Analysis of *xer* site-specific recombination *in vivo*

A thesis submitted for

the degree of

Doctor of Philosophy

at the University of Glasgow

by

Richard McCulloch

Institute of Genetics

University of Glasgow

Church Street

Glasgow

December 1992

ProQuest Number: 11007935

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 11007935

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

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

Thesis
9400
copy 1

GLASGOW
UNIVERSITY
LIBRARY

Dedicated to my mum and dad.

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.

CONTENTS

ABBREVIATIONS

Chemicals

ATP	adenosine triphosphate
CTP	cytidine triphosphate
DNA	2' deoxyribonucleic acid
dNTP	2' deoxy (nucleotide)
EDTA	ethylene diamine tetra-acetic acid (disodium salt)
IPTG	isopropyl B-D-thiogalactoside
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
Tris	tris (hydroxymethyl) amino ethane

Antibiotics

Ap	ampicillin
Cm	chloramphenicol
Km	kanamycin
Tet	tetracycline

Measurements

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

Miscellaneous

UV	ultra violet light
DNAase	deoxyribonuclease

ACKNOWLEDGMENTS

Firstly I thank Dave for all his help and encouragement during this thesis, and Mary for her patient advice and for showing where and what everything is in the lab. Thanks also to the "media ladies" for glassware, solutions and cakes and to Margaret, Linda and Diane.

Thanks to S.West and L.Coggins for the experiments they performed and helped me perform. Also, thanks to S.Lovett, R.Lloyd and G.Sharples for their kind gifts of strains and plasmids.

Special thanks go to Marshall for both reading this thesis and being kind enough to require only a few drinks in return for this horrendous experience. Thanks to everyone else in my lab for making the last three years enjoyable in work and, more importantly, outside of work - Dave, Jen, Gerhard, Stephen (thanks for 5 years in this case, and plenty of tequila during it), Garry, Nick, Martin, Sally, Shahnaz, Angela and the old *cer* workers (Sean and George). Important thanks go to the following groups of people: the football team, the multisport team (not unsurprisingly for the drinking rather than the sport) and all other friends I have spent productive hours in Tennants with. Thanks to C.Spence for the kind use of his computer (games), and Gillian and Andy for supplying me with t-shirts and loans.

The most important thank-you goes to Heather for all the times we have spent together, for magnificent "gourmet" Mexican rice and extra-curricular mambo.

SUMMARY

The *xer* site-specific recombination system is employed in two related biological processes. The first of these roles is in resolving multimers of high copy number, randomly partitioned plasmids in order to ensure their heritable stability. Plasmid multimers arise through homologous recombination. Their resolution to monomers by *xer* site-specific recombination requires a recombination locus named *cer* that was originally described in the naturally occurring plasmid ColE1, although similar sites have been isolated in many other plasmids (Summers and Sherratt, 1984; Summers *et al*, 1985). The second role of the *xer* system is in bacterial chromosome partition. A locus has been described, named *dif*, in the terminus region of the *Escherichia coli* chromosome that is a substrate for *xer* catalysed site-specific recombination (Blakely *et al*, 1991; Kuempel *et al*, 1991). It is believed *xer* recombination activity is required at *dif* because dimers of replicating chromosomes are produced by homologous recombination; these dimers cannot be partitioned into daughter cells as the host cell divides, and hence *xer* recombination converts the chromosomal dimers into monomers and allows cell division.

The proteins that act in *xer* site-specific recombination are encoded by the *E.coli* chromosome. At the start of this work three genes (*xer* genes) had been shown to be absolutely required for *cer* recombination, and one for *dif* recombination. *xerC* has been cloned and sequenced and is required for both *cer* and *dif* recombination (Colloms *et al*, 1990; Blakely *et al*, 1991). The protein encoded by *xerC* binds to both recombination sites *in vitro* and has sequence homologies to the λ integrase family of site-specific recombinases. *xerA* and *xerB* encode the proteins ArgR and aminopeptidase A (PepA) respectively; both are essential for *cer* recombination and are believed to be "accessory factors" in the reaction (Stirling *et al*, 1988 and 1989). While this work was being performed a fourth *xer* gene was identified, called *xprB*. Genetic evidence is presented that

the XprB is essential for both *cer* and *dif* recombination and that it shows amino acid sequence homologies to the λ integrase family of recombinases.

PepA is an amino-terminal exopeptidase whose role in the *cer* recombination reaction was not understood at the start of this work. Site-directed mutagenesis of the gene encoding PepA is described, and evidence presented that the mutant enzyme (named E354A) encoded by the altered *pepA* gene lacks any aminopeptidase activity both *in vitro* and *in vivo* but is still able to support *cer* recombination. The relevance of this observation to how PepA is employed by the *xer* site-specific recombination reaction is discussed.

The construction of an controllable *in vivo* recombination system and its use in analysing the mechanism of the *xer* recombination reaction is described. The system comprises a derivative of *E.coli* K12, named RM40, in which the expression of *xerC* is controlled by the *lac* promoter and operator sequences rather than its natural promoter sequences. Analysing the products derived from *cer*-mediated recombination of reporter plasmids demonstrated the existence of Holliday junction structures produced during the reaction. These putative recombination intermediates were shown to have arisen by the exchange of a specific pair of strands within the *cer* sites of the substrate DNA molecule used. These results suggest that *cer* site-specific recombination involves the same form of strand exchange mechanism that has been described *in vitro* for other members of the λ integrase family (Int, FLP and Cre; Stark *et al*, 1992). The Holliday junctions were isolated in anomalously large quantities when compared to the isolation of such reaction intermediates in other, related systems. The possible relevance of this to the *xer* recombination mechanism was analysed by comparing *cer* and *dif* recombination in RM40, and by altering the *cer* substrate molecules and reaction conditions employed.

Chapter 1

Introduction

1.1 General introduction

Site-specific recombination is a controlled process found in both eukaryotes and prokaryotes which generates precise rearrangements of DNA molecules at defined positions. The rearrangements are catalysed by a variety of proteins which are collectively termed "recombinases", and are achieved by cutting the DNA molecules at specific points and joining the ends to new DNA partners. Site-specific recombination is distinguished from homologous recombination in a number of ways. In homologous recombination the recombining DNA segments recognise their partners by comparing the DNA sequences through Watson-Crick base-pairing between the partner duplexes (a process which is mediated by RecA-like proteins), and therefore extensive homology is required between the two DNA molecules. In contrast, there is no requirement for extensive homology between DNA molecules during site-specific recombination because the recombination sites are brought together primarily through interactions between DNA binding proteins. Moreover, no base-pairing between the two recombining DNA duplexes is required before strand cleavage in site-specific recombination. The two processes are also distinguished by the fact that the proteins involved in the reactions are unrelated, and site-specific recombination is probably a more regulated process.

Site-specific recombination is related to another type of non-homologous recombination process called transposition. These processes differ in their biological and genetic consequences and in their reaction mechanisms. In site-specific recombination both recombining molecules are cleaved at two precise positions within a very short region of homology, and there is no synthesis or degradation of DNA during the strand exchange reaction (which in both processes involves a series of precise catalytic steps). In transposition there is no requirement for homology in the recombination sites and DNA synthesis is utilised during the strand breakage and reunion steps. During replicative transposition the entire transposing DNA segment is replicated, whereas in conservative (or "cut and paste") transposition the DNA

synthesis involves only a few bases of sequence. Reviews of transposition have been published by Craig and Kleckner (1987) and Derbyshire and Grindley (1987) and it will not be considered further here.

1.2 Consequences of site-specific recombination

Site-specific recombination alters the structure of DNA molecules in a number of ways which depend on the organisation of the recombination sites (see fig.1.1). Recombination between sites on different molecules is possible, and this causes the two substrate molecules (which could, for instance, be plasmids) to become fused; site-specific recombination of this sort is termed "intermolecular". Intramolecular recombination (between sites on the same molecule) is also possible, and has two possible outcomes depending on the relative orientations of the recombination sites. When the sites are in direct repeat in a circular substrate molecule, site-specific recombination results in deletion of the DNA intervening between the sites and generates two smaller product molecules; this reaction is termed "resolution" in this thesis, but has also been called "excision" in other systems (see below). In contrast, when recombination sites are in inverted repeat relative to each other, recombination causes the DNA sequences between the sites to invert in relation to the rest of the molecule. These reaction outcomes are a consequence of the fact that site-specific recombination sites are asymmetric, and the asymmetry is retained during recombination such that the left hand side of one site is always fused to right hand side of the other site, and *vice versa.* It should be noted, however, that in some circumstances this asymmetry can be overcome (see, e.g., Hoess *et al*, 1986).

Although the intramolecular reaction products were described above as being simple circles, it has been shown that topologically more complex products can be created during site-specific recombination reactions *in vitro.* This is illustrated, as an example, in Figure 1.1 by showing that catenated products can be generated during resolution reactions. Analysis of the topological nature of these, and other, *in vitro*

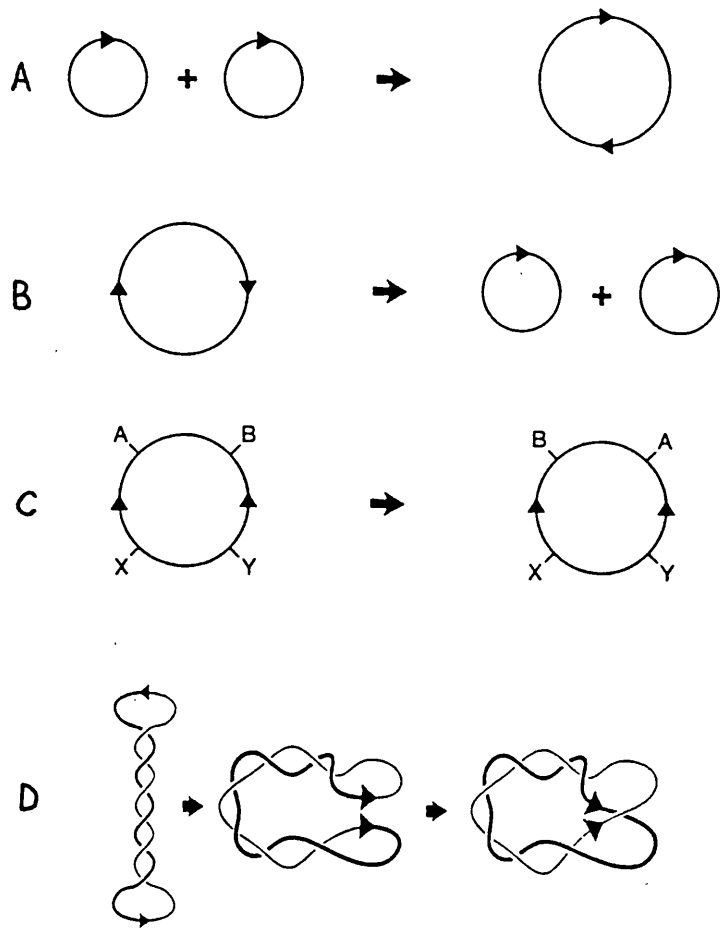


Figure 1.1. Physical consequences of site-specific recombination. Recombination sites are shown as arrows whose direction indicates the site's orientation. **A** Intermolecular recombination results in the fusion of two substrate DNA circles. **B** Intramolecular recombination in a circular substrate DNA molecule containing directly repeated sites results in two product circles; this is termed resolution in this thesis. **C** Intramolecular recombination of a substrate containing sites in inverted orientation results in inversion of the DNA between the sites. **D** Site-specific recombination *in vitro* in a supercoiled substrate molecule can generate topologically complex products; in this case resolution generates a multiply linked catenane because substrate supercoils were trapped during the reaction. (Adapted from Craig, 1988).

products has allowed information to be derived regarding the detailed mechanisms of various site-specific recombination reactions (this is discussed below).

The biological consequences of site-specific recombination are very diverse and incorporate the structural considerations described above. This is illustrated in the following examples:

(i) Integration and excision of temperate bacteriophages into and out of the host organism's chromosome. Integration of the phage generates a prophage, in which the phage genome becomes part of the host genome and is replicated by the host machinery. Excision is required to release the phage and allow it to spread horizontally, by infection, into other cells. Bacteriophage λ is notable because it was here that site-specific recombination was first described (reviewed by Landy, 1989), but other phages also employ this strategy (e.g. ϕ80 and P22; Leong *et al*, 1985).

(ii) Resolution of cointegrate intermediates during replicative, inter-replicon transposition of type II transposons. These Tn3-like transposons move by a replicative mechanism during which the donor and recipient replicons become fused. Resolution of these cointegrate structures is achieved by site-specific recombination of directly repeated transposon-borne *res* sites, and is catalysed by transposon-encoded recombinases called resolvases (Sherratt, 1989; Stark *et al*, 1989).

(iii) Control of gene expression by "switching" of DNA segments. The Hin, Gin, Cin and Pin systems encode a family of DNA invertase recombinases that execute inversions of DNA segments (either containing genes or the promoter sequences of genes), and thereby mediate the variation of flagellar antigens in *Salmonella typhimurium* and tail fibre proteins in bacteriophages Mu, P1 and relatives (reviewed by Glasgow *et al*, 1989). *fimB* and *fimE* encode *Escherichia coli* recombinases that are unrelated to the DNA invertases which catalyse the inversion of a small chromosomal region containing the promoter for the gene *fimA*, and therefore alter the fimbriation state of the cell (Klemm, 1986).

(iv) Stabilisation of low copy number plasmids by multimer resolution. Bacteriophage P1 does not integrate into the host's chromosome during lysogeny, but

instead exists as a low copy number circular plasmid. The phage-encoded recombinase Cre is required to resolve dimers of P1 which arise through homologous recombination to ensure that the lysogenic phage is efficiently segregated into daughter cells during host division (reviewed by Sadowski, 1986). Similar systems are employed in other low copy number plasmids (e.g. the plasmid R46 also encodes a recombinase; see Dodd and Bennett, 1987).

(v) FLP is a recombinase encoded by the 2-micron plasmid found in *Saccharomyces cerevisiae* which acts on plasmid-borne FRT sites to cause inversion of a large DNA segment and promote amplification of the plasmid to high copy numbers in the cell (Cox, 1989).

1.3 General reaction mechanisms in site-specific recombination

A number of site-specific recombination sytems have been reconstituted *in vitro*, using either crude cell extracts or purified proteins, simple buffer solutions and defined substrate molecules. The reactions that have been established in this way include those catalysed by λ Integrase (Int; Nash, 1975), yeast FLP (Vetter *et al*, 1983), bacteriophage P1 Cre (Abremski *et al*, 1983) and some of the resolvases and DNA invertases (Hatfull and Grindley, 1988). Analysing the reactions *in vitro* has shown that the basic reaction mechanisms of the different systems (described above) are similar, but the detailed biochemistry of the reactions can be divided into two classes; the integrase-like and the resolvase/invertase class of reactions (see below). The distinction between the reaction mechanisms is reflected in the fact that all recombinases can be assigned as being part of one of the two classes because the proteins are homologous within the two classes (see below).

The various recombination sites in the different systems share little sequence homology, and in fact the two recombining sites in a given system can be different in primary sequence. However, in all site-specific recombination reactions a small region

of sequence identity between the recombination sites is essential. This is where the strand exchanges that occur during the recombination reaction are found. The parental DNA molecules are broken at specific, fixed positions within these regions and the recombinational strand exchanges occur by joining the broken ends of one parental duplex to the ends of the other broken parental duplex. Recombinant DNA molecules generated in this way contain no gaps or nicks. The breakage positions in the top and bottom strands of both parental duplexes are staggered (see figures 1.3 and 1.7), and consequently the recombinant DNA molecules contain some heteroduplex DNA. The size of this stagger varies from 2-8 bp, and is one of the features that distinguish the two classes of site-specific recombination reaction (see below).

The strand exchange mechanism of all site-specific recombination reactions involves a transient, covalent linkage between the substrate DNA and recombinase proteins (see figs. 1.3 and 1.7). Correlating with this is the fact that in all recombination sites the recombinases bind to DNA sequences surrounding the points of strand exchange. The DNA sequence which encompasses the recombinase binding sites and staggered cleavage positions is termed the "crossover" region or "core" recombination site. The form that the covalently linked protein-DNA intermediate takes is different in the two classes of recombination reaction and is conserved within each class (see below). The function of this intermediate is to conserve the energy of the substrate DNA molecule's phosphodiester backbone, and hence no site-specific recombination reactions utilise high energy co-factors.

In some site-specific recombination systems the crossover region contains all the sequence information that is required in the reaction (e.g. the *loxP* site of Cre, the FRT site of FLP and the *attB* site of phage λ; below). In other cases, the recombination sites have a more more elaborate architecture, and can employ additional proteins and protein binding sites (e.g. in λ, the *attP*, *attL* and *attR* sites contain binding sequences for the proteins IHF, Xis and Fis). In all cases, however, the catalytic steps of cutting, exchanging and ligating the DNA are mediated by the recombinases, and the other so-called "accessory factors" have various supporting roles in the reactions.

Current understanding of the biochemical details of resolvase/DNA invertase and λ integrase-like site-specific recombination reactions, and the functions of the site-specific recombination proteins, are described separately below. It should be noted, however, that these families cannot be separated in a biological context, since similar reactions are performed by members of both classes.

1.4 The resolvase/DNA invertase class of site-specific recombination system

As suggested by the name, this family is divided into two sub-families on the basis of the type of reaction that is executed, despite the fact that they apparently share the same reaction mechanism. The resolvases catalyse resolution reactions and are primarily involved in recombining cointegrate structures (described above), although they have also been described in plasmid-stabilising monomerisation (e.g. R46; Dodd and Bennett, 1987). The resolvases encoded by the transposons Tn3 and γδ are the best studied *in vitro* (although resolvases from other transposons have also been described, e.g. Tn21, Tn552 and Tn1721; reviewed by Hatfull and Grindley, 1988; Stark *et al*, 1989 and Sherratt, 1989). The functions of the invertases Gin, Hin, Cin and Pin have been discussed (Section 1.2) and the sites of all these systems are interchangeable (Glasgow *et al*, 1989). Recombination of Gin and Hin have been extensively analysed *in vitro* (Kahmann *et al*, 1985; Bruist *et al*, 1987).

These recombinases are small proteins of highly conserved size (all approximately 180 amino acids) and sequence (see fig.1.2). The proteins appear to be comprised of two structural domains that can be separated by proteolytic digestion; a C-terminal domain of approximately 40 amino acids that is mainly responsible for sequence specific DNA binding (Abdel-Meguid *et al*, 1984), and an N-terminal domain (whose X-ray crystallographic structure has been solved; Sanderson *et al*, 1990) that contains the catalytic site and is responsible for protein-protein interactions.

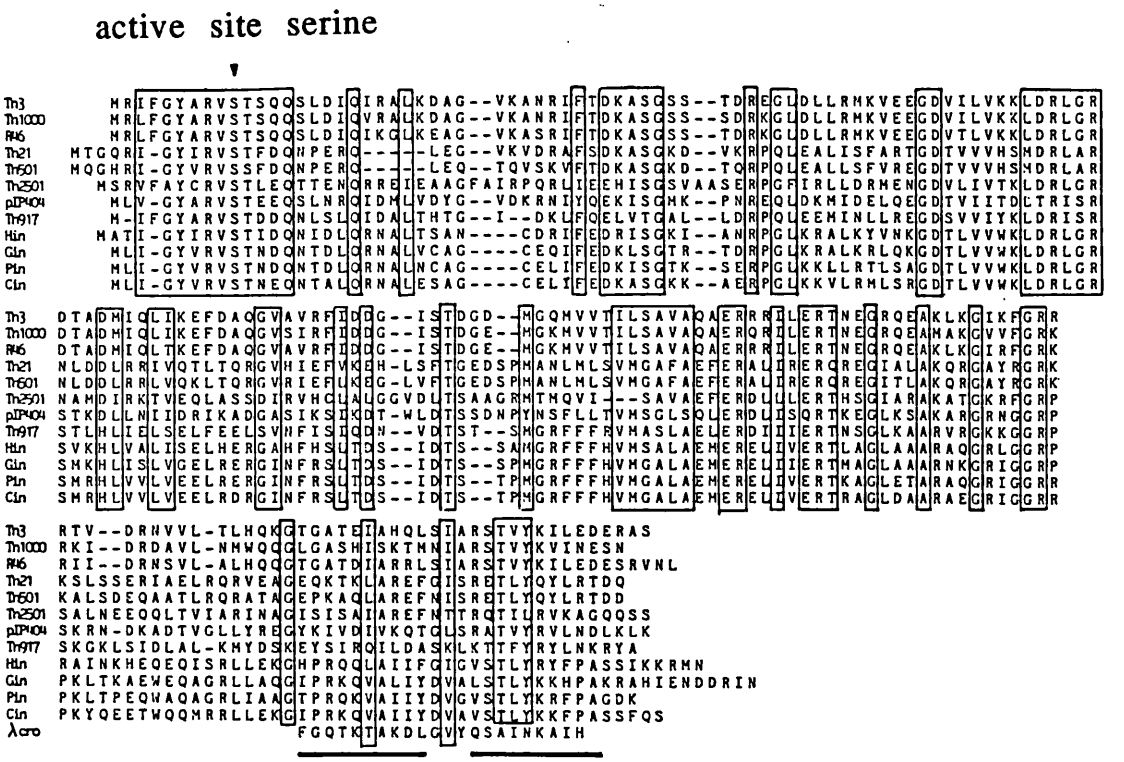


Figure 1.2. Amino acid sequence alignment of some members of the resolvase/DNA invertase family of site-specific recombinases. These proteins are believed to interact with their DNA binding sites *via* a helix-turn-helix motif, as shown. Also indicated is the proposed active site serine residue. (Taken from Sherratt, 1989).

The recombination site of Tn3 (called *res*) is approximately 120 bp in length and contains 3 sub-sites (named I, II and III), each of which has dyad symmetry and is believed to bind dimers of resolvase (Grindley *et al*, 1982; Kitts *et al*, 1983). Strand exchange occurs exclusively at the central AT sequence of sub-site I, but II and III are essential for recombination. All the different *res* sites have the same organisation, although the size of the subsites and spacings between them vary considerably (M.Boocock, pers.comm.). The different activities of resolvase when bound to the different subsites is probably due to differences in the size and sequences of the sub-sites altering the nature of the protein-DNA interactions. Resolvases subunits bound at sub-sites II and III might act as accessory factors that facilitate the formation of a nucleoprotein synaptic complex in which the recombination reaction occurs (Stark *et al*, 1989; see fig.1.4).

The recombination sites of the DNA invertases are 26 bp in length and, like *res* sub-sites, have dyad symmetry and probably bind dimers of the recombinase (Mertens *et al*, 1988). The strand exchange reactions occur at the central 2 bp of the sites (Klippel *et al*, 1988). They are only efficiently recombined, however, if an enhancer sequence (called *sis*), which binds the host protein Fis, is present in the same DNA substrate molecule (Kahmann *et al*, 1985). Fis therefore represents an accessory factor that possibly performs a similar function in the invertase reaction to resolvase at sub-sites II and III during resolvase recombination (Bednarz *et al*, 1990; see fig.1.4).

The recombination reactions catalysed by resolvases and DNA invertases are mechanistically very similar. The top and bottom strand cuts are staggered by 2 bp and create recessed 5' ends and overhanging 3'OHs (fig.1.3). Putative intermediates have been isolated using abnormal *in vitro* reaction conditions in which the substrate molecules have been cut at all four cleavage positions, suggesting that the recombination reactions proceed by double strand breaks, duplex rotation and ligation (Klippel *et al*, 1988; Reed and Grindley, 1981). The isolated intermediates are covalently linked at each broken 5' end to a resolvase/DNA invertase protein, and it is believed that an absolutely conserved serine residue present near the N-terminus of the

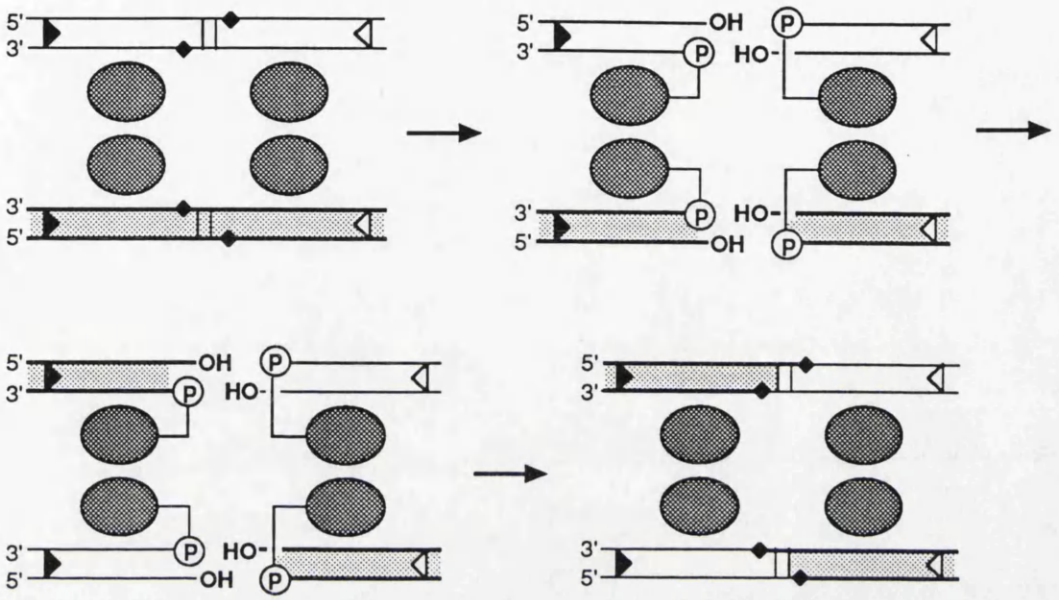


Figure 1.3. Events at the crossover site during resolvase/DNA invertase-catalysed site-specific recombination reactions. Recombinase subunits are depicted as shaded ovals; the ends of the crossover sites by inverted arrows; the two base pairs between the staggered cleavages (the overlap) as vertical lines; and the phosphates that are attacked by the recombinases as black diamonds. The four DNA strands (thick and thin lines) of the two crossover sites (differentiated by shading) are cleaved and covalently linked protein-DNA intermediates form. Recombination occurs by rotation of the left half-sites relative to the right half-sites and subsequent joining of the broken DNA ends to a new half site. (Adapted from Stark *et al*, 1992).

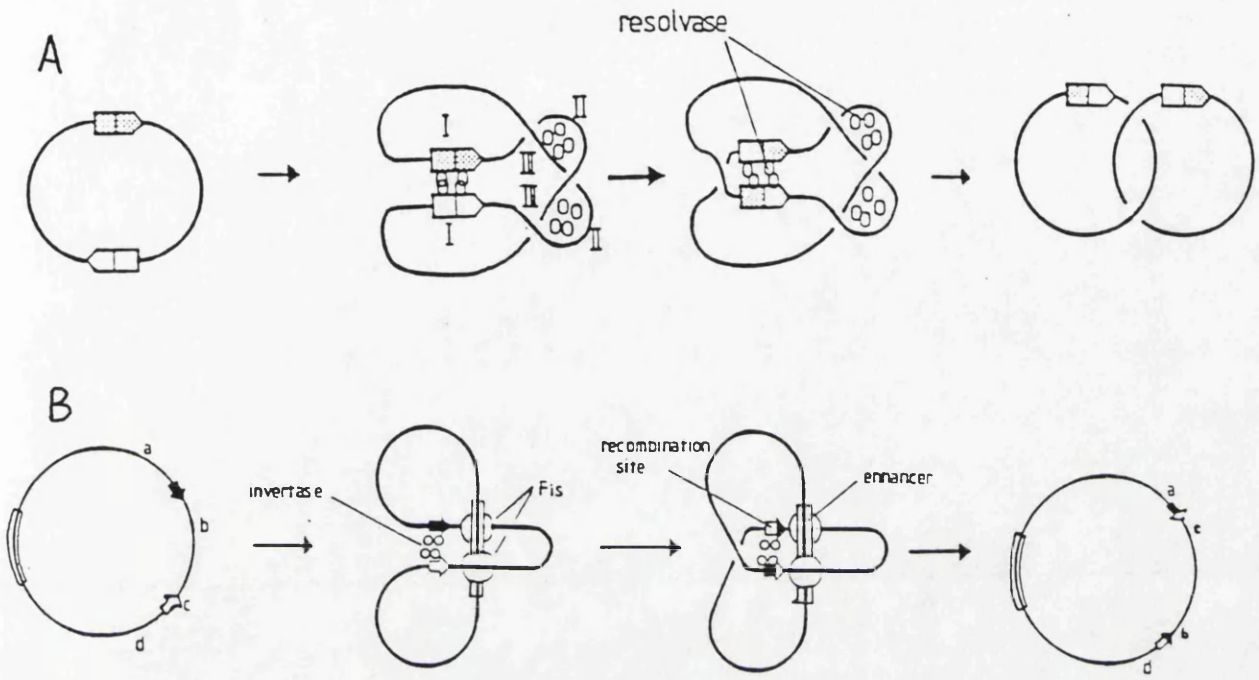


Figure 1.4. Topology of reactions catalysed by the resolvases and DNA invertases. **A** The resolution reaction catalysed by resolvases proceeds *via* a -3 synapse and produces a singly-interlinked (-2) catenane. The -3 synapse of the *res* sites required for the reaction may be specified and stabilised by interwrapping of subsites II and III around resolvase. **B** Inversion by DNA invertases proceeds *via* a -2 synapse and produces an unknotted circular product. The -2 synapse may be specified and stabilised by interactions between the enhancer element and bound Fis.

In **A**, the three subsites of the *res* site are shown as I, II and II and the resolvase subunits within the synapse are indicated. In **B**, a, b, c and d are markers to illustrate the inversion reaction; the DNA invertase and Fis proteins and their DNA binding sites (the recombination site and enhancer respectively) are indicated.

proteins (fig.1.2) is the catalytic residue involved in the phosphodiester linkage (Klippel *et al*,1988; Hatfull and Grindley, 1986).

Tn3 *res* sites are only efficiently recombined *in vitro* when they are present in direct repeat on the same, supercoiled DNA molecule (Kitts *et al*, 1983). Similarly, DNA invertase sites must also be present on the same, supercoiled substrate, but, in contrast to *res* sites, they must be in inverted repeat (Kahmann *et al*, 1985). In addition to this site orientation selectivity the resolvase/DNA invertase reactions *in vitro* have topological selectivities (summarised in fig.1.4): resolvase-catalysed deletion reactions normally produce a specific (-2) catenane product and involve a linkage change of +4 (Wassermann and Cozzarelli, 1985; Boocock *et al*, 1987); DNA invertase-catalysed inversion normally produces an unknotted circular product and has a +4 linkage change (Kahmann *et al*, 1987; Kanaar *et al*, 1988). These results have suggested that the strand exchange reactions only occur when the recombining sites form a specifically interwrapped synaptic complex, and that the strand exchange is a simple right-handed 180^{o} rotation of the cleaved DNA duplexes and subsequent ligation. The differences in the two reactions' product topologies can be accounted for by the different synaptic topologies required for strand exchange. Resolvase reactions occur *via* a -3 synapsis which is thought to be stabilised by resolvase subunits bound to the interwrapped accessory sites (Stark *et al*, 1989a and 1989b), whilst the inversion reactions occur *via* a -2 synapse that is stabilised by Fis bound to the *sis* enhancer (Bruist *et al*, 1987; Kanaar *et al*, 1989).

1.5 The λ integrase class of site-specific recombination system

The second class of site-specific recombination reaction is called the λ integrase class because of the sequence homologies of the recombinase proteins to λ Int. The functions of some of these reactions have been discussed (λ Int, FimB/E, FLP and Cre; see above), but many other λ integrase-like recombinases have been described; e.g.

TnpI of Tn4430 mediates cointegrate resolution (Mahillon and Lereclus, 1988) and the D protein encoded by the *E.coli* F factor is involved in plasmid-stabilising monomerisation (Lane *et al*, 1986 and O'Connor *et al*, 1986).

This family of recombinases has no sequence homologies with the resolvase/DNA invertases and the proteins display much greater variation in size and sequence. Two regions of amino acid similarity have been identified by comparing the recombinase sequences, termed domains 1 and 2 (Argos *et al*, 1986; Abremski and Hoess, 1992; see fig.5.17). Only four residues are completely conserved in all published λ integrases. The reactions catalysed by λ Int, P1 Cre and FLP are the only ones to have been analysed *in vitro*, and therefore only they will be described in detail (although it is likely that all the λ integrase-like systems have a common basic reaction mechanism).

The recombination sites of FLP and Cre (called FRT and *loxP* respectively) are simpler than the *att* sites of λ Int (see figs.1.5 and 1.6). *loxP* comprises inverted 13 bp Cre binding sites surrounding an 8 bp spacer sequence, and therefore it is simply a crossover site (Hoess and Abremski, 1985). FRT also comprises two 13 bp inverted recombinase binding sites surrounding an 8 bp spacer, but differs from *loxP* in that it has an additional 13 bp FLP binding site that is not essential for recombination (Gronostajski and Sadowski, 1985a; Andrews *et al*, 1987). Neither of these site-specific recombination reactions require any accessory factors. In contrast, the *att* sites of λ are complex in structure and Int-catalysed recombination requires accessory factors. Two pathways for λ recombination exist, one for phage integration and the other for prophage excision (fig.1.6). The integration reaction recombines the phage-borne *attP* site with the *E.coli* chromosomal *attB* site and generates *attL* and *attR* sites; it requires the host-encoded IHF accessory protein in addition to Int. Excisive recombination recombines *attL* and *attR* to regenerate *attP* and *attB*; it is not, however, the exact reverse since it requires Int, IHF, phage-encoded Xis and is stimulated by host-encoded Fis. The two reaction pathways are carefully regulated (reviewed by Landy, 1989). The crossover sites of *attP* and *attB* are not identical, but have a similar

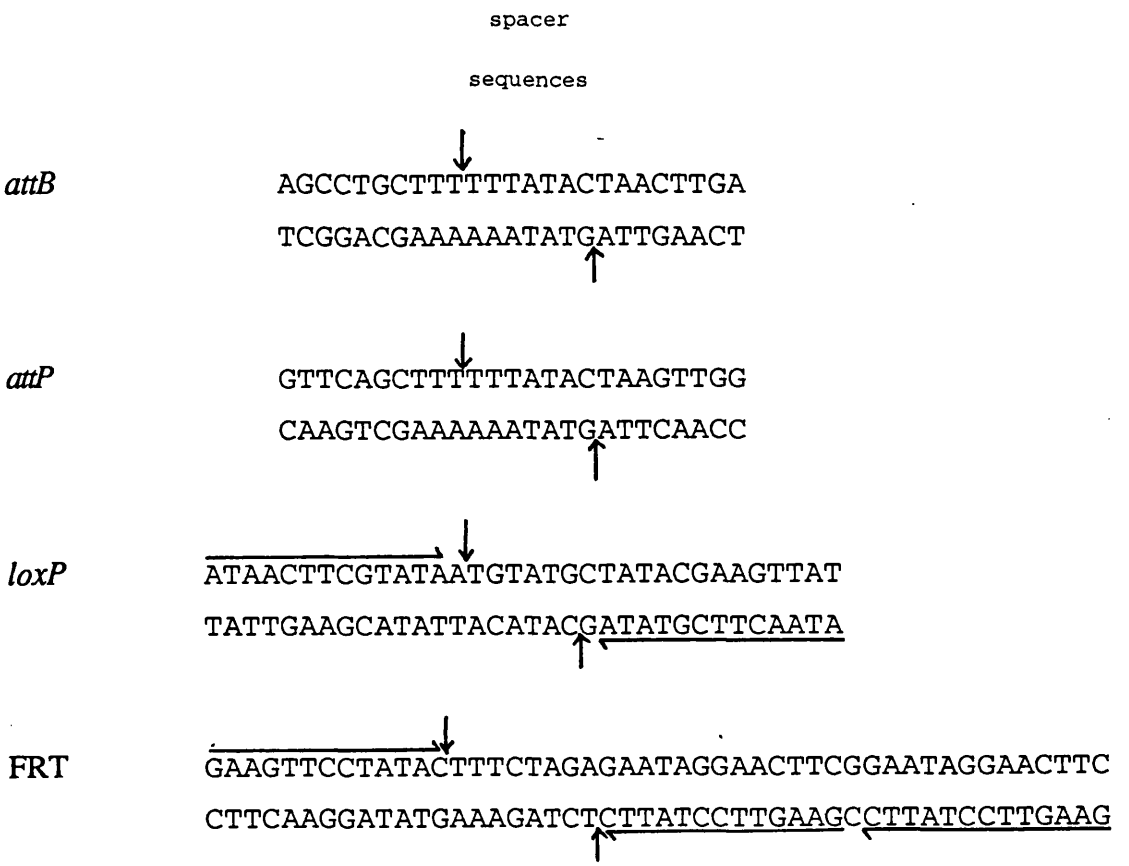


Figure 1.5. Comparison of the crossover sites for λ Int (*attP* and *attB*), FLP (FRT) and Cre (*loxP*). Each site consists of two binding sites for the recombinases (which are inverted repeats in *loxP* and FRT) surrounding the spacer sequence. In FRT the right-most FLP binding site can be deleted. The sites are cleaved by their respective recombinases on either side of the overlap, as shown.

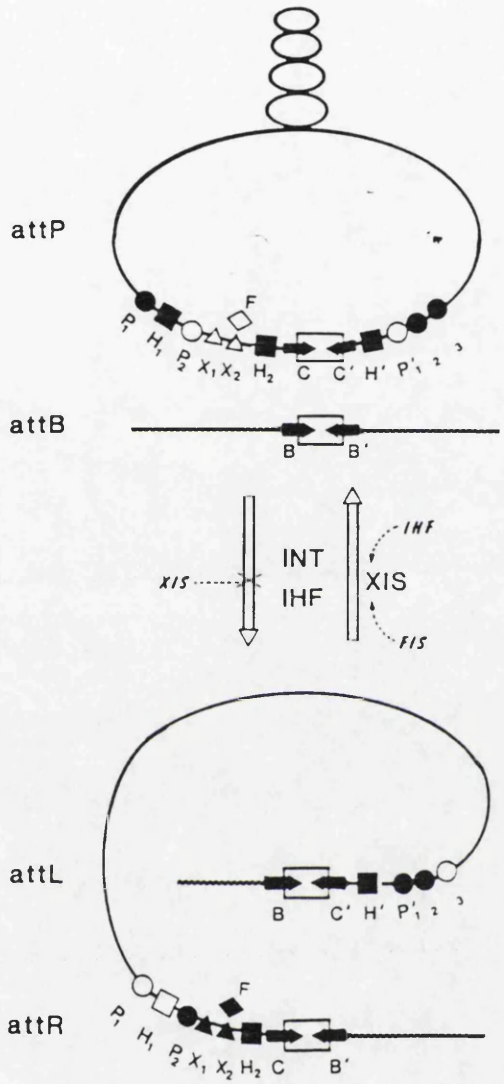


Figure 1.6. Phage λ integrative and excisive recombination pathways. Integration of phage λ into the host chromosome involves recombination between phage-borne *attP* and chromosomal *attB* and generates *attL* and *attR*, excision regenerates *attP* and *attB*. The proteins required for the two reactions are indicated, as are the proteins' binding sites: IHF (H), Xis (X), Fis (F), arm-type Int (P) and core-type Int (C or B).

organisation to *loxP* and FRT sites (fig.1.5) since they comprise two Int binding sites (so-called core sites; see below) around a 7 bp spacer sequence. The full *attP* site is 240 bp in size and is made up of, in addition to the crossover region, multiple binding sites for Int (arm-type binding sites; see below), IHF, Xis and Fis (Landy, 1989; fig.1.6). Int's ability to bind core and arm sequences is a consequence of the fact that the protein contains two DNA-binding domains which can be separated by proteolytic cleavage; a 7.5 kDa N-terminal fragment is made which binds the arm sites and a 32 kDa C-terminal fragment that binds the core sites (Moitoso de Vargas *et al*, 1988).

In all these recombination reactions the strand exchanges are made within the crossover sites, and the top and bottom strand cuts are made at fixed positions near the edges of the spacer sequences (fig.1.5). The size of the stagger between the cut positions is greater than in the resolvase/DNA invertase reactions and varies from 6 to 8 bp; the sequence between the cuts is called the "overlap" sequence. Notice that the spacer and overlap sequences are not necessarily equivalent, since in *loxP* the spacer between the Cre binding sites is 8 bp and the stagger between cleavage positions is 6 bp (fig.1.5). A further difference between the integrases and resolvese/invertases in the cleavage step of the reactions is that cleavage on both strands of integrase-like sites would generate a protruding 5'OH and a recessed 3' phosphate. In certain *in vitro* reaction conditions, the integrase recombinases become covalently linked to their substrates *via* a phosphodiester link involving the 3' end of the cleaved DNA and the absolutely conserved tyrosine found in domain 2 (for Int see Craig and Nash, 1983 and Pargellis *et al*, 1988; for FLP see Andrews *et al*, 1987 and Gronostajski and Sadowski, 1985b; for Cre see Hoess and Abremski, 1985).

The mechanism of the strand exchange reaction employed by the λ integrases is quite different to the double strand breakages used by the resolvase/invertase class. The reactions proceed *via* two independent pairs of strand exchanges, as shown in figure1.7. For the Int and Cre reactions it has been demonstrated that the "top" pair of strand exchanges is always made first (Kitts and Nash, 1985; Hoess *et al*, 1987), but this may not always be true in FLP recombination (Jayaram *et al*, 1988). In all cases,

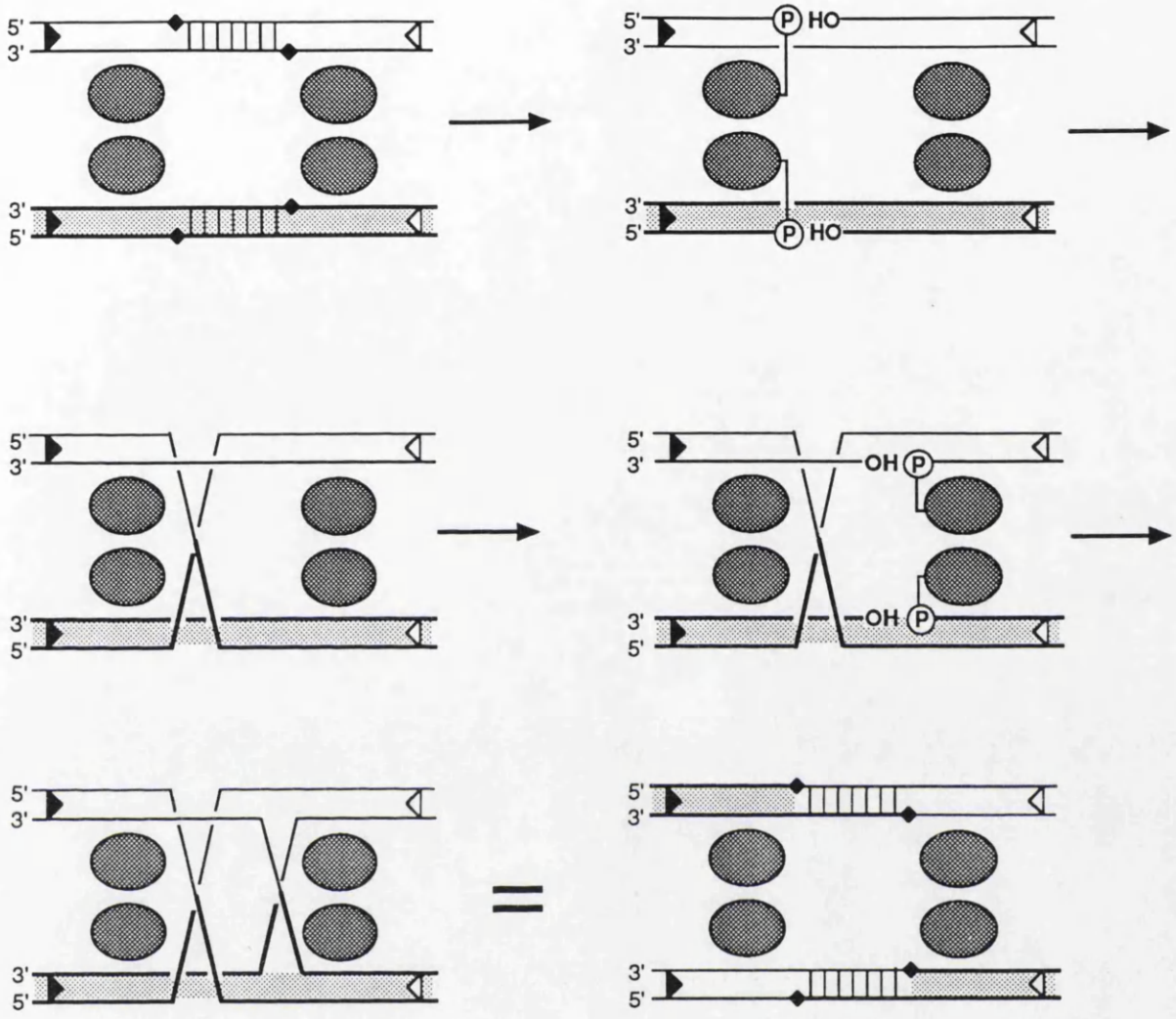


Figure 1.7. Events at the crossover site during recombination reactions catalysed by λ integrase-like recombinases. Recombinase subunits are depicted as shaded ovals; the ends of the crossover sites by inverted arrows; the base pairs between the staggered cleavages (the overlap region) as vertical lines (note the size of this can vary from 6-8 bp); and the phosphates that are attacked by the recombinases as black diamonds. A single DNA strand (thick or thin line) from each crossover site (the sites are differentiated by shading) is cleaved and covalently linked protein-DNA intermediates form. The strands are exchanged and a Holliday intermediate is created. Recombinant product is made by branch migration of the Holliday junction across the overlap and a second pair of strand exchanges; it is not intended that this figure imply that two 4-way junctions are present simultaneously. (Adapted from Stark *et al*, 1992).

exchange of the first pair of strands creates a Holliday junction intermediate that is converted to full recombinant product molecules by the second pair of strand exchanges. These Holliday intermediates have been isolated using specific *in vitro* reaction conditions and substrate molecules (For Int see Nunes-Duby *et al*, 1987 and Kitts and Nash, 1988; for FLP see Jayaram *et al*, 1988 and Meyer-Leon *et al*, 1988 and 1990; for Cre see Hoess *et al*, 1987). Resolution of the Holliday junctions by the second pair of strand exchanges requires that they branch migrate from the point that they are generated (by the first pair of strand exchanges) across the overlap sequence, as shown. It is proposed that this is the reason that the overlap sequences of the recombining sites must be homologous (see Stark *et al*, 1992 for review). Integrase-like recombination is therefore sub-divisible into two stages, each involving cutting and rejoining of the DNA strands.

The rigid substrate requirements described for resolvase/DNA invertase recombination is not reflected in integrase-like reactions. Neither FLP nor Cre require supercoiling of their substrate molecules *in vitro*, and both enzymes can catalyse both intermolecular and intramolecular (deletion and inversion) recombination (Vetter *et al*, 1984;Gronostajski and Sadowski, 1985c; Abremski and Hoess, 1984). Integrative recombination catalysed by Int does require a supercoiled *attP* substrate molecule (though not *attB*), whilst excisive recombination does not need supercoiled substrates (Landy, 1989). This lack of selectivity may suggest that the integrase family of site-specific recombination reactions does not have the same restrictions as the resolvase/invertase family in forming catalytically competent synapses.

1.6 Identification of the *xer* site-specific recombination system

The *xer* site-specific recombination system was identified during analysis of the mechanisms employed by the naturally occurring, high copy number plasmid ColE1 for its stable maintenance in growing cell cultures. Various plasmid functions have been

described which ensure the stable maintenance of plasmids during cell growth and division (reviewed by Nordstrom and Austin, 1989); the two basic requirements are that the rate of plasmid replication matches the rate of replication of the host's chromosome(s), and that after cell division each daughter cell receives at least one copy of the plasmid. The first of these requirements is satisfied by the different mechanisms used by plasmids to maintain their copy number at a set level in the host cell's cytoplasm. It was during analysis of the way that ColE1 fulfills the second requirement that the *xer* recombination system was discovered.

Plasmids can be partitioned by either an active or a random process as the host cell divides. There is evidence that low copy number plasmids employ the former partitioning mechanism. The F factor of *E.coli* and the plasmids R1 and P1 all have systems comprising *cis*-acting sites and plasmid-encoded *trans*-acting *par* proteins which interact with the host's segregation "machinery" and place a copy of the plasmids into each daughter cell as the host cell divides (reviewed by Austin and Nordstrom, 1990). In contrast, there is no evidence for active partitioning of high copy number plasmids such as ColE1 (and its relatives), and indeed all evidence suggests that these plasmids segregate into the daughter cells at random during host division (Summers and Sherratt, 1984). It is calculated that a random mechanism of this sort would generate plasmid-free segregants with a probability of 2^{1-n}, where n represents the number of independently segregating plasmid units. This means that an experimentally undetectable frequency of $<10^{-5}$ plasmid-free daughter cells per cell division is produced when the plasmid copy number is >18. Correlating with this is the observation that ColE1 has an estimated copy number of 30 at division (Timmis, 1981), and ColE1-free segregants have never been found in non-selective cell cultures.

Paradoxically, it was found that many commonly used cloning vectors (such as pACYC184) are lost from bacterial cultures at frequencies of 10^{-2}-10^{-5}, despite the fact that these vectors are derived from ColE1 (or its relative pMB1) and have copy numbers in excess of the naturally ocurring plasmids (Summers and Sherratt, 1984). Summers and Sherratt (1984) showed that in Rec^+ strains pACYC184 generates

plasmid multimers while ColE1 does not, and that this multimerisation is associated with an increase in the instability of the cloning vector. They also demonstrated that dimers of ColE1 introduced into these strains are rapidly converted to monomers while pACYC184 dimers are not, and that the multimers of pACYC184 are present at lower copy numbers in the cell. The conclusions derived from these results are two-fold:

(i) The instability of the high copy number cloning vectors is a result of homologous recombination causing multimerisation of the plasmids. Multimerisation results in an increase in the frequency of plasmid-free segregants because the multimers have numerous origins of replication, and, because copy number control mechanisms "count" the number of plasmid replication origins, the multimers are maintained at a lower copy number in the cell. The decrease in the number of independently segregating units increases the probability of generating daughter cells that are free of plasmid at host division (see above).

(ii) ColE1 contains a determinant, that is absent from cloning vectors, which acts to resolve dimers into monomers and therefore maximise the plasmid's copy number and stability.

1.7 Characterisation of the ColE1 *cer* recombination site

The determinant of ColE1 that causes dimer resolution was isolated by a combination of sub-cloning fragments of the plasmid, deletion mutagenesis and sequencing. It was originally defined as a 284 bp *Hpa*II-*Taq*I fragment, and is named *cer* (Summers and Sherratt, 1984; Summers *et al*,1985; Summers and Sherratt, 1985; see fig.1.8). A plasmid with two *cer* sites present in direct repeat acts as a substrate for site-specific recombination. When the *cer* sequence is deleted from ColE1 the plasmid becomes unstable and multimerises in Rec^+ strains. If it is cloned into pUC8, *cer* increases the stability of the cloning vector and stops it from multimerising in Rec^+ strains. The site-specific recombination reaction at *cer* sites is highly directional;

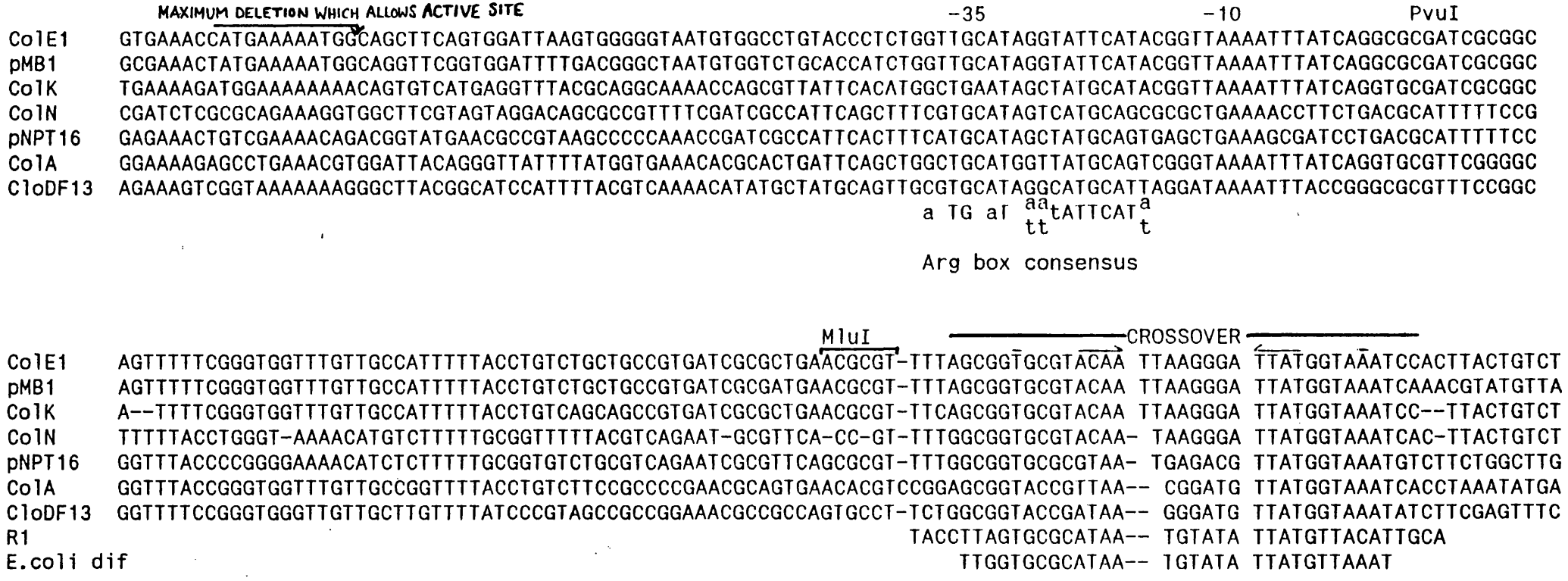


Figure 1.8. Alignment of the DNA sequence of ColE1 *cer* to homologous sites from the plasmids pMB1, ColK, ColN, pNTP16, ColA, CloDF13 and R1, and to the chromosomal *dif* sequence. The positions of the arg-box, promoter, proposed crossover and *Mlu*I sequences within *cer* are indicated. Also shown is the final deletion position at the left end of *cer* that still allows a functional site. Gaps (-) have been inserted to maximise the sequence homologies.

	Left arm recombinase binding site	Spacer	Right arm recombinase binding site
ColE1	GCGGTGCGTACAA	TTAAGGGA	TTATGGTAAAT
ColK	GCGGTGCGTACAA	TTAAGGGA	TTATGGTAAAT
pMB1	GCGGTGCGTACAA	TTAAGGGA	TTATGGTAAAT
ColN	GCGGTGCGTACAA	-TAAGGGA	TTATGGTAAAT
ColA	GCGGTGCGTACAA	--CGGATG	TTATGGTAAAT
NPT16	GCGGTGCGCGTAA	-TGAGACG	TTATGGTAAAT
ColE2	GGGGGGCGTACAA	--CGGGAG	TTATGGTAAAT
ColE3	GGGGTGCGTACAA	--CGGGAG	TTATGGTAAAT
pSC101	GCGGTGCGCGCAA	--GATCCA	TTATGTTAAAC
CloDF13	GCGGTACCGATAA	--GGGATG	TTATGGTAAAT
type I hybrid	GCGGTGCGTACAA	TTGGGATG	TTATGGTAAAT
type II hybrid	GCGGTGCGTACAA	--GGGATG	TTATGGTAAAT
dif	TTGGTGCGCATAA	--TGTATA	TTATGTTAAAT
R1	TTAGTGCGCATAA	--TGTATA	TTATGTTACAT

Figure 1.9. Comparison of the putative crossover sites from different *cer*-like sites. The conserved, putative left- and right-arm recombinase binding sites are indicated, as are the spacer sequences which, as shown, vary in their putative sequences and sizes (from 6-8 bp).This diagram was kindly supplied by J.Roberts.

plasmids containing directly repeated *cer* sites are resolved by intramolecular recombination and do not recombine intermolecularly (see e.g. Colloms, 1990 and this thesis). This directionality is expected from the site's function, since intermolecular multimerisation would lead to destabilisation and not stabilisation of ColE1.

Highly homologous sites have been identified from a large number of related, naturally occuring plasmids: ColK (Summers *et al*, 1985), CloDF13 (Hakkaart *et al*, 1984), pMB1 (Greene *et al*, 1981), ColA (Morlon *et al*, 1988), pNTP16 (P.Strike, pers.comm.) and ColN (Kolot, 1990) all contain sites that have homology with the entire 284 bp of *cer* (see fig.1.8); in addition the plasmid R1 contains a plasmid stabilisation site with homology to the proposed crossover region of *cer* (see below and fig.1.8; Clerget, 1984). It is likely that most, if not all, plasmids contain sites that function in resolving multimers which arise by homologous recombination, although in many cases they are not homologous in sequence to *cer*. For example, P1 encodes the Cre protein that acts at *loxP* sites (Austin *et al*, 1981; see above), the F factor encodes the D protein which acts at *rfsF* sites (O'Connor *et al*, 1986) and a resolvase recombinase acts at the R46 *per* site (Dodd and Bennett, 1987).

The position where the strand exchange reactions occur in *cer* have been coarsely mapped by sequencing the reaction products of recombination between ColE1 *cer* and the *cer*-like sites from ColK and CloDF13 (Summers *et al*, 1985; Summers, 1989). This suggested that the exchanges had occurred within a 35 bp region at the right-hand end of the *cer* site. Comparing the numerous *cer*-like sites in this region showed that it is the most conserved sequence within the sites, and suggested that it might be the crossover sequence of *cer* with a similar organisation to the crossover sequences of λ *att* sites, P1 *loxP* sites and the yeast 2-micron plasmid FRT sites (see figs.1.9 and 1.5). It is hypothesised that the *cer*-like crossover regions comprise imperfect and highly conserved, inverted repeat recombinase binding sites flanking a less conserved spacer sequence. Notice that if this analysis is correct then the recombinase(s) that act on the *cer*-like sites are able to perform strand exchange reactions on various sites that have spacer regions which vary in their sequence and in their size (from 6-8 bp). The

positions of strand cleavage within this presumptive crossover sequence are not known, but, by analogy with other λ integrase-like systems, it is hypothesised that the cleavages are within the spacer sequence (see Chapter 5).

Deletion analysis determined that 200 bp of the sequence upstream of the putative crossover region is all that is needed for *cer* recombination, therefore suggesting that approximately 250 bp of the 284 *Hpa*II-*Tap*I fragment comprise the actual *cer* site (Summers and Sherratt, 1988). Within this 200bp region is a conserved "arg-box" sequence that is essential for the recombination reaction (see fig.1.10 and below; Stirling *et al*, 1988). The arg-box is only some 18 bp in size, however, and it is therefore necessary to explain the function of the remaining 180 bp (approximately) of sequence upstreamof the *cer* crossover. The sequences upstream of the arg-box are not believed to bind any proteins, but instead to act as a "flexible" region; this is based on the observation that they can be replaced by unrelated sequences that have alternating AT and GC tracts which are found in bent DNA (Summers and Sherratt, 1988; Drew and Travers, 1985). The region between the arg-box and putative crossover site is more conserved in sequence than the "flexible region", and the distance between the two sequence motifs is also conserved, but despite this its function is not yet understood. A transcript is made in this region, but does not appear to be involved in *cer* recombination since a mutation that reduces the level of its expression 60-fold has no effect on *cer* recombination (Summers and Sherratt, 1988).

Recombination between ColE1 *cer* and CloDF13 *parB* sites is inefficient and generates two *cer*-like sites (termed the typeI and typeII hybrids; see fig.1.8) whose recombination characteristics have important implications (see sections 1.8.1 and 1.8.2, below). The typeII hybrid site supports both intramolecular and intermolecular recombination, unlike *cer* which recombines in an exclusively intramolecular direction (Summers, 1989). All the sequences necessary for the typeII hybrid recombination reactions are present in the region downstream of the *cer MluI* site, therefore offering further evidence that this is the crossover region. The reasons for these altered reaction properties are currently being investigated, and it is likely that they are due to the

differences between the *cer* and typeII hybrid sites in their overlap and left arm crossover sequences (J.Roberts, pers.comm.).

1.8 The *E.coli* chromosomally encoded *xer* genes

As stated above, the only sequences of ColE1 required for *xer* site-specific recombination are within the 250 bp of the *cer* site. Although a transcript is expressed from *cer*, it has very limited coding potential and is not conserved in the other *cer*-like sites, and it was therefore suggested that the protein(s) which recombine *cer* are encoded by the *E.coli* chromosome (Summers *et al*, 1985). For this reason the transposon Tn5 was used to mutagenise *E.coli*, and *xer* mutants were identified by their inability to resolve a *cer* reporter plasmid, pKS455 (see fig.4.20), that carries selectable drug resistence markers. Three genes were identified by this technique as being essential for *cer* recombination; *xerA*, *xerB* (Stewart, 1986; Stirling *et al*, 1988 and 1989) and *xerC* (Colloms *et al*, 1990). All have been cloned and sequenced, and are discussed below. (While work was being performed for this thesis a fourth *xer* gene (*xerD*/*xprB*) was identified by different methods; it is described in Chapter 5.)

1.8.1 *xerA*

The map position in the *E.coli* chromosome (70.5 mins) and nucleotide sequence of the cloned *xerA* gene were shown to be identical to the sequence and map position determined for *argR*, which encodes the arginine biosynthetic repressor (Stirling *et al*, 1988; Lim *et al*, 1987). ArgR (in conjunction with its corepressor, L-arginine) is a negative regulator of the expression of the genes involved in arginine biosynthesis (for review, see Glansdorff, 1987). Gel retardation and footprinting experiments using the purified protein have shown that ArgR binds to an 18 bp sequence in *cer* that has homologies with the operator sequences found in the promoter regions of the arginine biosynthesis operons (Stirling *et al*, 1988; see fig.1.10). It is intriguing that the *cer* site

appears to contain only one copy of this 18 bp arg-box sequence, whilst the arginine biosynthesis genes, as well as *argR* itself, contain two repeats of the loosely conserved arg-box. The *cer* footprinting data that are available do not detail whether ArgR/XerA binds to the region directly downstream of the arg-box (which would be the location of a second arg-box if it was present) because its sequence has innate resistance to DNAaseI cleavage, and therefore the possibility that this may reflect an alteration of ArgR binding to *cer* (when compared to other operators) that has significance in terms of *cer* recombination remains to be investigated.

A number of results show that ArgR is not the *cer* recombinase, but is an accessory factor in the reaction; it displays no sequence homologies to either class of recombinases, it binds approximately 100 bp from the proposed strand exchange positions and typeII hyrid sites (see above) are able to recombine in the absence of functional ArgR (Summers, 1989). It seems unlikely that the role of ArgR during *cer* recombination is to repress transcription from the promoter which overlaps with the *cer* arg-boxes. This can be said because alterations in the level of expression of the transcript have no effect on *cer* recombination unless the mutations also affect ArgR binding (Summers and Sherratt, 1988; it may in fact be argued that the existence of the transcript and promoter sequences in the *cer* site are simply a consequence of the incorporation of the argR binding site during *cer* evolution). The role of ArgR may be analogous to that of IHF during λ recombination, and is in the assembly of higher-order protein-DNA complexes in which *cer* recombination occurs. In this role ArgR (along with the upstream sequences and PepA; see below) could contribute to the intramolecular directionality of the *cer* recombination reaction.

1.8.2 *xerB*

Database searching using the predicted amino acid sequence of XerB (derived from the cloned gene) revealed 31% identity to bovine lens leucine aminopeptidase (Stirling *et al*, 1989; see fig.3.2). This raised the possibility that *xerB* may be an *E.coli* gene encoding an aminopeptidase. Analysis of the Xer phenotype of various *pep*

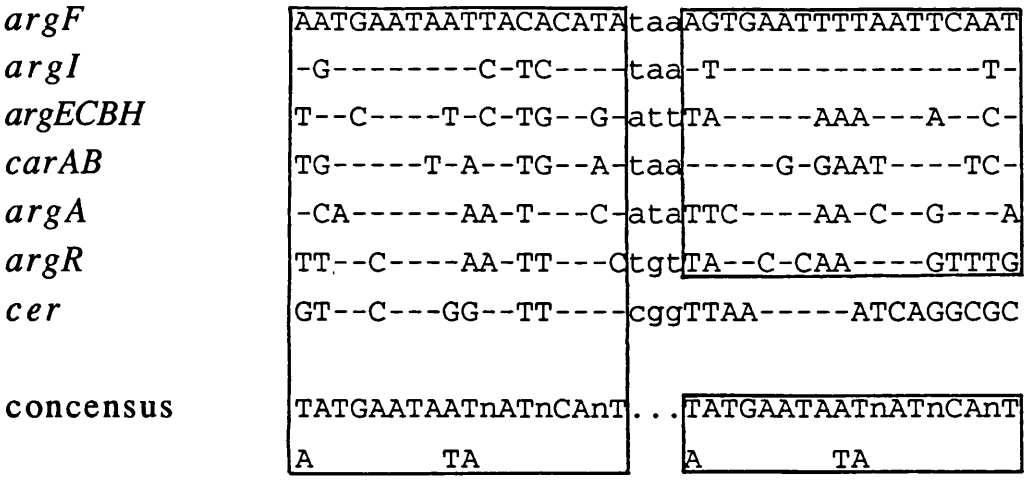


Figure 1.10. Comparison of the ArgR DNA binding sequence in *cer* to the arg-boxes from the arginine biosynthetic genes. The sequences are aligned against the arg-boxes found in the *argF* promoter region, bases unchanged are denoted by a dash and the boxes are framed. (Sequences taken from Glansdorff, 1987 and Stirling *et al*, 1988).

mutants of *S.typhimurium* suggested that *xerB* may be the *E.coli* equivalent of *pepA*; it was necessary to perform the genetic analysis in *S.typhimurium* because the *pep* genes of this organism have been more extensively mapped and analysed (Miller, 1987). It was further shown that a clone of the *S.typhimurium pepA* gene was able to complement an *xerB* mutation in *E.coli* and allow *cer* recombination (Colloms, 1990), therefore offering compelling evidence that *xerB* is *pepA* and encodes *E.coli* aminopeptidase A.

Aminopeptidase A (PepA) was previously purified from *E.coli* and named aminopeptidase I by Vogt (1970). Analysis of the purified enzyme *in vitro* has shown that it is heat stable (70 ^{o}C for 5 mins), has an approximate molecular weight of 52 kDa, aggregates to form a larger molecular weight species in low ionic strength buffer and that it is an exopeptidase which cleaves amino-terminal residues from various peptide substrates. *E.coli* and *S.typhimurium* encode a number of aminopeptidases (PepA, PepB, PepM, PepN and PepP) - as well as four dipeptidases (PepD, PepQ, PepE and PepG), a tripeptidase (PepT) and several carboxypeptidases - which all have non-specific substrate specificities (Miller and McKinnon, 1974; Miller and Schwartz, 1978; Miller, 1987). The exclusive requirement for PepA rather than the other aminopeptidases in *cer* site-specific recombination is therefore surprising, and is not yet understood, but it could function in either an enzymatic or structural capacity (this is considered in more detail in Chapter 3). It is clear, however, that PepA is an accessory factor (in conjunction with ArgR), since it has no sequence homologies to either class of recombinases and because type II hybrid sites will recombine in *pepA* strains (Summers, 1989).

1.8.3 *xerC*

This was the last *xer* gene to be identified, and has been mapped to the 85 minute region of the *E.coli* chromosome, between the genes for adenylate cyclase (*cya*) and DNA helicase II (*uvrD*; Colloms *et al*, 1990). The translated amino acid sequence of XerC has homologies to the λ integrase class of recombinases (see fig.5.17), and

partially purified preparations bind to the crossover region of the *cer* site (Colloms, 1990). This suggests that XerC is the recombinase responsible for the *xer* strand cleavage and exchange reactions, and it is likely that the recombination mechanisms will correspond to the scheme described for Int, FLP and Cre-catalysed recombination (see above).

xerC constitutes the third member of an operon which also contains the previously cloned gene *dapF*, encoding the enzyme diaminopimelate epimerase (Colloms *et al*, 1990; Richaud *et al*, 1987; Richaud and Printz, 1988). The co-transcribed genes in the operon are, in order, *dapF*, *orf235*, *xerC* and *orf238* (see fig.4.1). The roles of the proteins encoded by the two open reading frames are not known and no homologies to published proteins have been found. Why *xerC* should be part of an operon, and how its function may be related to, or regulated by, the other proteins is not understood.

1.9 A cellular role for *xer* site-specific recombination

Insight into the cellular role of the *xer* site-specific recombination system was gained through the observation that *xerC* mutant strains have a tendency to produce filaments and have aberrant and amplified nucleoids (Blakely *et al*, 1991). This suggested that *xerC* mutants, though viable, have defects in cell division and nucleoid segregation. The same phenotype is observed in strains carrying deletions in a region of the *E.coli* chromosome close to the terminus of replication called *dif* (Kuempel *et al*, 1991). Analysing the sequence of the *dif* region revealed a 33 bp sequence similarity to the crossover region of the *cer* site (see figs.1.8 and 1.9), and it was therefore proposed that *dif* is an *E.coli* chromosomal substrate for *xer* site-specific recombination.

The 33 bp *dif* sequence is sufficient to act as a substrate for site-specific recombination when cloned into plasmids, and XerC binds specifically to it in gel

retardation assays (Blakely *et al*, 1991). Recombination of *dif* does not require either of the *cer* accessory factors ArgR or PepA, which is consistent with the fact that *dif* and *cer* are only homologous at their crossover sites and *argR* and *pepA* strains do not have a filamentous phenotype. The *dif* site-specific recombination reaction, like recombination of the typeII hybrid (which can also function as a simple crossover site without accessory sequences), shows no directionality - i.e. it proceeds both intramolecularly and intermolecularly.

The above results have led to the hypothesis that the cellular role of the *xer* site-specific recombination system is in chromosome partitioning. Odd numbers of homologous recombinational exchanges between monomeric sister chromosomes (either during or after their replication) will generate chromosomal dimers. Dimers formed in this way cannot be partitioned into the daughter cells at cell division. Therefore it is proposed that *xer* site-specific recombination at *dif* acts to convert these chromosomal dimers into segregateable monomers, in an analogous function to plasmid stabilisation by recombination at *cer* sites (Blakely *et al*, 1991). It is believed that the positioning of the *dif* locus at the replication terminus allows all dimers to be resolved just prior to termination of replication, and it could minimise the decatenation required to separate the monomerised chromosomes. The apparent lack of directionality of the *dif* recombination reaction would create as well as resolve chromosome dimers, but it is possible that this reaction mechanism is needed because the sites are unable to determine when the chromosomes are dimeric, and therefore must rapidly recombine irrespective of the chromosomal configuration.

Support for the above hypothesis is provided by the fact that the filamentous phenotype and aberrant nucleoids of *xerC* strains are overcome when they are made *recA*, suggesting that without homologous recombination generating chromosome dimers *dif* recombination is not necessary (Blakely *et al*, 1991). It is also strengthened by the fact that *xerC* is widely distributed in bacteria (G.Blakely, pers.comm.). The hypothesis, however, still requires formal demonstration of *xer* recombination at the

chromosomal *dif* locus, perhaps by analysing the dimeric state of $XerC^+$ and $XerC^-$ cells.

Chapter 2

Materials and Methods

Table 2.1 Bacterial Strains

Strain	Genotype	Source/reference
AB1157	*thr*-1, *leuB6*, *hisG4*, *thi-1*, *ara-14*, Δ(*gpt-proA*)62, *argE3*, *galK2*, *supE44*, *xyl-5*, *mtl-1*, *tsx-33*, *lacY1*, *rpsL31*	Bachmann, 1972
DS941	AB1157, but *recF143*, *lacZ*ΔM15, *lacI*^{q}	D. Sherratt
DS942	DS941, but *lacZ*ΔH220 (*lacI*)	D. Sherratt
DS956	DS941, but *xerA9* (*argR*::fol)	D. Sherratt
CSX17	DS941, but *xerB1* (*pepA*::Tn5)	C.Stirling, 1989
HOM38b	DS941, but *pepA7*	H.O'Mara
DS980	DS941, but *orf235*::Tn5	S.Colloms
DS981	DS941, but *xerC*::Kan	P.Sykora
DS982	DS941, but *orf238Y2* (::mini Mu)	S.Colloms
DS984	DS941, but *xerCY17*(::miniTn10)	S.Colloms
DS9008	DS941, but *xprB*::miniTn10	M.Burke
STL116	AB1157, but *xprB*::miniTn10	S.Lovett
CS85	AB1157, but *ruvC53*, *eda-51*	B.Lloyd
JC7623	AB1157, but *recBC*, *sbcB*	C.Richaud
BMH 71-18	*thi*, *supE*, Δ (*lac-proAB*) *mutS*::Tn10 (F' *lacI*^{q} *lacZ*ΔM15)	Promega
EM1	DS941, but *mutS*	E.Morrell
RM10	DS941, but *lacPOxerC*, *dapF/orf238*::Kan	This work
RM20	"	"
RM30	"	"
RM40	"	"
RM41	RM40, but *xerA9*	"
RM42	HO'M38b, but *lacPOxerC*, *dapF/orf238*::Kan	"
RM43	RM40, but *ruvC53*	"
RM50	DS941, but *lacPOdapF*	"
RM60	DS942, but *xerC*::Kan	"
RM61	DS942, but *xerCY17*	"
RM62	DS942, but *xprB*::miniTn10	"

Table 2.2 Plasmids

Plasmid/resistance	Description	Source/Reference
pBR322 (Ap, Tet)	vector derived from pMB1	Sutcliffe, 1978
pUC18 (Ap)	vector derived from pBR322	Yanisch-Perron *et al*, 1985
pUC19 (Ap)	"	"
pUC9 (Ap)	"	"
pKK223-3 (Ap)	"	Pharmacia
pIC20R (Ap)	"	R.Wilson
pAT153 (Ap)	"	A.Twigg
pSELECT (Tet)	pBR322-derived mutagenesis vector	Promega
pUC71K (Ap, Km)	pUC19 + kanamycin resistence gene	Pharmacia
pCT1050 (Ap)	pAT153 + λP_L promoter	R.Thompson
pGP1-2 (Cm)	T7 polymerase expression vector	Tabor and Richardson, 1985
pGP1-2Km (Km)	vector derived from pGP1-2	S.Rowlands
pBAD (Ap)	pKK223-3 derived P_{tac} expresion vector	A.C.Boyd
pCB105/6 (Cm)	λdv-based vector	"
pKS455 (Ap, Cm)	pUC9-based 2-*cer* reporter plasmid	D.Summers
pCS202 (Cm, Tet)	λdv-based 2-*cer* reporter plasmid	C.Stirling, 1987
pCS126 (Ap)	pBAD + 1.9 kbp *Hind*III *pepA* fragment	"
pCS118 (Cm)	pCB106 + 1.9 kbp *Hind*III *pepA* fragment	"
pCS350 (Cm)	pCB106 + 920 bp *Sph*I-*Acc*I *argR* rfagment	"
pSD105 (Ap)	pBAD+1.2 kbp*Hind*III-*EcoR*I *xerC* fragment	S.Colloms, 1990
pSD102 (Ap)	pTZ18R + 3.8 kbp *Hind*III-*Bgl*II *xerC* fragment	"
pSD110 (Ap)	pBR322-based 2-typeII reporter plasmid	"
pSD113 (Ap)	pBR322 + 300 bp *cer* fragment in *EcoR*I-*Hind*III	"
pSD115 (Ap)	pSD113 + 300 bp *cer* in *Pvu*II	"
pSD124 (Ap, Km)	pUC18-based 2-*dif* reporter plasmid	Blakely *et al*, 1991
pSD126 (Ap, Km)	pBR322-based 2-*dif* reporter plasmid	"
pGS762 (Ap)	pUC18 + 1.12 kbp *ruvC* fragment	Sharples and Lloyd, 1991
pJC763 (Ap)	pBR322 + 3.8 kbp *xerB* fragment	Lovett and Kolodner, 1991
pRM10 (Ap)	pIC20R + 450 bp *Ssp*I-*Hind*III *rrn*B terminator from pKK223-3	This work
pRM11 (Ap, Km)	pRM10 + 1.3 kbp *Pst*I kanamycin (ex pUC71K)	"
pRM20 (Tet)	pSELECT + 1.9 kbp *Hind*III *pepA* fragment	This work
pRM21 (Ap, Tet)	pRM20, but containing *pepA*E354A	"
pRM40 (Ap)	pBAD + 1.9 kbp *Hind*III *pepA*E354A (ex pRM21)	"
pRM50 (Ap)	pCT1050 + 1.2 kbp *Hind*III-*EcoR*I *xerC* fragment	"

Table 2.2 (continued)

pRM60 (Cm)	pGP1-2 + 1.6 kbp *EcoR*I-*BamH*I λP_L-*xerC* fragment from pRM50	This work
pRM65 (Km)	pGP1-2Km + 1.6 kbp *EcoR*I-*BamH*I λP_L-*xerC* fragment from pRM50	"
pRM70 (Ap)	pUC18 + 2.1 kbp *Pvu*II-*Stu*I *xerC* frag.(ex pSD102)	"
pRM71 (Ap, Km)	pRM70 + 2.4 kbp *BamH*I fragment from pRM102	"
pRM80 (Cm)	pCB105 + 1.12 kbp *Hind*III-*EcoR*I *ruvC* fragment from pGS762	"
pRM90 (Ap)	*cer* *Mlu*I$^-$ derivative of pSD113	"
pRM91 (Ap)	pRM90 + wild type *cer* at *Pvu*II site (see pSD115)	"
pRM92 (Ap)	pSD113 + *Mlu*I$^-$ *cer* at *Pvu*II site (see pSD115)	"
pRM93 (Ap)	pRM90 + *Mlu*I$^-$ *cer* at *Pvu*II site (see pSD115)	"
pRM101 (Ap)	pSD102 + 230 bp *Pvu*II-*Sma*I *lac*PO in *Nru*I site	"
pRM102 (Ap, Km)	pRM101 + 1.8 kbp *EcoR*I Kanamycin-*rrn*B terminator cassette from pRM11	"
pRM120 (Ap)	GGG-AAT *cer* site derivative of pSD113	"
pRM121 (Ap)	pRM120 + wild type *cer* in *Pvu*II (see pSD115)	"
pRM122 (Ap)	pSD113 + GGG-AAT *cer* in *Pvu*II (see pSD115)	"
pRM123 (Ap)	pRM120 + GGG-AAT *cer* in *Pvu*II (see pSD115)	"
pRM130 (Ap)	pUC19 + 1.7 kbp *Hind*III-*EcoR*I *xprB* fragment from pJC763	"
pRM131 (Ap)	760 bp *Sac*II deletion derivative of pRM130	"
pRM132 (Ap)	pUC19 + 1.3 kbp *Hind*III-*Hae*II *xprB* fragment from pJC763	"
pRM133 (Ap, Gen)	pRM130 *Sac*II+ 2.0 kbp *Hind*III gentamycin resistance gene	"
pRM134 (Ap, Gen)	pRM130 + 2.0 kbp *Hind*III gentamycin resistance gene in *EcoR*V site	"
pRM135 (Cm)	pCB106 + 1.7 kbp *Hind*III-*EcoR*I *xprB* fragment from pJC763	"
pRM140 (Cm)	replacement of *xerC* in pRM60with 1.7 kbp *Hind*III-*EcoR*I *xprB* from pRM130	"

2.1 Bacterial strains and plasmids. The derivatives of *Escherichia coli* K-12 used in this work are listed in Table 2.1; new strains were constructed either by P1 transduction (Miller, 1972) or by linear transformation of mutant, cloned genes (Winans *et al*, 1985; see Chapter 4). The plasmids that were used and constructed for this work are listed in Table 2.2.

2.2 Chemicals

CHEMICALS	SOURCE
General chemicals, biochemicals and organic solvents	BDH, May and Baker, Sigma
Media	Difco, Oxoid
Agarose	BRL, Flowgen
IPTG	BRL
Radiochemicals	NEN
Nucleotides	Boehringer Mannheim, Promega
Antibiotics	Sigma
Val-Leu-NH_2	BioMac

2.3 Bacterial growth media and conditions. Bacteria were grown in standard L-broth (Miller, 1972) and L-agar at 37 ^{o}C, or in minimal agar (100 ml of which contained 25 ml D&M salts, 75 ml 2% agar in distilled water, 0.2% glucose, 20 ug/ml vitamin B1 and the following amino acids at 30 ug/ml: threonone, leucine, histidine, arginine and proline) at the same temperature. Antibiotics and other supplements were added where appropriate (see text). The bacterial strains were stored in 50% L-broth, 20% glycerol and 1% peptone at -20 ^{o}C, or in thiamine-containing slopes at room temperature.

The antibiotic concentrations used for both liquid and plate selection were as follows:

Antibiotic	Stock	Solution	Selective concentration
Ampicillin	5 mg/ml	water	50 ug/ml
Tetracycline	1.25 mg/ml	10 mM HCl	12.5 ug/ml
Chloramphenicol	2.5 mg/ml	ethanol	25 ug/ml
Kanamycin	5 mg/ml	water	50 ug/ml
Streptomycin	10 mg/ml	water	100 ug/ml
Trimethoprin	1 mg/ml	water	10 ug/ml
Gentamycin	1 mg/ml	water	10 ug/ml

All stocks were stored at 4 ^{o}C and were added to molten agar cooled to 55 ^{o}C.

2.4 Bacterial transformation. All transformations of plasmid DNA used standard $CaCl_2$ treatment of cells (Sambrook *et al*, 1989; Cohen and Hsu, 1972).

2.5 Isolation of plasmid DNA. Four methods for preparing plasmid DNA from bacterial cultures were used:

(i) Small scale DNA preparations used a modification of the boiling (STET) method of Holmes and Quigley (1981) that is in common use in our laboratory (Colloms, 1990); where stated, a phenol/chloroform extraction step was included before precipitation of the plasmid DNA with isopropanol. DNA was made from "patched" cultures of *E.coli* from L-agar plates, or by harvesting 3 ml of an L-broth culture at mid-log phase (e.g. in *in vivo* recombination assays) or at stationary phase.

(ii) Mid scale preparations used Qiagen columns and followed the manufacturers' instructions.

(iii) Large scale preparations for long-term storage of the DNA used a modification of the technique of Birnboim and Doly (1976) developed by C.Boyd (1985).

(iv) For rapid analysis of the plasmid content of a bacterial culture, the single colony lysis technique was used. In this, a 1 cm patch of a single transformant was collected and resuspended in 100-200 ul of single colony gel buffer (2% ficoll, 1% SDS, 0.01% bromophenol blue in Tris-acetate buffer). The cells were allowed to lyse at room temperature for 15 mins, cell debris and chromosomal DNA spun down in a microfuge for 30 mins at 4 ^{o}C and 30 ul of the supernatant loaded onto an agarose gel.

2.6 *in vitro* DNA manipulations. DNA manipulations were essentially as described in Sambrook *et al* (1989).

Restriction digestions were performed in 10-30 ul of 1X restriction buffer (purchased from BRL) and contained 0.3-3 ug of DNA. 5-30 units of restriction enzymes (purchased from BRL, Pharmacia, BIOLABS and Promega) were used in the digestions.

The ends of restricted DNA fragments were filled-in using either the Klenow fragment of *E.coli* DNA polymerase I or phage T4 polymerase (both purchased from BRL; see Sambrook *et al*, 1989).

DNA fragments were ligated for 2-5 hours at room temperature in 20 ul of 1X ligation buffer using 1-3 units of T4 DNA ligase (both purchased from BRL) before bacterial transformation.

End-labelling of DNA fragments was performed as described in Sambrook *et al* (1989) using the Klenow fragment of *E.coli* DNA polymerase I (purchased from BRL) and 10-20 uCi of $\alpha[^{32}P]$dATP/dCTP (purchased from NEN).

2.7 Site-directed mutagenesis of DNA. Two site-directed mutagenesis techniques were employed:

(i) The *pepA* gene was mutagenised after being cloned into pSELECT (purchased from Promega) following the instructions supplied. The repair-deficient strain EM1 was used in preference to BMH71-18.

(ii) Mutagenesis of the *cer* site in pSD113 followed a protocol developed by M.Stark (unpublished). Approximately 2 ug of plasmid DNA was nicked in 1X restriction buffer (containing 300 ug/ml ethidium bromide) using 2 ug/ml DNAaseI (Stark *et al*, 1991), purified by ethanol precipitation after phenol/chloroform extraction (Sambrook *et al*, 1989) and then denatured by boiling at 100 ^{o}C for 10 mins in 25 ul of 1X annealing buffer (purchased from Promega). The denatured DNA was placed on ice and 2 pmoles of both the mutagenic oligonucleotide and *Sca*I deletion oligonucleotide (kindly supplied by M.Stark) were added; annealing of the oligos was performed by incubating the DNA at 55 ^{o}C for 2 mins before it was cooled slowly to room temperature. The plasmid DNA was repaired from the oligonucleotides in 40 ul of 1X synthesis buffer (purchased from Promega) using T4 DNA polymerase and T4 ligase (purchased from BRL) following the procedures described for pSELECT mutagenesis (Promega). The repaired DNA was transformed into the either of the repair deficient strains EM1 or BMH71-18 and grown overnight. Qiagen preparations of the plasmid DNA were performed from overnight cultures and samples digested with a large excess of *Sca*I (25 units for 3 hours) before being transformed into DS941. Plasmid DNA was prepared from a number of the *Sca*I-resistant clones and the desired site-directed mutants selected by further restriction digestion.

The mutagenic oligonucleotides that were used are described in the text. The *Sca*I deletion oligonucleotide had the sequence: 5'GTGACTGGTGAGTATTCAACCAA GTCATTC 3'. All the oligos were synthesised on an Applied Biosystems PCR-mate oligonucleotide synthesiser and were not purified before use; they were phosphorylated using T4 kinase as described in the pSELECT mutagenesis protocol (Promega).

2.8 Gel electrophoresis of DNA. DNA was analysed on three kinds of horizontal agarose gels: standard 0.7-1.2% gels, 1.2% low melting point (LMP) gels and 1.2% alkaline denaturing gels. The DNA samples were applied to the gels in Ficoll-containing loading buffer (Sambrook *et al*, 1989). Standard and LMP gels were made and run in Tris-acetate buffer (40 mM Tris.Acetate [pH 8.2], 20 mM Na.Acetate, 1 mM EDTA). The preparation and running of the LMP gels was as described by the agarose manufaturers' instructions (FMC Bioproducts), whilst the standard and alkaline denaturing gels were as detailed in Sambrook *et al* (1989). Custom-made gel electrophoresis kits were used which made 23 cm long gels and contained 3 litres of buffer; the gels were routinely run for 15-18 hours at 1.5-2.5 volts/cm.

The DNA was visualised either by 254 n.m. UV illumination after staining with ethidium bromide, or by autoradiography (with and without vacuum drying of the gels). Ethidium bromide-stained gels were photographed on a Pentax 35 mm SRL camera fitted with a Kodak Wratten no.23A filter using Ilford HP5 film; autoradiographs used Fiji NIF RX Medical X-ray film.

2.9 Extraction of DNA from agarose gels. DNA was purified from standard agarose gels by low speed centrifugation through siliconised glass wool (Heery *et al*, 1990) and from LMP gels by phenol/chloroform extraction (following the agarose manufacturers' instructions).

2.10 Electron microscopy. Electron microscopic (EM) examination of purified DNA was performed essentially as described by Coggins (1987). The DNA was prepared for EM by spreading in 40% formamide, 1X TE (0.1 M Tris [pH 8.3], 10 mM EDTA), 0.1 mg/ml cytochrome c on a hypophase of 10% formamide, 0.1X TE. Contrast was enhanced by staining with uranyl.acetate, and low angle (8-10^{o}) shadowing was performed by vacuum evaporation of Platinum : Palladium (80 : 20) wire.

2.11 Gel electrophoresis of protein. PepA was analysed on 12% SDS-polyacrylamide gels as described by Laemmli (1970) using vertical kits purchased from BioRad. The protein samples were applied to the gel in a loading buffer that contained 5% SDS, 50% glycerol, 0.01% bromophenol blue, 50 mM Tris-HCl (pH 8.0) and 5% beta-mercaptoethanol. The proteins were visualised (after fixing in 40% methanol, 10% acetic acid) by staining with Coomassie blue in 50% methanol, 25% TCA.

2.12 Preparation of concentrated cell extracts for aminopeptidase assays. 100 ml cell cultures were harvested by centrifugation and resuspended in 20 mM Tris-HCl (pH 8.2), 100 mM KCl, 0.1 mM EDTA, 1 mM MgOAc. They were then lysed by the addition of T4 lysozyme at 4 ug/ml and repeated cycles of freezing in liquid nitrogen and thawing by incubation at 37 ^{o}C. Cell debris were removed by centrifugation.

2.13 *in vitro* aminopeptidase assays. The enzymatic substrate used was leucine-*p*-nitroanilide (purchased from Sigma). The assays were performed at 37 ^{o}C in 20 mM Tris-HCl pH 8.2, 100 mM KCl, 1 mM $MnCl_2$, 0.1 mM EDTA, 1 mM L-leucine-*p*-nitroanilide.1 ml samples were removed from a total assay volume of 14 ml at the times stated. The reaction was measured by the absorbance change at 400 n.m and the amount of reaction product (*p*-nitroaniline) in the assay was calculated using the extinction coefficient taken from the Handbook of Chemistry and Physics (53rd edition; $\log_{10}$ Extinction Coefficient is 4.19). Aminopeptidase units were expressed as nmoles of *p*-nitroaniline produced per milligram of protein per min.

Chapter 3

Analysis of the role of aminopeptidase A in *cer* site-specific recombination by site-directed mutagenesis

3.1 Introduction

The role that Aminopeptidase A (PepA) plays in the *cer* site-specific recombination reaction is not yet understood. The evidence that the protein is essential for *cer* recombination, and not required for *dif* recombination, is genetic. Strains containing Tn5 insertions within the *pepA* gene (e.g. CSX17) will not recombine reporter plasmids containing directly repeated *cer* sites, but will recombine similar plasmids if they contain *dif* sites (Stirling *et al*, 1989; Blakely *et al*, 1991); these Xer$^-$ mutations can be complemented by a minimal *pepA* clone. Because it has not been possible to recreate the *cer* recombination system *in vitro*, experiments to analyse the role of PepA have been limited (e.g. attempting to see if it has DNA binding properties; see below). This has meant that considerations regarding its function have been necessarily speculative, and have centred on the question of whether the protein is employed by the *cer* system for its catalytic properties (it is an amino-exopeptidase) or in a structural capacity (see Stirling *et al*, 1989). Definitive experimental examination of these alternative (though not exclusive) roles has so far proved elusive (G. Szatmari, H.O'Mara, pers.comms.; see Discussion).

Analysis of the proposed cellular functions of PepA does not offer any answers to these questions. Vogt (1970) isolated, characterised and crystallised a hexameric, 323 kDa aminopeptidase from *E.coli* whose catalytic activity was dependent on the divalent cation Mn^{2+}. This protein, called aminopeptidase I, is very likely to be PepA, since this heat-resistant peptidase activity is absent from DS941*pepA*::Tn5 strains (CSX17; Stirling *et al*, 1989) and *pepA* mutants selected by their resistance to the toxic dipeptide L-Valyl-L-Leucine-amide lack a protein of the size described for Pep I (Miller and Schwartz, 1978). The *in vitro* characterisation of PepA by Vogt demonstrated that it is an aminopeptidase that will processively cleave the amino-terminal residues from many different peptide molecules (in size and sequence) to produce free amino acids. This so-called "broad-specificity" of PepA is similar to other aminopeptidases from *E.coli* and *S.typhimurium* (PepN and PepB; see Chapter 1), but it should be noted that the

substrate profiles of these enzymes can be distinguished since they do not recognize peptides with every possible amino terminal residue (Miller, 1987). The aminopeptidases appear to be involved in the degradation of peptides derived from cleavage (by other enzymes) of abnormal or truncated proteins in the cell, and also in breaking down exogenous peptides which are transported into the cell so that they can be utilised as nutrient sources. This analysis is derived from metabolic examination of *S.typhimurium* strains mutant in aminopeptidases A, B, D and N; it is likely, however, that *E.coli* encodes the same set of enzymes (Miller, 1987; Miller and Schwartz, 1978).

Although there is no direct evidence, either from *in vitro* or *in vivo* experiments, that proteins are substrates for PepA, the possibility cannot be ruled out that it processes, and therefore activates, one of the *xer* proteins (i.e. ArgR, XerC, XprB or indeed a protein not yet identified). Activation would involve the cleavage of one or more N-terminal amino acids, and must be due to a substrate specificity for the Xer protein(s) possessed by PepA that the other cellular aminopeptidases do not have. No experiments have been able to identify any such amino-terminal cleavage of these proteins, but this has not been exhaustively analysed and remains possible. Since XerC and XprB are able to recombine type II hybrid and *dif* sites in the absence of functional PepA, this suggests that the putative recombinases are not the targets for PepA-activation. If this were how PepA functioned in *cer* recombination then the amino teminus of the target protein(s) must be free on the surface of the molecule to allow access to PepA; alternatively the cleavage may occur during the translation of the *xer* proteins.

An alternative target for aminopeptidase cleavage by PepA in *cer* recombination could be small peptide molecules, which are the natural substrates for this enzyme. These have not been identified, and are therefore purely hypothetical, but it is conceivable they could function either as activators that require cleavage or as inhibiors that are inactivated by cleavage. This possibility has the same problem as *xer* protein-cleavage in explaining why PepA alone of all the broad-specificity peptidases in the cell is active in this role.

Whilst there are difficulties in imagining how PepA acts in an enzymatic role during *cer* recombination, it is equally problematic to explain how it might act structurally. If PepA were involved in the nucleoprotein complex which forms during the recombination reaction (either directly binding to the *cer* site, or interacting with the other *xer* proteins) this would answer why PepA and no other peptidases are specifically involved in the reaction. However, no *cer*-specific DNA binding activity has been described for this protein (C.Stirling, pers.comm.), and no experiments have been performed to attempt to address the possibility of it forming protein-protein interactions with either ArgR, XerC or XprB (e.g. by gel binding assays). Clearly, however, the fact that these experiments yielded negative results cannot exclude these possible roles since they may be a consequence of having used the incorrect reaction conditions.

This chapter describes site-directed mutagenesis of the *pepA* gene in an attempt to analyse the above possibilities. Site-directed mutagenesis has been used to analyse the catalytic mechanisms of a number of enzymes (Leatherbarrow and Ferscht, 1896). The experiments in this study, however, were not intended as an analysis of the amino acid residues involved in the active site of PepA. Instead the idea was to target specific residues in PepA's active site in the hope of creating derivatives that lack any peptidase activity which could then be tested in their ability to support *cer* site-specific recombination. This experimental approach is therefore asking the question: is PepA's aminopeptidase activity required for *cer* recombination ?. The implicit assertion of this approach is that PepA acts as a structural component of the *xer* machinery, and does not utilise its enzymatic activity during the *cer* reaction. The reason that this hypothesis must be made is because a mutant PepA which is both Xer$^-$ and peptidase$^-$ is unable to show that PepA acts enzymatically, because a protein with both phenotypes may have structural perturbations as a result of the site-directed amino acid alterations.

3.1 Site-directed mutagenesis of *pepA*

The three dimensional structure of bovine lens leucine aminopeptidase (henceforth called BLLAP) published by Burley *et al* (1990) was used to design the mutagenesis strategy. The crystallographic structure of this enzyme (to 2.7A resolution) was reported at the start of this work, and later a more refined structure (as well as the structure of the protein complexed with the inhibitor bestatin) was reported (Burley *et al*, 1991 and 1992). BLLAP is an aminopeptidase which, like PepA, displays a wide substrate profile *in vitro*. The reason that the structure of this protein was believed to be the related to that of PepA is because of the high sequence conservation between the two enzymes (Stirling *et al*, 1989; see below). Described below are some of the important structural features of BLLAP (taken from the 3-D structure) which determined the PepA mutation that was made in this study.

BLLAP is active as a 324 kDa hexamer consisting of six identical subunits, each 487 amino acids in size. The protein contains 12 Zn^{2+} ions (two in each monomer) which are essential for its enzymatic activity; they can, however, be replaced by Mg^{2+}, Mn^{2+} or Co^{2+}, although each reduces the K_m of the enzyme. The N-terminal 137 amino acids of each BLLAP monomer can be cleaved using trypsin and the remaining C-terminal portions retain the hexamer structure and are catalytically active (van Loon-Klassen *et al*, 1979)

Isolated BLLAP monomers have an approximate "comma" shape, comprising an N-terminal and a C-terminal domain. The C-terminal domain is 327 amino acids in size, comprises the body of the comma and contains the proposed active site and its associated Zn^{2+} ions. This is consistent with photoaffinity labelling experiments that localised the active site to within the larger of the trypsin-cleaved fragments. The hexamer has 32 symmetry, in which the catalytic domains are clustered around the threefold symmetry axis. The 160 residue, N-terminal domains extend outwards, and away from, this catalytic core. It is believed that the residues responsible for trimerisation are within the C-terminal domain, whilst those involved in trimer-trimer

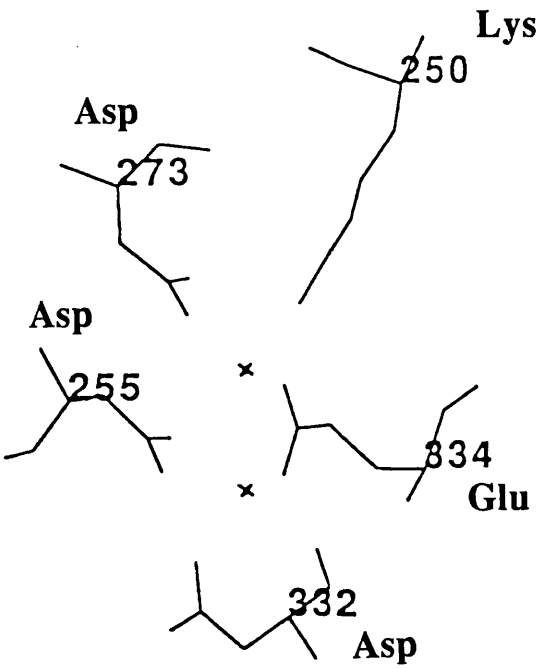


Figure 3.1. Stick figure representation of the active site residues of bovine lens leucine aminopeptidase. Each α-carbon atom is labelled and the two active site zinc ions are shown as crosses, the uppermost of which is in the readily exchangeable subsite (site 1).(Adapted from Burley *et al*, 1992).

interactions are in the N-terminal domain (this, however, disagrees somewhat with the fact that trypsin cleavage of the majority of the N-terminus does not destabilise the hexamer). The six active sites are located in the interior of the hexamer and line the walls of a disc-shaped cavity. Access to this interior cavity for substrate molecules is believed to be *via* three solvent channels that run along the 2-fold symmetry axis. The maximal dimensions of molecules wishing to gain access to the active sites are approximately 7A, meaning that it is that unlikely BLLAP could utilise proteins as substrates unless their N-termini are in an extended form; whether this would also be true for PepA depends on whether it is also active as a hexamer and whether it retains these structural features.

The active site residues of BLLAP (described below; fig.3.1) were identified on the basis of their proximity to the bound Zn^{2+} ions. Because two Zn^{2+} ions are used by each monomer the active site is described as consisting of two subsites. The subsites do not appear to be equivalent, however, since one (site 1, the activation site) will readily exchange the bound Zn^{2+} ion with a variety of other divalent cations (Mg^{2+}, Mn^{2+} and Co^{2+}), whereas the other (site 2) binds Zn^{2+} more tightly and will only accept other ions when it is unoccupied. The crystallographic structure suggested that the metal ion in site 1 is co-ordinated by one carboxylate oxygen atom from the residues Asp273 and Glu334, and by the side-chain amino group of Lys250 (see fig.3.1). In site 2 the co-ordination is achieved by one carboxylate oxygen from each of Asp255, Asp332 and Glu334, and by the carbonyl oxygen of Asp332. This illustrates that each Zn^{2+} ion is chelated by more than one amino acid residue, and that Glu334 is involved in the co-ordination at both subsites. Two positively charged residues (Lys262 and Arg336) are also described as being in the active site because they extend into the region around the Zn^{2+} ions. They are, however, not thought to be involved in metal ion chelation and their inclusion in the defined active site is mainly a consequence of their polar character and proximity.

It was important to ask at this stage whether the structure of BLLAP would allow a reasonable mutagenesis strategy to be extrapolated for PepA, for which no 3-D

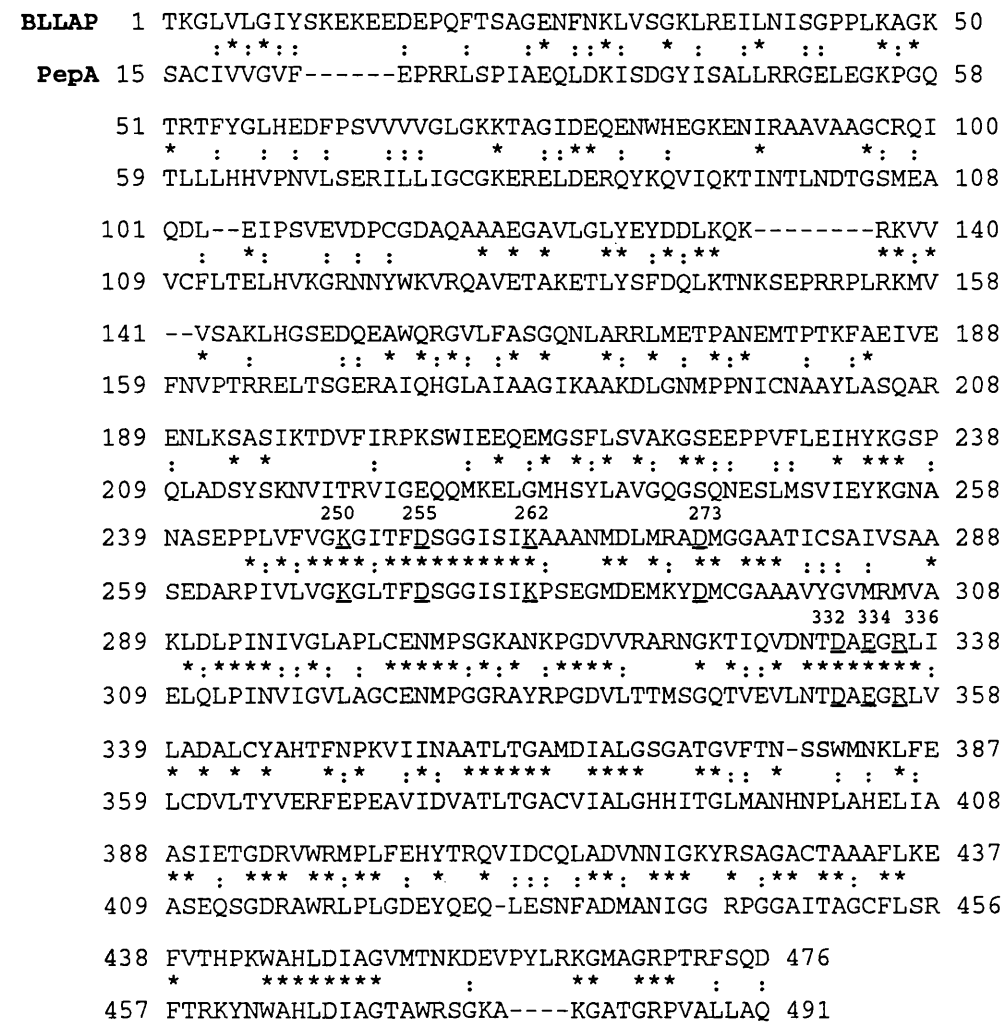


Figure 3.2. Comparison of the protein sequences of PepA and bovine lens leucine aminopeptidase. Gaps (-) have been introduced in the sequences to maximise homology. Identical residues are indicated by an asterisk (*) and conservative changes (within the exchange groups (V, L, I, F, Y, M, W) - (A, T, G, S, C) - (H, K, R) - (D, E, Q, N)) are shown by a colon (:). The seven active site residues are underlined and their amino acid position within the bovine enzyme's sequence detailed.

structural data are available (work is in progress to determine this, however; N.Isaacs and A.Lipscomb, pers.comms.). PepA has a molecular weight of approximately 55kDa (based on its constituent amino acids). In a low ionic strength buffer it aggregates and becomes too large to migrate into non-denaturing protein gels (Miller and Schwartz, 1978). This is consistent with the protein, like BLLAP, being hexameric in nature. An alignment of the protein sequences of PepA and BLLAP (see fig.3.2) reveals that they have an overall sequence identity of 31%, and this increases to 52% in the C-terminal portion of the proteins (residues 210 to 451 of BLLAP and 230 to 470 of PepA). This is a very high level of homology between proteins from such diverged organisms. The greater homology between the C-termini reflects the fact that N-terminal 160 residues of BLLAP can be cleaved off without compromising the activity of the enzyme, and may indicate that the catalytic core of PepA also resides in the C-terminus. Figure 3.2 also shows that all the seven residues implicated in the active site of BLLAP are conserved in PepA. It should be noted, however, that PepA is a Mn^{2+}-dependent aminopeptidase, and is inhibited by Zn^{2+} (Stirling *et al*, 1989); it must therefore be assumed that the residues (described above) which chelate Zn^{2+} in BLLAP perform the same function with Mn^{2+} in PepA. Finally, it has been shown that the compound bestatin is capable of allowing *E.coli* transformed with a PepA expression vector to grow on media containing the toxic dipeptide Val-Leu-NH_2 (G.Szatmari, pers.comm.; see section 3.5). Bestatin is a slow binding inhibitor of BLLAP *in vitro* and its co-structure with the enzyme has recently been solved (Burley *et al*, 1992); its ability to inhibit PepA further supports the belief that the structures, and perhaps catalytic mechanisms, of the two proteins are conserved.

It was decided that the residues in PepA that putatively chelate Mn^{2+} (which were implicated by the BLLAP Zn^{2+}-co-ordination residues) would be targetted in the site-directed mutagenesis. Lys262 and Arg336 were decided against because the detailed catalytic mechanism used by these enzymes had not been analysed, and consequently the involvement of these residues was speculative. In contrast, it was known that binding of metal cations was essential for the activity of both enzymes, and the co-

ordination residues therefore offered the best possibility of creating a peptidase-deficient PepA derivative.

The residue Glu354 in PepA was chosen as the single amino acid to be altered in this experiment. It was targetted because it is PepA's equivalent of Glu334 in BLLAP (fig.3.2). Since it is the only residue implicated in the chelation of both Mn^{2+} ions, its alteration was thought to be likely to cause the largest reduction in the catalytic activity of PepA. Only one mutant was made, for two reasons. This experimental approach relies on creating a PepA mutant which is peptidase-deficient, because a mutant that is Xer^+ but has detectable residual catalytic activity cannot determine that PepA is used in a purely structural role in the *cer* recombination reaction (as noted a $peptidase^-$, Xer^- mutant is a null result in this approach). Because multiple residues are used in the co-ordination of metal ions in PepA/BLLAP it is possible that mutating only one residue would be insufficient to make such a mutant, and therefore the E354 mutant of PepA could be used as a source to create double mutants with even greater reductions in catalytic activity. The other reason for this cautious approach was the fact that the PepA mutagenesis was based on the structural data of BLLAP. Since it is possible that the enzymes in fact do not have the the same structure, a single ("test") mutant was made which could be subsequently compared to mutations in the other residues if necessary.

The mutagenesis strategy is diagrammed in figure 3.3 and used the pSELECT plasmid and protocols developed by Promega (Materials and Methods). The gene encoding PepA was cloned as 1.9 kbp *Hind*III fragment from pCS126 (Stirling *et al*, 1989) into pSELECT, creating pRM20. This allowed the production of single-stranded DNA which served as a *pepA* template for mutagenesis using the oligonucleotide described (fig.3.3). The *pepA* insert in pRM20 was orientated by *EcoR*V restriction (which cuts once once in *pepA* and once in pSELECT) and, as shown, the gene is transcribed from the T7 Polymerase promoter. The mutagenesis reaction was performed following the pSELECT suppliers' instructions.

Mutation by the oligonucleotide in this experiment has two consequences. Firstly, it creates a novel *Pst*I site in the *pepA* gene. This meant that the mutation in the gene

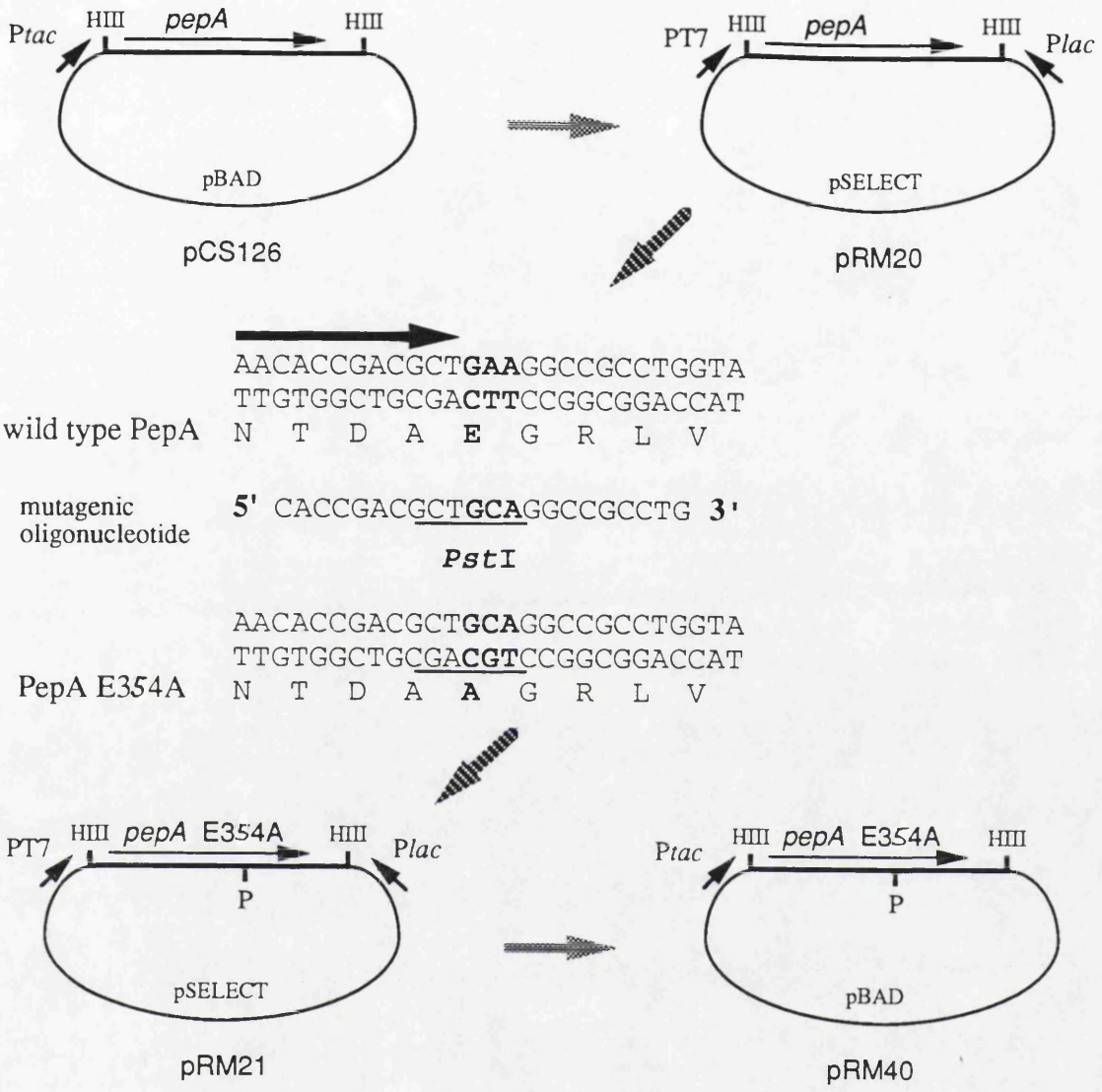


Figure 3.3. Strategy used in the site-directed mutagenesis of *pepA*. Cloning steps are represented by (➡), the mutagenesis step is represented by (➡). The details of the pSELECT and pBAD vectors are discussed in the text. The sequence shown represents part of the *pepA* gene and transcribed protein; it is taken from Stirling *et al* (1989) and centres on the region of the gene that has been mutagenised. The sequences in bold text are the codon that has been altered and the resultant PepA amino acid change. Note that the mutation made introduces a *Pst*I recognition site.

could be directly selected by *Pst*I restriction of the ampicillin resistant clones made during the mutagenesis, and it was not necessary to sequence the mutant pRM20 derivative that was chosen - called pRM21. The second consequence was that it changed the glutamic acid residue at position 354 in PepA to an alanine. Alanine was chosen as the replacement residue because the lack of a polar side chain at position 354 should mean that the active site of PepA becomes disrupted in its ability to chelate Mn^{2+}, and should therefore be reduced in its catalytic competence. Furthermore, the smaller size of alanine relative to glutamic acid should not structurally perturb the protein and therefore not compromise its putative structural role in *cer* recombination.

The 1.9 kbp insert of pRM21 was next cloned into *Hind*III-restricted pBAD, and its orientation selected (by *Eco*RI and *Pst*I digestion) such that *pepA* is transcribed from the vector's *tac* promoter. The plasmid that was created in this way is called pRM40 and is exactly equivalent to the expression vector pCS126 except that it expresses the E354A derivative of PepA rather than wild type (the *Pst*I restriction confirmed that pRM40 contained a *Pst*I recognition site at the correct position in *pepA*). This plasmid allowed the mutant PepA to be purified (by IPTG induction of *pepA* expression; see below) to determine if the Glu to Ala mutation had altered its catalytic activity, and allowed complementation assays to be performed to determine if PepA E354A is Xer^+ or Xer^-.

3.3 Determination of the aminopeptidase activity of PepA E354A *in vitro*

Wild type and mutant PepA (expressed from pCS126 and pRM40 respectively) were purified from the strain CSX17 (DS941*pepA*::Tn5) by Mary Burke, using an adaptation of the methods described by Vogt (1970) and Stirling *et al* (1989). A heat step (70 ^oC for 5 mins) can be used in these purifications to distinguish PepA activity from the other, non-heat stable, cellular aminopeptidases. However, because the Glu to

Figure 3.4. 12% SDS-polyacrylamide electrophoresis of purified wild type and mutant PepA. The proteins were over-expressed (from the *tac* promoter) by IPTG induction of the plasmids pCS126 and pRM40 in the strain CSX17. One sample of wild type PepA was purified following the technique described by Vogt (1970) and involved a 70 ^{o}C heat step. The other wild type PepA sample and the mutant (E354A) PepA sample used an adaptation (similar to Stirling *et al*, 1989) of this technique in which no heat step was used. The positions of the molecular size markers (in kDa) are indicated to the left.

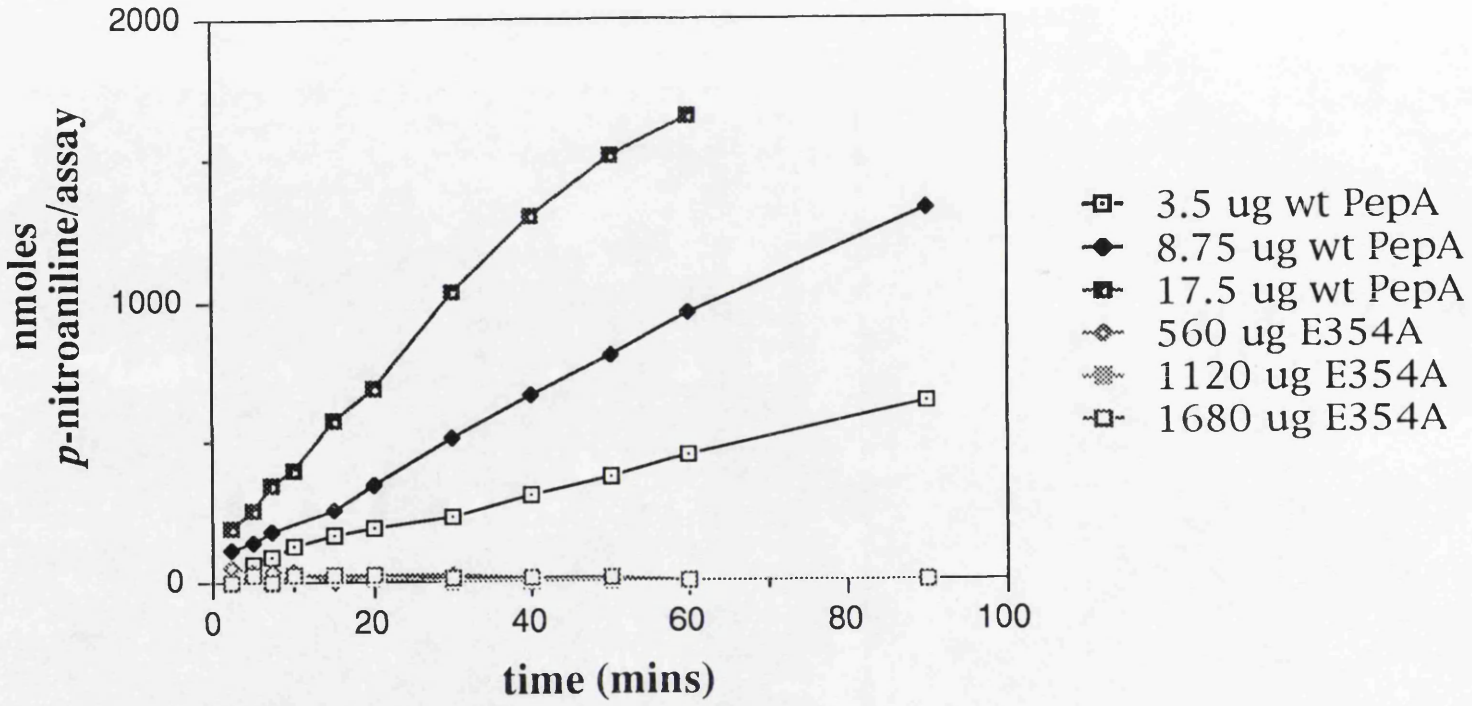


Figure 3.5. Comparison of the *in vitro* peptidase activities of purified wild type and mutant (E354A) PepA. The peptidase assays used leucine-*p*-nitroanilide as a substrate and the activities are shown as the amount of *p*-nitroaniline product made at three concentrations of the two enzymes. Product concentrations were spectrophotometrically calculated at 400 n.m. using the previously determined extinction coefficient for *p*-nitroaniline (1.55×10^4; 53rd edition of the Handbook of Chemistry and Physics).

nmoles of *p*-nitroaniline / assay

sample time (min)	amount of wild type PepA/assay			amount of PepA E354A/assay		
	3.5 ug	8.75 ug	17.5 ug	560 ug	1120 ug	1680 ug
2.5	N/D	120	190	50	0	0
5.0	60	140	250	40	20	20
7.5	90	180	340	40	0	0
10	130	N/D	400	40	20	20
15	160	260	570	0	20	20
20	190	350	690	30	10	20
30	230	510	1030	30	0	10
40	310	660	1300	10	0	10
50	370	800	1520	10	0	10
60	440	950	1650	0	0	0
90	640	1320	N/D	0	0	0

specific activity (nmoles of *p*-nitroaniline/ mg protein/ min)

2610	1980	2110	N/D	N/D	N/D

N/D = not determined

Figure 3.6. Determination of the specific activity of wild type PepA and PepA E354A. The upper table shows the amount of reaction product (*p*-nitroaniline) present at various times in *in vitro* peptidase assays using three concentrations of the two enzyme preparations, and leucine-*p*-nitroanilide substrate; it is graphed in figure 3.5. The lower table shows the specific activities of the three assays for each protein (expressed as nmoles of product per mg protein per min) calculated as an average of the specific activities from each time point and the slope of the graph in figure 3.5.

Ala change in PepA E354A may result in a temperature sensitive mutant no such step was utilised in this purification; the maximum temperature that the proteins were exposed to was 37 ^{o}C as the *E.coli* culture was grown. Figure 3.4 shows a SDS-polyacrylamide gel of the purified proteins, and compares them to PepA previously prepared following the Vogt protocol. This shows that both the wild type and mutant proteins were purified to near homogeneity in this preparation, and that they both appeared to be the same size (approximately 55kDa) as heat-purified PepA. This suggests that the mutation made in pRM40 has not introduced a stop codon in *pepA*'s open reading frame or a protease cleavage site in PepA, both of which would result in production of a truncated enzyme.

Aminopeptidase assays were performed on the purified enzymes using leucine-*p*-nitroanilide (purchased from Sigma) as a substrate (see Materials and Methods). Figure 3.5 shows a graph of the spectrophotometrically determined amounts of reaction product (*p*-nitroaniline) made over time in these assays. The assays were performed using three different concentrations of both wild type and mutant PepA, and, for comparison, the result is shown in tabular form along with the specific activities calculated from the assays (fig.3.6).

As expected, wild type PepA produced increasing amounts of *p*-nitroaniline with time, and the rate of this production increased as the concentration of the enzyme was increased. In contrast, no reaction product was detected in the assays involving the mutant PepA preparation. It should be noted that the concentrations of PepA E354A used in this experiment were substantially higher than the concentrations of wild type PepA: 0.25, 0.62 and 1.25 ug/ml of wild type; and 40, 80 and 120ug/ml of mutant (i.e. approximately 100X more mutant than wild type in each assay). The product concentrations from each reaction time point, and the slopes of the graph in figure 3.5, were then used to calculate specific activites for each concentration of wild type PepA. These were compared and the average specific activity for the wild type preparation was calculated as 2230U/mg of protein, which is higher than the activity determined previously for PepA (1300U/mg; Stirling *et al*, 1989, and Szatmari pers.comm.). The

lack of reaction products seen using PepA E354A meant it was not possible to determine a specific activity for this protein preparation. This experiment was repeated on two separate occasions by myself and independently by Mary Burke, and comparable specific activities for wild type PepA were determined in each case. Similarly, no detectable activity was seen using the mutant PepA preparation in these assays - even when the reaction was allowed to proceed for three hours.

These experiments suggest that the mutant PepA used in these assays is highly deficient in aminopeptidase activity. It was not possible to quantify the amount that the catalytic activity is reduced in this protein preparation since no reaction products could be detected in the assay used. The lack of reaction product suggests that PepA E354A is unable to catalyse the hydrolysis of leucine-*p*-nitroanilide, or at least that any residual activity in this enzyme preparation is beyond the limits of detection of this assay.

It seems unlikely that the inactivity of the E354A mutant PepA preparation is simply a result of the treatment of this particular purified sample. This can be concluded because the wild type enzyme was prepared in the same manner and at the same time and was highly active, and also because the enzyme appeared to be the correct size on protein gels. This therefore suggests that the lack of aminopeptidase activity in the mutant is a result of the site-directed alteration to PepA's amino acid sequence. The glutamic acid to alanine change (or indeed other chance, uncharacterised changes made during the mutagenesis; see discussion) could have this effect either by perturbing the structure of PepA or by specifically disrupting the integrity of the active site.

Structural perturbations of the protein could either be general conformational changes because E354 is involved in organising the tertiary structure of PepA, or could be local changes to the active site region which make the mutant unable to recognise leucine-*p*-nitroanilide. The latter possibility is a problem in this analysis because such local changes may simply be confined to leucine-*p*-nitroanilide as a substrate, and other (natural) substrates may still be readily processed by PepA E354A (this is considered more fully in later sections). Available evidence from other experiments suggests that conservative site-directed changes in protein sequence (where the new amino acid's

side chain occupies less space than the natural residue and does not form novel interactions) rarely cause widespread conformational alterations (see e.g. Leatherbarrow and Ferscht, 1986; Brodo and Argos 1990). However, since there is no satisfactory method of determining whether a given site-directed change has this effect (short of solving the 3-D structure of the mutant protein), this possibility cannot be excluded in explaining the reduction in peptidase activity of PepA E354A.

The importance of the above result as regards analysing the role of PepA in *cer* site-specific recombination is that the E354A mutant is as depleted in peptidase activity as can be detected in *in vitro* assays. For this reason it was not felt necessary to create further PepA mutants before analysing the Xer phenotype of this PepA derivative.

3.4 Determination of the *xer* phenotype of PepA E354A *in vivo*

Ideally the ability of this PepA E354A to support *cer* recombination would have been tested *in vitro*, since the same reaction conditions employed for determining its peptidase activity could have been used in *cer* recombination assays and a strong correlation between the two properties could have been derived. Unfortunately, the *cer* reaction has not yet been recreated *in vitro* and it was therefore necessary to determine the *xer* phenotype of PepA E354A *in vivo*. This was achieved by doubly transforming pRM40 and pCS202 (a *cer* reporter plasmid; see fig.4.21), or pCS126 and pCS202, into DS941*pep*A::Tn5 (CSX17). The transformants were selected on media containing both ampicillin and chloramphenicol, and their DNA was analysed by boiling preparations and agarose gel electrophoresis (fig.3.7). This showed that both pCS126 and pRM40 were able to complement the chromosomal *pepA*::Tn5 mutation and allow the complete recombination of pCS202 to its resolution product (pCS203, which is indicated). Restriction of the DNA samples with *PstI* confirmed that pCS126 contained wild type *pepA* and pRM40 contained *pepA*E354A. This suggests that the E354A PepA mutant, which is peptidase$^-$ *in vitro*, is Xer$^+$ *in vivo*.

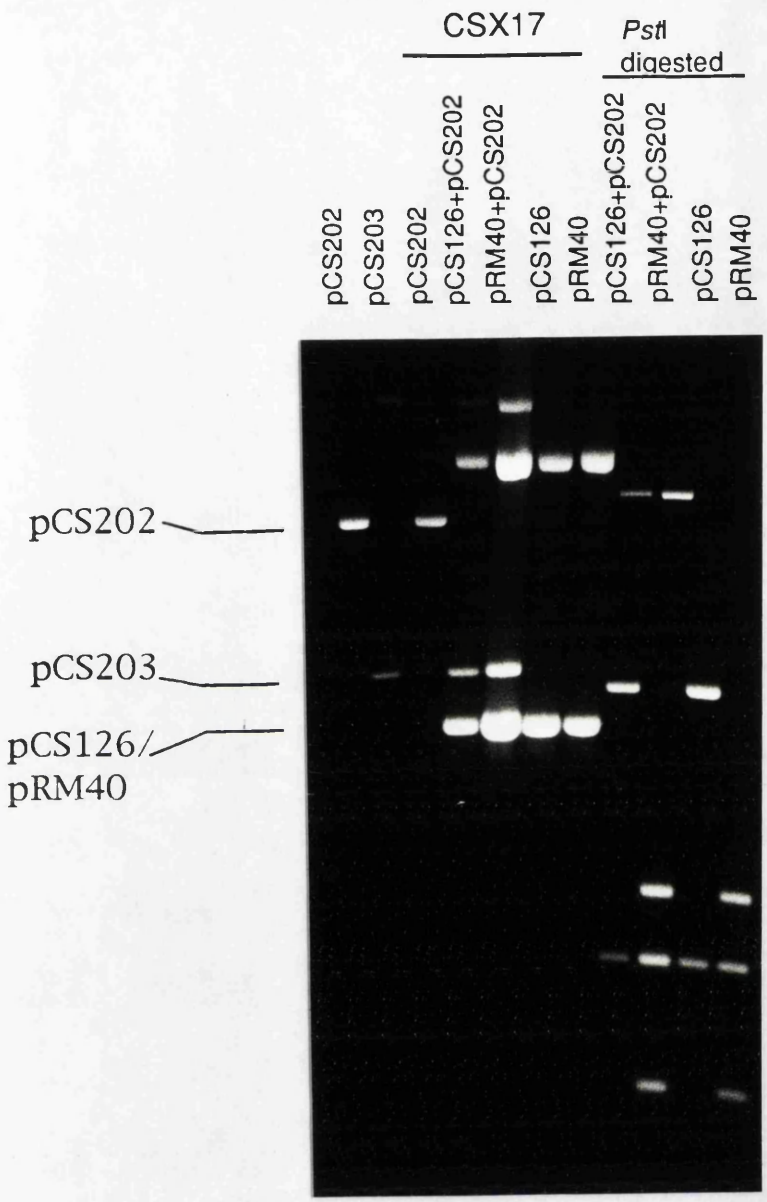


Figure 3.7. Complementation of an Xer⁻ *pepA*::Tn5 insertion by wild type PepA and PepA E354A. DS941*pepA*::Tn5 (CSX17) was transformed with pCS202 alone, with pCS202 and pCS126 or with pCS202 and pRM40. Plasmid DNA was isolated and run on a 1.2% agarose gel before and after *Pst*I restriction. The reporter plasmid pCS202 is resolved in Xer$^+$ strains to give the plasmid pCS203. pCS126 strongly expresses wild type PepA from the *tac* promoter; pRM40 is the same plasmid as pCS126, but expresses PepA E354A. *Pst*I restriction identifies the site-directed sequence alteration to the *pepA* gene that distinguishes the two plasmids.

Interpretation of this experiment is complicated by the numerous potential differences between the analysis of the activities of PepA E354A *in vitro* and *in vivo*. One possible interpretation of the fact that the mutant appears to be Xer^+ is that the peptidase deficiency described *in vitro* is reflected *in vivo*, and therefore the role that PepA plays in the *cer* recombination reaction does not require that PepA be active as an aminopeptidase, or that it has other catalytic activities. A number of objections to this are possible, however, and all would mean that this experiment could not determine whether the peptidase activity of PepA is required for the *cer* reaction:

(i) It is conceivable that the description of PepA E354A as being $peptidase^-$ is only due to the insensitivity of the leucine-*p*-nitroanilide assay, and *in vivo* there is sufficient enzymatic activity from the protein expressed from pRM40 to allow *cer* recombination.

(ii) The lack of activity *in vitro* when using PepA E354A may have been overcome in the different *in vivo* reaction conditions. This could, for instance, have been due to local structural alterations resulting from the glutamic acid to alanine change being overcome in the cell.

(iii) It is possible that the mutant PepA has been specifically altered in its ability to recognise the substrate analogue leucine-*p*-nitroanilide, and this does not reflect any change in the activity of the enzyme on other substrates. For instance, it is possible that PepA has a capacity to specifically recognise one of the *xer* proteins and this enzymatic activity is unaltered in the E354A mutant.

3.5 Determination of the peptidase activity of PepA E354A *in vivo*

*E.coli*K12 is sensitive to valine and valine-containing peptides when grown on minimal medium. This is because valine is feedback inhibitor of both its own synthesis and that of isoleucine, and therefore if isoleucine is not included in the medium the cells cannot grow (Miller, 1987). This property has been exploited by using various valine-

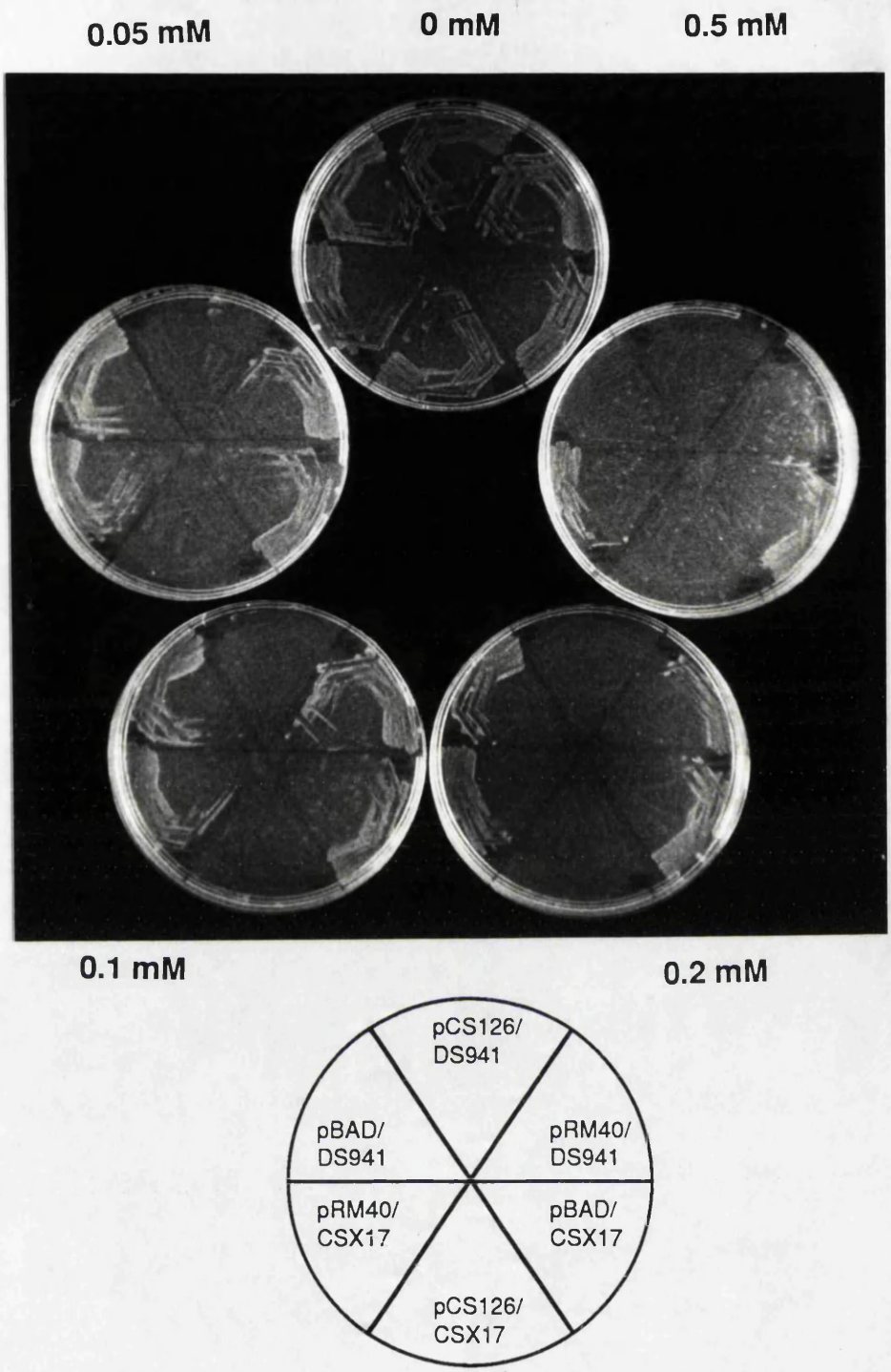


Figure 3.8. Determination of the peptidase activities of wild type PepA and PepA E354A *in vivo*. Both DS941 and DS941*pepA*::Tn5 (CSX17) were transformed with pBAD, pCS126 or pRM40 and grown for approximately 15 hours on minimal agar containing with either 0, 0.05, 0.1, 0.2 or 0.5 mM Val-Leu-NH_2. The organisation of each plate into six sectors is diagrammed and the concentration of Val-Leu-NH_2 in each plate is indicated. Val-Leu-NH_2 is cleaved by aminopeptidase A and releases valine, which is toxic to *E.coli*. pCS126 expresses wild type PepA, pRM40 expresses PepA E354A and pBAD is the vector used in these plasmids.

containing peptides to select for *E.coli* derivatives that were mutant in aminopeptidases and in peptide uptake (valine resistant mutants; Miller and Schwartz, 1978). The molecule Val-Leu-NH_2 was demonstrated to be a relatively specific substrate for aminopeptidase A in *E.coli* because all strains isolated which were resistant to this dipeptide but still sensitive to valine were mutant in *pepA*. Val-Leu-NH_2 has also been used more recently in our laboratory to attempt to select for peptidase$^-$, Xer$^+$/Xer$^-$ mutants (H.O'Mara, pers.comm.; see discussion). These experiments suggested that the dipeptide would be a suitable molecule to determine whether the peptidase deficiency of PepA E354A is seen *in vivo*, and therefore to address some of the considerations discussed in section 3.4.

This analysis was performed by transforming pCS126 (*pepA* wild type), pRM40 (*pepA*E354A) and pBAD (expression vector alone) into both DS941 and DS941*pepA*::Tn5 (CSX17) and growing the cells overnight in L-agar supplemented with ampicillin. Transformants from each plate were then streaked onto minimal media (containing ampicillin and the amino acid supplements detailed in materials and methods) containing the following concentrations of Val-Leu-NH_2: 0, 0.05, 0.1, 0.2 and 0.5 mM. The cells were allowed to grow overnight before being photographed (fig.3.8).

Without Val-Leu-NH_2 all the cells grew equally well. Increasing concentrations of Val-Leu-NH_2 were generally toxic to both DS941 and CSX17 cells containing pBAD, since they grew progressively more poorly. This suggests that the dipeptide can infact be cleaved in the absence of functional PepA, presumably by the other aminopeptidases. However, at 0.5 mM Val-Leu-NH_2, pBAD/CSX17 grew while pBAD/DS941 did not. This shows that this assay is able to detect the amount of PepA made from the single chromosomal *pepA* gene in wild type cells and distinguish that amount of activity from the peptidase activity in *pepA* mutant strains; it also illustrates that PepA is the enzyme that predominantly cleaves this dipeptide in the cell. Both strains grew very poorly when they contained pCS126; they did not grow detectably even at 0.05 mM Val-Leu-NH_2. This suggests that the amount of peptidase activity in

these cells is substantially higher than cells without the expression vector, and that the level of expression of PepA from pCS126 more than compensates for the loss of the enzyme in CSX17. Both DS941 and CSX17 containing pRM40 grew as well as the same cells containing the pBAD vector alone; there was no evidence of the expression of PepA from this plasmid causing the cells to efficiently cleave Val-Leu-NH_2 as in pCS126-containing cells. Significantly there was as much growth of pRM40/CSX17 at 0.5 mM Val-Leu-NH_2 as there was of pBAD/CSX17. This suggests that there is no more peptidase activity in cells containing pRM40 than there is in *pepA*::Tn5 mutants, despite the fact that PepA should be expressed to the same high levels as in the cells containing pCS126. Furthermore, three separate repetitions of this experiment showed some growth of pRM40/DS941 at 0.5 mM Val-Leu-NH_2 whilst there was no growth of pBAD/DS941. This may be a consequence of PepA E354A monomers expressed from pRM40 forming heterohexamers with wild type PepA monomers expressed from the chromosome, and resulting in a population of hexameric PepA in the cells which cannot cleave the toxic dipeptide.

This experiment confirms that the lack of PepA E354A-catalysed hydrolysis of leucine-*p*-nitroanilide seen *in vitro* reflects a more general reduction in the mutant's enzymatic activity. This is important because it rules out the possibility that the Xer^+ phenotype of this protein *in vivo* is a consequence of the *in vitro* $peptidase^-$ phenotype being overcome by the different cellular reaction conditions. The fact that PepA E354A is reduced in its ability to cleave two different substrates suggests that the mutation might have a general effect. This, however, cannot exclude the possibility that the mutant enzyme is still capable of processing other substrates (e.g. *xer* proteins).

A difficulty in correlating the reduction of enzymatic activity *in vitro* with the *xer* activity *in vivo* was the possibility that pRM40 expressed sufficient protein for the level of enzymatic activity to be high enough in the cells to allow *cer* recombination (assuming that PepA acts enzymatically in the *xer* reaction). Within the levels of its sensitivity, this experiment suggests that pRM40 expresses no more peptidase activity than a *pepA*::Tn5 strain (CSX17). This means that, at least as regards Val-Leu-NH_2, a

PepA protein (in this case PepAE354A) can be expressed in *in vivo* assays where it has no more peptidase activity than an Xer$^-$, *pepA* knockout strain and it can support *cer* recombination. The problem with concluding that this means that PepA has a structural role in *cer* recombination is that these assays may not be sensitive enough to detect the residual amount of peptidase activity that is allowing the *cer* reaction to proceed, and the fact that the mutant enzyme may still have full enzymatic activity when it is directed against substrates for which it has a higher specificity.

3.6 Complementation in CSX17 using pRM21

It was intended that this analysis would be concluded by replacing the wild type DS941 *pepA* gene with the *pepA*E354A site-directed mutant. Construction of such a strain would have allowed the mutant *pepA* to be transcribed from the gene's natural promoter, and therefore would have allowed *cer* recombination assays to be performed where the mutant enzyme was expressed at wild type levels in the cell. This would have meant that any residual peptidase activity from PepA E354A would be reduced below that made from pRM40, and therefore if the protein was being used as a enzyme in the *cer* reaction this may detect it. To do this the plasmids pRM41 and pRM45 were constructed. These are derivatives of pRM40 and pCS126 respectively, in which a kanamycin resistence gene is inserted into a unique *Stu*I recognition sequence between the *pepA* gene and downstream *orf13* (see Colloms, 1990 and Stirling *et al*, 1989 for details of this open reading frame). The plasmids were linearised by restriction with *BamH*I and transformed into the strain JC7623 (the use of JC7623 for strain construction is discussed in Chapter 4; see also Winans *et al*, 1985). It was hoped that this linear transformation would allow the plasmid-borne (kanamycin resistant) *pepA* genes to exchange with the chromosomal copies of the genes by homologous recombination, thereby creating a selectable strain containing the *pepA*E354A gene in the natural chromosomal location of *pepA*, as well as a control strain. Several attempts failed to yield any kanamycin resistant JC7623 derivatives (perhaps because of the

strain extract	specific activities (nmoles *p*-nitroaniline/mg extract/min) no heat treatment	70 oC heat treatment
DS941	16.6	15.4
CSX17	4.7	0
pCS126/CSX17	95.2	91.2
pRM40/CSX17	5.1	0
pRM20/CSX17	17.8	17.6
pRM21/CSX17	9.9	0

Figure 3.9. Aminopeptidase activities of cell extracts of *pepA* wild type and mutant strains, and strains containing plasmids expressing wild type PepA or PepA E354A. DS941*pepA*::Tn5 (CSX17) was transformed with pRM20, pRM21, pCS126 or pRM40. 100 ml cultures of these transformants, and of DS941 and CSX17, were grown overnight at 37 ^{o}C and concentrated cell extracts prepared. As in figures 3.5 and 3.6 the activities were determined using leucine-*p*-nitroanilide as a substrate and the amount of *p*-nitroaniline product was calculated spectrophotometrically. The specific activities are expressed as nmoles of product per mg of extract per min, and are the average that was determined for two concentrations of the various extracts, before and after heating to 70 ^{o}C for 5 mins.

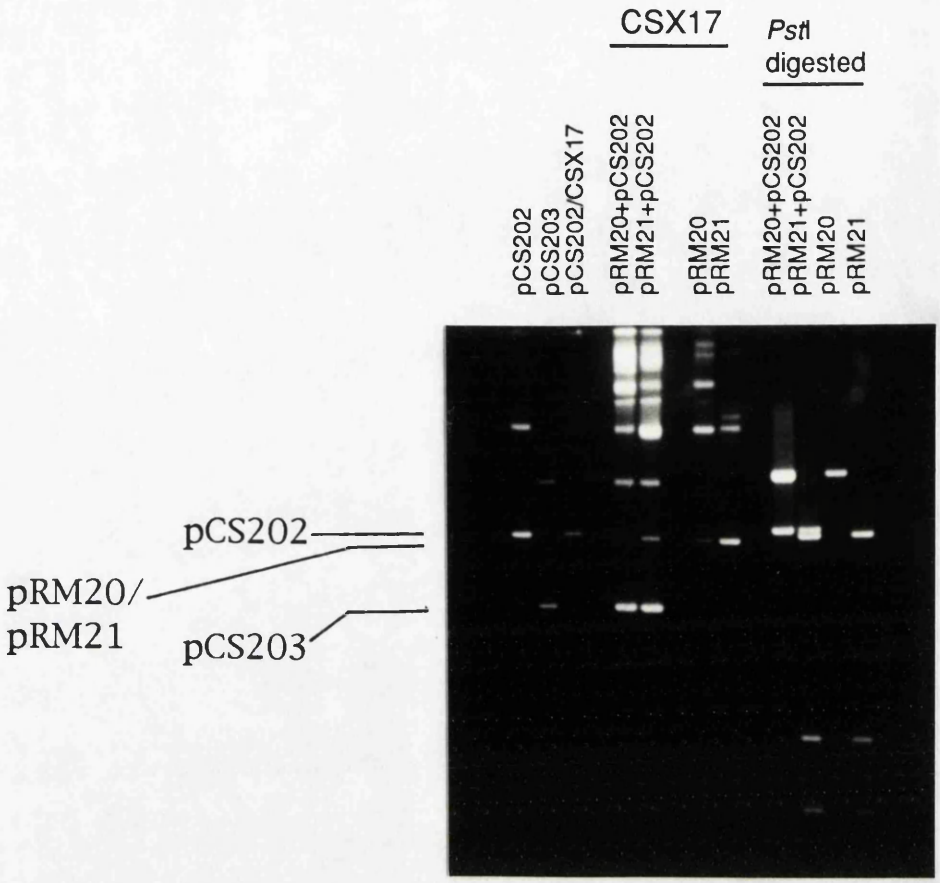


Figure 3.10. Complementation of an Xer⁻ *pepA*::Tn5 insertion by pRM20 and pRM21. DS941*pepA*::Tn5 (CSX17) was transformed with either pCS202, pRM20 or pRM21. The strains containing pRM20 and pRM21 were then transformed with pCS202. The plasmid DNA was run on a 1.2% agarose gel, before and after digestion with *Pst*I. pCS202 is resolved by *xer* recombination to give pCS203. pRM20 expresses wild type PepA, pRM21 expresses PepA E354A (the plasmids are distinguished by *Pst*I digestion); the genes encoding the two proteins are poorly transcribed from the T7 polymerase promoter in both plasmids, and therefore the levels of enzyme in the cells are lower than in cells containing pCS126 and pRM40 (see fig.3.9).

disruption to the expression of *orf13*, whose function is not understood) and therefore it was not possible to perform this experimental analysis.

In the plasmids pRM20 and pRM21 the expression of the *pepA* genes is driven by the promoter from bacteriophage T7 (fig.3.3). Because this promoter is not recognised by *E.coli* RNA polymerase (Tabor and Richardson, 1986) the transcription level of the *pepA* genes is likely to be very poor, and therefore these plasmids offer the opportunity to express PepA E354A at low levels in the cell and perform an equivalent experiment to the strain construction described above. To determine the level of PepA expressed from the plasmids, crude cell extracts were prepared from 100 ml cultures of the following strains: DS941, CSX17, pRM20/CSX17, pRM21/CSX17, pCS126/CSX17 and pRM40/CSX17 (Materials and Methods). Aminopeptidase assays were then performed (using leucine-*p*-nitroanilide as a substrate) on the extracts before and after heating to 70 ^{o}C for five mins. The specific activities determined from these assays are tabulated in figure 3.9.

CSX17 (DS941*pepA*::Tn5) displayed a low amount of peptidase activity which was reduced to 0 U/mg of extract after heating. This is because PepA is the only heat stable aminopeptidase in *E.coli*K12. The activity seen before heating represents cleavage of leucine-*p*-nitroanilide by the other cellular aminopeptidases. DS941 had approximately 15 U/mg extract after heating; this agreed with the activity determined previously (Stirling *et al*, 1989) and represents the activity derived from chromosomally encoded PepA. Both before and after heat treatment, the cells containing pCS126 showed a high level of peptidase activity (approx. 96 U/mg extract) that was absent from pRM40-containing cells (this activity is lower than the 600U/mg determined by Stirling *et al*). This is in agreement with the other experiments of this chapter which suggested that pCS126 expresses PepA to levels considerably exceeding that made from chromosomal *pepA*, and that pRM40 expresses PepA E354A which has undetectable peptidase activity. When wild type PepA was expressed from pRM20 the amount of activity was approximately 17 U/mg extract, which was approximately 10-fold lower than pCS126-containing cells and was comparable to DS941. This suggests

that the amount of PepA made from pRM20 is little more than is made from the chromosomal *pepA* gene, and consequently the level of any peptidase activity expressed from pRM21 will be 5-fold lower than from pRM40.

To determine the *xer* activities of pRM20 and pRM21 they were transformed into CSX17. After selecting for the presence of these plasmids by growth on tetracycline, the reporter plasmid pCS202 was transformed into the cells and double transformants were selected on media containing chloramphenicol and tetracycline. It was necessary to establish the expression vectors in CSX17 before transforming in pCS202 because the resistance determinant of the reporter plasmid that is lost during *cer* resolution is tetracycline. The DNA from the double transformants was prepared by the boiling technique and analysed by agarose gel electrophoresis (fig.3.10). Both pRM20 and pRM21 were able to complement the *pepA*::Tn5 mutation in CSX17 and allow the complete recombination of pCS202. This suggests that the PepA E354A is Xer^+ even when expressed at low levels that are comparable to the expression of PepA from the chromosome, and therefore if PepA acts enzymatically during the *cer* reaction the low levels of peptidase activity expressed from pRM21 are sufficient to achieve this function.

3.7 Discussion

The experiments in this chapter describe the catalytic properties of a single site-directed derivative of *E.coli* aminopeptidase A, and utilise this mutant enzyme to analyse the function of PepA in the *cer* site-specific recombination reaction. In the mutagenesis, the amino acid residue Glu354 was replaced by alanine. This specific alteration was based on the published structure of bovine lens leucine aminopeptidase (Burley *et al*, 1990 and 1992). Glu334 was implicated as one of the seven active site residues in BLLAP that was involved in the chelation of Zn^{2+} ions; Glu 354 was

chosen for mutation in PepA because it is the equivalent residue to Glu334 and is therefore believed to co-ordinate the binding of Mn^{2+}.

The mutant derivative of PepA (PepA E354A) is deficient in its ability to hydrolyse substrates both *in vitro* and *in vivo* (leucine-*p*-nitroanilide and Val-Leu-NH_2, respectively). This suggests that Glu354 might be at the catalytic site of the enzyme and may be important in the catalytic mechanism of PepA, since replacement of the residue with Ala has compromised the active site of the enzyme. However, discussion of the catalytic role of Glu334 and the implications of the mutagenesis experiment for considerations regarding the enzymatic mechanism of BLLAP and PepA are inappropriate. This is for a number of reasons: the lack of aminopeptidase activity may be due to tertiary structural alterations in PepA and not purely a result of the enzyme being unable to chelate Mn^{2+}; the *pepA*E354A gene was not sequenced and therefore the possibility of mutations in other active site residues cannot be excluded; and there is no available model for the mechanism of peptide hydrolysis by these enzymes.

The mutagenesis described above was not performed in order to analyse the catalytic mechanism of PepA, but instead to examine the protein's role in the *cer* site-specific recombination reaction. Explicitly, this experimental approach was attempting to create a peptidase-deficient PepA derivative which could be tested for its ability to support *cer* recombination. This was done because of our inability to examine the role of PepA in *in vitro* assays (e.g. it would be relatively simple *in vitro* to remove Mn^{2+} ions from the reaction buffer and determine whether recombination could proceed). The fact that PepA E354A displayed no detectable catalytic activity when using leucine-*p*-nitroanalide as a substrate *in vitro* or Val-Leu-NH_2 *in vivo*, means that it is as "good" a mutant as can be produced for this experimental approach. The mutant enzyme is Xer^+ when highly expressed from pRM40 (*tac* promoter) and when poorly expressed from pRM21 (unidentified promoter sequences in pSELECT). This represents suggestive, but not conclusive (see below), evidence that the aminopeptidase activity of PepA is not

required in *cer* site-specific recombination. Some independently performed experiments offer circumstantial support for this conclusion:

(i) Val-Leu-NH_2 was used to isolate *E.coli pepA* mutants which were subsequently tested for their Xer phenotype (H.O'Mara and G.Szatmari, pers.comm.). A large number of the strains isolated from this screening procedure were both Xer^- and $peptidase^-$, but one clone (named HOM38a) was resistant to Val-Leu-NH_2 (i.e. $peptidase^-$) and was Xer^+. The gene encoding this PepA mutant has been subcloned (plasmid pH31.9) as a 1.9 kbp *Hind*III fragment, but the mutation has not been characterised. The mutant strains which were Xer^- and $peptidase^-$ have also not been characterised, but could be explained as encoding truncated PepA derivatives.

(ii) It has been shown that *E.coli* grown in media containing high concentrations of bestatin are resistant to Val-Leu-NH_2 but are still able to recombine *cer* reporter plasmids (G.Szatmari, pers.comm.). This can be interpreted as suggesting that in these conditions PepA is not catalytically active but is still able to support *cer* recombination, and it is therefore not functioning as an enzyme during the reaction.

The first reason that none of these experiments can be considered conclusive is the fact that it is not possible to rule out any residual peptidase activity in the reaction conditions used (either from an incompletely mutant protein or a subpopulation of PepA escaping bestatin inhibition) causing the Xer^+ phenotype. This is primarily because the limits of detection of the peptidase assays that were used are not known. It is also because of the non-quantitative nature of the *in vivo cer* recombination assay: *in vitro* it would be possible to determine whether the reaction rate had been altered by the reduction in PepA's catalytic efficiency, whereas *in vivo* a reduction in the reaction rate might not be detected because the reporter plasmids are able to recombine to completion before sufficient cells have grown to prepare DNA samples. However, no aminopeptidase activity at all has been determined for PepA E354A and therefore, despite the above arguments, it is entirely possible that PepA does not act as an aminopeptidase in *cer* recombination.

A more difficult problem to explain in this experimental approach is the possibility that PepA E354A remains Xer^+ either because the enzyme retains full enzymatic activity when it acts on substrates for which it has a higher specificity (e.g. the *xer* proteins) than the analogues used in these experiments, or because it has other enzymatic activities unaffected by the mutation made. The former of these objections does not seem likely to me because the E354A alteration almost certainly perturbs the general catalytic activity of PepA, and not only its activity against some substrates. The second possibilty cannot be ruled out, although it is not obvious what alternative enzymatic roles PepA might have in *cer* recombination.

To determine conclusively the PepA does not act enzymatically, future experiments would need to take a more positive approach. For example, it should be possible to isolate (or make) Xer^-, $peptidase^+$ derivatives of PepA if the interpretation of the experiments described in this chapter is correct. This could be achieved either by classical genetic techniques, or perhaps by creating a PepA derivative that is truncated in the same way as trypsin-cleaved BLLAP (which retains its peptidase activity; see section 3.1). If these experiments suggested that PepA was not required to process the other Xer proteins or to cleave small peptide molecules during the recombination of *cer*-like sites then it would become necessary to consider in more detail how PepA functions as a structural component of the reaction machinery. Since there is no evidence that it interacts with the *cer* site directly, it may be possible to perform gel binding assays in combination with other Xer proteins to see if it is involved in protein-protein interactions. Ultimately, however, it is probable that the function of PepA in *cer* recombination will be unravelled by analysing the *cer* reaction *in vitro*.

Chapter 4

Analysis of *cer* site-specific recombination *in vivo* using the controllable strain RM40

4.1 Introduction

To be able to analyse the mechanism of the *xer* site-specific recombination reaction it is necessary to be be able to perform the reaction and examine the reaction products in a controlled manner. Ideally this would be achieved by recreating the system *in vitro*, since this would allow the concentrations of the protein and DNA components to be altered, as well as the reaction conditions and type of substrates used. Many site-specific recombination reactions have been reconstituted *in vitro* (see Chapter 1) and to achieve this for *xer*, ArgR and PepA have been purified (Stirling *et al*, 1988 and 1989) and XerC is being purified (G.May, pers.comm.). This has allowed the role of these proteins in *xer* recombination to be analyzed using the gel binding assay (see above refs. and Fried and Crothers, 1981). It has also made it possible for the proteins to be added to *cer* and *dif* reporter plasmids in various simple buffers to look for the products of recombination. Experiments of this kind, and also experiments using crude cell extracts, have so far failed to yield positive results (G.May, S.Colloms, G.Szatmari pers. comms.). This could be for a number of reasons; for instance, the precise conditions required for recombination *in vitro* may not have been developed yet, or alternatively some essential factors have not yet been identified.

In the absence of an *in vitro* assay, it was decided to develop a system *in vivo* in which *xer* recombination could be controllably analysed. Such a system would be unlikely to be as manipulable or as informative as an *in vitro* assay. However, experiences from other systems suggested that it could yield some mechanistic data.

Controlled, *in vivo* experiments have been performed in the analysis of the site-specific recombination reactions catalysed by λ Int (Bliska and Cozzarelli, 1986), Tn3 Resolvase (Bliska *et al*, 1991) and bacteriophage P1 Cre (Adams *et al*, unpublished). All these systems have relied on being able to control the expression of the recombinase gene. This allowed the recombination reaction to be switched on at a set time and allowed the reaction products to be examined at various times subsequently. To achieve this control of expression, the genes encoding λ Int (*int*) and Tn3 resolvase (*tnpR*)

were cloned downstream of the lambda leftwards promoter (λ P_L) in plasmid constructions also containing the gene *cI857*, which expresses the thermolabile λ P_L repressor CI. In this way the expression of the recombinases is turned off when grown at 28-30°C and turned on at 37-42°C, when the repressor is thermally inactivated. (The methods used in the P1 Cre *in vivo* system are not known to me.)

The experiments performed in all of these cases were topological analyses of the products of the recombination reactions, and it was shown that the products made *in vivo* are not the same as those observed *in vitro*. In the case of λ Int the reaction products are nearly entirely free circles and not the catenanes and knots seen *in vitro*; similarly Tn3 resolvase produces free circles and not singly-linked catenanes (see Chapter 1). It was demonstrated that the reason for these results is the action of the enzyme DNA gyrase, which efficiently unlinks knots and catenanes *in vivo*. When this enzyme activity was inhibited by the drug norfloxacin (a superpotent analogue of nalidixic acid) the patterns of reaction products seen *in vivo* matched those predicted by *in vitro* experiments. These experiments have been interpreted as suggesting that the reactions proceed by the same mechanism *in vivo* and *in vitro*. The results of the Cre studies appear not to be as simple as this, and may suggest that this reaction produces topologically more complex products *in vivo* from those observed *in vitro*.

The above discussion is not intended as an exhaustive description of the experiments performed in these systems, but simply to illustrate the types of experiment that have been performed using *in vivo* assays. The amount of manipulation that is possible is clearly not as extensive as can be achieved *in vitro*: the amounts of the component proteins and the physiological conditions of the reactions cannot be readily changed. Nevertheless, the experiments can be performed in a controlled manner and mechanistic information about the reactions can be obtained. The types of substrate plasmids used in these experiments can clearly be varied quite considerably (e.g. comparing *cer*/*dif* in *xer* recombination). An important difference exists between analysis of the *xer* recombination reaction and the analyses described above, however. Since no *in vitro* recombination experiments have been performed for *xer*

recombination, any mechanistic results that can be derived would not be confirmation of already existing data, as was the case for λ Int and Tn3 resolvase.

This chapter reports the development of a controllable *xer* recombination system and the use of this system in determining that the *xer* reaction involves Holliday junction structures, which have previously been implicated *in vitro* as reaction intermediates in λ integrase-type recombination.

4.2 Construction of strains RM10, RM20, RM30 and RM40

It was decided that the gene encoding XerC would be used to control the *xer* recombination reaction *in vivo*. It is conceivable that controlling the expression of *argR* or *pepA* could allow the *xer* recombination reaction to be tightly regulated in the same way that controlling *int* achieved for the λ Int system and *tnpR* for resolvase. However, because *dif* recombination does not require ArgR or PepA, and because XerC was believed to be the *xer* recombinase (see Chapter 5), *xerC* was chosen. It was also decided that the promoter used to control *xerC* expression would not be λP_L but the *lac* promoter and its associated operator. Expression from the *lac* promoter is repressed by the protein LacI, which binds an operator sequence (O) found downstream of the promoter and overlapping the transcription start point (Beckwith, 1987). The repressor is thought to act by stopping the binding of RNA Polymerase. Addition of glucose to the bacterial culture can enhance this repression by indirectly reducing the binding of the catabolite activator protein (CAP), which assists the binding of RNA Polymerase. LacI can be inactivated by the gratuitous activator IPTG , which acts by stopping the repressor binding its operator sequence, and therefore turns on the expression of genes controlled by *lac*PO. The decision to use *lac*PO was a result of two factors. Firstly, it is readily available for cloning in a number of vectors commonly used in our laboratory (e.g. pUC18/19). Secondly, the strain that has been ubiquitously used in previous *cer*/*dif* recombination experiments is DS941, which is a *lacI*^{q} mutant. Use of this strain

would therefore offer maximal repression since it makes more than normal amounts of repressor, as well as offering experimental continuity.

The cloning steps involved in constructing this controllable strain are detailed in figure 4.1. pSD102 (Colloms, 1990) contains a 3.6 kbp *Hind*III-*Bgl*II *E.coli* chromosomal fragment that contains the genes *dapF*, *orf235*, *xerC* and part of the fourth gene of the *xerC* operon *orf238*, as well as the upstream gene *cyaX*. The plasmid pRM101 was generated by inserting a 230 bp *Pvu*II-*Sma*I fragment from pUC19 containing *lac*PO into a unique *Nru*I site that is 96 bp upstream of the *xerC* start codon and is within the coding sequence of *orf235*. The orientation of this insert was determined by restriction with *Pst*I (*Pst*I cuts 10 bp from the *Sma*I side of the *lac*PO insert and also within pSD102).

It was intended that the controllable *xerC* construction would be inserted into the *E.coli* chromosome and replace the wild type *xerC* gene (see below). For this reason an antibiotic resistance determinant was required to select for this insertion. A transcriptional terminator was also required upstream of the *lac*PO insert; this is because *xerC* is part of an operon, and without a terminator transcription from the operon's promoter would read through the *lac* promoter and compromise the expression control. These requirements were achieved in three steps. Firstly, the strong ribosomal RNA terminator was cloned as a 450 bp *Ssp*I-*Hind*III fragment from pKK223-3 (Pharmacia) into *Bgl*II-*Hind*III restricted pIC20R, creating pRM10. Secondly, the kanamycin resistance gene from Tn903 was cloned from pUC71K as a 1.3 kbp *Pst*I fragment into *Pst*I digested pRM10, creating pRM11. *Eco*RI restriction of this plasmid releases a 1.8kbp fragment containing the terminator and kanamycin resistance genes. The fragment was cloned into *Stu*I digested pRM101, creating pRM102, in which approximately 440 bp of operonic sequence have been replaced by this insert. This deletes the final 180 bp of *dapF* (the C-terminal 58 amino acids of DapF) and the first 260 bp of *orf235* (the amino-terminal 85 residues of ORF235). Note that the orientation of the inserted fragment was determined by *BamH*I restriction (*BamH*I cuts within the *lac*PO insert and in the polylinker of pRM11, beside the kanamycin gene), and the

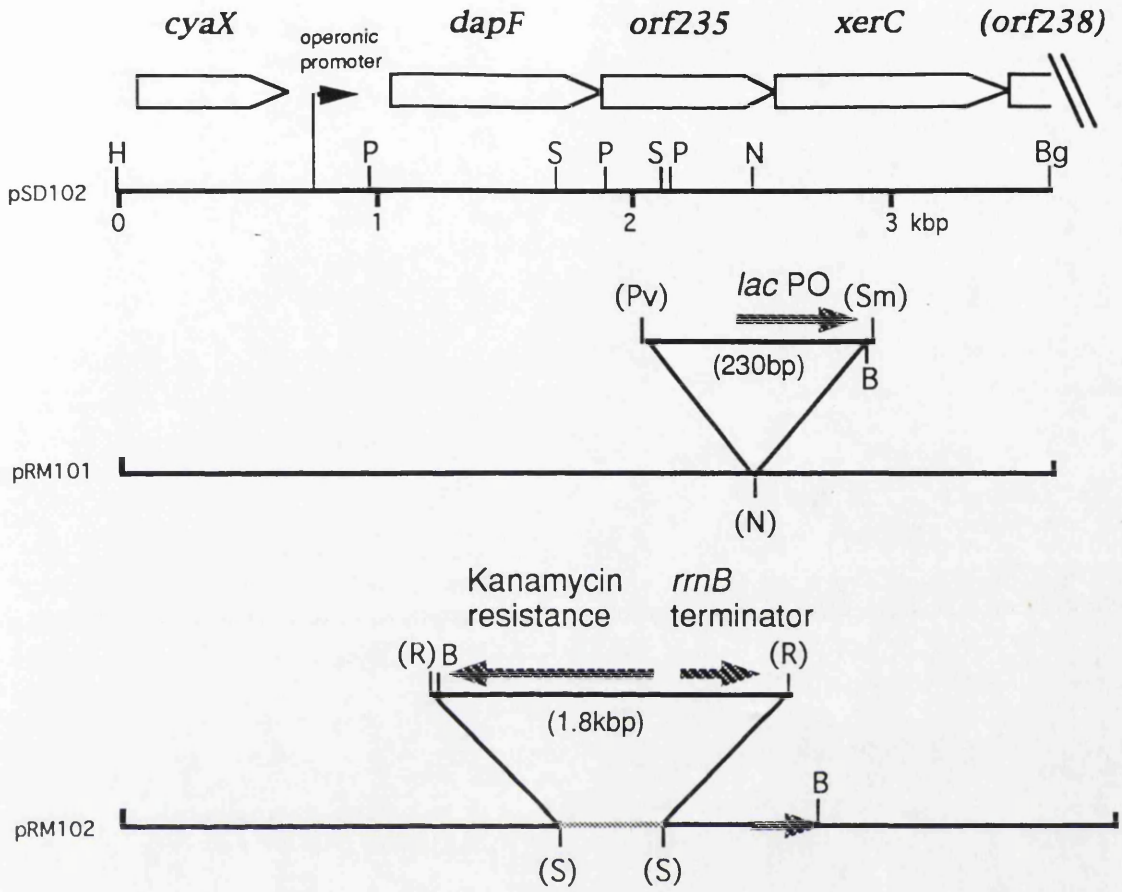


Figure 4.1. Plasmid clonings involved in the construction of a strain in which the transcription of *xerC* can be controlled. The top plasmid diagram is adapted from Colloms *et al* (1990) and shows only the *E.coli* chromosomal insert of the plasmid pSD102; it details the genes present in the insert and the important restriction sites. Below are the plasmids created by sequentially inserting (upstream of *xerC*) the *lac* promoter and operator of pUC19 (as a 230 bp *Sma*I-*Pvu*II fragment: making pRM101), and a 1.8 kbp *EcoR*I fragment carrying the kanamycin resistence gene of Tn903 and the *rrnB* transcriptional terminator (making pRM102).

Restriction enzyme recognition sites sites are shown as follows:

Bg=*Bgl*II H=*Hind*III N=*Nru*I P=*Pst*I Pv=*Pvu*II R=*EcoR*I S=*Stu*I Sm=*Sma*I B=*BamH*I

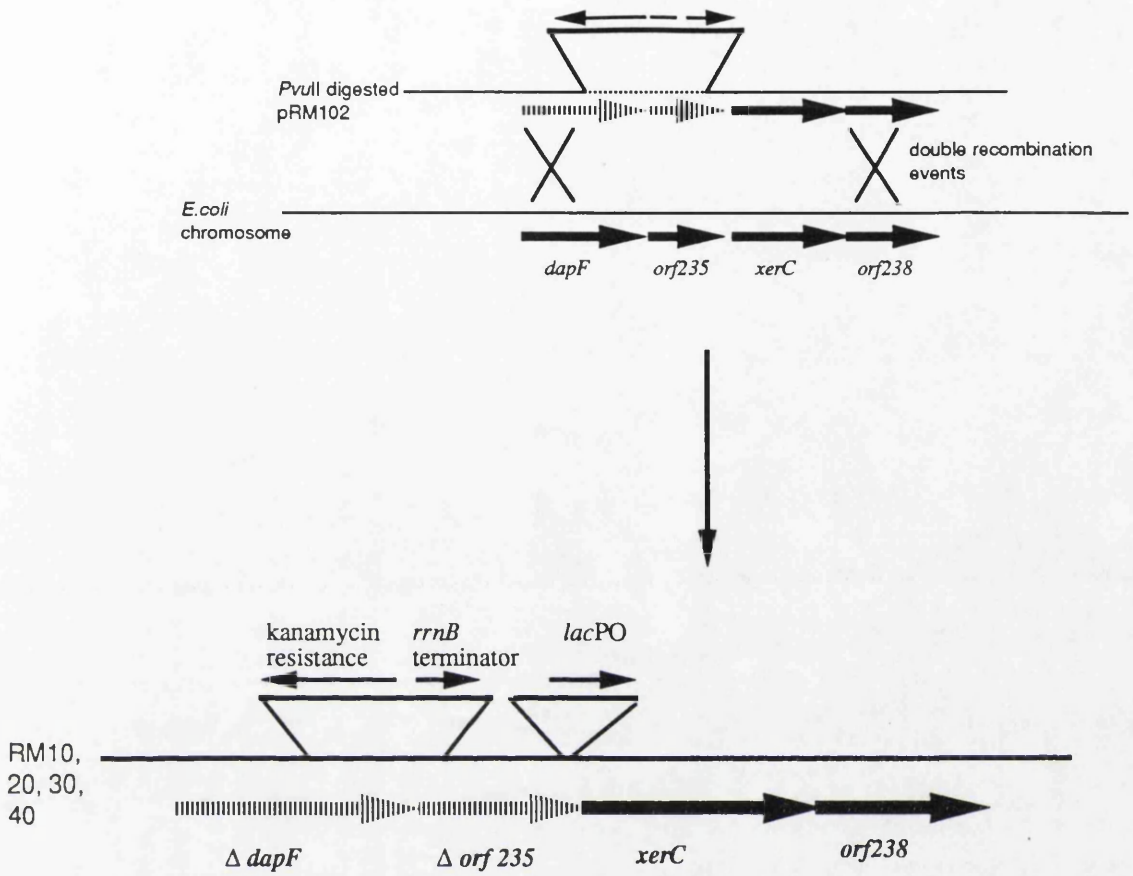


Figure 4.2. Construction of strains RM10, RM20, RM30 and RM40 by linear transformation of JC7623. Linear DNA fragments transformed into the strain JC7623 and containing a selectable copy of a gene with a homologue in the *E.coli* chromosome can exchange with that copy *via* a double recombination event (Winans *et al*, 1985). In this case the 4.4 kbp *Pvu*II DNA fragment contains the kanamycin resistence gene, and exchange generates strains in which the *xerC* gene is transcribed from the *lac* promoter rather than its natural operonic promoter. In addition to the inserted *lacPO* fragment, the kanamycin resistant strains RM10, 20, 30 and 40 contain insertions within *dapF* and *orf235* upstream of *xerC*, and are therefore likely to be mutant in these genes. These constructions have not been checked by Southern blotting, but are based on genetic characterisation of the strains.

kanamycin gene was previously orientated in pRM11 by *Hind*III restriction (which cuts within the polylinker and at one end of the inserted gene). As shown, the kanamycin gene is transcribed in the opposite direction to the operon and should not further burden the ribosomal RNA terminator.

Plasmid pRM102 was used to create a strain in which the chromosomal copy of *xerC* is expressed from the *lac* promoter rather than its natural promoter. This was achieved by transforming the linearized plasmid into the strain JC7623, which has the genotype *recB*, *recC*, *sbcB*. The *recB/C* mutations in this strain mean that it will not efficiently degrade linear DNA that has entered the cytoplasm, whilst the *sbcB* mutation suppresses the strain's recombinational defects. Winans *et al* (1985) were the first to use this technique of strain construction in *E.coli*, and they showed that if linear DNA transformed into JC7623 contains a copy of an *E.coli* gene it can exchange with the chromosomal copy by a double recombination event (see fig.4.2). In this way they used JC7623 to create various *E.coli* mutants. The technique was also used by Richaud *et al* (1987) to make mini-Mu insertional mutants in all the genes of the *xerC* operon (Colloms, 1990).

In this study, pRM102 was digested with *Pvu*II, which cuts twice within the vector but not within the insert, and transformed into JC7623 by standard $CaCl_2$ methods. Forty-three colonies were isolated and replica patched onto kanamycin and onto ampicillin-containing L-agar plates. Of these 43 colonies 42 were resistant to kanamycin but sensitive to ampicillin, suggesting that their chromosomal *xerC* had exchanged with the pRM102 *xerC* construction. Four of these colonies were then P1 transduced into DS941, creating strains RM10, RM20, RM30 and RM40.

Figure 4.2 shows the presumed chromosomal organization of these strains. This is based on the known restriction maps of the plasmid constructions, on evidence presented in the above papers that the linear DNA inserts by a double recombination event and on genetic characterisation of the strains (see below). Because of these considerations Southern blots were not performed to determine the status of the *xerC* region. However, it is probable that RM10, 20, 30, 40 are *dapF* and *orf235* mutants

because of the *lac*PO insertion into *orf235* and the deletions made in *dapF* and *orf235* during the plasmid constructions. *orf238* is downstream of *xerC* and is believed to be co-transcribed, therefore when XerC is expressed from the *lac* promoter it is likely that *orf238* will be transcribed also.

4.3 Controllable *in vivo* recombination in strains RM10, 20, 30 and 40.

To determine whether these strains are capable of controlling the expression of *xerC*, and therefore able to be used in *in vivo* recombination assays, they were transformed with the *cer* reporter plasmid pKS455 (Summers, 1989; see fig.4.20). The transformed bacteria were plated onto either L-agar plates containing ampicillin alone, or onto L-agar plates containing both ampicillin and 1mM IPTG. A number of colonies from these plates were pooled and their plasmid DNA content visualized by agarose gel electrophoresis.

Figure 4.3A shows the result of this experiment. Lanes 9 and 10 are marker lanes showing pKS455 and its resolution product (p456) that is produced after transformation into DS941. For each of RM10, 20, 30 and 40 this recombination only occurred when IPTG was present in the plates (lanes 2, 4, 6 and 8), suggesting that without IPTG induction there is little, or no, XerC expressed in these strains.

The transformed strains growing on the non-IPTG plates were then patched out onto another set of IPTG+ and - plates and their plasmid DNA was analysed in the same way as before (fig.4.3B). This experiment shows two things: the repression of *xerC* expression is not complete, since some pKS456 reaction product was visible in the uninduced (non-IPTG) lanes; and in these strains the repression of *xerC* can be overcome by addition of IPTG to the growing culture, since transfer to IPTG-containing plates resulted in resolution of pKS455.

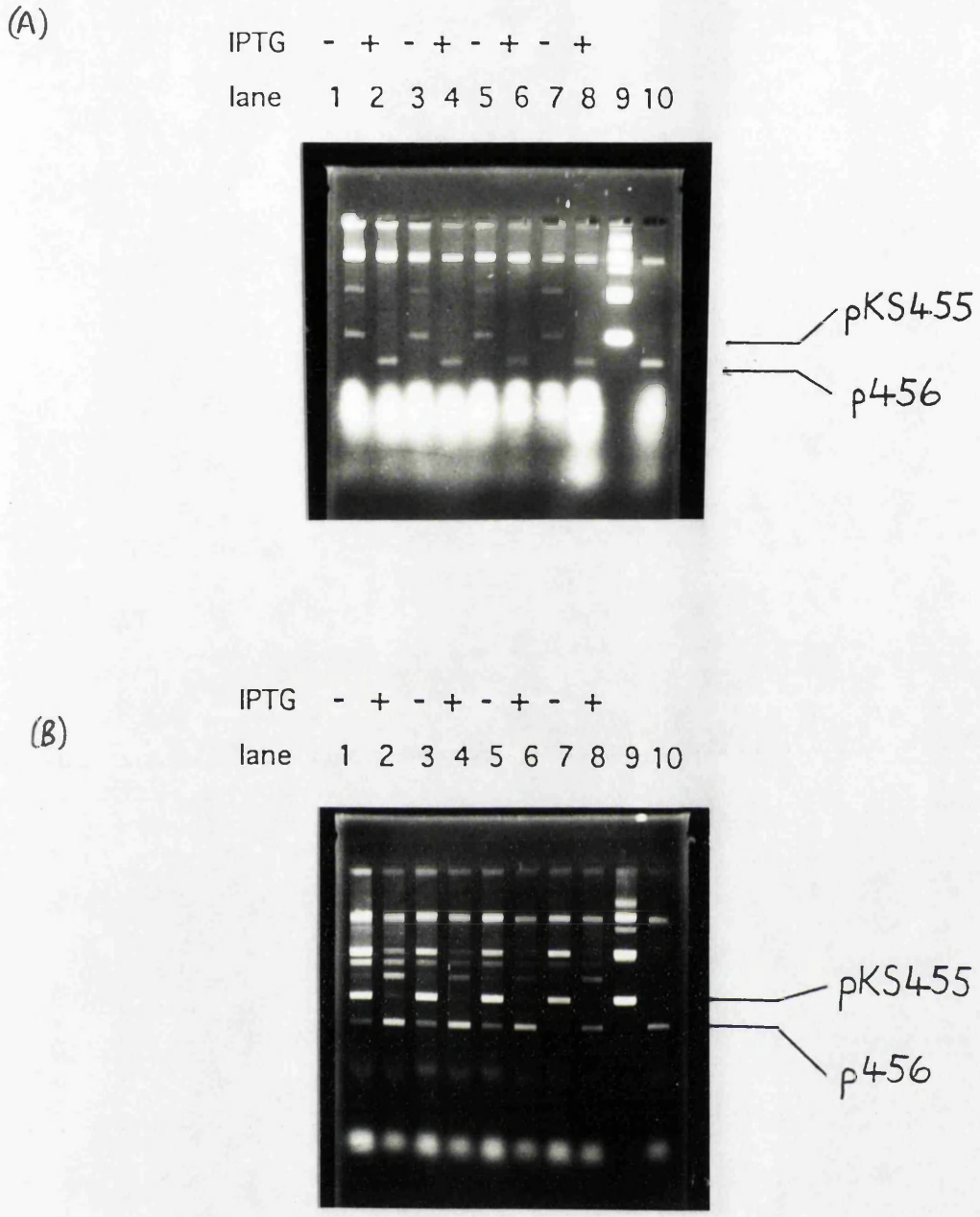


Figure 4.3. *cer* mediated site-specific recombination in strains RM10, RM20, RM30 and RM40. The strains were transformed with the *cer* reporter plasmid pKS455 and grown on L-broth with and without 1 mM IPTG. pKS455 is resolved by intramolecular recombination to produce p456. Plasmid DNA was prepared by the single colony lysis technique and run on a 1.0% gel. (**A**) DNA after overnight growth of the transformants. (**B**) DNA after repatching the transformants growing on IPTG- plates onto IPTG+ and IPTG- plates. For both **A** and **B**:

Lanes 1, 2: RM10; lanes 3, 4: RM20; lanes 5, 6: RM30; lanes 7, 8: RM40; lane 9: pKS455; lane 10: p456.

Together these two experiments show that the four strains meet the criteria demanded for a controllable *in vivo* system. Reporter plasmids can be introduced into the strains under repressed conditions such that they are not recombined, and at a set time subsequently the recombination reaction can be turned on.

4.4 Analysis of pSD115 recombination in RM40

Strain RM40 was transformed with the *cer* reporter plasmid pSD115 (Colloms, 1990; see fig.4.4) and plated onto L-agar supplemented with 1% glucose, 50ug/ml diaminopimelic acid and ampicillin. Glucose was included in this assay to increase the repression of *xerC*. Diaminopimelic acid (the enzymatic product of DapF) was included because it was observed that the transformed cells grew better on plates containing the acid than on plates without it; this may be a consequence of the mutation created in *dapF* during the construction of this strain.

A large number of the transformant colonies were pooled and inoculated into 40 ml of L-broth supplemented with 1% glucose and ampicillin and grown, shaking, at 37°C until the culture reached approximately mid-log phase (the OD_{600} was not determined). At this point the 40 ml culture was split into two separate 20 ml cultures; IPTG was added to one at a final concentration of 2 mM and then both allowed to continue growing. At the time of XerC induction and 30, 60, 90 and 120 minutes subsequently DNA was prepared from both cultures using standard boiling preparations, except that extraction with phenol/chloroform preceded isopropanol precipitation. The DNA prepared in this way was analysed by agarose gel electrophoresis (fig.4.5, lanes 1-10).

Without IPTG induction of *xerC* expression there was no recombination over the time course of the assay; this can be seen by the fact that the substrate, pSD115, was unchanged (lanes 1-5). A small amount of the replicative resolution product of the

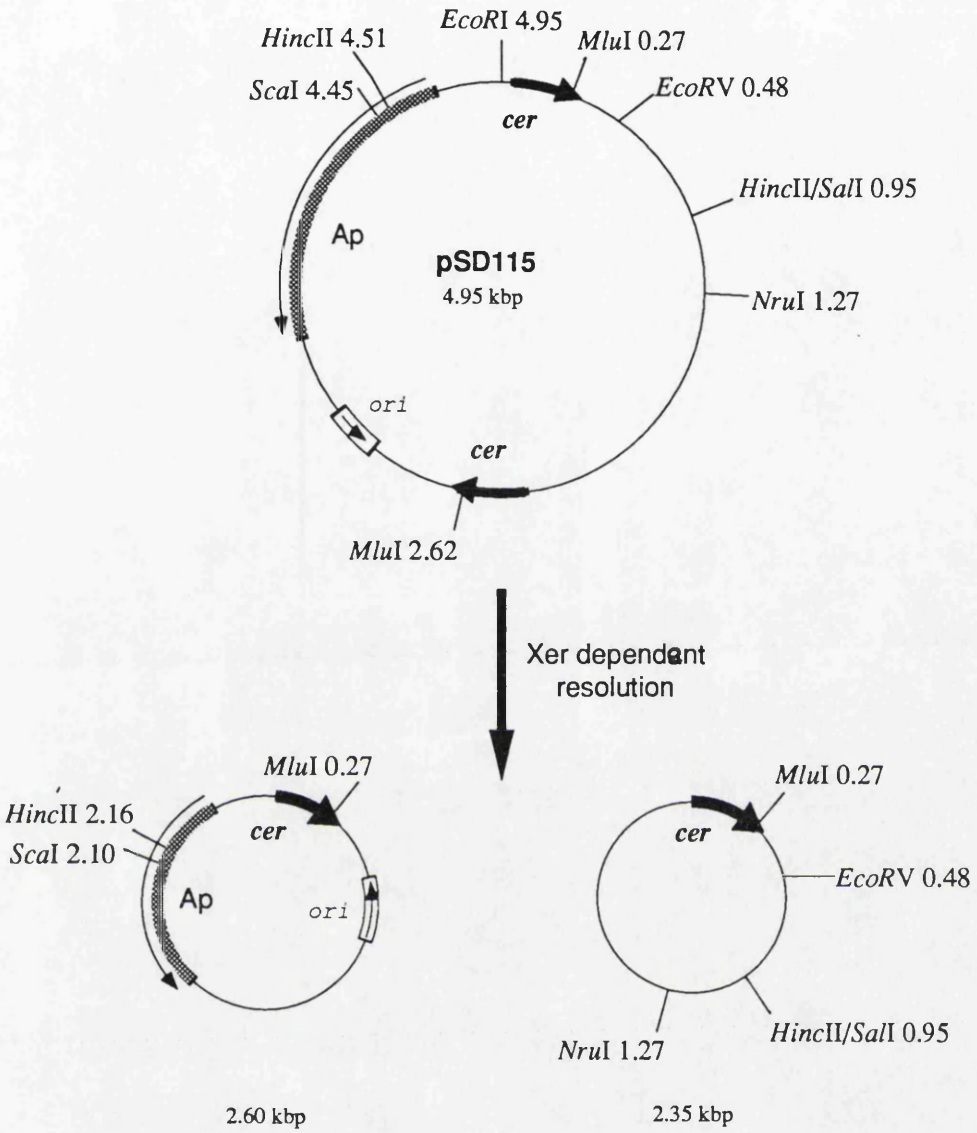


Figure 4.4. Diagrammatic representation of Xer-dependant resolution of the *cer* reporter plasmid pSD115. pSD115 was constructed in pBR322 by S.Colloms (1990). The plasmid is resolved in Xer^+ strains to yield 2.6 kbp and 2.35 kbp-sized, circular products. Note that only the 2.6 kbp-sized product contains a replication origin, and hence the smaller recombination product is rapidly lost from bacterial cultures.

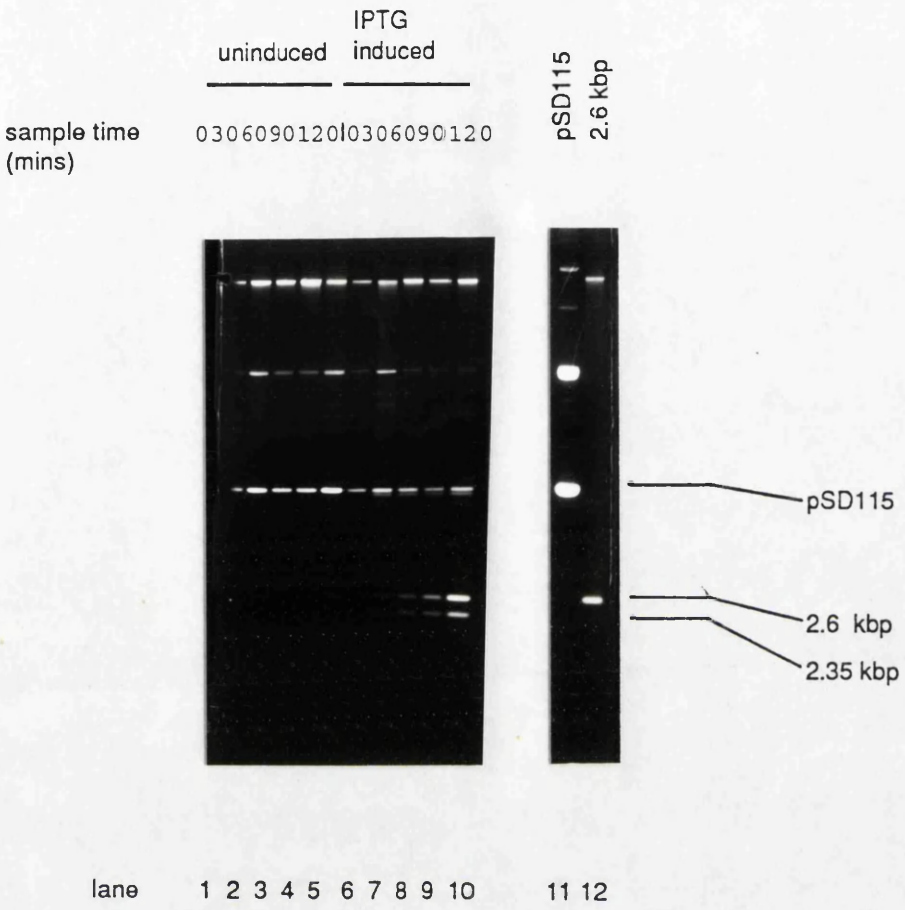


Figure 4.5. Time course of *cer* recombination *in vivo* using strain RM40. Two *in vivo* reactions were set up using pooled colonies of RM40 transformed with pSD115: in one (lanes 6-10) XerC expression was induced by addition of 2 mM IPTG to pSD115/RM40 at mid-log; in the other (lanes 1-5) the cells were grown on 1% glucose and XerC expression was not induced at the equivalent time. Time points refer to the time the plasmid DNA samples were made after the addition of IPTG. The DNA samples were made by boiling preparations (including phenol/chloroform extraction) and were run on a 1.2% agarose gel. In Xer^+ strains pSD115 resolution yields a 2.6 kbp, replicative product and a 2.35 kbp, non-replicative product (which is not observed in standard *cer* recombination assays; lane 12).

reaction was visible but the amount did not change during the assay, meaning that it probably arose during the transformation of RM40 and the overnight growth on the plate. In contrast, induction of XerC caused the appearance of two DNA species that increased in quantity from 30 to 120 mins (lanes 6-10). These correspond to the two products of pSD115 *xer* dependent resolution (see fig.4.4). The larger reaction product (2.6 kbp) is that normally seen when pSD115 is transformed into DS941, because it contains the origin of replication and is therefore maintained in the cell population. The smaller (2.35 kbp) product is normally not detected in recombination assays with wild type cells because it has no replication origin; the fact that it was detected in this recombination time course suggests that this is a more sensitive assay. Detection of the non-replicative product, and the increase in amounts of both products, shows that *xer* recombination continued throughout the assay.

The products of this reaction are unlinked circles. However, this cannot be said be to the initial topological structure of the products since no attempt has been made to inhibit DNA gyrase in these experiments. The *in vivo* experiments on λ Int and Tn3 resolvase-catalysed reactions described in the Introduction showed that without treatment with norfloxacin these reactions also produced free circles, despite these not being the products seen *in vitro*. Preliminary experiments using norfloxacin during *cer in vivo* assays suggested that the drug strongly inhibits the recombination reaction (data not shown). The meaning of this observation is not known, however, because these experiments were not followed up.

In figure 4.5 it can be seen that the pSD115 substrate band's appearance changed after the induction of XerC. This is caused by the presence of a second, *cer* recombination-dependent DNA species which had a slightly faster mobility and made the substrate band look like a doublet. This was analysed further by restriction digestion of the DNA samples with *EcoR*I, which cuts once within pSD115 and once within the 2.6 kbp product (see fig.4.4). An agarose gel of these restrictions is shown in figure 4.6.

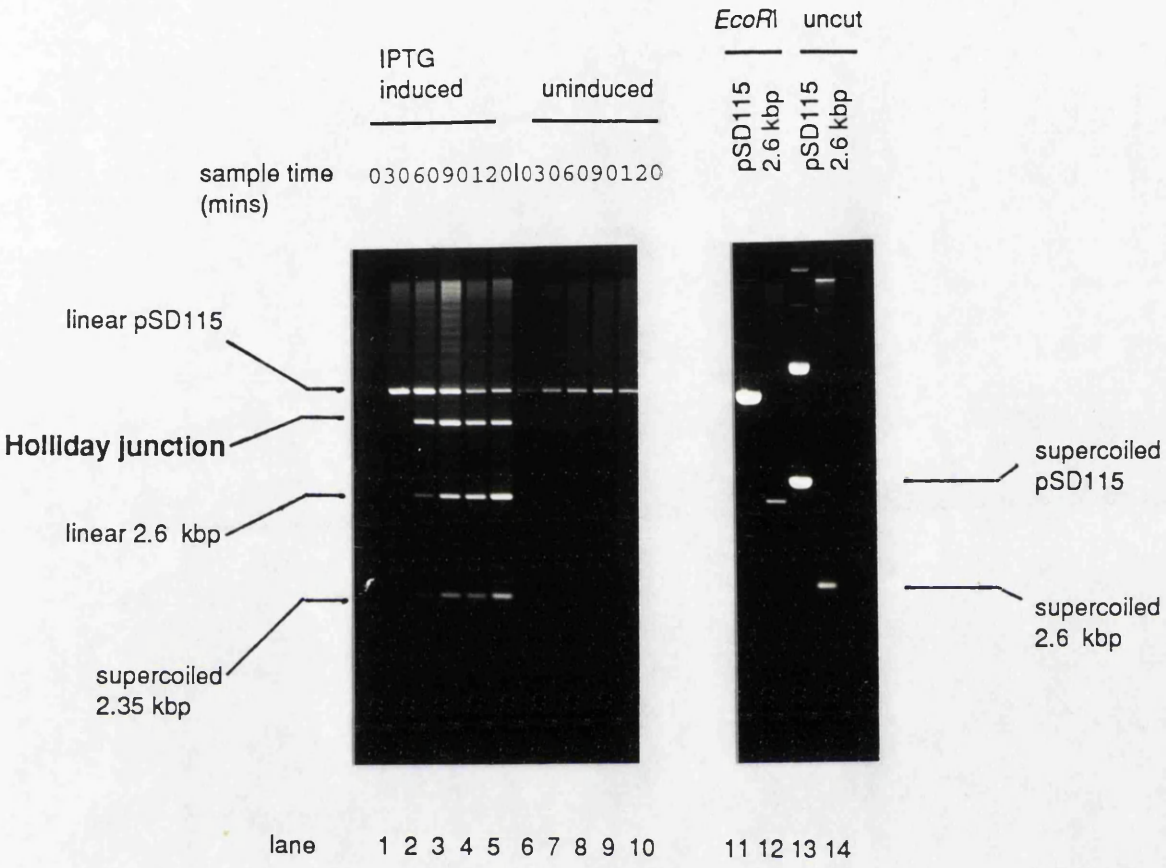


Figure 4.6. *EcoR*I restriction analysis of pSD115 *in vivo* recombination time course. Samples of the plasmid DNA from figure 4.5, representing *cer* mediated recombination of pSD115 in the strain RM40 over 120 mins, were digested with *EcoR*I and run on a 1.2% agarose gel. In this gel the IPTG-induced lanes are 1-5, the uninduced control samples are in lanes 6-10. A single *EcoR*I recognition site is present in pSD115 and in the 2.6 kbp, replicative product (see figure 4.4 and lanes 11-14); the 2.35 kbp product contains no *EcoR*I sites. The putative Holliday junction structure (α) is indicated.

Four DNA species with distinct electrophoretic mobilities were seen (lanes 1-5). The fastest migrating band had a slightly greater mobility than supercoiled 2.6 kbp product (compare to lane 14) and is therefore unrestricted 2.35 kbp product; the slower DNA species is linearized 2.6 kbp product (compare to lane 12). As expected, these products were only seen after IPTG induction and increased in quantity thereafter. The slowest migrating band was present throughout the time course and was the only DNA seen in the uninduced control lanes (6-10), therefore it is linearized pSD115 (compare to lane 11).The fourth DNA species had an electrophoretic mobility that was intermediate between supercoiled and linearized pSD115 substrate (compare to lanes 11-12), only appeared after XerC induction and was not present in the uninduced samples. The appearance of this DNA was therefore dependent on the induction of the *cer* recombination reaction, and consequently it may represent the restricted form of the DNA species that was barely distinguishable from pSD115 before *EcoR*I restriction (fig.4.5, lanes 6-10). The fact that it was not seen in the uninduced samples, nor in the 0' time point, strongly argues that this is not simply a contaminant DNA but is related to the *cer* recombination reaction.

A possible explanation for this DNA species is that it represents a Holliday junction, or DNA with similar physical characteristics. If a Holliday junction were formed between the two *cer* sites in pSD115 it would be confined to this 312bp of homology (if it were capable of branch migration), since the DNA between the *cer* sites is non-homologous pBR322 sequence (fig.4.4). This would therefore explain the increase in the mobility of the DNA when restricted by *EcoR*I, because the digestion converts a "figure-8" structure containing two supercoiled domains into an "α" structure which has a single supercoiled domain and two linear "tails".

Holliday junctions were initially proposed as being intermediates in homologous recombination (Holliday,1964). However, there is considerable evidence that equivalent structures are intermediates during λ integrase-like site-specific recombination reactions (see Chapter 1 for review and refs.). If this DNA species does indeed represent such a Holliday junction then it may represent an intermediate in the

cer recombination reaction, and offer an insight into the reaction's mechanism. The following sections detail experiments designed to determine whether this DNA species is indeed a Holliday junction.

4.5 Electrophoretic analysis of putative Holliday junctions

The hypothesis that this DNA species produced during *cer* mediated site-specific recombination in RM40 is a Holliday junction rests on the observed change in its electrophoretic mobility when restricted with one enzyme, *EcoR*I. The reason for this interpretation is the belief that the Holliday junction is embedded within regions of non-homology, causing the substrate to adopt a supercoiled figure-8 structure in which the two supercoiled domains (separated by the junction) are the size of the reaction products (2.6 and 2.35 kbp). Within these domains are a large number of restriction enzyme recognition sites, present only once (*Sca*I, *EcoR*I, *EcoR*V, *Sal*I, *Nru*I; see fig4.4), or twice (*Hinc*II, *Mlu*I) within the sequence of pSD115. If the DNA is in a Holliday junction conformation then digestion with these enzymes will generate either α structures of subtly different shapes when cut once, or χ structures when both domains are cleaved. The differences in the conformations of these derivatives should be reflected in alterations in their mobilities during gel electrophoresis.

RM40 transformed with pSD115 was used to set up a 20 ml culture (as described in section 4.4) and XerC expression was induced at mid-log phase with 2 mM IPTG. 90 mins after induction, plasmid DNA was prepared by the boiling method, including phenol/chloroform extraction. Figure 4.7 shows restriction digestions of this DNA, using the enzymes which cut only once, after agarose gel electrophoresis.

The bands corresponding to the putative α structure Holliday junctions are indicated, and it is clear that they had different electrophoretic mobilities in this gel. The mobilities of these structures might have been predicted to simply be a consequence of the size of the supercoiled domain, .i.e. those with the smaller domain remaining

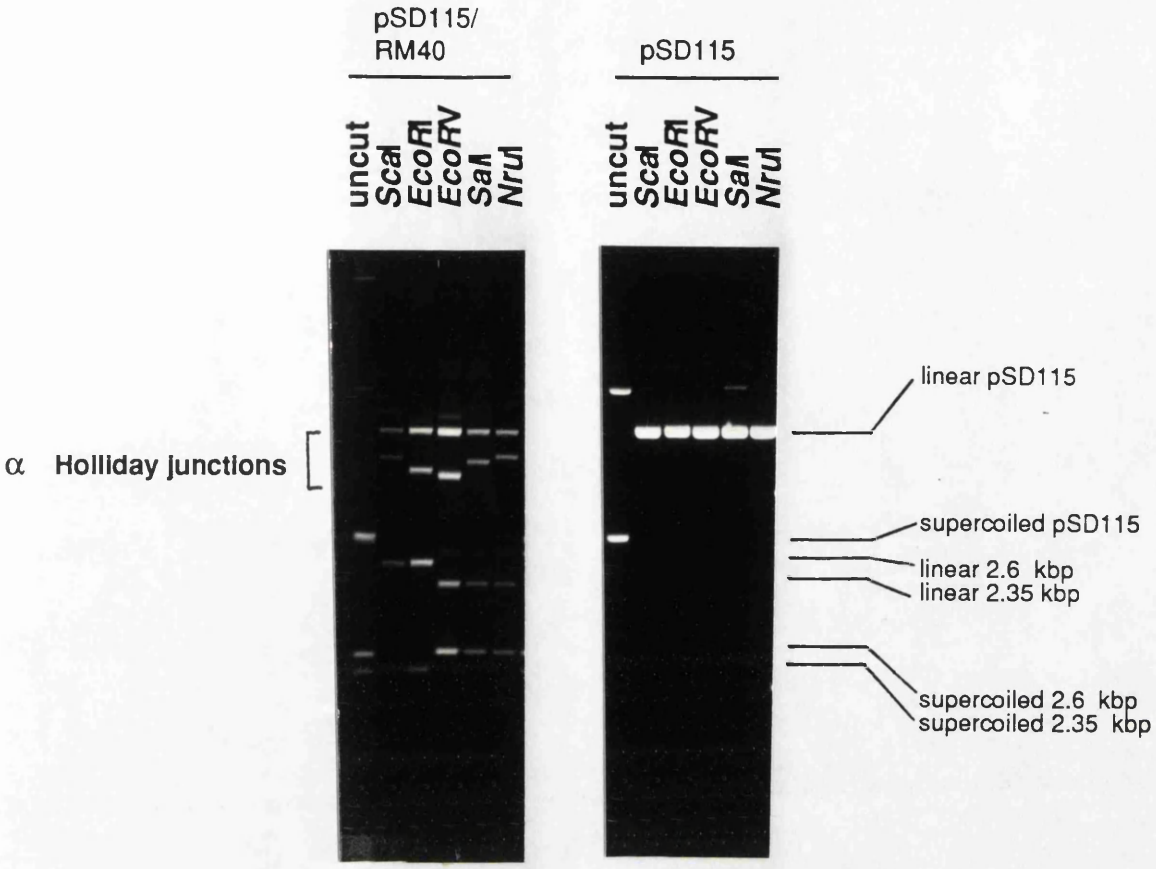


Figure 4.7. Analysis of the electrophoretic mobility of pSD115-derived α structures. RM40 transformed with pSD115 was grown to mid-log phase and XerC expression induced for 90 mins with 2 mM IPTG. Plasmid DNA samples were prepared, restricted with either *Sca*I, *Eco*RI, *Eco*RV, *Sal*I or *Nru*I and run on a 1.2% agarose gel. For comparison, pSD115 DNA digested with the same enzymes was also run on this gel. The positions of the restriction enzyme recognition sequences in pSD115 and how they are distributed into the *cer* mediated 2.6 and 2.35 kbp resolution products is detailed in figure 4.4. Supercoiled and linear substrate and product molecules are indicated, as are the putative α structures.

supercoiled after restriction digestion having equivalent mobilities and migrating faster than α molecules with a larger supercoiled domain. This is not the case, since *Sca*I and *Eco*RI α molecules ran differently in the gel, despite both being cleaved in the 2.35 kbp domain; similarly the three α molecules with 2.6 kbp supercoiled domains had different mobilities. These data suggest that the "arms" that extend from the α structures contribute significantly to the electrophoretic mobilities of the DNA. It appears that if one arm is very long and the other very short the structures have greater mobility than when the two arms are more equally sized. This is illustrated by comparing the α structures that were produced by *Eco*RV, *Sal*I and *Nru*I restriction. These enzymes cut progressively further from the *cer* site (174, 640 and 961bp respectively) and created α molecules with decreasing gel mobilities. This is confirmed by the slower migration of the *Sca*I α in the gel than the *Eco*RI α: *Sca*I cuts 515bp further from the *cer* site than does *Eco*RI.

Figure 4.8 shows an agarose gel of the same DNA restricted with *Hinc*II and *Mlu*I, both of which cut twice in pSD115. For reference it was again cut with *Eco*RI. Digestion with *Hinc*II linearised both of the product molecules and cleaved pSD115 into 3.55 and 1.4kbp fragments; it also resulted in a putative Holliday junction with a slower mobility than both pSD115 and the *Eco*RI α structure. This alteration in electrophoretic mobility is consistent with cleavage in both supercoiled domains, resulting in a Holliday junction that has a χ structure. *Mlu*I also linearised both products, but because it cuts within the *cer* sites it digested pSD115 to give product sized linear molecules. Despite cutting twice, as *Hinc*II does, *Mlu*I digestion did not result in the formation of χ structures (lane 4); there was no band corresponding to anything other than linear products and a very small amount of linear substrate (which was probably a consequence of slightly incomplete restriction). This is most readily explained by the Holliday junction being capable of branch migration throughout the *cer* site region of homology. *Mlu*I restriction would initially create a χ structure, but the Holliday junction would move to the ends of the molecule by branch migration and product-sized, linear molecules would be released.

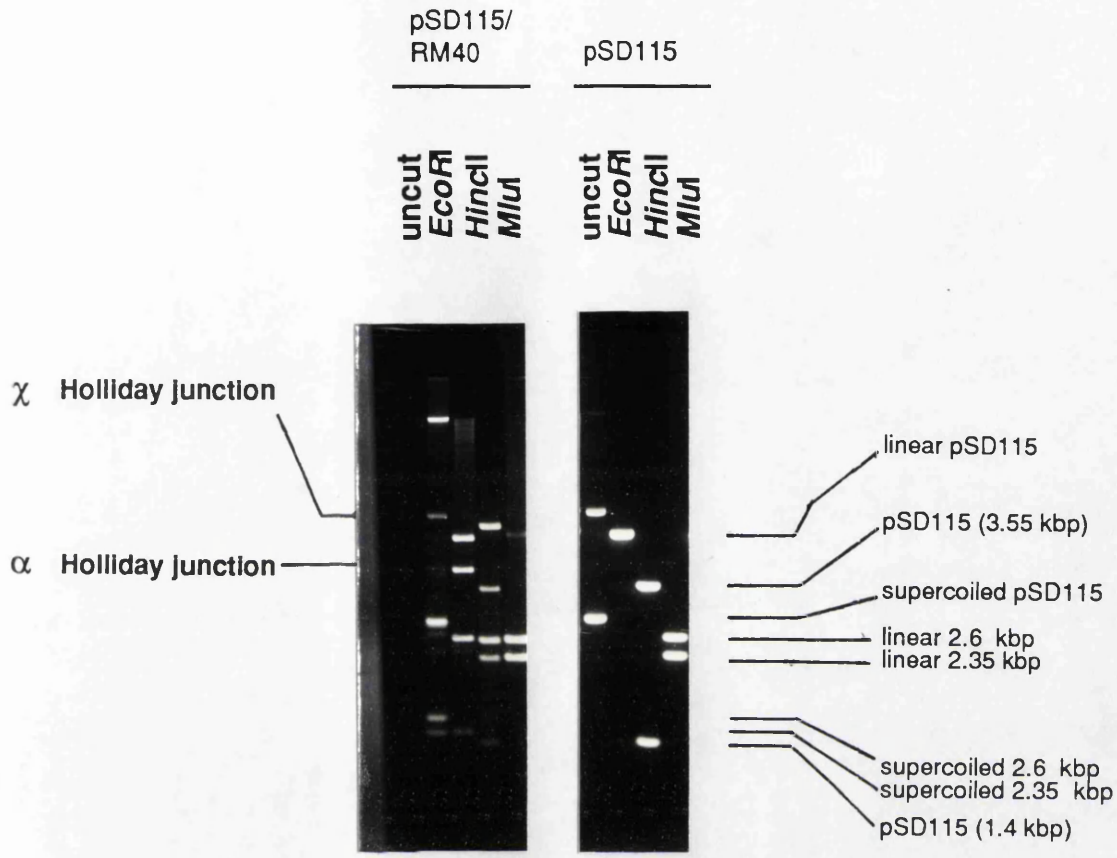


Figure 4.8. Restriction digestion of putative, pSD115-derived Holliday junctions in both supercoiled domains. Samples of the DNA shown in figure 4.7, representing pSD115 after 90 mins of *cer* mediated site-specific recombination in RM40, were digested with either *EcoR*I, *Hinc*II, or *Mlu*I and run on a 1.2% agarose gel. For comparison, pSD115 DNA digested with the same enzymes was run on the same gel. Figure 4.4 shows the position of the single *EcoR*I recognition site present in pSD115 and in the 2.6 kbp product, and the positions of the two *Hinc*II and *Mlu*I sites in pSD115 (which are found as single sites in each product). The putative α and χ Holliday junction structures are indicated, as well as supercoiled and linear substrate and product molecules and the pSD115 fragments generated by *Hinc*II digestion.

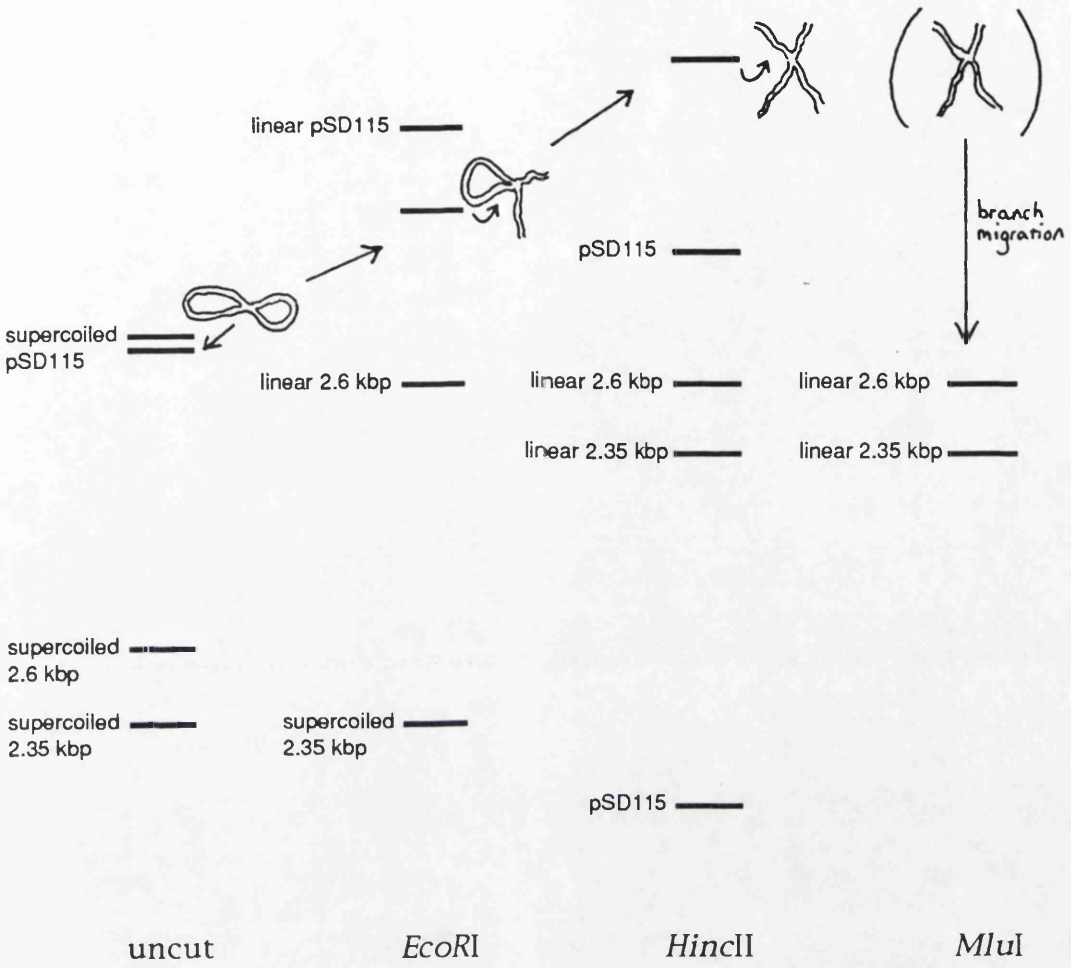


Figure 4.9. Diagrammatic representation of the alterations in the electrophoretic mobility of the putative Holliday junctions after restriction digestion. The proposed structures adopted by the Holliday junctions are shown, for simplicity, without supercoiling in their undigested domains; it is not intended that this imply that these domains are nicked, and it is likely that Holliday junctions that have been nicked in these ways would have different mobilities to those shown. Uncut, the Holliday junctions contain two supercoiled domains and adopt a "figure-8" structure whose electrophoretic mobility is nearly indistinguishable from supercoiled pSD115 substrate DNA. A single cleavage in either supercoiled domain generates a more retarded α structure. Cleavage in both supercoiled domains results in highly retarded χ structure molecules. Digestion within both *cer* sites allows the Holliday junction to branch migrate to the ends of the DNA and results in product-sized, linear DNA fragments.

All the above electrophoretic analysis is summarised in figure 4.9. The analysis is fully consistent with this recombination-dependent DNA species containing a Holliday junction. Restriction with *Mlu*I suggested that the structure is centred on the *cer* sites of pSD115 and is capable of branch migration. All of these DNA samples were stripped of protein by phenol/chloroform extraction, therefore excluding the possibility that these bands represent a protein-bound structure, for example synapsed *cer* sites.

4.6 The putative Holliday junctions are substrates for cleavage by RuvC *in vitro*

The experiments described above are consistent with this *cer* recombination-derived DNA species being a Holliday junction. It was, however, felt necessary to examine the conformation of the DNA further. This was examined by determining whether the species is a suitable substrate for an enzyme that specifically cleaves 4-way DNA junctions.

A number of enzymes have been described that can act as "junction-resolvases". The first to be identified was the product of bacteriophage T4's gene *49*, T4 endonuclease VII (Mizuuchi *et al*, 1982; Lilley and Kemper, 1984). This enzyme has a critical role in the life cycle of phage T4 because it cleaves branched DNA structures which form during the replication of the phage (Mosig, 1987). Analysis of the enzymatic activity of the purified protein showed that it is not specific for 4-way DNA junctions, since it will also cleave 3-way junctions, heteroduplex loops and extended single-stranded termini. Because of these diverse activites this protein was felt not to be a suitable enzyme to probe the structure of the *cer* recombination junctions.

Bacteriophage T7 encodes an enzyme (the product of gene *3*, T7 endonuclease I) that is utilised during the life cycle of this phage in a very similar way to how endonucleaseVII acts during T4 growth, indeed it has been demonstrated to be able to complement T4 gene *49* mutants. *in vitro* analysis of the purified enzyme (Parsons and

West, 1990) showed it to be slightly more specific than T4 endonucleaseVII. For instance, it footprints all four strands of a synthetic 4-way junction point and cleaves in both possible orientations. Again, however, this endonuclease will recognise and cleave 3-way junctions, making it unsuitable for this study.

Enzyme activities of the type described above have also been discovered using synthetic 4-way χ structures in extracts from eukaryotic cells. For example, three apparently distinct endonuclease activities have been partially purified from *S. cerevisiae* (Jensch *et al*, 1989). However, in this, and in other cases from higher eukaryotes, the specificities, and indeed the identities, of the enzymes are not yet clear, and are therefore not suitable for these experiments.

Recently Connolly and West (1990) used α structures which had been artificially generated using RecA to identify activity from *E.coli* cell extracts that would resolve Holliday junctions. They further showed that this activity was absent in extracts made from *ruvC* mutants (Connolly *et al*, 1991). The gene encoding RuvC has been cloned (Takahagi *et al*, 1991 and Sharples and Lloyd, 1991) and been found to be part of a "ruv" locus. *ruvC* is located upstream of two genes, *ruvA* and *ruvB*, which constitute a SOS-regulated operon; *ruvC* is itself part of a two-gene operon (the other gene is an unidentified open reading frame, *orf26*). Mutations in any of the 3 *ruv* genes confer a "ruv phenotype": they are deficient in homologous recombination in various genetic backgrounds. A single *ruv* mutant, however, shows only slightly reduced levels of recombination. This has been interpreted as meaning they are involved late in homologous recombination, e.g. in a resolution step (see e.g. Lloyd, 1991).

Biochemical analysis of the roles of these *ruv* proteins is under way and all three have been purified. The roles of RuvA and RuvB are not yet clear, but the function of RuvC has been determined. Purified RuvC will act *in vitro* to cleave RecA-generated Holliday intermediates without the need for RuvA or RuvB. It will also cleave synthetic 4-way junctions (Holliday junctions or cruciforms) in both possible orientations without RecA (Dunderdale *et al*, 1991, and Iwasaki *et al*, 1991). The studies also show that RuvC will not recognise 3-way junctions, or any of the other structures which

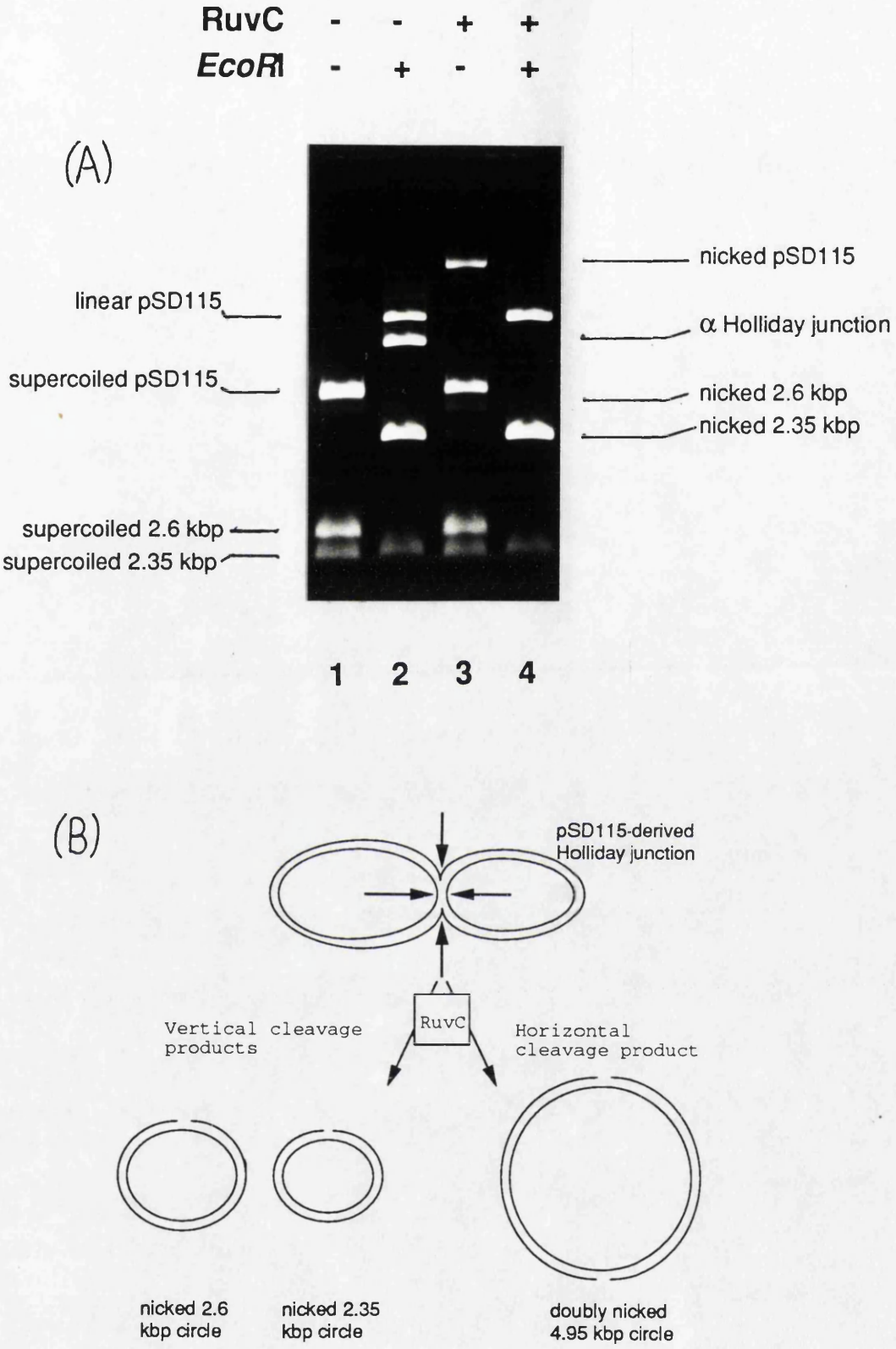


Figure 4.10 (A) Treatment of putative Holliday junctions *in vitro* with RuvC. RM40 transformed with pSD115 was grown to mid-log phase, XerC expression was induced for 90 mins with 2 mM IPTG and plasmid DNA was prepared from the culture and purified on a CsCl gradient. The DNA was then incubated *in vitro* with purified RuvC (lanes 3 and 4) and run on an agarose gel with (lane 4) and without (lane 3) subsequent digestion by *EcoR*I (this experiment was performed by S.West). Lanes 1 and 2 represent undigested (lane 1) and *EcoR*I digested (lane 2) controls in which the DNA has not been treated with RuvC. **(B) Schematic illustration of RuvC-catalysed resolution of supercoiled pSD115-derived Holliday junctions.** Cleavage by RuvC in a vertical orientation (as drawn) would generate nicked 2.6 and 2.35 kbp DNA species; cleavage horizontally would result in doubly nicked pSD115 substrate-sized DNA.

served as substrates for the phage endonucleases, in the reaction conditions employed in analysing 4-way junctions. Furthermore, the 4-way junctions must have at least 12 bp of homology before they act as substrates for RuvC.

These studies offer compelling evidence that RuvC is at least one of the *E.coli* enzymes which cleave Holliday junction intermediates that form during homologous recombination. The specificity of RuvC for 4-way junctions also made it the most informative enzyme for this study.

1.2% agarose gels (23cm in length) run for approximately 16 hours at 50 volts (e.g. see fig.4.5) give the best separation of supercoiled pSD115 and supercoiled "Holliday junction" that I have achieved. Because of this problem, and also because of the loss of material incurred when DNA is isolated from gels, it was not possible to purify supercoiled Holliday junctions, and therefore a CsCl preparation of pSD115 from RM40 after 90 mins induction was prepared by myself and used by S.West in *in vitro* assays using purified RuvC. Figure 4.10A shows an agarose gel of this assay; the *in vitro* conditions used for treatment with RuvC are published (Dunderdale *et al*, 1991).

Lanes 1-2 show the DNA preparation before exposure to RuvC. As stated, this was a mixture of pSD115 substrate, products and putative Holliday junctions. Without restriction the supercoiled substrate and Holliday junction were indistinguishable in this gel (lane 1). However, after restriction with *EcoR*I the α structure representing the Holliday junction was clearly visible (lane 2). Lanes 3-4 are the same samples after treatment with RuvC (note that *EcoR*I restriction followed RuvC treatment). *EcoR*I restriction showed that the α band was no longer present, whilst the substrate and product bands had not been depleted. This suggests that RuvC has specifically cleaved the supercoiled Holliday junction. Examination of the unrestricted DNA (lane 3) revealed an increase in the amount of nicked (open circle) form of pSD115 and both products. This would be consistent with the action of RuvC on supercoiled Holliday junctions, as illustrated in figure 4.10B. RuvC simply cleaves a 4-way junction in either of two possible orientations; it does not reseal the nicked DNA. For a supercoiled

molecule this would result in either singly nicked product-sized molecules (2.6 and 2.35 kbp in this case) or doubly nicked substrate molecules, depending on which orientation is used. If the nicked DNAs in this assay arose from the action of RuvC it suggests that RuvC cleaves these Holliday junctions in both orientations.

This experiment was compromised by the inability to distinguish the supercoiled Holliday junctions from the substrate. The apparent reduction in the amount of substrate band in the unrestricted sample (lane 3) must be assumed to have resulted from RuvC cleavage of the Holliday junctions present within that band. An alternative explanation could be that the increased amount of nicked material was a result of non-specific nicking during the RuvC treatment. Some observations argue against this, however: there was no concurrent reduction in the amounts of supercoiled products, as would be expected from non-specific nicking; after *EcoR*I digestion it was only the α structure that had been removed (if anything, the amounts of linear pSD115 and linear 2.6 kbp product appeared to have increased; compare lanes 4 and 2).

Although reservations must be acknowledged, this result is nevertheless interesting (and is strengthened when considered in conjunction with the previous electrophoretic analyses), and may be taken as further evidence that these structures are 4-way DNA junctions. If RuvC is able to recognise and cleave this DNA it suggests it has at least 12bp of sequence homology within the junction. Futhermore, the size of the nicked "resolution products" corresponded to the *cer* recombination products, suggesting that the junction is centred on the *cer* sites. Both of these interpretations are in agreement with *Mlu*I restriction analysis (see above).

4.7 Electron microscopy of isolated χ structures

This section describes direct visualisation of the structure of the putative Holliday junctions by electron microscopy. This was achieved by restricting the DNA prepared above (section 4.6) with *Hinc*II and isolating the presumptive χ stucture from an 1.2%

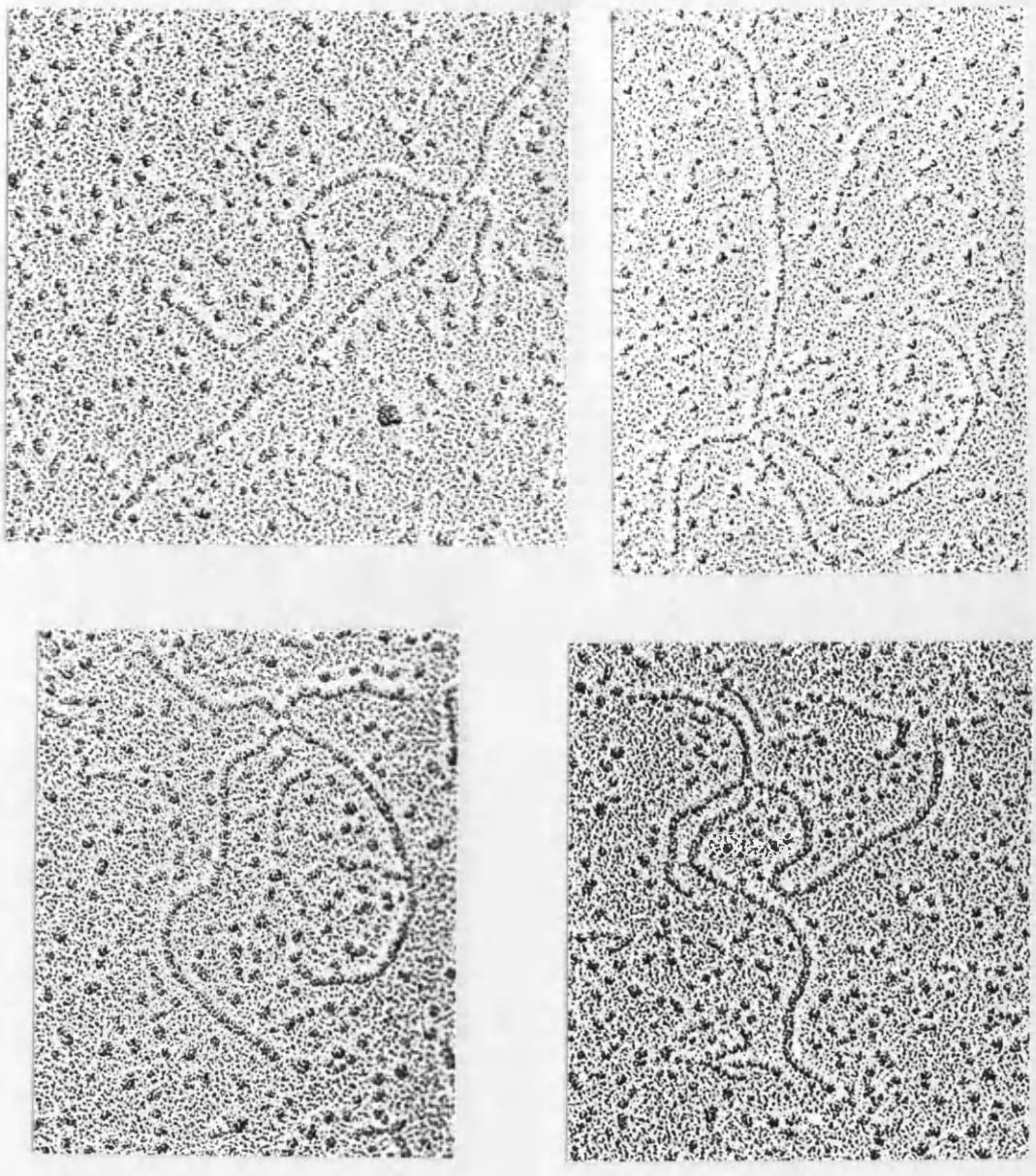


Figure 4.11. Electron micrographs of isolated χ structures. A Sample of the plasmid DNA shown in figure 4.10 was digested by *Hinc*II and run on a 1.2% agarose gel. DNA was purified from the band representing the *Hinc*II χ structure and spread, in 40% formamide, onto grids and shadowed with platinum : palladium for electron microscopy. The migrographs are at 90,000X magnification and the bar shown represents 0.33 um, which corresponds to approximately 1 kbp of DNA.

arm lengths in cm (percentage of total)

Holliday junction	long arm (1)	long arm (2)	short arm (1)	short arm (2)	total length of arms (cm)
1395a	6.3 (41.4)	5.0 (32.6)	2.2 (14.3)	1.8 (11.7)	15.3
1395b	6.0 (38.5)	5.4 (34.6)	2.1 (13.5)	2.1 (13.5)	15.6
1391a	5.4 (35.6)	5.3 (35.3)	2.2 (14.5)	2.2 (14.5)	15.1
1391b	6.7 (42.9)	4.8 (30.8)	2.3 (14.7)	1.8 (11.5)	15.6
1389B	6.1 (43.3)	4.5 (31.9)	2.2 (15.6)	1.3 (9.3)	14.1
1392	6.7 (42.4)	5.0 (31.6)	2.5 (15.8)	1.6 (10.1)	15.8
1388a	5.8 (38.7)	5.4 (36.0)	2.0 (13.3)	1.8 (12.0)	15.0
1388b	6.7 (42.9)	4.8 (30.8)	2.3 (14.7)	1.8 (11.5)	15.6
1388c	6.4 (42.1)	4.6 (30.2)	2.5 (16.4)	1.7 (11.2)	15.2
1393B	5.7 (37.5)	5.1 (33.6)	2.2 (14.5)	2.2 (14.5)	15.2
1394a	6.2 (41.3)	4.9 (32.7)	2.1 (14.0)	1.7 (11.3)	15.0
1394b	6.0 (40.5)	4.5 (30.4)	2.4 (16.2)	1.9 (12.8)	14.8
1393a	6.9 (45.1)	4.6 (30.1)	2.2 (14.4)	1.6 (10.5)	15.3
1393b	6.6 (41.5)	5.3 (33.3)	2.0 (12.6)	2.0 (12.6)	15.9
1388B	6.8 (44.1)	4.4 (28.6)	2.7 (17.5)	1.5 (9.7)	15.4
1390	6.0 (38.0)	5.5 (34.8)	2.1 (13.3)	1.9 (12.0)	15.8
1389a	6.0 (41.7)	4.5 (31.2)	2.4 (16.7)	1.5 (10.4)	14.4
1389b	5.6 (36.6)	5.6 (36.6)	2.1 (13.7)	2.0 (13.1)	15.3
1404	5.9 (38.3)	5.4 (34.7)	2.1 (13.5)	2.1 (13.5)	15.5
1405	6.3 (38.7)	5.6 (34.4)	2.5 (15.3)	1.9 (11.7)	16.3
1412	5.9 (38.3)	5.5 (35.7)	2.1 (13.6)	1.9 (12.3)	15.4
1410	6.1 (38.4)	5.6 (35.2)	2.2 (13.8)	2.0 (12.6)	15.9
1411	5.7 (37.5)	5.4 (35.5)	2.2 (14.5)	1.9 (12.5)	15.2
1409	6.2 (41.6)	4.5 (30.2)	2.6 (17.4)	1.6 (10.7)	14.9
1408a	6.1 (41.5)	4.8 (32.7)	2.2 (15.0)	1.6 (10.9)	14.7
1408b	6.1 (41.5)	4.7 (32.0)	2.2 (15.0)	1.7 (11.6)	14.7
1406	6.4 (41.8)	4.9 (32.0)	2.4 (15.7)	1.6 (10.5)	15.3
1407	5.9 (40.4)	4.8 (32.9)	2.1 (14.4)	1.8 (12.3)	14.6
1403B	6.1 (40.9)	4.9 (32.9)	2.4 (16.1)	1.5 (10.1)	14.9
1403a	5.7 (40.4)	4.5 (31.9)	2.0 (14.2)	1.9 (13.5)	14.1
1403b	5.9 (40.1)	4.7 (32.0)	2.1 (14.3)	2.0 (13.6)	14.7
1403c	6.4 (43.5)	4.2 (28.6)	2.5 (17.0)	1.6 (10.9)	14.7
1400	6.2 (43.4)	4.5 (31.5)	1.9 (13.3)	1.7 (11.9)	14.3
1402	5.8 (38.9)	5.3 (35.6)	2.0 (13.4)	1.8 (12.1)	14.9

average percentages	**40.6** (35.6-45.1)	**32.7** (28.6-36.6)	**14.3** (12.6-17.5)	**11.8** (9.3-14.5)	
measured sizes (bp)	**2010** (1762-2232)	**1619** (1416-1812)	**708** (624-866)	**584** (460-718)	
predicted sizes (bp)	**1990** (+/- 156)	**1565** (+/- 156)	**773** (+/- 156)	**620** (+/- 156)	

Figure 4.12. Arm lengths of individual χ structures. Contour lengths (in cm) of the arms of 36 χ structures (determined from the electron micrographs using a map measuring instrument) are shown in the upper table. There were four different sized arms in each structure, and the table is arranged in columns, with the largest arm size on the left and smallest on the right. The total sizes of the structures (determined by summing the arm lengths) is in the rightmost column; the percentage of total size that each arm represents is shown in brackets. The lower table details the four arm lengths (and variation around the figure) calculated using the average for each structure, based on the size (4950 bp) of pSD115. For comparison, the predicted arm lengths are also shown (note that the Holliday junction is capable of branch migration, and hence the figure varies +/- 156 bp).

agarose gel. Isolation of the DNA involved low speed centrifugation of the gel chip through siliconised glass wool (after visualising the DNA by ethidium bromide staining; see Materials and Methods), ethanol precipitation and resuspension in TE (pH8.0). The χ structures prepared in this way were next spread onto grids in 40% formamide, and shadowed for electron microscopic examination (see materials and methods and Coggins, 1987). This work was done in collaboration with Lesley Coggins. Figure 4.11 shows representative electron micrographs of the DNA seen by this technique.

The structures shown are clearly consistent with the DNA being a χ structure with a central Holliday junction. Formamide treatment can be seen to have spread the DNA out at the Holliday structure, causing it to appear as a "hole" where the four strands meet. The majority of the molecules seen on the grid could be determined to be χ structures, although in many cases the arms had become fixed such that they crossed over each other and the hole at the point of strand convergence was not always clearly visible. The grid also contained smaller pieces of linear DNA, which could have arisen by shearing of the structures during their preparation, or may simply have been background contaminant DNA that was accidentally isolated from the gel.

Figure 4.12 shows the results of measuring the arm lengths of 36 χ structures from the electron micrographs. Only molecules in which the path of each arm could be accurately determined, and where the Holliday structure could be rigorously assigned, were measured. The sizes of the arms are shown in centimetres, and the percentage of the total size that each arm represents is shown in brackets. The sizes suggested that each molecule contained two long and two short arms. Furthermore, in each case there appeared to be one larger and one smaller arm in both groups; this is reflected in the arrangement of the measurements into columns. The average total size was 15.1 cm, the maximum measured size was 16.3 cm and the minimum was 14.1 cm; these values are within the amount of variation expected by experimental error. The top row in the lower table shows the average percentage that each arm represents of the total molecule size, and the variation around that figure. From these averages the arm lengths of the

molecules were calculated, based on a molecule of 4.95 kbp - the size of pSD115. The calculated sizes were: 2010bp +222/-248; 1619bp +193/-203; 708bp +158/-84; 584bp +134/-124. These correspond well with the predicted arm lengths: 1990bp +/-156; 1565bp +/-156; 773bp +/-156; 620 +/-156. The predicted sizes are based on the positions of the *Hinc*II recognition sites within pSD115, assuming that the Holliday junction is centred on the *cer* site, and take into account the possibility that the junctions can branch migrate throughout the *cer* site.

These results confirm that these DNA structures are χ structures and therefore show that a Holliday junction that is produced during *cer* recombination in RM40. The measurements confirm the prediction, based on restriction analysis, that the Holliday junction is within the *cer* sequence of pSD115.

4.8 The nature of the first pair of strand exchanges

In the formation of Holliday junctions four possible combinations of strand exchanges can be made between the recombining DNA duplexes (this is illustrated in fig.4.13). The two *cer* sites are depicted as being aligned in parallel during the recombination reaction. There is, however, no experimental evidence available regarding how *cer* sites are aligned. This has been done simply to clarify the terms "top strand" and "bottom strand" when they are discussed.

Exchanges that are made between two top strands or two bottom strands result in Holliday junctions that are resolved by a second pair of strand exchanges between the two bottom and two top strands respectively. Both of these reaction mechanisms, when performed between directly repeated *cer* sites (as in pSD115), result in resolution of the substrate molecule to give two product circles (of 2.6 and 2.35 kbp in this case). Holliday junctions created by either of these strand exchange regimes would be indistinguishable by the experiments performed in the previous sections.

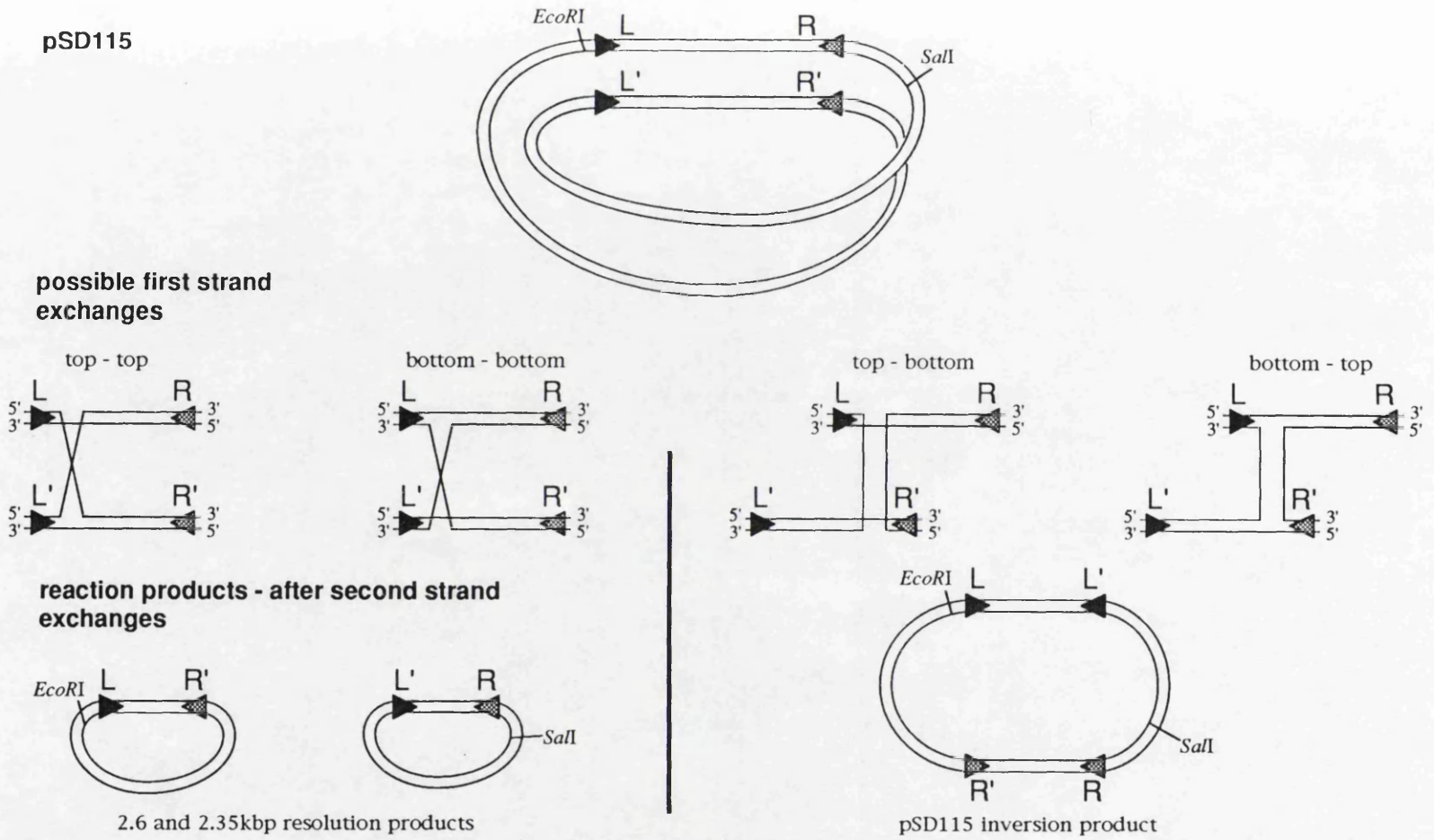


Figure 4.13. Diagram of the possible strand exchange combinations in the formation of Holliday junctions. The upper diagram shows the *cer* sites of pSD115 aligned in parallel before strand exchange. The crossover regions are represented by inverted arrows; the left and right arms of the crossover are denoted by L and R respectively; and one site is distinguished from the other by dashes at the arms (L' and R'). Four pairs of strand exchanges that would generate Holliday junctions are theoretically possible (as shown). The reaction products that would be generated by a second pair of strand exchanges (see text) on these structures are detailed.

It is also possible to create Holliday junctions by exchanging the top and bottom, or bottom and top strands of the recombining duplexes (fig.4.13). Resolution of these junctions would be by bottom to top, and top to bottom second strand exchanges respectively. Both these arrangements would fuse the left side of one recombining site to the left side of the recombining partner, and reciprocally right side to right side, resulting in inversion of the DNA intervening between the two *cer* sites of pSD115. Such a mechanism is unlikely because experimental evidence has shown that recombination of directly repeated *cer* sites always results in a resolution reaction and not inversion. However, formation of these junctions is theoretically possible and it is not clear whether the previous experimental evidence of this chapter can exclude their existence; for example, it is possible that the orientation of the *cer* sites is not detected during their recombination until after the formation of a Holliday junction.

In figure 4.13 the first pair of strand exchanges are shown as occurring towards the left-hand side of the *cer* core site. It should be noted that this is simply for convenience and is based on location of the point of first strand exchange being to the left of the spacer sequence in λ Int-catalysed reactions (if the *att* recombination sites are drawn in the same arrangement as the *cer* sites in figure 4.13: Kitts and Nash, 1988b; see Chapter 1). In the case of *cer* recombination the positions in the crossover sequence where the strand cleavages and exchanges take place are not known, and the definition of the spacer sequence (see fig.1.9) has been derived by comparing a number of *cer*-like sites. The first pair of strand exchanges may occur to the right of this presumptive overlap and be resolved, after branch migration, to the left. Indeed the initial strand exchanges could occur within any part of the "core" sequence and be resolved at any other position (even the same position) without altering the consequences of the four reaction schemes described, nor compromising the analysis of the nature of the first pair of strand exchanges described below.

Figures 4.14 and 4.15 show the details of the determination of the pair of strands exchanged in these *cer*-derived Holliday junctions. A 100ml culture of pSD115/RM40 was grown to mid-log phase and *xerC* expression induced by addition of 2mM IPTG.

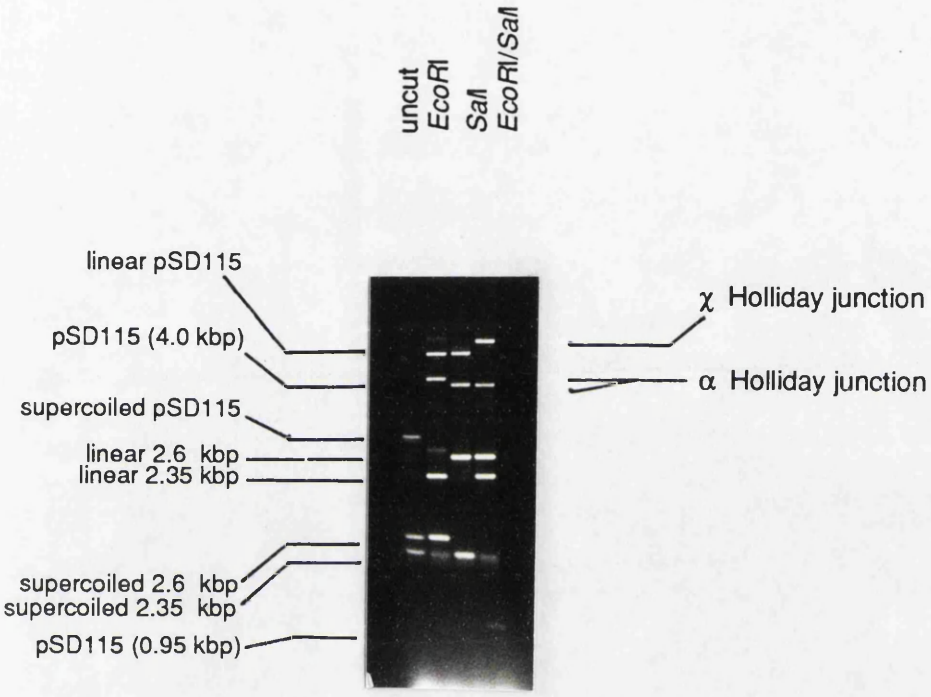
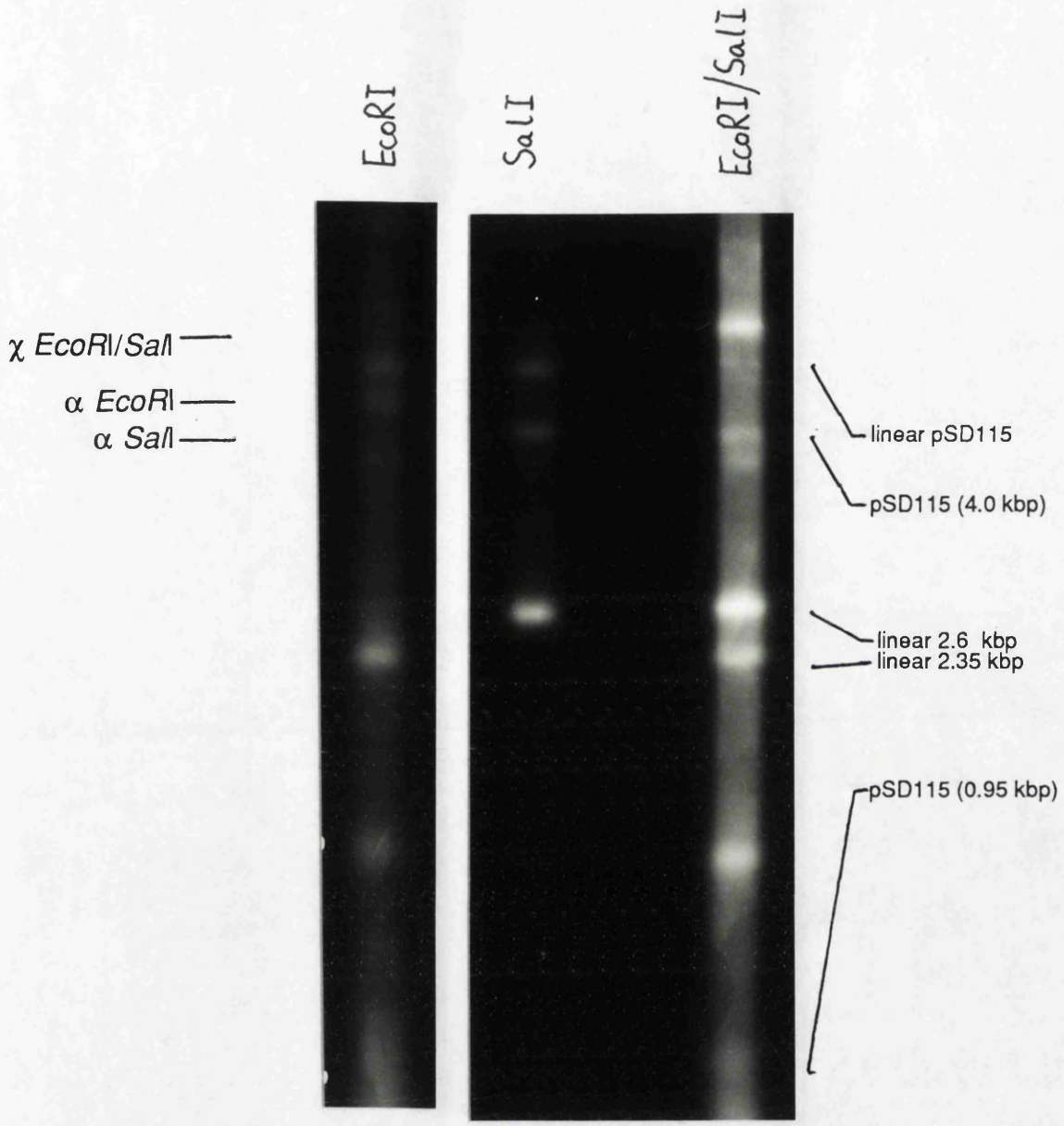


Figure 4.14. Asymmetry in the first pair of strand exchanges. (A) Generation of pSD115-derived Holliday structures by restriction digestion. Plasmid DNA representing the reporter plasmid pSD115 after 90 mins of *cer* mediated recombination in RM40 was restricted with either *EcoR*I, *Sal*I or both *EcoR*I and *Sal*I, and half the reaction volume was run on a 1.2% agarose gel. The α structures created by single restriction digestion, and the χ structures created by double digestion, are indicated.(Continued over page)



(B) End-labelling and purification of the Holliday junctions. The remaining halves of the digestions were end-labelled as described in the text and separated on a 1.2% Seaplaque low melting point agarose gel. DNA was visualised by autoradiography of the undried gel (it is shown here as a negative contact print), and the *EcoRI* α, *SalI* α and *EcoRI-SalI* χ structures (indicated) were then purified by phenol extraction.(Continued over page)

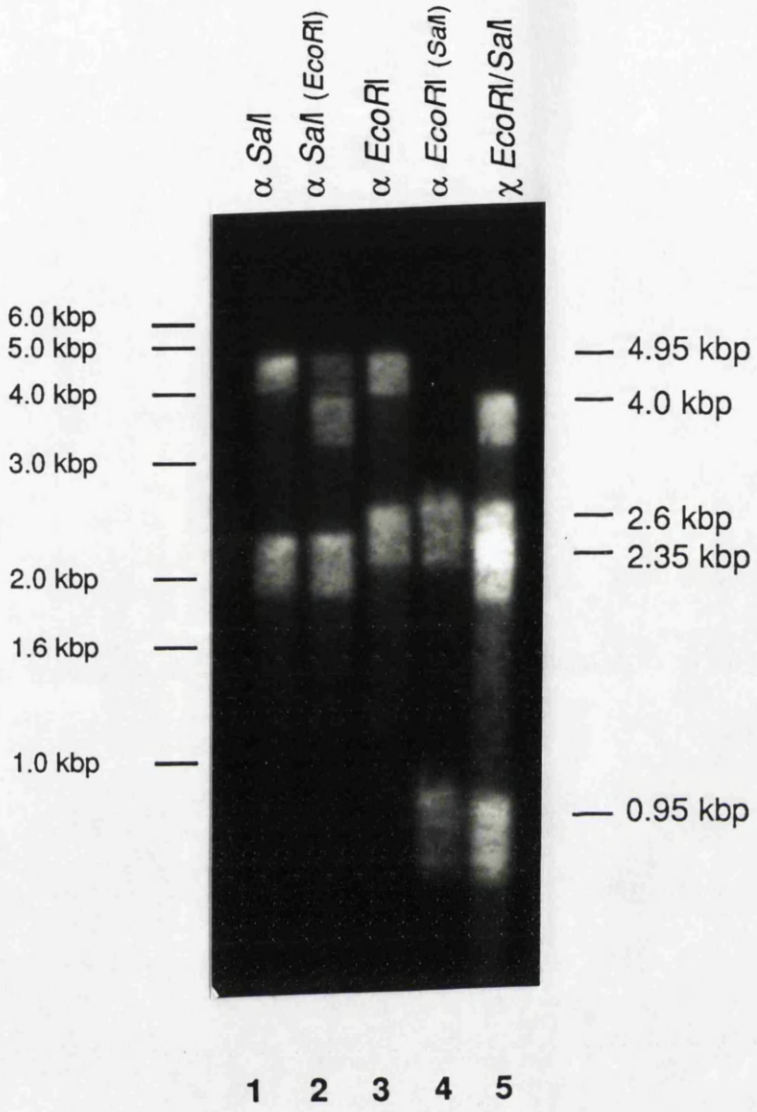


Figure 4.14(C) Analysis of the size of the single strands in isolated Holliday junctions. The purified, end-labelled Holliday junctions were run (with and without further restriction digestion) on a 1.2% alkaline, denaturing agarose gel and the DNA visualised by autoradiography after vacuum drying. The sizes of the labelled single strands are shown (they were derived from the 1 kbp BRL ladder markers on the same gel), and the predicted sizes for the four possible Holliday junctions are detailed in figure 4.15. The lane order was as follows:

Lane 1: *Sal*I α (undigested); lane 2: *Sal*I α (*EcoR*I digested); lane 3: *EcoR*I α (undigested); Lane 4: *EcoR*I α (*Sal*I digested); lane 5: *EcoR*I-*Sal*I χ (undigested).

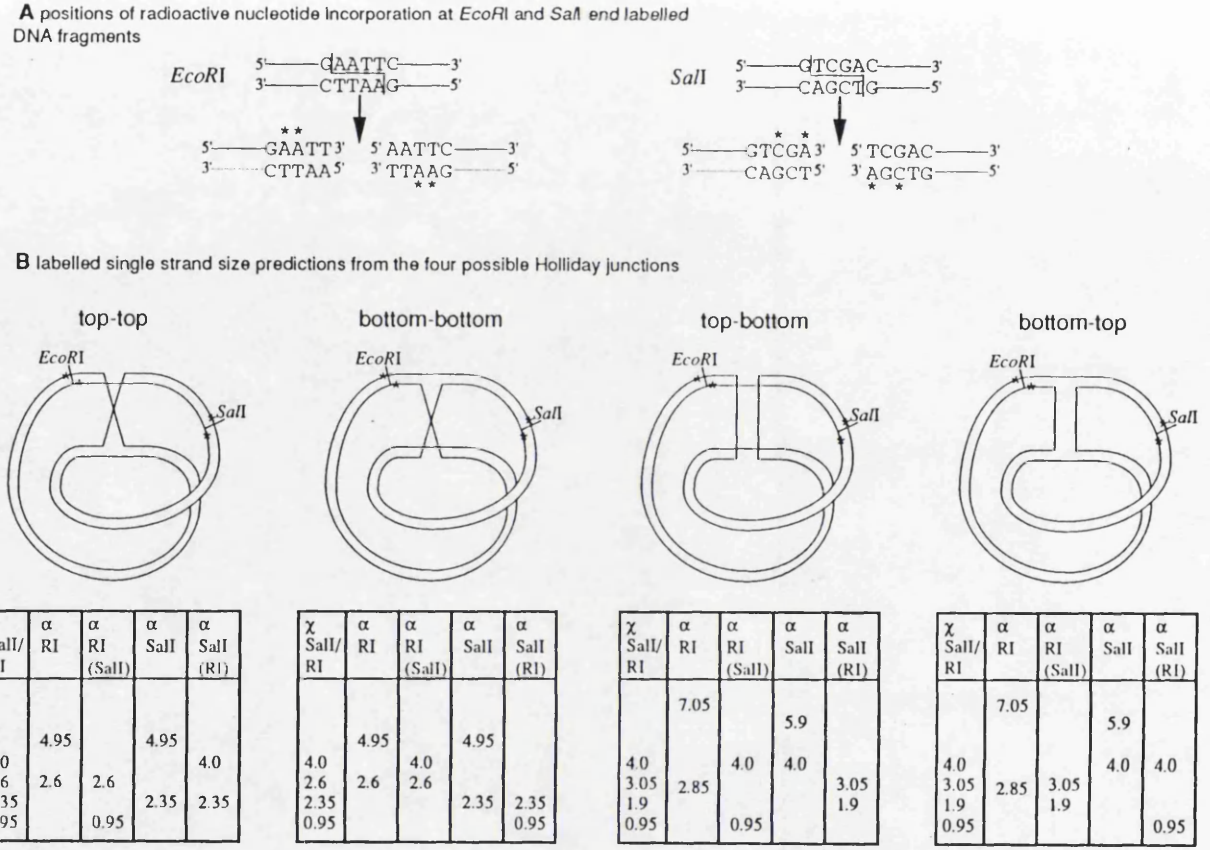


Figure 4.15. Single strand size predictions for the four possible Holliday junctions. **(A)** Diagrammatic representation of the position of incorporation of ^{32}P labelled nucleotides into *EcoR*I and *Sal*I digested DNA that has been end-labelled using the Klenow fragment of *E.coli* DNA polymerase I. The incorporated dATP and dCTP nucleotides are represented by an asterisk (*). **(B)** End-labelled, single strand sizes predictions for the four possible Holliday junctions detailed in figure 4.13. The organisation of the aligned *cer* sites and strand exchanges correspond to figure 4.13, and the ends of the labelled single strands are shown by an asterisk (*). The sizes of the strands are tabulated under the junctions and are shown in kbp for both the χ and α structures (before and after restriction digestion). *EcoR*I = RI and *Sal*I = SalI in this diagram.

After 90 mins of further growth the cells were harvested and DNA was prepared using Qiagen columns. 8ul of this DNA was restricted with either *EcoR*I, *Sal*I or *EcoR*I and *Sal*I, in a total volume of 20ul. Half (10ul) of these restrictions were separated by agarose gel electrophoresis (fig. 4.14A), the remainder was stored overnight at -20 ^{o}C. The gel in figure 4.14A showed the predicted restriction pattern for these DNA samples. *EcoR*I digestion created α structures and linearized pSD115 and the 2.6 kbp product; *Sal*I also made α structures (with different gel mobility) and linearized pSD115, but it linearized the smaller, 2.35 kbp, product; the double digest resulted in χ form Holliday junctions, linearized both products and digested pSD115 into 4.0 and 0.95 kbp fragments.

The stored restriction digests were then end-labelled in a total volume of 20ul. The *EcoR*I restricted DNA was labelled using α[^{32}P]dATP, *Sal*I and the double restricted DNAs used both α[^{32}P]dATP and α[^{32}P]dCTP. The way these end-labellings treat the ends of the DNA fragments are shown in figure 4.15A. Because the strand that incorporates the radioactive nucleotides can be assigned, the single strands within the four possible Holliday junctions that become end-labelled can be predicted (fig.4.15B). The end-labelled reactions were separated on a 1.2% "Sea plaque" low melting point agarose gel. Figure 4.14B shows an autoradiograph of this undried gel. For both the *EcoR*I and *Sal*I material the uncut product bands (2.35 and 2.6 kbp respectively) were not visible because they did not incorporate any label. The Holliday junctions which were cut from the gel are indicated: α bands from both the *EcoR*I and *Sal*I digests, and the χ band from the *EcoR*I and *Sal*I double digest. DNA was purified from these gel chips by phenol extraction, following the agarose manufacturer's instructions.

Half the purified *EcoR*Iα DNA was restricted with 5 units of *Sal*I at 37 ^{o}C for 1 hour, and half the *Sal*Iα was digested with *EcoR*I using the same reaction conditions. These samples were then separated, along with the unrestricted *EcoR*Iα, *Sal*Iα and *EcoR*I/*Sal*Iχ, on a denaturing agarose gel and the single-stranded DNA was visualized by autoradiography. BRL 1 kbp marker DNA (and pSD115 restricted with *EcoR*I and

*EcoR*I plus *Sal*I) was run on the same gel and visualized by ethidium bromide staining, thereby allowing the sizes of the single-stranded DNAs to be determined. The result of this experiment is shown in figure 4.14C.

The sizes of the single-strands from the unrestricted lanes (1, 3 and 5) showed that the isolated Holliday junctions resulted from either top-top or bottom-bottom first strand exchanges and not top-bottom or bottom-top (compare to the single-strand size predictions in fig.4.15B). This suggests that these stable Holliday junctions are not a consequence of aberrant "inversion" fusions between directly repeated *cer* sites in this *in vivo* assay. It may suggest that the *cer* site orientation is detected before the first pair of strand exchanges are executed, although the possibility of their reversal in the inversion Holliday junctions during pSD115 *cer* mediated recombination cannot be excluded.

Top strand and bottom strand exchanges were distinguished by *Sal*I restriction of the *EcoR*Iα and *EcoR*I restriction of the *Sal*Iα (lanes 2 and 4).The single strands produced by these restrictions were 2.6 and 0.95 kbp, and 4.0 and 2.35 kbp respectively. Both these results suggest that the top strands were exchanged to produce these Holliday junctions (see fig.4.15B). Even on longer exposures of this gel the fragment sizes that correspond to bottom strand exchanges were not seen, suggesting a very strong bias in *cer* recombination towards exchanging the top strands.

This ability, described in the previous sections, to detect Holliday junctions is extremely unusual. In other λ integrase-like recombination reactions they have only been detected under specific *in vitro* reaction conditions. This is almost certainly because they are intermediates in the reaction pathways (Craig, 1988), and are consequently short-lived. Because of the amount of DNA seen in the form of Holliday junction at a given time point during *cer* reaction in RM40, the following sections describe experiments that address whether their isolation is a result of an artefact of the *in vivo* system I have employed.

4.9 Construction of RM50, and *in vivo* recombination in this strain

In the construction of strain RM40, partial deletions were made in the genes *dapF* and *orf235* (section 4.2), probably making the proteins encoded by these genes inactive. The relationship, if any, between the role of XerC and the three other proteins encoded by the operon is not understood. No cellular roles nor homologies to other proteins have been identified for the proteins encoded by *orf235* and *orf238*. It is therefore possible that the large quantities of Holliday junctions seen during recombination assays in RM40 are a result of these mutations. If this were the case then a role for DapF, ORF235 or ORF238 in the *cer* recombination reaction may be assigned, thereby perhaps offering some insight into the reason why *xerC* is part of an operon. This section describes strain RM50, in which the *lac*PO and resistance determinants have been placed upstream of *dapF* so that on induction with IPTG the whole operon is transcribed.

The transcriptional start site of the operon has been mapped to 269 nucleotides upstream of the *dapF* start codon (Richaud and Printz, 1988). A *Pst*I recognition site 146 nucleotides upstream of the *dapF* strart codon was chosen as the site of insertion of *lac*PO, since it would then be downstream of the operonic promoter (see fig.4.16).

The first step in the construction of strain RM50 was subcloning of a 2.1 kbp *Pvu*II-*Stu*I fragment from pSD102 into *Pvu*II digested pUC18 (fig.4.16). This cloning replaced the pUC18 polylinker, *lac*PO region and part of the *lacZ* gene with the *E.coli* chromosomal region encompassing the *cyaX* gene, *xerC* operonic promoter and most of the *dapF* gene, as well as some plasmid sequences from pTZ18R. The plasmid that was created is called pRM70, and the purpose of the sub-cloning was to make the *Pst*I site unique, thereby facilitating insertion of the transcriptional control apparatus.

In this construction it was again felt necessary to include a transcriptional terminator upstream of the *lac* promoter to reduce transcriptional readthrough. Kanamycin was used as the antibiotic resistance to select for the strain after linear transformation, as in RM40. Restriction of pRM102 with *BamH*I released a 2.39 kbp

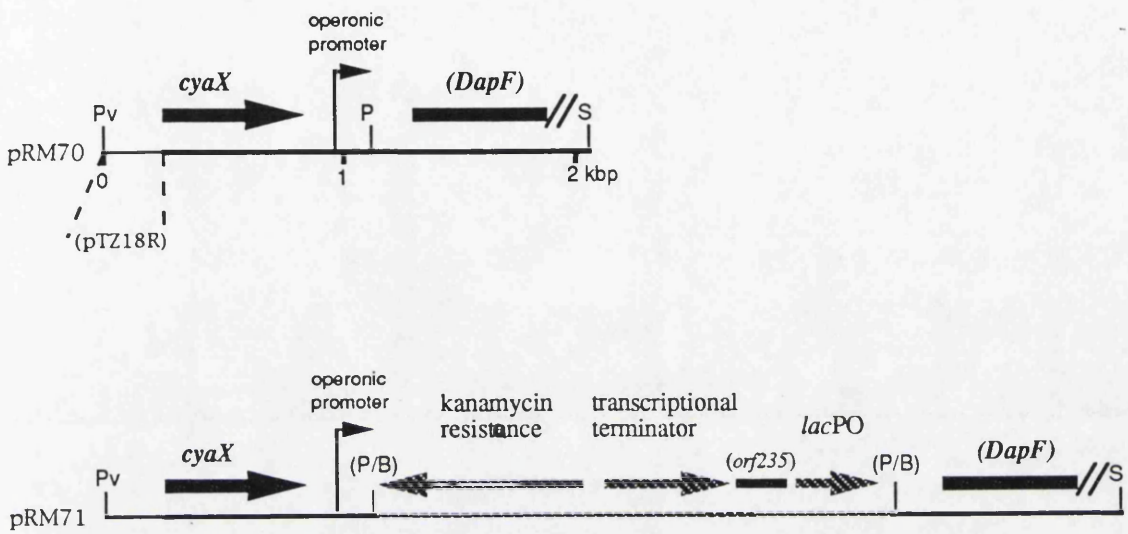


Figure 4.16. Plasmid clonings involved in the construction of strain RM50. The top plasmid diagram shows the 2.1 kbp *Pvu*II-*Stu*I fragment cloned from pSD102 (Colloms *et al*, 1990) into pUC18, creating pRM70. This plasmid contains the *E.coli* chromsomal gene *cyaX* and part of the *dapF* gene; a unique *Pst*I site is found between the transcriptional start site and open reading frame of *dapF*. The lower diagram shows the plasmid pRM71 that was created by inserting a 2.39 kbp *BamH*I fragment from pRM102 (this thesis; fig.4.1) into the unique *Pst*I site. The insert in pRM71 is orientated such that the *lac* promoter and operator sequences can direct transcription into *dapF*; the kanamycin resistance gene allows the selection of this DNA after linear transformation into the strain JC7623.

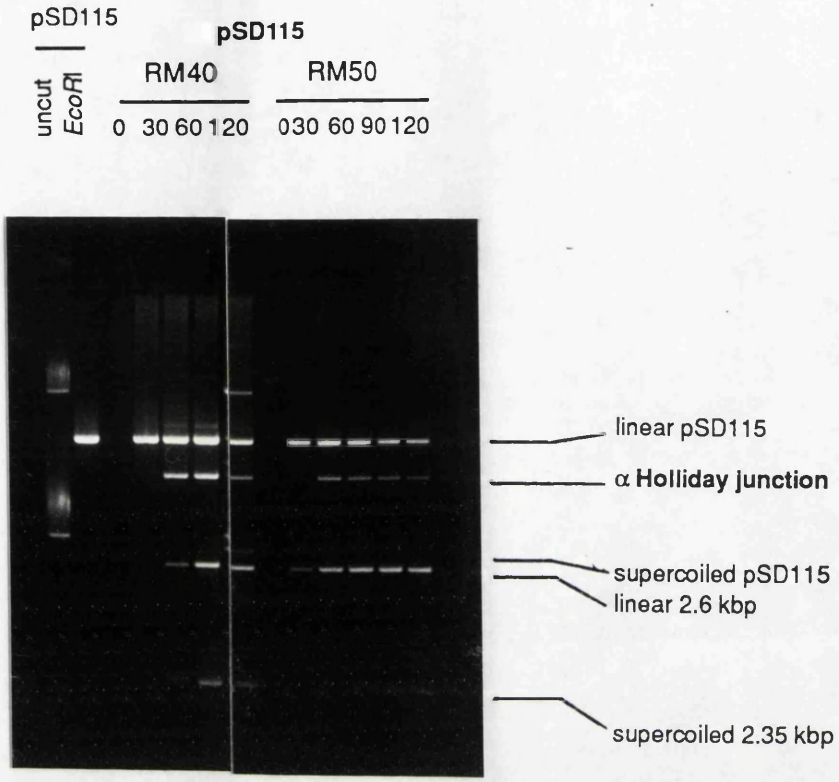


Figure 4.17. Comparison of pSD115 time course recombination *in vivo* using the strains RM40 and RM50. RM40 and RM50 were both transformed with pSD115, and XerC expression was induced with 2 mM IPTG in both cultures as they reached mid-log phase. DNA samples were prepared by boiling preparations at the times shown after induction, digested with *EcoR*I and run on a 1.2% agarose gel. Holliday junction structures are indicated (α). The plasmid pSD115 and its *cer* mediated recombination products are described in figure 4.4.

fragment containing *lac*PO, the kanamycin resistence gene and ribosomal RNA terminator (see section 4.2/ fig.4.3). The fragment also contained 350 bp of *orf235* sequence from between the *lac* promoter and kanamycin/terminator sequences, but it was felt this would not cause problems by homologous recombination (e.g. deletion) when inserted into the chromosome. The 2.39 kbp *BamH*I fragment was cloned into the *Pst*I site of pRM70 and the correct orientation, such that *lac* promoter-derived transcription would read into *dapF*, was selected by *EcoR*I restriction. The plasmid that was created in this way is called pRM71.

pRM71 was restricted with *Sst*I (which cuts uniquely in the pTZ18R sequence and therefore linearizes pRM71) and transformed into JC7623. Twenty-five clones were recovered by this transformation and all were resistant to kanamycin and sensitive to ampicillin, indicating that that the construct had recombined into the chromosome. Three were transduced into DS941 with bacteriophage P1, creating strains RM50, RM60 and RM70. Only RM50 was subjected to further analysis.

RM50 and RM40 were transformed with pSD115 and grown on L-agar plates supplemented with 1% glucose, 50 ug/ml diaminopimelic acid and ampicillin. Separate 20 ml cultures were set up in L-broth containing 1% glucose and ampicillin, using large numbers of colonies from the plates. The two cultures were grown to mid-log phase and IPTG was added at a final concentration of 2mM. DNA was prepared at the point of induction and every 30 mins subsequently for 2 hours. The DNA was analysed by agarose gel electrophoresis after *EcoR*I digestion (fig.4.17).

RM50 behaved in a similar manner to RM40, i.e. the amount of reaction products increased linearly after induction of the *xerC* operon. A greater amount of replicative product was observed at time 0' in RM50 than in RM40. This might be due to fortuitous transcription initiation downstream of the the *lac* promoter, although I have no evidence to support this contention. *EcoR*I restriction revealed the presence of α structures during RM50 recombination, which correspond to the Holliday junctions that are seen during *cer* recombination in RM40. Furthermore, there appeared to be similar quantities of the junctions during the RM50 time course and the RM40 time course.

These results suggest that when the entire operon is controlled in the same way as the *xerC* gene alone, the *in vivo* recombination reaction still involves production of stable Holliday junctions. This indicates that their production is not a consequence of the mutations in *dapF* or *orf235* present in RM40, since RM50 almost certainly has no such mutations. The roles these genes may have in *cer* recombination, if they indeed they have any roles, are not clarified by this experiment.

4.10 Controlled *in vivo* recombination using the plasmid pRM60

Another possible explanation for the stable Holliday junctions is that both RM40 and RM50 contain an unidentified mutation that causes the accumulation of reaction intermediates. Such a mutation could be in either *xerC* (the presumptive recombinase gene), or another gene, for instance the *cer* accessory factors.

in vitro recombination assays with Cre mutant proteins (Hoess *et al*, 1987 and Wierzbicki and Hoess, 1987), designated HA136, HA15 and NA14-2, demonstrated that a Holliday junction intermediate accumulated which was not seen in the same conditions using wild type Cre. The mutants were all capable of producing reaction products *in vitro* and the amount of DNA present in the intermediate varied. Although it was reported that the amount of recombination that the mutants supported *in vivo* varied, ranging from 100% in comparison to wild type Cre (NA14-2) to 0% (HA136), it was not discussed whether Holliday junctions could be observed *in vivo*. The amino acid changes in the three mutant Cre proteins are dispersed throughout the protein's sequence: positions 91, 159 and 312. Only the NA14-2 change (alanine312 to threonine) is within either of the conserved λ integrase-like domains (Abremski and Hoess, 1992), but it is not one of the better conserved amino acids within the domain (e.g. the equivalent amino acid in XerC is Leucine). Why these mutants accumulate reaction intermediates is not known, but it may be related to the fact that HA136 and HA15 (but not NA14-2) are reduced in their binding affinity for *loxP*.

The λ Int mutant derivative Int-h, which will undergo some recombination without the normally essential accessory factor IHF, has been shown to accumulate intermediates *in vitro* in which only one pair of strands have exchanged (Kitts and Nash,1987, 1988a and 1988b). In these experiments, however, the amounts of Holliday junction seen were very low and required the inclusion of phosphorothioate groups in the *att* site.

To determine whether mutations in RM40/50 are the cause of the stable Holliday junctions during *cer* recombination, another XerC expression vector, pRM60, was made. This plasmid was made separately, and from different starting DNA, to the plasmids involved in the construction of strains RM40 and 50, and is used in different strain backgrounds. Construction of pRM60 is detailed in figure 4.18.

A 1.2 kbp *xerC*-containing *Hind*III-*Eco*RI fragment from pSD105 (Colloms, 1990) was cloned into *EcoR*V restricted pCT1050 (a gift of R.Thompson), creating the plasmid pRM50. The correct orientation of insertion, such that the λP_L promoter would drive expression of *xerC*, was selected by *Sal*I restriction. pRM50 was then restricted with *EcoR*I and *BamH*I, releasing a 1.6kbp fragment containing the *xerC* gene and upstream λP_L promoter. This fragment was cloned into pGP1-2Cm (Tabor and Richardson, 1986) and pGP1-2Km (a gift from S. Rowland), both digested with *EcoR*I and *BamH*I. The only difference between the two plasmids generated in this way, called pRM60 and pRM65, is that they carry chloramphenicol and kanamycin resistance genes respectively. In both plasmids the T7 polymerase gene has been replaced by *xerC*, and in both *xerC* expression is driven by λP_L.The *cI857* gene in pRM60/65 encodes the temperature sensitve CI repressor, meaning that when they are grown at 28-30 ^oC CI is active and represses *xerC* expression, if grown at 37-42 ^oC CI is thermally inactivated and XerC is made. Because CI857 is expressed from the *lac* promoter in these plasmids it was felt that in DS941 strains (which are $lacI^q$) the control of XerC expression would not be "tight", therefore DS942 (which is $lacI^-$) *xerC* derivatives were made by P1 transduction of the *xerC*Y17 (choramphenicol resistant) and *xerC*PS6 (kanamycin resistant) alleles from the strains DS984 and DS981 respectively.

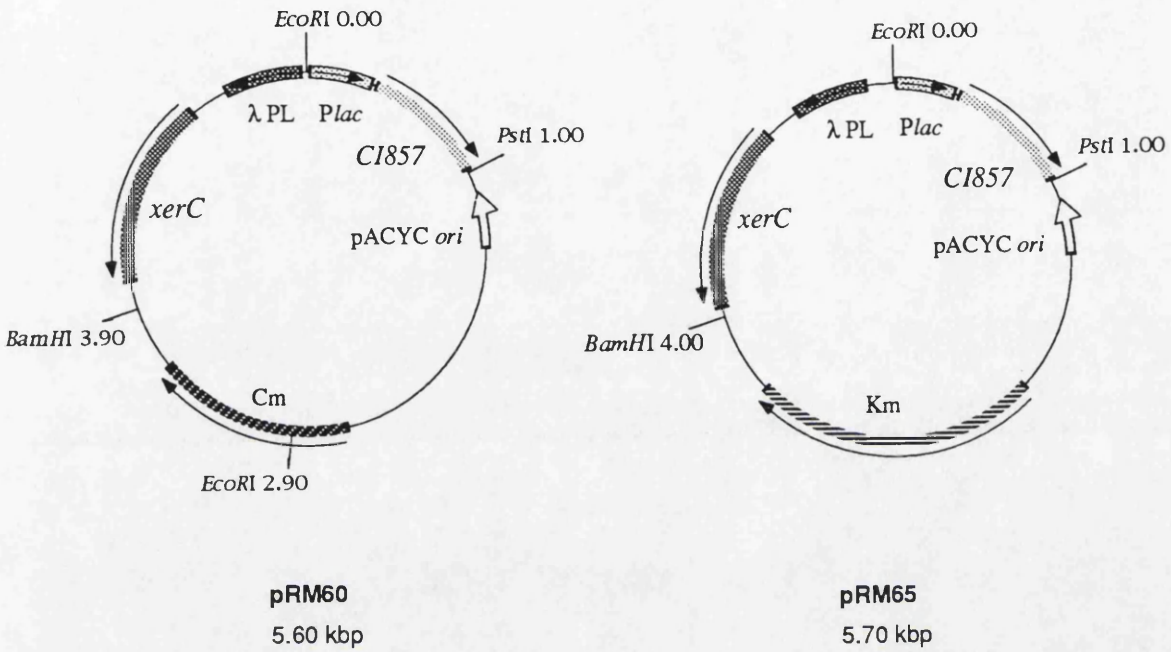


Figure 4.18. Diagrammatic representation of the plasmids pRM60 and pRM65. These plasmids are equivalent in all respects except the antibiotic resistence genes they carry (pRM60 is chloramphenicol resistant and pRM65 is kanamycin). *CI857* (expressed from the *lac* promoter) encodes a temperature-sensitive derivative of the λ CI repressor, thus the expression of *xerC* (driven from the λ P_L promoter) is switched off when grown at 28-30 ^{o}C and turned on at 37-42 ^{o}C.

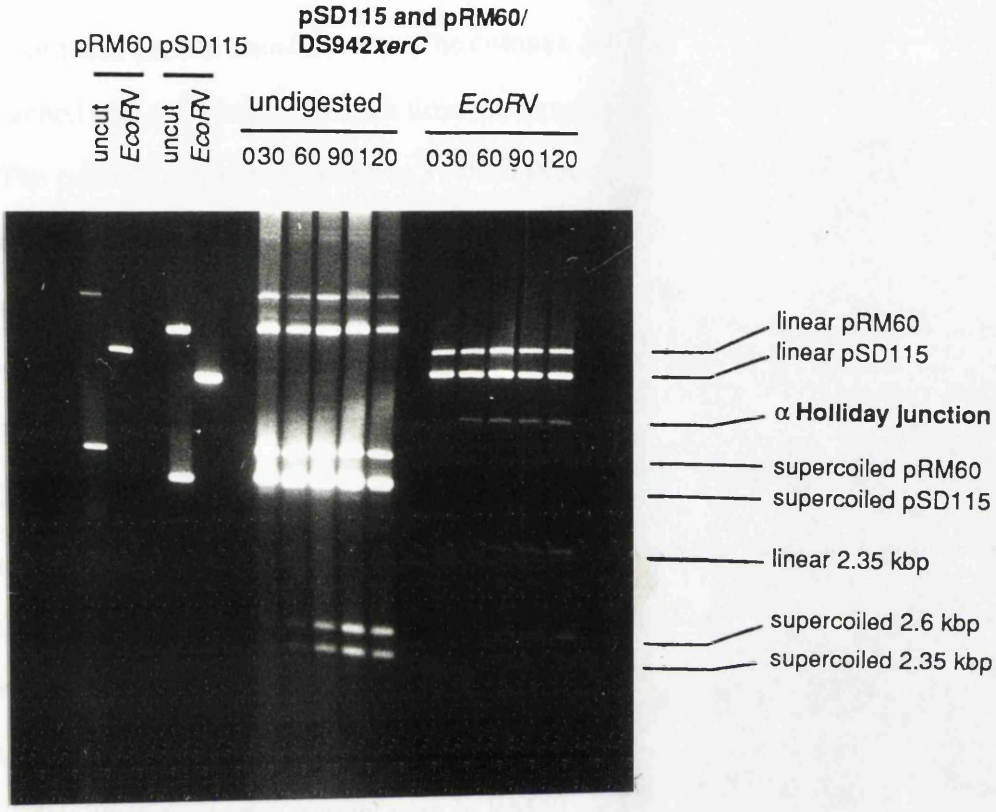


Figure 4.19 pSD115 *in vivo* recombination time course controlled by the plasmid pRM60. DS942*xerC*::Kan was transformed with pSD115 and pRM60 and was grown at 30 °C to mid-log phase. XerC expression was induced by elevating the culture temperature to 42 °C and plasmid DNA was prepared at the times subsequent to induction as shown. The DNA was run on a 1.2% agarose gel before and after restriction digestion by *EcoR*V. Maps of pRM60 are shown in figure 4.18, maps of pSD115 and its *cer* recombination-derived reaction products are shown in figure 4.4. Holliday junction structures made during the recombination time course are indicated (α).

DS942*xerC*PS6 was transformed sequentially by pRM60 and pSD115 and grown at 30 °C on L-agar plates containing chloramphenicol and ampicillin. 30 ml L-broth cultures containing chloramphenicol and ampicillin were then set up using a large number of these double transformants. The cultures were grown shaking at 30 °C until they reached mid-log phase, at which time the temperature was raised to 42 °C for 30 mins. The cultures were then moved to 37 °C. DNA samples were made by the boiling technique at the point of XerC induction and every 30 mins. afterwards until 120 mins. had elapsed. The DNA was analysed by gel electrophoresis without restriction and after restriction by *EcoR*V (fig.4.19).

The unrestricted time course (lanes 1-5) showed that this *in vivo* system works in a similar way to RM40, i.e. the two resolution products of pSD115 recombination increased in quantity with time, and before XerC induction there was little, or no, recombination. *EcoR*V restriction (lanes 6-10) linearized pRM60, pSD115 and the 2.35 kbp product. It also revealed the presence of the Holliday junctions by converting them into α structures. It is not really possible to compare the proportion of DNA in the form of Holliday junctions to the amount seen in RM40 assays since the DNA has been run on different gels, but it appeared to be similar.

The plasmids pRM102 and pRM60 were constructed at different times in separate experiments. The starting plasmids used in their construction are different and their DNA stocks were prepared separately. For this reason the use of pRM60 acts as a control for mutations in *xerC* that may have arisen during the construction of pRM102 (or subsequent steps in the construction of strains) which could result in a phenotype characterised by Holliday junction accumulation. It is conceivable that such mutants are very common and are present in both plasmids, but unlikely. The plasmid pSD105 (the progenitor of pRM60) was derived from pSD102 (the progenitor of pRM102), so it is possible that pSD102 could contain such a mutation which would thus be present in both plasmids. However, to determine if this were the case the *E.coli* chromosomal copy of *xerC* would have to be re-cloned, because pSD102 is a deletion derivative of the original *xerC* clone (pSD100) made by S. Colloms. Such an experimental approach

would be time-consuming and not guaranteed to yield a definitive result, because identifying a mutation of this kind is problematic (see above).

Because the strains used in this experiment are different, and were constructed separately, I believe that it is unlikely that the stable Holliday junctions are the result of mutations of another gene or genes. Furthermore, the method used to induce expression of XerC is different in both systems, suggesting that the means of induction (IPTG) is not resulting in these Holliday junctions in RM40.

The above experiment cannot rule out the possibility that this feature of *cer* recombination is a consequence of the systems being employed, but the fact that two distinct systems (in character and in construction) both show the presence of stable Holliday junctions suggests that this is a general feature of pSD115 recombination.

4.11 Do pKS455 and pCS202 make stable Holliday junctions?

All the recombination assays I have so far described have used the *cer* reporter plasmid pSD115 as a substrate. It is therefore possible that the Holliday junctions accumulate because of some feature of this plasmid. Altered recombination sites which have this effect *in vitro* have been described in other λ integrase-like systems. Nunes-Duby *et al* (1987) showed that recombination between a wild type *att* site and a site containing a nick in its overlap sequence accumulated single strand exchange intermediates; Cowart *et al* (1991) demonstrated that chemically crosslinking one *att* site had the same effect. Clearly, however, neither of these site alterations could be present in the *in vivo* recombination assays in this work. Kitts and Nash (1987, 1988) produced evidence that recombination between a wild type *att* site and certain "*saf*" *att* site mutants (which have altered overlap sequences) could result in accumulation of Holliday intermediates. It is conceivable that the DNA preparations of pSD115 used in the experiments described here have heterologous *cer* sites analogous to this, although

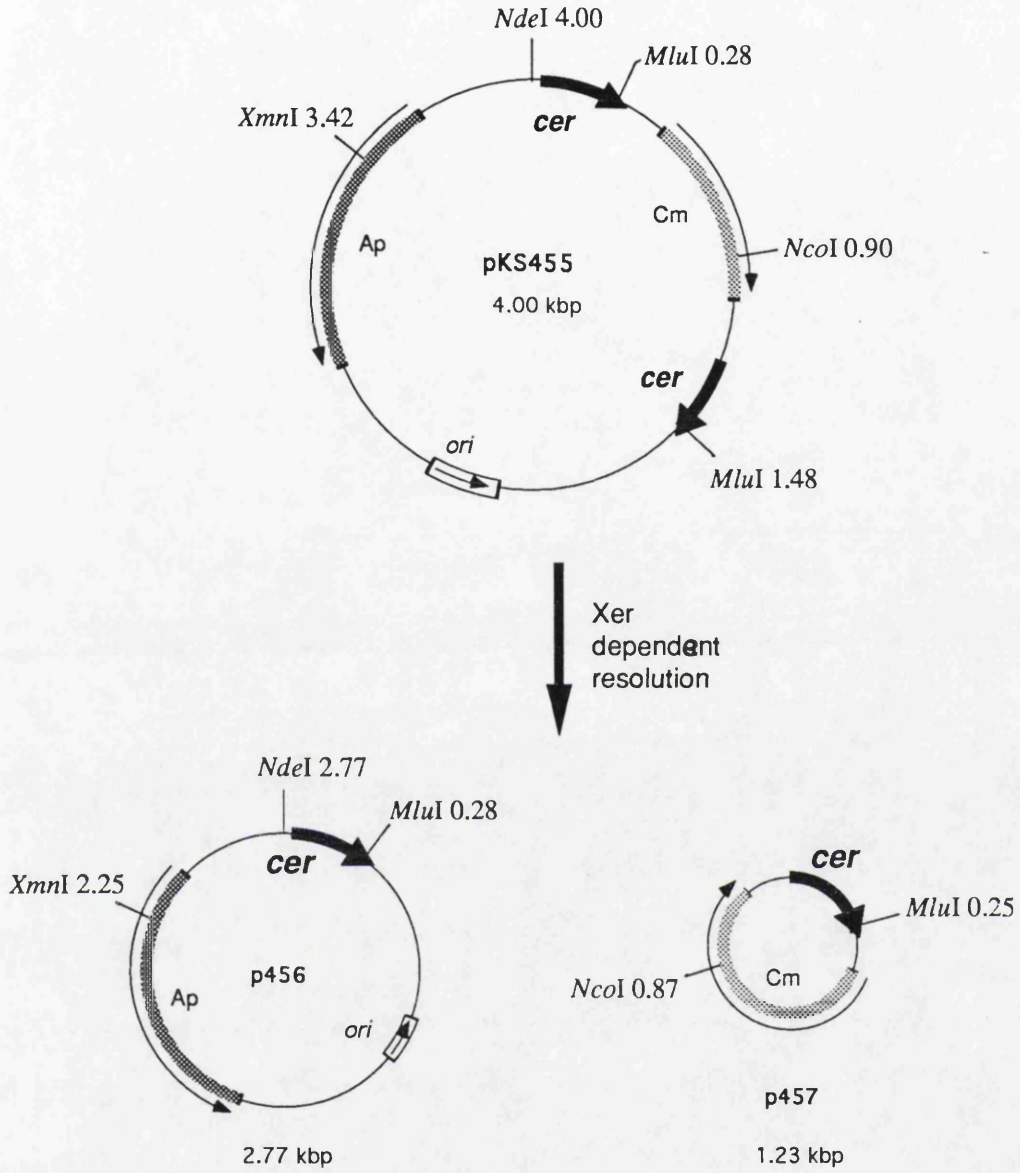


Figure 4.20. Diagrammatic representation of Xer-dependent resolution of the *cer* reporter plasmid pKS455. pKS455 was constructed in pUC8 by D.Summers (1989). The plasmid is resolved in Xer$^+$ strains to yield a 2.77 kbp product named p456 and a 1.23 kbp product mamed p457. Note that only p456 contains a replication origin, and hence p457 is rapidly lost from bacterial cultures.

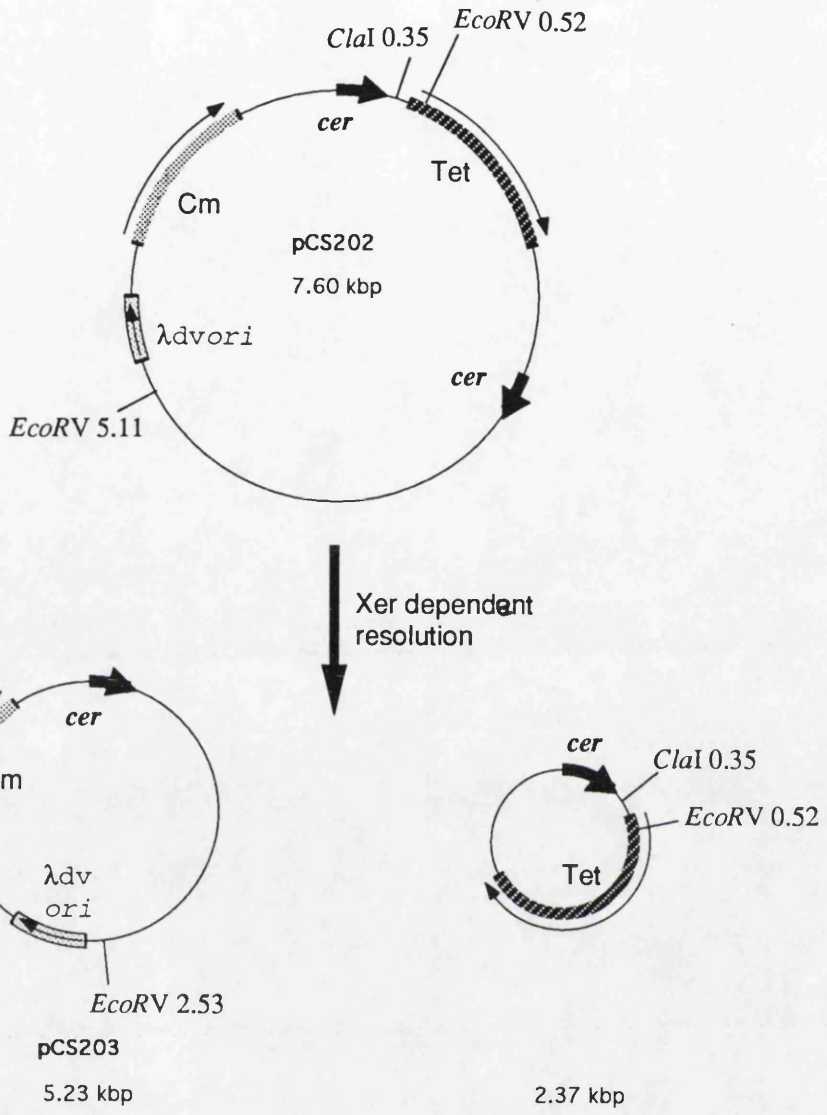


Figure 4.21. Diagrammatic representation of Xer-dependent resolution of the *cer* reporter plasmid pCS202. pCS202 was constructed by C.Stirling (1987) and contains a λdv origin of replication. The plasmid is resolved in Xer$^+$ strains to yield a 5.23 kbp product named pCS203 and an unnamed 2.37 kbp product. Note that only pCS203 contains a replication origin, and hence the smaller recombination product is rapidly lost from bacterial cultures.

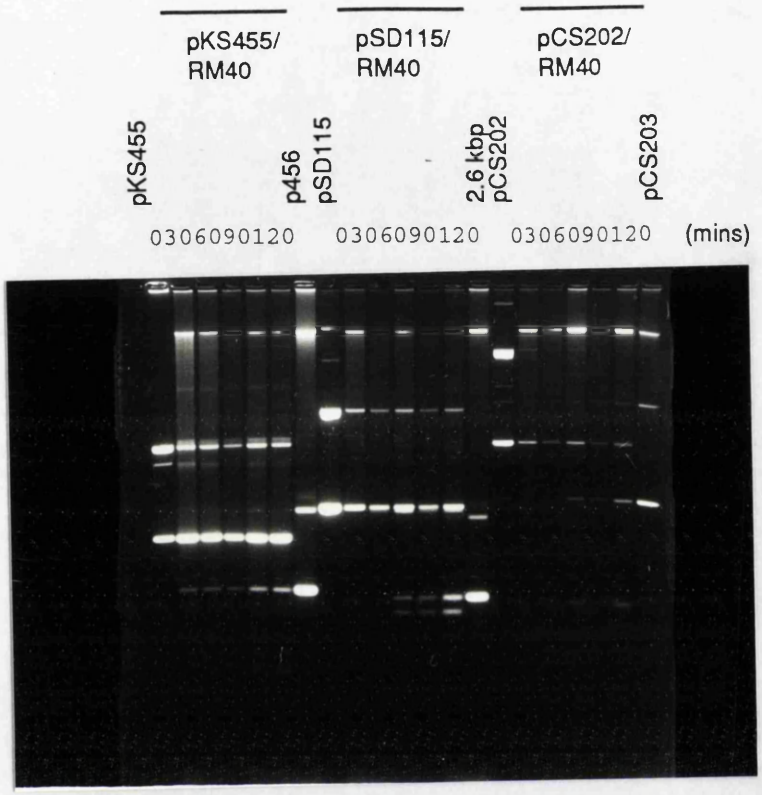


Figure 4.22. Comparison of pSD115, pKS455 and pCS202 recombination in RM40. RM40 was transformed with pSD115, pKS455 and pCS202 and *in vivo* time course recombination assays were performed using cultures set up from pooled transformants. XerC expression was induced using 2 mM IPTG as each culture reached mid-log phase. DNA from each assay was prepared by the boiling technique at the times subsequent to XerC induction shown and was run, unrestricted, on a 1.2% agarose gel. Resolution of all three reporter plasmids by *cer* mediated recombination yields two circular reaction products which are described in figures 4.4, 4.20 and 4.21 (pSD115, pKS455 and pCS202, respectively).

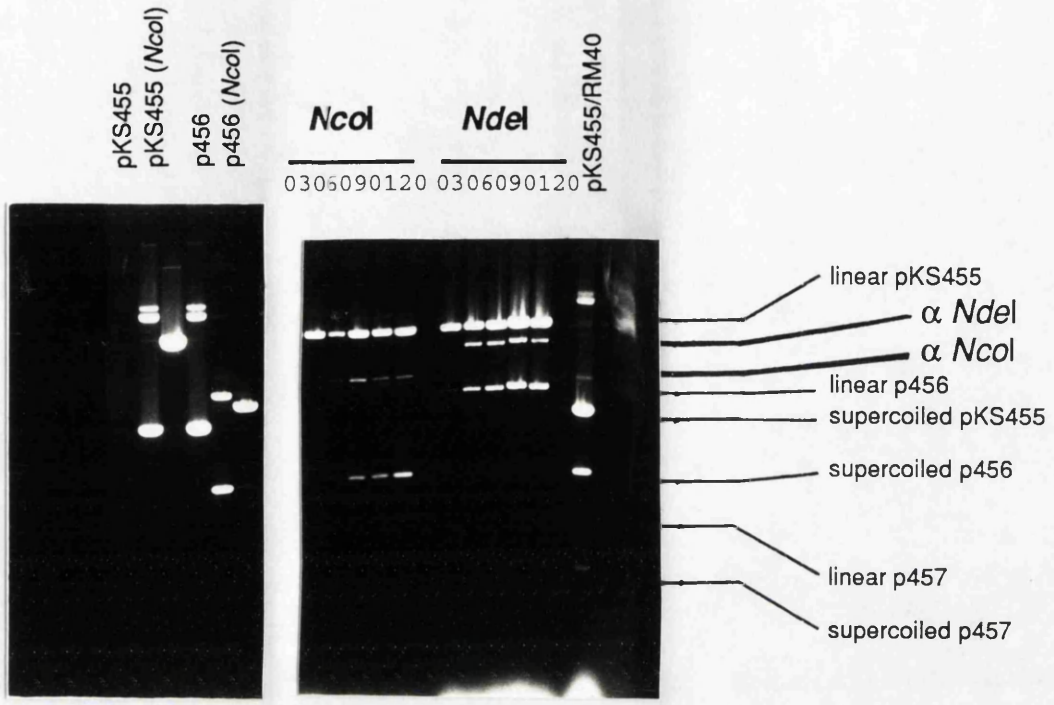
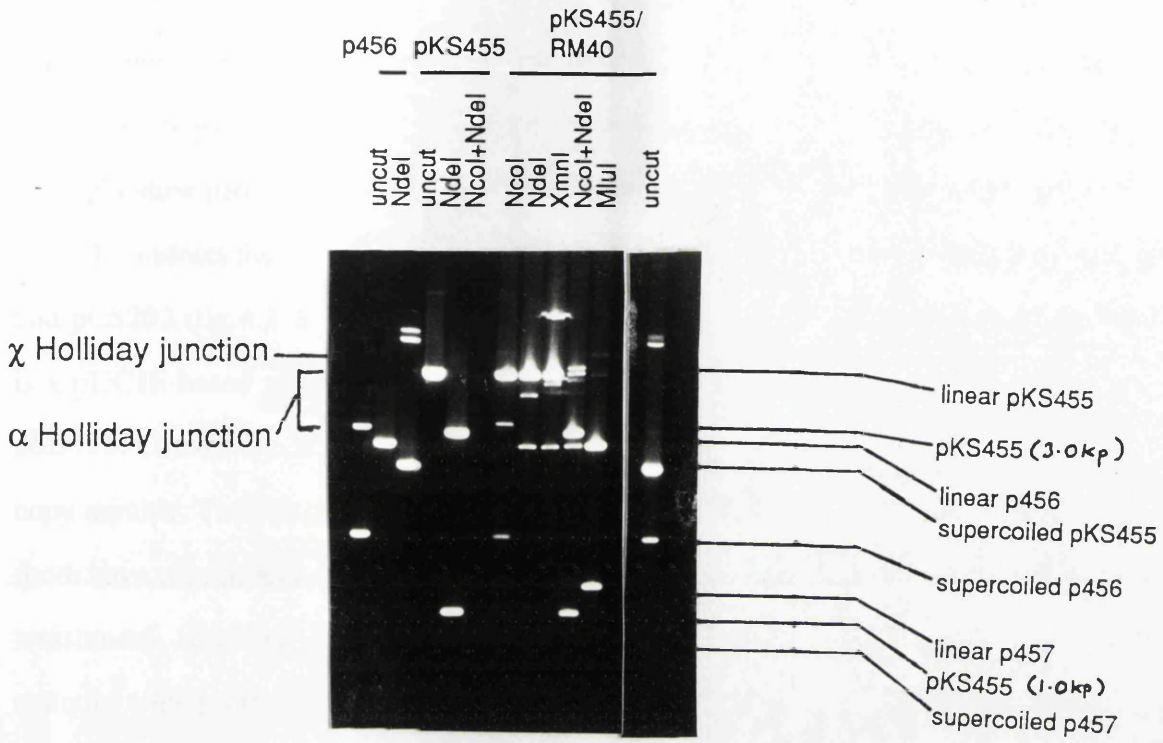


Figure 4.23. Restriction analysis of pKS455 *in vivo* recombination time course. (A) Generation of α structure Holliday junctions. DNA samples from the pKS455 recombination time course produced in the strain RM40 and shown in figure 4.22 were digested with either *Nco*I or *Nde*I and run on a 1.2% agarose gel. *cer* mediated resolution of pKS455 yields two products: a 2.77 kbp, replicative product (p456) and a 1.23 kbp, non-replicative product (p457). Restriction maps of pKS455 and its reaction products are described in figure 4.20. Holliday junction structures are indicated (α).



(B) Further restrictions of pKS455-derived Holliday junctions. DNA samples representing pKS455 after 135 mins of recombination in RM40 (see figure 4.22) were digested with either *Nco*I, *Nde*I, *Xmn*I, *Nco*I and *Nde*I, or *Mlu*I and run on a 1.2% agarose gel. Double restriction of pKS455 with *Nco*I and *Nde*I produces 3.0 kbp and 1.0 kbp linear DNA fragments (indicated as pKS455). α and χ structures corresponding to pKS455-derived Holliday junctions are indicated.

it should be noted that whether recombination using *att* site *saf* mutants generates stable Holliday junctions *in vivo* has not been reported.

The reason for the accumulation of *cer* Holliday junctions may not be a consequence of the *cer* sites themselves in pSD115, but could be because of some other feature of the plasmid. Possibilities could include the transcription pattern of pSD115, the replication pattern, or even some unusual secondary structure it could be adopting.

To address these possibilities two other *cer* reporter plasmids, pKS455 (fig.4.20) and pCS202 (fig.4.21), were used in *in vivo* recombination assays in RM40. pKS455 is a pUC18-based plasmid and therefore has a higher copy number in the cell than pSD115. Conversely, pCS202 has a λdv origin of replication and consequently a lower copy number. The plasmids also contain different antibiotic resistance genes to pSD115 (both have chloramphenicol resistance genes, and pCS202 has the gene for tetracycline resistance), resulting in different patterns of transcription. Furthermore, pKS455 contains a *lac* promoter that is induced by IPTG and results in transcription across one *cer* site.

The plasmids were transformed into RM40 and plated onto L-agar containing 1% glucose, 50 ug/ml diaminopimelic acid and their respective antibiotics. The recombination assays were performed in the same manner as previously described for pSD115. DNA was prepared at the time points shown in figure 4.22, which compares unrestricted time courses of the three plasmids. The differences in plasmid copy are reflected in the differing amounts of DNA visible in each assay. In all cases the amounts of reaction products (which are described in figs. 4.20 and 4.21) increase with time, and all produce the non-replicative product circles.

Figure 4.23A shows the same DNA samples for pKS455 as in figure 4.22 but this time restricted with *Nco*I (which cleaved once and linearized the substrate and small product) and *Nde*I (which linearized the substrate and large product). The bands corresponding to the α Holliday junctions are marked, and were only visible after IPTG induction. In *Nco*I restriction, they had a gel mobility almost indistinguishable from the nicked large product, but *Nde*I altered their mobility and they were readily

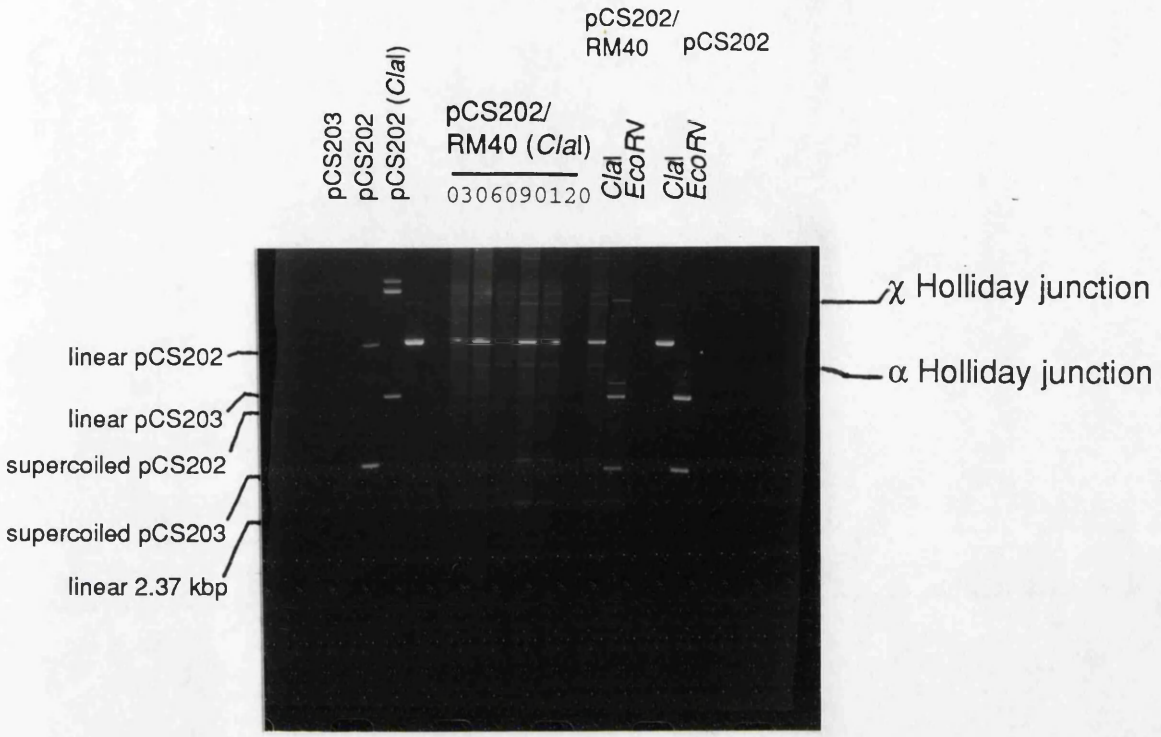


Figure 4.24. Restriction analysis of pCS202 *in vivo* recombination time course. DNA samples from the pCS202 recombination time course produced in the strain RM40 and shown in figure 4.22 were digested with *Cla*I and run on a 1.2% agarose gel; in addition a DNA sample representing pCS202 after 135 mins recombination was restricted with *EcoR*V and run on the same gel. *cer* mediated resolution of pCS202 yields two products: a 5.23 kbp, replicative product named pCS203 and an unnamed 2.37 kbp, non-replicative product. Restriction maps of pCS202 and its reaction products are described in figure 4.20. Digestion of pCS202 with *EcoR*V produces 5.0 kbp and 2.2 kbp linear fragments (indicated as pCS202). α and χ structures correponding to pCS202-derived Holliday junctions are indicated.

seen. This change in α structure mobility, and the fact that they migrated a distance that is intermediate between linear and supercoiled substrate, is consistent with pSD115 α Holliday junctions.

Figure 4.23B shows further restrictions of these DNA samples which confirm that pKS455 makes Holliday junctions during recombination in RM40. The figure shows the pKS455/RM40 135' sample restricted by *Nco*I, *Nde*I, *Xmn*I, *Nco*I and *Nde*I, and *Mlu*I. *Xmn*I altered the mobility of the α structure further in comparison to *Nco*I and *Nde*I digests, which is again consistent with previous analysis using pSD115. Doubly restricting the DNA with *Nde*I and *Nco*I linearized both products and cleaved pKS455 into 3.0 and 1.0kbp fragments; it also altered the mobility of the Holliday junction so that it migrated slower in the gel than linear pKS455, which is consistent with the DNA adopting a χ structure. The *Mlu*I digest also linearized both products, but cleaved pKS455 in product sized fragments because it cuts within the two *cer* sites. As seen with pSD115, *Mlu*I restriction allowed the χ structure Holliday junction to branch migrate free, resulting in linear product molecules. (Note that both the *Mlu*I restriction and the *Nde/Nco*I double restriction were not entirely complete and linear pKS455 could be seen in both, as well as some nicked Holliday junction in the *Mlu*I lane.)

Figure 4.24 shows similar restriction analysis done with the pCS202 time course. Restriction of the time course with *Cla*I linearized the substrate plasmid and small product of the reaction and converted the supercoiled Holliday junctions into α structures, which are indicated. pCS202 does not have as many useful sites as pKS455 or pSD115 (for instance *Mlu*I does not cut solely within the *cer* sites) and therefore the restriction analysis was not as exhaustive. *Eco*RV restriction of the 135' time point of pCS202/RM40 linearized both products and digested pCS202 into 5.0 and 2.2 kbp fragments. It also converted the α structure Holliday junction into a χ structure. The Holliday junctions were only visible after IPTG induction.

These experiments demonstrate that two other *cer* reporter plasmids, pKS455 and pCS202, also make stable Holliday junctions during RM40 recombination assays. The

amount of DNA present in the form of Holliday junction appeared to be comparable to that seen with pSD115. This strongly suggests that this is a general feature of *cer* recombination in this strain and is not a consequence of features inherent to the reporter plasmid pSD115. Furthermore, the three plasmids used were independently constructed by three people using different starting plasmids, making the chances of all containing the type of site mutants described in other systems extremely unlikely.

4.12 Discussion

This chapter describes a system that has been developed to analyse the reaction mechanism of *xer* site-specific recombination *in vivo*. This was done primarily because of our inability to recreate the *xer* recombination reactions *in vitro* (e.g. *cer* and *dif*). Developing the system involved constructing a strain in which the expression of *xerC*, the presumptive *xer* recombinase gene, could be tightly controlled. Experimental analysis using this strain involves initiating the recombination reaction at a set time, and preparing DNA samples of the ongoing reaction thereafter. In this respect it resembles the experimental approach adopted in *in vitro* assays.

The substrates used in this chapter have all been plasmids containing two directly repeated *cer* sites - so called *cer* reporter plasmids. Recombination assays performed using this *in vivo* system confirmed that these plasmids are always resolved to give two circular products. One of these product plasmids contains no origin of replication and these experiments represent the first visualisation of it, illustrating that this is a more sensitive assay than any previously utilised in the analysis of *xer* recombination.

Analysis of the product profile of the plasmid pSD115 by restriction digestion demonstrated that *cer* recombination *in vivo* involves the production of highly stable Holliday junctions, which account for approximately 25% of the DNA isolated from RM40 after 60 mins of recombination. Three experimental approaches were used to prove that these DNA species are indeed Holliday junctions:

(i) Restriction analysis. The mobility of the junction during gel electrophoresis varies depending on its structure. Uncut, the DNA is not readily separated from supercoiled substrate, but a single restriction generates an α structure that has a gel mobility intermediate between linear and supercoiled substrate. This mobility can be altered by changing the supercoiled domain that is cleaved and the position of the cleavage. Restriction of both domains generates a highly retarded χ structure. These experiments suggest that the junction is contained within a certain region of pSD115; *Mlu*I restriction demonstrated that it is contained within the homologous sequences around the *cer* sites (as would be expected for correct recombination) and that the junction is capable of branch migration.

(ii) Electron microscopy. Direct visualisation of isolated χ structures confirmed the predictions made on the basis of restriction digestion, and confirmed that the junction is positioned within, or close to, the *cer* sites of pSD115.

(iii) *in vitro* treatment with RuvC. The supercoiled "figure-8" Holliday junction was shown to be a suitable substrate for cleavage by RuvC, further suggesting that the structure is present within homologous sequence and is capable of branch migration. To my knowledge this experiment is the first demonstration that RuvC is able to recognize and act upon supercoiled Holliday junctions. Previous experiments have used artificial 4-way junctions made using oligonucleotides and *in vitro* generated, RecA coated intermediates (see e.g. Dunderdale *et al*, 1991). This analysis therefore offers further evidence that RuvC is at least one of the *E.coli* Holliday-resolving enzymes. The apparent lack of preference in its direction of cleavage may be of relevance to considerations regarding the mechanism of this enzyme. However, whether or not this reaction profile would be altered by the addition of RuvA or RuvB, or indeed RecA, would need to be examined.

The demonstration of Holliday junctions being produced during *cer* recombination has a number of implications. Their isolation suggests that the *cer* recombination reaction involves a strand exchange mechanism which is consistent with that determined for other members of the λ integrase family of site-specific

recombination reactions (for reviews see Craig, 1988; Stark *et al*, 1992). That this should be the case was predicted by the sequence homology between the recombinase proteins of this family and XerC (Colloms *et al*, 1990). Indirectly this therefore offers support for the belief that XerC is the *xer* recombinase (n.b. identification of *xprB* as an *xer* protein means that this interpretation is not necessarily correct, however: see Chapter 5). Furthermore, it is almost certain that recombination of the chromosomal site *dif*, which also requires XerC (Blakely *et al*, 1991), will involve the same reaction mechanism.

Holliday junction intermediates are a central part of the inferred mechanisms of reactions catalysed by the λ integrase family, exemplified by the λ Int, FLP and Cre systems. Analysis *in vitro* has determined that these reactions proceed by sequential pairs of single strand exchanges at fixed, staggered positions within the cross-over sites. A consequence of this is the generation of Holliday junction intermediates in which only one pair of strands has exchanged; this intermediate is thought to migrate across the distance between the strand exchange points - the "overlap". All previous experiments to isolate these Holliday junctions have been performed *in vitro* and, in most cases, have used altered sites or proteins to detect them. For instance, in λ Int (for which the largest body of evidence is available) they have been isolated in three types of experiment: (i) "suicide" substrates were used which contained a nick in the overlap sequence of one *att* site (Nunes-Duby *et al,* 1987); (ii) a thiophosphate residue was placed at the point of the second strand exchanges (Kitts and Nash, 1987, 1988a); (iii) wild type *att* sites were recombined with mutant sites containing heterologous residues near the point of the second strand exchanges (Nunes-Duby *et al*, 1987; Kitts and Nash, 1987 and 1988b). The yields of the Holliday junctions were very small in all these approaches, except the suicide substrates. This was interpreted as suggesting that the formation of the intermediate is a reversible process (i.e. the first pair of strand exchanges can be reversed), and both the "forward" and "backward" progression of the Holliday junction must be halted for it to accumulate (Nunes-Duby *et al*, 1987). The

inference is that this is most successfully achieved using nicked site substrates, but the reasons for this are not totally understood.

For Cre reactions, Holliday junctions have only been seen when using mutant proteins in *in vitro* reactions (Hoess *et al*, 1987; see also section 4.10). Exactly how these mutants cause the accumulation of otherwise undetectable intermediates is not known, and how these mutants relate to the above considerations is not clear.

In *in vitro* reactions using wild type FLP and wild type FRT sites, structures corresponding to Holliday junctions have been isolated by two groups (Meyer-Leon *et al*, 1988 and 1990; Jayaram *et al*, 1988). The amount of DNA present in this form was small in both cases and appeared to vary depending on the precise reaction conditions used. Why these intermediates should be detectable using wild type FLP, but apparently not with λ Integrase or Cre, is not addressed; but it is conceivable that it is a result of the *in vitro* conditions that have been employed.

The demonstration of *cer*-derived Holliday junctions expands and supports the *in vitro* analysis on the reaction mechanism of the λ integrase family. This is because the junctions have been visualised during a wild type recombination reaction *in vivo* (see below). Indirect evidence for the existence of these reaction intermediates during λ Int *in vivo* recombination has been published (Echols and Green, 1979; Enquist *et al*, 1979). They have also been seen during recombination of a reporter plasmid containing a ColE1 *cer* site and a CloDF13 *parB* site in direct repeat in DS941 (D. Summers, pers. comm.). However, evidence is presented in this chapter that they can be seen in recombination between two wild type *cer* sites and with wild type proteins, and consequently offers *in vivo* confirmation of *in vitro* experiments on related systems. It is noted, however, that one of the difficulties of *in vivo* analysis is attempting to rule out the possibility that mutations in the system being employed are affecting the results.

The ability to detect these structures, and the quantities that they are present in, suggest that there may be something highly unusual about the *cer* recombination reaction, or at least the reaction in this system. This is illustrated by comparison to the *in vivo* analysis of the λ Int catalysed reaction (Bliska and Cozzarelli, 1987).

Restrictions were done in this system that were similar to those employed in this study and would have revealed the existence of Holliday junctions had they been detectable; they were, however, not reported. The analysis made use of λP_L to control the recombination reaction, but this difference cannot explain the results presented here since a very similar approach has been taken (section 4.10), giving the same results. This comparison also rules out the possibility that the stable junctions are simply a consequence of differences between the *in vivo* and *in vitro* conditions; e.g. the ongoing replication of the substrate during the *in vivo* reaction. Although such highly stable junctions may be the natural intermediates in *cer* recombination, alternative explanations should be considered because of the conflicting evidence from other, better characterised, λ integrase-like reactions.

It is possible that the stable Holliday junctions are not intermediates in the *cer* reaction, but could instead be the reaction products. For instance, XerC may catalyse only the first pair of strand exchanges and never the second exchanges. If, after the first pair of exchanges was complete, the proteins involved in the reaction disassembled, this would generate Holliday junctions that would be substrates for resolution *in vivo* by RuvC or related proteins (this is considered further in Chapter 5). It is possible that this mechanism could be a result of the role that *dif* plays in chromosomal segregation; conceivably, stable Holliday junctions could be advantageous to this function (see Chapter 6). Alternatively, these junctions might only be present in *cer* recombination, and the *dif* recombination mechanism might involve normal, transient intermediates. If this were the case they may represent a specific adaptation of the role of *xer* in plasmid stability. This could be a consequence of the altered sequence of *cer* (e.g. this may explain the different overlap), indeed it is possible that the *cer* site contains specific features which block the normal resolution of the Holliday junctions (see Chapter 5). All of these speculations require experimental examination. This could be achieved by, for example, examining the recombination of *dif* reporter plasmids in RM40 (Chapter 5) or *in vitro* experiments to test whether XerC does indeed catalyse more than just one pair of strand exchanges.

Another possibility is that these Holliday junctions are genuine intermediates in *cer* recombination that have somehow "stalled". This could be achieved by the recombination complex "losing" the junction after the first pair of strand exchanges. Such a junction would become free to branch migrate outside the normal points of exchange, and would have to be "recaptured" or cleaved by cellular junction-resolvases to generate products. This scheme would explain why the first pair of strand exchanges are not reversed, without the need to invoke novel reaction mechanisms. There is no available evidence from the experiments I have performed to show that the junctions are intermediates and not reaction side-products, as these "stalled" junctions would be. They can, however, branch-migrate after isolation, suggesting they have not arisen by an aberrant strand exchanges; unfortunately, this is compromised by our lack of knowledge regarding precisely where the *cer* exchanges take place. This second possibility could be analysed by determining whether the Holliday junctions are free to branch migrate *in vivo* (see Chapter 5) This would also address the possibility that the junctions have stalled but remain bound by protein. Exactly what would cause such a situation to arise is not known, but might suggest that some unidentified factor in RM40 is altered in comparison to wild type cells.

All of the experiments described in this Chapter were performed before it was understood that a fourth *xer* gene (*xprB*) is essential for the *cer* recombination reaction. XprB has amino acid sequence homologies to the λ integrase recombinases and consequently it has implications for the *cer* reaction mechanism, most relevantly that it may explain the accumulation of *cer*-derived Holliday junctions in RM40 assays. This is considered in more detail in Chapter 5.

The isolation of these *cer* Holliday junctions has allowed the demonstration that the top strands of the recombining DNA duplexes have been specifically exchanged during their formation. This suggests that there is a bias, or asymmetry, in the strand exchange mechanism. This has previously been shown to be the case for both λ Integrase (Kitts and Nash, 1988b) and Cre (Hoess *et al*, 1987), although it might not always be true in FLP reactions (Jayaram *et al*, 1988). Theoretically, strand exchange

bias could result from a number of sources. Firstly, it is possible that all conceivable first strand exchanges do occur, but only the intermediates produced by the top strand exchanges accumulate, whilst the rest are rapidly resolved. This has not been rigorously excluded for λ Int or Cre, but the alternative explanation - that the bias resides in the formation of the intermediates - is favoured. Asymmetry in the formation of the Holliday junctions could be a consequence of an asymmetric overlap sequence exerting a local effect; for instance, adopting a specific conformation and altering cleavage efficiency. Alternatively, the recombinase binding sites around the spacer region may have a different sequence and exert this effect either directly through the DNA, or through their protein binding affinity. Finally, the bias might reside in the remainder of a complex binding site, in which the arms may act through protein-protein interaction or more directly.

In Cre the only asymmetry is within the *loxP* overlap and it is inferred that this must be the cause of the strand exchange bias (Hoess *et al*, 1987). Conversely, it has been demonstrated that the λ asymmetry resides in the *attP* site arms and not the core (Kitts and Nash, 1988b).

Analysis of the asymmetry in the first pair of *cer* strand exchanges is hampered by the lack of detailed biochemistry that has been performed. For example, the *cer* core site contains two asymmetric regions: the presumptive spacer and presumptive recombinase binding sequences. However, whether or not these truly are such sequences awaits experimental confirmation. The site also contains approximately 250bp of accessory sequence to the left of the core that is essential for recombination. This could be tested for its role in strand exchange bias by comparing the exchanged strands in Holliday junctions produced during *dif* and *t*ypeII hybrid recombination. Both these sites will recombine without the *cer* "arm" sequences (Blakely *et al*,1991; J.Roberts, pers. comm.) and therefore if bias was still present it must reside in the core.

Chapter 5

Further characterisation of *xer* site-specific recombination *in vivo*

5.1 Introduction

Chapter 4 described *in vivo* analysis of *cer* recombination using the strain RM40, and showed that highly stable Holliday junctions were formed. The possible reasons for the accumulation of these putative reaction intermediates were discussed. Accumulation of the junctions may represent a feature of *cer* recombination that has not been described in other λ integrase-like reactions (for reviews see Craig, 1989 and Stark *et al*, 1992), and therefore this chapter describes experiments designed to further the analysis of the junctions, and of the *in vivo* system being employed.

It is hypothesised that *xer* site-specific recombination has a cellular role in monomerising chromosomal dimers which arise as a result of general recombination during replication, therefore facilitating efficient segregation of the daughter chromosomes (Blakely *et al*, 1991). This is achieved by *xer*-catalysed site-specific recombination at the *dif* locus. Previous recombination assays have shown that this reaction has a number of features which distinguish it from the plasmid-borne *cer* system: the two sites have different sequences, *dif* recombination does not require either PepA or ArgR, and the reactions have different properties (Blakely *et al*, 1991; see chapter 1 for review). These considerations illustrate that the *dif* reaction not only deserves mechanistic analysis in its own right, but comparing the *cer* and *dif* reactions might explain the stable Holliday junctions and reveal other features of the *xer* reaction mechanism. The experiments in this chapter therefore also analyse the *dif* recombination reaction, using *dif* reporter plasmids in RM40.

Genetic characterisation of a fourth *xer* gene is presented in this chapter. This gene, called *xprB/xerD*, was not identified before the *cer* analysis in Chapter 4 was performed, nor before most of the experiments detailed in this chapter were executed. Identification of the necessity for XprB/XerD in both the *cer* and *dif* site-specific recombination reactions has implications regarding the production of stable Holliday junctions in RM40 assays, as well as for more general analysis of the *xer* reaction mechanisms, and experiments are presented in this chapter to address this.

5.2 *cer* recombination in *argR* and *pepA* derivatives of RM40

Previous experiments have revealed that *cer* recombination has an absolute requirement for both ArgR and PepA (Stirling *et al*, 1988 and 1989), although the exact manner in which these proteins act during the recombination reaction has not been determined. It is very likely that RM40 contains wild type copies of *argR* and *pepA*, but using this strain to analyse these accessory factors allowed information to be derived regarding their role in the *cer* recombination reaction; it was also possible to determine whether the *cer* reaction mechanism was altered in this strain. For instance, it was conceivable that the generation of stable Holliday junctions in RM40 reflected an altered reaction mechanism (perhaps as a result of the changed expression of XerC) in which the requirement for either, or both, of the proteins had been overcome. Alternatively, it was possible that they act at a stage in the *cer* reaction subsequent to the first pair of strand exchanges. If this were the case then the *argR* and *pepA* genotype of RM40 would not account for the accumulation of the Holliday junctions, but previous assays would not have been able to address this possibility. A reaction scheme of this sort would not necessarily mean that the *cer* reaction was proceeding aberrantly in RM40, and would offer further support to the belief that the *cer* reaction mechanism is quite different to those described for other members of the λ integrase family. These alternatives were tested by analysing the recombination of the *cer* reporter plasmid pSD115 (see fig.4.4) in *argR* and *pepA* derivatives of RM40.

RM41 (RM40*argR*) was constructed by P1 transduction of the *argR* allele from DS956 into RM40. This mutation is a replacement of the *argR* open reading frame by the *fol* gene from the IncW plasmid R388 (encoding trimethoprim resistance; Flynn *et al*, 1989), and consequently RM41 was selected as a kanamycin and trimethoprim resistant derivative of RM40. RM42 was constructed by P1 transduction of the kanamycin resistant determinants from RM40 into the strain HOM38b, which is a DS941 derivative that is both Xer^- and $peptidase^-$ (as judged by its inability to recombine the reporter plasmid pCS202 and its resistance to Val-Leu-NH_2) that was

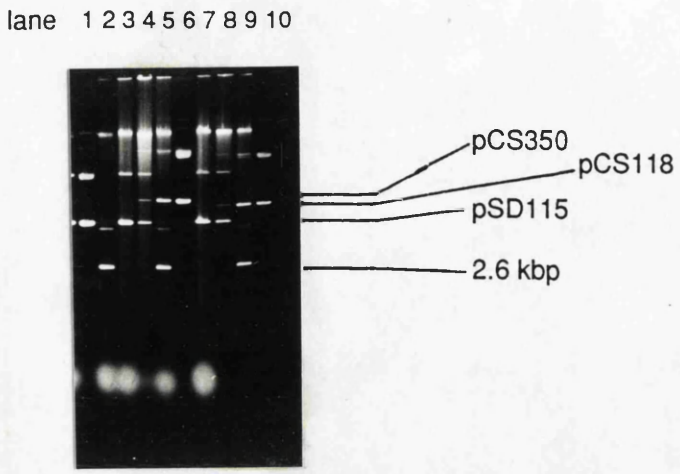
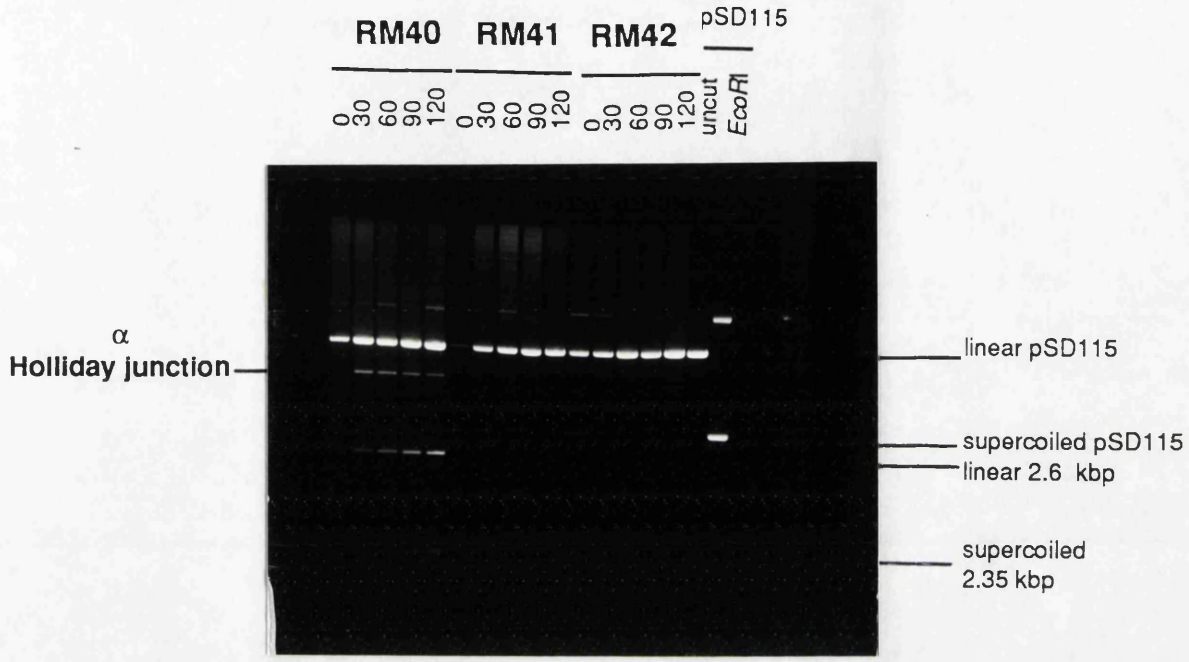


Figure 5.1. *cer* mediated site-specific recombination in *argR* and *pepA* derivatives of RM40. (A) Complementation of Xer⁻ *argR* and *pepA* RM40 strains by pCS350 and pCS118. RM41 (RM40*argR*) was transformed by either pSD115 alone or by pSD115 and pCS350, and RM42 (HOM38b*lac*PO*xerC*) was transformed by pSD115 or by pSD115 and pCS118; the transformants were grown on L-broth containing either 1% glucose or 1 mM IPTG. Plasmid DNA was isolated and run on a 1% agarose gel. The reporter plasmid pSD115 is resolved by *cer* site-specific recombination to give a 2.6 kbp, replicative plasmid and a 2.35 kbp, non-replicative plasmid (see fig.4.4). pCS350 and pCS118 express ArgR and PepA respectively and are compatible with pSD115 in their antibiotic resistances and replication origins.

Lane	Sample
1	pSD115
2	2.6 kbp product
3	pSD115/RM41 (IPTG)
4	pSD115 and pCS350/RM41 (Glucose)
5	pSD115 and pCS350/RM41 (IPTG)
6	pCS350
7	pSD115/RM42 (IPTG)
8	pSD115 and pCS118/RM42 (Glucose)
9	pSD115 and pCS118/RM42 (IPTG)
10	pCS118



(B) Comparison of pSD115 time course recombination in strains RM40, RM41 and RM42. The strains were transformed with pSD115, and pooled transformants were used to set up cultures in which XerC expression was induced with 2 mM IPTG at mid-log phase. DNA samples were prepared at the times shown and run on a 1.2% agarose gel after being digested with *EcoR*I. Linear substrate, linear 2.6 kbp and supercoiled 2.35 kbp DNA is indicated, as is α structure DNA corresponding to pSD115-derived Holliday junctions. A map of pSD115 and its *cer* recombination products is shown in figure 4.4.

isolated during a mutagenesis experiment (H. O'Mara, pers. comm.; see Chapter 3). The nature of the mutation in HOM38b has not been characterised, but because it is present in DS941 the strain (RM42) created here has the same genetic background as RM40 and RM41 and (although not constructed in RM40) it represents RM40*pepA*.

The genotypes of RM41 and RM42 were checked by complementation of the chromosomal mutations using the ArgR and PepA expression vectors pCS350 and pCS118 respectively (C. Stirling, unpublished). The strains were transformed with pSD115 and grown on L-agar plates supplemented with ampicillin and either 1% glucose or 1 mM IPTG. RM41 containing pSD115 and growing on glucose was then transformed with pCS350 and spread onto ampicillin, chloromphenicol plates containing either glucose or IPTG; RM42/pSD115 was similarly transformed with pCS118. DNA was then isolated from these strains by the single colony lysis technique and analysed by agarose gel electrophoresis (fig.5.1A). In both strains pSD115 was not resolved even when the expression of XerC was induced with IPTG, indicating that they contain other *xer* mutations. The reporter plasmid was resolved to give 2.6kbp product in RM41/pCS350 and in RM42/pCS118 when grown on IPTG, but not when grown on glucose. This was taken as confirmation that RM41 is ArgR$^-$ and RM42 is PepA$^-$, and both display the same controllable *xerC* expression as RM40.

To analyse the roles of ArgR and PepA in *cer* recombination and in Holliday junction formation, RM40, RM41 and RM42 were transformed with pSD115 and plated onto L-agar plates supplemented with ampicillin, 1.0% glucose and 50 ug/ml diaminopimelic acid. 20 ml L-broth cultures were set up from each plate and recombination assays were performed by induction of XerC with 2 mM IPTG when the cultures reached mid-log phase. Plasmid DNA samples were prepared by boiling preparations (including phenol/chloroform extraction) at the point of induction and every 30 mins thereafter for two hours. The samples were then digested with *EcoR*I and separated by agarose gel electrophoresis (fig.5.1B). The reaction time course seen in RM40 was the same as described previously: the 2.6 and 2.35 kbp products appeared after addition of IPTG and increased in quantity from then on; α structures

corresponding to Holliday structures also appeared after XerC production. In contrast, there were no reaction products made in the RM41 or RM42 time courses. This is consistent with results from other assays, which suggested that both accessory factors are essential for *cer* recombination. After *EcoR*I restriction digestion no α structures were visible in the RM41 and RM42 time courses. This shows that without functional ArgR or PepA *cer* recombination does not generate stable Holliday junctions in these *in vivo* assays. Two alternative interpretations can be placed on this result.

The first interpretation is that the role of ArgR and PepA is to stabilise Holliday junctions that are made by the first pair of strand exchanges; by functioning in this way ArgR and PepA could facilitate the subsequent resolution of the Holliday junctions by a second pair of strand exchanges. If this interpretation is correct then the results suggest that when the accessory proteins are absent, such as in RM41 and RM42, the first strand exchange is rapidly reversed and the substrate is reformed, meaning that no reaction products are made. However, the λ integrase reaction also has an absolute requirement for an accessory factor (IHF), and analysis of its function in the reaction does not support this interpretation of the results (Landy, 1989). Extensive *in vitro* experiments have suggested that the role of IHF is to facilitate the formation of nucleoprotein synaptic structures in which the strand exchange reactions take place (Landy, 1989), therefore implicating the accessory factor in the synapsis of the *att* sites. No evidence has been published suggesting that Int recombination in the absence of IHF, e.g. when using Int-h mutants, involves lower levels of Holliday junctions; indeed greater amounts of these intermediates are seen in Int-h catalysed reactions (Kitts and Nash, 1988). Another objection to this explanation of the *cer* results is the difficulty in understanding why XerC would exclusively reverse the first *cer* strand exchanges when ArgR and PepA are absent, and not sometimes perform the second strand exchanges. It should be noted, however, that plausible reasons for this reversal could be hypothesised and the roles of the *cer* accessory proteins are not understood, and therefore it remains possible that their functions do not conform to those previously determined for IHF, or other accessory factors.

The second explanation for this result is that ArgR and PepA are required for the formation, and not stabilisation, of the Holliday junctions. This would suggest that the accessory proteins act in *cer* recombination at a stage prior to the first pair of strand exchanges, and that without them this reaction cannot be catalysed. This interpretation may implicate the proteins in bringing together (synapsing) the recombining *cer* sites (which is consistent with the proposed function of IHF in λ Int recombination). Alternatively, it may mean that ArgR/PepA are needed to activate the *xer* machinery and allow the strand exchange reactions to proceed; this could be achieved either by protein-protein or protein-DNA interactions. The fact that neither Holliday junctions, nor products, are seen in either mutant strain suggests that both proteins function during *cer* recombination in these ways. However, whether they perform these functions together or separately (in different roles) is not addressed by this experiment.

This analysis suggests that the high levels of Holliday junctions seen in RM40 are not a consequence of *cer* recombination overcoming its need for accessory factors in this strain. Such a situation could have arisen as a result of RM40 containing a mutant *xerC* gene (perhaps analogous to *int-h*), or because of the altered *xerC* expression in RM40 (for instance higher levels than in DS941) allowed XerC or XerD to escape the binding or processing of ArgR/PepA, and illegitimately synapsed the *cer* sites and performed the first strand exchanges. Because the *xerC* gene present in RM41 should be the same as that in RM40, and the methods employed in its expression are the same in all three strains, it is unlikely that this is the case.

5.3 Site-directed mutational analysis of the putative *cer* overlap sequence

(A) Determining whether the GGG motif in the putative *cer* overlap sequence is the cause of Holliday junction accumulation

The overlap is the region of DNA between the staggered cleavage positions in integrase-like recombination, and it is believed that the Holliday intermediates branch migrate across this sequence during the strand exchange reaction (for reviews see Craig, 1988 and Stark *et al*, 1992; see also chapters 1 and 4). In λ the *att* site overlap has been experimentally defined as 7 bp in size, and has the sequence 5'TTTATAC3' (Craig and Nash, 1983; see fig.1.5). An *att* site mutant with alterations in this sequence has been described which has an *in vitro* phenotype with interesting parallels to the stable Holliday junctions seen during *cer* recombination (S. Nunes-Duby, pers.comm.). The altered overlap has the sequence 5'TTGGTAC3', and therefore has an increased G/C content by virtue of changing TA to GG. Substrates containing two copies of this site show very poor levels of recombination in *in vitro* assays, but do allow the first (top) strand exchanges. It has also been shown that artificial Holliday junctions containing this mutant *att* site have faster resolution kinetics (reversing the first strand exchanges). The proposed explanation for the lack of recombination is that the Holliday junction intermediates are unable to branch migrate across the mutant overlap because the higher G/C content creates an energy barrier to their movement across the sequence. No stable Holliday junctions have been isolated in these assays, but this can be explained by the increased rate of their resolution, which may in turn be a consequence of the placement of the "GG-barrier" near the point of top strand exchange, or alternatively a result of the *in vitro* conditions used.

The putative *cer* overlap has the sequence 5'TTAAGGGA3' (see figs. 1.9 and 5.2). Note that this sequence actually constitutes the proposed *cer* spacer sequence between the recombinase binding sites, and no analysis of the positions of strand

cleavage in the *cer* site have been performed. Despite this, it is assumed for the purposes of the experiments described below that the strand cleavages are made at the edges of the spacer and therefore the spacer and overlap sequences are equivalent in *cer*.The run of three Gs within this sequence may represent a barrier to Holliday junction branch migration that is similar to that described for the mutant λ *att* site. If this were correct then it is possible that *cer* naturally contains such a blockage, and that the accumulation of Holliday junctions has a functional role in the *xer* recombination reaction. The putative overlap sequences of all the plasmid-borne *cer*-like sites that have been described (fig.1.9) have a high G/C content, and all but pNTP16's contain a run of three Gs. This further supports the belief that the GGG overlap sequence may have some biological significance. Note, however, that the putative *dif* overlap contains only one G, and, since it has not yet been determined whether *dif* recombination produces stable Holliday junctions (nor has this been done for the *cer*-like sites above), it is conceivable that these stable junctions are involved in plasmid stability and not in chromosome segregation. This hypothesis cannot readily explain why the first pair of strand exchanges are not reversed in the stalled intermediates, but this may be due to the positioning of the GGG sequence or it could be coupled to other features of the *cer* reaction mechanism.

This hypothesis was tested by site-directed mutagenesis of the *cer* overlap sequence, changing the GGG to AAT (fig.5.2). Altering the overlap sequence in this way removes the presumptive G/C barrier to branch migration and should be reflected in a reduction, or even removal, of the stable Holliday junctions if the hypothesis is correct. This specific pattern of base changes was chosen because it creates an *Ssp*I recognition site and therefore allows rapid screening for the mutant sites. The mutagenesis was done using the plasmid pSD113 as a substrate and a technique, developed by Marshall Stark (see materials and methods for fuller description), which involved selection for the overlap's change with a second oligonucleotide that removes a unique *Sca*I site in pSD113. The method required no sub-cloning of the *cer* site into a mutagenesis vector (fig.5.2A), and it was therefore possible to directly select the

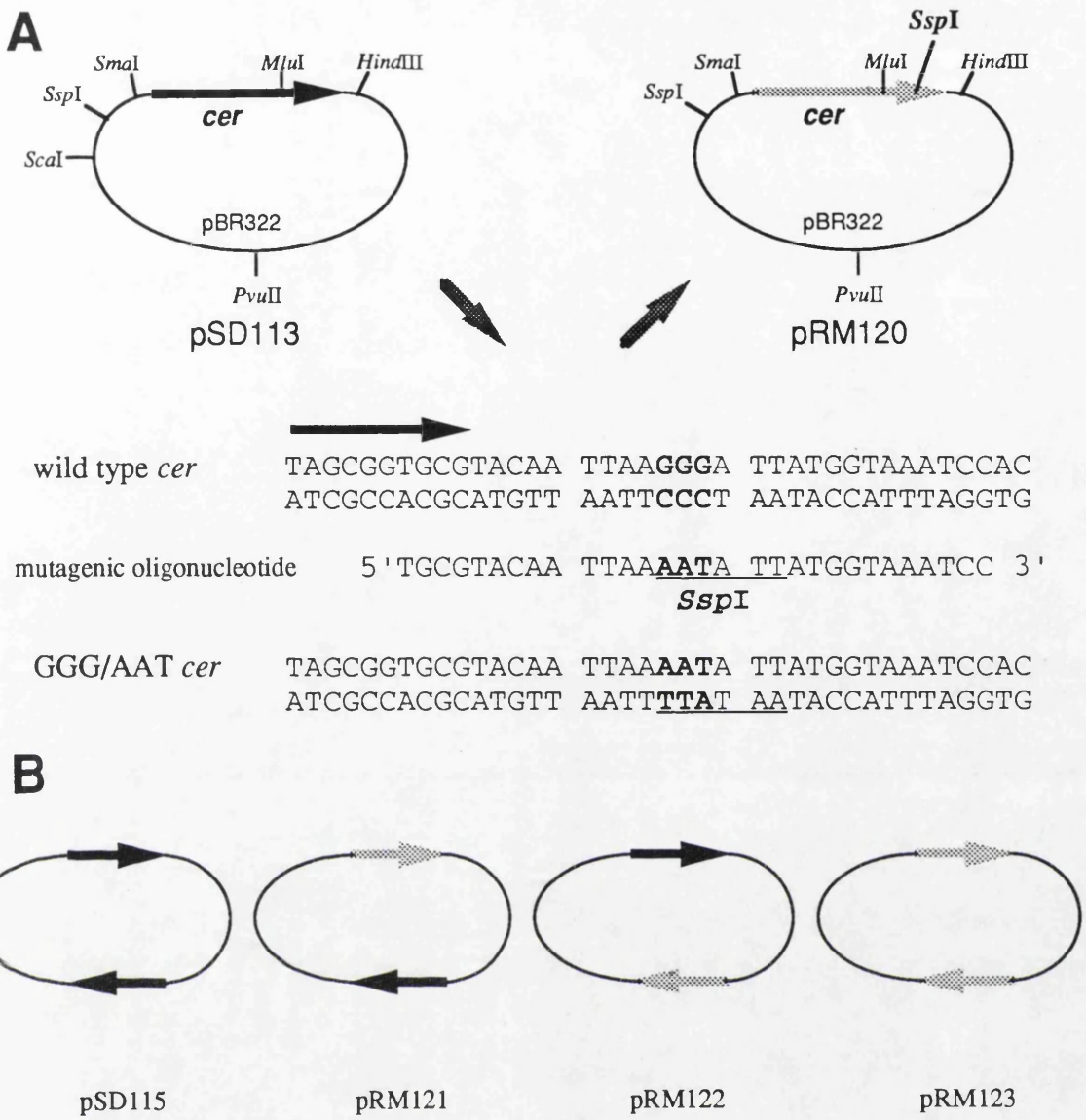


Figure 5.2. Site-directed mutagenesis of the putative *cer* spacer region. (A) Mutagenesis strategy. The mutagenesis involved a second oligonucleotide which deleted the *Sca*I recognition site in pSD113 and allowed the selection of the GGG to AAT change (Materials and Methods). The sequence of *cer* shown represents only the putative spacer region and surrounding sequences that are believed to bind the recombinase proteins. Detailed in bold face are the changes that have been made in *cer* using the oligonucleotide shown; note that they create an *Ssp*I recognition site.**(B) Reporter plasmids constructed.** pSD115 (fig.4.4) was constructed by cloning *cer* into the unique *Pvu*II site of pSD113 (Colloms, 1990).The reporter plasmids pRM121, pRM122 and pRM123 are of exactly the same design as pSD115 and contain directly repeated wild type and GGG-AAT mutant *cer* sites in the positions shown.

mutant pSD113 derivative (called pRM120) by its resistance to *Sca*I restriction, and its susceptibility to digestion with *Ssp*I. Because the GGG to AAT change creates an *Ssp*I site which was shown to cut at the correct position within *cer* (data not shown), it was not felt necessary to confirm the mutation by direct plasmid sequencing.

The way that pSD113 and pRM120 were then used to construct *cer* reporter plasmids is based on the way that pSD115 was constructed (Colloms, 1990) and is detailed in figure 5.2B. Wild type or mutant *cer* sites were inserted (as *Sma*I-*Hind*III fragments) into either pSD113 or pRM120 restricted by *Pvu*II. The three plasmids made in this way were called pRM121 ("top" *cer* site is mutant, "bottom" is wild type), pRM122 ("top" *cer* site is wild type, "bottom" is mutant) and pRM123 ("top" and"bottom" sites are mutant). The orientations of the sites were determined by restriction with *BamH*I, which cuts once within the pBR322 sequence and once at one end of both sites. The identities of the sites were checked by double restrictions using both *EcoR*I and *Ssp*I. The reporter plasmids thus constructed contain directly repeated *cer* sites and are identical to pSD115 in all respects except the sequence of their putative overlaps.

Recombination of pRM121, pRM122 and pRM123 was compared to that of pSD115 by transforming the plasmids into DS941 and growing them on L-agar containing ampicillin. Plasmid DNA samples were prepared from patched cultures of the transformants and were separated, without restriction, by agarose gel electrophoresis (fig.5.3). This showed that all four reporter plasmids recombined to completion in this assay; the only DNA species visible were the replicative, 2.6 kbp reaction products. No differences in the extents of resolution of the four plasmids could be detected, suggesting that in DS941 the changes made do not grossly affect the efficiency of *cer* recombination. This is the case whether it is two wild type *cer* sites, two mutant *cer* sites, or a wild type and a mutant *cer* site that are recombining.

To determine whether the GGG to AAT change has the proposed effect on the level of Holliday junctions during recombination, and to further test the effect that the altered overlaps have on the rates of pSD115 resolution (see section 5.3B, below), the

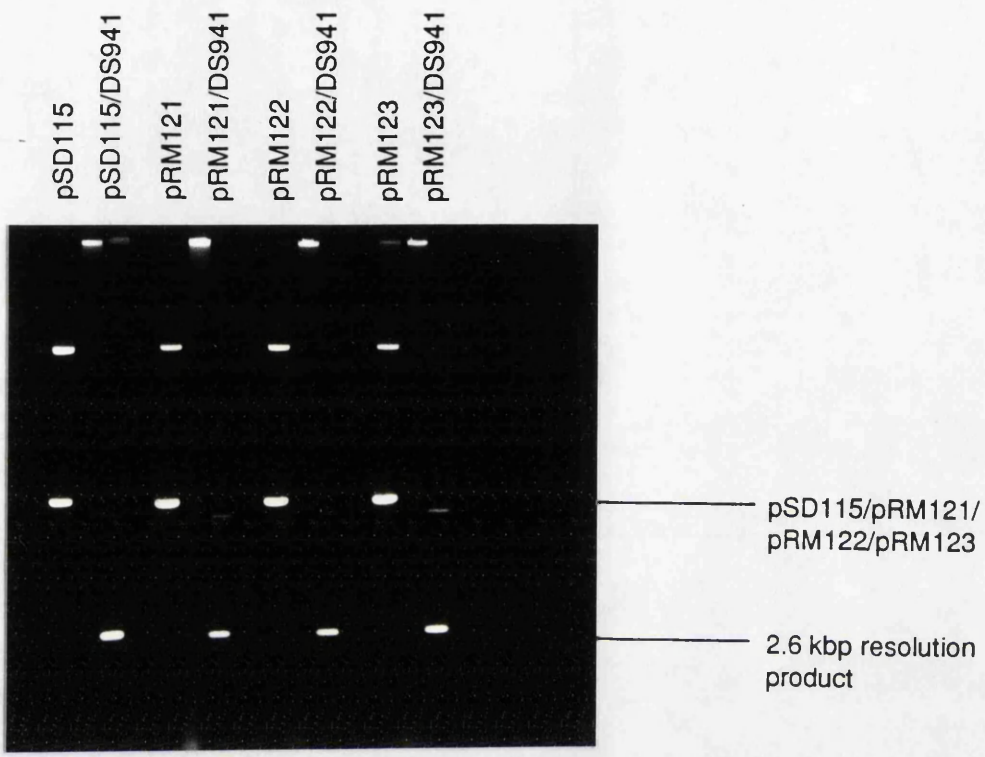


Figure 5.3. Recombination of plasmids pSD115, pRM121, pRM122 and pRM123 in DS941. Plasmid DNA was isolated (by the boiling method) from DS941 transformed separately with each of the four reporter plasmids and run on a 1.2% agarose gel. *cer* mediated site-specific recombination of all these reporter plasmids yields a replicative, 2.6 kbp product and a non-replicative, 2.35 kbp product, as shown in figure 4.4.

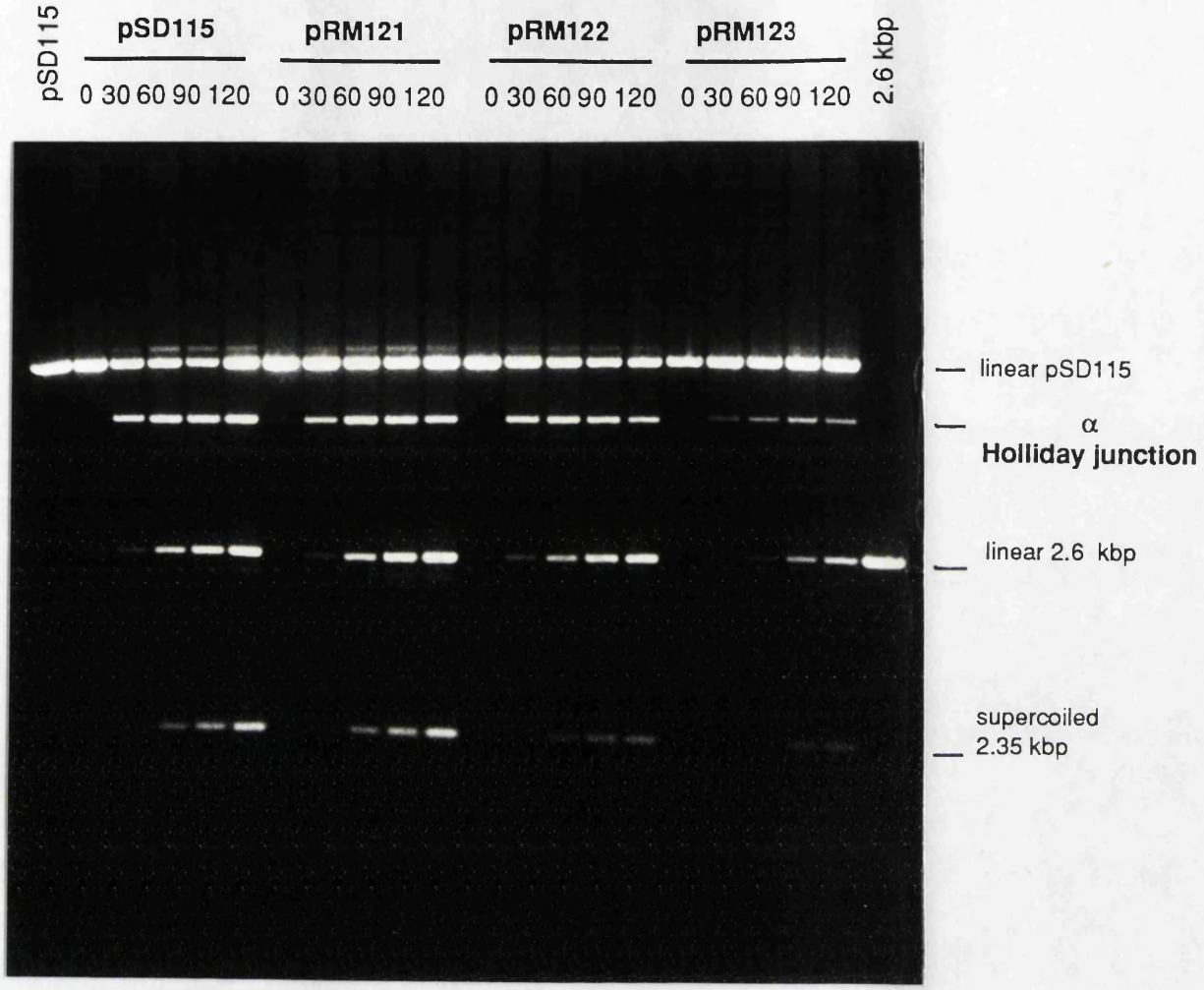


Figure 5.4. Determination of whether altering GGG to AAT reduces the accumulation of *cer* Holliday junctions. Pooled colonies of RM40 transformed with either pSD115, pRM121, pRM122 or pRM123 were used to perform *in vivo* recombination assays. XerC expression was induced at mid-log phase in each culture using 2 mM IPTG. DNA samples were prepared by the boiling method at the times subsequent to XerC induction shown and run on a 1.2% agarose gel. α structures corresponding to *cer* recombination-derived Holliday junctions are indicated.

four plasmids were transformed into RM40 and standard *in vivo* recombination assays were performed (see sections 4.4 and 5.2). Figure 5.4 shows an agarose gel of the DNA samples from this experiment after restriction with *EcoR*I.

*EcoR*I restriction showed that α structures corresponding to stable Holliday junctions were still present in the pRM123 time course, although at a slightly lower level. The fact that this plasmid appeared to be poorer in recombination than pSD115 (as judged by the lower amount of products over the time course in comparison to pSD115) suggests that this reduction in the amount of Holliday junction is not a consequence of the GGG to AAT change allowing the junctions to branch migrate across the previously blocked overlap, but instead is a reflection of the reduction in recombination rate in this substrate. If the GGG in wild type *cer* were the cause of the Holliday junction accumulation then the hypothesis would predict that the change to AAT would result in an increase in the rate of the reaction, because the Holliday intermediates would be able to cross the changed overlap sequence more readily and therefore allow the second pair of strand exchanges to occur more efficiently. These considerations suggest that the GGG sequence feature of the putative *cer* overlap is not the reason for the accumulation of Holliday junctions, and consequently some other explanation must be sought.

(B) Determining whether overlap sequence heterologies affect the efficiency of *cer* recombination

Experiments performed both *in vivo* and *in vitro* in other λ integrase-like systems have shown that DNA molecules (e.g. plasmids or phages) containing a wild type recombination site and a second site with an altered overlap sequence are very poor substrates for recombination. If the sites both contain altered overlaps, however, they can often recombine with normal efficiency. This has been interpreted as suggesting that the actual sequence of the overlap is unimportant, but homology between the sites is essential for efficient recombination. It is proposed that molecules containing sites

that have non-homologous overlaps are poor recombination substrates because sequence heterologies are encountered during branch migration across the overlaps and this reduces the ability of the recombinases to perform the second strand exchanges (see Craig, 1988 and Stark *et al*, 1992 for review). In Cre, a single base pair difference between two *loxP* sites severely reduces recombination, even when the change is at the centre of the overlap and therefore does not directly affect the strand exchange reaction (Hoess *et al*, 1986). Similar results have been reported using λ *att* site *saf* mutants (Landy, 1989) and in FLP FRT sites (Senecoff and Cox, 1986). The experiments performed here on *cer* recombination, however, do not appear to conform to the results from these other λ integrase-like systems.

Recombination of GGG/AAT *cer* sites with wild type *cer* sites (pRM121 and pRM122) appeared to proceed at the same rate in these *in vivo* assays as recombination between two wild type *cer* sites (pSD115), as judged by the amount of reaction products that were visible after induction of XerC expression. The rate of the reaction between two GGG/AAT sites (pRM123) seemed to be somewhat slower, although still fast enough to produce large amounts of products (and be indistinguishable from pSD115 recombination) in DS941 assays (fig.5.3). If the GGG sequence is within the *cer* overlap then these results suggest that *cer* recombination is not affected by having non-homologous sequence in this region. The experiment also suggests that the greatest reduction in recombination efficiency is seen when the mutant sequences are matched. Neither of these interpretations are in agreement with the results from other λ integrase-like systems (see above). The conclusions of these studies would have suggested that in pRM121 and pRM122 the branch migration of the Holliday intermediate would have been blocked by the 3 bp GGG/ATT heterology, reducing the efficiency of the second strand exchanges to the right of the putative overlap, and blocking recombination. They would also have suggested that this reduction in recombination would have been overcome in pRM123, where the overlaps are homologous. It should be noted that the 3 bp GGG/AAT heterology in the *cer* experiments described here is larger than many of

the heterologies analysed in other systems. The implications of these apparent differences are discussed, and tested, below.

It is possible that the overlap heterology created during recombination of plasmids pRM121 and 122 does not block the movement of the Holliday junctions because the intermediates made during *cer* recombination are able to branch migrate across base pair heterologies. Such a situation would be highly unusual in site-specific recombination and therefore must be tested; the creation of an *Ssp*I recognition sequence in the altered sites allowed this to be done. If the *cer* junction is capable of migrating across the GGG/AAT heterology and being resolved to the right, then the reaction products would contain a region of mis-matched bases around this sequence. This mismatch would only be partially recognised by *Ssp*I (at best) and would also be expected to be repaired in the cell, giving rise to a mixture of products containing both wild type and mutant sequence - resulting in reaction products that are partially restricted by *Ssp*I.

To perform this analysis the DNA samples prepared after transforming the four plasmids into DS941, and the 90 min time points from the four RM40 assays, were digested with a large excess (20 units) of *Ssp*I and separated by agarose gel electrophoresis (fig.5.5). It should be noted that all the reporter plasmids contain a pBR322-derived *Ssp*I site 430 bp from the GGG of the "top" *cer* site (within the 2.6 kbp domain), in addition to the sites created by the mutagenesis (see figs 5.2 and 5.6).

The restrictions of both pSD115 and pRM123 represent controls. In pSD115 the only *Ssp*I site is within pBR322, and this is why only the substrate and 2.6 kbp product were cleaved by *Ssp*I digestion (both were linearised: lanes 2, 4 and 6; the 2.35 kbp product was uncut). In pRM123 both *cer* sites contain *Ssp*I sites, which meant that the 2.35 kbp product was now linearised (lane 19), the 2.6 kbp product was cleaved twice (the 430 bp fragment left the gel; lanes 19, 21) and the substrate was cleaved three times (lanes 17, 19). In pSD115 *Ssp*I restriction generated an α form Holliday junction (lane 4), in pRM123 the *Ssp*I sites within *cer* allowed the junction to migrate to products (lane 19).(The analysis is diagrammed in fig.5.6.)

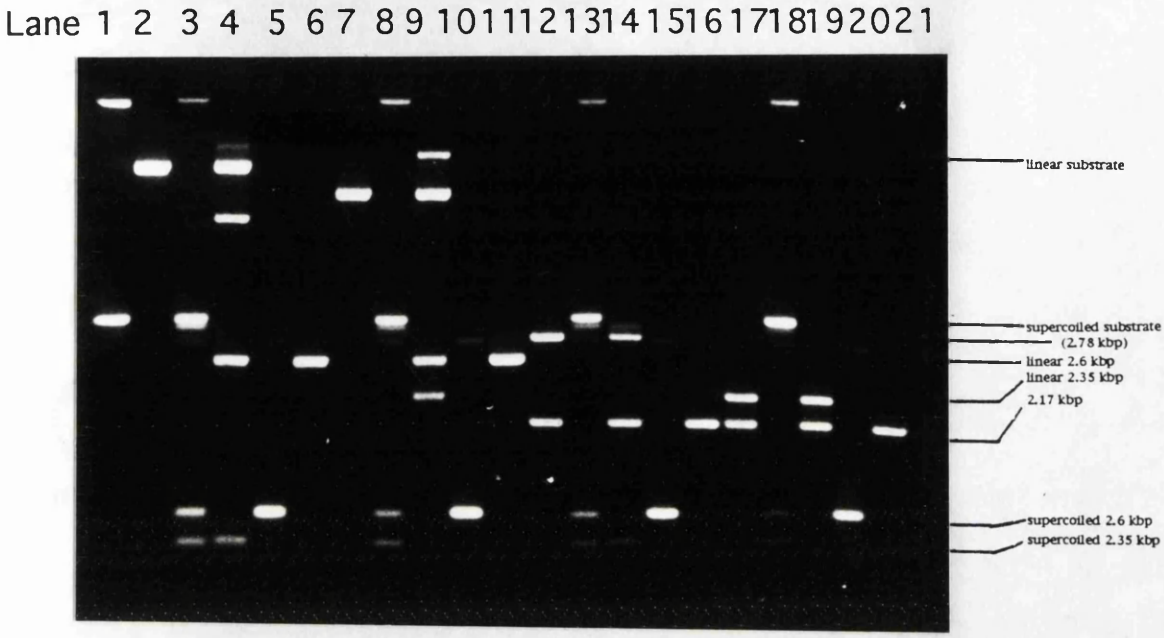


Figure 5.5. *Ssp*I restriction analysis of DS941 and RM40-derived pSD115, pRM121, pRM122 and pRM123 recombination products. The following DNA samples were run on a 1.2% agarose gel:

Lane	DNA sample	restriction digestion
1	pSD115	uncut
2	pSD115	*Ssp*I
3	pSD115 after 90 mins recombination in RM40	uncut
4	pSD115 after 90 mins recombination in RM40	*Ssp*I
5	pSD115 2.6 kbp resolution product from DS941	uncut
6	pSD115 2.6 kbp resolution product from DS941	*Ssp*I
7	pRM121	*Ssp*I
8	pRM121 after 90 mins recombination in RM40	uncut
9	pRM121 after 90 mins recombination in RM40	*Ssp*I
10	pRM121 2.6 kbp resolution product from DS941	uncut
11	pRM121 2.6 kbp resolution product from DS941	*Ssp*I
12	pRM122	*Ssp*I
13	pRM122 after 90 mins recombination in RM40	uncut
14	pRM122 after 90 mins recombination in RM40	*Ssp*I
15	pRM122 2.6 kbp resolution product from DS941	uncut
16	pRM122 2.6 kbp resolution product from DS941	*Ssp*I
17	pRM123	*Ssp*I
18	pRM123 after 90 mins recombination in RM40	uncut
19	pRM123 after 90 mins recombination in RM40	*Ssp*I
20	pRM123 2.6 kbp resolution product from DS941	uncut
21	pRM123 2.6 kbp resolution product from DS941	*Ssp*I

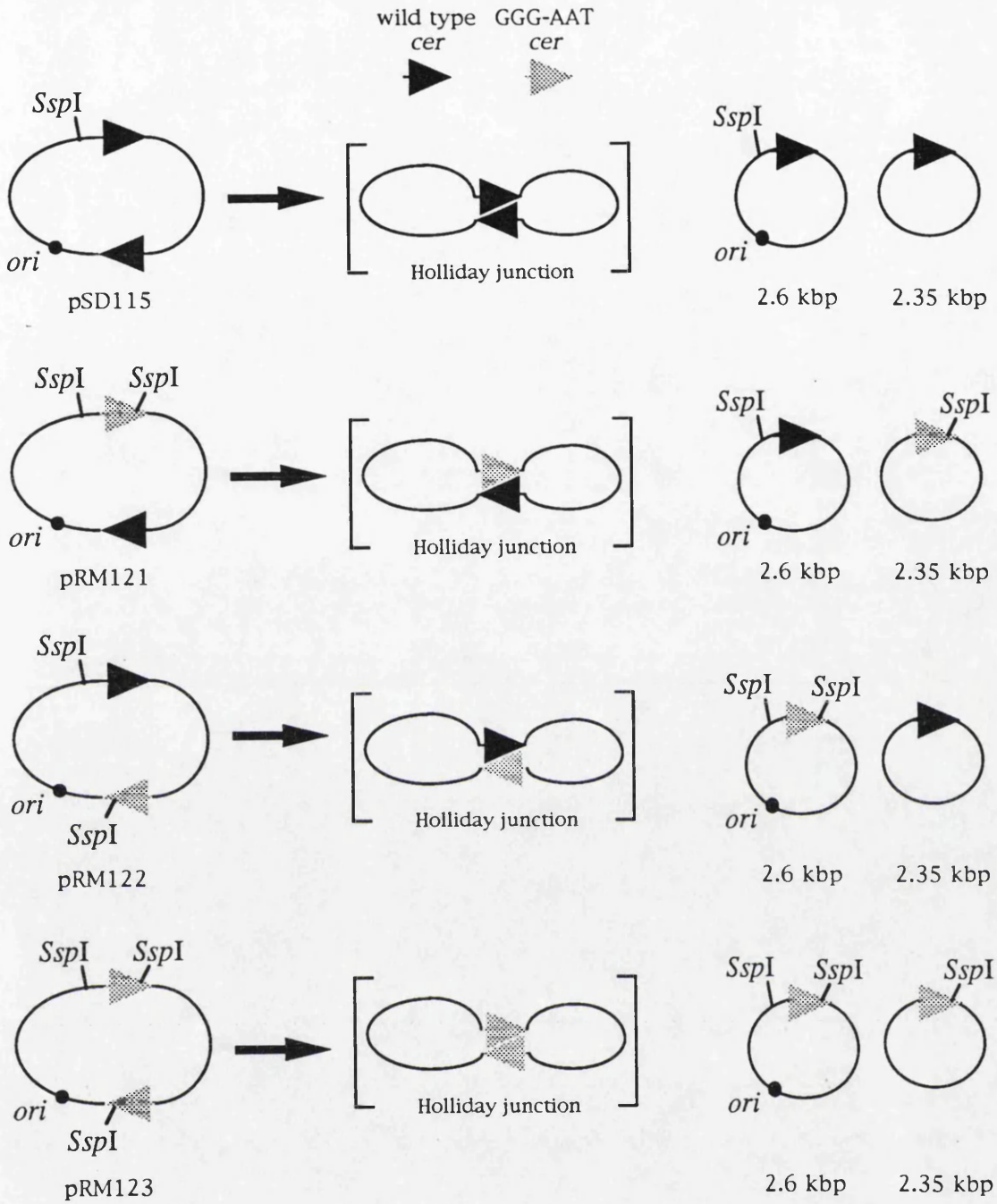


Figure 5.6. Diagrammatic representation of the segregation pattern of *Ssp*I recognition sites during *cer* mediated recombination of the reporter plasmids pSD115, pRM121, pRM122 and pRM123. In pSD115 neither of the *cer* sites contain an *Ssp*I recognition sequence, and therefore after *cer* mediated site-specific recombination only the replicative, 2.6 kbp resolution product contains a vector-derived *Ssp*I site. In pRM123 both of the resolution products contain *cer*-derived *Ssp*I sites after recombination. In pRM121 the "top" *cer* site contains an *Ssp*I site which is segregated into the non-replicative, 2.35 kbp resolution product. In pRM122 the "bottom" *cer* site contains an *Ssp*I site which is segregated into the replicative, 2.6 kbp resolution product.

In pRM121 the "top" *cer* site is mutant and contains the *Ssp*I site. Restriction of pRM121/DS941 (lane 11) showed that the 2.6 kbp product was simply linearised at the pBR322-derived *Ssp*I site; there was no evidence of the smaller product that would be generated by partial cleavage at the *cer-Ssp*I site. Restriction of the 90 min pRM121/RM40 sample confirmed this (lane 9), and also showed that the 2.35 kbp product was completely linearised - there was no sign of supercoiled DNA which would indicate partial cleavage. Together these results suggest that the recombination reaction completely segregates the *Ssp*I site into the smaller reaction product.(In this plasmid the substrate DNA was cleaved twice (it is 430 bp smaller:lanes 7, 9) and the Holliday junction was a nicked α structure, because it can branch migrate up to the single *cer Ssp*I site and release the supercoils in the small domain.)

In pRM122 the "bottom" *cer* site is mutant. In the products of this recombination there was again no evidence for partial *Ssp*I digestion: the 2.6 kbp product was 2.17 kbp in size after restriction (from DS941: lane 16; RM40 90': lane 14) and consequently was cleaved to completion at both the pBR322 and *cer Ssp*I sites, i.e. there was no evidence of linear that would indicate partial digestion; the 2.35 kbp product was not at all cleaved by *Ssp*I (lane 14). This shows that in this plasmid the *Ssp*I site is completely segregated into the larger reaction product.(The substrate DNA here was cleaved twice to give 2.17 and 2.78 kbp fragments (lanes 12, 14) and the Holliday junction had a highly unusual mobility because the majority of the large domain had been removed (it had a very similar mobility to supercoiled substrate: lane 14.)

These observations, that the *Ssp*I site is segregated into the small product when pRM121 recombines and into the large product when pRM122 recombines, can be interpreted in two different ways:

(1) The Holliday junctions may be able to efficiently branch migrate across the heterologous sequence and be resolved by a second pair of strand exchanges to the right of the presumptive overlap sequence. However, this can only be possible if it is proposed that the mismatched reaction products are repaired in favour of the top strands

(assuming the *cer* sites are drawn as in fig.5.2). A directional mismatch repair of this sort would result in the *Ssp*I site segregation described. For this to be true it must also be proposed that some feature of the products' top strands other than the mismatched bases define the directionality, since half the time they are repaired to GGG and half to AAT. Similar experiments in which the segregation of mismatches in phage λ attachment sites were analysed during phage integration discounted this type of directional mismatch repair in explaining their results (Bauer *et al*, 1989). Consequently I favour the explanation described below, but I have no experimental evidence which might support this.

(2) The segregation pattern of the *Ssp*I sites is consistent with the strand exchanges involved in *cer* recombination occuring to the left of the GGG sequence. This reaction may have been forced to occur at this position because the heterology blocked the branch migration of the Holliday junction to the right of the overlap, where it would normally be resolved by a second pair of strand exchanges. This is consistent with analysis reported in λ, where the exchange points of the reaction have been shown to be altered when mutant *att* sites (containing multiple sequence changes in their overlap) are recombined with wild type sites (Bauer *et al*, 1989). The *cer* experiment described above is quite different to this λ experiment, however, because the reaction between the different *cer* sites appeared to proceed with normal efficiency. In contrast, the reaction between the different *att* sites produced very low levels of reaction products when compared to the reaction between wild type *att* sites. The difference between λ and *cer* could be explained by suggesting that in *cer* recombination the second strand exchange is never made; consequently the recombination reaction between non-homologous sites would only be impaired when the heterology impeded the first strand exchanges. Alternatively it is possible that the putative *cer* overlap, as defined, does not represent the true strand exchange positions. This would mean that the second strand exchanges do not naturally occur to the right of GGGA, but instead somewhere to the left. This possibility economically explains the *Ssp*I segregation pattern and why pRM121 and pRM122 recombine as well as pSD115; it also does not require that the

cer contains a novel reaction mechanism, such as only performing one pair of strand exchanges. Since no biochemical experiments have been performed to test the cleavage positions in *cer* this remains a possibility. It cannot, however, offer any explanation for the accumulation of Holliday junctions.

5.4 Are the stable Holliday junctions protein-bound *in vivo*?

An important consideration in attempting to understand why *cer* recombination produces stable Holliday junctions is whether they are bound by protein(s) *in vivo*, or, alternatively, whether they have become dissociated from the *xer* proteins and are free to migrate throughout the *cer* site. This has relevance in attempting to design other experiments; for example, if they are no longer protein-bound this could mean that the *xer* reaction only performs a single strand exchange - a hypothesis that could be tested further. It has previously been demonstrated that when isolated *in vitro*, and therefore stripped of any protein that may have been bound, the junctions are capable of branch migration (section 4.5). This, however, does not address their state in the cell during the *cer* recombination reaction.

A number of approaches could be taken to attempt to answer this question. It may be possible to isolate protein-bound Holliday junctions using cell lysis techniques which maintain protein-DNA interactions. These techniques have been used successfully to analyse the higher-order structure of chromatin and its protein and DNA constituents (Worcel and Burgi, 1972; Schmid, 1988; Pettijohn, 1988). It is likely, however, that any protein-DNA complexes involving *cer* Holliday junctions would be substantially less stable than chromatin, and therefore more difficult to isolate intact. Another technique which could be employed is *in vivo* footprinting (see e.g. Becker and Wang, 1984). Again, however, this may yield ambiguous results because of the difficulty in distinguishing protein that is bound to *cer* from protein bound to the junction itself. The approach that was therefore adopted was to study recombination

between heterologous *cer* sites in order to determine whether the junction is capable of branch migration outside of the *cer* cross-over region.

Previous experiments have defined the cross-over region of the *cer* site (see Chapter 1, figs 1.8 and 1.9; Summers *et al*, 1985; Summers, 1989). If *cer* is split by *Mlu*I restriction then XerC will only bind to the downstream region in gel binding assays (Colloms, 1990; see figs. 1.8, 5.7). It has also been shown that the *dif* and typeII hybrid sites require only 33bp of their sequence that is homologous to the region in *cer* downstream of the *Mlu*I site to act as functional sites in recombination assays and to bind XerC (Blakely *et al*,1991; J.Roberts. pers.comm.). These results offer compelling evidence that the strand exchanges in *cer* recombination occur within this downstream core sequence, and it is therefore likely that the Holliday junctions in RM40 are generated within this region. For these reasons the *Mlu*I site was chosen as the position to make an alteration to the *cer* sequence which would allow the construction of reporter plasmids that could test the position of the Holliday junction *in vivo*.

The mutagenesis method employed in making this alteration was the same as used in analysing the *cer* overlap (section 5.2 and materials and methods) and pSD113 was again used as the substrate. A single C to T change was made which resulted in the abolition of the *cer Mlu*I site (fig. 5.7A). Only a single base was altered in the hope that this small heterology would not impede any branch migration of the Holliday junction. The plasmid pRM90 was selected as containing this mutation by virtue of the fact that it was resistant to both *Sca*I and *Mlu*I digestion. It was not felt necessary to sequence the plasmid because the lack of *Mlu*I restriction indicates a sequence alteration, and the nature of this change was not of primary importance. Three reporter plasmids were constructed using pSD113 and pRM90 in the same manner as described in making pRM121/122/123 (fig.5.7B). The plasmids are identical to pSD115 except in the following ways: the "top" *cer* site of pRM91 has no *Mlu*I recognition site, the "bottom" *cer* site of pRM92 has no *Mlu*I site and pRM93 contains no *Mlu*I sites. The orientations

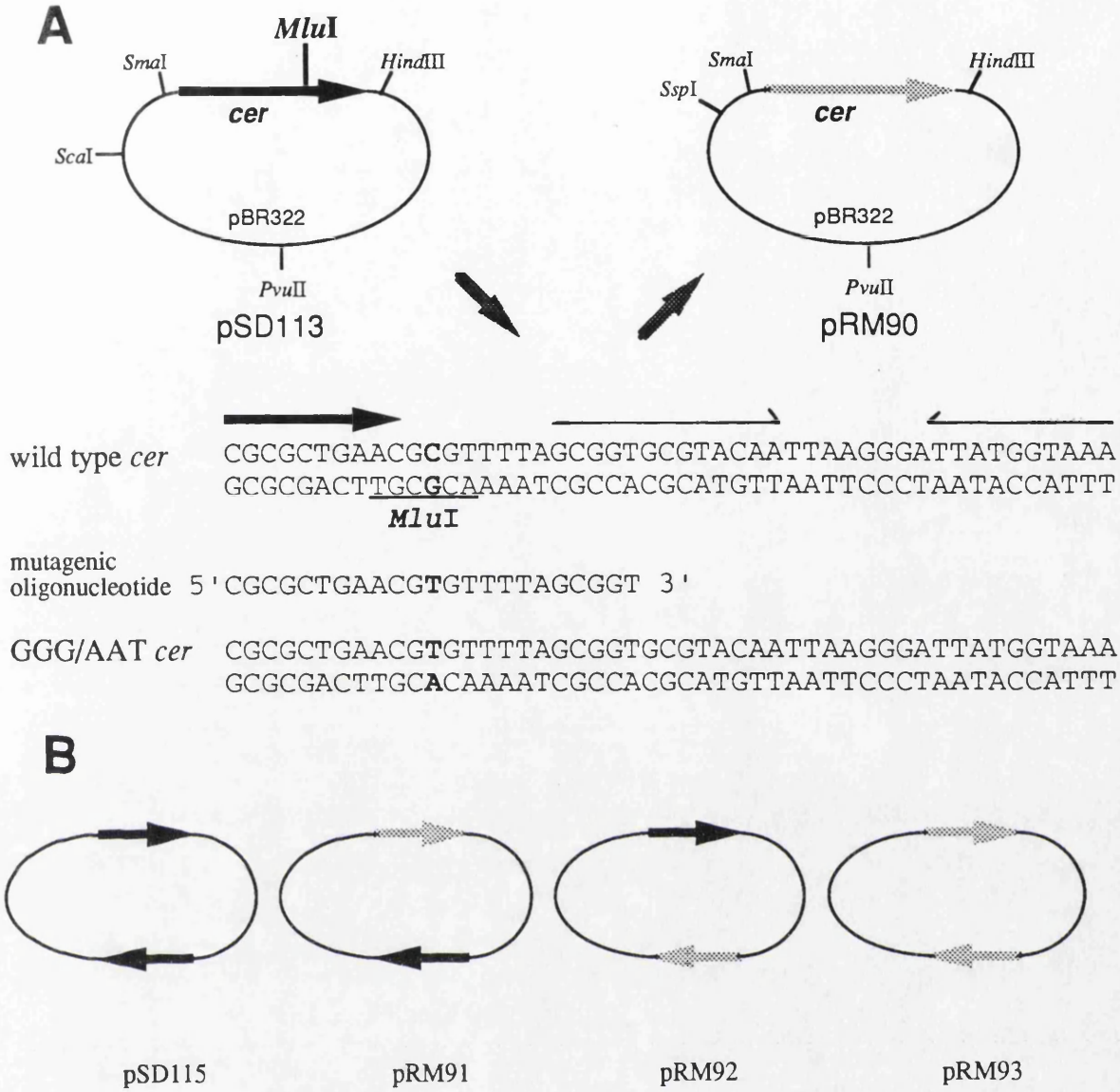


Figure 5.7. Site-directed mutagenic removal of the *Mlu*I recognition site in *cer*. (A) Mutagenesis strategy. The stategy that was employed to create the plasmid pRM90, in which the *cer* sequence's *Mlu*I site has been removed, is the same strategy that was used in altering the *cer* spacer sequence (fig.5.2 and Materials and Methods). The single base pair alteration induced by the mutagenic oligonucleotide is shown in bold face. Only the region between the *Mlu*I site and right-hand end of *cer* are shown (note that the sequences where the recombinase proteins are thought to bind are detailed as inverted arrows). **(B) Reporter plasmid constructions.** pRM91, pRM92 and pRM93 are reporter plasmids derived from pSD113 and pRM90 and are of exactly the same construction as pSD115 and pRM121/122/123. They contain directly repeated, wild type and mutant *cer* sites in the positions shown.

of the sites were determined by *BamH*I restriction, and the identities of the sites by *EcoR*I and *Mlu*I double restriction.

These reporter plasmids allow an indirect test of whether the Holliday junction is able to branch migrate *in vivo* or whether it is "fixed" by protein bound to it. In pRM91 and pRM92, if the junction moves upstream of the *Mlu*I sequence and is resolved outside the cross-over site then products containing heterologies at this position will be generated. These should be identifiable as being partially resistant to *Mlu*I cleavage, since mismatch repair should generate both *Mlu*I$^+$ and *Mlu*I$^-$ derivatives; notice that mismatch repair could occur either before or after resolution. Unfortunately, if the junctions are free to branch migrate *in vivo* but are always "recaptured" and resolved within the cross-over region (by the *xer* machinery) before repair occurs, then this experimental approach would be unable to detect this.

To perform this analysis the four reporter plasmids were transformed into DS941 and grown on L-agar plates containing ampicillin. The transformants were then patched out and their plasmid DNA analysed by agarose gel electrophoresis (fig.5.8). Figure 5.8A shows that the plasmids recombined to completion in this assay, and all resolved to give the same 2.6 kbp-sized product. This suggests that the site-directed mutation in the *Mlu*I site has little, or no, effect on the recombination efficiency. The DNA samples were restricted using a large excess of *Mlu*I (20 units) and examined by agarose electrophoresis (fig.5.8B). pSD115 and pRM93 act as controls: the reaction product derived from pSD115 was completely linearised by *Mlu*I restriction and pRM93's was not digested at all. For neither pRM91 nor pRM92 was there any evidence of partial *Mlu*I restriction, which would indicate a heterogeneous population of reaction products: pRM91/DS941 was unrestricted, pRM92/DS941 was linearised. This result suggests that in DS941 the strand exchanges have occured downstream of the *Mlu*I site during the recombination reaction. This could indicate that Holliday junctions which form during *cer* recombination do not branch migrate beyond the *Mlu*I site (and therefore may be protein-bound throughout the reaction), or that stable Holliday junctions are not made during *cer* reactions in DS941 (no experiment has demonstrated their existence in

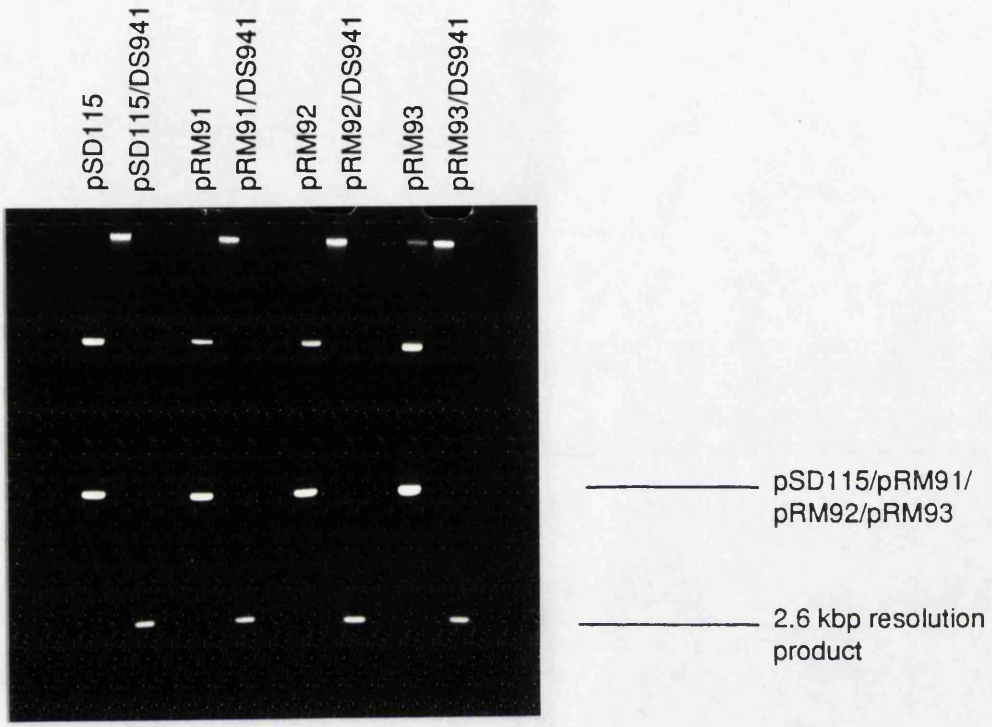
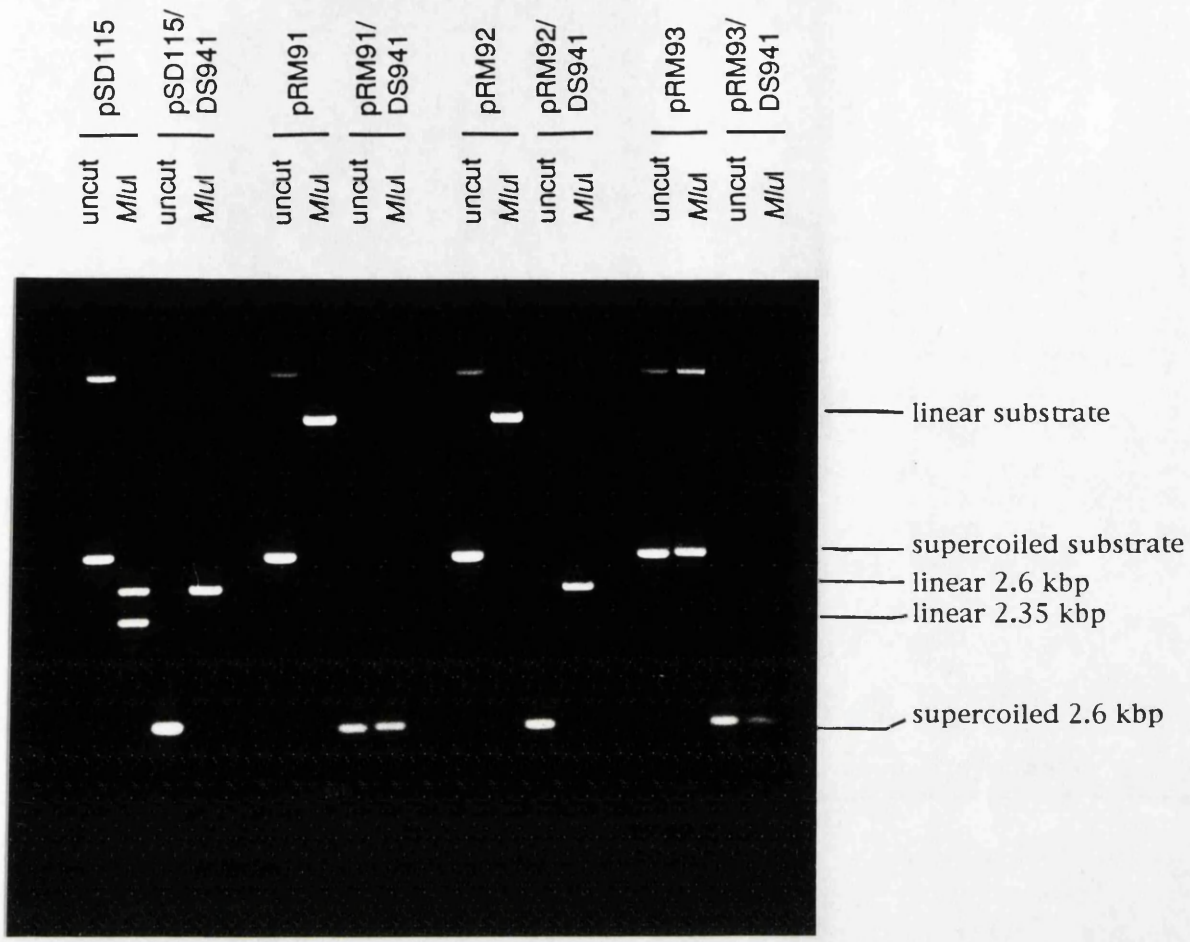


Figure 5.8. *cer* mediated site-specific recombination of reporter plasmids containing sequence heterologies at the *Mlu*I site. (A) Recombination efficiency in DS941. Plasmid DNA was isolated (by the boiling method) from DS941 transformed separately with the reporter plasmids pSD115, pRM91, pRM92 and pRM93 and run on a 1.2% agarose gel. Resolution of all these reporter plasmids yields a replicative, 2.6 kbp product and a non-replicative, 2.35 kbp product, as shown in figure 4.4.



(B) *Mlu*I restriction analysis of DS941-derived pSD115, pRM91, pRM92 and pRM93 recombination products. The reporter plasmids and their DS941-derived, 2.6 kbp resolution products (shown in figure 5.8A) were digested with *Mlu*I and run on a 1.2% agarose gel. In pSD115 both *cer* sites are wild type (*Mlu*I$^+$), in pRM91 both are mutant, in pRM91 the "top" *cer* site has no *Mlu*I site and in pRM92 the "bottom" *cer* site has no *Mlu*I site.

such assays). Possible objections to these interpretations are that junctions do form, and do migrate upstream of *Mlu*I, but repair always favours the top strand of the product, or that the heterology at the *Mlu*I site prevented the unbound junctions from migrating beyond it.

The analysis was followed up by looking at the reaction products made during recombination in RM40. The four reporter plasmids were transformed into RM40 and plated on L-agar plates containing ampicillin, 1.0% glucose and 50 ug/ml diaminopimelic acid. Liquid cultures were set up and XerC expression was induced at mid-log phase using 2 mM IPTG. 120 mins after induction, boiling preparations were performed and the DNA samples were run on a 1.2% Sea Plaque low melting point agarose gel. For each plasmid the 2.6 and 2.35 kbp reaction products were isolated from the gel by phenol/chloroform extraction (following the agarose manufaturer's instructions) and their *Mlu*I status was determined by restriction (using 20 units of *Mlu*I) and standard agarose gel electrophoresis (fig.5.9).

Both of the reaction products of pSD115 recombination were completely linearised (lanes 4, 6), whilst both from pRM93 were undigested (lanes 19, 21), by *Mlu*I. These restriction patterns are as expected for these plasmids, agree with the analysis from DS941 and show that the DNA prepared from the Sea Plaque gel is sufficiently clean to permit the analysis of pRM91 and 92. The replicative (2.6 kbp) reaction product of pRM91 made during recombination in RM40 was partially linearised by *Mlu*I in this gel (lane 9). This was unexpected because this product was uncut when taken from DS941 (see above), and therefore suggests that there may be a difference in the *cer* recombination reactions in the two strains. This apparent discrepancy between the two strains was confirmed by the restriction of pRM92's 2.6 kbp product isolated from RM40: the DNA was completely linearised when prepared from DS941, but in this assay a portion of the DNA remained supercoiled (lane 14), suggesting that *Mlu*I was only able to partially restrict it. These partial restrictions are mirrored by the non-replicative reaction products (which could not be isolated from the DS941 assays). For pRM91 the *Mlu*I site would be expected to segregate into the 2.35

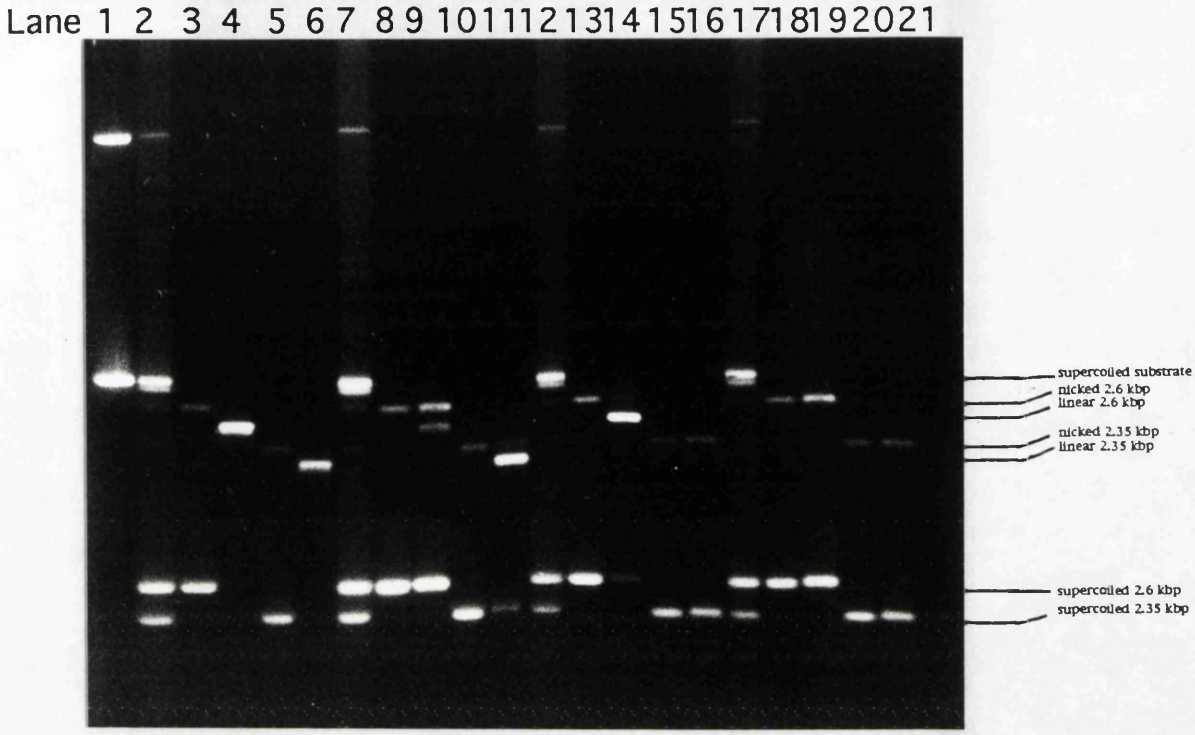


Figure 5.9. *Mlu*I restriction analysis of the pSD115, pRM91, pRM92 and pRM93 resolution products produced in RM40. Isolation of the replicative, 2.6 kbp and non-replicative, 2.35 kbp products from RM40 recombination of the four reporter plasmids is described in the text. The following DNA samples were run on a 1.2% agarose gel:

Lane	DNA sample	restriction digestion
1	pSD115	uncut
2	pSD115 after 90 mins recombination in RM40	uncut
3	pSD115 2.6 kbp resolution product from RM40	uncut
4	pSD115 2.6 kbp resolution product from RM40	*Mlu*I
5	pSD115 2.35 kbp resolution product from RM40	uncut
6	pSD115 2.35 kbp resolution product from RM40	*Mlu*I
7	pRM91 after 90 mins recombination in RM40	uncut
8	pRM91 2.6 kbp resolution product from RM40	uncut
9	pRM91 2.6 kbp resolution product from RM40	*Mlu*I
10	pRM91 2.35 kbp resolution product from RM40	uncut
11	pRM91 2.35 kbp resolution product from RM40	*Mlu*I
12	pRM92 after 90 mins recombination in RM40	uncut
13	pRM92 2.6 kbp resolution product from RM40	uncut
14	pRM92 2.6 kbp resolution product from RM40	*Mlu*I
15	pRM92 2.35 kbp resolution product from RM40	uncut
16	pRM92 2.35 kbp resolution product from RM40	*Mlu*I
17	pRM93 after 90 mins recombination in RM40	uncut
18	pRM93 2.6 kbp resolution product from RM40	uncut
19	pRM93 2.6 kbp resolution product from RM40	*Mlu*I
20	pRM93 2.35 kbp resolution product from RM40	uncut
21	pRM93 2.35 kbp resolution product from RM40	*Mlu*I

kbp product if recombination occurs downstream of this site; however, some of this DNA was uncut by *Mlu*I when isolated from RM40 (lane 11). Conversely, for pRM92, the 2.35 kbp product was slightly linearised by *Mlu*I digestion when it was expected to be *Mlu*I$^-$ (lane 16; this partial digestion was not as clear as previous cases, but can be accounted for by the lower amount of DNA sample in this restriction and the fact that in all cases the "parental" *Mlu*I status is the larger proportion of the DNA).

This result suggests that the Holliday junctions seen in RM40 are not protein-bound and are free to branch migrate beyond the *Mlu*I site of these plasmids; as noted this appears to be in conflict with the results from the analysis of the reaction in wild type cells (DS941). It should be noted that an assumption of this interpretation is that in RM40 the *cer* recombination reaction is correctly initiated within the core sequence downstream of the *Mlu*I site, and the result is not a consequence of the junctions being generated by a first pair of strand exchanges upstream of *Mlu*I and then being resolved downstream. This experiment cannot determine whether the heterogeneous reaction products have arisen through mismatch repair before or after the resolution of the Holliday junctions. For example, it is possible that in RM40 the junctions are all correctly resolved within the core sequence, but in some cases the *Mlu*I heterology is repaired before the resolution step. Alternatively, the Holliday junctions generated in RM40 may be resolved by cellular enzymes, generating reaction products which contain *Mlu*I heterologies that are then repaired to yield partially restrictable DNA.

The difference between the products of the DS941 and RM40 reactions is, I believe, significant, and can be interpreted in two ways. The first possibility is that the Holliday junctions seen in RM40 time courses are present at the same level during recombination in wild type cells, but only in RM40 are they freed of protein and can migrate outside their natural confines. This must therefore suggest that (for unidentified reasons) they remain bound by the *xer* machinery in DS941 but not in RM40. The second, more plausible, interpretation is that the *cer* recombination reaction is altered during RM40 time course experiments in a way which allows Holliday intermediates to accumulate to unnatural levels. This could occur because the reaction is somehow "de-

stabilised" after the first strand exchange and the *xer* proteins disassemble from the intermediate, releasing the Holliday junctions. This experiment does not explain why this should be happening, nor does it address whether these reaction by-products are eventually "recaptured" and resolved by *xer* site-specific recombination or whether they are processed by cellular enzymes to generate reaction products (it is even possible that replicating the Holliday junctions could generate products). Note that if they are not resolved by *xer* recombination their stability must be ascribed to the fact that they are inefficiently resolved by other means.

5.5 The effect of RuvC on *cer* recombination

The following section describes an experiment designed to be complementary to the above *Mlu*I analysis and to further explore the way the *cer* reaction proceeds in RM40 and wild type cells. It is not known from the previous experiments whether the products seen in RM40 time courses are derived from *xer*-catalysed resolution, or whether they are a result of cellular enzymes processing the stable Holliday junctions. For instance, it is possible that in RM40 the second pair of strand exchanges are never executed by the *xer* machinery. Analysis of pRM91 and pRM92 recombinant products suggests that this may not be the case in DS941 recombination assays, but it could explain why the junctions accumulate in RM40. It even remains possible that in DS941 the recombination reaction does generate Holliday junctions (because only one pair of strands are exchanged) and the above experiments require an alternative explanation.

The only *E.coli* Holliday junction-resolving enzyme that has been described in detail is RuvC, although there is some evidence that others may exist (see e.g. Taylor, 1992). Chapter 4 demonstrated that *cer*-derived Holliday junctions are suitable substrates for this enzyme *in vitro*. Because of these considerations it was decided to analyse how *cer* reporter plasmids recombine in *ruvC* strains. This tests whether the second pair of strand exchanges of the site-specific recombination reaction are catalysed by *xer* or by cellular enzymes. The strain used, CS85 (*ruvC53*::Tn10), was kindly

supplied by R.Lloyd and was transformed with pKS455 (fig.4.20) by Mary Burke. Replica patching showed that all the transformants were chloramphenicol sensitive, indicating that the strain is Xer^+. However, because *ruv* mutants display only a modest reduction in homologous recombination (one of the pieces of evidence suggesting that other *E.coli* proteins can resolve Holliday junctions) it was decided to attempt to determine whether the rate of the *cer* recombination reaction is reduced in *ruvC* cells.

A strain construction and a plasmid contruction were required in order to perform this analysis. The strain that was made was an *ruvC* derivative of RM40 called RM43; this was made by P1 transduction of the tetracycline resistant *ruvC53* allele from CS85 into RM40. The plasmid that was made (called pRM80) is an RuvC expression vector which is compatible, in its antibiotic resistance and replication origin, to the reporter plasmid pSD115. The cloning strategy was as follows: a 1.12 kbp *Hind*III-*Eco*RI fragment containing *ruvC* was cloned from pGS762 (kindly supplied by G.Sharples) into the expression vector pCB105, also restricted with *Hind*III and *Eco*RI. This insertion is directional and the *ruvC* gene is orientated so that its expression is driven by the *lac* promoter of pCB105. pRM80 is 5.4 kbp in size, has a λdv origin of replication and is chloramphenicol resistant. The *ruv* status of RM43 was determined by showing that it is UV sensitive, and pRM80 was confirmed as expressing RuvC by its ability to make RM43 UV resistant (data not shown; see also Sharples and Lloyd, 1991).

Both RM40 and RM43 were transformed with pSD115 and plated onto L-agar containing ampicillin, 1.0% glucose and 50 ug/ml diaminopimelic acid. RM40 was also doubly transformed with pSD115 and pRM80 and selected on L-agar plates with chloramphenicol in addition to the above supplements. L-broth cultures containing the appropriate antibiotics were set up using these transformants and were allowed to grow to mid-log phase before XerC was induced with 2 mM IPTG (note that this should also induce the expression of RuvC to higher levels in the cells containing pRM80). DNA samples were made every 45 mins and the cultures were stopped from reaching stationary phase by adding L-Broth supplemented with 2 mM IPTG and antibiotics after each DNA sampling. It was hoped that this would allow the recombination

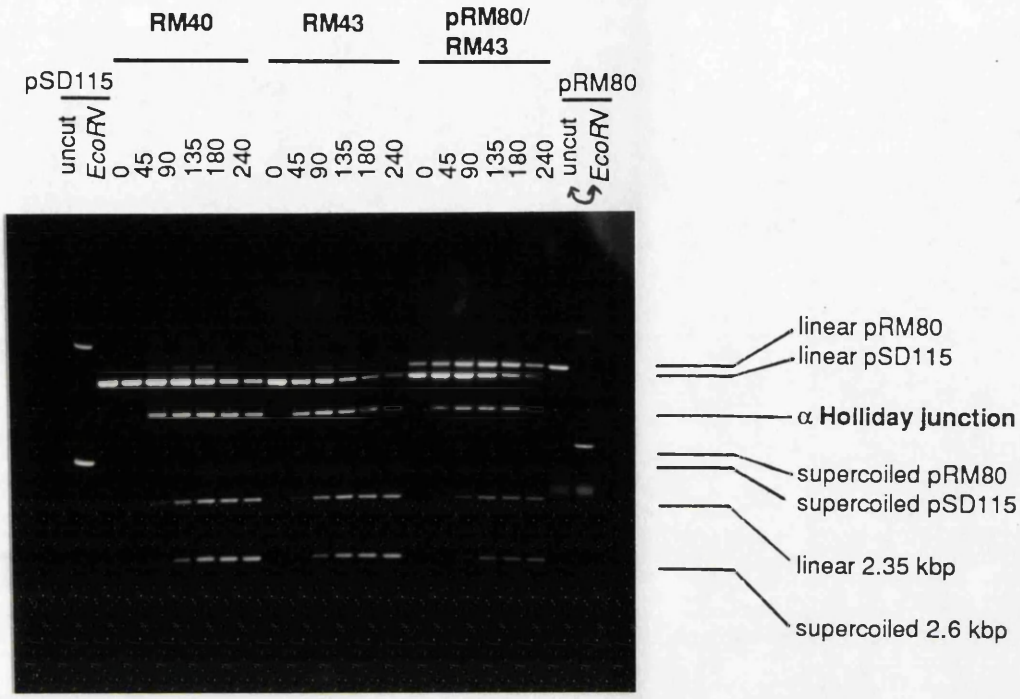


Figure 5.10. *in vivo* recombination of pSD115 in RuvC$^+$ and RuvC$^-$ derivatives of RM40 over 240 mins. *in vivo* recombination assays were performed using pSD115 transformed into the following strains: RM40, RM43 and RM40 transformed with pRM80. Plasmid DNA samples were prepared by the boiling method at the times shown after XerC induction and were digested with *EcoR*V before being run on a 1.2% agarose gel. *EcoR*V digestion linearises the RuvC expression vector pRM80 (as shown) and also linearises pSD115 and its 2.35 kbp resolution product (see fig.4.4). α structures corresponding to Holliday junctions are indicated.

reaction to proceed for a long period and perhaps allow any differences in the rates of the reaction in the three cultures to be seen. The DNA prepared in this way is shown in figure 5.10 after restriction with *EcoR*V (which cuts once in the smaller domain of pSD115 and linearises pRM80) and agarose gel electrophoresis.

In all three cultures the reaction did not go to completion, i.e. even at the latest time point there was still pSD115 substrate remaining. This may indicate that the ongoing plasmid replication was almost balancing the rate of substrate breakdown by *xer* recombination, and that the time given in this assay was insufficient to recombine all the substrate (in some of the original experiments in section 4.2 all the substrate had been converted to product). This, and the fact that at the earliest time points both Holliday junctions and products were visible, make comparisons of the apparent rates inconclusive. However, it is clear that neither the absence of active RuvC in RM43, nor the increased amount of RuvC from pRM80, caused a large alteration in the amount of Holliday junction resolution. This can be said because the amount of material present in α structures was comparable in all strains, and because the amount of reaction products did not appear to have been altered.

This experiment was repeated, except that the DNA samples were made at earlier time points after XerC induction (fig.5.11). This was done to attempt to determine if alterations in the rate of appearance of the Holliday junctions and/or reaction products could be discerned in the three strains. It appeared that in all cases the junctions were created very soon after XerC induction and before any reaction products arose. However, the times that the junctions were generated (approximately 2.5 mins after IPTG addition), and the times the first linear non-replicative products were seen (10 mins), appeared to be the same in all three cultures.

These experiments suggest that RuvC is not responsible for processing the stable *cer* Holliday junctions produced during RM40 recombination assays, since the junctions are seen in the same quantities, and reaction products are made as efficiently, in an *ruvC* derivative of RM40 (RM43) and in the presence of an RuvC expression vector (pRM80). Consequently, this implies that the reaction products in these assays are

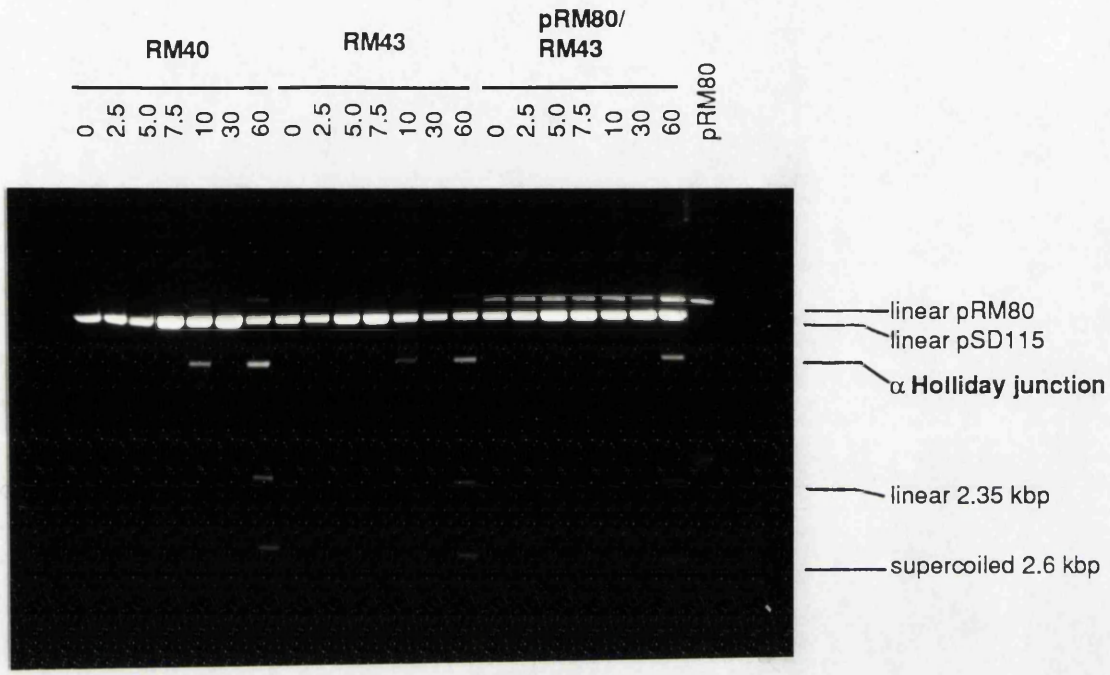


Figure 5.11. *in vivo* recombination of pSD115 in RuvC$^+$ and RuvC$^-$ derivatives of RM40 over 60 mins. *in vivo* recombination assays were performed using pSD115 transformed into the same strains as shown in figure 5.10: RM40, RM43 and RM40 transformed with pRM80. Plasmid DNA samples were prepared by the boiling method at earlier times than in figure 5.10 (as shown) and were again digested with *EcoR*V before being run on a 1.2% agarose gel. *EcoR*V digestion linearises the RuvC expression vector pRM80 (as indicated) and also linearises pSD115 and its 2.35 kbp resolution product (see fig.4.4). α structures corresponding to Holliday junctions are indicated.

derived from *xer*-catalysed recombination of the substrate, supporting the belief that both strand exchanges are executed by the *xer* machinery. Because the *Mlu*I experiments suggested that the Holliday junctions are free to migrate *in vivo*, and because this experiment suggests that these junctions are not processed by RuvC, it may be concluded that the junctions are "re-caught" by the *xer* machinery and resolved by a second strand exchange in the *cer* core. Furthermore, if this is correct, the *Mlu*I heterologies seen in the pRM91 and pRM92 reaction products must be accounted for by mismatch repair of the unbound junctions before they are resolved This analysis supports the contention that the junctions are an aberrant by-product of the *cer* reaction in RM40.

The above interpretation of these experiments can be contended in a number of ways. It is possible that the stable Holliday junctions are processed not only by RuvC, but also by other, unidentified resolvases (e.g. RecG; see Lloyd, 1990 and Taylor,1992). If this were the case then a mutation in *ruvC* alone may not alter the rate that the junctions are processed sufficiently for these conclusions to be drawn. It is also possible that overexpression of RuvC alone will not increase the rate of the junctions' resolution, perhaps because it requires other co-factors, for example RuvA, RuvB or even ORF26 (see section 4.6). The stability of the junctions, which the *Mlu*I analysis says are not protein-bound, suggests that *E.coli* does not contain much Holliday resolving activity; this has not been experimentally determined, and therefore may indicate a flaw in this experimental approach. A final objection is the fact that it is not possible in these experiments to test directly if the second pair of strand exchanges are catalysed by *xer* recombination to resolve the junctions; this could only be determined *in vitro* and therefore remains a supposition at this time.

5.6 Comparing *cer*, type II hybrid and *dif* recombination in RM40

This section describes the analysis of *dif* and type II hybrid recombination reactions in RM40, and compares them to the *cer* reaction. The reasons for performing these experiments have already been discussed (Chapter 4 and 5.1). The plasmid that was used to analyse *dif* recombination is called pSD126 (Blakely *et al*, 1991) and it contains two directly repeated 532bp *dif* sites. The plasmid pSD110 was also analysed (Colloms, 1990); this is a type II hybrid reporter plasmid containing directly repeated sites and is of very similar design to pSD115. This plasmid is of interest because the type II hybrid sites (henceforth simply called typeII sites) are only the "core" sequence downstream of the *Mlu*I site and their recombination characteristics have been shown to be very similar to *dif*, despite being derived from *cer* (Summers, 1989). Like *dif* reporter plasmids, pSD110 is not simply resolved by *xer* recombination, but sites on separate molecules will also recombine to cause plasmid mutlimerisation. Because these two recombination sites will recombine both intra- and inter-molecularly they are said to lack reaction "selectivity".

The three reporter plasmids were transformed into RM40 and *in vivo* recombination assays were performed in the same manner as described previously (sections 4.4 and 5.2). The plasmid DNA samples were restricted with enzymes that cleave the substrates and replicative resolution products once and separated by agarose gel electrophoresis (fig.5.12A).

The reaction profile of pSD110 in this time course showed both similarities and differences to the profile of pSD115 (which was the same as in previous experiments). Both resolution products were made in this assay, and were of the size expected; furthermore a band which almost certainly represents the pSD110-derived α Holliday junction was visible after induction. Where the reaction differs is that two DNA species were produced during recombination which had slower gel mobilities than the linearised substrate. The faster migrating species is the restriction product from one of the two possible *xer*-catalysed dimerisation reactions (diagrammed in fig.5.13). *Pst*I

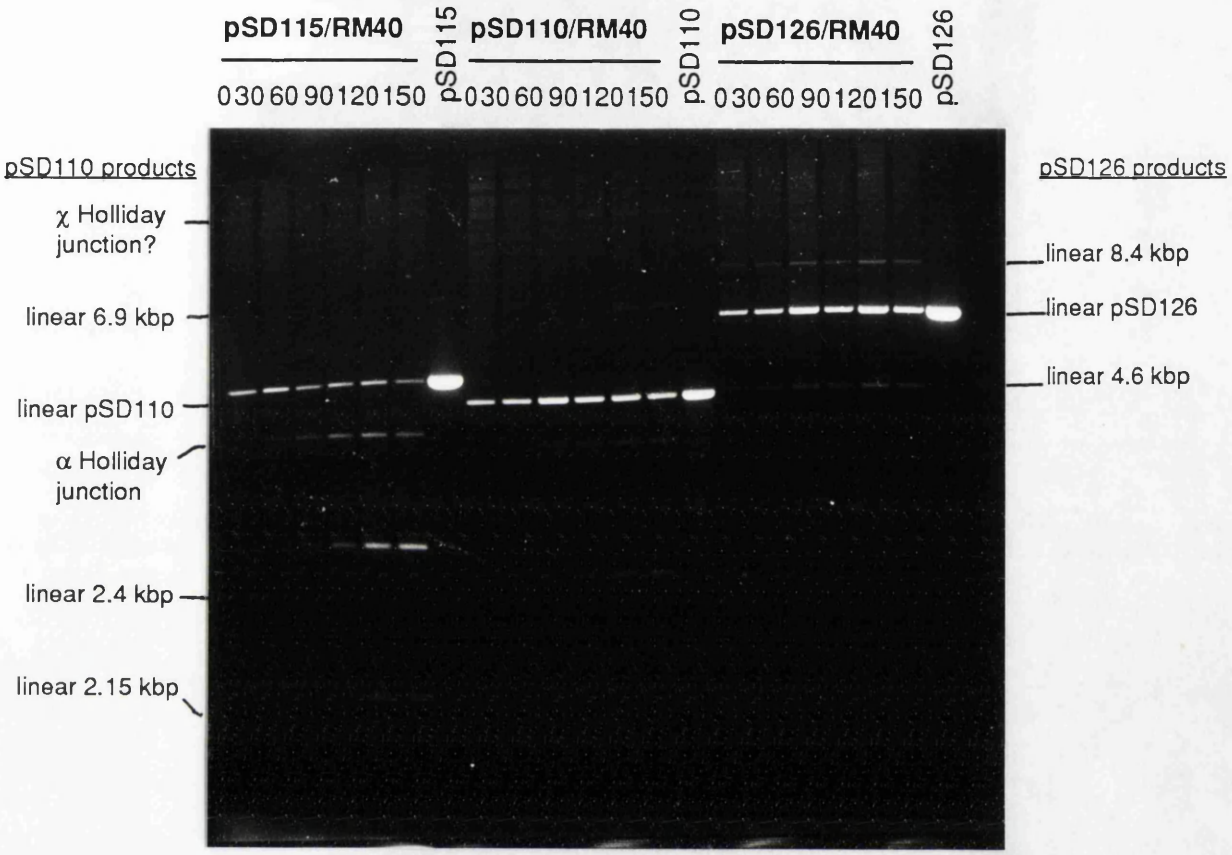


Figure 5.12.(A) Comparison of *cer*, *dif* and type II hybrid recombination in RM40 time course assays. RM40 was transformed with either pSD115, pSD110 or pSD126. Pooled colonies of the transformants were used to set up separate recombination assays in which XerC expression was induced using 2 mM IPTG when the cultures reached mid-log phase.Plasmid DNA samples were prepared by the boiling method at the times subsequent to induction shown and were digested with *EcoR*I (pSD115 and pSD126) or *Pst*I (pSD110) before they were run on a 1.2% agarose gel. Both pSD110 and pSD126 will recombine intramolecularly and intermolecularly in Xer^+ strains (see fig.5.13). Resolution of pSD110 yields a replicative, 2.4 kbp and a non-replicative, 2.15 kbp product; resolution of pSD126 yields a replicative, 4.6 kbp and a non-replicative, 1.9 kbp product (in both plasmids restriction digestion linearises the larger, replicative product). Restriction of the intermolecular recombination products creates a 6.9 kbp (pSD110) and a 8.4 kbp (pSD126) linear DNA fragment. α Holliday junctions are indicated.

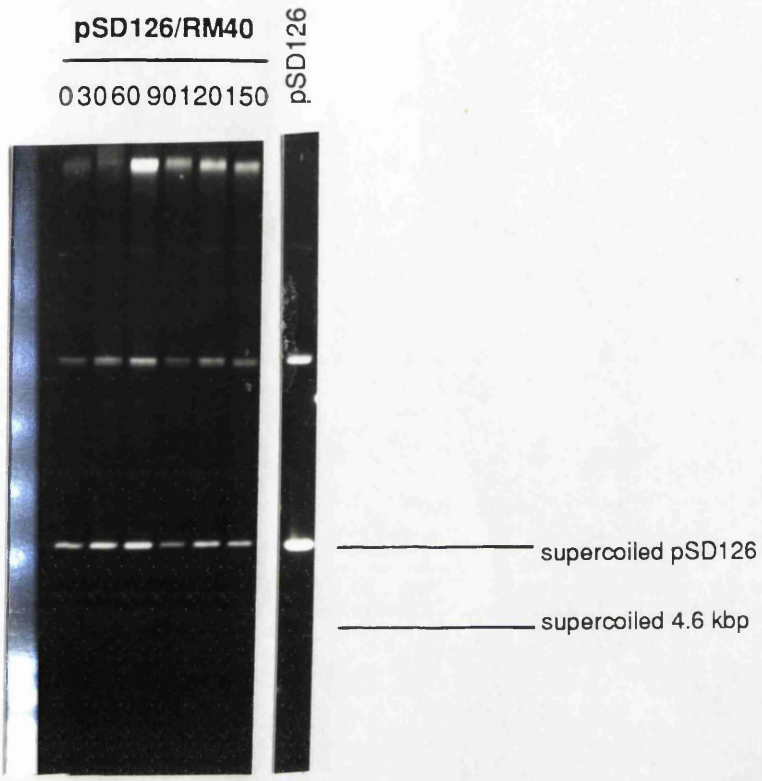


Figure 5.12 (B) Unrestricted pSD126 *in vivo* recombination time course. The DNA samples from the pSD126 time course shown after *EcoR*I digestion in figure 5.12A were run without restriction digestion on a 1.2% agarose gel. Intermolecular, *dif* mediated site-specific recombination in Xer^+ strains produces mutimers of pSD126 (for example, see fig.5.22).

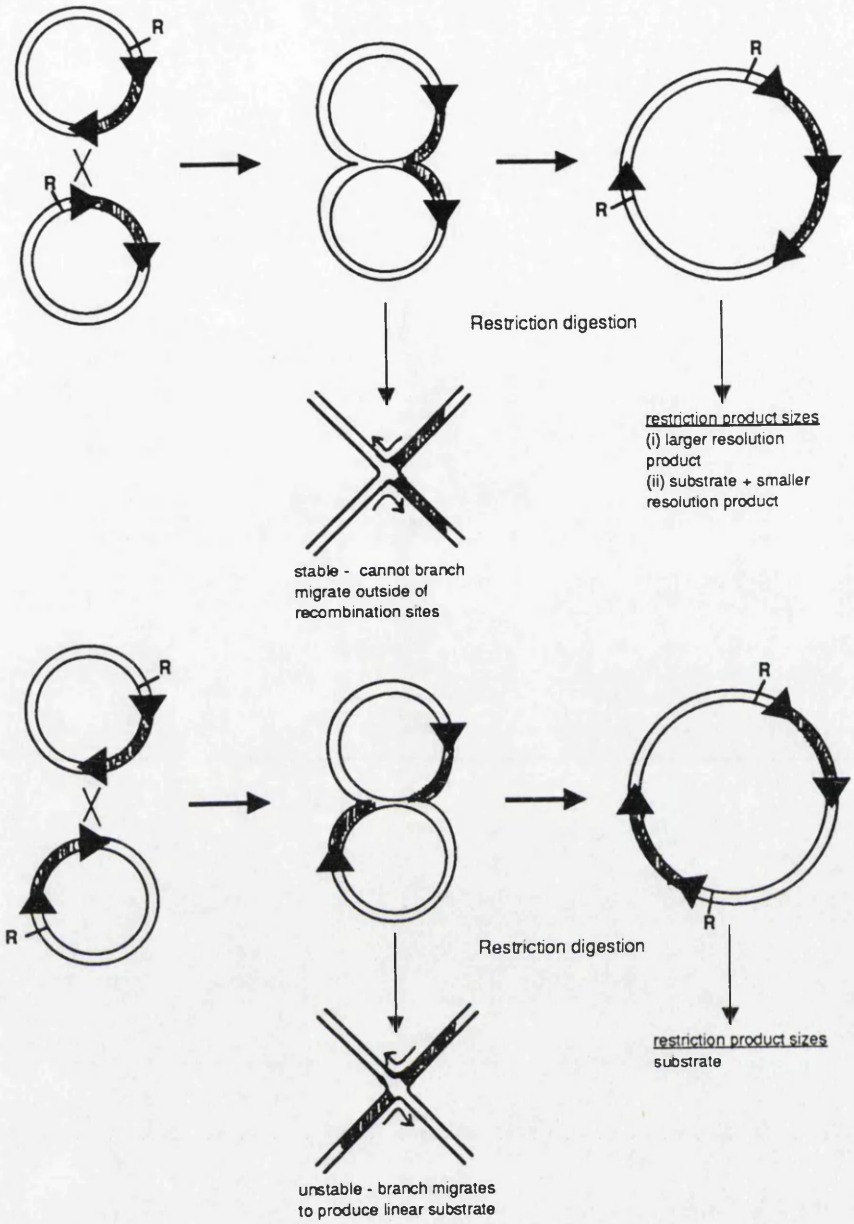


Figure 5.13 Diagrammatic representation of intermolecular site-specific recombination of reporter plasmids containing two recombination sites. (Restriction enzyme recognition site is shown as R). Two intermolecular recombination reactions are possible that produce different products which can be distinguished by the different DNA fragments they create after restriction digestion. Intermolecular recombination of both *dif* and the type II hybrid is expected to proceed *via* a Holliday junction; one junction is free to branch migrate to yield linear DNA fragments after restriction digestion, digestion of the other forms a χ structure.

restriction of one dimer yields two substrate-sized linear molecules. Restriction of the other dimer yields a linear, replicative product-sized molecule, and second linear whose size represents two non-replicative products plus one replicative product (hence it is larger than the substrate). Restriction of higher multimers, e.g. trimers or tetramers, will yield the same sized DNA fragments. The slowest migrating band may represent the χ form DNA species that would be produced by *PstI* restriction of the supercoiled Holliday junction of the latter type of dimer described above; unrestricted this molecule would have a mobility indistinguishable from supercoiled pSD110 dimer. (Note that restriction of the former dimer's Holliday junction would result in linear substrate molecules because the junction would be free to branch migrate to the ends of the χ.) Comparing the amount of reaction products seen in pSD110 recombination to pSD115 may suggest that *cer* is more efficient than typeII. However, this is complicated by the distribution of the typeII reaction products between dimerisation (half of whose products are not seen after *PstI* restriction) and resolution.

In contrast to these results, pSD126 did not appear to have recombined during the time course of this assay. At all time points, including the 0' sample, some reaction products (4.6 kbp linear resolution and 8.4 kbp dimerisation products) were visible; they did not, however, increase in quantity after XerC induction. In addition, there was no sign of the 1.9 kbp, non-replicative resolution product. Figure 5.12B shows the unrestricted DNA samples of the pSD126 time course, illustrating that the reaction had not exclusively recombined in the direction of multimerisation. These facts suggest that *dif* recombination did not respond to the induction of XerC. An explanation for this could be that the lack of accessory sequences makes *dif* a poorer site for *xer* recombination than *cer*. This seems unlikely, however, because pSD110 contains "minimal" typeII sites and recombined efficiently. The lack of recombination may therefore suggest that an essential factor for *dif* recombination was missing in this experiment, or that some feature of the *in vivo* assay was inhibiting the reaction. Because the reaction did not proceed, *EcoRI* restriction could not determine whether *dif* recombination involves the production of stable Holliday junctions.

5.7 The effect of mutations in *dapF*, *orf235* and *orf238* on *dif* recombination

The lack of pSD126 recombination in RM40 may be a consequence of the *dapF* and *orf235* mutations in this strain (section 4.2), despite the fact that these mutations were shown not to be the cause of Holliday junction accumulation during *cer* recombination. It is possible that the proteins that are co-expressed with XerC have a role in the chromosomal site-specific recombination reaction that has been overcome in the plasmid-based *cer* reaction (perhaps by the use of ArgR and PepA). If the proteins do have this type of function it is not clear how they might act. It seems unlikely they directly interact with the *dif* site, since it has been demonstrated that the 33 bp core sequence is sufficient for recombination of reporter plasmids (Blakely *et al*, 1991). However, this does not exclude the possibility that they could have a role in activation or regulation of XerC, or indeed in interacting with a larger, undefined *dif* site.

This possibility was first tested by analysing pSD126 recombination in *orf235* and *orf238* mutant strains, and comparing this recombination to that in RM40 and to pSD115. This was done because the Xer phenotype of these mutants has only previously been determined for *cer* and typeII recombination (Colloms, 1990; Colloms *et al*, 1990). The strains used were: DS980 (*orf235*::Tn5), DS981 (*xerC*::Kan) and DS982 (*orf238*::miniMu, chloramphenicol resistant); a suitable insertional mutant of *dapF* was not available. pSD126 and pSD115 were transformed into DS941, DS980, DS981 and DS982 and plated on ampicillin L-agar; the plasmids were also transformed into RM40 and plated on ampicillin L-agar supplemented with either 1.0% glucose or 2 mM IPTG. In addition, the reporter plasmids and the XerC expression vector pSD107 were doubly transformed into DS980 and DS981 and grown on L-agar containing both ampicillin and chloramphenicol. The plasmid DNA from these transformants was analysed by agarose gel electrophoresis (fig.5.14).

Both pSD115 and pSD126 recombined in DS982, suggesting that *orf238* (downstream of *xerC*) is not required for *dif* recombination. The Tn5 insertion in

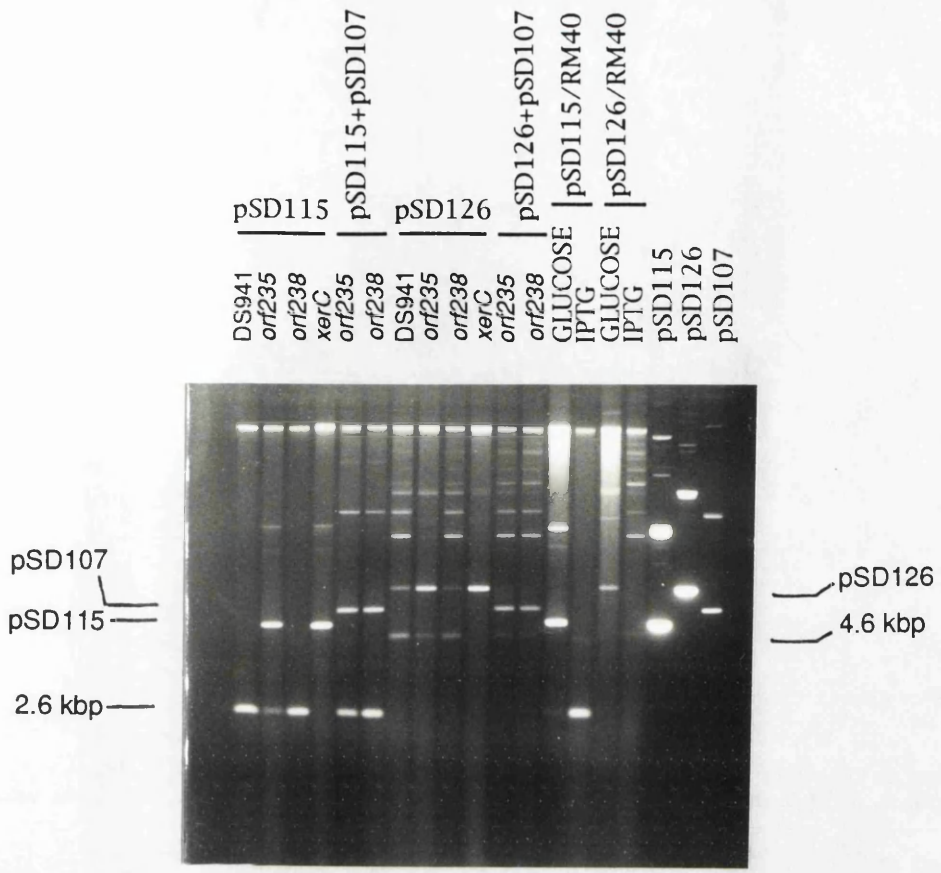


Figure 5.14. *cer* and *dif* site-specific recombination in *orf235*, *xerC* and *orf238* insertional mutants, and in RM40. The DS941 derivatives shown (described in the text) were transformed with pSD115 or pSD126 alone or in conjunction with the XerC expression vector pSD107. RM40 was transformed with pSD115 or pSD126 and grown on either 1% glucose ot 1 mM IPTG. Plasmid DNA was prepared by the single colony lysis method from patched cultures of the transformants and was run on a 1.0% agarose gel. In Xer^+ strains pSD115 is resolved to produce a 2.6 kbp plasmid whilst pSD126 is recombined both intermolecularly and intramolecularly (producing a 4.6 kbp plasmid). pSD107 contains a minimal complementing clone of *xerC* and is compatible with both reporter plasmids in its antibiotic resistance and replication origin.

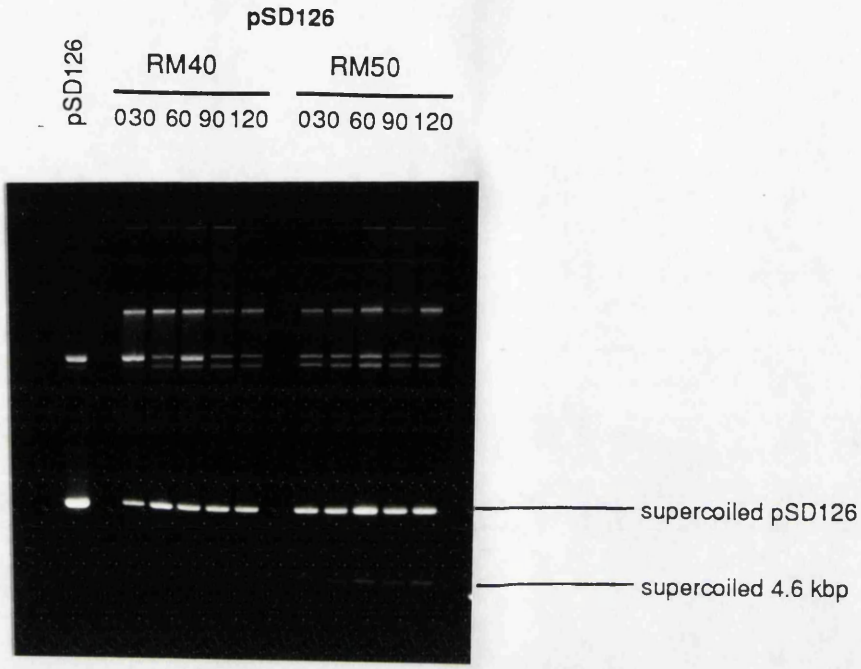


Figure 5.15. Comparison of pSD126 time course recombination *in vivo* using the strains RM40 and RM50. RM40 and RM50 were both transformed with pSD126, and XerC expression was induced with 2 mM IPTG in both cultures as they reached mid-log phase. DNA samples were prepared by boiling preparations at the times shown after induction and run on a 1.2% agarose gel.

DS980 has been shown to have a "leaky" Xer$^-$ phenotype (Colloms, 1990); this is illustrated by the partial resolution of pSD115 in this strain compared to the lack of resolution in DS981, suggesting that a small amount of XerC is expressed in DS980 but none in DS981. This pattern was also seen for *dif*: incomplete recombination of pSD126 occured in DS981 (compared to the large amount of multimers/resolution products in DS941). The partial mutation was complemented by pSD107, which contains a minimal *xerC* clone. Both these results suggest that *orf235* (upstream of *xerC*) is not involved in *dif* recombination. This experiment also showed that when RM40 was grown on IPTG-containing plates pSD126 was recombined to the same extent as in DS941, suggesting that the lack of *dif* recombination during time course assays is a result of some subtle effect, perhaps due to growth in liquid culture.

Despite this analysis it remains possible that the lack of pSD126 recombination seen previously in RM40 is a consequence of the fact that the operonic genes have a subtle, and not absolute, effect on the *dif* reaction. Furthermore, these results cannot rule out the possibility that *dapF*, which is also mutant in RM40, has a role in the reaction. For these reasons it was decided to examine pSD126 recombination in the strain RM50 (see 4.9; fig.4.16). Both RM40 and RM50 were transformed with pSD126 and treated in exactly the manner described in section 5.6 to perform recombination assays. Figure 5.15 shows the unrestricted DNA from these assays. As previously described, the repression of *xerC* is not as complete in RM50 and therefore some recombination products were visible at time 0'. After induction of XerC there did not appear to be any greater increase in reaction products in RM50 than in RM40, suggesting that the *dapF* and *orf235* mutations are not the cause of the lack of *dif* recombination in the last RM40 time course assay. Since there was no detectable pSD126 recombination it cannot be said with certainty that these genes have no role in *dif* recombination, but the previous experiments suggest this may be the case.

5.8 pSD126 recombination controlled by pRM60

The construction and use of this plasmid is addressed in section 4.10. This experiment was performed because it tested two possibilities: (i) the lack of *dif* recombination may be a result of unidentified mutations in the strain RM40, (ii) recombination of pSD126 may require a different XerC induction regime; it is likely, but not certain, that pRM60 expresses XerC to a higher level than RM40 because the λ P_L promoter is thought to be stronger than the *lac* promoter (Reznikoff and McClure, 1986), and in pRM60 *xerC* is present in multiple copies.

DS942*xerC*PS6 was doubly transformed with pRM60 and pSD126 and grown at 30 °C on ampicillin, chloramphenicol L-agar plates. A number of transformants were then used to set up a 20 ml L-Broth culture which was grown, shaking, at 30 °C until it reached mid-log phase. At this point a sample of DNA was prepared using standard boiling preparations and the temperature of the remaining culture was elevated to 42 °C for 30 mins. DNA was again sampled, the culture was put to 37 °C, and further samples prepared every 30 mins for another 90 mins. The assay performed in this way was analysed, before and after restriction digestion with *EcoR*I, by agarose gel electrophoresis (fig.5.15).

In both the restricted and unrestricted time courses, using pRM60 did not result in *dif* recombination after XerC induction, and this result appears to confirm that seen in RM40 assays. The implications of this are similar to those discussed for *cer* in section 4.10; the lack of recombination in the assay is unlikely to be due to mutations in *xerC* or in other genes, but instead has more general implications for the mechanism of the recombination reaction. It cannot be said with certainty how much XerC was expressed from pRM60 in this experiment, but changing the method, and perhaps the level, of XerC expression did not alter the pattern of *dif* recombination in these assays.

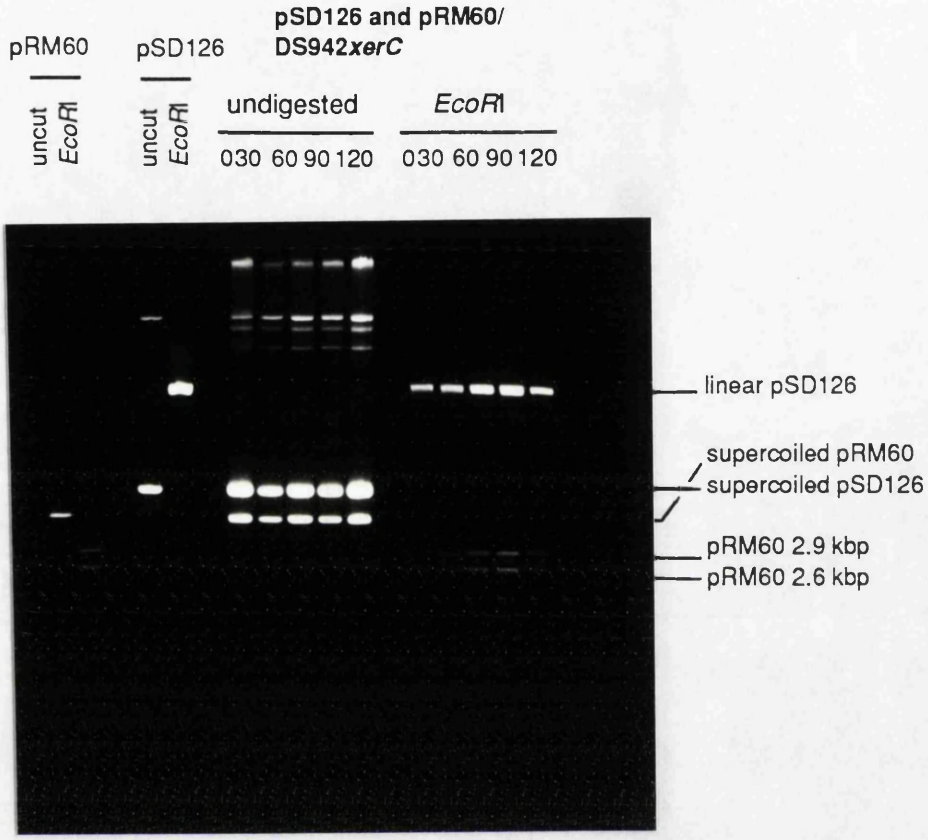


Figure 5.16 pSD126 *in vivo* recombination time course controlled by the plasmid pRM60. DS942*xerC*::Kan was transformed with pSD126 and pRM60 was grown at 30 °C to mid-log phase. XerC expression was induced by elevating the culture temperature to 42 °C and plasmid DNA was prepared at the times subsequent to induction as shown. The DNA was run on a 1.2% agarose gel before and after restriction digestion by *EcoR*I. *EcoR*I restriction of pRM60 produces two linear DNA fragments of 2.6 and 2.9 kbp; restriction of the recombinant products of pSD126 is described in figure 5.12 and examples of pSD126 *dif* mediated multimerisation are shown in figure 5.14.

5.9 Comparison of pSD126 and pSD124 recombination in RM40

It has been determined previously that the 33bp of *dif* which are homologous to the *cer* core sequence are sufficient to support *dif* recombination in DS941 (Blakely *et al*, 1991). The reporter plasmid pSD126 contains two copies of the 532bp *dif* sequence initially cloned from the *E.coli* terminus by Kuempel *et al* (1991). Despite the fact that this plasmid is also efficiently recombined in DS941, it is conceivable that the uncharacterised DNA around the 33bp core could contain sequences that serve to inhibit, or regulate, *dif* recombination, and that this is only apparent in short term assays. There is no published evidence that this should be the case, nor is it clear how these sequences might act (or what proteins might bind them), but it remains a possibility that could explain the above observations.

This was tested by comparing the recombination of the plasmids pSD126 and pSD124 in RM40. pSD124 is a reporter plasmid of the same design as pSD126, but constructed in pUC18 and containing directly repeated minimal *dif* sites (Blakely *et al*, 1991). The plasmids were transformed separately into RM40 and selected on L-agar, ampicillin, 1.0% glucose, 50 ug/ml diaminopimelic acid plates. 20 ml L-Broth cultures were set up from these plates and *in vivo* recombination assays were performed. Plasmid DNA samples were prepared after XerC induction at the times shown, restricted by *EcoR*I and separated by agarose gel electrophoresis (fig.5.17).

Recombination of pSD124 (which is 4.4kbp in size) should both multimerise and resolve the plasmid. After *EcoR*I restriction multimers would be seen as a 6.1 kbp DNA fragment, and the resolution products as a linear 2.7 kbp fragment (the replicative product) and a supercoiled 1.7 kbp species (the non-replicative product). However, the plasmid did not recombine substantially over the 180 mins of this assay, since none of these species increased in quantity after the induction of XerC. This suggests that pSD126 does not contain a "silencer" element within the larger *dif* site, and supports the belief that this lack of recombination in RM40 assays may be a significant feature of *dif*'s reaction mechanism.

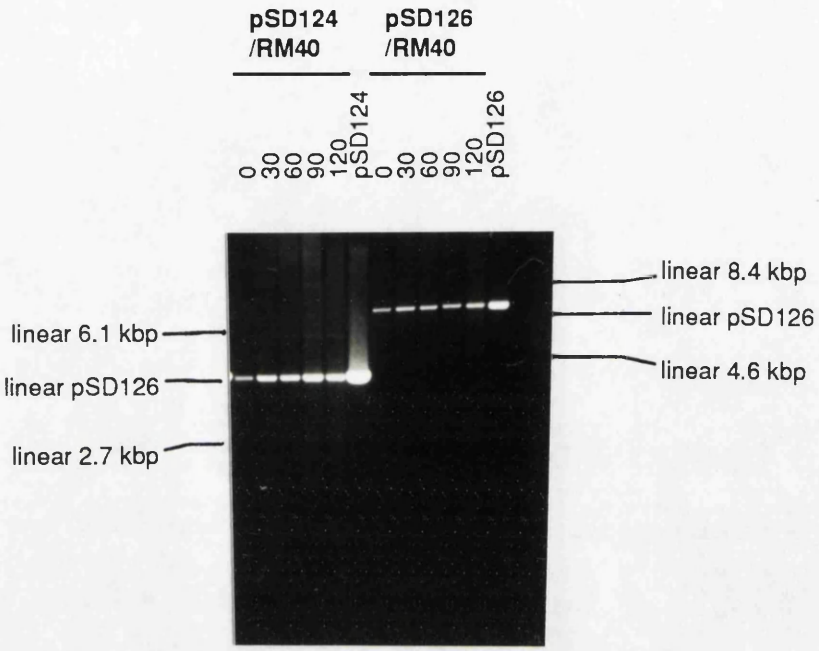


Figure 5.17. Comparison of 532 bp *dif* and minimal, 33 bp *dif* site-specific recombination in RM40. The reporter plasmids pSD126 and pSD124 were transformed into RM40 and *in vivo* recombination assays were performed using a number of colonies from each transformation. Plasmid DNA was isolated by the boiling method at the times subsequent to XerC induction shown and was run on a 1.2% agarose gel after *EcoR*I restriction digestion. pSD124 contains two directly repeated 33 bp *dif* sites and is a pUC18-based plasmid (Blakely *et al*, 1991; the products of *dif* mediated site-specific recombination of pSD124 are described in the text). See figure 5.12 for a description of pSD126's *dif* mediated recombinant products.

5.10 Genetic characterisation of a fourth *xer* gene: *xprB*

While the above analysis was being performed, our attention was drawn to an *E.coli* gene, *xprB*, which encodes a protein that was shown to have sequence homology to the λ integrase family of site-specific recombinases (Lovett and Clark, 1985; Lovett and Kolodner, 1991). *xprB* is situated upstream of *prfB* (which encodes the translation release factor RF2) and has been genetically mapped to 62 mins on the *E.coli* chromosomal map. It is thought to be the first gene in an operon containing two other members - *xprA* and *recJ*. The belief that these genes are co-translated is based on their close proximity, as revealed by DNA sequencing, and the fact that Tn10 insertions within either *xprB* or *xprA* exert a polar effect on the expression of RecJ. It is possible, however, that these insertions are simply disrupting the *recJ* promoter, since the translation start point(s) of these genes have not yet been determined. The naming of the two *xpr* genes was a reflection of the fact that their cellular roles were unknown. Maxicell labelling techniques were used, however, to demonstrate that they encoded proteins of the predicted sizes. Homology searches using the putative sequence of XprA have so far failed to yield any results, although there is some evidence to suggest that it is exported to the periplasm or outer membrane and is a membrane-associated protein. RecJ has been shown to be involved in the RecF pathway of homologous recombination and the purified protein has single-strand, DNA-specific exonuclease activity. The fact that the putative protein sequence of XprB shows some homology to the λ integrase family of recombinases prompted us to consider whether it may be involved in the *xer* site-specific recombination reactions.

Figure 5.18 shows a protein sequence alignment between XprB, XerC and some relevant members of the λ integrases. The two domains detailed have been demonstrated to be the only regions conserved within this highly divergent family (Argos *et al*, 1986; Hoess and Abremski, 1992). The alignment illustrates that even within these domains there is considerable sequence variation, indeed only the four amino acid residues in bold are found in all identified members of the family.

Domain 1

Lambda Int	TGQ**R**VGDLCEMKWSD
Flp	NCG**R**FSDIKNVDPKS
Cre	TLL**R**IAEIARIRVKD
P22 Int	TGL**R**RSNIINLEWQQ
R46 ORF	TGM**R**ISEGLQLRVKD
Tn554 TnpA	GGL**R**IGEVLSLRLED
Tn554 TnpB	CGM**R**ISELCTLKKGC
FimB	HGF**R**ASEICRLRISD
FimE	HGM**R**ISELLDLHYQD
XerC	AGL**R**LSELVGLDIKH
XprB	TGL**R**VSELVGLTMSD

Domain 2

Lambda Int	**H**EL**R**SLSA-RLYEKQ-ISDKFAQHLLGHKS-DTMA-SQ**Y**RD
Flp	**H**IG**R**HLMTSFLSMKGLTELTNVVGNWSDKRASAVART**TY**TH
Cre	**H**SA**R**VGAARDMARAG-VSIPEIMQAGGWTN-VNIV-MN**Y**IR
P22 Int	**H**DL**R**HTWASWLVQAG-VPISVLQEMGGWES-IEMV-RR**Y**AH
R46 ORF	**H**TL**R**HSFATALLRSG-YDIRTVQDLLGHSD-VSTT-MI**Y**TH
Tn554 TnpA	**H**ML**R**HTHATQLIREG-WDVAFVQKRLGHAHVQTTL-NT**Y**VH
Tn554 TnpB	**H**AF**R**HTVGTRMINNG-MPQHIVQKFLGHES-PEMT-SR**Y**AH
FimB	**H**ML**R**HSCGFALANMG-IDTRLIQDYLGHRN-IRHT-VW**Y**TA
FimE	**H**NL**R**HACGYELAERG-ADTRLIQDYLGHRN-IRHT-VR**Y**TA
XerC	**H**KL**R**HSFATHMLESS-GDLRGVQELLGHAN-LSTT-QI**Y**TH
XprB	**H**VL**R**HAFATHLLNHG-ADLRVVQMLLGHSD-LSTT-QI**Y**TH

Figure 5.18. Alignment of the protein sequences of some members of the λ integrase family. Not all of the published integrase-like proteins are included, and only the two most conserved regions (domains 1 and 2) are shown. The residues that are found to be conserved in all known proteins with homology to this family are shown in bold face. The sequence of XprB is from Lovett and Kolodner (1991); the sequences of the other proteins (except domain 1 of Cre and FLP) are taken from Colloms *et al* (1990). The alignment of domain 1 (and the sequences of Cre and FLP in this domain) is adapted from Abremski and Hoess (1992), whilst domain 2 is adapted from Colloms *et al* (1990). The positions of these domains within the proteins' sequences vary, but domain 1 is always found nearer the N-terminus than domain 2.

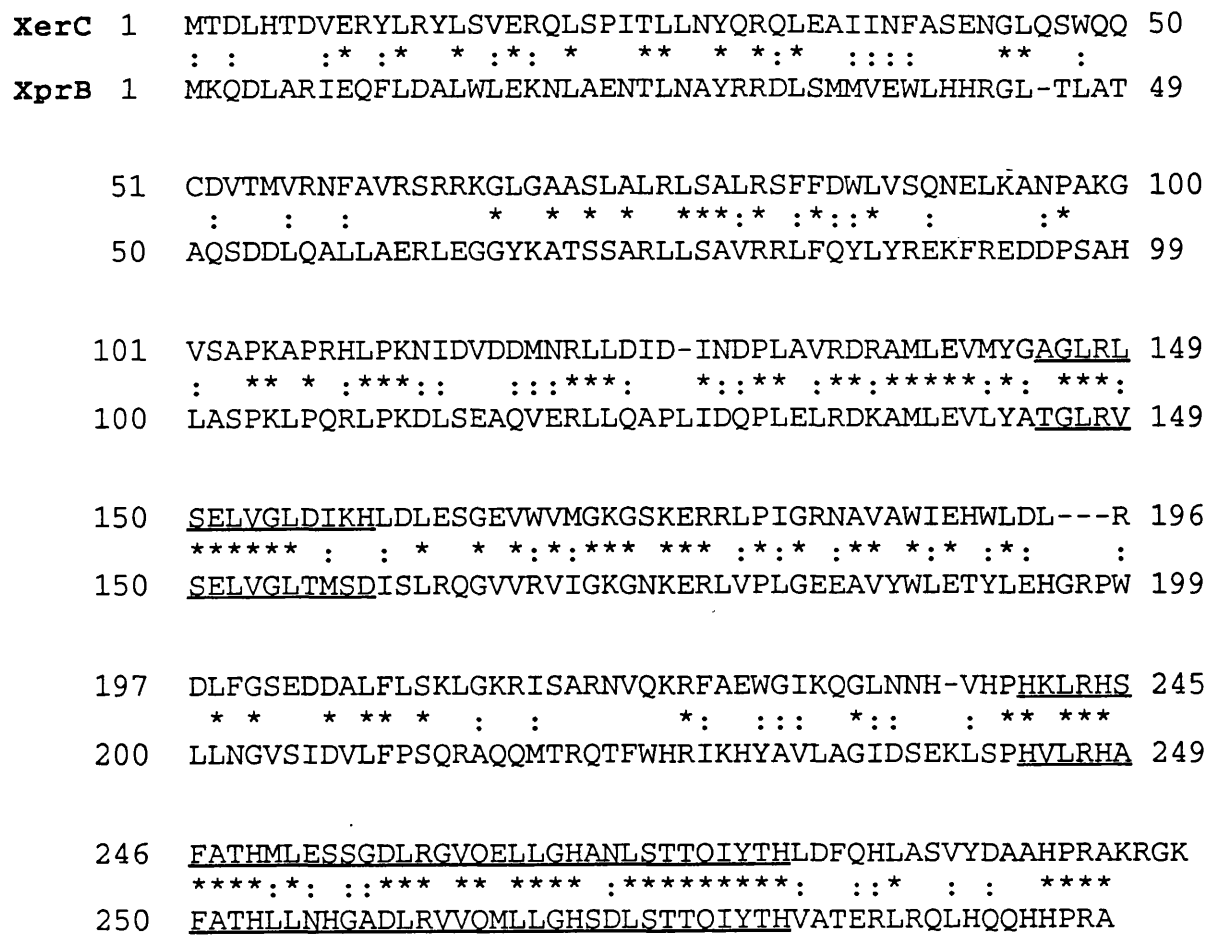


Figure 5.19. Comparison of the protein sequence of XerC with the sequence of XprB. Gaps (-) have been introduced in the sequences to maximise the homologies. Identical residues are indicated by an asterisk (*) and conservative changes (within the exchange groups (V, L, I, F, Y, M, W) - (A, T, G, S, C) - (H, K, R) - (D, E, Q, N)) are shown by a colon (:). The residues that are homologous to the two conserved domains found in the λ integrase family of site-specific recombinases are underlined.

Experimental evidence exists which implicates these four invariant residues in the catalytic mechanism employed by these proteins. The conserved tyrosine of λ Int has been shown to make a phosphodiester bond with the cleaved *att* site during strand exchange in certain *in vitro* reaction conditions (Pargellis *et al*, 1988), and a similar covalently bound intermediate has been isolated between the tyrosine of FLP and the FRT site (Gronostajski and Sadowski, 1985b; Evans *et al*, 1990). Random mutagenesis of Cre has isolated proteins with alterations in all three conserved domain 2 residues, and in all cases the proteins will still bind *loxP* but are unable to support recombination (Wierzicki *et al*, 1987; this study also isolated mutants in other residues within domain 2 and many showed the same phenotype); in addition, the conserved domain 1 arginine has been specifically altered and the mutant protein is recombination defective (Abremski and Hoess, 1992). In FLP, mutageneses of all four residues have been carried out (Parsons *et al*, 1988; 1990) and mutation of any of them results in arrest of the recombination reaction at the strand cleavage or exchange steps. The fact that XprB contains all of these conserved residues and that they are suitably spaced lends support to the contention that it may be a λ integrase-like recombinase, and that it is capable of catalysing a site-specific recombination reaction.

Figure 5.19 shows an alignment between the entire sequences of XerC and XprB, and highlights the conserved domains. This demonstrates that the proteins appear to have the same number of constituent amino acids and the homology between them extends beyond domains 1 and 2. The λ integrases are divergent not only in sequence but also in their sizes, which is probably a reflection of the variation in their functions and the accessory factors they employ. 37% of amino acids are identical between XerC and XprB; this is higher than the identities between either XerC or XprB and either of the other *E.coli* chromosomal recombinases, FimB and FimE (the most similar pair are FimE and XerC which have 28% identical residues; FimB and E have 50% identity). The homologies that exist between these four proteins extend across the region spanning domains 1 and 2. XerC and XprB show most pronounced similarity to each other, and differences to other proteins, in the region upstream of, and including,

domain 1. This displays 60% amino acid identity, whilst the most homologous other protein is the *E.coli* F factor's recombinase D (Lane *et al*, 1986) which is 40% identical to XprB. Domain 2 appears more generally conserved between all the λ integrases than domain 1, and here XerC/XprB are no more related (70% identity) than XprB is to R46's ORF. This analysis suggests that XerC and XprB are somewhat more related to each other than to other members of the family but its significance is not clear, e.g. it may be a result of both being resident in the *E.coli* chromosome (hence the overall similarities to FimB/E) or it may have some functional relevance.

In order to test whether *xprB* has a role in the *xer* system S.Lovett kindly supplied us with the strain STL116 (*xprB*::miniTn10 kanamycin) and the plasmids pJC763 and pRDK168 (Lovett and Kolodner, 1991; see fig.5.20). Mary Burke transformed STL116 with the plasmids pCS202, pKS455 and pSD124 and showed, by agarose gel electrophoresis of the isolated DNA, that none recombined - indicating that the strain contains an *xer* mutation. It was further shown that pRDK168 (which contains the 1.7 kbp *Hind*III-*Eco*RI fragment from pJC763) would complement the STL116 mutation and allow resolution of pCS202, whilst pSD105, pX3 and pCS126 (vectors expressing XerC, ArgR and PepA respectively) would not. This indicated that, at least for *cer*, mutations in either *xprB* or *xprA* are responsible for the Xer^- phenotype of STL116.

To further this analysis the sub-clones detailed in figure 5.20 were constructed from pJC763, and Mary Burke transduced the miniTn10 insertion from STL116 into DS941, creating DS9008. The first question that was asked was whether it is XprB or XprA that is required for *xer* recombination. To attempt to answer this, the plasmids pRM130, 131 and 132 were transformed into DS9008 with pCS202 and selected on L-agar plates containing ampicillin and chloramphenicol. The transformants were then patched out and their DNA analysed by agarose gel electrophoresis (fig.5.21). pRM130 contains the same 1.7 kbp *Hind*III-*Eco*RI fragment (encompassing *xprB* and almost all of *xprA*)that is present in pRDK168, but here it is cloned into pUC19 such that the translation of the insert is driven by the *lac* promoter. It was hoped this plasmid would

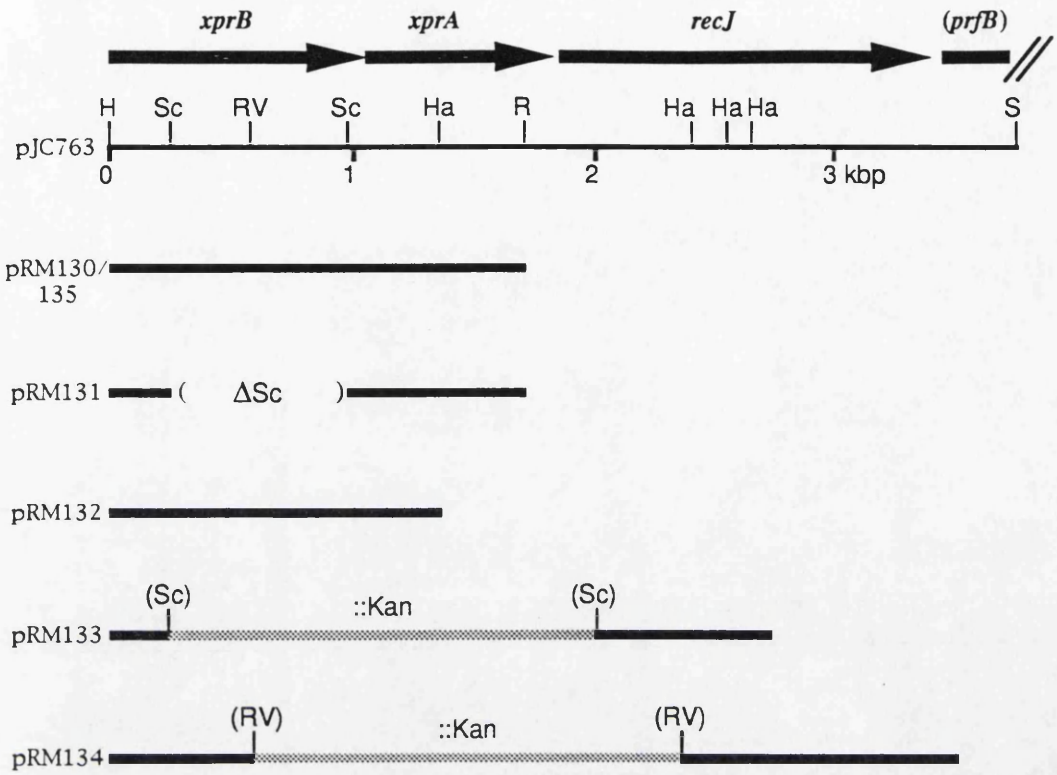


Figure 5.20. Plasmid constructions used in the genetic analysis of *xprB*'s role in *xer* site-specific recombination. The upper diagram shows a restriction map of the *E.coli* chromosomal insert present in the plasmid pJC763 (Lovett and Kolodner, 1991). The genes that are contained in this fragment are indicated above (arrows denote the direction of their transcription); it is thought that *xprB*, *xrpA* and *recJ* are co-expressed. Only a part of the downstream gene *prfB* is present within this insert. Below this map is shown the extent of chromosomal fragments carried by various plasmids. The chromosomal DNA in the plasmids pRM130, 131, 132, 133 and 134 is cloned into pUC19; the insert in pRM135 is in pCB106. In all the plasmids the inserts are orientated such that they can be transcribed by the vector's *lac* promoter. Δ denotes a deletion of a part of the chromosomal DNA, and ::Kan denotes the insertion of the Tn903 kanamycin resistance gene.

Restriction enzyme recognition sites sites are shown as follows:

H=*Hind*III R=*Eco*RI S=*Sal*I Sc=*Sca*I Ha=*Hae*II RV=*Eco*RV

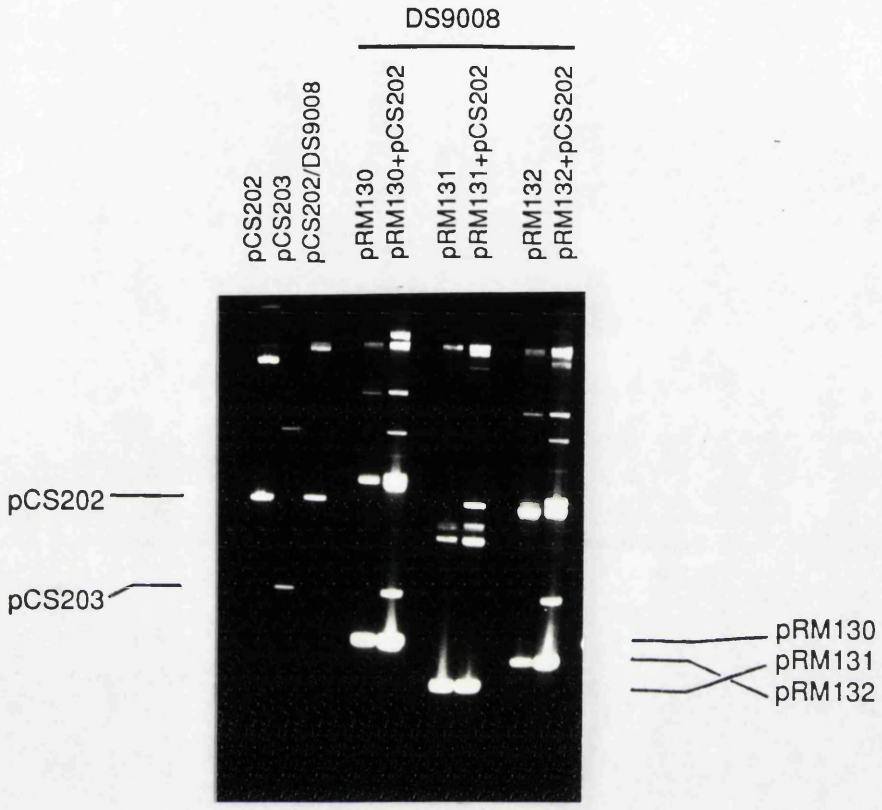


Figure 5.21. Complementation of *xprB*::miniTn10 with deletion derivatives of pRM130. DS9008 (DS941*xprB*::miniTn10) containing pCS202 was transformed with the plasmids pRM130, pRM131 and pRM132. Plasmid DNA was isolated by the boiling method and visualised on a 1.2% agarose gel. pCS202 is a *cer* reporter plasmid (fig.4.21) that is resolved in Xer^+ strains to give the plasmid pCS203.

allow the overexpression of XprB for the purposes of purification; pRDK168 is unsuitable for this function because the expression of the two genes relies on their own, undefined promoter sequences. All of *xprA* except the final 11 codons is included in pRM130 (therefore it is possible that active XprA is expressed), and therefore the plasmid could not determine whether XprB or XprA was complementing the *xer* mutation in DS9008. The plasmid pRM131 is a *Sac*II deletion derivative of pRM130 and was not able to complement DS9008; the deletion is of approximately 760 bp and encompasses almost all of *xprB* (it leaves only the amino-terminal 42 amino acids of XprB). Since this deletion leaves *xprA* intact it suggests that expressing XprA alone is insufficient to complement the mutation in DS9008. pRM132 is a second deletion derivative of pRM130: the region cloned is the 1.3 kbp *Hind*III-*Hae*II fragment of pJC763 shown (cloned in the same orientation in pUC19 as pRM130). This plasmid did allow resolution of pCS202, and, because it only encodes the amino-terminal 104 amino acids of XprA (it deletes the C-terminal 131 residues), it suggests that XprB alone is capable of complementing the DS9008 *xer* mutation.

This experiment illustrates that a minimal *xprB* clone, but not a minimal *xprA* clone, is able to complement DS9008 and make it Xer^+. The experiment therefore offers further genetic evidence that a fourth gene is essential for *cer* site-specific recombination. If the Tn10 insertion in this strain is polar on *xprA* then the fact that pRM131 was unable to complement suggests that it is the mutation in *xprB*, and not in *xprA*, that is making these cells Xer^-. This was expected, since it is XprB that has sequence homology to the λ integrases. A role for XprA cannot, however, be ruled out by these experiments since it is possible that the protein is expressed to sufficient levels in DS9008 for it to function. Analysis of the *xer* phenotype of insertional mutants in *xprA* would be required to rigorously rule this out.

Because both XerC and XprB have sequences compatible with them being recombinases it was decided to test whether overexpression of one protein could overcome a chromosomal defect in the other. This was analysed for both *cer* and *dif* recombination. The plasmid pRM135 contains the same insert as pRM130, also

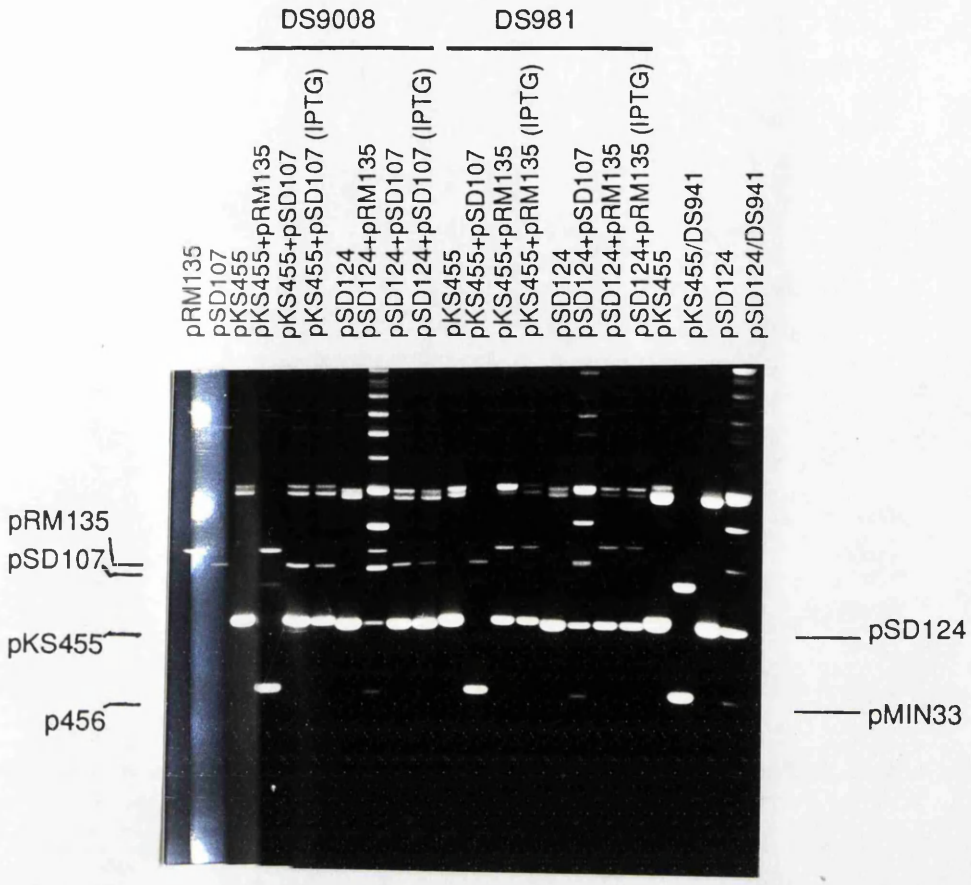


Figure 5.22. Complementation of *xprB* and *xerC* insertional mutants with XprB and XerC expression vectors. DS9008 (DS941*xprB*::miniTn10) and DS981 (*xerC*::Kan) were transformed with pKS455 and pSD124 alone or in conjunction with pRM135 or pSD107; where indicated the cells were grown on media containing 1 mM IPTG. Plasmid DNA was isolated and run on a 1.2% agarose gel. The plasmids pRM135 and pSD107 (expressing XprB and XerC respectively) both encode chloramphenicol resistance and have λdv replication origins. pKS455 is a *cer* reporter plasmid (fig.4.20) that is resolved in Xer^+ strains to give p456; pSD124 is a *dif* reporter plasmid (Blakely *et al*, 1992) that recombines intramolecularly (to give a 2.7 kbp product called pMIN33) and intermolecularly (causing multimers) in Xer^+ strains.

expressed from the *lac* promoter, cloned into the vector pCB106 (which is chloramphenicol resistant and has a λdv origin of replication). This plasmid is essentially the XprB equivalent of the XerC expression vector pSD107 (*xerC* cloned into pCB106; Colloms, 1990), and is compatible with the reporter plasmids pKS455 and pSD124 (fig.4.20 and section 5.9). DS9008 and DS981 (*xer*C::kanamycin; constructed by P.Sykora) were transformed with pKS455 and with pSD124 and grown on L-agar containing ampicillin. These four strains were then transformed with pRM135 and pSD107, selected on ampicillin and chloramphenicol and patched out onto appropriate plates; they were also patched onto L-agar supplemented with 1 mM IPTG in order to induce the expression of XerC/XprB to higher levels. DNA was prepared by standard boiling preparations and separated by agarose gel electrophoresis (fig.5.22).

This analysis demonstrates that XerC cannot complement an *xprB* mutation and XprB cannot complement an *xerC* mutation. This is true for both *cer* and *dif* recombination: both pKS455 and pSD124 would only recombine in DS9008 if pRM135 was present, and in DS981 if pSD107 was present; neither plasmid would recombine in DS9008 when complemented with pSD107 or in DS981 with pRM135 - even when the proteins were over-expressed using IPTG. This offers evidence that both XerC and XprB are required for *xer* site-specific recombination, and having an abundance of only one of the putative recombinases is insufficient for either the *cer* or *dif* reactions.

Identification of a fourth *xer* gene, which may encode a second recombinase, has implications regarding the *xer* reaction mechanism. Of direct relevance is the possibility that XprB may explain the observations described in this chapter and in chapter 4. This is examined in the following section.

5.11 Analysis of *cer* and *dif* recombination in RM40/pRM135

The experiments performed until now in analysing *dif* recombination have suggested that the reaction does not respond to induction of XerC expression in RM40 with IPTG, despite the fact that the strain has been shown to be capable of supporting *dif* recombination when grown on plates. Similarly, pSD115 (and other *cer* reporter plasmids) when recombining in RM40 produce large amounts of Holliday junctions which may not be present during recombination in wild type cells and are unusual in comparison to other integrase-like reactions. All these experiments have relied on being able to control the expression of XerC, because this was assumed to be the *xer* recombinase responsible for the strand cleavage and exchange steps of the reaction. However, the results described in section 5.10 show that another, previously unidentified, protein is required for *cer* and *dif* recombination, and that this protein has a sequence consistent with it being a recombinase also. Gel binding analysis performed by G.May (pers.comm.) demonstrated that XprB purified from *xerC*, *xprB* cells containing pRM135/130 will specifically bind the 33 bp cross-over core site of *dif*. In addition, purified XerC and purified XprB when combined will bind both *cer* and *dif* sites and create a more retarded complex than they make when bound individually. These results cannot be interpreted as proof that both proteins are active recombinases, but they do suggest that both are essential for the recombination reactions and bind the presumptive cross-over region of *dif* (therefore probably of *cer* also). When taken together with the sequence of the proteins, these data suggest it is a definite possibility that both are involved in the strand cleavage and exchange steps.

Isolation of this new *xer* gene offers a single hypothesis that is capable of explaining the results described in *dif* and *cer* reactions in RM40, and with pRM60. The hypothesis, tested below, is that the Holliday junction accumulation is a consequence of controlling the expression of only one of the recombinases: when XerC is induced the concentrations of XerC and XprB are "imbalanced", occasionally causing the nucleoprotein intermediate to disassemble and the *cer* reaction to terminate

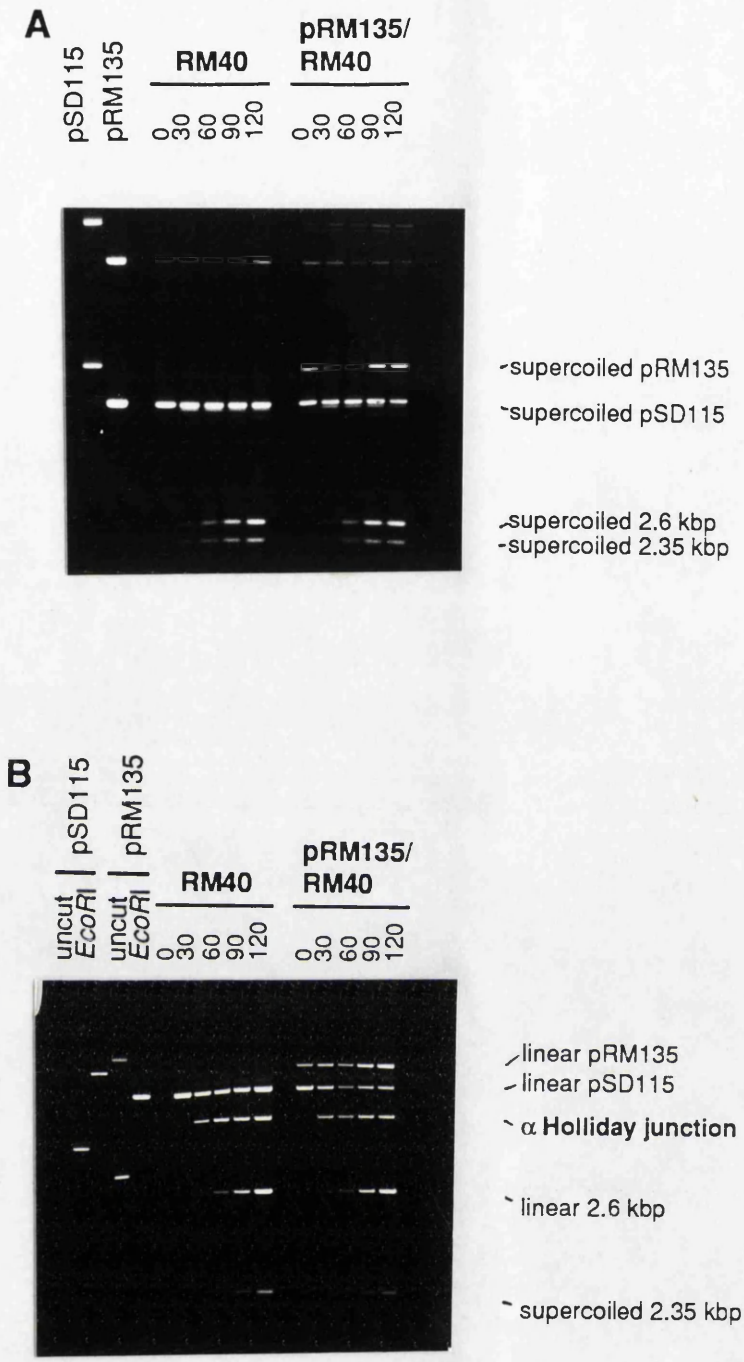


Figure 5.23. Time course of *cer* recombination *in vivo* using strain RM40 transformed with pRM135. pSD115 was transformed into either RM40 or RM40 already carrying the XprB expression vector pRM135. *in vivo* recombination assays were performed using these transformants (see text) and plasmid DNA was prepared by the boiling method at the times shown after induction of XerC expression with 2 mM IPTG. The plasmid DNA was run without restriction digestion (**A**), or after digestion with *EcoR*I (**B**), on a 1.2% agarose gel. pRM135 is linearised by digestion with *EcoR*I and pSD115-derived α structure Holliday junctions are indicated (refer to fig.4.4 for details of pSD115 and its *cer* site-specific recombination mediated resolution products).

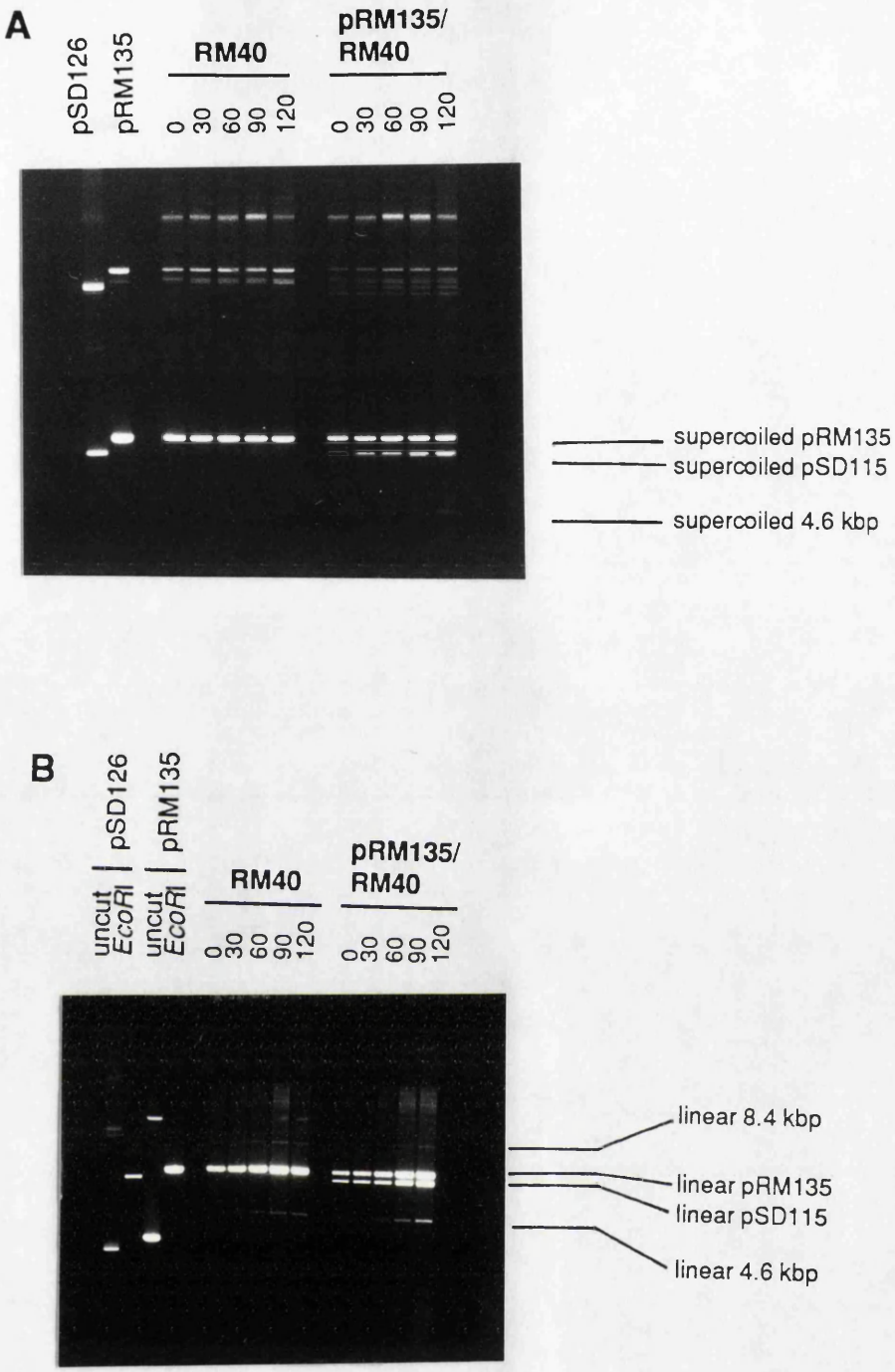


Figure 5.24. Time course of *dif* recombination *in vivo* using strain RM40 transformed with pRM135. pSD126 was transformed into either RM40 or RM40 already carrying the XprB expression vector pRM135. *in vivo* recombination assays were performed using these transformants (see text) and plasmid DNA was prepared by the boiling method at the times shown after induction of XerC expression with 2 mM IPTG. The plasmid DNA was run without restriction digestion (**A**), or after digestion with *EcoR*I (**B**), on a 1.2% agarose gel. pRM135 is linearised by digestion with *EcoR*I. Refer to figures 5.12 and 5.13 for a description of of pSD126's *dif* site-specific recombination mediated products.

after the first pair of strand exchanges. The same explanation could be true for *dif*, except that the consequences are different: the imbalance of the recombinases results in termination of the reaction before any strand exchanges can occur; alternatively the first strand exchanges are efficiently reversed in these conditions, meaning that all the Holliday junctions revert rapidly to substrate. The reason that *dif* recombination proceeds in RM40 grown on plates could either be due to differences in the level of XerC induction, a different level of XprB in the cells, or simply that the cells are allowed to grow for longer and reaction products are eventually made. The reasons why an "imbalance" of XerC and XprB may cause these effects can only be guessed at.

To test this hypothesis, RM40 was doubly transformed with pSD115 and pRM135, or with pSD126 and pRM135. In both cases the transformants were grown on L-agar plates supplemented with ampicillin, chloramphenicol, 1.0% glucose and 50 ug/ml diaminopimelic acid. 20 ml L-Broth cultures were then set up in conjunction with pSD115/RM40 and pSD126/RM40 cultures prepared from similar plates. After reaching mid-log phase 2 mM IPTG was added to the four cultures (note that this would be expected to induce the expression of XprB from pRM135 as well as XerC from the chromosome) and DNA samples were prepared at the time points shown in figures 5.23 and 5.24.

Figure 5.23 shows agarose gels of the unrestricted (5.23A) and *EcoR*I digested (5.23B) time courses of pSD115 recombination, in the presence and absence of pRM135. The amount of reaction products made in these assays did not appear to be altered by the presence of pRM135, suggesting that this had not changed the rate of the *cer* reaction in RM40. *EcoR*I restriction revealed α structures in both time courses; the presence of the XprB expression vector had not decreased or increased the amount of Holliday junction made during RM40 *cer* recombination.

Figure 5.24 shows similar gels of the pSD126 recombination assays. pRM135 is visible below supercoiled pSD126 in figure 5.234, but its presence did not seem to have caused the *dif* reaction to proceed; there was no increase in the amount of resolution products, nor any evidence of intermolecular recombination. *EcoR*I

restriction confirmed this (fig.5.24B), and also showed that the presence of pRM135 had not caused the accumulation of pSD126-derived Holliday junctions (which would be visible as α structures migrating faster than linear pSD126 and as highly retarded χ structures). These results can be interpreted in two ways:

(i) It is conceivable that these assays are giving a misleading result because pRM135 is not expressing XprB in these cells. This seems unlikely because G.May (pers.comm.) has used pRM135 to overexpress XprB (by IPTG induction) in various strains, allowing the protein to be largely purified. This shows that pRM135 is capable of expressing XprB; however, it remains possible that it is in this specific experiment that it is not being made. To confirm that the expression vector was present in RM40 in conjunction with the reporter plasmids the remains of the assay cultures (which had been stored overnight at 4 ^{o}C) were spread onto non-selective agar and then individual colonies were replica patched onto L-agar plates containing either ampicillin or chloramphenicol. 40 colonies were analysed from both assays and all were resistant to both antibiotics, indicating that the cells contained both plasmids.

It was decided to attempt to make an *xprB*$^{-}$ derivative of RM40. Recombination assays performed in this strain (with and without pRM135) would not only address the question of whether XprB is made from pRM135 in RM40, but would also determine whether any of the *cer/dif* reaction steps are possible in the absence of XprB. To do this the plasmid pRM134 was constructed, in which a gentamycin resistance gene (isolated as 1.7 kbp *Hind*III fragment from the plasmid pGM160) has been inserted into the *EcoR*V site present within the *xprB* open reading frame of pRM130 (fig.5.19). Complementation assays, performed by Mary Burke, showed that this plasmid will not complement the chromosomal mutation in DS9008. The plasmid was therefore linearised by *EcoR*I restriction and transformed into JC7623, and gentamycin resistant, ampicillin sensitive colonies were selected. These strains, in which the natural *xprB* gene has been replaced by the inactive pRM134-derived copy, are in the process of being P1 transduced into DS941 and RM40 for further analysis. For this reason it cannot be said whether XprB is genuinely being expressed in the above assays.

(ii) The assumption of these experiments is that the *cer* and *dif* reaction features in RM40 are a result of there being too much XerC after induction with IPTG, and that this would be overcome by expressing XprB in these cells. It is, however, possible that the amount of XerC expressed from the *lac* promoter is insufficient to match the wild type level. If this were the case then expression of XprB from pRM135 would not "rescue" the phenotypes since the "imbalance" causing them is a result of there being too much XprB in RM40 assays rather than in assays performed in wild type cells. This interpretation would mean that if the assays were performed by controlling the expression of XprB rather than XerC then the observed phenotypes might be altered.

Attempting to perform this experiment is, unfortunately, not simple. This is because the clones of *xprB* that are available (fig.5.19) do not have enough DNA upstream of the translational start site to insert the *lac*PO control apparatus used in the construction of RM40/50. Furthermore, the positions of the promoter and transcriptional start of *xprB* have not been defined. This is illustrated by an attempt that was made to create an XprB expression vector (pRM140) which is equivalent to pRM60 and would allow controllable expression of the protein. The plasmid construction involved cloning the 1.7 kbp *Hind*III-*Eco*RI fragment from pRM130 into *Hpa*I-*BamH*I digested pRM60, thereby replacing *xerC* with *xprB* and making its expression be driven by λ P_L. However, even when grown at 28 ^oC pRM140 still complemented an *xprB*::miniTn10 insertion in DS942 (constructed by P1 transduction), suggesting that the promoter sequences are present within the cloned *xprB* fragment. Because of these difficulties it has not been possible to address this question further.

5.12 Discussion

The genetic characterisation performed here and by Mary Burke shows that *xprB* is essential for both the *cer* and *dif* site-specific recombination reactions. This means that, in total, four proteins have been described which are required for *cer*

recombination, and two for *dif*. The possibility that other factors may be needed for either reaction cannot be excluded, however, as illustrated by the fact that *xprB* was not isolated in previous screens for *xer* proteins (Stirling, 1988; Colloms, 1990). Examination of *xprB* strains revealed that these cells have the same tendency to produce filaments that has been described for *xerC* strains, therefore confirming that the lack of pSD124 recombination in these strains reflects a mutation in cellular function and suggesting that XprB is needed for chromosome segregation (Blakely *et al*, 1991; D.Sherratt, pers.comm.).

Analysis of the putative amino acid sequence of XprB revealed that it belongs in the λ integrase class of recombinases (Lovett and Kolodner, 1991) and that it contains the four invariant residues which have been implicated in the catalytic mechanism of these proteins. Gel binding assays performed using purified XprB have shown that it specifically binds to *dif* and *cer* sites, and that it forms a distinct complex when combined with XerC (G.May, pers.comm.). Other experiments of this sort suggest that each protein binds to a different "arm" of the *dif* site, and that the two proteins bind in a co-operative manner (G.Blakely, pers.comm.). It is tempting to speculate from this evidence that the *xer* recombination utilises two recombinase proteins in its mechanism; however, the possibility cannot yet be excluded that either XerC or XprB is inactive and has a purely structural role. Site-directed mutageneses of the key amino acid residues (see above) are being performed and this should allow this question to be addressed (L.Arciszewska, pers.comm.). These mutant proteins will also make it possible to address other mechanistic questions that arise if it is found that both XerC and XprB are indeed active recombinases. For example, if they bind exclusively to separate arms of the *dif*/*cer* core sites then it will be possible to determine which protein catalyses which strand exchange steps (see Chen *et al*, 1991).

Why *xer* recombination should require two recombinases, or indeed why the reaction might use an inactive recombinase, is an intriguing question. The fact that both proteins appear to be required for both the *cer* and *dif* reactions excludes a number of possible explanations. For instance, it is not the case that one protein acts in plasmid-

borne reactions and the other on chromosomal recombination; similarly two separate recombination pathways do not exist, one catalysed by XerC and one by XprB, which are differentially active depending on the relative concentrations of the two proteins. The proteins are encoded by separate genes which are located in different regions of the *E.coli* chromosome, and both appear to be part of transcriptional operons. Why they should both be co-expressed with other proteins is not clear, and is not explained by analysing the other genes in the operons, since their functions have either not been determined (*orf235*, *orf238*, *xprA*) or are not obviously related to site-specific recombination (*dapF*, *recJ*). It is conceivable that the separate expression of XerC and XprB allows the *xer* recombination reactions to be regulated. For instance, the two proteins may be expressed at different times during the *E.coli* cell cycle, and therefore the reaction only occurs at a specific stage when their concentrations are matched. However, there is no evidence available which suggests that the expression of either operon is regulated in any manner. Furthermore, it is by no means certain that the activities of either of the two *xer* recombination reactions would need to be controlled in this sort of way.

Two other systems have been described which appear to utilise more than one integrase-like recombinase. Phase variation of type 1 fimbriae in *E.coli* is controlled by the proteins encoded by *fimB* and *fimE*, both of which display sequence homology to the λ integrases (Klemm, 1986; Dorman and Higgins, 1987). Tn554 is a transposon that has a number of properties which seem to be closely related to integrative bacteriophages, and it has been shown to contain two genes (*tnpA* and *tnpB*) whose products are essential for transposition and have sequence homology to the λ integrases (Murphy *et al*, 1985; Bastos and Murphy, 1988). Evidence exists that both FimB and FimE are able to catalyse the *fim* site-specific inversion reaction on their own (McClain *et al*, 1991). This shows that both are active recombinases, but contrasts with the situation for *xer* recombination where both XerC and XprB must be present for the reaction to occur; it also shows that the ways the recombinases are employed by the two systems are almost certainly not comparable.In the case of Tn554, it appears that both

TnpA and TnpB are essential for transposition and, like *xer*, a mutation in one cannot be complemented by overexpression of the other. However, these two proteins are of considerably different sizes (unlike XerC and XprB), and no characterisation of how they interact with the transposon ends or the transposon's chromosomal target site has been reported. These facts illustrate that the mechanisms of XerC and XprB in *xer* recombination may be quite different to the roles of TnpA and TnpB in the Tn554 transposition reaction, and comparing the two systems is probably premature at this time.

This chapter described a number of experiments designed to attempt to understand why *cer* recombination produces stable Holliday junctions during assays in RM40, and latterly why *dif* recombination appears to be inhibited in the same assays. It was considered that the Holliday junctions may constitute evidence that the *cer* reaction involved only a single pair of strand exchanges, and that this suggested that the reaction had a unique mechanism within the λ integrase family. No evidence from this chapter directly supports such a scheme. *cer* recombination was analysed in cells expressing either no active RuvC or containing an abundance of RuvC, and the reaction was unaltered, indicating that products can be made in the absence of at least one cellular Holliday junction-resolving enzyme. Experiments involving the *cer* reporter plasmids pRM91 and pRM92 suggest that the Holliday junctions made in RM40 may be free of proteins, which would be consistent with the belief that the *cer* reaction terminates after a single strand exchange. This result, however, appears to conflict with analysis of recombination in DS941. This fact alone suggests that the Holliday junctions seen in RM40 recombination assays are a consequence of this strain and do not reflect the reaction in wild type cells, and are therefore a result of changes to the reaction conditions. This is supported by the fact that *dif* recombination appears not to proceed in these assays. The attempts described (e.g. using pRM60 and minimal *dif* sites) did not explain this apparently aberrant *dif* reaction. Other experiments (changing the osmolarity of the L-broth to alter the supercoiling density of the *dif* substrate plasmids,

and allowing the reaction to proceed to stationary phase) which were not described also yielded negative results.

If the accumulation of Holliday junctions and the lack of *dif* recombination is explained by an alteration in the reaction conditions in RM40 relative to DS941 then it is likely that the cause of these observations lies in the altered expression of XerC in this strain. The assays performed using strains RM41 and RM42, and the fact that *dif* recombination does not require ArgR and PepA, suggest that these results are not due to changing the expression of XerC in comparison to its accessory factors.

Because the involvement of *xprB* in *xer* recombination was not identified before the above experiments were performed, and because XprB may act as a second recombinase, it is tempting to speculate that the Holliday junctions and lack of *dif* recombination are a consequence of the changed amounts of XerC relative to XprB in RM40. Since only one experiment has attempted to analyse this possibility (and since there are many reasons for its negative result, as discussed) this remains a feasible explanation. It is difficult, however, to understand why changing the relative amounts of the two recombinases should make the *cer* reaction terminate after the first pair of strand exchanges, and why the *dif* reaction should be strongly inhibited. This is a problem for two reasons: (i) it is not clear why the *cer* reaction would be de-stabilised while it is proceeding and (ii) why *dif* would respond to the changed protein concentrations in a different way to *cer*. There are no obvious answers to these problems, but it should be noted that all site-specific recombination reactions which have been mechanistically analysed *in vitro* and *in vivo* utilise only one recombinase, and therefore no precedents for this situation exist. Furthermore, differences between *dif* and *cer* in, for instance, their relative binding affinities for the two putative recombinases may explain their different responses to RM40 recombination assays.

Exploring this possibility further would require the ability to perform *in vivo* recombination assays using a regulatable *xprB* expression system, or alternatively, *in vitro* reconstitution of the reactions. It may, for example, be possible to make artificial *cer*/*dif* Holliday junctions and incubate them with purified XerC and/or XprB to see

how they are processed. Following the reaction profile of *cer* and *dif* reporter plasmids *in vitro* when the concentrations of the two proteins are systematically varied would also be informative.

Site-directed mutagenesis of the GGG motif in the putative *cer* overlap sequence suggested that this is not the cause of Holliday junction accumulation, but revealed other features of the *cer* reaction. Recombination between a wild type *cer* site and a GGG/AAT mutant was as efficient as recombination between two wild type sites. Although this is in apparent conflict with experiments performed in other λ integrase-like reactions (see section 5.3) it is in agreement with some other work on *xer* recombination. It has been shown that ColE1 *cer* is able to recombine with numerous other *cer*-like sites; for example, the sites from pNTP16 and ColE1 will recombine with apparently normal efficiency in DS941 assays (G. Szatmari, pers.comm.), and ColE1 *cer* and CloDF13 *parB* will also recombine, though much less efficiently (Summers, 1989). These examples, and recombination between *cer* sites from many different sources (A.Anjan, pers.comm.), might suggest that the *cer* recombination reaction is relatively unconstrained in its spacer/overlap sequence requirements. This analysis should be treated with caution, however, since to my knowledge the apparent rates of the reactions between these various sites have not been determined (as they were in this work), and recombination between a wild type *dif* site and mutant with a single base pair change in its spacer has been shown to be undetectable in DS941 assays (G.Blakely, pers.comm.).

A number of interpretations can be made of these data. It is conceivable that *cer* recombination is indeed different to other integrase-like reactions and can act on sites with different overlaps. This could be explained by the reaction resolving the Holliday junction intermediates without any branch migration. It is also possible that this reflects the different functions of *cer* and *dif*, and the *dif* reaction involves a stereotyped branch migration step. Alternatively these results may simply mean that the way that the overlap has been defined is incorrect; for example it may be a smaller number of bases or may be positioned differently within the core sites (see fig.1.9). I believe that this

question will only be resolved by a rigorous mutational analysis of the presumptive overlap sequences of both *cer* and *dif*, and by developing an *in vitro* assay.

Chapter 6

Concluding remarks

The principal objective of the work presented in this thesis was to develop an *in vivo* system that would allow controlled *xer* recombination in order that the mechanism of the site-specific recombination reaction could be analysed. In practice this meant that a strain, RM40, was developed in which the expression of *xerC* (which was believed at the outset of this work to be the gene encoding the *xer* recombinase protein) could be switched on at a set time and the reaction products subsequently analysed. Holliday junction structures were identified during analysis of *cer* reporter plasmid recombination in this strain. This suggests that the *cer* site-specific recombination reaction proceeds *via* the same basic strand exchange mechanism that has been described for other members of the λ integrase family of recombination systems. By extrapolation, it seems probable that *xer* catalysed site-specific recombination of the *dif* locus will involve a similar strand exchange reaction, although this requires confirmation not only using *dif* reporter plasmid substrates but also by analysis of *dif* recombination in its natural chromosomal location. Isolation of the *cer* derived Holliday junctions meant that it was possible to establish that the top strands of the recombining DNA duplexes had been exchanged in their production. Strand exchange bias of this sort has been described for both Cre and λ Int-catalysed recombination. It remains to be examined what features of *cer* causes this asymmetry and whether it is reflected in the recombination of *dif* and other *cer*-like sites.

It was considered that the ability to detect Holliday junction reaction intermediates in such large quantities may represent a highly novel feature of the *xer* reaction mechanism where only one pair of strand exchanges are executed, or alternatively that the intermediates are inefficiently resolved for functional reasons. These considerations can be incorporated into a model for chromosome segregation that is diagrammed in figure 6.1 (Sherratt *et al*, unpublished). Previous work on *xer*-catalysed *dif* recombination suggested that the reaction (which proceeds both intramolecularly and intermolecularly) involves rapid/frequent exchanges at the *dif* locus during breakdown of chromosomal dimers generated by homologous recombination. This was because the system might be unable to determine whether the *E.coli* chromosome is in a dimeric or

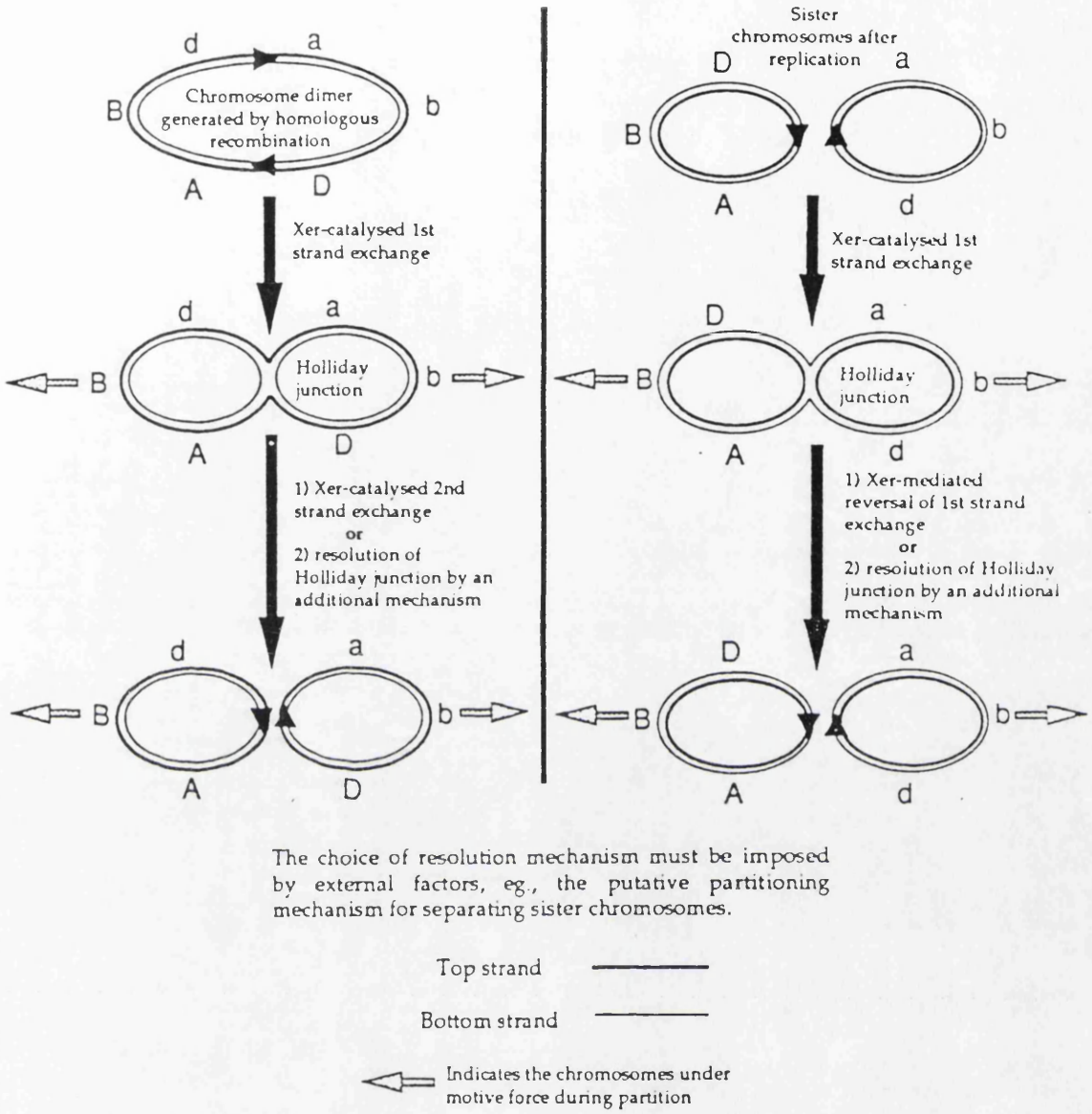


Figure 6.1. Diagrammatic representation of the role of stable, *dif*-derived Holliday junctions in the segregation of replicated bacterial chromosomes.

a monomeric configuration after completion of replication (see Blakely *et al*, 1991 and Chapter 1). The new model avoids the need for these frequent exchanges. If stable Holliday junctions are generated irrespective of the daughter chromosome configuration, then they can be resolved to yield monomers by either reversing the first pair of strand exchanges (when the replicated chromosome is monomeric) or executing the second exchanges (when the replicated chromosome is dimeric). If the choice between these alternative resolution mechanisms was dictated by external cellular factors (for instance the cellular "machinery" that separates sister chromosomes after replication) then this strategy could ensure the successful segregation of replicated chromosomes. Notice that the resolution of the *dif*-derived Holliday junctions could result from the activities of another enzyme(s) (e.g. RuvC?).

The above model has not been experimentally tested and it therefore remains speculation at this time. It rests on the observation of stable Holliday junctions during *cer* site-specific recombination in RM40. Attempts to isolate Holliday junctions using *dif* reporter plasmids in the same strain were unsuccessful, and, moreover, the recombination of these substrates appeared not to proceed in these reaction conditions. An alternative explanation for the accumulation of Holliday junctions was therefore presented after the identification of a fourth *xer* gene, *xprB*, which is essential for both *cer* and *dif* site-specific recombination (see Chapter 5). The protein encoded by *xprB* appears to be a λ integrase-like recombinase, and this suggested that the *xer* recombination reaction may involve the co-ordinated use of two recombinases - XerC and XerD (n.b. after the completion of this thesis it was agreed by S.Lovett and D.Sherratt that *xprB* should be renamed *xerD*; this nomenclature is adopted below). It should be noted that the requirement for XerD in chromosomal site-specific recombination does not necessarily mean that the model decribed above is incorrect, and it is conceivable that the involvement of the protein could be explained in the context of the recombination mechanism described in the model, perhaps as controlling the direction of resolution.

If it is assumed that both XerC and XerD are active recombinases, and that both are involved in the strand exchange reactions during recombination, then it is clear that the mechanism of *xer* recombination is more complex than was previously envisaged when only *xerC* had been identified. Gel binding experiments have suggested that XerD binds to the right "arm" of the *dif* crossover site (see fig.1.9) and XerC binds to the left (G.Blakely, pers.comm.). This may also be the case for *cer*, and consequently this represents a mechanism for the correct alignment of sites (e.g. directly repeated sites aligned so that they result in resolution recombination), although the possibility of other asymmetric features of the sites being used cannot be excluded. It is also conceivable that during *xer* recombination one pair of strand exchanges are executed by one recombinase and the second pair by the other. This is currently being investigated by examining reaction products generated, *in vitro* and *in vivo*, using combinations of wild type and site-directed mutants (in the active site residues) of XerC/D, and looking at the products generated by either XerD or XerC acting on artificial *dif* Holliday junctions *in vitro* (L.Arciszewska, pers.comm.).

During site-specific recombination it is assumed that the complete reaction requires four subunits of recombinase, i.e. one subunit bound to each half-site. In some reactions (e.g. Int or FLP) all these subunits are the same protein, but in *xer* it is possible that two subunits of XerD and two of XerC are bound. Recent experiments on λ integrase-like recombination have attempted to address the question of whether a given subunit bound to one crossover site catalyses strand cleavage in the DNA duplex to which it is bound or in the other DNA duplex. Experiments using combinations of FLP proteins mutant in their putative active site residues and wild type FLP have suggested that residues from more than one FLP subunit contribute to the enzyme's active site (the "fractional active site" hypothesis), and that cleavage at one crossover site is by a tyrosine nucleophile from a FLP subunit bound at the other site (termed *trans* cleavage; Chen *et al*, 1992). This result appears to be in conflict with analysis from λ Int reactions, which proposed that subunits cleave and become covalently linked to the crossover site to which they are bound (Kim *et al*, 1990; see Stark *et al*, 1992 for

review). It should be possible to examine this question in *xer* recombination also, using the mutants of XerC and XerD described above.

Identification of XerD may offer an explanation for the failure of previous attempts to recreate the *xer* recombination reaction *in vitro* using either purified *xer* proteins or crude cell extracts containing over-expressed XerC, ArgR and PepA. However, subsequent experiments using partially purified XerC and XerD and *dif* substrate plasmids have also not yielded *in vitro* recombination (G.May, pers.comm.). This may mean that a refinement of the reaction conditions being employed is necessary, or it may indicate that more *xer* proteins are involved in the recombination reaction.

One of the original reasons for developing the strain RM40 was to perform a topological analysis of the reaction products of *cer* site-specific recombination and compare them to *dif* and type II hybrid products. This stemmed from considerations regarding why *cer* recombination displays resolution selectivity and *dif* does not. It is possible that *cer* is poor substrate and can only recombine once a specific synaptic complex (presumably requiring ArgR and PepA) is made; in contrast, *dif* may be a good substrate and is able to form stable synaptic complexes without accessory factors and sequences. The *cer* synaptic complex could act as a "topological filter", favouring intramolecular recombination over intermolecular recombination, whilst *dif* (and typeII hybrid) recombination displays no such topological selectivity. The reasons for this are not clear, but could result from differences in the strength or conformation of XerC and XerD interactions with the *cer* and *dif* crossover sites, which could in turn result from differences the sites' sequences or spacers (J.Roberts, pers.comm.). A prediction from this hypothesis that could be tested (using RM40) is that the products of *cer* recombination, but not of *dif* recombination, have a specific topology (see Chapter 1, section 1.4). This experiment was not attempted, but it has been shown that *in vivo* analysis of this sort can be performed if the cellular enzyme DNA gyrase is inhibited, e.g. using the drug norfloxacin (see Chapter 4), and therefore it remains feasible. Preliminary experiments showed that *cer* recombination is inhibited by norfloxacin and

it may therefore be necessary to alter the XerC induction conditions in RM40 or the means of inhibition of DNA gyrase. Clearly, however, this experimental approach would not be required if the *xer* recombination reaction could be reconstituted *in vitro*.

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 R. (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.

Abremski K. E. and Hoess R. H. (1992). Evidence for a 2nd conserved arginine residue in the integrase family of recombination proteins. Protein Engineering **5**: 87-91.

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

Andrews B. J., Beatty L. G. 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., 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 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

Bastos M.and Murphy E. (1988). Transposon Tn554 encodes three products required for transposition. EMBO J **7**: 2935-2941.

Bauer C. E., Gardner J. F., Gumport R. I. and Weisberg R. A. (1989). The effect of attachment site mutations on strand exchange in bacteriophage lambda site-specific recombination. Genetics **122**: 727-735.

Becker M.M. and Wang J.C. (1984). Use of light for footprinting DNA *in vivo*. Nature. **309**: 682-687.

Beckwith J. (1987). The lactose operon. In: *Salmonella typhimurium* and *Escherichia coli*: cellular and molecular biology. Ed. Neidhardt F.C. pp 1439-1452

Blakely G., Colloms S.D., May G., Burke M. and Sherratt D.J. (1991). *Escherichia coli* XerC recombinase is required for chromosomal segregation at cell division. New Biologist **3**: 789-798

Bliska J.B. and Cozzarelli N.R. (1987). Use of site-specific recombination as a probe of DNA sructure and metabolism *in vivo*. J. Mol. Biol. **194**: 205-218.

Bliska J.B., Benjamin H.W. and Cozzarelli N.R. (1991). Mechanism of Tn3 resolvase recombination *in vivo*. J. Biol. Chem. **266**: 2041-2047.

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.

Bordo D. and Argos P.(1991). Suggestions for "safe" residue substitutions in site-directed mutagenesis. J. Mol. Biol. **217**: 721-720

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. (1990). Molecular structure of leucine aminopeptidase at 2.7 angstrom resolution. Proc. Natl. Acad. Sci. USA **87**: 6878-6882.

Burley S.K., David P.R. and Lipscomb W.N. (1991). Leucine aminopeptidase: bestatin inhibition and a model for enzyme-catalysed peptide hydrolysis. Proc. Natl. Acad. Sci. USA **88**: 6916-6920.

Burley S.K., David P.R., Sweet R.M., Taylor A. and Lipscomb W.N. (1992). Structure determination and refinement of bovine lens leucine aminopeptidase and its complex with bestatin. J. Mol. Biol. **224**: 113-140.

Chen J. W., Lee J. and Jayaram M. (1992). DNA cleavage in *trans* by the active site tyrosine during FLP recombination: switching partners before exchanging strands. Cell. **69**: 647-658.

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

Coggins L.W. (1987). Preparation of nucleic acids for electron microscopy. In: Electron microscopy in molecular biology, a practical approach; IRL Press, Oxford pp 1-29

Cohen S.N. and Hsu L. (1972). Non chromosomal antibiotic resistence in bacteria: genetic transformation of *Escherichia coli* by R factor DNA. Proc. Natl. Acad. Sci. USA **69**: 2110-2114.

Colloms S.D. (1990). Isolation and characterisation of *E.coli xerC*, a gene required for *cer* site-specific recombination. Phd thesis: University of Glasgow.

Colloms S.D., Sykora P., Szatmari G. and Sherratt D.J. (1990). Recombination at ColE1 *cer* requires the *Escherichia.coli xerC* gene product, a member of the lambda integrase family of site-specific recombinases. J. Bacteriol. **172**: 6973-6980.

Connolly B. and West S.C. (1990). Genetic recombination in *E.coli*: Holliday junctions made by RecA protein are resolved by fractionated cell-free extracts. Proc. Natl. Acad. Sci USA. **87**: 8476-8480.

Connolly B., Parsons C.A., Benson F.A., Dunderdale H.J., Sharples G.J., Lloyd R.G. and West S.C. (1991). Resolution of Holliday junctions in vitro requires the *Escherichia coli ruvC* gene product. Proc. Natl. Acad. Sci USA. **88**: 6063-6067.

Cowart M., Benkovic S. J. and Nash H. A. (1991). Behavior of a cross-linked attachment site - testing the role of branch migration in site-specific recombination. J. Mol. Biol. **220**: 621-629.

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.

Craig N.L. and Kleckner N.(1987). Transposition and site-specific recombination. In: *Salmonella typhimurium* and *Escherichia coli*: cellular and molecular biology. Ed. Neidhardt F.C. pp 1054-1070

Craig N. L. (1988). The mechanism of conservative site-specific recombination. Ann Rev Genet **22**: 77-105.

Derbyshire K.M. and Grindley N.D.F. (1987). Replicative and conservative transposition in bacteria. Cell. **47**: 325-327.

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., Higgins C.F.. (1987). Fimbrial phase variation in *Escherichia coli*: dependance 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

Dunderdale H.J., Benson F.E, Parsons C.A., Sharples G.J., Lloyd R.G. and West S.C. (1991). Formation and resolution of recombination intermediates by *E.coli* RecA and RuvC proteins. Nature **354**: 506-510.

Echols H. and Green L. (1979). Some properties of site-specific and general recombination inferred from Int-initiated exchanges by bacteriophage lambda. Genetics. **93**: 297-307.

Enquist L. W., Nash H. and Weisberg R. A. (1979). Strand exchange in site-specific recombination. Proc. Natl. Acad. Sci USA. **76**: 1363-1367.

Evans B. R., Chen J. W., Parsons R. L., Bauer T. K., Teplow D. B. and Jayaram M. (1990). Identification of the active-site tyrosine of FLP recombinase: possible relevance of its location to the mechanism of recombination. J. Biol. Chem. **265**: 18504-18510.

Flynn H., Burke M., Stirling C.J. and Sherratt D.J. (1989). Use of gene replacement to construct *Escherichia coli* strains carrying mutations in two genes required for stability of multicopy plasmids. J. Bacteriol. **171**: 2241-2243.

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

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 *EcoR*I endonuclease and methylase. J. Biol. Chem. **256**: 2143-2153

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). Determination of DNA sequences essential for FLP-mediated recombination by a novel method. J. Biol. Chem. **260**: 2320-2327.

Gronostajski R. M. and Sadowski P. D. (1985b). 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. (1985c). The Flp protein of the 2-micron plasmid of yeast - intermolecular and intramolecular reactions. J. Biol. Chem. **260**: 2328-2335.

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

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. 857-867

Heery D.M., Gannon F. and Powell R. (1990). A simple method for sub-cloning DNA fragments from gel slices. TIG **6**: 173.

Hoess R. H. and Abremski K. (1984). Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. Proc. Natl. Acad. Sci. USA **81**: 1026-1029.

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 characterization of intermediates in site-specific recombination. Proc. Natl. Acad. Sci. USA **84**: 6840-6844.

Holliday R. (1964). A mechaism for gene conversion in fungi. Genet. Res (Cambridge). **5**: 282-304.

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

Iwasaki H., Takahagi M., Shiba T., Nakatu A. and Shinagawa H. (1991). *Escherichia coli* RuvC is an endonuclease that resolves the Holliady structure. EMBO J. **10**: 4381-4389.

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.

Jensch K., Kosak H., Seeman N.C. and Kemper B. (1989). Cruciform cutting endonucleases from *Saccharomyces cerevisiae* and phage T4 show conserved reactions with branched DNA. EMBO J. **8**: 4325-4334.

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

Kim S.H., de Vargas L.M, Nunes-Duby S.E. and Landy A. (1990). Mapping of a higher-order protein-DNA complex:two kinds of long range interaction in λ *attL*. Cell **63**: 773-781

Klippel A., Mertens G., Patschinsky T., and Kahmann R. (1988). 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.

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. (1987). Homology-dependent interactions in phage-lambda site-specific recombination. Nature **329**: 346-348.

Kitts P. A. and Nash H. A. (1988a). An intermediate in the phage-lambda site-specific recombination reaction is revealed by phosphorothioate substitution in DNA. Nucl. Acids Res. **16**: 6839-6856.

Kitts P. A. and Nash H. A. (1988b). Bacteriophage-lambda 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 1 fimbriae in *Escherichia coli*. EMBO J. **5**: 1389-1393.

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

Kuempel P.A., Henson J.M., Dircks L., Tecklenburg M. and Lim D.F. (1991). *dif*, a recA-independant recombination site in the terminus region of the chromosome of *Escherichia coli*. New Biologist **3**: 799-949

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. Ann. 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.

Leatherbarrow R.J. and Ferscht A.R. (1986). Protein engineering. Protein engineering **1**:7-16

Leong J. M., Nunes-Duby S., Oser A., Lesser C. F., Youderian P., Susskind M. M. and Landy A. (1984). Site-specific recombination systems of phages-phi-80 and P22: binding-sites of integration host factor and recombination-induced mutations. Cold Spring Harbor Symp. Quant. Biol. **49**: 707-714.

Lilley D.M.and Kemper B.. (1984). Cruciform-resolvase interactions in supercoiled DNA. Cell **36**: 413-422.

Lim D., Oppenheim J.D., Eckhart T. and Mass W.K. (1987). Nucleotide sequence of the *argR* gene of *Escherichia coli* K12 and isolation of its product, the arginie repressor. Proc. Natl. Acad Sci USA. **84**: 6697-6701.

Lloyd R.G. (1991). Conjugational recombination in resolvase-deficient *ruvC* mutants of *Escherichia coli* K12 depends on *recG*. J. Bacteriol. **173**: 5414-5418.

Lovett S.T. and Clark A.J. (1985). Cloning of the *Escherichia coli recJ* chromosomal region and identification of its encoded proteins. J. Bacteriol. **162**: 280-285.

Lovett S.T. and Kolodner R.D. (1991). Nucleotide sequence of the *Escherichia coli recJ* chromosomal region and construction of RecJ-overexpression plasmids. J. Bacteriol. **173**: 353-364.

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.

McClain M., Blomfield I.C. and Eisenstein B.I. (1991). Roles of *fimB* and *fimE* in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. J. Bacteriol. **173**: 5308-5314.

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.

Meyer-Leon L., Huang L. C., Umlauf S. W., Cox M. M. and Inman R. B. (1988). Holliday intermediates and reaction by-products in Flp protein-promoted site-specific recombination. Mol. Cell. Biol. **8**: 3784-3796.

Meyer-Leon L., Inman R. B. and Cox M. M. (1990). Characterization of Holliday structures in Flp protein-promoted site-specific recombination. Mol. Cell. Biol. **10**: 235-242.

Miller J. (1972). Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

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 K., Kemper B. and Weisberg J.R. (1982). T4 endonuclease VII cleaves Holliday junctions. Cell **29**: 357-365.

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

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., Howyler L. and 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.

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.

Pargellis C. A., Nunes-Duby S. E., Devargas L. M. and 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.

Parsons R. L., Prasad P. V., Harshey R. M. and Jayaram M. (1988). Step-arrest mutants of Flp recombinase: implications for the catalytic mechanism of DNA recombination. Mol. Cell. Biol. **8**: 3303-3310.

Parsons C.A. and West S.C. (1990). Specificity of binding to four-way junctions in DNA by bacteriophage T7 endonuclease I. Nucl. Acid Res. **18**: 4337-4384.

Parsons R. L., Evans B. R., Zheng L. and Jayaram M. (1990). Functional-analysis of arg-308 mutants of Flp recombinase: possible role of arg-308 in coupling substrate binding to catalysis. J. Biol. Chem. **265**: 4527-4533.

Pettijohn D.E. (1988). Histone-like proteins and bacterial chromosome structure. J. Biol. Chem. **263**: 12793-12796.

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.

Reznikoff W.S. and McClure W.R. (1986). *Escherichia coli* promoters. In: Maximising gene expression. Eds Reznikoff W. and Gold L.

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

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

Sanderson M.R., Freemont P.S., Rice P.A., Goldman A., Hatfull G.F., Grindley N.D.F. and Steitz T.A. (1990). The crystal structure of the catalytic domain of the site-specific recombination enzyme γδ resolvase at 2.7A resolution. Cell. **63**:1323-1329

Schmid M.B. (1988). Structure and function of the bacterial chromosome. TIBS. **13**: 131-135.

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.

Sharples G.J. and Lloyd R.G. (1991). Resolution of Holliday junctions in *Escherichia coli*: identification of the *ruvC* gene product as a 19-kilodalton protein. J. Bacteriol. **173**: 7711-7715.

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

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

Stark W.M., Grindley N.D.F., Hatfull G.F. and Boocock M.R. (1991). Resolvase-catalysed reactions between *res* sites differing in the central dinucleotide of subsite I. EMBO J. **10**: 3541-3548

Stark W.M., Boocock M.R., and Sherratt D.J. (1992). Catalysis by site-specific recombinases. TIG. **8**:

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

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

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 ColK and ColE1: 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.

Tabor S. and Richardson C.C. (1985). A bacteriophage T7 polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA. **82**: 1074-1078.

Takahagi M., Iwasaki H., Nakata A. and Shinegawa H. (1991). Molecular analysis of the Escheichia coli ruvC gene, which encodes a Holliday junction-specific endonuclease. J.Bacteriol. **173**: 5747-5753.

Taylor A.F. (1992). Movement and resolution of Holliday junctions by enzymes from *E.coli.* Cell. **69**: 1063-1065.

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

Van Loon-Klassen L., Cuypers H.T. and Bloemendal H. (1979). Limited tryptic digestion of bovine eye lens leucine aminopeptidase. FEBS Letters. **107**: 366-370

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

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

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.

Winans S.C., Elledge S.J., Krueger J.H. and Walker G.C. (1985). Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli.* J. Bacteriol. **161**: 1219-1221.

Worcel A. and Burgi E. (1971). On the structure of the folded chromosome of *Escherichia coli.* J. Mol. Biol. **71**: 127-147.

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.

GLASGOW UNIVERSITY LIBRARY