CLONING AND CHARACTERISATION OF *E. COLI xer* GENES

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

by

Colin John Stirling

Institute of Genetics
Church Street
Glasgow

October 1987

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

CONTENTS **PAGE**

Abbreviations

(i) Chemicals

Ac - acetate

APS - ammonium persulpate

ATP - adenosine triphosphate

BSA - bovine serum albumin

DNA - deoxyribonucleic acid

d(NTP) - 2'-deoxy (nucleotide)

DTT - dithiothreitol

EDTA - ethylene diamine tetra-acetic acid (disodium salt)

EtBr - ethidium bromide

EtOH - ethanol

FSB - final sample buffer

SCFSB - single colony final sample buffer

IPTG - isopropyl B-D thiogalactopyranoside

RNA - ribonucleic acid

RNase - ribonuclease A

SDS - sodium dodecyl sulphate

TEMED - NNN' N' tetramethyl ethylenediamine

Tris - tris (hydroxymethyl) amino ethane

X-gal - 5-bromo, 4-chloro, 3-indolyl, B-D galactoside

(ii) Antibiotics

Ap - ampicillin

Cm - chloramphenicol

Km - kanamycin

Nal - naladixic acid

Rif - rifampicin

Str - streptomycin

Tet - tetracycline

(iii) Phenotype

X^r - resistance to X

X^s - sensitivity to X

oriV - replication origin

1^m - monomer

2^m - dimer

(iv) Measurements

mA - milliamps (10^{-3} A)

bp - base pair

kb - kilobase pair (10^3 bp)

kDa - kilodalton (10^3 dalton)

^oC - degrees centigrade

Ci - curie

uCi - microcurie (10^{-6} curie)

g - centrifugal force equal to gravitational acceleration

g - gramme

mg - milligramme (10^{-3} g)

ug - microgramme (10^{-6} g)

l - litre

ml - millilitre

ul - microlitre

M - Molar

mM - millimolar

uM - micromolar

m - metre

cm - centimetre (10^{-2}m)

mm - millimetre (10^{-3}m)

mins - minutes

pH - acidity ($-\log_{10}$ [molar concentration H^+ ions])

sec - seconds

v - volts

w - watts

(v) Miscellaneous

ln - natural log

CAT - chloramphenicol acetyl transferase

UV - ultra violet light

Tn - transposon

pfu - plaque forming units

moi - multiplicity of infection

aa - amino acids

rt - room temperature

% - percentage

LMP - low melting point

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dave Sherratt for his inspiring confidence and enthusiasm. Special thanks to Gill Stewart for her help and advice during our time together as the "Xer group". Also to Chris Boyd for the provision of numerous plasmids, ideas, and computer software. Thanks also to Derek Blake, Vroni Knott, and John Fothergill for their valuable contributions to this project. Many thanks to Mary Burke for maintaining the "Sixth floor machine", and to everyone in our group for tolerating my occasional outbursts at inanimate objects determined to obstruct me. Most of all I must thank Fiona for her love, support and tolerance throughout.

SUMMARY

The heritable stability of the natural multicopy plasmid, ColE1, can be greatly reduced by the formation of plasmid multimers *via* homologous recombination. Wild-type ColE1 carries a 233bp site, *cer*, which when present in direct repeat (as is the case in a ColE1 dimer), acts as a substrate for a site-specific recombinase which efficiently resolves plasmid multimers to monomers. This monomerisation function is strongly correlated with an increase in plasmid stability. Current evidence suggests that ColE1 segregates randomly, and that *cer*-specific monomerisation serves to enhance stability by maximising the number of independently segregating plasmid copies (Sherratt *et al*, 1984; Summers & Sherratt, 1984). The 233bp *cer* site represents the only ColE1 sequence necessary for *cer*-specific recombination, however the available evidence makes it appear highly improbable that this locus also encodes the *cer*-specific recombinase; on this assumption the recombinase would require to be host encoded. A procedure was devised for the selection of *E. coli* K12 mutants defective in *cer*-specific recombination (Xer^-). Using this procedure a total of 11 independent *xer* mutants have been isolated from Tn5-mutagenised populations; of the 11 *xer* lesions, six are genetically linked to an inserted copy of Tn5, whilst two appear to be unlinked spontaneous mutations, the remaining 3 mutants are uncharacterised.

This collection of mutants has been employed to clone the wild-type *xer* genes by genetic complementation. This has revealed that the 11 mutants fall into two complementation groups; two mutants are complemented by the cloned *xerA* gene, whilst the remaining nine are complemented by a clone of the *xerB* gene. High resolution mapping of the *xerA* and *xerB* loci indicates that they are quite discrete, mapping to 70.5 and 96.5 minutes respectively on the *E. coli* K12 linkage map.

The *xerA* gene was delineated to within a 920bp *Sph*I/*Acc*I fragment, and the DNA sequence of this fragment determined. Sequence analyses revealed the presence of a 156 codon open reading frame (ORF) which would encode a polypeptide with a predicted molecular weight equivalent to 17.2KDa; the expression of this ORF has been confirmed

insertion of Tn5 in the mutant allele xerA3 which, coupled with a deletion analysis of the locus, provides convincing evidence that this 156 codon ORF must represent the xerA gene.

The xerB locus (a 1.92 Kb HindIII fragment) has also been sequenced and the xerB gene identified as a 503 codon open reading frame which encodes a polypeptide with a predicted molecular weight equal to 55.3KDa. Interestingly there are several features of the DNA sequence flanking the xerB gene which indicate that it may be transcribed as part of a polycistronic mRNA, with at least one other gene immediately downstream of xerB; whether or not this downstream gene encodes an Xer-related function remains to be determined. The XerB protein has also been visualised in minicell expression studies, and has been over-expressed in whole cells to a level of 5% of total cell protein. The XerB protein has been purified to homogeneity and its identity confirmed by N-terminal peptide sequencing. This purified material has been used to raise anti-XerB antisera which might now be employed to investigate several features of XerB including its expression, cellular location(s), and function(s).

The sequences of the xerA and xerB genes (and their products), coupled to the genetic map positions of both loci, indicate that we have identified two entirely novel E. coli functions involved in site-specific recombination at the ColE1 cer locus. The precise role(s) the two proteins play in cer-specific recombination, and in the biology of E. coli remain to be determined.

CHAPTER 1

INTRODUCTION

1.1 GENERAL INTRODUCTION

Bacterial plasmids may be defined as extrachromosomal covalently closed circles of duplex DNA, capable of autonomous replication, and stable maintenance within their host. However, beyond this basic similarity they are a very heterogenous group. As a class, they vary in size from less than 5Kb to more than 100Kb, and in copy number from 1 or 2 per chromosome equivalent to over 100. They also vary enormously at the genetic level, both in terms of the number of genes they carry and the types of functions they encode. Many plasmid encoded functions are considered to be non-essential for either plasmid maintenance or host viability, such esoteric functions may however confer a selective advantage to the host under bizarre environmental conditions; for example the ability to utilise toluene and/or xylene as carbon and energy substrates as conferred by the TOL plasmid of Pseudomonas putida, or the resistance to various antibiotics conferred by the R-plasmids (Broda,1979). On the other hand, a number of plasmid encoded functions may be regarded as having a direct affect on the maintenance of a plasmid within a population even in the absence of any obvious environmental pressure. For example, the conjugal transfer mechanisms of many large plasmids, eg. F, R100 etc., permits horizontal spread of the replicon into plasmid free cells. Another example being plasmid based bacteriocin production, and reciprocal bacteriocin immunity, which enables plasmid containing cells to select against closely related plasmid free cells by the release of a bacteriocin to the environment.

Despite their enormous diversity all plasmids share certain functional requirements, the most obvious of which being that they must possess a site which is recognised, either by plasmid or host specified factors, as an origin for initiation of DNA replication. It is this property of natural plasmids that has been exploited in the construction of many of the cloning vectors commonly used today, for example pBR322 (Sutcliffe, 1979) and pUC8 (Vieira and Messing, 1982) which are both recombinant plasmids whose origin of replication is derived from the natural plasmid pMB1 (a close relative of ColE1). A general trend in the construction of many plasmid-derived cloning vectors has been to delete, either specifically or at random, as much DNA sequence as possible which is non-essential for replication *per se*. A striking consequence of this approach is the inherent instability of most cloning vectors; plasmid stability being a

measure of the frequency with which plasmid-free segregants are generated from a population of plasmid-containing cells when growing under non-selective conditions. In marked contrast natural plasmids are extremely stable under the same growth conditions. Thus, by inference, most natural plasmids appear to encode functions which promote their stable inheritance by all daughter cells at division. Such stability functions are of considerable biological interest since they may aid our understanding of fundamental plasmid biology, and in turn may pave the way for a new generation of stable plasmid cloning vectors which could potentially yield significant economic (and environmental) benefits, especially in large scale industrial fermentations.

1.2 STABLE PLASMID MAINTENANCE

Whilst I have previously described "plasmid stability" as a measure of the frequency with which plasmid free segregants arise, it is instructive to differentiate this measurable parameter from the concept of "stable maintenance" of a natural plasmid such as ColE1, F, or pSC101 etc.. The former requires only that every cell carries one or more plasmids, whilst the latter asserts that a plasmid is maintained at a relatively constant copy number (N per chromosome-equivalent) throughout the population, the number maintained being dependent on the plasmid concerned, for example for F N=1, and for pSC101 N=5 (Timmis, 1981) whilst for ColE1 N=15 (Durkacz and Sherratt, 1973). Stable plasmid maintenance has two key components;

1. Copy number control in phase with cell growth and division.

2. Segregation of plasmid copies between daughter cells.

1.3 COPY NUMBER CONTROL

As previously intimated, a given plasmid species tends to be maintained at a characteristic mean copy number within a population. In order to maintain this (presumed) "Optimal" copy number it appears that plasmids have evolved complex replication control mechanisms capable of monitoring copy-number and reacting to correct any deviations from the norm. In all systems which have been extensively characterised the control of replication is at the level of inititiation and involves a negative feedback loop, often employing a

number of negative regulators. Several classes of negative regulators have been identified, these include;

1) Small RNA molecules e.g. RNA1 (ColE1; Tomizawa et al, 1981), or CopA (R1; Stougaard et al, 1981).
2) Small polypeptides e.g. Rom (ColE1; Twigg & Sherratt, 1980; Tomizawa & Som, 1984), or CopB (R1; Molin et al, 1981).
3) Accessory binding sites which apparently titrate-out replication initiation factors e.g. the direct repeats found near the replication origins of F and P1 (Tolun & Helinski, 1981) and ; Chattaraj et al, 1984).

A negative-feedback circuit provides a means whereby, in principle, the plasmid copy number may be monitored via a simple "Inhibitor dilution" model. Briefly this model asserts that the constitutive expression of a plasmid-borne repressor of replication provides a measure of the plasmid copy concentration (via gene dosage); too few copies results in a low inhibitor concentration which permits replication, whereas too many copies produce a level of inhibitor which prevents further initiation of replication. Perhaps the best studied example of such a control circuit is that of the ColE1 family of small multicopy plasmids.

1.4 REPLICATION CONTROL IN ColE1

Replication of ColE1 (6646bp) is unidirectional from a single replication origin. Replication from this origin does not require any plasmid encoded proteins; initiation is dependent on an RNA primer synthesised by host RNA polymerase, whilst DNA synthesis per se is initiated by host DNA polymerase I. The process also requires a number of other host functions including DNA polymerase III, plus the products of the dnaB,-C, -G, and -Z genes (for review see Scott, 1984).

The initiation process itself is dependent on the synthesis of pre-primer RNA. This transcript is initiated at a point 555 bp upstream from the origin of DNA synthesis, this transcript may continue through the replication origin resulting in several sizes of transcript (Scott, 1984). However, some of the nascent pre-primer transcripts hybridise to the DNA template near the origin, the resultant DNA-RNA hybrid then serves as a substrate for RNAase H which

cleaves the pre-primer RNA to generate Primer-RNA (RNAII), DNA polymerase I then initiates DNA synthesis on the free 3` hydroxyl group of RNAII (Itoh and Tomizawa, 1980; Masukata and Tomizawa, 1984).

The formation of the DNA-RNA hybrid near the origin is critical for RNAase H processing, and thus for the initiation of DNA synthesis. However, this hybridisation is under the negative control of a 108 nucleotide ColE1 encoded RNA, termed RNA1. RNA1 is transcribed from the opposite strand to primer-precursor RNA, but from wholly within the primer-precursor sequence (see Fig. 1.1), therefore RNA1 is complementary to a 108 nucleotide portion of primer-precursor RNA , thus the two RNA species are capable of inter-molecular base pairing along the entire length of RNA1. The basis of RNA1 inhibition of DNA replication appears to be that the binding of RNA1 to the nascent primer-precursor RNA in some way alters the secondary structure of the nascent transcript such that it can no longer form a stable DNA-RNA hybrid near the origin, thus RNAase H processing is prevented (Lacatena and Cesareni, 1983; Tomizawa and Itoh, 1982).

1.4.2 THE INTERACTION BETWEEN RNA1 AND PRIMER-PRECURSOR RNA

The RNA1 sequence, and by necessity the complementary 108 nucleotide portion of primer-precursor RNA, are capable of extensive *intra* molecular base pairing to form a tRNA-like hairpin-loop structure (shown diagrammatically in Fig. 1.1). The existence of this structure has been demonstrated *in vitro* by mapping the sites at which various double-, and single-, strand specific ribonucleases cut purified RNA1 (Tomizawa, 1984). Furthermore, the significance of this tertiary structure *in vivo* has been demonstrated using a series of point mutations which destabilise the stem structures and result in defective RNA1, incapable of inhibiting RNAase H processing (Lacatena and Cesareni, 1983). In addition, such mutational analyses have demonstrated that as few as seven nucleotides, all present in the loops of the RNA1 secondary structure, may determine the specificity of the RNA1/Primer-precursor RNA interaction. A number of independent point mutations in the loop structures of RNA1 result in an RNA1 capable of interacting effectively with it s homologous primer-precursor RNA, but not capable of inhibiting processing of a wild-type primer-precursor, despite 107 out of 108 nucleotides being complementary. Such mutant RNA1 sequences result in ColE1 derivatives which are compatible with one another. In addition to RNA1, a second

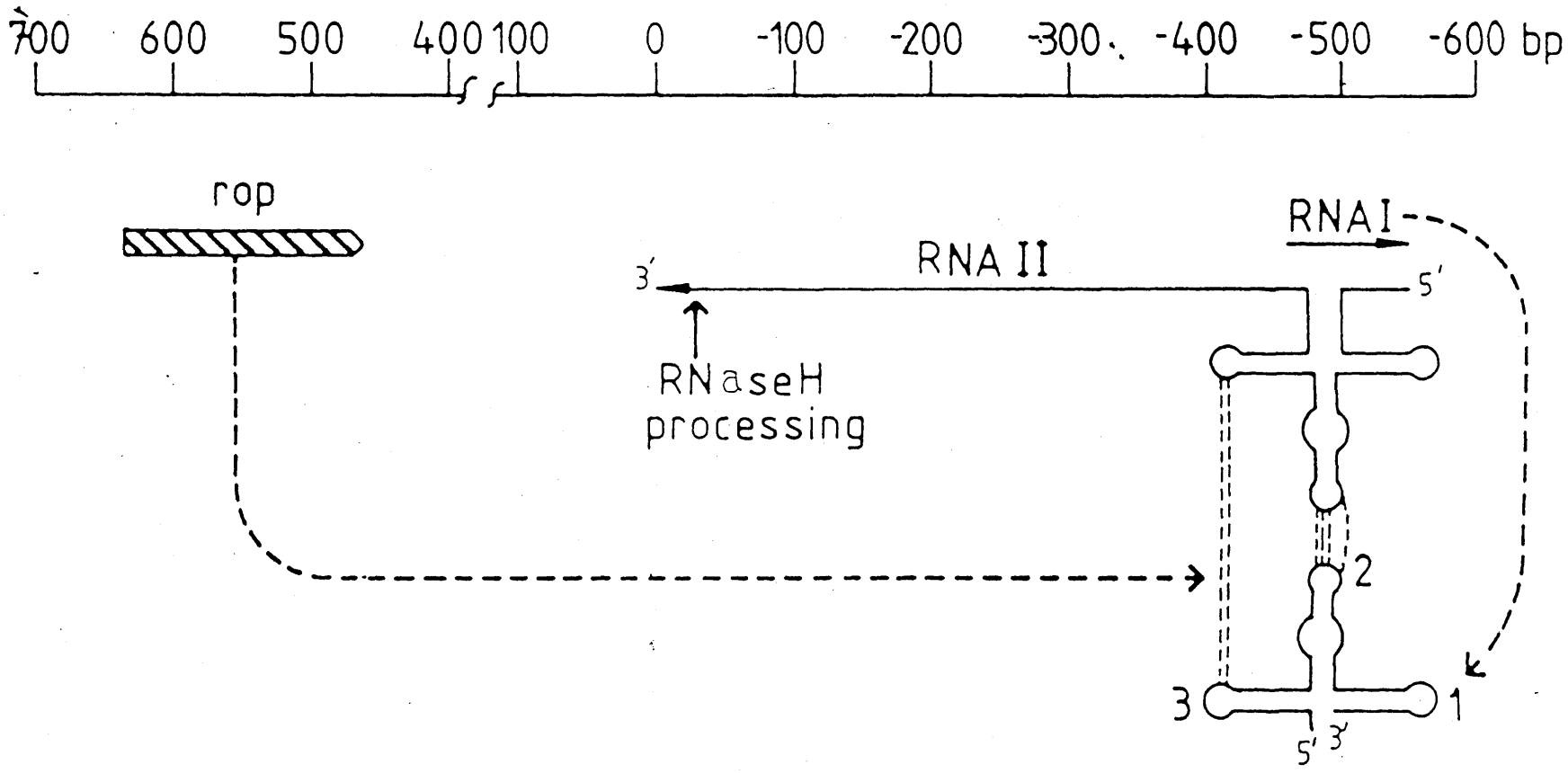


Figure 1.1 The replication control region of ColE1. Interaction between the complementary RNA I and RNA II molecules prevents primer processing by RNaseH to form the active replication primer. The Rom protein catalyses the interaction between RNA I and RNA II, thereby enhancing the negative control mechanism (Scott, 1984).

negative regulator of copy number is encoded by ColE1 type plasmids; the 63 amino-acid RNA One Modulator (Rom) protein (previously designated "Rop"). Deletion of the pMB1 rom gene results in a 5-7 fold increase in plasmid copy number (Twigg and Sherratt, 1980). ROM catalyses the initiation of stable binding between RNA1 and primer-precursor (Fig 1.1), but has no influence on the subsequent rate of propagation of ribonucleotide base-pairing (Tomizawa and Som, 1984).

1.4 PLASMID SEGREGATION

A dichotomy is apparent between low copy-number plasmids and multicopy plasmids with respect to the mode of plasmid segregation employed, therefore these two classes will be considered seperately.

1.4.2 SEGREGATION OF LOW COPY NUMBER PLASMIDS

The remarkable stability of plasmids with copy numbers as low as 1 or 2 per chromosome suggests that such plasmids may be actively partitioned between daughter cells at division (Jacob, 1963). Indeed there now exists evidence for such "Active partition" in a number of plasmids including NR1 (Mika et al, 1980), R1 (Nordstrom et al, 1980), F (Ogura and Hiraga, 1983), and pSC101 (Meacock and Cohen, 1980). In the case of pSC101 a plasmid locus designated par (partition) has been found to complement partition deficient plasmid mutants when present in cis, but not in trans (Meacock and Cohen, 1980); the locus is also capable of increasing the stability of unrelated plasmids derived from p15A and R1 replicons (Meacock and Cohen, 1980; Nordstrom et al, 1981). The requirement for a cis locus is a common feature of all active partitioning systems reported. The par sequence is presumed to interact with some secondary component which is itself actively partitioned; the identity of this structure is uncertain, however candidates include the chromosomal body on the basis that several plasmids have been observed to co-sediment with it (Worcel and Burgi, 1974; Kline and Miller, 1975; Kline et al, 1976). Alternatively the par site may interact with a membrane attachment site, a hypothesis supported by the observation that the par region of pSC101 associates with the outer membrane fraction of Escherichia coli, as does that of the plasmid R1 (Gustafsson et al, 1983). There is as yet no evidence that any trans acting plasmid encoded products are required for the active partitioning of pSC101 (Tucker et al, 1984). However the situation is more complex in the E. coli sex factor, F. This plasmid

has also been shown to carry a specific site (sopC), required in cis for active partitioning, however in this case the presence of the sopC locus is not in itself sufficient to confer a par^+ phenotype. Ogura and Hiraga (1983) have demonstrated that a further two plasmid loci are required, sopA and sopB, both of which specify trans acting elements. The products of the sopA and sopB loci are believed to be polypeptides of molecular weights equal to 41-44KDa and 36-37KDa respectively (Ogura and Hiraga,1983; Wehlmann and Eichenlaub, 1980; Komai et al, 1982). Ogura and Hiraga have reported that the 37KDa protein encoded by the sopB locus binds specifically to the sopC DNA sequence together with two host proteins of 75KDa and 33KDa respectively. On this basis they have proposed a putative model for the active partitioning of F plasmid (Ogura and Hiraga, 1983). However the identification of the 75KDa and 33KDa host proteins, with respect to their cellular role and location, may prove to be a major step forward in the understanding of active partitioning mechanisms.

The active partitioning of a unit copy number plasmid, such as F, poses no conceptual difficulties, it simply requires one round of plasmid replication after which the products are partitioned between the daughter cells. However many actively partitioned plasmids have copy numbers greater than 2 at cell division (e.g. pSC101), and in these cases the net effect of the active mechanism is uncertain. Novick et al (1975) proposed the Equipartition model in which every individual plasmid within the dividing cell is a substrate for the partition apparatus, and that the mechanism itself ensures an equal distribution between daughter cells; this hypothesis being based on the curing kinetics of a plasmid mutant temperature-sensitive for replication, under restrictive conditions of temperature. However, using a similar experimental principle, but modified detection techniques, Hashimotoh-Gotoh and Ishii (1982) obtained curing kinetics which led them to propose the Single Site Inheritance model for partition. In this model only two members of the cell's plasmid complement are actively segregated, the remainder being free to segregate randomly by diffusion. These alternative hypotheses are of some considerable consequence with respect to our understanding of the process(es) involved in active partitioning, however the experimental data available to date cannot unequivocally discriminate between them (see Nordstrom, 1984).

The process of active partitioning, as we currently understand it, cannot ensure plasmid stability unless plasmid replication can keep pace with cell division. Whilst unit copy number plasmids would be most vulnerable to such instability, current evidence suggests that at least one such plasmid has evolved counter measures of a most dramatic nature. The F plasmid encodes two polypeptides which act antagonistically to prevent host cell division until after plasmid replication has taken place. The first polypeptide (the LetD protein), acts as an inhibitor of cell division, whilst the second (LetA) is synthesised after plasmid replication and acts as a suppressor of LetD, thus triggering cell division (Miki et al, 1984a; Miki et al, 1984b). It has been suggested that the plasmid R1 might exert a similar control on cell division however this is as yet unsubstantiated (Gustafsson et al, 1983).

1.4.3 PARTITION OF MULTICOPY PLASMIDS

There have been no substantiated reports of active partitioning systems in high copy number plasmids, and the available data are consistent with the "Random segregation" model proposed by Durcacz and Sherratt (1973). This model proposes that the plasmid population within a dividing cell segregates in a purely random fashion as a consequence of free diffusion of all plasmid copies throughout the cytoplasm. Clearly the probability of generating a plasmid free segregant (P_o) can then be calculated using the binomial distribution and is equal to 2^{1-s} (where s is equal to the number of independently segregating units present at cell division). If a single plasmid molecule represents the segregating unit, then a copy number of 15 per chromosome equivalent should ensure stable inheritance since at cell division a minimum of 30 plasmid copies should be present per cell, resulting in $P_o = 2^{-29}$ (i.e. 1.9×10^{-9} plasmid-free segregants per cell division). The obvious prediction of this model is that as copy number increases then there should be a concomitant increase in stability. Paradoxically, many plasmid cloning vectors have copy numbers far in excess of that of ColE1 and yet are grossly unstable (Timmis, 1981; Summers & Sherratt, 1984). This anomaly is accounted for, at least in part, by the formation of plasmid multimers *in vivo via* homologous recombination. Evidence suggests that plasmid copy number is controlled *via* an "origin-counting" mechanism which does not differentiate between say 20 monomers and 10 dimers (Summers &

Sherratt, 1984; Jones, 1985) . Therefore the formation of plasmid multimers does not reduce the number of plasmid copies (replication origins) present, but it does reduce the number of independently segregating units. Summers & Sherratt (1984), have demonstrated a direct correlation between multimerisation and instability for the cloning vector pACYC184.

In a recombination-proficient strain plasmid cloning vectors are extensively multimerised to produce dimers, trimers, tetramers etc.. On the other hand ColE1 remains predominantly monomeric even in hyper-recombinogenic strains. It has been shown that ColE1 carries a site, designated cer (ColE1 Resolution), at which a site-specific recombinase acts to promote the efficient resolution of plasmid multimers to monomers, thus maximising the number of independently segregating plasmid copies. This site has been shown to effectively stabilise heterologous cloning vectors (such as pACYC184) when present in cis; such stabilisation being strongly correlated with an increase in the proportion of monomeric plasmid forms (Summers & Sherratt, 1984). The functional cer locus has been localised to within 279bp HpaII/TaqI fragment (ColE1 co-ordinates 3687-3966bp) which represents the sole ColE1 sequence necessary for cer-specific recombination (Summers et al, 1985). Similar sites to cer have been identified in other closely related multicopy plasmids including ClODF13 (crl; Haakaart et al, 1984), ColK (ckr; Summers et al, 1985), pMB1 (Sherratt et al, 1986), and ColA (Morlon et al, manuscript in preparation); all of which share considerable sequence similarity to cer (Fig. 1.2). This sequence similarity implies that these five multicopy plasmids are likely to share very similar multimer-resolution mechanisms. Analagous resolution systems have also been identified in a number of low copy-number plasmids including P1 (loxP/Cre; Austin et al, 1981), R46 (per; Dodd & Bennet, 1986), and F (rfsF/D-protein; Lane et al, 1986).

1.5 SITE-SPECIFIC RECOMBINATION

The significance of site-specific DNA re-arrangements in, and between, the genomes of a diverse range of organisms has only recently begun to be appreciated (for review see Sadowski, 1986). Such re-arrangements represent important steps in the development of many organisms ranging from bacteriophages to higher mammals. The range of functions dependent upon site-specific recombination is continually

```
        3731
COLE1   :  GTGAAACCATGAAAAATGGCAGCTTCAGTGGATTAAGTGGGGGTAATGTGGCCTGTACCC
PMB1    :  .C.....T.............G...G.......TT.AC...C........T...C...A
COLK    :  TGA...GATG......AAA...TG...TGA.G..T.CGCA..CA..ACCA..G.TATT.A
COLA    :  .GA...GAGCCTG...C.TGGAT.A...G.TTA.TTTAT..T.A..CAC.CA...ATT.A
CLODF13:  AGA..GT.GGT.....AA.GGCT.A.G.CATCCATTT.AC.TCA..ACATATGCTATG.A
                -35
          3793  3800      P1          -10
COLE1   :  TCTGGTTGCATAGGTATTCATACGGTTAAAATTTATCAGGCGCGATCGCGGCAGTTTTTC
PMB1    :  ............................................................
COLK    :  CA...C..A....C...G.....................T.................-.
COLA    :  C....C.....G.T...G..GT...G.............T...T...G...G....AC.
CLODF13:  GT..CG........C..G...TA..A.........C.G......T.TC....G.....C.

COLE1   :  GGGTGGTTTGTTGCCATTTTTACCTGTCTGCTGCCGTGATCGCGCTGAACGCGTTT-TAG
PMB1    :  ..........................................A...........-...
COLK    :  ............................A..A.....................A...-C..
COLA    :  ..............GG............T.C...CC..A...AG.....A...CCGG..
CLODF13:  ......G.......TTG.....T.CCGTA..C....GA.A...CGCC.GT..C..AC.G.

                CROSS-OVER REGION                                  3966
COLE1   :  CGGTGCGTACAATTAAGGGA--TTATGGTAAATCCACTTACTGTCTGCCCTCGTAGCCATC
PMB1    :  ...................--...........A.ACGTA...TAAT.TATCG.CAT..G
COLK    :  ...................--...........--...................T.....
COLA    :  ....-----..CG....C...TG...........AC..A.A.A.GAC.GT.AC.TT.TTC.
CLODF13:  ....-----..CGA.......TG...........ATCT.CGAGT.TCAGGTA.AA.AAA.G.
```

Figure 1.2 Optimal Alignment of ColE1 *cer* with the analagous sites from pMB1, ColK, ColA, and CloDF13; (Greene *et al*, 1981; Summers *et al*, 1985; Leung *et al*, 1985; Morlon *et al*, in preparation; Hakkaart *et al*, 1984). The *cer* sequence begins at ColE1 co-ordinate 3731bp and ends at the *taq*I site at 3966bp. (.) indicates sequence identity; (-) indicates a gap inserted to maximise homology. The cross-over region, promoter P1, and several ColE1 co-ordinates are indicated.

expanding and includes;

1) Integration and excision of temperate phage genomes e.g. Lambda (Weisberg & Landy, 1983)
2) The resolution of co-integrate structures formed during the transposition of class II transposons such as Tn3 (Arthur & Sherratt, 1979; Reed, 1981).
3) The switching of flagellar antigens in S.typhimurium (Simon et al, 1980).
4) Control of type I fimbriation in E. coli (Klemm et al, 1986).
5) Switching of host ranges in bacteriophages Mu and P1 (van de Putte et al, 1980; Iida et al, 1982).
6) Maintenance of faithful segregation in for example P1 (Austin et al, 1981), and ColE1 (Summers and Sherratt, 1984). Also in the maintenance of the multicopy 2 micron circle of S. cerevisiae (Futcher, 1986; Volkert & Broach, 1986).
7) The expression of nitrogen fixation genes during heterocyst differentiation in cyanobacteria (Golden et al, 1985).
8) The generation of diversity of antibodies (Bernard et al, 1978), and of the T-cell receptors of the immune system (Gascoigne et al, 1984).

Site-specific recombination at a functional level is quite independent of generalised (Homologous) recombination. The process requires two interacting sites which may be identical, as in the case of Tn3 resolvase (Grindley et al, 1982), or may be heterologous as in the case of the Lambda Int system (Weisberg & Landy, 1983). Recombination between these "Special sites", requires a "site-specific" recombinase (and in some cases additional accessory factors). In all systems characterised in detail the recombination reaction is conservative, involving no net synthesis or loss of DNA, (Sadowski, 1986), and does not require any high energy co-factors. Instead the recombinase forms a covalent intermediate with one of the DNA termini upon strand-breakage (either a phospho-serine of phospho-tyrosine linkage) and in this the recombinase appears to conserve the energy of the phospho-diester bond for subsequent re-ligation of the phospho-diester backbone. Finally, all known site-specific recombinases appear to act stoichiometrically, as opposed to catalytically, for reasons which remain uncertain.

Site-specific recombination may be either inter- or intra-molecular. Intermolecular recombination results in the fusion of DNA sequences, e.g. phage lambda integration into the bacterial chromosome. The product(s) of an intramolecular reaction depend upon the relative orientation of the participating sites: most sites, though superficially symmetrical, tend to possess an inherent asymmetry which imparts a polarity to the site, and thus a directionality to the recombination reaction. When two sites are present direct repeat then recombination between them results in the excision of the intervening sequence plus one copy of the site (this reaction is often termed "Deletion" or "Resolution"; e.g. the resolution of co-integrates by Tn3 resolvase). Alternatively, recombination between two sites in inverted repeat results in the inversion of the sequence between the sites (e.g. the Mu Gin system). Some recombination systems exhibit a remarkable discrimination in the directionality of the reactions they will perform; e.g. the transposon resolvases will catalyse the intramolecular resolution reaction between directly repeated res sites, but are unable to promote either inversion between inverted copies of res, or inter-molecular fusion reactions (Krasnow & Cozzarelli, 1983; Reed, 1981). This bias towards the resolution reaction is consistent with the biological function of the resolvases (Arthur & Sherratt, 1979). The range of complexity found in various site-specific recombination systems is best illustrated by consideration of a few examples.

1.5.2 LAMBDA INTEGRATION AND EXCISION

Lambda integrative recombination was the first site-specific recombination system to be reproduced in vitro (Nash, 1975). The recombining sites, attP (240bp) and attB (25bp), are distinctly heterologous; they share a 15bp core sequence (O), but this is flanked by unique arms in each att site designated P, P', B, and B' (attP= POP', attB= BOB'; Weisberg & Landy, 1983). Strand exchange occurs within the 15bp of homology to generate two hybrid sites attR (BOP'), and attL (POB') (Mizuuchi et al, 1975). Integrative recombination requires the phage encoded Int recombinase (40KDa; Nash, 1975), plus the E. coli "Integration Host Factor" (IHF) (Nash et al, 1977; Kikuchi & Nash, 1978). The interactions between Int, IHF and the att sites are complex; Int has been shown to specifically bind to 4 seperate regions within attP (Hsu et al, 1980), whilst IHF binds

to a further 3 discrete locations (Craig & Nash, 1984).

Excisive recombination occurs between the hybrid sites and requires both Int and IHF, plus the phage encoded Xis protein (Guarneros & Echols, 1970). Xis has been shown to bind the P arm of both attR and attP (Yin et al, 1985). Binding of Xis to attP inhibits integrative recombination.

The role of IHF in Int-specific recombination appears to be as an accessory factor since low-efficiency recombination can proceed in its absence (Lange-Gustafson & Nash, 1984). IHF is a hetero-dimeric DNA binding protein, comprising two polypeptide sub-units in equimolar ratios; the alpha peptide (the 10.5KDa product of the himA gene which maps at 38 mins), and the beta peptide (9.5KDa product of the hip gene located at 20.5 mins), for review see Weisberg and Landy (1983). Whilst IHF was first discovered via its role in Lambda Int-specific it has subsequently been found to be intimately involved in an apparently diverse range of biological functions including the regulation of expression of a number of genes and operons including Lambda cII (Hoyt et al, 1982), the early operon of phage Mu (Goosen & van de Putte, 1984), and also the val, ilv, and xyl genes of E. coli (Friden et al, 1984; Friedman et al, 1984). IHF has also been found to be involved in the conjugal transfer of R100, in that it appears to be required for the normal expression of the plasmid tra genes (Dempsey, 1987). Gamas et al (1986) have also reported that IHF is involved in the stable maintenance of pSC101, perhaps at the level of promoting initiation of DNA replication. In the cases of pSC101 replication and Lambda integration, it has been suggested that IHF operates by so affecting the local topology of the region as to favour the subsequent/concomitant interaction of other proteins, in the formation of complex protein:DNA assemblies necessary to promote efficient recombination or initiation of DNA replication.

1.5.3 THE loxP/Cre SYSTEM OF P1

P1 in the prophage state behves as an autonomously replicating episome, and is maintained at a copy number of 1-2 per chromosome equivalent (Ikeda & Tomizawa, 1968). The stable maintenance of P1 is due, at least in part, to a site-specific resolution function analagous to ColE1 cer. This recombination system requires only two copies of the loxP site plus the phage encoded Cre protein (35KDa; Austin et al, 1981). No other phage or host components are required

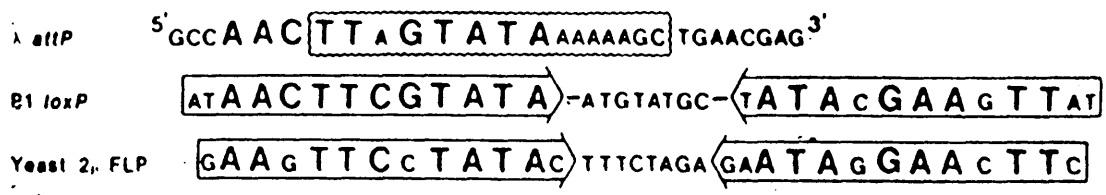


Figure 1.3 Comparison between P1 loxP, Lambda attP, and the yeast FLP sequence. The core sequence of lambda attP (o) where crossing-over takes place is indicated by the boxed region ([ ]). The inverted repeats of loxP and FLP are highlighted by boxed arrows. Homologous bases are indicated by large type (after Hoess and Abremski, 1984).

for recombination either *in vivo* or *in vitro* (Hoess & Abremski, 1984). The 34bp *loxP* site is relatively simple, comprising two 13bp inverted repeats hyphenated by an 8bp spacer region. The polarity of the site, and thus the directionality of the reaction, is conferred by this 8bp region. Cre has been shown to bind to two domains within *loxP*, each comprising one 13bp repeat plus the contiguous 4bp of the spacer region (Hoess & Abremski, 1984). Occasionally P1 undergoes Cre-mediated integration into the bacterial chromosome at a specific site designated *loxB* (Sternberg *et al*, 1981).

Despite the apparent difference in complexity between *lox*/Cre and the lambda *int* systems the *loxP* site exhibits significant sequence similarity to lambda *attP*, and even more so to the Flp site in the yeast 2-micron circle (Fig. 1.3; Hoess & Abremski, 1984). Furthermore, significant similarities have been identified in the amino-acid sequences of the lambda Int, Cre and Flp proteins, together with those of the integrases from phages P2, 186, P22, Phi 80, and P4 (Argos *et al*, 1986; Dorman & Higgins, 1987). The observed homologies might indicate that this group of recombinases may be analagous, if not homologous; the group has been termed the "Integrase family" (Argos *et al*, 1986).

1.5.4 THE INVERTASES

Site-specific DNA inversions are used as biological switches in a number of systems. The most fully characterised being the G-segment inversion in phage Mu, and the H-segment inversion in *S. typhimurium*; at the molecular level these two systems share many common features. The H-segment inversion is catalysed *in vitro* by the 21KDa Hin recombinase plus two additional factors, Fis (Factor for Inversion Stimulation) and HU (the major Histone-like protein) (Johnson *et al*, 1986; Johnson & Simon, 1987; Kahmann *et al*, 1985). The consequence of inversion *in vivo* is to connect, or disconnect, the promoter for the H2 flagellin gene, thus serving to activate or inactivate H2 gene expression (Fig. 1.4). The *cis* requirements for Hin inversion include two 26bp recombination sites, *hixL* and *hixR*, plus a 60bp enhancer sequence which stimulates recombination 200 fold (Johnson & Simon, 1987). Intriguingly, the enhancer element is present within the coding sequence of the *hin* gene, and is thus situated within 200bp of *hixL* (see Fig. 1.4). However the enhancer-like properties of this region can be demonstrated by the fact that it continues to enhance recombination

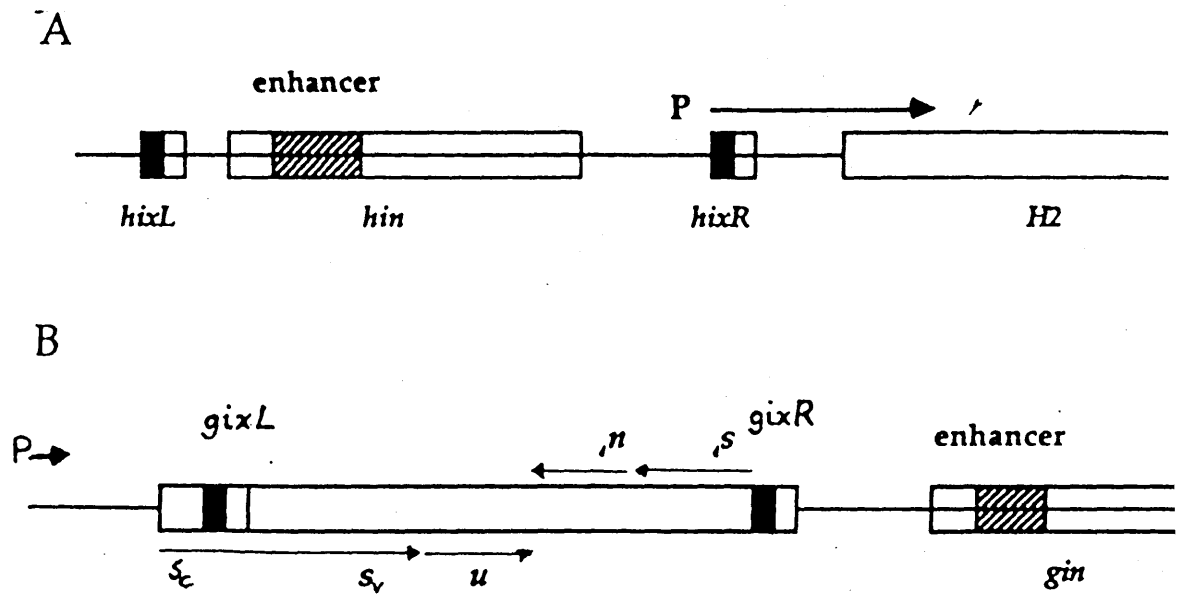


Figure 1.4 Schematic representation of the Hin and Gin inversion systems.

A. The H2-promoter is shown in the **ON** position in which it initiates transription of the H2 gene. Hin mediated inversion between hixL and hixR uncouples H2 from its promoter, thus switching H2 expression **OFF**.

B. The S and U genes are transcribed from promoter P. The S gene has a constant 5' region (S_c), but may have one of two different 3' coding regions (S_v or S'). Inversion between gixL and gixR above would fuse the S' variable region to the S_c constant region, thus creating an alternative C-terminal coding sequence. Inversion also serves to switch on either U or U' by transcription from promoter P.

when present up to 4Kb from the *hix* sites, and does so in either orientation (Johnson & Simon, 1987). The Fis protein (12KDa) binds to two sites within the enhancer sequence and is essential for enhancer function; Fis has many similarities to IHF and may stimulate recombination in an analagous fashion (Kahmann *et al*, 1985). The HU protein is a second accessory factor which stimulates Hin recombination *in vitro* by a further 10-fold. HU is a heterodimeric protein which binds DNA non-specifically. It has been suggested that HU may coat DNA and may, in doing so, ease the torsional stresses inherent in the bending of DNA to bring two recombination sites into juxtaposition (Johnson *et al*, 1986; Johnson & Simon, 1987). HU is also required for the initiation of strand exchange during Mu transposition *in vitro* (Craigie *et al*, 1985), and has also been shown to be involved in *oriC*-specific DNA replication (Dixon & Kornberg, 1984). In both these cases a role for HU in the formation of topologically complex protein:DNA intermediates has been suggested.

The H-segment appears to belong to a family of closely related invertible segments including the Mu G-segment, the P1 and P7 C-segments (all of which change alter the phage's host-range), and the cryptic P-segment on the *E. coli* chromosome. The respective invertases Gin, Cin, and Pin, all share extensive sequence similarity to Hin at the amino-acid level, and all cross complement one another. The sequence similarity extends to the Tn3 and Gammadelta resolvases, but these do not functionally complement the inversion systems (Plasterk & van de Putte, 1984; Simon *et al*, 1980). The *gin* gene also contains an enhancer region whose function depends upon Fis binding, this sequence has been designated *sis* (sequence for inversion stimulation; Kahmann *et al*, 1985).

The formation of a highly ordered nucleo-protein complex as an intermediate in site-specific recombination has been proposed for a number of systems. The formation of such a structure may play a crucial role in maintaining the fidelity (and perhaps the directionality) of the recombination reaction (Boocock *et al*, 1987; Gellert & Nash, 1987).

1.5.5 TYPE I FIMBRIATION

The expression of type I fimbriae in *E. coli* is subject to phase variation, i.e. a cell is either completely fimbriated or "bald". The phase variation being the result of a genetic switch occurring at a

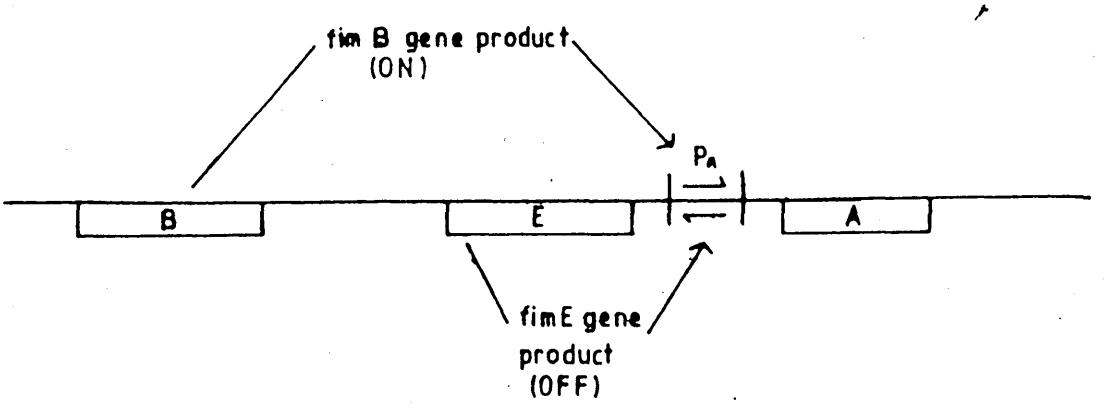


Figure 1.5 Organization of the fim genes.

The fimB gene product controls the switching of the 300bp region into the on position with the promotor pointing into fimA, whereas the fimE gene product is responsible for switching the 300bp region into the off position with the promoter pointing away from fimA.

frequency of 10^{-3} per generation (Eisenstein, 1981). This genetic switch has recently been discovered to be the result of a site-specific inversion of a 300bp segemnt of the *E. coli* chromosome (Abraham *et al*, 1985; Klemm, 1986). The invertible segment contains the promoter for the *fimA* gene which encodes the major structural component of type I fimbriae. Thus *fimA* expression is switched on or off in a fashion analagous to H2 gene expression in *S. typhimurium* (Fig. 1.5). The inversion is apparently controlled by two proteins whose genes are located immediately upstream of the invertible fragment (see Fig. 1.5). The products of these two genes regulate the directionality of the inversion switch; FimB and FimE directing the switching of the *fimA* promoter into the "On" and "Off" orientations respectively (Klemm, 1986). The *fim* genes map to 98 mins. on the K-12 linkage map (Brinton *et al*, 1961). The *flu* gene (43 mins.) also apears to be involved in the *fimA* switch, however its precise role remains obscure (Diderichsen, 1980; Klemm, 1986).

The *fimB* and *fimE* genes are highly homologous at the DNA sequence level; they encode polypeptides of 200 and 198 residues respectively, which in-turn exhibit extensive homology in their amino-acid sequences (Klemm, 1986). On the basis of this homology it has been proposed that the two genes arose out of a gene-duplication event (Klemm, 1986). It has been subsequently noted that the FimB and FimE proteins share striking similarities to the major conserved regions identified amongst the "Integrase family" of site-specific recombinases (Dorman & Higgins, 1987). It therefore seems probable that the FimB and FimE proteins are infact recombinases; if so then this use of a pair of antagonistic invertases would represent a quite novel system.

1.6 *cer*-SPECIFIC RECOMBINATION

Very little is currently understood regarding the mechanism(s) of *cer*-specific recombination. At a functional level, the system requires the presence of two directly repeated copies of the 279bp (*Hpa*II/*Taq*I) *cer* site. The *cer* site has more recently been de-lineated to within a 233bp region of ColE1 running from co-ordinate 3733bp to the *Taq*I site at 3966bp; the actual cross-over site having been further defined to within a 35bp region of *cer* (3906-3942bp) (Summers *et al*, 1985; Summers & Sherratt, manuscript in preparation). Recent evidence suggests that at least part of this 233bp region need be present in only one of the two participating sites; progressive

ExoIII deletions into the left-hand end of cer produced a series of deleted-sites which will recombine with a copy of the wild-type site, but not with another copy of the deleted-site. Progressive deletions into the 233bp cer region show a concomitant decrease in the efficiency with which they will recombine with a wild-type site; the smallest fragment which retains this partial activity being D3793 (a 173bp region from 3793-3966bp), the next deletion in the series, D3800 (3800-3966bp), removes a further 7bp at which point recombination with a wild-type site is abolished (see Fig. 1.2; Sherratt et al, 1986; Summers & Sherratt, manuscript in preparation). It has been noted that the region of cer required in only one copy (i.e. from 3733-3793bp), contains a series of A+T and G+C rich regions which alternate with a periodicity of approximately 10bp; this periodicity is conserved in the sites of related plasmids even though the primary sequence of this region shows considerable "divergence" (see Fig. 1.2). Such sequence periodicity underlies the bending of DNA around nucleosomes (Drew & Travers, 1985); it has been suggested that the loss of this "bendable" sequence may be responsible for the observed reduction in recombination efficiency (Summers & Sherratt, manuscript in preparation).

Finally note that the cer site contains a consensus promoter sequence, P1 (see Fig. 1.2); transcriptional fusion experiments have confirmed that this promoter is indeed active in vivo. The significance, if any, of such transcription is unclear, but it is interesting to note that the loss of the -35 region of P1 in D3800 is correlated with the complete abolition of recombination at that site. However, a site-directed mutation in the -10 region of P1 which reduces transcription 18-fold, has no apparent affect on recombination efficiency (Summers & Sherratt, manuscript in preparation); an alternative explanation for the affect of the 7bp difference between D3793 and D3800 will be presented later.

1.6.2 THE cer-SPECIFIC RECOMBINASE

The small size and limited coding potential of the cer site make it seem unlikely that the recombinase is encoded by it. The largest potential cer-encoded polypeptide would be 70 residues in length, encoded by a potential gene running from 3966-3756bp; this compares with the smallest known site-specific recombinases which are at least 2.5 fold larger (e.g. the 185 residue Tn3 resolvase). Leung et al

(1985), introduced a frame-shift mutation into this potential "gene" at position 3901bp (22 codons into the reading frame), but found no reduction in recombination efficiency. Furthermore, when one compares the DNA sequence of the cer site to those of the highly "conserved" loci from related plasmids (Fig. 1.2), then one finds a number of 1 and 2 bp insertions/deletions which indicate that no protein coding gene has been conserved amongst all five sites. This led to the proposal that the cer-specific recombinase might be encoded by the E. coli chromosome, at a locus designated xer (chromosomally encoded recombinase; Summers et al, 1985). Whilst a number of host-encoded accessory factors (e.g. IHF, Fis etc.) are involved in various site-specific recombination systems, in all cases known the recombinase per se is encoded by a locus adjacent to the site at which it acts, thus a host encoded cer-specific recombinase would represent a novel system.

The existence of E. coli Xer function(s) was first demonstrated by Gill Stewart (1986), who isolated two mutants defective in cer-specific recombination (DSX1 and DSX2). The mutant selection procedure involved the transformation of a mutagenised population with the chimaeric plasmid pKS441; this plasmid is essentially a pseudo-dimer consisting of one replication origin, but two copies of cer with a resistance marker (cat) inserted between them. When a wild-type strain (Xer^+) is transformed with this plasmid then site-specific recombination between the cer sites results in the resolution of pKS441 into two discrete circular products, p442 and p443; this effectively isolates the cat gene from the plasmid replication origin (Fig. 1.6). Thus Xer^+ transformants remain chloramphenicol sensitive, whereas Xer^- transformants fail to resolve pKS441 and can therefore be selected by virtue of their resistance to chloramphenicol. Note that the pKS441 used in these experiments was recovered from the strain W5445; this strain was found to exhibit a "partial" Xer phenotype, i.e. it did not completely resolve a 2-cer plasmid after transformation and thus maintained a mixture of resolved and unresolved plasmid. The 2-cer form could therefore be gel-purified for use in the mutant selection procedure: this extremely laborious task was of course overcome after the isolation of the first Xer^- mutant (Stewart, 1986).

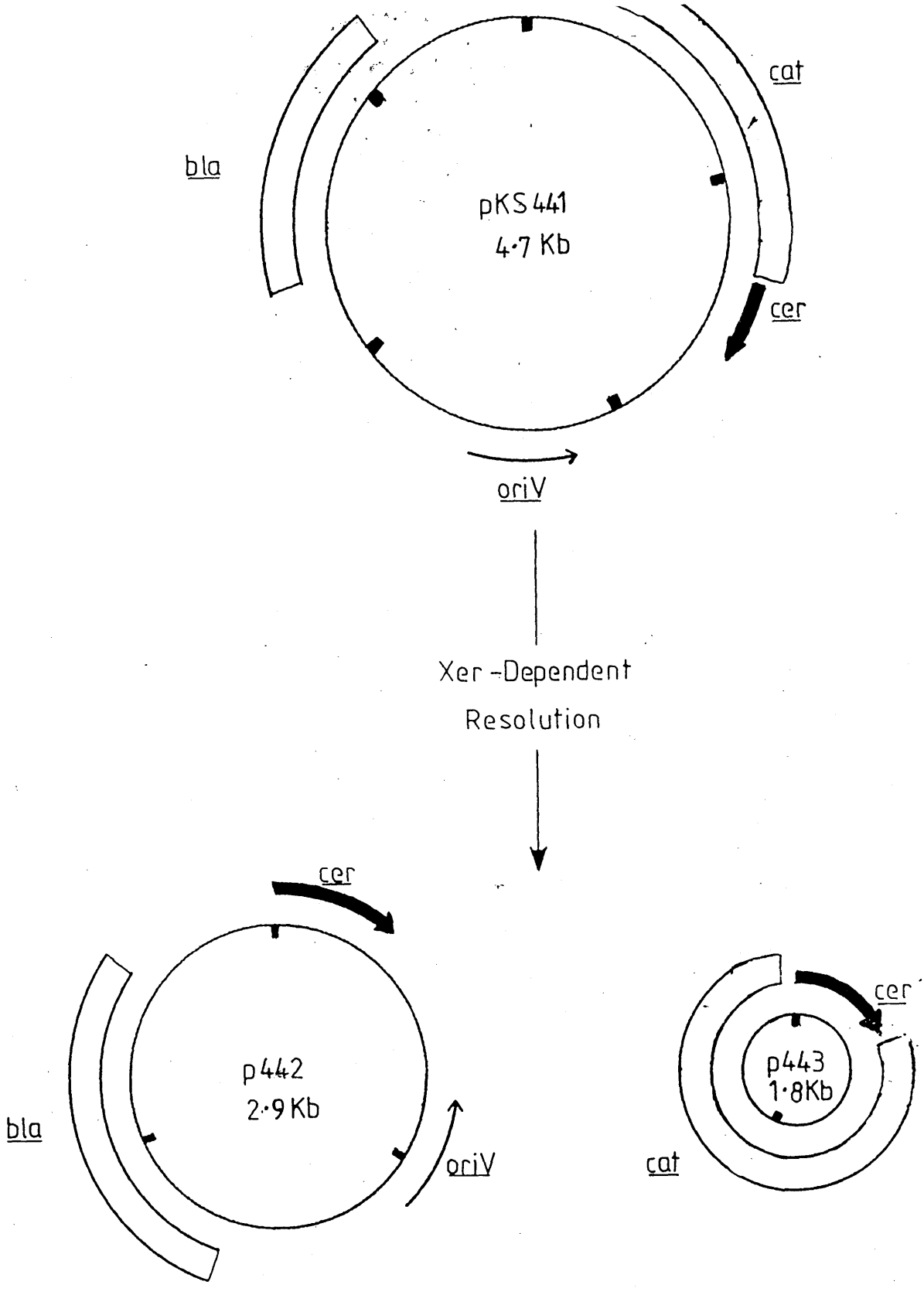


Figure 1.6 Diagrammatic representation of the Xer-Dependent Resolution of the 2-cer reporter plasmid pKS441

1.7 PROJECT AIMS

1. To increase the number of independently isolated *xer* mutants.

2. To use these mutants to clone the *xer* gene(s) by complementation and thus identify a minimum number of complementation groups.

3. To characterise the cloned gene(s), and investigate the nature of the encoded function(s).

CHAPTER 2

MATERIALS AND METHODS.

Table 2.1 Bacterial Strains

Strain	Genotype	Source/Reference
DS903	*recF*143, *thi*, *leuB*6, *his*4, *thr*1, *proA*2 *argE*, *galK*2, *supE*44, *tsx*3, *rpsL*31, *ara*14	D.J. Sherratt
DS902	As DS903 but *recA*13, $ArgE^+$, $RecF^+$	D.J. Sherratt
DS941	As DS903 but $lacI^q$ ZΔM15	D.J. Sherratt
DS942	As DS903 but Δ(*lacpro*)	D.J. Sherratt
DS947	As DS942 but SupO	D.J. Sherratt
CSH26	*ara*, *thi*, Δ(*lacpro*)	Miller (1972)
JM101	*supE*, *thi*, Δ(*lacpro*), (F', *traD*36, *proAB*, $lacI^q$ ZΔM15)	Yanisch-Perron *et al* (1985)
DSX1	As DS903 but *xerB*8 (formerly *xer*1)	Stewart (1986)
DSX2	As DS903 but *xerB*9 (formerly *xer*2)	Stewart (1986)
CSX1	As CSH26 but *xerA*3 (formerly Xer3)	This work
CSX2	As CSH26 but *xerA*3 (" X3.2)	"
CSX3	As DS903 but *xerA*3 (" X3.D)	"
CSX4	As DS947 but *xerB*1	"
CSX5	" " " *xerB*2	"
CSX6	" " " *xerB*3	"
CSX7	" " " *xerB*4	"
CSX8	" " " *xerB*5	"
CSX9	" " " *xerB*6	"
CSX10	" " " *xerB*7	"
CSX11	" " " *xerA*5	"
CSX16	As JM101 but *xerA*3	"
CSX17	As DS941 but *xerB*1	"
CSX18	As JM101 but *xerB*1	"
CSX20	As DS941 but *xerB*2	"
CSX21	" " " *xerB*4	"
CSX22	" " " *xerB*5	"

Table 2.2 Plasmids

Plasmid	Marker(s)	Description	Source/Ref.
ColE1	Colicin E1	Naturally occurring	D. Sherratt
pBR322	Ap^r, Tet^r	Derived from pMB1	Sutcliffe (1979)
pUC8	Ap^r	Derived from pBR322	Vieira & Messing (1982)
pUC9	"	"	"
pUC18	"	"	Norrander et al (1983)
pUC19	"	"	"
pACYC184	Cm^r, Tet^r	Derived from p15A	Chang & Cohen (1978)
pCB101	Cm^r	Lambda dv replicon	Boyd & Sherratt (1986)
pCB18B	Cm^r	Derived from pCB101	A.C. Boyd
pCB19B	"	"	"
pBAD	Ap^r	ptac expression vector derived from pKK223	A.C. Boyd
pMC1871	Tet^r	contains lacZ gene cartridge	Shapira et al (1983)
pKS450	Ap^r	pUC9 plus 377bp HpaII cer fragment	D.K. Summers
pKS451	"	"	"
pKS492	"	pUC18 plus 279bp cer fragment	"
pKS493	"	"	"
pKS455	Ap^r, Cm^r	Derivative of pKS450 containing 2 cer sites flanking the cat gene	"
pGS30	Ap^r	pAT223 + EcoRI/HindIII xerA fragment	Stewart (1986)
pGS36	"	pBR322 + SphI xerA fragment	Stewart (1986)
pGS38	"	pUC19 + SphI/AccI xerA fragment	Stewart (1986)
pCS80	Ap^r, Tet^r, Km^r	pBR322::Tn5	Chapter 4
pCS100	Ap^r	306bp HpaII Rom^+ fragment from ColE1 cloned into NarI site of pUC8	This work
pCS110	Ap^r	pCS100 + xerB as a 3Kb BamHI fragment	Chapter 5
pCS111	"	pUC18 + xerB as a 2.8KB SphI fragment	"
pCS112	"	pUC18 + xerB as a 1.9Kb HindIII fragment	"
pCS118	Cm^r	pCB19B + 1.9KB HindIII fragment from pCS112	This work
pCS119	"	" " " " " "	"
pCS120	"	pCB18B + 1647bp EcoRI/HindIII fragment	Chapter 5
pCS121	"	pCB18B + 274bp EcoRI/HindIII fragment	"
pCS124	Ap^r	pUC19 + 829bp HindIII/MluI fragment	"
pCS125	"	pUC19 + 1092bp HindIII/MluI fragment	"
pCS126	"	pBAD + 1.9Kb HindIII fragment from pCS112	Chapter 7
pCS127	"	" " " " " "	"
pCS130	"	AccI deletion derivative of pCS112 ($xerB^+$)	Chapter 5
pCS131	"	KpnI deletion derivative of pCS112	"
pCS132	"	HincII deletion derivative of pCS112	"
pXB1	Ap^r, Km^r	5.2Kb BamHI fragment from xerB1 cloned into pBR322	"

Table 2.2 cont.

Plasmid	Marker(s)	Description	Source/Ref.
pCS202	Cm^r, Tet^r	Lambda dv-based 2 cer reporter plasmid: deletion marker = Tet^r	Chapter 3
pCS203	Cm^r	Xer-deletion derivative of pCS202	"
pCS208	Tet^r, $IPTG^s$	pAT153 based 2 cer reporter plasmid: deletion marker = ptac-mob5	This work
pCS209	Tet^r, $IPTG^r$	Xer-deletion derivative of pCS208	
pCS210	Tet^r, $LacZ^+$	pACYC184 based 2 cer reporter plasmid: deletion marker = lacZ	Chapter 4
pCS211	Tet^r	Xer-deletion derivative of pCS210	"
pCS300	Ap^r, Tet^r, Km^r	pBR322 + the xerA3 allele	"
pCS349	Cm^r	pCB19B + SphI/AccI xerA fragment	"

Table 2.3 Bacteriophages

Phage	Description	Source/Reference
$P1_{kc}$	Generalised transducing phage	M. Masters
Lambda NK467	Suicide vector for Tn5 mutagenesis	N. Kleckner
M13mp18	Cloning vector derived from M13	Norrander et al (1983)
M13mp19	" " " " "	"
mCS301	M13mp19 + 3.4Kb SphI/PstI fragment from pCS300	Chapter 6
mCS304	M13mp18 + 1Kb SphI/PstI fragment from pCS300	"
mCS305	M13mp19 + 0.8Kb DraI/PstI fragment from mCS304	"
mCS320	M13mp19 + SphI/AccI xerA fragment from pGS36	"
mCS112	M13mp19 + 1.9Kb HindIII fragment from pCS111	Chapter 7
mCS113	" " " " " "	"
mCS114	" + 0.9Kb HindIII/SstI fragment from pCS124	"
mCS115	" + 0.7Kb HindIII/BglII fragment from pCS112	"
mCS116	HincII deletion derivative of mCS113	"
mCS117	KpnI deletion derivative of mCS113	"
mCS118	AccI deletion derivative of mCS113	"

2.1 Bacterial strains. The bacterial strains used were all derivatives of Escherichia coli K-12 and are listed in table 2.1. Genotype and phenotype symbols are those recommended by Bachman et al (1976) and Novick et al (1976).

2.2 Plasmids. The plasmids used and constructed in this study are listed in table 2.2 and their nomenclature follows that of Novick et al (1976).

2.3 Bacteriophages. The bacteriophages used and constructed during this work are listed in table 2.3.

2.4 Chemicals.

CHEMICALS	SOURCE
General chemicals	B.D.H., Hopkins and Williams, Kochlight Laborotories, May and Baker
Media	Difco, Oxoid
General biochemicals	Sigma, Pharmacia
Agarose	BRL
X-gal, IPTG	BRL
Radiochemicals	NEN
10 X core buffer	BRL
Antibiotics	Sigma
Restriction enzymes	BRL, Boerhinger mannheim
En^{3}hance	Dupont

2.5 Culture media.

L-Broth: 10g tryptone, 5g yeast extract, 5g NaCl, 1g Glucose, 20mg thymine, made up to 1 litre in distilled water and adjusted to pH7.0 with NaOH.

L-Agar: As L-Broth without glucose and the addition of 15g/l agar.

Iso-sensitest Broth: 23.4g Iso-sensitest Broth made up to 1 litre with distilled water.

Iso-sensitest Agar: 31.4g Iso-sensitest Agar made up to 1 litre with distilled water.

Minimal Agar: 7g K_2HPO_4, 2g KH_2PO_4, 4g NH_4SO_4, 0.25g trisodium citrate, 0.1g $MgSO_4.7H_2O$, 17.5g agar, made up to 1 litre in distilled water.

Davis-Mingioli (D&M) Salts (X4): 28g K_2HPO_4, 8g KH_2PO_4, 4g $(NH_4)_2SO_4$, 1g sodium citrate, 0.4g $MgSO_4.7H_2O$, made up to 1 litre with distilled water.

D&M Minimal Medium: 25ml D&M salts, 5ml 20% casamino acids, 250ul 20% glucose, 0.5ml vitamin B1 (1mg/ml), made up to 100ml with distilled water.

M9 Salts (X10): 6g Na_2HPO_4, 3g KH_2PO_4, 0.5g NaCl, 1g NH_4Cl in 1 litre of distilled water.

M9 Minimal Medium: 10ml M9 salts, 2ml glucose (20% w/v), 0.1ml 1M $MgCl_2$, 0.1ml 100mM $CaCl_2$, 0.1ml (1mg/ml) vitamin B1, 2.5ml 20% casamino acids, made up to 100ml with distilled water.

3XD Minimal Media: 150ml phosphate concentrate, 10ml 10% NH_4Cl, 6ml 10% $MgSO_4.7H_2O$, 10ml 0.1% Glycerol, 3ml 0.05M $CaCl_2$, 15g casamino acids, 797ml distilled water. Phosphate concentrate and $CaCl_2$ added slowly after autoclaving.

210-Medium: D&M minimal medium, casamino acids (1%), glucose (0.2%), X-Gal (25mg.l^{-1}), Tetracycline (7mg.l^{-1}).

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

Phosphate concentrate: 30g KH_2PO_4 (anhydrous), 70g Na_2HPO_4 (anhydrous) made up to 1 litre in distilled water.

Supplements: When required, supplements were added to minimal media at the following concentrations:

glucose 2mg/ml	amino acids 40ug/ml	thymine 50ug/ml
thiamine vitamin B1 20ug/ml		casamino acids 1%

2.6 Sterilisation. All growth media were sterilised by autoclaving at 120°C for 15 minutes; supplements, gelatin solution and buffer solutions at 108°C for 10 minutes and $CaCl_2$ at 114°C for 10 minutes.

2.7 Buffer solutions.

Electrophoresis

10 X E Buffer: 48.4g Tris-HCl, 16.4g NaAc, 3.6g $Na_2EDTA.2H_2O$, made up to 1 litre in distilled water, pH adjusted to 8.2 with acetic acid.

10 X TBE Buffer pH8.3: 109g Tris-HCl, 55g Boric Acid, 9.3g $Na_2EDTA.2H_2O$, made up to 1 litre in distilled water; pH is 8.3.

10 X Laemmli gel buffer: 144g Glycine, 30g Tris-HCl made up to 1 litre in distilled water. When diluted, 10ml 10% SDS (w/v) was added for every 1 litre of 1 X concentration.

Single Colony Gel Loading Buffer: 2% Ficoll, 1% SDS, 0.01% Bromophenol blue, 0.01% Orange G in Buffer E.

Protein loading buffer: 5% SDS, 50% Glycerol, 0.01% Bromophenol blue, 50mM Tris (pH 6.8), made up to 5% beta-mercaptoethanol immediately prior to use.

Horizontal Agarose and Polyacrylamide Gel Loading Buffer: 25% (w/v) sucrose, 0.06% Bromophenol Blue, 10mM Tris-HCl (pH8.0).

Vertical Agarose Gel Loading Buffer: 1% Ficoll, 0.5% SDS, 0.06% Bromophenol Blue, 0.06% Orange G in 1 X TBE.

Restriction and Ligation Buffers:

10 X Low salt: 100mM Tris-HCl pH7.5, 100mM $MgSO_4$, 10mM DTT. Stored at 4^oC.

10 X Medium salt: 500mM NaCl, 100mM Tris-HCl pH7.5, 100mM $MgSO_4$, 10mM DTT. Stored at 4^oC.

10 X High salt: 1M NaCl, 500mM Tris-HCl pH8.0, 100mM $MgCl_2$, 10mM DTT. Stored at 4^oC.

10 X SmaI Buffer: 200mM KCl, 100mM Tris-HCl pH8.0, 100mM $MgCl_2$, 10mM DTT. Stored at 4^oC.

10 X Ligation Buffer: 660mM Tris-HCl pH7.6, 66mM $MgCl_2$, 100mM DTT. Stored at -20^oC.

4mM ATP: Dissolve 60mg of ATP in 0.8ml distilled water. Adjust to pH7.5 with 0.1M NaOH, made up to 1ml with distilled water; Stored at -20^oC in aliquots and thawed once only.

TE Buffer: 10mM Tris-HCl, 1mM EDTA; pH8.0.

All these buffers were stored for long term at -20^oC.

Protein sample buffer: 10% Glycerol, 0.01% Bromophenol blue, 5% B-mercaptoethanol, 3% SDS, 0.625M Tris-HCl pH8.0. Stored at room temperature.

10 X Nick Translation Buffer: 500mM Tris-HCl pH7.2, 100mM $MgSO_4$, 1mM DTT, 500ug/ml BSA; stored at -20^oC

20 X SSC: 3M NaCl, 300mM trisodium citrate pH7.0

Wash Buffer: 5mM NaH_2PO_4, 1mM Na_2EDTA, 0.2% SDS pH7.0

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

Birnboim-Doly I: 50mM Glucose, 25mM Tris-HCl pH8.0, 10mM EDTA; add lysozyme to 1mg/ml immediately before use.

Birnboim-Doly II: 0.2M NaOH, 1% SDS; made fresh.

Birnboim-Doly III: 2.5M KAc pH4.8; mix equal volumes of 3M CH_3COOK and 2M CH_3COOH, pH should be 4.8.

STET Buffer: 8% Sucrose, 5% Triton X-100, 50mM EDTA, 50mM Tris-HCl pH8.0

Buffered Saline Gelatin (BSG): 0.85% NaCl, 0.03% KH_2PO_4, 0.06% Na_2HPO_4, 100ug/ml Gelatin.

5% Sucrose BSG: BSG made in the presence of 5% Sucrose.

20% Sucrose BSG: BSG made in the presence of 20% Sucrose.

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

Name	Selective conc^n	Stock sol^n	Storage temp.(^oC)
Ampicillin (Ap)	50ug/ml	5mg/ml (water)	+4
Tetracycline (Tet)	10ug/ml*	1mg/ml (50% EtOH in water)	+4
Chloramphenicol (Cm)	25ug/ml	2.5mg/ml (EtOH)	-20
Kanamycin (Kan)	50ug/ml**	5mg/ml (water)	+4
Trimethoprim (Tp)	50ug/ml	5mg/ml (50% EtOH in water)	+4

When required, antibiotic stock solutions were added to molten agar pre-cooled to 55°C.

* 7ug/ml was employed with pCS210 and pCS211.

** 25ug/ml was employed when selecting chromosomal Tn5 insertions.

2.9 Indicators. X-gal (5-Bromo-4-chloro-3-indolyl-B-galactoside) was used in conjunction with the host strains DS941 and JM101 plus the pUC and M13 vectors providing a screen for plasmids with inserts in the polylinker region. Clones containing inserts are generally white; clones lacking inserts are blue. X-gal was stored at a concentration of 20mg/ml in dimethylformamide (DMF) at -20°C and added to L-agar

plates to a final concentration of 20ug/ml.

2.10 Growth conditions. Liquid culture for transformation, DNA preparations etc. were routinely grown in L-broth, or when Tp selection was required in isosensitest broth, at 37°C with vigorous shaking. Growth on plates was on L-agar, minimal medium plus supplements, or on isosensitest agar for selection on Tp. Antibiotics were used as required. Plates contained 25ml of agar solution and were incubated at 37°C overnight unless otherwise stated. All dilutions were carried out in D&M salts.

Bacterial strains were stored in 50% L-broth, 20% glycerol and 1% peptone at -20°C, and on L-agar slopes (room temp.). Inocula from these stocks were applied to L-agar plates and grown, prior to replating on selective plates.

2.11 Preparation of P1 lysates: The donor strain was grown up in L-broth to late log phase, then the cells harvested by centrifugation (10,000g,5mins). The cells were resuspended in the same volume of fresh L-broth, then 100ul of cells mixed with 100ul $CaCl_2$ (50mM), 100ul $MgSO_4$ (100mM), and 100ul of a P1 phage dilution sufficient to produce an m.o.i of approximately 0.001. The mixture was then vortexed, then incubated at 37°C, 25mins, to permit phage adsorbtion and infection. The cells were then added to 3mls molten soft-agar (47°C), then plated onto a fresh (undried) L-agar plate (containing 0.5mM $CaCl_2$). After overnight incubation (37°C), 2.5mls of phage-buffer was added to the plate and incubated for 15 mins. The soft agar layer plus phage buffer was then harvested and placed in a centrifuge tube. A few drops of chloroform were added to the tube, then the mixture vortexed thoroughly for 30 secs. This was then allowed to stand for 30 mins (room temp.) before re-vortexing, then spinning out the debris by centrifugation (12,000g, 10 mins). The resultant supernatant was then titred for phage concentration; the method routinely recovering 10^9-10^{10} p.f.u.ml^{-1}.

2.12 P1 transduction: The recipient strain was grown up to mid-log phase in L-broth, then harvested and resuspended in the same volume of fresh L-broth. To 100ul of cells was added 100ul $CaCl_2$ (50mM), 100ul $MgSO_4$ (100mM), and sufficient P1 lysate to provide an m.o.i. of

approximately 0.01. After 20 mins incubation at 37°C, phage infection was stopped by the addition of 200ul 1M sodium citrate (filter sterilised). After the addition of 500ul of L-broth, the mixture was incubated at 37°C, 30 mins, to permit expression. Then the cells plated onto a suitable medium.

2.13 Transformation with plasmid DNA. Plasmids were introduced to different strains by genetic transformation. An overnight culture of the recipient was diluted 1 in 100 into 20ml L-broth and was grown to a density of approximately 10^8 cells/ml (about 90min - 2 hours). The cells were harvested (12000g, 5min, 4°C) and resuspended in 10ml of cold 50mM $CaCl_2$. The cells were pelleted again, resuspended in 1ml of cold 50mM $CaCl_2$ and kept on ice for at least 15min before use. 200ul aliquots of the competent cells were added to the plasmid DNA, mixed gently and left on ice for up to 1 hour. The cells were heat shocked (2min, 42°C) and returned to the ice for a further 15min. An equal volume of L-broth was added and the cells incubated at 37°C for 1 hour to allow expression of the plasmid resistance genes. The cells were plated out on the appropriate selections. For transformation to ampicillin resistance, no expression time was necessary.

2.14 Transfection of JM101 with M13: This followed exactly that procedure described above (2.13), except that after heat-shock the cells were incubated on ice for 15 mins, after which 200ul of fresh, log phase JM101 culture was added. To this mixture was then added 30ul of IPTG (15mg/ml) and 30ul X-Gal (40mg/ml). The cells were then mixed, added to 3ml of molten soft-agar (pre-cooled to 47°C), and then plated onto thoroughly dried plates containing D&M minimal agar (+ B1 and Glucose).

2.15 UV sensitivity test. This technique was used to rapidly confirm the recombination status of strains. Stationary phase cultures of the strain to be tested and control strains were spotted on to L-agar plates at 10^{-2}, 10^{-4} and 10^{-6} dilutions. These plates were exposed to UV radiation (17.5ergs/sec/M^2) for 30 and 60 seconds and incubated overnight in the dark. *rec*$^+$ strains usually grow after 60 seconds exposure while *rec*A$^-$ strains show reduced growth after 30 seconds and do not grow after 60 seconds exposure. *rec*F$^-$ strains show an intermediate phenotype. Periodically all strains were checked for their

recombination status.

2.16 Plasmid DNA preparation. Two methods were used to obtain DNA from cells.

Birnboim and Doly (1979) DNA preparation- Large Scale: 200 ml cultures of stationary phase cells were harvested by centrifugation (12000g, 5min at 4°C). The pellet was resuspended in 4ml of Birmboin-Doly I solution and incubated on ice for 5min. 8ml of Birmboin-Doly II solution were added and the solution left on ice for 5min before 6ml of cold Birmboin-Doly III solution was added, gently mixed and left on ice for a further 5min. The cell debris and most chromosomal DNA was removed by centrifugation (32000g, 5min at 4°C) and the plasmid DNA precipitated by addition of an equal volume of isopropanol followed by centrifugation at 39200g for 15min. This DNA was further purified by banding on a CsCl/EtBr gradient. The DNA was resuspended in 2.09ml of water and added to 270ul of a 15mg/ml EtBr solution. 5g of CsCl were dissolved in 3ml of water and added to the DNA/EtBr solution. The gradients were centrifuged in a Beckman Ti70 fixed angle rotor at 200,000g for 16 hours at 20°C. Two bands were visible, a lower supercoiled plasmid band and an upper chromosomal and relaxed plasmid DNA band. The lower band was removed using a 1ml syringe and the EtBr removed by repeated butanol extractions (using water saturated butanol). The salts were removed by dialysis in 2 X 500ml 1 X TE. The DNA was then ready for use.

Mini DNA preparation using the method of Holmes and Quigley (1981): 1.5ml of an overnight culture containing the plasmid of interest was harvested by centrifugation in a 1.5ml eppendorf tube and resuspended in 350ul of STET buffer. 25ul of STET buffer containing lysozyme at a concentration of 10mg/ml was added and the tube vortexed briefly. This solution was boiled for 40sec and centrifuged in an eppendorf microfuge for 15min at 4°C. The pellet was discarded using a toothpick and 40ul of 3M NaAc and 400ul of cold isopropanol was added, followed by microcentrifugation for 7min which prcipitated the nucleic acid. The pellet was washed twice in 70% ethanol and dried briefly in a vacuum drier before being resuspended in 20ul - 50ul 1 X TE. This DNA was suitable for digestion and other <u>in vitro</u> manipulations.

2.17 Chromosomal DNA preparation. 100ml of stationary phase culture was harvested by centrifugation (12000g, 10min, 4°C) and resuspended in 3.3ml 25% sucrose. 0.67ml of a 5mg/ml lysozyme solution in 0.25M Tris-HCl pH8.0 was added and the mixture left on ice for 5min. 1.3ml of 0.25M EDTA was added followed by 5.3ml of 2% Sarkosyl (made up in 50mM Tris-HCl pH8.0 and 50mM EDTA pH8.0). Lysis was allowed to proceed for 5min on ice. Treatment of the lysed cells with RNase (final concentration 50ug/ml) and then with Proteinase K (final concentration 50ug/ml) broke down the RNA and proteins. The DNA was purified initially by repeated phenol extractions until the interface was clear. The addition of 1.5ml of 5M NaCl to every 20ml of solution prior to phenol extraction helped clear the interface. Chloroform and ether extractions removed any traces of phenol or chloroform. The DNA was precipitated using ethanol, washed in 70% ethanol and further purified by banding in a CsCl/EtBr gradient (see above for this procedure).

2.18 Ethanol precipitation of DNA. The DNA solution was made 0.3M in NaOAc and 2 volumes of cold ethanol added. After mixing, the DNA was precipitated by cooling on ice for up to 1 hour and pelleted by centrifugation (27000g, 15min, 4°C for large volumes or 12000g, 15min, 4°C for small volumes in eppendorf centrifuge tubes). The pellet was usually washed in 70% ethanol and dried briefly in a vacuum drier.

2.19 Restriction of DNA. Restrictions were usually performed in a total volume of 20ul containing between 0.25ug and 1ug of DNA, 2ul of 10X restriction buffer and 1 unit/ug DNA of enzyme, the volume being made up with distilled water. For larger scale restrictions the volumes were scaled up accordingly. The reactions were allowed to proceed for 1 to 2 hours at the appropriate temperature. Reactions were stopped either by the addition of gel loading buffer or by rapid heating to 70°C for 5 min followed by rapid cooling on ice.

2.20 Ligation of DNA fragments. The restriction fragments to be ligated were mixed such that the insert was in 3 times excess over the vector (10 times excess for blunt end ligations) and made up to 20ul by the addition of 2ul 10 X ligation buffer, 2ul 4mM ATP and distilled water. T4 DNA ligase was added (0.01 units/ug DNA for 'sticky' end ligation and 1 unit/ug DNA for 'Blunt' end ligation) and the solution

ligated for 1 hour at room temperature (overnight at 16°C for 'blunt' end ligations). Aliquots of the ligation mix were used to transform competent cells.

2.21 Calf Intestinal Phosphatase (CIP) treatment. To increase cloning efficiency CIP was used to remove the 5'-terminal phosphate groups from the linearised vector to prevent recircularisation of the vector. This technique is particularly useful if there is no direct selection for the insert. CIP works in high, medium or low salt buffers and was used at a concentration of 1-2 units/ug DNA. It was added directly to the restriction digest for the final 15min of the incubation and was heat killed in the manner described for restriction enzymes.

2.22 Single colony gel analysis. This technique enables the plasmid content of a colony to be observed without the purification of plasmid DNA. A single colony was patched out (1cm square) and grown overnight. The patch was scraped off the plate using a toothpick and suspended in 150 - 250ul of single colony gel buffer. The cells were left to lyse at room temperature and centrifuged in a microfuge (12000g, 4°C) for at least 15min. 30ul of the supernatant were loaded onto an agarose gel which did not contain ethidium bromide.

2.23 Rapid whole-cell protein preparation for SDS-PAGE: Take 400ul of late log phase culture and harvest cells by microfugation (20 secs). Remove all but 5-10ul of supernatant, then vortex thoroughly to generate a thick cell slurry. Add 80ul of Protein loading buffer (with 5% beta-mercaptoethanol added immediately prior to use). Vortex thoroughly, then incubate in a boiling water bath, 5-10 mins. Load 20-30ul of sample; the remainder may be stored (-20°C) for future use.

2.24 Gel electrophoresis. Both vertical and horizontal agarose gels were used. Vertical gels were used mainly for single colony analysis while horizontal agarose gels were used for restriction analysis of plasmids and for southern analysis. Unless otherwise stated 1% agarose gels were used.

Horizontal gels. Two types were commonly run.

(1) 100ml gels - 100ml of molten agarose was poured into a 11 X 19cm

perspex gel former with a 13 space teflon well former. After the gel had set, the comb was removed and the gel placed in a horizontal gel tank, submerged in E buffer and loaded with 20 - 30ul of sample. Gels were usually run overnight at 20V and stained in ethidium bromide (0.5ug/ml) for 30mins. The stained DNA was photographed on a 254nM wavelength UV transilluminator. Some gels were made up with ethidium bromide to a concentration of 0.5ug/ml.

(2) 200ml gels - These gels were made by pouring 200ml of molten agar into a 16.5 X 23cm gel former with a 20 space well former. The gels were run in E buffer in a gel tank with a buffering capacity of 3 litres overnight at 20V. The gels were stained with ethidium bromide (0.5ug/ml) and visualised on a 254nM wavelength transilluminator.

Vertical gels: The gel kits held two 16 X 15cm glass plates separated by 3mm spacers. After sealing the edges with molten agar the agar was precooled to 55^{o}C and poured between the glass plates. The comb (10 or 15 teeth) was inserted and the gel allowed to set. The top and bottom wells were filled with E buffer and the comb removed. Samples were loaded and the gel run, generally at 5 V/cm for about 4 hours prior to staining in 0.5ug/ml ethidium bromide. The gel was photographed on a 254nM wavelength transilluminator.

Vertical gels were used mainly for single colony analysis and the interpretation of the DNA band seen followed that of Dugaiczyk et al (1975). The fastest migrating and generally most abundant band was the supercoiled plasmid monomer. Behind this ran the open cicular plasmid band often comigrating with the supercoiled plasmid dimer. Open circular dimers and other higher forms ran higher up the gel. Plasmid linears could sometimes be detected running between the supercoiled monomer and the open circle monomer. Sheared fragments of chromosomal DNA ran as a thick band toward the top of the gel. Large plasmids like R388 run above the chromosomal band.

Photographing of gels. Ethidium bromide stained gels were viewed on a 254nm UV transilluminator and photographed using Polaroid type 67 land film or using a Pentax 35mm SLR loaded with Ilford HP5 film. Both cameras were fitted with a Kodak Wratten filter No.9.

2.25 Extraction of DNA from agarose gels using GENECLEAN. After staining, the gel was placed on a long wave transilluminator (300nM - 360nM) and the band of interest excised. The agarose chip was added to 2-3 volumes (usually 800ul) of "NaI" soln', then heated to 45-55°C for a few minutes until the agarose had completely dissolved. 5ul of "Glassmilk" suspension was then added, rapidly mixed, then placed on ice for 5 mins. Spin (5 secs, microfuge), and carefully discard the supernatant. Wash pellet three times in ice-cold "New" soln' by adding 700ul, resuspending the pellet, re-spinning, and discarding the supernatant. After final wash, take care to remove all traces of New soln'. Elute DNA from glass beads by adding 20ul T.E (pH 8.0), and incubating at 50°C for 3 mins. Spin 10 secs to pellet glass beads, then carefully remove the supernatant. Re-spin the supernatant to ensure the complete removal of all glass beads. The DNA is now suitable for restriction, ligation, end-labeling etc.. The compositions of the "NaI", "Glassmilk", and "New" solutions are undisclosed.

2.26 Preparation of polypeptides produced from purified minicells. The plasmid to be examined was transformed into the minicell strain DS944. 200ml of L-Broth was inoculated with this transformed strain and grown upto stationary phase at 37°C. The minicells were easier to purify if the culture was in stationary phase though it did reduce the level of incorporation acheived. Partially purified minicells were obtained by low speed centrifugation (5mins, 2000g at room temperature) which preferentially pelleted the whole cells. The supernatant was carefully decanted and the minicells pelleted (16000g, 10min, 4°C) and resuspended in 1ml of BSG. This was loaded onto a 5%-20% sucrose BSG gradient, poured using a peristaltic pump and a gradient former. The sucrose gradient was centrifuged at 1600g for 30min at 4°C. The minicells form a creamy band in the top third of the gradient. These were carefully drawn-off using a pasteur pipette, and then harvested by centrifugation (16000g, 10min, room temperature), and resuspended in 0.5ml BSG. The sucrose gradient purification step was then repeated. After the second sucrose gradient the minicells were harvested and resuspended in about 600ul of M9 minimal media supplemented with glucose and vitamin B1. The minicells were split into 2 aliquots and IPTG added to one at a final concentration of 0.1mM (to derepress the tac promoter) and incubated with shaking, 37°C, 20 mins. After incubation 20uCi of ^{35}S-methionine was added and

the minicells incubated at 37°C with shaking for a further hour. 1.5ml of L-Broth was added and the tubes incubated at 37°C for 30 minutes. The minicells were centrifuged at 16000g at room temperature for 5min and washed once in phage buffer prior to being lysed with about 50ul of protein sample buffer.

2.27 Protein Gel Electrophoresis: The electrophoresis of proteins followed the procedure of Laemmli (1970). Unless specifically stated all SDS-polyacrylamide gels were 10% running gel with a 4% stacking gel. The gel plates were separated by 0.8mm spacers and the gel was pre-sealed using molten 0.6% agarose in H_2O. The gels were poured according to the table below;

	Running Gel	Stacking Gel
Acrylamide/Bis (30 ; 0.8)	10ml (10%)	2ml (4%)
Tris-HCl	11.25ml (1M; pH8.8)	1.9ml (1M; pH6.8)
10% SDS	0.3ml	0.15ml
TEMED	0.01ml	0.01ml
APS (made fresh; 100mg/ml)	0.15ml	0.15ml
dH_2O	8.3ml (10%)	10.9ml (4%)

The running gel was poured leaving about 1.5-2.0cm between the top of the running gel and the bottom of the comb. Isopropanol was layered on top of the running gel to create a sharp interphase. Once the running gel had set, the stacking gel was poured after removing the isopropanol and washing away any residue with distilled water. The comb was pushed into the stacking gel firmly. After the gel had set the comb was removed, and running buffer poured into the gel tank covering the wells. Residue acrylamide was washed out of the slots using running buffer prior to loading the samples. The gel was run overnight at about 6mA until the bromophenol blue dye had reached the bottom of the gel.

The gel was then either stained with Coomassie to visualise protein bands, or prepared for autoradiography as follows; the gel was fixed in 30% methanol/10% glacial acetic acid for 1 hour before being soaked in EN^{3}HANCE for a further hour. The EN^{3}HANCE was removed and the gel soaked in water for a further 30 minutes before being dried in a Biorad vacuum drier at 80°C for at least 1 hour. The dried gel was placed against Blue sensitive film (Kodak Xomat S) and kept in the

dark at -70°C for a variable time (up to 1 month) to allow exposure of the film.

2.28 Southern hybridisation. Hybridisation of labelled DNA to filter bound chromosomal or plasmid borne DNA followed the procedure of Southern (1975) as modified by Reed and Mann (1985). Horizontal agarose gels were run with the required samples. The gel was photographed after ethidium bromide staining. The gel was soaked in 0.25M HCl for 15 minutes, soaked in 0.4M NaOH for 15 minutes, then placed on a glass plate covered by 2 layers of 3MM paper which overhung the edges and dipped into a tray containing the transfer buffer (0.4M NaOH). Pall Biodyne Nylon membrane cut to the size of the gel was pre-soaked in water and placed on top of the gel ensuring that no air bubbles were trapped. This was covered by further layers of 3MM paper and by paper towels until an 8cm stack was formed. The stack was topped with a glass plate weighed down with a 500ml bottle of water. Transfer was allowed to proceed overnight. The nylon membrane was briefly washed in 2 X SSC and was then ready for pre-hybridisation and hybridisation.

Pre-hybridisation was carried out using a solution of 1.5 X SSC, 0.5% Blotto (w/v non fat powdered milk) and 1% (w/v) SDS. Heat denatured sheared salmon sperm DNA was added at a concentration of 100ug/ml. About 1ml per 4cm^2 of filter was added, sealed in a plastic bag, and gently agitated for 1 hour at 65°C.

Hybridisation was also carried out at 65°C and in the same solution. The probe was labelled by nick translation according to Rigby *et al* (1977) and denatured by boiling for 5 minutes in 1% SDS and 1.5 X SSC and chilling on ice. The probe was added to the filter by opening up the bag and resealing. The solution was mixed by agitation prior to incubating with gentle agitation overnight at 65°C.

The filter was washed in wash buffer sucessivily at room temperature, 42°C, 55°C and 65°C until no, or very few, counts were found in the wash buffer. The filter was sealed wet in a plastic bag and autoradiographed for a variable length of time depending on the number of counts retained on the filter.

2.29 Nick Translation of DNA. Labelling of plasmid DNA with ^{32}P followed the procedure of Rigby et al (1977). The nick translation reaction was set up according to the following table:

10 X Nick translation Buffer	5ul
DNA	1ug
unlabelled dNTP's	1ul of 1mM stock
[gamma - ^{32}P] dATP	25uCi (2.5ul)
dH_2O	up to 44ul

This solution was chilled to 0°C

A 10^{-4} dilution of a 1mg/ml DNase I solution was prepared in ice cold nick translation buffer and 50% glycerol (stored at -20°C). 0.5ul of this solution was added to the nick translation reaction along with 5 units of E.coli DNA polymerase I. The reaction was incubated at 16°C for 1 hour and stopped with 2ul 0.5M EDTA pH8.0.

The unincorporated counts were separated from the labelled probe using a spun minicolumn (Maniatis et al, 1982). Sephadex G-50 was prepared in 1 X TE buffer and used to fill a 1ml syringe, one end of which had been blocked using siliconised glass wool. The syringe was spun for 4 minutes at 1500g at 4°C. 100ul of TE was added to the column which was centrifuged at 1500g for 4 minutes at 4°C. THe nick translation reaction was diluted to 100ul with TE and spun through the column in the same manner. The void volume was collected in a 1.5ml eppendorf tube and contained the incorporated label. The unincorporated counts remained in the column.

2.30 DNA sequencing: All DNA sequencing reactions were performed on single-stranded M13 templates using the di-deoxy chain termination technique. All template preparations, solutions and reaction conditions were as described in the "M13 Cloning/dideoxy sequencing Instruction Manual", published by Bethesda Research Laboratories with the following exceptions:

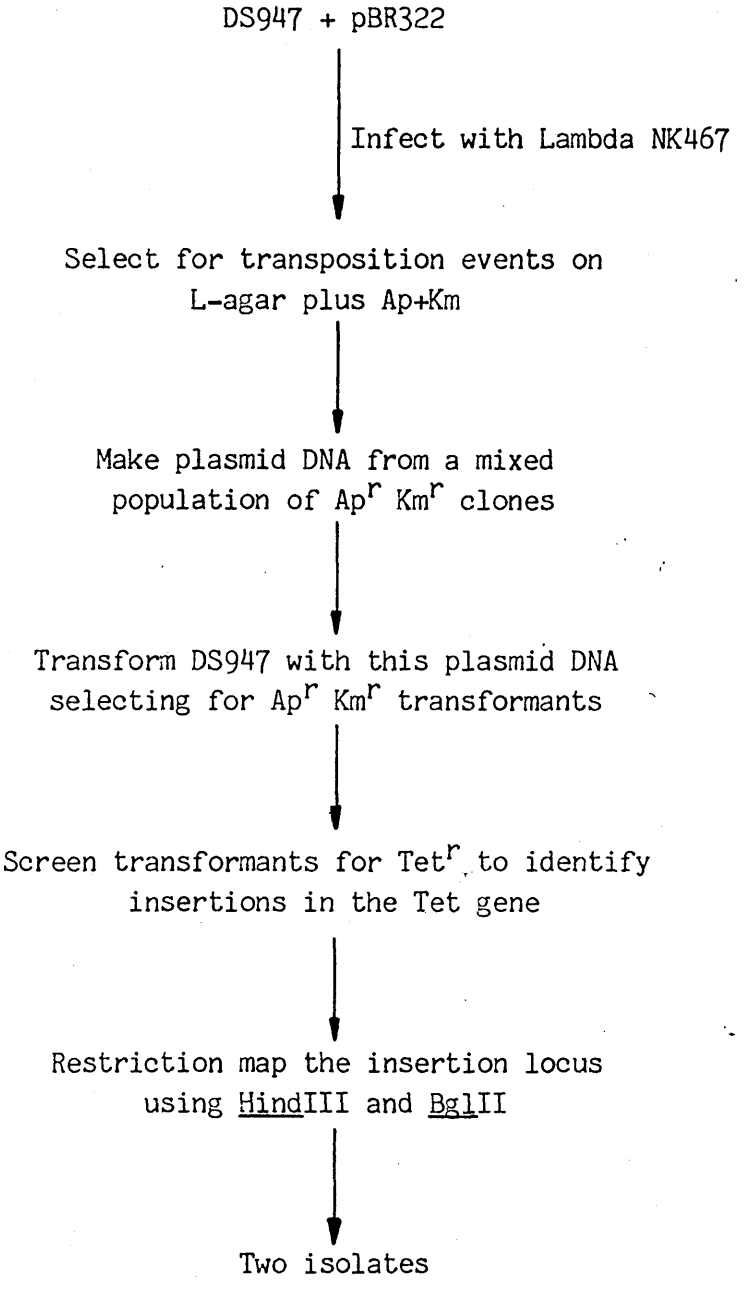
1. Working dideoxynucleotide concentrations:

0.25mM	ddATP
0.25mM	ddCTP
0.5 mM	ddGTP
1.0 mM	ddTTP

2. Final reaction volume was reduced to eliminate wastage.
   The primer/template hybridisation mix volume was reduced to 8ul
   from 12.4ul: composition = 5ul template (1ug)
   2ul M13 sequencing primer (4ng)
   1ul 10x polymerase buffer

3. Samples were electrophoresed on 8% polyacrylamide/urea gels as described in the BRL manual. When necessary, gels were prepared in 30% formamide, this has the advantage of resolving regions of extreme secondary structure in the DNA sequence.

2.31 Generation of pBR322::Tn5 derivatives: pBR322::Tn5 derivatives were generated by <u>in vivo</u> transposition of Tn5 from lambda NK467 into pBR322. The procedure for the selection of such events is illustrated in figure 2.1; and the structures of two independent isolates pCS80 and pCS81 are indicated in figure 2.2. For lambda NK467 description and infection conditions, see chapter 3.



pCS80 (10.2Kb) - Encoding resistance to Ap, Km and Tet
pCS81 (10.2Kb) - Encoding resistance to Ap and Km

FIGURE 2.1 Flow diagram illustrating the procedure used in the selection of the pBR322::Tn5 derivatives pCS80 & pCS81.

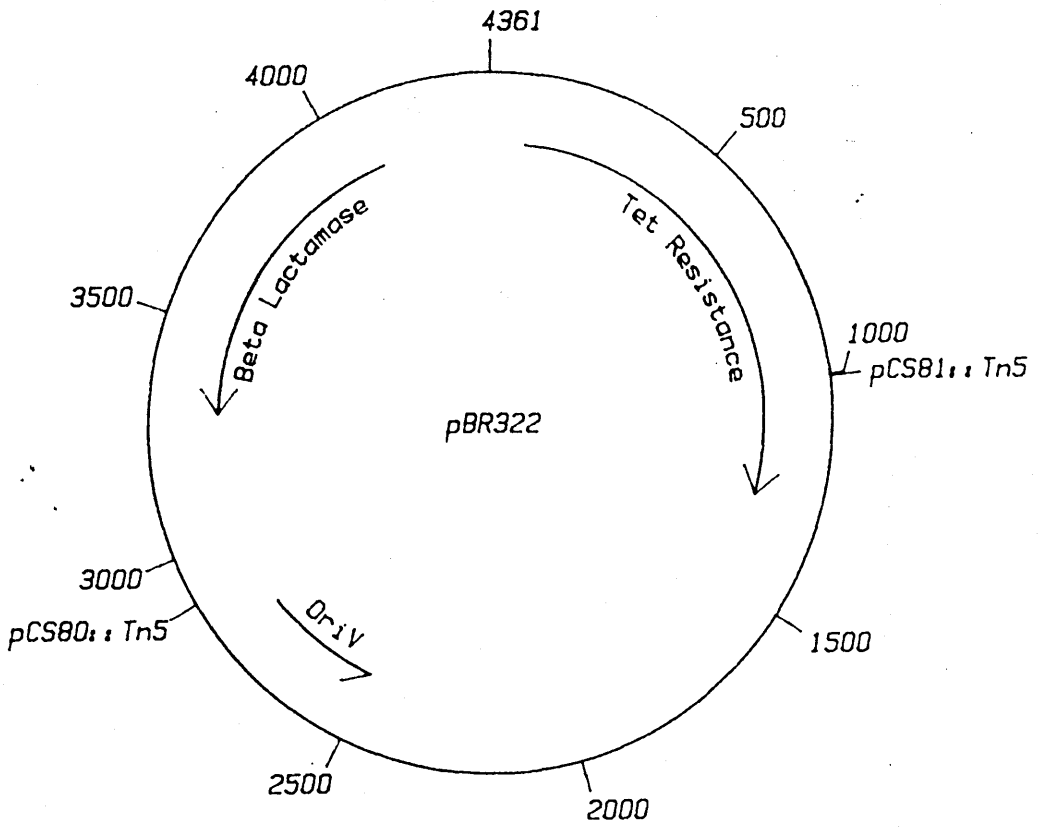


Figure 2.2 Structures of the pBR322::Tn5 Derivatives pCS80 & pCS81
The site of Tn5 insertion in pCS80 (2.9Kb), and pCS81 (1Kb), are indicated. The two plasmids are 10.2Kb in size; both encode resistance to Ap and Km, whilst pCS80 also encodes Tet^r.

CHAPTER 3

THE ISOLATION OF E. coli xer MUTANTS

3.1 INTRODUCTION

The only ColE1 sequence necessary for *cer* specific recombination, is the 279bp *cer* site itself (Sherratt *et al*, 1984). This sequence has limited coding potential and is incapable of encoding a polypeptide of similar size to known site-specific recombination enzymes. This fact led to the suggestion that the *cer*-specific recombinase might be host encoded, at a locus designated *xer* (Summers *et al*, 1985). The existence of such host functions was first demonstrated by Gill Stewart, who isolated two independent mutants defective in *cer*-specific recombination (Xer$^-$). However, whether these mutants correspond to one or more complementation groups, and indeed, whether either represents a lesion in the recombinase gene itself, was undetermined. My first goal was therefore to increase the number of independently isolated *xer* mutants, such that a minimum number of *xer* complementation groups might ultimately be determined. Thus it was necessary to optimise both a mutagenesis protocol and an appropriate selection procedure.

3.2 CHOICE OF MUTAGEN

Since their relatively recent discovery, transposable elements have proven to be extremely useful tools in the field of bacterial genetics (For review see Kleckner *et al*, 1977). In particular the transposon Tn5 has been used extensively in genetic analyses in a variety of species, including *E. coli* (Kleckner *et al*, 1977; Berg & Berg, 1983). Tn5 is a composite transposon comprising two terminal 1535bp IS50 elements (IS50L and IS50R), flanking a 2.7Kbp central portion (Fig. 3.1; Auerswald *et al*, 1981; Beck *et al*, 1982; Mazodier *et al*, 1985). The central portion of Tn5 encodes the Neomycin phosphotransferase II (*NPTII*) gene whose expression confers resistance to the aminoglycoside antibiotics neomycin and kanamycin. In addition, two other genes confer resistance to CL990 (a member of the bleomycin family of antibiotics), and to streptomycin respectively. However the streptomycin resistance gene is cryptic in *E. coli* (Berg *et al*, 1982; Berg & Berg, 1983; Mazodier *et al*, 1985). Some of the properties of transposable elements as mutagens are listed below, with particular reference to Tn5.

1. Tn5 is ranked amongst those transposable elements with the lowest insertional specificities (Kleckner *et al*, 1977). It inserts into many different locations around the *E. coli* chromosome (Shaw & Berg,

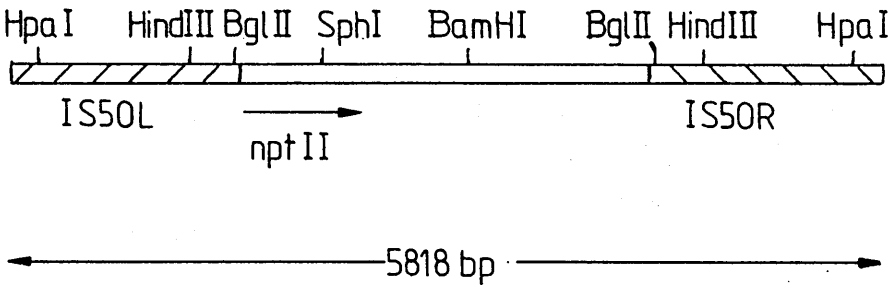


Figure 3.1 Diagrammatic Representation of Tn5

1979), and into a many different sites within particular genes (Miller *et al*, 1980; De Bruijn *et al*, 1983). Nonetheless, Tn5 does exhibit a bias towards insertion in certain target sites, so-called "Hotspots". However, of those hotspots studied in detail, no target site consensus sequence is apparent. In the case of two such hotspots in the plasmid pBR322, short stretches of sequence with limited homology to IS50 ends are present in close proximity. It has been suggested that these homologous sequences in the target DNA may play some role in promoting insertions in the nearby hotspot (Berg & Berg, 1983; De Bruijn & Lupski, 1983).

2. Insertion of a transposable element into a gene usually results in the complete inactivation of that gene, such mutations being generally non-leaky and stable. Reversion of *lacZ*::Tn5 insertion mutants occurs at frequencies of approximately 10^{-6} per generation (Berg, 1977), such reversion being the result of a transposase-independent excision event (Egner & Berg, 1981). Secondary transposition (transposase dependent) of the Tn5 element, giving rise to two chromosomal copies, can occur at frequencies of up to 0.1-1% (Berg, 1977). Clearly this phenomenon can complicate the interpretation of subsequent results. The frequency of secondary IS50 transpositions is less readily measured, but is of reduced significance.

3. Tn5 insertions into multigenic operons usually exert highly polar effects on downstream genes, such polarity being independent of the relative orientation of Tn5 insertion. Indeed such polarity has been exploited in determining the genetic organisation of a number of operons, for example, the complex *nif* gene cluster of *K. pneumoniae* (Merrick *et al*, 1980). In a minority of cases reported, Tn5 insertion does not exert a strict polar suppression of downstream genes (Berg *et al*, 1980, De Bruijn *et al*, 1983). Some heterogeneity in polar affects is not at all unexpected.

4. Transposon mutagenesis results in the introduction of a novel physical and genetic marker at the insertion locus. This has a number of very considerable advantages over classically induced genetic lesions, for example;

a) Insertion of the *NPTII* gene permits the pre-selection of a randomly mutagenised population prior to screening for a desired

mutant. This can considerably reduce the number of clones which have to be screened.

b) The inserted transposon sequence, with its resistance determinants, provide markers which facilitate both physical and genetic linkage analyses. One can also exploit the 100% linkage between a mutant allele and its inserted copy of Tn5 in the process of strain construction, a process which can be extended to the co-transduction of neighbouring alleles (Kleckner *et al*, 1977).

c) An insertion locus of interest can be readily cloned by virtue of its linkage to a selectable transposon marker.

5. Finally, the use of Tn5 as a mutagen is aided by the availability of highly engineered, transposon delivery vectors. These vectors can greatly increase the ease and efficiency with which mutant populations are generated.

3.2.2 LAMBDA NK467 as a Tn5 Delivery vector

Ideally, one would like to introduce a selectable transposable element on a "suicide vector", that is, one which fails to replicate in the host strain, such that when subsequently selecting for a transposon marker, one recovers clones arising predominantly from transposition events. The vector lambda NK467 satisfies this criterion. This vector, a derivative of phage lambda, was constructed by N. Kleckner, and is reviewed in detail by De Bruijn & Lupski (1983). The phage genotype is as follows; delta b221, *rex*::Tn5, *O*am29, *P*am80. The b221 deletion removes the *attP* site along with the phage *int* and *xis* genes (Davidson & Szybalski, 1971). This phage is therefore unable to efficiently integrate into the host genome. The *O* an *P* genes are both essential for vegetative phage replication, therefore lambda NK467 can propagate only in an amber suppressing strain. Consequently when this phage infects any lambda^{S}, Su^{O} strain, it is unable to replicate or integrate. Thus the vast majority of Km^{r} clones which arise upon selection, must be the result of Tn5 transposition events.

3.3.1 Selection of xer mutants

My selection of xer mutants was based on the same principle as previously described (chapter one); that is, the transformation of mutant populations with a 2-cer pseudodimeric plasmid, the resolution of which results in the loss of a selectable marker. However the

plasmid employed previously for this purpose, pKS441 (Fig 1.6), resulted in a surprisingly leaky selection. It was observed that transformation of pKS441 into a wildtype host gave rise to transformants which were resistant to both ampicillin and chloramphenicol. However, when plasmid DNA from such transformants was analysed by agarose gel electrophoresis, the parental form (pKS441) was undetectable with only its resolution product, p442, being present (G. Stewart, and my own observations (data not shown)). Therefore it is clear that when using pKS441, chloramphenicol resistance (Cm^r), does not correlate well with the Xer phenotype. In order to explain the persistent chloramphenicol resistance observed, one is forced to conclude that the cat gene is being replicated, albeit perhaps at some low level. The maintenance of the cat gene on a replicon might come about in a number of ways, for example;

a) A low level of undetected pKS441 may still be present in the cells. This could represent either plasmid which is yet to be resolved, or indeed, which is unresolvable. Alternatively it may represent a steady state concentration of pKS441 arising from the cer mediated fusion of p442 and p443.

b)Conversely, p443 may be replicated by becoming integrated, perhaps transiently, into the host's chromosome.

c) Finally, the small circular form, p443, may be capable of a low level of autonomous replication.

Each of these possible explanations implies that the cat gene dosage would be considerably reduced, and that the Cm^r phenotype presumably depends upon high levels of expression from the few remaining copies of cat. I therefore decided to investigate the potential of an alternative test plasmid, pKS455, which is identical to pKS441, except that it carries a promoterless copy of the cat gene from pACYC184 (Fig. 3.2). We presume that expression of cat in this construct is dependent on transcription known to emanate from within the cer site (D. Summers, pers. comm.). With a reduced level of cat gene transcription one might expect that phenotypic expression of chloramphenicol resistance might require high gene dosage. Table 3.1 illustrates the results obtained when pKS455 is transformed into either an Xer^- strain (CSX10), or into a wildtype strain DS947. The observation that CSX10 transformed with pKS455 is unable to grow under chloramphenicol selection immediately after transformation, but is able to do so after a period of growth, may be explained in terms of

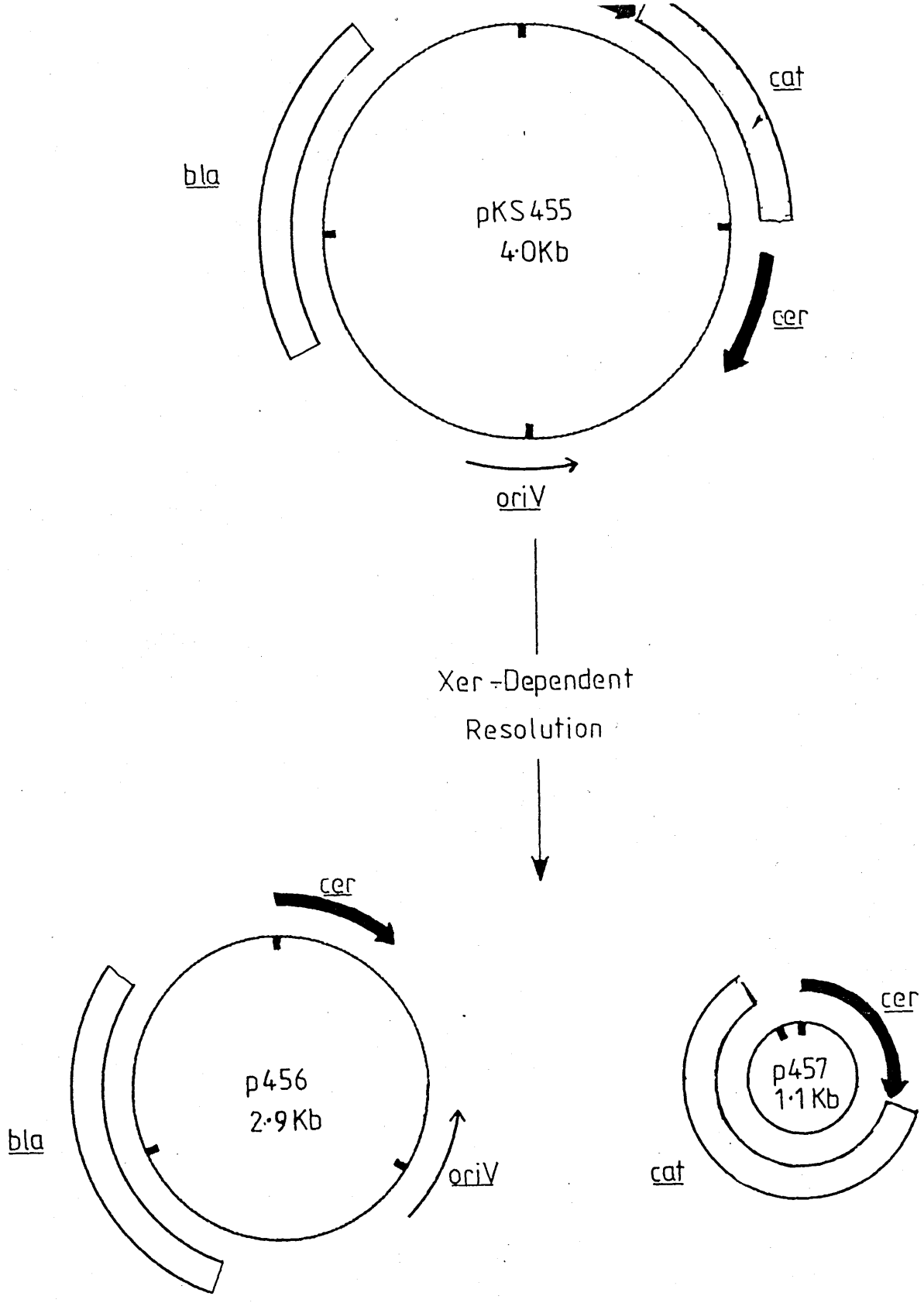


Figure 3.2 Diagrammatic representation of the Xer-Dependent Resolution of the 2-*cer* reporter plasmid pKS455

TABLE 3.1 Transformation of DS947 or CSX10 With pKS455

STRAIN	Antibiotic Selection	Transformants.ml^{-1} of Transformation Mix	Percentage of clones which grow when replica plated onto Cm
DS947	Ap	3×10^3	<0.04
"	Cm	0	-
CSX10	Ap	1×10^4	100
"	Cm	0	-

Strains DS947 and CSX10 were made competent as described in materials and methods, after which they were transformed with pKS455, and equal aliquots plated onto L-agar plus either Ap (50 mg.l^{-1}), or Cm (30mg.l^{-1}). After overnight incubation at 37^{o}C, colonies were counted and the number of transformants per ml calculated. The Ap^r colonies from each transformation were then replica-plated onto L-agar plus Cm (30mg.ml^{-1}), incubated overnight (37^{o}C), then the proportion of growing colonies scored.

TABLE 3.2 Frequency of Tn5 Transposition From lambda NK467

Strain	p.f.u. of NK467	Total N^{o} of Km^r clones	Frequency of Transposition per p.f.u. of NK467
DS947	2×10^9	2.2×10^5	1.1×10^{-4}
"	0	0	-

gene dosage. It seems unlikely that significant plasmid DNA replication will take place until a transformed cell has recovered, and re-initiated its cell cycle. It may then take a number of generations before the plasmid copy number is such that the low level of *cat* gene expression is sufficient to confer a Cm^r phenotype.

Clearly pKS455 produces a more reliable relationship between Cm^r and the Xer phenotype of a transformed strain; provided that strain is permitted a short period of growth prior to selection. Accordingly, this plasmid became the centrepiece of my mutant isolation strategy. A flow diagram depicting this strategy is shown in Figure 3.3.

3.3.2. GENERATION OF MUTANT POPULATIONS

Large populations of *E. coli*::Tn5 insertion mutants were generated as follows: To approximately $1x10^{11}$ stationary phase cells, of either CSH26 or DS947, were added $2x10^9$ plaque forming units (p.f.u.) of lambda NK467. The resultant culture was incubated statically for two hours, to permit expression of Km^r. The total number of viable Km^r clones generated within this population was then determined by plating small aliquots onto L-agar+ Km. The frequency of transposition is then expressed as the number of Km^r clones recovered per infectious phage particle added (Table 3.2). The results shown in Table 3.2 were highly reproducible, indicating that large populations ($2x10^5$) of independent Tn5 insertion mutants can be generated with ease. A mutant population of this size could, in principle, saturate the *E. coli* chromosome with an insertion every 15bp or so.

In order that the mutagenised population could then be transformed with pKS455, it had to be amplified. This was done by inoculating the entire primary population (10^{11} cells of which $2x10^5$ are Km^r mutants) into 100mls of L-broth plus Km, then growing this culture overnight. From this culture a 0.5ml inoculum (approximately $2.5x10^9$ Km^r cells) was subcultured and made competent for transformation. The final number of Km^r cells present in the transformation protocol was determined to be approximately $2.5x10^{10}$. This population represents the progeny from an initial inoculum of $2x10^5$ independent mutants. If one assumes that each mutant clone is fairly represented, then each mutant will be present as approximately $1x10^5$ siblings. Is this sufficient to ensure the transformation of at least one sibling of each?

In order to answer this question I calculated the proportion of

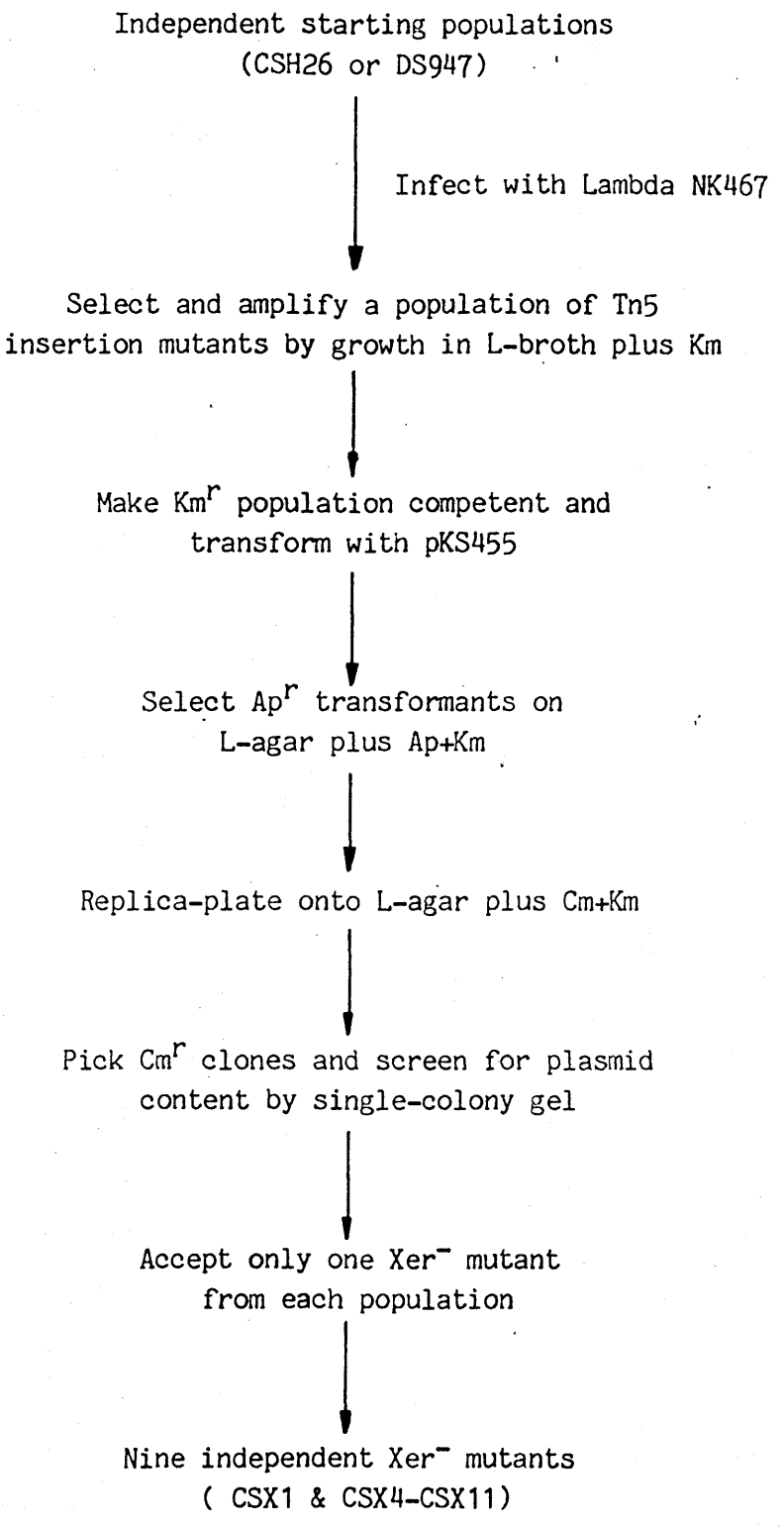


FIGURE 3.3 Flow diagram illustrating the mutant isolation strategy

cells made competent by my standard protocol (see Materials and Methods). This was determined by transforming a small aliquot (0.1ml) of competent cells with an excess of plasmid (pKS455) DNA, then plating those cells onto selective and non-selective media to determine the relative viable counts (Table 3.3). The slightly better results obtained with CSH26 over DS947 correlates with the generally fitter phenotype of the former. Using this information one can calculate that when transforming a population comprising $2x10^5$ mutant clones, each present as $1x10^5$ siblings, then one can hope to transform 45 siblings of each mutant clone. This calculation is based on two major assumptions:

1. That all viable Tn5 insertion mutants are of equal fitness, thus maintaining an even representation within the population throughout growth.
2. That each mutant clone can be transformed with an efficiency equal to that of the parental strain.

Whilst these assumptions are patently untenable, the calculations based upon them nontheless provide a useful guide to the scale of the experiment. I should add that one could circumvent the requirement for population amplification after mutagenesis by introducing the 2-cer test cartridge via either sex factor mating, or phage infection, both of these routes of genetic transfer being capable of achieving extremely high efficiencies. Such a system would be of great advantage in screening for any Xer^- mutants which exhibit a significant reduction in fitness. As indicated in Fig. 3.3, the transformed populations were then plated onto L-agar+Km+Ap, followed by replica plating onto L-agar +km+Ap+Cm to select for Cm^r mutants. The proportions of Cm^r clones recovered from DS947 and from two independent DS947::Tn5 populations are shown in table 3.4.

3.4.1 CHARACTERISATION OF Cm^r MUTANTS

As a consequence of the selection procedure, one would expect to recover two major classes of Cm^r mutants:

1. Chromosomal Xer mutants
2. Plasmid mutants (including both cer mutants and plasmid rearrangements).

In order to distinguish between these possibilities, plasmid DNA was recovered from Cm^r clones then retransformed into the parental (Xer^+) strain. In those cases where the recovered plasmid was resolved,

TABLE 3.3 Percentage of Cells Made Competent

Strain	Viable Count Prior to Transformation (c.f.u. ml^{-1})	N^{o} of Ap^{r} Transformants	% of Input Cells Transformed
CSH26	5.4×10^{9}	3.8×10^{6}	0.07
DS947	4.8×10^{9}	2.0×10^{6}	0.04

TABLE 3.4 Proportion of Cm^{r} Clones Isolated After Transformation of a Population of DS947::Tn5 Mutants With pKS455

Population	N^{o} of Ap^{r} clones	N^{o} of Ap^{r} Cm^{r} clones	Frequency of Cm^{r} clones
DS947::Tn5 *1	3.4×10^{4}	14	4.1×10^{-4}
DS947::Tn5 *2	3.0×10^{4}	20	6.7×10^{-4}
DS947	2.1×10^{4}	4	1.9×10^{-4}

DS947::Tn5 *1 and *2 represent two independently mutagenised populations of DS947.

giving rise to Cm^S transformants carrying only p456, it was assumed to have come from a *bona fide* *xer* mutant; by way of examples, the results of such analyses are presented for four mutants identified (CSX3, CSX4, CSX5, & CSX11; Fig.3.4). This definition will have precluded any *xer* mutants which so affect the 2-*cer* substrate as to render it unresolvable. Using this criterion I have isolated a total of nine independent *xer* mutants, each from quite seperate populations of Tn5 insertion mutants (Table 3.5).

3.4.2 COTRANSDUCTION OF THE *xer* LESION WITH Tn5

In order to determine whether or not any of the newly isolated *xer* mutants were caused by transposon insertion, I carried out a series of P1 transduction analyses. If an *xer* mutation is the direct result of a Tn5 insertion, then one would expect to observe a high degree of linkage between the *xer* mutation and the Tn5 NPTII gene. Whilst in principle the two ought to display 100% linkage, secondary transposition events can result in apparently reduced linkage.

Generalised transducing lysates were prepared by growing P1*kc* on various independent mutants (Lennox, 1955). Transductions were performed as described in Materials and Methods, with Tn5 transductants being selected by their resistance to Km. These Km^r transductants were then transformed with the 2-*cer* reporter plasmid pCS202 (see Fig. 3.6). Transformants were selected on chloramphenicol (the pCS202 maintenance marker), then screened for their resistance to tetracycline (the pCS202 deletion marker). Plasmid DNA was also recovered from these transformants and analysed by agarose gel electrophoresis (Fig. 3.7). Those transductants which, when transformed with pCS202, were Tet^r and carried a plasmid comigrating with pCS202 were scored as being Xer^-. Those which were Tet^S and carried a plasmid comigrating with the pCS202 deletion derivative, pCS203 (see Fig. 3.6), were scored as being Xer^+. The frequency of cotransduction of Xer^- with Km^r for several independent mutants was calculated in this way (Table 3.6). From the results in table 3.6, one can conclude that there is a copy of Tn5 tightly linked with each *xer* lesion tested. In the case of the *xerA*3 allele, the data strongly suggest that this lesion is a direct result of transposon insertion into the *xerA* gene. Insufficient data are available to make a strong case for causal insertion in the *xerB*2, *xerB*3, *xerB*4 and *xerA*5 alleles. The 85% cotransduction observed between Tn5 and *xerB*1 might reflect transposon

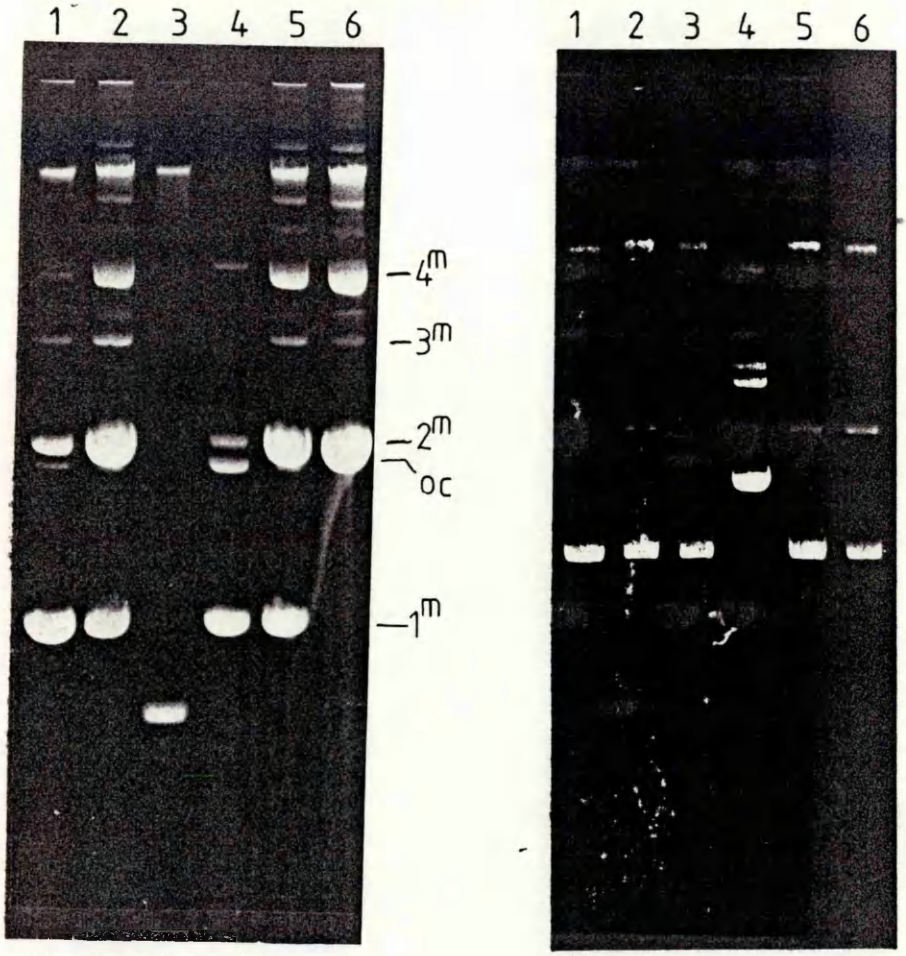


Figure 3.4 The recovery of plasmid DNA from xer mutants and its subsequent resolution in a wild-type strain. Mini-prep's of plasmid DNA were prepared from CSX3, CSX4, CSX5 & CSX11; all of which were found to contain various multimeric forms of pKS455. (a) CSX3, lane 1; CSX4, lane 2; p456 marker, lane3; pKS455 marker, lane 4; CSX5, lane 5; CSX11, lane 6. However, after transformation of these DNA prep's (pCSX3, pCSX4, pCSX5 & pCSX11) into DS947, plasmid DNA was again recovered and found to contain the resolution product p456. (b) pCSX3, lane 1; pCSX4, lane 2; p456 marker, lane 3; pKS455 marker, lane 4; pCSX5, lane 5; pCSX11, lane 6.

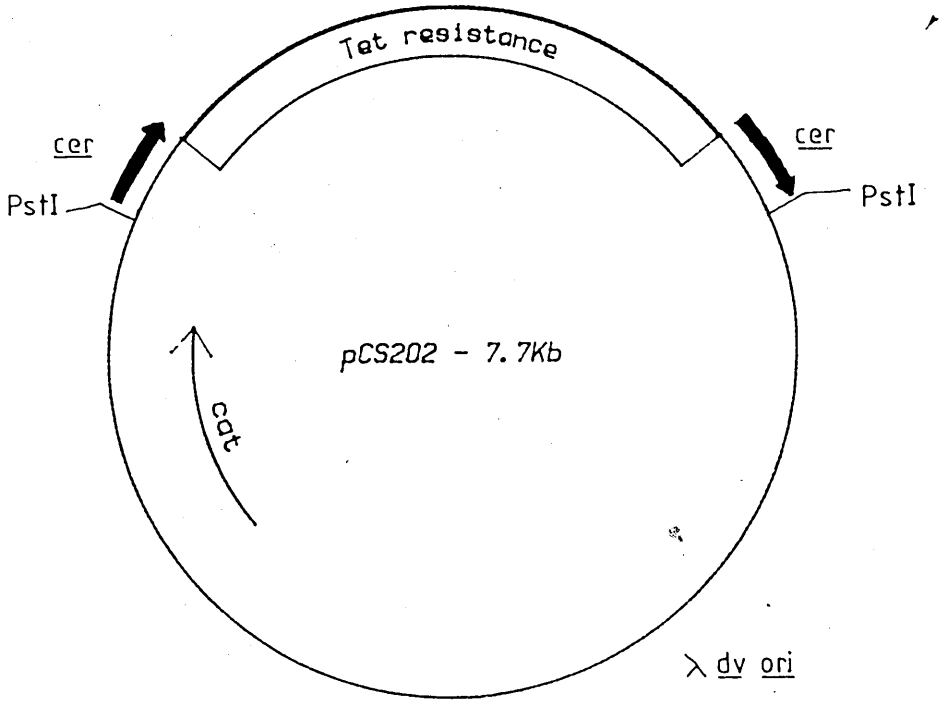


Figure 3.6 Structure of the 2-cer reporter plasmid pCS202: pCS202 was constructed by cloning the 2.7Kb cer-tet-cer "cartridge" as a PstI fragment (from pCS201) into the unique PstI site in the vector pCB101. pCS202 encodes resistance to both Cm and Tet when maintained in an Xer⁻ host. However, Xer-dependent resolution of pCS202 deletes the tet gene to produce pCS203 (5.3Kb; encoding resistance to Cm).

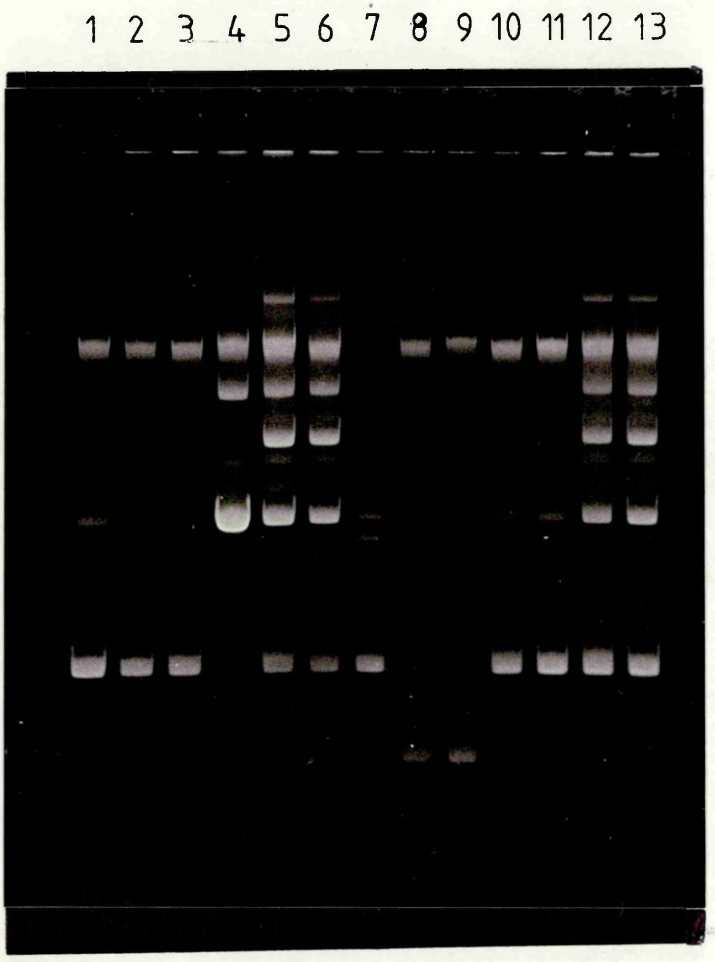


Figure 3.7 Co-transduction of the xerA3 allele with Tn5. Using a P1-lysate from CSX1, Km^r was transduced into DS903. Ten independent transductants were transformed with pCS202, then plasmid DNA recovered and analysed on a single colony gel; these ten clones are shown in lanes 1-6 and 10-13 (incl.). Markers are pCS202, lane 7; pCS203, lanes 8 + 9.

TABLE 3.5 Derivation and Nomenclature of Xer⁻ Mutants

Parental strain	Mutagenised Population	Xer⁻ Mutant Strain Name	Mutant Allele
CSH26	CSH26::Tn5 *3	CSX1	xerA3
DS947	DS947::Tn5 *1	CSX4	xerB1
"	" " *2	CSX5	xerB2
"	" " *3	CSX6	xerB3
"	" " *4	CSX7	xerB4
"	" " *5	CSX8	xerB5
"	" " *6	CSX9	xerB6
"	" " *7	CSX10	xerB7
"	" " *8	CSX11	xerA5

The strains listed above are the original isolates of each independent mutant; the original lab-names for each, plus all their subsequent derivatives are listed in Appendix I.

TABLE 3.6 Frequencies of Co-transduction of Xer⁻ with Km^r

P1kc Lysate of Strain	xer Lesion	Nº of Km^r Transductants	Nº of Xer^-/Km^r Co-transductants	Co-transduction frequency
CSX1	xerA3	20	20	1.0
" 4	xerB1	26	22	0.85
" 5	xerB2	2	2	1.0
" 7	xerB4	1	1	1.0
" 8	xerB5	3	3	1.0
" 11	xerA5	1	1	1.0

insertion at a locus tightly linked to a spontaneous xerB lesion. If this were the case, the genetic distance between the two loci would have to be less than 0.5 mins (Taylor & Trotter, 1967); the maximum distance between loci which can be cotransduced by phage P1 being approximately 1.8 mins (Miller, 1972). Alternatively, xerB1 might be the genuine result of Tn5 insertion, but the degree of linkage observed may have been reduced by a low level of secondary transposition events. Subsequent data, derived from the cloning of both the xerB and xerB1 alleles, provides compelling evidence that this xer lesion does carry an inserted Tn5. It seems likely that secondary transpositions will only have a significant affect on the co-transduction frequency when they have occurred in the donor population prior to making the transducing lysate. Indeed, I have since made a fresh lysate from CSX1 which co-transduces Xer^- with Km^r at a frequency of only 50%. This clearly illustrates the major disadvantage of this type of mutagenesis.

3.4.3 CAN ColE1 COMPLEMENT THE RECOMBINATION DEFICENCY IN xer MUTANTS?

The 279bp cer site represents the only ColE1 sequence necessary, in an Xer^+ host, to promote multimer resolution and enhance plasmid stability (Summers & Sherratt, 1984, & D.K. Summers, pers. comm.). However, previous attempts to identify any plasmid encoded recombinase may have been thwarted by the existence of complementing chromosomal functions. If so, one might expect ColE1, in turn, to complement xer mutations. This has not been observed. Wildtype ColE1 has been shown to multimerise in an xerA3 strain, CSX1, whereas it is predominantly monomeric in the isogenic parent CSH26. Furthermore, ColE1 is rapidly lost from CSX1, but is stably maintaind in CSH26 (Stewart, 1986). This latter observation also demonstrates that xer functions play a role in ColE1 stability, thus corroborating the observed correlation between monomerisation and plasmid stability (Summers & Sherratt, 1984).

The observation that ColE1 is rapidly lost from an Xer^- strain is an intriguing one. It had previously been assumed that ColE1 might exert a positive selection on its own maintenance within a population, via colicin E1 production. Colicin production is a lethal event, but is normally initiated in only a small fraction of the population, a situation somewhat analagous to phage bursts within a lysogenic population (Durkacz et al, 1974). A level of colicin is thus built

up, within the environment, which is lethal to non-immune cells. ColE1 encodes reciprocal colicin immunity, and so ColE1$^+$ cells are presumed to have a selective advantage within the population. The rapid appearence of plasmid-free segregants within an Xer$^-$ population might suggest that colicin production, in itself, cannot serve to stably maintain ColE1. The possibility exists that all such segregants are in fact colicin E1 resistant mutants. Such resistant strains can arise as a result of mutation of the cell surface receptor for colicin. The sensitivity of our plasmid free segregants to colicin E1 has not been determined. Furthermore, I must add that the level of colicin production from an Xer$^-$, ColE1$^+$ host has not been determined. It is possible that significantly reduced levels of colicin are produced by an Xer$^-$ population, either as a direct consequence of reduced *cea* gene expression or perhaps indirectly as a result of the rapid loss of ColE1 from the population.

3.4.4. OTHER PHENOTYPES OF Xer$^-$ STRAINS

ColE1 has a number of close relatives, identified by DNA sequence homology, and overall genome organisation. These include ColK, pMB1, ColA, and CloDF13. The CloDF13 *crl* site, and the ColK *ckr* site, have been shown to be functionally analagous to *cer*, promoting both monomerisation and increased stability (Hakkaart *et al*, 1984, Summers *et al*, 1985). The *crl* and *ckr* sites exhibit considerable sequence homology to *cer*, centered around a 150bp "core" of maximal homology. This core homology is maintained by sites in both pMB1 and ColA whose recombinational properties have yet to be investigated (See Fig. 1.2). The sequence homology led us to ask whether or not the ColK and CloDF13 sites also depend upon host *xer* functions. Evidence suggests that indeed they do. It has been shown that *crl* and *ckr* multimers are not recombined in either *xerA* or *xerB* mutants (G. Stewart, 1986; D. Summers, pers. comm.). Furthermore, *crl* and *ckr* plasmid derivatives which exhibit very stable phenotypes in wildtype strains, are highly unstable in the *xerB* mutant DSX1 (Stewart, 1986).

No other phenotype of *xer* mutants has yet been observed. They exhibit indistinguishable growth rates from their isogenic parents (Stewart, 1986), and appear quite normal with respect to colony and cellular morphology.

3.5 DISCUSSION AND CONCLUSIONS

Previous evidence has demonstrated that the 279bp *cer* site is the only ColE1 sequence required to facilitate multimer resolution. This monomerisation activity is correlated with, and indeed indistinguishable from, the role of *cer* as a plasmid stability determinant. There is to date, no direct evidence for any trans acting products of the *cer* locus. Nontheless it seems highly probable that *cer* is transcribed, at least in part, *in vivo*. It has been suggested that transcription of the *crl* locus is essential to its function as a recombination site (Hakkaart *et al*, 1984). Whether this represents a requirement for transription *per se*, or the need for a functional *crl*-encoded RNA or peptide, is unclear. A role for transcription in the stimulation of recombination at the *HOT1* site in the *S. cerevisiae* rRNA gene cluster has recently been reported by Voekel-Meiman *et al* (1987).

The existence of an *E. coli* K12 chromosomal function, essential to *cer* site-specific recombination, was first demonstrated by Gill Stewart, who isolated two spontaneous *xerB* mutants DSX1 and DSX2. Using a similar selection principle, but with a more reliable system, I have isolated a further nine independent *xer* mutants. Of these nine *xer* mutations, six have been tested for their genetic linkage to the transposable element, Tn5. In all six cases a high degree of linkage was observed (*via* P1 cotransduction), indicating that in each case a copy of Tn5 is inserted at, or very near, the mutant *xer* locus. These closely linked , perhaps causal, Tn5 insertions can now be exploited in a number of ways.

1. In the process of strain construction, these insertions have proven useful in the transduction of mutant *xer* alleles into a variety of genetic backgrounds (Appendix 1).
2. They also provide a route by which one can readily clone the mutant *xer* loci, *via* their linkage to a selectable Tn5 marker. The cloned chromosomal sequences then provide important information regarding the structure of the *xer* locus. Furthermore, the cloned *xer* sequence can be used as a probe to identify clones, within libraries, which carry a wild type copy of that gene. This approach to the cloning of wild type genes, is invaluable in the absence of a geneticselection for the wild type clone.
3. In principle, these insertions also provide a very useful means with which to map the *xer* lesions on the *E. coli* chromosome. This,

however, proved to be extremely problematic. The mapping of *xer* genes is discussed in Chapter 4.

These mutants provide a base from which to launch a detailed investigation of host genes, and gene products involved in *cer* specific recombination. I decided to concentrate on this analysis, rather than continue isolating more mutants. Clearly the 11 idependent mutants isolated to date do not represent a large enough sample, to give us confidence that all *xer* complementation groups are represented. Therefore an important extension to this work would be a more rigorous mutant isolation programme; this would ideally include the use of alternative mutagens. Furthermore, the development of a strategy capable of detecting conditionally lethal *xer* mutants is crucial if we are to identify *xer* genes whose products are essential to host viability.

All *xer* mutants isolated to date have been selected by virtue of their ability to maintain a 2-*cer* plasmid in its multi-*cer* state. Whilst we interpret this as being indicative of the loss of the resolution activity, there is an alternative explaination. It is plausible that *cer*-specific resolution is unaffected in our mutants, but that the reverse reaction, that of *cer* mediated fusion, is enhanced. Such fusions could be catalysed by either a site-specific, or homologous, recombination mechanism. One might then argue that the *xer* functions identified, are in fact merely negative regulators of the fusion pathway. This possibility is seriously undermined by the observation that *cer*-mediated fusions are infact inhibited in an Xer^- strain. It has been shown that two co-existing cer^+ plasmids (which carry no sequence homology outwith *cer*), will fuse *in vivo*, forming a "cointegrate" structure, present at only very low levels. Cointegrate formation occurs independently of RecA, but is dependent on XerA, with no fusions being detected in an $XerA^-$ strain (Summers, pers. comm.). It would therefore appear probable that both the fusion, and resolution reactions are debilitated by the same *xer* lesion.

Finally, it is interesting to note that we have observed no phenotype of *xer* mutants other than those associated with episomal site-specific recombination and concomitant plasmid stability. Indeed, the *xer* functions are, apparently, the host's contribution to the

major stability determinant of ColE1 and its relatives. It seems probable that these functions perform some other role in *E. coli*, the exposition of which may shed considerable light on the nature and evolutionary origin of the *cer*/*xer* interaction.

CHAPTER 4

CLONING THE *xerA* GENE

4.1 INTRODUCTION

In order to carry out a detailed investigation of the *xer* genes identified by our mutant selection procedure, it was decided to clone them in plasmid vectors. Initially no positive selection was available for the Xer^+ phenotype, therefore the availability of an *xerA*::Tn5 insertion mutant, *xerA3*, was central to the cloning strategy. This strategy involved firstly the cloning of the mutant *xerA3* allele, exploiting its linkage to the *nptII* gene of Tn5. Restriction information derived from the cloned chromosomal DNA flanking this insertion locus was then used to facilitate the cloning of the wildtype *xerA* gene.

4.2.1 CLONING THE xerA3 MUTANT ALLELE FROM CSX1

It has been shown that the *xerA3* allele from strain CSX1 exhibits 100% linkage to Tn5 in P1 transductional mapping (Table 3.6). Hence, this particular *xer* lesion must have Tn5 inserted either in, or very near, the *xerA* gene. Therefore by cloning out the Tn5 element plus flanking chromosomal DNA, one can expect to recover either the disrupted *xerA* gene sequence, or chromosomal DNA from a locus tightly linked to *xerA*.

To this end, chromosomal DNA made from CSX1 was cut to completion with the restriction endonuclease *Eco*RI, for which there are no recognition sites within Tn5. The resultant digest was ligated to pBR322 vector DNA which had previously been cut with *Eco*RI and 5'-dephosphorylated with CIP. The resultant ligation mix was transformed into the $RecA^-$ strain DS902, and Ap^r, Km^r transformants selected on L-agar containing the appropriate antibiotics. In this way the 19.4Kb plasmid, pCS300, was isolated.

4.3 RESTRICTION MAPPING OF pCS300

The restriction map of pCS300 (Fig. 4.1) was derived via a relatively standard approach. Various restriction digests of pCS300, were size fractionated by gel electrophoresis through either agarose or polyacrylamide gels, different percentages of each gel matrix being employed, where appropriate, to provide optimal resolution of fragments in different size ranges (Maniatis *et al*, 1982). The electrophoretic mobilities of all restriction fragments were measured by hand, as were those of marker fragments of known molecular weights; routinely lambda cI857, or pBR322 (Daniels *et al*, 1983; Sutcliffe,

1979). These mobilities were entered into the computer program ZGELFIT (A.C. Boyd, pers. comm.), together with the molecular weights of each standard fragment. This program then calculates the approximate molecular weight of each unknown fragment, by comparing its mobility to the spectrum of mobilities of standard fragments, according to the algorithm of Dugglesby, Kinns, & Rood (1981). This assumes that for each DNA fragment , the electrophoretic mobility is inversely proportional to the molecular weight. This particular algorithm has been shown, to be superior to methods which plot mobility versus log[molecular weight] (Boyd, 1982). Computer calculations of molecular weights were only accepted for fragments whose mobilities lay within the spectrum of standards available on each gel. Furthermore, the standard deviation between observed and expected molecular weights of the marker fragments (i.e. ZGELFIT output vs. ZGELFIT input) were calculated. This provides an estimate of the range within which, the relationship between molecular weight and moblilty is reliable for any given gel; only those unknowns whose sizes lay within this range were accepted.

The data generated in this way were used to assemble, on paper, the restriction map of pCS300 shown in figure 4.1. The logical construction of this map was greatly aided by the presence in the plasmid of two large stretches of known DNA sequence, namely pBR322 and Tn5 (complete restriction maps of these known sequences were generated using the automatic string search facility of the computer program ZSEARCH (Boyd. A.C. pers. comm.)). In cases where this information proved insufficient for the ordering of sites, various double restriction digests were also employed. Only a representative sample of those gels generated in the course of this restriction mapping are shown (Figs. 4.2, 4.3, & 4.4). The restriction endonucleases employed are listed in Table 4.1, together with a breakdown of the number of cuts each is observed to make in the various components of pCS300 (i.e. in pBR322, Tn5, or in the chromosomal DNA).

The restriction map of pCS300 indicates that it is approximately 19.4 Kb in size, comprising 4.3Kb of pBR322 DNA plus an insert of approximately 15.0Kb. This insert represents approximately 9.2Kb of E. coli chromosomal DNA punctuated by a centrally located copy of Tn5 (5.8Kb). This restriction map provides valuable information regarding the structure of the insertion locus.

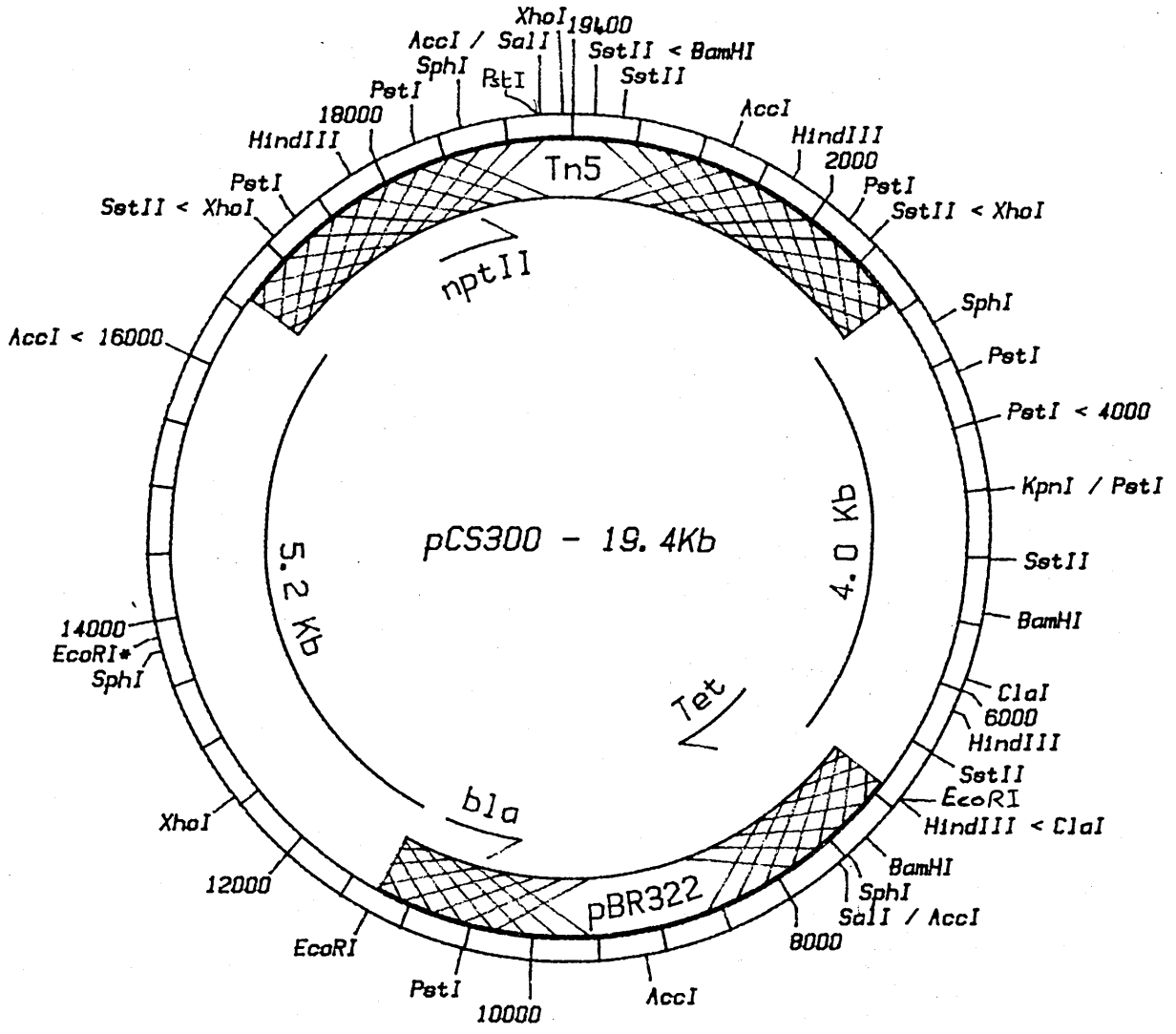


Figure 4.1 The Restriction Map of pCS300

The presence of a "/" between enzymes labels indicates that the two enzymes cut at the same site. The presence of a "<" between labels indicates that the enzymes cut at very closely linked sites and cannot be resolved on the map-scale shown: note that the order of the KpnI and PstI sites at 4500bp has not been resolved. Note the presence of a hyper-sensitive EcoRI* site at 13.9Kb. For a linear map of the pCS300 insert, see Fig. 4.9.

Table 4.1 The number of Recognition Sites Identified in pCS300 for Various Restriction Endonucleases

Enzyme	Nº of Cuts in pCS300	Predicted Nº of Sites in- pBR322	Tn5	Chromosomal DNA
AccI	5	2	2	1
BamHI	3	1	1	1
BglII	2	0	2	0
ClaI	2	1	0	1
EcoRI	2	1	0	2*
HindIII	4	1	2	1
HpaI	6	0	2	5
KpnI	1	0	0	1
NotI	2	0	2	0
PstI	8	1	4	3
SalI	2	1	1	0
SmaI	1	0	1	0
SphI	4	1	1	2
SstII	6	0	4	2
XbaI	0	0	0	0
XhoI	4	0	3	1

* The 2 EcoRI sites represent the boundaries between the vector sequences and the cloned insert.

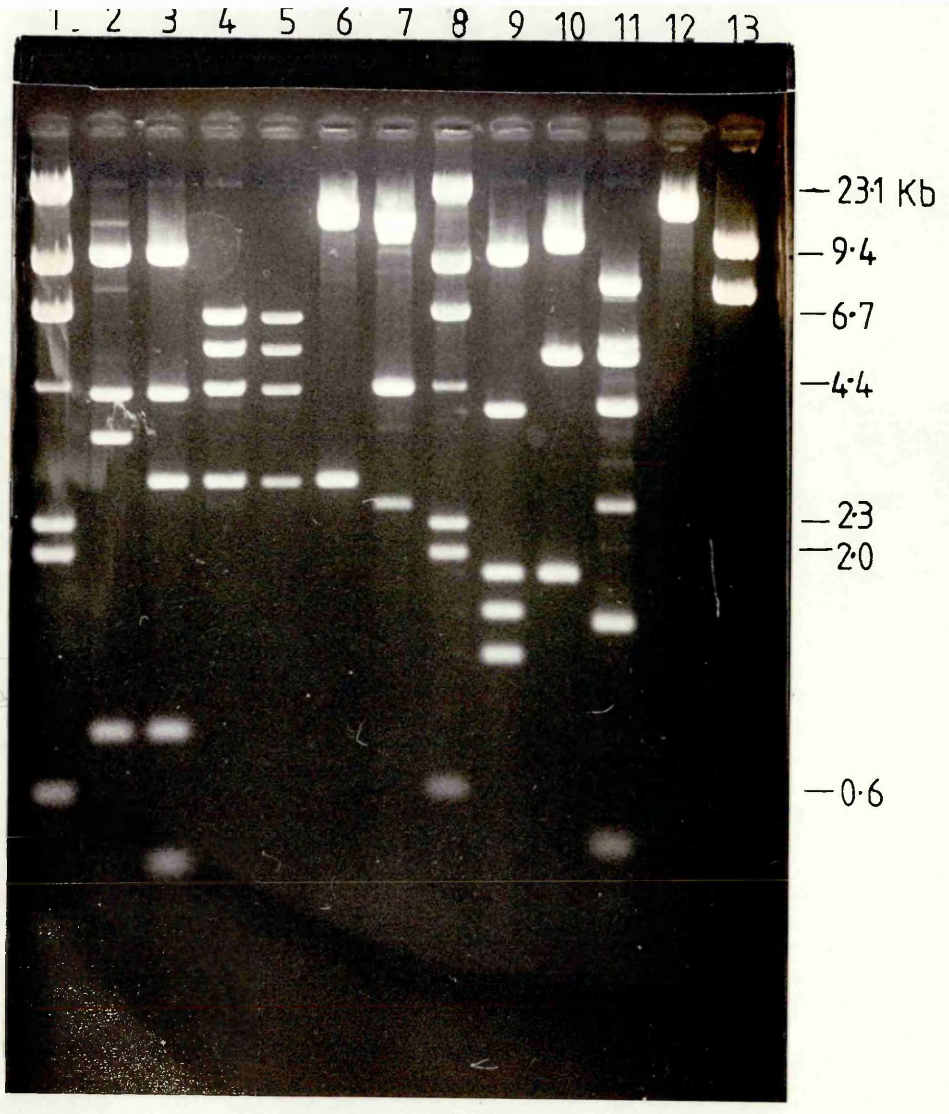


Figure 4.2 Restriction analysis of pCS300.

Lane	Sample	Lane	Sample
1	Lambda/HindIII	8	Lambda/HindIII
2	pCS300/HindIII	9	pCS300/BamHI/BglII
3	pCS300/HindIII/BglII	10	pCS300/BamHI
4	pCS300/EcoRI/BglII	11	pCS300/BamHI/EcoRI
5	" " "	12	pCS300/SmaI
6	pCS300/BglII	13	pCS300/SalI
7	pCS300/EcoRI		

Samples were run on a 0.8% agarose gel. Note the presence of two very prominent EcoRI* partial restriction products in lane 7 (2.5 and 12.6 Kb) and lane 11 (2.5 and 5.8kb). These are the result of a hypersensitive EcoRI* site located at co-ordinate 13.9Kb (see Fig. 4.1). This site has been only slightly restricted in the EcoRI/BglII digests (lanes 3 & 4) since extreme precautions were taken to limit the amount of EcoRI employed.

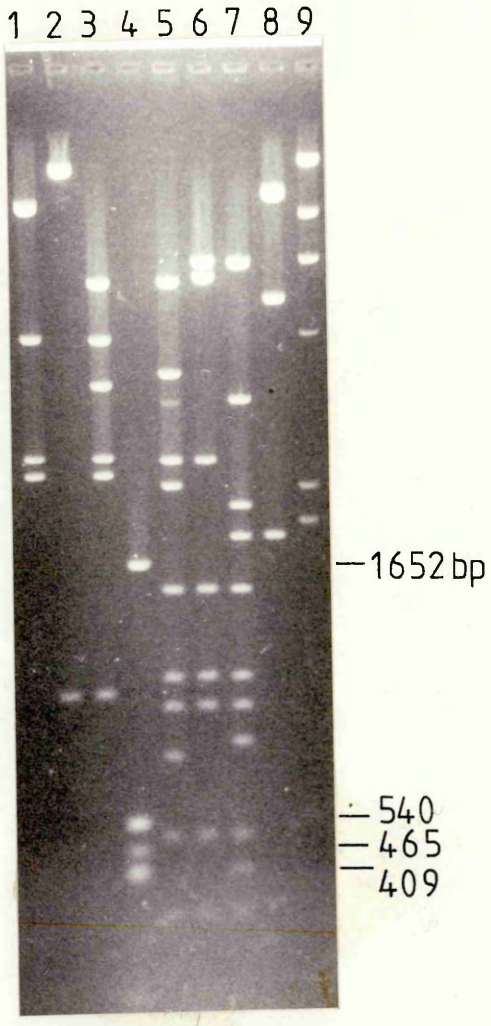


Figure 4.3 Restriction analysis of pCS300

Lane	Sample	Lane	Sample
1	pCS300/XhoI	6	pCS300/PstI
2	pCS300/ClaI	7	pCS300/BamHI/PstI
3	pCS300/XhoI/ClaI	8	pCS300/BamHI
4	pBR322/DdeI	9	Lambda/HindIII
5	pCS300/EcoRI/PstI		

Digests run on a 1% agarose gel.

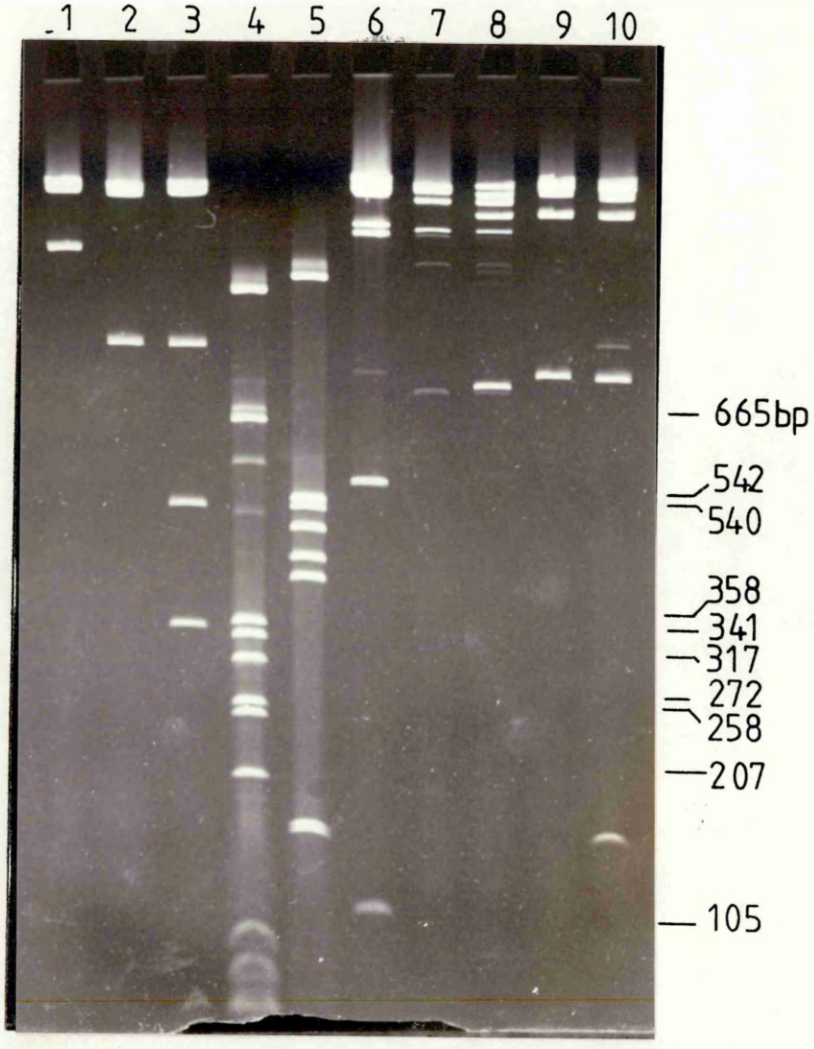


Figure 4.4 High resolution restriction analysis of pCS300

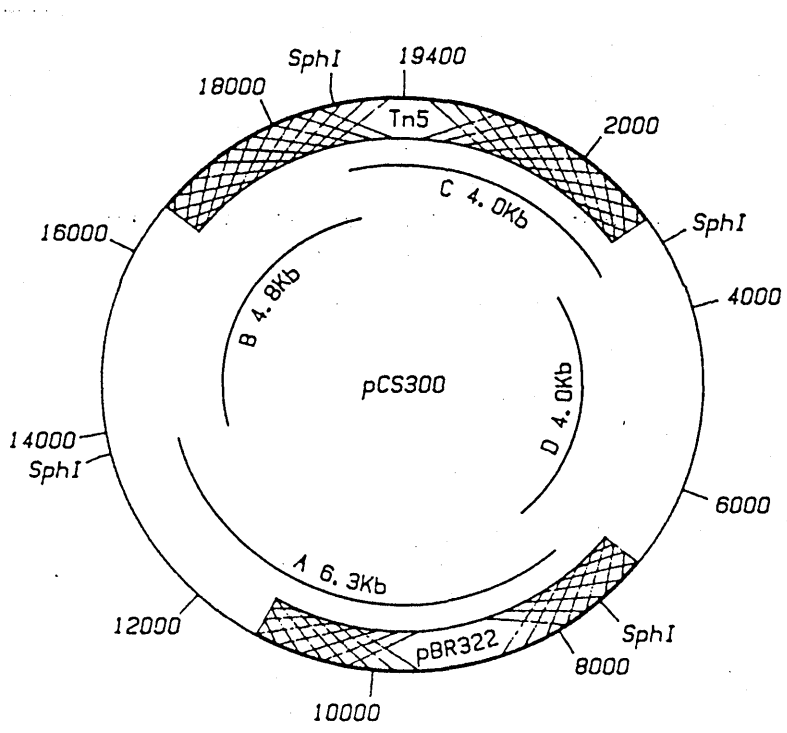
Lane	Sample	Lane	Sample
1	pCS300/BamHI	6	Lambda/HindIII
2	pCS300/ClaI	7	pCS300/EcoRI
3	pCS300/BamHI/ClaI	8	pCS300/EcoRI/HindIII
4	pBR322/SauA	9	pCS300/HindIII
5	pBR322/DdeI	10	pCS300/ClaI/HindIII

Digests run on a 3.5% polyacrylamide gel. The sizes of marker bands are given in base pairs.

4.4 DOES THE INSERT CLONED IN pCS300 CORRESPOND TO THE xerA3 ALLELE?

Despite the fact that P1 transduction data has identified a single Tn5 insertion, in CSX1, inextricably linked to the xerA3 allele, we cannot simply conclude that it is this locus which has been cloned in pCS300. It is inevitable that secondary transpositions will occur (Berg, 1977), giving rise to alternative Tn5 insertion loci in (at least initially), a small proportion of the population. It is plausible that the primary insertion locus is on an EcoRI fragment which is less readily cloned than one such secondary insertion locus; this might be simply due to the very large size of the fragment, or perhaps as a consequence of one, or more, of the genes that it carries being detrimental to host viability when present in multicopy. In this event, the sequence cloned in pCS300 might have proven to be a time consuming "Red herring". In order to obviate this possibility, I compared the flanking restriction patterns of the cloned Tn5 in pCS300, with those of the primary insertion, in situ, in the chromosomes of CSX1, and its xerA3 transduced derivatives, CSX3 and CSX14. This was achieved by cutting pCS300 and chromosomal DNA (made from CSX1, CSX3, and CSX14) with either SphI or BamHI; both BamHI and SphI cut once within Tn5 (at positions 3056 and 2084 respectively, Fig. 1.1), therefore each unique insertion will yield two highly characteristic Tn5 containing fragments. These digests were resolved on 0.7% agarose gels, then Southern blotted onto nylon membrane filters. The blots were then probed with radiolabelled Tn5 sequences generated by the nick-translation of pCS80; this plasmid is a pBR322::Tn5 derivative generated in vivo by transposition from lambda NK467 into pBR322, the procedure for selecting such events, and the structure of two pBR322::Tn5 isolates (pCS80 & pCS81) are shown in Figs 2.1 and 2.2 respectively).

On digestion with SphI , pCS300 gives rise to four fragments, A, B, C, and D, of approximately 6.3, 4.8, 4.0, and 4.0 Kb respectively (Fig. 4.6). Whilst all four fragments will hybridise to pCS80 sequences, clearly fragments C and D will co-migrate and therefore appear as a single band; note that only fragments B and C would be expected from a chromosomal digest of this same Tn5 insertion locus in situ. SphI digestion of CSX1, CSX3 and CSX14 chromosomal DNA generates, in each case, only two discrete SphI fragments indicative of the presence of a single major copy of Tn5. The two fragments detected in these chromosomal digests are approximately 4.8 and 4.0 Kb



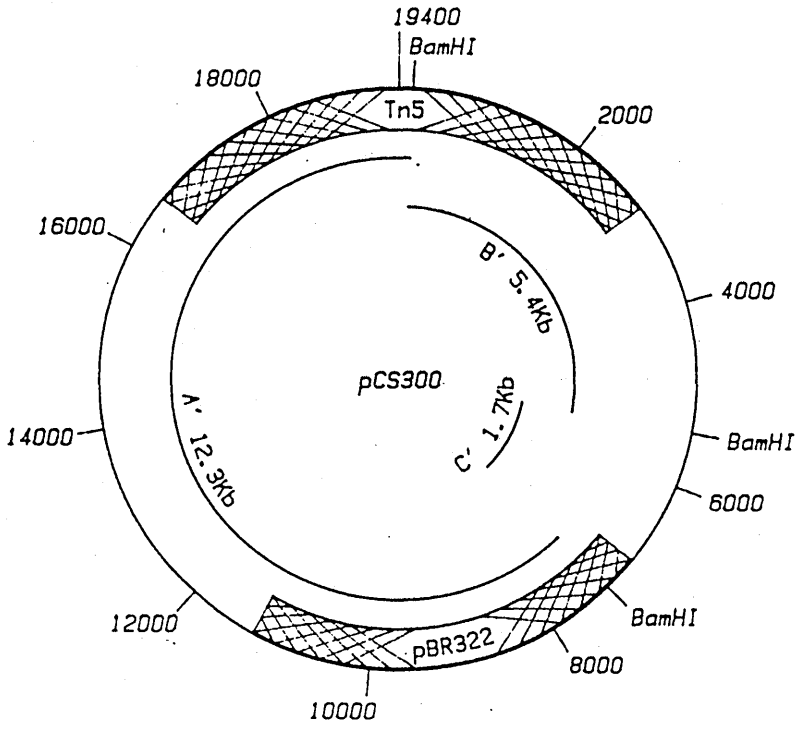


Figure 4.6 Restriction Patterns Generated by SphI or BamHI Digestion of pCS300

respectively. They co-migrate with fragments B and C from the pCS300/SphI digest, and are therefore indistinguishable from the restriction pattern predicted for the *in situ* chromosomal digestion of the Tn5-insertion locus which has been cloned in pCS300 (Fig. 4.7).

A similar analysis of BamHI restriction digests corroborates the SphI results. Upon digestion with BamHI, pCS300 yields three restriction fragments A', B', and C' of 12.3, 5.4 and 1.7Kb respectively (Fig. 4.6). As shown in figure 4.8, all three of these fragments hybridise to pCS80 DNA used as a probe (the relatively weak signal from fragment C' reflects the small amount of pBR322 sequence that it carries). As shown in figure 4.6, the insert cloned in pCS300 carries a BamHI site in the right hand arm of the flanking DNA, but it does not include the nearest chromosomal BamHI site to the left hand side of the insertion. From this one can predict that BamHI digestion of chromosomal DNA containing the corresponding Tn5 insertion *in situ*, will generate two fragments, one of 5.4Kb and the other of greater than or equal to 8.5Kb. The blot shown in figure 4.8 confirms that all three mutants contain a single copy of Tn5 (generating two bands which hybridise to the pCS80 probe), and that all three strains share the same major insertion locus. The sizes of the two bands generated are 5.4Kb and >12.3Kb. These compare with the predicted sizes, based on the pCS300 map, of 5.4 Kb and >8.5Kb.

From the combined SphI and BamHI restriction data presented, one can conclude that all three mutant strains, CSX1 and its two *xerA3* transduction derivatives, CSX3 and CSX14, all carry a single major primary copy of Tn5. Furthermore, that in all three strains, that Tn5 copy is present in an identical location which thus identifies it as the transposon genetically linked to the *xerA3* allele. Finally, the SphI and BamHI restriction map of the insertion locus cloned in pCS300, is entirely consistent with that insert being the *xerA3* linked locus from strain CSX1.

I should stress that the chromosomal bands hybridising to pCS80 sequences are assumed to be due to homology to Tn5, as opposed to the pBR322 portion of pCS80. The justification being that each strain is Km^r, and so must surely carry Tn5 sequences, however none of the strains contain episomal plasmid DNA. Indeed, whilst CSX1 had been cured of pKS455 (by a passive non-selective growth period), the transductants CSX3 and CSX14 had never been transformed with any plasmid whatsoever. Obviously had any signals been detected from the

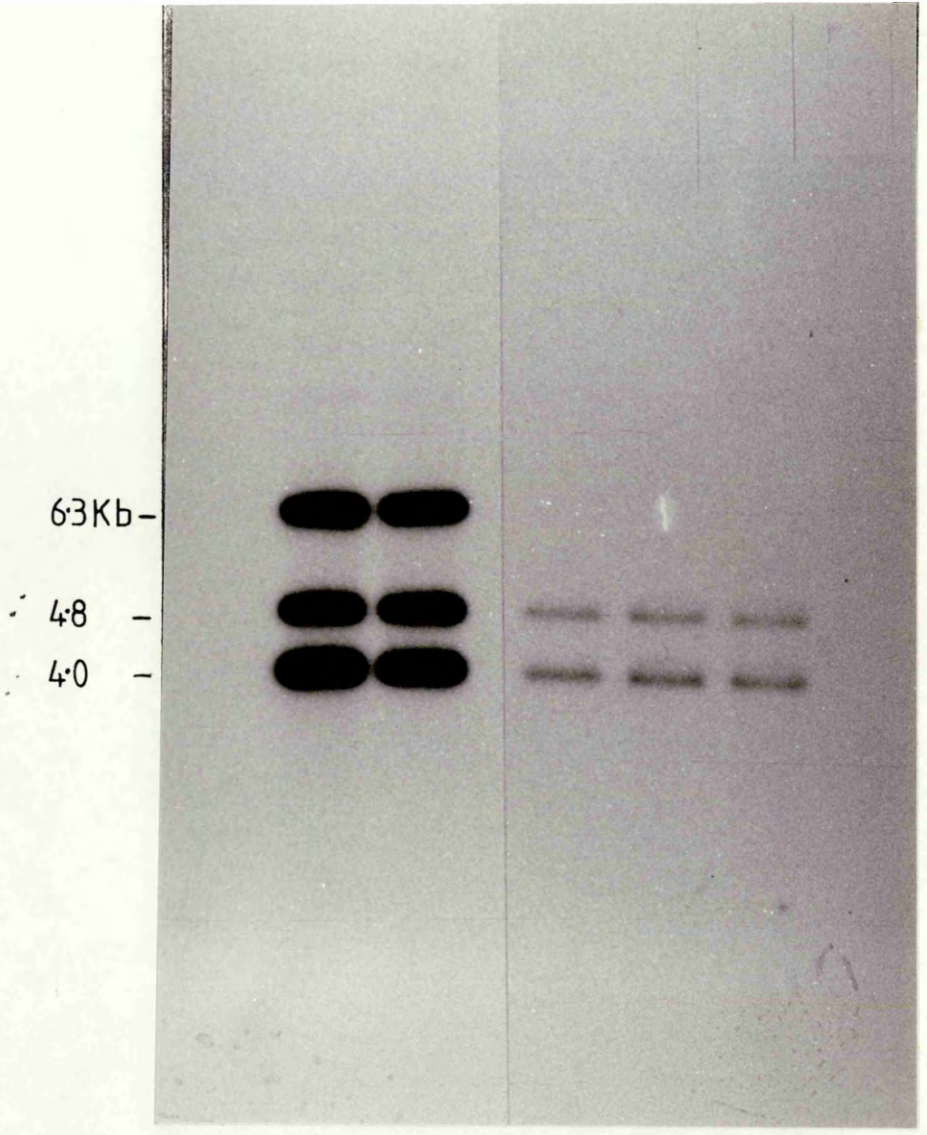


Figure 4.7 SphI-restricted chromosomal DNA from CSX1, CSX3, & CSX14, probed with nick-translated pCS80. The photograph presented above is a composite of two autoradiograms of different exposure times.

Lane	Sample
1	pCS300/SphI
2	" "
3	CSX1/SphI
4	CSX3/SphI
5	CSX14/SphI

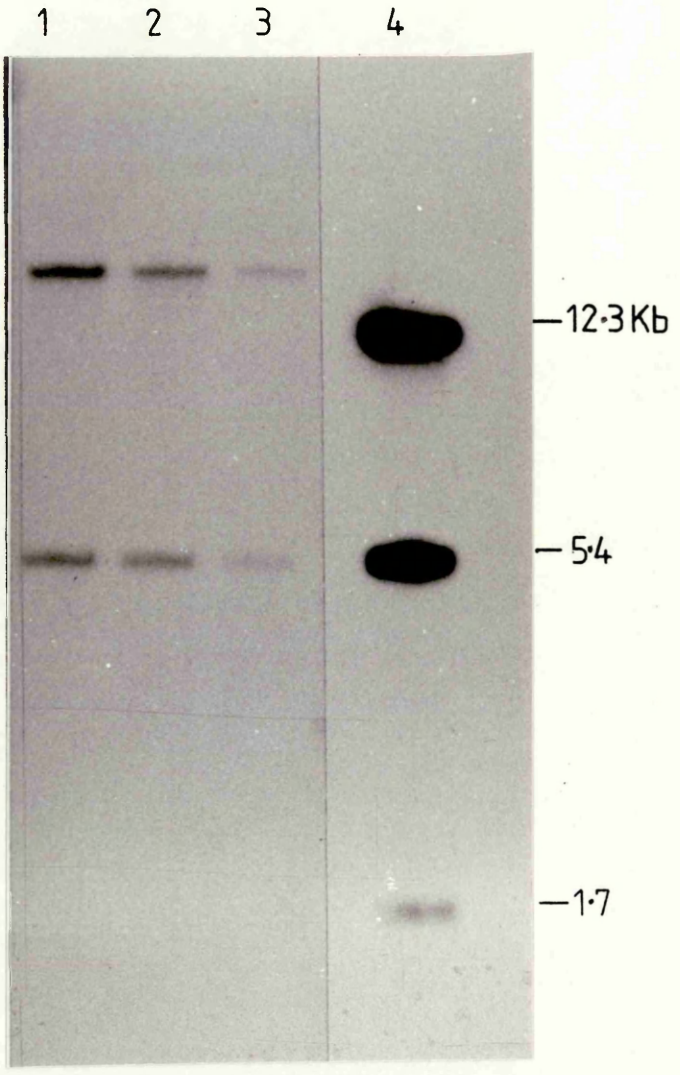


Figure 4.8 BamHI-restricted chromosomal DNA from CSX1, CSX3, & CSX14, probed with nick-translated pCS80 As in the case of Fig 4.7, the photograph shown above is a composite of two autoradiograms of different exposure times.

Lane	Sample
1	CSX1/BamHI
2	CSX3/BamHI
3	CSX14/BamHI
4	pCS300/BamHI

control $Tn5^-$ strain, or multiple signals from the $Tn5^+$ strains, then the blots would have been re-probed with labelled pBR322 DNA.

4.5 CLONING THE WILD-TYPE xerA GENE

The cloned xerA3 allele in pCS300 could have been used as a probe for homologous sequences in a genebank prepared from a wild-type E. coli strain (for example, in colony or plaque hybridisations). However, since 5.2Kb of chromosomal DNA to the left of the Tn5 insertion, plus 4.0Kb to the right, had been cloned with the mutant locus, it seemed likely that the entire xerA gene sequence would be present within the cloned EcoRI fragment; we therefore adopted a simpler, though less direct, approach to the cloning of the wild-type xerA gene.

Having cloned and restriction mapped the mutant xerA3 allele, we deduced the restriction map of the wild-type locus by simply deleting Tn5 from the pCS300 map (this assumes that no significant rearrangements have taken place at the target locus during, or after, Tn5 insertion). Thus by cloning size-selected restriction fragments from digests of wild-type chromosomal DNA, we aimed to create a "mini-genebank", greatly enriched for the specific sequence of interest. We chose to target in this way the predicted 8.45Kb EcoRI/HindIII fragment, which runs from the HindIII site 3.15Kb to the right of Tn5 in pCS300, to the EcoRI site 5.2Kb to its left (Fig. 4.1). This particular fragment has the advantage over the predicted 9.2Kb EcoRI fragment, of having heterologous ends, thus potentially enhancing the cloning efficiency.

Chromosomal DNA was prepared from strain DS902 (Xer^+), carrying F'4315 (Bachmann, B., pers. comm.). This F' (carrying approximately mins. 89-91) was initially believed to complement the xerA3 mutation, thus by inference, was thought to carry the xerA gene (Stewart, 1986). DNA was made from this strain in the belief that it was diploid for xerA, thus further enriching for the desired clone. However, subsequent attempts to repeat F'4315 complementation of xerA3 have failed (M. Burke and D. Sherratt, pers. comm.). The reasons for this anomaly are unclear, but might reflect the notorious instability of Hfr4315 and its derived F'. It should be added that the mapping of both the xerA3 and xerA5 mutations has proven to be problematic in many other respects, a phenomenon to be discussed more fully later.

DNA prepared from DS902 + F'4315 was cut to completion with EcoRI and HindIII. The resultant double digest was size fractionated by electrophoresis through a 0.4% agarose (low melting temperature grade) gel. DNA fragments in the 7-13Kb size range were recovered from the gel matrix, and ligated to the vector pAT223, also cut with EcoRI and HindIII. The resultant ligation mix was then transformed into DS941, and Ap^r transformants selected. The plasmid pAT223 is an expression vector similar to pKK223 (Brosius, et al, 1981) but based on pAT153 and encoding resistance to ampicillin (Boyd, A. C., pers. comm.; Twigg and Sherratt, 1980). Recombinant clones recovered were screened for their insert size by single colony gel analysis. Clones with inserts of approximately 8.5Kb were then tested for their ability to complement the xerA3 mutation in CSX3. This was done by transforming each test plasmid into CSX3 carrying the reporter plasmid pCS202. Since pAT223 and pCS202 are based on compatible replicons, with independently selectable resistance markers, transformants carrying both plasmids can be readily recovered by selecting on L-agar plus Ap+Cm. Transformants were then either restreaked onto the same medium, and grown overnight prior to plasmid DNA recovery, or where stated, plasmid DNA was recovered from "pooled transformants", (i.e. 30-50 colonies scraped off the primary transformation plate). Plasmid DNA from such transformants was analysed, on single colony gels, to determine whether pCS202 persisted in its 2-cer form, or had been resolved to its deletion derivative, pCS203. In this way Gill Stewart isolated pGS30, a plasmid which functionally complements pCS202 resolution in a xerA3 background (see Fig 4.10, section 4.6). This plasmid also effects complementation in a $RecA^-$, $XerA^-$ strain (DSX360), therefore can be said to complement xerA3 in trans (Stewart, 1986). The restriction mapping of pGS30 revealed that it carries an EcoRI/HindIII insert of approximately 8.45Kb. Furthermore, the restriction map of this insert corresponds closely (for all enzymes tested), with that predicted from the flanking DNA cloned in pCS300 (Fig 4.9); using the same assay system, pCS300 does not complement pCS202 resolution in CSX3 (data not shown). Therefore we conclude that we have cloned the wild-type equivalent (presumably from DS902), of part of the sequence present in pCS300, and that this wild type sequence includes a functional xerA gene.

4.6 DETERMINING THE MINIMUM pGS30 SEQUENCE WHICH COMPLEMENTS xerA3

It seemed probable that the xerA gene might be smaller than 8.45Kb, therefore Gill Stewart generated a number of pGS30 subclones, and tested each for xerA3 complementation. The three SphI fragments from pGS30 were cloned into the SphI site in pBR322. Of these three subclones only pGS36 (carrying the 3.1Kb SphI fragment from the centre of the pGS30 insert) proved capable of complementing a xerA3 mutant. In addition the 0.93Kb SphI/AccI fragment from pGS36 was cloned into pUC19 (Yanisch-Perron et al, 1985) to generate pGS38 (Stewart, 1986; See Fig 4.9). The subclone pGS38 was found to partially complement resolution of pCS202 to pCS203 in CSX3; that is to say, that on single colony gel analysis of plasmid DNA recovered from a test strain, pGS38 is observed together with both pCS202 and pCS203. A gel illustrating the relative abilities of pGS30, pGS36, and pGS38 to complement pCS202 resolution in CSX3 is shown in figure 4.10. The gel shows that pGS30 and pGS36 both efficiently complement the Xer$^-$ phenotype of CSX3, as evidenced by the absence of any detectable levels of pCS202, coupled with the appearence of pCS203. On the other hand pGS38 complements partially, whilst the control plasmid pCS20 does not complement at all. From these data, one can conclude that the 3.1Kb SphI fragment, present in pGS36, contains sufficient sequence information to encode an xerA gene product capable of efficiently complementing the xerA3 mutation. The partial complementation by pGS38 was initially interpreted as indicative of the deletion of some pGS36 sequences, which are required for efficient complementation. However, data to be presented shortly suggests this interpretation to be false.

This type of gel assay for complementing activity has been used extensively, with pCS202 and with analagous reporter plasmids, throughout this thesis. In order to limit the number of similar gels presented in the thesis, I will in general refer only to the results of such assays. The results of such "gel assays" for Xer complementation will be expressed in three qualitative terms:

1. "Positive" or "efficient" complementation; i.e. No pCS202 detectable, but its resolution product plus the complementing (test) plasmid being clearly visible. Thus pGS30 and pGS36 are described as efficiently complementing pCS202 resolution in CSX1 (Fig. 4.10).

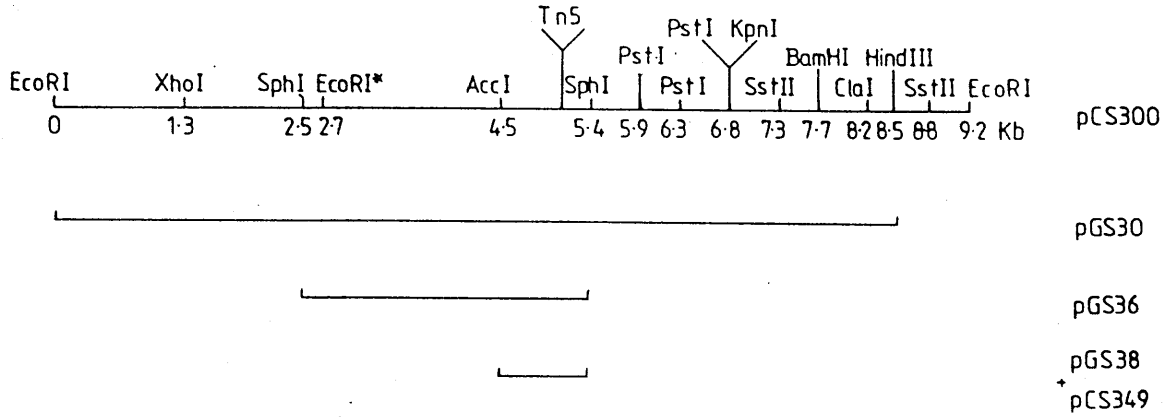


Figure 4.9 Linear Restriction Map of the pCS300 Insert, Indicating the Extent of the Wild-Type Sequences Present in a Number of Plasmid Sub-Clones.

The site of Tn5 insertion in xerA3 is indicated at 5.1Kb on the pCS300 map. The restriction map of the pGS30 insert is indistinguishable from that of the chromosomal sequences present in pCS300.

Figure 4.10 Complementation of pCS202 resolution in CSX3 by pGS30, pGS36 and pGS38.

Lane	Sample	Lane	Sample
1	pGS30	5	CSX3/pCS202 + pGS30
2	pGS36	6	" " + pGS36
3	pGS38	7	" " + pGS38
4	pCS20	8	" " + pCS20
		9	pCS202
		10	pCS203

"CSX3/pCS202 + pGS30", indicates that strain CSX3 was transformed first with pCS202, and then subsequently with pGS30. Plasmid pCS20 was used as a negative control and represents pUC8 carrying an 850bp insert (specifically the *rom* gene of ColE1).

2. A "Negative" result indicates the absence of any detectable reporter-plasmid resolution product, with only the 2-cer form plus the test plasmid being visible. This is the case with pCS20 (Fig. 4.10)

3. "Partial Complementation" (+/-) is used to describe the situation observed in pGS38 complementation tests (Fig. 4.10). In this case both resolved and unresolved forms of the reporter plasmid are present, plus of course the complementing plasmid.

4.6.2 THE PARTIAL COMPLEMENTATION PHENOTYPE OF pGS38

The partial complementation phenotype of pGS38 was intriguing, especially since sequence data for the 0.93Kb SphI/AccI fragment, inferred that the entire xerA gene was likely to be present in pGS38 (Chapter 6). Therefore the complementation characteristics of pGS38 were investigated further. The samples run on the gel shown in figure 4.10 are derived from pooled transformants. It was therefore necessary to investigate the clonal phenotype of individual transformants to ascertain whether or not single clones also exhibit mixed phenotypes with respect to pCS202/pCS203. Individual clones were streaked out onto L-agar plus Ap+Cm and incubated overnight before analysis of their plasmid content on single colony gels. By repeated sub-culturing of each clone, the persistence of the mixed phenotype was also determined. After the first overnight subculture (approximately 30 doublings), of 14 individual transformants of CSX3 carrying pCS202 plus pGS38, 13 out of 14 clones contained both pCS202 and pCS203 in roughly equivalent amounts. In one case out of 14, no pCS202 was apparent, only pGS38 and pCS203. After one further sub-culture the pattern was quite different; 13 out of 14 clones now contained pCS203 plus pGS38, but no detectable pCS202 (by ethidium bromide staining). In only one case out of 14 did the mixed phenotype persist. These results indicate that the "partial complementation" phenotype is a genuinely clonal one, and not an artefact of having made plasmid DNA from a mixture of transformants. The gradual elimination of pCS202 in each clone ostensibly suggests that there is a slow rate of resolution ongoing in these strains; one which takes on average between 30-60 generations to resolve all available pCS202 into its derivative pCS203. However, one must be very cautious in interpreting such resolution data in quantitative terms since pCS202 and pCS203 are

mutually incompatible; whilst this in itself imposes a serious complication on any "kinetic" analysis, there would be an additional complication should one form have some unknown replication and/or segregational advantage over the other. The problem posed by incompatibility can be explained as follows: Consider a CSX3 cell carrying pGS38, pCS202 and pCS203. Since pGS38 is compatible with both pCS202 and pCS203, then it can be assumed, for the moment, to be inherited by all progeny. However, since pCS202 and pCS203 are incompatible, then at cell division progeny will arise, at some indeterminate frequency, which have received **either** pCS202 **or** pCS203, whilst the remainder of the progeny will have received **both.** This provides an alternative route <u>via</u> which $pCS202^{-}/pCS203^{+}$ clones can arise, independent of Xer activity. Conversely, any CSX3 (pGS38) segregants which receive only pCS202, would be capable of re-establishing a mixed phenotype. It is interesting to note that no clones were detected (in the limited sample of 14), which carried only pGS38+pCS202. Such clones would have been expected to appear if complementation with pGS38 resulted in a transient period of Xer activity, followed by its complete shutdown.

It is plausible that the observed mixed phenotype is the result of gross segregational instability of pGS38. In this scenario, after transformation of CSX3+pGS38 with pCS202, the incoming plasmid replicates prior to its resolution into pCS203. If pGS38 is extremely unstable, then during cell division a segregant might arise which carries pCS202 (+/- pCS203), but no pGS38. Although such a segregant would be phenotypically Ap^{S}, it would be very likely to survive and grow, within the protective body of a predominantly Ap^{r} colony (due to the depletion of the local Ap concentration as a consequence of beta-lactamase activity). Such Ap^{S} clones would of course be Xer^{-}, and might be the entire source of pCS202 in plasmid preps. Whilst within a single colony this phenomenon might occur to some degree, it cannot explain the persistence of the mixed phenotype after streaking to single colonies, since each founder cell must have carried pGS38. The partial complementation phenotype of the pGS38/pCS202 plasmid combination was found to be independent of the order in which the two starting plasmids are transformed into CSX3. In other words, CSX3 carrying pGS38 then transformed with pCS202, behaves indistinguishably from CSX3 carrying pCS202 then transformed with pGS38, both with respect to the appearance, and persistence of the mixed plasmid

phenotype (data not shown).

4.6.3 EFFICIENT COMPLEMENTATION OF xerA3 BY pCS349

It should be stressed that pGS36 and pGS38 differ not only in the size of the chromosomal fragment that they carry, but also in that they are based on different vectors, pBR322 and pUC19 respectively. Whilst these two vectors are closely related, they have very different copy numbers; pUC19 being present at approximately 90 copies per chromosome equivalent, compared to 40 copies for pBR322 (Jones, 1985; Timmis, 1981). I have found that when the equivalent insert to that present in pGS38, is provided on a lower copy number plasmid, pCS349, then highly efficient *xerA* complementation is observed. Plasmid pCS349 was constructed as follows; the equivalent to the 0.93Kb *Sph*I/*Acc*I fragment from pGS36 was cut out of mCS320 as an *Eco*RI/*Hind*III fragment (mCS320 is an M13 mp19 derivative containing the 0.93Kb fragment from pGS36 cloned into the polylinker region for the purposes of DNA sequence determination (see Chapter 6)). The *Eco*RI/*Hind*III fragment from mCS320 was ligated to the *Eco*RI/*Hind*III restricted vector, pCB19B, to generate pCS349 (See Fig 4.11). Vectors pCB18B and 19B are improved lambda dv derivatives of the plasmid pCB101 (Boyd & Sherratt, 1986); they lack a number of restriction sites present in their predecessor, pCB101, thus rendering more of the polylinker sites unique (including *Eco*RI and *Hind*III). Both vectors encode Cm^r, and they represent a "matched pair", differing only by the orientation of their poly-linkers with respect to the *lac* promoter and *lacZ'* sequence; pCB18B carrying the mp18 version (as found in pUC18), and pCB19B carrying the mp19 version (Boyd A.C., pers. comm.; Yanisch-Perron *et al*, 1985). Plasmid pCS349 was tested for its ability to complement pCS210 resolution in strain CSX11 (*xerA*5). The reporter plasmid pCS202 could not be used with pCS349 since both are based on lambda dv replicons, and so are mutually incompatible. The reporter plasmid pCS210 offers compatibility with pCS349 (since the former is based on the p15A replicon), plus an independently selectable marker, i.e. Tet^r for pCS210 and Cm^r for pCS349. Furthermore, the deletion marker of pCS210 is the *E. coli lacZ* gene, enabling resolution to be monitored indirectly *via* LacZ activity on X-gal plates (See Fig 4.12). Strain CSX11 was used since it is chromosomally deleted for the *lac* operon (being delta *lac*/*pro*), therefore the LacZ activity of this strain is solely dependent on that encoded by pCS210. Complementation

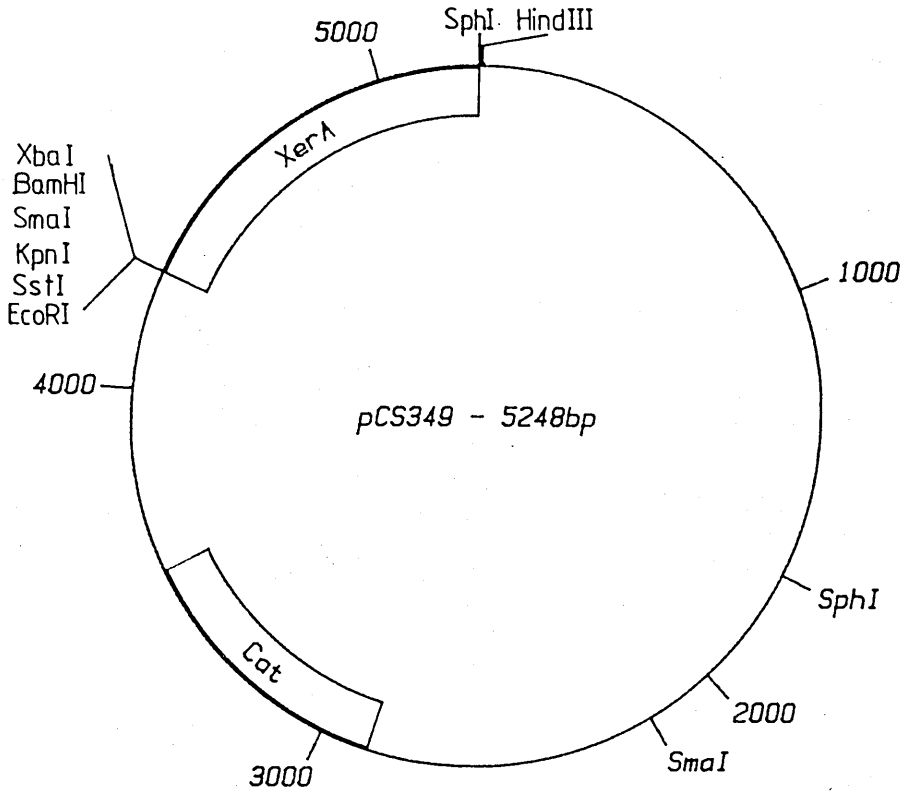


Figure 4.11 Structure of the xerA minimal-clone pCS349

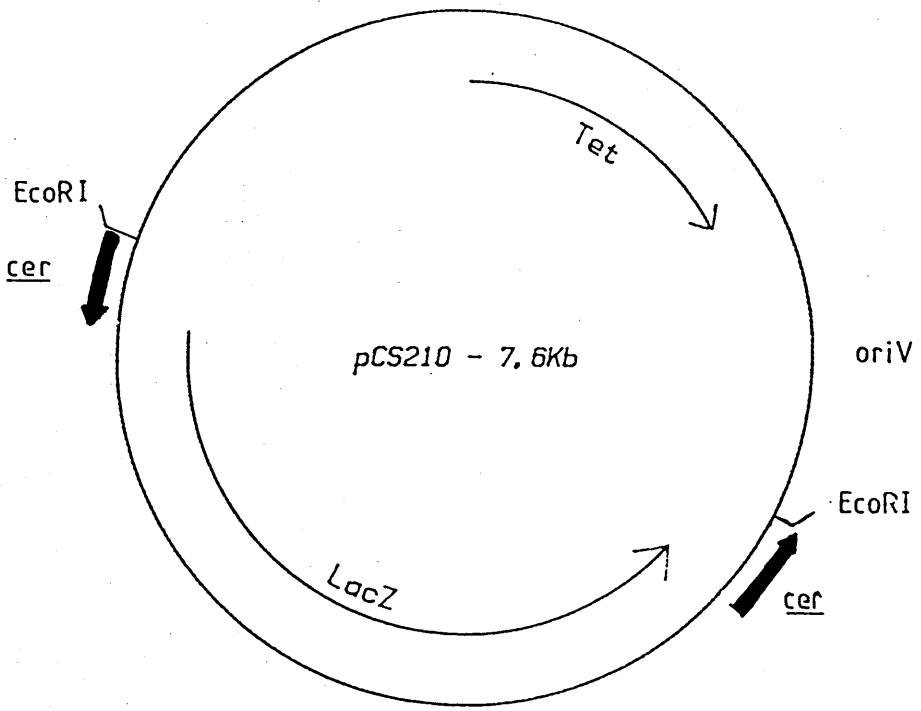


Figure 4.12 Structure of the 2-cer reporter plasmid pCS210: pCS210 carries a 3.6Kb cer-lacZ-cer "cartridge" inserted into the unique EcoRI site in the vector pACYC184. The resultant plasmid encodes beta-galactosidase plus resistance to Tet when maintained in an Xer⁻ host; however, Xer-dependent resolution of pCS210 deletes the lacZ gene to generate pCS211 (4.3Kb; encoding resistance to Tet).

tests were performed following the same principle as before, where CSX11 carrying pCS210 was transformed with various test plasmids, after which the transformants were plated out onto DM minimal medium, containing casamino acids (1%), glucose (0.2%), X-gal (25mg.l^{-1}), tetracycline (7mg.l^{-1}), and a suitable antibiotic for the selection of cells containing the test plasmid. After overnight incubation, transformant colonies were scored for their colony phenotype with respect to the blue/white colour test, and in most cases, plasmid DNA from pooled transformants was analysed on single colony gels. Using this assay system I tested plasmids pGS36, pGS38, pCS349 plus the vector controls pUC19 and pCB19B, and found that pGS36 and pCS349 both complement pCS210 resolution efficiently, whereas pGS38 exhibits a similar partial complementation phenotype to that observed in previous experiments; pUC19 and pCB19B do not complement at all (Table 4.2; Fig. 4.13). Colour photographs of the primary transformation plates from pCB19B, pCS349 and pGS38 complementation tests are shown in figure 4.14. The colony colours listed in table 4.2 correlate very well with the observed plasmid content of each strain tested. In the cases of pGS36 and pCS349, the white colonies reflect the absence of any detectable pCS210 in the pooled transformants (Fig 4.13, tracks 6 and 8). Note that in tracks 5 and 7 some dimeric pCS211 is present which does not appear in the pCS211 control track (Fig 4.13 track 4). The two major forms present in track 4 migrating above the monomer supercoiled pCS211, are presumed to be linear monomer and open circle monomer forms; this is supported by comparing the relative migration rates of each form in track 4 (pCS211) and track 1 (pCB19B). In the plasmid DNA isolated from the pCB19B and pUC19 complementation tests (tracks 5 + 7), no monomeric pCS210 or pCS211 are present. There is, however, a band in both cases which comigrates with pCS210 trimer. As can be seen in the CsCl purified pCS210 DNA sample, such multimerisation of pCS210 in an Xer^- strain is common. On pooling plasmid DNA from the blue and white colonies which appeared in the pGS38 complementation, both pCS210 and pCS211 were apparent in single colony gels in similar ratios to those previously observed with pCS202:pCS203 (data not shown). When the primary transformant colonies are re-streaked onto fresh medium, pGS36 and pGS349 transformants re-streak as purely white clones, whereas pUC19 and pCB19B transformants re-streak as exclusively blue clones. In the case of pGS38, both the blue and pale blue transformant colonies (Fig.

Table 4.2 Complementation Tests in CSX11 + pCS210 With Various *xerA* Sub-Clones

Strain	Test Plasmid	Colony Phenotype on X-Gal	XerA Complementation
CSX11+pCS210	pGS36	White	+
"	pGS38	Blues, Pale Blues and Whites*	+/-
"	pCS349	White	+
"	pUC19	Blue	-
"	pCB19B	Blue	-

Complementation was scored as either Efficient (+), Partial (+/-), or Negative (-).

* The ratio of Blues: Pale Blues: Whites being approx. 20:60:20. Note that all Blues and Pale Blues re-streak to give rise to exclusively white colonies.

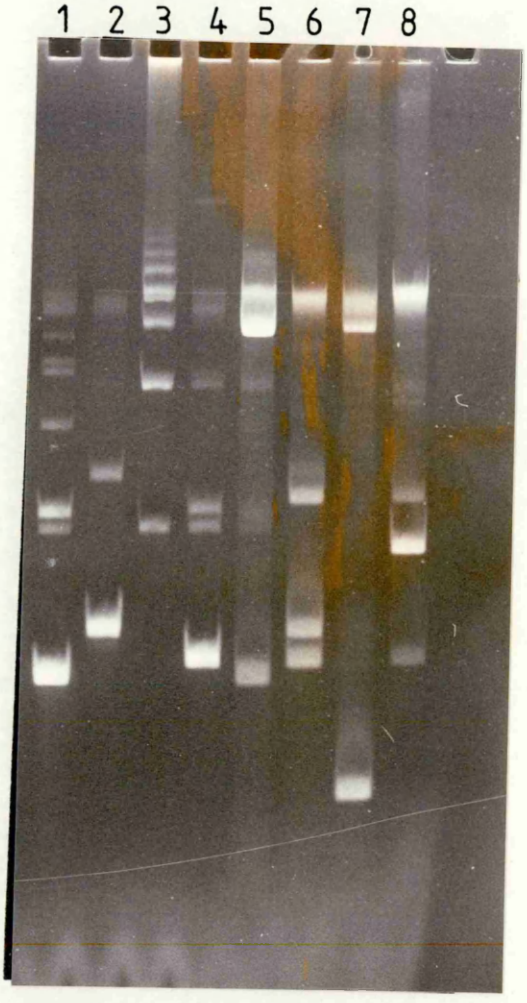
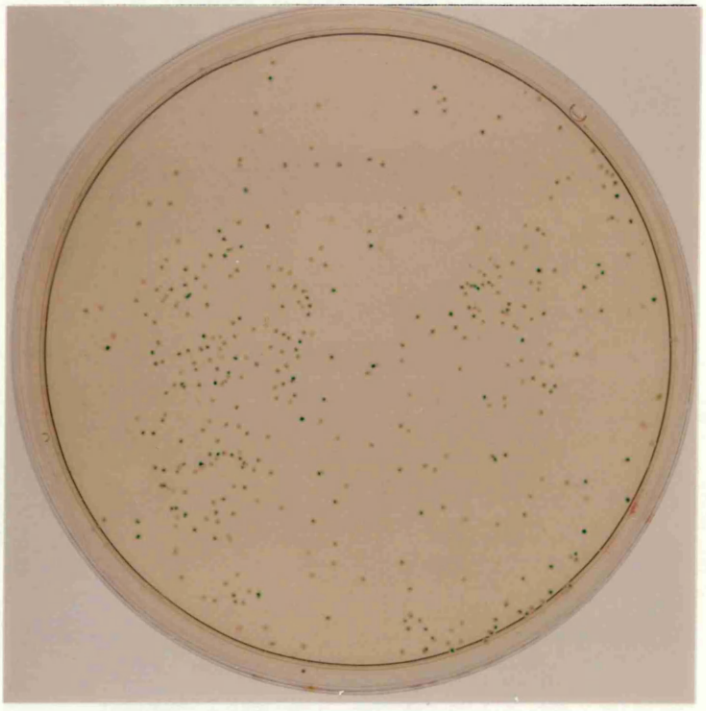


Figure 4.13 Complementation of pCS210 resolution in CSX11 by pCS349 and pGS36.

Lane	Sample	Lane	Sample
1	pCB19B	5	CSX11/pCS210 + pCB19B
2	pCS349	6	" " + pCS349
3	pCS210	7	" " + pUC19
4	pCS211	8	" " + pGS36



b.

c.

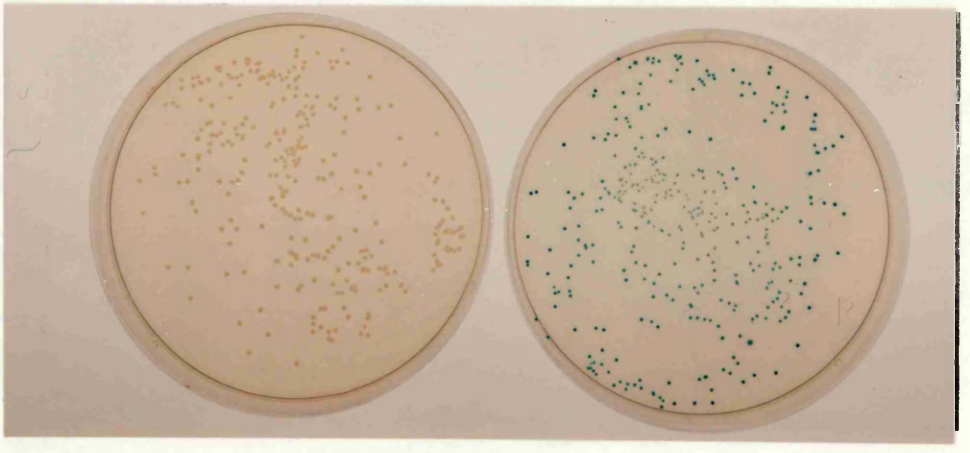


Figure 4.14 Complementation of pCS210 resolution in CSX11 by pCS349 and pGS38 as assayed by the pCS210 colony-colour test.
(a) CSX11/pCS210 + pGS38; (b) CSX11/pCS210 + pCS349; (c) CSX11/pCS210 + pUC8. The negative control for complementation was pUC19; I should stress that the CSX11 + pUC19 gives rise to white colonies in the absence of pCS210.

4.14), re-streak to give predominantly white colonies (indicative of the resolution of pCS210 to pCS211). A few pale blue secondary colonies do arise, but these too restreak as exclusively whites in the next sub-culture (white colonies from the pGS38 complementation restreak as whites).

The pCS210 resolution assay in a delta lac/pro background provides a colour test which accurately reflects the pCS210/pCS211 status of a clone. Therefore in a number of future complementation tests, the colour test has been deemed sufficient evidence as to the Xer phenotype of a strain; where stated, such "pCS210 resolution colour tests" have also been confirmed by "pCS210 gel assays" in which the clones plasmid content was directly visualised.

From these complementation assays there is now conflicting data as to the efficacy of the 0.93Kb SphI/AccI fragment in xerA complementation. The pCS349 data infers that this fragment does indeed contain the entire xerA gene, an inference corroborated by DNA sequence analysis of this region (Chapter 6). The question is therefore raised as to whether the mixed phenotype observed in pGS38 complementation is a consequence of poor complementation per se, or is infact the result of some inhibitory influence of pGS38 on Xer activity. In order to test for any such inhibitory role, the effects of pGS38 on the Xer phenotype of wild-type hosts was determined.

4.7.1 INHIBITION OF pCS202 RESOLUTION IN DS902 BY pGS38

The partial complementation observed for pGS38, relative to pCS349 (carrying the same cloned chromosomal sequence), seems to imply that either pGS38 is not efficiently expressing xerA (perhaps as a result of mutation), or that the presence of pGS38 has a direct inhibitory affect on Xer dependent pCS202 resolution. To test this latter possibility, DS902 was transformed with pGS38, then DS902 (pGS38) was transformed with pCS202 to determine the Xer phenotype. DS902 is Xer^+ in a pCS202 resolution assay; after transformation with pCS202, single colony gel analysis of plasmid DNA recovered from pooled Cm^r transformants reveals that only pCS203 is present (data not shown). A single DS902 (pGS38) clone was transformed with pCS202, selecting the desired transformants on L-agar +Ap+Cm. Five independent transformants were then subcultured overnight, prior to being analysed on a single colony gel (Fig. 4.15). In this case, significant levels of pCS202 persist, indicative of an inhibitory effect of pGS38 on the Xer

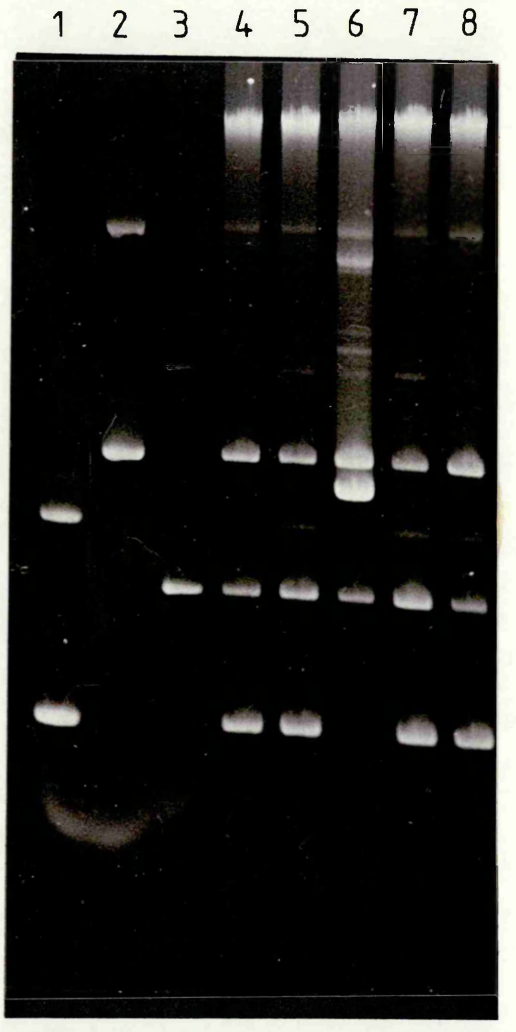


Figure 4.15 Inhibition of pCS202 resolution in DS902 by pGS38

Lane	Sample
1	pGS38
2	pCS202
3	pCS203
4-8	DS902/pGS38 + pCS202 5 independent isolates.

activity of DS902. That the cloned *xerA* gene does not exert a similar effect on DS902 when cloned on a lower copy number vector was demonstrated using the plasmid pCS350. This plasmid is based on a derivative of pCB101, namely pGS301; pGS301 is 5.8Kb, and encodes resistance to tetracycline. Plasmid pGS301 has a unique *Pvu*II site, into which was cloned the *Eco*RI/*Hind*III *xer*A fragment from mCS320, after blunting the 5'-protruding ends by "Filling-in" with Klenow in the presence of all 4 dNTPs. The resultant plasmid, pCS350 (6.75Kb), is therefore in the lambda dv compatibility group, encodes Tet^r, and carries the *E. coli xerA* gene. This plasmid was constructed for the purpose of complementing the resolution of the reporter plasmid, pKS455, in $XerA^-$ strains and was found to do so (data not shown). DS902 with or without pCS350, was transformed with pKS455, and pooled transformants were analysed on single colony gels. DS902 was found to efficiently resolve pKS455 into its deletion derivative, p456, irrespective of the presence or absence of pCS350 (data not shown). Therefore, one can conclude that the presence of plasmid-borne copies of the *xerA* gene does not in itself exert an inhibitory influence on DS902 Xer activity; the phenomenon seems to be peculiar to the high copy plasmid pGS38.

4.7.2 EFFECTS OF pGS38 AND pBAD ON Xer ACTIVITY OF DS941

The next major question to be investigated was whether or not the inhibitory influence of pGS38 was specific to the presence of the *xerA* insert, and not just a feature of the vector. To this end DS941 (*lacI*^{q}), was transformed with either pGS38 or the vector pBAD. Clearly the ideal control would have been pUC19, however, pBAD is very similar to pUC19 in size, copy number, and sequence; the major difference being that pBAD carries the *tac* promoter- *rrnB* terminator cartridge of pAT223, as opposed to the *lac* promoter and *lacZ'* sequence of pUC19 (A.C Boyd pers. comm.). DS941 carrying either pGS38 or pBAD, was then transformed with pCS202, and transformants selected on L-agar plus Ap+Cm. Single colony gel analysis of subcultured transformants showed that no inhibition of pCS202 resolution was apparent by either pBAD or pGS38 (Fig. 4.16). It seemed possible that the observed difference between pGS38 inhibition in DS902 and DS941 might reflect the *lacI*^{q} status of the latter; the *lacI*^{q} mutation increases the level of lac repressor 10 fold over the wild-type *lacI* gene (Muller-Hill *et al*, 1968). DS941 is thus capable of more effective repression of multicopy

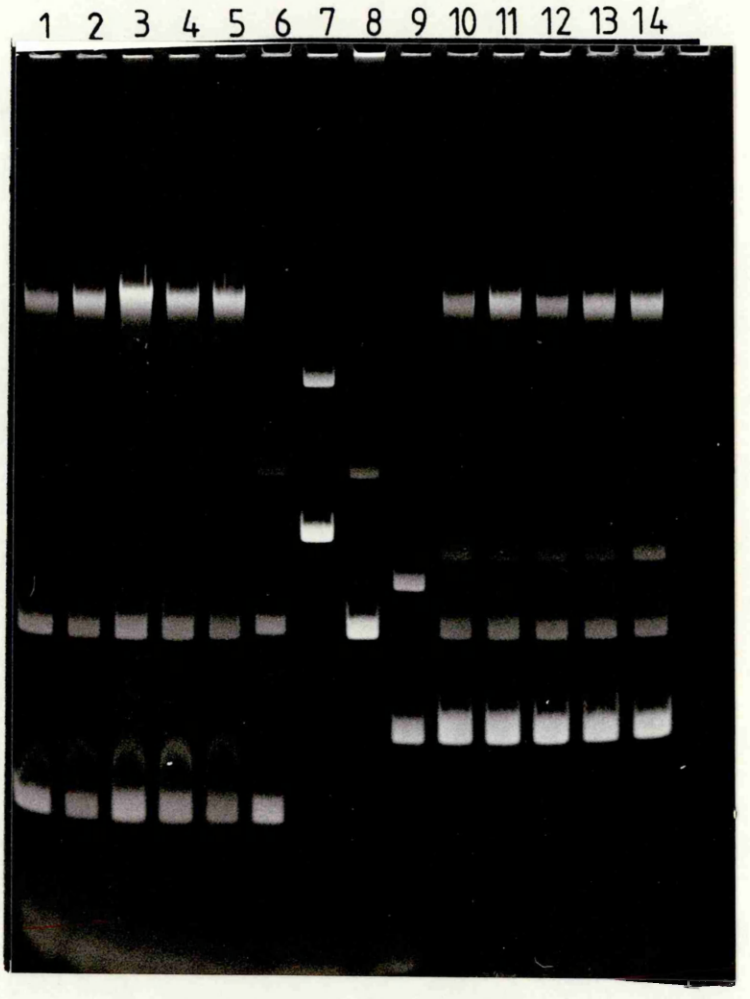


Figure 4.16 Effects of pBAD and pGS38 on pCS202 resolution in DS941 when grown in the absence of IPTG. Five independent clones of DS941/pBAD + pCS202 were analysed (lanes 1-5), together with 5 clones of DS941/pGS38 + pCS202 (lanes 10-14). Markers: pBAD, lane 6; pCS202, lane 7; pCS203, lane 8; pGS38, lane 9.

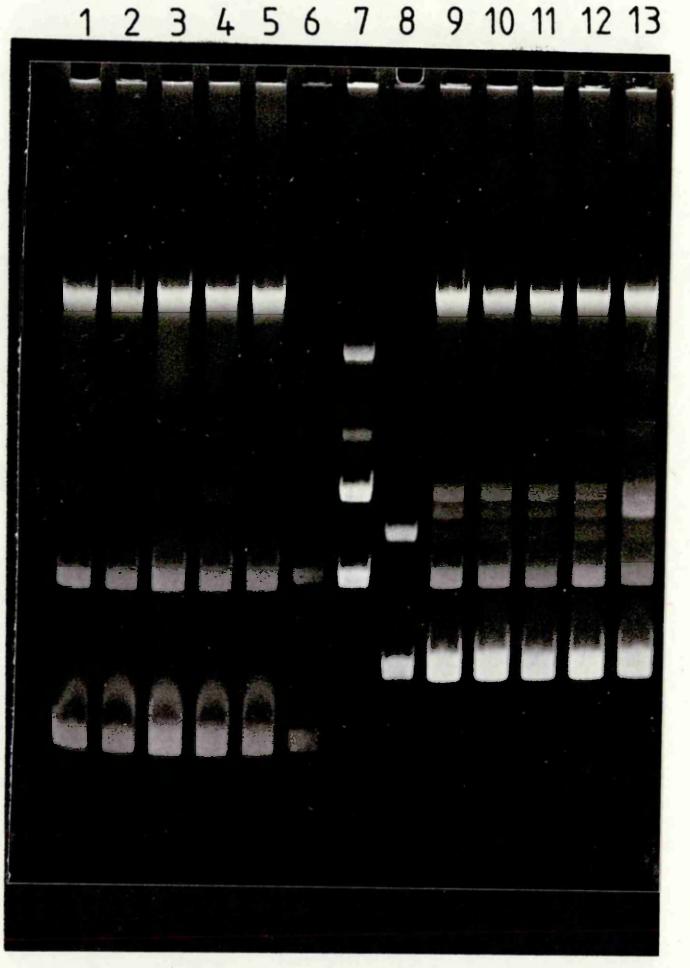


Figure 4.17 Effects pBAD and pGS38 on the resolution of pCS202 in DS941 when grown under IPTG-induced conditions. As in figure 4.16, five independent clones of DS941/pBAD + pCS202 were analysed (lanes 1-5), as were 5 clone of DS941/pGS38 + pCS202 (lanes 9-13). Markers: pBAD, lane 6; pCS202+pCS203 mixture, lane 7; pGS38, lane 8.

lac promoters in pGS38, than is DS902 (lacI) (this contention is supported by observations made on the expression of cloned genes; this laboratory) . Therefore, assuming that overexpression of the xerA gene from the lac promoter in pGS38 was essential to the inhibition of Xer activity, I repeated the DS941 experiment in the presence of the gratuitous lac inducer IPTG. DS941 containing either pBAD or pGS38, was grown up for transformation in the presence of IPTG ($20mg.l^{-1}$). After transformation with pCS202, transformants were selected on L-agar plus Ap+Cm+IPTG. Surprisingly, single colony gel analysis of subcultured clones revealed that both pGS38 and pBAD produced a significant inhibitory effect on the resolution of pCS202 (Fig. 4.17). Previous experiments with DS941 and the reporter plasmid pCS208 suggest that the DS941 Xer activity is not inhibited by IPTG per se (Chapter 3).

These results, summarised in table 4.3, appear to indicate that the presence of a high copy number plasmid does not, in itself, inhibit Xer activity, but that the affect requires the induction of a strong plasmid promoter, irrespective of the gene being driven from that promoter. The need to induce pGS38 with IPTG in order to observe inhibition in DS941 is not inconsistent with the observed inhibitory effects, during complementation tests, in other strains in the absence of IPTG; the reason being that the strains employed in pGS38 complementation tests, were all either lacI, or delta lac/pro; therefore none would have been able to effectively repress pGS38 (c.f. DS902). Obviously, to determine whether or not "partial complementation" and "Xer^+ inhibition" are one and the same phenomenon, one would like to test pGS38 complementation in an $XerA^-$, $lacI^q$ strain in the presence and absence of IPTG. For this purpose, a suitable strain could be constructed by transduction of the xerA3 allele into DS941.

4.8 DOES THE CLONED xerA GENE COMPLEMENT ANY OTHER xer MUTANTS?

The 0.93Kb SphI/AccI fragment appears to carry all the sequence necessary for the complementation of the xerA3 and xerA5 mutants. Therefore this subclone was tested for its ability to complement the remaining nine independently isolated mutants (xerB1-xerB9). Complementation tests were performed as described previously, and involved the test and reporter plasmid pairings of either pGS38+pCS202, or pCS350+pKS455; in no case was complementation

Table 4.3 Summary of Plasmid Inhibition Results

Strain	Resident Plasmid	IPTG Induction	Reporter Plasmid Resolution
DS902	pGS38	-	Inhibited
"	pCS350	-	Efficient
DS941	pGS38	+	Inhibited
"	"	-	Efficient
"	pBAD	+	Inhibited
"	"	-	Efficient

observed (data not shown). Furthermore, the larger wild-type clone, pGS30, has been tested for its ability to complement pCS202 resolution in various xerB mutants (including xerB1, B2, B4, B5, B8, and B9), and as before no complementation was observed (G. Stewart, 1986). Therefore it seems clear that the xerB mutations cannot be complemented by the cloned xerA gene, and that the (intact) xerB gene is not present within the pGS30 insert (later evidence will confirm this conclusion).

4.9 CHROMOSOMAL MAP POSITION OF THE xerA LOCUS

As already intimated, the attempts to map the xerA locus using conventional ("recombinational") techniques, have proven to be problematic; with different experiments yielding quite different results (Stewart, 1986; D. Sherratt & M. Burke, pers. comm.). The reasons for these inconsistencies remains unclear, but might reflect some biological function of XerA (or a closely linked gene), e.g. in the recombination processes involved during Hfr crosses etc..

A completely novel, indirect route for mapping a gene's position on the E. coli chromosome has recently been made possible by the publication of the complete restriction map of the K-12 chromosome for the restriction endonucleases HindIII, BamHI, EcoRI, EcoRV, BglI, KpnI, PstI & PvuII (Kohara et al, 1987). Obviously, one can scrutinise the chromosomal map for close-fits to the restriction pattern of a cloned gene. The only match identified to the restriction pattern surrounding xerA (derived from pCS300 & pGS30), is found at 3455Kb on the Kohara et al map, equivalent to 70.5 mins on the conventional E. coli K-12 linkage map (Fig. 4.19). The close corespondence between the two restriction maps makes it seem highly likely that xerA does indeed map to 70.5 mins; in order to confirm this conclusion two of Kohara's clones from this region have been requested, and will be tested for the presence of the xerA gene. Only two other genes have currently been mapped within this immediate vicinity, those being argR (70.5; regulatory gene), and mdh (70.4; malate dehydrogenase: Bachmann, 1983). According to the Kohara et al map, both argR and mdh should be present in pCS300 (and perhaps in pGS30), however neither function has previously been implicated in a recombination system.

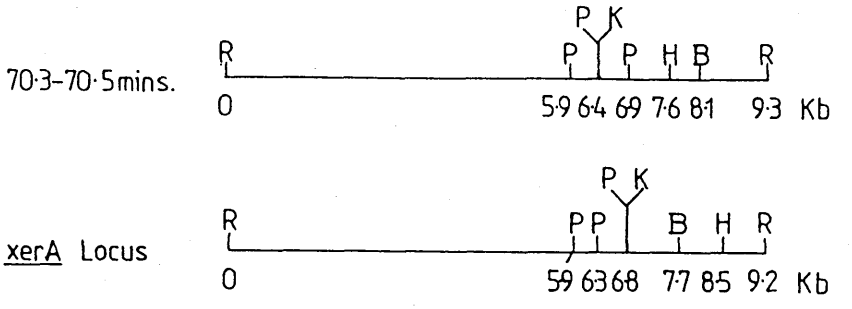


Figure 4.20 **Comparison Between the Restriction Map of the *xerA* Locus and that of the 70.4 min. Region on the *E. coli* K-12 Linkage Map.**

Code: B= BamHI E= EcoRI
H= HindIII K= KpnI
P= PstI

4.0 DISCUSSION AND CONCLUSIONS

The Tn5 insertion tightly linked to the xerA3 allele in CSX1, has been cloned as an EcoRI fragment, together with 9.2Kb of flanking chromosomal DNA, in the plasmid pCS300. The restriction map of this plasmid was determined, and used to confirm that the cloned locus does indeed correspond to xerA3.

Using restriction data derived from pCS300, the wild-type xerA gene was cloned as an 8.45Kb EcoRI/HindIII fragment into the vector pAT223, the resultant clone being identified by virtue of its complementation of pCS202 resolution in CSX3. The complementing function has been narrowed down to a 0.93Kb SphI/AccI fragment, which complements two out of the 11 independent mutants isolated; the mutant alleles in question were thus designated xerA3 and xerA5. This minimal clone fails to complement any of the remaining nine mutants, all of which have since been shown to belong to a single complementation group, xerB. The larger clone, pGS30, was found to be unable to complement any of the six xerB mutants tested, indicating that a functional xerB gene is not present. The DNA sequence of the 0.93Kb SphI/AccI fragment has been determined (see chapter 6).

An inhibitory affect on pCS202 resolution by the transcriptional induction of the multi-copy plasmid, pBAD, was uncovered. In DS941, the presence of pBAD reduces the apparent rate of resolution of incoming pCS202, but does so only in the presence of IPTG. No such affect is apparent in the absence of IPTG. A similar IPTG-dependent affect has been demonstrated for pGS38, however my data cannot exclude the possibility that an overexpressed xerA gene may play a contributory role in this case. The nature of the observed inhibitory affect of induced pBAD on Xer-dependent resolution of pCS202 is intriguing. It seems likely that the induction of strong promoters on such high copy plasmids, will exert considerable stress on the transcriptional/translational apparatus of the cell; but how such a "stress" might affect Xer activity is unclear. Possible causes might include the induction of a "stress" response (such as the "heat shock" response (for a recent review see Lindquist, 1986)); one of the direct, or indirect, effects of which being a reduction in apparent Xer activity. It is possible that total Xer activity in "pCS202 resolution inhibited" strains may not be reduced, but that Xer factors (and/or co-factors), may be sequestered to perform other activities in the cell, thereby diminishing the rate of the only Xer

reaction that we can monitor. It is interesting to speculate that the reduction in *cer*-specific recombination during "transcriptional stress" might reflect the potential involvement of transcription of the *cer* locus in promoting recombination (see Chapter 1). If this were so, then the sequestration of transcriptional factors by multiple, induced *lac* or *tac* promoters, might exert a competitive inhibition on recombination at *cer* sites. Other less exciting possibilities might include a "stress" related reduction in the expression of *xerB* and/or any other unidentified *xer* gene(s). It seems unlikely that the induction of pBAD has a deleterious affect on gene expression in general, since the growth rate of induced pBAD-containing DS941 is indistinguishable from that of an uninduced culture. However, I should add that the presence of an additional plasmid, namely pCS202, might further stress a pBAD induced cell; the growth rates of DS941+pBAD+pCS202 have not been measured.

CHAPTER 5

CLONING THE XerB GENE

5.1 INTRODUCTION

The *xerA* gene was cloned by exploiting an *xerA*-linked Tn5 insertion. The same route was available for the cloning of *xerB* via the Tn5 insertion mutant *xerB1*. However, the reporter plasmid pCS210 (described in chapter 4), offers an alternative approach to the cloning of *xer* genes by providing a reliable colony colour test for a strain's Xer phenotype (via the deletion of the *lacZ* gene). This plasmid therefore provides a means with which to rapidly screen an *E. coli* genebank for complementation of an Xer^- mutant. The archetypal "$XerB^-$" mutant used in the screening was CSX4, carrying the Tn5-linked lesion, *xerB1*. Direct screening for *xerB1* complementation obviates the requirement for first cloning the *xerB1* allele in order to target the wild-type sequence. Nonetheless *xerB1* was cloned and ultimately provided useful information as to the structure of the *xerB* locus and the nature of the *xerB1* lesion.

5.2 CLONING THE XerB1 ALLELE FROM CSX4

The Tn5 element in the chromosome of CSX4 was found to be present on very large (>30KB) *Eco*RI, *Cla*I, and *Sst*I restriction fragments via Southern blot analyses (data not shown). Such large fragments could not be readily cloned into a plasmid vector (although of course suitable lambda vectors are available). Therefore it was decided to clone flanking DNA to only one side of the transposon insertion. The restriction enzyme *Bam*HI cuts once in Tn5 (at position 3056bp), thus generating two Tn5 fragments, one of which carries the intact NPTII gene. In a southern blot of *Bam*HI-restricted CSX4 DNA, two fragments were detected which hybridised to labelled pCS80 sequences. The two fragments were sized relative to size standards which also hybridise to pCS80 (namely pCS300/*Bam*HI and pCS300/*Eco*RI). The calculated sizes of the two CSX4/*Bam*HI fragments being approximately 5.2Kb and 3.6Kb respectively (data not shown). Since *Bam*HI cuts Tn5 into two portions of 3056bp and 2762bp respectively, then one can calculate that in order to produce two chromosomal fragments of 5.2 and 3.6 Kb, the transposon must have inserted into a 3Kb *Bam*HI fragment, and must be situated either 0.6Kb or 0.8Kb from one end.

*Bam*HI-restricted CSX4 DNA was ligated to *Bam*HI/CIP treated pBR322, transformed into DS902, and transformants selected on L-agar +Ap+Km. In this way the plasmid pXB1 was isolated. Strain DS902 carrying pXB1 was found to be resistant to Ap and Km, but is sensitive to Tet,

consistent with the inactivation of the pBR322 tet gene by an insert in the BamHI site.

5.2.2 RESTRICTION MAPPING OF pXB1

Plasmid pXB1 was restriction mapped using the same approach described in detail in section 4.2.2, "Restriction mapping of pCS300". Again a sample of those gels used to generate pXB1 restriction data are shown (Figs. 5.1, 5.2, 5.3); determination of the order of the three closely spaced HindIII sites, and the three closely spaced AccI sites, required the sizing of small fragments, from various single and double digests, on polyacrylamide gels (c.f. Figs. 5.2 & 5.3). The restriction map of pXB1, plus a list of all the restriction endonucleases tested for restriction of pXB1, is presented in figure 5.4. The restriction map of pXB1 reveals that the plasmid carries a a 5.2Kb BamHI fragment inserted at the BamHI site of pBR322; the size of this insert is identical to one of the two Tn5-containing fragments observed in the Southern analysis of CSX4/BamHI DNA discussed above. Thus it seems likely that the insert in pXB1 does indeed represent part of the major Tn5 insertion locus of CSX4. This 5.2 Kb insert comprises 3056bp from the left end of Tn5, plus approx. 2.1Kb of flanking chromosomal DNA. Whilst this represents the flanking DNA to only one side of the insertion, it nonetheless provides important information with regards to the structure of the xerBI locus; information which ultimately demonstrated that the xerBI lesion is indeed the result of a Tn5 insertion into the xerB structural gene.

5.3 CLONING THE WILD-TYPE XerB GENE

The strategy for cloning xerB involved screening an E. coli K12 library for complementation of reporter-plasmid resolution in CSX4. The cosmid library employed was a generous gift from Derek Blake, (University of Oxford). The library was generated from the canonical E. coli strain, OP-Phi, by Sau3A partial digestion and phosphatase treatment of genomic DNA. The partial digestion products were size-fractionated on a sucrose gradient, ligated to BamHI-restricted cosmid vector DNA, pDVcos, and then packaged, in vitro, into infective phage particles (Derek Blake, pers. comm.). The cosmid vector pDVcos is a 5Kb derivative of pTCF (Grosveld et al, 1982); its other salient features are that it is based on the pMB1 replicon, encodes resistance to Ap, has a unique BamHI site, and carries two tandemly arranged

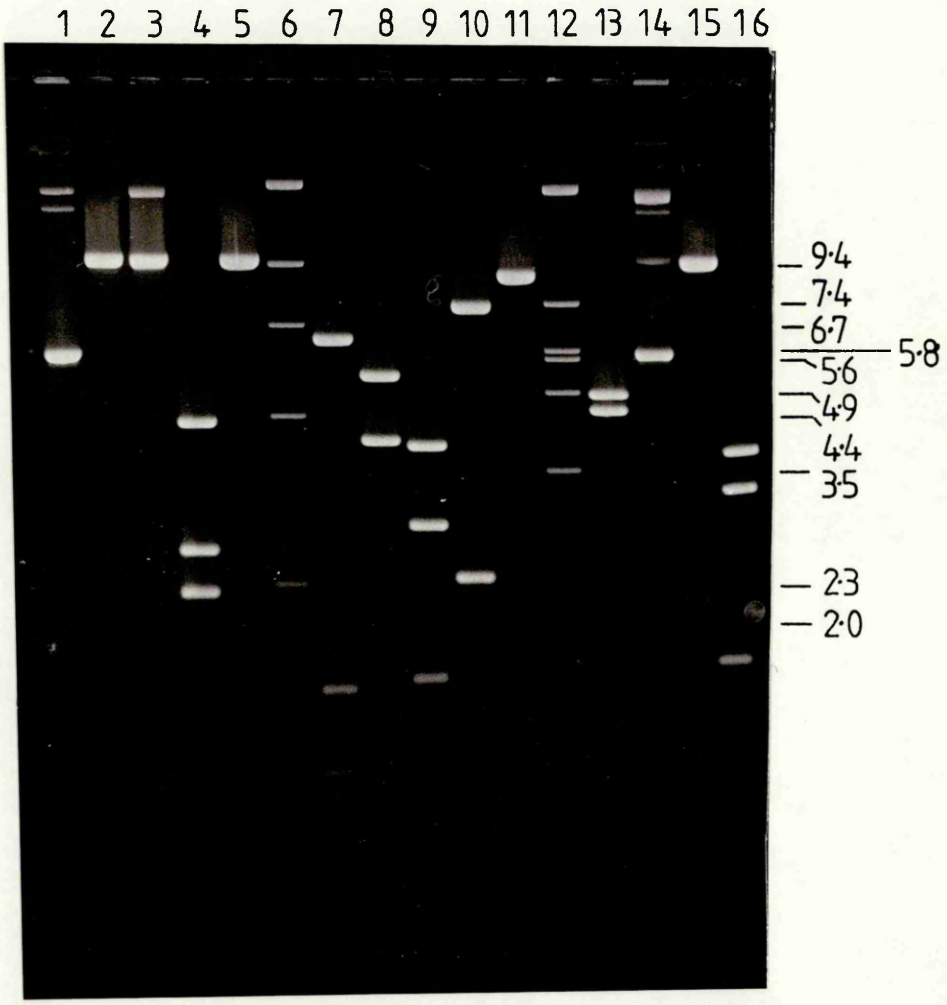


Figure 5.1 Restriction analysis of pXB1

Lane	Sample	Lane	Sample
1	pXB1/uncut	9	pXB1/AccI
2	pXB1/EcoRI	10	" /XhoI
3	" /SstII	11	" /HpaI
4	" /HindIII	12	Lambda/EcoRI
5	" /ClaI	13	pXB1/SalI
6	Lambda/HindIII	14	" /SstI
7	pXB1/PstI	15	" /SmaI
8	" /SphI	16	" /PvuII

Digests were run on a 1% agarose gel. The sizes of marker bands are given in Kb.

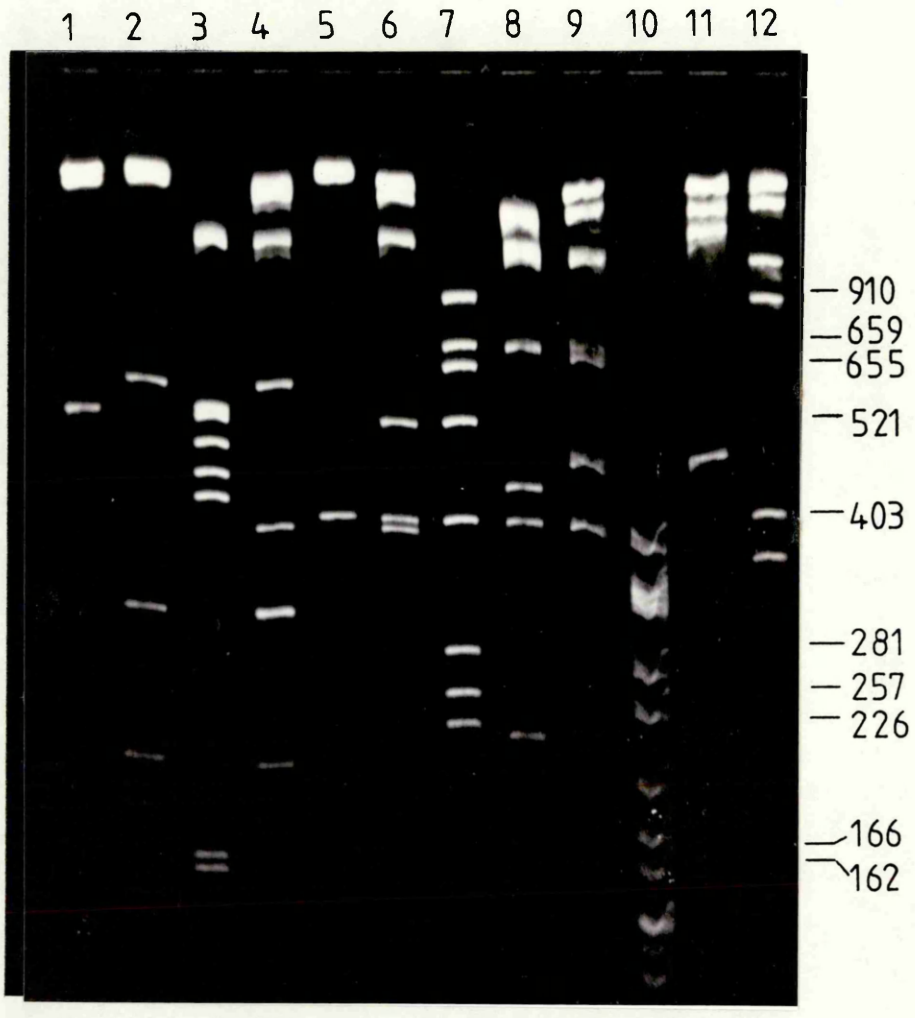


Figure 5.2 High resolution restriction analysis of pXB1

Lane	Sample	Lane	Sample
1	pXB1/SphI	7	pBR322/AluI
2	" /SphI/SalI	8	pXB1/AccI/HindIII
3	pBR322/DdeI	9	" /HindIII/SalI
4	pXB1/SphI/AccI	10	pBR322/HhaI
5	" /SalI	11	pXB1/HindIII
6	" /AccI	12	" /HindIII/SphI

Digests were run on a 5% polyacrylamide gel. The sizes of bands are given in base pairs,

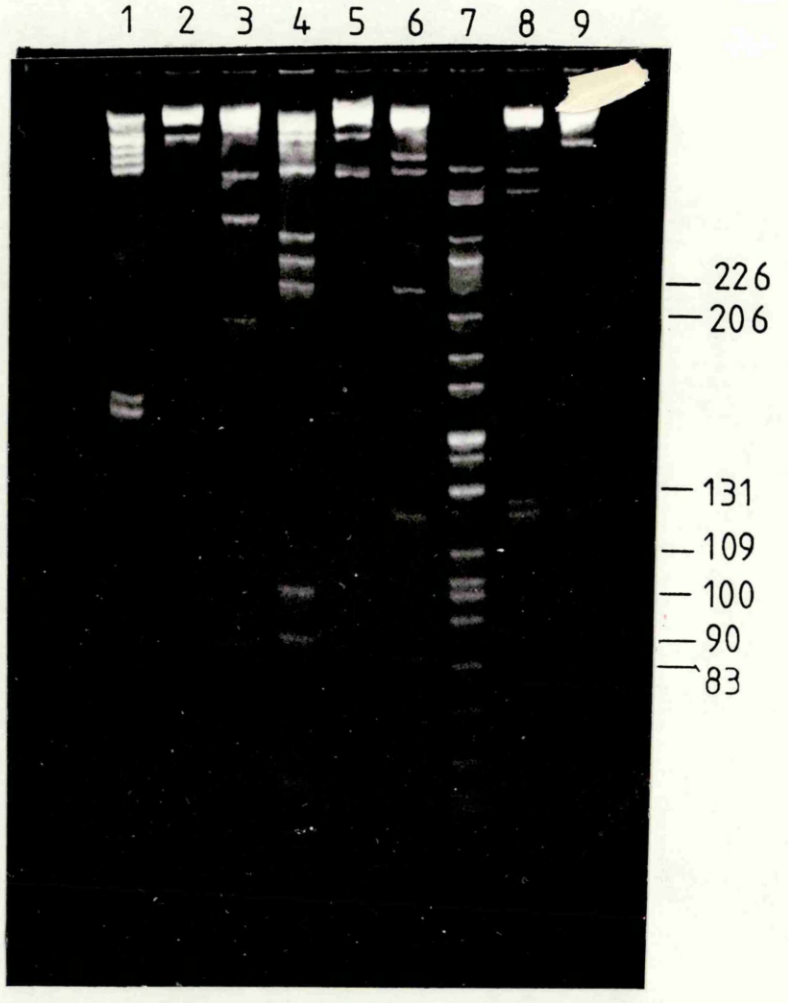


Figure 5.3 High resolution restriction analysis of pXB1

Lane	Sample	Lane	Sample
1	pBR322/DdeI	6	pXB1/AccI/HindIII
2	pXB1/SphI	7	pBR322/HhaI
3	" /SphI/AccI	8	pXB1/HindIII/SphI
4	pBR322/AluI	9	" /HindIII
5	pXB1/AccI		

Digests were run on an 8% polyacrylamide gel. The sizes of fragments are given in base pairs.

Enzyme	Nº of Cuts in pXB1	Predicted Nº of Sites in- pBR322	Tn5*	Chromosomal DNA
AccI	6	2	1	3
BamHI	2	1	1	1**
BglII	2	0	1	1
ClaI	1	1	0	0
EcoRI	1	1	0	0
HindIII	5	1	1	3
HpaI	3	0	1	2
PstI	4	1	3	0
SalI	3	1	1	1
SmaI	1	0	1	0
SphI	3	1	1	1
SstI	1	0	0	1
SstII	1	0	1	0
XhoI	2	0	2	0

*Note that only the left-hand end of Tn5 is present in pXB1 (1-3056bp).

** The 2 BamHI sites in pXB1 correspond to the vector/insert boundaries.

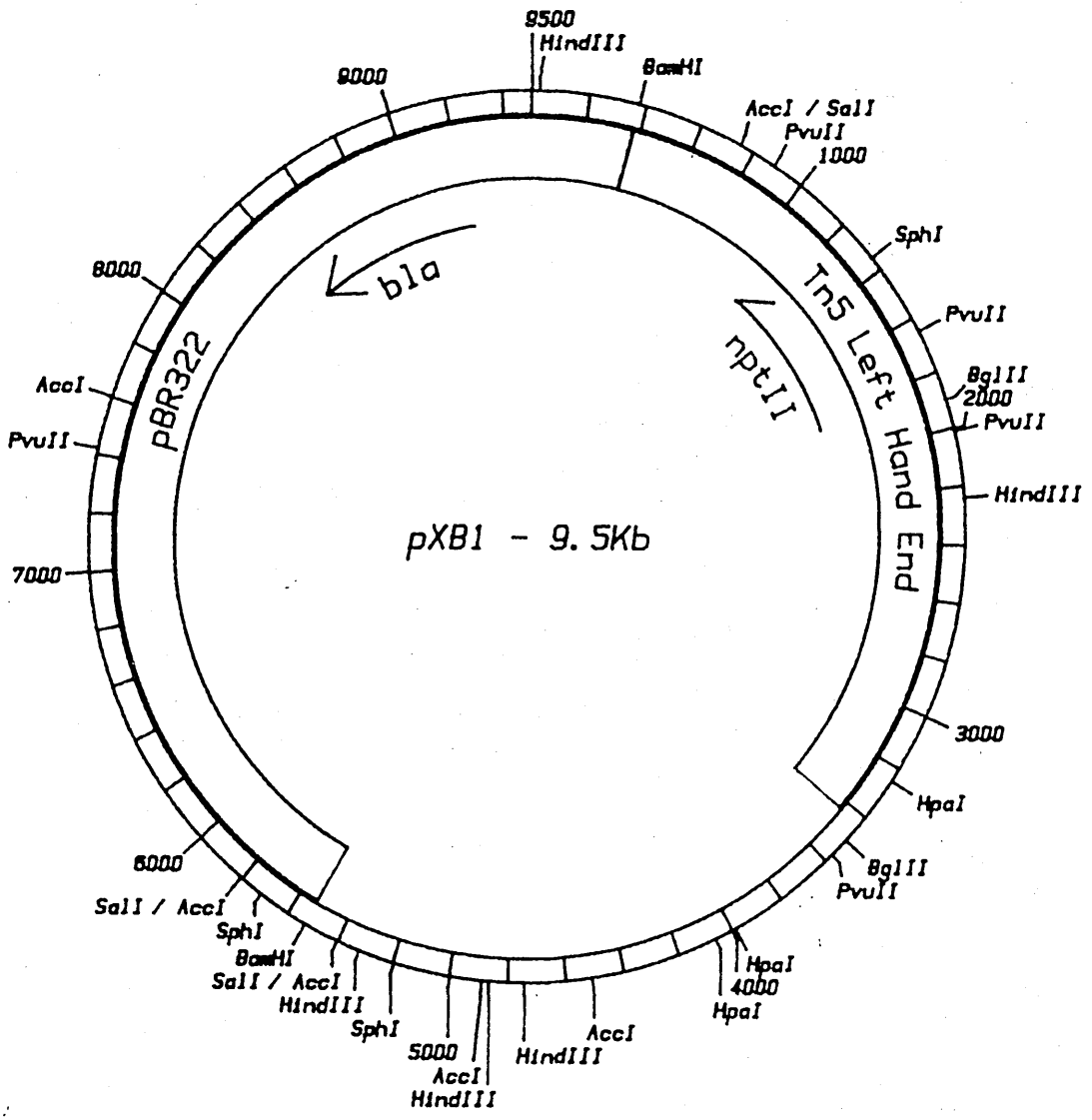


Figure 5.4 Restriction Map of pXB1

Note that only those enzymes which have sites within the cloned chromosomal sequence have been represented. For a tabular summary of all enzymes tested with pXB1 see figure 5.4b.

copies of the lambda cos site (Derek Blake, pers. comm.). A cosmid-based library offers the considerable advantage that large genomic fragments (40-50Kb) may be cloned, thus reducing the total number of individual clones required to yield fair representation of the entire genome. Efficient lambda packaging requires that the cos ends be separated by approx. 37-52Kb (i.e. 78-105% of the size of the lambda genome) (Feiss & Becker, 1983; Murray, 1983). Therefore, when using a 5Kb cosmid vector, the process of encapsidation selects for insert sizes of approx. 30-45Kb.

The pMB1 replicon of pDVcos renders this cosmid compatible with the pCS210, therefore I proceeded to screen the cosmid library for $XerB^+$ clones, by testing for the deletion of lacZ from pCS210 on X-Gal plates: Obviously any XerB complementing cosmid which also encodes the E. coli lacZ gene would produce a "Biologically false negative" in the 210-colour test, and so may go undetected. Clearly this possibility could be averted by making a library from a $LacZ^-$ Xer^+ E. coli strain. Despite this potential problem, the library was screened as follows; CSX4 (pCS210) was grown to saturation, shaking at 37°C, in 20mls L-broth + Tet ($7mg.l^{-1}$) + 0.2% maltose. The culture was harvested by centrifugation (6000g, 2 mins.), and resuspended in 10mls 20mM Tris-HCl, pH 7.5, 10mM $MgCl_2$. Ten microlitres of packaged phage suspension was then added to 0.2ml of these cells and this mixture incubated at 37°C for 30 minutes. After this phage adsorbtion and infection phase, 0.2ml L-broth was added, and incubation continued at 37°C for a further 30 minutes to permit expression of the cosmid bla gene. The total transfection mix was then plated onto 4 plates of "210 medium" + Ampicillin ("210 medium" comprising, DM minimal medium plus casamino acids (1%), glucose (0.2%), X-Gal ($25mg.l^{-1}$), and tetracycline ($7mg.l^{-1}$)). After two days incubation at 37°C, the colonies were counted and scored for their blue/white colour. A total of 582 colonies were recovered of which one was white, the remaining 581 varying in colour from pale to dark blue. The single white clone (B4), plus 20 pale blues, were restreaked onto 210 medium + Ap, and incubated for two days at 37°C; clone B4 maintained its white phenotype, whilst the 20 pale blue clones gave rise to both pale and dark blues. Therefore, in one case out of 582 there had been a complete loss of LacZ activity, whilst maintaining resistance to tetracycline. A STET Mini-prep of plasmid DNA was made from clone B4, analysed on agarose gels, and found to contain a high molecular weight species

(cosmid pB4), plus a plasmid co-migrating with pCS211 (the pCS210 deletion derivative) (data not shown). The cosmid pB4 was clonally purified by transformation of this STET DNA, into DS902 under conditions of dilute DNA concentration. Transformants were selected on L-agar+ Ap, then screened for co-transformation with pCS211 by patching out onto L-agar +Ap, or L-agar+ Ap+Tet. Two out of four transformants tested were Ap^r, Tet^s and a large scale plasmid DNA prep' was made from each; the two isolates of this cosmid clone were labelled pB4 and pB4'. The abilities of pB4 and pB4' to complement the xerB1 lesion were then re-tested by transformation into CSX4 (pCS202), selecting transformants on L-agar+Ap+Cm. STET DNA prepared from pooled transformants was run on a 0.85% agarose gel, and was found to contain a cosmid species plus pCS203, with no pCS202 apparent; identical results were obtained with both pB4 and pB4' (Fig 5.6a). Therefore one can conclude that pB4 does indeed complement the xerB1 lesion, indicating that either the xerB gene, or a suppressor has been cloned. The cosmid pB4 has also been found to complement pCS202 resolution in strains DSX1 (xerB8), DSX2 (xerB9), CSX20 (xerB2), CSX21 (xerB4) and CSX22 (xerB5); however, it fails to complement CSX3 (xerA3) in the same assay (Fig. 5.6b), and likewise it fails to complement CSX11 (xerA5) in a 210 colony-colour assay, indicating that a functional xerA gene is not present on the same pB4 cosmid insert.

5.3 TARGETING pB4 SEQUENCES FOR SUBCLONING via HOMOLOGY TO pXB1

Restriction mapping of pB4 was not undertaken due to its large size, crudely estimated from various digests to be approximately 45Kb (see Fig. 5.7a). Clearly the restriction pattern of pB4 is quite complex, therefore I decided to "Home-in" on the xerB locus by exploiting the flanking DNA cloned in pXB1. I could not predict the sizes of wild type restriction fragments from the map of pXB1, since only one flank of the insertion locus is present. However, by probing Southern blotted pB4 digests with nick-translated pXB1, then those cosmid fragments bearing the equivalent locus could be detected. Obviously, since pDVcos is derived from pBR322, then cosmid vector sequences will also be detected by the pXB1 probe. One way to differentiate any genomic signals from vector signals is to run analogous restrictions of chromosomal DNA alongside the pB4 digests. The chromosomal DNA should, in theory, only hybridise to the genomic

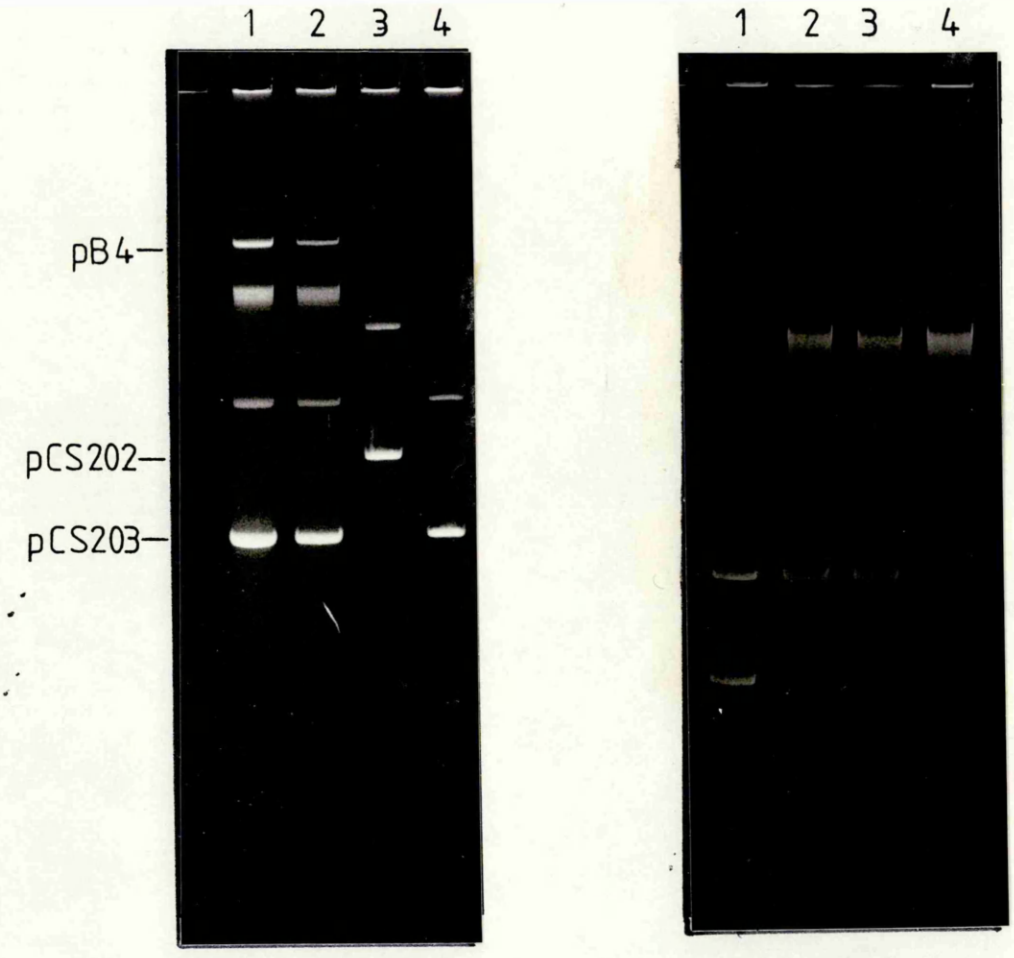


Figure 5.6 pB4 complementation tests in CSX3 and CSX4. Panel (A) illustrates the successful complementation of pCS202 resolution in strain CSX4 by both pB4 and pB4': CSX4/pCS202 + pB4, lane 1; CSX4/pCS202 + pB4', lane 2; pCS202, lane 3; pCS203, lane 4. Panel (B) illustrates that neither cosmid complements the Xer^- phenotype of CSX3: pCS202 + pCS203 mixture, lane 1; CSX3/pCS202 + pB4, lane 2; CSX3/pCS202 + pB4', lane 3; CSX3/pCS202, lane 4.

sequences from pXB1; this assumption being supported by the previous failure to detect any specific hybridisation of pCS80 to blots of wild type chromosomal DNA (see Fig. 4.7, section 4.4). Any pB4 restriction fragments detected which are identical in size to bands detected in the analagous chromosomal digest are therefore very likely to represent genomic fragments from around the xerB locus. Such fragments could then be preferentially sub-cloned and tested for XerB complementation. To this end pB4 and DS947 chromosomal DNA were each restricted with SalI, HindIII, PstI, SphI, SstII, or XhoI. The digests were then run on a 0.7% agarose gel which was stained and photographed prior to Southern blotting onto Nylon membrane (Fig. 5.7a). The following samples were run as molecular weight size markers; plasmid pCS80 linearised with EcoRI (10.2Kb), pCS300 linearised with SmaI (19.5Kb), and pCS300 cut with EcoRI (15.1Kb and 4.3Kb plus the 12.6Kb and 2.5Kb partials). All of these bands, with the exception of the 2.5Kb pCS300/EcoRI partial, should yield signals when probed with pXB1 by virtue of their hybridising to pBR322 and/or Tn5 sequences. As additional markers, a lambda cI857/HindIII digest was also run on the gel. The blot was then probed with a mixture of nick-translated pXB1 plus nick-translated lambda DNA (in a ratio of 9:1 respectively); all procedures were as stated in materials and methods. This somewhat cavalier approach to the probing of a filter (i.e. labelled pBR322, Tn5, lambda, and xerB1 sequences) could have been tempered by subsequently reprobing the same filter (stripped of initial label) with labelled pCS80 + lambda, thus enabling true genomic signals to be identified by negative inference. This was deemed unnecessary since pB4 fragments "Targeted" in this way were rapidly subcloned and shown to complement xerB1 mutants. The autoradiogram generated from this pXB1/lambda probed blot is quite complex (Fig. 5.7b), however it is clear that only HindIII and SphI generate small genomic fragments with (stong) homology to pXB1, therefore these two restrictions will be considered in detail, with reference to any predictions made from the restriction map of pXB1 (see Fig. 5.4):

HindIII: The three genomic HindIII sites in pXB1 predict hybridisation to four DS947/HindIII fragments, one of 120bp, one of 460bp and two boundary fragments of unknown sizes. One of these latter two bands is expected to yield the strongest signal due to its 1.4Kb of homology with pXB1. It is this strong band which must carry at

least part of the wild type xerB locus. In practice this strong signal was found to originate from a 1.9Kb fragment (Fig 5.7b, lane 12), with the predicted 460bp fragment also visible. The predicted 120bp fragment would have run off the bottom of the gel, but may have been barely detectable in any case. The fourth fragment is expected to produce a signal of 40% the intensity of that of the 460bp band (due to its 208bp homology to pXB1) and consequently is not clearly visible. A candidate band of approximately 8Kb can be seen on the original autoradiogram. The salient points regarding the pB4/HindIII digest are that the 1.9Kb, and 460bp genomic fragments appear to be present. The 1.9Kb fragment was therefore subcloned and tested for XerB complementation. It is interesting to note that there is no signal from an 8Kb fragment. Therefore, if the 8Kb band in the DS947 digest is real, then this would indicate that the xerB locus is situated within 10Kb (8+2) of the end of the pB4 insert.

SphI: One SphI site in the genomic sequence of pXB1 predicts that two DS947 fragments should be detected. The position of the site predicts one strong, and one weak signal, with the strong signal emanating from the fragment encompassing the Tn5 insertion site in xerB1. In fact four signals were observed whose approximate sizes are 2.7, 2.8, 5.5 and 6.6 Kb respectively (Fig. 8, lane 14). The major signal is clearly from the 2.8Kb fragment, therefore it is this fragment which was targeted for subcloning from pB4. The two larger bands are faint and are presumed to be partial digestion products. The pB4/SphI digest clearly produces a strong signal from a 2.8Kb fragment, however whether or not the 2.7Kb fragment is present is unclear. Significantly, pB4/SphI generates the same two faint bands (5.5 and 6.6Kb) as observed in the DS947 digest; that these two pB4 fragments are partial digestion products is clear from the original gel photograph (Fig. 5.7a, lane 5).

The remaining enzymes tested all generated a high molecular weight chromosomal band exhibiting strong homology to PXB1. However, notice that SstII digestion also generates a small chromosomal restriction fragment (4 Kb) which yields a relatively weak signal; this band also appears in the pB4 digest. The appearence of two bands in the chromosomal digest cannot be easily interpreted since there are no SstII sites within the chromosomal sequences present in pXB1 (see Fig. 5.4b); it may indicate the presence of a second chromosomal locus with

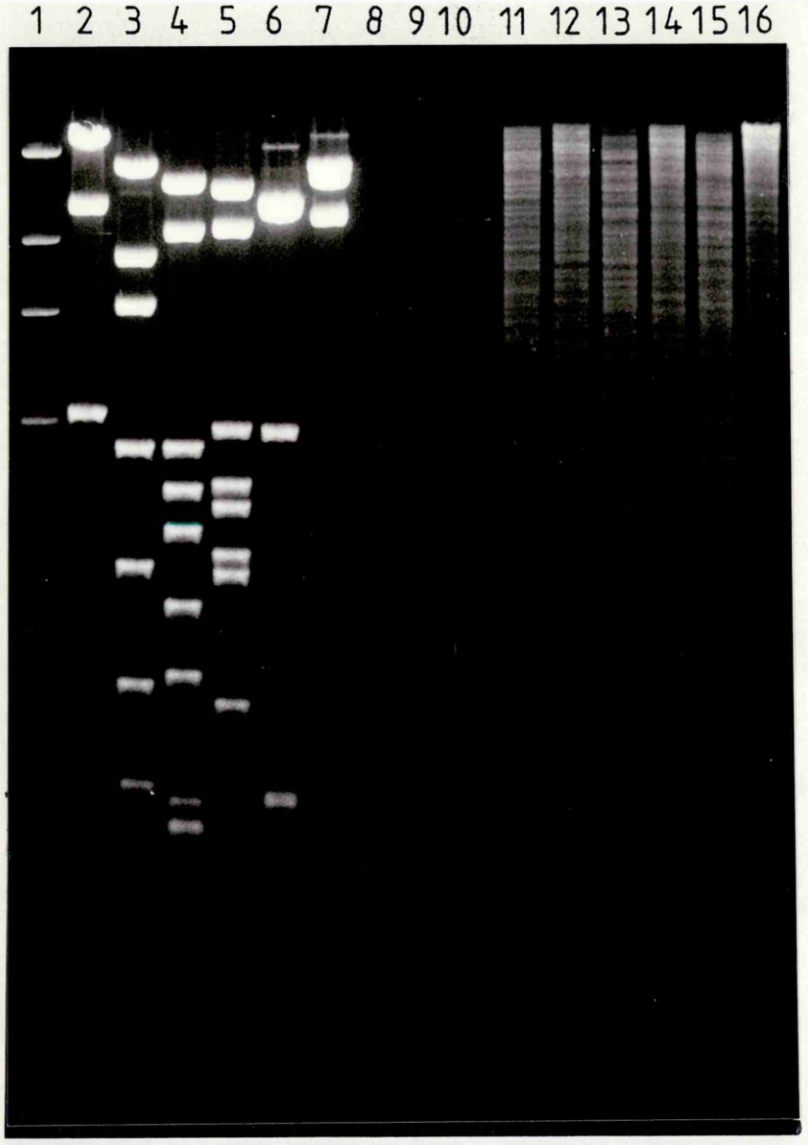


Figure 5.7 Southern hybridisation of parallel restrictions of cosmid pB4 and DS947 chromosomal DNA, probed for homology to pXB1.
Panel A) Ethidium bromide stained gel photographed prior to blotting..

Lane	Sample	Lane	Sample
1	Lambda/HindIII	10	pCS80/EcoRI
2	pB4/SalI	11	DS947/SalI
3	" /HindIII	12	" /HindIII
4	" /PstI	13	" /PstI
5	" /SphI	14	" /SphI
6	" /SstII	15	" /SstII
7	" /XhoI	16	" /XhoI
8	pCS300/SmaI		
9	pCS300/EcoRI		

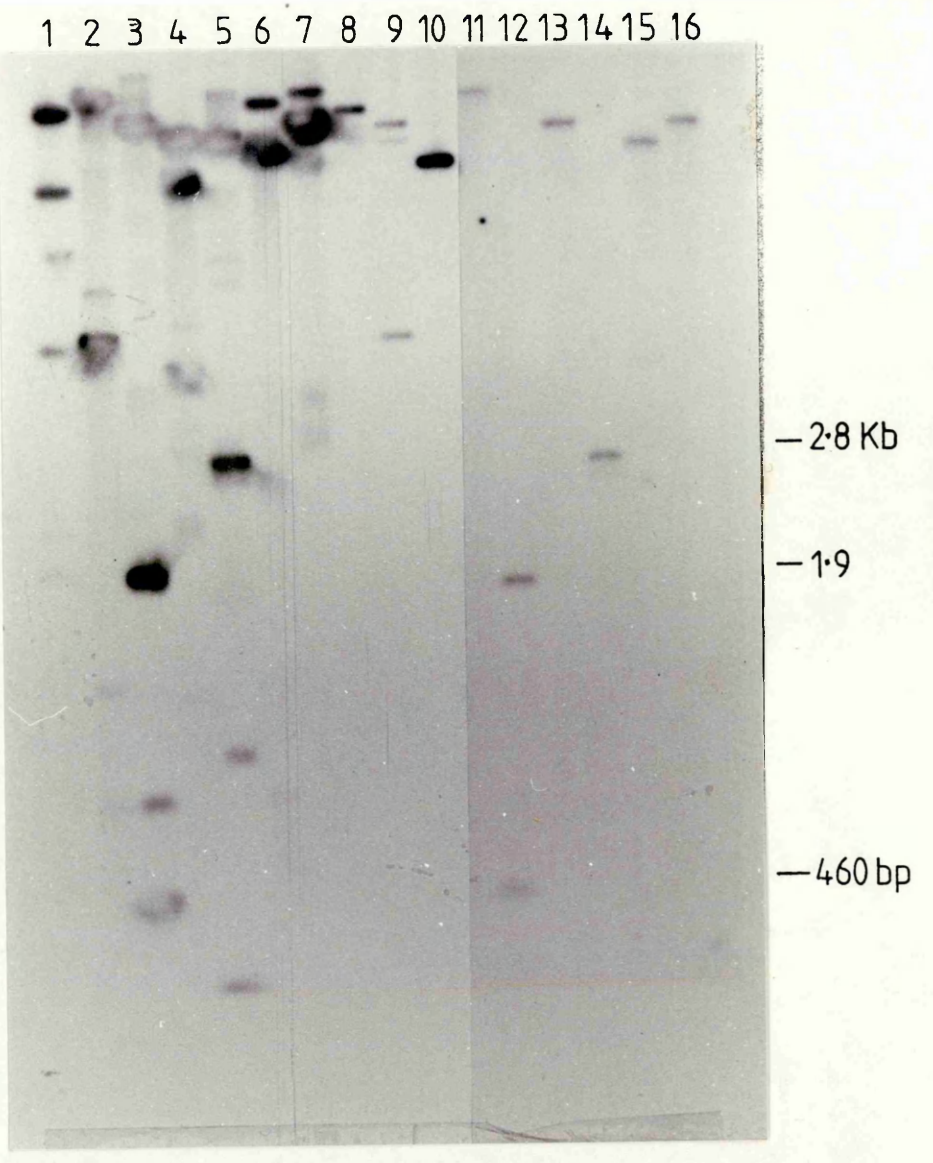


Fig. 5.7 Panel B) Southern blot probed with nick-translated pXB1 plus nick-translated lambda DNA. Notice that the photograph is again a composite of two seperate autoradiograms exposed for different periods of time.

homology to xerB, however further work would be necessary to confirm this.

In summary, the combined HindIII and SphI data strongly suggest that the genomic sequence present in pXB1 is also present in pB4. The DS947 data suggests that the wild type sequence encompassing the site at which Tn5 is inserted in xerB1, is borne on a 2.8Kb SphI restriction fragment, or a 1.9Kb HindIII restriction fragment. Both of these fragments appear to be present in pB4, therefore both were subcloned into plasmid vectors and tested for XerB complementation. In addition a 3Kb fragment identified in pB4/BamHI digests was also subcloned (pB4/BamHI digestion generates only three fragments of approximately 24, 18 and 3 Kb respectively; data not shown). This small BamHI fragment is of special interest since previous blotting data had indicated that the xerB1-linked copy of Tn5 , is inserted into a 3Kb BamHI fragment in the chromosome of DS947 (see section 5.2).

5.3.2 ISOLATION OF XerB$^+$ SUBCLONES FROM pB4

Having identified certain small pB4 restriction fragments which are likely to carry at least part of the xerB gene, I set about subcloning them, but in order to guard against erroneous prejudice I proceeded as follows: pB4 was cut with either BamHI or SphI, treated with CIP, then these digests ligated to either BamHI-restricted pCS100, or SphI-restricted pUC18 as appropriate: the vector pCS100 was constructed by cloning the rom gene from ColE1, as a 309bp HpaII fragment, into the NarI site of pUC8. Both ligations were then screened for XerB$^+$ clones by transformation into CSX4 (pCS210), followed by plating out onto 210-medium + Ap. Obviously, any clones which effect complementation should be readily identifiable by their white colony colour. Of four white clones recovered from the BamHI/pCS100 ligation, all were found to contain two plasmid species; one comigrating with pCS211, and the other being approximately 6Kb in size. The "targeted" 3Kb BamHI fragment from pB4, when cloned into pCS100 (3Kb), would of course generate a plasmid of 6Kb. The 6Kb plasmid from one such clone was designated pCS110, and was clonally purified from the reporter plasmid by transformation of STET DNA into DS902, selecting for Ap^r transformants and screening for Tet^s clones. The structure of pCS110 was then determined by digestion with BamHI, SphI and with HindIII and EcoRI. The restriction map of the 3Kb BamHI insert from pCS110, plus its orientation relative to the

pCS100 polylinker, are shown later in Fig 5.15.

With regards to the SphI subcloning, from a total of seven white clones analysed, all appeared identical with respect to their plasmid content, comprising pCS211 plus a second species of approximately 5.5 Kb. The 5.5Kb plasmid from one such clone was designated pCS111, and was clonally purified as described for pCS110. The restriction map of pCS111 has been determined for various enzymes including SphI, BamHI, HindIII, SalI etc. (See Fig. 5.15). Both pCS110 and pCS111 were then re-tested to confirm their $XerB^+$ status; in the standard CSX4 (pCS210) colony colour assay, both scored as $XerB^+$, whereas pUC18, pCS100, and pAT153 all fail to complement.

5.3.3 DEFINING THE MINIMUM COMPLEMENTING SEQUENCE

From restriction analyses of pCS110 and pCS111, it appears that their inserts overlap considerably (see Fig. 5.15), and that both contain the 1.9Kb HindIII fragment originally targeted for subcloning in section 5.3.1 (b). Therefore this HindIII fragment was sub-cloned from pCS111 as follows; HindIII-restricted pCS111 was run on a 0.8% gel, and the 1.9Kb fragment recovered from the gel as described in materials and methods. The purified fragment was then ligated to HindIII restricted pUC18, and this ligation was transformed into DS941 (lacZ delta M15), and then plated out onto L-agar+ Ap, X-Gal, and IPTG (the use of X-Gal in this case being to screen for pUC18 clones with inserts which disrupt their lacZ' coding sequence; such disruption generally resulting in a failure to complement the chromosomal lacZ delta M15 mutation leading to a phenotypically $LacZ^-$ clone). A number of Ap^r, $LacZ^-$ clones were screened for size by single colony gels, then their insert size and orientation confirmed by restriction with HindIII, AccI, and by HindIII/AccI double digestion. In this way pCS112 was isolated (See Figs. 5.11 and 5.15). This plasmid was also found to complement pCS210 resolution in CSX4 according to the 210-colony colour test. At this point I initiated the DNA sequence analysis of this 1.9Kb HindIII fragment (see chapter 7), rather than attempting to further reduce the complementing function "Blind". The resultant DNA sequence predicted a number of useful restriction sites, which I confirmed empirically, then exploited for the generation of various pCS112 sub-clones which were then tested for complementation in CSX4. The strategies employed to construct each subclone are detailed explicitly in Figs. 5.11, 5.12, and 5.13. The

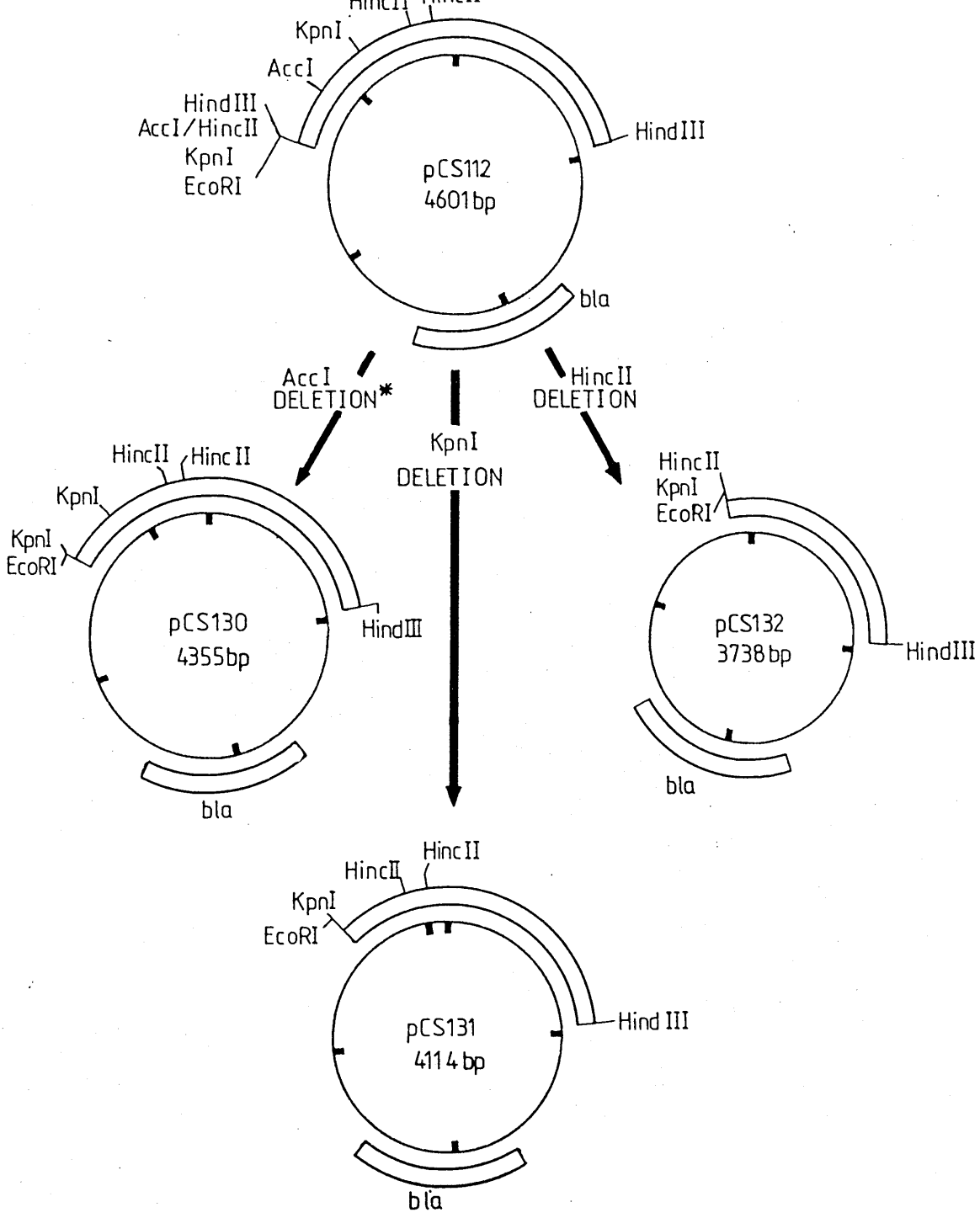


Figure 5.11 Construction of pCS130, pCS131 and pCS132 by Deletion of sequences from pCS112

The KpnI and HincII deletions were achieved by restriction followed by re-ligation at low DNA concentrations. * The AccI deletion required "Filling-in" of non-compatible AccI ends with Klenow + dNTPs; the resultant blunt-ends were then ligated to generate pCS130.

pCS130	carries the	1687bp	HindIII/Acc	fragment	from pCS112
pCS131	"	" 1467bp	HindIII/KpnI	"	" "
pCS132	"	" 1070bp	HindIII/HincII	"	" "

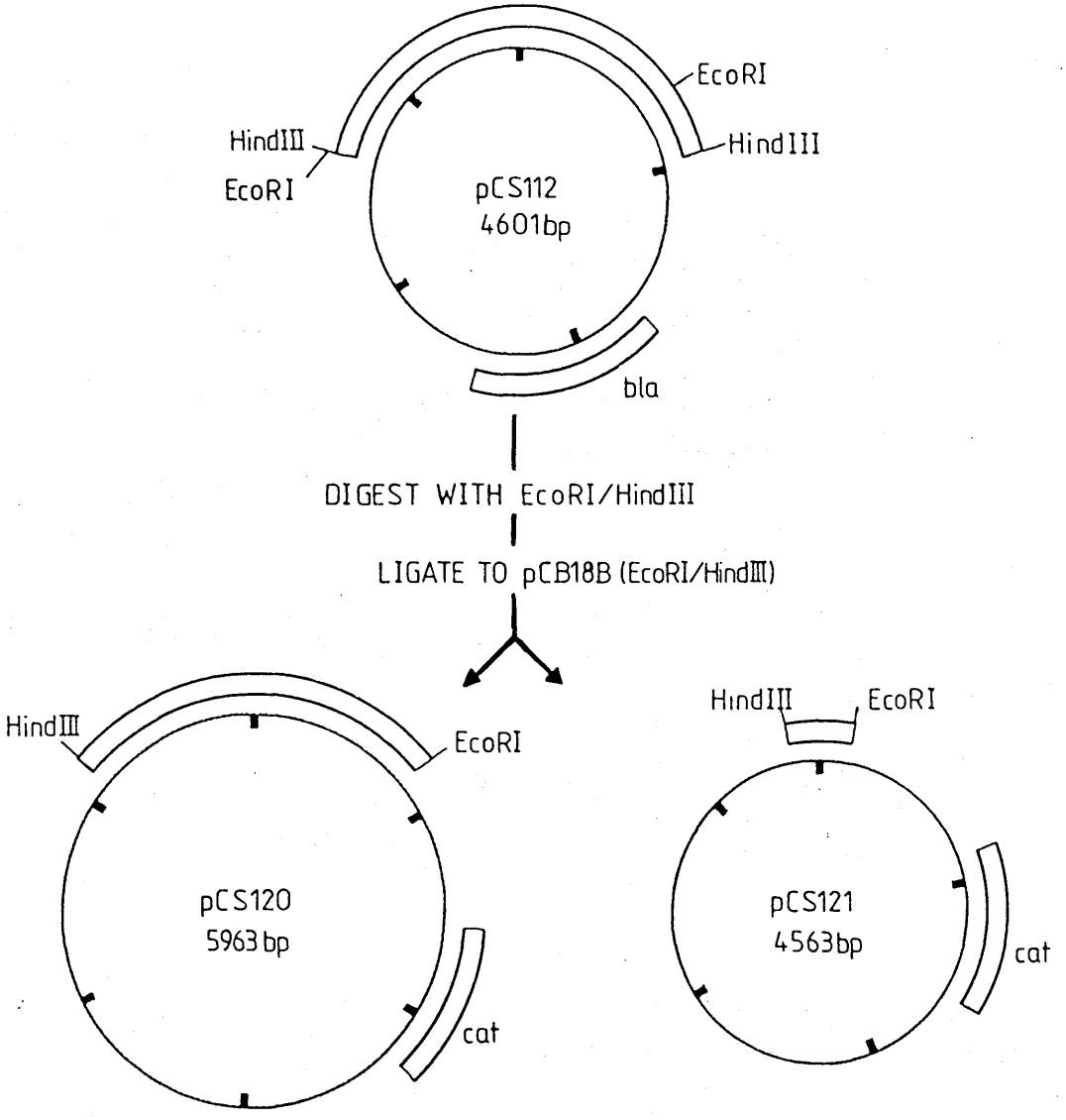


Figure 5.12 Construction of pCS120 and pCS121

pCS120 carries the 1641bp EcoRI/HindIII fragment from pCS112, whilst pCS121 contains the 274bp fragment.

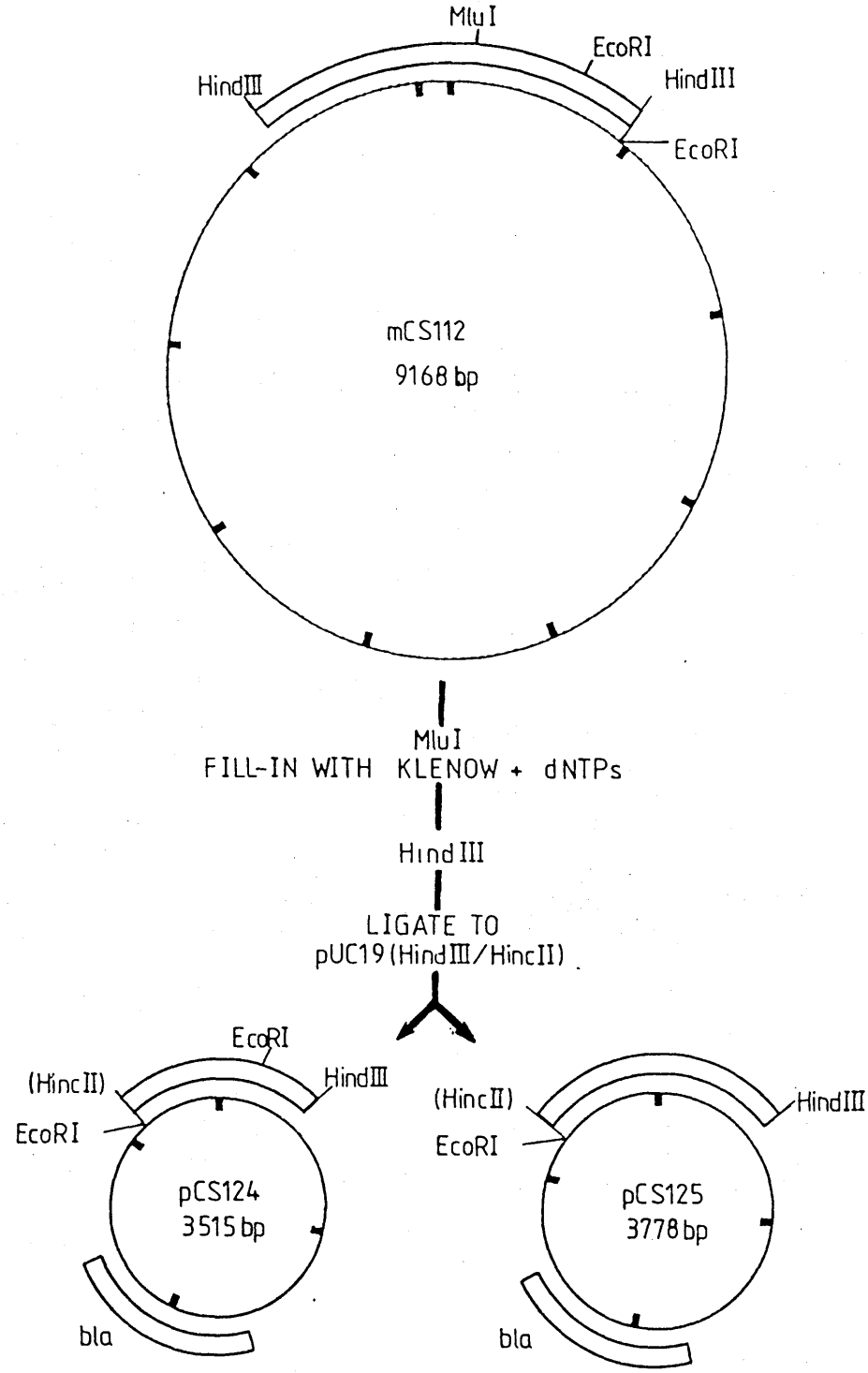


Figure 5.13 Construction of pCS124 and pCS125

mCS112 is an M13mp19 derivative containing the 1.9Kb HindIII fragment from pCS111 cloned into the polylinker for the purpose of DNA sequence determination. There is a single MluI site in mCS112 situated 829bp from the RIGHT-HAND HindIII site. The sub-clone pCS124 contains the 829bp fragment, whilst pCS125 carries the 1092bp HindIII/MluI fragment.

subclones pCS120, pCS121, pCS124, pCS125, pCS130, pCS131, and pCS132 were each tested for complementation of pCS210 resolution in CSX4 *via* a colony-colour test assay. The results of these complementation tests are summarised in Table 5.1; in each case the Xer phenotype conferred by each sub-clone has been confirmed in a pCS202 gel resolution assay (data not shown). The results clearly show that pCS110, 111, 112 and pCS130 all complement in CSX4, but that the remaining sub-clones do not. It is very important to note here that the observed efficiency of pCS112 and pCS130 complementation depends upon the medium onto which the test-plasmid transformation is plated; on L-agar+Ap+Tet partial complementation is observed in pooled transformants, but on DM minimal agar + Glucose, Casamino acids, Ap + Tet efficient complementation is observed, this phenomenon will be discussed shortly. Unless otherwise stated, all future complementation tests were plated onto a DM/Casamino acids medium. The results of this complementation analysis are summarised in figure 5.15. The minimum sequence identified as retaining the complementing function is the 1687bp *Hind*III/*Acc*I fragment in pCS130. Furthermore the data suggests that the functional *xerB* locus must extend beyond the *Eco*RI site at one end, and beyond the *Kpn*I site at the other; i.e. it must be a minimum of 1187bp in size. The correspondence between the pXB1 restriction map and those of pCS111, and pCS112 (Fig. 5.15) confirms that the complementing function cloned is, by definition, the *xerB* gene as opposed to a suppressor allele. I should add that all of these complementation tests were performed in a $RecA^+$ background therefore one cannot state unequivocally that complementation is effected *in trans*. Obviously recombination between the chromosomal and plasmid borne sequences can occur, such recombination could regenerate a functional *xerB* gene even when both chromosomal and plasmid copies are defective. Whilst homologous recombination does proceed in CSX4, it does so at low efficiences due to its $RecF^-$, $RecE^-$ status. Furthermore, given the relative position of the Tn5 insertion in *xerB*1 (see pXB1), one would expect recombination to be almost as efficient at regenerating an *xerB* gene from either pCS120 and/or pCS131 as it is from pCS130. This is patently not the case, therefore it seems reasonable to conclude that pCS130 complementation in CSX4 is not dependent on recombination with the *xerB*1 locus, and thus is likely to be effected *in trans* by a functionally intact *xerB* locus.

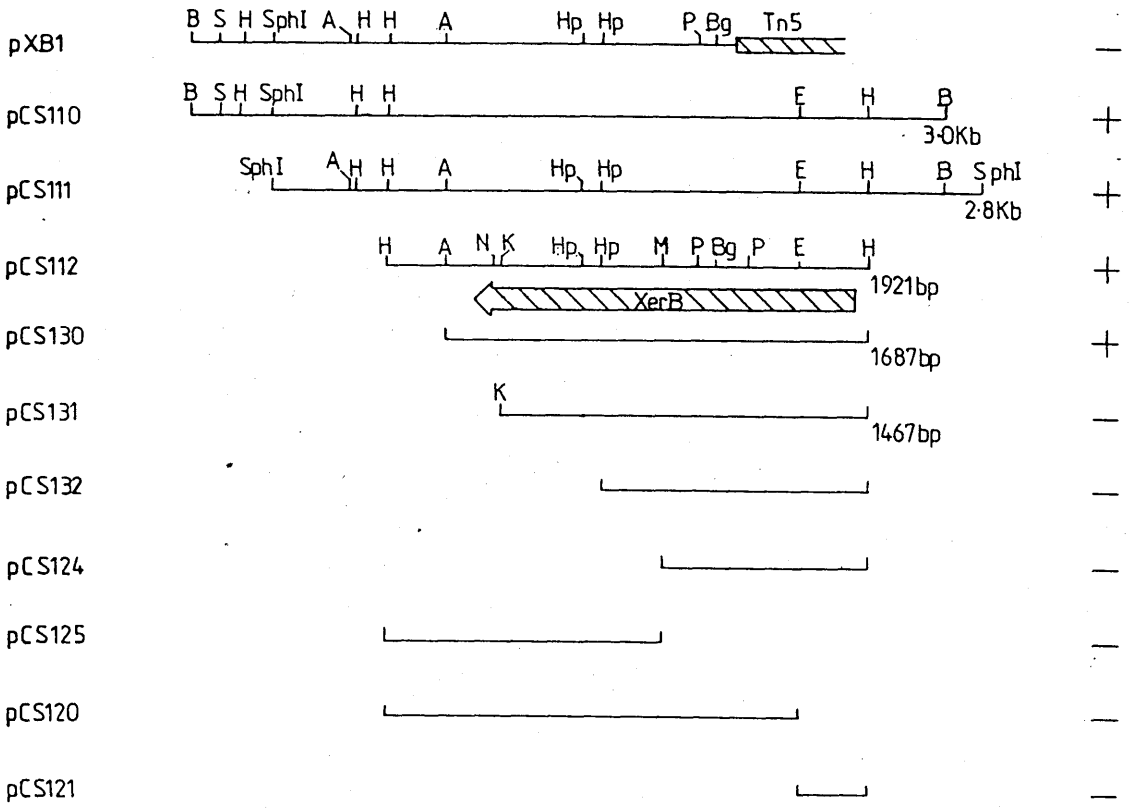


Figure 5.15 Complementation of xerB1 by Various Sub-Clones

Each sub-clone was tested for complementation of pCS210 resolution in CSX4: "+" indicates complementation whereas a "-" indicates failure to complement. The complete insert is shown in each case, with the exception of pXB1 where only a portion of the Tn5 sequence is represented. Only those restriction-sites which have been confirmed empirically are shown:

ENZYME CODE:	A= AccI	H= HindIII	N= NarI
	B= BamHI	Hp= HpaI	P= PvuII
	Bg= BglII	K= KpnI	S= SalI/AccI
	E= EcoRI	M= MluI	SphI= SphI

Note that the extent of the xerB coding sequence (identified by DNA sequence analysis- see Chapter 7) is indicated by a broad-hatched arrow. The sizes of the inserts in pCS110, 111, 112, 130 & 131 are as indicated.

FOR INFORMATION ONLY: The relative orientations of each insert with respect to the vector's polylinker (and plac) was determined for pCS110 & pCS111. In the case of pCS110 the vector's lac promoter drives transription rightward through the insert as shown above, i.e. from SalI to SphI. The opposite is true in pCS111, with plac initiated transription running from right to left (i.e. from BamHI to EcoRI).

Table 5.1 Complementation of *xerB1* by Various Plasmid Sub-clones

Test Strain	Test Plasmid	Colony Phenotype	Xer Phenotype
CSX4 + pCS210	pCS110	White	+
"	" 111	"	+
"	" 112	"	+
"	" 120	Blue	-
"	" 121	"	-
"	" 124	"	-
"	" 125	"	-
"	" 130	White	+
"	" 131	Blue	-
"	" 132	"	-

5.4 DOES THE MINIMUM xerB CLONE COMPLEMENT ANY OTHER xer MUTANTS. ?

The cloned xerA gene was found to complement two out of the 11 mutants isolated, thus consigning nine mutants to a "non-XerA" group. The cosmid pB4 complements the Xer^- phenotypes of the six members of this group against which it has been tested. This indicates either that all of these mutants are $XerB^-$, or that pB4 carries a number of other xer genes. In order to help distinguish between these possibilities, it was necessary to test the minimum $xerB^+$ sequence against all of the other mutants. Ideally the 1687bp HindIII/AccI fragment should have been used, however, at the time of the analysis this had not yet been identified thus it was the 1921bp HindIII fragment which was tested. The complementation assay was the standard transformation of pCS112 into a mutant strain carrying pCS202, followed by agarose gel analysis to determine the status of the reporter plasmid. All nine "non-XerA" mutants were tested in this way, and in each case, efficient complementation was observed (Fig. 5.16). It would therefore appear likely that all the xer genes identified by our mutant isolation programme have now been cloned.

5.5 MAPPING THE xerB LOCUS ON THE E. coli K12 LINKAGE MAP

A long term goal in a number of laboratories is to develop a characterised library of the E. coli K12 genome; i.e. a set of restriction mapped, overlapping clones which represent the entire chromosome. Such a library should allow us to "home-in" on a map location and thus rapidly clone a gene whose map position is known. Conversely, one should be able to match cloned genes to within the library, and thus determine a very accurate map position for the cloned locus. Dr. Knott in Oxford is currently developing a characterised cosmid library, and generously agreed to attempt to match pB4 to the clones characterised to date. The procedure involves sizing all cosmid restriction fragments generated by HinfI digestion, then feeding this "HinfI fingerprint" into a computer program which searches for comparable restriction spectra amongst a database of clones. Fortunately, the search for pB4 like sequences proved successful, with pB4 mapping to a region from which eight contiguous cosmid clones had been characterised; indeed pB4 served to join two smaller contiguous sequences (contigs), to form Contig 48 (Fig. 5.17). Dr. Knott had previously found that cosmids 157/1 and 296/57 both hybridised to a synthetic oligonucleotide prepared against the pyrB

Figure 5.16 Complementation of Independent Xer⁻ Mutants by pCS112

Strain	Mutant Allele Designation	pCS112 Complementation
CSX4	*xerB*1	+
" 5	*B*2	+
" 6	*B*3	+
" 7	*B*4	+
" 8	*B*5	+
" 9	*B*6	+
" 10	*B*7	+
DSX1	*B*8	+
DSX2	*B*9	+
CSX3	*A*3	-
CSX11	*A*5	-

Complementation was scored in a pCS202 gel resolution assay.

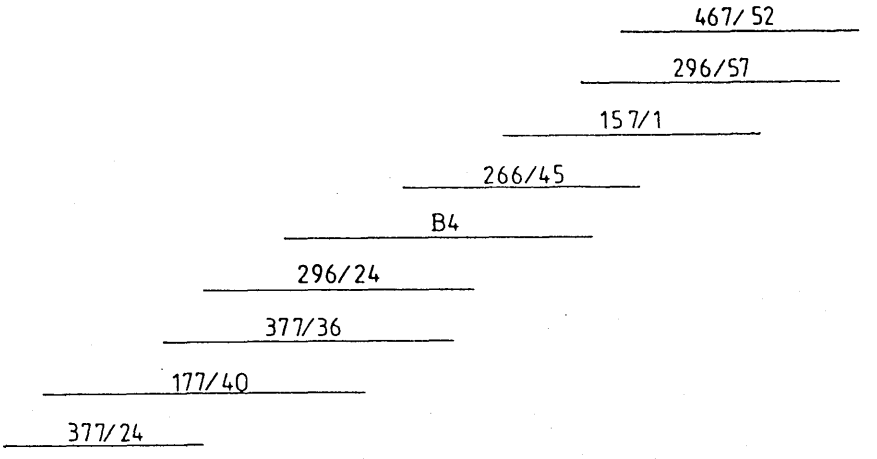


Figure 5.17 Overlapping Cosmid Inserts Corresponding to Contig 48

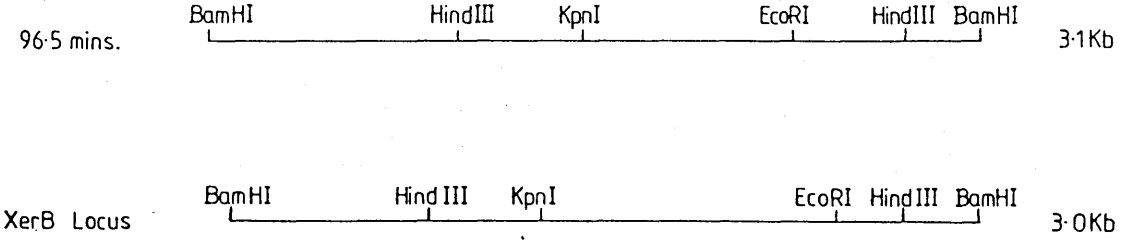


Figure5.18 Comparison Between the Restriction Map of the Cloned *xerB* Locus With That of the 96.5min Region of the K-12 Chromosome.
In addition note that there are no *PstI* sites in either fragment.

gene sequence. The pyrB locus (encoding aspartate carbamoyl-transferase) has been mapped to 96.5 mins (see Bachmann, 1983), thus contig 48 had been tentatively assigned to this region of the E. coli K12 linkage map. Dr. Knott kindly supplied me with her clones from contig 48 which I proceeded to test for the presence of the xerB, pyrB, and argI genes by means of genetic complementation. I failed to recover cosmid 266/45 from the material received, however, as will become clear, this particular clone would probably yield redundant information. Cosmids were tested for xerB by complementation of pCS210 resolution in CSX4, as assayed by the 210 colour test. The results indicate that only pB4 and 157/1 encode a functional xerB gene, the remaining six cosmids all fail to complement CSX4 (Table 5.2). Each cosmid was then tested for its abiliy to complement the uracil requirement of a pyrB mutant strain. In addition I tested each clone for the presence of the argI gene, whose map position is estimated as 96.6mins (Bachmann, 1983). The argI gene encodes ornithine carbamoyl transferase, as does its duplicate gene argF (7 mins.), therefore I tested for each cosmid's ability to complement the arginine requirement of an argI, argF double mutant. Strain NGX2 is pyrB, argI, and argF, and is therefore a suitable host strain for both complementation tests. NGX2 was transformed independently with each of the eight cosmids, or with pBR322 as a control, transformants were then selected on DM minimal agar plus Glucose, casamino acids, uracil and Ap. Eight colonies from each transformation were patched out sequentially onto three media composed of DM minimal agar + glucose and Ap, plus the following supplements;

1) Proline, Leucine, Arginine, and Uracil ("Ura+Arg" medium)
2) Proline, Leucine, and Arginine ("Arg" medium)
3) Proline, Leucine, and Uracil ("Ura" medium)

The plates were then incubated at 37°C for two days before being scored for the presence or absence of growth. The results are summarised in table 5.3 from which it seems clear that pB4, 157/1 and 296/57 all complement both the uracil and arginine requirements of NGX2, and thus would appear very likely to carry the pyrB and argI genes. Of the remaining 5 cosmids tested, 3 can be said to complement neither requirement, those being 377/24, 177/40, and 467/52. No conclusions can be drawn with regards to 296/24, or 377/36. The

Table 5.2 Complementation of <u>xerB1</u> by Various Cosmids from Contig 48

Test Strain	Cosmid	Colony Phenotype	XerB Phenotype
CSX4 + pCS210	p377/24	Blue	-
"	p177/40	"	-
"	p377/36	"	-
"	p296/24	"	-
"	pB4	White	+
"	p157/1	White	+
"	p296/57	Blue	-
"	p467/52	Blue	-
"	pBR322	Blue	-

Table 5.3 Complementation of Arginine and Uracil Prototrophy in NGX2 by Cosmids from Contig 48

Strain	Cosmid	Growth on DM minimal + supplements			Strain Phenotype	
		Ura+Arg	Ura	Arg	PyrB	ArgI
NGX2	p377/24	+	-	-	-	-
	p177/40	+	-	-	-	-
	p377/36	-	-	-	N.D	N.D
	p296/24	-	-	-	N.D	N.D
	pB4	+	+	+	+	+
	p157/1	+	+	+	+	+
	p296/57	+	+	+	+	+
	p467/52	+	-	-	-	-
	pBR322	+	-	-	-	-

NGX2 was transformed with each cosmid, selecting Ap^r clones on DM minimal, Cas-amino acids + Ap. Several colonies were then patched onto each test medium. The failure of p377/36 and p296/24 transformants to grow on even Ura+Arg medium seems to reflect the genetic instability of these cosmids since neither strain would re-grow when patched onto DM, Cas-amino acids + Ap, but would grow in the absence of Ap; (N.D.= Not Determined).

Table 5.4 Cosmid Genotypes Established by Genetic Complementation

Cosmid	Established Genotype
pB4	*xerB*, *pyrB*, *argI*
p157/1	*xerB*, *pyrB*, *argI*
p296/57	*pyrB*, *argI*

established genotypes of each cosmid from contig 48 are summarised in table 5.4. These genotypes would suggest that the overlaps between "cosmid fingerprints" shown in figure 5.17 are misleading, since 296/57 and pB4 do not apear to overlap at all, and yet both express PyrB and ArgI. Such errors from the fingerprints may arise as a result of HinfI-fragment truncation by cosmid vector sequences; thus two inserts may share considerable overlap at their ends, but this might not be obvious if their terminal fragments have been truncated to different lengths. Obviously this method of fingerprinting can only detect overlaps which share at least one **intact** HinfI fragment, however, in practice multiple common-fragments are required to assign an overlap.

The genetic data strongly suggests that xerB must be located in very close proximity to pyrB and argI, a very conservative estimate would be to within 0.5 mins, or 15-25Kb. However, from this data we cannot distinguish the order of genes on the linkage map, save to say that it must be either xerB-pyrB-argI or pyrB-argI-xerB; therefore I concluded that xerB maps to between 96-97 minutes on the K-12 linkage map.

This work had been completed at the time of publication of the restriction map of the E. coli K-12 chromosome (see chapter 4). If one searches this map for restriction "fingerprints" compatible with that of the xerB locus, then the only convincing match is found at 4560Kb, equivalent to 96.5 mins on the K-12 linkage map (Fig 5.18). This comparison confirms the genetically established map position, and thus highlights the potential of the Kohara map in the prediction of chromosomal map positions from simple restriction data.

5.6 DISCUSSION AND CONCLUSIONS

The xerB gene has been cloned, and defined genetically to within a 1.69Kb HindIII/AccI fragment. The slightly larger 1.92Kb HindIII fragment cloned in pCS112 has been found to complement all nine mutants in the "non-XerA" complementation group. On the assumption that this fragment encodes a single xer gene (an assumption supported by sequence data, chapter 7), all nine of these mutants have been assigned to the xerB complementation group, and designated alleles xerB1-xerB9 (Appendix 1). However, as will be discussed in chapter 7, I cannot, at this stage, formally preclude the existence of other small xer genes within this fragment. Had more time been available,

then all those sub-clones listed in figure 5.15, would also have been tested for their abilities to complement xerB2-xerB9. Comparisons of the restriction maps of the cloned xerB and xerB1 alleles, coupled with the xerB1 complementation data (Fig. 5.15), makes it quite clear that the Tn5 insertion in strain CSX4 is within the functional xerB locus.

An apparent inconsistency exists between the inhibitory affects on Xer activity observed with pGS38 and pBAD (chapter 4), compared with the highly efficient complementation effected by pCS112. However, as alluded to earlier, complete resolution of pCS202 to pCS203 is observed only when pooled transformants of CSX4 (pCS202) + pCS112 are taken off DM minimal agar, whereas a partial resolution phenotype is observed when L-agar cultures are used. The 210 colony colour test is affected in a similar fashion; i.e. pCS112 transformants grow as uniform white colonies on 210-medium (DM minimal based), but grow as mixed blues and whites on L-agar based medium, the basis of this medium affect would seem unlikely to also cause "partial complementation" with pGS38, since in this case mixed blues and whites are observed regardless of the medium employed. Whilst these complementation tests are fraught with the possibility that clonal segregation of test and reporter plasmid may lead to a mixed phenotype, it seems irrefutable that pBAD and pGS38 can exert an inhibitory affect on the cer-specific recombination in wild type strains. Clearly one would now wish to re-test these inhibitory affects on minimal medium, and also to test for any inhibitory affects of pCS112 in a wild type strain which are not apparent in the complementation assay. If such tests do not resolve the anomaly, then one might investigate other possible explainations for the difference. These might include:

1) That pCS112 exerts less "Stress" on its host than either pBAD or pGS38, perhaps as a simple consequence of the plasmid construction, or perhaps as a result of the function(s) it expresses.

2) Increased levels of XerB protein, resulting from high gene dosage, might act to suppress the inhibitory influence of the vector; the simplest route for which would be if vector-inhibition is effected via a reduction in host xerB gene expression. In order to test this hypothesis, one might supply the cloned xerB gene on a plasmid

compatible with pBAD, then test for their combined affect in a wild type host. Some of these experiments may help to unravel this intriguing phenomenon.

Finally, the XerB gene has been mapped to 96.5 mins. on the E. coli K12 linkage map. The interval between 96-97 mins. contains only eight previously identified loci, none of which are obviously associated with a DNA recombination function; they are fdp (Fructose diphosphatase, 96.0 mins), corB (Mg^{++} transport system I, 96.4 mins), pyrB and pyrI (aspartate carbamoyltransferase catalytic and regulatory subunits respectively, 96.5 mins), argI (96.6 mins), sbaA (regultion of serine and branched chain amino acid metabolism, 96.7 mins), valS (96.8 mins), and leuX (leucine tRNA, 97.0 mins) (Bachmann, 1983).

CHAPTER 6

SEQUENCE OF THE *xerA* GENE

6.1 INTRODUCTION

The *xerA* gene has been delineated genetically to within a 0.93Kb *Sph*I/*Acc*I fragment, the determination of whose DNA sequence was considered a priority for several reasons:

1. The *xerA* sequence may be compared against sequence databases, chiefly to ascertain whether or not the locus has been sequenced previously having been identified *via* some other host function.
2. The DNA sequence can be used to predict the precise structural organisation of the *xerA* gene, plus the primary structure of its polypeptide product (assuming that the gene product is indeed a protein).
3. The DNA and inferred protein sequences may then be searched for sequence "motifs" associated with particular functions; e.g. for the presence of a "*cer*-like" site in the DNA, or for a "Helix-Turn-Helix" DNA binding motif in the polypeptide, or for similarities to other known recombinases etc..

6.2 SEQUENCING STRATEGY

The sequencing strategy centered upon the generation and cloning of a series of nested deletions which enabled the ordered, and thus rapid generation of contiguous sequence data. The approach employed was that of Henikoff (1985), which exploits two very useful features of *E. coli* Exonuclease III (ExoIII):

1. Processive digestion of DNA ends at a uniform rate.
2. The 3'-5' exonucleolytic activity of ExoIII is dependent on a "recessed" 3'-OH end, i.e. one which is part of a duplex. It therefore fails to initiate exonucleolytic degradation at a DNA end with a 4bp 3' protrusion.

Thus if a DNA duplex is digested with two different restriction enzymes such that it has a "protected" 3' protrusion at one end, and a "susceptible" 5' protrusion at the other, then subsequent ExoIII digestion will proceed in a uni-directional fashion. The importance of this is that a vector with a cloned insert can be subjected to ExoIII deletion in such a way as to delete into the inserted sequence whilst protecting the vector, thus deleted molecules may be cloned after simple re-circularisation. By performing an ExoIII deletion time-course, stopping samples at specific intervals, then a series of nested deletions are generated; the technique is illustrated in a flow diagram in figure 6.1. Obviously this approach has the considerable advantage of enabling a given strand to be sequenced in a defined

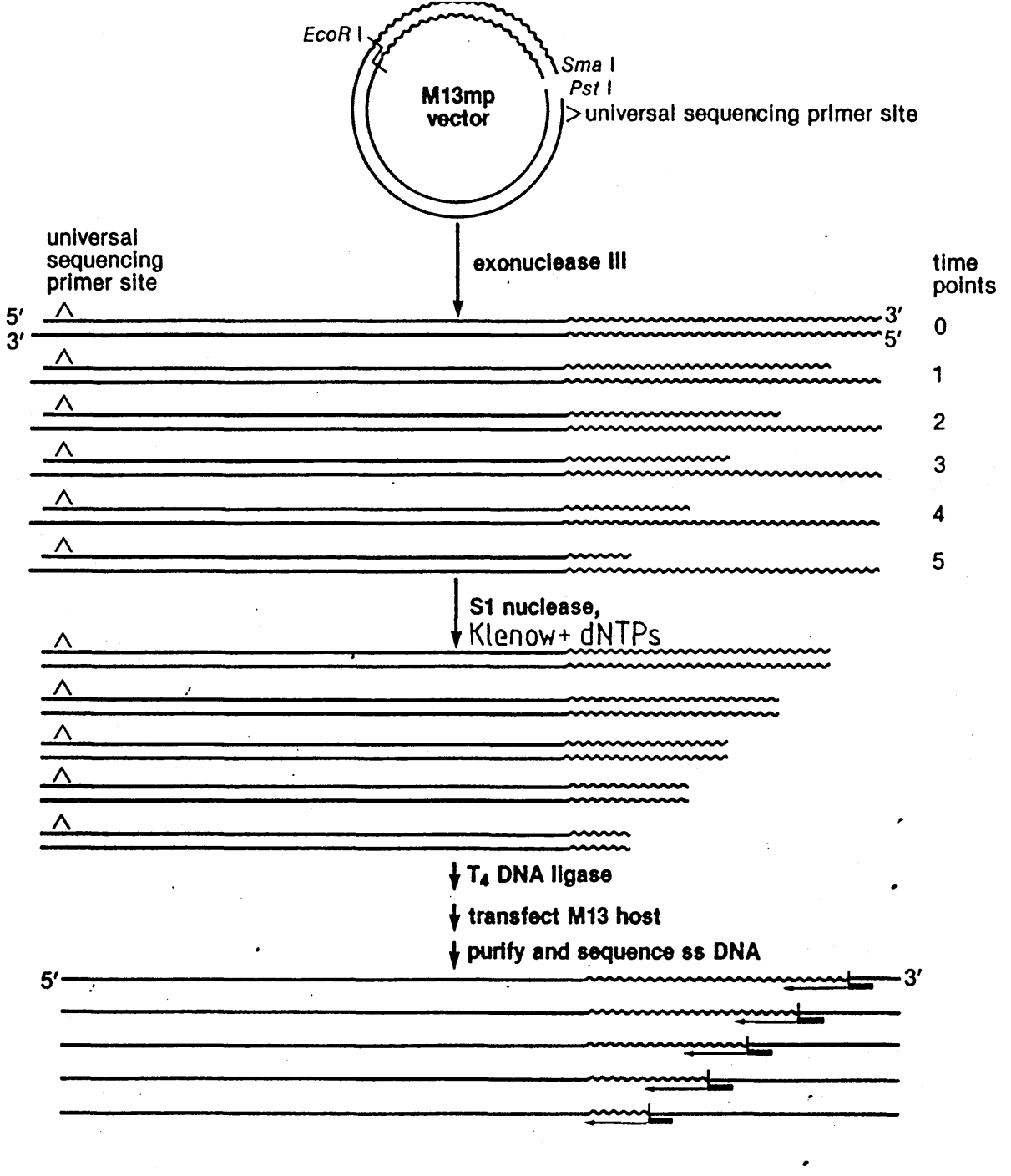


Figure 6.1 Flow Diagram Illustrating the ExoIII-Deletion Technique

order, thus the complete single strand sequence can be determined in a highly efficient manner. In order to obtain the second strand sequence, one may clone the same fragment in the opposite orientation with respect to the "Primer hybridisation site", then begin deleting inwards from the alternative end. In addition to their use as sequencing templates, deletions generated in this fashion may be employed to further delineate the cloned function *via* complementation tests.

One strand of the XerA sequence was determined using ExoIII deletion derivatives of the wild type *Sph*I/*Acc*I fragment as templates, whilst the second strand was determined from deletion derivatives of the *xerA*3::Tn5 insertion locus originally cloned in pCS300; this had the advantage of precisely defining the Tn5 insertion sequence. The two substrates used for ExoIII deletion were mCS320 (wild type; Fig. 6.2), and mCS301 (*xerA*3; Fig.6.2). In addition two specific subclones were required to complete the *xerA*3 strand, those being mCS304 (Fig. 6.3a) and mCS305 (Fig. 6.3b). The deletion protocol employed was as published by Henikoff (1985), the sole modification being that ExoIII digestions were performed at 30°C, as opposed to 37°C; this lower reaction temperature was found to result in a more uniform deletion of the available substrate. The rate was calibrated by a deletion time-course on an *Sst*I/*Bam*HI-restricted mCS320 substrate, followed by S1 digestion, after which samples were run on an agarose gel, and the mode fragment size for each time point calculated from the electrophoretic mobility of the major band (Fig. 6.4). In this way the deletion rate was calibrated to between 130-150bp.min^{-1} per susceptible end at 30°C. Therefore in order to generate suitable overlaps for the purposes of accurate sequence determination, time-points were taken at 2 min intervals. After S1 digestion, end-repair, and re-ligation, the ligation mix was transfected into JM101 (Yanisch-Perron *et al*, 1985) then plated out as described in materials and methods. The phage plaques recovered were picked and resuspended in 0.5ml 2YT medium for long term storage at 4°C. Individual clones were grown up, by infecting 2.5ml of faintly turbid JM101 with 50 microlitres of stored phage, followed by incubation with vigorous shaking at 37°C overnight. From these cultures a 1.5ml aliquot was microfuged, 5 mins., and the whole-cell pellet seperated from the phage containing supernatant. Replicative form DNA (RF DNA) was recovered from the cell pellet *via* the STET procedure, then linearised by restriction with *Hind*III and sized on 0.7% agarose gels. Clones

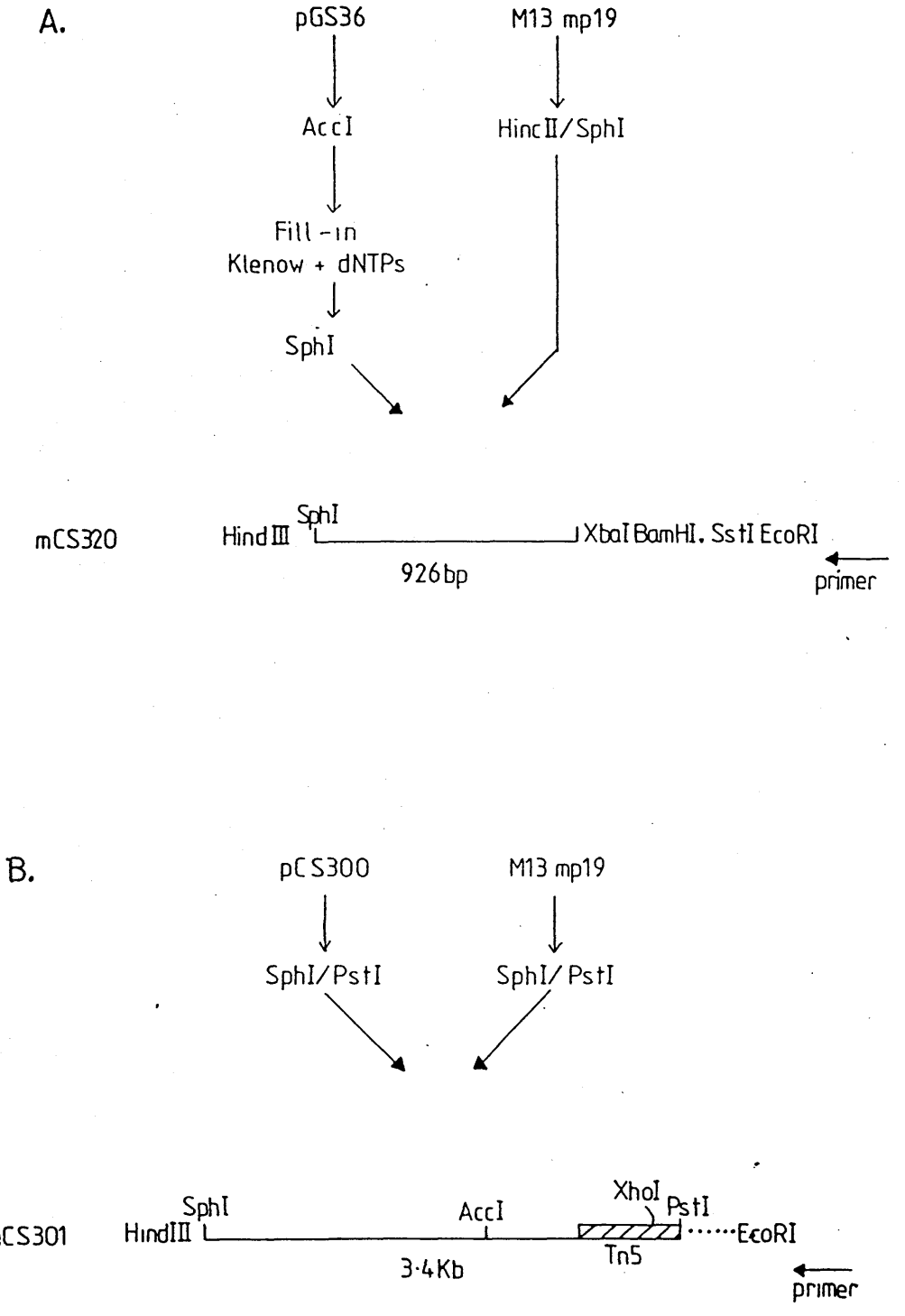
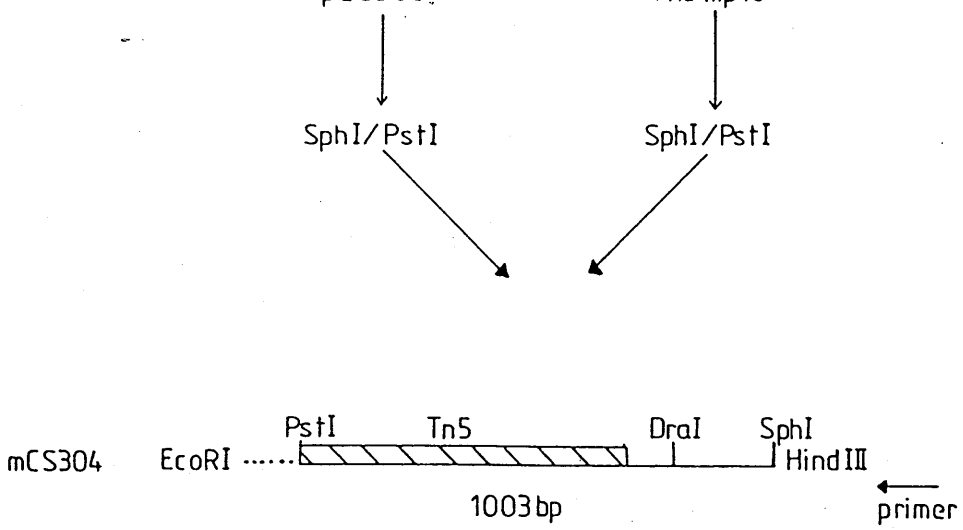


Figure 6.2 Construction of mCS320 and mCS301

Only those polylinker sites of some significance are shown, other polylinker sites present are indicated by a period (.).



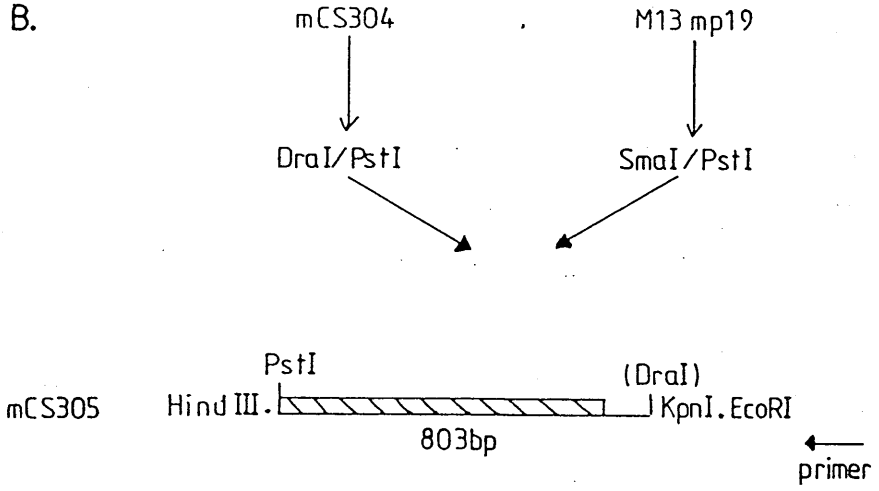


Figure 6.3 Construction of mCS304 and mCS305

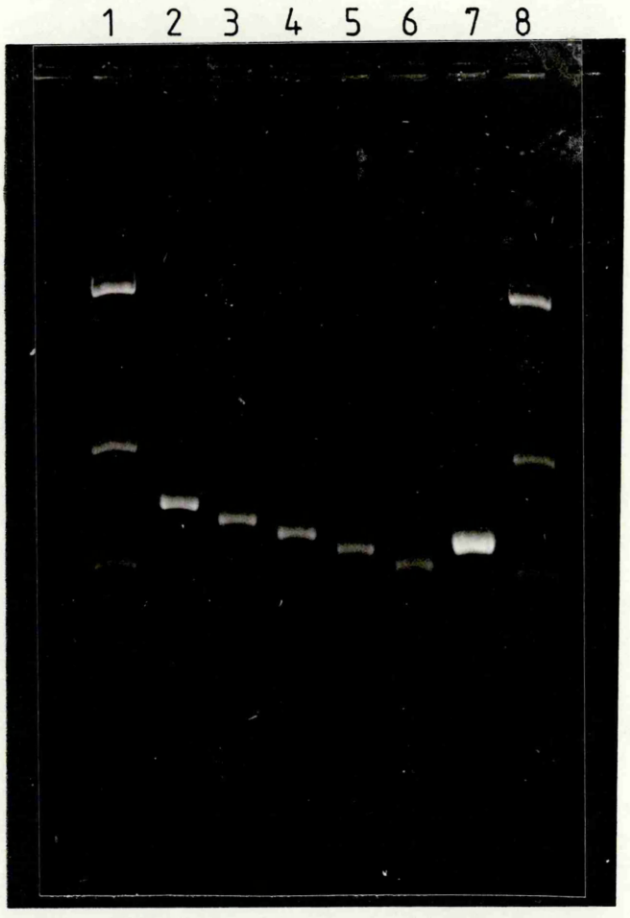


Figure 6.4 ExoIII deletion time-course of mCS320. Time-points were taken at 2 minute intervals as described in the text, then sized by agarose gel electrophoresis. Loading order: T_0 (time zero), Lane 2; T_2 (2 minute time-point), lane 3; T_4, lane 4; T_6, lane 5; T_8, lane 6; Size markers: M13mp19 linearised with HincII, lane 7; Lambda/HindIII lanes 1 + 8.

with probable overlapping deletions were identified, and then single-stranded DNA (s.s DNA) was recovered from the appropriate culture supernatant (as described in materials and methods), for use as template in the di-deoxy chain termination sequencing procedure of Sanger et al, (1977). In total, four deletion derivatives of mCS320 were sequenced (plus of course mCS320 itself), those being mCS320 Δ21, Δ22, Δ25 and Δ44. These five clones provided 924bp of contiguous sequence running from the AccI site to the SphI site (5'-3').

Exonuclease III deletions of mCS301 were performed in an identical fashion, except that the substrate DNA was initially restricted with PstI and XhoI to provide protected and susceptible ends respectively. Suitable clones were again size- selected prior to sequencing; the sequenced clones being mCS301 Δ213, Δ47, Δ48, Δ49, Δ82, Δ86, plus mCS304 and mCS305. The mCS304 and 305 sequences run (5'-3') from the equivalent of the SphI site in pCS300 into the left end of Tn5, whereas the remaining sequences all run, out of the right end of Tn5, towards the AccI site. All sequencing reactions and associated gels were performed and run as described in materials and methods. All autoradiograms were interpreted by eye, and the sequence read on at least two seperate occaisions. The sequence was then entered into computer files using the sequence assembler/editor program ZALIGN (Boyd, A. C., pers. comm.). Each sequence was entered twice into two independent files, then those files cross-compared to eliminate "input-errors". Finally, each computer file was double-checked against a final proof-reading of the autoradiograms. The derivation of the XerA sequence is shown in tabular form (including the database file names; Table 6.1). The complete 926bp annotated sequence (including the intact AccI site not available from mCS320) is presented in figure 6.6.

6.2.2 DETERMINATION OF THE xerA3::Tn5 BOUNDARY SEQUENCE

The sequences of the right- and left-hand boundaries between the Tn5 element and the xerA gene were sequenced in clones mCS305 and mCS301 Δ213 respectively. The boundaries were identified by aligning the sequences to the outside ends of IS50R and IS50L (Krebs & Reznikoff, 1986; Fig. 6.7). Note that transposition of Tn5 creates a 9bp "Target duplication" at the site of insertion; that is to say that 9bp of target DNA are replicated during insertion of the element such that the replicated copies flank the transposon in direct repeat (

Table 6.1 XERA112 Sequence Derivation

Plus Strand

Clone	File-Name	XerA112 Co-ordinates	Length of Sub-sequence	Gel(s)
mCS304	*C.MCS304	1-319	319	4,12
mCS305	*C.MCS305	198-319	121	7
mCS301Δ213	*C.MCS301213	292-409	117	6
" 49	" 49	322-558	237	5,4
" 48	" 48	484-621	138	5
" 47	" 47	592-828	237	5
" 86	" 86	695-827	133	2
" 82	" 82	807-943	137	2

Minus Strand

Clone	File-Name	XerA112 Co-ordinates	Length of Sub-sequence	Gel(s)
mCS320	*C.MCS320	926-671	255	7,10
mCS320Δ22	*C.MCS32022	750-425	305	8,9
" 21	" 21	619-221	398	8,9
" 25	" 25	558-208	350	8,9
" 44	" 44	280-1	280	10,11

Figure 6.6 Annotated XERA112 Sequence

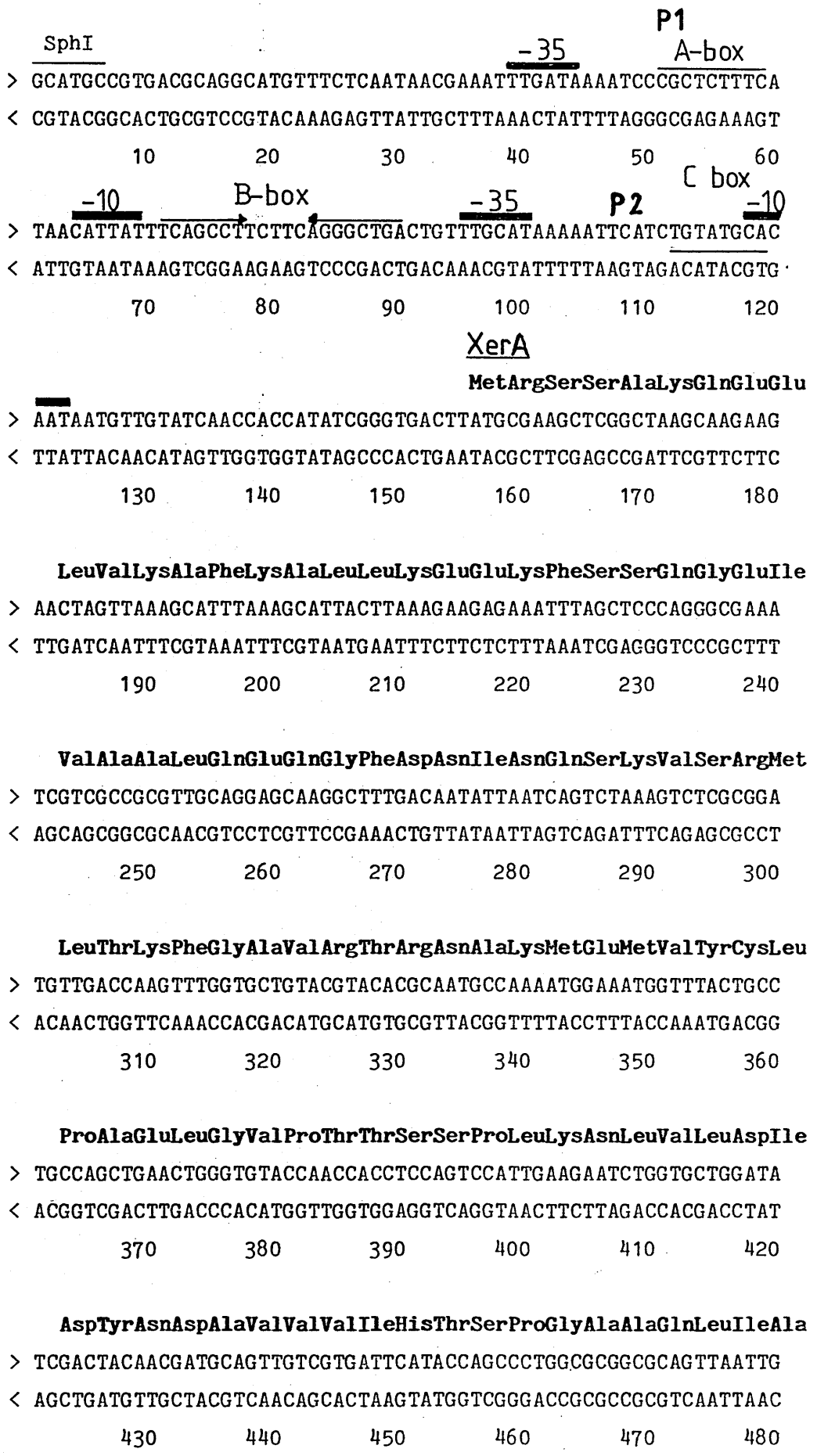


Fig. 6.6 cont.

```
     ArgLeuLeuAspSerLeuGlyLysAlaGluGlyIleLeuGlyThrIleAlaGlyAspAsp
> CTCGCCTGCTGGACTCACTGGGCAAAGCAGAAGGTATTCTGGGCACCATCGCTGGCGATG
< GAGCGGACGACCTGAGTGACCCGTTTCGTCTTCCATAAGACCCGTGGTAGCGACCGCTAC
        490        500        510        520        530        540

     ThrIlePheThrThrProAlaAsnGlyPheThrValLysAspLeuTyrGluAlaIleLeu
> ACACCATCTTTACCACCCCTGCTAACGGTTTCACAGTCAAAGACCTGTACGAAGCGATTT
< TGTGGTAGAAATGGTGGGGACGATTGCCAAAGTGTCAGTTTCTGGACATGCTTCGCTAAA
        550        560        570        580        590        600

     GluLeuPheAspGlnGluLeu***
> TAGAGCTGTTCGACCAGGAGCTTTAATCTCTGCCCCGTCGTTTCTGACGGCGGGGAAAAT
< ATCTCGACAAGCTGGTCCTCGAAATTAGAGACGGGGCAGCAAAGACTGCCGCCCCTTTTA
        610        620        630        640        650        660

> GTTGCTTATCCCTCTCAACCCCCTGCTTTCCCCTGCGATTAATTTAACGAATAGTGCGTT
< CAACGAATAGGGAGAGTTGGGGGACGAAAGGGGACGCTAATTAAATTGCTTATCACGCAA
        670        680        690        700        710        720

> TTACTGCGACATGTCATTCACACAATGAATACATAAGGTAAAAAAAGCACATTATGCAAA
< AATGACGCTGTACAGTAAGTGTGTTACTTATGTATTCCATTTTTTTCGTGTAATACGTTT
        730        740        750        760        770        780

> ATTCATTATCTAATTGAAAAAACTAGAATTAACGATAAATAACCGTATTTTTAATTCTTT
< TAAGTAATAGATTAACTTTTTTGATCTTAATTGCTATTTATTGGCATAAAAATTAAGAAA
        790        800        810        820        830        840

> TTTGTTATTAAAATTCACATTTTTAACACTTAGTATCAACTGAAACAGTTAGCGCGGTAT
< AAACAATAATTTTAAGTGTAAAAATTGTGAATCATAGTTGACTTTGTCAATCGCGCCATA
        850        860        870        880        890        900

                   AccI
> TAATTAGCTCAATAATTAGTGTATAC
< ATTAATCGAGTTATTAATCACATATG
        910        920    926
```

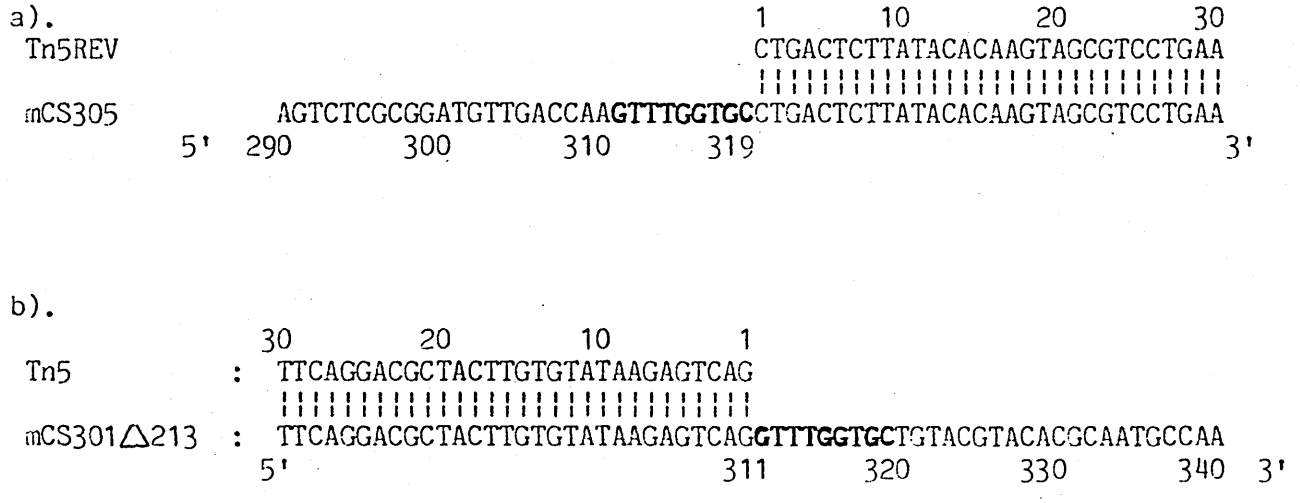


Figure 6.7 Alignment of Tn5 Ends With the mCS305 and mCS301△213 Sequences

The top strand numbering refers to the Tn5 sequences, where base-pair 1 corresponds to the first base of the transposon. The bottom strand numbering refers to the XERA112 sequence co-ordinates as shown if figure 6.6. The 9 base-pair duplication is highlighted in bold-type. Note that in the case of mCS305 a total of 66bp of Tn5 sequence was determined, whereas in mCS301△213 208bp of the inserted transposon sequence was determined. The derived Tn5 sequence was found to be identical to that published by Krebs & Reznikoff (1986).

Schaller, 1978). In this case the duplicated sequence is clearly 5'-GTTTGGTGC-3', corresponding to base pairs 311-319 in Fig. 6.6. When one excises the Tn5 sequence, "in computero", by splicing the two direct repeats such that only one copy remains, then the xerA3 derived sequence is the perfect complement of that derived from mCS320.

6.3 DNA SEQUENCE ANALYSIS

An astonishing wealth of DNA sequence information has been generated in recent years, the availability of which has allowed sequence comparisons of regions of DNA with known biological functions, e.g. protein coding sequences (PCSs), promoters, ribosome binding sites etc.. Comparative studies of this sort have uncovered "consensus sequences" and other sequence "motifs" correlated with genes as a whole, and with specific associated functions. Armed with information of this nature one can then analyse novel sequence data and distinguish probable protein coding genes from intergenic non-coding DNA. The features associated with E. coli genes upon which I shall focus are;

SEQUENCE MOTIF	PROGRAM
1. Open Reading Frames (ORFs).	HYPOTHETICAL
2. Gene Content:	
a. Periodicity in base composition	ZFICK
b. Codon bias	CODONPREFERENCE
3. Transcriptional initiation sites.	ZPROM
4. Translational initiation sites.	ZPERTRON
5. Transcriptional termination sites.	TERMINATOR

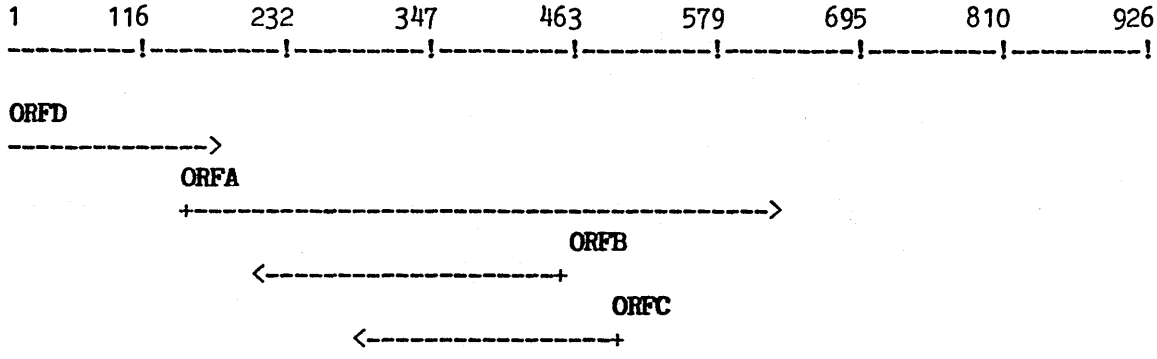
Before proceeding with a discussion of computer assisted sequence analysis it is pertinent to define the terms "Window" and "Weight Matrix". A "Window" is a fixed length of sequence, e.g. 200bp, within which a statistical calculation is performed. The window is then shifted along the sequence by some small increment, e.g. 3bp, after which the calculation is repeated, re-considering the bulk of the previous window in the light of the new increment. Clearly this approach has powerful applications in the identification of statistical trends within a sequence.

A "Weight matrix" is employed when one is searching for a sequence which is similar to a given consensus sequence. The consensus sequence is, by definition, derived from a number of similar sequences, and therefore is not in itself a useful comparative tool since it gives no appreciation of the relative importance of each residue within the sequence; however, a weight matrix can offer such discrimination. The typical matrix is a 2-dimensional array of values and sequence, with a column number equal to the consensus sequence length, and a row number equal to the number of possible residues to be found in the sequence (e.g. 4 rows in DNA weight matrices). The "weighted" matrix is then constructed empirically by scoring the frequencies with which any given residue occurs at each position within the sequence. In general, programs employing such matrices compare a sequence window against the matrix, calculate a score for each window, then print out those scores above a user defined threshold. Obviously the reliability of the weight matrix, plus the definition of a realistic threshold value are the keynotes to this type of analysis.

6.3.2 OPEN READING FRAMES

An open reading frame (ORF) is classically defined as the series of "sense" codons between two in-frame "non-sense" codons (i.e. UAA, UGA, or UAG). However, only those ORFs which include an in-frame initiation (or "start") codon may be considered as potential protein-coding genes. The AUG start codon initiates translation in 91% of 300 characterised *E. coli* proteins, whereas GUG initiates in 8% of cases; Two natural *E. coli* proteins are initiated by the UUG codon, whereas there is a single example of AUU initiation (Gold *et al*, 1984; Stormo, 1986). The identification of potential Protein Coding Sequences (PCSs) by the identification of ORFs has been further complicated by studies on the *E. coli rf2* gene, whose protein product (Release Factor 2), is only expressed when the translating ribosome commits a frame-shifting "error" (Craigen & Caskey, 1986).

The program HYPOTHETICAL displays ORFs in a schematic fashion illustrating all "hypothetical" genes which are initiated by either an AUG or GUG codon and which extend beyond a defined minimum length: a minimum "Gene" length of 50 codons was chosen on the grounds that most *E. coli* protein-coding genes are longer than 50 codons, whilst most fortuitous ORFs are less than 50. A total of four potential genes were identified in this way, of which only three are wholly contained within XERA112, the fourth spanning the *Sph*I site (Fig. 6.8). Of the



ORF	START CODON	CO-ORDINATES	STRAND	CODONS	PREDICTED MOLECULAR WEIGHT (KDa)
A	AUG	156-623	+	156	17.3
B	AUG	452-207	-	82	9.0
C	GUG	498-289	-	70	7.7
D	-	*-169	+	56*	-

Figure 6.8 HYPOTHETICAL genes identified in the XERA112 sequence.

three intact ORFs (A, B, and C), all three would be disrupted by the Tn5 insertion in xerA3 (310bp). It is important to note that minicell analyses have demonstrated that the plasmid pGS38 encodes an insert specific polypeptide of apparent molecular weight equal to 16.5KDa (Stewart, 1986); from figure 6.8 it is evident that only ORF-A could encode such a polypeptide, the predicted molecular weight being 17.2KDa. It therefore seems highly probable that ORF-A represents a genuine PCS (PCS-A) which is expressed in vivo, since this gene is disrupted in the xerA3 mutant then it represents a very strong candidate for the xerA gene itself. Having identified several ORFs in the XERA112 sequence I attempted to distinguish bona fide PCSs, including PCS-A, from fortuitous ORFs using the computer programs ZFICK and CODONPREFERENCE.

6.3.3 ZFICK ANALYSIS

An objective statistical analysis of coding and non-coding DNA sequence has revealed that the base usage pattern of coding sequence has a characteristic periodicity, arising as a consequence of the fact that any given base (i.e A, C, G, or T) is most often found in the same codon position (i.e. 1^{st}, 2^{nd}, or 3^{rd}); there is no such periodicity in non-coding sequence (Fickett, 1982). The TESTCODE algorithm statistically assesses periodicity within a 200bp sequence window, a "Testcode Indicator" (T.I.) value is then assigned to the window centrepoint, then the calculation window is shifted by 3 bp. The T.I. value gives a measure of periodicity within the sequence, from which its "Probability of coding" may be inferred. The threshold values listed below were determined empirically by comparing the T.I. values of coding and non-coding sequences. Using these thresholds TESTCODE was applied to independent test data, and found to make erroneous "Coding" predictions in only 3% of all 200bp windows tested. Erroneous "Non-coding" predictions were 6%, whilst "No Opinion" was expressed for 18% of coding sequences (Fickett, 1982).

T.I. VALUE	PROBABILITY OF CODING	PREDICTION
< 0.74	< 0.29	Non-Coding
0.74 - 0.95	0.40 - 0.77	No Opinion
> 0.95	> 0.92	Coding

The program ZFICK uses the TESTCODE algorithm and plots the T.I. values graphically (Boyd. A.C., pers. comm.). ZFICK analyses of the

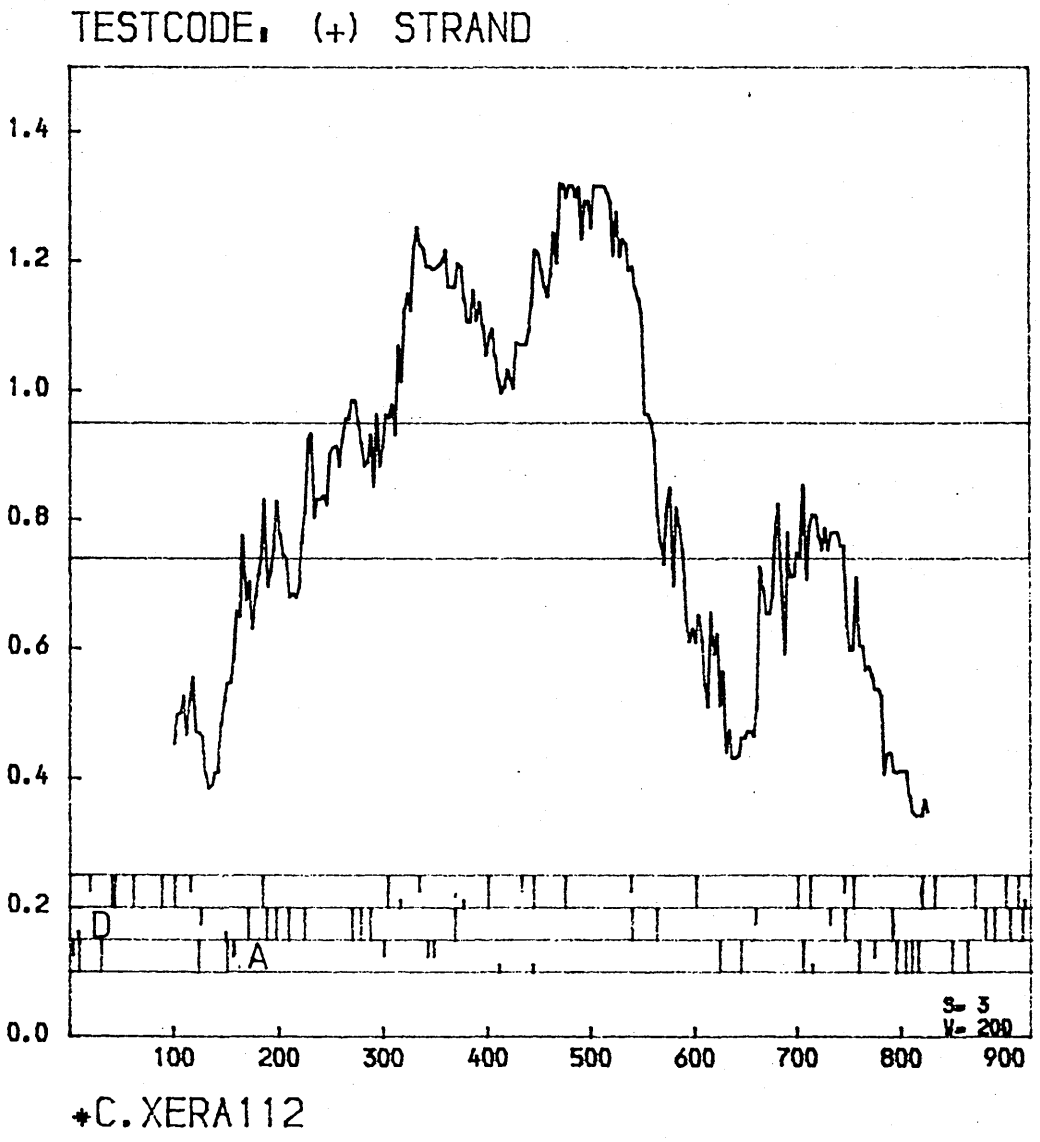


Figure 6.9 ZFICK Analysis of the XERA112 Plus Strand

Below the TESTCODE output there is diagrammatic representation of the three reading frames running from left to right. Code: T = ATG codon; t = GTG codon; I = Non-sense codon.

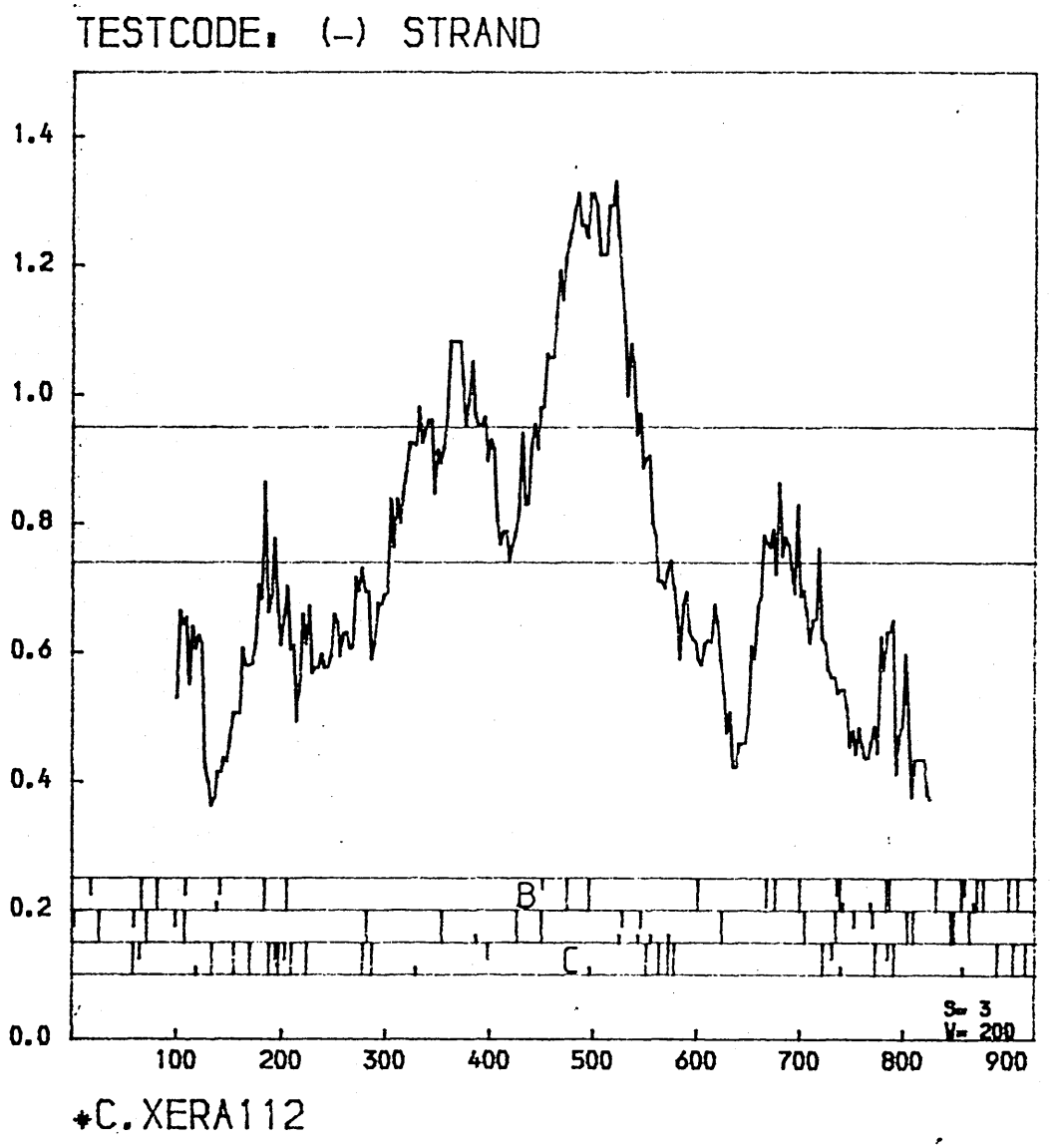


Figure 6.10 ZFICK Analysis of the XERA112 Minus Strand

Note that the reading frames on the Minus strand output run from right to left.

XERA112 plus strand sequence (Fig. 6.9), reveals considerable correspondence between PCS-A and a region of predicted coding sequence. It should be noted that a "tailing-off" towards the end of a PCS is expected due to the randomising affect of adjacent non-coding sequence as the analysis window spans the PCS boundaries. With this in mind it should be obvious that a small PCS is likely to yield a relatively poor TESTCODE (this is true for the 189bp ColE1 rom gene; data not shown). Since TESTCODE is detecting a periodic signature within a sequence, then the two complementary DNA strands are expected to yield very similar ZFICK plots as can be seen by comparing Figs. 6.9 and 6.10. This reflects a major limitation of ZFICK, this being that whilst it can identify a coding region, it does not differentiate between frames within that region. Therefore, whilst ZFICK supports PCS-A, we must also consider ORFs B and C as potential genes. One way to statistically compare the contents of each frame independently is to examine the codon bias, this was done using the CODONPREFERENCE program.

6.3.4 CODONPREFERENCE ANALYSIS

Due to the degeneracy of the genetic code, most amino acids are specified by more than one codon (synonomous codons). The analysis of codon usage patterns within known PCSs has revealed that synonomous codons are not used at equal frequencies, their relative frequencies being dependent on both the gene and the organism. In E. coli the observed codon bias strongly correlates with the relative abundance of the cognisant tRNAs (Ikemura, 1981a; 1982). Highly expressed E. coli genes (e.g. for the ribosomal protein RpoC, or outer membrane protein OmpA) almost exclusively use those codons which are recognised by the most abundant tRNAs; a phenomenon believed to reflect their requirement for efficient translation (Post et al, 1979; Post & Nomura, 1980; Ikemura, 1981b; Grosjean & Fiers, 1982). On the other hand, genes whose protein products are expressed at low levels (e.g. Lac repressor, or DnaG), tend to use synonomous codons in rough proportion to the abundance of the corresponding tRNA, thus resulting in a smaller codon preference (Grantham et al, 1980). In striking contrast, non-coding regions of E. coli DNA exhibit no pronounced preference for any tri-nucleotide sequences (Gribskov et al, 1984). The CODONPREFERENCE program determines a "Codonpreference statistic" (CPS) for windows of 25 codons and in all three reading frames. The CP statistic is based on the likelihood that any given codon will be

found in either coding or non-coding sequence, as defined by an empirically determined "Codon frequency table" (Gribskov et al, 1984; Grantham et al, 1981). In practice the CODONPREFERENCE algorithm performs well in the detection of genes which are either highly or moderately expressed (producing a CPS of between 1.0 - 2.0), but less well against poorly expressed genes (e.g. dnaG scores a CPS of between 0.5 - 1.0). The average CPS expected from random sequence is 0.45, but this varies with base composition. The program plots the CPS value against the sequence for all three frames in the strand under analysis; superimposed upon each plot are all those ORFs identified by HYPOTHETICAL as potential genes. Reading frames in both the plus and minus strand outputs run from left to right, thus the minus strand numbering does not directly compare to the plus strand. In addition the plots include the positions of all "rare" codons (i.e. those whose usage represents 5% or less of the usage of synonomous codons), it has been noted that there is paucity of rare codons in most genes (noticeable even in weakly expressed ones), whereas they are common in non-coding sequence (Gribskov et al, 1984).

CODONPREFERENCE plots of the XERA112 sequence are presented in Figs. 6.11 and 6.12. From consideration of both plots it is clear that PCS-A exhibits considerable codon bias, although it is interesting to note that a shorter ORF, ORF-A', would exhibit a higher mean CPS than does PCS-A. However ORF-A' would only encode a polypeptide of approximately 12KDa, much smaller than that observed in mini-cells. Therefore one is forced to conclude that the N-terminal 1/3rd of PCS-A has a more restricted codon choice than does the remainder of the coding sequence, this might reflect other selection pressures exerted on this part of the sequence e.g. the need to maitain a specific nucleotide sequence in this region. Alternatively, the poor codon usage might serve to limit the rate of translation of PCS-A. Open reading frame D exhibits a CPS value typical of random sequence, and on this basis is unlikely to represent a protein coding gene. On the other hand ORFs B and C coincide with regions of above average codon preference; it should be noted however, that both of these ORFs are wholly internal to PCS-A and that ORF-B is "in-phase" with PCS-A (Two reading frames may be described as in-phase when they are encoded on opposite strands by the same codon triplets). Furthermore the codon preference plot for PCS-A and ORF-B are in fact mirror images of one another (Figs. 6.11, 6.12). The phenomenon of bias codon usage on one strand being mirrored in-phase on the complementary strand has been

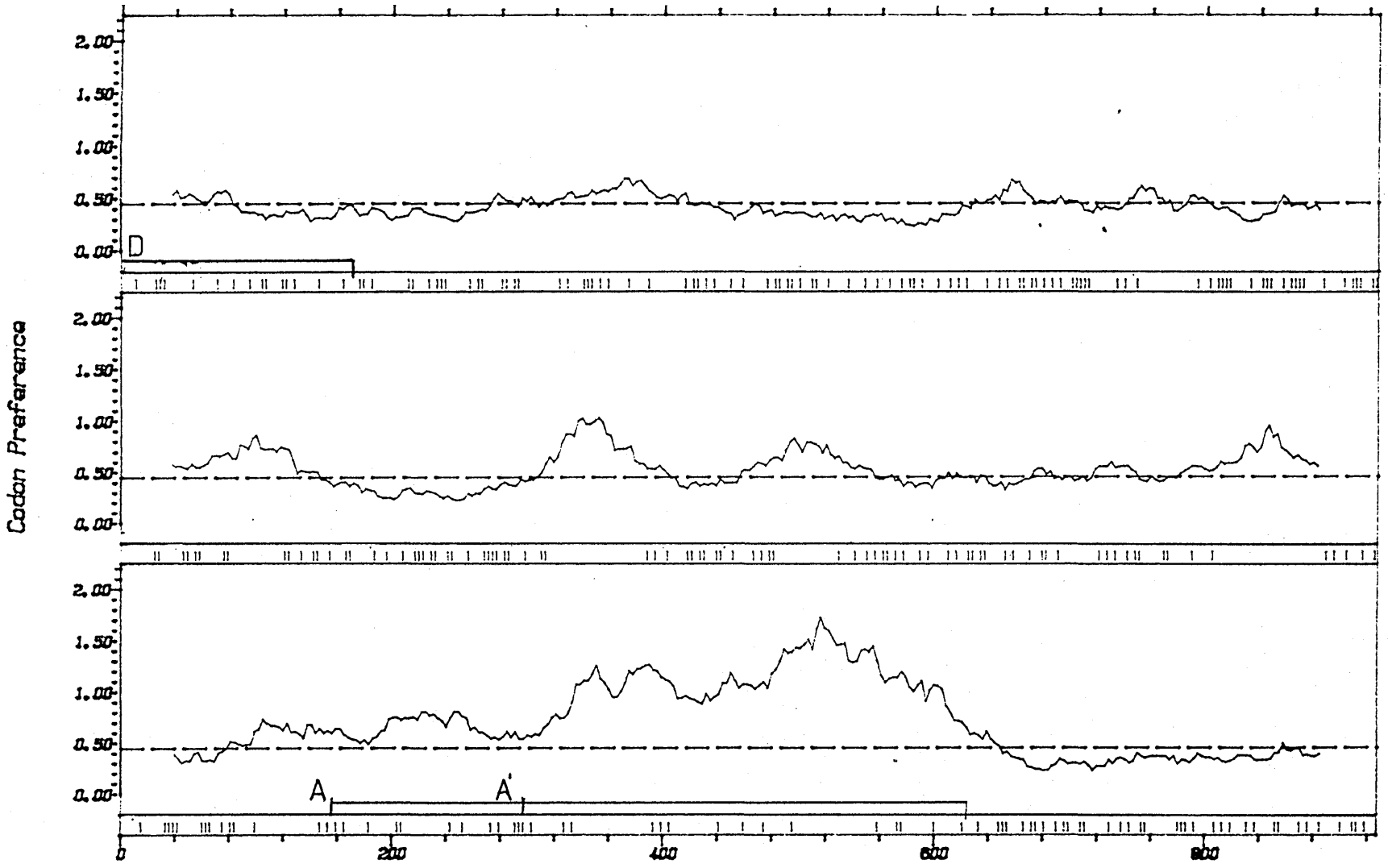


Figure 6.11 CODONPREFERENCE Analysis of the XERA112 Plus Strand: The three reading frames are presented in parallel. Each rare codon is indicated by a vertical dash (¦) below the graphical output.

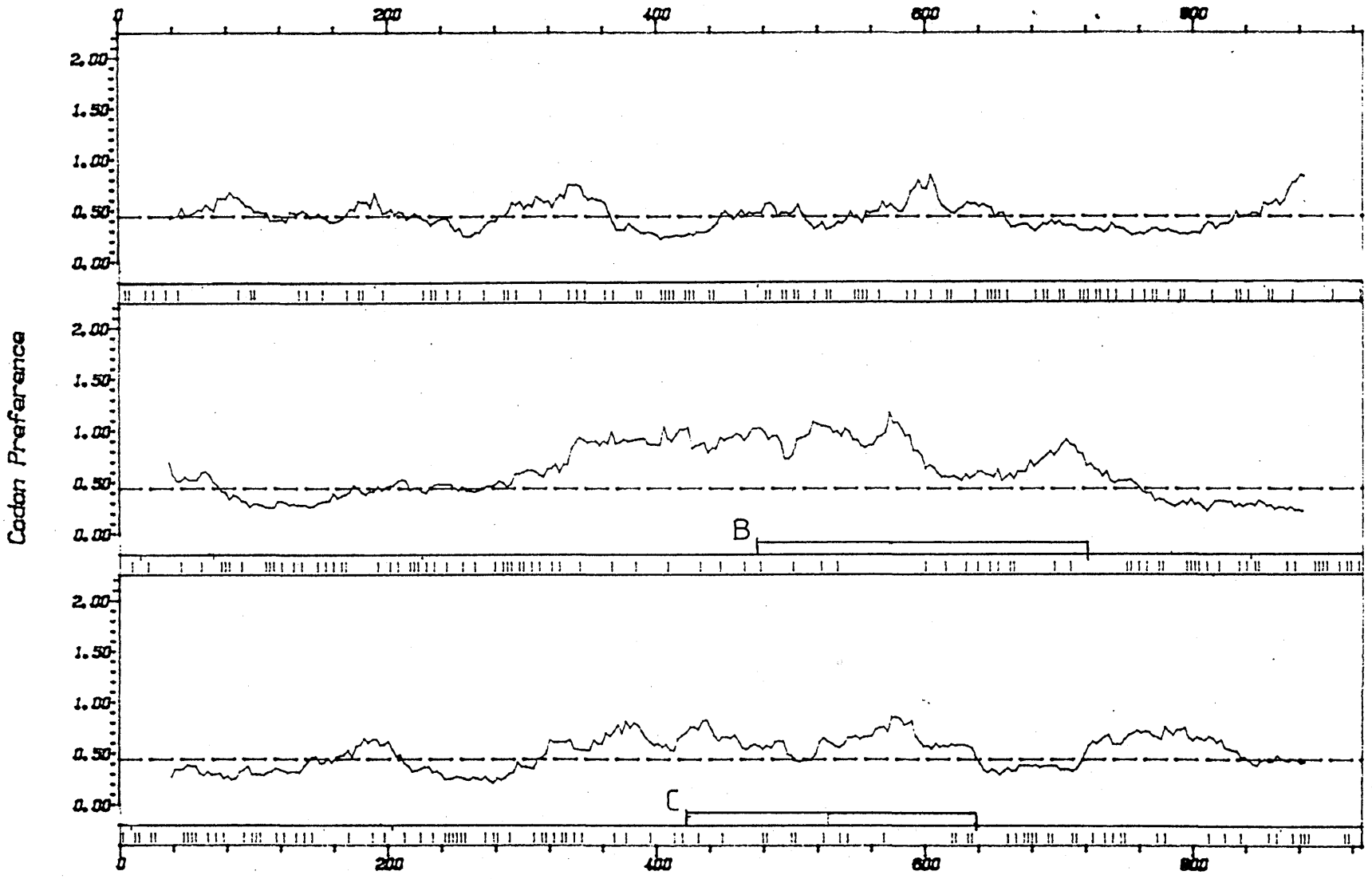


Figure 6.12 CODONPREFERENCE Analysis of the XERA112 Minus Strand

previously noted in surveys of E. coli coding sequences (Aota-Steinberger, 1984). The explanation for this appears to lie in the "choice" of codons which E. coli prefers. Of the 22 preferred codons (de Boer, 1986), 10 can be split into two groups of complementary pairs, the consequence of which being that when one of these 10 codons appears on one DNA strand, its complemetary preferred codon appears in phase on the other strand: e.g.

	Lys	
5'	CUG	3'
3'	GAC	5'
	Gln	

The evolution of a preference for complementary codons is apparent, and yet the selection pressure for this is less so; it might reflect an exploitation of certain "optimal" codon-anticodon interaction energies (Grosjean & Fiers, 1982), or alternatively it might be a consequence of a requirement to optimise double stranded coding capacity. Therefore, whilst this analysis clearly supports the contention that ORF-A corresponds to a protein coding sequence, ORFs B, and C are less readily assigned; they both exhibit poorer codon bias than PCS-A, but this might simply reflect their relative levels of expression.

6.3.5 TRANSLATIONAL INITIATION SITES: ZPERTRON ANALYSIS

An initiation codon (i.e. AUG, GUG, or more rarely UUG or AUU), is not in itself sufficient to promote efficient initiation of translation, thus inferring that additional sequence(s) are required. However a consensus sequence for translational initiation sites (often called "Ribosome binding sites" (RBSs)), has remained elusive, with only the central "U" of the initiation codon being absolutely conserved in all known translational starts (Stormo et al, 1982). One relatively common feature of ribosome binding sites is the presence of a short region of three or more purine nucleotides, located 5' to the start codon, which is complementary to part of the 3' terminus of the E. coli 16S RNA (Shine & Dalgarno, 1974; 1975; Godson et al, 1978). The 16S RNA forms part of the 30S ribosomal subunit which interacts with mRNA in concert with ribosomally associated proteins and initiation factors. Therefore Shine and Dalgarno postulated that the complementarity between 16S RNA and the so-called "Shine and

Dalgarno" (S.D.) sequence on the mRNA might trigger translational initiation at the nearby start codon; there is now a considerable body of evidence to support this hypothesis (for review see Stormo, 1986). However whilst a S.D. sequence can promote the initiation of translation it is not a pre-requisite, since a number of genes lack an S.D.-like sequence and yet are efficiently translated, e.g. phage fd genes VII and IX, and the Lambda cI gene. Furthermore, sequences 3' to the start codon have also been demonstrated to influence the efficiency with which translation is initiated (Looman et al, 1987).

Since so little is known about the sequence requirements of ribosome binding sites, a user defined weight matrix would be unlikely to prove succesful in identifying such sites in novel sequence, thus the powerful "perceptron" algorithm, with its inherent "learning" capabilities was applied. This algorithm was "introduced" to 124 known starts in an E. coli mRNA library, which it then scrutinised automatically searching for recurrent sequence patterns, thus constructing its own weight matrix. The optimal matrix was found to be W101, comprising 101 nucleotides of sequence stretching from -60 to +40 relative to the start codon. As with all matrix analyses, a significance level must be set such that a minimum number of spurious results are obtained; using W101 Stormo et al (1982) found that 124 real gene starts within a 78,000 nucleotide library all scored above +2, whilst all other sequences scored less than +1. The ZPERTRON program was employed to search for start sites using the W101 weight matrix (Boyd, A. C., pers. comm.), both strands of XerA112 were analysed and those sites which scored above -100, and which would initiate translation of a sequence at least 20 codons in length, are listed in table 6.2. Only two sites returned perceptron ratings of >2, those being the PCSA start (b.p 156) and an AUG codon at position 659 on the plus strand; this latter start would initiate a 29 amino-acid polypeptide (3.2 KDa). On the minus strand the AUG start of ORF-B scored -35, whilst that of ORF-C went undetected. From this we can conclude that the AUG start of PCSA is situated within a sequence context similar to those of 124 known translational starts, therefore one would predict that it is likely to act as an initiation site for the translation of PCSA in vivo.

Table 6.2 ZPERTRON Analysis of XERA112

Plus Strand

Start coord.	5' ---	CONTEXT	--- 3'	Stop coord.	Length (codons)	Est. MW (kDa)	PERCEPTRON rating
156	CGGGTGACTT	ATG CGA ...	CTT TAA	626	156	17.2K	48
659	GGCGGGGAAA	ATG TTG ...	CAA TGA	748	29	3.2K	46
115	ATTCATCTGT	ATG CAC ...	AAC TAG	186	23	2.5K	-87
342	CAATGCCAAA	ATG GAA ...	CTT TAA	626	94	10.3K	-95

Minus Strand

Start coord.	5' ---	CONTEXT	--- 3'	Stop coord.	Length (codons)	Est. MW (kDa)	PERCEPTRON rating
452	CAGGGCTGGT	ATG AAT ...	AAG TAA	204	82	9.0K	-35
574	GTCTTTGACT	GTG AAA ...	GTA TGA	449	41	4.5K	-67

6.3.6 TRANSCRIPTIONAL INITIATION SITES: ZPROM ANALYSIS

Escherichia coli transcriptional initiation sites (Promoters), have been highly characterised at the DNA sequence level; comparisons made between 112 known promoters has produced a "Promoter consensus sequence" comprising two relatively well conserved 6bp sequences, the so-called "-35" and "-10" regions, separated by an optimal distance of 16-18bp (Hawley and McClure, 1983):

"-35" "-10"
TTGACA- 16-18bp -TATAAT

Direct analysis of 16 mutations in the lac promoter region suggests that promoter "Up mutations" tend to increase similarity to the consensus sequence, whereas "Down mutations" reduce it (Bujard, 1980; Reznikoff, 1984). Two weight matrices were composed by comparing 100 known promoters and determining the frequency with which any given base appears at each of the 6 positions within the -35 and -10 regions respectively. The program ZPROM then searches for promoters by scoring for two statistical parameters; "Statistic I" scores for the degree of similarity to the -35 and -10 regions by using the weight matrices, then "Statistic II" incorporates an adjustment designed to provide a measure of the distance between the promoter homologies (the spacing in known E. coli promoters varies from between 15-19bp and this has been suggested to be important for recognition by RNA polymerase (Stefano & Grallan, 1982)). Therefore statistic II is more stringent in identifying promoter-like sequences. Using a control array of promoters, the significance levels of Stat' I >0.002 and Stat' II >0.0002 were found to identify all real promoters with a minimum number of false positives (Harr et al, 1983). ZPROM analysis was performed on both strands of XERA112 and the results printed for all those sequences which scored above threshold for at least one statistic (Table 6.3). A score of 1 in both columns would be obtained by the "perfect" consensus sequence TTGACA-17-TATAAT. A total of 16 potential promoters were found on the plus strand, with a further 10 present on the minus strand. The best match by far is that starting at co-ordinate 39 on the plus strand (promoter P1), the position of P1 would enable it to drive transcription of PCSA; a second candidate for a weaker PCSA promoter (P2) is located at co-ordinate 95. The XERA112 sequence is unusually rich in promoter-like sequences with a total of 26 scoring sites in 926bp, compared with for example a total of 16 in

Table 6.3 ZPROM Analysis of XERA112

Plus Strand

Pos.	-35	-10	Statistic I	Statistic II	Distance
21	GTTTCT	CAATAACGAAATTTGA___ TAAAAT.	0.0017265	0.0002590	16
38	TTTGAT	AAAATCCCGCTCTTTCA__ TAACAT.	0.0002097	0.0002097	17
39	TTGATA	AAATCCCGCTCTTTCA___ TAACAT.	0.0313752	0.0047063	16
39	TTGATA	AAATCCCGCTCTTTCAT__ AACATT.	0.0006129	0.0006129	17
39	TTGATA	AAATCCCGCTCTTTCATAA CATTAT.	0.0063827	0.0001277	19
95	TTGCAT	AAAAATTCATCTGTATG__ CACAAT.	0.0029424	0.0029424	17
195	TTTAAA	GCATTACTTAAAGAAGA__ GAAATT.	0.0013198	0.0013198	17
490	TGGACT	CACTGGGCAAAGCAGAAGG TATTCT.	0.0041449	0.0000829	19
599	TTTAGA	GCTGTTCGACCAGGAGC__ TTTAAT.	0.0006488	0.0006488	17
720	TTTACT	GCGACATGTCATTCA____ CACAAT.	0.0033957	0.0000679	15
731	ATGTCA	TTCACACAATGAATACA__ TAAGGT.	0.0002153	0.0002153	17
782	TTCATT	ATCTAATTGAAAAAAC___ TAGAAT.	0.0084931	0.0012740	16
804	TAGAAT	TAACGATAAATAACCG___ TATTTT.	0.0015508	0.0002326	16
826	TATTTT	TAATTCTTTTTTGTTAT__ TAAAAT.	0.0002305	0.0002305	17
828	TTTTTA	ATTCTTTTTTGTTAT____ TAAAAT.	0.0048770	0.0000975	15
848	TTAAAA	TTCACATTTTTAACACT__ TAGTAT.	0.0008238	0.0008238	17

Minus Strand

Pos.	-35	-10	I	II	Distance
48	TTGATA	CTAAGTGTTAAAAATGT__ GAATTT.	0.0013842	0.0013842	17
114	GTTAAT	TCTAGTTTTTTCAATTA__ GATAAT.	0.0005922	0.0005922	17
127	TTTTCA	ATTAGATAATGAATTTTG_ CATAAT.	0.0037054	0.0005558	18
128	TTTCAA	TTAGATAATGAATTTTG__ CATAAT.	0.0015212	0.0015212	17
129	TTCAAT	TAGATAATGAATTTTG___ CATAAT.	0.0045043	0.0006756	16
198	TCGCAG	TAAAACGCACTATTCGT__ TAAATT.	0.0004638	0.0004638	17
301	TTAAAG	CTCCTGGTCGAACAGCTC_ TAAAAT.	0.0019863	0.0002979	18
383	GTGTCA	TCGCCAGCGATGGTGCC__ CAGAAT.	0.0013019	0.0013019	17
618	TGGTCA	ACATCCGCGAGACTT____ TAGACT.	0.0023882	0.0000478	15
907	ATGCCT	GCGTCACGGCATGCGTA__ TACACT.	0.0005082	0.0005082	17

Note "Distance" indicates the number of base pairs between the -35 and -10 regions.

5804bp of the lac operon, or 14 in 2676bp of pUC8. This density of consensus-like sites would appear to be at least partly due to the exceedingly A:T rich nature of the XERA112 sequence from 700-926bp, coupled of course with the A:T richness of the consensus sequence itself; the XERA112 sequence is 74% A:T between base-pairs 700-926, compared with 52% A:T in the remainder of the sequence. A total of 13 out of the 26 sites scored by ZPROM are situated in this 226bp of sequence. Furthermore, of the remaining 13 sites scored, four correspond to variants of P1, illustrating the extent of its conformity to the consensus sequence. The significance of the A:T richness of the region 3' to PCS-A is unclear.

6.3.7 TRANSCRIPTIONAL TERMINATION SITES: TERMINATOR ANALYSIS

Unlike transcriptional promoters, terminators exhibit rather diverse sequence characteristics, therefore our ability to predict them from sequence data alone is somewhat limited. Of the two major classes of terminator, those being factor-dependent and factor-independent, only the latter have a discerned sequence motif, that being a short palindrome followed by a T rich sequence on the non-sense strand (Platt, 1981). Using a similar approach to ZPERTRON, Brendel and Trifonov (1984), designed a program (TERMINATOR) which searches for sequences similar to factor-independent terminators. The perceptron algorithm was "taught" from only 30 characterised termination sites, thus the predictive power of the derived weight matrix is limited. The authors found that a P statistic (scoring for primary sequence) of >3.5 was optimal for discriminating real terminators from random sequences. The TERMINATOR analysis of XERA112 identified a total of 9 sites scoring above threshold; 5 on the plus strand, and 4 on the minus strand (Fig. 6.13). The highest scoring site (terminator T1, at co-ordinate 54) is upstream of PCSA, but overlapping with the promoter sequence P1. It is interesting to speculate that T1 is unlikely to attenuate transcription from the juxtaposed promoter P1, since the predicted P1 transcript will lack all, or most, of one arm of the dyad symmetry which defines T1. On the other hand T1 might be expected to terminate transcription emanating from upstream of P1, such organisation could in principle serve to protect downstream genes from expression via any transcription other than that initiated at P1. Similar overlapping promoter/terminator sequences have been identified in other situations including phage fd and phi X174 (Rosenberg & Court, 1979). I will return to the

Figure 6.13 TERMINATOR Analysis of XERA112

Plus Strand

```
          -40  -35  -30  -25  -20  -15  -10   -5  -1+  +5        p      s
           .    .    .    .    .    .    .    .   ..   .
  54=> TCTTTCATAACATTATTTCAGCCTTCTTCAGGGCTGACTGTTTGCATAAAA    5.14     22
                            --------   --------
 751=> ACATAAGGTAAAAAAAGCACATTATGCAAAATTCATTATCTAATTGAAAAA    4.03     12
                               -- ---      --- --
 794=> TTGAAAAAACTAGAATTAACGATAAATAACCGTATTTTTAATTCTTTTTTG    4.42     18
                                  -----    -----
 804=> TAGAATTAACGATAAATAACCGTATTTTTAATTCTTTTTTGTTATTAAAAT    4.70     14
                     -- -- --     -- -- --
                       -----    -----
 874=> TATCAACTGAAACAGTTAGCGCGGTATTAATTAGCTCAATAATTAGTGTAT    3.63      0
```

Minus Strand

```
          -40  -35  -30  -25  -20  -15  -10   -5  -1+  +5        p      s
           .    .    .    .    .    .    .    .   ..   .
  74=> TTTTAATAACAAAAAAGAATTAAAAATACGGTTATTTATCGTTAATTCTAG    3.62     31
                                  -----    -----
                               -- -- --     -- -- --
  83=> CAAAAAAGAATTAAAAATACGGTTATTTATCGTTAATTCTAGTTTTTTCAA    3.71     26
                     -- - -----    ----- - --
                     -- -- --     -- -- --
 600=> TACGTACAGCACCAAACTTGGTCAACATCCGCGAGACTTTAGACTGATTAA    4.05      0
 856=> AAATAATGTTATGAAAGAGCGGGATTTTATCAAATTTCGTTATTGAGAAAC    3.58      4
                                  ----     ----
```

Primary structure (p) threshold = 3.50
Secondary structure (s) threshold = 0

discussion of T1 shortly. This analysis also identified 3 clustered terminator-like sequences 3' to PCSA (at co-ordinates 751, 794, and 804), any or all of which might act to terminate transcription running through PCSA. However, TERMINATOR failed to identify a very striking region of dyad symmetry immediately 3' to the end of PCSA (co-ordinate 631bp, fig 6.6).

6.3.8 PCS-A CORESPONDS TO THE xerA GENE

In order to formally eliminate ORFs B & C as candidates for the xerA gene, the 725bp SstI/DraI fragment from mCS320 was cloned into pUC19 (SstI/HincII), thus generating pCS370. Plasmid pCS370 carries the XERA112 sequence from 198-923bp and thus includes both the intact ORF-B and ORF-C, however it lacks the N-terminal 14 codons of PCS-A. pCS370 fails to complement pCS202 resolution in CSX3 (as defined in a gel resolution assay), on which basis it can be said to lack a functional xerA gene. Since PCS-A is the only ORF which is both disrupted by Tn5 in xerA3 **and** is truncated in pCS370, then one can conclude that it is very likely to represent the xerA gene; from this point onward PCS-A will be referred to as the xerA gene.

Sequence analyses have revealed that the xerA gene has all the hallmarks commonly associated with E. coli genes; i.e. a promoter, ribosome binding site, periodic base usage, biased codon usage, and even a number of potential transcriptional terminators downstream of the coding sequence. None of the other ORFs exhibit all of these features, and none appear to be strong candidates for computer-predicted genes; however, such sequence analyses could not, and should not, be used to make strong predictions regarding the "non-expression" of any ORF. Therefore I cannot, at this stage, formally preclude the possibility that one (or all) of the smaller ORFs might represent additional genes, perhaps expressing Xer-related functions.

6.4 DNA SEQUENCE COMPARISONS

The XERA112 sequence has been screened for any sequence identities within either the EMBL or GENBANK databases (using the rapid-scanning WORDSEARCH program (Devereux et al, 1984)), but no matches above random background were found. However, more painstaking one-to-one sequence comparisons have also been performed using ZDOTMAT, BESTFIT and GAP (Boyd A.C., pers comm.; Devereux et al, 1984), in order to pin-point both small regions of high similarity, and also any larger regions of only limited sequence similarity. This type of analysis is

of course limited by the choice of sequences to which one compares XERA112 in detail.

6.4.1 COMPARISON TO ColE1: THE PRESENCE OF A CER-LIKE SITE

An obvious comparison to perform was to screen for any "homologies" between XERA112 and ColE1; this uncovered a short region of XERA112 which exhibits significant similarity to part of the *cer* locus. The optimal alignment includes 61bp of XERA112 (304-364bp) and a 61bp portion of *cer* which, with the insertion of two single basepair gaps into each sequence, results in a total of 40 out of 61 matches (66%; See Fig. 6.14). The most striking feature of this alignment is that the 61bp portion of *cer* sequence in question encompasses the 34bp region known to contain the cross-over site (Summers *et al*, 1985). Recall that over the 34bp cross-over region one can align the sites from ColE1, pMB1, ColK, ColA, and CloDF13 such that 28 residues are perfectly "conserved" amongst all five sites; If one compares the *xerA* sequence to this "consensus" then 19 out of these 28 residues are found to be present (Fig 6.14). This obviously raised the possibility that the "homologous" site found in the *xerA* gene might also act as a substrate for Xer dependent site-specific recombination. In order to test for intramolecular recombination at the "XerA-site" (xas) of a type similar to that observed at *cer*, I constructed two pairs of test plasmids: the first carrying two copies of the XerA-site in either direct or inverted orientation (pCS368 and pCS367 respectively, Fig 6.15), the second pair of plasmids carrying a single XerA-site plus a *cer* site again in either direct or indirect relative orientation (pCS361 and pCS360 respectively, Fig. 6.16). On the assumption that any Xer dependent recombination at xas would also be dependent on XerB, the plasmids described above were initially propagated in CSX4 (*xerB*1). Plasmid DNA was recovered and their structures confirmed by digestion with *Eco*RI, *Hind*III, *Acc*I and *Sph*I. The plasmid DNA was then transformed into DS902 (RecA$^-$, Xer$^+$), then individual transformants propagated overnight prior to the recovery of plasmid DNA. The structure of each plasmid after being passaged through DS902 was then compared with the "Parental form" recovered directly from CSX4 to ascertain whether or not any XerB-dependent rearrangements had occured. The products of any deletion reactions ought to be clearly visible as plasmid species of increased electrophoretic mobilities. Analysis of plasmid DNA recovered from DS902, revealed the presence of a single plasmid species indistinguishable in size from the parental

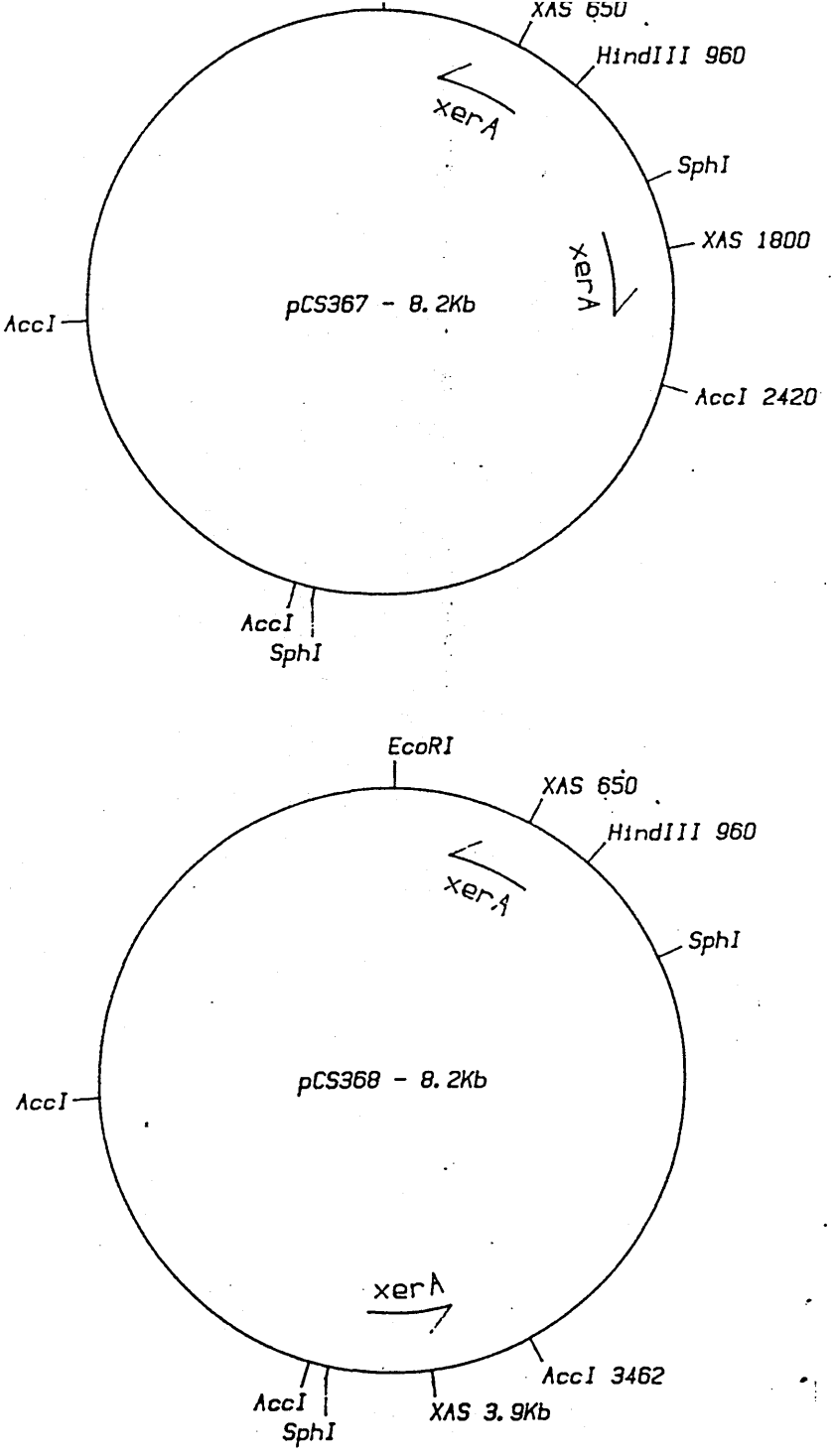


Figure 6.15 Construction of pCS367 and pCS368

The 2.9Kb SphI xerA$^+$ fragment from pGS30 was cloned into the SphI site in pBR322, the resultant plasmids pCS365 and pCS366 corresponding to the opposite orientations of the inserted fragment. A second xerA gene was then added by cloning the 959bp EcoRI/HindIII fragment from mCS320 into the unique EcoRI and HindIII sites in pCS365 and pCS366 to generate pCS367 and pCS368 respectively. The structures of pCS367 and pCS368 are shown above; copies of the xerA structural gene are indicated by arrows, with the approximate co-ordinates of the embedded XAS sites clearly labelled. Obviously pCS367 carries two copies of XAS in inverted repeat, whilst pCS368 contains two copies in direct repeat.

```
        304                                                          364
COLE1   :  TGAACGCGTTT-TAGCGGTGCGTACAATTAAGGGA--TTATGGTAAATCCACTTACTGTCTGCC
XERA    :  ...C.AA....GGT..T..A......CGC..T.CC-AAA....-....GG-T......C.....

PMB1    :  ..........-.....................--...........A.ACGTA...TAAT.
COLK    :  .......A...-C.....................--...........--..........
COLA    :  .....A...CCGG......----..CG....C...TG...........AC..A.A.A.GAC.G
CLODF13:   CC.GT..C..AC.G.....----..CGA.......TG...........ATCT.CGAGT.TCAGG
```

FIGURE 6.14 Optimal alignment between the XERA112 sequence (304-364bp) and the cross-over region of cer. The sequences from pMB1, ColK, ColA, and CloDF13 are also presented in the same alignment with ColE1 as shown in Fig. 1.2. (.) indicates homology to ColE1; (-) indicates a gap inserted to maximise homology.

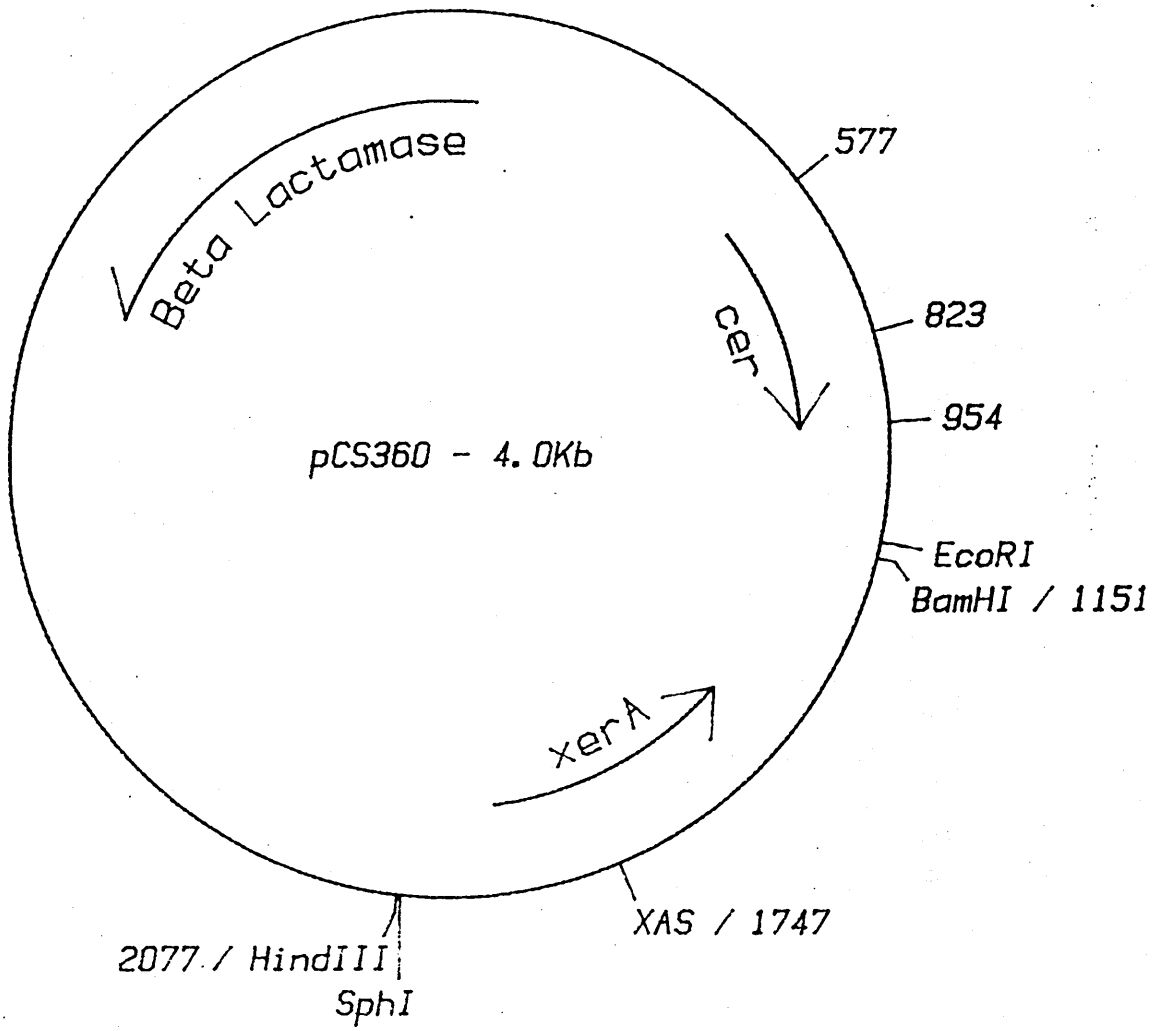


Figure 6.16 Construction of pCS360 and pCS361

The 959bp XerA$^+$ EcoRI/HindIII fragment from mCS320 was cloned into the polylinker of pKS451 and pKS450 to generate pCS360 and pCS361 respectively. The plasmids pKS451 and pKS450 were constructed by cloning the 377bp Cer$^+$ HpaII fragment from ColEI into the NarI site of pUC9; the two derivatives corresponding to opposite orientations of the inserted cer fragment (D.K. Summers, pers. comm.). The orientation of cer fragment in pCS360 is given by an arrow as before. The co-ordinates of the BamHI and HindIII sites are given, as are the approximate co-ordinates of the cer cross-over region (823bp) and the XAS region (1747bp). Obviously pCS361 is identical to pCS360 except for the orientation of the cer fragment which thus places the cer cross-over region at approximately 708bp. The cer and XAS regions are in inverted repeat in pCS360, and in direct repeat in pCS361.

Table 6.4 Restriction predictions for various XAS-specific inversion products.

		Predicted Restriction pattern (Kb)	
Plasmid	Enzyme(s)	Parental plasmid	Inversion product
pCS360	BamHI/HindIII	3.1, 0.95	3.3, 0.7
pCS361	" "	3.1, 0.95	3.2, 0.8
pCS267	AccI/HindIII	2.1, 1.6, 3.1, 1.5	2.1, 1.6, 3.7, 0.9
pCS368	SphI	2.9, 4.8, 0.5	7.0, 0.5, 0.7

plasmid as isolated from CSX4; this was true for pCS360, 361, 367, and 368 (data not shown). In order to test for cer/xas-specific inversion, each DS902 derived plasmid prep' was digested with suitable restriction enzymes, then its restriction profile compared with that of the parental construct. Table 6.4 lists the predicted restriction profiles for each of the four test plasmids, plus those of each putative cer/xas-specific inversion product. The plasmid restriction digests in question were electrophoresed through either agarose or poly-acrylamide gels as appropriate, whereupon EtBr stained bands were visualised under U.V.. In no case were there any bands detected other than those predicted for the parental form of each plasmid (data not shown). There is therefore no evidence for any intra-molecular recombination between either two copies of the xas locus, or between xas and cer. However, rare recombination events might not have been detected in the assay system employed.

6.4.2 IHF BINDING SITE

E. coli IHF is a heterodimeric DNA binding protein which has been implicated in an apparently diverse range of processes including; enhancing site specific recombination, conjugal DNA transfer, and the regulation of gene expression (see Introduction). One region of XerA112 (261-273bp), exhibits strong "homology" (8 out of 9 bp) to the consensus IHF binding site T/CAANNNNTTGATA/T (Leong et al, 1985).

IHF Consensus Sequence		T/CAANNNNTTGATA/T
XerA112	261-273bp	CAANNNNTTGAcA

It is particularly intriguing that such a close match to the IHF binding site should occur within 30bp of the so-called XerA-site (302-364), representing a similar spatial arrangement to that found between the lambda attP site and its proximal IHF binding sites H2 and H' (Weisberg & Landy, 1983).

6.4.3 A nut-LIKE SITE 5' TO THE xerA CODING SEQUENCE

A somewhat surprising finding was that of the presence of a Lambda nut-like site between the -35 region of promoter P1, and the start of the xerA structural gene. The lambda nut sites (nutL and nutR), are of course locations at which the lambda anti-termination protein, N, plus certain host factors, including NusA, become associated with transcribing RNA polymerase (RNAP), and in doing so

render it resistant to downstream termination sites (Friedman & Gottesman, 1983). This phenomenon is crucial to the lambda life-cycle since it delays the expression of certain phage genes situated beyond termination sites, until after sufficient lambda N protein has been synthesised to overcome those terminators. A number of phage nut-like sites have been identified in various bacteriophages including Lambda, P21 and P22, where in each case the sites can be aligned by virtue of three sequence motifs, namely the A-box, B-box, and C-box:

1. The consensus A-box is common to all locations at which the E. coli NusA factor is known to act; it is therefore believed to represent the NusA recognition sequence. However, recent evidence has demonstrated that the A-box can be deleted from one such location without affecting the capacity for NusA interaction (Zuber et al, 1987).
2. The B-box is a region of hyphenated dyad symmetry whose sequence is variable; this sequence is apparently recognised specifically by the appropriate anti-termination protein, e.g. Lambda N recognises lambda nut, whilst lambda Q recognises lambda qut etc..
3. The so-called C-box is highly conserved amongst all known nut-like sites, with the sole exception of lambda qut. This sequence conservation infers a functional role for the C-box but this role remains to be uncovered.

Candidate A-, B-, and C- boxes are present in the XERA112 sequence which compare very well with the lambda nut sites ; it is particularly striking to note that the xerA B-box is in fact an almost perfect inverse of nutL B-box with a total of 14 out of 15 matching bases (Fig. 6.18). Interestingly, the XERA112 B-box region is the same as that identified as the potential transcriptional terminator, T1 (section 6.3.7). Significantly, many E. coli terminators are preceeded by a NusA box including those of the crp and trp operons, plus the attenuators of the his, trp, and ampC operons etc. (Friedman and Gottesman, 1983). The NusA boxes from a number of such loci, plus those from phage nut-like sites, all appear to fit the consensus sequence T/CGCTCTT(T)A, to which the XER112 A-box conforms very well:

Consensus A-box	T/CGCTCTTTA
XERA112 A-box	CGCTCTTTc

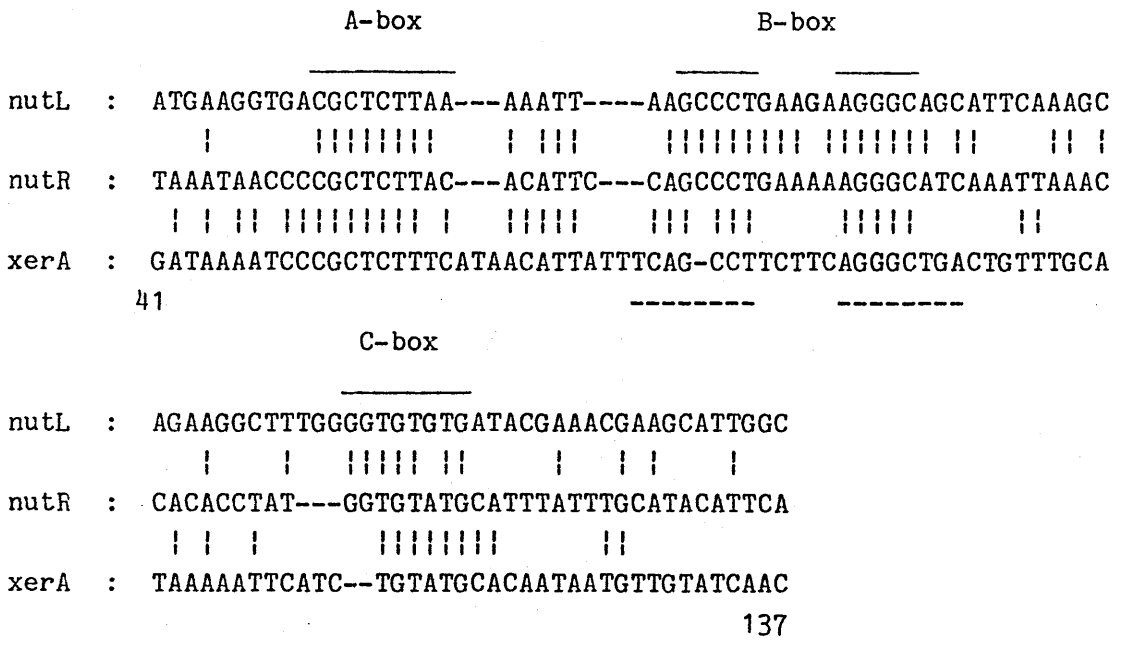


Figure 6.18 Comparison Between XERA112 and Phage Lambda nut Sites

The numbers refer to the XERA112 co-ordinates. Regions of hyphenated dyad symmetry representing the B-box are highlighted; notice that the sequence corresponding to the loop of the nutL B-box (GAAGA) is the inverse of that present in the xerA B-box (TCTTC). (-) represents an single-base gap inserted to maximise homology.

The role of NusA at terminators appears to be one involving the NusA dependent pausing of RNA polymerase, the stalled RNAP:NusA complex is then manifestly more susceptible to subsequent interactions with other termination factors, including Rho (Greenblatt *et al*, 1981). A similar NusA-induced pausing may also be essential at *nut* sites during the association of the phage antitermination protein. Whilst the XerA T1 site resembles a terminator, the additional presence of a consensus C-box provides the signature of an anti-termination locus. Similar signatures have been identified in all six of the *E. coli rrn* operons, and in the case of *rrnG* the site has been shown to function as an anti-terminator (Li *et al*, 1984). If we consider the location of the XERA112 site, then it is clear that it overlaps with the two previously identified promoter consensus sequences, P1 and P2; thus the NusA box would be outwith both the presumptive P1 and P2 transcripts (See Fig. 6.6). There is to date no precedent for NusA functioning *via* a recognition box which is outwith a transcript: indeed there is some evidence to suggest that NusA actually recognises the RNA (Nakamura, 1986). Thus any termination or anti-termination activity at the XerA site would have to either affect transcription initiated upstream of P1, or else involve a novel role for NusA during transcriptional initiation; it is interesting to note that NusA has been found to compete with the *E. coli* sigma subunit in binding to RNAP *in vitro*, however there is no evidence for the involvment of RNAP:NusA complexes in the initiation of transcription (Greenblatt & Li, 1981). If the XerA site has any *nut*-like function, and acts upon transcription emanating from upstream, then it might not affect the expression of XerA *per se*, since the lambda *nut* sites are not in themselves terminators, but simply serve to obviate **downstream** terminators. Whilst there are no obvious terminator-like structures within the *xerA* coding sequence, there are a number of potential terminators to its 3' side; however there are no candidate genes, within the available sequence, which might require anti-termination for their expression. Clearly the transcriptional organisation of the *xerA* locus is worthy of detailed investigation.

6.5 AMINO ACID SEQUENCE COMPARISONS

The *xerA* gene product is predicted to be a 156 amino acid polypeptide with a molecular weight of 17.2KDa. The entire primary structure of this protein can be inferred from the DNA sequence, and was presented in figure 6.6: the amino acid composition of XerA

predicts a hydrophilic protein (30% charged residues), which is probably acidic in nature (containing a ratio of acidic:basic residues equal to 21:17; Table 6.5). The precise function of the XerA protein, with respect to cer-specific recombination is unknown; it may be the cer-specific recombinase, but on the other hand, might be either an accessory, or regulatory component of the system. In the absence of a direct assay for XerA activity, I sought to uncover some clues as to the nature of its function by comparing its primary structure to those of proteins with characterised biological activities. The XerA sequence was screened for similarity to existing protein sequences in the NBRF database, however no sequences were identified with above random sequence identities. More specifically I then scrutinised the XerA sequence for any similarities to the conserved sequence motifs identified within functionally related protein "families"; for example, there are no matches to the highly conserved "nucleotide binding site" motif associated with proteins which bind, and in most cases hydrolyse, ATP (Walker et al, 1982; Higgins et al, 1986):

Nucleotide binding site motif: G----GKT/S

Other obvious motifs to screen for are the highly conserved regions found amongst "families" of site-specific recombinases. The resolvases from transposons Tn3, Tn21, Tn501, Tn1000, and Tn1721 plus the Hin, Gin, Cin, and Pin invertases, are all roughly equivalent in size (approximately 20KDa), and all share several highly conserved regions (Diver et al, 1983; Hiestand-Nauer & Iida, 1983; Newman & Grindley, 1984; Michels et al, 1987). The most prominent conserved sequence being GlyTyr-X-ArgValSer (X= Val, Ile, or Ala), where the Serine has been identified as the "Active residue" in Tn1000 resolvase, i.e. that which is covalently modified during the recombination reaction (Reed & Moser, 1984). However, none of the conserved regions observed amongst this group of recombinases have convincing "homologues" in the XerA sequence.

Another group of site-specific recombinases, the so-called "Integrase family", comprises the integrase proteins of phages P2, P4, 186, P22, Lambda, Phi80, the Cre protein of P1, the E. coli FimB and FimE proteins, and the Flp protein encoded by the S. cerevisiae 2-micron circle. This family of recombinases is more diverse than the resolvase/invertase group, and exhibit only limited sequence

Table 6.5 Amino Acid Composition of the XerA Polypeptide

Aromatic	Hydrophobic	Weakly Hydrophobic	Acidic	Amide	Basic	Cysteine
F = 7	L = 19	P = 5	D = 9	N = 6	K = 11	C = 1
W = 0	I = 9	A = 16	E = 12	Q = 7	R = 5	
Y = 3	V = 11	G = 10			(H = 1)	
	M = 4	S = 10				
		T = 10				
10	43	51	21	13	16	1
					(1)	

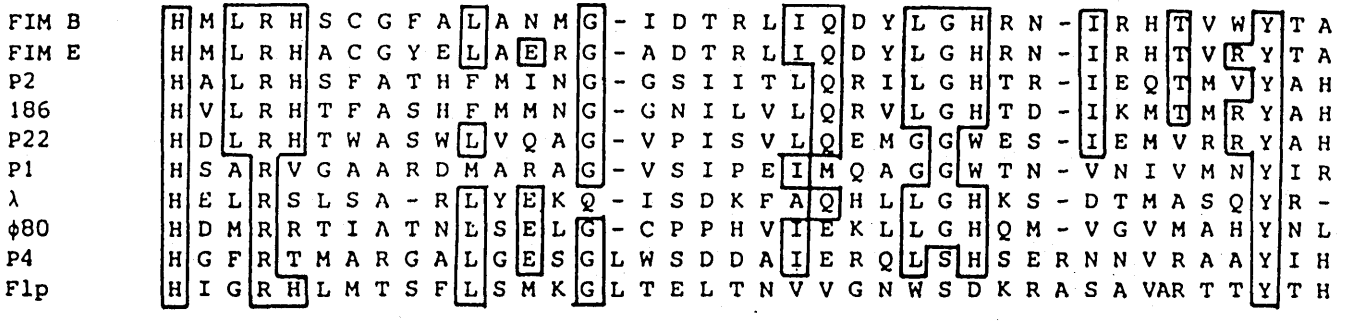


Figure 6.19 Homologies between the C-terminal regions of the FimB, FimE, Flp, and a number of phage integrases. After Dorman and Higgins (1987).

similarity over a 40 amino acid interval (Fig. 6.19). Once again however this interval is of particular importance since it includes a completely conserved Tyrosine which is postulated to be the "Active residue" in this class of recombinases (Argos et al, 1986; Dorman & Higgins, 1987). The XerA amino acid sequence contains no significant matches to the major "conserved" positions identified amongst the integrase family.

6.5.2 DNA BINDING MOTIFS

If the XerA protein is a site-specific recombinase then one might expect it to possess a DNA binding domain. A considerable amount of work has been done on the structure and properties of a number of DNA binding proteins; the most fully characterised to date being Lambda Cro, Lambda cI, and the E. coli CAP protein; for each of which there is a substantial body of genetic and physical data which identifies the particular amino-acids involved in DNA binding. The availability of crystal structures for Cro, cI and CAP has enabled the structure of each DNA binding domain to be ascertained, and in each case it comprises two alpha-helices linked by a tight turn, the physical parameters of the "Helix-Turn-Helix" structure being virtually identical in all three proteins (for review see Pabo & Sauer, 1984). By aligning the amino acid sequences of these three proteins with those of 18 other DNA binding proteins, Pabo & Sauer (1984), have proposed a Helix-Turn-Helix "sequence motif" spanning 20 residues. More recently Dodd and Egan (1987), have extended the number of Helix-turn-helix sequence motifs to a total of 37, whose alignment has been used to develop a weight-matrix with which to systematically assess the significance of any match found in novel protein sequences. The XerA sequence was searched by eye for "bestfits" to the consensus sequence, whereupon the statistical score for each was calculated according to Dodd & Egan; however, no sequences were identified which are statistically similar to the Lambda Cro-like DNA binding domain.

An alternative type of DNA binding motif, the so-called "Cysteine-zinc DNA binding finger", has been identified amongst a wide variety of eukaryotic transcriptional regulatory proteins (Berg, 1986). Whilst it seems quite reasonable to suppose that prokaryotes might also employ this means of achieving faithful DNA:Protein interaction, there is no such motif in XerA.

6.6 SECONDARY STRUCTURE PREDICTIONS

The putative secondary structure of the XerA polypeptide was calculated by PEPTIDESTRUCTURE (Devereux et al, 1984), according to the method of Garnier et al (1978). Because of the extremely speculative nature of such predictions, the data will be discussed only very briefly. The XerA secondary structure prediction has been summarised in a graphical form by PLOTSTRUCTURE (Devereux, et al, 1984; Fig. 6.20), with a additional graphical representation of the proteins hydropathy profile, generated by PEPPLOT, according to the method of Kyte and Doolittle (1982; Fig. 6.21). From consideration of figures 6.20 and 6.21, one can summarise the main characteristics of XerA as follows. The protein's hydropathy profile is distinctly dichotomous, the N-terminal 65 amino-acids being very hydrophilic, whereas the remainder of the protein is weakly hydrophobic. This profile is entirely consistent with XerA being a water soluble, globular protein, whose major surface determinants are formed by the N-terminal half of the polypeptide, with the bulk of the remainder likely to be buried internally (Kyte and Doolittle, 1982). The PLOTSTRUCTURE output suggests that the N-terminal end of XerA (the 65 hydrophilic amino-acids), are likely to form two major alpha-helices, linked by a number of turns and a short stretch of Beta-sheet. The remainder of the protein is then predicted to adopt a predominantly Beta-sheet conformation with a third major alpha-helix at the extreme C-terminus (Fig.6.20).

6.7 CONCLUSIONS AND DISCUSSION

Perhaps the single most tantalising fact to arise from DNA sequence comparisons was that of the presence of a site very similar to the cer cross-over region. The existence of this site (plus its associated IHF binding site ?), may be indicative of a chromosomal activity of the Xer recombination system, perhaps involving site-specific chromosomal re-arrangements; if so then recombination may occur between heterologous sites, thus explaining the failure of two XerA-sites to participate in efficient recombination. It would seem probable that any such re-arrangement between two chromosomal sites would involve inversion as opposed to deletion; this assumption being based upon a prejudice towards the conservation of chromosomal sequence: However, one should bear in mind the possibility that the xerA gene is infact part of some unknown lysogenised/integrated genetic element, in which case a "deletion" reaction might serve to

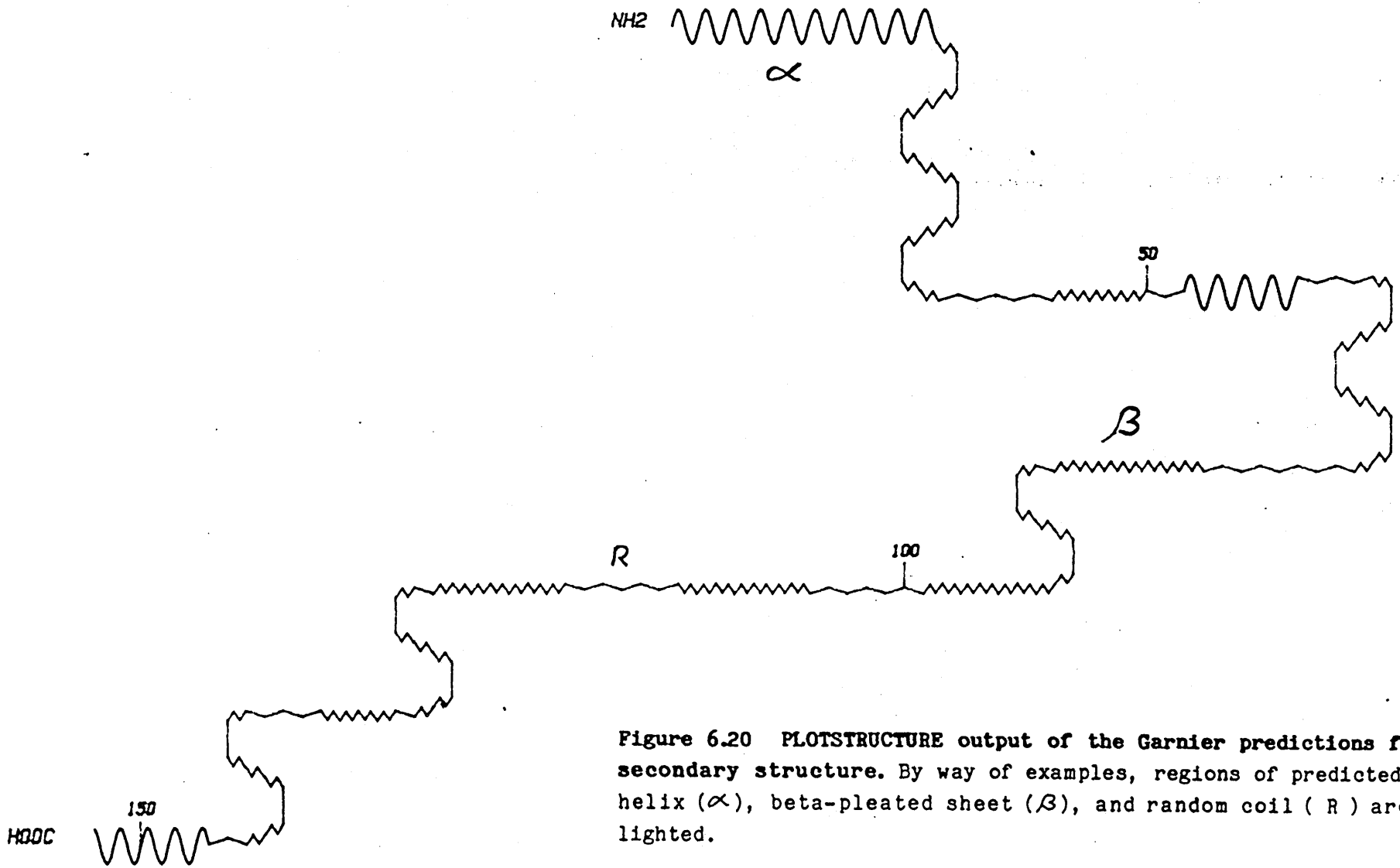


Figure 6.20 PLOTSTRUCTURE output of the Garnier predictions for XerA secondary structure. By way of examples, regions of predicted alpha-helix (α), beta-pleated sheet (β), and random coil (R) are highlighted.

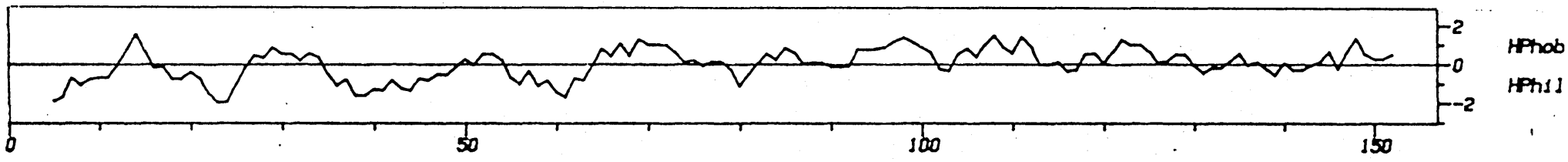


Figure 6.21 Hydropathy profile of XerA: As calculated by PEPPLOT (Gribskov et al, 1986, according to the method of Kyte and Doolittle, 1982). The plot represents the mean hydropathy calculated over a window of 7 residues.

excise that element. One could test for any specific re-arrangements centered on the *xerA* locus by screening for "Restriction Fragment Length Polymorphisms", using pGS30 as a probe against Southern blots of various restrictions of wild type chromosomal DNA.

It would seem apparent that recombination at the XerA-site would disrupt the XerA coding sequence, unless of course the partner-site with which it recombines can re-generate the *xerA* structural gene. Precedents exist for inversion systems with a site within a structural gene, such that inversion generates an altered coding sequence; e.g. the Mu *gin* system, or the *Salmonella hin* system (van de Putte *et al*, 1980; Simon *et al*, 1980). However, the evidence suggests that XerA is essential to Xer-mediated recombination, such that an inversion which disrupted, or altered its coding sequence might be irreversible. Thus one might suppose that any inversion at the XerA site would re-generate a gene whose product could serve to promote the reverse reaction. Interestingly, the reverse reaction might be catalysed by a variant form of XerA; by way of a precedent, the *E. coli fimA* inversion system requires both the *fimB* and *fimE* gene products, one of which promotes recombination in one direction, whereas the other promotes the reverse reaction. The FimB and FimE proteins are highly homologous at the amino-acid level, and are encoded by tandem genes believed to be the products of a gene duplication event (Klemm, 1986). An inversion system which alters the coding sequence of the invertase itself would represent an entirely novel DNA re-arrangement system.

Should inter-molecular recombination between a *cer* site and the chromosomal site be possible, then this could serve to integrate a cer^{+} plasmid into the chromosome, a phenomenon which might contribute to the stabilising affect of *cer*. One could test this possibility by probing for chromosomally integrated copies of ColE1 in Southern blots.

An alternative interpretation as to the significance of the *xer*-site can be proposed which is of some considerable interest. The Xas sequence might not be a recombination site *per se*, but might represent a recombinational enhancer-like element analagous to the Fis binding sites identified in the Gin and Hin systems (see Introduction). This possibility is particularly intriguing since the Gin and Hin enhancer sequences are present within the coding sequences of the respective recombinase genes. Given the Gin/Hin precedents, one ought to consider the possibility that the *xerA*-site stimulates recombination

at a nearby site, perhaps outwith the xerA gene.

The DNA and protein sequence analyses have failed to produce any clues as to the biochemical function of the XerA polypeptide. If XerA is indeed a site-specific recombinase then it may represent a novel class of such enzymes since it bears little or no similarity to either of the two curently recognised families of site-specific recombinases.

CHAPTER 7

SEQUENCE OF THE xerB GENE

7.1 INTRODUCTION

The 1.9Kb HindIII fragment encoding the xerB gene was sequenced for the same major reasons as those outlined for xerA in chapter 6. Once again the generation of ExoIII deletions played a crucial role in the sequencing strategy; the deletion substrates being the M13 mp19 derivatives mCS112 and mCS113, corresponding to the two alternative orientations of the HindIII fragment with respect to the vector's polylinker (Fig 7.1). Exonuclease III deletion time courses were performed on BamHI/SstI restricted mCS112 or mCS113 as described in chapter 6; again deletion clones were screened for their insert size by agarose gel electrophoresis after restriction of RF DNA with HindIII, suitably sized clones were then selected for DNA sequence determination. The complete "First" strand was determined from mCS112 plus 13 of its deletion derivatives, whereas the second strand was derived from mCS113, plus 6 specific sub-clones mCS114, 115, 116, 117, 118, and 119 (Fig 7.2), then completed with two mCS113 ExoIII deletion clones (Table 7.1). The complete sequence of the HindIII fragment, XERBCON1, is 1921bp in length, and is represented in a fully annotated form in figure 7.4.

7.2 MINICELL EXPRESSION STUDIES

Prior to a discussion of the sequence data it is instructive to consider the XERBCON1-specific polypeptides expressed in minicells. The 1.9Kb HindIII fragment mCS112 was cloned into the unique HindIII site in the ptac expression vector pBAD; pCS126 and pCS127 coresponding to the opposite orientations of the cloned fragment with respect to the vector's tac promoter (as identified by EcoRI restriction; Fig 7.3a). Plasmid encoded proteins were then identified by virtue of their de novo synthesis in minicells purified from DS944 carrying either pBAD, pCS126, or pCS127. Minicells were purified, then incubated in the presence of ^{35}S methionine, either in the presence or absence of IPTG (see materials and methods). Figure 7.3b (lanes 6 + 7) clearly shows that pBAD encodes three protein species, of 33, 29 and 27 KDa respectively. These three proteins correspond to different processed forms of beta-lactamase (Dougan et al, 1979). As expected, both pCS126 and pCS127 were also found to express beta-lactamase (Fig. 7.3b, lanes 1-4). The pCS127-containing minicells are relatively heavily contaminated with whole cells resulting in the appearence of a broad spectrum of protein species; however, none of the proteins present exhibit any detectable increase in intensity in

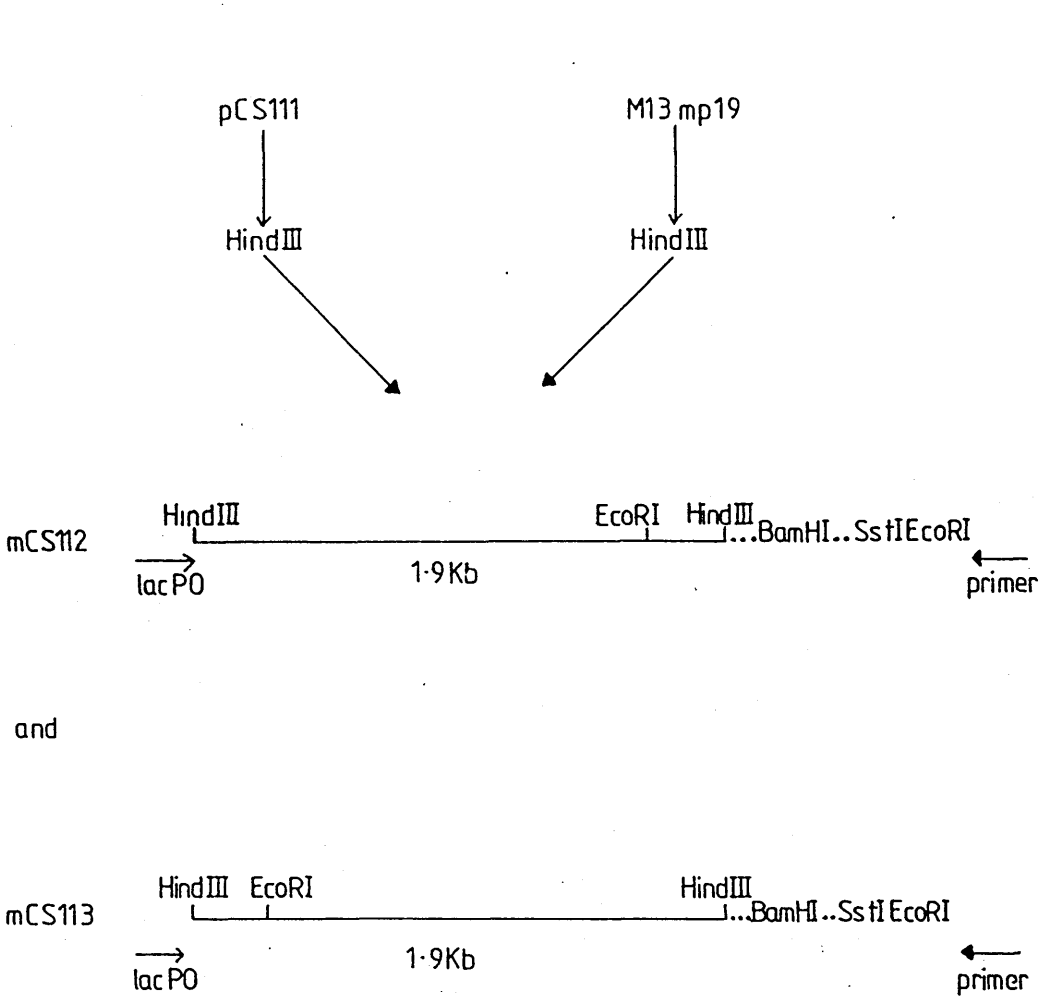


Figure 7.1 Construction of mCS112 and mCS113

As before periods (.) are used to indicate the presence of additional polylinker sites.

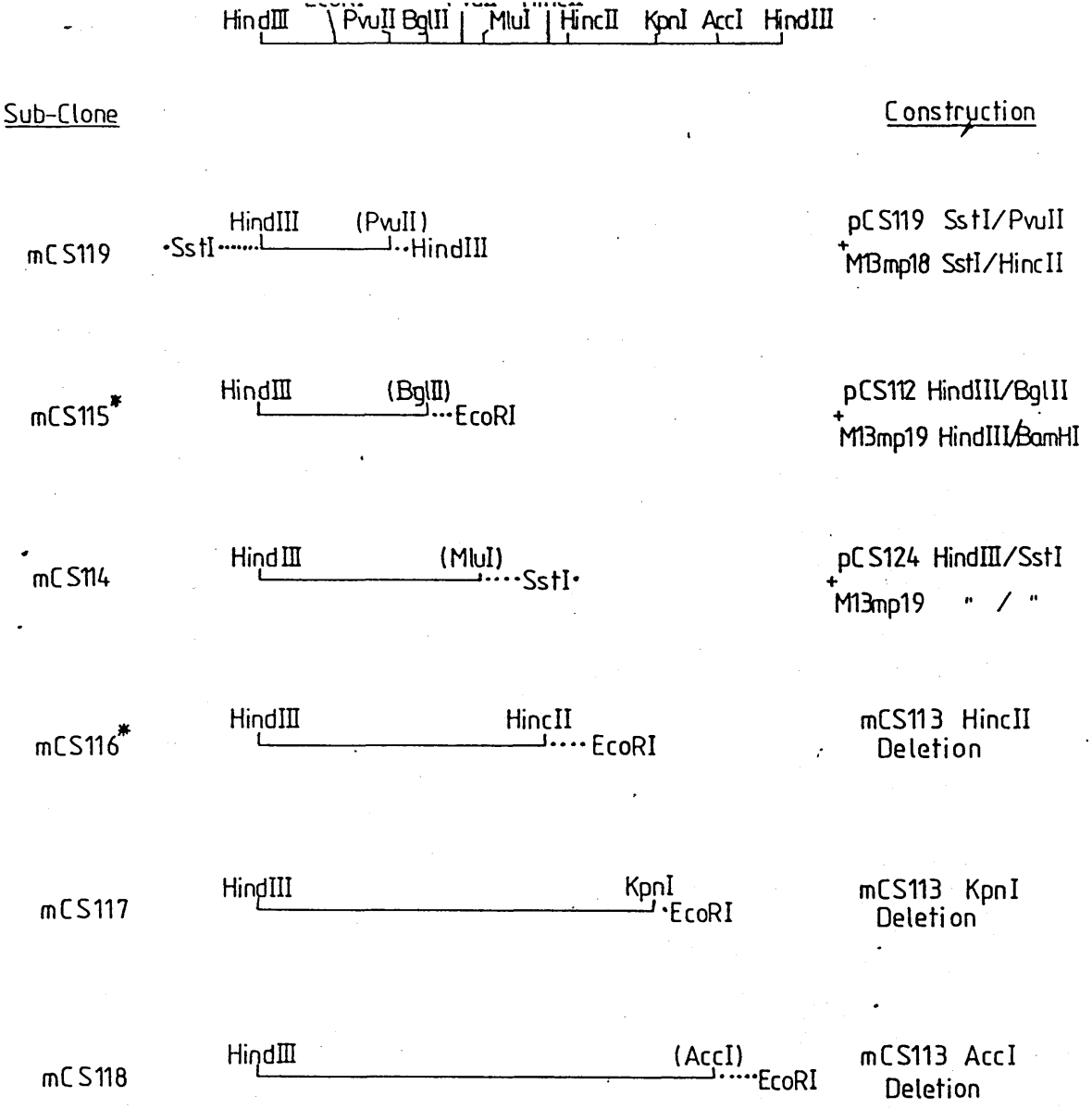


Figure 7.2 Construction of a Range of Specific Sub-Clones for the Purposes of DNA Sequence Determination.

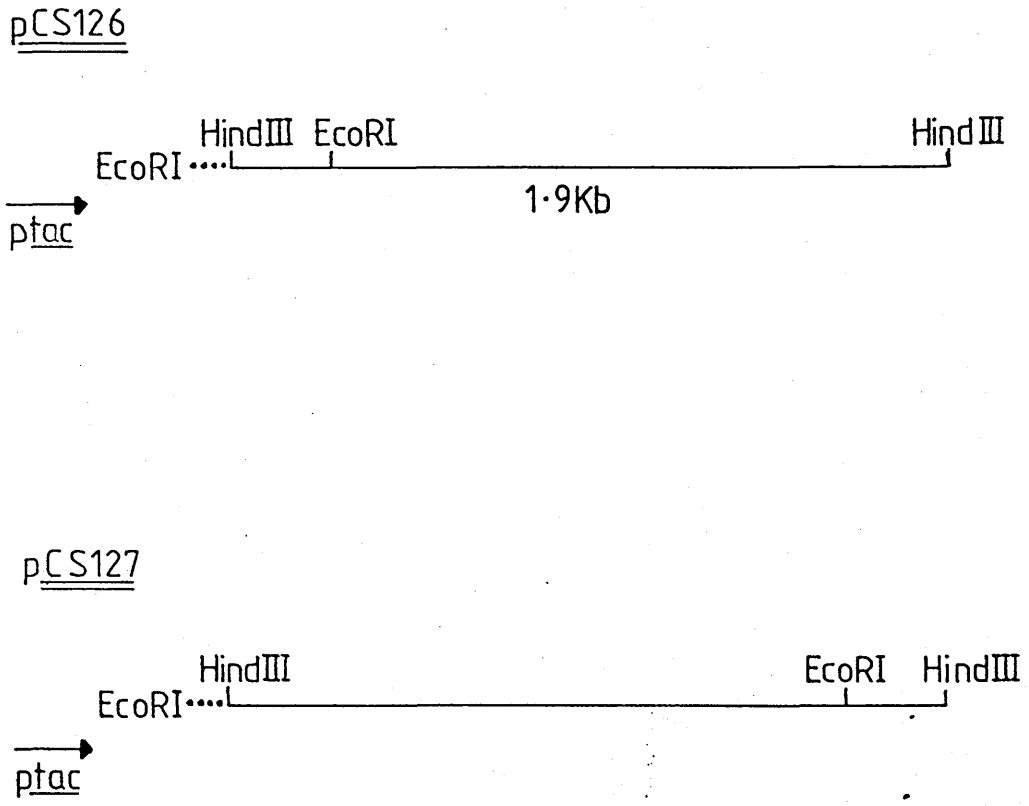


Figure 7.3a: Orientation of the 1.9Kb HindIII fragment relative to ptac in pCS126 and pCS127.

Table 7.1 Derivation of the XERBCON1 Sequence

Plus Strand

Clone	File-Name	XERBCON1 Co-ordinates	Sub-sequence Length	Gel(s)
mCS112	XERB.MCS112	1-208	208	G,H
mCS112Δ2.3	XERB.MCS112D23	175-375	201	G,H
" 5.2	" D52	269-455	187	G
" 7.5	" D75	303-492	190	G,H
" 3.5	" D35	398-571	174	G,H
" 5.3	" D53	475-743	268	I,K,L
" 5.1	" D51	620-868	249	K,L
" 6.1	" D61	859-1048	190	B,I,L
" 7.11	" D711	915-1225	311	O
" 8.3	" D83	1121-1415	297	C,I,J
" 9.2	" D92	1239-1480	241	K,L
" 9.15	" D915	1368-1565	203	M,N
" 7.7	" D77	1524-1786	268	M,N
" 8.1	" D81	1641-1921	280	I,J

Minus Strand

Clone	File-Name	XERBCON1 Co-ordinates	Sub-sequence Length	Gel(s)
mCS113	MCS113.MCS113	1921-1623	299	P,Q
mCS118	" .DACC	1687-1453	235	S,T,Z
mCS117	" .DKPN	1467-1237	231	S,T
mCS113Δ6.9	" .D69	1346-1013	334	P,R,U
mCS116	" .DHINCII	1070-815	256	U,V,X'
mCS114	" .MLUI	832-618	215	U,V
mCS115	" .BGLII	626-430	197	W
mCS119	" .PVUII	477-227	251	W,X'
mCS113Δ4.2	" .D42	255-1	255	P

Figure 7.4 The Annotated XERBCON1 Sequence

```
                                                         XerB
 HindIII                                          S.D.       Met
> AAGCTTGCAATTCTATCTGTAGCCACCGCCGTTGTCTTTAAGATTCAGGAGCGTAGTGCA
< TTCGAACGTTAAGATAGACATCGGTGGCGGCAACAGAAATTCTAAGTCCTCGCATCACGT
          10        20        30        40        50        60

   GluPheSerValLysSerGlySerProGluLysGlnArgSerAlaCysIleValValGly
> TGGAGTTTAGTGTAAAAAGCGGTAGCCCGGAGAAACAGCGGAGTGCCTGCATCGTCGTGG
< ACCTCAAATCACATTTTTCGCCATCGGGCCTCTTTGTCGCCTCACGGACGTAGCAGCACC
          70        80        90       100       110       120

   ValPheGluProArgArgLeuSerProIleAlaGluGlnLeuAspLysIleSerAspGly
> GCGTCTTCGAACCACGTCGCCTTTCTCCGATTGCAGAACAGCTCGATAAAATCAGCGATG
< CGCAGAAGCTTGGTGCAGCGGAAAGAGGCTAACGTCTTGTCGAGCTATTTTAGTCGCTAC
         130       140       150       160       170       180

   TyrIleSerAlaLeuLeuArgArgGlyGluLeuGluGlyLysProGlyGlnThrLeuLeu
> GGTACATCAGCGCCCTGCTACGTCGGGGCGAACTGGAAGGAAAACCGGGGCAGACATTGT
< CCATGTAGTCGCGGGACGATGCAGCCCCGCTTGACCTTCCTTTTGGCCCCGTCTGTAACA
         190       200       210       220       230       240

                                      EcoRI
   LeuHisHisValProAsnValLeuSerGluArgIleLeuLeuIleGlyCysGlyLysGlu
> TGCTGCACCATGTTCCGAATGTACTTTCCGAGCGAATTCTCCTTATTGGTTGCGGCAAAG
< ACGACGTGGTACAAGGCTTACATGAAAGGCTCGCTTAAGAGGAATAACCAACGCCGTTTC
         250       260       270       280       290       300

   ArgGluLeuAspGluArgGlnTyrLysGlnValIleGlnLysThrIleAsnThrLeuAsn
> AACGTGAGCTGGATGAGCGTCAGTACAAGCAGGTTATTCAGAAAACCATTAATACGCTGA
< TTGCACTCGACCTACTCGCAGTCATGTTCGTCCAATAAGTCTTTTGGTAATTATGCGACT
         310       320       330       340       350       360

   AspThrGlySerMetGluAlaValCysPheLeuThrGluLeuHisValLysGlyArgAsn
> ATGATACTGGCTCAATGGAAGCGGTCTGCTTTCTGACTGAGCTGCACGTTAAAGGCCGTA
< TACTATGACCGAGTTACCTTCGCCAGACGAAAGACTGACTCGACGTGCAATTTCCGGCAT
         370       380       390       400       410       420
```

Fig. 7.4 cont.

```
     AsnTyrTrpLysValArgGlnAlaValGluThrAlaLysGluThrLeuTyrSerPheAsp
> ACAACTACTGGAAAGTGCGTCAGGCTGTCGAGACGGCAAAAGAGACGCTCTACAGTTTCG
< TGTTGATGACCTTTCACGCAGTCCGACAGCTCTGCCGTTTTCTCTGCGAGATGTCAAAGC
        430       440       450       460       470       480
     GlnLeuLysThrAsnLysSerGluProArgArgProLeuArgLysMetValPheAsnVal
> ATCAGCTGAAAACGAACAAGAGCGAACCGCGTCGTCCGCTGCGTAAGATGGTGTTCAACG
< TAGTCGACTTTTGCTTGTTCTCGCTTGGCGCAGCAGGCGACGCATTCTACCACAAGTTGC
        490       500       510       520       530       540

     ProThrArgArgGluLeuThrSerGlyGluArgAlaIleGlnHisGlyLeuAlaIleAla
> TGCCGACCCGCCGTGAACTGACCAGCGGTGAGCGCGCGATCCAGCACGGTCTGGCGATTG
< ACGGCTGGGCGGCACTTGACTGGTCGCCACTCGCGCGCTAGGTCGTGCCAGACCGCTAAC
        550       560       570       580       590       600

     AlaGlyIleLysAlaAlaLysAspLeuGlyAsnMetProProAsnIleCysAsnAlaAla
> CCGCCGGGATTAAAGCAGCAAAAGATCTCGGCAATATGCCGCCGAATATCTGTAACGCCG
< GGCGGCCCTAATTTCGTCGTTTTCTAGAGCCGTTATACGGCGGCTTATAGACATTGCGGC
        610       620       630       640       650       660

     TyrLeuAlaSerGlnAlaArgGlnLeuAlaAspSerTyrSerLysAsnValIleThrArg
> CTTACCTCGCTTCACAAGCGCGCCAGCTGGCTGACAGCTACAGCAAGAATGTCATCACCC
< GAATGGAGCGAAGTGTTCGCGCGGTCGACCGACTGTCGATGTCGTTCTTACAGTAGTGGG
        670       680       690       700       710       720

     ValIleGlyGluGlnGlnMetLysGluLeuGlyMetHisSerTyrLeuAlaValGlyGln
> GCGTTATCGGCGAACAGCAGATGAAAGAGCTGGGGATGCATTCCTATCTGGCGGTCGGTC
< CGCAATAGCCGCTTGTCGTCTACTTTCTCGACCCCTACGTAAGGATAGACCGCCAGCCAG
        730       740       750       760       770       780

     GlySerGlnAsnGluSerLeuMetSerValIleGluTyrLysGlyAsnAlaSerGluAsp
> AGGGTTCGCAAAACGAATCGCTGATGTCGGTGATTGAGTACAAAGGCAACGCGTCGGAAG
< TCCCAAGCGTTTTGCTTAGCGACTACAGCCACTAACTCATGTTTCCGTTGCGCAGCCTTC
        790       800       810       820       830       840

     AlaArgProIleValLeuValGlyLysGlyLeuThrPheAspSerGlyGlyIleSerIle
> ATGCACGCCCAATCGTGCTGGTGGGTAAAGGTTTAACCTTCGACTCCGGCGGTATCTCGA
< TACGTGCGGGTTAGCACGACCACCCATTTCCAAATTGGAAGCTGAGGCCGCCATAGAGCT
        850       860       870       880       890       900
```

Fig 7.4 cont.

LysProSerGluGlyMetAspGluMetLysTyrAspMetCysGlyAlaAlaAlaValTyr

> TCAAGCCTTCAGAAGGCATGGATGAGATGAAGTACGATATGTGCGGTGCGGCAGCGGTTT

< AGTTCGGAAGTCTTCCGTACCTACTCTACTTCATGCTATACACGCCACGCCGTCGCCAAA

910 920 930 940 950 960

Hydrophobic

GlyValMetArgMetValAlaGluLeuGlnLeuProIleAsnValIleGlyValLeuAla

> ACGGCGTGATGCGGATGGTCGCGGAGCTACAACTGCCGATTAACGTTATCGGCGTGTTGG

< TGCCGCACTACGCCTACCAGCGCCTCGATGTTGACGGCTAATTGCAATAGCCGCACAACC

970 980 990 1000 1010 1020

GlyCysGluAsnMetProGlyGlyArgAlaTyrArgProGlyAspValLeuThrThrMet

> CAGGCTGCGAAAACATGCCTGGCGGACGAGCCTATCGTCCGGGCGATGTGTTAACCACCA

< GTCCGACGCTTTTGTACGGACCGCCTGCTCGGATAGCAGGCCCGCTACACAATTGGTGGT

1030 1040 1050 1060 1070 1080

SerGlyGlnThrValGluValLeuAsnThrAspAlaGluGlyArgLeuValLeuCysAsp

> TGTCCGGTCAAACCGTTGAAGTGCTGAACACCGACGCTGAAGGCCGCCTGGTACTGTGCG

< ACAGGCCAGTTTGGCAACTTCACGACTTGTGGCTGCGACTTCCGGCGGACCATGACACGC

1090 1100 1110 1120 1130 1140

Hydro-

ValLeuThrTyrValGluArgPheGluProGluAlaValIleAspValAlaThrLeuThr

> ACGTGTTAACTTACGTTGAGCGTTTTGAGCCGGAAGCGGTGATTGACGTGGCGACGCTGA

< TGCACAATTGAATGCAACTCGCAAAACTCGGCCTTCGCCACTAACTGCACCGCTGCGACT

1150 1160 1170 1180 1190 1200

-phobic

GlyAlaCysValIleAlaLeuGlyHisHisIleThrGlyLeuMetAlaAsnHisAsnPro

> CCGGTGCCTGCGTGATCGCGCTGGGTCATCATATTACTGGTCTGATGGCGAACCATAATC

< GGCCACGGACGCACTAGCGCGACCCAGTAGTATAATGACCAGACTACCGCTTGGTATTAG

1210 1220 1230 1240 1250 1260

LeuAlaHisGluLeuIleAlaAlaSerGluGlnSerGlyAspArgAlaTrpArgLeuPro

> CGCTGGCCCATGAACTGATTGCCGCGTCTGAACAATCCGGTGACCGCGCATGGCGCTTAC

< GCGACCGGGTACTTGACTAACGGCGCAGACTTGTTAGGCCACTGGCGCGTACCGCGAATG

1270 1280 1290 1300 1310 1320

Fig. 7.4 cont.

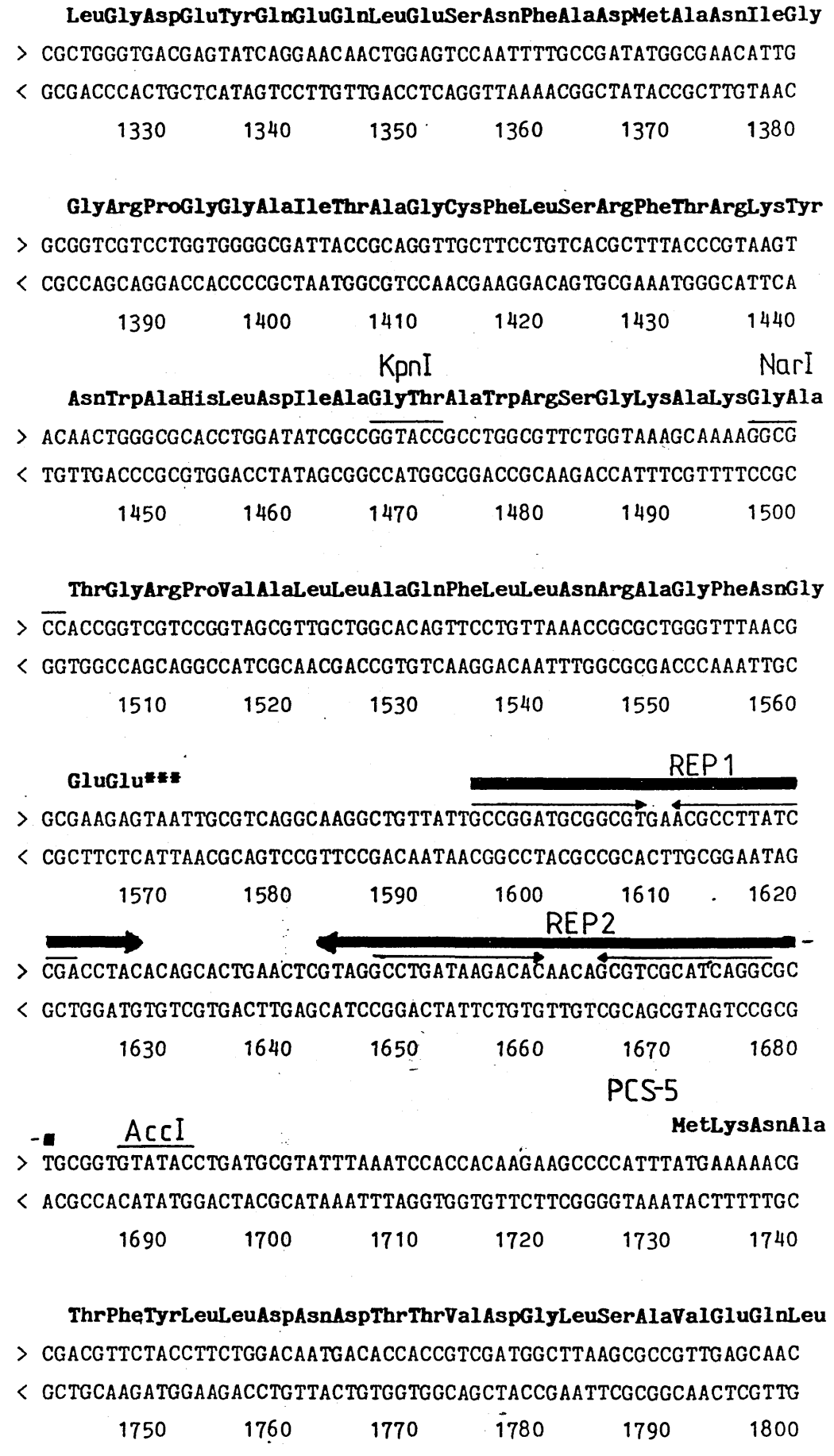

LeuGlyAspGluTyrGlnGluGlnLeuGluSerAsnPheAlaAspMetAlaAsnIleGly
> CGCTGGGTGACGAGTATCAGGAACAACTGGAGTCCAATTTTGCCGATATGGCGAACATTG
< GCGACCCACTGCTCATAGTCCTTGTTGACCTCAGGTTAAAACGGCTATACCGCTTGTAAC
1330 1340 1350 1360 1370 1380
GlyArgProGlyGlyAlaIleThrAlaGlyCysPheLeuSerArgPheThrArgLysTyr
> GCGGTCGTCCTGGTGGGGCGATTACCGCAGGTTGCTTCCTGTCACGCTTTACCCGTAAGT
< CGCCAGCAGGACCACCCCGCTAATGGCGTCCAACGAAGGACAGTGCGAAATGGGCATTCA
1390 1400 1410 1420 1430 1440
KpnI
NarI
AsnTrpAlaHisLeuAspIleAlaGlyThrAlaTrpArgSerGlyLysAlaLysGlyAla
> ACAACTGGGCGCACCTGGATATCGCCGGTACCGCCTGGCGTTCTGGTAAAGCAAAAGGCG
< TGTTGACCCGCGTGGACCTATAGCGGCCATGGCGGACCGCAAGACCATTTCGTTTTCCGC
1450 1460 1470 1480 1490 1500
ThrGlyArgProValAlaLeuLeuAlaGlnPheLeuLeuAsnArgAlaGlyPheAsnGly
> CCACCGGTCGTCCGGTAGCGTTGCTGGCACAGTTCCTGTTAAACCGCGCTGGGTTTAACG
< GGTGGCCAGCAGGCCATCGCAACGACCGTGTCAAGGACAATTTGGCGCGACCCAAATTGC
1510 1520 1530 1540 1550 1560
REP1
GluGlu***
> GCGAAGAGTAATTGCGTCAGGCAAGGCTGTTATTGCCGGATGCGGCGTGAACGCCTTATC
< CGCTTCTCATTAACGCAGTCCGTTCCGACAATAACGGCCTACGCCGCACTTGCGGAATAG
1570 1580 1590 1600 1610 1620
REP2
> CGACCTACACAGCACTGAACTCGTAGGCCTGATAAGACACAACAGCGTCGCATCAGGCGC
< GCTGGATGTGTCGTGACTTGAGCATCCGGACTATTCTGTGTTGTCGCAGCGTAGTCCGCG
1630 1640 1650 1660 1670 1680
PCS-5
AccI
MetLysAsnAla
> TGCGGTGTATACCTGATGCGTATTTAAATCCACCACAAGAAGCCCCATTTATGAAAAACG
< ACGCCACATATGGACTACGCATAAATTTAGGTGGTGTTCTTCGGGGTAAATACTTTTTGC
1690 1700 1710 1720 1730 1740
ThrPheTyrLeuLeuAspAsnAspThrThrValAspGlyLeuSerAlaValGluGlnLeu
> CGACGTTCTACCTTCTGGACAATGACACCACCGTCGATGGCTTAAGCGCCGTTGAGCAAC
< GCTGCAAGATGGAAGACCTGTTACTGTGGTGGCAGCTACCGAATTCGCGGCAACTCGTTG
1750 1760 1770 1780 1790 1800

Fig. 7.4 cont.

```
     ValCysGluIleAlaAlaGluArgTrpArgSerGlyLysArgValLeuIleAlaCysGlu
>    TGGTGTGTGAAATTGCCGCAGAACGTTGGCGCAGCGGTAAGCGCGTGCTCATCGCCTGTG
<    ACCACACACTTTAACGGCGTCTTGCAACCGCGTCGCCATTCGCGCACGAGTAGCGGACAC
          1810       1820       1830       1840       1850       1860

                                                              HindIII
     AspGluLysGlnAlaTyrArgLeuAspGluAlaLeuTrpAlaArgProAlaGluSer
>    AAGATGAAAAGCAGGCTTACCGGCTGGATGAAGCCCTGTGGGCGCGTCCGGCAGAAAGCTT
<    TTCTACTTTTCGTCCGAATGGCCGACCTACTTCGGGACACCCGCGCAGGCCGTCTTTCGAA
          1870       1880       1890       1900       1910       1921
```

Figure 7.4: A number of pertinent restriction sites are indicated, as is a potential Shine-Dalgarno (S.D.) sequence present upstream of the xerB initiation codon. Two regions of extreme (mean) hydrophobicity within the XerB primary structure are highlighted (see sect. 7.5.3). REP sequences are shown as broad arrows, whilst other regions of dyad symmetry are indicated by thin arrows.

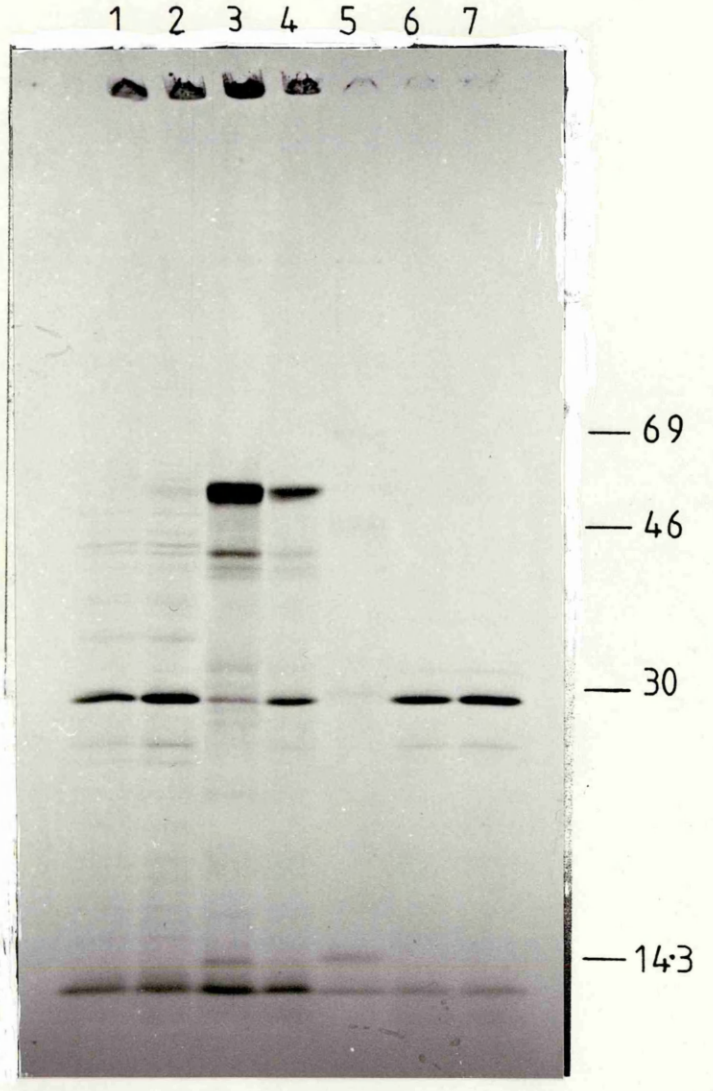


Figure 7.3 b Polypeptides expressed by pCS126 and pCS127 in minicells.

Lane	Sample	Lane	Sample
1	pCS127 (+)	5	Molecular weight markers
2	pCS127 (-)	6	pBAD (+)
3	pCS126 (+)	7	pBAD (-)
4	pCS126 (-)		

+/- in parentheses indicates the presence or absence of IPTG during the labelling of minicells. The sizes of bands are given in KDa.

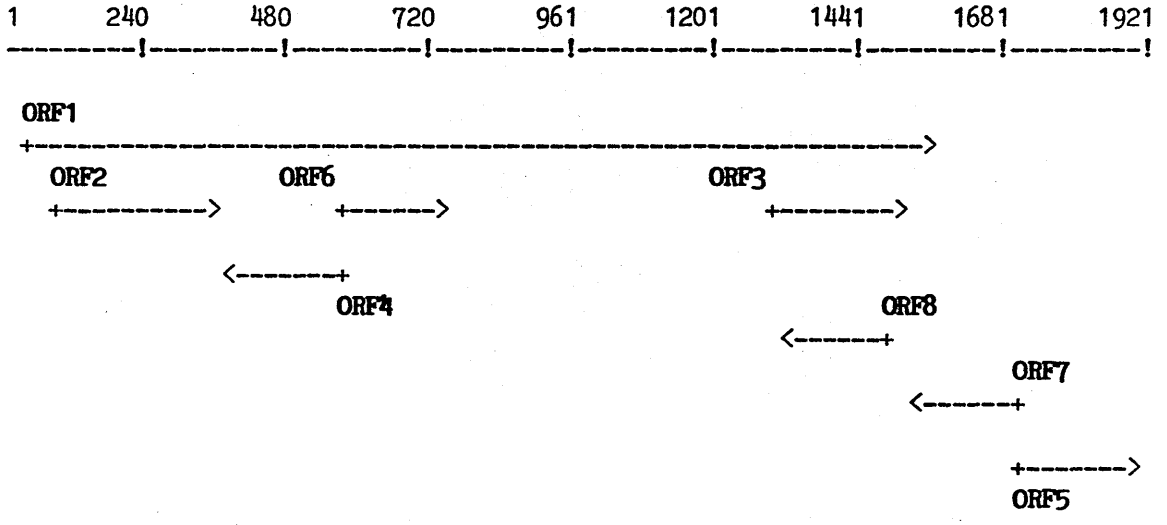
the presence of IPTG. On the assumption that IPTG induction of these minicells resulted in an increased level of transcription of the 1.9Kb insert in pCS127, then it would seem unlikely that the insert encodes any protein-coding genes on the transcribed strand: note that any proteins of less than 14KDa would not have been detected in this experiment. On the other hand, IPTG induction of pCS126-containing minicells results in an increase in the expression of a number of protein species. The major band having an apparent molecular weight equivalent to 54KDa; there are then at least nine other induced bands of 42, 40, 39, 33, 28.5, 28, 25, 15, & 14 KDa respectively (Fig. 7.3b; lanes 3 & 4). This data will be discussed in the light of potential genes identified by DNA sequence analyses.

7.3 DNA SEQUENCE ANALYSES

The same analytical procedures used in the study of the *xerA* gene sequence were applied to the XERBCON1 sequence; having discussed the underlying principles in some detail in chapter 6, I shall re-introduce each only briefly.

7.3.2 OPEN READING FRAMES

The HYPOTHETICAL analysis of ORFs within XERBCON1 revealed the presence of one large ORF, ORF-1, which extends from an ATG initiation codon to a TAA stop codon between co-ordinates 60-1569bp respectively, and which would encode a 503 amino acid polypeptide with a predicted molecular weight of 55.3KDa (Figs. 7.4, 7.5). The strand encoding ORF-1 (the designated "plus" strand), is that whose transcription from *ptac* in pCS126 results in the translation of a 54KDa protein in considerable quantities. Therefore the minicell data presented in figure 7.3c provides convincing evidence that ORF-1 is expressed *in vivo*; ORF-1 has therefore been designated "Protein Coding Sequence-1", (PCS-1). There are **no** ORFs outwith PCS-1 which could encode a polypeptide larger than 9.4KDa (see ORF-2; Fig. 7.5), and therefore it would appear likely that all those XERBCON1-specific proteins observed in mini-cells must be products of ORF-1: however this assumes that no major glycosylation of small polypeptides had taken place, a process which can result in a considerable decrease in a protein's electrophoretic mobility. One could test for the presence of glyco-proteins by screening for any shifts in observed electrophoretic mobility after treatment with an endoglycosidase. If some, or all, of the smaller polypeptides observed in mini-cells are indeed products of ORF-1, then



ORF	START CODON	CO-ORDINATES	STRAND	CODONS	PREDICTED MOLECULAR WEIGHT (KDa)
1	AUG	60 - 1568	+	503	55.3
2	GUG	103 - 357	+	85	9.4
3	GUG	1300 - 1515	+	72	7.9
4	GUG	587 - 375	-	71	7.8
5	AUG	1731 - *	+	63*	-
6	GUG	568 - 741	+	58	6.4
7	AUG	1728 - 1558	-	57	6.3
8	GUG	1504 - 1340	-	55	6.1

Figure 7.5 HYPOTHETICAL genes identified in the XERBCON1 sequence.

they might either represent specific cleavage products of the 54KDa protein, or might correspond to the products of either translational initiation, or termination, at secondary sites within PCS-1. One could potentially identify degradative products by performing a mini-cell labelling experiment under "pulse-chase" conditions. Whereas any proteins produced by translational initiation at sites within PCS-1 could be identified in minicell-analyses of various XERBCON1 subclones which lack the primary initiation codon. The significance of these smaller polypeptides with respect to XerB activity is unknown.

The only other ORF for which there is evidence of expression is ORF-5 (Fig. 7.5), the evidence coming not from mini-cells, but from the fact that strain JM101 (lacZ delta M15) carrying mCS113 gives rise to blue plaques on X-Gal containing medium; this suggests that the inserted sequence has fortuitously generated an in-frame translational fusion with the vector's lacZ' coding sequence, which gives rise to a fusion polypeptide capable of effective intermolecular complemention. From inspection of the XERBCON1 sequence it became clear that it is ORF-5 which is fused in-frame to lacZ' in mCS113, and therefore the expression of the $LacZ^+$ phenotype requires translational initiation at some site within ORF-5 *in vivo*; for this reason ORF-5 is also considered to be a *bona fide* protein coding sequence, PCS-5.

The complementation data discussed in chapter 5, which identified a 1680bp HindIII/AccI fragment as the minimum xerB1 complementing clone, provides convincing evidence that PCS-1 must represent the xerB gene (Fig. 5.15). This conclusion being further corroborated by the restriction map of pXB1 (Chapter 5) from which one can estimate that in the xerB1 mutant locus, Tn5 is inserted between XERBCON1 co-ordinates 490-550bp, and would thus disrupt PCS-1.

7.3.3 BASE PERIODICITY: ZFICK ANALYSIS

The TESTCODE analysis of XERBCON1 was performed using the program ZFICK (Figs. 7.6, 7.7). The results indicate that the xerB gene (PCS-1), has a base periodicity characteristic of *E. coli* coding sequence (Fig 7.6). Furthermore, the TESTCODE statistic also begins to peak in correspondence with the start of PCS-5, implying that it too may represent coding sequence. There is also a third region in which TESTCODE predicts coding sequence, that being the region between xerB and PCS-5; an ORF on the minus strand, ORF-7, fits closely to this peak.

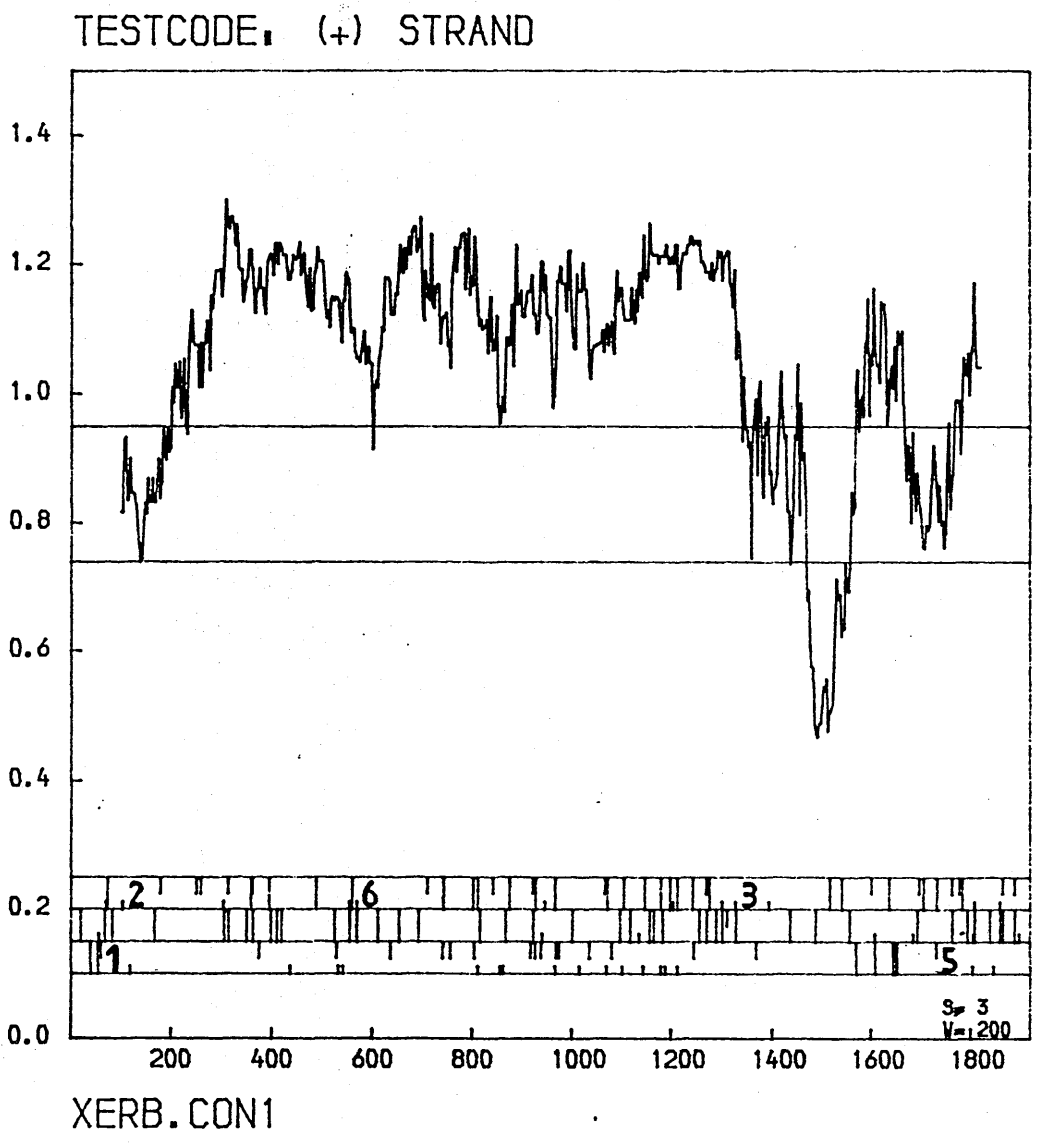


Figure 7.6 ZFICK Analysis of the XERBCON1 Plus Strand

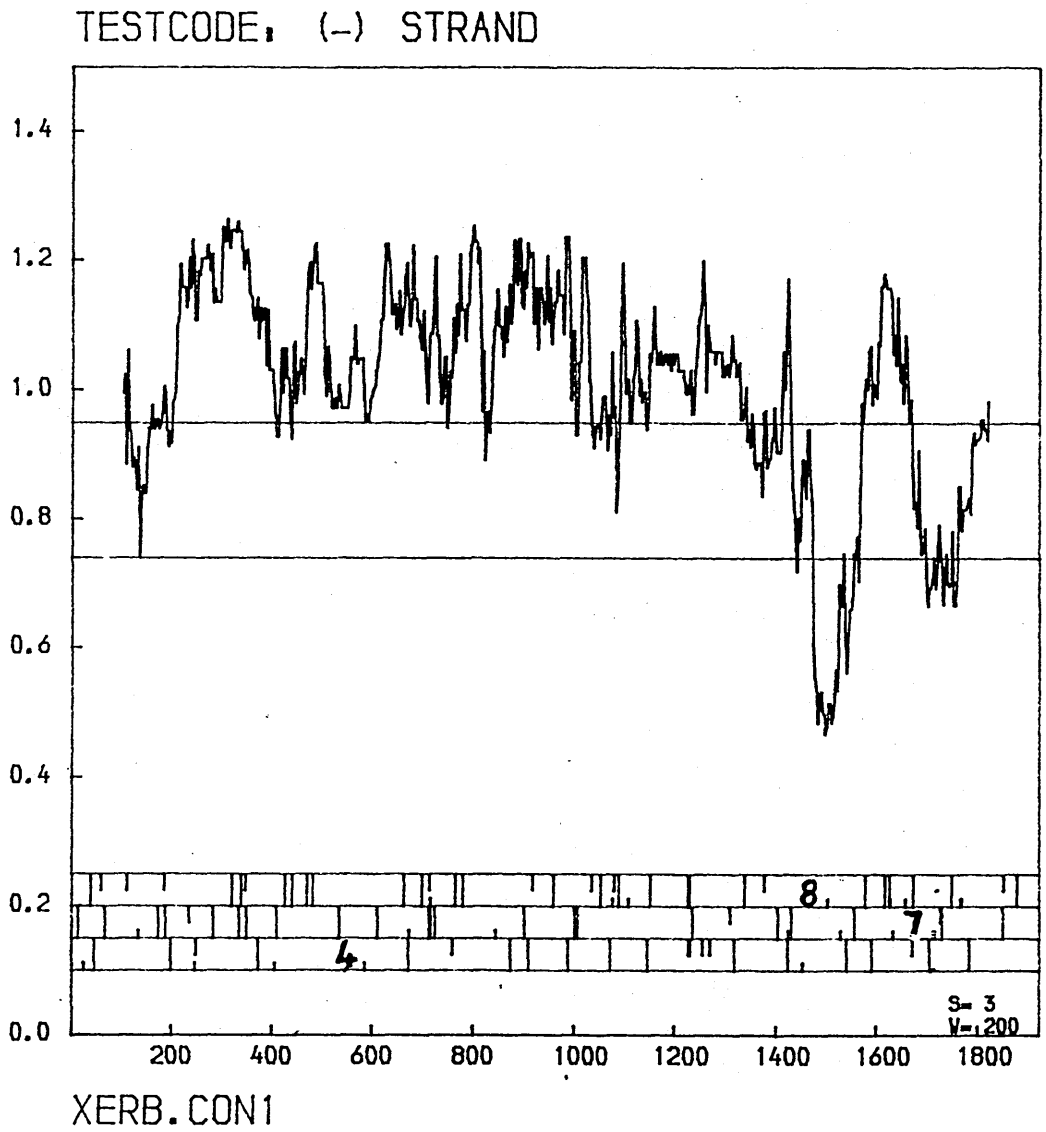


Figure 7.7 ZFICK Analysis of the XERBCON1 Minus Strand

7.3.4 CODONPREFERENCE ANALYSIS

The CODONPREFERENCE output for XERBCON1 reveals that both xerB, PCS-5 exhibit considerable codon bias, coupled with a striking scarcity of rare codons compared to all other reading frames (Figs. 7.8, 7.9). It is clear from figure 7.9 that ORF-7 resembles random sequence; note that even poorly expressed genes tend to exhibit above-random codon preference (see chapter 7). Obviously ORF-4 contains predominantly preferred codons, but as before, this must be tempered with the fact that ORF-4 is in-phase with xerB, and that the corresponding region of xerB exhibits a higher codon preference than does ORF-4, thus the former would appear to represent the dominant "partner".

It would therefore appear that ZFICK and CODONPREFERENCE contradict one another with respect to ORF-7; of course one can never escape the caveat that such poor codon bias might indicate that expression of ORF-7 must be strictly curtailed, or that the DNA/RNA sequence is under some secondary selection pressure to maintain a given nucleotide sequence. As will be discussed shortly, there is reason to believe that ORF-7 may be a fortuitous ORF, since this sequence (i.e. between xerB and PCS-5), exhibits some of the characteristics of an intergenic space.

7.3.5 TRANSLATION INITIATION SITES: ZPERTRON ANALYSIS

The most common variant of the Shine and Dalgarno sequence found in naturally occurring E. coli genes is 5'-AGGAGG-3', situated 5-9 nucleotides upstream of the initiation codon (Stormo et al, 1986). The xerB gene has a very good candidate S.D sequence in the pentanucleotide 5'-AGGAG-3', situated 9 nucleotides upstream of the predicted AUG start codon. In addition ZPERTRON analysis (using the W101 weight matrix) identifies this particular ATG codon (60bp), as a probable translational start with a PERTRON score of +50 (note that with W101, a score of > +2 is expected for genuine starts; Table 7.3). In addition the AUG "start" of ORF-2 (178bp), scores a noteable +1. Conversely, the highest PERTRON score returned for any of potential start codons within PCS-5 was found to be -145 for the AUG (1731bp; data not shown), and yet the evidence that PCS-5 is expressed in vivo is quite compelling. This highlights the major limitation of ZPERTRON analysis, this being that whilst it provides useful corroborative evidence to support a translational start, it cannot be used to rule-out any other sites.

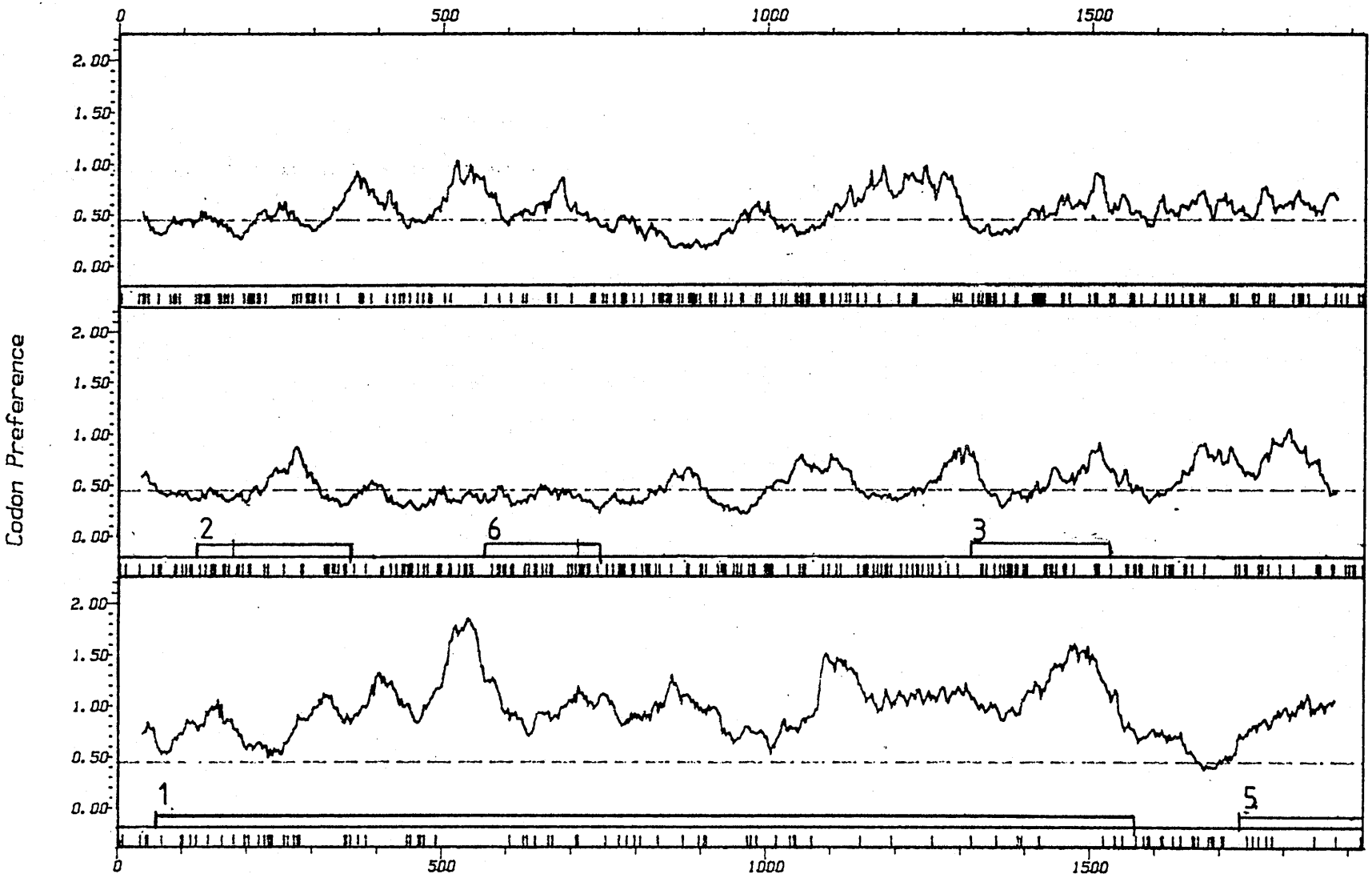


Figure 7.8 CODONPREFERENCE Analysis of the XERBCON1 Plus Strand

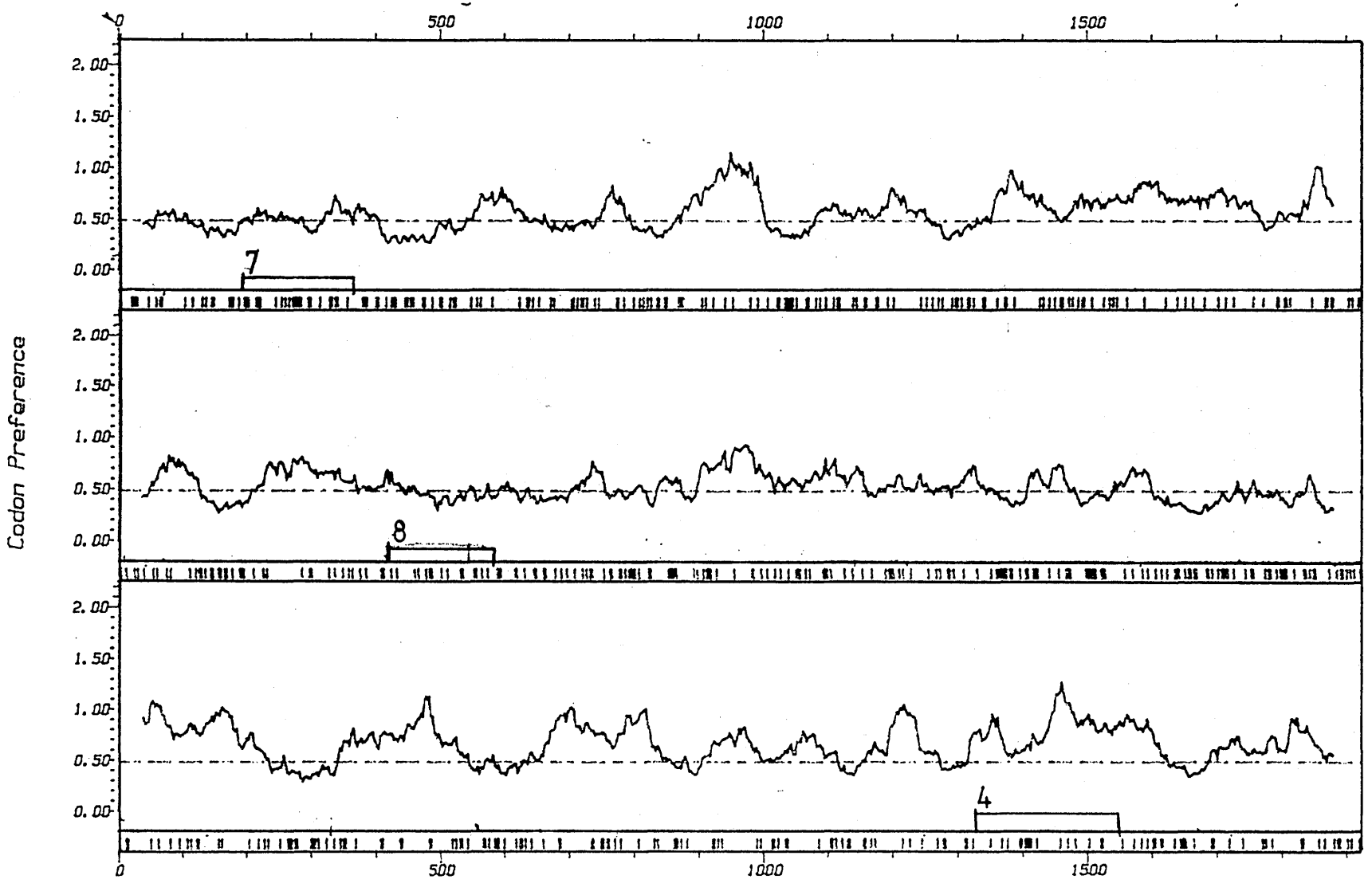


Figure 7.9 CODONPREFERENCE Analysis of the XERBCON1 Minus Strand

Table 7.3 ZPERTRON Analysis of XERBCON1

Plus Strand

Start coord.	5' --- CONTEXT --- 3'	Stop coord.	Length (codons)	Est. MW (kDa)	PERCEPTRON rating
60	AGCGTAGTGC ATG GAG ... GAG TAA	1571	503	55.3K	50
1080	GTTAACCACC ATG TCC ... GAG TAA	1571	163	17.9K	35
178	AAAATCAGCG ATG GGT ... CGC TGA	360	60	6.6K	1
1035	CTGCGAAAAC ATG CCT ... GAG TAA	1571	178	19.6K	-25
804	CGAATCGCTG ATG TCG ... GAG TAA	1571	255	28.1K	-55

Minus Strand

Start coord.	5' --- CONTEXT --- 3'	Stop coord.	Length (codons)	Est. MW (kDa)	PERCEPTRON rating
1256	CCAGCGGATT ATG GTT ... AGT TAA	1146	36	4.0K	-46

It is interesting to note that three alternative ATG codons within xerB are among the highest scores returned by ZPERTRON; this may help to explain at least some of the minor translational products observed in the mini-cell analysis described in chapter 5. Whether or not any such "internal" initiation events are of any significance *in vivo* under wild-type conditions remains to be determined.

7.3.6 PROMOTERS: ZPROM ANALYSIS

Recall that ZPROM scrutinises DNA sequence windows and scores for two statistical parameters which reflect the degree of coincidence, for any given sequence window, with the consensus promoter sequence TTGACA-17-TATAAT; the empirically determined significance levels for these two parameters being

STAT' I > 0.002

STAT' II > 0.0002

On the basis of these significance levels, ZPROM detected no strong matches to the promoter consensus sequence on either strand of XERBCON1 (Table 7.4). There are certainly no candidates for a promoter which could drive transcription of the xerB gene; there is therefore a sound argument for extending the XERBCON1 sequence further upstream of the xerB structural gene, 440bp 5' to the HindIII site are already sub-cloned in pCS111. Determination of this sequence might be considered a priority in future work on this project since it ought to either identify the xerB gene promoter, or alternatively would indicate whether or not there is a 5' coding sequence with which xerB might be co-transcribed. There are no strong candidates for a PCS-5 promoter, a fact which one might also construe as being indicative of xerB/PCS-5 co-transcription. However one must temper this with the fact that ZPROM can detect only those promoters which are similar to the consensus sequence; whilst this should identify the majority of *E. coli* promoters, it should be noted that many of the positively regulated promoters identified to-date tend to exhibit very poor "homology" to the -35 box, and thus would score poorly in ZPROM (Raibaud & Schwartz, 1984).

7.3.7 TRANSCRIPTIONAL TERMINATION SITES: TERMINATOR ANALYSIS

Computer searching of the XERBCON1 sequence using the TERMINATOR program identified no convincing candidates for rho-independent terminator sites on the plus strand. The only strong candidate throughout the entire sequence being situated at 1532bp on the minus

Table 7.4 ZPROM Analysis of the XERBCON1 Sequence

Plus Strand

Pos.	-35	-10	Statistic I	Statistic II	Distance
143	TTCTCC GATTGCAGAACAGCTCGA_	TAAAAT.	0.0031208	0.0004681	18
144	TCTCCG ATTGCAGAACAGCTCGA__	TAAAAT.	0.0004015	0.0004015	17
448	TCGAGA CGGCAAAAGAGACGCTC__	TACAGT.	0.0009392	0.0009392	17
598	TTGCCG CCGGGATTAAAGCAGCA__	AAAGAT.	0.0002919	0.0002919	17
814	TTGAGT ACAAAGGCAACGCGTCG__	GAAGAT.	0.0025560	0.0025560	17
1204	GTGCCT GCGTGATCGCGCTGGGT__	CATCAT.	0.0005273	0.0005273	17

Minus Strand

Pos.	-35	-10	I	II	Distance
1488	TTTCCA GTAGTTGTTACGGCCTT__	TAACGT.	0.0009632	0.0009632	17
1844	TTTTTA CACTAAACTCCATGCAC__	TACGCT.	0.0002676	0.0002676	17
1846	TTTACA CTAAACTCCATGCAC____	TACGCT.	0.0040584	0.0000812	15
1884	AAGACA ACGGCGGTGGCTACAGA__	TAGAAT.	0.0005112	0.0005112	17

Figure 7.10 TERMINATOR Analysis of XERBCON1

Plus Strand

```
          -40  -35  -30  -25  -20  -15  -10   -5  -1+  +5        p      s
           .    .    .    .    .    .    .    .   ..   .
 742=> TGAAAGAGCTGGGGATGCATTCCTATCTGGCGGTCGGTCAGGGTTCGCAAA    3.72      0
1707=> AATCCACCACAAGAAGCCCCATTTATGAAAAACGCGACGTTCTACCTTCTG    3.81      0
```

Minus Strand

```
          -40  -35  -30  -25  -20  -15  -10   -5  -1+  +5        p      s
           .    .    .    .    .    .    .    .   ..   .
 390=> CTGTGCCAGCAACGCTACCGGACGACCGGTGGCGCCTTTTGCTTTACCAGA    3.51     65
                     ---------    ---------
1171=> AGCTCTTTCATCTGCTGTTCGCCGATAACGCGGGTGATGACATTCTTGCTG    3.67      0
1807=> ACGATGCAGGCACTCCGCTGTTTCTCCGGGCTACCGCTTTTTACACTAAAC    3.58      0
```

Primary structure (p) threshold = 3.50
Secondary structure (s) threshold = 0

strand (Fig.7.10). Note that this "terminator" would be ideally placed to terminate transcription of ORF-7 (stop codon, 1555bp); the presence of this motif associated with gene expression provides circumstantial evidence that ORF-7 might be expressed, but there is no direct experimental evidence to support this contention at this stage. Nonetheless, I should stress that the xerB1 complementation data discussed in chapter 5 rules out ORFs 2, 7, 5 as encoding the xerB1 complementing function; however, the evidence does not preclude the possibility that any/all of these reading frames might encode associated Xer functions. Whilst no other sites were identified in TERMINATOR analysis, a number of noteable regions of dyad-symmetry were identified in the region between xerB and PCS-5 (i.e within ORF-7); Fig. 7.4).

7.3.8 xerB:lacZ TRANSLATIONAL FUSIONS

As already intimated, a translational fusion between PCS-5 and the mp19 lacZ' coding sequence resulted in effective complementation of the chromosomal lacZ delta-M15 mutation in JM101. As described in figure 7.3, six specific sub-fragments from XERBCON1 were cloned into M13. Of these sub-clones two (mCS115 and mCS116) were also found to result in a blue-plaque phenotype in JM101 (on X-Gal containing medium), indicative of the expression of a complementing fusion polypeptide; analysis of the insert/vector boundary sequences revealed that in both mCS115 and 116 it is the xerB coding sequence (ORF-1), which is fused in-frame with lacZ'. The mCS115 insert boundary is the xerB BglII site (623bp), thus fusing 188 codons of the xerB gene to the LacZ alpha fragment. Whereas, in mCS116 the boundary site is the HincII site (1070bp), thus fusing 337 codons of xerB in-frame with lacZ'. The remaining sub-clones generated for sequencing, mCS114, 117, 118, and 119 all produced in white plaque phenotypes; in all but mCS118, the fusion boundary is within the xerB coding sequence, but is in an alternative reading frame.

In addition to the fortuitous translational fusions generated above, I also specifically fused the xerB coding sequence at the NarI site (1497bp), to each of the three translational fusion vectors, pNM480, pNM481, and pNM482; these vectors carry the entire lacZ coding sequence but lack an in-frame start codon and ribosome binding site; the three vectors are identical to one another except for the relative phase of their polylinker with respect to the LacZ reading frame (Minton et al, 1984). The 1.5Kb BamHI/NarI fragment from pCS119 was

cloned into the compatible "sticky-ends" created by BamHI/AccI restriction of pNM480, 481 and 482. Each ligation was transformed into DS947 (delta lac/pro, SupO) and clones selected on L-agar+ AP. The desired clones were first identified by size on single-colony gels, then suitable candidates were confirmed by restriction of STET DNA with either PstI or HindIII. Thus the plasmids pCS140, pCS141, and pCS142 were isolated, corresponding to the 1.5Kb BamHI/NarI fragment from pCS119 cloned into pNM480, 481 and 482 respectively. From examination of the vector/insert boundaries it was predicted that xerB would be fused in-frame to lacZ in pCS142. The LacZ phenotypes of DS947 carrying either pCS140, 141 or 142 were then tested by plating each out onto L-agar + Ap+ X-Gal, followed by overnight incubation at 37°C. Colonies arising from DS947+ pCS140, or DS947+ pCS141 were white, whereas those from DS947+ pCS142 were dark blue, thus confirming that the xerB reading frame is translated through the NarI site (1497bp). The pCS142-encoded fusion polypeptide is predicted to include the first 480 amino-acids of XerB, (lacking only 23 residues from the C-terminus), and thus has a predicted molecular weight of approximately 165KDa. Interestingly, pCS140, 141 and 142 were all found to be unable to complement pCS202 resolution in CSX4, indicating that the C-terminal 23 residues of XerB are of vital significance to its activity.

These gene-fusion results provide conclusive evidence that the xerB reading-frame is translated at three positions well seperated along its length; those being the BglII site (623bp, mCS115), the HincII site (1070bp, mCS116) and the NarI site (1497bp, pCS142).

7.4 DNA SEQUENCE COMPARISONS

The XERBCON1 sequence was screened against the GENBANK and EMBL sequence databases using the WORDSEARCH program, but no files were identified in either database with significant sequence "homology". The XERBCON1 sequence has also been compared in fine detail with both XERA112 and ColE1, but again no significant sequence identities were found; more specifically, no equivalents to cer-, nut-, or IHF- like sites of XERA112 are present in XBCON1.

7.4.2 IDENTIFICATION OF REP SEQUENCES

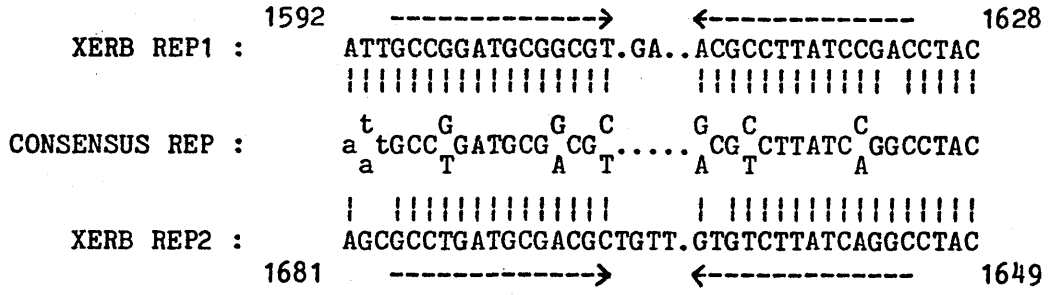
Two copies of the so-called REP (Repetitive Extragenic Palindromic) sequence are present in the XERBCON1 sequence, both are situated between the end of xerB and the start of PCS-5, with the two

copies, REP-1 and REP-2, being present in inverted orientation relative to one another (Fig. 7.4): both REP-1 and REP-2 exhibit extensive "homology" to the consensus REP sequence (Gilson et al, 1987; Fig. 7.12).

To date 118 copies of the REP sequence have been identified within "databased" E. coli sequences, their observed incidence predicts the presence of 500-1000 copies throughout the genome, thus accounting for some 1% of E. coli DNA. All known REP sequences are within transcription units, but **none** are within known coding sequences; they are present either between genes within an operon (inter-cistronic), or in the 3' untranslated region of mRNA (post-cistronic) (Stern et al, 1984; Gilson et al, 1987). In all cases where REPs are present between two genes whose relative expression levels are known, the upstream gene (promoter proximal), is invariably expressed at a higher level than is the downstream gene (Stern et al, 1984). This phenomenon may be explained, at least in part, in the light of recent findings which suggest that REP sequences act to stabilise upstream mRNA (apparently by providing protection from 3'-5' exonucleolytic activity). The effective concentration of an mRNA has been found to be increased by up to 10-fold in the presence of a 3' REP sequence (Newbury et al, 1987). It has also been suggested that REP sequences might be involved in chromosomal organisation, perhaps providing nucleation points for condensation of the nucleoid. Alternatively they have been implicated in a potential "gene-shuffling" mechanism involving recombination between REPs within different operons. The presence, and position, of the REP sequences found in XERBCON1 have a number of major implications.

1. They suggest an "operonic" organisation between xerB and PCS-5; such an organisation could explain the apparent absence of a PCS-5 specific promoter sequence. If indeed xerB and PCS-5 are two cistrons within an operon, then it is very probable that PCS-5 encodes an Xer related function; there is therefore a strong case for extending the XERBCON1 sequence into PCS-5. A role for PCS-5 in Xer recombination could then be investigated by gene-disruption in vitro, followed by the introduction of the mutagenised PCS-5 allele into the chromosome via gene-replacement, whereupon the Xer phenotype of the engineered mutant could be determined. If PCS-5 does encode an Xer function, then the apparent failure to isolate any PCS-5 mutants to date, might indicate that it encodes an essential function, in which case any attempted gene-replacements using a completely disrupted copy of PCS-5

Figure 7.12 Comparisons Between the REP Consensus Sequence and XERBCON1



Regions of hyphenated dyad symmetry within each element are highlighted. The numbers refer to base-pair co-ordinates in XERBCON1.

would be likely to fail. Note that 8 out of 9 of the "XerB" mutants which are complemented by pCS112 (1.9Kb HindIII), have not been tested with any of the pCS112 sub-clones; it is therefore plausible that some of these eight might in fact be complemented by an alternative xer gene(s) within pCS112, possibly the truncated PCS-5 product.

2. Since there are no precedents to-date for REP sequences within a coding sequence, then it would appear that ORF-7 does not resemble a typical E. coli gene.

3. Finally, the REP sequences are likely to aid in the stabilisation of xerB mRNA; this may prove a useful feature in facilitating the overexpression of XerB protein for the purpose of purification.

7.5 PROTEIN SEQUENCE ANALYSES

The inferred primary structures for XerB and for the N-terminal portion of PCS-5 are listed in figure 7.4. As previously intimated, XerB is 503 amino acids in length, whilst only the first 63 amino-acids of PCS-5 can currently be predicted from the available DNA sequence. The amino acid composition of XerB reveals a considerable number of charged residues from which one would predict it to be a weakly acidic protein (Table 7.5).

The amino acid sequences of XerB, and that derived from PCS-5, were then employed in protein sequence comparisons. Once again, no significant matches to either sequence were found by WORDSEARCH when scanning the NBRF protein database. Detailed scrutiny of each sequence revealed no similarities to the highly conserved regions of either the "Integrase" or "Resolvase/Invertase" families of site-specific recombinases, nor does either exhibit a nucleotide binding site motif (see chapter 6).

The XerA, XerB and PCS-5 sequences were compared with one another using ZDOTMAT, GAP and BESTFIT, but no highly conserved sequences were identified either between all three proteins, or between any pair of proteins. Whilst a number of minor comparisons could be drawn, these relied heavily upon the presence of chemically similar amino-acids (as defined by Dayhoff et al, 1978), with very few exact matches. In no case was the observed degree of similarity any greater than would be expected by chance; indeed better quality

Table 7.5 Amino Acid Composition of the XerB Polypeptide

Aromatic	Hydrophobic	Weakly Hydrophobic	Acidic	Amide	Basic	Cysteine
F = 12	L = 50	P = 20	D = 19	N = 23	K = 24	C = 9
W = 4	I = 27	A = 48	E = 38	Q = 21	R = 32	
Y = 14	V = 34	G = 49			(H = 10)	
	M = 16	S = 29				
		T = 24				
30	127	170	57	44	56 (10)	9

alignments were achieved after "randomising" each sequence by shuffling their respective amino-acid compositions into novel (random) primary structures (data not shown).

7.5.2 DNA BINDING MOTIFS

Using the same procedures as described in chapter 6, the XerB primary sequence was searched for the presence of either a Lambda Cro-like DNA binding domain, or a Cysteine-zinc DNA binding finger, however no significant matches to either sequence motif were found.

7.5.3 XerB HYDROPATHY PROFILE AND SECONDARY-STRUCTURE PREDICTIONS

A striking feature of the protein's hydropathy profile is the presence of clusters of either hydrophobic or hydrophilic residues (Figs. 7.13). Obviously one would expect the hydrophilic clusters to represent surface determinants of the protein, i.e. those which interface with the aqueous environment. In addition to the concentrations of hydrophilic residues, there are two major clusters of hydrophobic amino acids situated between positions 300-320 and 375-390 respectively (Fig. 7.13). Kyte and Doolitle (1982) studied the mean hydropathies of 20 residue segments from 30 water soluble, globular proteins, from which they chose the 20 most hydrophobic segments as representative of the extremes found in globular proteins. These 20 extremes were then compared to various membrane-spanning regions from a number of proteins in an effort to determine hydropathy parameters which could identify membrane-spanning regions within novel sequences. These comparisons led them to conclude that when the mean hydropathy of a 19 residue segment exceeds +1.6, then there is a high probability that it represents a membrane-spanning sequence; using these stricter parameters the XerB protein's hydropathy was re-assessed: regions with extreme mean-hydropathies were superimposed upon a PLOTSTRUCTURE output of the Chou and Fasman predictions for the XerB secondary structure (Fig 14a, 14b). This analysis clearly shows that there are two regions in XerB which exhibit the extreme mean hydrophobicity characteristic of membrane-spanning sequences (Fig. 7.14b); similar analyses revealed no such regions in XerA (156 residues), Tn1000 transposase (1003 residues), Tn3 resolvase (185 residues), or LacZ (1018 residues), whereas analagous hydrophobic regions were identified in the integral membrane protein, LacY (418 residues; data not shown). Examination of the primary structures of these two hydrophobic regions reveals that both contain an

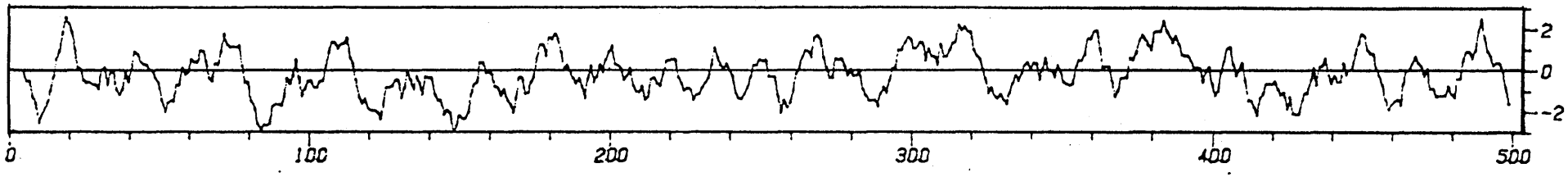


Figure 7.13 Hydropathy profile of XerB: As calculated by PEPPLOT over a window of seven residues.

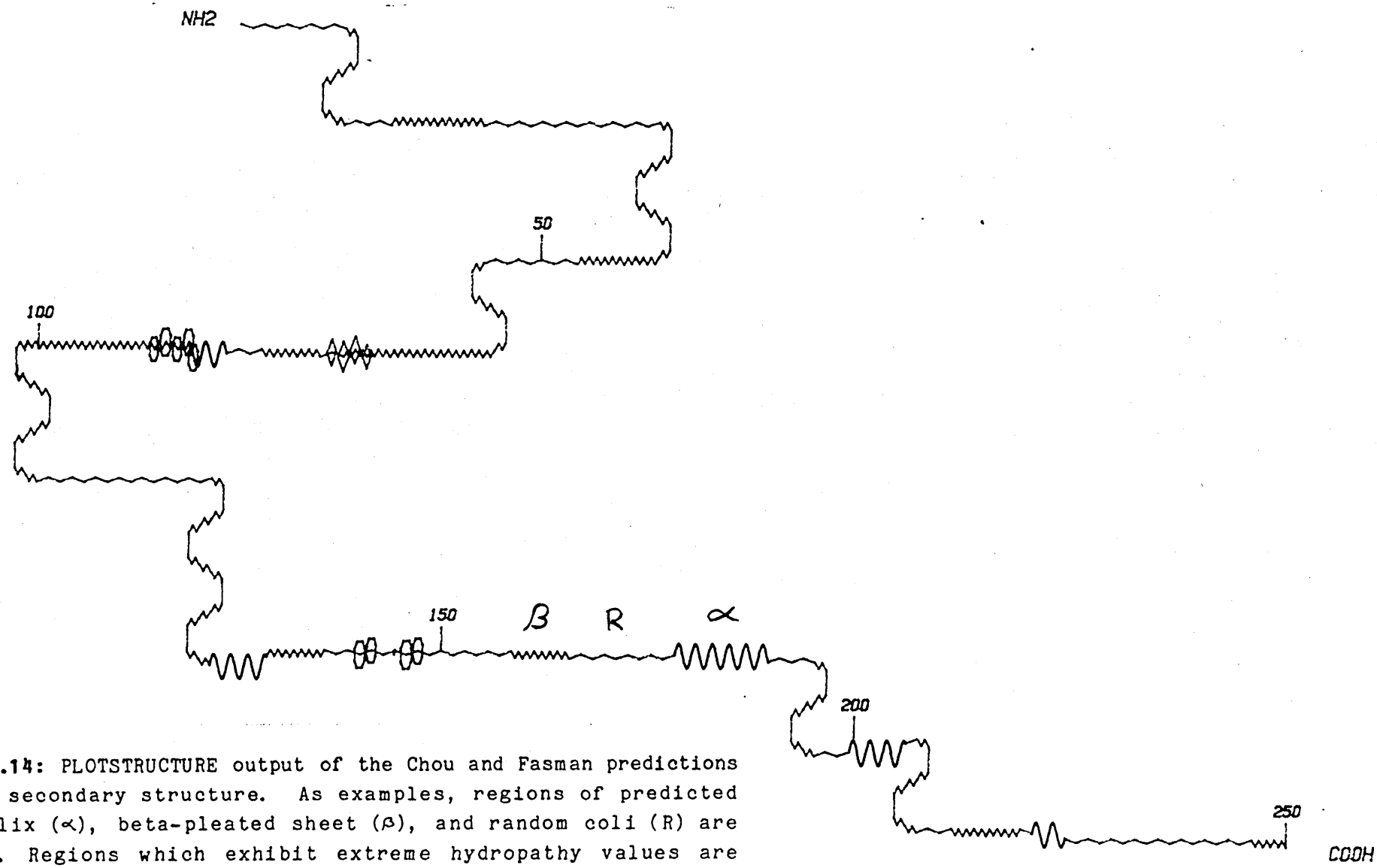


Figure 7.14: PLOTSTRUCTURE output of the Chou and Fasman predictions for XerB secondary structure. As examples, regions of predicted alpha-helix (∝), beta-pleated sheet (β), and random coli (R) are labelled. Regions which exhibit extreme hydropathy values are indicated (i.e. those whise mean hydropathy exceeds +/- 1.6 over a window of 19 residues); ◊ Hydrophobic; () Hydrophilic.
(a) Predictions for residues 1-250; (b) Predictions for residues 250-503.

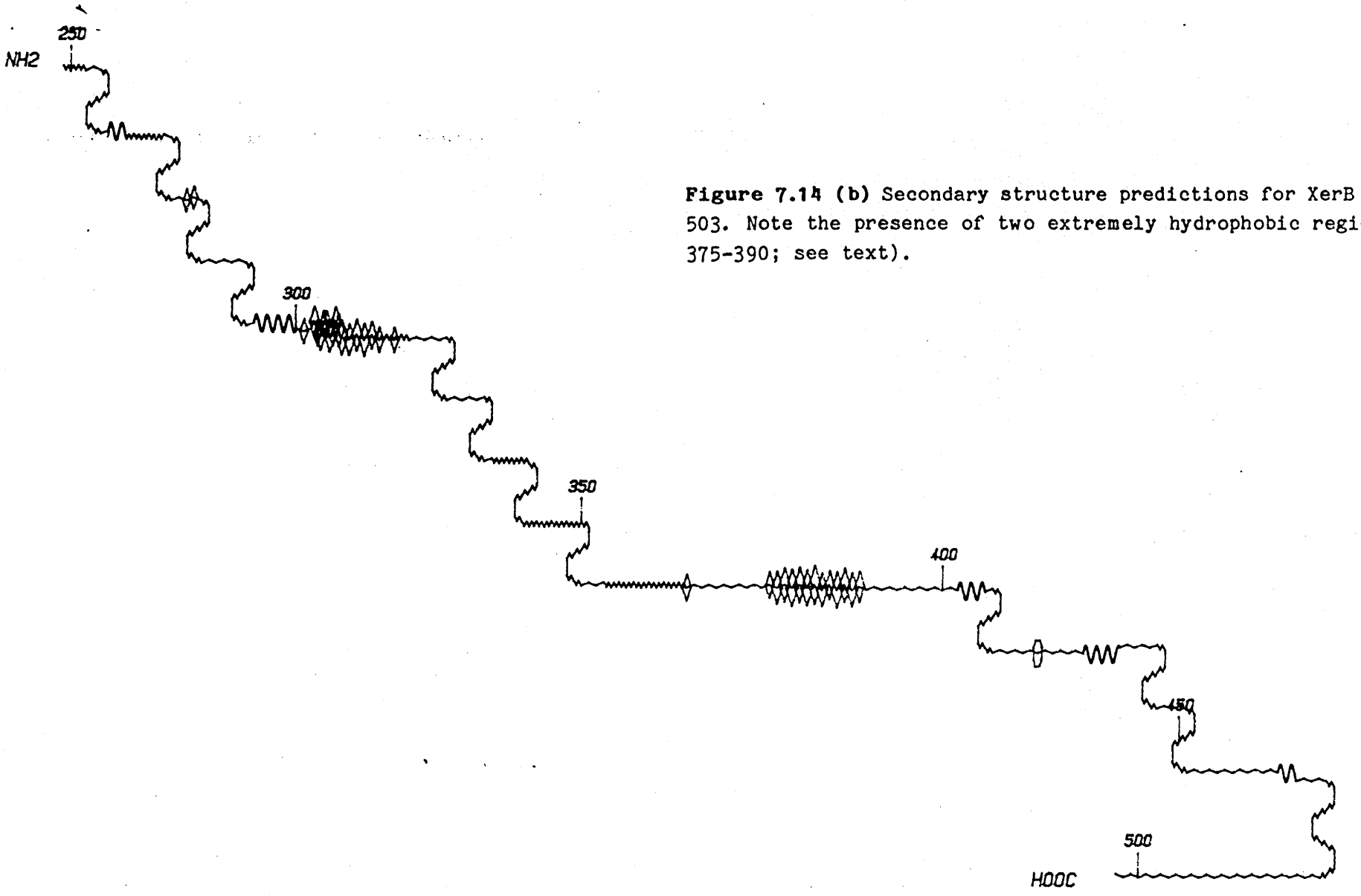


Figure 7.14 (b) Secondary structure predictions for XerB residue
503. Note the presence of two extremely hydrophobic regions (3
375-390; see text).

uninterrupted stretch of hydrophobic amino-acids of at least 11 residues (see Fig. 7.4). In principle, 11 amino-acids in Beta-sheet conformation would be just sufficient to span the average lipid bi-layer (3.6nm), whereas it requires 23 residues in alpha-helical conformation. The secondary-structure predictions for these two hydrophobic regions certainly do not rule out the possibility that both might span a lipid bi-layer in beta sheet conformation.

7.6 CONCLUSIONS IN BRIEF

Having already discussed the results in this chapter in some detail I shall summarise them only very briefly. The sequence analyses suggest that the XerB gene product is a 503 residue polypeptide, which appears likely to be capable of efficient translation due to the gene's very good ribosome binding site, high codon-bias, and the presence of REP sequences which may stabilise the mRNA. It seems quite probable that the *xerB* gene is transcribed as part of an operon, but there is as yet no evidence that the downstream gene, PCS-5, is an *xer* gene.

CHAPTER 8

PURIFICATION OF THE XerB PROTEIN

It rapidly became clear that the XerB protein could be readily purified by virtue of two contributory factors:

1. The IPTG induction of DS941 carrying pCS126 resultes in the overexpression of XerB to a level approaching 5-10% of total cell protein.
2. The XerB protein is insoluble at neutral pH except at high ionic strengths.

8.1.2 OVEREXPRESSION OF XerB

It became clear in the mini-cell analyses described in chapter 7 that the IPTG induction of mini-cells carrying pCS126 resulted in a high level of expression of the XerB protein; this has also been found to be the case with DS941 whole cells. An induction time-course performed on DS941+pCS126 revealed that near maximum levels of XerB were accumulated within two hours of induction, and that only a small increase in total XerB was observed if induction was continued overnight (Fig. 8.1); given the level of XerB accumulated by induced cells it is surprising to note that there is no significant difference in the growth rates of induced *versus* uninduced DS941+pCS126, nor is there any difference between the growth rates of DS941+pCS126 *versus* plasmid-free DS941 (data not shown). The gel shown in figure 8.1 also includes protein extracts from DS941 and from CSX17 (DS941 *XerB*1; Appendix 1); there is, however, no obvious difference between the two protein profiles, suggesting that XerB is not a major cellular protein under standard growth conditions.

8.1.3 PURIFICATION PROTOCOL

The XerB protein was purified from DS941+pCS126 after a two hour induction with IPTG as shown in figure 8.2. The protocol is primitive, and yet extremely effective, and can be summarised as follows; Cells are lysed by sonication in "KPM" buffer (see Fig. 8.2), in which XerB is relatively insoluble. The XerB precipitate sediments, together with the membrane fraction and cellular debris, upon centrifugation at 46,000g, 30 mins.. The resultant pellet is then washed by vigorous resuspension in the same buffer (KPM), in order to remove contaminating KPM-soluble proteins. The suspension is then re-spun, and the pellet resuspended in $Tris_{1000}$ buffer (see Fig. 8.2); this treatment solubilises XerB, and will also strip the membrane fraction of its peripherally associated proteins. After one further spin, the

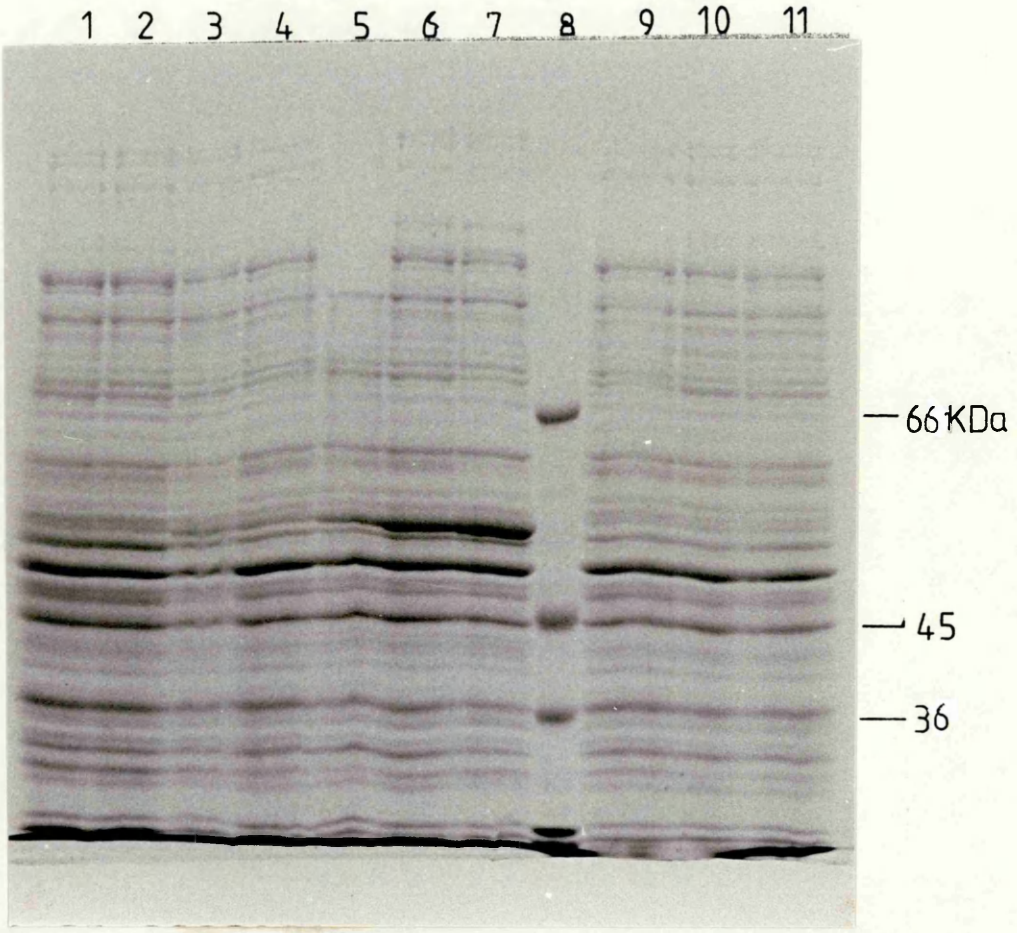


Figure 8.1 Induction time course of pCS126-specified xerB expression. 20ml cultures grown with shaking at 37°C to an OD_{600}= 0.6, the culture was then split into two at T= zero, to one half was added IPTG (50 $ug.ml^{-1}$), then both cultures were returned to 37°C and samples taken after various incubation times. 400ul samples were taken at each time point and their OD_{600} determined, then the sample diluted as appropriate so as to standardise the amount loaded per track. Samples were prepared for electrophoresis as previously described (2.23); then run on an 8% Laemmli gel.

Lane	Strain	IPTG	Incubation (minutes)
1	DS941	-	o/n
2	CSX17	-	o/n
3	DS941+pCS126	-	o/n
4	" "	-	120
5	" "	+	70
6	" "	+	120
7	" "	+	o/n
8	Mol. wt. markers		
9	DS941+pBAD	-	120
10	" "	+	70
11	" "	+	120

o/n signifies overnight incubation.

Figure 8.2 XerB purification protocol: The numbers in brackets refer to the fraction number. (o/n = overnight).

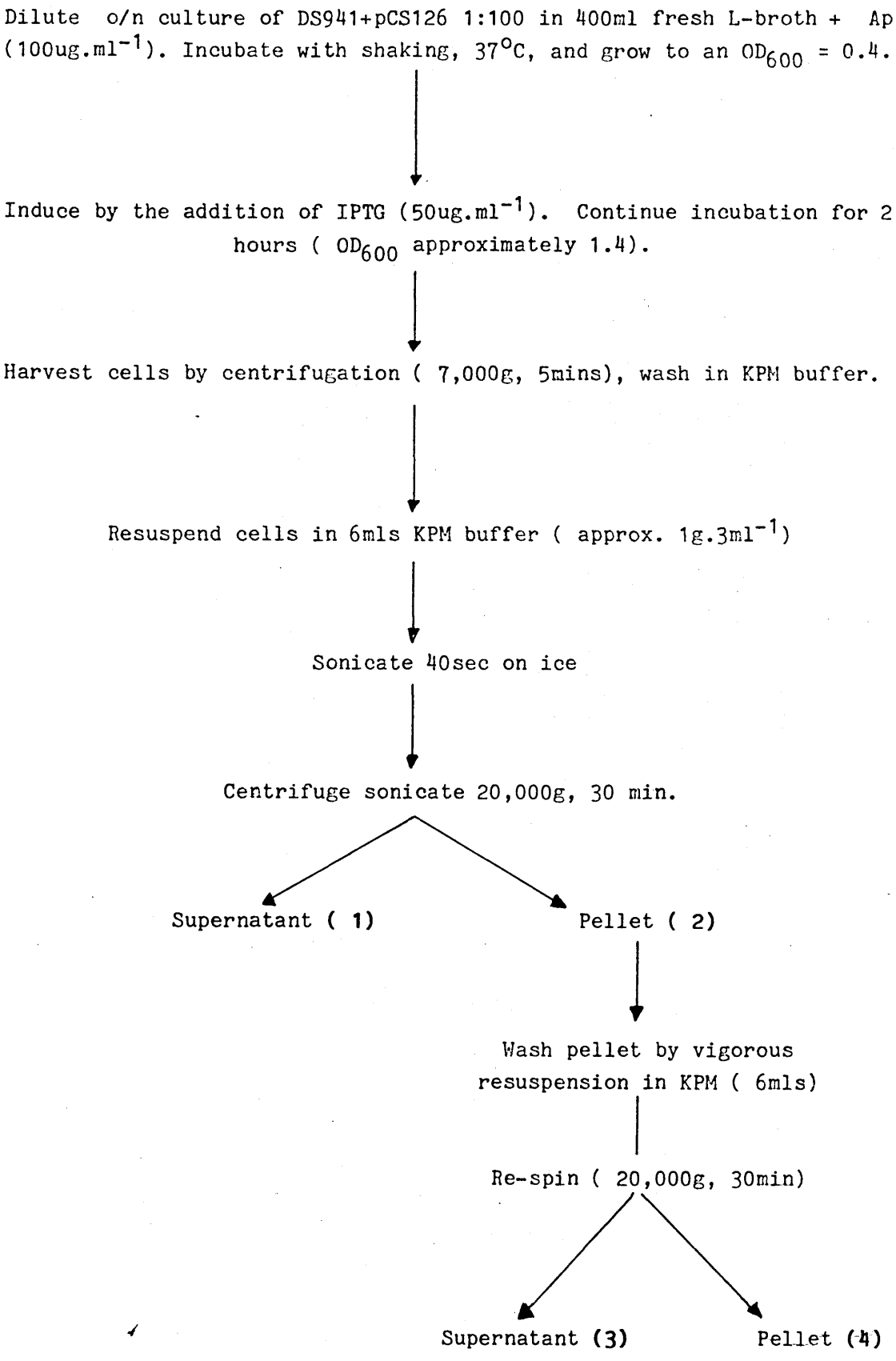
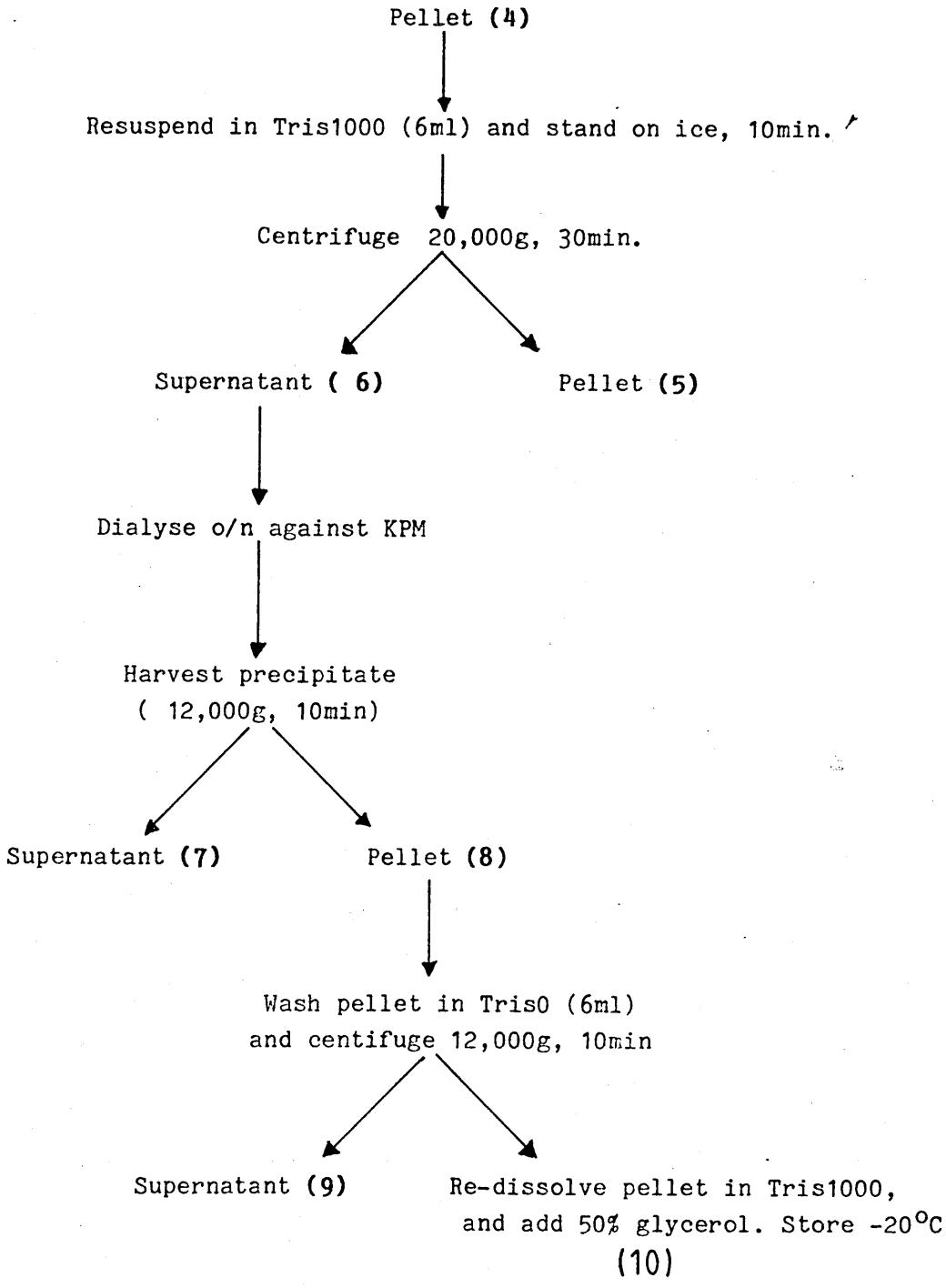


Fig. 3.2 cont.



Buffers:

KPM : 25mM KPi (pH7.0), 5mM $MgCl_2$, 1mM EDTA, 1mM DTT, 1.2mM PMSF, and 1mM Benzamidine

Tris0 : 50mM Tris.HCl (pH8.5), 5mM $MgCl_2$, 1mM EDTA, 1mM DTT, 1.2mM PMSF, and 1mM Benzamidine.

Tris1000 : As Tris0 but with 1M NaCl.

bulk of XerB protein (70-80% of the total) is present in a soluble form in the supernatant, and represents approximately 80-90% of the total protein present within that fraction (as estimated by eye from Coomassie-stained Laemmli gels; Fig. 8.3). To remove the majority of those proteins contaminating the $Tris_{1000}$ soluble fraction, it is dialysed against fresh KPM buffer (overnight, 4°C), this dialysis brings about the precipitation of XerB, after which the precipitate may be harvested by centrifugation at 12,000g, 10 mins., then the pellet washed in KPM buffer and re-spun, whereupon the pellet may be readily solubilised in fresh $Tris_{1000}$ buffer. The resultant material is adjudged to be electrophoretically pure by Coomassie stained gel analysis (Fig. 8.4). Quantitatively, the procedure repeatedly recovers 5-10mg (10-20 nanoMoles) of highly purified XerB from 1.5g (wet weight) of cells; Table 8.1 indicates the total protein present in each fraction as determined by Bradford Assay (Bradford, 1976). The protocol outlined above clearly relies on the physical presence of the 55KDa protein, in SDS-PAGE, as an assay with which to monitor purification; acording to this assay, it would appear as though approximately 10% of the XerB present in whole cells is lost in the initial KPM soluble phase (Fig. 8.3, lane 9), whereas a further 10% remains in the insoluble pellet even after treatment with $Tris_{1000}$ (Fig. 8.3, lanes 10 and 12). The 10% of xerB protein which remains in the membrane/debris fraction could of course be integrally associated with membrane (c.f. the potential membrane spanning sequences identified in chapter 7); however, the properties of the bulk of the XerB present would suggest that the 10% lost within the debris is probaly "stuck" there, either inaccessible to $Tris_{1000}$, or in an irreversibly denatured form. Whilst one might argue that XerB could be peripherally membrane-associated due to its co-fractionation with membranes in KPM buffer, such an association with membrane is not required to explain the observed co-sedimentation since purified XerB precipitates spontaneously in KPM buffer in the (presumptive) absence of significant quantities of lipid.

8.1.4 SOLUBILITY CHARACTERISTICS OF PURIFIED XerB

A simple qualitative analysis was performed to gauge the degree to which XerB is soluble in various buffers. Equal quantities of precipitated XerB (approx. 50ug) were collected in microfuge tubes by centrifugation of a fine suspension of XerB in KPM buffer. To each tube was then added 100 microlitres of one of a number of buffers,

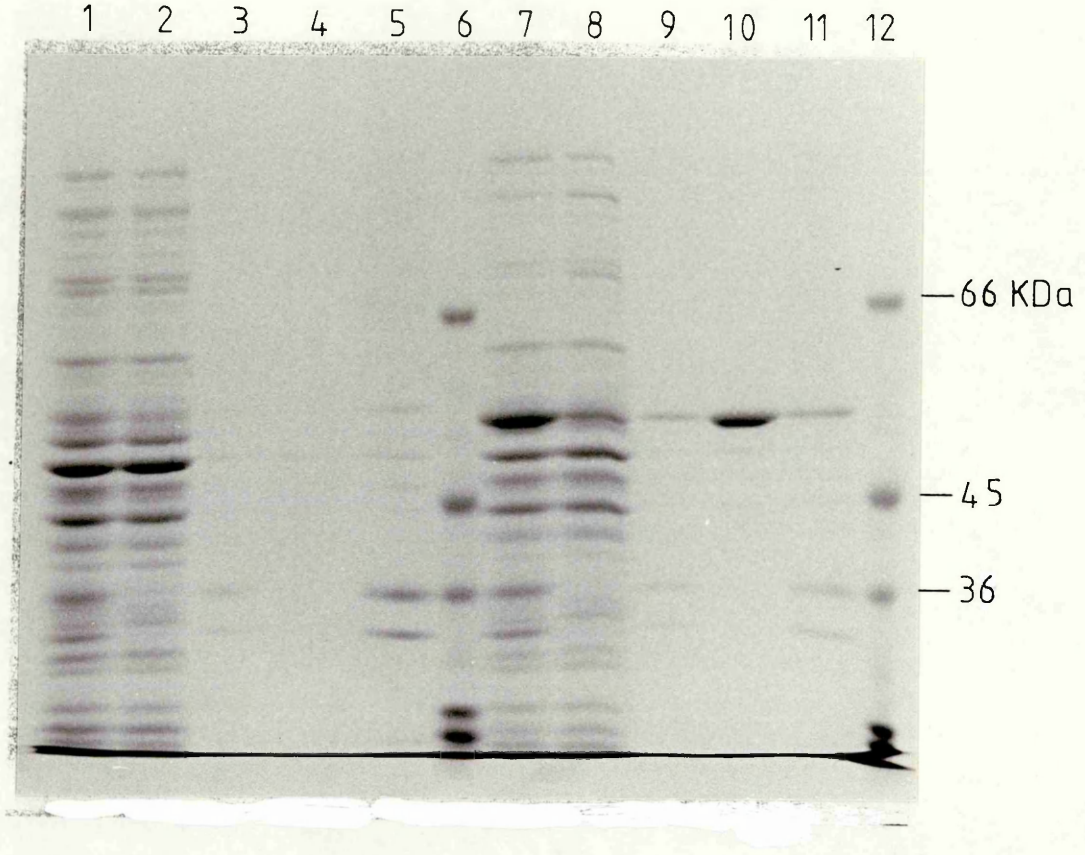


Figure 8.3 Purification of XerB. This gel follows the purification of XerB as described in figure 8.2:

Lane	Strain/source	Fraction (see Fig. 8.2)	Description
1	DS941+pBAD		Whole cells
2	DS941+pBAD	1	KPM soluble phase
3	" "	5	Tris1000 insoluble
4	" "	6	Tris1000 soluble
5	" "	5	Tris1000 insoluble
6	Mwt. Markers		
7	DS941+pCS126		Whole cells
8	" "	1	KPM soluble phase
9	" "	5	Tris1000 insoluble
10	" "	6	Tris1000 soluble
11	" "	5	Tris1000 insoluble
12	Mwt. Markers		

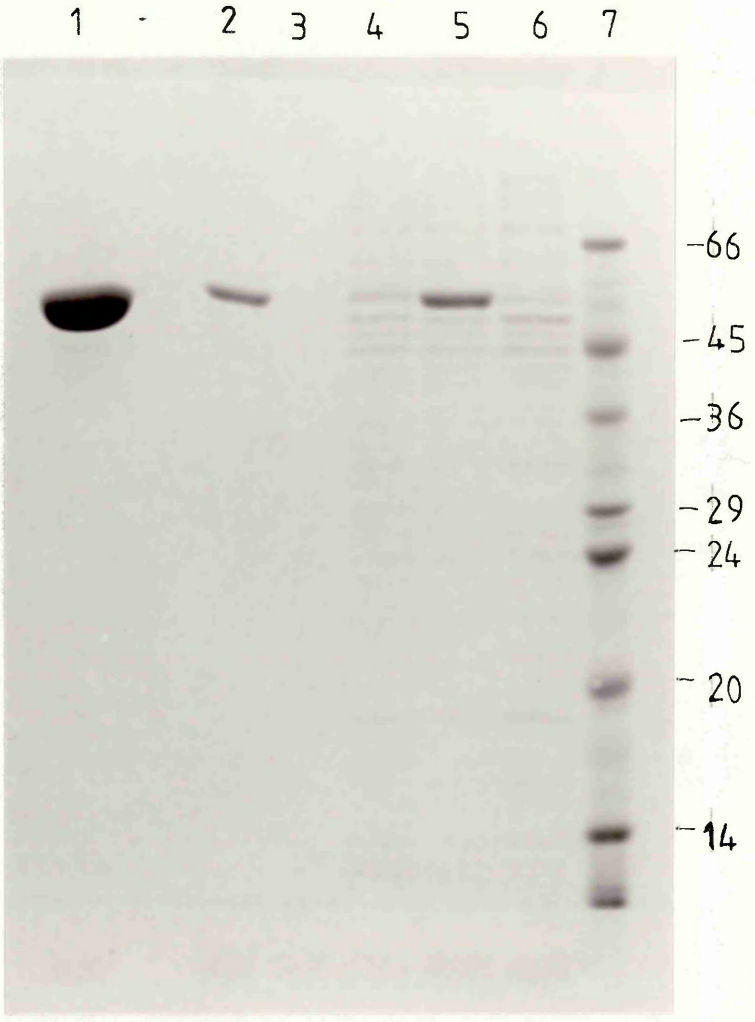


Figure 8.4 Purification of XerB by precipitation in KPM:

Lane	Strain/source	Fraction	Description
1	DS941+pCS126	10	Tris1000 dissolution of KPM pellet (10 ug)
2	DS941+pCS126	10	" " " " (1 ug)
3	" "	9	S/N from Tris0 wash of KPM pellet
4	" "	7	S/N after KPM precipitation
5	" "	6	Tris1000 soluble phase
6	DS941+pBAD	6	Tris1000 soluble phase
7	Mwt. Markers		

S/N= Supernatant

Table 8.1 Quantitation of XerB purification: For fraction number, see purification protocol (Fig. 8.2).

Fraction Number	Description	Vol.(ml)	conc.(mg/ml)	Total Protein (mg)
	Crude sonicate	7.0	40.7	285
1	KPM soluble	5.5	40.0	220
3	KPM wash	6.0	2.3	14.0
6	Tris1000 soluble	6.0	1.3	7.9
7	KPM dialysate	7.3	0.2	1.5
10	Tris1000 soluble	2.0	3.4	6.8

Table 8.2 Qualitative analysis on the solubility of XerB

a)

Buffer	Degree of dissolution with varying [NaCl]mM 0	500	1000	1500
25mM KPi pH 6.0	-	+/-	+/-	+/-
25mM Tris.HCl pH 8.0	-	+++	+++	+++
25mM Gly/NaOH pH 10.0	+	+++	+++	+++

all buffers were made 1mM EDTA, 1mM DTT.

b)

Buffer	Degree of dissolution with varying [NaCl]mM 0	50	100	150	200
25mM Tris.HCl pH8.5 1mM EDTA, 1mM DTT	-	+	++	+++	+++

Code: +++ = pellet completely dissolved

after which the pellet was resuspended by vortexing thoroughly. After 5 mins. on ice, the tubes were re-spun (12,000g, 5 mins.), and the pellet size used to estimate the degree of protein dissolution. The various buffers tested and the solubility of XerB in each are listed in tables 8.2a, and 8.2b. It is clear that XerB is highly soluble in a Tris buffer at pH8.5, in the presence of 150mM NaCl, but is virtually insoluble at pH7.0, or pH6.0.

If whole cells are sonicated in $Tris_{1000}$ buffer, then XerB is recovered entirely within the soluble fraction after centrifugation of the initial sonicate (data not shown). On the other hand, when cells are sonicated in $Tris_{500}$ then the vast majority of XerB is found as a precipitate, despite the fact that the purified protein is highly soluble in this buffer; this discrepancy between the solubilities of XerB in purified form *versus* crude extracts may simply be the result of the presence of many additional factors in the crude sonicates (e.g. a very high total protein concentration etc.).

8.2 THE N-TERMINAL AMINO ACID SEQUENCE OF XerB

XerB protein has been purified to a sufficient degree, and in sufficient quantities to enable its N-terminal sequence to be determined. The amino-acid sequence of the purified protein can furnish two key pieces of information; firstly it can unequivocally identify the purified protein as being the product of the *xerB* gene, and secondly it can identify the actual translational start of the gene. A 5 nanoMole sample of XerB was dialysed free from buffer salts against 1mM DTT, a procedure which results in the precipitation of the protein as expected. The precipitate was washed in the same buffer, after which it was harvested by centrifugation, drained, then vacuum dried overnight. The dried-down sample was then forwarded to, and sequenced by, J. E. Fothergill at the Department of Biochemistry, University of Aberdeen.

The first 20 residues of XerB were analysed, 13 of which could be identified unequivocally, five positions were assigned ambiguously, and two residues were unresolved. The five ambiguous positions were each assigned to one of three groups of residues, His,Met & Pro; Ser & Thr; or Arg & Phe; the ambiguity arising out of the fact that the residues within each group all produce highly similar elution profiles, which could not be distinguished under the conditions employed. When one compares the determined protein sequence with that inferred from the DNA sequence, then one finds that all 13

Figure 8.5 N-terminal amino acid sequence of XerB: derived vs. inferred.

```
                Pro                          Pro
                His   Arg                ThrHis        Arg
Derived Seq.  : MetGluPheSerValLysSerGlySerMetGluLysGlnPhe ? Ala ? IleValVal

Inferred Seq. : MetGluPheSerValLysSerGlySerProGluLysGlnArgSerAlaCysIleValVal

                1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

unequivocally identified residues have precise correspondents in the inferred amino-acid sequence (Fig. 8.5). Furthermore, at the five "ambiguous" positions the group of two or three unresolved alternatives always includes the residue predicted by the DNA sequence (e.g. residue 9 is identified as being either Ser or Thr, in turn the DNA sequence predicts a serine). The amino acid residues at positions 15 and 17 could not be identified chromatographically. The DNA sequence predicts a Ser at position 15, and a Cys at 17, both Ser and Cys being residues which are notoriously difficult to identify chemically; however I should add that the elution profiles obtained for residues 15 and 17 are not inconsistent with their being Ser and Cys respectively.

8.3 DNA BINDING ACTIVITY OF PURIFIED XerB

The natural extension of this work is to ascertain whether XerA and/or XerB play direct or indirect roles in promoting recombination between *cer* sites. Ideally, this goal would be pursued *via* the development of an *in vitro* recombination assay, however, due to the limitations of time I have not attempted to develop such a system. An alternative approach to ascribing some role to the two Xer proteins is to determine whether or not either, or both, exhibit a DNA binding activity, particularly one which is specific to the *cer* sequence. A very simple, yet elegant, technique for the detection of DNA:protein interactions is that of "Gel Retardation", i.e. the observed reduction in the electrophoretic mobility of a small DNA fragment, as a consequence of its association with protein molecule(s). This approach has been successfully employed in a number of systems including CAP, Fis, Lac Repressor, & Tn3 Resolvase (Fried & Crothers, 1981; Koch & Kahmann, 1987; Kramer *et al*, 1987; M. Boocock, pers. comm.).

In the short time available to me, I concentrated on gel retardation experiments using XerB. Purified XerB was found to retard an end-labelled 431bp fragment carrying the complete *cer* site, however it apparently does so non-specifically since the retardation is eliminated by the addition of small quantities of sonicated Calf Thymus DNA (C.T. DNA) as a non-specific competitor (Fig. 8.6). Furthermore, similar retardation profiles are observed when XerB is incubated with various end-labelled "non-cer" fragments (from pUC8; data not shown). These results would suggest that XerB may be capable of binding DNA non-specifically (under the conditions tested), but do not indicate whether or not XerB is a true DNA binding protein, or

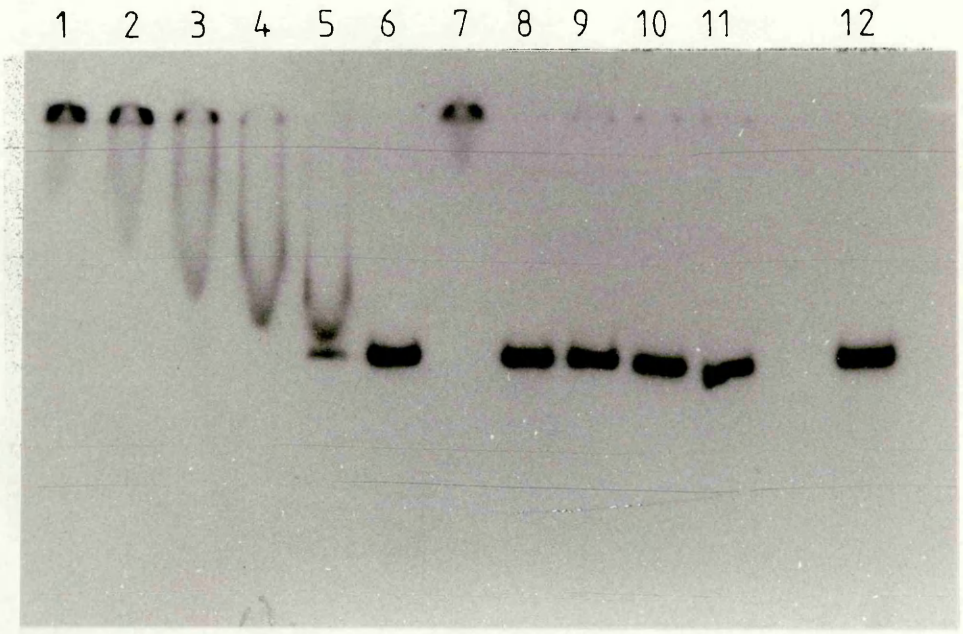


Figure 8.6 Non-specific binding of XerB to a 431bp cer$^+$ fragment. The 431bp EcoRI/PvuII cer$^+$ fragment from pKS492 was gel-purified and then end-labelled with alpha-^{32}P dATP using Klenow. All binding reactions were performed in 20ul volumes in the following reaction buffer 25mM Tris pH8.5, 50mM NaCl, 10% glycerol, 1ng end-labelled fragment, 500ug/ml BSA. Reactions were initiated by the addition of protein, after which they were mixed then incubated at 37^{o}C for 10mins. Samples were then electrophoresed on a 5% polyacrylamide gel as described by Bushman et al (1984). Lanes 1-6 represent varying amounts of XerB protein in the absence of any carrier DNA: 1ug, lane 1; 500ng, lane 2; 250ng, lane 3; 125ng, lane 4; 63ng, lane 5; No XerB added, lane 6. The samples loaded in lanes 7-11 were all incubated in the presence of 1ug of XerB protein, but with varying amounts of sonicated calf-thymus DNA: 0, lane 7; 50ng, lane 8; 100ng, lane 9; 150ng, lane 10; 200ng, lane 11. The sample in lane 12 was incubated with 0 XerB, 200ng calf thymus DNA.

simply a protein which "sticks" to large poly-anions. Furthermore, I should emphasise the relatively large amounts of XerB required to effect the observed retardation; it requires 63ng of protein to partially retard 1ng of end-labelled DNA fragment (equivalent to a >300-fold molar excess of XerB:DNA). This might indicate that XerB binding is relatively weak, or that only a small proportion of the protein preparation is responsible for the observed retardation; if the latter is true then this could either signify the presence of a minor contaminating protein species which effects retardation, or that only a small fraction of the XerB protein has been recovered in an active state. Finally, one should note that in most cases where significant gel retardation of a Protein:DNA complex has been observed, it is almost certainly due to a conformational change (bend) in the DNA fragment as a consequence of protein binding, and apparently not a simple function of increased molecular weight. Therefore, in cases where protein binds DNA but does not induce a significant bend, then one might not expect to observe gel retardation. If this were the case with XerB, then specific DNA binding may occur but go undetected, the observed retardation being the result of non-specific protein:DNA aggregation at very high protein concentrations (in this regard note that whilst at low protein concentration Tn3 Resolvase binds specifically to the *res* sequence, at high concentration it exhibits a relatively non-specific binding activity which results in protein:DNA aggregates which fail to enter the gel (M. Boocock, pers. comm.)). It would therefore be prudent to examine protein binding *via* an alternative route which does not require a conformational change, e.g. DNAaseI footprinting, or co-immunoprecipitation of Protein:DNA complexes.

8.4 RAISING REAGENT ANTI-BODIES TO PURIFIED XerB

Since the XerB purification protocol involves the use of an over-producing strain from which the protein is recovered in an insoluble state, and in the absence of an assay for XerB activity, then it would seem wise to question the active status of the purified protein; for example the Hin recombinase may be purified in an active form when over-expressed to a level of 0.1% of total cell protein, however increased levels of over-expression lead to less active extracts (Johnson *et al*, 1986). Furthermore, *cer*-specific recombination requires **at least** one other protein (XerA), and perhaps many others together with unknown co-factors, in the absence of which XerB might be

rendered impotent; such problems might be surmounted by the (initial) analysis of Xer activity in crude extracts from wild-type cells. The study of XerB at wild-type levels could be greatly assisted by the availability of anti-XerB anti-serum; to this end two New Zealand white rabbits have been immunised with purified XerB as described in figure 8.7.

On the assumption that high avidity anti-XerB antibodies are (ultimately) raised, then they may be used in a number of very interesting, and potentially highly informative experiments e.g.:

1. Determination of the cellular location of XerB in wild-type cells, i.e. does it segregate with the soluble or membrane fractions.

2. The assay of XerB protein levels in each XerB$^-$ mutant (with the two spontaneous mutants, DSX1 and DSX2, being of particular interest). The levels of XerB expressed in XerA$^-$ mutants is also of key interest.

3. One could assay XerB binding to the cer sequence in wild-type extracts by immunoprecipitation of XerB in the presence of end-labelled DNA; co-immunoprecipitation of nucleic acid with a specific binding protein (by, of course, anti-protein anti-serum) has been observed in a number of systems including the replication initiator protein of R6K (Bastia & Patel, in press). This approach to XerB:DNA "binding" has the advantage that it does not necessarily require that XerB binds DNA directly, since one might reasonably expect to detect any stable association between XerB and DNA mediated by other proteins.

4. In an extension of the preceeding experiment, one might seek to identify any proteins with which XerB forms stable complexes, by a similar co-immunoprecipitation assay; either from crude extracts of proteins labelled in vivo, or from unlabelled extracts whereupon precipitated proteins might be visualised after I^{125} iodination.

5. One could also employ anti-XerB anti-serum to conduct a survey of other enterobacteriacea in order to determine the ubiquity of XerB-like polypeptides.

Figure 8.7 Immunisation Schedule:

18/7/87 : Prior to immunisation, collect 20ml whole blood from each rabbit and recover serum after clotting (37^{o}C, 1hour; then 4^{o}C o/n), and centrifugation (3,000g, 10min). Store serum frozen -20^{o}C.

Day 1 (28/7/87) : Immunise two New Zealand white rabbits with 250ug each of purified XerB in 1ml Freund's complete adjuvant. Inject 0.5ml of suspension (125ug XerB) intramuscularly into each hind limb.

Week 5 (1/9/87) : Boost both rabbits with 200ug XerB in Freund's incomlete adjuvant (1ml); again given in two intramuscular injections

Week 7 : Collect 40ml of whole blood from each rabbit and recover and store serum as described above.

A protocol has been established which enables the rapid preparation of substantial quantities of electrophoretically pure XerB. The purified material has an apparent molecular weight equal to 55KDa (as measured by SDS-PAGE), and thus would appear to represent the full-length translational product of the XerB gene. The N-terminal peptide sequence of the first 20 residues of this protein has been determined, the derived sequence confirms that:

1. The protein preparation is homogeneous.
2. The purified protein is the product of the putative XerB gene identified by DNA sequence analysis.
3. The XerB translational start is indeed at the AUG codon identified by ZPERTRON as the most probable initiation codon (60bp; chapter 7).
4. The purified protein has undergone no N-terminal proteolytic cleavage events (e.g. Signal Peptide cleavage); however, given the extent of XerB over-expression, one cannot preclude the possibility that the 55KDa species represents an accumulated un-processed form. Any significant post-translational modifications might be better identified in wild-type extracts, using specific anti-serum to detect the XerB protein.

There is to-date no functional assay for the purified XerB protein; whilst it appears capable of binding DNA there is no evidence of any cer-sequence specific DNA binding. Thus the biochemical activity if XerB remains elusive. It is hoped that rabbit anti-serum raised against the purified protein will react, with high avidity, to the native XerB protein in crude extracts from wild-type cells; If so, then some of the experiments outlined in section 8.4 could be undertaken, the results of which might provide some long-awaited clues as to the expression and function(s) of XerB. I have since recovered anti-sera which immunoprecipitate purified XerB in a standard gel-immunoprecipitation assay, indicating that anti-XerB antibodies have been raised (data not shown).

CHAPTER 9

CONCLUDING REMARKS

independent *xer* mutants, then to use these mutants to clone and characterise the wild-type *xer* genes by genetic complementation. In the course of this work, I have isolated a total of nine independent *xer* mutants, bringing the total number of mutants in our collection to 11. Of the nine novel mutants isolated, six of the *xer* lesions are genetically linked to an inserted copy of the mutagen Tn5, the remaining three mutants have yet to be tested for such linkage.

The 11 mutants have been found to fall into two complementation groups, designated *xerA* and *xerB*, with two mutants belonging to the *xerA* group, and the remaining nine comprising the *xerB* group. The *xerA* gene was cloned, in collaboration with Gill Stewart, by taking advantage of information derived from the *xerA*3::Tn5 mutant allele cloned in pCS300. On the other hand, *xerB* was cloned by screening directly for complementation using the highly sensitive pCS210 colony-colour test for Xer phenotype.

The *xerA* Gene and its Product

The *xerA* complementing function has been de-lineated to within a 920bp *Sph*I/*Acc*I fragment. The DNA sequence of this fragment has been determined, and the *xerA* gene identified as a 156 codon protein-coding gene which encodes a polypeptide of predicted molecular weight equal to 17.2 KDa. The polypeptide product of this gene has been observed in minicell expression studies and has an apparent molecular weight of 16.5KDa. The *xerA* locus maps to 70.5mins. on the K12 linkage map, as does *argR* (arginine repressor; Bachmann, 1983). The *argR* sequence has not been published to date, however since writing this thesis I have received a copy of a pre-print from Lim *et al*, in which they present the sequence of the *argR* gene. This has revealed that the *xerA* and *argR* genes are in fact one and the same. Lim *et al* report the sequencing of an 880bp *Sph*I/*Acc*I fragment (compared to the 920bp *Sph*I/*Acc*I restriction fragment cloned in mCS320), and they present only 818bp of this sequence running from the *Sph*I site towards the *Acc*I site. This 818bp of sequence is identical to the appropriate portion of the XERA112 sequence with the exception of 5 additional base-pairs in the XERA112 sequence at co-ordinates 766, 781, 803, 805, and 808 respectively; each position has been triple-checked and the extra base confirmed on both strands of the XERA112 sequence. There

are no sequence differences within the *xerA* structural gene. Lim *et al* have unequivocally identified the product of the 156 codon ORF as being the arginine repressor. ArgR (plus the co-repressor, L-arginine), negatively regulates a number of genes in the arginine regulon by binding to "Operator" sites which usually overlap with the repressed gene's promoter. In all known cases, the Arg operator comprises two loosely conserved copies of an 18bp "ArgR box" (for review see Glansdorf, 1987). There are two such ArgR boxes upstream of *xerA/argR*, which overlap with the promoter designated P2 (chapter 6), and to which ArgR (+ arginine) has been shown to bind (Lim *et al*, in press). Intriguingly, there is also a very close match to an ArgR box within *cer*; this sequence aligns very well with the consensus ArgR box with 15/18 matches, this compares with the two ArgR binding sites identified upstream of *xerA/argR* which exhibit 15/18 and 14/18 matches respectively:

```
                                  -35
                                 ____
                          3795   3800      3812
Cer (3795-3812)  :          GTTGCATAGGTATTCATA
                            ||||||| ||||||||
ArgR-Box consensus :        aNTGNaTN[aa/tt]tATTCAT[a/t]
```

Note that the ArgR box within *cer* overlaps with the -35 region of the promoter P1, this might indicate that transcriptional initiation at P1 is regulated by ArgR. It is most interesting to recall the data on *cer* deletion derivatives (chapter 1), in which the sequence from ColE1 co-ordinates 3793-3966bp maintained the ability to recombine with a wild-type copy of *cer*, but the sequence from 3800-3966bp did not; note that a major portion of the ArgR box has been deleted in the latter case. Obviously determining whether or not ArgR/XerA binds to this site within *cer* would be an immediate priority in any future work on this project. If indeed it does, then its affect upon recombination might be mediated through regulation of P1 transcription, and/or might be a function of binding *per se*. It seems rather unlikely that ArgR is the *cer*-specific recombinase, however its binding at this site within *cer* may be involved in the assembly of a higher order structure necessary to promote efficient strand exchange in the cross-over

region some 100bp downstream. With this in mind it is interesting to note that, in its native state, ArgR is found as an oligomer. This oligomer usually binds two adjacent copies of the ArgR box present at Arg operators. An interesting speculation is that an ArgR oligomer may bind the single ArgR box within *cer* and then, in a plasmid dimer, might bring the two *cer* sites into a synaptic complex by association with the ArgR box in the second *cer* site.

It would appear that a novel role in *cer*-specific recombination has been uncovered for the *E. coli* ArgR transcriptional repressor protein. This is particularly interesting since the same "duality" of function has been noted for IHF which regulates the *val*, *ilv*, and *xyl* genes etc., but was first discovered as an accessory factor in lambda Int-specific recombination. This might indicate that episomal elements such as ColE1 and phage Lambda, have evolved to exploit the properties of pre-existing host DNA-binding proteins as accessory elements in their site-specific recombination systems.

The *xerB* Gene and its Product

A comprehensive deletion analysis has unequivocally identified the *xerB* gene as a 1509bp sequence which encodes a 503 residue polypeptide with a predicted molecular weight of 55.3KDa; the minimum functional *xerB* clone being a 1687bp *Hind*III/*Acc*I fragment in pCS130. DNA sequence analyses have identified the start of a second gene, PCS-5, 160bp downstream of the *xerB* stop codon; the proximity of the two genes, coupled with the presence of two REP sequences in the intergenic region provides the signature of an operon. Whether or not PCS-5 encodes an Xer-related function is a question which ought to be addressed in any extension to this work.

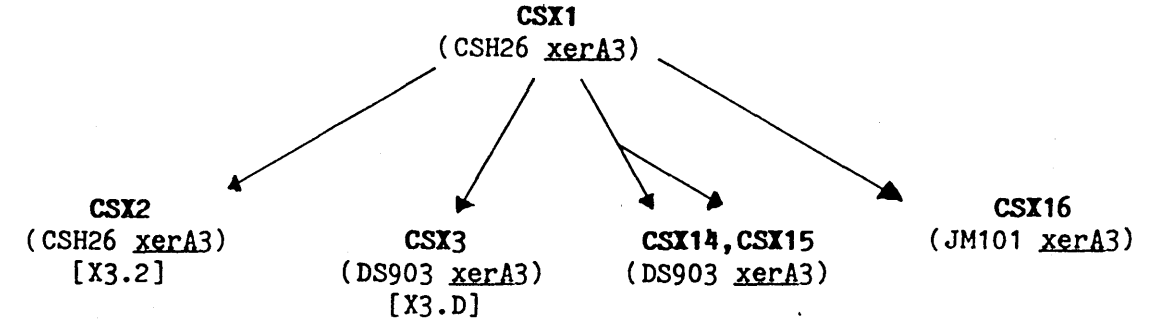
A protocol has been established for the purification of XerB to homogeneity. The N-terminal sequence of the purified material has confirmed its identity, and in turn has identified the translational start of the *xerB* gene. There are, as yet, no clues as to the role of XerB in *cer*-specific recombination. Attempts to demonstrate *cer*-sequence specific DNA binding using a gel-retardation assay have proved unsuccessful; however, under the conditions tested, the purified protein does appear to exhibit a non-specific DNA binding activity. In order to study XerB at wild-type levels the purified protein has been used to raise anti-XerB antisera which provide a

antisera may be used to investigate a number of important features of XerB including its expression, processing, and cellular location(s). High avidity antiserum might also prove a powerful tool in the investigation of any intermolecular complexes with which XerB associates (e.g. protein:DNA, or protein:protein etc.).

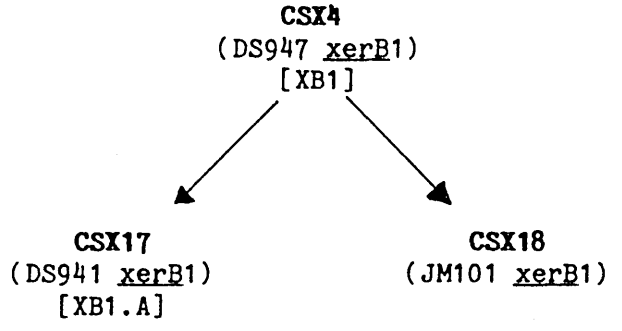
This work has identified and characterised two host encoded functions which are essential to *cer*-specific recombination; one of these functions has been identified as the transcriptional repressor ArgR, thus exposing an entirely novel role for this extensively studied protein. Assuming a direct involvement for ArgR/XerA and/or XerB in the recombination process, then their identification may greatly assist in the development of an *in vitro cer* recombination system. Such a system could, in turn, elucidate the precise role(s) that each protein performs, and thus enhance our understanding of an apparently novel site-specific recombination mechanism.

Strain constructions: An arrow between strains indicates the P1 co-transduction of *xer*$^-$ with Km^r. The transductanct strain name is given, followed by the parental strain plus mutant allele designation in brackets (). Finally the original lab-name, where appropriate is given in square brackets [].

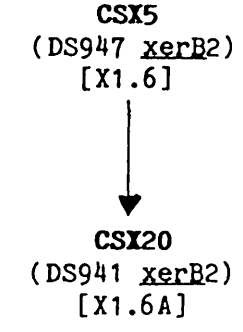
1. xerA3



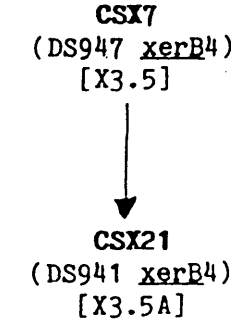
2. xerB1



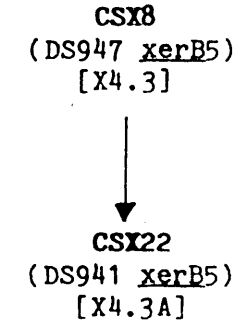
3. xerB2



4. xerB4



5. xerB5



BIBLIOGRAPHY

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

Abraham, J.M., Freitag, C.S., Clements, J.R., and Eisentstein, B.I. (1985). Proc. Natl. Acad. Sci. U.S.A. **82**, 5724-5727

Abremski, K., Hoess, R. and Sternberg N. (1983). Studies on the properties of the 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.

Alff-Steinberger, C. (1984). Evidence for a coding pattern on the non-coding strand of the E. coli genome. Nucl. Acids Res. **12**, 2235-2241.

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

Archer, J.A.C., (1985). Structure and function of the plasmid ColK. Ph.D. Thesis. University of Glasgow.

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

Arthur, A. and Sherratt D.J. (1979). Dissection of the transposition process: a transposon-encoded site-specific recombination system. Mol. Gen. Genet. **175**, 267-274.

Auerswald, E. A., Ludwig, G., and Schaller, H. (1981). Structural analysis of Tn5. Cold Spring Harbor Symp. Quant. Biol. **45**, 107-113.

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

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

Bachman, B.J. (1983). Linkage map of E. coli K-12. Microbiol. Rev. **47**, 180-230.

Barik, s., Ghosh, B., Whalen, W., Lazinski, D., and Das, A. (1986). Site-specific entry and assembly of a regulatoru component into the elongating transcription apparatus

detected by immunoprint. In. Molecular Genetics of Bacteria and Phages, p159. Cold Spring Harbor Laboratory.

Berg, D.E. (1977). Insertion and excision of the transposable kanamycin resistance determinant Tn5, in: DNA Insertion Elements, Plasmids and Episomes. Cold Spring Harbor Laboratory 1977: Bukhari, A.I., Shapiro, J.A. and Adhaya, S.L.(Eds). pp205-212.

Berg, D.E. and Berg, C.M. (1983). The prokaryotic transposable element Tn5. Biotech. **1**, 417-435.

Berg, J. (1986). Potential metal-binding domains in nucleic acid binding proteins. Science **232**, 485-487.

Bernard, O., Hozumi, N. and Tonegawa, S. (1978). Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell **15**, 1133-1144.

Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**, 1515-1523.

Boocock, M., Brown, J. L., and Sherratt, D. J. (1987). Topological specificity in Tn3 resolvase catalysis. In: ICN-UCLA Symp.: DNA replication and recombination 703-718.

Boyd, A. C. (1982). Ph.D. Thesis.

Boyd, A. C., and Sherratt, D. J. (1986). Polar mobilisation of the Escherichia coli chromosome by the ColE1 transfer origin. Mol. Gen. Genet. **203**, 496-504.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of ug quantities of protein utilising the principle of protein-dye binding. Anal. Biochem. **72**, 248-254.

Brendel, V., and Trifonov, E. N., (1984). A computer algorithm for testing potntial prokaryotic terminators. Nucl. Acids Res. **12**, 4411-4427.

Brinton, C. C., Gemski, P., Falkow, S., and Baron, L. S. (1961). Location of the piliation factor on the chromosome of Escherichia coli. Biochem. Biophys. Res. Commun. **5**, 293-298.

Broda, P. (1979). Plasmids. Oxford and San Francisco. W.H. Freeman.

Brosius et al, (1984). Expresion vector pKK223-3. Analects **12**, 1-2. Published by Pharmacia P-L Biochemicals.

Broyles, S.S. and Pettijohn, D.E. (1986). Interaction of the Escherichia coli Hu protein with DNA: evidence for formation of nucleosome-like structures with altered DNA helical pitch. J. Mol. Biol. **187**, 47-60.

Bujard, H. (1980). The interaction of E. coli RNA polymerase with promoters. T.I.B.S. **5**, 274-278.

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

Cesareni, G. (1982). The RNA primer promoter, as defined *in vitro*, is essential for pMB1 plasmid replication *in vivo*. J. Mol. Biol. **160**, 123-126.

Cesareni, G., and Banner, D. W. (1985). Regulation of plasmid copy number by complementary RNAs. T.I.B.S. **10**, 303-306.

Cesareni, G., Cornelissen, M., Lacatena, R. M., and Castagnoli, L. (1984). Control of pMB1 replication: initiation of primer formation by Rop requires RNA1. EMBO J. **3**, 1365-1369.

Cesareni, G., Meusing, M.A. and Polisky, B. (1982). Control of ColE1 DNA replication: the *rop* gene product negatively affects transcription from the replication primer promoter. Proc. Natl. Acad. Sci. (USA) **79**, 6313-6317.

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

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

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

Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. Biochemistry **13**, 222-238.

Clark, S.P., Yoshikai, Y., Taylor, S., Siu, G., Hood, L. and Mak, T.W. (1984). Identification of a diversity segment of human T-cell receptor B-chain and comparison with the analogous murine element. Nature **311**, 387-389.

Craig, N.L. (1985). Site-specific inversion: enhancers, recombination proteins and mechanisms. Cell **41**, 649-650.

Craig, N. and Nash, H. (1983). The mechanism of phage site-specific recombination: site-specific breakage of the DNA by Int topoisomerase. Cell **35**, 795-803.

Craig, N.L. and Nash,H.A. (1984). *E.coli* integration host factor binds to specific sites in the DNA. Cell **39**, 707-716.

Craigen, W. J., and Caskey, C. T. (1986). Expression of peptide chain release factor 2 requires high-efficiency frameshift. Nature **322**, 273-275.

Craigie, R., Arndt-Jovin, D. J., and Mizuuchi, K. (1985). A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: Protein and DNA substrate requirements. Proc. Natl. Acad. Sci. U.S.A. **82**, 7570-7574.

Craigie, R., and Mizuuchi, K. (1986). Role of DNA topology in Mu transposition: Mechanism of sensing the relative orientation of two DNA segments. Cell **45**, 793-800.

Daniels, D. M., Schroeder, J. L., Szybalski, W., Sanger, F., Blattner, F. R., Coulsen, A. R., Hong, G. F., Hill, D. F., and Pearson, G. B. (1983). A molecular map of coliphage lambda, complete annotated lambda sequence. In: Lambda II, Hendrix, Roberts, Stahl, and Weisberg (eds.). Cold Spring Harbor Monograph.

Das, A., and Wolska, K. (1984). Transcription anti-termination *in vitro* by Lambda N Gene product: Requirement for phage *nut* site and the products of host *nusA*, *nusB*, and *nusE* genes. Cell **38**, 165-173.

Davidson, N., and Szybalski, W. (1971). In: The Bacteriophage Lambda. 113-138. Hershey, A. D., (ed.). Cold Spring Harbor Laboratory, Cold Spring Harbor.

Dayhoff, M. O., Schwartz, R. M., and Orcott, B. L. (1978). In: Atlas of protein sequence and structure. Dayhoff, M. O. (ed.). National Biomedical Reseach Foundation Washington, D.C. **5**, 345-352.

De Boer, H.A., Comstock, L.J. and Vasser, M. (1983). The *tac* promoter: A hybrid derived from the *trp* and *lac* promoters. Proc. Natl. Acad. Sci. (USA) **80**, 21-25.

De Boer, H. A., and Kastelein, R. A. (1986). Biased codon usage: An exploration of its role in optimization of translation. In: Maximising gene expression, 225-285. Reznikoff, W., and Gold, L. (eds.). Butterworths publ.

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

De Bruijn, F.J., Stroke, I.L., Marvel, D.J., and Ausubel, F.M. (1983). Construction of a correlated physical and genetic map of the *Klebsiella pneumoniae hisDGO* region using transposon Tn5 mutagenesis. EMBO J. **2**, 1831-1838.

Della Latta, P., Bouanchaud, D. and Novick, R.P. (1978). Partition kinetics and thermosensitive replication of pT169, a naturally occurring multicopy tetracycline resistance plasmid of *Staphylococcus aureus*. Plasmid **1**, 366-375.

Dempsey, W.B., (1987). Integration host factor and conjugative transfer of the antibiotic resistance plasmid R100. J. Bacteriol. **169**, 4391-4392.

Devereux, J., Haeberli, P., and Smithies, D. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12**, 387-395.

Diderichsen, B. (1980). *flu*, a metastable gene controlling surface properties of *Escherichia coli*. J. Bacteriol. **141**, 858-867.

Diver, W.P., Grinsted, J., Frtzinger, D.C., Brown, N.L., Altenbuchner, J., Rogowsky, P., and Scmitt, R. (1983). DNA sequences of and complementation by the *tnpR* genes of Tn21, Tn501, and Tn1721. Mol. Gen. Genet. **191**, 189-193.

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

Dodd, I.B., and Egan, J.B. (1987). Systematic method for the detection of potential lambda Cro-like DNA-binding regions in proteins. J. Mol. Biol. **194**, 557-564.

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

Dougan, G., Saul, M., Twigg, A., Gill, R. and Sherratt, D. (1979). Polypeptides expressed in *Escherichia coli* K-12 minicells by transposition elements Tn1 and Tn3. J. Bacteriol. **138**, 48-54.

Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. J. Mol. Biol. **186**, 1-18.

Dugglesby, R.G., Kinns, H., and Rood, J.H. (1981). A computer program for determining the size of DNA restriction fragments. Anal. Biochem. **10**, 49-55.

Durcacz, B.W., Kennedy, C.K., and Sherratt, D.J. (1974). Plasmid replication and the induced synthesis of colicins E1 and E2 in *Escherichia coli*. J. Bacteriol. **117**, 940-946.

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

Egner, C., and Berg, D.E. (1981). Excision of transposon Tn5 is dependent on the inverted repeats but not on the transposase function of Tn5. Proc. Natl. Acad. Sci. U.S.A. **78**, 459-463.

Eisenstein, B.I. (1981). Phase variation of type I fimbriae in *Escherichia coli* is under transcriptional control. Science **214**, 337-339.

Feiss,M., and Becker, A. (1983). DNA packaging and cutting. In: Lambda II, 305-330. Hendrix, Roberts, Stahl, and Weisberg (eds). Cold Spring Harbor Laboratory.

Fickett, J.W. (1982). Recognition of protein coding regions in DNA sequences. Nucl. Acids Res. **10**, 5303-5318.

Fishel, R.A., James, A.A. and Kolodner, R. (1981). *rec*A-independent general genetic recombination of plasmids. Nature **294**, 184-186.

Flamm, E.L. and Weisberg, R.A. (1985). Primary structure of the *hip* gene of *Escherichia coli* and of its product, the B-subunit of integration host factor. J. Mol. Biol. **183**, 117-128.

Friden, P., Voelkel, K., Sternglanz, R., and Freundlich, M. (1984). Reduced expression of isoleucine and valine enzymes in integration host factor mutants of *E. coli*. J. Mol. Biol. **172**, 573-579.

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.

Friedman, D.I., and Gottesman, M. (1983). Lytic mode of lambda development. In: Lambda II, 21-52. Hendrix, Roberts, Stahl, and Weisberg (eds). Cold Spring Harbor Laboratory.

Friedman, D.J., Olson, E.J., Carver, D., and Gellert, M. (1984). Synergistic effect of *himA* and *gyrB* mutations: evidence that Him functions control expression of *ilv* and *xyl* genes. J. Bacteriol. **157**, 484-489.

Futcher, A.B. (1986). Copy number amplification ot the 2-micron circle plasmid of *Saccaromyces cerevisiae*. J. Theor. Biol. **119**, 197-204.

Gamas, P., Burger, A.C., Churchward, G., Caro, L., Galas, D. and Chandler, M. (1986). Replication of pSC101: effects of mutations in the *E. coli* DNA binding protein IHF. Mol. Gen. Genet. **204**, 85-89.

Garnier, J., Osguthorpe, D.J., and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**, 97-120.

Gascoigne, N., Chien, Y., Becke, D., Kavaler, J., and Davis, M. (1984). Genomic organization and sequence of T-cell receptor beta-chain constant- and joining- region genes. Nature **310**, 387-391.

Gellert, M., and Nash, H. (1987). Communication between segments of DNA during site-specific recombination. Nature **325**, 401-404.

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

Godson, N.G., Fiddes, J.C., Carvell, B.G., and Sanger, F. (1978). Comparitive DNA sequence analysis of G4 and phi-X174 genomes.

In: Single stranded DNA phages. Denhardt, Dressler, and Ray (Eds.).

Gold, L., Stormo, G., and Saunders, R. (1984). Escherichia coli translational initiation factor IF3: A unique case of translational regulation. Proc. Natl. Acd. Sci. U.S.A. **81**, 7061-7065.

Golden, J.W., Robinson, S.J. and Haselkorn, R. (1985). Rearrangement of the nitrogen fixation genes during heterocyst differentiation in the cyanobacterium, Anabena. Nature **314**, 419-423.

Goosen, N., and van de Putte, P., (1984). Regulation of Mu transposition: I. Localization of the presumed recognition sites for HimD and ner functions controlling bacteriophage Mu transcription. Gene **30**, 41-46.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pave, A. (1980). Codon catalog usage and the genome hypothesis. Nucl. Acids Res. **8**, r49-r62.

Greenblatt, J., and Li, J. (1981). Interaction of the sigma factor and the NusA gene protein of E.coli with RNA polymerase in the initiation-termination cycle of transcription. Cell **24**, 421-428.

Greenblatt, J., McLimont, M., and Hanly, S. (1981). Termination of transcription by nusA gene protein od E.coli. Nature **292**, 215-220.

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

Gribskov, M., Burgess, R.R., and Devereux, J. (1986). PEPPLOT, a secondary structure analysis program for the UWGCG sequence analysis software package. Nucl. Acids Res. **14**, 327-334.

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

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

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

Grosjean, H., and Fiers, W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**, 199-209.

Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H.H.M., and Flavell, R.A. (1982). The construction of cosmid libraries which can be used to transform eukaryotic cells. Nucl. Acids Res. **10**, 6715-6733.

Guarneros, G., and Echols, H. (1970). New mutants of bacteriophage Lambda with a specific defect in the excision from the host chromosome. J. mol. Biol. **47**, 565-574.

Gustafsson, P. and Nordstrom, K. (1980). Control of plasmid R1 replication: kinetics of replication in shifts between different copy number levels. J. Bacteriol. **141**, 106-110.

Gustafsson, P., Wolf-Watz, H., Lind, L., Johansson, K-E. and Nordstrom, K. (1983). Binding between the par region of the plasmid R1 and pSC101 and the outer membrane fraction of the host bacteria. EMBO J. **2**, 27-32.

Hakkaart, M.J.J., Van den Elzen, P.J.M., Veltkamp, E. and Nikjamp, H.J.J. (1984). Maintenance of multicopy plasmid ClODF13 in E.coli cells: evidence for site-specific recombination of parB. Cell **36**, 203-209.

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

Hashimoto-Gotoh, T. and Ishii, K. (1982). Temperature-sensitive replication plasmids are passively distributed during cell division at non-permissive temperature: A new model for replicon duplication and partitioning. Mol. Gen. Genet. **187**, 523-525.

Hawley, D.K., and McClure, W.R. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. Nucl. Acids Res. **11**, 2237-2255.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**, 351-359.

Hiestand-Nauer, R. and Iida, S. (1983). Sequence of the site-specific recombinase gene cin and of its substrates serving in the inversion of the C segment of bacteriophage P1. EMBO J. 2, 1733-1740.

Higgins, C.F., Hiles, I.D., Salmond, G.P., Gill, D.R., Downie, J.A., Evans, I.J., Holland, I.B., Gray, L., Buckel, S.D., Bell, A.W., and Hermodson, M.A. (1986). A family of related ATP-binding sub-units coupled to many distinct biological processes in bacteria. Nature **323**, 448-450.

Hoess, R. 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. 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., Ziese, M. and Sternberg, N. (1982). P1 site-specific recombination: Nucleotide sequence of the recombining sites. Proc. Natl. Acad. Sci. (USA) **79**, 3398-3402.

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

Hoyt, M.A., Knight, D.M., Das, A., Miller, H.I., and Echols, H. (1982). Control of phage Lambda development by stability and synthesis of cII protein: role of viral cIII and host hfl, himA, and himD genes. Cell**31**, 565-573.

Hsu, P.L., Ross, W., and Landy, A. (1980). The Lambda phage att site: Functional limits and interaction with Int protein. Nature **285**, 85-91.

Iida, S., Meyer, J., Kennedy, K. and Arber, W. (1982). A site-specific conservative recombination system carried by bacteriophage P1. EMBO. J. **1**, 1445-1453.

Ikeda, H. and Tomizawa, S.I. (1968). Prophage P1, an extrachromosomal replication unit. Cold Spring Harbor Symp. Quant. Biol. **33**, 791-798.

Ikemura, T. (1981a). Correlation between the abundance of E.coli tRNAs and the occurrence ot the respective codons in its protein genes. J. mol. Biol. **146**, 1-21.

Ikemura, T. (1981b). Correlation between the abundance of E.coli tRNAs and the occurrence ot the respective codons inits protein genes: A proposal for a synonomous codon choice that is optimal for the E.coli translation system. J. Mol. Biol. **151**, 389-409.

Itoh, T., and Tomizawa, J. (1980). Formation of RNA primer for the initiation of replication of ColE1 DNA by ribonuclease H. Proc. Natl. Acad. Sci. U.S.A. **77**, 2450-2454.

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the regulation of DNA replication in bacteria. Cold Spring Harbor Symp. Quant. Biol. **28**, 329-348.

Johnson, R.C., Bruist, M.F., Glaccium, M.B. and Simon, M.I. (1984). In vitro analysis of Hin-mediated site-specific recombination. Cold Spring Harbor Symp. Biol. **49**, 751-760.

Johnson, R.C., Bruist, M.F. and Simon, M.I. (1986). Host protein requirements for in vitro site-specific DNA inversion. Cell **46**, 531-539.

Johnson, R.C. and Simon, M.I. (1985). Hin-mediated site-specific recombination requires two 26bp recombination sites and a 60bp recombinational enhancer. Cell **41**, 781-791.

Johnson, R.C., and Simon, M.I. (1987). Enhancers of site-specific recombination in bacteria. T.I.G. **3**, 262-267.

Johnson, R.C., Yin, J.C.P., and Reznikoff, W.S. (1982). Control of Tn5 transposition in E.coli is mediated by a protein from the right repeat. Cell **30**, 873-882.

Jones, C.J. (1985). Factors affecting the stability of E.coli plasmid vectors. Ph.D thesis, University of Glasgow.

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., Rudt,F. and Mertens, G. (1984). Substrate and enzyme requirements for in vitro site-specific recombination in bacteriophage Mu. Cold Spring Harbor Symp. Quant. Biol. **49**, 285-294.

Kanaar, R., Van de Putte, P. and Cozzarelli, N.R. (1986). Purification of the Gin recombination protein of Escherichia coli phage Mu and its host factor. Biochem. Biophys. Acta **866**, 170-177.

Kikuchi, A., Flamm, E. and Weisberg, R. (1985). An E.coli mutant unable to support site-specific recombination of bacteriophage lambda. J. Mol. Biol. **183**, 129-140.

Kikuchi, Y. and Nash, H.A. (1978). The bacteriophage lambda int gene product. A filter assay for genetic recombination, purification of Int and specific binding to DNA. J. Biol. Chem. **253**, 7149.

Kleckner, N., Roth, J. and Botstein, D. (1977). Genetic engineering in vivo using translocatable drug-resistance elements. New methods in bacterial genetics. J. Mol. Biol. **116**, 125-159.

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

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

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

Komai, N., Nishizawa, T., Hayakawa, Y., Murotsu, T., and Matsubara, K. (1982). Detection and mapping of six miniF-encoded proteins by cloning analysis of dissected miniF segments. Mol. Gen. Genet. **186**, 193-203.

Kramer, H., Niemoller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B., and Muller-Hill, B. (1987). lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. EMBO J. **6**, 1481-1491.

Krasnow, M.A., and Cozzarelli, N.R. (1983). Site-specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented res sites. Cell **32**, 1313-1324.

Krebs, and Reznikoff, W.S. (1986). Transcriptional and translational initiation sites of IS50: control of transposase and inhibitor expression. J. Mol. Biol. **192**, 781-791.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**, 105-132.

Lacatena, R.M., and Cesareni, G. (1983). Interaction between RNA1 and the primer precursor in the regulation of ColE1 replication. J. mol. Biol. **170**, 635-650.

Lacatena, R.M., Banner, D.W., Castagnoli, L., and Cesareni, G. (1984). Control of initiation of pMB1 replication: purified Rop protein and RNA1 affect primer formation in vitro. Cell **37**, 1009-1014.

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

Landy, A. and Ross, W. (1977). Viral integration and excision: structure of the lambda att sites. Science **197**, 1147-1160.

Lane, D., de Feyter, R., Phua, S-H., and Semon, D. (1986). D protein of miniF acts as a repressor of transcription and as a site-specific resolvase. Nucl. Acids. Res. **14**, 9713-9729.

Lange-Gustafsson, B. and Nash, H. (1984). Purification and properties of Int-h a variant protein involved in site-specific recombination of bacteriophage lambda. J. Cell Chem. **259**, 12724-12732.

Lennox, E.S. (1955). Transduction of linked genetic characters by bacteriophage P1. Virology **1**, 190.

Leong, J.M., Nunes-Duby, S., Lesser, C.F., Youderian, P., Susskind, M.M, and Landy, A. (1985). The phi 80 and P22 attachment sites: primary structure and interaction with E.coli integration host factor. J. Biol. Chem. **260**, 4468-4477.

Leung, D.W., Chen, E., Cachianes, G., and Goeddel, D.V. (1985). Nucleotide sequence of the partition function of E.coli plasmid ColE1. DNA **4**, 351-355.

Li, S., Squires, C., and Squires, C. (1984). Antitermination of E.coli rRNA transcription is caused by a control segment containing lambda nut-like sequences. Cell **38**, 851-860.

Linquist, S. (1986). The heat shock response. Ann. Rev. Biochem. **55**, 1151-1191.

Looman, A.C., Bodlaender,J., Comstock, L.J., Eaton, D., Jhurani, P., de Boer, H.A., and Knippenberg, P.H. (1987). Influence of the codon following the initiation codon on the expression of a modified lacZ gene in E.coli. EMBO J. **6**, 2489-2492.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Masukata, H., and Tomizawa, J. (1984). Effects of point mutations on formation of the RNA primer for ColE1 replication. Cell **36**, 513-522.

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

Meacock, P.A. and Cohen, S.N. (1980). Partitioning of bacterial plasmids during cell division: a cis-acting locus that accomplishes stable plasmid inheritance. Cell **20**, 529-542.

Michiels, T., Cornelis, G., Ellis, K., and Grinsted, J. (1987). Tn2501, a component of the lactose transposon tn951, is an example of a new category of class II transposable elements. J. Bacteriol. **169**, 624-631.

Miki,T., Chang, Z. and Horiuchi, T. (1984a). Control of cell division by sex factor F in E. coli II. J. Mol. Biol. **174**, 627-646.

Miki, T., Easton, A.M. and Rownd, R.H. (1980). Cloning of the replication, incompatiblity and stability functions of R plasmid NR1. J. Bacteriol. **141**, 87-99.

Miki, T., Yoshioka, K. and Horiuochi, T. (1984b). Control of cell division by sex factor F in Escherichia coli I. J. Mol. Biol. **174**, 605-625.

Miller, J.H. (1972). Experiments in Molecular Genetics. Cold Spring Laboratory, Cold Spring Harbor Press, New York.

Miller, H.I. (1984). Primary structure of the himA gene of Escherichia coli: homology with DNA-bending protein HU and association with the phenylalanyl-tRNA synthetase operon. Cold Spring Harbor Symp. Quant, Biol. **49**, 691-698.

Miller, J.H., Calos, M.P., Galas, D., Hofer, M., Buchel, D.C., and Muller-Hill, B. (1980). Genetic analysis of transpositions in the lac region of E.coli. J. Mol. Biol. **144**, 1-18.

Miller, H.I. and Nash, H.A. (1981). Direct role of the himA gene product in phage lambda integration. Nature **290**, 523-526.

Mizuuchi, M., abd Mizuuchi, K. (1980). Integrative recombination of bacteriophage lambda: Extent of DNA sequences involved in attachment site function. Proc. Nat. Acad. Sci. U.S.A. **77**, 3220-3224

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

Muller-Hill, B., Crapo, L., and Gilbert, W. (1968). Mutants that make more lac repressor. Proc. Natl. Acad. Sci. U.S.A. **59**, 1259.

Murray, N.E. (1983). Phage lambda and moleculat cloning. In: Lamda II, 395-432. Hendrix, Roberts, Stahl, and Weisberg (Eds.). Cold Spring Harbor Laboratory.

Nakamura, Y., Mizusawa, S., Court, D.L., and Tsugawa, A. (1986). Regulatory defects of a conditionally lethal nusAts mutant of E.coli. Positive and negative modulator roles of NusA protein in vivo. J. Mol. Biol. 189, 103-111.

Nash, H.A., Mizuuchi, K., Weisberg, R., Kikuchi, Y., and Gellert, M. (1977). Integrative recombination of bacteriophage lambda-The biochemical approach to DNA insertions. In: DNA insertion elements, plasmids, and episomes, p363. Bukhari et al, Eds. Cold Spring Harbor Laboratories. Cold Spring Harbor, New York.

Newman, B.J., and Grindley, N.D.F. (1984). Mutants in the gammadelta resolvase: A genetice analysis of the recombination function. Cell **38**, 463-469.

Nordstrom, K. (1984). Equipartition and other modes of partition: on the interpretation of curing kinetics using rep(ts) plasmids. Mol. Gen. Genet. **198**, 185-186.

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

Norrander, J., Kempe, T., and Messing, J. (1983). Improved M13 vectors using oligonucleotide-directed mutagenesis. Gene **26**, 101-106.

Novick, R.P., Clones, R.C., Cohen, S.N., Curtiss, R., Datta, N. and Falkow, S. (1976). Uniform nomenclature for bacterial plasmids: a proposal. Bact. Rev. **40**, 168-189.

Novick, R.P., Wyman, L., Bouanchaud, D., and Murphy, E. (1975). Plasmid life cycles in Staphylococcus aureus. In: Microbiology-1974. 105-109. Schlessinger, D. (ed.). American society for microbiology, Washington D.C.

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

Pabo, C.O., and Sauer, R.T. (1984). Protein-DNA recognition. Ann. Rev. Biochem. **53**, 295-321.

Plasterk, R.H.A., Brinkman, A.D. and Van de Putte, P. (1983). DNA inversions in the chromosome of E. coli and in phage Mu: relationship to other site-specific recombination systems. Proc. Natl. Acad. Sci. (USA) **80**, 5355-5358.

Plasterk, R.H.A., Van de Putte, P. (1984). Genetic switches by DNA inversions in prokaryotes. Biochim. Biophys. Acta. **782**, 111-119.

Platt, T. (1981). Termination of transcription and its regulation in the Tryptophan Operon of E. coli. Cell **24**, 10-23.

Post, L.E., and Nomura, M. (1980). DNA sequences from the str operon of E.coli. J. Biol. Chem. **255** 4460-4466.

Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H., and Dennis, P.P. (1979). Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in E.coli. Proc. Natl. Acad. Sci. U.S.A. **76**, 1697-1701.

Proctor, G.N. and Rownd, R.H. (1982). Rosanilins: indicator dyes for Cm^r enterobacteria containing chloramphenicol acetyl transferase. J. Bacteriol. **150**, 1375-1382.

Raibaud, O., and Schwartz, M. (1984). Positive control of transcription initiation in bacteria. Ann. Rev. Genet. **18**, 173-206.

Reed, K.C., and Mann, D.A., (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acids Res. **13**, 7207-7221.

Reed, R.R. (1981). Transposon-mediated site-specific recombination: a defined in vitro system. Cell **25**, 713-719.

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

Reznikoff, W.S. (1984). In: The microbe 1984 II. Procaryotes and Eucaryotes. 36th Symp. Soc. Gen. Microbiol. Kelly, D.P., and Carr, N.G. (eds.).

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarrington, G., Stephens, P., Eaton, M., and Humphreys, G. (1984). Codon usage can affect efficiency of translation of genes in E.coli. Nucl. Acids Res. **12**, 6663-6671.

Rodier, F., Gabarro-Apra, J., Ehrlich, R., and Reiss, C. (1982). Key for protein coding sequence identification: computer analysis of codon strategy. Nucl. Acids Res. **10**, 391-402.

Rosenberg, M., and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genet. **13**, 319-353.

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

Schaller, H. (1978). The intergenic region and the origins for filamentous phage DNA replication. Cold Spring Harbor Symp. Quant. Biol. **43**, 401-408.

Scott, J.R. (1984). Regulation of plasmid replication. Microbiol. Rev. **48**, 1-23.

Shapira, S.K., Chou. J., Richaud, F.V., and Casadaban, M.J. (1983). New versatile plasmid cloning vectors for expression of hybrid proteins coded by a cloned gene fused to *lacZ* gene sequences encoding an enzymatically active carboxy-terminal portion of beta-galactosidase. Gene **25**, 71-82.

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

Sherratt, D., Dyson, P., Boocock, M., Brown, L., Summers, D., Stewart, G. and Chan, P. (1984). Site-specific recombination in transposition and plasmid stability. Cold Spring Harbor. Symp. Quant. Biol. **49**, 227-233.

Sherratt, D.J., Summers, D.K., Boocock, M., Stirling, C., and Stewart, G. (1986). Novel recombination mechanisms in the maintenance and propagation of plasmid genes. Banbury Report **24**: Antibiotic resistance genes; Ecology, transfer and expression. pp 263-273. Cold Spring Harbor Laboratories.

Shine, J. and Dalgarno, L. (1974). The 3' terminal sequence of *E.coli* 16s rRNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71**, 1342-1346.

Shine, J., and Dalgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. Nature **254**, 34-38.

Som, T., and Tomizawa, J. (1983). Regulatory regions of ColE1 that are involved in determination of plasmid copy number. Proc. Natl. Acad. Sci. U.S.A. **80**, 3232-3236.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**, 503-517.

Stefano, J.E., and Gralla, J.D. (1982). Specific and non-specific interactions of mutant *lac* promoters. In: Promoters; structure and function, 69-79. Rodriguez and Chamberlain (eds.)

Stern, M.J., Ames, G., Smith, N.H., Robinson, E.C., and Higgins, C.F. (1984). Repetitive extragenic palindromic sequences: A major component of the bacterial genome. Cell **37**, 1015-1026.

Sternberg, N., Hamilton, D. and Hoess, R. (1981). Bacteriophage P1 site-specific recombination II. Recombination between *lox*P and the bacterial chromosome. J. Mol. Biol. **150**, 487-507.

Stewart, G. (1986). ColE1, cer site-specific recombination. Ph.D. Thesis. Glasgow University.

Stewart, V., and Yanofsky, C. (1985). Evidence for transcription antitermination of tryptophanase operon expression in Escherichia coli K12. J. Bacteriol. **164**, 731-740.

Stormo, G.D. (1986). Translation initiation. In: Maximising gene expression, 195-224. Reznikoff, and Gold (eds.).

Stormo, G.D., Schneider, T.D., Gold, L., and Ehrenfeucht, A. (1982). Use of the perceptron algorithm to distinguish translational initiation sites in E.coli. Nucl. Acids Res. **10**, 2997-3011.

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

Summers, D.K. and Sherratt, D.J. (1984). Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. Cell **36**, 1097-1103.

Summers, D.K. and Sherratt, D.J. (1985). Bacterial Plasmid Stability. Bioessays **2**, 209-211.

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

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

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

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

Tomizawa, J. (1984). Control of ColE1 PLasmid Replication: the process of binding of RNA I to the primer transcript. Cell **38**, 861-870.

Tomizawa, J. (1985). Control of ColE1 plasmid replication: initial interaction of RNA I and the primer transcript is reversible. Cell **40**, 527-535.

Tomizawa, J. and Itoh, T. (1981). Plasmid ColE1 incompatibility determined by interaction of RNA I with primer transcript. Proc. Natl. Acad. Sci. (USA) **78**, 6096-6100.

Tomizawa, J., Itoh, T., Selezer, G. and Som, T. (1981). Inhibition of ColE1 primer formation by a plasmid-specified small RNA. Proc. Natl. Acad. Sci. (USA) **78**, 1421-1425.

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

Tucker, W.T., Miller, C.A., and Cohen, S.N. (1984). Structural and functional analysis of the *par* region of the pSC101 plasmid. Cell **38**, 191-201.

Twigg, A. and Sherratt, D.J. (1980). *Trans* complementable copy number mutants of plasmid ColE1. Nature **283**, 216-218.

Van de Putte, P., Plasterk, R. and Kuijpers, A. (1984). A Mu *gin* complementing function and an invertible DNA region in *E. coli* K-12 are situated on the genetic element e14. J. Bacteriol. **158**, 517-522.

Van der Ende, A., Baker, T.A., Ogana, T., and Kornberg, A. (1985). Initiation of enzymatic replication at the origin of *E.coli* chromosome: primase as the sole priming enzyme. Proc. Natl. Acad. Sci. U.S.A. **82**, 3954-3958.

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

Voelkel-Meiman, K., Keil, R.L., amd Roeder, G.S. (1987). Recombination-stimulating sequences in yeast rDNA correspond to sequences regulating transcription by RNA polymerase I. Cell **48**, 1071-1079.

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

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha and beta subunits of ATP synthase, myosin, kinases, and other ATP requiring enzymes and a common nucleotide binding motif. EMBO J. **1**, 945-951.

Wehlmann, H., and Eichenlaub, R. (1980). Plasmid mini-F encoded proteins. Mol. Gen. Genet. **180**, 205-211.

Weisberg, R.A. and Landy, A. (1983). Site-specific recombination in phage lambda. In: Lambda II. Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (Eds). Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Wu, H.M. and Crothers, P. (1984). The locus of sequence-directed and protein induced DNA bending. Nature **308**, 509-513.

Yanish-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of M13mp18 and pUC19 vectors. Gene **33**, 103-119.

Yin, S., Bushman, W., and Landy, A. (1985). Interaction of the lambda site-specific recombination proteins Xis with attachment site DNA. Proc. Natl. Acad. Sci. U.S.A. **82**, 1040-1044.

Zieg, J. and Simon, M. (1980). Analysis of the nucleotide sequence of an invertible controlling element. Proc. Natl. Acad. Sci. (USA) **77**, 4196-4200.

Zuber, M., Patterson, T.A., and Court, D.L. (1987). Analysis of nutR, a site required for transcriptional antitermination in phage lambda. Proc. Natl. Acad. Sci. U.S.A. **84**, 4514,4518.