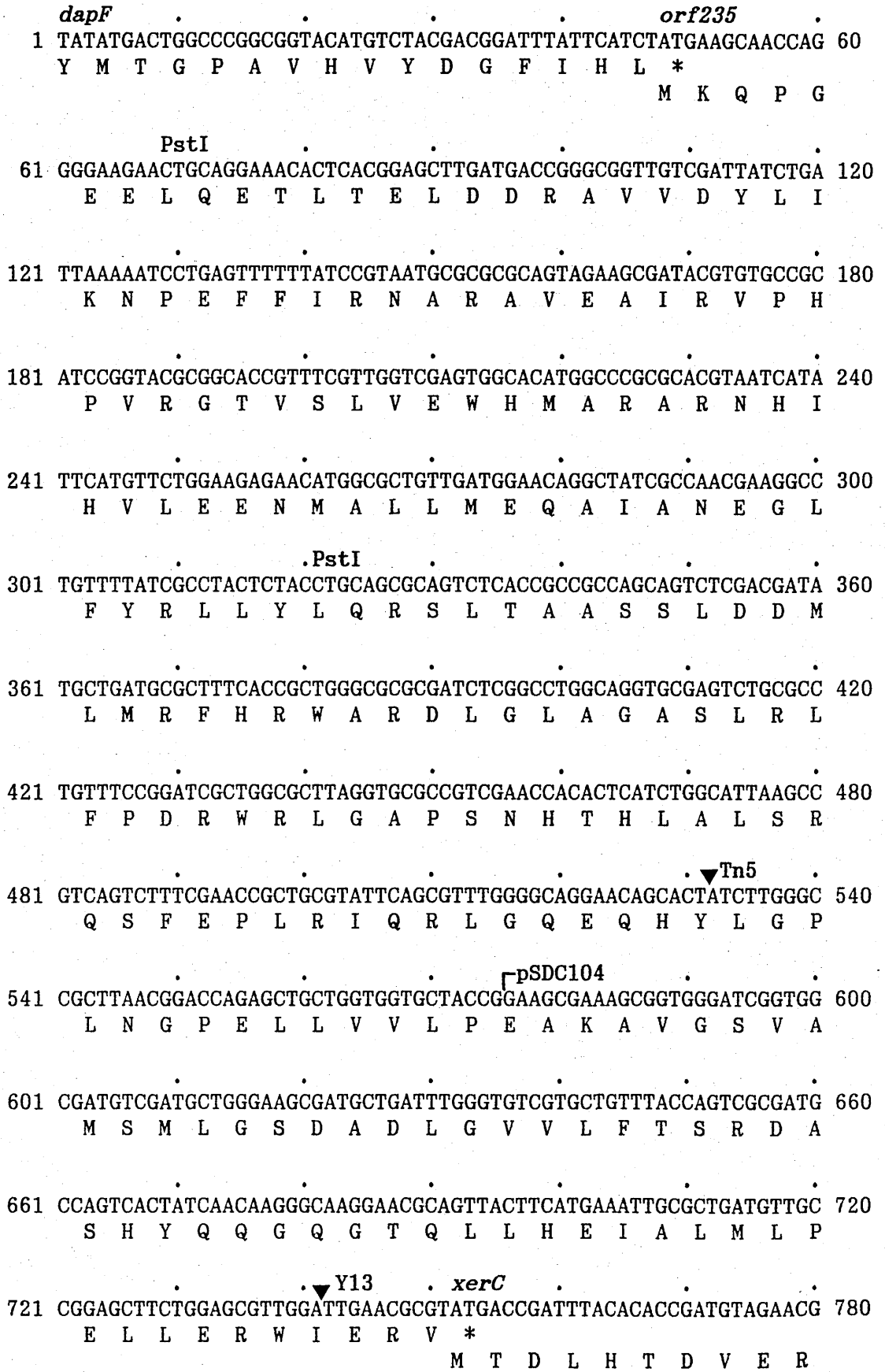
I am therefore fairly confident that the sequence determined in this work is correct. The complete sequence of the *xerC* region, extending from the 3' end of *dapF* to the 5' end of *uvrD* is shown in Fig 5.5.

All gels were read on at least two occasions and entered directly into the computer using the SEQED programme. Readings from gels were then compared to the sequence determined from the other strand using the REVERSE and GAP programmes. Conflicts between readings of the two strands were resolved by referring back to all gels covering the region of interest. The sequence shown in Fig 5.5 was double checked by a final proof-reading of all autoradiographs just prior to writing this thesis.

5.3 Analysis of the *xerC* sequence

The *xerC1* mutation is complemented by the 1,232 bp chromosomal fragment contained within pSDC104, but not by the shorter fragment contained within pSDC123, which extends 962 bp to the left of the BglII site at nucleotide 1,807 (see Fig 5.5). Plasmids that contain inserts which start at the SalI site within the minimal complementing fragment and extend either to the left or to the right from this site (pSDC106 and pVMK42, see Fig 4.9) do not complement *xerC1*. Therefore, it seemed likely that the *xerC* reading frame would be wholly contained within the 1,232 bp chromosomal fragment carried by pSDC104, but would be disrupted by the deletion of a further 270 bp. The complementation data also suggested that the *xerC* reading frame would span the SalI site in pSDC104.

The DNA sequence of the *xerC* region was searched for protein coding sequences using the CODONPREFERENCE program (Fig 5.6 A and B). This program indicates any open reading frame beginning with an AUG start codon. It also plots the occurrence of rare codons (those that occur at less than 5% of the frequency of synonymous codons) in all three reading frames. Above this is plotted a codon preference statistic, calculated over a window of 25 codons (Gribskov *et al*, 1984). *E.coli* uses synonymous codons on a non-random basis, which correlates to the abundance of the


dapF
orf235
PstI
PstI
Tn5
pSDC104
Y13
xerC

```
                .         .         .         .         .         .
781 CTACCTACGTTATCTGAGCGTGGAGCGCCAGCTTAGCCCGATAACCCTGCTTAACTACCA 840
    Y  L  R  Y  L  S  V  E  R  Q  L  S  P  I  T  L  L  N  Y  Q

      pSDC123
841 GCGTCAGCTTGAGGCGATCATCAATTTTGCCAGCGAAAACGGCCTGCAAAGCTGGCAGCA 900
    R  Q  L  E  A  I  I  N  F  A  S  E  N  G  L  Q  S  W  Q  Q

                .         .         .         .         .         .
901 ATGTGATGTGACGATGGTGCGCAATTTTGCTGTACGCAGTCGCCGTAAAGGGCTGGGAGC 960
    C  D  V  T  M  V  R  N  F  A  V  R  S  R  R  K  G  L  G  A

                .         .         .         .         .         .
961 AGCAAGTCTGGCGTTACGGCTTTCTGCGCTACGTAGCTTTTTTGACTGGCTGGTCAGCCA 1020
    A  S  L  A  L  R  L  S  A  L  R  S  F  F  D  W  L  V  S  Q

                .         .         .         .         .         .
1021 GAACGAACTCAAAGCTAACCCGGCGAAAGGTGTTTCGGCACCGAAAGCGCCGCGTCATCT 1080
     N  E  L  K  A  N  P  A  K  G  V  S  A  P  K  A  P  R  H  L

                .    SalI .         .         .         .         .
1081 GCCGAAAAACATCGACGTCGACGATATGAATCGGCTGCTGGATATTGATATCAATGATCC 1140
     P  K  N  I  D  V  D  D  M  N  R  L  L  D  I  D  I  N  D  P

                .         .         .         .         .         .
1141 CCTCGCTGTACGCGACCGTGCAATGCTGGAAGTGATGTACGGCGCGGGTCTGCGTCTTTC 1200
     L  A  V  R  D  R  A  M  L  E  V  M  Y  G  A  G  L  R  L  S

                .         .         .         .         .         .
1201 TGAGCTGGTGGGGCTGGATATTAAACACCTCGACCTGGAGTCTGGTGAAGTGTGGGTTAT 1260
     E  L  V  G  L  D  I  K  H  L  D  L  E  S  G  E  V  W  V  M

                .         .         .         .         .         .
1261 GGGGAAAGGCAGCAAAGAGCGCCGCCTGCCGATTGGTCGCAACGCTGTGGCGTGGATTGA 1320
     G  K  G  S  K  E  R  R  L  P  I  G  R  N  A  V  A  W  I  E

                .         .         .         .         .         .
1321 GCACTGGCTTGATTTGCGCGACCTGTTTGGTAGCGAAGACGACGCGCTTTTTCTGTCGAA 1380
     H  W  L  D  L  R  D  L  F  G  S  E  D  D  A  L  F  L  S  K

                .         .         .         .         .         .
1381 ACTGGGCAAGCGTATCTCCGCGCGTAATGTGCAGAAACGCTTTGCCGAATGGGGCATAAA 1440
     L  G  K  R  I  S  A  R  N  V  Q  K  R  F  A  E  W  G  I  K

                .         .         .         .         .         .
1441 ACAAGGGCTGAATAATCACGTTCATCCGCATAAATTACGTCACTCGTTCGCCACGCATAT 1500
     Q  G  L  N  N  H  V  H  P  H  K  L  R  H  S  F  A  T  H  M

                .         .         .         .         .         .
1501 GCTGGAGTCGAGCGGCGATCTTCGTGGTGTGCAGGAGCTGCTGGGTCATGCCAACCTCTC 1560
     L  E  S  S  G  D  L  R  G  V  Q  E  L  L  G  H  A  N  L  S

                .         .         .         .         .         .
1561 CACCACGCAAATCTATACTCATCTTGATTTTCAACACCTTGCCTCGGTGTACGATGCGGC 1620
     T  T  Q  I  Y  T  H  L  D  F  Q  H  L  A  S  V  Y  D  A  A
```

Y17

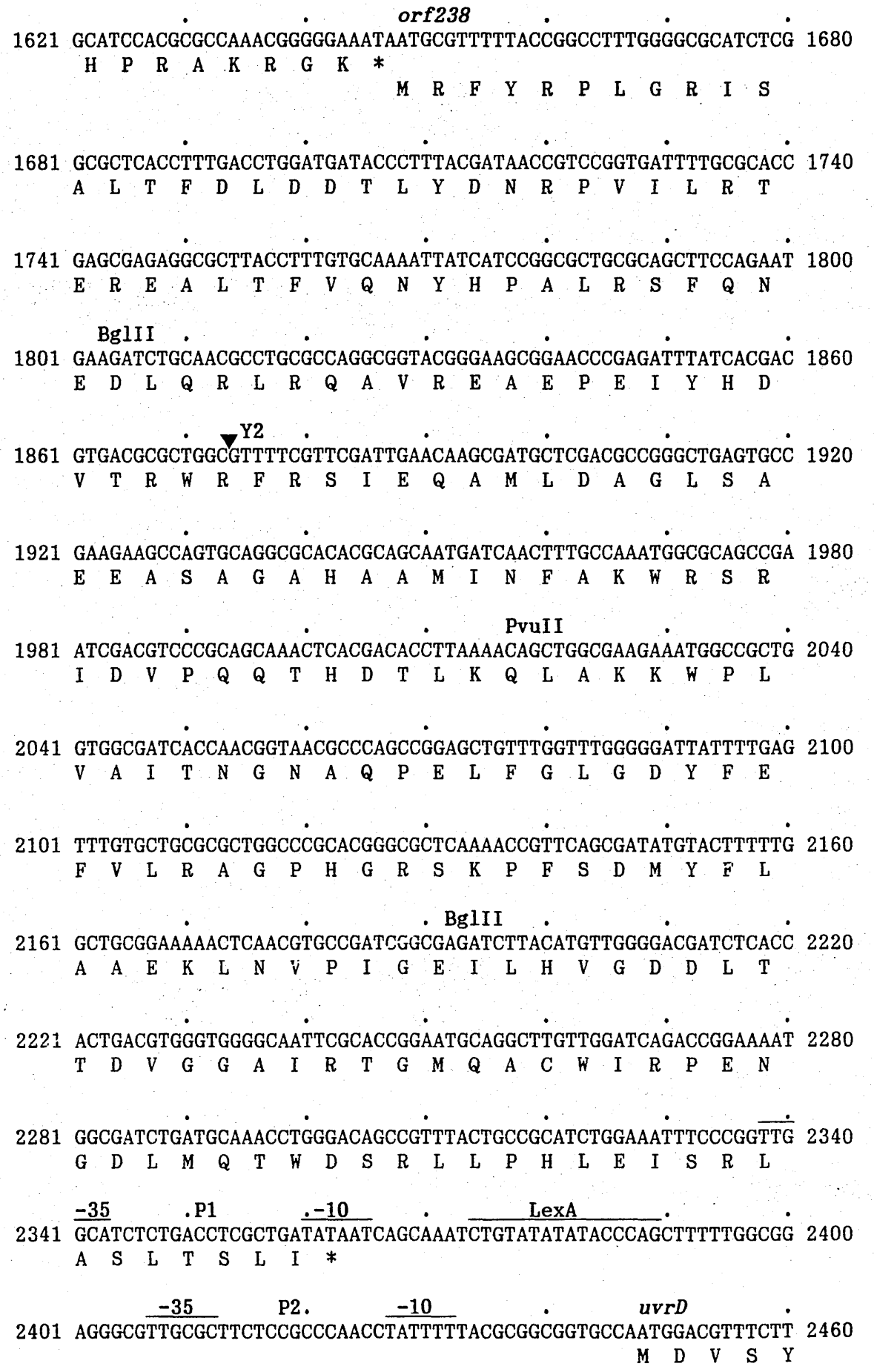

orf238
1621 GCATCCACGCGCCAAACGGGGGAAATAATGCGTTTTTACCGGCCTTTGGGGCGCATCTCG 1680
H P R A K R G K *
M R F Y R P L G R I S
1681 GCGCTCACCTTTGACCTGGATGATACCCTTTACGATAACCGTCCGGTGATTTTGCGCACC 1740
A L T F D L D D T L Y D N R P V I L R T
1741 GAGCGAGAGGCGCTTACCTTTGTGCAAAATTATCATCCGGCGCTGCGCAGCTTCCAGAAT 1800
E R E A L T F V Q N Y H P A L R S F Q N
BglII
1801 GAAGATCTGCAACGCCTGCGCCAGGCGGTACGGGAAGCGGAACCCGAGATTTATCACGAC 1860
E D L Q R L R Q A V R E A E P E I Y H D
Y2
1861 GTGACGCGCTGGCGTTTTCGTTCGATTGAACAAGCGATGCTCGACGCCGGGCTGAGTGCC 1920
V T R W R F R S I E Q A M L D A G L S A
1921 GAAGAAGCCAGTGCAGGCGCACACGCAGCAATGATCAACTTTGCCAAATGGCGCAGCCGA 1980
E E A S A G A H A A M I N F A K W R S R
PvuII
1981 ATCGACGTCCCGCAGCAAACTCACGACACCTTAAAACAGCTGGCGAAGAAATGGCCGCTG 2040
I D V P Q Q T H D T L K Q L A K K W P L
2041 GTGGCGATCACCAACGGTAACGCCCAGCCGGAGCTGTTTGGTTTGGGGGATTATTTTGAG 2100
V A I T N G N A Q P E L F G L G D Y F E
2101 TTTGTGCTGCGCGCTGGCCCGCACGGGCGCTCAAAACCGTTCAGCGATATGTACTTTTTG 2160
F V L R A G P H G R S K P F S D M Y F L
BglII
2161 GCTGCGGAAAAACTCAACGTGCCGATCGGCGAGATCTTACATGTTGGGGACGATCTCACC 2220
A A E K L N V P I G E I L H V G D D L T
2221 ACTGACGTGGGTGGGGCAATTCGCACCGGAATGCAGGCTTGTTGGATCAGACCGGAAAAT 2280
T D V G G A I R T G M Q A C W I R P E N
2281 GGCGATCTGATGCAAACCTGGGACAGCCGTTTACTGCCGCATCTGGAAATTTCCCGGTTG 2340
G D L M Q T W D S R L L P H L E I S R L
-35 .P1 .-10 . LexA .
2341 GCATCTCTGACCTCGCTGATATAATCAGCAAATCTGTATATATACCCAGCTTTTTGGCGG 2400
A S L T S L I *
-35 P2. -10 . uvrD .
2401 AGGGCGTTGCGCTTCTCCGCCCAACCTATTTTTACGCGGCGGTGCCAATGGACGTTTCTT 2460
M D V S Y

Figure 5.5 DNA sequence of the dapF to uvrD region of the E. coli chromosome. The sequence presented here joins the previously published sequences of dapF (Richaud and Printz, 1988) and uvrD (Easton and Kushner, 1983; Finch and Emerson, 1983; 1984). The inferred protein sequences of Orf235, XerC and Orf238 are shown as are those of the C-terminus of DapF and the N-terminus of UvrD. The positions of the in-frame mini-Mu insertions (Y13, Y17 and Y2), as determined by DNA sequencing are shown. The position of the Tn5 insertion in xerC1 is indicated as is the SalI site at which the Km^r cassette was inserted to produce the xerC2 mutation. The conserved domains 1 and 2 found in the lambda integrase family of site-specific recombinases are underlined within XerC. The $XerC^+$ plasmid pSDC104 carries the fragment from the position marked at nucleotide 576 to the BglII site at 1807. The $XerC^-$ plasmid pSDC123 carries the fragment from the position marked at nucleotide 845 to the BglII site at 1807. Two potential promoters for uvrD (P1 and P2) and a LexA binding site (Easton and Kushner, 1983; Finch and Emerson, 1984) are shown. Nucleotide 1 is enumerated 1101 in Richaud and Printz (1988) and in Colloms et al (1990). The sequence from the PvuII site at 2017 is presented in Finch and Emerson (1983). Nucleotide 2192 is enumerated 1 in Finch and Emerson (1984).

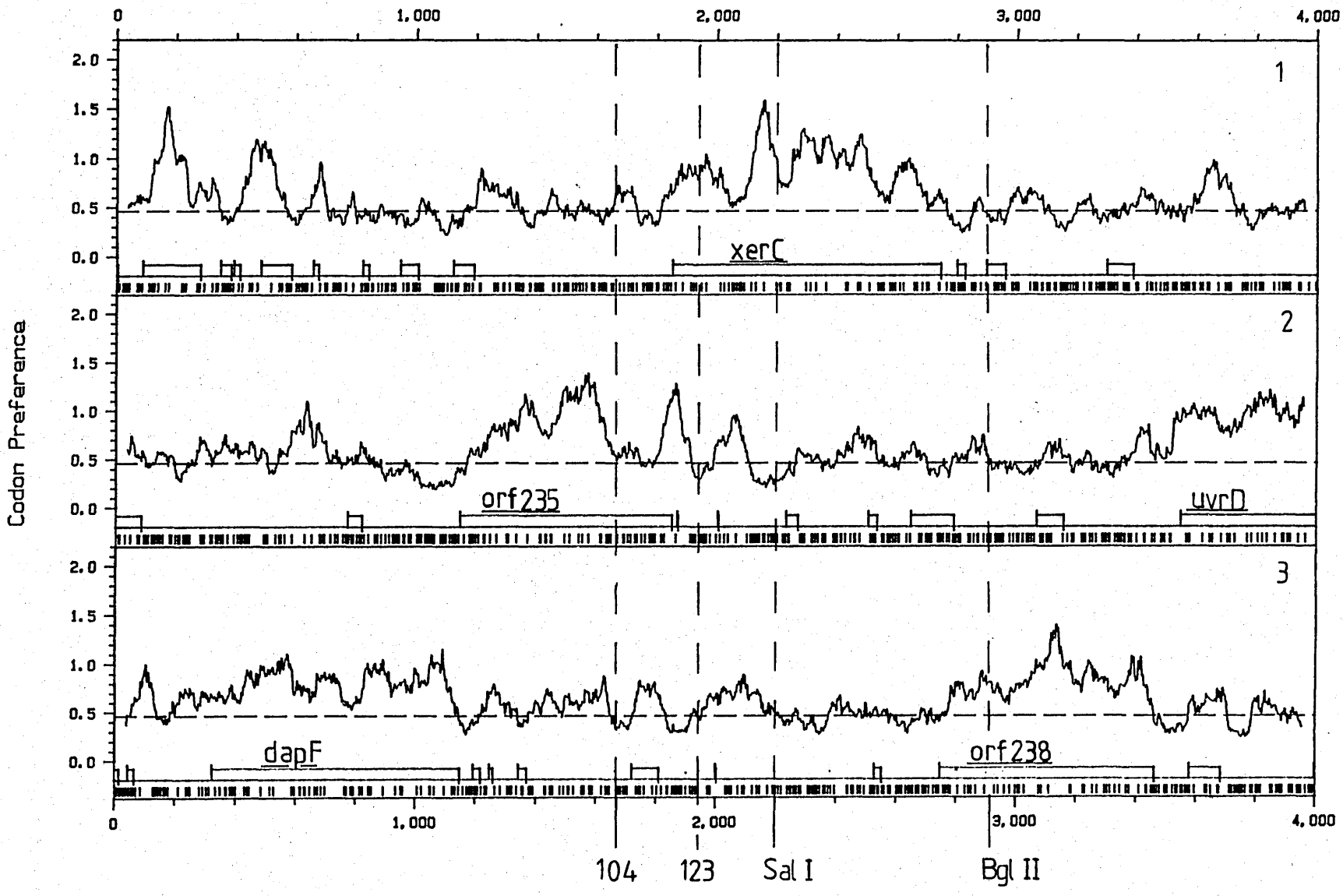
cognate tRNAs in *E. coli* (Ikemura, 1982). The codon preference statistic is calculated using a codon usage table generated from known *E. coli* coding sequences. In practice it is found that highly expressed to moderately expressed *E. coli* genes have a codon preference statistic of between 1.0 and 2.0, and poorly expressed *E.coli* genes score from 0.5 to 1.0. Random sequence is expected to score about 0.45 (though this varies with base composition), and in practice most non-coding sequences do score about this level.

Examination of Fig 5.6 reveals that there are three open reading frames which span the SalI site in pSDC104. One of these is encoded by the top strand in frame 1, the other two are encoded by the bottom strand in frames 5 and 6. All three of these reading frames are wholly contained within the 1,232 bp minimal complementing fragment. However, only the one in frame 1 is disrupted by the deletion carried in pSDC123. The open reading frame in frame 1 has a paucity of rare codons and has a codon preference statistic well above that expected for non-coding sequence. It is also in the correct orientation to be transcribed from the pTZ18R *lac* promoter in pSDC104 and in all of the other pSDC102 deletion derivatives. The other two open reading frames have a much lower codon preference statistic and are probably not genuine protein coding sequences. Taken together, the complementation data and the codon preference data strongly suggest that the reading frame encoded by frame 2 corresponds to the *xerC* gene. This open reading frame could encode a protein of 298 amino acids with a calculated molecular weight of 33,800.

5.4 Analysis of the *xerC1* Tn5 insertion

The site of Tn5 insertion in the *xerC1* allele, determined by DNA sequencing, lies outside the minimal complementing fragment and 5' of the *xerC* open reading frame (see Fig 5.5). Insertions of Tn5 are known to exert polar effects in most cases (Berg and Berg, 1983). The leaky phenotype of *xerC1* and the complementation data

A)


0
1,000
2,000
3,000
4,000
Codon Preference
2.0
1.5
1.0
0.5
0.0
1
2
3
xerC
orf235
uvrD
dapF
orf238
104
123
Sal I
Bgl II

B)

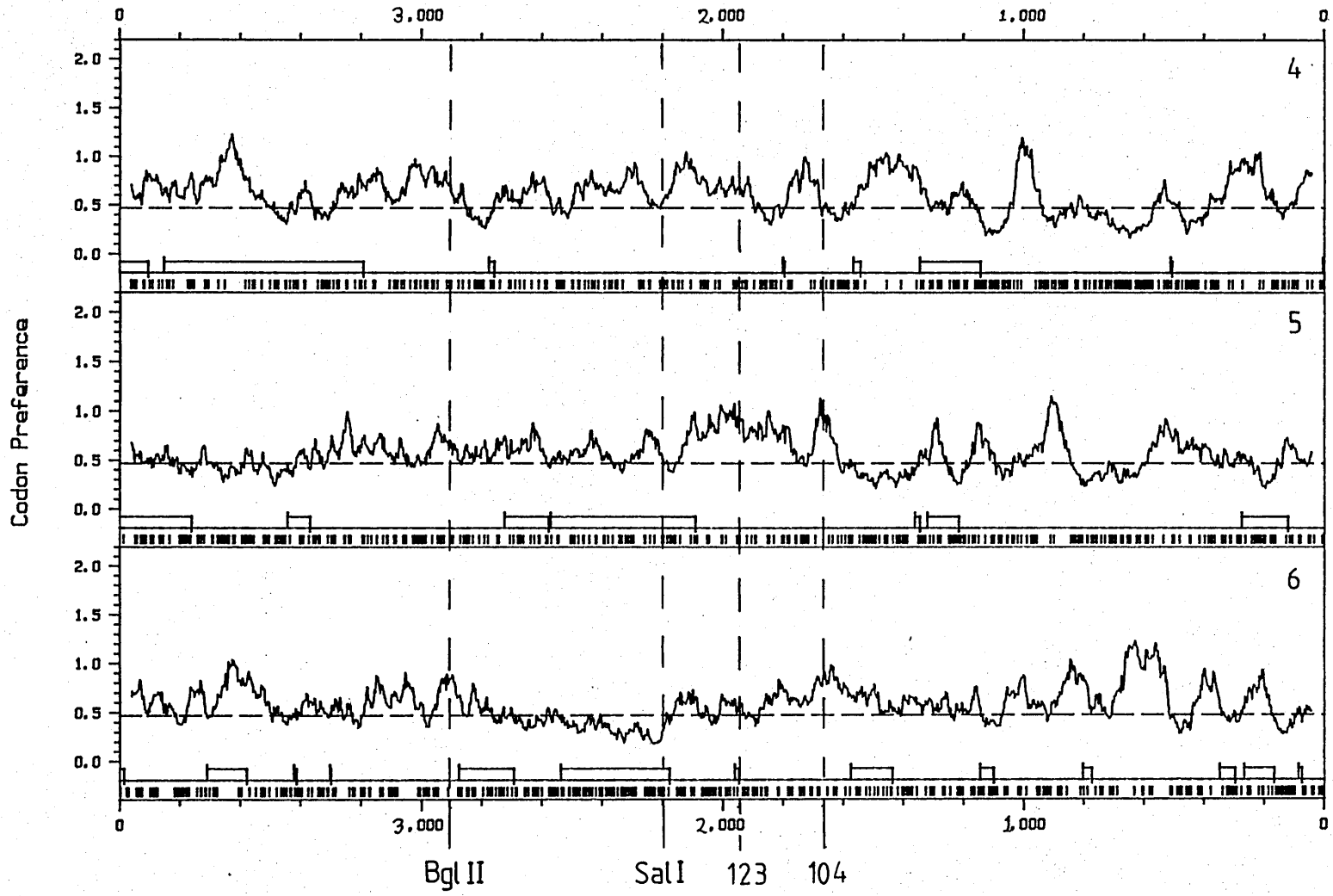


Figure 5.6 Codonpreference plot of the dapF - xerC - uvrD sequence. The dapF, orf235, xerC, orf238 and uvrD reading frames are indicated. The plot extends only 450 bp into the 2,160 bp uvrD coding sequence. The numbering is the same as that in Colloms et al (1990); nucleotide 1101 on this plot is nucleotide 1 in Fig 5.5. The three forward frames (1, 2 and 3) are shown in (A), the three reverse frames (4, 5 and 6) are shown in (B). The boundaries of the inserts carried by pSDC104 and pSDC123 are indicated.

could therefore be explained by a reduction in transcription of *xerC* brought about by the polar upstream insertion.

5.5 Other open reading frames in the *xerC* region.

Examination of the sequence upstream of *xerC* reveals an open reading frame (*orf235*) which could encode a protein of 235 amino acids (Fig 5.5). Downstream of *xerC*, between *xerC* and *uvrD*, there is an open reading frame (*orf238*) which could encode a protein of 238 amino acids (Fig 5.5). Examination of Fig 5.6 reveals that both of these reading frames have a codon preference statistic well above that expected by chance, making it likely that they are in fact genuine protein coding sequences.

The *dapF*, *orf235*, *xerC* and *orf238* open reading frames are all in very close proximity. At both the *dapF* - *orf235* and the *orf235* - *xerC* junctions, the sequence ATGA provides the stop codon of one reading frame and (with a -1 frameshift) the start codon of the next. At the *xerC* - *orf238* junction, the sequence TAATG forms the *xerC* stop codon and the *orf238* start codon (a +2 frameshift). This arrangement leaves little room for a *xerC* promoter unless it lies upstream of *dapF* or is contained within the coding regions of either *orf235* or *dapF*. It is known that *dapF* does have its own promoter. Its transcriptional start has been mapped to a position 269 nucleotides upstream of the *dapF* start codon (Richaud and Printz, 1988). It seems likely that all four open reading frames are expressed from the same promoter on one polycistronic mRNA. The close proximity of successive stop and start codons suggests that some form of translational coupling could occur.

5.6 Insertion mutations in *dapF*, *orf235*, *xerC* and *orf238*

Catherine Richaud carried out a mutation analysis of the cloned *dapF* region using Mu *d*II PR13 (Fig 5.7). This mini Mu produces translational fusions to a copy of *lacZ* lacking a translational start. Plasmid pDF3, containing

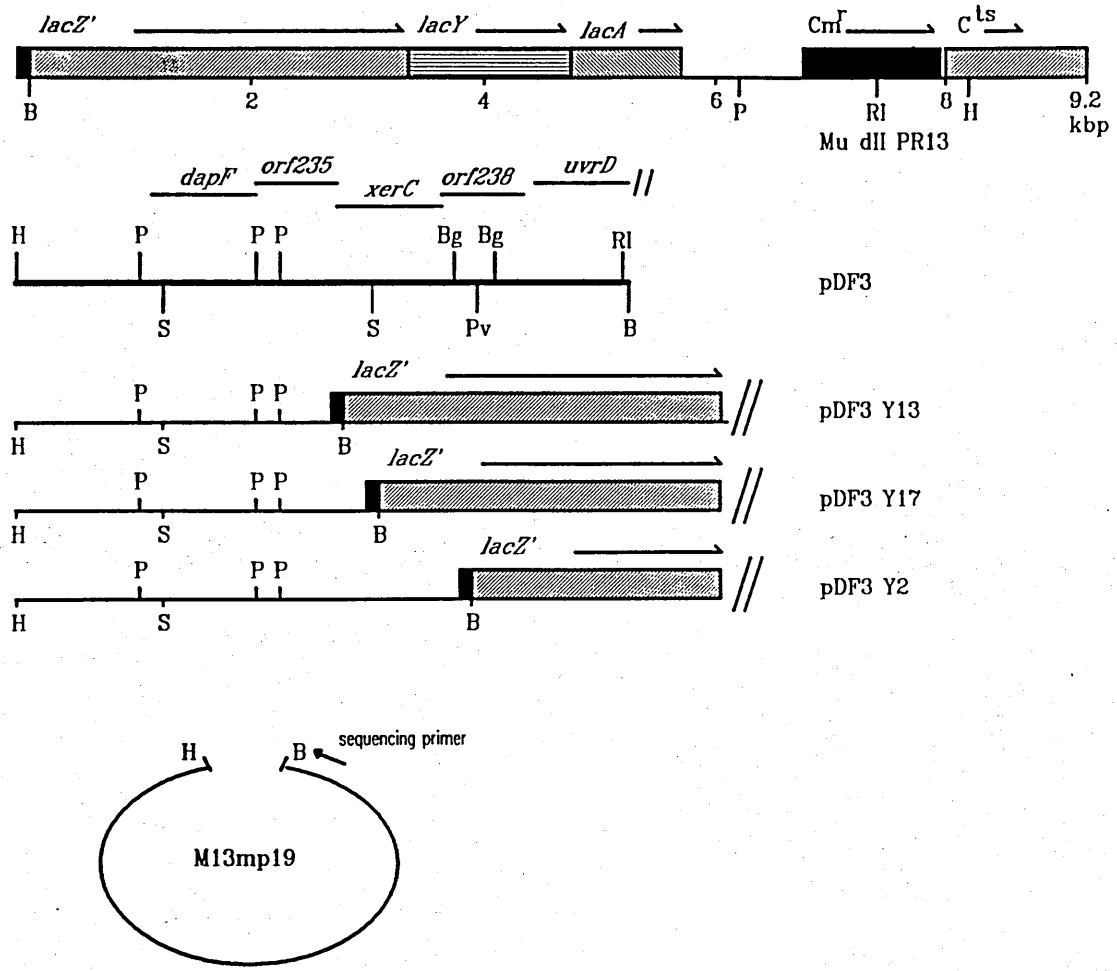


Figure 5.7 Determination of the sites of insertion of Mu dII PR13 in the xerC region. The top of the figure shows the organisation of the 9.2 kbp Mu dII PR13 element. The bottom of the figures shows a map of the 5.3 kbp region contained within pDF3, which was subjected to mutagenesis with Mu dII PR13 by Richaud et al (1987). Below this are shown the relevant sections of the three pDF3::Mu dII PR13 derivatives from which the HindIII - BamHI fragments were cloned into M13mp19 for sequencing.

the HindIII - BamHI fragment which extends from the 3' end of *cya* to the 5' end of *uvrD*, was subjected to insertion mutagenesis with Mu *d*II PR13. Derivatives of pDF3 containing Mu *d*II PR13 *lacZ* translational fusions were selected as blue colonies on medium containing X-gal (Richaud *et al*, 1987). Mutations were then transferred to the chromosome by transformation of JC7623 with linearised plasmid DNA. Catherine Richaud supplied us with plasmids and strains carrying four separate Mu *d*II PR13 insertions. One of these (here called Y30) was known to map within *dapF*, and cause a partial Dap^- auxotrophic phenotype. The other three insertions were known to map downstream of *dapF*.

The downstream insertions had not been accurately mapped by Richaud. To accurately map these insertions, plasmids containing the Y13, Y17 and Y2 mini Mu insertions were cleaved with HindIII and BamHI. Mu *d*II PR13 has a unique BamHI site 117 bp from the *lacZ* end. Fragments of the correct size to carry mini Mu - pDF3 boundaries (2.8 kbp, 3.1 kbp and 3.9 kbp respectively for Y13, Y17 and Y2) were excised from an agarose gel, purified with Geneclean and cloned into M13mp19 to give mSD17, mSD18 and mSD19 (Fig 5.7). Sequencing of these recombinant phage gave identical mini Mu sequence for 117 bp followed by pDF3 sequences. These pDF3 sequences could be located on the *xerC* bottom strand sequence (Table 5.1). This allowed the determination of the exact position at which each copy of the mini Mu had inserted (Fig 5.5) and defined which reading frame was fused translationally to *lacZ*.

It turned out that each insertion was into a different open reading frame. Y13 is an insertion into *orf235* at nucleotide 741, the Y17 insertion is into the *xerC* reading frame at nucleotide 1,058 and Y2 is an insertion into *orf238* at nucleotide 1,873 (Fig 5.5). All three of these insertions had placed *lacZ* in frame with the corresponding open reading frame previously identified by computer. Since all three insertions are blue on X-gal this provided evidence that these open reading frames are in fact translated *in vivo*.

Audrey McLaughlin (a project student) found that all

four chromosomal fusions expressed beta galactosidase at approximately the same level, about 1-5% the level of the fully derepressed chromosomal *lacZ* gene in ML308. Audrey measured beta galactosidase activities in exponentially growing cultures grown in L broth or minimal medium in the presence and absence of lysine and diaminopimelic acid. She was unable to detect any regulation acting on expression of these fusions under these conditions.

5.7 Xer phenotypes of mini Mu insertions

JC7623 carries the *recBC sbcBC* mutations, rendering it an unsuitable host for ColE1 related plasmids (Bassett and Kushner, 1984) In order to study the Xer phenotypes caused by the mini Mu insertions in the *xerC* region the four insertion mutations were transduced into DS941 using P1. Transductants were selected by virtue of the mini Mu encoded resistance to chloramphenicol.

DS941 derivatives carrying mini Mu insertions in *dapF* (Y30), *orf235* (Y13), *xerC* (Y17) and *orf238* (Y2) were transformed with the reporter plasmid pKS455. DS941 *dapF* transformants were plated onto medium supplemented with 20 ug/ml DAP. Plasmid DNA was prepared from transformants by the boiling method and analysed by agarose gel electrophoresis (Fig 5.8). As can be seen the insertions into *dapF* and *orf235* prevented complete resolution of pKS455. The extent of resolution varied slightly from experiment to experiment, but some unresolved pKS455 always remained. This Xer phenotype is reminiscent of the phenotype caused by the *xerC1* mutation. The most likely explanation for this phenotype is that insertions between the *dapF* promoter and the *xerC* coding sequence are polar on expression of *xerC*. A reduction in transcription of *xerC*, caused by the polar mini Mu insertions, will lead to a lower cellular XerC protein concentration and thus reduce the rate of resolution of pKS455.

The Y17 mini Mu insertion into the *xerC* coding sequence completely abolished resolution of pKS455 (Fig 5.8). Even after continued culture for many generations no resolution of pKS455 in DS941 *xerC* Y17 was detected. Peter

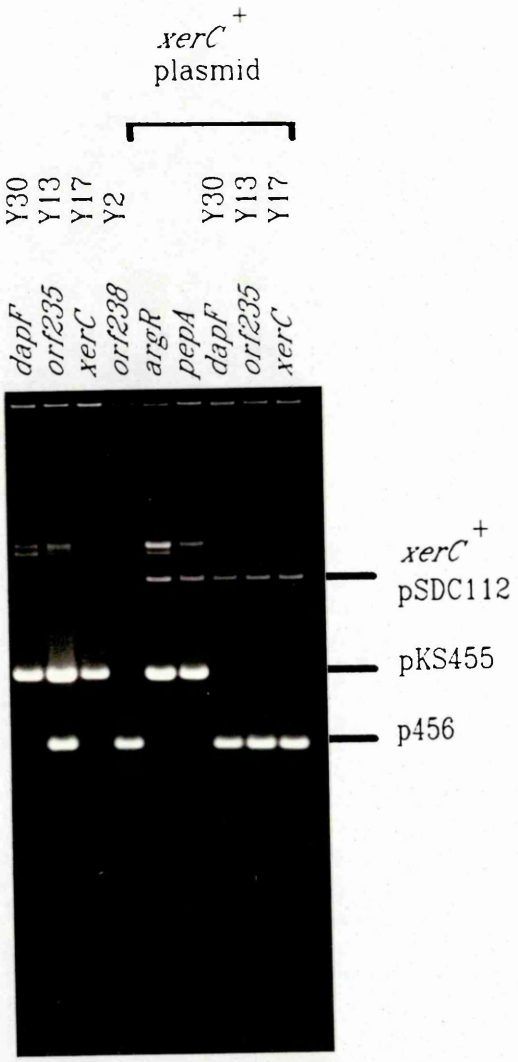


Figure 5.8 Complementation of Xer⁻ mini-Mu insertions in dapF, orf235 and xerC by a plasmid containing only the xerC open reading frame. The DS941 derivatives shown were transformed either with pKS455 alone or with pKS455 and pSDC112. Plasmid DNA was isolated by the boiling method and run on a 1.2% agarose gel. The reporter plasmid pKS455 is resolved by cer site-specific recombination in Xer^+ strains to give plasmid p456. pSDC112 contains the same 1,232 bp $XerC^+$ fragment as pSDC104 in a Km^r lambda dv-based replicon.

Sykora in our lab constructed a strain (DS941 *xerC2*) that carries a Km^r gene cloned into the SalI site within the chromosomal *XerC* gene. This strain shows an identical $XerC^-$ phenotype to DS941 *xerC* Y17 (see Fig 5.9).

The Y2 insertion, which is downstream of *xerC*, within *orf238*, gave no detectable phenotype. Resolution of pKS455, pCS202 and pSDC110 proceeded at the same rapid rate as in wild type strains. DS941 *orf238* Y2 had a growth curve very similar to DS941 in both rich medium (L broth) and in minimal medium supplemented with DS941 requirements (A. McLaughlin, personal communications). Colonies on plates also appeared normal, as did cell morphology under the microscope.

The Xer^- phenotypes of the *dapF* Y30, *orf235* Y13 and *xerC* Y17 insertion mutations were complemented by the 1,232 bp minimal *xerC* complementing fragment. DS941 derivatives carrying these mutations were doubly transformed with the plasmids pKS455 and pSDC112. Transformants were selected for resistance to both ampicillin and kanamycin. DNA was recovered by the boiling method and analysed on an agarose gel. The reporter plasmid pKS455 was found to be completely resolved to its deletion derivative p456 in all cases (Fig 5.8). (pSDC112 carries the $xerC^+$ EcoRI - HindIII fragment from pSDC104 on a Km^r lambda *dv* - based plasmid.)

Recombination at the type II ColE1 *cer* - CloDF13 *parB* hybrid site was totally abolished by both the *xerC2* (Km^r) and the *xerC* Y17 mutations. DS941 *xerC2* and DS941 *xerC* Y17 were transformed with pSDC110, which contains two directly repeated copies of the type II site. Analysis of plasmid DNA by agarose gel electrophoresis revealed that no resolution of pSDC110 had occurred (Fig 5.9 and not shown). Plasmid DNA was also found to be non recombinant by restriction even after many generations of growth. The *dapF* Y30 and *orf235* insertions were also found to reduce intramolecular type II hybrid recombination in pSDC110 to a level similar to that found in DS941 *xerC1*.

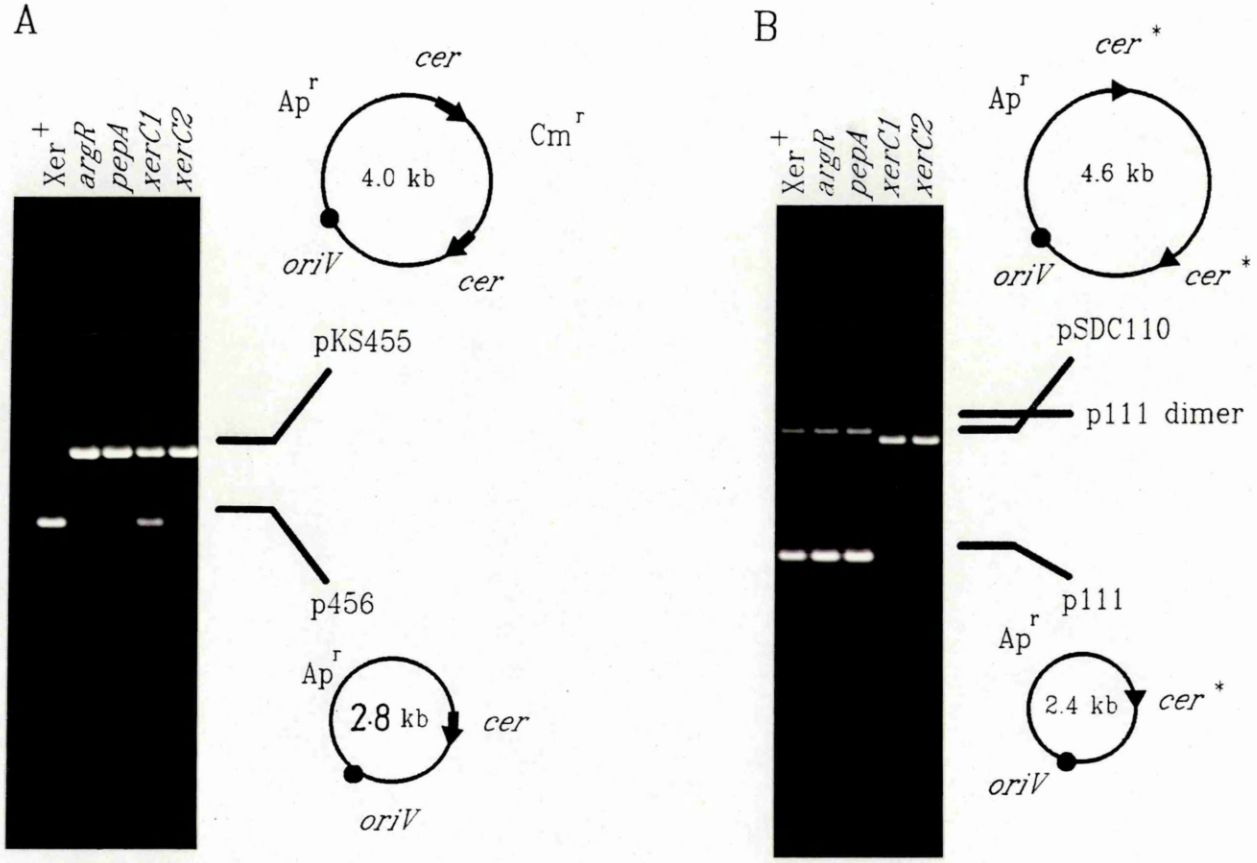


Figure 5.9 cer-mediated site-specific recombination in Xer$^+$ and Xer$^-$ E. coli strains. A) DS941 E. coli Xer$^+$ and Xer$^-$ strains were transformed with the reporter plasmid pKS455, which contains two copies of the wild type cer site. cer-mediated recombination yields p456. B) DS941 E. coli Xer$^+$ and Xer$^-$ strains were transformed with the reporter plasmid pSDC110, which contains two copies of the type II (ColE1 cer - CloDF13 parB) hybrid site (Sau3A - AluI fragment, cer*). Intramolecular cer site-specific recombination acting on pSDC110 yields p111. Plasmid DNA was isolated using the boiling technique and run on a 1.2% agarose gel.

5.8 XerC belongs to the lambda integrase family of site-specific recombinases.

The inferred protein sequence of XerC was searched by eye for similarities to either the resolvase/invertase class or the lambda integrase class of site-specific recombinases. A convincing match to the most highly conserved region of the lambda integrase family, the so called domain 2. was easily found. This match to domain 2 was found near to the C-terminus of the XerC sequence, where it is found in all other members of the lambda integrase family of site-specific recombinases, including the yeast FLP protein (Argos *et al*, 1986).

The lambda integrase domain 2 contains three residues, histidine, arginine and tyrosine which are totally conserved throughout the lambda integrase family (Argos *et al*, 1986). Other residues are highly conserved but are not present in all members of the family. XerC contains all three totally conserved amino acids and has matches to many of the highly conserved positions, all with the correct spacing. Homology to the other highly conserved region of the integrase family, domain 1, was also found in the XerC sequence. Alignments of domains 1 and 2 of XerC to the other members of the integrase family is shown in Fig 5.10.

An exhaustive search using the Edinburgh ICL 64 x 64 distributed array processor (Collins *et al*, 1988) was carried out by John Collins to compare the XerC sequence to 17,425 protein sequences. This search revealed regions of similarity to many members of the lambda integrase family of recombinases as summarised in Fig 5.11. It can be seen that sequence similarity to other members of the integrase family is centred on the classical domain 1 and domain 2 regions but extends out from these in many cases. Of particular note are the alignments with the *E. coli* proteins FimB and FimE (Klemm, 1986; Dorman and Higgins, 1987) and the alignment with TnpI from *Bacillus thuringiensis* Tn4430 (Mahillon and Lereclus, 1988). These alignments span the region between domains 1 and 2 and cover about 160 amino acids. An alignment between amino

Domain 1

		1
consensus		elay tG RisEll L didl
XerC		RAMLEVMYGAGLRLSELVGLDIKHLDLE
R46	Orf3	RLFAQLLYGTGMRISEGLQLRVKDLDFD
FimE		YCLILLAYRHGMRISELLDLHYQDLDLN
FimB		YCLTLLCFIHGFRASEICRLRISDIDLK
Tn4430	TnpI	YAIATLLAYTGVRISEALSIKMNDFNLQ
Tn554	TnpA	KLILMLMYEGGLRIGEVLSLRLEDIVTW
Tn554	TnpB	ATMTMIVQECGMRISELCTLKKGCLLED
F:D	Prot	KMLLATLWNTGARINEALALTRGDFSLA
P1	Cre	TAGVEKALSLGVTKLVERWISVSGVADD
P22	Int	KSVVEFALSTGLRRSNIINLEWQQIDMQ
Lambda	Int	RLAMELAVVTGQRVGDLCEMKWSDIVDG

Domain 2

consensus		H LRHs at l e G- ir vq llGH n-i tt- Yth
XerC		HKLRHSFATHMLESS-GDLRGVQELLGHAN-LSTT-QIYTH
R46	Orf3	HTLRHSFATALLRSG-YDIRTVQDLLGHSD-VSTT-MIYTH
FimE		HMLRHACGYELAERG-ADTRLIQDYLGHRN-IRHT-VRYTA
FimB		HMLRHSCGFALANMG-IDTRLIQDYLGHRN-IRHT-VWYTA
Tn4430	TnpI	HQLRHFFCTNAIEKG-FSIHEVANQAGHSN-IHTT-LLYTN
Tn554	TnpA	HMLRHTHATQLIREG-WDVAFVQKRLGHAHVQTTL-NTYVH
Tn554	TnpB	HAFRHTVGTRMINNG-MPQHIVQKFLGHES-PEMT-SRYAH
F:D	Prot	HTFRHSYAMHMLYAG-IPLKVLQSLMGHKS-ISST-EVYTK
P1	Cre	HSARVGAARDMARAG-VSIPEIMQAGGWTN-VNIV-MNYIR
P22	Int	HDLRHTWASWLVQAG-VPISVLQEMGGWES-IEMV-RRYAH
Lambda	Int	HELRSLSA-RLYEKQ-ISDKFAQHLLGHKS-DTMA-SQY-R
Flp		HIGRHLMTSFLSMKGLTELTNVVGNWSDKRASAVARTTYTH

Figure 5.10 Alignment of the XerC protein sequence to other site-specific recombinases of the lambda integrase family. Only the two most conserved regions (domains 1 and 2) are shown. The regions of XerC shown are those underlined in Fig 5.5. The consensus line shows residues present in at least 4 of the proteins in lower case and those present in at least 8 proteins in upper case. Gaps (-) have been introduced to maximise the homology. The sequence of R46 Orf3 is from Hall and Vockler (1987). The sequences of the chromosomally encoded E. coli recombinases FimB and FimE are from Klemm (1986) and Dorman and Higgins (1987). TnpI comes from the Bacillus thuringiensis transposon Tn4430 (Mahillon and Lereclus, 1988). The sequences of the Tn554 transposition proteins TnpA and TnpB are from Murphy et al. (1985).The D protein is a recombinase from the E. coli F factor (Lane et al, 1986). The sequences of the P1, P22, lambda and Flp recombinases are taken from the data of Argos et al. (1986).

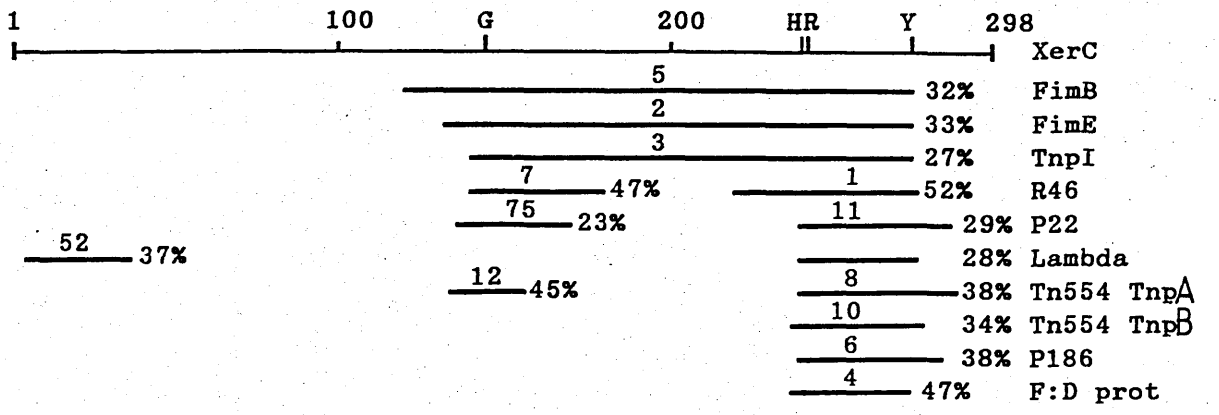


Figure 5.11 Summary of the alignments found between the XerC protein and other members of the lambda integrase family of site-specific recombinases. A protein sequence data-base containing 17,425 protein sequences was searched with the 64 x 64 Distributed Array Processor using the "Best Local Similarity" algorithm of Smith and Waterman (1981) as modified by Collins et al (1988). All alignments to known members of the lambda integrase family of site-specific recombinases out of the top 90 scoring alignments are summarised above. The rank according to score is shown above each alignment. The percentage of identical amino acids within each alignment is shown. The alignment to domain 2 of lambda integrase did not score in the top 90 but is shown for information. The positions in XerC of the conserved glycine residue in domain 1 and the conserved histidine, arginine and tyrosine residues in domain 2 are shown.

acids 4 - 36 of XerC and 76 - 107 of lambda Int was also found. This region of lambda Int can not be aligned with any of the other members of the integrase family of site specific recombinases: P2 Int, 186 Int, P22 Int, P1 Cre, Phi80 Int, and P4 Int (Argos *et al*, 1986). It is also outside of the 7KDa N-terminal domain of lambda Int which binds to the arm type sites of *attP* (Moitoso de Vargas *et al*, 1988). The spacing between this region and domains 1 and 2 is very similar in XerC and lambda Int. However, the biological significance of this alignment, if any, is not known.

The same type of computer search was done to look for proteins similar to the inferred protein sequences of Orf235 and Orf238. No convincing similarities were found between these sequences and any others in the data base.

5.9 A cellular role for *xerC*?

It was noticed that DS941 *xerC2* and DS941 *xerC* Y17 produced smaller colonies than DS941 on L agar plates. However, when growth rates were measured in liquid media there was no apparent difference in exponential growth phase doubling times. The small colony size on plates might have reflected a slowness in recovery from stationary phase, since it was particularly noticeable when these strains were recovered from storage. Examination of DS941 *xerC2* and DS941 *xerC* Y17 under the microscope revealed that these *xerC*$^-$ strains have a tendency to filament. DS941 *dapF* Y30 (medium was supplemented with DAP), DS941 *orf235* Y13, DS941 *xerC1* and DS941 *orf238* Y2 did not show these phenotypes. Both the tendency to form filaments and the small colony size were complemented by the plasmid pSDC105, which contains the *xerC*$^+$ 1,232 bp fragment from pSDC104 transcribed from the strong P_{tac} promoter. The *xerC* Y17 mutation was transduced with P1 into CSH50 and CSH50 delta *fimB - fimH*. Examination of the two strains produced revealed that CSH50 delta *fimB - fimH*, *xerC* Y17 has a greater tendency to filament than CSH50 *xerC* Y17; and both form filaments to a greater extent than their *xerC*$^+$ parental strains (not

shown).

Nucleoids were visualised in various *xerC*$^+$ and *xerC*$^-$ strains by D. Sherratt and G. Blakely, using the method of Hiraga *et al* (1989). Cells were first treated with chloramphenicol to inhibit protein synthesis, stained with DAPI and visualised by fluorescence microscopy. Most of the filaments present in *xerC* mutant strains had aberrant nucleoids. In many cases a single, highly fluorescent (possibly amplified) nucleoid was present in the middle of the filament (D. Sherratt and G. Blakely, pers. comm.).

5.10 Discussion

The work reported in this chapter provides the DNA sequence of the *xerC* gene, a gene whose function is absolutely required for *cer* site-specific recombination. It also provides the DNA sequence that spans the gap between the published *E. coli dapF* (Richaud and Printz, 1988) and *uvrD* (Easton and Kushner, 1983; Finch and Emerson, 1983; 1984) sequences. The sequence determined here overlaps for 181 bp with the published *uvrD* sequences (Finch and Emerson, 1983; 1984; Easton and Kushner 1983) with no differences. The *xerC* sequence overlaps at the other end with the published *dapF* 3' flanking sequence (Richaud and Printz, 1988) for 138 bp and for a further 534 bp with the unpublished *dapF* 3' flanking sequence (C. Richaud, personal communications).

There are differences between the sequence determined in our lab and that determined by Richaud. Although these differences are minor, they are crucial to the existence of *orf235*, since some of them produce frame shifts. These differences were therefore checked very carefully on the relevant sequencing gels. Furthermore the sequence reported here is consistent both with the high codon preference statistic throughout the *orf235* reading frame and with the fact that *orf235* is known to be translated at its 3' end.

The sequencing data would best be confirmed by the visualisation of a protein of the correct size to be encoded by *orf235*. An attempt to do this using minicells containing pSDC121 (see Materials and Methods) has so far

failed. One possible explanation for this is that *orf235* may be missing part of its translational initiation signals in pSDC121. A similar experiment using minicells containing pSDC120 (see Materials and Methods) has verified that *orf238* does indeed encode a protein with an apparent molecular weight of approximately 25,000 (data not shown).

The *xerC* gene appears to be transcribed as part of a multicistronic unit that has *dapF* as the first gene and *xerC* as the third gene. The second and fourth genes in this transcription group (*orf235* and *orf238*) are of unknown functions but appear to be translated at similar levels to *dapF* and *xerC*. The fact that these four genes are co-transcribed suggests that their functions might in some way be related.

All genes so far identified in this region, namely the *cya* operon (Aiba *et al*, 1984), the *dapF - xerC* operon and *uvrD* (Easton and Kushner, 1983; Finch and Emerson, 1983; 1984) are transcribed in the same direction, clockwise on the *E. coli* map. This fits in with the observation that the majority of *E. coli* genes are transcribed away from *oriC* (84 min), in the same direction as the direction of DNA replication (Brewer, 1988).

Insertion mutations 5' of the *xerC* open reading frame, within either the *dapF* or *orf235* open reading frames, substantially reduce site-specific recombination at *cer*. The defect caused by at least one of these (a Tn5 insertion in *orf235* to give *xerC1*) is not complemented by the plasmid pSDC106 which carries the complete *orf235* reading frame but only part of *xerC* (see Fig 4.9). All of the insertion mutations upstream of *xerC* are, however, complemented by a plasmid carrying only the intact *xerC* open reading frame. This suggests that upstream insertions reduce *cer* recombination only by reducing transcription of *xerC*, and that the *dapF* and *orf235* gene products are not themselves required for *cer* site-specific recombination. An insertion in *orf238* (downstream of *xerC*) does not abolish recombination at *cer* showing that the product of this gene is not required for *cer* recombination. It is known that *dapF* has its own promoter (Richaud and Printz, 1988), and it seems probable that *dapF*, *orf235*, *xerC* and

orf238 are all transcribed on the same mRNA from this promoter.

Strains carrying insertions within the *xerC* open reading frame are totally defective in recombination between *cer* sites. They are also totally defective in recombination at a variant *cer* site which does not require the function of *argR* or *pepA* for recombination. This variant *cer* site is functionally much smaller than the wild type *cer* site, requiring at most only 50 bp. around the crossover region (Summers, 1989). This suggests that XerC acts at or near the crossover site of *cer*, and - taking into account the similarity to other known recombinases - it is almost certainly the *cer* recombinase.

The XerC protein has substantial amino acid sequence similarity to the lambda integrase family of site-specific recombinases, especially at the two most conserved domains. The second of these domains contains the absolutely conserved tyrosine residue that is known to become linked to the recombination site DNA in lambda integrase (Pargellis *et al*, 1988) and in the FLP protein of the yeast 2-micron plasmid (Gronastajski *et al*, 1985). The other parts of domain 2, including the other two absolutely conserved residues, might also be elements of the recombinase active site (Argos *et al*, 1986). Given this similarity to the lambda integrase family it seems likely that XerC catalyses recombination at *cer* with a mechanism similar to that of these other recombinases. This would involve making two staggered nicks within the *cer* crossover region generating transient 5' protruding ends and 3' phosphodiester protein-DNA linkages to tyrosine 275 of XerC.

In the region between the two conserved domains the XerC sequence is most similar to the sequences of the *E. coli* chromosomally encoded FimB and FimE proteins (Klemm, 1986; Dorman and Higgins, 1987) and to the TnpI protein of *Bacillus thuringiensis* Tn4430 (Mahillon and Lereclus, 1988). FimB and FimE are two *E. coli* proteins which act to reversibly invert a small segment of the *E. coli* chromosome to bring about type I fimbrial phase variation. TnpI probably acts to resolve cointegrate intermediates, formed by the transposition of the class II transposon

Tn4430 (Mahillon and Lereclus, 1988). Within the conserved 38 amino acid domain 2 the XerC sequence is closest to a possible recombinase from plasmid R46 (Hall and Vockler, 1987). The significance of these similarities is not known.

The two *E. coli* genes previously identified as being required for *cer* site specific recombination, *pepA* and *argR*, are known to have other roles in *E. coli* apart from ensuring plasmid stability. It seems likely that *xerC* also has some other role in *E. coli*. This is supported by the observation that *xerC* mutants have altered growth characteristics and cell morphology, which are complemented *in trans* by *xerC*. The cellular role of *xerC* might involve acting at an *E. coli* chromosomal recombination site or sites. XerC could either act (like FimB and FimE) to invert the DNA between two chromosomal recombination sites, or it could act at a single recombination site to resolve any chromosome multimers that arise by homologous recombination. The second of these two possibilities is consistent with the phenotype seen in *xerC* mutant strains. Chromosomal multimers should fail to partition and could give filaments with amplified DNA. If XerC is involved in chromosome monomerisation, it would appear that FimB and FimE can substitute, at least in part, for this function; probably by acting at their normal recombination sites.

Most site-specific recombinases are found in very close proximity to their sites of action. However, by using both sequence comparison to *cer* and a functional (dimer resolution) assay, we have not been able to detect any such site in the *cya* - *xerC* - *uvrD* chromosomal region. A *cis*-acting locus (*dif*) has recently been identified in the terminus region of the *E. coli* chromosome. Disruption of this locus leads to a phenotype very similar to that seen in *xerC* mutants, giving aberrant nucleoids and causing the cells to filament (Kuempel, pers. comm.). The *dif* locus has been sequenced and contains a small region of sequence similarity to the crossover region of *cer*. Garry Blakely is currently working in collaboration with Kuempel *et al* to ascertain whether *dif* does indeed contain a *cer*-like recombination site.

CHAPTER SIX

PARTIAL PURIFICATION OF XerC AND INVESTIGATION OF ITS *cer* BINDING ACTIVITY

6.1 Introduction

One of the long term aims of the *cer* / *xer* project is to develop an *in vitro* recombination system. This would help to clarify the role of PepA and ArgR in *cer* site-specific recombination and might also allow the identification of other factors, such as IHF, Fis and transcription, which could be involved in *cer* recombination. *In vitro* recombination sytems exist for the lambda Int / *att* system (Nash, 1975), the yeast 2-micron FLP system (Vetter *et al*, 1983), the bacteriophage P1 Cre / *lox* system (Abremski *et al*, 1983) and other lambda integrase class recombination systems as well as for many members of the Resolvase / DNA invertase family of recombinases (eg Reed and Grindley, 1981; Kahmann *et al*, 1985). All of these *in vitro* recombination systems employ simple buffered solutions containing some NaCl and either Mg^{2+} ions or spermidine. They all require the recombinase protein and some also require the addition of accessory proteins such as IHF, Fis or Xis.

In the absence of an *in vitro* recombination assay, it was decided to develop an *in vitro* DNA binding assay for XerC. If XerC is indeed the *cer* recombinase, then it should display a DNA binding activity which is specific to the *cer* sequence, probably in the crossover region of *cer*. Once a binding assay was developed it could be used to follow the purification of XerC. Purified XerC could be used in an *in vitro* recombination system, once such a system is developed.

A particularly sensitive technique for the detection of protein binding to specific DNA fragments is the gel binding assay (Fried and Crothers, 1981; Garner and Revzin, 1981). This involves electrophoresis of protein-DNA complexes through non-denaturing polyacrylamide gels. Binding of protein to a radioactive (end-labelled) DNA molecule is detected by a reduction in mobility of the DNA fragment. This reduction in mobility is caused either by an increase in the apparent size and/or bending of the DNA fragment. Specific protein-DNA complexes can be formed in the presence of a large excess of unlabelled "non-

specific" DNA. This competes out the binding of non-specific DNA-binding proteins to the labelled DNA fragment.

The gel binding assay was first used to study the binding of purified CAP protein and *lac* repressor to the regulatory sequences of the *lac* promoter (Garner and Revzin, 1981; Fried and Crothers, 1981). Gel binding assays have since been used to study the binding of many purified proteins to their DNA binding sites (eg ArgR binding to *cer*: Stirling *et al*, 1988b; MetJ binding to *met* operon promoters: Phillips *et al*, 1989). The gel binding assay has also been used to detect DNA binding activities in crude extracts, and to follow the purification of DNA binding proteins from these crude extracts (eg Strauss and Varshavsky, 1984; Morrell, 1990).

The gel binding assay has several advantages over the filter binding assay which was traditionally used to assay binding of proteins to DNA molecules. The filter binding assay can only distinguish two classes of DNA molecules; those bound by protein and those not bound. The gel binding assay not only separates bound from unbound DNA but also resolves the protein-bound DNA into distinct complexes. From the number and pattern of complexes produced it is often possible to make inferences about the nature of each complex. This type of approach has been used to study the binding of Tn3 resolvase to the Tn3 *res* site (Bednarz, 1989). Yet more can be learned about the nature of protein - DNA interactions in the various complexes by carrying out footprinting reactions on complexes isolated by gel electrophoresis.

Gel binding assays have been used to study the binding of several members of the lambda integrase family of site-specific recombinases to their recombination sites. Lambda Int binds strongly to the arm-type sites in attP but only weakly to the core-type binding sites (Thompson and Landy, 1988). Binding of lambda Int to the core-type sites can be detected in a gel binding assay, but the pattern of complexes produced is unclear (Thompson and Landy, 1988).

Binding of Flp to its recombination site (FRT) can

also be detected in gel binding assays. The FRT site contains two 13 bp inverted repeat sequences flanking an 8 bp spacer region, with an additional 13 bp repeat located to one side (Andrews *et al*, 1985). Flp binds to FRT and gives three distinct complexes in gel binding assays (Andrews *et al*, 1987). Footprinting experiments carried out on these complexes showed that each complex represented a unique Flp - DNA association. Flp produces a single complex in gel binding assays when it binds to a DNA fragment containing only one 13 bp repeat sequence (Andrews *et al*, 1987). It seems likely that the three complexes result from the binding of one, two or three Flp molecules to the intact FRT site (Andrews *et al*, 1987).

The recombination site for the protein Cre, *lox*, consists of two 13 bp inverted repeat sequences flanking a spacer region of 8 bp. Cre binds *lox* to produce three complexes which can be separated by non-denaturing gel electrophoresis (Wierzbicki *et al*, 1987). The stoichiometry of these complexes has been determined. The protein:DNA ratio is approximately 1:1 for complex 1, 2:1 for complex 2 and 3:1 or 4:1 for complex 3 (Hoess, pers Comm.). There is evidence, however, that complex 3 is not a simple Cre - *lox* complex, and may be some sort of recombination intermediate (Hoess, pers comm.).

This chapter reports the development of a gel binding assay to detect the binding of XerC to the *cer* site and the use of this assay to follow XerC through a partial purification.

6.2 Over expression of *xerC*

The plasmid pSDC105 was constructed in order to over express xerC. This plasmid contains the 1,232 bp $XerC^+$ fragment from pSDC104 in the vector pBAD. The insert in pSDC105 is in the correct orientation for *xerC* to be transcribed from the strong IPTG-inducible P_{tac} promoter. The following preliminary experiments, using the plasmid pSDC105, were carried out by George Szatmari.

Cultures of DS941 carrying pSDC105 were grown overnight in L broth plus ampicillin, with or without 1mM

IPTG. Total cellular protein was run on an SDS-polyacrylamide gel. An extra protein of apparent molecular weight 32,000 was present in the IPTG induced culture. This protein appeared to make up 2-5% of total cellular protein in the induced culture. The apparent size of this protein is in good agreement with the predicted molecular mass of XerC, 33.8 kDa.

A culture of DS941 / pSDC105 was grown overnight in the presence of IPTG, the cells were harvested and sonicated in a low salt buffer. The sonicate was then centrifuged to produce a cleared lysate. The majority of the 32 kDa protein was insoluble and was found in the insoluble pellet fraction. The 32 kDa protein was resuspended from this fraction in a buffer containing 1M NaCl. On storage, the 32 kDa protein precipitated and a fraction consisting largely of the 32 kDa protein was obtained by centrifugation. This fraction was run on a SDS-polyacrylamide gel, the 32 kDa protein was eluted from a gel slice and precipitated with acetone by the method of Hager and Burgess (1980). The purified protein was then subjected to Edman degradation on an Applied Biosystems automated peptide sequencer to obtain the N-terminal protein sequence. The sequence obtained (20 amino acids) agreed entirely with that predicted from the DNA sequence. This confirms that the 32 kDa protein seen in pSDC105 containing, IPTG-induced cultures is indeed XerC.

The precipitated XerC protein could only be dissolved with the use of harsh treatments, such as boiling in the presence of SDS. No DNA binding or recombinase activity could be demonstrated for this precipitated form of XerC. The reasons for the precipitation of this inactive form of XerC are not clear. In case it had anything to do with the very large amounts of XerC produced by overnight expression, a different expression protocol was developed for the subsequent work carried out by myself. Overnight cultures carrying pSDC105 were diluted 1:80 in fresh L broth containing ampicillin. They were grown with good aeration at 37^{o}C to a density giving an A_{600} of 0.5 and were induced by the addition of IPTG to a final concentration of 1mM. The cultures were then grown a

further three hours with good aeration at 37°C. A (pSDC105 specific) polypeptide of apparent molecular weight 32,000 was visible after SDS-polyacrylamide gel electrophoresis of total cellular protein. However this protein made up less than 1% of total cellular protein. Cultures treated in this way were used to produce protein extracts for the experiments described in the rest of this chapter.

6.3 Development of a gel binding assay

A culture of DS941 *xerC2* containing pSDC105 was grown and induced as described above. Cells were harvested by centrifugation and washed once in 10 mls of a buffer containing 10 mM Tris-HCl pH 7.5 and 100 mM NaCl. The wet cell pellet was weighed and then frozen at -70°C.

The cell pellet (weighing 0.80 g) was thawed and resuspended in 3 mls of buffer B (50 mM Tris-HCl pH 7.5, 1.0 M NaCl, 1 mM EDTA and 0.1 mM DTT). 0.8 mg of PMSF (phenylmethylsulphonyl fluoride, a protease inhibitor) dissolved in 50 ul of ethanol was added dropwise to the unbroken cells. The cells were then broken by sonication with a Dawe sonicator using the button probe. Seven pulses of ten seconds, each followed by one minute to allow cooling, were carried out on ice.

The sonicate was then cleared by centrifugation for 15 minutes at 18,000 rpm in a JA21 rotor (37,000 x g) at 4°C. The cleared sonicate was then diluted in buffer B to make a two-fold dilution series and then used in a gel binding assay as described in Chapter 2. The extract was incubated with either the *cer* containing 304 bp BamHI - HindIII fragment from pKS492 (see Fig. 6.8) or with the "non-specific" 235 bp DdeI fragment from pUC19. Complexes were run on a polyacrylamide Tris-glycine pH 9.4 gel and, after autoradiography, the result shown in Fig 6.1 was obtained.

As can be seen, the crude extract did not retard the DdeI fragment from pUC19, but did retard the *cer* containing fragment. At the lower protein concentrations one distinct retarded complex was visible. At the highest protein concentration nearly all of the *cer* containing

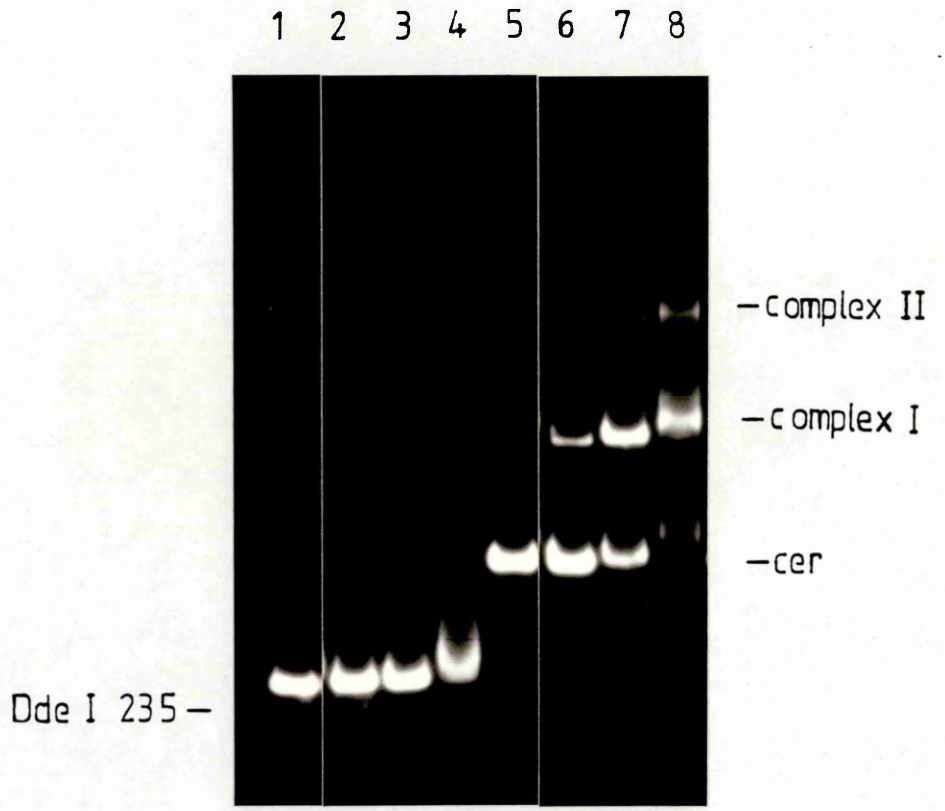


Figure 6.1 Gel binding assay on a XerC$^+$ crude extract using non-specific and cer fragments.
A DS941 xerC2/pSDC105 culture was induced as described in the text and sonicated in a buffer containing 50mM Tris-HCl pH 7.5, 1mM EDTA, 0.1mM DTT and 1.0 M NaCl. A cleared lysate was produced by centrifugation (12,000g, 10 min). The lysate was diluted in the same buffer to give a two-fold dilution series. Approximately 1 ng of end-labelled fragment was mixed with 1 ul of each protein dilution and incubated at 37^{o}C for 10 min. Reactions were then loaded on a non-denaturing polyacrylamide gel as described in Chapter 2. Lanes 1-4 contain the 235 bp DdeI fragment from pUC18. Lanes 5-8 contain the cer site cut out as a 304 bp BamHI - HindIII fragment from pKS492.

Lanes 1 and 5	0	protein
Lanes 2 and 6	1/8	ul protein
Lanes 3 and 7	1/4	ul protein
Lanes 4 and 8	1/2	ul protein

(The undiluted crude extract contained of the order of 15 mg/ml total protein.)

fragment was bound and two distinct complexes were visible.

It appeared from this experiment that a protein in the crude extract was binding specifically to *cer* DNA. An experiment was then carried out to ascertain whether this retardation of *cer* was dependent upon the expression of *xerC*.

Two parallel cultures were grown and induced with IPTG as described above. One of these cultures contained DS941 *xerC2*/pSDC105, the other contained DS941 *xerC2*/pBAD. Extracts from both of these cultures were produced by sonication in buffer B, followed by centrifugation, as described above. DS941 *xerC2*/pBAD contains no fuctional *xerC* gene, so the extract produced from this strain contains no XerC. The extract produced from DS941 *xerC2*/pSDC105 contained significant quantities of XerC (not shown).

These two extracts were then assayed for their ability to bind *cer* DNA. A gel binding assay was carried out using a mixture of two labelled DNA fragments. One of these fragments was the *cer*-containing 304 bp BamHI - HindIII fragment from pKS492 (see Fig 6.8), the other was the 235 bp DdeI fragment from pUC19. The results of this experiment are shown in Fig 6.2.

The sonicate produced from DS941 *xerC2*/pBAD produced no detectable retardation of either the "non-specific" or the *cer*-containing fragments. The XerC-containing extract from DS941 *xerC2*/pSDC105 retarded the *cer*-containing DNA fragment to give the same pattern seen before, but appeared not to bind to the DdeI pUC19 fragment.

All of these gel binding experiments contained a large excess (about 200 ng) of unlabelled, supercoiled pUC18 carrier DNA to compete out binding to the labelled fragments of any proteins that bind non-specifically to DNA. Therefore the choice of the 235 bp DdeI fragment from pUC19 as a control "non-specific" fragment was perhaps not ideal. Since this fragment came from pUC19, its sequence was represented in approximately 20 fold excess in the unlabelled carrier DNA (the reactions contained about 1 ng of 235 bp fragment and 200 ng of 2686 bp pUC18).

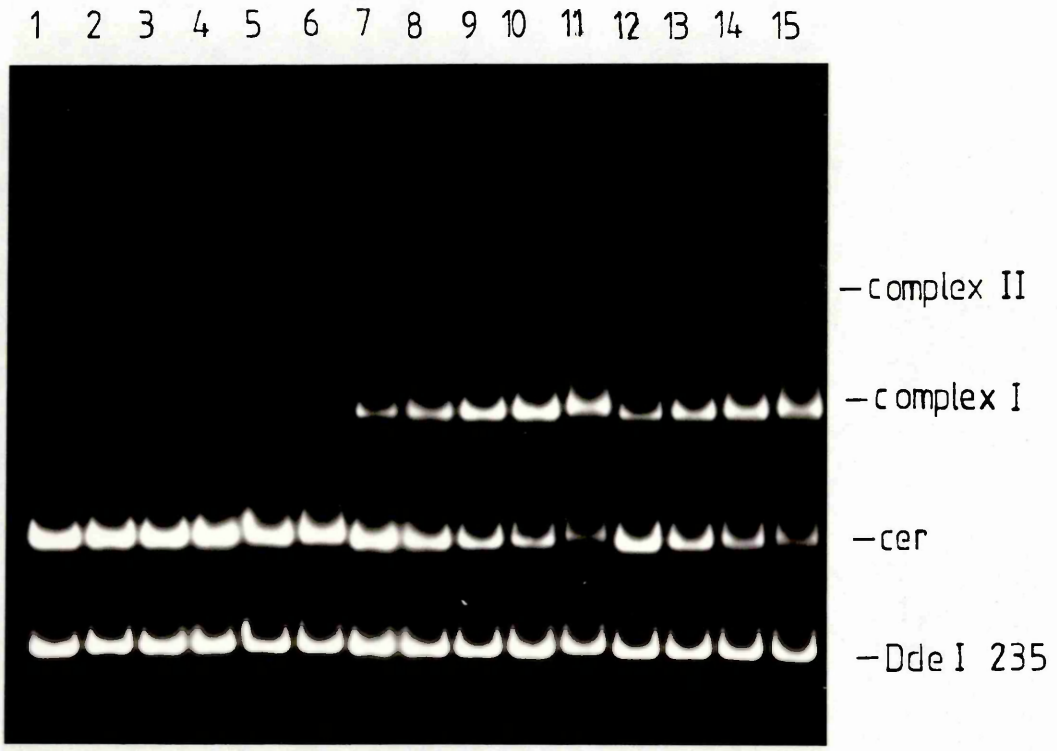


Figure 6.2 Gel binding assay on XerC$^+$ and XerC$^-$ crude extracts using a mixture of two fragments. A wild type cer site (pKS492, HindIII-BamHI) was mixed with the 235 bp DdeI fragment from pUC18 and incubated with different dilutions of crude extract for 10 min at 37^{o}C. Crude extracts were prepared from DS941 xerC2/pBAD (lanes 2-6) or DS941 xerC2/pSDC105 (lanes 7-15). The cells were sonicated in a buffer containing 50mM Tris-HCl pH7.5, 1mM EDTA, 0.1mM DTT and either 1M NaCl (lanes 2-11) or 100mM NaCl (lanes 12-15). Lane 1 contains no added protein.

Lanes 2, 7 and 12	1/32 ul protein
Lanes 3, 8 and 13	1/16 ul protein
Lanes 4, 9 and 14	1/8 ul protein
Lanes 5, 10 and 15	1/4 ul protein
Lanes 6 and 11	1/2 ul protein

(the undiluted crude extracts contained of the order of 15 mg/ml total protein.)

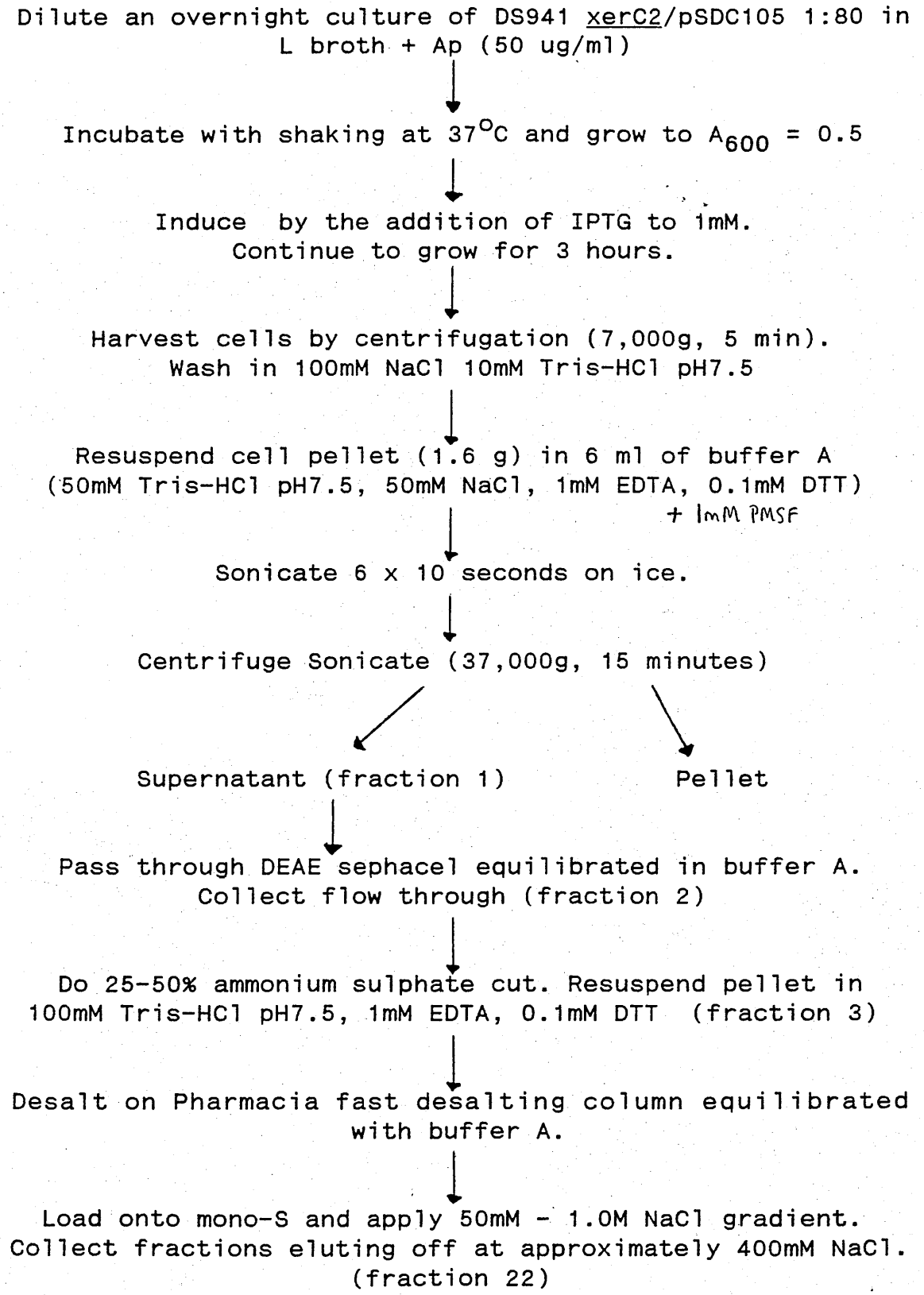
6.4 Partial purification of XerC

The gel binding assay was then used to follow the partial purification of the XerC protein. In order to quantitate this purification, the XerC binding unit was defined as the amount of XerC which retards 50% of a *cer*-containing DNA fragment, when added to a standard gel binding assay. The effects of varying the amount of *cer* fragment in the gel binding assay were never investigated. However, if the same amount of fragment (1 ng of a 300 bp fragment) is always used, this assay should be reasonably quantitative for XerC.

The effect of salt concentration in the cell breaking buffer was investigated. IPTG-induced DS941 *xerC2*/pSDC105 cell pellets were sonicated in 50 mM Tris-HCl pH7.5, 1 mM EDTA, 0.1 mM DTT containing either 1.0 M NaCl, 100 mM NaCl or 50 mM NaCl. The concentration of NaCl present had little effect on either the recovery of XerC binding activity (Fig 6.2 and not shown), or on the recovery of XerC polypeptide in the soluble fraction, as visualised by SDS-polyacrylamide gel electrophoresis (not shown).

A purification protocol was arrived at by repeatedly preparing sonicates from cultures of DS941 *xerC2*/pSDC105, and experimenting with various protein fractionation steps. The fractions were run on SDS-polyacrylamide gels and stained with Coomassie blue to visualise proteins. This allowed the XerC polypeptide to be followed through the purification. The insoluble XerC which had been purified by G. Szatmari was used as a size marker on these gels. Small samples from each fraction were also diluted 50% in glycerol and stored at -20^{o}C. These were used at a later time to assay the various fractions for XerC binding activity. Thus the purification was followed both by the presence of XerC-sized proteins and by XerC binding activity. However, the objective of the purification was to purify the binding activity rather than just the 32 kDa protein.

The purification protocol finally arrived at is summarised in Fig 6.3. A typical protein preparation is described below.



Figur 6.3 Flow chart for the purification of XerC.

A cleared sonicate (fraction 1) was produced as above from a 1.8 g cell pellet of IPTG-induced DS941 *xerC2*/pSDC105 suspended in 6 ml of buffer A (50 mM Tris-HCl pH7.5, 50 mM NaCl, 1 mM EDTA, 0.1 mM DTT). All subsequent purification steps were carried out at 4^{o}C.

A 6ml DEAE sephacel column was poured in a 10 ml syringe and equilibrated with buffer A. The cleared sonicate was loaded onto this column and allowed to drip slowly through by gravity. The column was then washed with further buffer A. The flow-through was collected in fractions of approximately 1 ml. The first three fractions were found to contain very little protein (as assayed by Bradford reagent) and were discarded. The next eight 1 ml fractions contained the bulk of the XerC binding activity and were pooled to give fraction 2. The column was washed with further buffer A and then proteins which had bound to the column were washed off with a buffer containing 50 mM Tris HCl, 1.0 M NaCl, 1mM EDTA and 0.1 mM DTT. The buffer A flow-through fraction was found to contain at least one protein of the expected size for XerC, whereas the 1M NaCl wash fraction contained no proteins of this size.

The DEAE step was originally designed to remove nucleic acids from the sonicate. However, as can be seen from Fig 6.4, a number of proteins were also removed by this step. The yield of XerC binding activity from this step was near to 50% (Table 6.1). The NaCl concentration was kept as low as possible (50mM NaCl) to maximise the number of proteins binding to the DEAE column.

The DEAE flow through (fraction 2) was further purified by ammonium sulphate fractionation. Proteins precipitating in the ranges 0% - 25%, 25% - 50% and 50% - 70% saturation ammonium sulphate were re-dissolved in 1 ml of a buffer containg 100 mM Tris-HCl pH 7.5, 1 mM EDTA and 0.1 mM DTT.

The 25% - 50% ammonium sulphate fraction (fraction 3) was found to contain the bulk of proteins of the correct size to be XerC (Fig 6.4). The 0% - 25% ammonium sulphate fraction contained some protein of the correct size to be XerC, but this material was very difficult to resuspend (Fig 6.4). The 25% - 50% ammonium sulphate fraction was

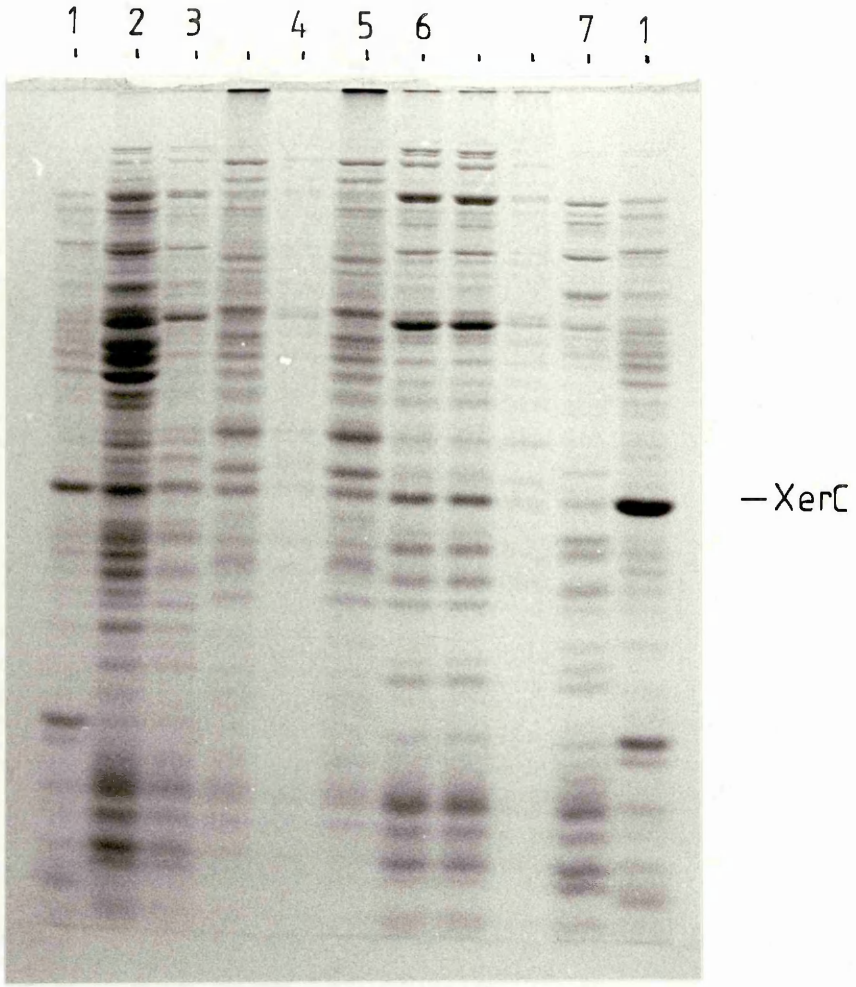


Figure 6.4 SDS-polyacrylamide electrophoresis of fractions obtained from DEAE Sephacel and ammonium sulphate precipitation.

Lane 1 XerC size marker (from G. Szatmari)
Lane 2 Cleared sonicate from DS941 xerC2/pSDC105 (50mM NaCl)
Lane 3 DEAE Sephacel flow through
Lane 4 0% - 25% ammonium sulphate fraction (soluble fraction)
Lane 5 0% - 25% ammonium sulphate fraction (insoluble)
Lane 6 25% - 50% ammonium sulphate fraction (soluble)
Lane 7 50% - 70% ammonium sulphate fraction (soluble)

the only fraction that contained any significant XerC binding activity (Fig 6.7).

The 25% - 50% ammonium sulphate fraction was then passed through a fast desalting column (Pharmacia), which had previously been equilibrated with buffer A. The run was carried out at 2 ml/min on a Waters FPLC system. The protein was recovered in 13 mls of buffer A.

The desalted protein was loaded onto a mono-S cation exchange column which had been equilibrated with buffer A. The column was then developed with a 17.5 ml gradient running at 0.5 ml/min. The gradient started with 100% buffer A (50 mM Tris-HCl pH7.5, 50 mM NaCl, 1 mM EDTA, 0.1 mM DTT) and ended with 100% buffer B (50 mM Tris-HCl pH 7.5, 1.0 M NaCl, 1 mM EDTA and 0.1 mM DTT). Fractions of 0.5 ml were collected. Elution of proteins from the column was monitored by measuring the absorbance at 280 nM (Fig 6.5).

The pooled flow-through fraction and also fractions that eluted during the gradient were analysed by SDS-polyacrylamide gel electrophoresis (Fig 6.6). No proteins of the correct size for XerC were present in the flow through. Two proteins of approximately the same size as XerC eluted from the column at approximately 400 mM NaCl (fractions 20 - 23) and another eluted at approximately 800 mM NaCl (fractions 31 - 32, not shown).

The fractions were assayed for XerC activity using the gel binding assay (Fig 6.7). Binding activity was found in the fractions that coincided with the presence of the larger of the two XerC-sized proteins eluting at 400 mM NaCl (fractions 21 - 23). No XerC activity was found in the flow-through fraction or in the fractions that eluted between 700 mM and 900 mM NaCl (fractions 28 - 33) (not shown).

The purification is summarised in Table 6.1. Protein concentrations were measured using Bradford reagent. XerC binding activity was estimated from the gel binding assay shown in Fig 6.7. In some lanes of Fig 6.7, the fraction being assayed contained XerC binding activity but the amount of protein added did not not retard 50% of the *cer* containing fragment. In these cases the number of binding

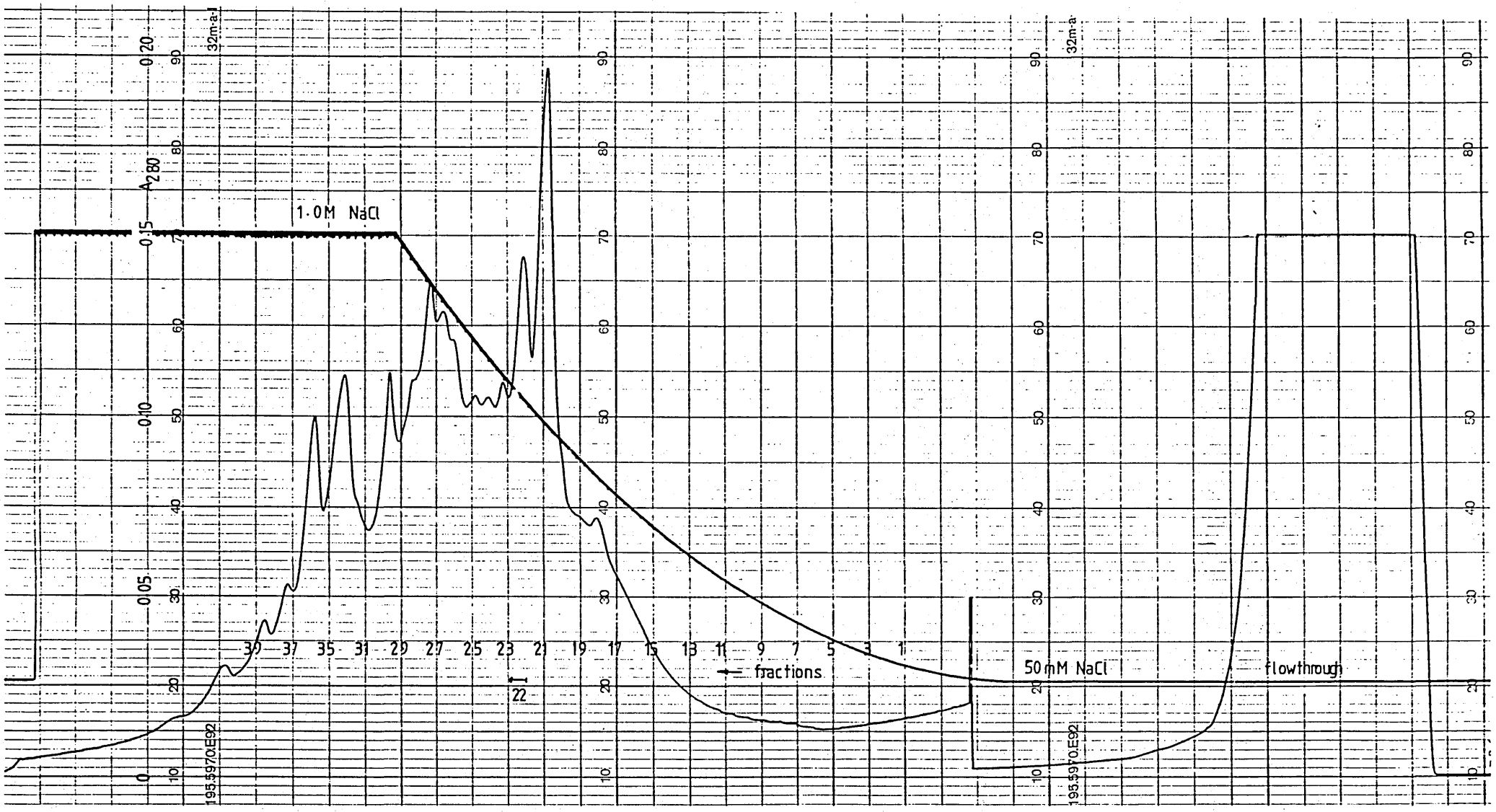


Figure 6.5 Elution of proteins from the mono-S ion exchange column as monitored by absorbance at 280nM. The desalted 25-50% ammonium sulphate fraction was loaded on to the column in buffer A (containing 50mM NaCl). The column was washed with 10 ml of buffer A and then developed with a 17 ml gradient from 50 mM NaCl to 1.0 M NaCl. The gradient was applied at a rate of 0.5 ml/minute. Each square represents 2 min (1ml). Note that the plot of A_{280} lags behind the plot of the gradient by approximately 3 ml.

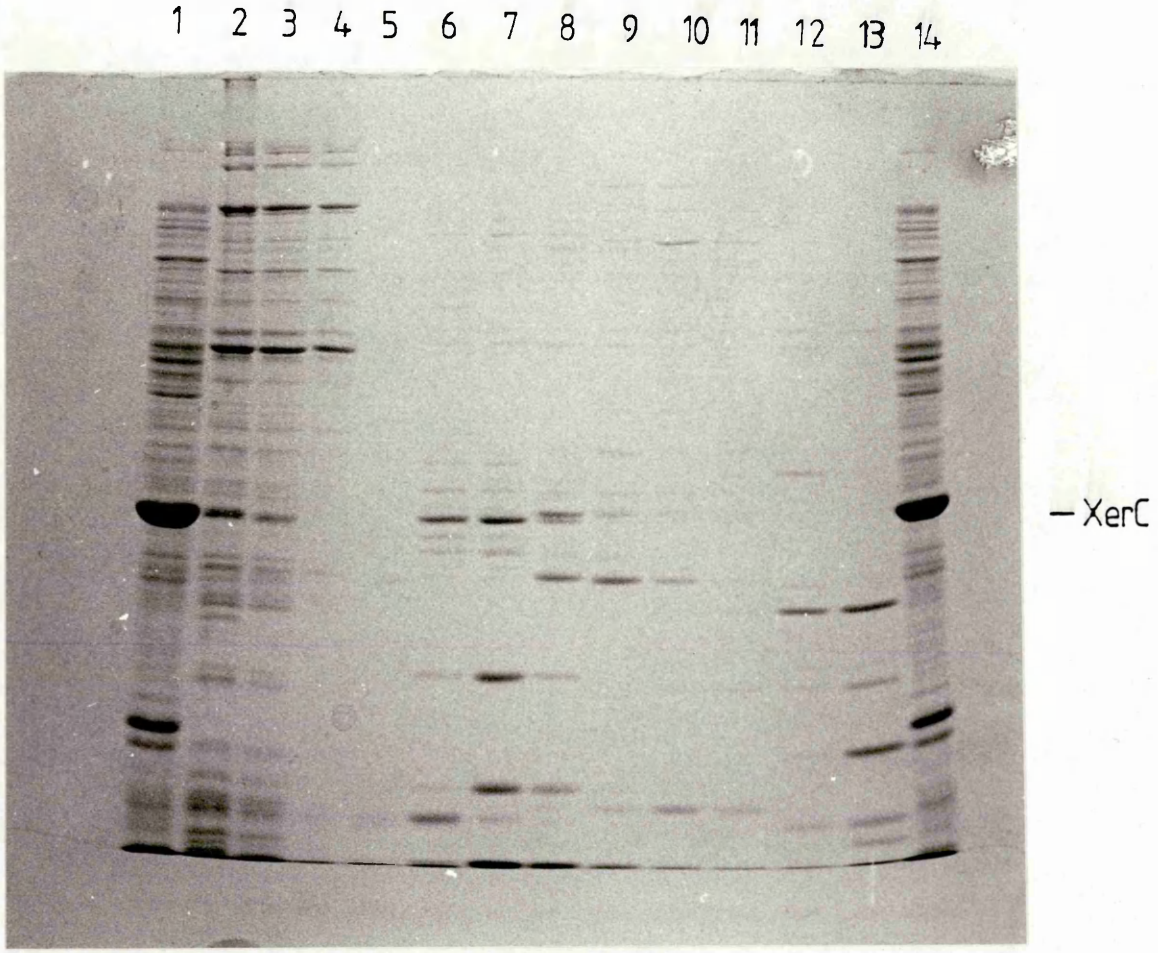


Figure 6.6 SDS-polyacrylamide electrophoresis of fractions obtained from mono-S ion exchange chromatography.

Lanes 1, 14	XerC size marker (from G. Szatmari)
Lane 2	25-50% ammonium sulphate fraction
Lane 3	25-50% " " " desalted
Lane 4	Mono-S flow-through
Lane 5	Mono-S fraction 19
Lane 6	Mono-S fraction 20
Lane 7	Mono-S fraction 21
Lane 8	Mono-S fraction 22
Lane 9	Mono-S fraction 23
Lane 10	Mono-S fraction 24
Lane 11	Mono-S fraction 25
Lane 12	Mono-S fraction 26
Lane 13	Mono-S fraction 27

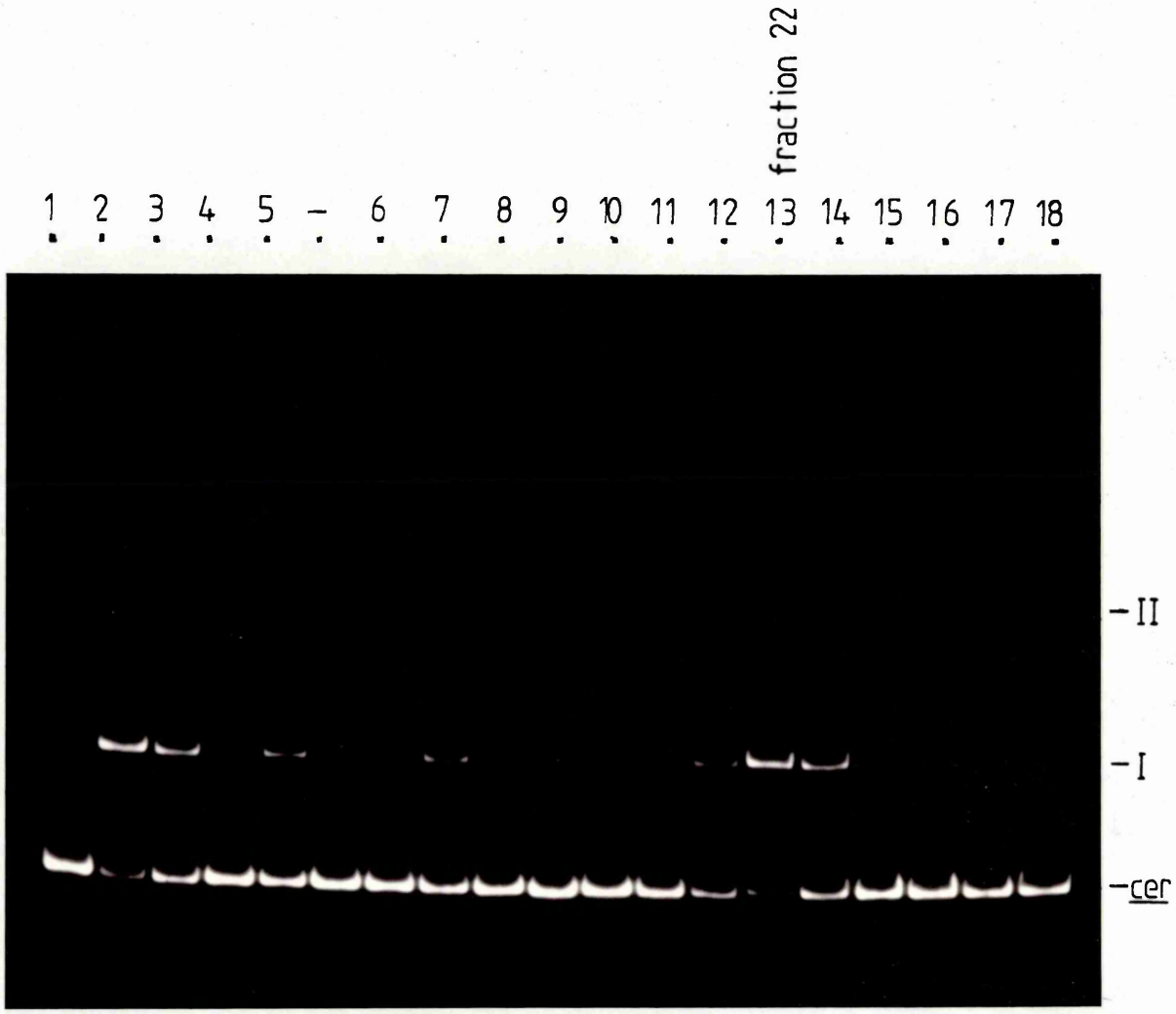


Figure 6.7 Gel binding assay on various fractions obtained in the purification of XerC. The following amounts of protein were incubated with approximately 1ng of end labelled cer DNA for 10 min at 37°C and then loaded on a non-denaturing polyacrylamide gel.

Lane	1	0 protein
Lane	2	1/8 ul cleared sonicate
Lane	3	1/8 ul DEAE flow through
Lane	4	1/10 ul 0-25% ammonium sulphate
Lane	5	1/20 ul 25-50% ammonium sulphate
Lane	6	1/20 ul 50-70% ammonium sulphate
Lane	7	1/10 ul 25-50% ammonium sulphate (spun)
Lane	8	1/2 ul desalted 25-50% ammonium sulphate
Lane	9	1 ul Mono-S flow through
Lanes	10-18	1 ul each of Mono-S fractions 19-27

Table 6.1 Purification of XerC

Fraction	Volume	Total Protein	Total Activity	Yield	Specific Activity	Purification
Cleared sonicate	5.5 ml	210 mg	88 kU	100%	410 U/mg	1 x
DEAE flow through	5.8 ml	19 mg	46 kU	52%	2,300 U/mg	5.5 x
25-50% AS cut	1 ml	4.2 mg	10 kU	11%	2,400 U/mg	5.8 x
25-50% AS (spun before desalting)	1 ml	/	7.5 kU	8.5%	/	/
Desalted	10 ml	3.5 mg	7.5 kU	8.5%	2,100 U/mg	5.2 x
Mono-S 21	0.5 ml	75 ug	1 kU	1.1%	13,500 U/mg	33 x
Mono S 22	0.5 ml	65 ug	2 kU	2.3%	31,000 U/mg	75 x
Mono-S 23	0.5 ml	65 ug	0.5 kU	0.6%	7,700 U/mg	18 x
Mono-S (Total)	1.5 ml	205 ug	3.5 kU	4.0%	/	/

units present was estimated by extrapolation from the results obtained in other XerC binding experiments (eg Fig 6.1 and Fig 6.2). It is estimated that the values given for the XerC binding activities in Table 6.1 are probably accurate to within a factor of two.

Fraction 22 contained the highest concentration of XerC binding activity. As can be seen on Fig 6.6, fraction 22 contained one major polypeptide of 32 kDa and six or seven other major bands. This 32 kDa protein had co-purified with the XerC binding activity and was the same size as a protein which had been positively identified as XerC by N-terminal sequencing. It was estimated that the XerC - sized protein made up about 15% of the protein in fraction 22.

The approximate concentration of XerC required to retard *cer*-containing DNA fragments can now be estimated. The protein concentration of fraction 22 is 130 ug/ml. 1 ul of fraction 22 was added to the 10ul reaction shown in lane 13 of Fig. 6.7. The total protein concentration in the binding reaction was therefore 13 ug/ml, so the concentration of XerC-sized protein was approximately 2 ug/ml, which is 6×10^{-8} M. The amount of *cer* DNA in the reaction is approximately 1 ng which corresponds to a concentration of 5×10^{-10} M.

6.5 XerC binds to the crossover region within *cer*

Gel binding experiments were carried out to localise the binding site or sites of XerC within the *cer* region. These experiments used fragments containing only part of the *cer* site. The derivation of these fragments is shown in Fig 6.8.

The PstI - BamHI fragment from pKS493 was end-labelled with ^{32}P using the Klenow fragment of DNA polymerase I and alpha $\langle^{32}P\rangle$ dATP. This labelled only the end closest to the crossover site in *cer*. The labelled fragment was then cleaved with MluI, splitting the *cer* site into two fragments. One of these (73 bp in length) contained the crossover region and was labelled at one end. The other fragment (219 bp) contained the rest of *cer*

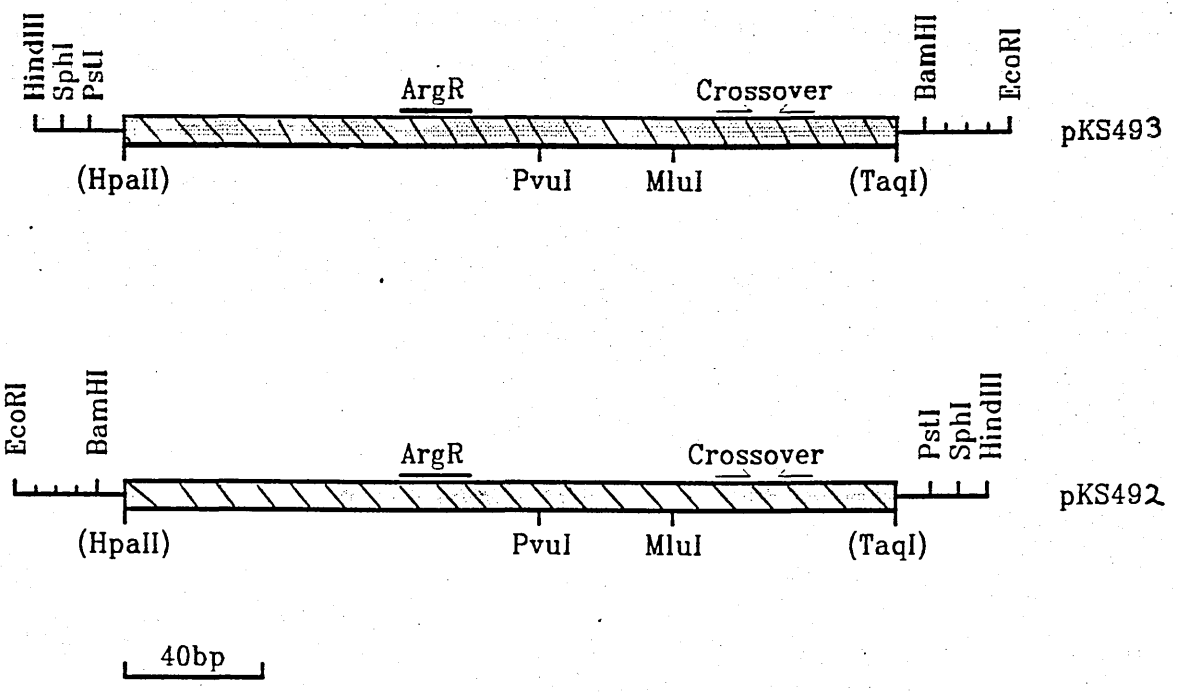


Figure 6.8 Derivation of the fragments used in gel binding assays.

and was unlabelled. This mixture of fragments was incubated with the most highly purified fraction of XerC, fraction 22 from the mono-S column. The reactions were then run on a non-denaturing polyacrylamide gel as described in Chapter 2. Reactions were also carried out to assay the binding of the same protein dilutions to the intact *cer*-containing DNA fragment. The result obtained after autoradiography is shown in Fig 6.9.

At the same protein concentration that produced two complexes on a full *cer* site, the 73 bp crossover region was also bound to a similar extent. The full *cer* site gave the previously observed pattern of two major complexes, with the least retarded complex predominating. The 73 bp crossover region gave only one major complex and a (less retarded) minor complex.

In another experiment a *cer*-containing fragment was labelled at both ends and then cleaved with MluI. This mixture of labelled fragments was then used in a gel binding experiment. As assayed by the disappearance of the unbound fragments, only the 73 bp crossover containing region of *cer* was bound significantly by XerC (not shown). However, no complexes could be seen in this experiment. This may have been because the complex produced from the crossover region comigrated with the unbound large fragment.

A *cer* containing fragment, labelled at both ends and then cleaved at the PvuI site was also used in a gel binding experiment. The results of this experiment are shown in Fig 6.10. Note that the cleavage with PvuI was partial. As assayed by dissappearance of unbound bands, only the smaller, crossover site-containing fragment was bound by XerC. One band, which probably corresponds to a bound form of this smaller fragment, can be seen on Fig 6.10.

These experiments show that, when the *cer* site is cleaved into two fragments, XerC binds only to the fragment containing the crossover site. XerC binds to these crossover site-containing fragments with approximately the same affinity as it binds to the intact site. However, the pattern of complexes produced by

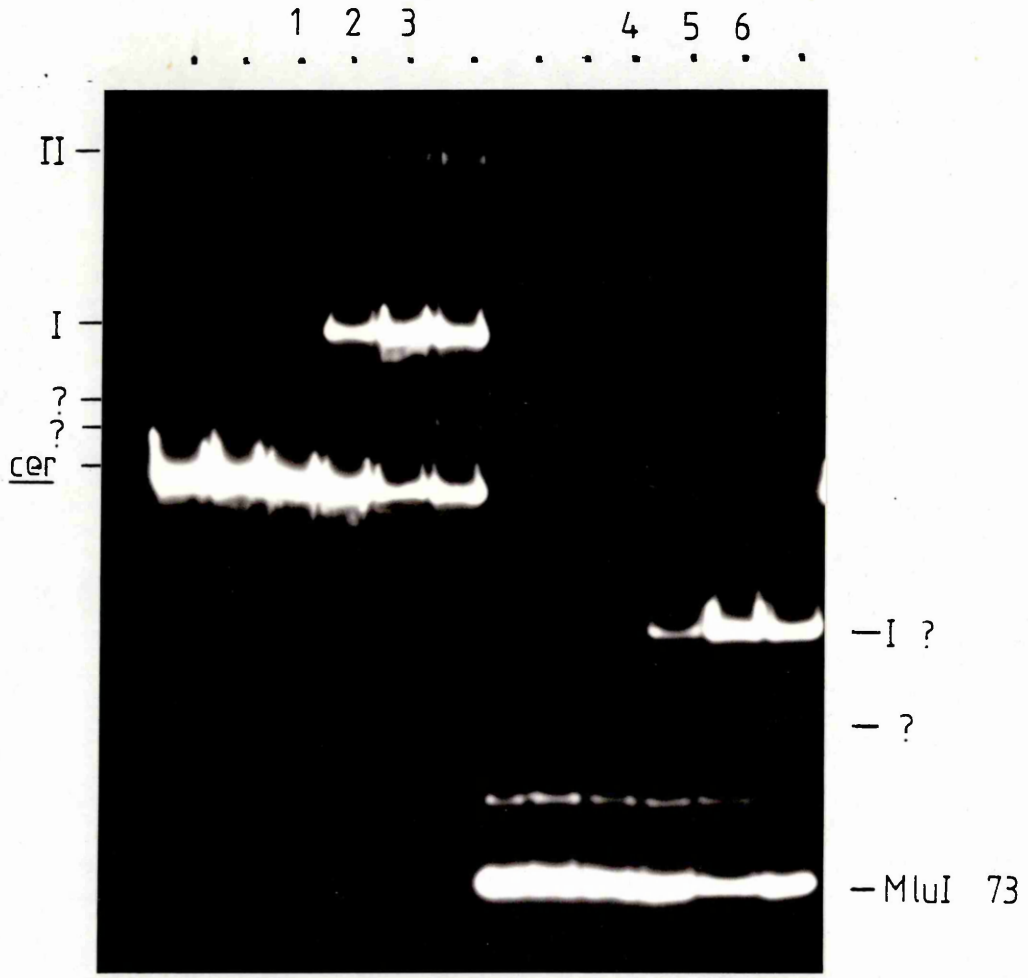


Figure 6.9 Gel binding assays on the mono-S fraction 22 using the crossover region of cer and the full cer site.
Lanes 1-3 contain the purified BamHI - PstI fragment of pKS493 labelled at the BamHI end. Lanes 4-6 contain the same fragment cleaved with MluI. Only the 73 bp fragment containing the crossover region is labelled.

Lanes 1 and 4	0	protein
Lanes 2 and 5	1 ul	fraction 22
Lanes 3 and 6	1/2 ul	fraction 22

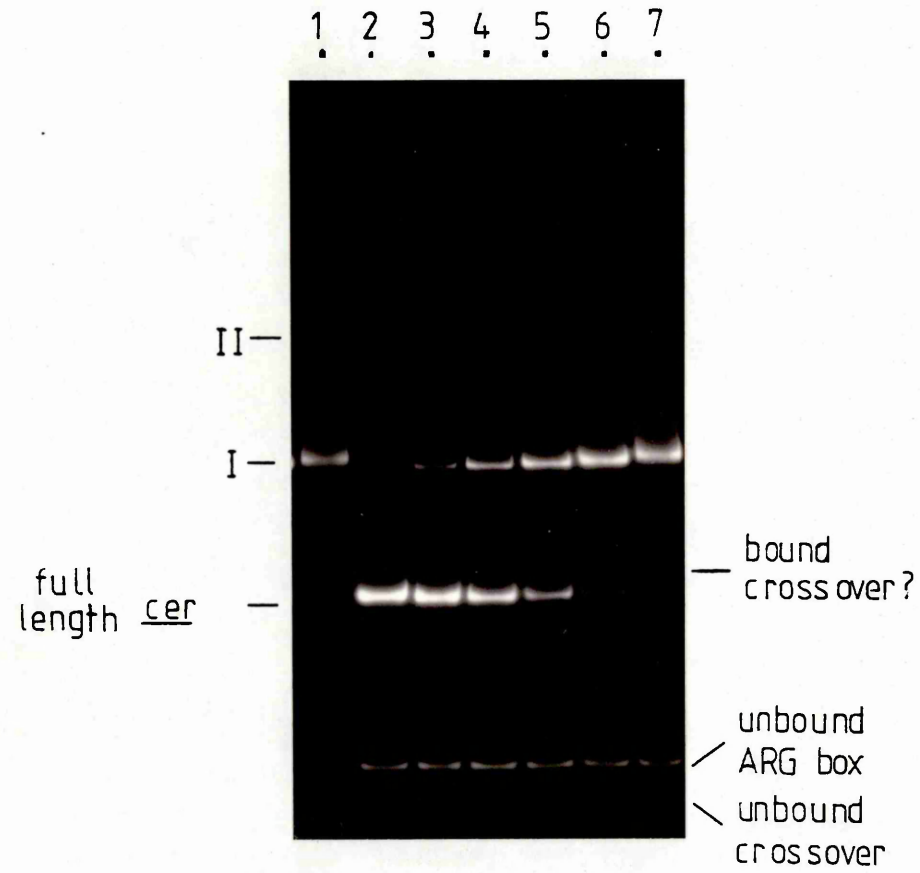


Figure 6.10 Gel binding assay using a cer fragment which has been cleaved with PvuI. End labelled fragments were incubated with different amounts of 25% - 50% ammonium sulphate fraction (fraction 3) containing XerC. Lanes 2-7 contain the 304 bp HindIII - BamHI fragment from pKS492 labelled at both ends and then cleaved with PvuI. Note that the PvuI cleavage was only partial. The DNA in lane 1 was not treated with PvuI.

Lanes 1 and 7	1/2 ul	protein
Lane 2	0	protein
Lane 3	1/32 ul	protein
Lane 4	1/16 ul	protein
Lane 5	1/8 ul	protein
Lane 6	1/4 ul	protein

(Undiluted fraction 3 contained approximately 3 mg/ml total protein.)

binding to the isolated crossover site appears to be different to the pattern produced by binding to the intact site.

6.6 Discussion

A gel binding assay has been developed to detect proteins which bind specifically to the *cer* site. Fragments containing *cer* were not retarded in this assay by an extract prepared from DS941 *xerC2*/pBAD. They were, however, retarded by an extract prepared from DS941 *xerC2*/pSDC105. The fact that the retarded complexes were only seen using extracts from the *xerC* over expressing strain demonstrates that this retardation is dependent on the expression of *xerC*.

The gel binding assay was then used to follow *cer* binding activity through a purification protocol. This protocol employed a negative DEAE step, an ammonium sulphate precipitation step and chromatography on a cation exchange column and gave a 75-fold purification of the binding activity over all.

Over expression of *xerC* led to the accumulation of a protein of apparent molecular weight 32,000. This protein was purified in an inactive form by G. Szatmari, and was shown by N-terminal sequencing to correspond to XerC. A protein of this size co-purified with the XerC binding activity, throughout the purification.

The purest fraction obtained contained a mixture of several proteins, one of which, making up about 15% of total protein, was the same size as XerC (as purified by G. Szatmari). It is not absolutely certain however that this protein is responsible for binding to *cer* in the gel binding assay. Nor is it absolutely certain that this protein is indeed the XerC protein. To check this, the *cer* binding protein would have to be further purified, possibly by a size fractionation step. That the purified protein did indeed correspond to XerC could then be verified by N-terminal sequencing.

Obviously much remains to be done to improve the XerC purification protocol. If XerC could be over expressed in

a soluble and active form, to the high levels seen by G. Szatmari, the purification of large amounts of XerC would be greatly aided. However, from the work of G. Szatmari, it would appear that the XerC protein is not soluble at low NaCl concentrations when it is expressed at these high levels.

When the *cer* site is split at either the PvuI or the MluI sites, only the crossover containing region is bound by XerC. XerC appears to bind to the isolated crossover site of *cer* and to the intact *cer* site with similar affinities. However the pattern of complexes produced by binding to the isolated crossover site is different from that produced by binding to the full site. It is possible that XerC can interact with regions of *cer* distal from the crossover site, but only after it has bound to the crossover site. It is also possible that XerC binds to sequences in *cer* that are disrupted by cleavage at the MluI and PvuI sites.

The nature of the complexes seen in the gel retardation experiments remains unclear. Possible explanations for some of them are: intermediates in binding (with one, two or more XerC molecules bound) as seen for Flp (Andrews *et al*, 1987) and Cre (Hoess, pers. comm), synaptic complexes containing two *cer* fragments, alternative conformations of complexes, or complexes with contaminating proteins in the XerC-containing fractions.

The protein - DNA interactions involved in the formation of XerC - *cer* complexes could be investigated using further gel binding assays on different regions of the *cer* site. Footprinting experiments would also be very useful for these studies. These could be carried out in solution on XerC - *cer* complexes. Alternatively complexes could be isolated in a non-denaturing polyacrylamide gel and cleaved in the gel before being analysed by denaturing gel electrophoresis. The sequences within *cer* to which XerC binds could be located more accurately by these experiments.

Hopefully purified XerC will soon prove useful in an *in vitro cer* site-specific recombination system. Such a system has, however, so far remained elusive.

CHAPTER SEVEN

CONCLUDING REMARKS

The main objective of this work was to isolate further *E. coli* mutants deficient in *cer* site-specific recombination, in the hope that the gene encoding the *cer* recombinase might be identified. The isolation of one such mutant led to the identification of *xerC*, expression of which is absolutely required for *cer* site-specific recombination. The XerC protein binds *in vitro* to the crossover region of *cer*, and its predicted amino acid sequence displays significant similarity to the lambda integrase family of site-specific recombinases. XerC is almost certainly the recombinase that acts at *cer*.

Members of the lambda integrase family of site-specific recombinases (eg FLP, Cre and lambda Int) cleave their recombination sites to give staggered nicks at either end of a spacer region of 6-8 bp. Given that XerC is the *cer* recombinase, and that it is a member of the lambda integrase class of recombinases, it seems probable that XerC will cleave *cer*, to produce staggered nicks in the crossover region. These nicks are predicted to produce protruding 5' OH ends and recessed 3' phosphates, covalently linked to tyrosine 275 of XerC. These nicks should be at either end of a 6-8 bp spacer region flanked by two inverted repeat XerC binding sites.

Where does XerC cleave *cer*? There is a highly conserved 11 bp sequence at the right hand end of the *cer*-like sites (see Fig 1.2 and Fig 7.1). To the left of this is a poorly conserved sequence of from 6-8 bp, which is flanked by a fairly well conserved sequence with some inverted repeat symmetry to the highly conserved 11 bp sequence (Fig 1.2 and Fig 7.1). Current thinking is that these two sequences are XerC binding sites, and that XerC probably cleaves *cer* and related sites at the boundaries of the 6-8 bp spacer. The nature of the interactions between XerC and the *cer* site could be further investigated by various footprinting techniques. The sites of strand cleavage could be identified if conditions were found in which XerC cleaves the *cer* site *in vitro*. Cleavage products could be run on a denaturing polyacrylamide gel to identify the exact phosphodiester bonds broken by XerC.

R1 contains a plasmid stabilising recombination site which displays sequence similarity to *cer* only in the crossover region (Clerget, 1984; Fig 1.2). This site acts *in vivo* as a substrate for both intermolecular and intramolecular site-specific recombination. This recombination is dependent on the presence of a functional *xerC* gene (Clerget pers. comm.). The R1 recombination site requires at most 44 bp for function (Clerget, pers. comm.) and contains no obvious ArgR binding site. It seems likely that recombination at this site will be independent of ArgR and PepA. The *E. coli dif* region, which may act to resolve chromosomal multimers, contains a 30 bp sequence which is almost identical to the R1 recombination site (Fig 7.1). Gary Blakely, in our lab, has constructed a synthetic 31 bp *dif* site, which appears to be a substrate *in vivo* for both intermolecular and intramolecular site-specific recombination. Recombination at this site will almost certainly require the presence of XerC, but not ArgR or PepA.

cer	GCGGTGCGTACAA	TTAAGGGA TTATGGTAAAT
parB	GCGGTACCGATAA	GGGATG TTATGGTAAAT
type II	GCGGTGCGTACAA	GGGATG TTATGGTAAAT
R1	TTAGTGCGCATAA	TGTATA TTATGTTACAT
dif	TTGGTGCGCATAA	TGTATA TTATGTTACAT

Fig 7.1 Alignment of the plasmid R1 and *E. coli* chromosomal *dif* recombination sites with the crossover regions of ColE1 *cer*, ClODF13 *parB*, and the type II (*cer-parB*) hybrid .

The plasmid R1 and *E. coli* chromosomal sites appear to behave in a similar way to the type II *cer-parB* hybrid. All three are substrates for both intermolecular and intramolecular recombination *in vivo*. This recombination requires only a region of about 30-50 bp which includes the two putative XerC binding sites.

An interesting question remains as to why the type II hybrid and the R1 and *dif* sites appear to have no selectivity for intramolecular reactions and probably can

all recombine in the absence of PepA and ArgR. It is possible that the wild-type *cer* site is a poor substrate for XerC, and can only recombine after a specific synaptic complex has been formed. This synaptic complex would be dependent on the upstream regions of *cer* and also on the ArgR and PepA proteins, and could act as a "topological filter", favouring intramolecular recombination over intermolecular recombination. The type II hybrid, *dif* and the R1 recombination sites could be better substrates for XerC mediated recombination and hence have no requirement for PepA and ArgR. This would account for their lack of topological selectivity.

If the wild-type *cer* site specific recombination reaction does proceed *via* a specific synaptic complex, one would predict that the reaction products would have a specific topology (compare with the resolvases and the DNA invertases). The topological features of the *cer* site-specific recombination reaction, and the involvement of PepA and ArgR in a "topological filter" could best be studied if an *in vitro* recombination system was developed. Such a system has so far remained elusive. We have looked for ArgR-independent, PepA-independent recombination *in vitro*, using XerC-containing crude extracts and DNA substrates with two directly repeated copies of the type II *cer-parB* hybrid site. We have also looked for recombination *in vitro* using purified ArgR and PepA along with XerC-containing extracts acting on DNA substrates with directly repeated wild-type *cer* sites. However, no recombination has so far been detected. This failure to detect recombination *in vitro* could reflect a requirement for further proteins accessory factors or the need for different reaction conditions.

The *dif* site is more symmetrical than the type II *cer-parB* hybrid. This might make recombination at the the *dif* site more efficient than recombination at the type II hybrid. If recombination at *dif* is independent of ArgR and PepA *in vivo*, it might be worthwhile to assay for *dif* site-specific recombination *in vitro* using semi purified XerC or crude extracts containing XerC.

If this fails we could look for resolution of

artificial Holliday junction intermediates by XerC. Resolution of *att* site Holliday junctions requires only the integrase protein and the core type Int binding sites (Hsu and Landy, 1984), so it is predicted that XerC would be the only protein required for Holliday junction resolution.

Hopefully the development of an *in vitro cer* site-specific recombination system will enhance our understanding of the mechanism of *cer* site-specific recombination. The possible role of XerC in chromosome multimer resolution is intriguing, and also deserves further study.

BIBLIOGRAPHY

Abdel-Meguid S.S., Grindley N.D.F., Templeton N.S., and Steitz T.A. 1984. Cleavage of the site-specific recombination protein gamma-delta resolvase: the smaller of the two fragments binds DNA specifically. Proc. Natl. Acad. Sci. USA 81: 2001-2005.

Abremski K., Hoess R., and Sternberg N. 1983. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. Cell 32: 1301-1311.

Abremski K., and Hoess R. 1984. Bacteriophage P1 site-specific recombination: purification and properties of the Cre recombinase protein. J. Biol. Chem. 259: 1509-1514

Aiba H., Mori K., Tanaka M., Ooi T., Roy A., and Danchin A. 1984. The complete nucleotide sequence of the adenylate cyclase gene of *Escherichia coli*. Nucleic Acids Res. 12: 9427-9440.

Aldea M., Maples V., and Kushner S. 1988. Generation of a detailed physical and genetic map of the *ilv-metE-udp* region of the *Escherichia coli* chromosome. J. Mol. Biol. 200: 427-438.

Allen J.G., Atherton F.R., Hall M.J., Hassall C.H., Holmes S.W., Lambert R.W., Nisbet L.J., and Ringrose P.S. 1978. Phosphonopeptides, a new class of synthetic antibacterial agents. Nature 272: 56-58

Andrews B.J., Proteau G.A., Beatty L.J., and Sadowski P.D. 1985. The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequence. Cell 40: 795-803

Andrews B.J., Beatty L.J., and Sadowski P.D. 1987. Isolation of intermediates in the binding of the FLP recombinase of the yeast plasmid 2 micron circle to its target sequence. J. Mol. Biol. 193: 345-358

Argos P., Landy A., Abremski K., Egan J.B., Haggard-Ljungquist E., Hoess R.H., Kahn M.L., Kalionis B., Narayana S.V.L., Pierson L.S., Sternberg N., and Leong J.M. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. 5: 433-440.

Austin S., Ziese M., and Sternberg N. 1981. A novel role for site-specific recombination in maintenance of bacterial replicons. Cell 25: 729-736

Austin S., and Abeles A. 1983. Partition of unit-copy miniplamids to daughter cells: II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. J. Mol. Biol. 169: 373-387.

Austin S., and Nordstrom K. 1990. Partition mediated incompatibility of bacterial plasmids. Cell. 60: 351-354

Bachmann B.J., 1972. Pedigrees of some mutant strains of Escherichia coli K12. Bacteriol. Reviews 36: 525-557

Bachmann B.J. 1990. Linkage map of *Escherichia coli* K12, edition 8. Microbiological reviews 54: 130-197

Bassett C.L., and Kushner S.R. 1984. Exonucleases I, III and V are required for stability of ColE1 related plasmids in *Escherichia coli*. J. Bacteriol. 157: 661-664

Beatty L.G., Babineau-Clary D., Hogrefe C., and Sadowski P.D. 1986. FLP site-specific recombinase of yeast 2-micron plasmid. Topological features of the reaction. J.Mol.Biol. 188: 529-544

Bednarz A.L. 1989 Dissection of the Tn3 resolution site. PhD thesis: University of Glasgow

Berg D.E. 1977. Insertion and excision of the transposable kanamycin resistance determinant Tn5. In: DNA inserion elements. Eds. Bukhari A.F., Shapiro J.A. and Adhya S.L. pp205-212

Berg D.E., Weiss A., and Crossland L. 1980. The polarity of Tn5 insertion mutations in *Escherichia coli*. J. Bacteriol. 142: 439-446

Berg D.E., and Berg C.M. 1983. The transposable element Tn5. Biotechnology 1: 417-435

Birnboim H.C., and Doly J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7: 1513-1523.

Bolivar F. 1978. Construction and characterisation of new cloning vehicles. III. Derivatives of pBR322 carrying unique EcoRI sites for selection of EcoRI generated recombinant DNA molecules. Gene 4: 121-136

Boocock M.R., Brown J.L., and Sherratt D.J. 1987. Topological specificity in Tn3 resolvase catalysis. UCLA Symposia on Molecular and Cellular Biology 47: 703-718.

Brewer B.J. 1988. When polymerases collide: replication and the transcriptional organisation of the *E. coli* chromosome. Cell 53: 679-686

Broda P. 1979. Plasmids. W.H. Freeman. Oxford and San Fransisco

Bruist M.F., Glasgow A.C., Johnson R.C., and Simon M.I. 1987. Fis binding to the recombinational enhancer of the Hin DNA inversion system. Genes and Development 1: 762-772

Burley S.K., David P.R., Taylor A., and Lipscomb W.N. 1982. Molecular structure of leucine aminopeptidase at 2.7 angstrom resolution. Proc. Natl. Acad. Sci. USA 87: 6878-6882

Bushman W., Yin S., Thio L.L., and Landy A. 1984. Determinants of directionality in lambda site-specific recombination. Cell 39: 699-706.

Chan P.T., Ohmori H., Tomizawa J.I. and Lebowitz J. 1985. Nucleotide sequence and organisation of ColE1 DNA. J. Biol. Chem. 260: 8925-8935

Chang A.C.Y., and Cohen S.N. 1978. Construction and characterisation of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J. Bacteriol. 134: 1141-1156

Chattoraj D., Cordes K., and Abeles A. 1984. Plasmid P1 replication: negative control by repeated DNA sequences. Proc. Natl. Acad. Sci. USA 81 6456-6460

Clerget M. 1984. A 140 bp DNA segment from the kanamycin resistance region of plamid R1 acts as an origin of replication and promotes site-specific recombination. J. Mol. Biol. 178: 35-46

Collins J.F., Coulson A.F.W., and Lyall A. 1988. The significance of protein sequence similarities. Comput. Appl. Biosci. 4: 67-71.

Colloms S.D., Sykora P., Szatmari G., and Sherratt D.J. 1989. Recombination at ColE1 *cer* requires the *Escherichia coli xerC* gene product, a member of the lambda integrase family of site-specific recombinases. J. Bacteriol. In press

Cox M., 1989. DNA inversion in the 2-micron plasmid of *Sacharomyces cerevisiae*. In: Mobile DNA. Ed. Berg D., and Howe M. pp661-670

Craig N.L., and Nash H.A. 1983. The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell 35: 795-803.

Cuypers H.T., van Loon-Klaasen L., Egberts W.V., de Jong W., and Bloemendal H. 1982. The primary structure of leucine amino peptidase from bovine eye lens. J. Biol. Chem. 257: 7077-7085

Davidson N., and Szybalski W. 1971. In: The bacteriophage lambda. ed. Hershey A.D. pp113-138

Davis M.A., and Austin S.J. 1988. Recognition of the P1 plamid centromere analog involves binding of the ParB protein and is modified by a specific host factor. EMBO J. 7: 1881-1888.

de Bruijn F.J., and Lupski J.R. 1984. The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids - a review. Gene 27: 131-149.

Dodd H.M., and Bennett P.M. 1986. Location of site-specific recombination system of R46: a function necessary for plasmid maintenance. J.Gen.Microbiol. 132: 1009-1020.

Dodd H.M., and Bennett P.M. 1987. The R46 site-specific recombination system is a homologue of the Tn3 and gamma-delta (Tn1000) cointegrate system. J.Gen. Microbiol. 133: 2031-2039.

Dorman C.J. and Higgins C.F. 1987. Fimbrial phase variation in Escherichia coli: dependence on integration host factor and homologies with other site-specific recombinases. J. Bacteriol. 169: 3840-3843.

Drew H., and Travers A. 1985. DNA bending and its relation to nucleosome positioning. J. Mol. Biol. 186: 773-790

Durkacz B., and Sherratt D.J. 1973. Segregation of colicinogenic factor ColE1 from a bacterial population temperature sensitive for DNA polymerase I. Mol. Gen. Genet. 121: 71-75.

Easton A.M., and Kushner S.R. 1983. Transcription of the *uvrD* gene of Esherichia coli is controlled by the *lexA* repressor and by attenuation. Nucleic Acids Res. 11: 8625-8640.

Finch P.W., and Emerson P.T. 1983. Nucleotide sequence of the regulatory region of the *uvrD* gene of *Escherichia coli*. Gene 25: 317-323.

Finch P.W., and Emerson P.T. 1984. The nucleotide sequence of the *uvrD* gene of *Escherichia coli*. Nucleic Acids Res. 12: 5789-5799.

Fried M., and Crothers D.M. 1981. Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucl. Acids Res. 9: 6505-6525

Garner M.M., and Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucl. Acids Res. 9: 3047-3060

Gerdes K., Rasmussen P.B. and Molin S. 1985. Unique type of plasmid maintenance function: Post segregational killing of plasmid free cells. Proc. Natl. Acad. Sci. USA 83: 3116-3120

Gibson M., Price M., and Higgins C.F. 1984. Genetic characterisation and molecular cloning of the tripeptide permease (*tpp*) genes of *Salmonella typhimurium*. J. Bacteriol. 160: 122-130

Gilson E., Clement J.M., Perrin D., and Hofnung M. 1987. Palindromic units: a case of highly repetitive DNA sequences in bacteria. TIG 3: 226-230

Gilson E., Perrin D., and Hofnung M. 1990. DNA polymerase I and a protein complex bind specifically to *Escherichia coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organisation. Nucleic Acids Res. 18: 3941-3952

Glansdorff N. 1987. Biosynthesis of arginine and polyamines. In: *Salmonella typhimurium* and *Escherichia coli*: cellular and molecular biology. Ed. Neidhardt F.C. pp 321-344

Glasgow A.C., Kelly H.T., and Simon M. 1989. Bacterial DNA inversion sytems. In: Mobile DNA. Ed. Berg D., and Howe M. pp637-660

Greene P.J., Gupta M., Boyer H.W., Brown W.E., and Rosenberg J.M. 1981. Sequence analysis of the DNA encoding EcoRI endonuclease and methylase. J. Biol. Chem. 256: 2143-2153

Gribskov M., Devereux J., and Burgess R.R. 1984. The codon-preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. 12: 539-549

Grindley N.D.F., Lauth M.R., Wells R.G., Wityk R.J., Salvo J.J., and Reed R.R. 1982. Transposon-mediated site-specific recombination: identification of three binding sites for resolvase at the *res* sites of gamma-delta and Tn3. Cell 30: 19-27.

Gronostajski R.M., and Sadowski P.D. 1985a. The FLP recombinase of the Saccharomyces cerevisiae 2 micron plasmid attaches covalently to DNA via a phosphotyrosyl linkage. Mol.Cell.Biol. 5: 3274-3279.

Gronostajski R.M., and Sadowski P.D. 1985b. Determination of DNA sequences essential for FLP mediated recombination by a novel method. J. Biol. Chem. 260: 12320-12327.

Gronostajski R.M., and Sadowski P.D. 1985c. The FLP recombinase of the 2 micron circle of yeast. Inter- and intramolecular reactions. J. Biol. Chem. 260:12328-12335.

Hager D., and Burgess R.R. 1980. Elution of proteins from SDS-polyacrylamide gels and renaturation of enzymatic activitiy: results with the sigma subunit of *Escherichia coli* RNA polymerase, wheat germ topoisomerase and other enzymes. Anal. Biochem. 109: 70-80.

Hakkaart M.J., Van Den Elzen P., Veltkamp E., and Nijkamp H.J.J. 1984. Maintenance of multicopy plasmid CloDF13 in *Escherichia coli* cells: evidence for site-specific recombination at *parB*. Cell 36: 203-209

Hall R.M., and Vockler C. 1987. The region of the IncN plasmid R46 coding for resistance to beta-lactam antibiotics, streptomycin / spectinomycin and sulphonamides is closely related to antibiotic resistant segments found in IncW plasmids and Tn21-like transposons. Nucleic Acids Res. 15: 7491-7501.

Haffter P., and Bickle T. 1988. Enhancer-independent mutants of the *cin* recombinase have a relaxed topological specificity. EMBO J. 7: 3991-3996.

Harr R., Haggstrom M., and Gustafsson P. 1983. Search algorithm for pattern match analysis of nucleic acid sequences. Nucleic Acids Res. 11: 2943-2957

Hartlein M., Frank R., and Madern D. 1987. Nucleotide sequence of the *Escherichia coli* valyl-tRNA synthetase gene, *valS*. Nucleic Acids Res. 15: 9081-9082

Hatfull G.F., and Grindley N.D.F. 1986. Analysis of gamma-delta resolvase mutants *in vitro*: evidence for an interaction between serine 10 of resolvase and site 1 of *res*. Proc. Natl. Acad. Sci. USA 83:5429-5433.

Hatfull G.F., and Grindley, N.D.F. 1988. Resolvases and DNA invertases: a family of enzymes active in site-specific recombination. In: Genetic Recombination. Ed. Kucherlapati R., and Smith G.R. pp 357-396

Heck J.D., and Hatfield G.W. 1988a. Valyly-tRNA synthetase gene of *Escherichia coli* K12: molecular genetic characterisation. J. Biol. Chem. 263: 857-867

Heck J.D., and Hatfield G.W. 1988b. Valyly-tRNA synthetase gene of *Escherichia coli* K12: primary structure and homology within a family of aminoacyl-tRNA synthetases. J. Biol. Chem. 263: 868-877

Henikoff S. 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods in enzymology 155: 156-165.

Higgins C.F., McLaren R.S., and Newbury S. 1988. Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? - a review. Gene 72: 3-14

Hiraga S., Niki H., Ogura T., Ichinose C., Mori H., Ezaki B., and Jaffe A. 1989. Chromosome partitioning in *Escherichia coli*: novel mutants producing anucleate cells. J. Bacteriol 171: 1496-1505

Hoess R.H. and Abremski K. 1985. Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. J. Mol. Biol. 181: 351-362.

Hoess R.H., Wierzbicki A., and Abremski K. 1986. The role of the *loxP* spacer region in P1 site-specific recombination. Nucl. Acids Res. 14: 2287-2300.

Hoess R.H., Wierzbicki A., and Abremski K. 1987. Isolation and characterisation of intermediates in site-specific recombination. Proc. Natl. Acad. Sci. USA 84: 6840-6844.

Holmes D.S., and Quigley M. 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal.Biochem. 114: 193-197.

Hsu R.L., and Landy A. 1984. Resolution of synthetic *att*-site Holliday structures by the integrase protein of bacteriophage lambda. Nature 311: 721-726.

Ikemura T. 1982. Correlation between the abundance of *Escherichia coli* tRNAs and the occurence of the respective codons in its protein genes: A proposal for a synonomous codon choice that is optimal for the *Escherichia coli* translation system. J. Mol. Biol. 151: 389-409

Jayaram M., Crain K.L., Parsons R.L., and Harshey R.M. 1988. Holliday junctions in FLP recombination: resolution by step-arrest mutants of FLP protein. Proc. Natl. Acad. Sci. USA 85: 7902-7906.

Johnson R.C., and Bruist M.F. 1989. Intermediates in Hin mediated DNA inversion: a role for Fis and the recombination enhancer in the strand exchange reaction. EMBO J. 8: 1581-1590

Kahmann R., Rudt F., Koch C, and Mertens G. 1985. G inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. Cell 41: 771-780.

Kahmann R., Mertens G., Klippel A., Brauer B., Rudt F., and Koch C. 1987. The mechanism of G inversion. UCLA Symposia on Molecular and Cellular Biology 47: 681-690.

Kanaar R., van de Putte P., and Cozzarelli N.R. 1988. Gin-mediated DNA inversion: product structure and the mechanism of strand exchange. Proc. Natl. Acad. Sci. USA 85: 752-756.

Kanaar R., van de Putte, P. and Cozzarelli N.R. 1989. Gin mediated recombination of catenated and knotted DNA substrates: implications for the mechanism of interaction between *cis*-acting sites. Cell 58: 147-159

Kelln R.A., and Zak V.L. 1978. Arginine regulon control in a *Salmonella typhimurium-Escherichia coli* hybrid merodiploid. Mol. Gen. Genet. 161: 333-336

Kitts P.A., Symington L.S., Dyson P., and Sherratt D.J. 1983. Transposon-encoded site-specific recombination: nature of the Tn3 DNA sequences which constitute the recombination site *res*.EMBO J. 2: 1055-1060.

Kitts P.A., and Nash H.A. 1988. Bacteriophage lamda site-specific recombination proceeds with a defined order of strand exchanges. J.Mol.Biol. 204: 95-107

Klemm P. 1986. Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type I fimbriae in *Escherichia coli*. EMBO J. 5: 1389-1393.

Klippel A., Mertens G., Patschinsky T., and Kahmann R. 1988a. The DNA invertase Gin of phage Mu: formation of a covalent complex with DNA via a phosphoserine at amino acid position 9. EMBO J. 7: 1229-1237.

Klippel A., Cloppenborg K., and Kahmann R. 1988b. Isolation and characterization of unusual *gin* mutants. EMBO J. 7: 3983-3989.

Koch C., and Kahmann R. 1986. Purification and properties of the *Esherichia coli* host factor required for inversion of the G segment in bacteriophage Mu. J.Biol.Chem. 261: 15673-15678.

Koch C., Vandekerckhove J., and Kahmann R. 1988. *Escherichia coli* host factor for site-specific DNA inversion: cloning and characterization of the *fis* gene. Proc. Natl. Acad. Sci. USA 85: 4237-4241.

Kohara Y., Akiyama K., and Isono K. 1987. The physical map of the whole *Escherichia coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell 50: 495-508.

Kolot M.N. 1990. Par site of the ColN plasmid: Structural and functional organisation. Mol.Gen.Genet. 222: 77-80

Laemmli U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Landy A. 1989. Dynamic, structural and regulatory aspects of lambda site-specific recombination. Annu. Rev. Biochem. 58: 913-949

Lane D., de Feyter R., Kennedy M., Phua S-H., and Semon D. 1986. D protein of mini F plasmid acts as a repressor of transcription and as a site-specific resolvase. Nucl. Acids Res. 14: 9713-9728.

Lim D., Oppenheim J.D., Eckhardt T., and Maas W.K. 1987. Nucleotide sequence of the *argR* gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor. Proc. Natl. Acad. Sci. USA 84: 6697-6701

Mahillon J., and Lereclus D. 1988. Structural and functional analysis of Tn4430: identification of an integrase-like protein involved in the co-integrate resolution process. EMBO J. 7: 1515-1526.

Mazodier P., Cossart P., Giraud E., and Gasser F. 1985. Completion of the nucleotide sequence of the central region of Tn5 confirms the presence of three resistance genes. Nucleic Acids Res. 13: 195-205

Mertens G., Klippel A., Fuss H., Blocker H., Frank R., and Kahmann R. 1988. Site-specific recombination in bacteriophage Mu: characterization of binding sites for the DNA invertase Gin. EMBO J. 7: 1219-1227.

Miller C.G., and MacKinnon K. 1974. Peptidase mutants of Salmonella typhimurium. J. Bacteriol. 120: 355-363

Miller C.G. 1975. Genetic mapping of *Salmonella typhimurium* peptidase mutations. J. Bacteriol. 122: 171-176

Miller C.G., and Schwartz G. 1978. Peptidase-deficient mutants of Escherichia coli. J. Bacteriol 135: 603-611.

Miller C.G. 1987. Protein degradation and proteolytic modification. In: *Salmonella typhimurium* and *Escherichia coli*: cellular and molecular biology. Ed. Neidhardt F.C. pp 680-691

Mizuuchi M., and Mizuuchi K. 1979. Integrative recombination of bacteriophage lambda: *in vitro* study of the intermolecular reaction. Cold Spring Harbor Symp. Quant. Biol. 43: 1111-1114

Moitoso de Vargas L., Pargellis C.A., Hasan N.M., Bushman E.W., and Landy A. 1988. Autonomous DNA binding domains of lambda integrase recognizs two different sequence families. Cell 54: 923-929.

Molin S., Stougaard P., Light J., and Nordstrom K. 1981. Isolation and characterisation of new copy number mutants of plasmid R1 and identification of a polypeptide involved in copy number control. Mol. Gen. Genet. 181: 123-130

Morrell E. 1990. Transposition of Tn7. PhD thesis university of Glasgow.

Mori H., Mori Y., Inchonose C., Niki H., Ogura T., Kato A., and Hiraga S. 1989. Purification and characterisation of SopA and SopB: proteins essential for F plasmid partitioning. J. Biol. Chem. 264: 15535-15541

Morlon J., Chartier M., Bidaud M., and Lazdunski C. 1988. The complete nucleotide sequence of the colicinogenic plasmid ColA. High extent of homology with ColE1. Mol. Gen. Genet. 211: 231-243

Murphy E., Huwyler L., and do Carmo de Freire Bastos, M. 1985. Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. EMBO J. 4: 3357-3365.

Nash, H. A. 1975. Integrative recombination of bacteriophage lambda DNA *in vitro*. Proc. Natl. Acad. Sci. USA 72: 1072-1076.

Newbury S.F., Smith N.H., Robinson E.C., Hiles I.D., and Higgins C.F. 1987. Stabilisation of translationally active mRNA by prokaryotic REP sequences. Cell 48: 297-310

Nordstrom K., Molin S., and Aagaard-Hansen H. 1980. Partitioning of plasmid R1 in *Escherichia coli*: kinetics of loss of plasmid derivatives deleted of the *par* region. Plasmid 4: 215-227

Nordstrom K., and Austin S.J. 1989. Mechanisms that contribute to the stable segregation of plasmids. Annu. Rev. Genet. 23: 37-69

Nunes-Duby S.E., Matsumoto L., and Landy A. 1987. Site-specific recombination intermediates trapped with suicide substrates. Cell 50: 779-788.

O'Connor M.B., Kilbane J.J., and Malamy M.H. 1986. Site-specific recombination in the oriVI region of the F factor: DNA sequences involved in recombination and resolution. J. Mol. Biol. 189: 85-102.

Ogura T., and Hiraga S. 1983a. Partition mechanisms of F plasmid: two plasmid gene encoded products and a *cis* acting region are involved in partition. Cell 32: 351-360

Ogura T., and Hiraga S. 1983b. Mini-F plasmid genes that couple host-cell division to plasmid proliferation. Proc. Natl. Acad. Sci. USA 80: 4784-4788

Pargellis C.A., Nunes-Duby S.E., Moitoso de Vargas L., Landy A. 1988. Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. J.Biol.Chem. 263: 7678-7685.

Phillips S.E.V., Manfield I., Parsons I., Davidson B.E., Rafferty J.B., Somers W.S., Margarita D., Cohen G.N., Saint-Giron I., and Stockley P.C. 1989. Cooperative tandem binding of *met* repressor of *Escherichia coli*. Nature 341: 711-715

Pollock T.J., and Nash H.A. 1983. Knotting of DNA caused by a genetic rearrangement. Evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda. J.Mol.Biol. 170: 1-18.

Prasad P.V., Young L.J., and Jayaram M. 1987. Mutations in the 2 micron circle site-specific recombinase that abolish recombination without affecting substrate recognition. Proc. Natl. Acad. Sci. USA 84: 2189-2193

Reed R.R. 1981. Resolution of cointegrates between transposons gamma-delta and Tn3 defines the recombination site. Proc. Natl. Acad. Sci. USA 78: 3428-3432.

Reed R.R., and Grindley N.D.J. 1981. Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. Cell 25: 721-728.

Reed R.R., and Moser C.D. 1984. Resolvase-mediated recombination intermediates contain a serine residue covalently linked to DNA. Cold Spring Harbor Symp. Quant. Biol. 49: 245-249.

Richaud C., Higgins W., Mengin-Lecreulx D., and Stragier P. 1987. Molecular cloning, characterisation and chromosomal localisation of *dapF* , the *E. coli* gene for diaminopimelate epimerase. J. Bacteriol. 169: 1454-1459.

Richaud C., and Printz, C. 1988. Nucleotide sequence of the *dapF* gene and flanking regions from *Escherichia coli* K12. Nucleic Acids Res. 16: 10367-10367.

Richet E., Abcarian P., and Nash H.A. 1986. The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda recombination. Cell 46: 1011-1021.

Richet E., Abcarian P., and Nash H.A. 1988. Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. Cell 52: 9-17

Sadowski P. D. 1986. Site-specific recombinases: changing partners and doing the twist. J.Bact. 165: 341-347.

Sanderson K.E. and Hurley S.A. 1987. Linkage map of *Salmonella typhimurium*. In: *Salmonella typhimurium* and *Escherichia coli*: cellular and molecular biology. Ed. Neidhardt F.C. pp 877-918

Senecoff J.F., and Cox, M.M. 1986. Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. J.Biol.Chem. 261: 7380-7386.

Shaw K.J., and Berg C.M. 1979. *Escherichia coli* K12 auxotrophs induced by insertion of the transposable element Tn5. Genetics 92: 741-747

Sherratt D.J. 1989. Tn3 and related transposable elements. In: Mobile DNA. Ed. Berg D., and Howe M. pp163-184.

Skogman S.G. and Nilsson J. 1984. Molecular cloning and characterisation of the gene for *Escherichia coli* valyl-tRNA synthetase. Gene 30: 219-226

Smith T.F., and Waterman M.S. 1981. Identification of common molecular subsequences. J. Mol. Bio 147: 195-197.

Stark W.M., Sherratt D.J., and Boocock M.R. 1989a. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. Cell 58: 779-790

Stark W.M., Boocock M.R., and Sherratt D.J. 1989b. Site-specific recombination by Tn3 resolvase. TIG 5: 304-309

Stewart G. 1986. ColE1 *cer* site-specific recombination. PhD thesis: University of Glasgow.

Stirling C.J. 1987. Cloning and characterisation of *Escherichia coli xer* genes. PhD thesis: University of Glasgow.

Stirling C.J., Stewart G., and Sherratt, D.J. 1988a. Multicopy plasmid stability in *Escherichia coli* requires host-encoded functions that lead to plasmid site-specific recombination. Mol. Gen. Genet. 214: 80-84.

Stirling C.J., Szatmari G., Stewart G., Smith M.C.M., and Sherratt D.J. 1988b. The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus. EMBO J. 7: 4389-4395.

Stirling C.J., Colloms S.D., Collins J.F., Szatmari G., and Sherratt D.J. 1989. *xerB*, an *Escherichia coli* gene required for plasmid ColE1 site-specific recombination, is identical to pepA, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. EMBO J. 8: 1623-1627.

Stougaard P., Molin S., and Nordstrom K. 1981. RNAs involved in copy number control and incompatibility of plasmid R1. Proc. Natl. Acad. Sci. USA 78: 6008-6012.

Strauss F., and Varshavsky A. 1984. A protein binds to a sattelite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell 37: 889-901

Summers D.K., and Sherratt, D.J. 1984. Multimerisation of high copy number plasmids causes instability: ColE1 encodes a determinant for plasmid monomerisation and stability. Cell 36: 1097-1103.

Summers D.K., Yaish S., Archer J., and Sherratt D.J. 1985. Multimer resolution systems of ColE1 and ColK: localisation of the crossover site. Mol. Gen. Genet. 201: 334-338

Summers D.K., and Sherratt, D.J. 1988. Resolution of ColE1 dimers requires DNA sequences implicated in the three dimensional organisation of the *cer* site. EMBO J. 7: 851-858.

Summers D.K. 1989. Derivatives of ColE1 *cer* show altered topological specificity in site-specific recombination. EMBO J. 8: 309-315.

Sutcliffe J. G. 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. 43: 77-90.

Thompson J.F., Moitoso de Vargas L., Koch C., Kahmann R., and Landy A. 1987. Cellular factors couple recombination with growth phase: chacterization of a new component in the lambda site-specific recombination pathway. Cell 50: 901-908.

Thompson J.F., and Landy A. 1988. Empirical estimation of protein induced bending angles: application to lambda site-specific recombination complexes. Nucleic Acids Res. 16: 9687-9705

Thompson J.F., and Landy A. 1989. Regulation of bacteriophage lambda site-specific recombination. In: Mobile DNA. Ed. Berg D., and Howe M. pp1-22

Timmis K.N. 1981. Gene manipulation *in vitro*. In: 31st Symp. Soc. Gen. Microbiol. Ed. Glover S.W., and Hopwood D.A. pp49-109

Tolun A., and Helinski D.R. 1981. Direct repeats of the F plasmid *incC* region express F incompatibility. Cell 24: 687-694

Tomizawa J., Itoh T., Selezer G., and Som T. 1981. Inhibition of primer formation by a plasmid specified small RNA. Proc. Natl. Acad. Sci. USA 78: 1421-1425

Tomizawa J., and Som T. 1984. Control of ColE1 plasmid replication: enhancement of binding of RNAI to the primer transcript by the Rom protein. Cell 38: 871-878

Tso J.Y., Hemodson M.A., and Zalkin H. 1982. Glutamine phosphoribosylpyrophosphate amidotransferase from cloned *Escherichia coli purF*. J. Biol. Chem. 257: 3532-3536

Vetter D., Andrews B.J., Roberts-Beatty L., and Sadowski P.D. 1983. Site specific recombination of yeast 2 micron DNA *in vitro*. Proc. Natl. Acad. Sci. USA 80: 7284-7288

Vieira, J. and Messing J. 1982. The pUC plasmids, an M13mp7 derived system for insertion mutagenesis with synthetic universal primers. Gene 19: 259-268

Vieira, J. and Messing J. 1987. Production of single-stranded plasmid DNA. Methods in enzymology 153: 3-11.

Vogt V.M. 1970. Purification and properties of an aminopeptidase from *Escherichia coli*. J. Biol. Chem. 245: 4760-4769

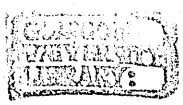
Volkert F.C., and Broach J.R. 1986. Site-specific recombination promotes plasmid amplification in yeast. Cell 46: 541-550.

Wasserman S.A.. and Cozzarelli N.R. 1985. Determination of the stereostructure of the product of Tn3 resolvase by a general method. Proc. Natl. Acad. Sci. USA 82: 1079-1083.

Wierzbicki A., Kendall M., Abremski K., and Hoess R. 1987. A mutational analysis of the bacteriophage P1 recombinase Cre. J. Mol. Biol. 195: 785-794

Yang Y. and Ferro-Luzzi Ames G. 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. Proc. Natl. Acad. Sci. USA 85: 8850-8854

Yanisch-Perron, C., Vieira, J., and Messing, J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.



THE END