H24/3510

MONASH UNIVERSITY
THESIS ACCEPTED IN SATISFACTION OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
ON ..... 13 May 2003 .....
.....[redacted].....
Sec. Research Graduate School Committee

Addendum and Errata

- Throughout this thesis Christiansen has been miss-spelt as "Chritiansen, Jannière as "Janière" and deletent as "deletant".
- p23, paragraph 1, line 5: change " nucleophilic attack of the phosphodiester...." to "nucleophilic attack on the phosphodiester....".
- p23, paragraph 2, line 11: change "Conservative replication..." to "Conservative transposition...".
- p33, paragraph 1, line 10: change "...from *Bacillus subtilis*, which does not bind to...." to "...from *Bacillus subtilis*, which also binds to....".
- p57, paragraph 3, line 3: change "deprotected by incubation at 55°C..." and add "deprotected by incubation in saturating concentrations of ammonia at 55°C...".
- p82, paragraph 2, line 10: change "overexpress difficult proteins..." to "overexpress toxic proteins..'
- p83, end of paragraph 1, add: Purification using this protocol resulted in protein that was approximately 90% pure and relatively free of degradation. The protein yield from 1 L of culture was typically in the order of 1 mg of $TnpX_{1-707}$.
- p83, paragraph 2, end of line 6: "....that suggests that the protein is not completely globular, or that the skewed peak shape is due to slow dissociation of dimers during chromatography."
- p93, paragraph 1, line 6: change "....topoisomerase activity exhibited by TnpX." to "....topoisomerase activity in the TnpX preparation."
- p104, paragraph 1, line 6: change "insertion frequency 100-fold higher" to "insertion frequency 40-fold higher", paragraph 2, line 6: add "or reduce expression" to the end of the sentence.
- p112, paragraph 1, line 8: change "illicit" to "elicit".
- p112, paragraph 2, line 5: change "very high concentrations of.." to "very large amounts of".
- p128, paragraph 1, line 5: change "there maybe a slight decrease" to "there was a slight decrease".
- Fig 3.5, figure legend, last line: delete "The no protein control lane is labelled A."
- Fig 4.2 change "*lacZ*" and "*ΔlacZ*" to "*lacZ'* " and "*ΔlacZ'* ".

FUNCTIONAL ANALYSIS OF THE CLOSTRIDIAL LARGE RESOLVASE TnpX

Vicki Adams B.Sc. (Hons.),

Department of Microbiology,

Monash University

Thesis submitted for the degree of

Doctor of Philosophy,

Monash University,

Melbourne, Australia

January, 2003

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 5

CHAPTER 6

LIST OF TABLES

SUMMARY

Chloramphenicol resistance in *Clostridium perfringens* and *Clostridium difficile* is often encoded by *catP* genes located within the 6.3 kb transposons, Tn*4451* and Tn*4453*, respectively. This family of transposons is capable of mobilization into a recipient cell in the presence of another conjugative element. Transposition has been demonstrated in *Escherichia coli* and excision of the transposon is mediated by the large resolvase TnpX, which results in the production of a circular molecule that is the transposition intermediate. It has been shown in this thesis that TnpX is the only protein required for the excision reaction *in vitro* and that TnpX is the only transposon-encoded protein required for insertion. Furthermore, the DNA binding sites of TnpX have been identified at the ends of the linear transposon, the joint of the circular intermediate and several insertion/deletion sites. The results have shown that TnpX binds with high affinity to the transposon ends and the circular intermediate joint but with a very low affinity to the insertion and deletion sites.

TnpX is 707 aa long and a member of the recently established large resolvase subgroup of the resolvase/invertase family of site-specific recombinases. The large resolvase proteins show a high level of similarity to resolvase proteins but are 2 to 3 times larger than traditional resolvases. In this thesis deletion analysis has shown that the last 110 aa of TnpX were not essential for function although a reduction in excision activity was observed when they were removed. A TnpX DNA binding site was localised to a region between amino acids 493-597 and the dimerization site was localised to a region close to amino acids 324-360. Random mutagenesis of the *tnpX* gene led to the identification of 22 TnpX mutants. Many of the mutations were located within the resolvase domain of TnpX, which confirmed that the resolvase domain was important

for both the excision and insertion reactions. Several mutants were identified that were located outside the resolvase domain and purification of these mutants allowed investigation of their DNA binding characteristics. It was found that they bound aberrantly to DNA, suggesting that the DNA binding domain may have been affected.

Based on results obtained from deletion studies of TnpX and random mutagenesis of the TnpX sequence, a model has been suggested that involves a dimerization site, possibly mediated by zinc binding, a bipartite DNA binding domain and a putative multimerization site. Work presented in this thesis has led to the conclusion that the excision and insertion reactions catalyzed by TnpX are different in both their DNA and protein requirements. The DNA substrates for excision had different properties to the insertion substrates and a deletion derivative of TnpX capable of significantly higher levels of insertion showed reduced excision levels. Based on these data a model is presented for the transposition mechanism mediated by TnpX.

DECLARATION OF AUTHENTICITY

I, Vicki Adams, declare that this thesis contains no material that has been accepted for the award of any other degree or diploma in any university and, to the best of my knowledge, contains no material which has previously been published or written by another person, except where due reference is made in the text of this thesis.

It should be noted that the protein purification, gel filtration and CD spectral analysis in Chapters 3, 4 and 5 were performed by Dr Isabelle Lucet, Department of Microbiology, Monash University. Although the screening procedure for the identification of TnpX random mutants (Chapter 5) was designed by the candidate the actual screening was performed by Jonathan Kim and Pauline Howarth, Department of Microbiology, Monash University.

Vicki Adams

ACKNOWLEDGEMENTS

I would really like to thank Julian Rood, my supervisor, (look commas!) for the opportunity to undertake a Ph.D in his laboratory. I have enjoyed working with you and the other members of the Rood laboratory immensely, the camaraderie within the lab is considerable. Julian, you have always had time for trivial queries or downright panic. You have worked very hard so that I can get this thesis finished and for that I thank you especially (and tell Meryl that I apologize for the amount of work you have done especially over the holidays). Another thank-you for trying to teach me how to write English despite the fact that I had apparently been taught this for 13 years, however, in that case it was called creative writing, which is what I'm still doing! So thanks Julian for all the thats, this' and whichs and I will not forget that overexpress is not hyphenated I promise (just wait until we get to the papers!).

My heartfelt thanks go to Dena Lyras, my other supervisor, for her boundless support on all manner of things including science, family, which restaurants to eat at and what to wear! Still working on that one. I hope that I will one day become a scientist of your calibre; you have taught me so much, including how to be a good supervisor. Thank-you so much Dena.

I would also like to thank my pseudosupervisor Isabelle who came onto the project two years ago and has shown us time and time again that if you want to do biochemistry get a biochemist to do it not a geneticist. Thank-you so much Isabelle for all your time, hard work and expertise when we were sorely in need of it.

Thanks also go to the rest of the Rood lab both past and present for stimulating and amusing conversation. These include Ruth Kennan, Trudi Bannam, the original TnpX girl (one day you will recover), Milena Awad, queen of the cutting reply, Ishara, Dane, Jenny, Xiaoyan and Anthony. I would like to thank Pauline Howarth for her work on the random mutants and generally superb technical expertise. Thanks also to Jonathan Kim for work conducted on the random mutants, I hope your medical career goes well. Thanks to our honours students for hours of entertainment, particularly Martina, who also struggled with TnpX. I would especially like to thank my past and present Ph.D students Priscilla, Jackie, Kylie and Sheena. It's always good to go through an ordeal with someone! The same goes for the adoptive Roodies, Lisa and Fiona. Lots of good times enjoyed by all.

I would also like to thank Jamie Rossjohn for attempting to crystalize TnpX, James Whisstock for being excited about TnpX as can only be seen in his brilliant mind and particularly Rob Pike who showed me how to perform gel filtration before the arrival of Isabelle.

Thanks go to the various netball team members, there have been a few over the years, usually at the same time. Netball helped to keep me sane although often the umpires were trying to make me more frustrated! May everyone I play with currently have many good years of netball left before all their knees give out!

I must also thank my family, Mum, Dad, Chris and Paul and Karen Hall. Thank-you for all the love, support and laughter over the years I would not have been where I am now without your help. I would also like to thank my grandparents Frank and Jean Hall for the hotdogs and the faith that I will eventually get there. Thank-you to Ben for all your support, particularly over the last 5 years. Thank-you for the dinners, house cleaning,

food shopping and dog-walking you've done when I'm too tired. And thank you so much for being able to sort out all my computer problems when I had given up.

ABBREVIATIONS

A	deoxyadenosine
A	ampere
aa	amino acids
acc. no.	accession number
AP	ammonium peroxodisulfate
Ap	ampicillin
~	approximately
ATP	adenosine triphosphate
bp	base pair(s)
BSA	bovine serum albumin
C	deoxycytosine
CCC	covalently closed circle (supercoiled DNA)
°C	degrees Celsius
CD	circular dichroism
CI	circular intermediate
Cm	chloramphenicol
°	degrees
Δ	denotes a deletion in genotype descriptions
dATP	deoxyadenosine triphosphate
dH_2O	distilled water
dsDNA	double-stranded DNA
DIG	digoxigenin
DIG-dUTP	digoxigenin-11-dUTP
DTT	1,4-dithiothreitol
dUTP	deoxyuracil triphosphate

DNA	deoxyribonucleic acid
dNTP(s)	deoxynucleotide triphosphate(s)
EDTA	ethylenediaminetetra-acetic acid
G	deoxyguanidine
g	gram(s)
g	gravitational constant
h	hour(s)
HCl	hydrochloric acid
6 x His	hexahistidine
His	histidine
Ω	denotes an insertion event in genotype descriptions
IPTG	isopropyl-β-D-thiogalactoside
IR	inverted repeat
IS	insertion sequence
kb	kilobase pair(s)
kDa	kilodalton(s)
Kn	kanamycin
kV	kilovolt(s)
L	litre(s)
LB	Luria-Bertani media
M	molar
mA	milliamperes
μF	microfarad(s)
μg	micrograms(s)
mg	milligram(s)
min	minute(s)

µL	microlitre(s)
mL	millilitre(s)
µm	micrometre(s)
µM	micromolar
mm	millimetre(s)
mM	millimolar
MOPS	3-[N-Morpholino]propane sulfonic acid
mRNA	messenger ribonucleic acid
#	number
NaCl	sodium chloride
NaOH	sodium hydroxide
ng	nanogram(s)
nM	nanomolar
nm	nanometre(s)
NMR	nuclear magnetic resonance
OC	open circular
ORF(s)	open reading frame(s)
ori	origin of replication
%	percent, percentage
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pH	$\log_{10}[H^+]$
^{R}	resistant
RBS	ribosome binding site

res	resolvase binding site
Rif	rifampicin
RNA	ribonucleic acid
RNase	ribonuclease
RP	reverse primer
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	second(s)
SOB	tryptone-yeast extract-NaCl
SOC	SOB + glucose
SSC	sodium salt citrate buffer
ssDNA	single-stranded DNA
T	deoxythymidine
TAE	Tris Acetic Acid EDTA
TBS	Tris buffered saline
TE	Tris EDTA
TEMED	*N*, *N*, *N'*, *N'*-tetramethylethylenediamine
Tn	transposon
Tris	Tris[hydroxymethyl]aminomethane
tRNA	transfer ribonucleic acid
$Tween_{20}$	polyoxyethylenesorbitan monolaurate
u	unit (enzyme activity)
UP	universal primer
U.S.A	United States of America
UV	ultraviolet
V	volts
vol.	volume(s)

v/v	volume per volume
w/v	weight per volume
YT	yeast extract tryptone

AMINO ACID ABBREVIATIONS

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartate
E	Glu	glutamate
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

Chapter 1

Introduction

The Clostridia

The genus *Clostridium* represents a diverse range of organisms that are grouped together as Gram positive, anaerobic rods capable of producing heat resistant endospores, but that otherwise may be unrelated (Stackebrandt and Rainey, 1997). Many of the clostridial species produce a large number of extracellular toxins and enzymes (Rood, 1998; Stackebrandt and Rainey, 1997). The major pathogens within the genus include the neurotoxigenic clostridia, *C. botulinum* and *C. tetani*, the enterotoxigenic *C. difficile* and the enterotoxic and histotoxic *C. perfringens* (Rood, 1998).

a) Clostridium perfringens

C. perfringens is the causative agent of gas gangrene, the pathogenesis of which involves the production and release of α toxin (Ellemor *et al.*, 1999). This species produces fourteen different toxins and enzymes, although not all are expressed in any one strain (McClane *et al.*, 2000). *C. perfringens* is also responsible for various gastrointestinal diseases ranging from food poisoning to enteritis necroticans. The former is caused by enterotoxin-producing strains (McClane *et al.*, 2000), the latter caused by isolates that produce β-toxin (Lawrence, 1997).

Tetracycline and erythromycin resistance represent the most common antibiotic resistance phenotypes found in *C. perfringens*, however, resistance to chloramphenicol is also observed (Lyras and Rood, 1997). Erythromycin resistance is mediated most

commonly by the erm(Q) gene (Berryman *et al.*, 1994) but is also encoded by the *erm*(B) gene, which is located on the 63 kb non-conjugative plasmid pIP402 (Berryman and Rood, 1995; Brefort *et al.*, 1977). Tetracycline resistance is more widely distributed and is encoded by the *tetA*(P) gene, which is located on the chromosome or on large conjugative plasmids such as pCW3 (Lyras and Rood, 1997; Lyras and Rood, 2000a; Sloan *et al.*, 1994). Plasmids like pCW3 also encode a second tetracycline resistance gene *tetB*(P), as part of the inducible *tet*(P) operon (Johanesen *et al.*, 2001; Sloan *et al.*, 1994). The protein encoded by the *tetA*(P) gene, TetA(P), is an efflux protein that actively pumps tetracycline out of the cell (Sloan *et al.*, 1994). TetB(P) is a ribosomal protection protein and has considerable similarity to the widely distributed TetM protein (Sloan *et al.*, 1994). Tetracycline resistance in *C. perfringens* is constitutive when encoded by the *tet*(M) gene, which is located on the chromosome within a defective Tn*916*-like determinant (Roberts *et al.*, 2001b).

Chloramphenicol resistance in *C. perfringens* is mediated by chloramphenicol acetyltransferase proteins encoded by *cat* genes (Lyras and Rood, 1997). These enzymes act by acetylating chloramphenicol, rendering the molecule incapable of binding to bacterial ribosomes and consequently resulting in loss of its antibacterial activity (Lyras and Rood, 1997). Resistance to chloramphenicol in *C. perfringens* is not as common as tetracycline or erythromycin resistance but is generally encoded by the *catP* gene, which is located on large conjugative plasmids related to pCW3 (Abraham and Rood, 1987; Abraham *et al.*, 1985; Lyras and Rood, 1997). In at least two of these plasmids the *catP* genes are found within related mobilizable transposons, namely Tn*4451* and Tn*4452* (Abraham and Rood, 1987). Chloramphenicol resistance is also mediated by the *catQ* gene, which is chromosomally located, shows limited similarity to *catP*, and is not thought to be transposon-associated (Lyras and Rood, 1997).

There are many different plasmids found in *C. perfringens*, encoding genes for antibiotic resistance, bacteriocin production and toxin production (Lyras and Rood, 2000a; Rood, 1998). The plasmid pIP404 encodes bacteriocin production (Garnier and Cole, 1986), and the beta, epsilon and iota toxins and enterotoxin structural genes can also be plasmid borne (Lyras and Rood, 2000a). *C. perfringens* is the only clostridial species that has been shown to contain conjugative tetracycline resistance plasmids, such as pCW3 (Lyras and Rood, 2000a). It has been shown that there are at least 13 other plasmids derived from various human, porcine and environmental isolates that show a significant similarity to pCW3 (Abraham *et al.*, 1985). Several of these plasmids, such as pIP401 and pJIR27, also encode chloramphenicol resistance by carrying Tn*4451* or Tn*4452* (Abraham and Rood, 1987; Abraham *et al.*, 1985).

Several potentially mobile genetic elements have been identified in *C. perfringens*. These elements include IS*1151*, IS*1469*, IS*1470* and Tn*5565* (Lyras and Rood, 2000a). IS*1151* is often associated with the epsilon toxin gene, *etx*, and sometimes the enterotoxin gene, *cpe*, when the latter is plasmid encoded (Brynestad *et al.*, 1997). A second element, IS*1469*, is located upstream of the *cpe* gene (Brynestad *et al.*, 1997). In food poisoning isolates this chromosomal arrangement is flanked by two copies of IS*1470*, that forms a putative transposon, Tn*5565*, which may be capable of excision as a circular molecule (Brynestad and Granum, 1999). Finally, *C. perfringens* may carry the chloramphenicol resistance transposon Tn*4451*, which is found on pIP401 (Abraham and Rood, 1987; Brefort *et al.*, 1977). Excision of Tn*4451* is mediated by the large resolvase TnpX, which is the subject of this thesis.

b) Clostridium difficile

C. difficile causes several gastrointestinal diseases ranging from antibiotic associated diarrhoea (AAD) to pseudomembranous colitis (PMC), which can be lethal if untreated (McClane *et al.*, 2000). These are predominantly nosocomial infections, commonly occurring in the elderly following antibiotic treatment (McClane *et al.*, 2000). The causative *C. difficile* strain is not necessarily resistant to the antibiotic being used in treatment but is able to sporulate and subsequently colonise the gut after treatment ceases, when the normal flora has been reduced (McClane *et al.*, 2000). The organism is then able to release large amounts of Toxins A and B into the lumen, causing characteristic lesions to develop (McClane *et al.*, 2000).

Resistance to tetracycline, erythromycin and chloramphenicol has also been found in *C. difficile* (Lyras and Rood, 2000a). Resistance to tetracycline is mediated by the *tet*(M) gene, which is located on the conjugative transposon Tn*5397* (Roberts *et al.*, 1999; Wang and Mullany, 2000). This element shows high levels of similarity to Tn*916* from *Enterococcus feacalis* and the Tet(M) determinant from *C. perfringens* (Roberts *et al.*, 2001b). A unique feature of this transposon is that it includes a functional Group II intron (Roberts *et al.*, 2001a). Tn*5397* encodes a large resolvase, TndX, which has been shown to be the protein responsible for excision of the element from the chromosome (Wang and Mullany, 2000).

Erythromycin resistance in *C. difficile* is mediated by *erm*(B) genes (Farrow *et al.*, 2000), which are found on the mobile element Tn*5398* (Farrow *et al.*, 2001). This element encodes no identifiable transposase or recombinase genes but is nevertheless excised and transferred by conjugation to recipient cells (Farrow *et al.*, 2001). Finally, chloramphenicol resistance is chromosomally located in *C. difficile* and is encoded by *catP* genes that are present on mobile elements (Tn*4453a* and Tn*4453b*). These

transposons are very similar to Tn*4451* from *C. perfringens* (Lyras and Rood, 2000b; Lyras *et al.*, 1998).

Mobile Genetic Elements in Bacteria

Genetic diversity in microbial populations is maintained in a number of ways, from random mutation during replication or under adverse environmental conditions, to complex pre-programmed cell differentiation. One mechanism contributing to genetic change is the movement of mobile genetic elements. These elements are highly varied in structure, mechanism of movement, and composition. Some elements, such as many of the insertion sequences, are small and cryptic, encoding only the genes required for excision and insertion of the element. Others, such as bacteriophages can be very large and encode not only proteins that allow genetic movement but many structural proteins as well.

a) Insertion Sequences (IS).

Insertion Sequence (IS) elements are small mobile elements that usually only encode one or two open reading frames (ORFs) responsible for translocation or transposition of the element and/or the regulation of this process (Mahillon and Chandler, 1998). The protein responsible for transposition, the transposase, acts on the ends of the element, catalysing strand exchange (Mahillon and Chandler, 1998). These ends often consist of inverted repeats (IR) (Mahillon and Chandler, 1998) and by convention the IRs upstream and downstream of the transposase gene are designated inverted repeat left (IRL) and inverted repeat right (IRR), respectively (Mahillon and Chandler, 1998).

IS transposition involves the transposase protein catalyzing strand cleavage by a conserved process, although the transposition products vary depending on the mechanism of second strand processing (Mahillon and Chandler, 1998). With two notable exceptions, the IS*91* family and some uncharacterised IS, the transposase proteins contain a conserved, acidic amino acid motif, known as the DDE motif, which has been shown to be involved in catalysis (Fig 1.1) (Mahillon and Chandler, 1998).

Control of the transposition process is necessary to prevent damage to the host genome by mutations caused by the insertion events (Mahillon and Chandler, 1998). Control is inherent in the manner in which most transposase proteins function, that is, they generally operate *in cis* due to the instability of both the mRNA transcripts and the protein itself (Mahillon and Chandler, 1998). Movement of co-resident IS elements is therefore unlikely and the frequency of transposition is reduced (Mahillon and Chandler, 1998). Control of transposition is not always achieved in such a manner and there are other mechanisms employed (Mahillon and Chandler, 1998).

IS elements include a diverse range of genetic elements that are separated into several families based on sequence similarity of terminal IRs, the nature of the transposase protein, the length of direct repeats produced via target site duplication (DR) and the transposition mechanism (Mahillon and Chandler, 1998). There are approximately 18 families, with some families further divided into subgroups (Mahillon *et al.*, 1999). There are too many IS families to describe in great detail here so discussion will be limited to the more relevant families.

Members of the IS*91* family are unlike other IS elements because they encode transposase proteins showing greater similarity to rolling circle replication proteins than to DDE transposases (Mendiola *et al.*, 1994). These elements use rolling circle

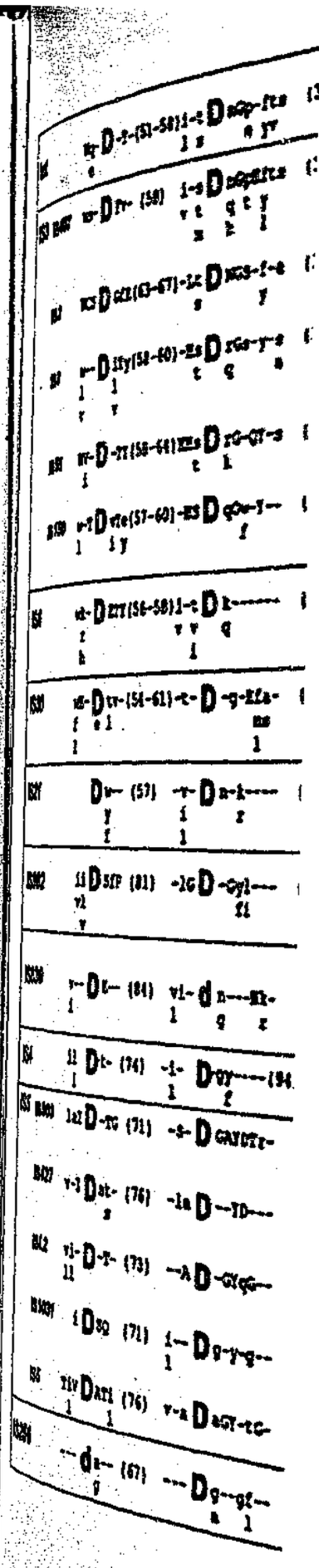
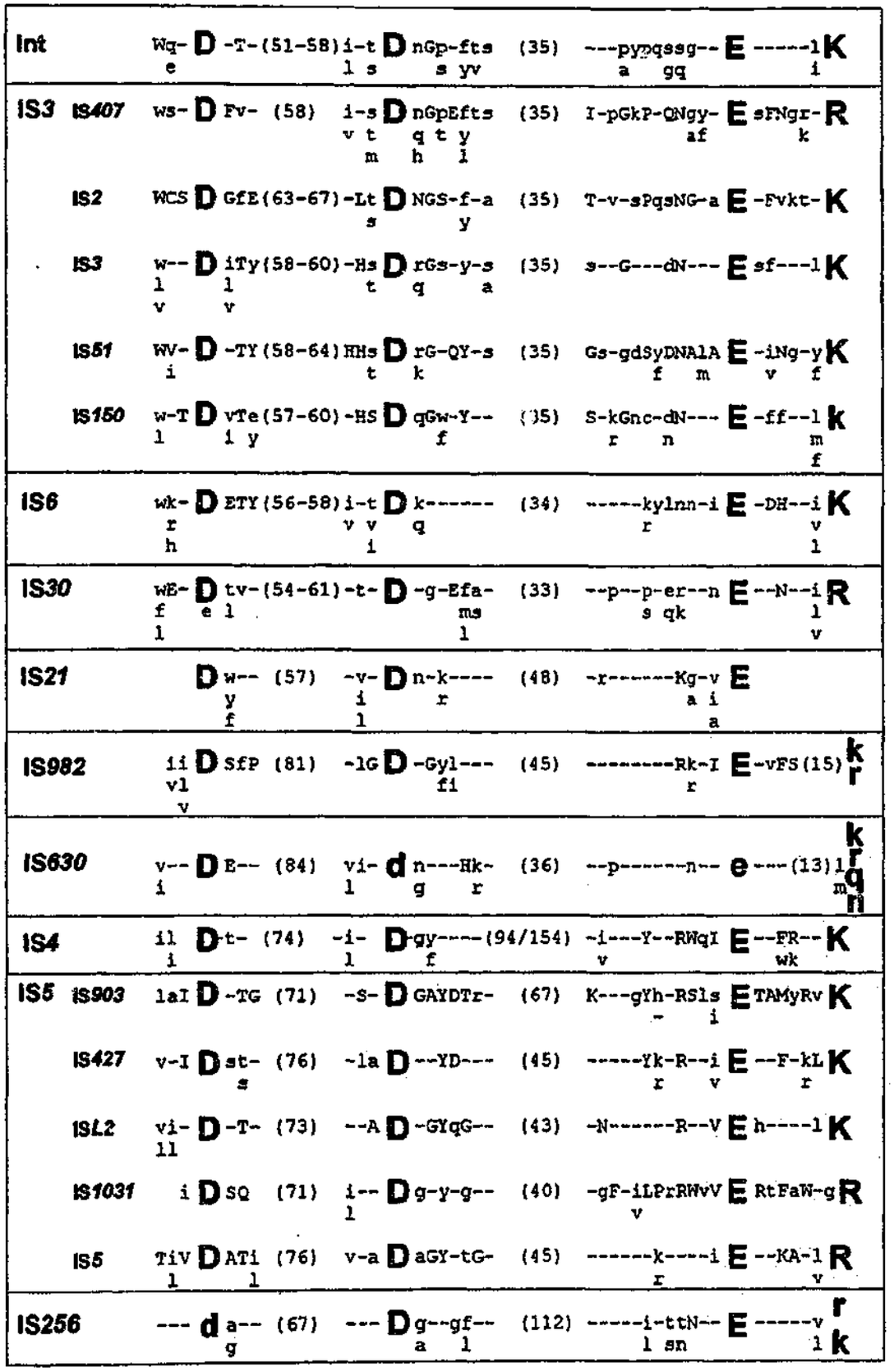


Figure 1.1 Alignment of DDE transposases

Transposase families are seperated by horizontal lines. The DDE residues are shown in large bold type. Capital letters indicate residues that show conservation within a family, lower case letters indicate residues that predominate but are not completely conserved. The numbers indicate amino acid spaces between residues. The large font IS names are family group names, small font IS names represent specific elements. Modified from Mahillon and Chandler (1998).



shown in
in a family,
onserved.
names are
ied from

replication to accomplish replicative transposition (Mahillon and Chandler, 1998). This family is unusual in other ways due to the nature of the transposition process (Tavakoli *et al.*, 2000) and they have no terminal repeats, target site specificity, or target site duplications (Mahillon and Chandler, 1998; Mendiola *et al.*, 1994; Tavakoli *et al.*, 2000). The IS*91* transposition protein is capable of working *in trans* (Mendiola *et al.*, 1994), while another member, IS*1294*, encodes a transposition protein proficient *in cis* (Tavakoli *et al.*, 2000).

The left end of IS*91*-like elements is not directly involved in the transposition process but acts as a replication terminator, while strand transfer begins precisely with the right end (Mendiola *et al.*, 1994; Tavakoli *et al.*, 2000). Movement of these elements to the insertion site is thought to involve a ssDNA pseudocircle, that is, a non-covalently linked DNA circle held together by the transposase protein (Tavakoli *et al.*, 2000). Depending on the efficiency of the terminator region, the transposition process can also result in the movement of DNA sequences downstream of the left end, for example, the IS*1294*-mediated translocation of the *aph* gene from pUB2380 (Tavakoli *et al.*, 2000).

The IS*110* family also has no terminal IRs, and does not generate direct repeats on insertion (Mahillon and Chandler, 1998). Two members of this family have been relatively well studied, IS*117* from *Streptomyces coelicolor*, and IS*492* from *Pseudoalteromonas atlantica*. IS*117*, also known as the minicircle, exists as a linear form, where it resides in either of two genomic locations, and as a circular molecule (Henderson *et al.*, 1989). It is thought that since the circular form of the element is capable of a higher frequency of integration than the linear form, then the circular form may be the transposition intermediate (Henderson *et al.*, 1989). The element encodes three putative ORFs, ORF1 is the largest and shows similarity to ORF1215 of IS*110*

(Henderson *et al.*, 1989), and may encode one of a novel subgroup of DDE transposases, possessing a DED or DDD acidic motif (Tobiason *et al.*, 2001).

IS*492* plays an important regulatory role in *Pseudoalteromonas atlantica*. This marine organism is capable of living as part of a biofilm on a solid support or in the open ocean. IS*492* is responsible for the phase variation of adhesion molecules on the bacterial surface (Perkins-Balding *et al.*, 1999). IS*492* also forms circles and was thought to produce DRs until it was found that one copy of the 5 bp sequence found at either end of the integrated element was derived from the target site, while the other formed the joint of the circular form (Perkins-Balding *et al.*, 1999). Unlike some IS transposition mechanisms, the IS*492* donor molecule is repaired, clearly a factor of significance if the element is involved in an important host regulatory process (Perkins-Balding *et al.*, 1999). The nature of the insertion process, the lack of direct repeats, and the presence of identical sequences at the joint and target sites suggests a novel transposition mechanism. The transposase protein MooV, shows some similarity to the site-specific invertase Piv, which also contains an acidic triad (Tobiason *et al.*, 2001).

A feature of IS*492* and many other IS families, is the production of a strong promoter at the joint of the circular form. This promoter is capable of driving expression of the transposase protein, thereby increasing the likelihood of insertion of the circular molecule (Perkins-Balding *et al.*, 1999). IS*492* is not only found within the adhesion locus but there are several copies throughout the genome. These IS*492* elements have all inserted into very similar target sites, with conservation of the 5 bp repeat where catalysis occurs (Perkins-Balding *et al.*, 1999).

The IS*1* family includes some of the smallest IS elements (Turlan and Chandler, 1995), including IS*1*, which was the first IS element identified (Mahillon and Chandler, 1998).

IS*1* is 768 bp long, includes 26 bp terminal IRs, produces a 9 bp DR on insertion and has two consecutive ORFs (Mahillon and Chandler, 1998). Transcription from the promoter, which is partially contained within IRL, results in the expression of InsA, the major but shorter product and InsAB', which is produced through a programmed translational frameshift (Mahillon and Chandler, 1998). InsA is capable of binding to IRL and down-regulating transcription by blocking the binding site of RNA polymerase, as well as by blocking the binding site of the transposase protein InsAB' (Turlan and Chandler, 1995). Transcription of the transposase gene can also be affected by the presence of a Rho-dependent transcriptional terminator at the end of the *insA* gene (Mahillon and Chandler, 1998). The transposase protein of IS*1* has conserved residues found in the λ Int family of site-specific recombinases (Mahillon and Chandler, 1998).

IS*1* is capable of both simple insertion and co-integration or replicon fusion, either by replicative or non-replicative means; there is also some evidence that IS*1* is capable of forming circular molecules, which may lead to simple insertion (Turlan and Chandler, 1995). Transposition has been shown to be down-regulated by transcription from genes outside of the IS (Mahillon and Chandler, 1998).

The large IS*3* family ranges from 1.2 kb to 1.5 kb in size, typically exhibiting IRs of 20-40 bp in length (Mahillon and Chandler, 1998). These elements generally consist of two overlapping reading frames and, as with the IS*1* family, produce multiple gene products through translational frameshifting. Unlike IS*1*, IS*3* elements can produce three proteins, OrfA, OrfAB and OrfB (Mahillon and Chandler, 1998). OrfA is capable of binding IRL, decreasing transcription of the transposase gene, *orfAB* (Mahillon and Chandler, 1998). IS*3* is capable of forming circular or linear molecules and makes

staggered cuts at either end of the element, resulting in 3 bp overhangs. These molecules are capable of simple insertion into a new target site (Sekine *et al.*, 1996).

IS*911*, another member of the IS*3* family, is 1.25 kb long, carries 36 bp IRs and generates a 3 or 4 bp DR upon insertion (Polard *et al.*, 1994). IS*911* is capable of circularization and at the joint of the circular molecule forms a strong promoter that is capable of driving transposase transcription (Ton-Hoang *et al.*, 1997). The transposase protein is sensitive to temperature, being less active above 30°C (Haren *et al.*, 1997). Many transposase proteins show limited stability, including that of IS*903*, which shows marked sensitivity to the Lon protease (Mahillon and Chandler, 1998). The inherent instability of these proteins is thought to be part of the reason many transposase proteins function only *in cis* (Mahillon and Chandler, 1998).

Transposition of IS*911* is highly variable, with the production of many different intermediates, including circular (Ton-Hoang *et al.*, 1997) and linear forms (Ton-Hoang *et al.*, 1999), the use of surrogate IRs to form pseudo-cointegrates (Polard *et al.*, 1994), and the production of tandem IS dimers that lead to the formation of normal cointegrates (Turlan *et al.*, 2000). The OrfAB transposase is responsible for each of these pathways, the outcomes of which are dependent on the arrangement of the IR and IR-like sequences. Neither of the pathways are replicative, illustrating that IS*911* transposition occurs by a conservative mechanism (Turlan *et al.*, 2000).

The IS*21* family includes some of the largest IS elements, ranging from 2 kb to 2.5 kb in size (Mahillon and Chandler, 1998). They generally encode two proteins, IstA and IstB (Mahillon and Chandler, 1998). IS*21* forms a tandemly repeated dimer that contains a promoter located between the two abutted transposon ends. This promoter drives transcription of the downstream IS genes (Reimmann *et al.*, 1989). The *istA* gene

encodes the DDE transposase (Reimmann and Haas, 1990), whereas IstB appears to stimulate transposition (Schmid *et al.*, 1998). The IS*21* tandem repeat $(IS21)_2$, is a highly active transposition substrate, being processed to form non-replicative cointegrates at high frequency (Reimmann and Haas, 1990). Simple insertion of IS*21* is catalysed by the full length IstA protein, while integration of the $(IS21)_2$ element is most efficient when orchestrated by the cointegrase (Schmid *et al.*, 1998). The transposase and cointegrase are both encoded by the *istA* gene, but the transposase consists of the full length IstA protein and the cointegrase begins at an alternative start site, resulting in a protein that lacks the N-terminal eight amino acids (Schmid *et al.*, 1998).

The IS*30* family ranges from 1 kb-1.2 kb with 20-30 bp terminal IRs and produces a 2-3 bp DR on insertion (Mahillon and Chandler, 1998). IS*30* consists of one reading frame, *orfA*, (Kiss and Olasz, 1999) and shows some target site specificity (Caspers *et al.*, 1984). It is capable of forming circular molecules with a 2 bp junction sequence; these molecules are capable of subsequent simple insertion (Kiss and Olasz, 1999). Like IS*21*, a higher degree of transposition is possible if tandemly repeated copies of IS*30* are produced, again with an intervening 2 bp sequence (Olasz *et al.*, 1993). The production of the $(IS30)_2$ dimer can be brought about by the deletion of all but 2 bp of intervening DNA between two directly repeated elements (Kiss and Olasz, 1999; Olasz *et al.*, 1993). This structure is generally thought to be a transposition intermediate as it frequently undergoes further processing to form other recombination products such as deletion, inversion or integration (non-replicative cointegrate formation) derivatives (Kiss and Olasz, 1999; Olasz *et al.*, 1993). The 2 bp intervening sequence is lost from either the tandem repeat or the circular form, with the introduction of direct repeats on insertion into new target DNA (Kiss and Olasz, 1999).

b) Transposons

Transposons, like IS elements, are capable of movement from one position to another and encode the proteins necessary to catalyze movement. However, unlike IS elements, transposons may encode additional genes whose function is unrelated to transposon movement. There are many different forms of transposons and they are capable of variable degrees of movement, from intramolecular transposition to intercellular transfer (Craig, 1996). Transposons also vary greatly in size from 2-3 kb to greater than 150 kb (Salyers *et al.*, 1995a; Sherratt, 1989; Tan, 1999). Some transposons are defective, whereas others appear to be intact but have never been observed to move. Some elements, while still independent, form part of more complex, multifactorial mobile genetic elements.

Compound transposons are very diverse, but all contain two IS elements that flank an otherwise immobile DNA segment (Fig 1.2). Movement of these elements can involve the whole transposon, or one or both IS elements. Recombination can occur at four possible sites. Each IS element may retain two functional terminal IRs (Fig 1.2), leading to variable transposition products. The mechanism of transposition is dependent on the IS element involved, as the transposase protein is encoded by one or both of the IS elements bounding the transposon. The intervening DNA segment often encodes an antibiotic resistance gene, or genes for the catabolism of unusual substrates. Well studied examples include Tn*10* (Kleckner, 1989; Pribil and Haniford, 2000) and Tn*5* (Berg, 1989; Bhasin *et al.*, 1999; Reznikoff *et al.*, 1999).

Tn*10* (9.3 kb) is bounded by two inverted copies of IS*10*, known as IS*10L* and IS*10R* (Pribil and Haniford, 2000). IS*10L* encodes a defective transposase protein but both terminal repeats are intact (Kleckner, 1989). IS*10R* encodes a fully functional transposase and is capable of movement independently from the rest of Tn*10* (Kleckner,

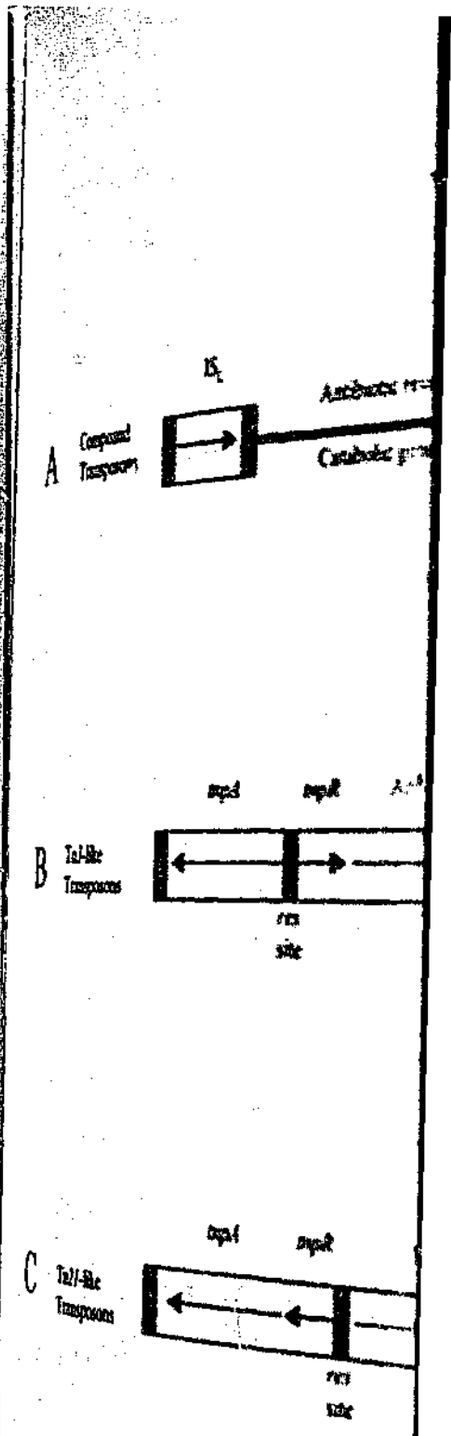
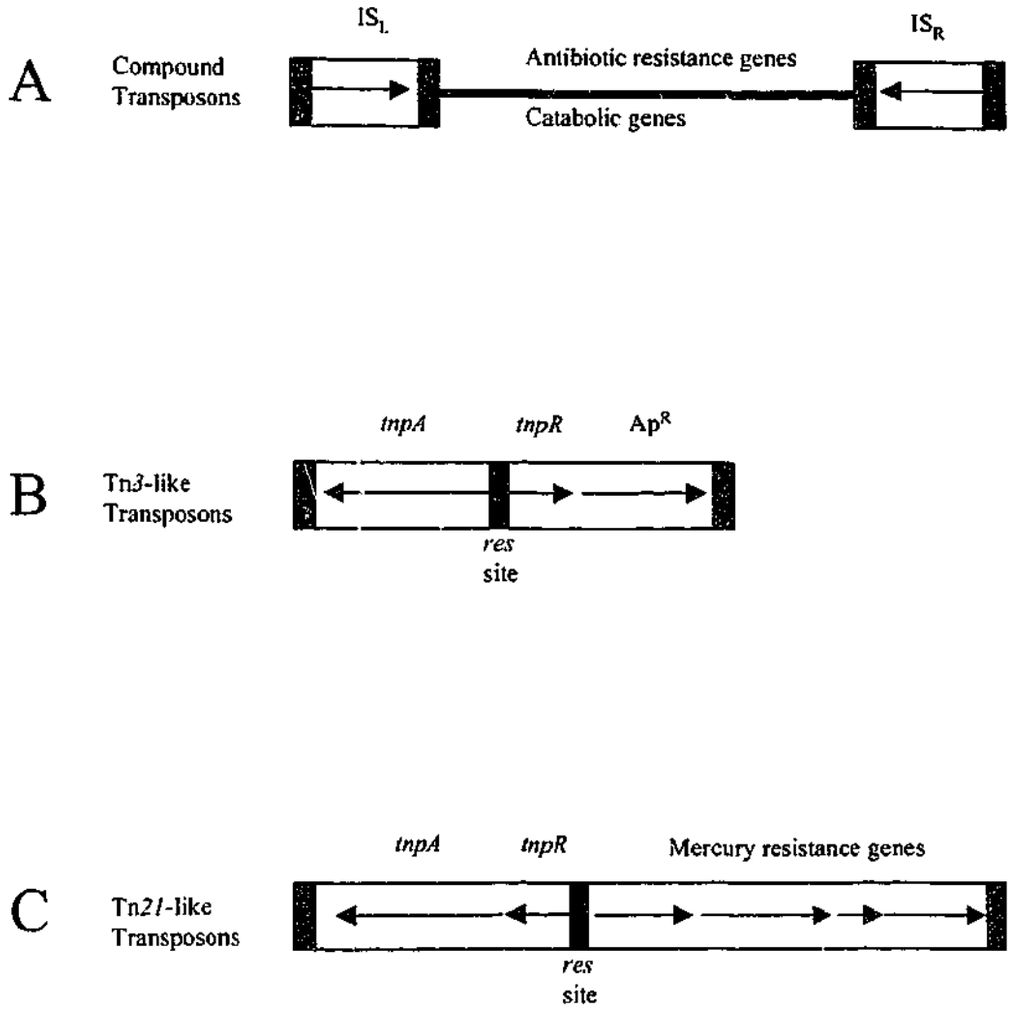


Figure 1.2 Major types of transposons

A Simplified structure of a compound transposon. Flanking IS elements have 'captured' the intervening DNA (Craig, 1996). The transposase genes are indicated by the blue arrows and the terminal repeats by the blue boxes.

B Structure of Tn*3*. Note the divergent expression of the transposase (*tnpA*) and resolvase (*tnpR*) genes (Craig, 1996). The ends of the transposon are shown in blue and the *res* site is coloured purple.

C Structure of Tn*21*. Transcription of the transposase and resolvase genes occurs in the same direction (Craig, 1996). The mercury resistance operon is indicated by orange arrows.


A
Compound
Transposons
IS$_L$
IS$_R$
Antibiotic resistance genes
Catabolic genes
B
Tn*3*-like
Transposons
tnpA
tnpR
Ap^R
res
site
C
Tn*21*-like
Transposons
tnpA
tnpR
Mercury resistance genes
res
site

1989). The IS*10R* transposase is also capable of moving IS*10L in trans*, although once removed from IS*10R*, IS*10L* is immobile (Kleckner, 1989). Tn*10* encodes tetracycline resistance, transposes using a cut and paste mechanism and exhibits some target site specificity. It introduces a 9 bp target site duplication on insertion (Pribil and Haniford, 2000), consistent with IS*10*-mediated transposition (Mahillon and Chandler, 1998). Transposition of Tn*10* can result in multiple transposition end products (Kleckner, 1989), including inverse transposition where recombination between the inside termini of Tn*10* produces a new Tn*10* transposon with different intervening DNA and IS*10* elements in reverse orientation (Kleckner, 1989).

The Tn*3* family of transposons do not contain IS elements (Fig 1.2). Common features include a similarity in structure, terminal inverted repeats, the *tnpA* and *tnpR* genes, and a conserved region known as the *res* site (Sherratt, 1989). The DDE transposase, TnpA, is responsible for the replicative transposition of these elements, resulting in cointegrate formation (Sherratt, 1989). The cointegrate is further processed by a resolvase protein, TnpR, that catalyses resolution into individual replicons, both of which retain a copy of the transposon (Sherratt, 1989). TnpR-mediated recombination, unlike the transposition reaction, involves not the transposon ends, but two internal *res* sites that are present in a directly repeated manner within the cointegrate structure (Sherratt, 1989; Watson *et al.*, 1996).

Members of the Tn*3* family fall into one of two subgroups depending on the orientation and organization of the *tnpR* and *tnpA* genes and the position of the *res* site. Tn*3*-like elements show divergent transcription of the *tnpA* and *tnpR* genes (Fig 1.2) (Sherratt, 1989). Many of the members of this group also encode ampicillin resistance (Sherratt, 1989). The second subgroup includes the mercury resistance transposons, Tn*21* and Tn*501* (Liebert *et al.*, 1999) and the catabolic transposons Tn*4655* and Tn*4653* (Tan,

1999). In these elements transcription of the *tnpA* and *tnpR* genes occurs in the same direction (Fig 1.2) (Sherratt, 1989).

There are other transposons which do not show sequence similarity to Tn*3*-like elements, and do not contain repeated IS elements. One such element is Tn*7*. This transposon shows considerable target site specificity, transposing to a single site in the *E. coli* chromosome, *att*Tn*7* (Craig, 1989). Tn*7* is 17 kb in length, encodes resistance to trimethoprim, streptomycin and spectinomycin, and includes five genes, *tnsABCDE,* that are involved in the transposition process (Craig, 1989). Transposition can result in insertion into *att*Tn*7* at high frequency, or into a multitude of sites at low frequency. The *tnsABC* genes are required for both kinds of transposition, however, no movement will occur without either *tnsD* or *tnsE* (Peters and Craig, 2001b). The TnsD protein mediates specific insertion into *att*Tn*7* (Peters and Craig, 2001b). TnsE binds preferentially to DNA structures associated with DNA replication and primarily directs integration into sites where DNA replication terminates within the bacterial chromosome or into conjugative plasmids (Peters and Craig, 2001a; Rao and Craig, 2001). TnsA and TnsB are transposase proteins responsible for cleavage of the DNA at the transposon ends (Rao and Craig, 2001). TnsC is a regulatory protein that mediates contact between the transposase proteins (TnsA and B) and the target site proteins (TnsD or E). TnsC is capable of binding and hydrolysing ATP and transposition is activated when ATP is bound by TnsC (Stellwagen and Craig, 2001). It is postulated that binding of TnsC to TnsD stabilises the TnsC-ATP state and allows formation of the recombination complex (Stellwagen and Craig, 2001). Tn*7* will not integrate into a target that already contains Tn*7*, a process known as transposition immunity, and may result from ATP hydrolysis mediated by interactions between TnsC and TnsB molecules already bound to the integrated transposon (Stellwagen and Craig, 2001).

Insertion of transposons at high frequency would not benefit the host organism because interruption of important host functions could result. Consequently, mechanisms have been developed to control unrestricted transposition. Tn*7* avoids this problem by preferentially inserting into *att*Tn*7*, which has little or no detrimental effect on the host. Further control of transposition occurs through the TnsB protein, which binds to the *tnsAB* promoter within the right end of Tn*7* (Craig, 1989) and may mediate transposition immunity (Stellwagen and Craig, 2001).

Another unclassified element is Tn*554*, which shares some similarity to Tn*7* (Murphy, 1989). Tn*554* integrates with high specificity and frequency into a single site, *att554*, in *Staphylococcus aureus* (Murphy, 1989; Murphy *et al.*, 1991). However, Tn*554* contains no terminal repeats and insertion, which is generally in the same orientation, does not lead to target site duplication (Murphy, 1989). Tn*554* encodes two proteins, *tnpA* and *tnpB*, both of which have similarity to the λ integrase family of site-specific recombinases (Murphy *et al.*, 1991). Both Tn*554* and Tn*7* insertion therefore resemble bacteriophage insertion at the mechanistic level. Tn*554* also encodes the small protein, TnpC, which is involved in orientation specificity on insertion (Murphy *et al.*, 1991), and macrolide-lincosamide-streptogramin B (MLS) and spectinomycin resistance genes (Murphy, 1989). Tn*554* has been found within type II and III SCC*mec* elements (Ito *et al.*, 2001; Ito *et al.*, 1999).

c) Conjugative Transposons

Conjugative transposons are highly mobile elements that encode their own transposition and conjugative transfer functions; therefore they are capable of moving both intracellularly and intercellularly (Clewell *et al.*, 1995; Rice, 1998; Salyers *et al.*, 1995b; Scott and Churchward, 1995; Waters, 1999). They have been shown to

contribute very significantly to the spread of antibiotic resistance genes between many bacterial genera (Waters, 1999). Transposition of conjugative transposons is usually mediated by an integrase protein rather than a DDE transposase (Rudy *et al.*, 1997). The elements range in size from Tn*916* at 18 kb (Rice, 1998) to *Bacteroides* conjugative transposons in excess of 120 kb (Smith *et al.*, 1998). With a few exceptions, conjugative transposons encode tetracycline resistance, usually carrying *tet*(M)-like resistance genes (Rice, 1998). Additional antibiotic resistance genes can also be present, such as those encoding erythromycin or kanamycin resistance (Rice, 1998; Smith *et al.*, 1998).

Most of the proteins encoded by these elements are involved in conjugative transfer, they include proteins responsible for DNA nicking and mobilization. Like conjugative plasmids these elements have an *oriT* site, where the circular form is nicked and relaxed to allow single stranded transfer to the recipient cell (Fig 1.3) (Rice, 1998). Other proteins encoded within the transfer regions are known to be important for transfer but their specific role is as yet unknown (Clewell *et al.*, 1995).

Tn*916* is the best studied conjugative transposon (Clewell *et al.*, 1995; Rice, 1998). Transposition is mediated by the products of the *int* and *xis* genes, which are located at one end of the transposon (Rice, 1998). Int belongs to the integrase family of site-specific recombinases and is required for insertion of Tn*916*, while both Int and Xis are required for excision (Jia and Churchward, 1999). Excision of Tn*916* results in the formation of a circular intermediate with a coupling sequence of 6 bp of mismatched bases at the joint (Jia and Churchward, 1999). This heteroduplex structure is resolved either by single-stranded transfer into the recipient cell, or by mismatch repair or replication (Fig 1.4) (Clewell *et al.*, 1995; Rice, 1998). Insertion is relatively specific,

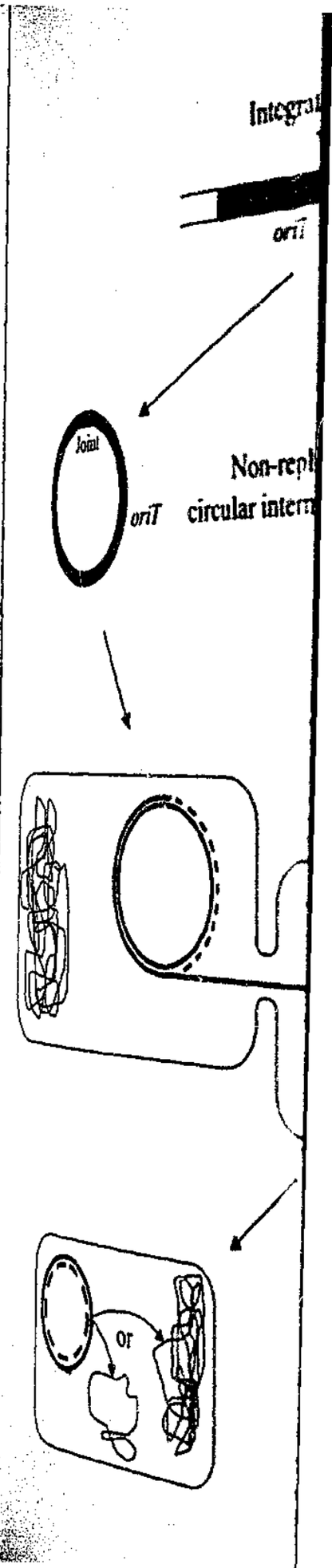
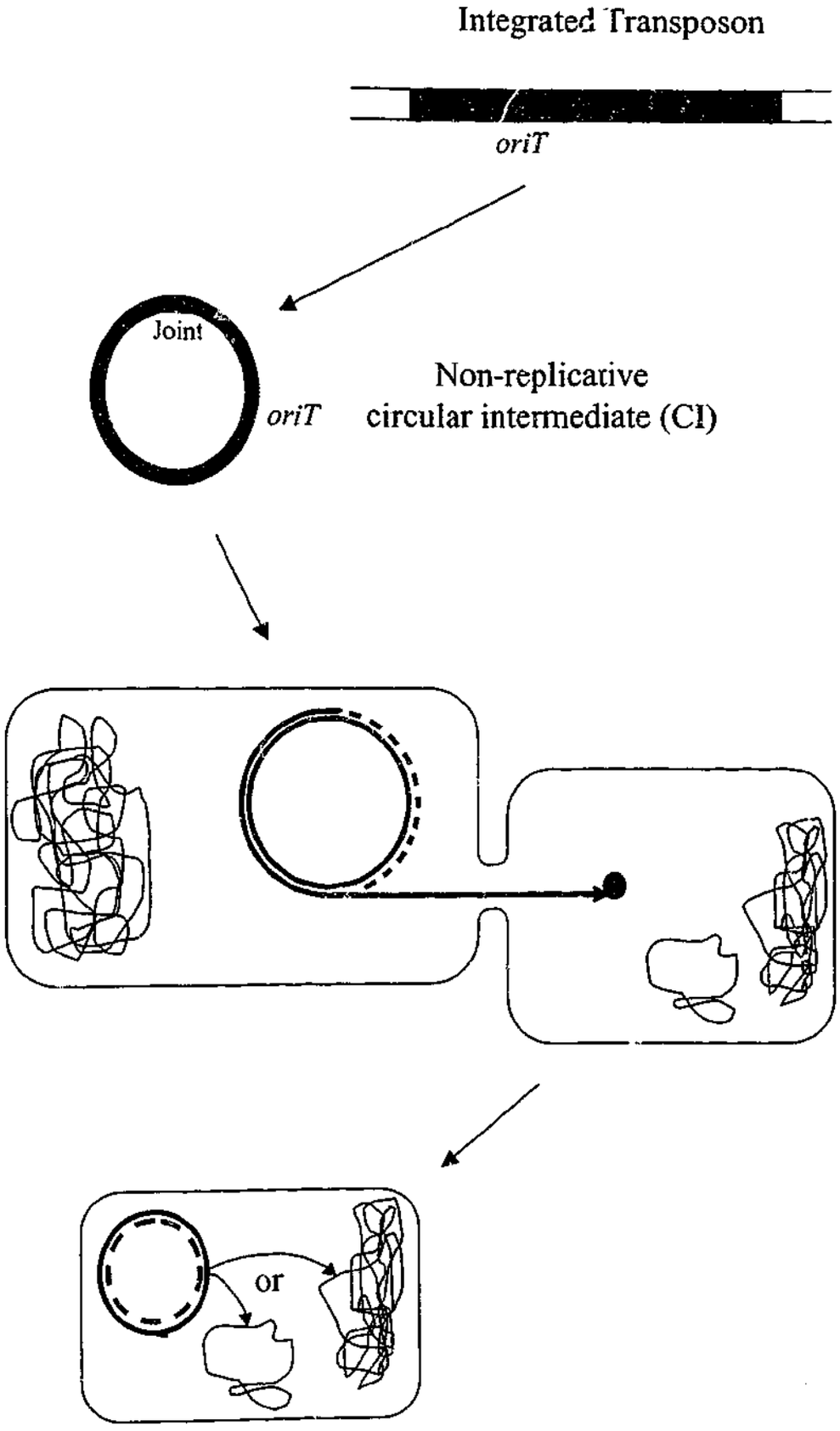


Figure 1.3 Conjugative transposition

The integrated transposon (blue rectangle) is excised at the transposons ends (red boxes) as a circular molecule. Nicking of the circular intermediate occurs at the *oriT* site (green box) and the single-stranded DNA is transferred across the mating bridge complexed to mobilization protein(s) (purple circle) to a recipient cell. The single-stranded DNA is replicated (dashed blue line), both in the host and the recipient and the new double stranded circular molecules must integrate into either the chromosome or a plasmid if they are to be stably inherited (red lines denote potential target sites).


Integrated Transposon
oriT
Joint
oriT
Non-replicative
circular intermediate (CI)
or

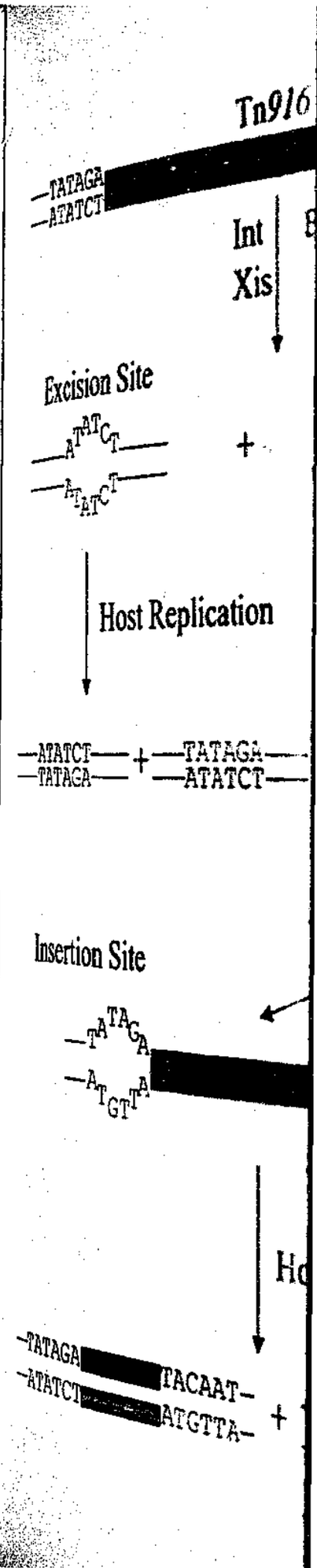
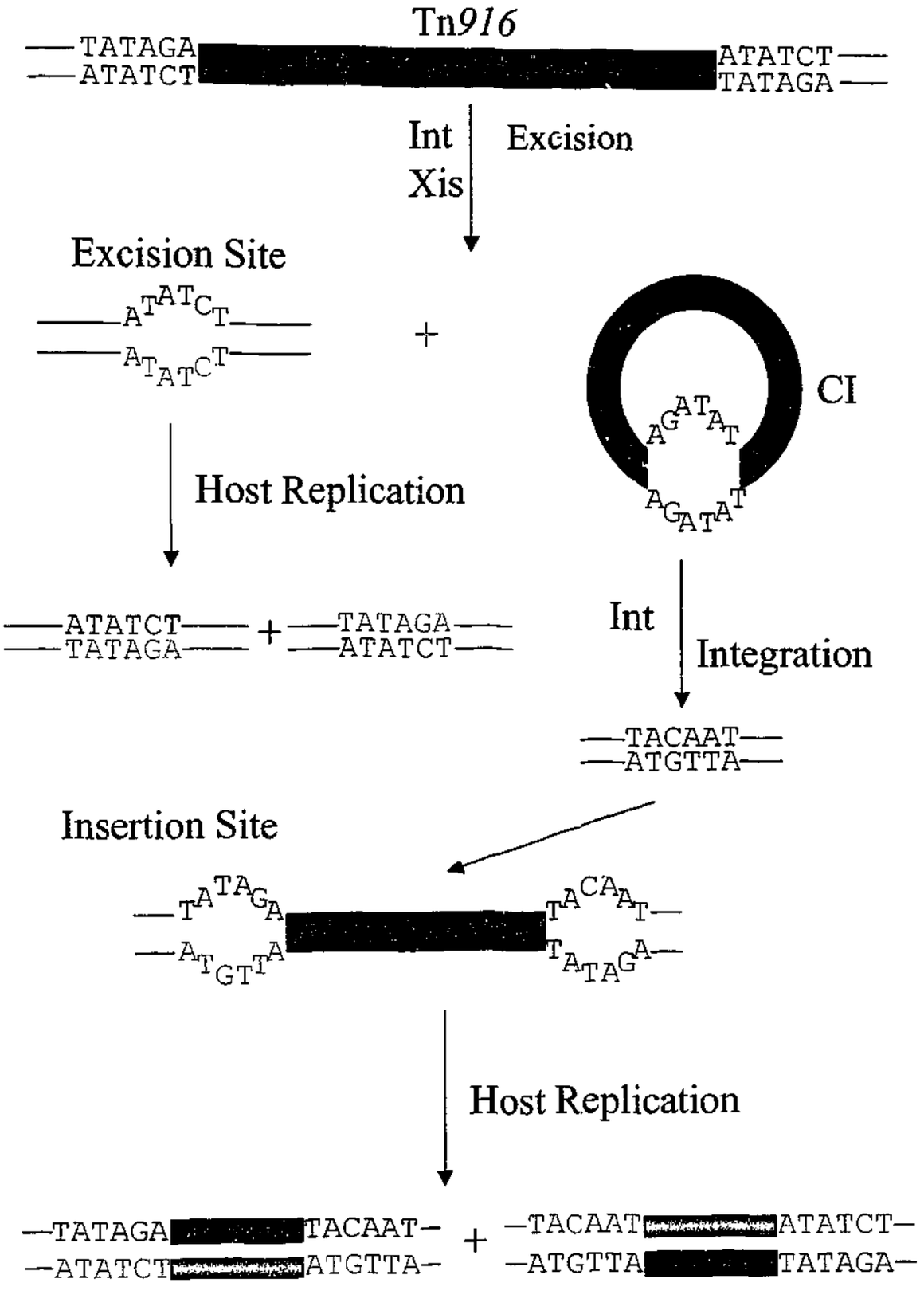


Figure 1.4 Heteroduplex formation at the joint of the circular intermediate

The conjugative transposon (green rectangle) is excised from the flanking DNA by the action of the Int and Xis proteins, producing a 6 bp mismatched region, or heteroduplex, at the joint of the circular intermediate (CI) and the excision site. The deletion site heteroduplex is resolved during host replication or by mismatch repair. The transposon heteroduplex is resolved after integration at a new target site. Modified from Scott and Churchward (1995).


Tn916
TATAGA
ATATCT
ATATCT
TATAGA
Int
Xis
Excision
Excision Site
CI
Host Replication
ATATCT
TATAGA
TATAGA
ATATCT
Int
Integration
TACAAT
ATGTTA
Insertion Site
Host Replication
TATAGA
TACAAT
ATATCT
ATGTTA
TACAAT
ATATCT
ATGTTA
TATAGA

occurring preferentially in AT rich regions, at sequences resembling the joint of the circular form (Rice, 1998).

The *Bacteroides* conjugative transposon CTnDOT, encodes tetracycline and erythromycin resistance and integrates into only seven sites within the *Bacteroides* chromosome (Cheng *et al.*, 2000). Transposition of this element, and others like it, is positively regulated by tetracycline, through the action of the transposon-encoded regulatory proteins RteA, RteB and RteC (Cheng *et al.*, 2000). It has recently been shown that excision is the regulated step, tetracycline induction resulting in a 1,000 to 10,000 fold increase in transfer frequency (Cheng *et al.*, 2000). These elements are also unusual as they are not only capable of transferring other DNA segments *in trans*, such as co-resident non-replicating *Bacteroides* units (NBUs), but also non-mobilizable plasmids *in cis* (Cheng *et al.*, 2000; Salyers *et al.*, 1995a). Mobilization of NBUs is also regulated by tetracycline in the presence of conjugative transposons, although the exact mechanism is unclear (Smith *et al.*, 1998).

There are several other novel conjugative transposons that will be mentioned briefly. The *Bacteroides* element, XBU4422, unlike other conjugative transposons, is cryptic and does not encode an antibiotic resistance gene (Salyers *et al.*, 1995b). Tn*5276* also does not carry antibiotic resistance genes, but encodes genes for sucrose metabolism and resistance to the bacteriocin, nisin (Salyers *et al.*, 1995a). Tn*5397* is a conjugative tetracycline resistance element from *Clostridium difficile* that, unlike other conjugative transposons, encodes a large resolvase that facilitates excision, the formation of a circular molecule and integration (Roberts *et al.*, 1999; Wang and Mullany, 2000). This element is otherwise very similar to Tn*916* (Roberts *et al.*, 2001b).

d) Mobilizable Transposons

Mobilizable transposons are capable of intercellular movement but do not encode all of the proteins necessary for bacterial conjugation (Smith *et al.*, 1998). They generally carry antibiotic resistance genes (Wang *et al.*, 2000b) and encode mobilization proteins, which allow their transfer into a recipient cell (Vendantam *et al.*, 1999). Mobilizable transposons are capable of cell-to-cell movement only if there is a co-resident conjugative plasmid or transposon also present in the cell (Crellin and Rood, 1998; Shoemaker *et al.*, 2000; Smith *et al.*, 1998; Tribble *et al.*, 1999a; Vendantam *et al.*, 1999; Wang *et al.*, 2000b).

Mobilizable elements range in size from Tn*5520*, at 4.7 kb (Vendantam *et al.*, 1999) to Tn*4555* at 12.1 kb (Tribble *et al.*, 1999b). Many of these elements are found in *Bacteroides* species (Smith *et al.*, 1998), including NBU1, NBU2, NBU3, Tn*4399*, Tn*4555*, and Tn*5520* (Smith *et al.*, 1998; Vendantam *et al.*, 1999). NBU1, NBU2 and Tn*4555* have been shown to form a circular intermediate that is capable of single stranded transfer into a recipient cell after nicking at *oriT* by an element-encoded mobilization protein (Shoemaker *et al.*, 2000; Tribble *et al.*, 1999a; Wang *et al.*, 2000b). Tn*4399* is capable of mobilizing non-conjugative plasmids *in cis* and carries the *mocA* and *mocB* genes, which are required for recognition and nicking of the transposon at *oriT* (Vendantam *et al.*, 1999). Tn*4555* encodes cefoxitin resistance and has been shown to produce a circular intermediate during transposition (Vendantam *et al.*, 1999). The *mobA* gene encodes the mobilization protein and the recombinase responsible for excision and integration of the element and belongs to the λ integrase family (Tribble *et al.*, 1997).

NBU1 and NBU2 show significant nucleotide similarity to Tn*4555* in the *oriT* and the mobilization gene regions, although the transposition mechanisms of these elements are

different (Shoemaker *et al.*, 2000; Tribble *et al.*, 1997; Tribble *et al.*, 1999a; Wang *et al.*, 2000b). The Mob proteins are responsible for both *oriT* recognition and nicking activity unlike many other mobilization processes, which require two proteins for these tasks (Salyers *et al.*, 1995b). Conjugative mobilization is induced by sub-inhibitory concentrations of tetracycline when the conjugative helper element is one of the conjugative tetracycline resistance transposons from *Bacteroides* (Smith *et al.*, 1998). The regulatory proteins produced by the conjugative transposons also act on the mobilizable transposons to regulate excision and transfer (Smith *et al.*, 1998).

Another *Bacteroides* mobilizable element, cLV25, has recently been discovered. This element encodes two mobilization proteins similar to those of Tn*4399*. However, the integrase gene shows similarity to *int* encoded on Tn*4555*. cLV25 has been shown to transpose in *E. coli* and is also active in the native host (Bass and Hecht, 2002).

e) Integrons

Integrons are comprised of specialised gene capture and expression elements (Hall and Collis, 1995). They consist of an integrase gene, *intI*, a recombination site known as *attI*, a promoter, and one or more gene cassettes, which are mobile elements in their own right (Brown *et al.*, 1996; Hall and Collis, 1995; Recchia and Hall, 1995). The gene cassette normally consists of a single gene, usually an antibiotic resistance gene, and a second recombination site known as a 59 base element (59 be) or *attC* site (Brown *et al.*, 1996; Hall and Collis, 1995; Recchia and Hall, 1995). Site-specific recombination between *attI* and a 59 be site located on a non-replicating, circular cassette molecule results in integration of the cassette into the integron (Brown *et al.*, 1996; Hall and Collis, 1995; Recchia and Hall, 1995).

There are four major types of integrons, designated as groups I-IV (Bennett, 1999; Recchia and Hall, 1995). The basic integron, In0, does not include any gene cassettes and consists of a 5' conserved sequence (5'CS), including the *intI* gene, an *attI* site, and the promoter for cassette expression, followed by a 3'conserved sequence (3'CS), which for group I elements consists of a sulphonamide resistance gene, *sulI*, and several complete or partial ORFs of unknown function (Bennett, 1999). Group I integrons such as In2 from Tn*21* encode IntI1 and one or more gene cassettes (Brown *et al.*, 1996). Many group I elements have been shown to excise and integrate their gene cassettes (Recchia and Hall, 1995). Group II integrons consist of elements encoding a defective integrase gene, *intI2*, and are found at the left end of Tn*7*-like elements. They also lack the 3'CS region (Bennett, 1999). There is one group III element identified to date but it has not been fully sequenced (Recchia and Hall, 1995). This element encodes the *intI3* gene, which is similar to *intI1*, but not the entire element. Superintegrons comprise the final group. They are very large and may contain over one hundred integrated cassettes (Rowe-Magnus *et al.*, 1999). *Vibrio cholerae* contains one such element that makes up approximately 10% of the genome (Rowe-Magnus *et al.*, 1999). This integron is 126 kb in length and consists of at least 179 cassettes (Rowe-Magnus *et al.*, 1999). An integrase gene, *intI4*, is located upstream of the first cassette along with an *attI* site and a promoter; the same basic structure as the other three integron groups (Rowe-Magnus *et al.*, 1999). The cassettes are bordered by 59 be sites with no more than two genes appearing without an intervening 59 be (Rowe-Magnus *et al.*, 1999). Superintegrons have been identified in two different *V. cholerae* strains, both have similar cassettes but have a different order to the cassettes, indicating that superintegrons remain dynamic in terms of their genetic organization (Rowe-Magnus *et al.*, 1999). Superintegrons are not limited to *V. cholerae*, they have also been identified in *V. metschnikovii*, *Pseudomonas*, *Xanthomonas*, *Shewanella* and *Nitrosomonas* (Rowe-Magnus *et al.*, 1999).

f) Bacteriophage Mu

Mu is an unusual bacterial virus that replicates using transposition (Pato, 1989). The phage DNA, in the form of a non-covalently closed circle, is held together by the MuN protein, which helps to protect the ends of the incoming DNA from nuclease degradation by the host (Pato, 1989). The invading DNA is integrated into the host chromosome by the phage-encoded proteins, MuA and MuB (Chaconas *et al.*, 1996; Pato, 1989). MuA is a multidomain member of the DDE family of transposases (Chaconas *et al.*, 1996) and is responsible for the initial integration event and subsequent replicative transposition, both of which result in a 5 bp target site duplication (Pato, 1989). MuB is responsible for binding to potential target sites and for recruitment of the nucleoprotein complex in which cleavage and strand exchange take place (Chaconas *et al.*, 1996; Pato, 1989). The integrated genome is capable of both intra- and intermolecular movement, resulting in deletion or inversion events and cointegrate formation, respectively (Pato, 1989).

Mu is 36.7 kb in length and has ends with limited similarity to Tn*3*-like transposons (Morgan *et al.*, 2002). The genes encoding MuA and MuB are located close to the left end of the element, with most of the genome taken up with head and tail genes. Towards the right end of the element is a segment of invertible DNA encoding alternative tail fibres that determine host specificity. The *gin* gene encodes the invertase responsible for the inversion reaction (Pato, 1989).

The life cycle of Mu begins with the initial integration of the genome, thus initiating either a lysogenic or lytic cycle. The lytic cycle occurs in the majority of infections and sees the genome replicated, through transposition, approximately 100 times. Packaging of the phage head prior to release results in inclusion of a DNA fragment larger than

37 kb and therefore includes more than the phage genome. After assembly of the phage particles they are released from the cell to renew the cycle (Pato, 1989).

It has been postulated that Mu may have evolved from the integration of a Tn*3*-like transposon into a λ-like phage, since there are areas of similarity in both elements (Pato, 1989). Mu is not the only transposing phage, for example, the element D108 shows similarity to Mu and elements such as VcA1, VcA2 and some *Pseudomonas* phages have similar properties to Mu (Benevides *et al.*, 1994; Johnson and Romig, 1981).

Recombination Mechanisms

There are many different types of mobile elements and there is also considerable diversity in the mechanisms by which these elements move. However, there are only three main families of recombinases commonly used by mobile elements; these are the DDE transposases, and two groups of site-specific recombinases, the λ integrase family and the resolvase/invertase family.

a) Transposase Mediated Transposition

Members of the DDE transposase family belong to a large superfamily of proteins, the phosphoryltransferases, which all have a similar structural fold (Aravind *et al.*, 2000; Grindley and Leschziner, 1995; Haren *et al.*, 1999; Turlan and Chandler, 2000). DDE transposases are commonly associated with mobile elements including most of the major IS families, many transposons and bacteriophage Mu (Chaconas *et al.*, 1996; Hallet and Sherratt, 1997; Mahillon and Chandler, 1998). The DDE catalytic site is conserved in these proteins, which may otherwise be quite variable in primary sequence (Fig 1.1) (Aravind *et al.*, 2000).

The acidic DDE residues are brought together within the active site and form the catalytic core, co-ordinating two divalent metal ions (Allingham *et al.*, 1999; Craig, 1996; Hallet and Sherratt, 1997; Mizuuchi, 1997). The metal ions in the catalytic pocket interact with the phosphate group to be cleaved, maintaining it in a pentavalent form (Hallet and Sherratt, 1997). This arrangement allows nucleophilic attack of the phosphodiester bond (Fig 1.5) (Craig, 1996; Hallet and Sherratt, 1997; Turlan and Chandler, 2000). The nucleophile usually involved in this step is water (Craig, 1996; Grindley and Leschziner, 1995; Hallet and Sherratt, 1997; Mizuuchi, 1997; Turlan and Chandler, 2000).

DDE-mediated transposition involves characteristic target site duplication, which is a direct result of the catalytic mechanism. The reaction proceeds in a two-step fashion, initially involving cleavage of the 3' end of the element (Craig, 1996; Hallet and Sherratt, 1997; Turlan and Chandler, 2000). Treatment of the 5' end of the element depends on which transposase is involved (Craig, 1996; Hallet and Sherratt, 1997; Turlan and Chandler, 2000). Some elements exhibit replicative transposition while others use a conservative or non-replicative method (Craig, 1996; Hallet and Sherratt, 1997; Turlan and Chandler, 2000). During replicative transposition, second strand cleavage does not occur until after host replication and is performed by host proteins, resulting in the formation of a cointegrate molecule (Fig 1.6A) (Craig, 1996; Hallet and Sherratt, 1997; Turlan and Chandler, 2000). Conservative replication can result in a multitude of different structures.

Non-replicative intramolecular transposition can result in inversions and deletions, while intermolecular transposition results in simple insertions (Craig, 1996; Turlan and Chandler, 2000), generally resulting in the loss of the donor molecule unless rescued by host repair processes (Craig, 1996). When 5' cleavage of the element occurs, it is

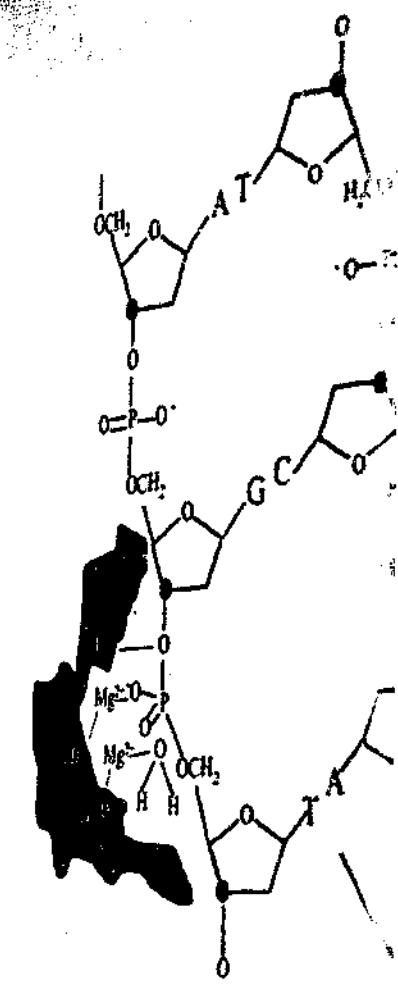
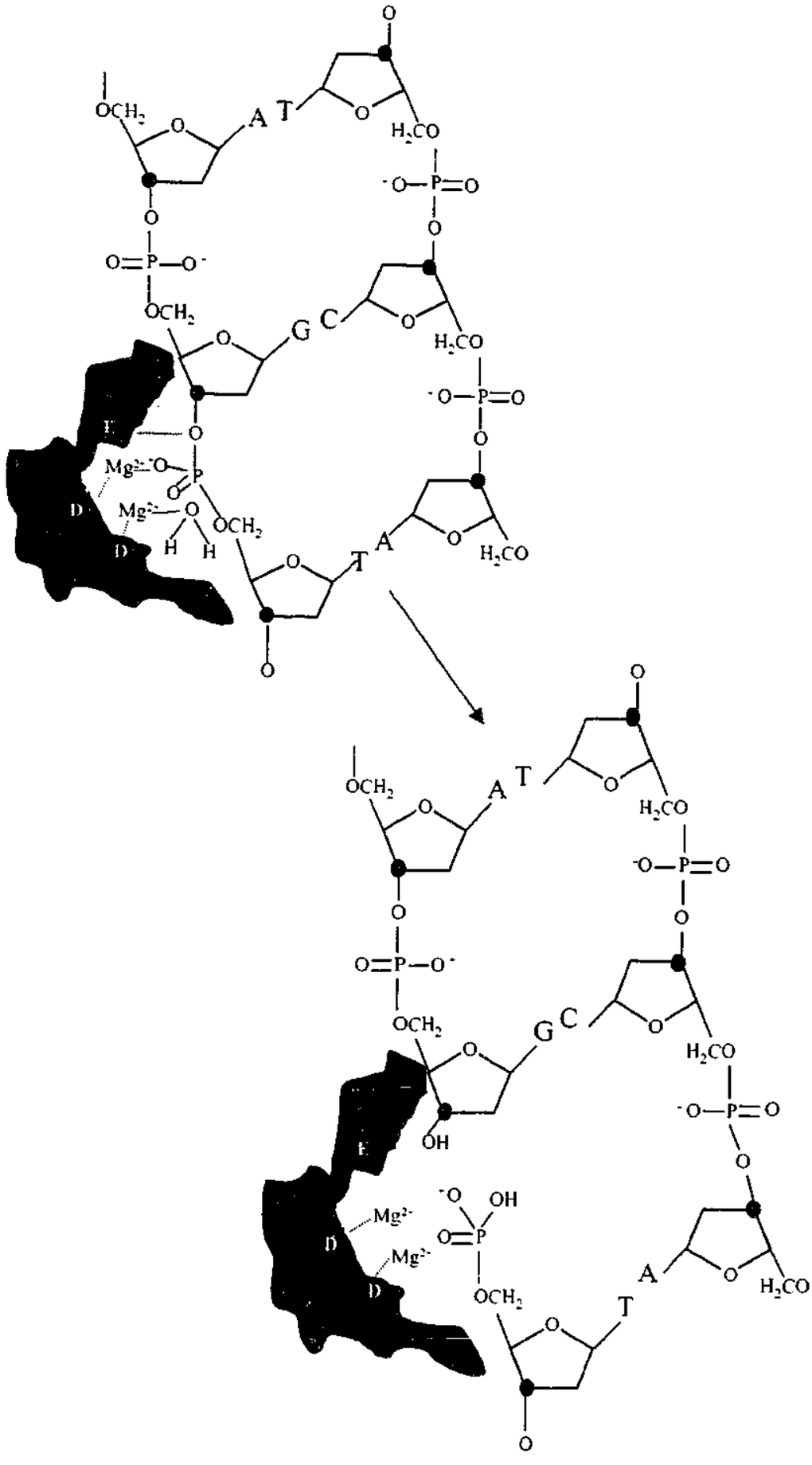


Figure 1.5 Transposase Mechanism

The transposase protein (purple) binds the DNA (red C's represent 5' carbons and blue circles represent 3' carbons). The DDE triad (white) forms a complex with two (or one) magnesium ions (blue dashed lines) and helps to stabilise the phosphate group in a pentavalent structure to allow attack by a water molecule (green). The phosphate group is hydrolyzed breaking the DNA backbone and producing a 5' phosphate group and a 3' hydroxyl group. Cleavage of the opposite DNA strand is accomplished differently by different elements and often the 3'OH is involved either in cleavage of the other DNA strand (Tn*10*) and/or the target site (Kleckner, 1989). Based on Mizuuchi (1997).


A
T
G
C
T
A
Mg2+
Mg2+
D
D
D
E
OH
OH
H
H

catalysed by the transposase but the nucleophile is the 3'OH group produced during the initial cleavage, not water, and results in the formation of a covalent phosphodiester bond (Fig 1.6) (Craig, 1996; Hallet and Sherratt, 1997; Turlan and Chandler, 2000). The 3'OH may attack the 5' end of either strand in non-replicative transposition and is responsible for cleavage of the target site in a staggered fashion, which produces the characteristic direct repeats found at insertion sites (Fig 1.6B) (Craig, 1996; Hallet and Sherratt, 1997; Turlan and Chandler, 2000).

DDE catalysis at the 3' end of the element utilises water as the nucleophile, producing a single strand nick (Craig, 1996; Grindley and Leschziner, 1995; Hallet and Sherratt, 1997; Mizuuchi, 1997; Turlan and Chandler, 2000). During replicative transposition of Tn*3* or Mu, the cleaved 3'OH ends proceed to attack the target site without further cleavage of the donor strand, this produces a replication fork and becomes a substrate for DNA polymerase I (Turlan and Chandler, 2000).

There are several elements that do not usually undergo replicative transposition, including Tn*7*, IS*911*, and Tn*10* (Haren *et al.*, 1999). Tn*7* encodes two transposase proteins, TnsA and TnsB. TnsB binds specifically to the element ends while TnsA binds DNA non-specifically and is brought to the ends of the element by its interaction with TnsB (Fig 1.6C) (Turlan and Chandler, 2000). TnsB cleaves the 3'ends of the element, while TnsA cleaves the 5' ends, introducing a 3 bp overhang, which is removed by host proteins during gap repair of the staggered target site (Haren *et al.*, 1999). The 3'OH ends of the element attack the target site, which has already been recruited to the nucleoprotein complex (Haren *et al.*, 1999; Turlan and Chandler, 2000).

Tn*10* utilises a different transposition mechanism, using a single transposase protein (Turlan and Chandler, 2000). Transposase monomers cleave the 3' ends, which then

Figure 1.6 DDE-mediated transposition

A Co-integrate Formation

Cleavage of the 3' end of the transposon (purple), followed by strand transfer and cleavage of the target molecule (red), results in the fusion of the two replicons (red and blue) and co-integrate formation following host replication (green).

B Target Site Cleavage

The cleaved 3'OH groups still attached to the transposase monomers (red) attack the target site in a staggered fashion. Integration of the cleaved element requires host gap repair (red text), producing the characteristic target site duplication.

C Tn*7* 5' End Treatment

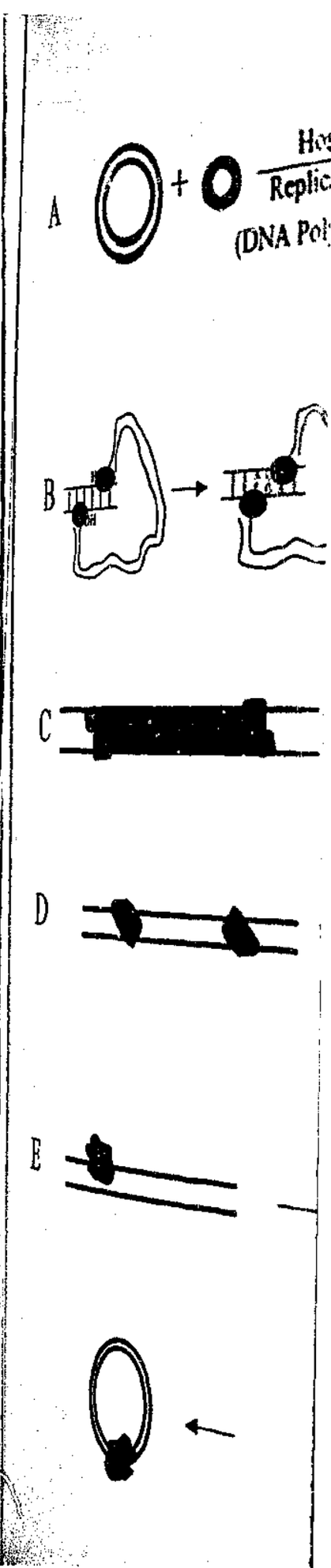
TnsB (red) binds and cleaves the 3' ends of the element, exhibiting sequence specific DNA binding. TnsA (green) is a non-specific DNA binding protein that interacts with TnsB enabling TnsA to bind and cleave the 5' ends of the element (Turlan and Chandler, 2000).

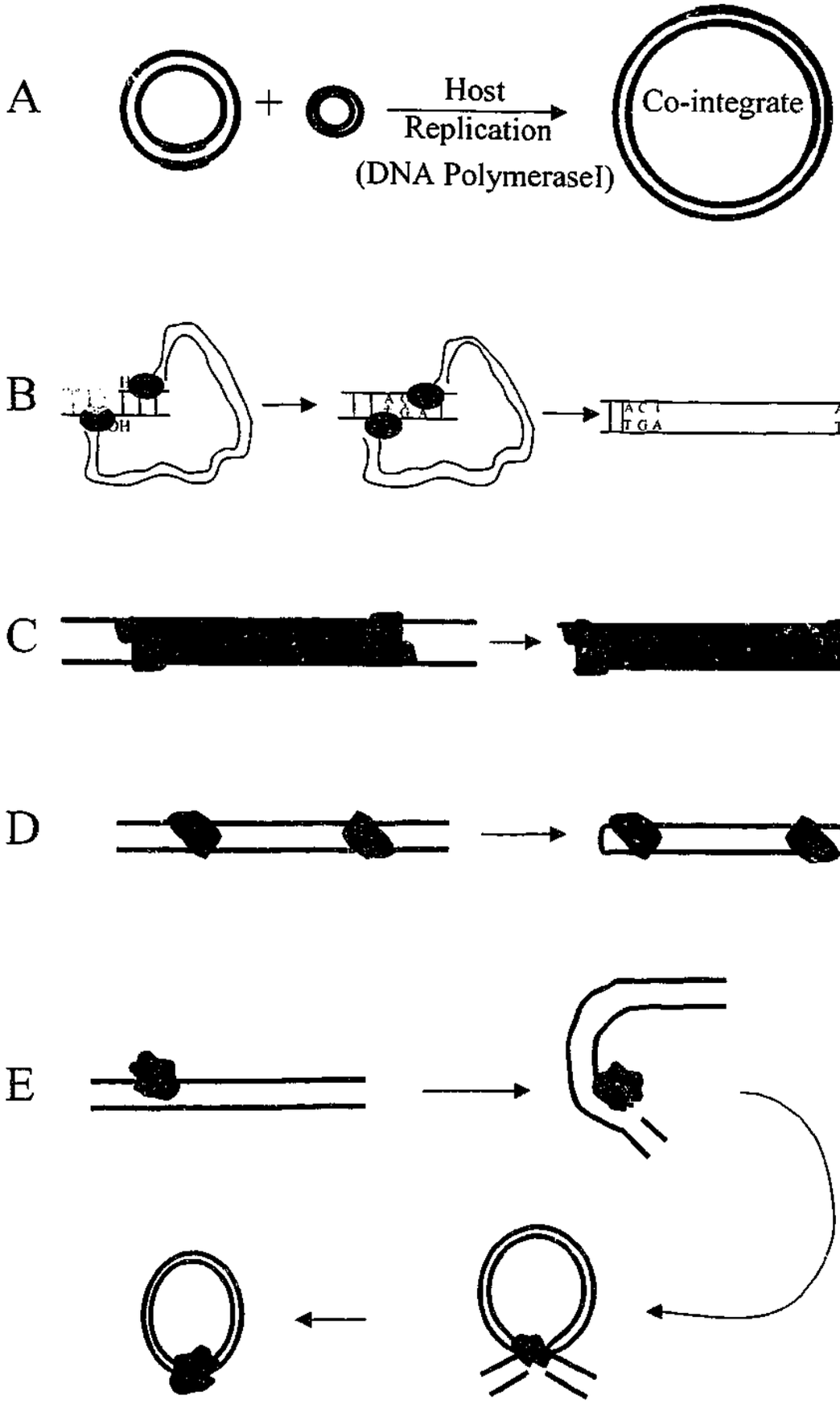
D Tn*10* 5' End Treatment

Tn*10* transposase (purple) binds to the 3' ends and cleaves the DNA. The 3'OH ends that are generated attack the adjacent 5' ends of the opposite strand, producing covalent bonds between the ends of the element (Turlan and Chandler, 2000).

E IS*911* Circle Formation

IS*911* transposase (magenta) binds to the top strand and cleaves the 3' end. The 3'OH is then bent back to attack the 5' end of the same DNA strand, forming a "figure 8" molecule. Host processing is thought to be involved in production of an independent circular molecule (Turlan and Chandler, 2000).




A
+
Host
Replication
(DNA PolymeraseI)
Co-integrate
B
OH
A C T
T G A
A C T
T G A
C
D
E

attack the 5' ends of the opposite strand to produce hairpin structures (Fig 1.6D) (Turlan and Chandler, 2000). The element is completely removed from the donor strand before synapsing with the target strand. The hairpin structures are again cleaved by the transposase and the 3'OH that is produced attacks the synapsed target site, resulting in simple insertion (Turlan and Chandler, 2000).

IS*911* is capable of producing a large range of transposition products (Polard *et al.*, 1994; Ton-Hoang *et al.*, 1999; Turlan *et al.*, 2000), including transposon circles (Ton-Hoang *et al.*, 1997). The IS*911* transposase, IstAB, cleaves the 3' end of the top strand only, which then attacks the 5' end of the same strand, 3 bp from the end of the element, forming a "figure 8" molecule (Fig 1.6E) (Turlan *et al.*, 2000). This structure is processed further, probably by host proteins, to produce a circular intermediate with a strong promoter that spans the joint of the circular molecule. The end result is increased expression of the *istAB* genes (Turlan *et al.*, 2000). The 3'ends of the circle joint are again cleaved and attack the target site with loss of the intervening donor DNA (Turlan *et al.*, 2000). Other elements such as IS*1* and IS*30* are capable of forming circular molecules (Kiss and Olasz, 1999; Shiga *et al.*, 1999). IS*911* circles can also produce linear forms that are capable of simple insertion, although at a lower frequency than intact circles (Ton-Hoang *et al.*, 1999). IS*911* can form tandem IS dimers where the ends of two copies of IS*911* are abutted (Turlan *et al.*, 2000). This structure acts in a similar fashion to a circle joint and produces a cointegrate through simple insertion (Turlan *et al.*, 2000). Other elements shown to form tandem IS dimers include IS*30* and IS*21* (Kiss and Olasz, 1999; Schmid *et al.*, 1998).

In summary, transposase mediated genetic rearrangements are highly variable but all have several common features. The first stage involves transposase-mediated cleavage of the 3' end or ends and the nucleophile is water. The process of strand transfer to the

target and joining of the element to the target site are also mediated by the transposase, with the 3'OH acting as the nucleophile and introducing a staggered break in the DNA, which is repaired by host proteins. The process does not require ATP or other high-energy co-factors and the cleavage process results in the loss of an ester bond (Mizuuchi, 1997). Processing of the 5' ends is either not mediated by the element (replicative transposition) or treated differently by different elements.

b) Site-specific Recombinases

The λ Integrase Family

The integrase family includes a large and diverse set of proteins (Hallet and Sherratt, 1997; Nunes-Düby *et al.*, 1998) with varied functions, including phage genome integration and excision, inversion of DNA segments and resolution of plasmid and chromosome dimers (Hallet and Sherratt, 1997; Nash, 1996). Although diverse, these proteins share two regions of homology known as box I and box II (Nunes-Düby *et al.*, 1998). Within these regions there are four almost invariant residues, RHRY, the signature sequence of the λ integrase family (Fig 1.7) (Gopaul and Duyne, 1999; Grainge and Jayaram, 1999; Hallet and Sherratt, 1997; Nunes-Düby *et al.*, 1998). The tyrosine residue is the nucleophile responsible for strand cleavage and formation of a covalent protein-DNA bond (Crisona *et al.*, 1999; Gopaul and Duyne, 1999; Grainge and Jayaram, 1999; Hallet and Sherratt, 1997; Mizuuchi, 1997; Nunes-Düby *et al.*, 1998). The two arginine residues are thought to activate the phosphodiester bond to be cleaved, by co-ordination in a bipyrimidal pentavalent form for attack by the tyrosine hydroxyl (Fig 1.8) (Mizuuchi, 1997). Note that in addition to these residues, there are other conserved polar and acidic residues flanking the catalytic pocket (Nunes-Düby *et al.*, 1998).

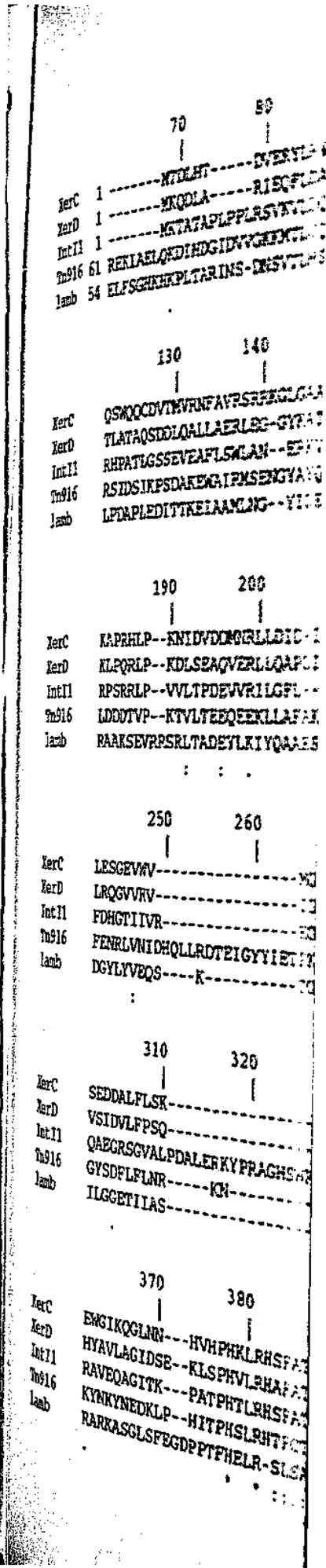
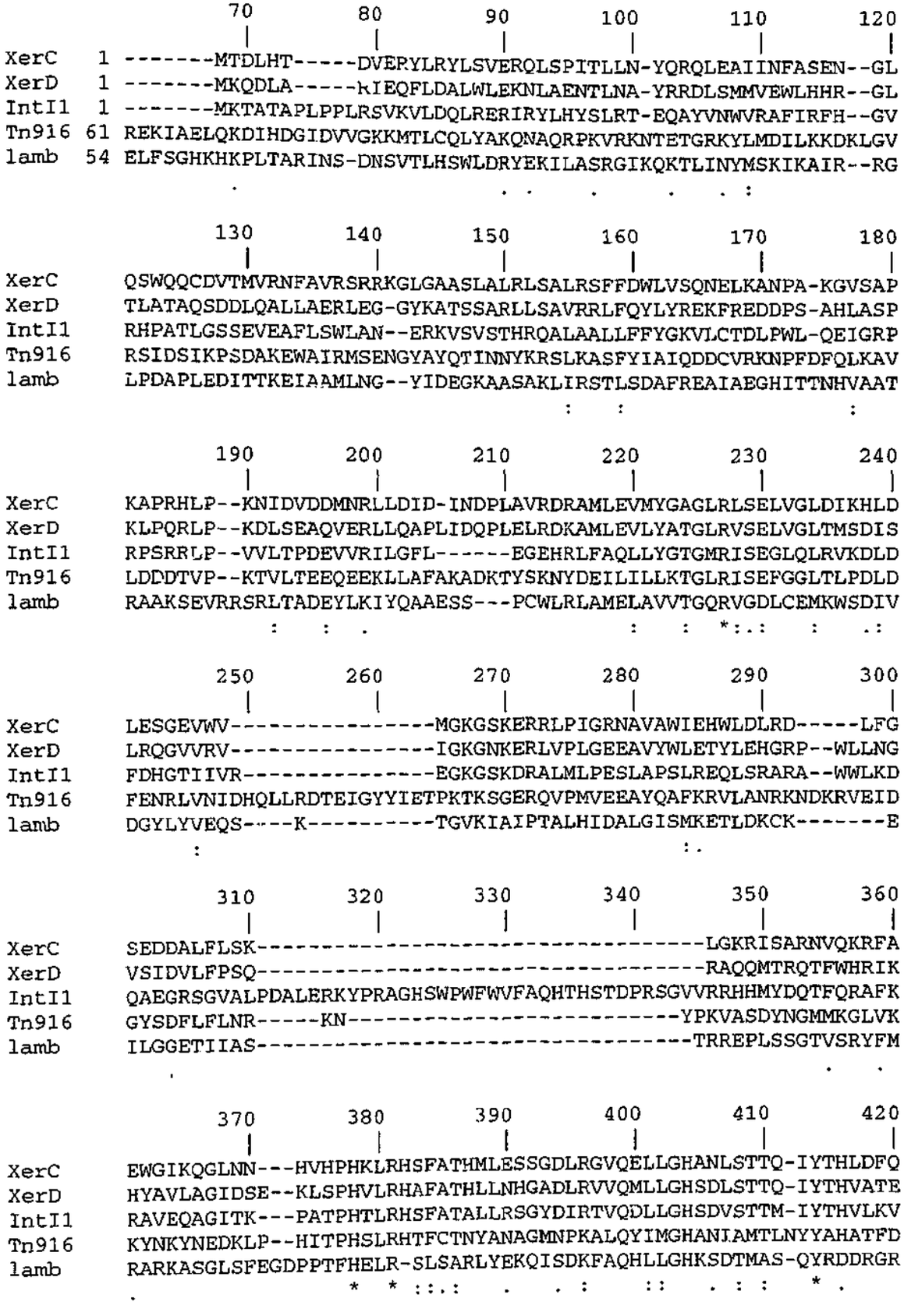


Figure 1.7 Alignment of integrase proteins

An alignment of several integrase family members shows a high degree of variability, but almost without exception, these proteins contain the tetrad RHRY (denoted by red asterisks) (Nunes-Düby et al., 1998). Residues that are conserved are indicated in red; varying degrees of similarity are indicated by two dots for high similarity and one dot for low similarity. The protein sequences are not complete, the alignment shows the area containing the RHRY tetrad and regions in between. The proteins shown in the alignment are λ integrase (designated lamb (Genbank accession no. P03700)), integrase from Tn*916* (designated Tn*916* (P22886), IntI1 integrase from *Klebsiella pneumoniae* (AAM89398) and the *E. coli* integrase proteins XerC (P22885) and XerD (P21891).


70 80 90 100 110 120
XerC 1 -------MTDLHT-----DVERYLRYLSVERQLSPITLLN-YQRQLEAIINFASEN--GL
XerD 1 -------MKQDLA-----KIEQFLDALWLEKNLAENTLNA-YRRDLSMMVEWLHHR--GL
IntI1 1 -------MKTATAPLPPLRSVKVLDQLRERIRYLHYSLRT-EQAYVNWVRAFIRFH--GV
Tn916 61 REKIAELQKDIHDGIDVVGKKMTLCQLYAKQNAQRPKVRKNTETGRKYLMDILKKDKLGV
lamb 54 ELFSGHKHKPLTARINS-DNSVTLHSWLDRYEKILASRGIKQKTLINYMSKIKAIR--RG
130 140 150 160 170 180
XerC QSWQQCDVTMVRNFAVRSRRKGLGAASLALRLSALRSFFDWLVSQNELKANPA-KGVSAP
XerD TLATAQSDDLQALLAERLEG-GYKATSSARLLSAVRRLFQYLYREKFREDDPS-AHLASP
IntI1 RHPATLGSSEVEAFLSWLAN--ERKVSVSTHRQALAALLFFYGKVLCTDLPWL-QEIGRP
Tn916 RSIDSIKPSDAKEWAIRMSENGYAYQTINNYKRSLKASFYIAIQDDCVRKNPFDFQLKAV
lamb LPDAPLEDITTKEIAAMLNG--YIDEGKAASAKLIRSTLSDAFREAIAEGHITTNHVAAT
190 200 210 220 230 240
XerC KAPRHLP--KNIDVDDMNRLLDID-INDPLAVRDRAMLEVMYGAGLRLSELVGLDIKHLD
XerD KLPQRLP--KDLSEAQVERLLQAPLIDQPLELRDKAMLEVLYATGLRVSELVGLTMSDIS
IntI1 RPSRRLP--VVLTPDEVVRILGFL------EGEHRLFAQLLYGTGMRISEGLQLRVKDLD
Tn916 LDDDTVP--KTVLTEEQEEKLLAFAKADKTYSKNYDEILILLKTGLRISEFGGLTLPDLD
lamb RAAKSEVRRSRLTADEYLKIYQAAESS---PCWLRLAMELAVVTGQRVGDLCEMKWSDIV
250 260 270 280 290 300
XerC LESGEVWV----------MGKGSKERRLPIGRNAVAWIEHWLDLRD-----LFG
XerD LRQGVVRV----------IGKGNKERLVPLGEEAVYWLETYLEHGRP--WLLNG
IntI1 FDHGTIIVR----------EGKGSKDRALMLPESLAPSLREQLSRARA--WWLKD
Tn916 FENRLVNIDHQLLRDTEIGYYIETPKTKSGERQVPMVEEAYQAFKRVLANRKNDKRVEID
lamb DGYLYVEQS----K----------TGVKIAIPTALHIDALGISMKETLDKCK-------E
310 320 330 340 350 360
XerC SEDDALFLSK----------LGKRISARNVQKRFA
XerD VSIDVLFPSQ----------RAQQMTRQTFWHRIK
IntI1 QAEGRSGVALPDALERKYPRAGHSWPWFWVFAQHTHSTDPRSGVVRRHHMYDQTFQRAFK
Tn916 GYSDFLFLNR-----KN----------YPKVASDYNGMMKGLVK
lamb ILGGETIIAS----------TRREPLSSGTVSRYFM
370 380 390 400 410 420
XerC EWGIKQGLNN---HVHPHKLRHSFATHMLESSGDLRGVQELLGHANLSTTQ-IYTHLDFQ
XerD HYAVLAGIDSE--KLSPHVLRHAFATHLLNHGADLRVVQMLLGHSDLSTTQ-IYTHVATE
IntI1 RAVEQAGITK---PATPHTLRHSFATALLRSGYDIRTVQDLLGHSDVSTTM-IYTHVLKV
Tn916 KYNKYNEDKLP--HITPHSLRHTFCTNYANAGMNPKALQYIMGHANIAMTLNYYAHATFD
lamb RARKASGLSFEGDPPTFHELR-SLSARLYEKQISDKFAQHLLGHKSDTMAS-QYRDDRGR

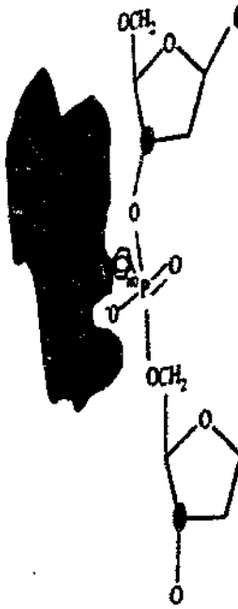
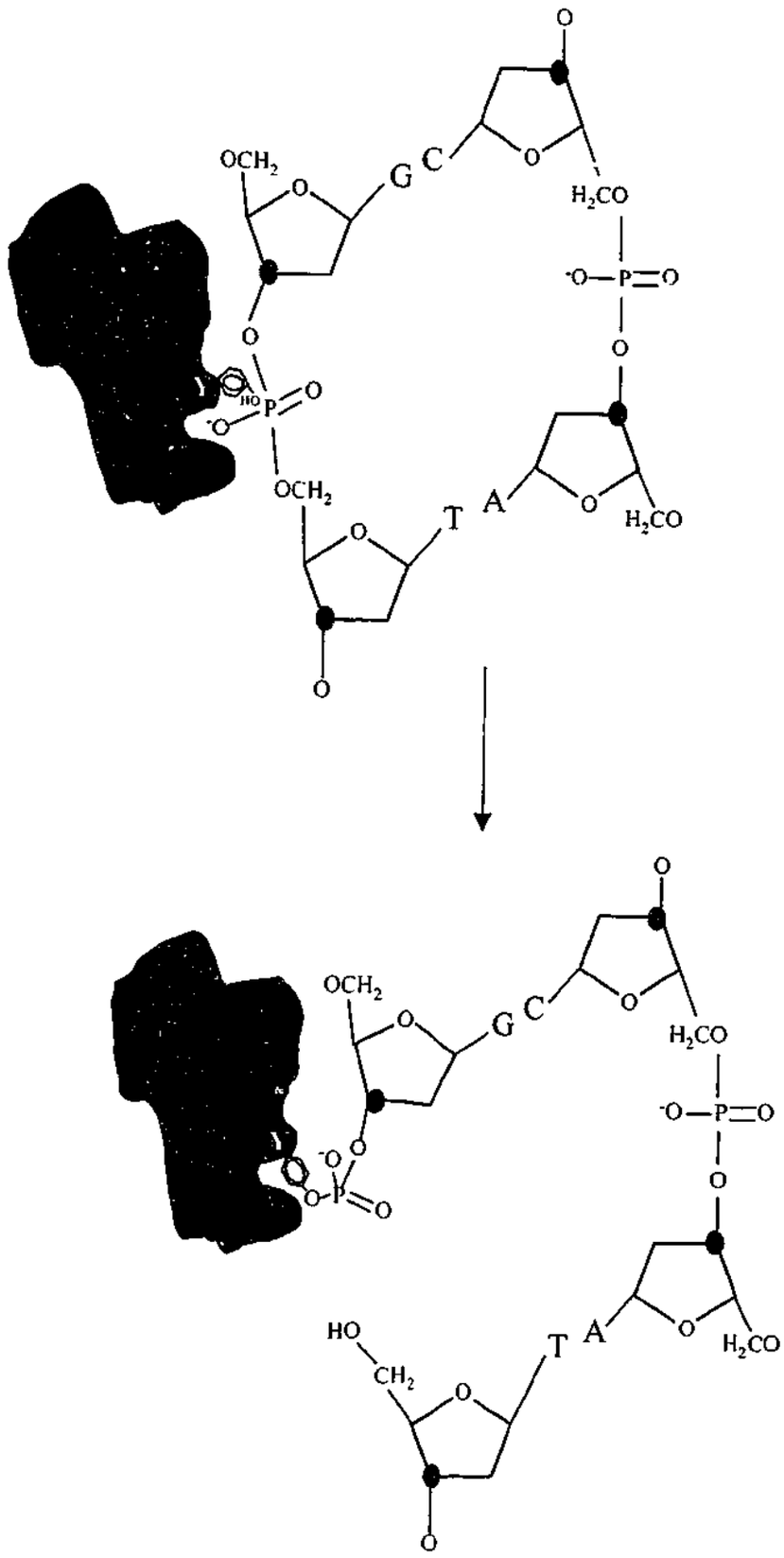


Figure 1.8 Mechanism of integrase-mediated site-specific recombination

The integrase protein (green) binds the DNA (red C's represent 5' carbons, blue circles represent 3' carbons). The tyrosine hydroxyl (magenta text) attacks the phosphate bond resulting in a 3' phosphotyrosyl covalent link and a 5'OH group (Hallet and Sherratt, 1997).


O
OCH2
G
C
H2CO
-O—P=O
O
HO
P=O
-O
OCH2
T
A
H2CO
O
-O
HO
CH2
T
A
H2CO
O

Site-specific recombination reactions do not require ATP or high energy co-factors due to the conservation of the reaction energy through the protein-DNA link (Hallet and Sherratt, 1997; Mizuuchi, 1997). Site-specific recombination also does not involve DNA replication, since restoration of the recombination site sequences occurs as part of the recombination process, a feature that sets this type of reaction apart from the DDE transposase mechanisms (Gopaul and Duyne, 1999; Hallet and Sherratt, 1997; Mizuuchi, 1997; Nash, 1996; Stark *et al.*, 1992). The use of a hydroxyl group from a suitable amino acid as the nucleophile responsible for catalysis, rather than hydrolysis, also maintains the number of ester bonds on the phosphate group, unlike the DDE transposases (Figs 1.5 & 1.7) (Mizuuchi, 1997). Another key factor in the site-specific recombination process is the importance of the orientation of the binding sites, as orientation dictates the type of reaction that will take place (Crisona *et al.*, 1999; Nash, 1996).

Recombination mediated by the integrase (Int) from bacteriophage λ promotes the integration and excision of the phage from the *E. coli* chromosome at a specific site, *attB* (Crisona *et al.*, 1999). The reactions involve multiple binding sites for both the Int protein, and other accessory proteins, both host and phage encoded (Cho *et al.*, 2000; Crisona *et al.*, 1999). The accessory proteins, as well as Int, are involved in the formation of the synaptic complex, which bring together the DNA strands and the proteins required for catalysis, in the correct topological structure (Gopaul and Duyne, 1999; Hallet and Sherratt, 1997). Synaptic complex formation also serves to allow regulation of the processes, particularly as excision can only occur if the phage genome has been integrated and there is an additional protein, Xis, which is only required for excision and inhibits the integration reaction (Cho *et al.*, 2000). This regulation process is important for the phage life-cycle and is tightly controlled (Cho *et al.*, 2000).

Catalysis is carried out by four Int proteins bound to recombination half-sites around a 6-8 bp core DNA sequence (Fig 1.9) (Crisona *et al.*, 1999; Gopaul and Duyne, 1999; Hallet and Sherratt, 1997). The unique feature of the tyrosine recombination system is the two step process of strand exchange (Gopaul and Duyne, 1999; Hallet and Sherratt, 1997; Mizuuchi, 1997). The reaction proceeds after synapsis and cleavage of only two of the four DNA strands followed by re-joining of the two strands to their recombinant partners, leading to the formation of a Holliday junction (Fig 1.9) (Gopaul and Duyne, 1999; Hallet and Sherratt, 1997; Mizuuchi, 1997). The Holliday junction intermediate can only be formed if the DNA is bent to facilitate strand exchange (Gopaul and Duyne, 1999). The next step involves the resolution of the Holliday junction and occurs when the junction is isomerized (Gopaul and Duyne, 1999). Isomerization bends the other two strands, thereby allowing nucleophilic attack by the corresponding tyrosine residues and the subsequent second round of strand exchange (Fig 1.9) (Gopaul and Duyne, 1999; Hallet and Sherratt, 1997; Mizuuchi, 1997; Nash, 1996).

The resolvase/invertase Family

The resolvase/invertase family is divided into four subgroups according to function; namely the invertases, resolvase-invertases, resolvases and the large resolvases (Smith and Thorpe, 2002). The catalytic mechanism is the same although the formation of a synaptic complex in each subgroup differs in both structure and composition, leading to alternative reaction products (Alonso *et al.*, 1996). The reaction mechanism and the structure of the synapse assembled by these enzymes leads to both topological and energetic restrictions on the reactions that can be catalysed, limiting them to intramolecular recombination (with the exception of the large resolvases), the outcome of which is dependent on the orientation of the recombination sites and the enzyme involved (Mizuuchi, 1997; Nash, 1996). The small resolvases and the invertases can only catalyse resolution and inversion reactions respectively, while the large resolvases

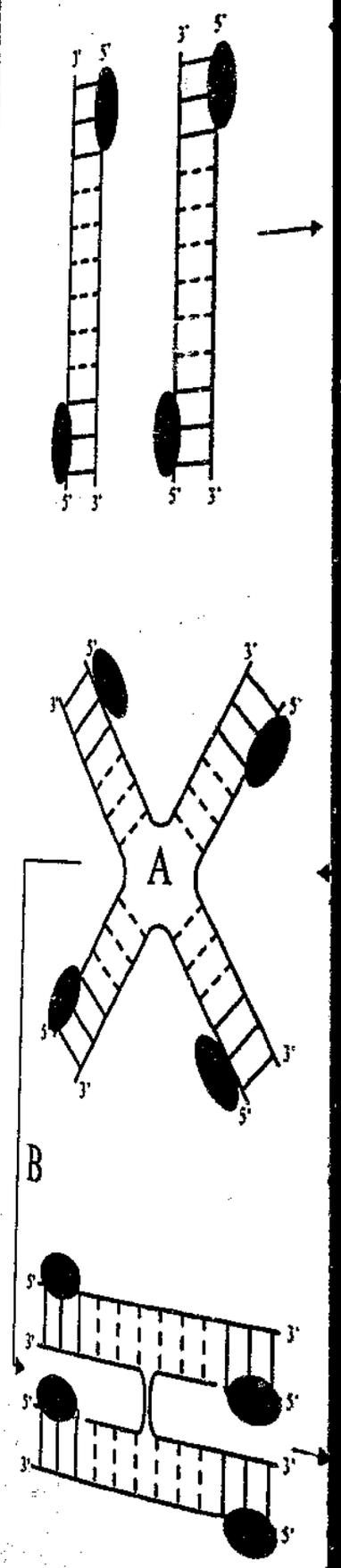
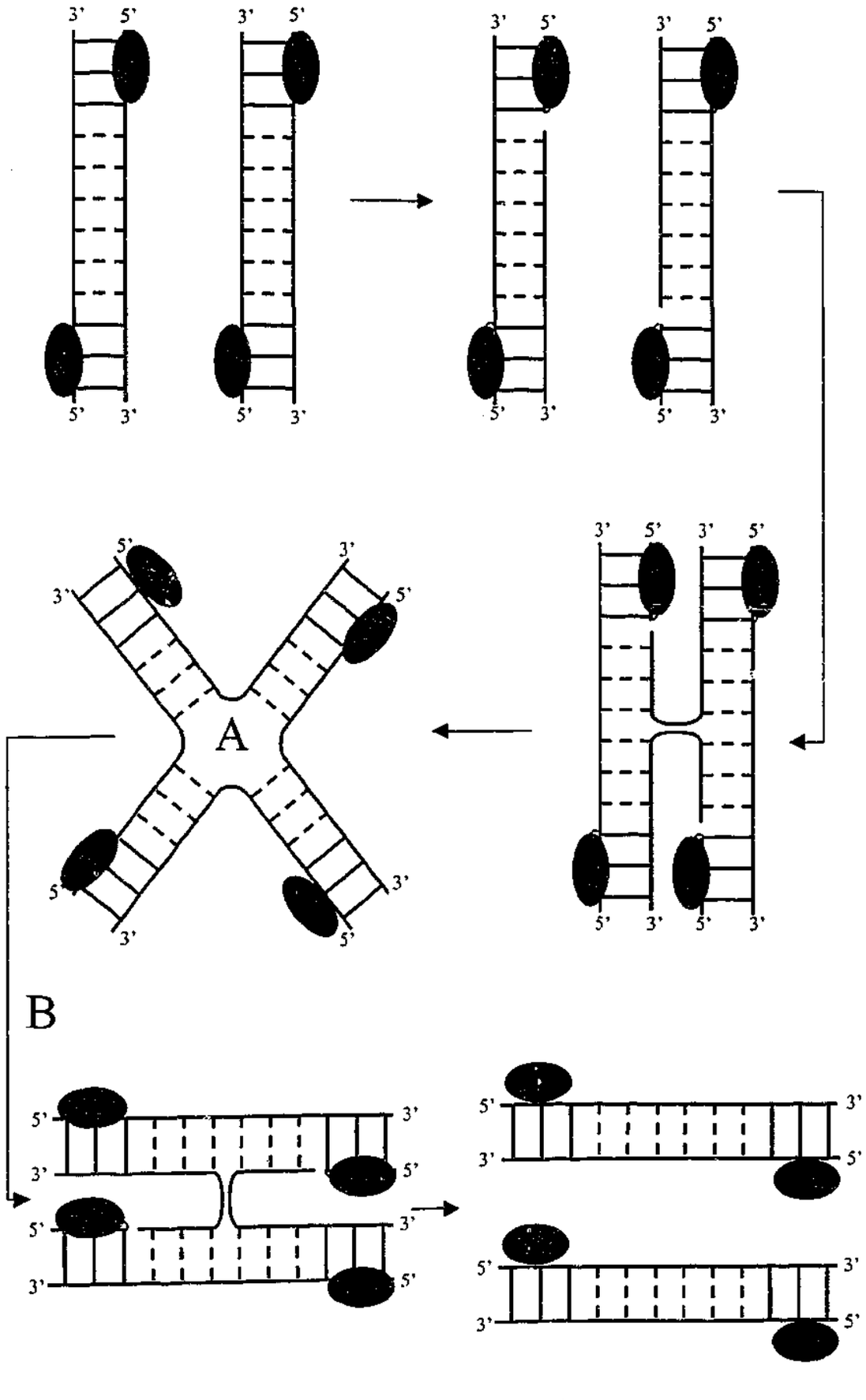


Figure 1.9 Integrase strand transfer

Integrase (green) binds to the target DNA. One integrase monomer from each strand cleaves the sugar-phosphate backbone (Fig 1.8) and strand exchange occurs between these two overhangs (black dashed lines). The DNA strands are re-ligated to the opposite strand leading to the formation of the Holliday junction intermediate, denoted A. Isomerisation (B) brings the next two strands to be exchanged into close proximity, second strand cleavage and exchange occurs and the resultant strands are re-joined. Adapted from Hallet and Sherratt (1997)


3' 5'
5' 3'
A
B

are capable of catalysing both excision and integration (Hallet and Sherratt, 1997; Nash, 1996; Smith and Thorpe, 2002; Stark *et al.*, 1992). The resolvase/invertase family shares common features with the integrase family, as no high energy co-factors are required, no DNA synthesis is involved, and the number of ester bonds is maintained throughout the reaction (Mizuuchi, 1997).

The majority of resolvase/invertase family members show significant sequence similarity (Alonso *et al.*, 1996; Smith and Thorpe, 2002). The domain structure of these proteins is reasonably simple, consisting of a large N-terminal domain of approximately 140 aa and a small C-terminal domain of about 40 aa (Smith and Thorpe, 2002). The N-terminal domain contains highly conserved, charged residues thought to play a part in catalysis, as well as several conserved regions involved in protein-protein interactions, including dimerization and multimerization (Fig 1.10) (Murley and Grindley, 1998; Oram *et al.*, 1995). The C-terminal domain is responsible for sequence specific DNA binding via a helix-turn-helix motif (Smith and Thorpe, 2002). Resolvase proteins typically range in size from 180 aa to 200 aa, with a corresponding molecular size of approximately 20 kDa.

The common mechanism employed by these enzymes involves DNA cleavage that is mediated by nucleophilic attack on the phosphate backbone by the hydroxyl side chain of a conserved serine residue, located at around residue 10 of these proteins (Fig 1.10) (Alonso *et al.*, 1996; Hallet and Sherratt, 1997; Oram *et al.*, 1995; Smith and Thorpe, 2002; Stark *et al.*, 1992). DNA cleavage (Fig 1.11) results in a covalent phosphoserine linkage between the enzyme and the 5' end of the cleaved DNA strand (Hallet and Sherratt, 1997; Stark *et al.*, 1992). Simultaneous cleavage of the complimentary DNA strand leads to 2 bp 3'OH overhangs with protein-DNA links at their 5' ends (Fig 1.11B) (Hallet and Sherratt, 1997; Stark *et al.*, 1992).

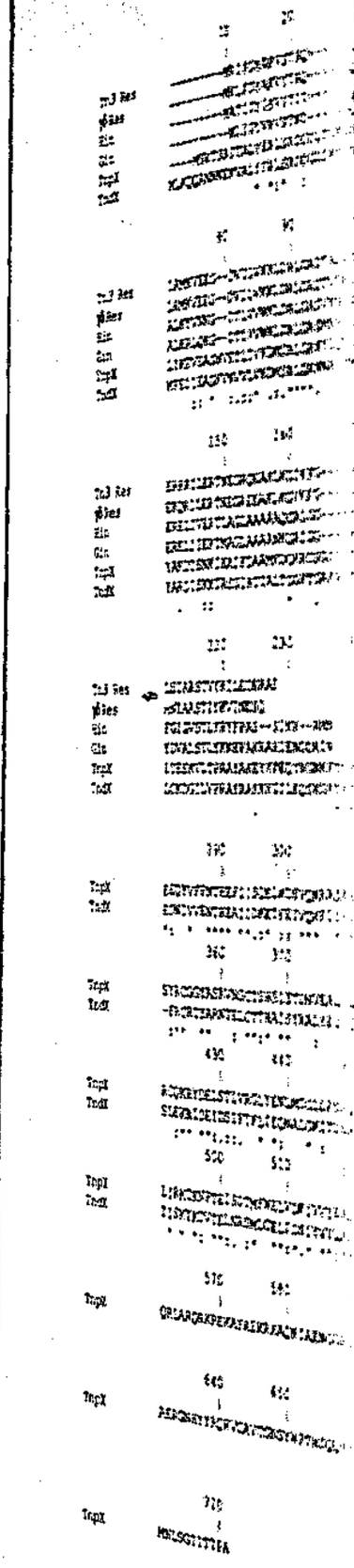
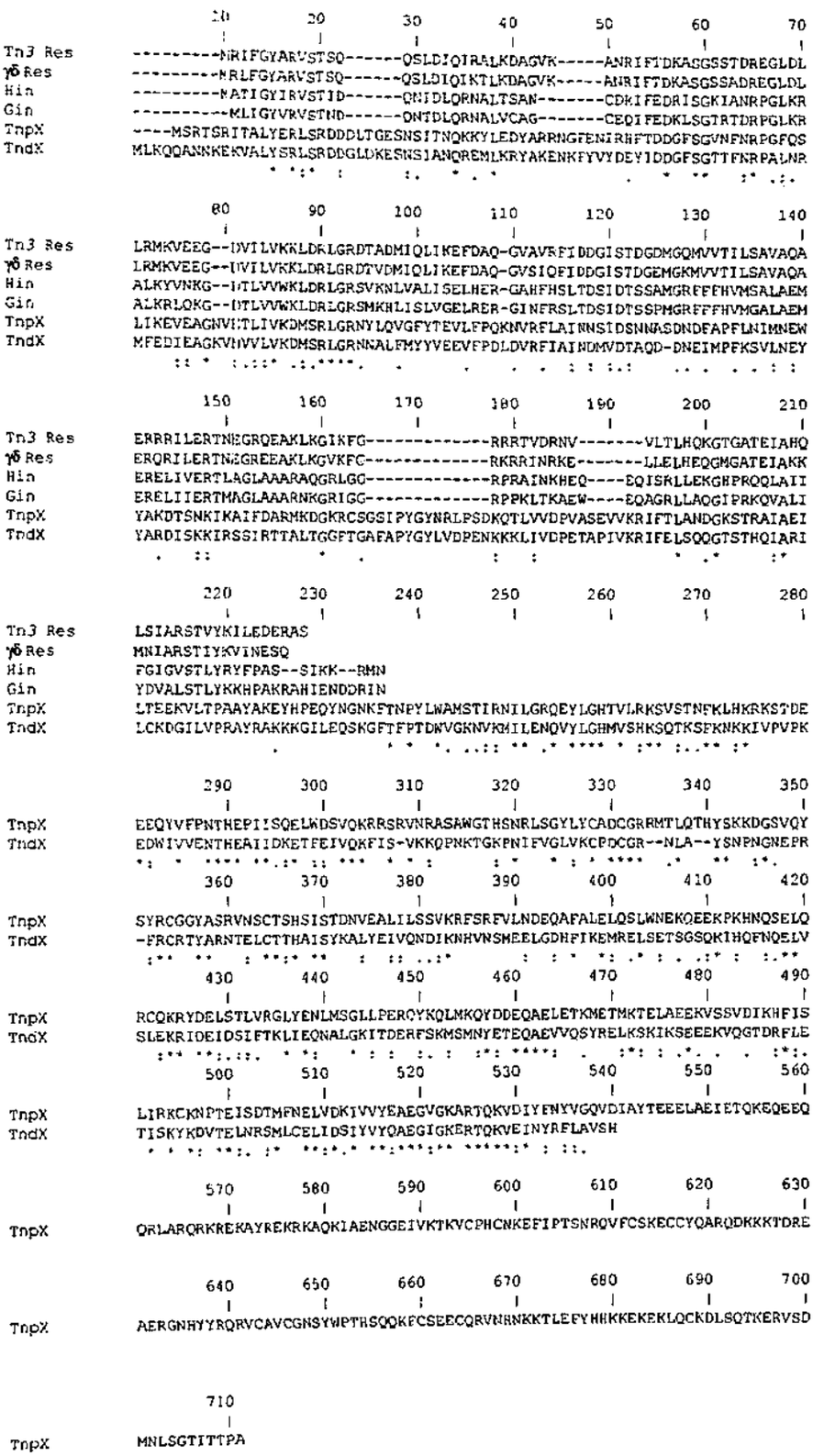


Figure 1.10 Alignment of members of the resolvase/invertase family

An alignment of six members of the resolvase/invertase family is shown. The protein sequences aligned are Tn*3* resolvase (Genbank accession no. P03011) and γδ resolvase (P03012), the invertase proteins Gin (NP050655) and Hin (P03013), and two large resolvase proteins, TnpX (S78538) and TndX (AAF35174). Conserved residues are shown in red. Asterisks denote complete conservation, two dots indicate high conservation and one dot indicates similar residues. The serine at position 19 in this alignment is the catalytic residue for this family.


Tn3 Res
γδ Res
Hin
Gin
TnpX
TndX

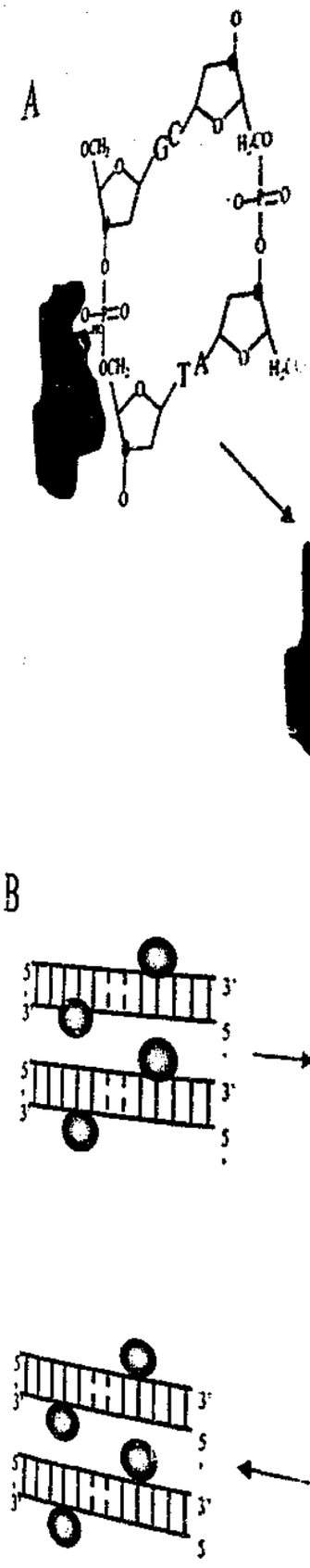


Figure 1.11 Mechanism of resolvase/invertase-mediated site-specific recombination

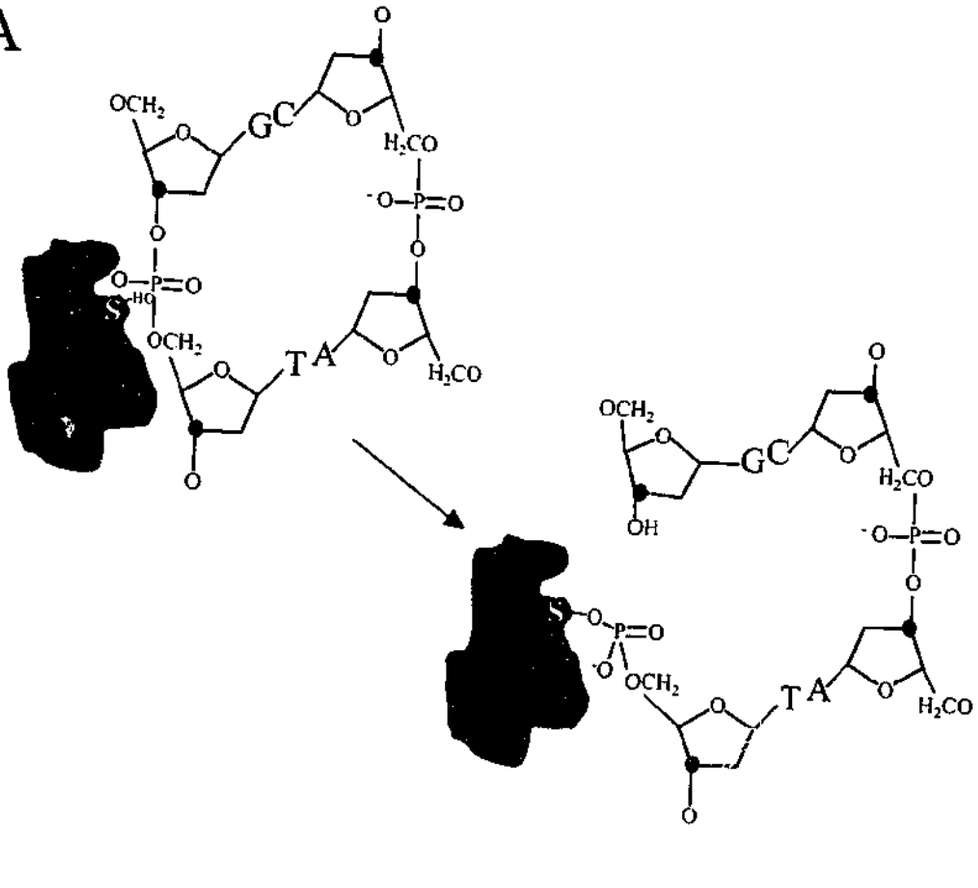
A Cleavage Reaction

The resolvase protein (purple) binds the DNA (red C's represent 5' carbons and blue circles represent 3' carbons). The serine (white) hydroxyl (magenta) group attacks the phosphate bond resulting in the formation of a covalent bond at the 5' end and the production of a 3'OH.

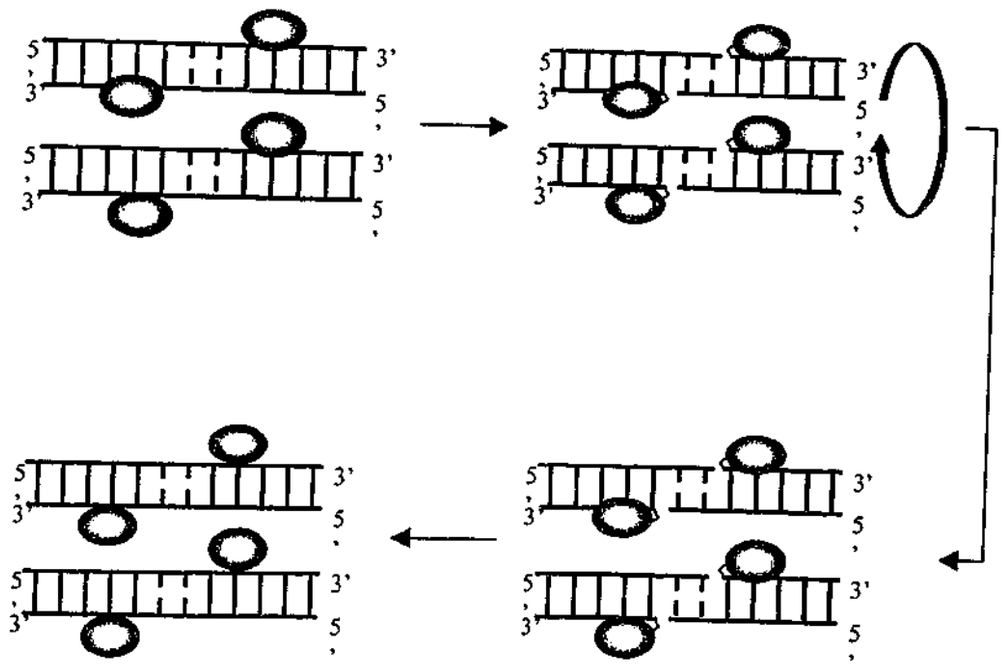
B Strand Transfer Reaction

Resolvase dimers (yellow) bind to the *res* site (only site I is shown) on the DNA. All four DNA strands are cleaved at once, with the production of four phosphoserine links (small black lines). One halfsite is rotated 180° (circular arrow) with respect to the other and the sugar-phosphate backbone is re-joined to form the recombinant. Based on Hallet and Sherratt (1997).

A



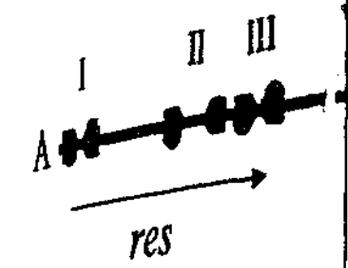
B



In contrast to strand exchange catalysed by the λ Int family of recombinases, both DNA strands are transferred simultaneously, by a 180° right-handed rotation of one set of strands with respect to the other (Fig 1.11B) (Hallet and Sherratt, 1997; Mizuuchi, 1997; Nash, 1996; Stark *et al.*, 1992). This process involves simultaneous cleavage of all four DNA strands, and is accomplished by four recombinase proteins, bound as dimers at the site of catalysis or core site (Mizuuchi, 1997; Stark *et al.*, 1992). Resolvase/invertase proteins require the DNA substrate to be supercoiled, which aids in the formation of the synapse and may also contribute to the rotation of the DNA strands during strand transfer (Murley and Grindley, 1998; Nash, 1996). The completion of the reaction involves the rejoining of the DNA molecules and is accomplished through nucleophilic attack by the 3'OH of the recombinant strand (Fig 1.11B), which results in the release of the resolvase monomers and resealing of the DNA backbone (Hallet and Sherratt, 1997; Mizuuchi, 1997).

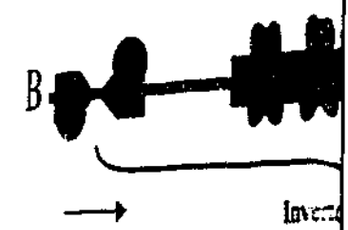
Resolvase proteins catalyse the resolution of cointegrates produced by replicative transposition (Hallet and Sherratt, 1997). They are encoded by mobile elements, such as members of the Tn*3* family of transposons, and show extensive sequence identity (Smith and Thorpe, 2002). Tn*3* and γδ resolvase are functionally interchangeable, as are the resolvase proteins from Tn*21* and Tn*1721* (Halford *et al.*, 1985). Resolvases do not require accessory proteins to be functional (Alonso *et al.*, 1996). The composition of the resolvase binding site, *res*, is quite complex, consisting of three subsites – site I, site II and site III (Fig 1.12A) (Alonso *et al.*, 1996; Hallet and Sherratt, 1997; Nash, 1996; Oram *et al.*, 1995). Each subsite consists of a dimer binding region in the form of 12 bp imperfect repeats, with a variable number of base pairs between the repeats (Alonso *et al.*, 1996; Hallet and Sherratt, 1997; Nash, 1996; Oram *et al.*, 1995). The distance between subsites also varies and is important for the resolution reaction (Kilbride *et al.*,

Figure 1.12 Resolvase/invertase recombination sites



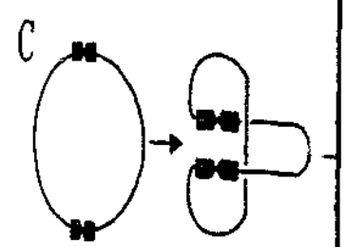
A The resolvase binding or *res* site

The *res* site consists of three inverted repeats showing variable spacing. Resolvase proteins (light blue ovals) bind to the repeats as dimers. Recombination will only occur if the *res* sites are directly repeated and strand transfer occurs between the repeats at site I (pink and green arrows) (Alonso et al., 1996). Blue arrows represent the inverted repeats at site II and the purple arrows represent the inverted repeats forming site III. The thin purple arrows indicate the orientation of the binding sites (Alonso et al., 1996).



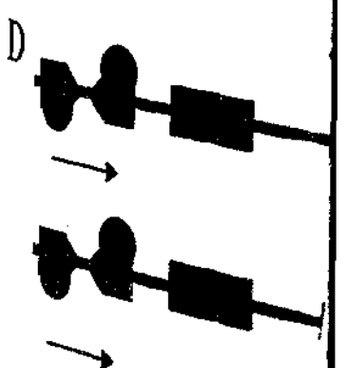
B Invertase Binding Site

Organisation of the invertase binding site. The blue ovals represent invertase subunits, purple represent Fis bound to the enhancer site. Recombination occurs between the half sites (pink and green arrows). The thin purple arrows indicate the orientation of the binding sites (Alonso et al., 1996).



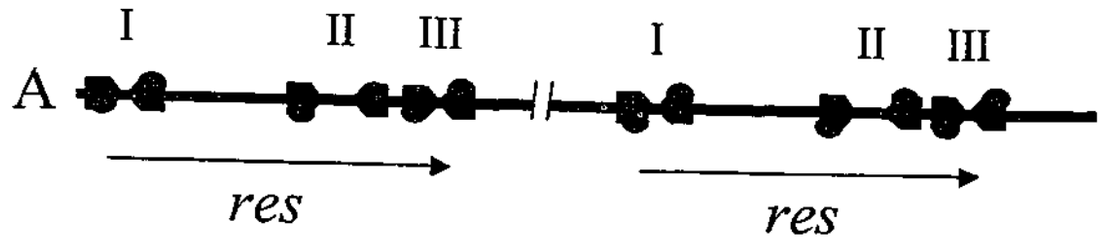
C Invertasome Formation

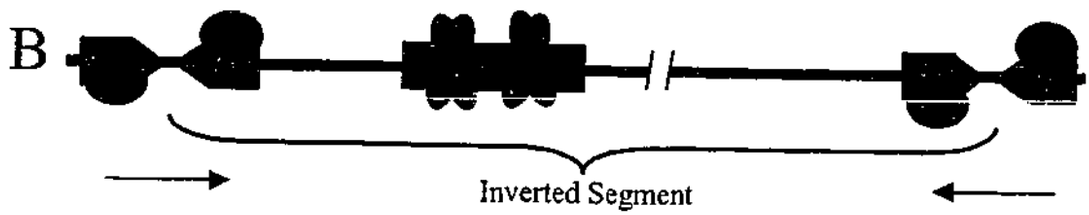
Synapse involving invertase and its binding sites results in the formation of an invertasome. Strand transfer between the recombination half sites (pink and green arrows) results in production of an inverted DNA segment (Stark et al., 1992).

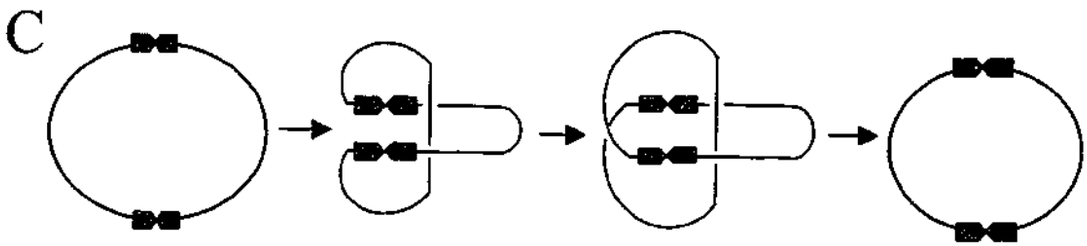


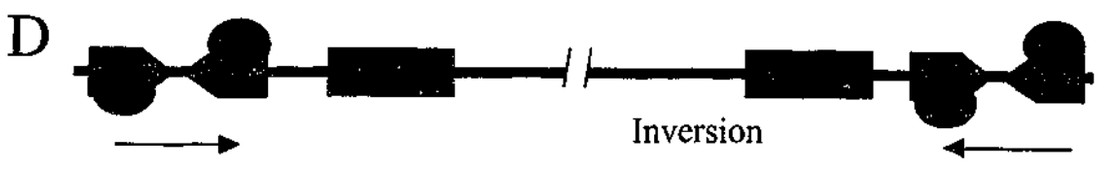
D Resolvase-invertase Binding Site

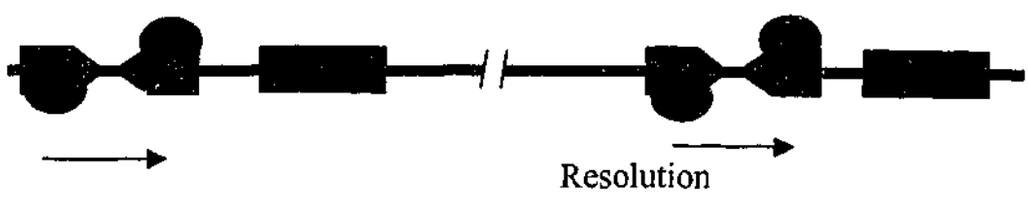
Whether resolution or inversion is catalysed by resolvase-invertase proteins depends on the orientation of the binding sites (Alonso et al., 1996). If in an inverted orientation (top diagram), inversion occurs, if directly repeated (bottom diagram), resolution occurs. The purple rectangle indicates a low affinity resolvase-invertase binding site that isn't an inverted repeat (Alonso et al., 1996). Purple arrows indicate the orientation of binding sites and blue ovals represent resolvase-invertase proteins.


A
I
II
III
I
II
III
res
res


B
Inverted Segment


C


D
Inversion


Resolution

1999). Site I is where DNA cleavage and strand exchange occurs, at a dinucleotide between the dimer binding regions (Stark *et al.*, 1991). The function of sites II and III is not clear but they are thought to be responsible for both synapse assembly and activation of the dimers bound at site I (Kilbride *et al.*, 1999). The accessory sites are essential for both catalysis and topological control of the reaction, by dictating the structure of the synaptic complex (Kilbride *et al.*, 1999).

The synaptic structure involves interactions between monomers bound at the accessory sites as well as those found at the catalytic site (Kilbride *et al.*, 1999) and is dependent on supercoiling for proper formation (Benjamin and Cozzarelli, 1988; Stark *et al.*, 1991). The resolvase synaptic complex traps three negative supercoils within its structure which leads to a build up of pressure as the coils are trapped within a relatively short DNA sequence (Benjamin and Cozzarelli, 1988). It is thought that this tension may drive the rotation of the DNA sequences to be exchanged and is consistent with a 180°, right-handed rotation (Fig 1.13) (Benjamin *et al.*, 1996; Hallet and Sherratt, 1997; Mizuuchi, 1997; Nash, 1996; Stark *et al.*, 1991). The resolvase reaction produces two circles linked as a simple, 2-noded catenane (Fig 1.13) with a corresponding loss of supercoiling (Benjamin *et al.*, 1996; Stark *et al.*, 1991). The sequence of the dinucleotides to be exchanged must be identical, otherwise a further 180° rotation in the same direction occurs to return the sequences to their non-recombinant partners (Stark *et al.*, 1991). This reaction can be detected as it produces knotted products, depending on the number of rotations (Fig 1.13) (Stark *et al.*, 1991).

Members of the invertase subgroup catalyse the inversion of defined DNA segments and are often involved in the regulation of alternative gene products, such as the expression of flagella or phage tail fibres (Nanassy and Hughes, 1998; Pato, 1989). Consequently, invertases are often chromosomally encoded, although the Gin invertase

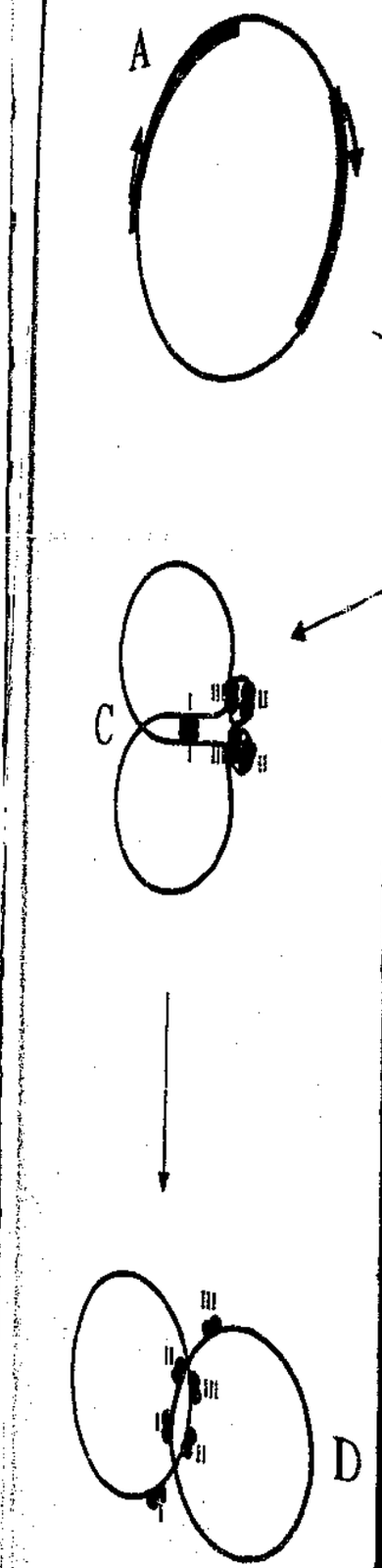
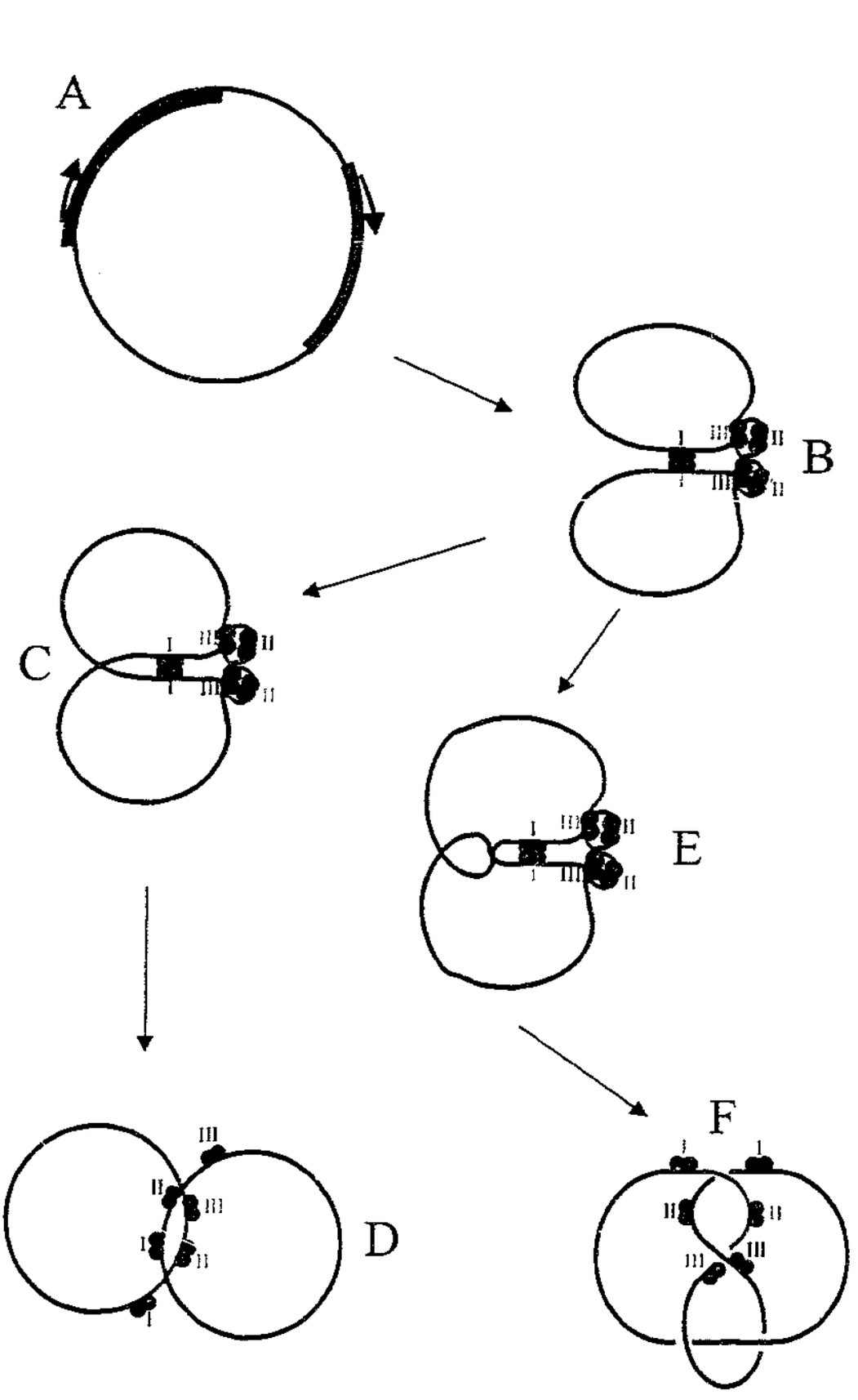


Figure 1.13 Model of the resolvase reaction

Purple and green rectangles or lines represent two directly repeated *res* sites. Small circles represent resolvase proteins, colour coded depending on the *res* site with which they are associated (purple or green). Sites are numbered in roman numerals and colour coded depending on the respective *res* site. Diagram A shows the starting substrate for resolvase activity, a co-integrate and the blue arrows indicate the orientation of the *res* sites. A synaptic complex is formed (B), resulting in the alignment of the site I regions for strand transfer. When the cleavage of the DNA is followed by a 180° rotation, structure C results (hybrid site Is are denoted by a black I), which leads to the formation of a singly-linked, 2-noded catenane (D). If structure B contains non-identical nucleotides at the 2 bp overlap region, synaptic complex formation will still occur (B). Cleavage will also occur but will not result in a simple 180° rotation (C), but will continue going around in the same direction (360°), restoring the non-recombinant arrangement (E). This produces a circular molecule with a twist - a 4-noded knot (F). Further rotation is possible if the synaptic complex remains stable - 720°, 1080°, etc, resulting in 8- and 12-noded knots respectively. Adapted from Stark *et al* (1992).


A
B
C
D
E
F
I
II
III

is encoded by the phage Mu (Pato, 1989). The organization of the recombination site is very important, inversion will only occur if the recombination sites are present on the same supercoiled DNA molecule in an inverted orientation (Alonso *et al.*, 1996; Haykinson *et al.*, 1996; Nanassy and Hughes, 1998; Nanassy and Hughes, 2001). The recombination sites consist of two monomer binding sites of 12 bp each, which flank the central dinucleotide where cleavage occurs (Fig 1.12B) (Alonso *et al.*, 1996; Nanassy and Hughes, 1998). Synapsis of the two recombination sites involves an additional site, known as the enhancer site, which binds two Fis dimers to form a productive invertasome (Alonso *et al.*, 1996; Nanassy and Hughes, 2001). Within this structure strand cleavage, transfer and rejoining occurs (Nanassy and Hughes, 2001). The invertasome structure traps two nodes (Benjamin *et al.*, 1996) and produces a free circular product with the loss of four negative supercoils (Fig 1.12C) (Nanassy and Hughes, 2001). The energy in the supercoiled DNA drives the reaction (Nanassy and Hughes, 2001). The strict organization of the synapse leads to the specificity of the substrate required and dictates what products are formed (Nash, 1996).

Interaction between the enhancer bound Fis proteins and the recombinase is thought to activate the complex, allowing cleavage of the DNA (Nanassy and Hughes, 2001). Another protein involved in the reaction, particularly if the enhancer site is very close to one of the recombination sites, is HU, a non-specific DNA binding protein that introduces bends to the DNA (Nanassy and Hughes, 2001). The organization of the recombination and the enhancer sites leads to a strict topological restraint on the reaction, as only inverted recombination sites on the same DNA molecule produce a productive synapse. Control of recombination in this manner is important, as unconstrained inversion, such as intermolecular reactions, would be highly detrimental to the host (Nash, 1996).

The resolvase-invertase subgroup shows significant similarity to both resolvase and invertase proteins but members are capable of both inversion and resolution, depending on the orientation of the recombination sites (Alonso *et al.*, 1996). Resolvase-invertase proteins are involved in resolution of plasmid multimers and in the inversion of replication origins of theta-replicating plasmids from Gram positive bacteria (Alonso *et al.*, 1996; Canosa *et al.*, 1997; Janière *et al.*, 1996; Petit *et al.*, 1995). The recombination site includes a dimer binding site of two inverted repeats on either side of the cleavage site, known as site I, and a second site, site II, to which the recombinase exhibits low affinity binding (Fig 1.12D) (Canosa *et al.*, 1997; Petit *et al.*, 1995). Recombination also requires another protein such as HU, or Hbsu from *Bacillus subtilis*, which does not bind to the recombination site (Alonso *et al.*, 1996; Janière *et al.*, 1996; Petit *et al.*, 1995).

The large resolvase proteins represent a relatively new subgroup within the resolvase/invertase family. The members of this subgroup are significantly larger than any other resolvase/invertase members, some with primary sequences in excess of 700 aa (Smith and Thorpe, 2002). Many new members have been identified from genome sequencing projects and therefore are of unknown function. The better studied members of this group exhibit a greater variety of reactions than normal resolvases, as some are capable of both excision (resolution) and integration (intermolecular recombination) (Chritiansen *et al.*, 1996; Matsuura *et al.*, 1996; Smith and Thorpe, 2002; Thorpe and Smith, 1998; Wang and Mullany, 2000). Table 1 summarises the members of this group, including their size and function if known.

Large resolvases that have been studied include XisF, SpoIVCA, the Int proteins from phages TP901-1, ϕC31, ϕR4, and the clostridial large resolvases, TnpX and TndX. XisF is responsible for the excision of a 55 kb segment of DNA from developing

Table 1.1 Large resolvase proteins

Name	Element	Organism	Function	Length	Accession Number
SCC88.14		*Streptomyces coelicolor*	Unknown	518	CAB75384
SCD12A.23		*Streptomyces coelicolor*	Unknown	549	CAB93414
SC8F4.15c		*Streptomyces coelicolor*	Unknown	468	CAB70641
SC2E1.37		*Streptomyces coelicolor*	Unknown	796	Murphy, L., Harris, D., Parkhill, J., Barrell, B. G., Rajandream, M. A. (1998) Unpublished
SCH10.38c		*Streptomyces coelicolor*	Unknown	547	T36550
SCD78.04c		*Streptomyces coelicolor*	Unknown	495	T36043
SC3C8.24		*Streptomyces coelicolor*	Unknown	672	T34899
CcrA	SCC*mec*	*Staphylococcus aureus*	Movement of SCC*mec* element	449	T44092
CcrB	SCC*mec*	*Staphylococcus aureus*	Movement of SCC*mec* element	542	BAA88759
RV1586C	ϕRv1	*Mycobacterium tuberculosis*	Excision and insertion of ϕ Rv1	469	D70542
Hyp.Prot.		*Synechocystis sp.*	Unknown	442	S76073
Y4BA/Y4PH		*Rhizobium sp.*	Unknown	694	P55368

Name	Element	Organism	Function	Length	Accession Number
XisF		*Anabaena*	Excision of *fdxN* element	514	AAA16762
Orf2	Tn*1207.1*	*Streptococcus pneumoniae*	Unknown	370	AAG12996
YBCK		*E. coli*	Unknown	508	P77698
Int	ϕ FC1	*Enterococcus faecalis*	Integration of ϕ FC1	464	AAAD26564
Int	ϕ 105	*Bacillus subtilis*	Unknown	474	IMBP4
YokA	ϕ SPβc2	*Bacillus subtilis*	Unknown	545	NP_046553
Orf1	ϕ TP901-1	*Lactococcus sp.*	Integration/excision of TP901-1	485	CAA59475
Gp31	ϕ WO	*Wolbachia sp.*	Unknown	514	BAA89656
Orf31	ϕ A118	*Listeria monocytogenes*	Excision?	452	CAB53817
Orf469	ϕ R4	*Actinomyces sp.*	Excision and integration of ϕ R4	469	BAA07372
Int	ϕ C31	*Streptomyces sp.*	Integration and excision of ϕ C31	613	NP_047974.1
SpoIVCA		*Bacillus subtilis*	Excision of the skin element	500	AAA22314.1
TndX	Tn*5397*	*Clostridium difficile*	Excision and integration of Tn*5397*	533	AAF35174.1
TnpX	Tn*4453*	*Clostridium difficile*	Excision and integration of Tn*4453*	707	AAF66226.1
TnpX	Tn*4451*	*Clostridium perfringens*	Excision and integration of Tn*4451*	707	AAB51419.1

heterocysts in *Anabaena* sp (Carrasco *et al.*, 1994). *Anabaena* are capable of nitrogen fixation but only in an anaerobic environment (Golden *et al.*, 1988; Golden *et al.*, 1985). Cell differentiation allows the production of specialised cells known as heterocysts where nitrogen fixation occurs (Golden *et al.*, 1988; Golden *et al.*, 1985). Carbon compounds are transported into the heterocyst from surrounding vegetative cells and nitrogenous compounds are transported from the heterocyst to the other cells (Golden *et al.*, 1988; Golden *et al.*, 1985). Differentiation only occurs under conditions of limited nitrogen availability and is consequently a highly regulated event (Carrasco *et al.*, 1994; Golden *et al.*, 1988; Golden *et al.*, 1985). During differentiation two deletion events occur, involving 11 kb (Golden *et al.*, 1985) and 55 kb (Golden *et al.*, 1988) of heterocyst chromosomal DNA. These rearrangements result in the reconstitution of two genes involved in the fixation of nitrogen, *nifD* and *fdxN*, respectively (Carrasco *et al.*, 1994). Both events are catalysed by site-specific recombinases, however, only the 55 kb deletion is catalysed by XisF (Carrasco *et al.*, 1994). The supply of XisF is under the control of a different promoter in vegetative cells and does not result in excision of the 55 kb region of DNA (Carrasco *et al.*, 1994). The presence of *xisH* and *xisI* located downstream of *xisF* within the *fdxN* element have been shown to be essential to the excision event mediated by XisF and are thought to be involved in regulation of the *fdxN* element excision (Ramaswamy *et al.*, 1997).

Sporulation is a mechanism employed by some species of bacteria to survive unfavourable environmental conditions. This process also involves cell differentiation and in *Bacillus subtilis* it occurs in response to nutrient starvation (Sato *et al.*, 1990). Sporulation involves multiple stages and differential enzyme production to synthesise and release a mature spore. The process involves a DNA deletion event in the mother cell, which results in the fusion of two truncated reading frames, *spoIVCB* and *spoIIIC*

(Stragier *et al.*, 1989), to produce a mother cell-specific sigma factor encoded by the reconstituted *sigK* gene (Sato *et al.*, 1994). This deletion event involves 48 kb of DNA and is catalysed by the SpoIVCA protein, which is encoded within the 48 kb *skin* element (Sato *et al.*, 1996; Takemaru *et al.*, 1995). The *skin* element is excised as a circular molecule, consistent with a resolvase mechanism, but the element is lost along with the rest of the mother cell chromosome during lysis to release the mature spore (Sato *et al.*, 1994). It is thought that since the *skin* element has similarity to the phage ϕ105, it may have been derived from a phage that integrated into the *B. subtilis* genome (Takemaru *et al.*, 1995).

TP901-1 is a temperate lactococcal phage that integrates via the phage *attP* site into a single chromosomal site, known as *attB* (Chritiansen *et al.*, 1996). Deletion and mutational analysis has shown that integration is mediated by the product of the *orf1* gene located adjacent to the *attP* site, producing the hybrid sites *attL* and *attR* (Chritiansen *et al.*, 1996). Excision of the integrated phage, or recombination between *attL* and *attR*, has also been shown to require Orf1, and an additional protein, Orf7 (Breüner *et al.*, 1999). Orf1 alone can only mediate a low level of excision, increasing over 3000-fold in the presence of Orf7 (Breüner *et al.*, 1999). These studies clearly show that the large resolvase, Orf1 or Int, is capable of novel resolvase reactions, that is *attP/attB* as well as *attL/attR* recombination. The minimal *attP* and *attB* sites have been determined as 56 bp and 31 bp respectively and a TC dinucleotide is likely to be the position of cleavage (Breüner *et al.*, 2001; Stoll *et al.*, 2002). Recently partially purified TP901-1 integrase has been demonstrated to catalyze integration *in vitro* as well as in human cells (Stoll *et al.*, 2002).

Similar to Orf1 from TP901-1, Sre (ORF469), from ϕR4, is responsible for phage integration and excision, but an as yet unknown protein is also required for excision (Matsuura *et al.*, 1995; Matsuura *et al.*, 1996). Recently, two other phage or phage-like elements have been characterized and shown to encode functional large resolvase proteins required for movement of their respective elements (Bibb and Hatfull, 2002; Yang *et al.*, 2002). The enterococcal temperate phage ϕFC1 encodes the large resolvase *mjl*, which is capable of catalyzing integration of the phage *attP* site into the bacterial *attB* site in the absence of other phage encoded genes both in the natural host and in *E. coli* (Yang *et al.*, 2002). Additionally, a prophage like element from *Mycobacterium tuberculosis*, ϕRv1 encodes a functional large resolvase, *Rv1586c* (or integrase), which is able to catalyze integration between ϕRv1 *attP* and the chromosomal *attB* site *in vivo* and which requires the directionality factor Rv1584c for subsequent excision (Bibb and Hatfull, 2002).

Another well studied phage system is that of the *Streptomyces* temperate phage ϕC31 (Kuhstoss and Rao, 1991; Thorpe and Smith, 1998; Thorpe *et al.*, 2000). The ϕC31 genome sequence has recently been determined, showing that the phage is 41.5 kb in length and carries a multitude of genes including the large resolvase gene, *int* (Kuhstoss and Rao, 1991; Smith *et al.*, 1999; Thorpe and Smith, 1998; Thorpe *et al.*, 2000). It has been shown *in vivo* (Kuhstoss and Rao, 1991) and *in vitro* (Thorpe and Smith, 1998) that Int catalyses recombination between the *attP* and *attB* sites located on the phage and the *Streptomyces* chromosome, respectively, leading to integration of the phage genome. Minimal *attP* and *attB* sites sufficient for integration *in vivo* have been shown to be 39 bp and 34 bp, respectively (Groth *et al.*, 2000). Deletion of the chromosomal *attB* site has lead to the identification of pseudo-*attB* sites into which an integration proficient vector was able to insert at a reduced frequency (Combes *et al.*, 2002).

Analysis of *attB* sites from different *Streptomyces* species has indicated that the dinucleotide where cleavage occurs is likely to be a conserved TT dinucleotide (Combes *et al.*, 2002).

Supercoiling of substrate DNA is not necessary for ϕC31 integrase-mediated integration to occur *in vitro* (Thorpe and Smith, 1998), indicating that the Int resolvase has relaxed topological requirements compared to other resolvase proteins, both in its requirement for supercoiling as well as in the substrates involved in the reaction (intermolecular recombination). No other proteins are required for integration, although it is thought that another protein may be required for excision (Thorpe and Smith, 1998). The Int protein exists predominantly as a dimer in solution and forms a stable synapse with *attP/B* sites, but not with *attL/R* (Thorpe *et al.*, 2000). However, Int is able to bind to all *att* sites (Thorpe *et al.*, 2000), although there are only three completely conserved residues, including the crossover dinucleotides between all four DNA fragments (Kuhstoss and Rao, 1991). The excision process appears to require different reaction conditions than the integration reaction, including other proteins and starting DNA substrates (ie *attL/R* sites) (Thorpe *et al.*, 2000).

The *Staphylococcus* Cassette Chromosome *mec* (SCC*mec*) elements confer methicillin resistance in *Staphylococcus aureus* (Ito *et al.*, 1999). They encode two large resolvase proteins, CcrA and CcrB (Ito *et al.*, 1999). Both proteins are required for *in vivo* excision and integration of the SCC*mec* element (Katayama *et al.*, 2000). Further study of these proteins is necessary, but it may be that the CcrA/B system shows some similarity to the XerC/XerD system. XerC and XerD belong to the λ Int family of site-specific recombinases and are responsible for resolving chromosome dimers produced during genome replication. In this system the DNA strands (top and bottom) are

cleaved by the two different integrase proteins such that XerC cleaves the top strand and XerD cleaves the bottom DNA strand, accompanied by strand transfer and rejoining (Blakely *et al.*, 1997).

The conjugative transposon Tn*5397* from *C. difficile* encodes tetracycline resistance and shows a high level of similarity to Tn*916* (Roberts *et al.*, 2001b). There are two major differences between the two elements. Firstly, Tn*5397* contains a functional group II intron within a homolog of *orf14* of Tn*916* (Roberts *et al.*, 2001a). Secondly, Tn*5397* encodes the large resolvase, TndX, in place of the *int* and *xis* genes which in Tn*916* are required for integration and excision of the transposon (Roberts *et al.*, 2001b; Wang *et al.*, 2000a). Tn*5397* has been shown to be capable of conjugative transfer between *C. difficile* strains with the transposon integrating into a specific chromosomal site. In contrast, matings between a *C. difficile* donor containing Tn*5397* and a *B. subtilis* recipient results in random insertion of Tn*5397* into the *B. subtilis* chromosome (Mullany *et al.*, 1990; Wang *et al.*, 2000a). Transfer of Tn*5397* has also been demonstrated using an oral biofilm model (Roberts *et al.*, 1999).

Tn*5397* produces a circular molecule, which has been detected by PCR amplification from *C. difficile* strains containing the integrated transposon (Wang *et al.*, 2000a). All insertion sites contain a GA dinucleotide, which is also present at either end of the integrated element, at the excision site and the joint of the circular form (Wang *et al.*, 2000a). It has been shown by allelic exchange that the *tndX* gene is essential for transfer of the transposon and production of the circular form (Wang *et al.*, 2000a). Work performed in *E. coli* using mini-transposon derivatives and overexpressed TndX led to the detection of excision and insertion reactions by PCR amplification. This work indicated that TndX was capable of catalyzing both reactions in the absence of any other Tn*5397* proteins (Wang and Mullany, 2000) in contrast to excision catalyzed by other

large resolvases (Bibb and Hatfull, 2002; Breüner *et al.*, 1999; Matsuura *et al.*, 1996; Thorpe and Smith, 1998). Furthermore, TndX-catalyzed insertion shows the same target site selectivity demonstrated during conjugation into either *C. difficile* or *B. subtilis* target sites (Wang and Mullany, 2000).

The Tn4451/Tn4453 Family of Transposons

Tn*4451* and Tn*4452* were the first members of this novel transposon family to be characterised (Abraham *et al.*, 1985). They were identified as ~ 6 kb segments of DNA that carried chloramphenicol resistance and were lost after conjugative transfer from *C. perfringens* strains containing pIP401 and pJIR27, respectively (Abraham and Rood, 1987; Abraham *et al.*, 1985; Brefort *et al.*, 1977). Tn*4451* and Tn*4452* were shown by heteroduplex analysis to be almost identical, with one ~400 bp region of non-homology at the right ends of the elements (Abraham and Rood, 1987). Subsequent studies in *E. coli* showed that Tn*4451* excises precisely from multicopy plasmids (Abraham and Rood, 1988) and later it was shown that Tn*4451* also excises precisely in *C. perfringens* from a multicopy plasmid (Bannam *et al.*, 1995). Characterization of the ends of the element showed some sequence conservation and the presence of 12 bp imperfect inverted repeats (Abraham and Rood, 1988).

Tn*4451* is 6338 bp in length, encodes six genes (Fig 1.14), each with a consensus ribosome binding site (Bannam *et al.*, 1995) and like Tn*5397*, is bounded by directly repeated GA dinucleotides. *tnpX* is the first gene encoded by Tn*4451* and is discussed in detail below. The second gene, *tnpV*, has limited homology to the *xis* gene from phage λ, and, although it was postulated to be involved in excision (Bannam *et al.*, 1995), subsequent deletion analysis showed that it was not required for this phenotype; the function of this protein is as yet unknown (Bannam *et al.*, 1995; Crellin and Rood,

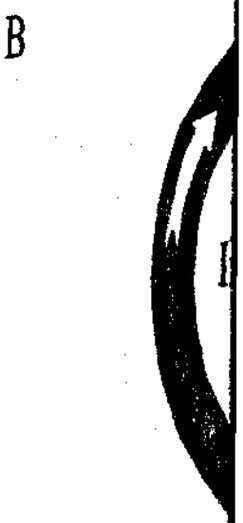


Figure 1.14 Genetic organization and transposition of the Tn*4451*/Tn*4453* transposon family

A Structural organization of the Tn*4451*/Tn*4453* family of elements (Bannam et al., 1995).

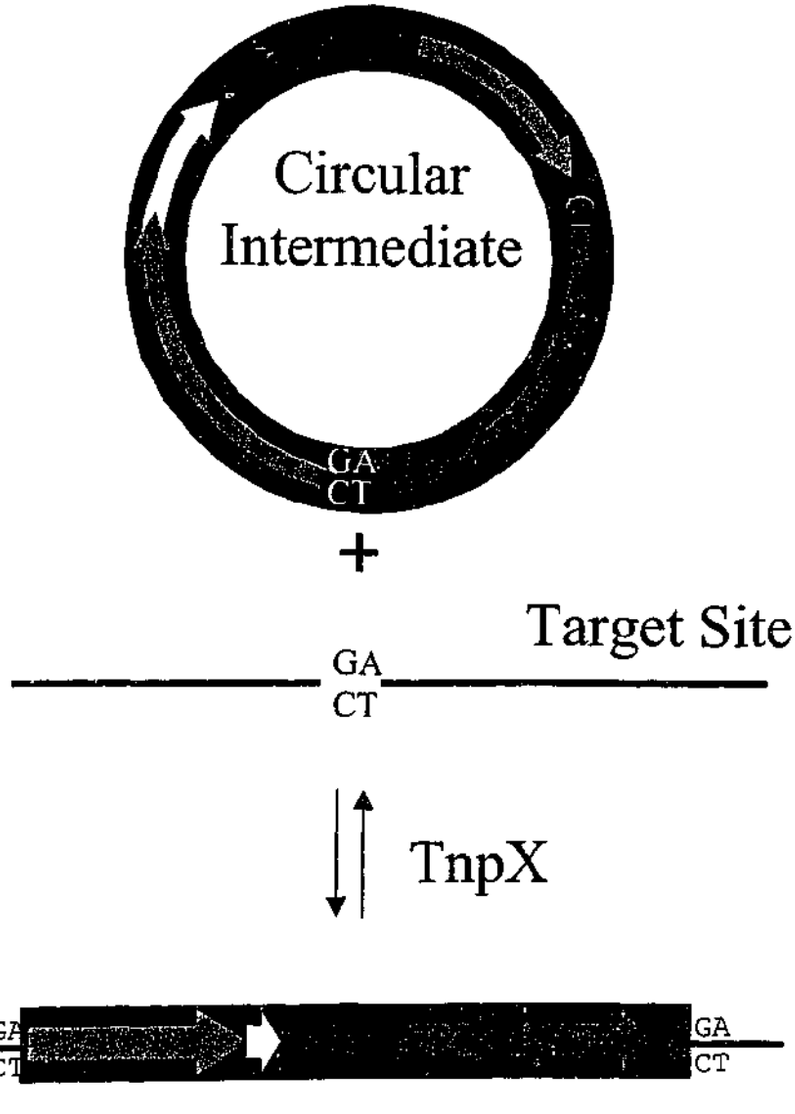
B Transposition Model

The Tn*4451*/Tn*4453* family of elements excise as a circular molecule, which is the transposition intermediate. The circular intermediate can be mobilized to a new cell in the presence of another conjugative element. Alternatively, or subsequently, the circular intermediate integrates at a new location by interaction with TnpX at the joint of the circular form and at the new target site (Crellin and Rood, 1997).

A

tnpX *tnpV catP* *tnpY* RS_A *tnpZ* *tnpW*

B



1997). The next gene, *catP*, encodes a chloramphenicol acetyltransferase and is the gene responsible for chloramphenicol resistance (Bannam *et al.*, 1995). The *tnpY* gene encodes a protein that contains a putative Walker A, box which is often found in ATP binding proteins (Bannam *et al.*, 1995), however, the function of *tnpY* is unknown. The next gene, *tnpZ*, encodes the Mob/Pre family protein, TnpZ (Crellin and Rood, 1998). Mob/Pre proteins are responsible for plasmid mobilization and function by nicking the DNA at a specific site, known as the RS_A site, which in Tn*4451* is located upstream of the *tnpZ* gene (Crellin and Rood, 1998). Both the TnpZ protein and the RS_A site have been shown to be functional, allowing Tn*4451* to be mobilized from one cell to another in the presence of another conjugative element (Fig 1.15) (Crellin and Rood, 1998). The final gene, *tnpW*, is very small and encodes a protein consisting of only 62 aa (Bannam *et al.*, 1995). This protein has no significant similarity to any genes of known function (Bannam *et al.*, 1995).

The product of the *tnpX* gene is responsible for the excision of Tn*4451* from multicopy plasmids in both *E. coli* and *C. perfringens* (Bannam *et al.*, 1995). It also increases the transposition frequency when added *in trans*, implying that the TnpX concentration is the rate-limiting step in transposition (Lyras and Rood, 2000b). The *tnpX* gene is 2.1 kb in length and encodes a protein of 707 aa, with a corresponding molecular size of 82 kDa (Bannam *et al.*, 1995). TnpX shows amino acid similarity, in the first ~ 100 aa, to members of the resolvase/invertase family (Fig 1.10). In addition, a region at the C-terminus was thought to have similarity to the integrase family of site-specific recombinases (Bannam *et al.*, 1995). Site-directed mutagenesis of residues S15, R13 and R89, which are conserved resolvase family residues (Fig 1.10), results in *tnpX* alleles that are unable to restore an excision phenotype to a Tn*4451* derivative containing an internal *tnpX* deletion (Crellin and Rood, 1997). Wild-type *tnpX* is able to complement the *tnpX* mutation and consequently excise the defective transposon,

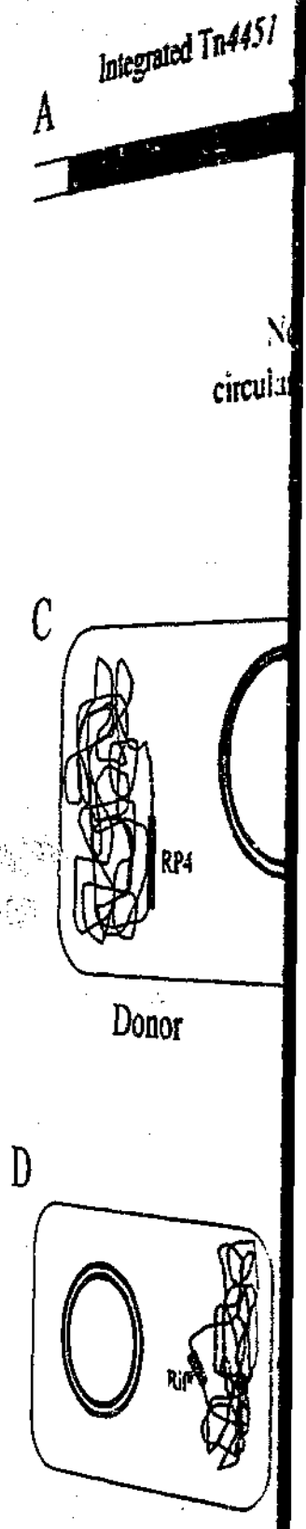
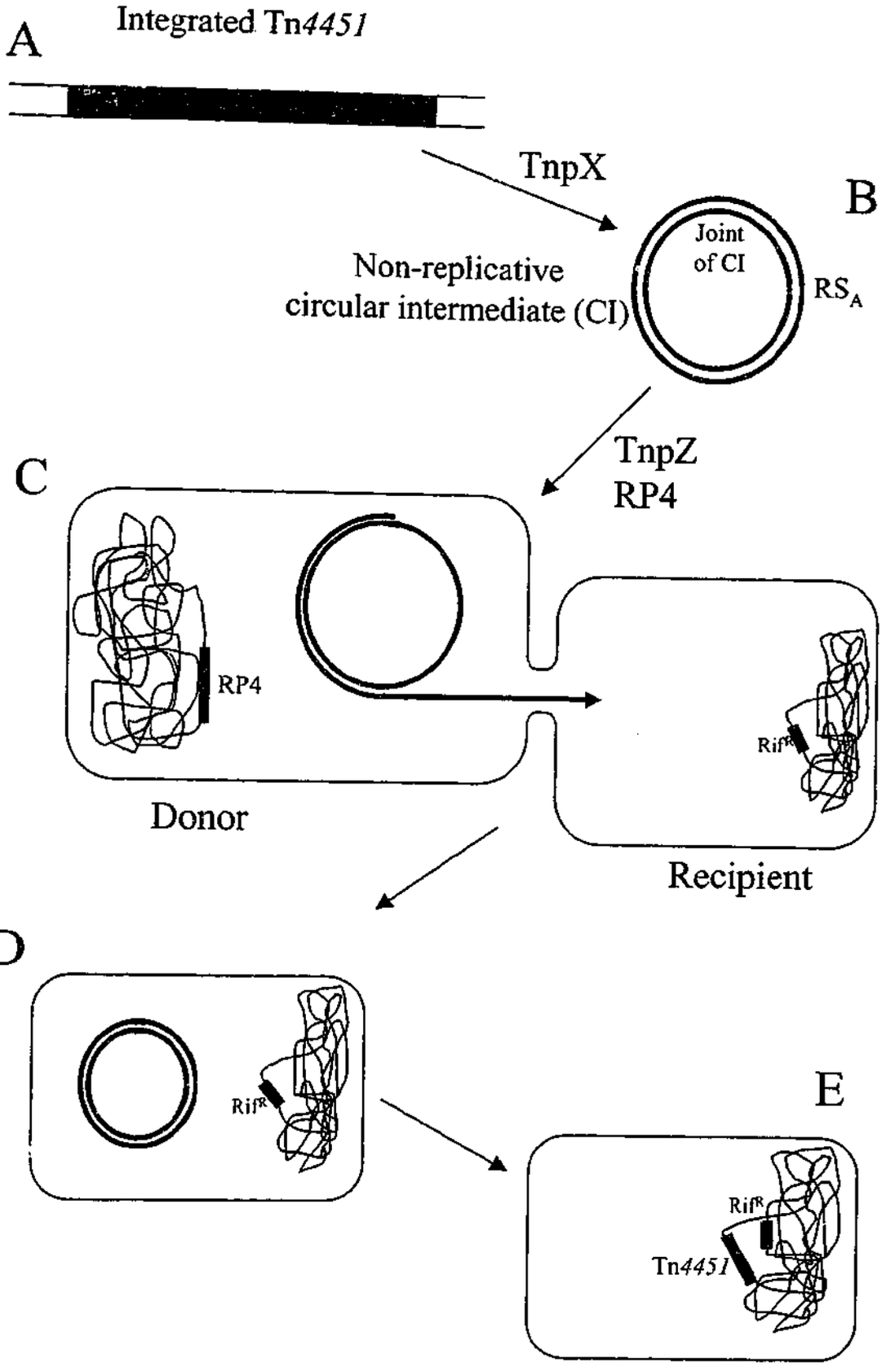


Figure 1.15 Mobilization of Tn*4451*-like transposons

Integrated Tn*4451*/Tn*4453* elements (A) (blue rectangle, red rectangles denote ends) can be excised as a circular molecule through the action of TnpX. The circular intermediate (B) can then be nicked by TnpZ at the RS_A site (green rectangle), and in the presence of a conjugative element, such as RP4, can be mobilized to another cell (C) where it integrates into the chromosome (D). The mobilisation assay uses an *E. coli* strain containing an RP4 derivative (pink rectangle) stably integrated onto the chromosome and demonstrates transfer of the transposon to a rifampicin resistant (Rif^R) recipient (E).


A
Integrated Tn4451
TnpX
B
Joint of CI
Non-replicative circular intermediate (CI)
RS$_A$
TnpZ
RP4
C
RP4
Rif^R
Donor
Recipient
D
Rif^R
E
Rif^R
Tn4451

thereby forming the basis of an excision assay (Crellin and Rood, 1997). Mutants of the conserved residues of the putative integrase domain, H632 and Y673, are able to complement the same Tn*4451* derivative in an excision assay. Taken together these results show that the reaction catalysed by TnpX is resolvase-mediated (Crellin and Rood, 1997). It is not known whether other Tn*4451* proteins, or other cellular proteins, are required for excision or integration but TnpX is essential for excision (Lyras and Rood, 2000b).

Two other members of this transposon family have been identified from a chloramphenicol resistant *C. difficile* isolate, namely Tn*4453a* and Tn*4453b* (Lyras *et al.*, 1998). Tn*4453a* has been completely sequenced and has 89% nucleotide sequence identity to Tn*4451* (Lyras and Rood, 2000b). All of the Tn*4451* genes are present. The highest degree of divergence occurs between the two TnpW proteins (24.2% non-identity), and the next highest between the two TnpX proteins at 12.3% non-identity (Lyras and Rood, 2000b). Interestingly, the TnpV and CatP proteins are completely conserved (Lyras and Rood, 2000b). The frequency of transposition of these elements has been determined using *E. coli* strains containing chromosomal insertions. Transposition is detected by mating to a suitable *E. coli* recipient, a process mediated by the conjugative plasmid pVS520. An additional *tnpX* allele is supplied *in trans* on a second plasmid to increase the level of TnpX within the cell, and therefore the transposition levels, or to allow phenotypic testing of *tnpX* mutants. The transposition frequencies of Tn*4451*, Tn*4453a* and Tn*4453b* are different, with Tn*4453a* showing the highest degree of transposition and transposition of Tn*4453b* being undetectable unless additional TnpX is supplied *in trans* (Lyras and Rood, 2000b). The TnpX proteins from each of the elements are functionally interchangeable. The transposition frequency is a characteristic of the transposon rather than being dependent on which TnpX protein is supplied *in trans* (Lyras and Rood, 2000b). For example, Tn*4453b* has the lowest

transposition frequency, but when the *tnpX* gene from Tn*4453b* is supplied to Tn*4453a* *in trans*, the element that has the highest transposition frequency, this strain still exhibits a high transposition frequency (Lyras and Rood, 2000b). The resolvase domain mutants derived from the Tn*4451* TnpX protein exhibit transposition frequencies below the limit of detection for both Tn*4451* and Tn*4453a*, which is expected since these proteins are unable to catalyze excision (Lyras and Rood, 2000b). In contrast, the integrase domain mutants have no significant effect on the transposition frequency of either element, illustrating that the integrase domain is not involved in the insertion process (Lyras and Rood, 2000b).

The transposition mechanism employed by these elements involves the formation of a circular intermediate (Lyras and Rood, 2000b). This circular molecule was first observed with excision assays carried out on Tn*4451* in *E. coli* (Bannam *et al.*, 1995). Its formation is consistent with a resolvase mechanism as this reaction essentially constitutes cointegrate resolution. In this situation the *res* sites are located at the ends of the transposon rather than in the middle of two duplicated elements (Fig 1.15). The formation of the circular intermediate also results in the generation of a strong promoter at the joint of the circular intermediate. This promoter is capable of facilitating expression of the *tnpX* gene (Lyras and Rood, 2000b). Since the circular molecule is non-replicative, the formation of this promoter presumably results in a high level of *tnpX* expression when the transposon needs to integrate for its survival (Bannam *et al.*, 1995; Crellin and Rood, 1997; Lyras and Rood, 2000b).

The joint of the circular intermediate resembles the ends of the element and the re-constituted target site since they all contain a GA dinucleotide (Crellin and Rood, 1997; Lyras and Rood, 1997). Extensive mutagenesis was carried out on the GA dinucleotides at the left and right ends of Tn*4451* (Crellin and Rood, 1997). This study

showed that when the GA dinucleotide at the left end was altered, excision produced deletion sites that fell into two categories: those with a GA, and those with the mutated dinucleotide (Crellin and Rood, 1997). These results are consistent with an excision mechanism involving a resolvase-mediated 2 bp staggered cut at the terminal GA dinucleotides of the transposon, resulting in a 2 bp heteroduplex at the deletion site and at the joint of the circular intermediate (Crellin and Rood, 1997). Mismatch repair or DNA replication would resolve the heteroduplex at the deletion site (Crellin and Rood, 1997). Dinucleotide mutant analysis suggests that the GA derived from the right end of the transposon constitutes the joint of the circular molecule and that if this dinucleotide is not present at the right end the circular molecule may not be formed, although the element is still excised (Crellin and Rood, 1997). The substitution of a T for the G in the GA dinucleotide at either or both ends of the element increases the excision frequency, suggesting that a TA dinucleotide is the preferred substrate for recombination (Crellin and Rood, 1997).

After formation of the circular intermediate, the action of TnpZ may result in nicking at the RS_A site, allowing transfer of a single strand of the non-replicative molecule to another cell if a conjugative element is present. Following transfer, or in lieu of transfer, the element must reintegrate into a replicating DNA molecule to avoid being diluted out of the bacterial population. It is assumed that integration is the reverse of the excision process (Fig 1.14B) although integration alone has only recently been shown experimentally (D. Lyras and J. I. Rood, unpublished results). The function of the other transposon genes, *tnpV*, *tnpY* and *tnpW*, and the role they may play in the transposition process is yet to be determined although it has been shown that they are not essential for transposition (D. Lyras and J. I. Rood, unpublished results).

Aims and Objectives

The overall objective of this project is to develop a detailed understanding of the novel transposition mechanism exhibited by the Tn*4451*/Tn*4453* family of elements. The specific aims of this thesis include determining the reaction conditions required for TnpX mediated excision and insertion. The results of this work show that *in vitro* purified TnpX is capable of catalyzing excision of a Tn*4451* derivative in the absence of other transposon or cellular proteins and that this reaction is enhanced by the addition of Ca^{2+} ions. The insertion reaction seems to be more complicated.

Another objective of this study was to identify the TnpX binding site. The binding site of traditional resolvase proteins is located within a transposon to facilitate recombination between the *res* sites after replicative transposition has occurred (Sherratt, 1989). The resolvase protein catalyzes recombination between the two sites to resolve the dimer formed by transposition (Craig, 1996). TnpX uses a resolvase mechanism to catalyze transposition. This suggests that the TnpX binding sites must be located at the transposon ends. The studies presented in this thesis show that the TnpX binding sites are indeed located at either end of the transposon, primarily within the transposon sequence itself, and also at the joint of the circular intermediate. Binding sites were also located at several deletion and insertion sites.

TnpX is considerably larger than the well studied resolvase proteins, showing significant similarity to these proteins within the first 100 amino acids (the resolvase domain). As a consequence, the final aims of this study were to locate functional domains of TnpX and to identify specific TnpX residues important for function. It was found that TnpX existed as a dimer in solution and that the last 110 aa were not essential for function. The DNA binding domain of TnpX was localized to the region

between amino acids 493-597 and the dimerization domain between amino acids 100-356. The results showed that many of the residues in the first ~150 amino acids of TnpX were important for both excision and insertion. Two amino acids within the 250-260 aa region were also found to be important and *in vitro* analysis indicated that this region may also be involved in DNA binding or recognition. One mutant was found that was capable of insertion but defective for excision indicating for the first time that the insertion and excision reactions can be separated at the mechanistic level.

Chapter 2

Materials and Methods

Strains and Culture Conditions

All bacterial strains were grown in 2 x YT broth (1.6% tryptone, 1% yeast extract and 0.5% NaCl, Oxoid) or on 2 x YT agar plates (1.5% bacteriological agar, Oxoid) supplemented with appropriate antibiotics (Sigma). *E. coli* strains were grown at 37°C (with shaking at 150 rpm for broth cultures), except where otherwise indicated. Strain characteristics are shown in Table 2.1. All media were sterilized by autoclaving at 121°C for 20 min. Long term storage of *E. coli* strains was in glycerol storage broth (3.7% (w/v) BHI broth, 50% glycerol) at –20°C or as freeze-dried cultures at room temperature.

Recombinant Plasmids and Cloning Vectors

Cloning experiments were generally carried out using one of the *E. coli* vectors pUC18 (Yanisch-Perron *et al.*, 1985), pSU39 (Bartolomé *et al.*, 1991), pWSK29 (Wang and Kushner, 1991) or temperature sensitive derivatives of the pWSK series, pTSK29 and pTSC29 (Phillips, 1999). Protein over-expression and purification required the use of the *E. coli* expression plasmid pET22b+ (Novagen), containing the T7 promoter, *lac* operator and a C-terminal hexa-His tag. Plasmid characteristics are shown in Table 2.2.

Table 2.1 Characteristics and origin of *E. coli* strains

Strain	Relevant Characteristics	Reference/Origin
DH5α	F⁻ ϕ80 d*lacZ*ΔM15Δ(*lacZYA-argF*)U169 *endA1 recA1 hsdR17* ($r_k^-m_k^-$)*deoR thi-1 supE44- gyrA96 relA1*	Bethesda Research Laboratories
DH12S	*mcr*Δ(*mrr-hsdRMS-mcrBC*) ϕ80 d*lacZ* D*mis* Δ*lacx74 dcoR recA1 araD139* Δ(*ara, leu*) *7697 galU galK* λ⁻ *nupG*/F' *proAB*⁺ *lacI*^{q}Z Δ*mis*	Bethesda Research Laboratories
C43(DE3)	BL21(DE3) with uncharaterized mutation(s)	(Miroux and Walker, 1996)
XL1-Red	*endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac mutD5 mutS mutT* Tn*10* Tc^R	Stratagene
LT101	F⁻ *hsdS20* ($r_B^-m_B^-$) *leu supE44 ara14 galK2 lacYI proA2 rpsL20 xyl-5 mtl-1 recA13 mcrB* Sm^R, Rif^R derivative of HB101	(Palombo *et al.*, 1989)
MM294	*supE*44 *hsdR end*A1 *pro thi*	(Meselson and Yuan, 1968)
JIR862	Rif^R derivative of MM294	(Abraham and Rood, 1987)
JIR888	JIR862::Tn*4451*, Rif^R Cm^R	(Abraham and Rood, 1987)
JIR891	JIR862::Tn*4451*, Rif^R Cm^R	(Abraham and Rood, 1987)
JIR2343	DH5α (pJIR683), Ap^RCm^R	(Bannam *et al.*, 1995)
JIR2463	DH5α (pJIR773), Kn^RCm^R	(Bannam *et al.*, 1995)
JIR5724	MM294::Tn*4453a* (pVS520), Tc^R Cm^R	(Lyras and Rood, 2000)
JIR7171	DH12S (pJIR1880), Ap^RCm^R	This Study
JIR7909	DH5α (pJIR2376), Kn^R	Lyras, D., unpublished
JIR7998	DH5α (pJIR2421), Cm^R	This Study

Antibiotic resistant phenotypes are denoted as follows, ampicillin resistance (Ap^R), chloramphenicol resistance (Cm^R), kanamycin resistance (Kn^R) and tetracycline resistance (Tc^R).

Table 2.2 Characteristics and origin of recombinant plasmids

Plasmid	Relevant Characteristics	Reference/Origin
pUC18	Cloning vector, ColE1 origin, 2.7 kb, Ap^R	(Yanisch-Perron *et al.*, 1985)
pSU39	Cloning vector, p15A origin, 2.7 kb, Kn^R	(Bartolomé *et al.*, 1991)
pWSK29	Cloning vector, pSC101 origin, 5.4 kb, Ap^R	(Wang and Kushner, 1991)
pTSK29	Cloning vector, pSC101 $origin^{TS}$, 5.4 kb, Kn^R	(Phillips, 1999)
pTSC29	Cloning vector, pSC101 $origin^{TS}$, 5.1 kb, Cm^R	(Phillips, 1999)
pT7Blue-3	Cloning vector, ColE1 origin, 3.8 kb, Ap^RKn^R	Novagen
pET22b+	Expression plasmid, ColE1 origin, C-terminal His tag, 5.5 kb, Ap^R	Novagen
pVS520	Conjugative, RP1 origin, 50 kb, Tc^R	(Palombo *et al.*, 1989)
pCB182	Promoter-probe vector, promoterless *lacZ* and *galK* genes, 6.8 kb, Ap^R	(Schneider and Beck, 1986)
pJIR47	pJIR45ΔTn*4451*, contains deletion site from pIP401, Ap^R	(Abraham *et al.*, 1985)
pJIR639	pBluescript *PstI*/*Eco*RV Ω (*PstI*/*Eco*RV: pJIR52, 3.2 kb, contains $tnpX^+$, $tnpV^+$, $\Delta catP$), Ap^R	(Bannam *et al.*, 1995)
pJIR683	pJIR641 *Eco*RV/*Asp*718 Ω (*Eco*RV/*Asp*718: pJIR45, 5.1 kb, contains Tn*4451tnpXΔ1*), Ap^RCm^R	(Bannam *et al.*, 1995)
pJIR773	pSU39 *PstI*/*Asp*718 Ω (*PstI*/*Asp*718: pJIR740, 8.0 kb, contains Tn*4451tnpXΔ1*), Kn^RCm^R	(Bannam *et al.*, 1995)
pJIR1452	pSU39 *PstI*/*Asp*718 Ω (*PstI*/*Asp*718: pJIR740, 8.0 kb, contains Tn*4451tnpXΔ1*, TA-----TA^b), Kn^RCm^R	(Crellin and Rood, 1997)
pJIR1860	pUC18 *PstI*/*Eco*RI Ω (pJIR639 UP/10233 PCR product, 1.6 kb contains $tnpX_{1\text{-}356}$), Ap^R	Recombinant
pJIR1861	pUC18 *PstI*/*Eco*RI Ω (pJIR639 UP/10246 PCR product, 2.0 kb contains $tnpX_{1\text{-}492}$), Ap^R	Recombinant
pJIR1862	pUC18 *PstI*/*Eco*RI Ω (pJIR639 UP/10232 PCR product, 2.4 kb contains $tnpX_{1\text{-}597}$), Ap^R	Recombinant

pJIR1872	pSU39 *Bam*HI/*Asp*718 Ω (*Bam*HI/*Asp*718: pJIR639, 3.2 kb, contains *tnpX*$^+$, *tnpV*$^+$, *ΔcatP*), Ap^R	Recombinant
pJIR1880	pWSK29 *PstI*/*Asp*718 Ω (*PstI*/*Asp*718: pJIR1452, 8.0 kb, contains Tn*4451tnpXΔ1*, TA------TA^b), Ap^RCm^R	Recombinant
pJIR1898	pUC18 *PstI*/*Eco*RI Ω (pJIR639 UP/7364 PCR product, 2.7 kb contains *tnpX*$^+$), Ap^R	Recombinant
pJIR1918	pJIR1872($tnpX_{T285I}$), Kn^R	Random Mutant
pJIR1919	pJIR1872($tnpX_{K157E}$), Kn^R	Random Mutant
pJIR1920	pJIR1872($tnpX_{K260E}$), Kn^R	Random Mutant
pJIR1921	pJIR1872($tnpX_{A138T}$), Kn^R	Random Mutant
pJIR1922	pJIR1872($tnpX_{S66P}$), Kn^R	Random Mutant
pJIR1923	pJIR1872($tnpX_{M83I}$), Kn^R	Random Mutant
pJIR1945	pJIR1872($tnpX_{A138V}$), Kn^R	Random Mutant
pJIR1946	pJIR1872($tnpX_{G161S}$), Kn^R	Random Mutant
pJIR1947	pJIR1872($tnpX_{D83G}$), Kn^R	Random Mutant
pJIR1948	pJIR1872($tnpX_{N131S}$), Kn^R	Random Mutant
pJIR1984	pJIR1872($tnpX_{S261P}$), Kn^R	Random Mutant
pJIR1998	pET22b+ *NdeI*/*Xho*I Ω (pJIR639 12882/12873 PCR product, 2.1 kb, $tnpX_{1-707}$), Ap^R	Recombinant
pJIR1999	pET22b+ *NdeI*/*Xho*I Ω (pJIR639 12882/13200 PCR product, 1.8 kb, $tnpX_{1-597}$), Ap^R	Recombinant
pJIR2011	pJIR1872($tnpX_{S85N}$), Kn^R	Random Mutant
pJIR2040	pJIR1872($tnpX_{D140N}$), Kn^R	Random Mutant
pJIR2041	pJIR1872($tnpX_{E251K}$), Kn^R	Random Mutant
pJIR2049	pJIR1872($tnpX_{R89Q}$), Kn^R	Random Mutant
pJIR2063	pJIR1872($tnpX_{S26F}$), Kn^R	Random Mutant
pJIR2067	pJIR1872($tnpX_{G56D}$), Kn^R	Random Mutant
pJIR2068	pJIR1872($tnpX_{C324R}$), Kn^R	Random Mutant

pJIR2085	pJIR1872($tnpX_{S15L}$), Kn^R	Random Mutant
pJIR2087	pJIR1872($tnpX_{G88R}$), Kn^R	Random Mutant
pJIR2094	pJIR1872($tnpX_{L34F}$), Kn^R	Random Mutant
pJIR2100	pJIR1872($tnpX_{R61L}$), Kn^R	Random Mutant
pJIR2159	pCB182 *SmaI*/*Hin*dIII Ω (PCR product 8187/4675, 1.0 kb, contains CI joint, TA derivative), Ap^R	Recombinant
pJIR2239	pET22b+ *NdeI*/*Xho*I Ω (pJIR639 12882/15962 PCR product, 1.0 kb, contains $tnpX_{1-356}$), Ap^R	Recombinant
pJIR2240	pET22b+ *NdeI*/*Xho*I Ω (pJIR639 12882/15963 PCR product, 1.5 kb, contains $tnpX_{1-492}$), Ap^R	Recombinant
pJIR2292	pT7Blue-3 *Eco*RV Ω (PCR product 4675/4676, 1.0 kb, contains CI joint, TA derivative), $Kn^R Ap^R$	Recombinant
pJIR2327	pET22b+ *NdeI*/*Xho*I Ω (pJIR1920 12882/12873 PCR product, 2.1 kb, contains $tnpX_{K260E}$), Ap^R	Recombinant
pJIR2329	pET22b+ *NdeI*/*Xho*I Ω (pJIR2068 12882/12873 PCR product, 2.1 kb, contains $tnpX_{C324R}$), Ap^R	Recombinant
pJIR2362	pET22b+ *NdeI*/*Xho*I Ω (pJIR2041 12882/12873 PCR product, 2.1 kb, contains $tnpX_{E251K}$), Ap^R)	Recombinant
pJIR2376	pTSK29 *Eco*RI Ω (*Eco*RI: pJIR2340, 0.25 kb contains CI joint, GA derivative), Kn^R	Recombinant
pJIR2421	pTSC29 *Eco*RI Ω (*Eco*RI: pJIR2340, 0.25 kb contains CI joint, GA derivative), Cm^R	Recombinant
pJIR2501	pJIR1452 Δ Tn*4451tnpXΔ1*, Kn^R	Recombinant

Antibiotic resistant phenotypes are denoted as follows, ampicillin resistance (Ap^R), chloramphenicol resistance (Cm^R), kanamycin resistance (Kn^R) and tetracycline resistance (Tc^R). The temperature sensitive origin of replication present on the pTSK and pTSC plasmids is denoted by $origin^{TS}$. CI denotes the circular intermediate.

Transformation Procedures

a) Preparation and Transformation of Pipes Competent *E. coli* Cells

The *E. coli* strains C43(DE3) and DH5α were prepared as described previously (Inoue *et al.*, 1990). Strains were subcultured on 2 x YT agar and grown at 37°C for 16-20 h. A single colony was used to inoculate 5 mL of 2 x YT broth and incubated overnight at 37°C with shaking. One mL of the overnight culture was then transferred into 125 mL of SOB (10 mM NaCl, 2.5 mM KCl, 2% (w/v) Tryptone, 0.5% (w/v) Yeast Extract) (Sambrook *et al.*, 1989) broth in a 1 L flask. The culture was incubated at 18°C with shaking until the culture reached a turbidity at 600 nm of 0.6. The culture was then chilled on ice for 10-15 min. The cells were pelleted by centrifugation at 1,500 *g* for 15 min at 4°C, the supernatant was discarded and the cell pellet was resuspended in 20 mL of transformation buffer (TB) (10 mM Pipes, 15 mM $CaCl_2$, 250 mM KCl (pH6.7), 55 mM $MnCl_2$, filter sterilized). The cells were incubated on ice for 10 min followed by centrifugation at 1,500 *g* for 15 min at 4°C. The cell pellet was resuspended in 10 mL of TB to which DMSO (dimethyl sulfoxide) was added to a final concentration of 7% and mixed by swirling. The cell suspension was incubated on ice for a further 10 min and dispensed as 100 µL aliquots into pre-chilled microcentrifuge tubes and snap frozen in a dry ice/ethanol bath. The cells were stored at -70°C until use.

Transformation experiments were conducted by pre-thawing Pipes-competent cells on ice prior to the addition of plasmid DNA. The plasmid/competent cell mixture was incubated on ice for 30 min. The cells were heat shocked at 42°C for 30 seconds and then immediately chilled on ice for 2 min. To allow the expression of antibiotic resistance genes, 900 µL of 2 x YT broth was added and the cell mixture was incubated, with

shaking, at 37°C for 1-2 hours. The transformation mixture was spread, as 100 µL aliquots, onto 2 x YT agar supplemented with appropriate antibiotics.

b) Preparation and Electroporation of Electrocompetent *E. coli* Cells

The following method was used for *E. coli* when a higher transformation efficiency than that obtained by the previous method was required. Electrocompetent DH12S (pJIR1880) cells, used for random mutant screening, were prepared using an established procedure (Smith *et al.*, 1990) with several modifications. A single *E. coli* colony was used to inoculate 10 mL of SOB broth supplemented with appropriate antibiotics. The cells were grown overnight at 37°C and 1 mL of the resulting culture was used to inoculate 500 mL of SOB broth. The culture was grown at 37°C with moderate agitation until the turbidity at 600 nm was 0.8. The cells were harvested by centrifugation at 4,000 *g* for 10 min at 4°C. The supernatant was discarded and the pellet resuspended in 250 mL of ice-cold 10% (v/v) glycerol. The cells were again pelleted, the supernatant discarded and the cell pellet resuspended in 250 mL of ice-cold 10% (v/v) glycerol. The cells were collected by centrifugation as before and the cell pellet resuspended in a final volume of 2 mL of ice-cold 10% (v/v) glycerol. The cells were snap frozen as 100 µL aliquots in a dry/ice ethanol bath and were stored at -70°C until use.

Prior to use the cells were thawed on ice. Ethanol or isopropanol precipitated DNA was added to a 20 µL aliquot of electrocompetent cells and mixed gently. The mixture was transferred to a pre-chilled 0.1 cm electroporation cuvette (Bio-Rad Laboratories, California, U.S.A.), which was subsequently placed in a Bio-Rad Gene Pulser™ and subjected to electroporation at 1.8 kV/cm, a resistance of 200 Ω and a capacitance of 25 µF. The cells were then subcultured into 2 mL of SOC broth (SOB broth supplemented with 20 mM glucose) and incubated at 37°C with shaking for 1 hour. The transformation

mixture was then plated, in 100 µL aliquots, onto 2 x YT agar plates supplemented with appropriate antibiotics and incubated.

c) Preparation and Transformation of $CaCl_2$-Competent *E. coli* Cells

Any *E. coli* strain that already contained a plasmid and required the introduction of another plasmid, as required in the construction of the *in vivo* assay strains, was rendered competent using this small scale method. Several colonies were inoculated into 5 mL of 2 x YT broth containing appropriate antibiotics and grown at 37°C with shaking until the culture had reached mid log phase (several hours depending on the amount of inoculum). The cells were pelleted at 12,000 *g* for 2 min, the supernatant decanted and the cell pellet resuspended in 1 mL of ice cold 10 mM $CaCl_2$. The cells were resuspended by vortexing and chilled on ice. Cells were pelleted by centrifugation at 12,000 *g* for 15 seconds, the supernatant decanted or removed with a pipette and the cells resuspended in 0.5 mL of ice cold 10 mM $CaCl_2$ by vortexing. The process was repeated twice and the cells resuspended in 0.3 mL of ice cold 75 mM $CaCl_2$ and finally 0.2 mL of 75 mM $CaCl_2$. The resulting suspension was divided into two 100 µL aliquots and 2-5 µL of plasmid DNA was added to one aliquot for transformation. The second aliquot of cells were used as a no DNA control.

The transformation procedure entailed incubation on ice for 30 min after addition of plasmid DNA, followed by incubation at 42°C for 30 seconds, incubation on ice for 2 min, addition of 0.9 mL of 2 x YT broth and finally incubation at 37°C for one hour with shaking. The transformation mixtures were subsequently plated onto 2 x YT plates containing appropriate antibiotics and incubated at 37°C.

E. coli *DNA Isolation and Manipulation*

a) Ethanol and Isopropanol Precipitation of DNA

DNA was precipitated by the addition of 0.1 volumes of 3 M sodium acetate (pH 4.6) and either one volume of 100% isopropanol or 2.5 volumes of 100% ethanol. The mixture was incubated at –70°C for 30 min followed by centrifugation at 12,000 *g* and 4°C for 15 min to pellet the precipitated DNA. The DNA pellet was washed with ice cold 70% ethanol, vacuum dried and resuspended in an appropriate volume of water, TE or a Tris-HCl (pH 8.0) buffer.

b) Phenol/Chloroform Extraction of DNA

An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the DNA solution and vortexed vigorously. The phases were separated by centrifugation at 12,000 *g* for 5 min. The aqueous phase was removed to a new tube and an equal volume of chloroform:isoamyl alcohol (24:1) was added, mixed and separated as before. The aqueous layer was removed to a new tube and the DNA recovered by isopropanol or ethanol precipitation.

c) Small Scale Plasmid DNA Isolation

Alkaline Lysis Method

E. coli plasmid DNA was routinely extracted using an alkaline lysis method (Morelle, 1989). The appropriate *E. coli* strain was grown overnight at 37°C in 5 mL of 2 x YT broth supplemented with the relevant antibiotics. The cells were harvested by centrifugation at 1,500 *g* for 3 min at room temperature and the supernatant discarded. The cell pellet was resuspended in 200 µL of lysis buffer (50 mM glucose, 25 mM Tris-HCl

(pH 8.0), 10 mM EDTA), and 400 µL of freshly prepared alkaline solution (0.2 M NaOH, 1% (w/v) SDS) was added. The solution was mixed gently by inversion and then incubated on ice for 5 min. The protein and high molecular weight DNA and RNA was precipitated by addition of 300 µL of 7.5 M ammonium acetate (pH 7.8) and the solution was again mixed by inversion. The solution was incubated on ice for 10 min prior to centrifugation at 12,000 *g* for 10 min. The supernatant was transferred to a new microcentrifuge tube and RNase A (Sigma Chemical Co., Missouri, U.S.A.) was added to a final concentration of 20 µg/mL. Isopropanol (0.6 volumes) was added to precipitate the DNA and the solution was incubated at room temperature for 10 min prior to centrifugation at room temperature for 10 min at 12,000 *g* to pellet the DNA. The DNA pellet was washed with cold 70% (v/v) ethanol, dried under vacuum using a Savant Speedvac SVC100 for approximately 5 min, and then resuspended in 100 µL of deionized water (dH_2O). The DNA preparation was incubated at 37°C for 10 min, then centrifuged at 12,000 *g* for 3 min at room temperature. The supernatant containing the purified plasmid DNA was transferred to a fresh tube and was stored at -20°C until use.

Applied Biosystems Method

T o produce high quality plasmid DNA for automated sequencing or cloning the DNA was extracted using the modified alkaline lysis/PEG precipitation procedure outlined in the PRISM™ Ready Reaction Dye Deoxy™ Terminator Cycle Sequencing Kit manual (Applied Biosystems, California, U.S.A.). The appropriate *E. coli* strain was grown overnight at 37°C in 5 mL of 2 x YT broth supplemented with the relevant antibiotics. The cells were harvested by centrifugation at 1,500 *g* for 3 min at room temperature and the supernatant discarded. The cell pellet was resuspended in 200 µL of lysis buffer, and 300 µL of freshly prepared alkaline lysis solution was added as before. The sample was mixed by inversion and then incubated on ice for 5 min before adding 300 µL of 3.0 M

potassium acetate, pH 4.6. The contents were mixed by inverting the tube and incubated on ice for 5 min. Cellular debris was removed by centrifugation at 12,000 *g* for 10 min at room temperature and the clear supernatant was transferred to a fresh tube. RNase A was added to a final concentration of 20 μg/mL and the solution was incubated at 37°C for 20 min. The supernatant was extracted twice with 400 μL of chloroform, where the layers were mixed for 30 seconds by inversion following each extraction. The sample was centrifuged at 12,000 *g* for 1 min at room temperature to separate the phases and the aqueous phase was transferred to a new tube. An equal volume of isopropanol was added to precipitate the DNA and the sample was immediately centrifuged at 12,000 *g* for 10 min at room temperature. The DNA pellet was washed with 500 μL of cold 70% ethanol, dried under vacuum, and then resuspended in 32 μL of deionized water. The plasmid DNA was then precipitated by the addition of 8.0 μL of 4 M NaCl and 40 μL of 13% (w/v) PEG_{8000} (Amresco). The sample was incubated on ice for 20 min and then centrifuged at 12,000 *g* for 15 min at 4°C. The supernatant was removed and the plasmid DNA pellet washed with 500 μL of cold 70% ethanol. The pellet was dried under vacuum and then resuspended in 30-100 μL of dH_2O. The DNA preparation was stored at -20°C until required.

d) Large Scale Plasmid DNA Isolation

CsCl-Gradients

CsCl-purified plasmid DNA was obtained from *E. coli* after pelleting an overnight 200 mL culture of the appropriate bacterial strain at 5,800 *g* for 10 min at room temperature. The bacterial pellet was resuspended in 3 mL of a 0.1 M Tris (pH 8.0)/25% (w/v) sucrose solution. Lysozyme was added to a final concentration of 1 mg/mL and following addition of 0.2 mL of 0.5 M EDTA the mixture was incubated at 37°C for 5 min, with shaking. Sterile water (3 mL) and 4 mL of triton lysis mixture (50 mM Tris-HCl pH8.0, 12.5 mM

EDTA and 10% (v/v) Triton X-100) were added and mixed briefly. The lysed bacterial culture was centrifuged at 26,900 *g* for 1 h at 4°C. The supernatant was very carefully decanted and the pellet discarded. CsCl (8 g) was disolved in the supernatant, the volume adjusted to 11 mL and loaded into Quick-Seal Ultracentrifuge Tubes (Beckman, California, U.S.A.) and overlayed with parafin oil. Ethidium bromide (250 µL of a 10 mg/mL stock) was added and the tubes were heat sealed, balanced and centrifuged at 260,000 *g* for at least 20 h at 20°C. DNA bands were visualized under UV light and the plasmid band was extracted using a 18 gauge needle and 5 mL syringe. The solution was extracted three times by at least double the volume of NaCl-saturated isopropanol in TE buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA) to remove the ethidium bromide. The sample was then dialyzed in two changes of TES buffer (10 mM Tris (pH 8.0), 1 mM EDTA and 0.4 M NaCl), one change of TE buffer and two changes of Tris-HCl buffer (20 mM, pH 8.0). The DNA solution was pervaporated to reduce the volume and then stored at –20°C. The DNA was quantitated by visualization on an ethidium bromide stained agarose gel against DNA standards of known concentration.

e) Small Scale Chromosomal DNA Isolation

Chromosomal DNA was isolated from *E. coli* (Ausubel *et al.*, 1994) using a 5 mL overnight culture of the appropriate strain. The culture was pelleted at 12,000 *g* for 2 min and the supernatant was discarded. The pellet was resuspended in 567 µL of TE to which SDS and proteinase K were added to final concentrations of 0.5% and 100 µg/mL, respectively. The suspension was mixed thoroughly and incubated at 37°C for 1 h. After incubation 100 µL of 5 M NaCl and 80 µL of 10% CTAB (cetyltrimethylammonium bromide) in 0.7 M NaCl were added and the mixture was incubated at 65°C for 10 min. An equal volume of chloroform:isoamyl alcohol (24:1) was added and the sample thoroughly agitated. The solution was centrifuged at 12,000 *g* for 5 min and the aqueous

layer removed to a new microcentrifuge tube. To this solution an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the aqueous layer was removed as before. To precipitate the DNA, 0.6 volumes of isopropanol was added to the extracted aqueous layer and the solution was centrifuged at 12,000 *g* for 5 min. The pellet was washed with 70% ethanol to remove salt, dried and resuspended in 100 µL of TE buffer.

Recombinant DNA Techniques

a) Restriction Endonuclease Analysis

DNA was digested using a variety of restriction endonucleases under conditions recommended by the manufacturers (Roche Diagnostics Australia, New South Wales, Australia or New England Biolabs Inc. (NEB), Massachusetts, U.S.A.). Restriction digests were performed in a volume of 10-30 µL for between 1-2 h at an appropriate temperature. PCR products were digested overnight when subsequently used for cloning experiments. Restriction reactions were terminated either by heat inactivation (when appropriate, either 65°C for 15 min or 80°C for 20 min), phenol/chloroform extraction or by addition of an appropriate amount of 6 x stop mix (0.1 M EDTA, 0.05% (w/v) bromophenol blue (Progen Industries Ltd., Queensland, Australia), 50% (w/v) sucrose (pH 7.0)), if restriction fragments were to be gel purified.

DNA samples were analyzed by agarose gel electrophoresis using 0.7-2% agarose gels (FMC Bioproducts, Maine, U.S.A.) in either TAE buffer (1 mM EDTA, 38.2 mM Tris-HCl (pH 7.8), 16.6 mM sodium acetate) for larger fragments or TBE buffer (89 mM Tris-HCl (pH 8.0), 89 mM boric acid, 2 mM EDTA) for smaller fragments (TBE was generally only used for gel mobility shift assays performed using 2% agarose gels). Electrophoresis was

carried out at between 10-100 V for between 0.5-16 h. Following electrophoresis, the gels were stained for 1-5 min in 10 μg/mL ethidium bromide and destained in deionised water. The DNA was visualized under UV light using a Spectroline Ultraviolet Transilluminator (Medos Company Pty. Ltd., Victoria, Australia). The size of DNA fragments was estimated by comparison to *Hin*dIII digested λ*cI857* DNA (Promega Corporation) or PCR markers (Promega Corporation) diluted in an appropriate volume of 6 x stop mix, and using the SEQAID II program (D. Rhoads and D. Roufa, Kansas State University, U.S.A.).

b) Isolation of DNA from Agarose Gels

Agarose containing the relevant DNA fragments was excised using a scalpel blade. The DNA was extracted from the gel using the QIAquick Gel extraction Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions and eluted using 30-50 μL of 5-10 mM Tris-HCl (pH 8.0) buffer.

c) DNA Ligation

Ligation reactions were carried out using 3 units of T4 DNA ligase (3 u/μL, Promega Corporation), 0.1 volumes of 10 x ligase buffer (Promega Corporation) and appropriate amounts of vector and insert DNA to an approximate ratio of 1:5. The reaction was performed at 16°C overnight and terminated by heat inactivation at 65°C for 10 min. The ligated DNA was isopropanol precipitated prior to transformation.

d) Southern Hybridization Analysis

Southern hybridization anlaysis was performed with several modifications from the original procedure (Southern, 1975). The DNA was subjected to electrophoresis alongside

DIG-labelled λ *Hin*dIII DNA molecular size markers (Roche Diagnostics Australia), stained with ethidium bromide (10 μg/mL), and the agarose gel photographed. To prepare the DNA for transfer to a nylon membrane, the DNA fragments of high molecular weight were first depurinated by incubating the agarose gel (at a slight angle such that the high molecular weight DNA end of the gel was fully submerged, but the low molecular weight DNA end of the gel was out of the solution), in 0.25 M HCl for 10-15 min at room temperature with gentle agitation. The gel was rinsed briefly in dH_2O and then incubated in denaturation solution (0.2 M NaOH, 0.6 M NaCl) for 30 min. The gel was again rinsed briefly in dH_2O and was then neutralized by incubating twice in neutralization solution (165 mM tri-sodium citrate, 1.5 M NaCl, 0.25 M Tris-HCl (pH 7.5)) for 20 min. Each incubation was carried out at room temperature on a shaker.

The DNA was then transferred to a Hybond-N+ membrane (Amersham Pharmacia Biotech, Buckinghamshire, U.K.) overnight on a support covered with a wick of Whatman 3MM chromatography paper placed within a container that held 500-600 mL of 10 x SSC (165 mM tri-sodium citrate, 1.5 M NaCl). The gel was placed upside down on the wick and the nylon membrane, which was presoaked in 2 x SSC, was placed over the gel. Two pieces of Whatman 3MM paper presoaked in 2 x SSC, a stack of paper towels and a weight (approximately 500 g) were then placed over the membrane. The DNA was allowed to transfer from the gel to the nylon membrane by capillary action overnight. Following transfer, the nylon membrane was removed and air-dried. The DNA was cross-linked to the membrane by exposure to UV light (312 nm) for 3 min.

The membrane was pre-hybridized for a minimum of 3 hours at 65°C in pre-hybridization solution (83 mM tri-sodium citrate, 0.75 M NaCl, 0.05% (w/v) skim milk powder, 1% (v/v) N-laurylsarcosine, 0.02% (v/v) SDS). Probe DNA (50 ng) was added to 5 mL of pre-hybridization solution (to create the hybridization solution) and denatured by boiling

for 10 min followed by immediate incubation on ice for 5 min. Following pre-hybridization, the pre-hybridization solution was removed from the membrane and the hybridization solution was added. Hybridization was carried out overnight under high stringency conditions (65°C). After hybridization, the membrane was washed twice for 5 min at room temperature in 2 x SSC, 0.1% SDS, and twice for 15 min at 65°C in 0.2 x SSC, 0.1% SDS.

DNA-DNA hybrids were identified by chemiluminescent detection using CDP-*Star*™ (Roche Diagnostics Australia) in accordance with the manufacturer's specifications. When required, following detection the membranes were stripped and reprobed as follows. The membrane was rinsed for 1 min in sterile H_2O, followed by two 10 min washes at 37°C in a freshly prepared solution of 0.2 M NaOH, 0.1% SDS. The membrane was then rinsed thoroughly for 5 min in 2 x SSC, placed back into pre-hybridization solution and then hybridized to a different probe.

Polymerase Chain Reaction (PCR)

a) Synthesis of Oligonucleotide Primers

Oligonucleotide primers used in this study are listed in Table 2.3 and were synthesized using a 392 DNA/RNA Synthesizer (Applied Biosystems). Following synthesis the oligonucleotide primers were deprotected by incubation at 55°C for 2 hours and dried in a Heto Maxi-Dry Plus vacuum concentrator. Oligonucleotide primers were resuspended in 100 µL of deionised water and quantitated by measuring the absorbance at 260 nm. An absorbance reading of 1.0 at 260 nm was equated to a single stranded DNA concentration of 33 µg/mL (Sambrook *et al.*, 1989). Both dry and aqueous oligonucleotide primers were stored at –20°C.

Table 2.3 Oligonucleotide Primers

Primer	5'-3' Sequence	Charateristics/Reference/Use
UP	GTTGTAAAACGACGGCCAGT	Universal Primer, PCR-C, SQ
169	TTTCTGACAGCGTTGCA	Internal to *tnpX*, CI PCR
1561	GCGCTGAATATCAGTACC	JIR888 target site, PCR-GS
1562	CTGGCTGAAGACGGTGAATC	JIR888 target site, PCR-GS
1982	CTTCGCCGGTTATCCAGATC	JIR891 target site, PCR-GS
1983	GCGCGGTTCGCTGGTGG	JIR891 target site, PCR-GS
2395	CGATTGGGGCTAAATTATCTGCAAG	Internal to *tnpX*, SQ
4675	AATCAAGCAGGACAAGGAACGA	Internal to *tnpZ*, CI PCR
4676	TGAAGTCCTTGACATTTTCTTA	5' end Tn*4451*, CI PCR
7364	GGGGGAATTCTTAAGCAGGGGTAGTA ATAG	5' end *tnpX*, PCR-C
8186	GACAGATGTTGCTACAGGAG	Flanking Tn*4451*, PCR-GS
8187	GCTACGGGCATAATCTTCGA	Internal to *tnpX*, CI PCR-GS
10232	GGCGAATTCCTACTCTTTGTTGCAATG AGG	3' end $tnpX_{1\text{-}597}$, PCR-C
10233	GGCGAATTCCTATCTGCTTGCATACCC ACC	3' end $tnpX_{1\text{-}356}$, PCR-C
10246	GGCGAATTCCTATTTACACTTGCGTAT CAG	3' end $tnpX_{1\text{-}492}$, PCR-C
10860	TGAATGTCTTTCTGGCTTAG	Internal to *tnpX*, SQ
11096	AATATTTTAGGCAGGCAGGA	Internal to *tnpX*, SQ
11216	GGAACTGCAATCTCTTTGGA	Internal to *tnpX*, SQ
11217	CTTGCCGAGATAGAAACACA	Internal to *tnpX*, SQ
12873	TCACTCGAGAGCAGGGGTAGTAATAG	3' end $tnpX_{1\text{-}707}$, PP
12882	GGGAATTGCATATGTCAAGGACTTC	5' end *tnpX*, PP
13200	TCACTCGAGCTCTTTGTTGCAATGAG	3' end $tnpX_{1\text{-}597}$, PP
13692	ATCCCGCGAAATTAATAC	pET22b+ vector, SQ

15894	GACTCTATGAAAGAACAGAAACA	Internal to *tnpW*, CI PCR-GS
15962	TCTCTCGAGTCTGCTTGCATACCCACC G	3' end $tnpX_{1\text{-}356}$, PP
15963	TCTCTCGAGTTTACACTTGCGTATCAG CGAAA	3' end $tnpX_{1\text{-}492}$, PP
16208	CATTAACATCGAAGCTTGA	3' flanking Tn*4451*, PCR-GS

Oligonucleotide primer purpose or uses are designated as follows: primers used to generate PCR products for cloning (PCR-C), primers used for sequencing (SQ), primers used to generate gel mobility shift DNA fragments (PCR-GS), primers used to generate PCR products for cloning into the protein expression vector pET22b+ (PP), and primers used to generate various PCR products encompassing the joint of the circular intermediate (CI).

b) Polymerase Chain Reactions

Standard PCR

PCR amplification was carried out using *Taq* DNA polymerase (5 units/μL, Roche Diagnostics Australia) in a total volume of 50-100 μL. Each reaction consisted of 0.1 volumes of the 10X reaction buffer supplied by the manufacturer, 0.2 mM dNTP's, 0.4 μM of each primer, between 0.05 and 1.0 μg of template DNA and sterile dH_2O. Amplification was performed in a GeneAmp PCR System 2400 (Perkin Elmer Corporation). The DNA template was first denatured at 94°C for 3 min, then the temperature held at 70°C, at which time 5 units of *Taq* DNA polymerase was added if not already present in the reaction mixture. Amplification was carried out for 30 cycles consisting of a 94°C denaturation step (1 min), 50°C annealing step (2 min) and a 72°C extension step (3 min). The final cycle consisted of ? min annealing at 50°C and 5 min of extension at 72°C. When required, the specificity of the reaction was increased by changing the annealing temperature to 55°C or 60°C and decreasing the extension time to 1 min. PCR products were detected by agarose gel electrophoresis of 5-10 μL samples.

Capillary PCR Reactions

A capillary PCR method was occasionally used for initial screening of recombinant clones as an alternative to extracting plasmid DNA or performing standard PCR reactions. Crude cell extracts were obtained by resuspending selected colonies in 50 μL of sterile dH_2O, vortexing the suspension vigorously for several sec and then lysing the cells by boiling for 10 min. The cellular debris was then removed by centrifugation at 12,000 *g* for 10 min at room temperature. The clear supernatant was removed and stored on ice or at -20°C until required.

A PCR master mixture was prepared and aliquoted into 18 µL reaction mixtures that consisted of 2 µL of 10X PCR reaction buffer, 0.2 mM dNTP's, 1 unit of *Taq* DNA polymerase, 0.4 µM of each primer and the appropriate amount of sterile dH_2O. To each reaction, 2 µL of the cellular extract was added as the DNA template and the reaction mixture was then drawn up in a capillary PCR tube by use of a CP-1 Cycle Prep Auto Gun (Corbett Research, New South Wales, Australia). The tubes were then placed in a FTS-1 Thermal Sequencer (Corbett Research) and the products were PCR amplified. The PCR program consisted of an initial cycle of denaturation at 94°C for 5 min, annealing at 50°C for 2 min and extension at 72°C for 3 min. The program then continued for a further 30 cycles of denaturation at 94°C for 1 min, annealing at 50°C for 1 min and extension at 72°C for 2 min. A final cycle consisted of denaturation at 94°C for 1 min, annealing at 50°C for 1 min and extension at 72°C for 5 min. The PCR products were detected by agarose gel electrophoresis of the entire 20 µL sample.

c) Purification of PCR Products

PCR products were purified for cloning or sequencing either directly using Magic™ PCR Preps Purification System (Promega Corporation), were extracted from agarose gels using the QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's instructions, or were isopropanol precipitated.

d) Preparation of Digoxigenin (DIG) Labelled PCR DNA Probes

PCR fragments to be used as probes were labelled with digoxigenin-11-dUTP (DIG-dUTP) during PCR amplification. PCR amplification was carried out using 5 units of *Taq* DNA polymerase (5 units/µL, Roche Diagnostics Australia) in a total volume of 100 µL. Each

reaction consisted of 10 µL of the 10X reaction buffer supplied by the manufacturer, 0.2 mM dNTP's, 0.4 µM of each primer, between 0.05 and 1.0 µg of template DNA and sterile dH_2O. In addition, 2 µL of DIG DNA Labelling Mix (1 mM dATP, 1 mM dCTP, 1 mM dGTP, 0.65 mM dTTP, 0.35 mM DIG-dUTP)(Roche Diagnostics Australia) was added to each reaction. The DNA was amplified in a GeneAmp PCR System 2400 as previously described. Amplification of DNA was confirmed by electrophoresis in agarose gels and the PCR product was then purified directly from the reaction using the Magic™ PCR Preps Purification System. The efficiency (ng labelled DNA/µL) of the labelling reaction was determined using the quantitation protocol supplied by the manufacturer (Roche Diagnostics Australia).

Nucleotide Sequencing

a) Automated Sequencing

Plasmid DNA to be sequenced was generated using the Applied Biosystems method already described. Cycle sequencing reactions were carried out on a GeneAmp PCR System 2400, using a PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase, FS (Applied Biosystems), according to the manufacturer's instructions. Excess dye terminators were removed by ethanol/sodium acetate precipitation using a protocol supplied by the manufacturer. Sequencing samples were resolved and analyzed on a 373 DNA STRETCH Sequencer (Applied Biosystems).

b) Manual Sequencing

Manual sequencing reactions were carried out to provide a reference sequence in order to determine the position of protection during DNaseI footprinting experiments. Sequencing

reactions were performed using the T7 Sequencing™ kit (USB) using [α-^{35}S] dATPαS (Perkin Elmer). Appropriate plasmid DNA templates (1 pmol) were used in conjunction with oligonucleotide primers (5 mM) used to generate the gel mobility and DNaseI protection assay DNA targets. Sequencing gels consisted of 8% acrylamide (acrylamide:*N,N'*-methylenebis-acrylamide, 19:1) (Astral Scientific, Gymea, Australia) and 0.5% (w/v) urea in 1 x TBEseq (100 mM Tris-HCl (pH 8.3), 2 mM EDTA and 100 mM boric acid). The gel was polymerized by addition of 300 µL of 10% (w/v) ammonium peroxodisulfate (MERCK, Darmstadt, Germany) and 34 µL of *N, N, N', N'*-tetramethylethylenediamine (TEMED) (BDH, Poole, England) The acrylamide solution was poured between two glass plates with 0.2 mm spacers and left to polymerize overnight. The gel was run using a Sequencing Gel Electrophoresis System Model S2 (Bethesda Research Laboratories, Gaithersburg, U.S.A.). Sequencing reactions (2 µL) were heated to 80°C for 3 min before loading onto the sequencing gel. The gel was run at a constant temperature of 50°C (75-80 W) for several hours before being lifted using chromatography paper and vacuum dried at 80°C for 1-2 hours using a Drygel Sr. Slab Gel Dryer Model SE 1160 (Hoefer Scientific Instruments, San Francisco, U.S.A.). The dried gels were exposed to Fuji X-ray film (Fuji, Tokyo, Japan) for up to 2 months and were developed using a Fuji FMP-100A film processor.

c) Sequence Analysis

Nucleotide sequences obtained from automated sequencing were compiled and compared using Sequencher™ 3.1 software (GeneCodes Corporation, Michigan, U.S.A.). Amino acid similarity searches were performed using the BLAST program at the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). Primary amino acid sequences were analyzed using the consensus secondary structure prediction and coiled–coil prediction programs from the Network Protein Sequence @nalysis (NPS@)

(http://npsa-pbil.ibcp.fr/). Multiple amino acid sequence alignments were produced using a CLUSTAL W program (Thompson *et al.*, 1994).

Random Mutagenesis of tnpX

a) XL1-Red Mutagenesis

Mutagenesis of pJIR1872 was achieved by passage of the plasmid through the *E. coli* mutator strain XL1-Red (Stratagene, California, U.S.A.) according to the manufacturer's instructions, with a few modifications. The Epicurian Coli® XL1-Red competent cells were thawed on ice, 3.4 µL of the β-mercaptoethanol provided with the kit was added to 200 µL of the thawed cells and mixed gently by inversion. The cells were dispensed into 40 µL volumes in pre-chilled 15 mL Falcon polypropylene tubes and the contents of the tubes were swirled gently every two min, while incubating on ice for 10 min. pJIR1872 DNA (25 ng), which had been isolated using the Applied Biosystems method, was added to the cells, the contents were swirled gently, and the tube incubated on ice for 30 min. The cells were heat shocked at 42°C for 50 seconds and immediately chilled on ice for 2 min. Pre-heated (42°C) SOC broth (0.9 mL) was added and the transformation mixture was incubated at 37°C for 1 h with moderate agitation. The transformation mixture was concentrated by centrifugation at 12,000 *g* for 10 min at room temperature, the supernatant discarded and the resulting cell pellet resuspended in 100 µL of fresh SOC broth. The entire transformation mixture was plated onto a single LB (1% (w/v) NaCl, 1% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 2% (w/v) agar, pH 7.0) agar plate, which was supplemented with kanamycin (20 µg/mL), and incubated for 24-30 h at 37°C.

b) Screening Procedure

Approximately 200 of the resultant transformants were used to inoculate 10 mL of LB broth supplemented with kanamycin (20 μg/mL) and were grown overnight at 37°C. Plasmid DNA was extracted from the overnight culture using the Applied Biosystems method and 2.5 μL of the DNA was used to transform *E. coli* strain JIR7171 to kanamycin resistance. Appropriate dilutions of the transformation mixture were made and 100 μL aliquots were plated on 2 x YT agar supplemented with kanamycin (20 μg/mL) and chloramphenicol (40 μg/mL), and incubated at 37°C for no longer than 18 h.

The largest colonies on the transformation plates were replated onto 2 x YT agar containing kanamycin (20 μg/mL) and chloramphenicol (40 μg/mL). From these plates, single colonies were used to inoculate 1 mL 2 x YT broths containing appropriate antibiotics and incubated at 37°C. Plasmid DNA was extracted using the alkaline lysis method and used to perform PCR to detect the presence of circular intermediate DNA, using oligonucleotide primers 4675 and 8187. Plasmid DNA producing very little or no circular intermediate PCR product was subsequently diluted 1 in 10 and used to transform DH5α to kanamycin resistance. The resultant colonies were cross-patched onto ampicillin and kanamycin 2 x YT plates to confirm the absence of the pJIR1880 plasmid. Five mL of 2 x YT broth containing kanamycin was subsequently inoculated from a kanamycin resistant, ampicillin sensitive colony and incubated overnight at 37°C with shaking. Plasmid DNA was extracted from the culture using the Applied Biosystems method and used for sequencing and reconfirmation of the chloramphenicol resistant phenotype by transformation of JIR7171 to ampicillin resistance and detection of the presence or absence of the circular intermediate by PCR amplification as before.

In vivo *Assays for TnpX Function*

a) Tn*4451tnpXΔ1* Stability Assay

Truncated tnpX *Derivatives*

Assay strains were constructed by introducing plasmids containing the truncated *tnpX* derivatives, and positive and negative controls, into strain JIR2463 using the $CaCl_2$-competency method of transformation. JIR2463 contains pJIR773, which carries Tn*4451tnpXΔ1*. Transformants were selected on 2 x YT plates containing kanamycin (10 μg/mL), chloramphenicol (10 μg/mL) and ampicillin (100 μg/mL). Stability assays were performed immediately after strain construction by inoculating 5 mL of 2 x YT broth containing kanamycin (20 μg/mL) and ampicillin (100 μg/mL) with each assay strain and incubating overnight at 37°C, with shaking. Plasmid DNA was extracted (Alkaline Lysis method) from the overnight cultures and used to transform DH5α cells to kanamycin resistance (20 μg/mL). The kanamycin resistant colonies were subsequently patched onto 2 x YT agar containing either chloramphenicol (30 μg/mL) or kanamycin (20 μg/mL). After overnight incubation at 37°C, the percentage of chloramphenicol resistant colonies was calculated and the plasmid DNA extracted earlier was used as a template to perform a PCR designed to detect the circular intermediate using the oligonucleotide primers 8187 and 4675. The negative control was the vector alone, pUC18, and the positive control was pJIR1898, containing wild-type *tnpX*. These were included in every experiment.

TnpX Random Mutants

Stability assays were performed as described above with some modifications. The host strain was JIR2343, which carried the Tn*4451tnpXΔ1* plasmid pJIR683. The plasmid DNA was extracted as before and used to transform DH5α cells to ampicillin resistance

(100 μg/mL). The ampicillin resistant colonies were subsequently patched onto chloramphenicol (30 μg/mL) and ampicillin plates (100 μg/mL) and the percentage of chloramphenicol resistant colonies was calculated. The negative control contained pSU39 and the positive control contained pJIR1872, as before the positive and negative control strains were included in each experiment.

b) Insertion Assay

Truncated tnpX *Derivatives*

The base strain for truncated *tnpX* derivative insertion assays was JIR7909, which carried the temperature sensitive CI plasmid pJIR2376 (Table 2.2). This strain was kanamycin resistant (20 μg/mL) at the permissive temperature of 30°C. The various *tnpX* plasmids were introduced into this strain using the $CaCl_2$-competency method with a few modifications. Due to the temperature sensitive origin of replication of pJIR2376, the transformation procedure was altered slightly by conducting the expression step at 30°C for 2 h with shaking rather than 1 h at 37°C. The transformation mixtures were also incubated at 30°C, rather than 37°C, on 2 x YT plates containing kanamycin (20 μg/mL) and ampicillin (100 μg/mL).

Insertion assays were conducted by using a single colony to inoculate 5 mL of 2 x YT broth containing kanamycin (20 μg/mL) and ampicillin (100 μg/mL) and incubating overnight at 30°C with shaking. The cultures were then diluted 1 in 10 into fresh 2 x YT broths (5 mL) with appropriate antibiotics and grown to mid-late log phase at 30°C (3-4 h) with shaking. The cultures were diluted, plated in duplicate onto 2 x YT agar containing kanamycin (20 μg/mL) and ampicillin (100 μg/mL), and incubation at either 30°C or 42°C for two days. The insertion frequency was defined as the ratio of the viable count at 42°C

over the viable count at 30°C. The positive control contained the plasmid pJIR1898 and the negative control strain contained pUC18.

TnpX Random Mutants

Random mutant insertion assays were performed as above except that the host strain used was JIR7998, containing the temperature sensitive chloramphenicol resistance plasmid pJIR2421. The antibiotics added to both 2 x YT broths and plates was kanamycin (20 μg/mL) and chloramphenicol (30 μg/mL). The insertion frequency was calculated as above. In this case the negative control strain contained pSU39 and the positive control strain contained pJIR1872.

c) Tn*4453a* Transposition Assay

Truncated tnpX *Derivatives*

Transposition donor strains were constructed using the base strain JIR5724, which carried Tn*4453a* on the chromosome and harboured the conjugative tetracycline resistance plasmid pVS520. The truncated *tnpX* derivatives were introduced into JIR5724 using the $CaCl_2$-competency method, selecting on 2 x YT plates containing tetracycline (10 μg/mL), chloramphenicol (10 μg/mL) and ampicillin (100 μg/mL). The second strain required for the transposition assay was the recipient strain LT101 which is resistant to rifampicin (150 μg/mL).

Overnight cultures of both donor and recipient strains were diluted 1 in 20 into 5 mL of fresh 2 x YT broth containing the appropriate antibiotics and grown at 37°C with shaking until the cultures had reached mid-log phase (about 3 h). Filter matings were performed by mixing 0.5 mL of both donor and recipient in a 2.5 mL sterile syringe and introducing the

mating mixture onto a sterile 0.45 μm nitrocellulose filter. The filters were placed onto 2 x YT plates without antibiotics and incubated at 37°C for 2 h. After incubation, the bacteria present on the filters were resuspended by vortexing in 2 mL of 2 x YT broth. The suspensions were diluted and plated onto 2 x YT plates containing rifampicin (150 μg/mL) and tetracycline (10 μg/mL) or rifampicin (150 μg/mL) and chloramphenicol (20 μg/mL). The transposition frequency was defined as the ratio of the viable counts on the rifampicin and chloramphenicol medium to the viable count on the rifampicin and tetracycline medium. The positive control strain contained pJIR1898 and the negative control strain contained pUC18.

TnpX Random Mutants

The random *tnpX* mutants were tested in almost the same way as the truncated *tnpX* derivatives. The donor strains were constructed as above except that the strains were selected using 2 x YT plates containing tetracycline (10 μg/mL), chloramphenicol (10 μg/mL) and kanamycin (20 μg/mL) instead of tetracycline, chloramphenicol and ampicillin. The donor strains were subsequently incubated in 2 x YT broth containing tetracycline (10 μg/mL), chloramphenicol (10 μg/mL) and kanamycin (20 μg/mL), not ampicillin. As with all random mutant assays the positive control strain contained pJIR1872 and the negative control strain contained pSU39.

In vitro *Assays using Purified TnpX Protein*

a) Excision Assay

Excision assays were performed using extensively dialyzed CsCl-purified pJIR1452 DNA and purified $TnpX_{1-707}$. Reactions were carried out in 30 μL volumes unless otherwise

stated and typically contained 10 mM Tris-HCl buffer (pH 8.0), 100 mM NaCl and 1 mM β-mercaptoethanol. Approximately 200 ng of plasmid DNA was added so that reactions could be readily visualized on an ethidium bromide stained 0.7% agarose gel. Varying amounts of $TnpX_{1-707}$ were added and the reactions incubated at 37°C for 30 min unless otherwise indicated, followed by heat inactivation at 80°C for 20 min to denature and remove $TnpX_{1-707}$ from the DNA. An aliquot was then taken for PCR analysis and $MgCl_2$ was added to the remaining mixture, to a final concentration of 6 mM, to allow for subsequent restriction enzyme digestion of the DNA with *Eco*RI and *Eco*RV for 2 h at 37°C. Stop mix was added to the digested products which were separated by electrophoresis on a 0.7% agarose gel at 15 V for approximately 16-18 hours after which the DNA was visualized by staining with ethidium bromide. PCR analysis was performed using the aliquot removed previously as a template, with oligonucleotide primers 15894 and 8187 in a 50 µL volume using a 60°C annealing temperature and a 1 min extension time.

DNA preparation for supercoiling specificity

Plasmid pJIR1452 DNA was prepared for assays to determine supercoiling specificity as follows. Open circular DNA was produced by treating CsCl-purified DNA with 1 unit of DNaseI in the presence of 360 µg/mL of ethidium bromide in a total volume of 50 µL, incubated at 30°C for 15 min. Linear DNA was generated by digestion with the restriction enzyme *Bgl*II. Open circular, linear and super-coiled DNA species to be used in further excision assays were purified from a 0.8% agarose gel using the Qiaquick gel extraction kit. Southern hybridization analysis was performed as described using a randomly DIG labelled PCR product (oligonucleotide primers 8187/15894) encompassing the joint of the circular intermediate of Tn*4451* as a probe.

b) Insertion Assay

Plasmid Substrate

In vitro insertion assays were performed using CsCl-purified plasmid DNA. The plasmids used were pJIR2159 and pJIR2292 or pJIR2159 and pJIR2501. Reactions were performed in 30 μL volumes and contained 10 mM Tris-HCl buffer (pH 8.0), 100 mM NaCl, 5 mM $MgCl_2$, 10 mM $CaCl_2$ and 1 mM β-mercaptoethanol. Sufficient DNA was added so that the DNA could be readily visualized on an ethidium bromide stained 0.7% agarose gel (100-200 ng of DNA). Approximately 1 μg of $TnpX_{1-707}$ was added to reactions unless otherwise stated. Typically, reactions were incubated at 37°C for 3 h then heat inactivated at 80°C for 20 min. Small aliquots were removed from the reactions for PCR analysis, stop mix was added and the reactions loaded on to an agarose gel and subjected to electrophoresis at 15 v overnight, after which the gel was stained and the DNA visualized.

PCR Substrate

In vitro insertion assays were performed using gel extracted PCR products. The products were generated using oligonucleotide primer pairs CI1: 169/4675, CI2: 8187/15894 and T: 8186/16208 (Table 2.3). Reactions were performed in 15 μL volumes and contained 10 mM Tris-HCl buffer (pH 8.0), 100 mM NaCl, 5 mM $MgCl_2$, 10 mM $CaCl_2$ and 1 mM β-mercaptoethanol. Sufficient DNA was added so that the DNA could be readily visualized on an ethidium bromide stained 1% agarose gel (100-200 ng of DNA). The same amount of protein was added as previously indicated. The reactions were incubated in the same fashion as before, separated via electrophoresis as described and the gel subjected to Southern hybridization analysis as for *in vitro* excision assays.

Analysis and Detection of the TnpX Protein

a) Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

To visualize purified proteins and whole cell lysates, protein samples were separated by electrophoresis in 0.75 mm thick, 12% denaturing polyacrylamide gels (Laemmli, 1970). Two minigels were prepared in a mini Protean II casting stand (Bio-Rad Laboratories), which was assembled as described by the manufacturer. The 12% resolving gel solution consisted of 4 mL of resolving gel buffer (1.5 M Tris-HCl, 0.4% (w/v) SDS, pH 8.8), 4.8 mL of 40% acrylamide solution (Amresco, Ohio, U.S.A.) (acrylamide: N,N'-methylene bis-acrylamide, 37:5:1) and 7.2 ml of dH_2O. To catalyze polymerization, 60 µL of 10% ammonium peroxodisulfate (AP) (MERCK KgaA, Darmstadt, Germany) and 17.5 µL of *N, N, N', N'*-tetramethylethylenediamine (TEMED) (Amresco) were added to the acrylamide solution, which was subsequently poured between the glass plates and overlaid with H_2O.

Once the resolving gel had set, a 4% stacking gel was prepared from 1.25 mL of stacking gel buffer (0.5 M Tris-HCl, 0.4% (w/v) SDS, pH 6.8), 500 µL of 40% acrylamide solution and 3.25 mL of dH_2O. To initiate polymerization, 30 µL of 10% AP and 5 µL of TEMED were added to the stacking gel solution. The mixture was then poured over the resolving gel, and a 10 well Teflon comb was inserted between the glass plates into the stacking gel solution.

While the stacking gel was polymerizing, the protein samples were prepared by the addition of an equal volume of 2X sample buffer (0.013% bromophenol blue, 10% β-mercaptoethanol, 20% glycerol, 2.5% SDS, 4 mL stacking gel buffer). Cell lysates were

denatured by boiling for 5 min, followed by centrifugation at 12,000 *g* for 2 min to pellet the cellular debris. Samples from protein purification fractions were incubated at 37°C for 10 min immediately before loading the gel, as the acid labile bonds in samples containing imidazole are partially hydrolyzed when boiled (QIAexpressionist Protein Purification Manual, Qiagen).

Protein samples were applied to the gel, which was then subjected to electrophoresis in Tris-glycine protein gel electrophoresis buffer (12.5 mM Tris-HCl, 100 mM glycine, 0.05% SDS, pH 8.3) in a mini Protean II Electrophoresis Cell (Bio-Rad Laboratories) at 200 V until the bromophenol blue dye front had reached the bottom of the gel. To estimate protein sizes, 5 µL of low molecular size standards from the LMW Electrophoresis Calibration Kit (Amersham Pharmacia Biotech), or 15 µL of SeeBlue™ pre-stained standards (Invitrogen), were run alongside the protein samples for coomassie stained gels.

To visualize the protein bands, the gels were stained with coomassie brilliant blue (0.2% coomassie brilliant blue (BDH Chemicals Australia Pty. Ltd., Victoria, Australia), 25% (v/v) isopropanol, 7% (v/v) glacial acetic acid) for 30 min at room temperature, followed by destaining in 33% (v/v) methanol, 7% (v/v) glacial acetic acid.

Purification of C-terminal His-tagged TnpX and Derivatives

a) Growth Conditions, Induction and Preparation of *E. coli* Cell Lysate

Each time a purification was to be carried out freshly transformed C43(DE3) cells were used. A fresh 2 x YT (ampicillin 100 µg/mL) plate was inoculated with a single transformant and incubated overnight at 37°C. The growth from this plate was used to inoculate 2 x 10 mL of 2 x YT broths supplemented with ampicillin (100 µg/mL). After

incubation at 37°C for 1-2 hours, the cultures were used to inoculate 2 x 500 mL of the same medium. These cultures were grown at 37°C until they reached a turbidity at 600 nm of 0.5, transferred to 30°C and induced with 0.5 mM IPTG for 4 h. Cells were pelleted by centrifugation at 4000 *g* for 10 min. The pellet was washed using PBS (phosphate-buffered saline) and stored at –70°C until needed.

When required the cells were resuspended in 40 mL of cold buffer A (50 mM Tris-HCl (pH7.0), 500 mM NaCl) containing 2 mM PMSF (phenylmethanesulfonyl fluoride) and 0.1% Triton X-100 and disrupted twice in a pre-cooled French Press. The lysate was sonicated using 2 x 30 sec bursts on ice and centrifuged for 45 min at 17,000 *g* at 4°C.

b) Purification of His-tagged Proteins

The supernatant was incubated for 30 min, rotating at 4°C, with 2 mL of Talon (Clontech) slurry pre-equilibrated in buffer A. The slurry was then packed into a 12 mL column at 4°C. The column was washed sequentially with 100 mL of buffer A, 20 mL of buffer A containing 5 mM imidazole, 10 mL of buffer A containing 10 mM imidazole and 10 mL of buffer A containing 25 mM imidazole. His–tagged protein was eluted from the column using either a 30 mL linear gradient of 25-400 mM of imidazole in buffer A or a step gradient using 4 mL of buffer A containing 50 mM, 100 mM, 200 mM and then 400 mM imidazole. Fractions of 1 mL were collected and 15 µL of each fraction were subjected to 10% SDS-PAGE and the gel was stained with coomassie brilliant blue. Recombinant TnpX was eluted in a broad peak between 50 and 400 mM imidazole. The purest fractions were pooled and dialyzed into 50 mM Tris-HCl buffer (pH 7.0) containing 350 mM NaCl, 2 mM DTT (dithiothreitol) and 50% glycerol.

c) Gel Filtration

Further purification was accomplished after Talon chromatography by gel filtration using a Superdex 200 16/60 Pharmacia column equilibrated in 50 mM Tris-HCl buffer (pH 7.0) containing 2 mM DTT and 350 mM NaCl. A protein sample (1-2 mg) was loaded onto the column, which was then eluted with the same buffer at a flow rate of 0.4 mL/min. Fractions of 1 mL were collected and purity assessed by 12% SDS-PAGE of 15 µL samples. The cleanest and most concentrated fractions were pooled and concentrated using vivaspin (Vivascience) 30 kDa exclusion limit columns. The size of native proteins was calculated after calibration of the gel filtration column, performed using standard proteins (Gel filtration Calibration Kit, Bio-Rad). Gel filtration data are presented as elution volumes. The apparent molecular size of each protein was determined by interpolation from the standard curve of log molecular size *versus* K_{av}. Purified protein was stored after dialysis into 50% glycerol at –70°C.

d) Determination of Protein Concentration

Protein concentration was calculated by determining the absorbance of the protein solution at 280 nm. The extinction coefficient for each protein was calculated by averaging the reduced and non-reduced (cys residues) extinction coefficient values using the ProtParam Tool from the ExPASy protein analysis site (http://au.expasy.org/). This value was used to calculate the protein concentration from the absorbance reading using the following formula:

$$\frac{A_{280} \text{ x dilution factor}}{\text{extinction coefficient}} = \text{Protein concentration (M)}$$

Analysis of DNA-Protein Interactions

a) Derivation of Target DNA

The target DNA used for gel mobility shift and DNaseI protection assays was generated using PCR amplification. The left end fragment was 360 bp in length, consisting of flanking DNA followed by the GA dinucleotide, approximately in the middle of the fragment, and then the transposon-derived sequence. The left end PCR product was produced using oligonucleotide primers 8186 and 8187. The right end fragment was 343 bp in length and was generated by PCR amplification using the oligonucleotide primers 15894 and 16208. Similar to the left end fragment, this fragment included flanking DNA and transposon derived sequences with the GA dinucleotide positioned approximately in the middle of the DNA fragment. The circular intermediate fragment was generated using the oligonucleotide primers 15894 and 8187 and PCR amplification to produce a fragment 405 bp in length, and again the GA dinucleotide was approximately in the middle of the fragment.

Insertion sites from *E. coli* strains JIR888 and JIR891, previously identified and partially sequenced (Abraham and Rood, 1987; Crellin and Rood, 1997) were genereated by PCR amplification from genomic DNA, using the oligonucleotide primer pairs 1561/1562 and 1982/1983. A fragment containing a deletion site originally from *C. perfringens* was generated using oligonucleotide primers 8186 and 16208 using pJIR47 as the DNA template.

b) Labelling of Target DNA for Gel Mobility Shift Assays

DNA targets were labelled in one of two ways using the DIG Gel Shift Kit (Roche Bohringer Mannheim, Australia). The first involved labelling with DIG-11-dUTP at the 3'

end of the DNA fragment while the alternative was for quantitative analysis and required the use of [^{32}P]-ATP. Labelling was accomplished in the same way. The labelling reaction (20 µL) contained 4 µL of 5 x concentrated labelling buffer (1 M potassium cacodylate, 125 mM Tris-HCl and 1.25 mg/mL bovine serum albumin, pH 6.6), 5 mM $CoCl_2$, 50 µM DIG-11-ddUTP (or 1 µL of α–[^{32}P]-ATP (6000 Ci/mmol, 10 mCi/mL (NEN)) for radiolabelled DNA), 1 unit of terminal transferase, and approximately 1 µg of PCR product. The reaction was incubated for 15 min at 37°C. The labelled DNA was precipitated with 0.1 volumes of 4 M LiCl and 2.5 volumes of 100% ethanol and incubation at –70°C for 30 min. The DNA was pelleted at 12,000 *g* for 15 min at 4°C. The DNA pellet was washed three times with 0.5 mL of cold 70% ethanol. The pellet was dried and the DNA resuspended in 10 µL of TEN buffer (10 mM Tris-HCl, 1 mM EDTA and 1 M NaCl, pH 8.0).

The newly labelled DNA was compared to a DIG-labelled DNA solution of known concentration provided in the kit. The DIG-labelled DNA was diluted and 1 µL aliquots of DNA solution spotted onto a nylon membrane. The DIG-labelled DNA was detected as described below. The intensity of newly labelled DNA spots was visually compared to the known standards and the concentration of the newly labelled DNA was estimated. The membrane was washed for 5 min in washing buffer consisting of buffer 1 (0.1 M maleic acid, 0.15 M NaCl, pH 7.5) containing 0.3% (v/v) of Tween 20. The membrane was subsequently blocked in buffer 2 (buffer 1 containing 1% (w/v) skim milk powder) for 30 min with shaking at room temperature. The DIG-labelled DNA was detected by incubating the membrane in anti-DIG-AP fragment diluted 1/10 000 in buffer 2 for 30 min followed by 2 x 15 min washes (washing buffer), equilibration in buffer 3 (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5) for 5 min. The substrate, 100 x concentrated CSPD® (Roche) diluted 1 in 100 in buffer 3, was then added. The membrane was incubated for 5 min in a

hybridization bag, transferred to a clean hybridization bag, and incubated for 15 min at 37°C as directed by the manufacturer before exposure. The membrane was exposed to Fuji X-ray film for an appropriate length of time (30 s to 30 min). The concentration of the labelled DNA was estimated from the DIG Gel Shift Kit (Roche) standards and the DNA was adjusted to a concentration of 7.5 fmol/μL.

Radiolabelled DNA was quantitated after determining the absorbance at 260 nm. An absorbance reading of 1.0 at 260 nm was equated to a double-stranded DNA concentration of 66 μg/mL (Sambrook *et al.*, 1989). The DNA concentration was adjusted to 7.5 fmol/μL and the amount of radioactivity was determined by counting 1 μL of DNA solution using a scintillation counter.

c) Gel Mobility Shift Assays

Gel mobility shift assays were performed essentially as described in the DIG Gel Shift Kit. Reactions typically contained 0.2 volumes of 5 x binding buffer (100 mM Hepes (pH 7.6), 5 mM EDTA, 50 mM ammonium sulfate, 5 mM DTT, 10% (w/v) Tween 20 and 150 mM potassium chloride), 1 μg of poly[d(I-C)], 1 μg poly-Lysine, 15 fmol of labelled DNA and a known amount of protein. The reaction volume was adjusted to 20 μL with sterile water. After addition of protein, the reaction was allowed to proceed for 15 min at room temperature after which 5 μL of loading buffer (60% 0.25 x TBE and 40% glycerol) was added and the tubes were placed on ice. The binding reactions were loaded onto a 4-6% polyacrylamide gel in 0.5 x TBE buffer (1 x TBE contains 89 mM Tris-HCl, 89 mM Boric acid and 2 mM EDTA, pH 8.0). The gel was pre-electrophoresed in 0.5 x TBE electrophoresis buffer at 21 mA before the binding reactions were loaded. After loading, electrophoresis continued at 4°C until the tracking dye (60% 0.25 x TBE, 40% glycerol and 0.2% (w/v) bromophenol blue) had run off the end of the gel.

After electrophoresis, the gel was removed and subjected to capillary transfer if the DNA was DIG labelled or, if radiolabelled DNA was used, vacuum dried and exposed to autoradiographic film or a phosphor screen (Molecular Dynamics, Sunnyvale, U.S.A.) for quantitation. The phosphor screen was scanned using the STORM™ 690 Phosphorimager (Molecular Dynamics). Quantitative data was obtained using ImageQuant software (Molecular Dynamics) to determine the volume of pixels within a standard shape placed around the bound and unbound DNA within the scanned image of the gel. The capillary transfer of DIG-labelled reactions was accomplished using 0.5 x TBE buffer and the DNA was transferred overnight to a nylon membrane. The DNA was fixed to the membrane by exposure for 3 min to UV light and the DNA was detected as described for DNA quantitation.

Due to the large size of the circular intermediate DNA fragment and the size of the complexes formed, circular intermediate binding reactions were loaded onto a large 2% agarose gel and subjected to electrophoresis in 0.5 x TBE at 4°C and 200 V until the tracking dye neared the bottom of the gel. This gel was then treated similarly to the polyacrylamide gels used for the other DNA targets, depending on whether the DNA was DIG-labelled or radiolabelled.

d) DNaseI Protection Assays

Labelling of Target DNA

The target DNA was labelled on one strand only using the following procedure. One oligonucleotide primer was labelled prior to PCR amplification through the action of polynucleotide kinase (Promega). The labelling reaction contained 1 μL of 10 x polynucleotide kinase buffer, 40 pmol of oligonucleotide primer, 5 μL γ-[^{32}P] ATP (6000

Ci/mmol, 10 mCi/mL (NEN)), 1 u of polynucleotide kinase and water to 10 μL. The reaction was mixed and incubated at 37°C for 30 min followed by heat inactivation (95°C for 3 min). The labelled oligonucleotide primer was subsequently used in a PCR reaction containing the following constituents: 5 μL of 10 x PCR buffer, 1 μL of template DNA (a gel purified PCR product diluted 1 in 50), 10 μL of labelled oligonucleotide primer, 40 pmol of the second oligonucleotide primer, 4 μL of 5 mM dNTPs, 1 μL of *Taq* DNA polymerase (5 units) and sterile water to a total volume of 50 μL. The PCR program had an annealing temperature of 60°C and an extension time of 1 min. The resultant PCR product was purified using the Wizard® PCR Preps DNA Purification System (Promega) according to the manufacturer's instructions and eluted in 50 mL of 5 mM Tris-HCl (pH 8.0) buffer and counted using a scintillation counter and the DNA was diluted to 18,000 cpm/μL.

DNaseI Protection Assays

DNaseI protection assay reactions were carried out as follows. Reactions contained labelled DNA (36,000 cpm), 16 μL of 5 x modified gel mobility shift binding buffer (100 mM Hepes (pH 7.6), 50 mM ammonium sulfate, 5 mM DTT, 10% (w/v) Tween 20 and 150 mM potassium chloride), 5 mM $MgCl_2$, 1 μg poly[d(I-C)], 3-6 μg of purified TnpX (or 10 μg of BSA for no TnpX controls) and sterile water to 80 μL. The reactions were incubated at room temperature for 15 min after which 2 units of RQ1 RNase-free DNase (Promega) was added and the samples incubated at room temperature for exactly 2 min. Subsequently, 100 μL of phenol/chloroform was added to stop digestion and the contents of the tubes were mixed. The aqueous volume was adjusted to 100 μL and 1 μL of 10 mg/mL tRNA was added prior to centrifugation at 12,000 *g* for 5 min to separate the phases. The aqueous layer was removed to a new tube and the DNA fragments were precipitated with the addition of 0.1 volumes of 3 M sodium acetate (pH 4.6) and 2.5

volumes of ice cold 100% ethanol. The reactions were incubated at –20°C for 30-45 min and the DNA was pelleted by centrifugation at 12,000 *g* at 4°C for 20 min. The DNA pellet was washed twice with 0.25 mL of ice cold 70% ethanol, air dried for 10 min and the DNA resuspended in 3 µL of loading buffer from the T7 Sequencing™ Kit (Amersham Pharmacia Biotech).

The partially digested DNA fragments were separated on an 8% polyacrylamide sequencing gel alongside appropriate sequencing reactions, as described previously. The gel was pre-electrophoresed in 1 x TBEseq buffer at 75-80 W until the gel reached a temperature of 50°C. Footprinting reactions (3 µL) and manual sequencing reactions (2 µL) were loaded onto the pre-electrophoresed gel and electrophoresis was performed such that the temperature of the gel remained at approximately 50°C (usually 76-78 W), until the bromophenol blue dye front had run off the gel. The gel was then removed onto chromatography paper and vacuum dried for 45 min to 1 h at 80°C before exposure to X-ray film.

Chapter 3

Purification of TnpX and *In Vitro* Assays

Introduction

Previous studies have shown that deletion of a small internal portion of the *tnpX* gene from Tn*4451* renders the transposon incapable of excision (Crellin and Rood, 1997). Provision of a functional *tnpX* gene *in trans* restores the excision phenotype (Crellin and Rood, 1997). Other data from studies of both Tn*4451* and Tn*4453a* have shown that provision of multiple copies of the *tnpX* gene increases the transposition frequency (Lyras *et al.*, 1998). These data provide evidence that TnpX is the protein responsible for movement of the Tn*4451/53* family of elements.

Two of the other five genes encoded by this family of elements have been asigned functions: *catP* confers chloramphenicol resistance and *tnpZ* has been shown to encode a functional mobilization protein which, in conjunction with an upstream RS_A site, catalyzes mobilization of the circular transposon intermediate to another cell in the presence of a conjugative element (Crellin and Rood, 1998; Lyras *et al.*, 1998).

The remaining genes have no known function. Genetic analysis of transposon derivatives in which *tnpV, tnpY* and *tnpW* have been separately deleted and replaced with an erythromycin resistance gene have indicated that none of these genes is essential for either excision or transposition in *E. coli* (D. Lyras and J. I. Rood, unpublished results). As a consequence it was postulated that TnpX was the sole protein required for excision and insertion of the transposon. To examine this hypothesis it was decided to overexpress and purify TnpX and to carry out *in vitro* studies on the resultant protein.

Results

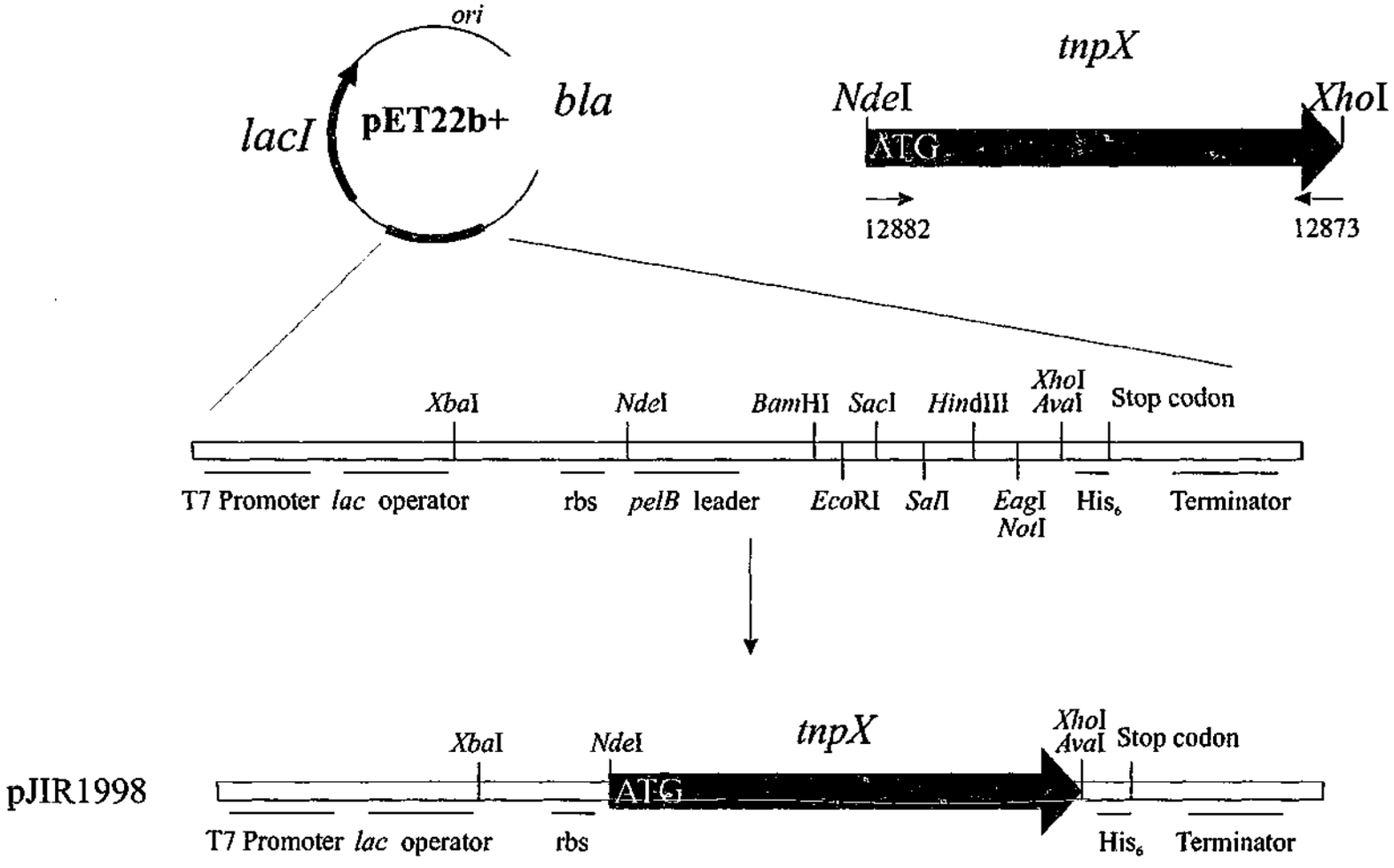
a) Construction of a TnpX Expression Plasmid.

Considerable analysis of transposition mechanisms can be performed within a cellular context. However, fine details of the transposition process can often be more easily dissected in an acellular environment. To this end purified protein must be obtained. Resolvase proteins have historically been difficult to purify as they exhibit a tendency to aggregate and become insoluble (Glasgow *et al.*, 1989; Sherratt, 1989). It was therefore decided that purification of TnpX could be most easily accomplished by creating a tagged fusion protein.

Preliminary work involved the cloning of the *tnpX* gene into the N-terminal hexa-His tag vector pRSETA (Stratgene) but resulted in poor expression levels and a high degree of protein degradation (data not shown). The expression plasmid pET22b+ was therefore chosen for expression of the recombinant TnpX protein as it allowed fusion to a C-terminal hexa-His tag and did not result in as many additional amino acids as pRSETA. To clone the *tnpX* gene behind the T7 promoter it was necessary to introduce two restriction sites, *Nde*I and *Xho*I, immediately upstream and downstream, respectively, of the *tnpX* coding sequence. PCR amplification using oligonucleotide primers 12882 and 12873 was used to generate the required sequence, which included an *Nde*I site at the beginning of the gene and an *Xho*I site in place of the *tnpX* stop codon. Replacement of the *tnpX* stop codon allowed the incorporation of the hexa-His tag at the C-terminus of the translated protein (Fig 3.1). The resultant 2.1 kb PCR product was purified by isopropanol precipitation and digested overnight with *Nde*I and *Xho*I. Following purification the digested PCR product and the vector DNA were then ligated and the mixture used to transform DH5α to ampicillin resistance. The resultant recombinant, pJIR1998, was identified by restriction

Figure 3.1 Construction of the TnpX expression plasmid, pJIR1998

Tne *tnpX* sequence was generated via PCR using oligonucleotide primers 12882 and 12873, which incorporated *Nde*I and *Xho*I sites. The *pelB* leader sequence was removed during cloning but the T7 promoter, *lac* operator, the *tnpX* coding sequence (blue arrow) and the hexa-His tag were present in the recombinant plasmid.


ori
lacI
pET22b+
bla
tnpX
NdeI
XhoI
ATG
12882
12873
XbaI
NdeI
BamHI
SacI
HindIII
XhoI
AvaI
Stop codon
T7 Promoter
lac operator
rbs
pelB leader
EcoRI
SalI
EagI
NotI
His6
Terminator
pJIR1998
XbaI
NdeI
tnpX
XhoI
AvaI
Stop codon
ATG
T7 Promoter
lac operator
rbs
His6
Terminator

analysis and then the insert completely sequenced to ensure that no errors had been introduced by PCR amplification and that the fusion junctions were intact. This construct was subsequently used for the purification of the TnpX-His_6 protein.

b) Purification of TnpX

TnpX is a 707 amino acid protein (82 kDa) that shows similarity to the resolvase family of site-specific recombinases and belongs to the recently delineated subgroup of large resolvases (Smith and Thorpe, 2002). pJIR1998 encodes a TnpX protein that includes two non-TnpX amino acids derived from the *Xho*I site used for cloning, a leucine and a glutamic acid residue, and a hexa-His tag at the C-terminus. For simplicity, this protein will be referred to as $TnpX_{1-707}$ in this thesis. Overexpression of $TnpX_{1-707}$ was attempted in three different strains, BL21(DE3) (Stratagene), and two derivatives of this strain, C41(DE3) and C43(DE3) (Miroux and Walker, 1996). C41(DE3) and C43(DE3) were isolated by selecting for mutants capable of expressing toxic proteins and membrane proteins at high levels and consequently can be used to overexpress difficult proteins (Miroux and Walker, 1996). Highest levels of expression were obtained in the C43(DE3) background and this strain was used for subsequent experiments.

$TnpX_{1-707}$ was purified using a Talon metal affinity resin to which the C-terminal His tag was able to bind in the absence of the binding competitor, imidazole. The loaded Talon column was then washed extensively with low but increasing amounts of imidazole to remove cellular contaminants. Purified tagged protein was eluted using either a 30 mL linear gradient of 25-400 mM imidazole or a step gradient with 4 mL steps of 50, 100, 200 and 400 mM imidazole. $TnpX_{1-707}$ eluted from the column in a broad peak from 50 mM-400 mM imidazole. The buffer used for resuspension of the bacterial pellet contained 500 mM NaCl and 0.1% Triton X-100 to increase the solubility of the protein during cell lysis,

since $TnpX_{1-707}$ precipitated readily when present at high protein concentrations and in low ionic strength buffers. Purified $TnpX_{1-707}$ was subsequently maintained in buffers with moderate salt concentrations (300-500 mM NaCl), although the Triton X-100 was removed from the protein preparation during washing of the column and protein elution. Degradation was also a problem, which was partially alleviated by the addition of PMSF during the purification process (Fig 3.2).

Small resolvase proteins form dimers as their activity is dependent on binding of two resolvase proteins to the active site, site I, to allow cleavage and subsequent covalent attachment of the two DNA strands to the two protein molecules (Hallet and Sherratt, 1997). To determine whether $TnpX_{1-707}$ existed as a monomer or dimer in solution, gel filtration was performed using a Pharmacia Superdex 200 16/60 column. $TnpX_{1-707}$ eluted as an asymmetrical peak (Fig 3.3) that suggests that the protein is not completely globular. The elution volume was used to determine the molecular weight of the protein complex by comparison with protein standards of known molecular weight. The elution volume of $TnpX_{1-707}$ corresponds to a molecular size of 196 kDa (Table 3.1). The calculated monomeric molecular weight of $TnpX_{1-707}$ is 83 kDa, consequently the expected molecular weight of a $TnpX_{1-707}$ dimer would be 166 kDa, while a trimer would be 249 kDa. The observed molecular size of 196 kDa is larger than the expected value for a dimer but smaller than the expected value for a trimer. It is thought that the discrepancy between the expected and observed molecular weights is due to the asymmetry of the elution peak. Since purified $TnpX_{1-707}$ was soluble and not aggregated and capable of dimerization, it was used for subsequent functional experiments. The purified protein was also subjected to CD spectral analysis and the protein was shown to contain a significant proportion of α-helical structure, also strongly suggesting that $TnpX_{1-707}$ was properly folded (I. Lucet and J. I. Rood, unpublished results).

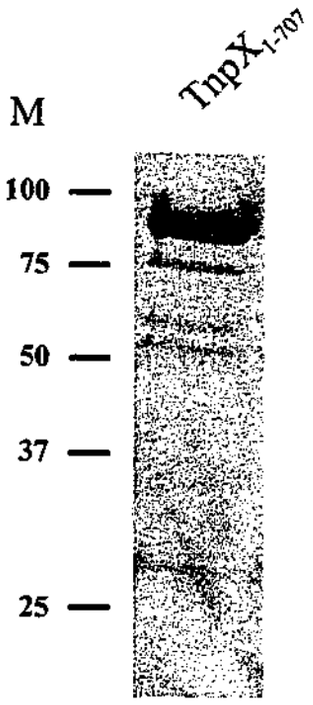


Figure 3.2 SDS-PAGE of purified recombinant $TnpX_{1-707}$

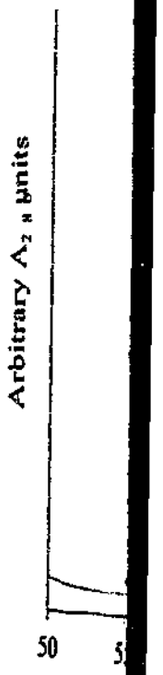
Strain C43(DE3)(pJIR1998) was induced for 4 hours in 1 L of 2 x YT medium. The cells containing $TnpX_{1-707}$ were disrupted using a French pressure cell and sonication as described in Chapter 2. The lysate was applied to the Talon column to allow the tagged protein to bind. The column was washed with 150 mL of wash buffer with an increasing imidazole concentration from 0-25 mM. $TnpX_{1-707}$ was eluted from the column in a broad peak between 50-400 mM imidazole. The figure shows a sample of Talon purified $TnpX_{1-707}$ after electrophoresis on a 12% SDS-PAGE gel and staining with Coommassie brilliant blue.

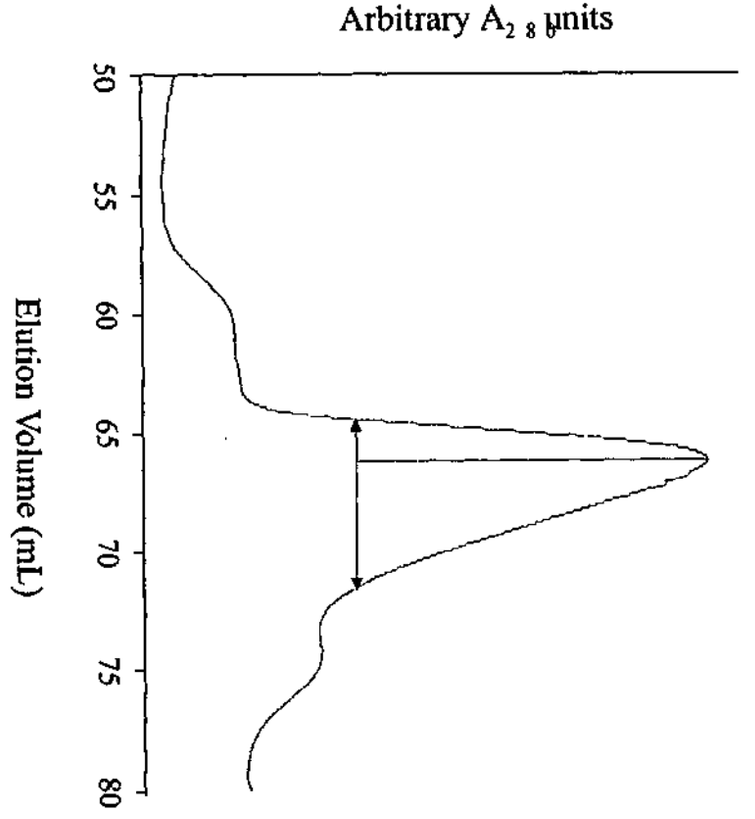
Table 3.1 Gel filtration data for purified TnpX

	Calculated Molecular Weight (kDa)	Elution vol. (mL)	Apparent Molecular Size (kDa)
$TnpX_{1-707}$	83	67	196

Figure 3.3 Determination of native molecular size of $TnpX_{1-707}$

Purified $TnpX_{1-707}$ was applied to a Superdex 200 16/60 gel filtration column (Pharmacia) and eluted. Fractions were collected and analyzed by SDS-PAGE. The elution volume was determined from the UV absorbance peak (panel A, red trace) and shown by SDS-PAGE to correspond to $TnpX_{1-707}$ protein. The molecular size of $TnpX_{1-707}$ was calculated by use of a standard curve produced by calibrating the column using proteins of known molecular weight (Gel Filtration Calibration Kit, Biorad). The asymmetry of the elution peak is indicated by the black double-headed arrow and the blue line indicates the top of the peak.




Arbitrary A_{280} units
Elution Volume (mL)
50
55
60
65
70
75
80

c) $TnpX_{1-707}$ Catalyzes Transposon Excision *In Vitro*

With the availability of purified $TnpX_{1-707}$ it was possible to determine exactly which components were required for the excision reaction. An *in vitro* excision assay was designed using purified $TnpX_{1-707}$ and a known *in vivo* excision substrate, pJIR1452. Plasmid pJIR1452 contains a Tn*4451* derivative that has a small, 185 bp, deletion internal to the *tnpX* gene. This plasmid had previously been used for *in vivo* excision assays where the deletion mutation was complemented by a second plasmid encoding the *tnpX* gene (Crellin and Rood, 1997). An additional modification was also made to the transposon encoded on pJIR1452. Tn*4451* is normally flanked by GA dinucleotides. TnpX catalyses a 2 bp staggered cleavage at the GA residues, facilitating the formation of a circular intermediate and reconstitution of the target site (Crellin and Rood, 1997; Lyras and Rood, 2000b), which both contain a single GA dinucleotide. The Tn*4451* derivative in pJIR1452 is not flanked by GA dinucleotides but by TA dinucleotides that were introduced by site-directed mutagenesis (Crellin and Rood, 1997). This derivative was shown to exhibit a significant increase in excision frequency (Crellin and Rood, 1997), and consequently was thought to be an ideal substrate for attempting to demonstrate excision *in vitro*.

The excision assay was initially carried out using a similar buffer to that used for protein purification but of lower ionic strength (100 mM NaCl) since the protein was diluted and more likely to remain in solution. The *in vitro* excision buffer was based on buffers used for the subsequent restriction enzyme digests. The pJIR1452 DNA was purified by CsCl buoyant density gradient ultracentrifugation as it was anticipated that a large amount of plasmid DNA of high purity would be required. The CsCl-derived DNA was dialysed over three days with four 2 L buffer changes to remove as much of the CsCl as possible in case the excision reaction was inhibited by cesium ions. After incubation of the DNA with $TnpX_{1-707}$, to detect if excision had occurred, the reaction mixture was digested with the

restriction enzymes *Eco*RI and *Eco*RV. Digestion of pJIR1452 would produce 5.8 kb and 3.7 kb fragments. If excision occurred 6.2 kb and 3.3 kb fragments would also be detected (Fig 3.4). The 6.2 kb fragment would correspond to the linearized circular intermediate whereas the 3.3 kb fragment would represent the linear deletent plasmid (Fig 3.4). PCR amplification of the joint of the circular intermediate (CI) was also performed to confirm that an intact CI joint had been formed (Fig 3.4).

Initial experiments suggested that a very small amount of excision was occurring (data not shown). The bands could barely be seen on a stained gel but excision was confirmed by PCR amplification of the joint of the circular intermediate from the assay mixture, which is a more sensitive test. In an attempt to increase the excision frequency the assay was repeated in the presence of different metal ions (Fig 3.5). Addition of 10 mM $MgCl_2$, $MnCl_2$, KCl, LiCl or EDTA resulted in no change in excision activity. In contrast, the addition of 10 mM $CaCl_2$ resulted in a significant increase in excision levels, to the point that the relevant bands could clearly be seen on an ethidium bromide stained gel (Fig 3.5). Addition of 10 mM $ZnCl_2$ resulted in a band migrating at an equivalent position to open circular DNA in an untreated plasmid preparation, and inhibited subsequent digestion by *Eco*RI and *Eco*RV. O-phenantroline is a strong chelating agent that can be used to remove tightly bound Zn^{2+} ions from proteins. PCR analysis showed that no circular intermediate was produced in reactions containing either $ZnCl_2$ or o-phenantroline (Fig 3.5). The PCR assay for these reactions was not inhibited since a PCR product representing amplification around the complete substrate plasmid, pJIR1452, was obtained (data not shown).

To further delineate the optimal conditions for *in vitro* excision, the assays were repeated using various incubation times, calcium concentrations, protein concentrations and temperatures (Fig 3.6, 3.7 and 3.8). DNA bands derived from the circular intermediate and deletent forms were visible after just 10 minutes incubation at 37°C, although the 0 and 5

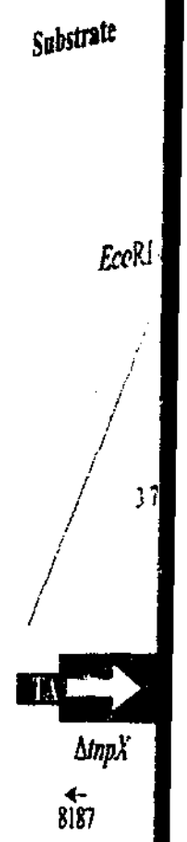
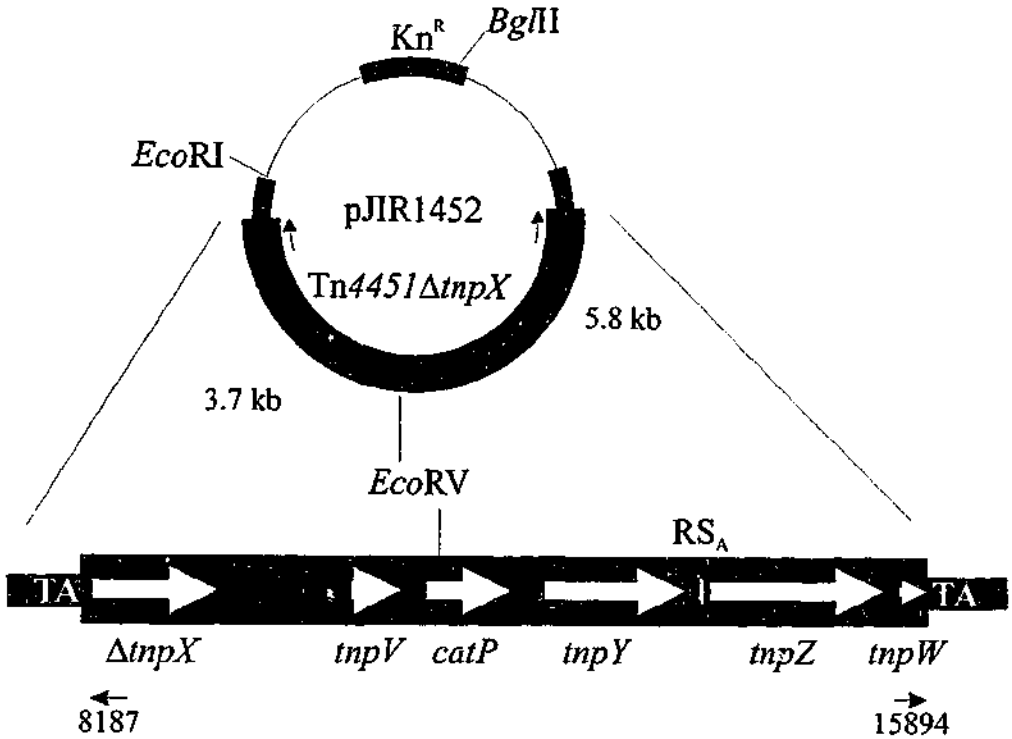


Figure 3.4 *In vitro* excision assay substrates and products

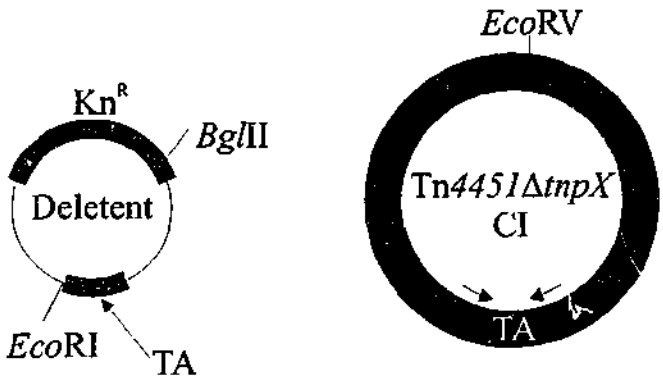
The substrate plasmid, pJIR1452, contains a copy of Tn*4451* (blue block arc and blue rectangle with white arrows indicating transposon genes) containing a deletion within the *tnpX* gene. This element also has TA dinucleotides in place of the GA dinucleotides found at the end of the wild-type transposon. This plasmid also encodes kanamycin resistance (red arc). The green rectangle and arc represent the target DNA derived from *C. perfringens*. Relevant restriction sites are marked. The products include the circular intermediate (blue circle) and the deletent plasmid, which contains the reconstituted target site (green arc) within the plasmid backbone. Also shown are the positions of the oligonucleotide primers, 8187 and 15894, which were used in PCR reactions to detect the formation of the joint of the circular intermediate (black arrows). Both the substrate and the circular intermediate product will produce a PCR product using the oligonucleotide primers indicated but the sizes will be very different, 5.4 kb and 909 bp, respectively.

Substrate



TnpX

Products

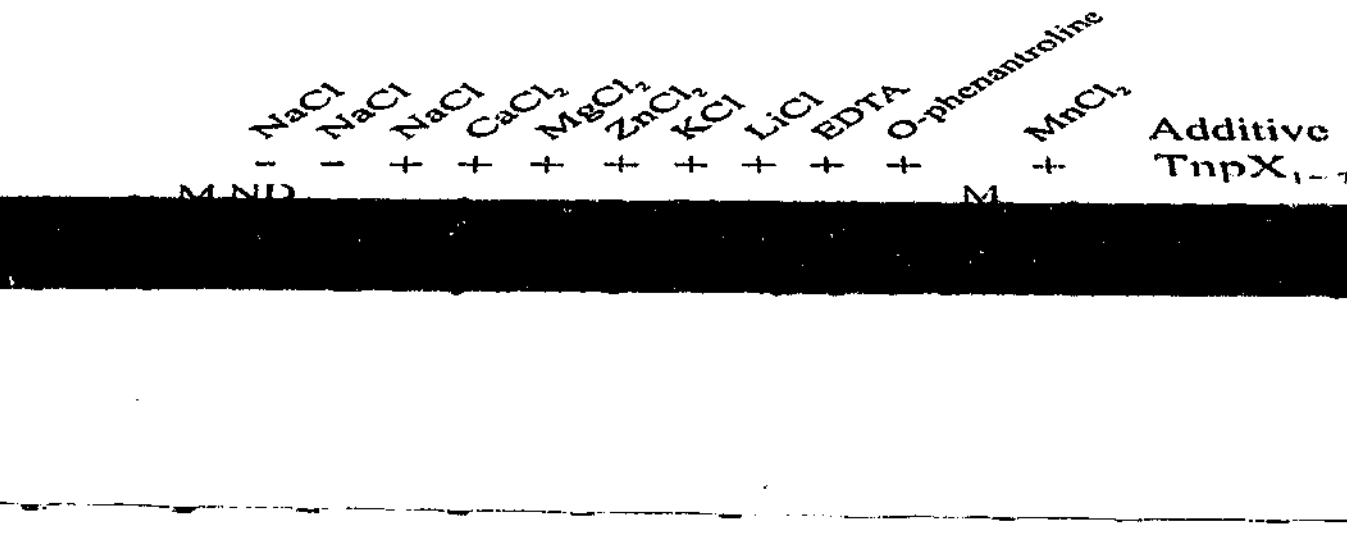


3.3 kb

6.3 kb

Figure 3.5 Optimization of the *in vitro* excision assay – requirement for metal ions

$TnpX_{1-707}$ (0.09 μM) was added to pJIR1452 DNA (200 ng) and incubated at 37°C for 30 minutes in the presence of 10 mM concentrations of each additive (labelled at top of figure). After incubation with $TnpX_{1-707}$, either $MgCl_2$ or EDTA was added to the reactions to produce suitable conditions for digestion by *Eco*RI and *Eco*RV. The reactions were subjected to electrophoresis using a 0.7% agarose gel and DNA visualized by staining with ethidium bromide. The presence of supercoiled (CCC), linear (L) and open circular (OC) DNA is indicated and all lanes contain digested reactions except for lane 1 (not digested (ND)). The starting substrates (S) and recombination products (P) are also indicated. PCR reactions were performed using oligonucleotide primers 8187 and 15894. The size of PCR products was ~400 bp as expected for the joint of the circular intermediate (405 bp) in each case. Molecular size markers (labeled M) were *Hin*dIII digested λ DNA. PCR markers (Promega) were used when PCR reactions were subjected to agarose gel electrophoresis. The no protein control lane is labelled A.

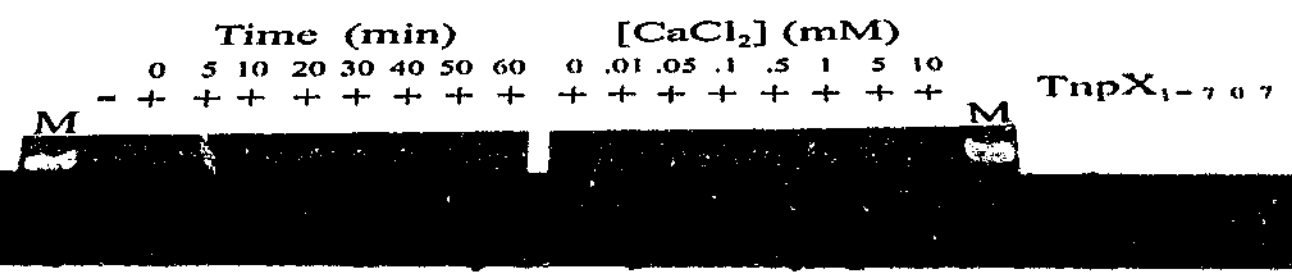


minute samples were PCR positive since the 0 time point was taken immediately after addition of $TnpX_{1-707}$ (Fig 3.6A). The PCR result indicates that some conversion of plasmid into circular intermediate and deletion derivative occurs very quickly. A reasonably high calcium concentration was required to enhance excision, as visible excision products were only produced at 5 mM and 10 mM $CaCl_2$ (Fig 3.6B). The 10 mM $CaCl_2$ concentration was subsequently shown to be optimal as the use of 20 and 50 mM $CaCl_2$ concentrations did not increase the levels of excision (data not shown). However, calcium concentrations as low as 10 nM were PCR positive (Fig 3.6B). Protein concentrations as low as 33 nM still gave a PCR positive result (Fig 3.7). However, a concentration of 0.2 μM resulted in maximal excision (Fig 3.7). Excess $TnpX_{1-707}$ did not result in a higher level of excision but led to the DNA becoming trapped in the wells, presumably due to the formation of higher order protein-DNA complexes (Fig 3.7). The reaction temperatures tested were 30°C, 20°C and 0°C and although all temperatures gave positive PCR results at all time points, only the 30°C reactions gave visible excision product bands (Fig 3.8). These results suggest that the optimal temperature for $TnpX_{1-707}$ activity was around 30°C or greater.

Typical resolvase proteins are more efficient at catalysing the excision of supercoiled substrates (Stark and Boocock, 1995). To determine if this situation also applied to TnpX-mediated excision, both open circular and linear forms of pJIR1452 were produced and then tested in the excision assay. The pJIR1452 DNA was linearized using the unique restriction enzyme site recognized by *Bgl*II (Fig 3.4), which cleaves on the opposite side of the plasmid to the transposon, well away from any DNA binding sites or catalytic sites. Open circular pJIR1452 DNA was produced by single strand nicking of the plasmid DNA by DNaseI, which led to topological relaxation and loss of plasmid supercoiling. All three plasmid forms, supercoiled (CCC), open circular (OC) and linear (L), were purified from an agarose gel before use in the excision reactions. Purification of the supercoiled and

Figure 3.6 Optimization of the *in vitro* excision assay – reaction time and calcium concentration

Optimal reaction times and calcium concentrations were determined. The excision reactions were performed as described in Fig 3.5. The reaction time experiments (panel A) were conducted using a single reaction with a total volume of 240 µL. Samples of 30 µL were taken at the specified time points and immediately transferred to 80°C for heat inactivation. The optimal calcium concentration (panel B) was determined by incubating the protein and DNA at the calcium concentrations indicated. The concentration of $TnpX_{1-707}$ in each reaction was 0.3 µM. The starting substrates (S) and recombination products (P) are indicated. PCR reactions were performed using oligonucleotide primers 8187 and 15894. The size of PCR products was ~400 bp as expected for the joint of the circular intermediate (405 bp) . Molecular weight markers (labeled M) were *Hin*dIII digested λ DNA. PCR markers (Promega) were used when PCR reactions were subjected to agarose gel electrophoresis.



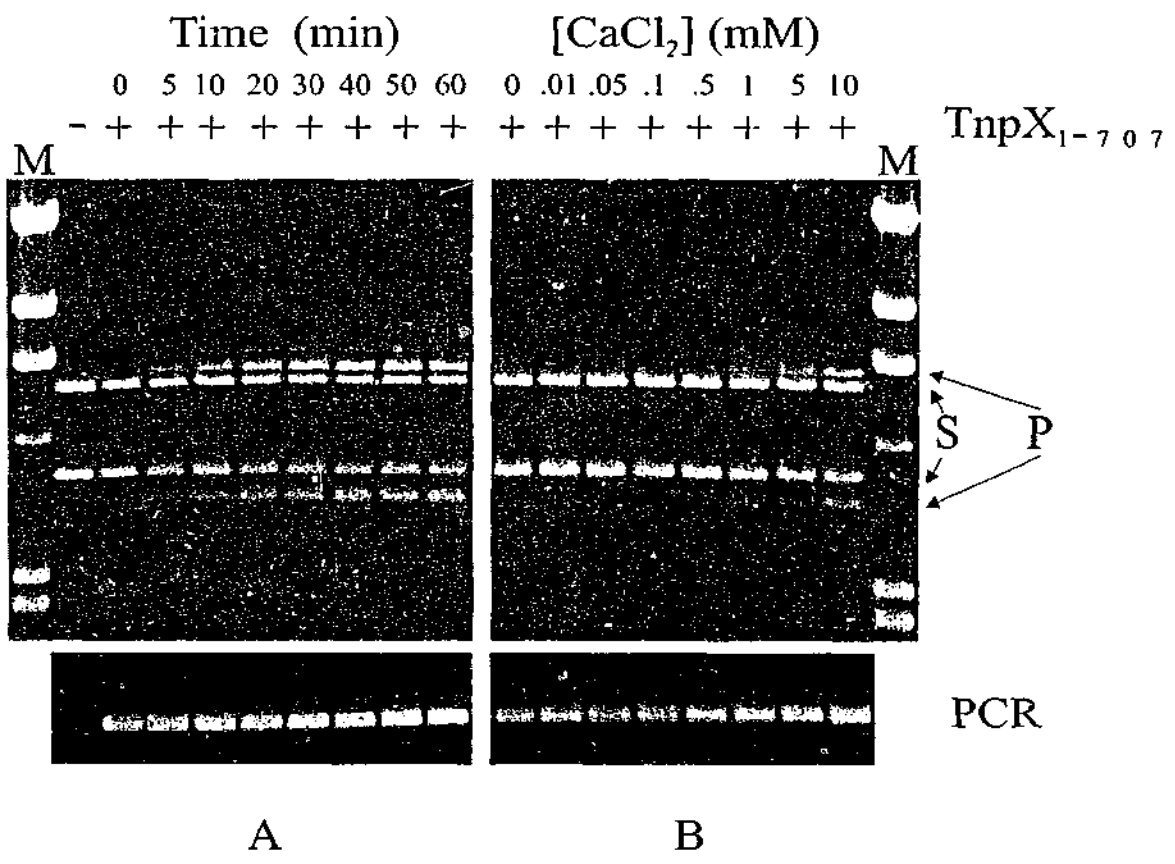
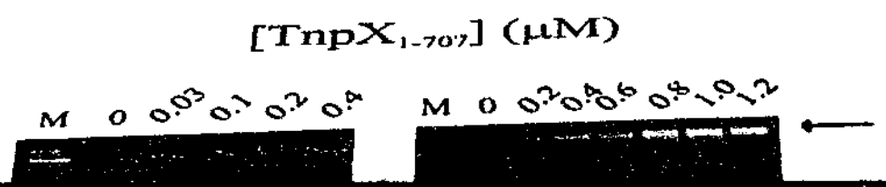


Figure 3.7 Optimization of the *in vitro* excision assay - $TnpX_{1-707}$ concentration

The effect of $TnpX_{1-707}$ concentration was examined both at low concentrations (panel A) and at high concentrations (panel B). Reactions contained 10 mM $CaCl_2$ and were incubated at 37°C for 30 minutes. High concentrations of protein resulted in increasing amounts of DNA being trapped in the wells (red arrow). The starting substrates (S) and recombination products (P) are indicated. PCR reactions were performed using oligonucleotide primers 8187 and 15894. The size of PCR products was ~400 bp as expected for the joint of the circular intermediate (405 bp) . Molecular weight markers (labeled M) were *Hin*dIII digested λ DNA. PCR markers (Promega) were used when PCR reactions were subjected to agarose gel electrophoresis.



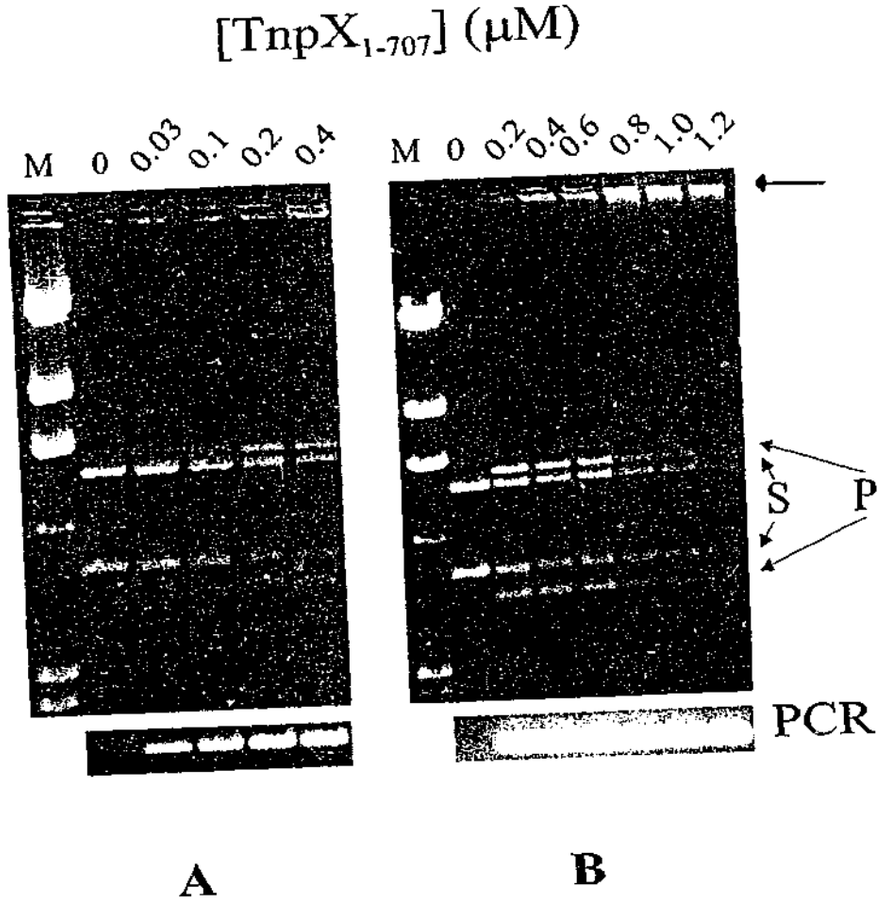
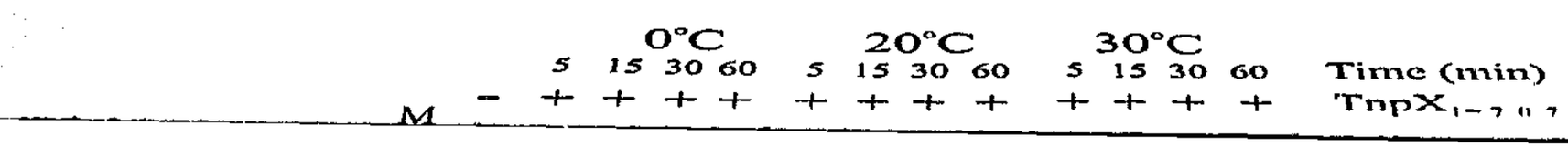
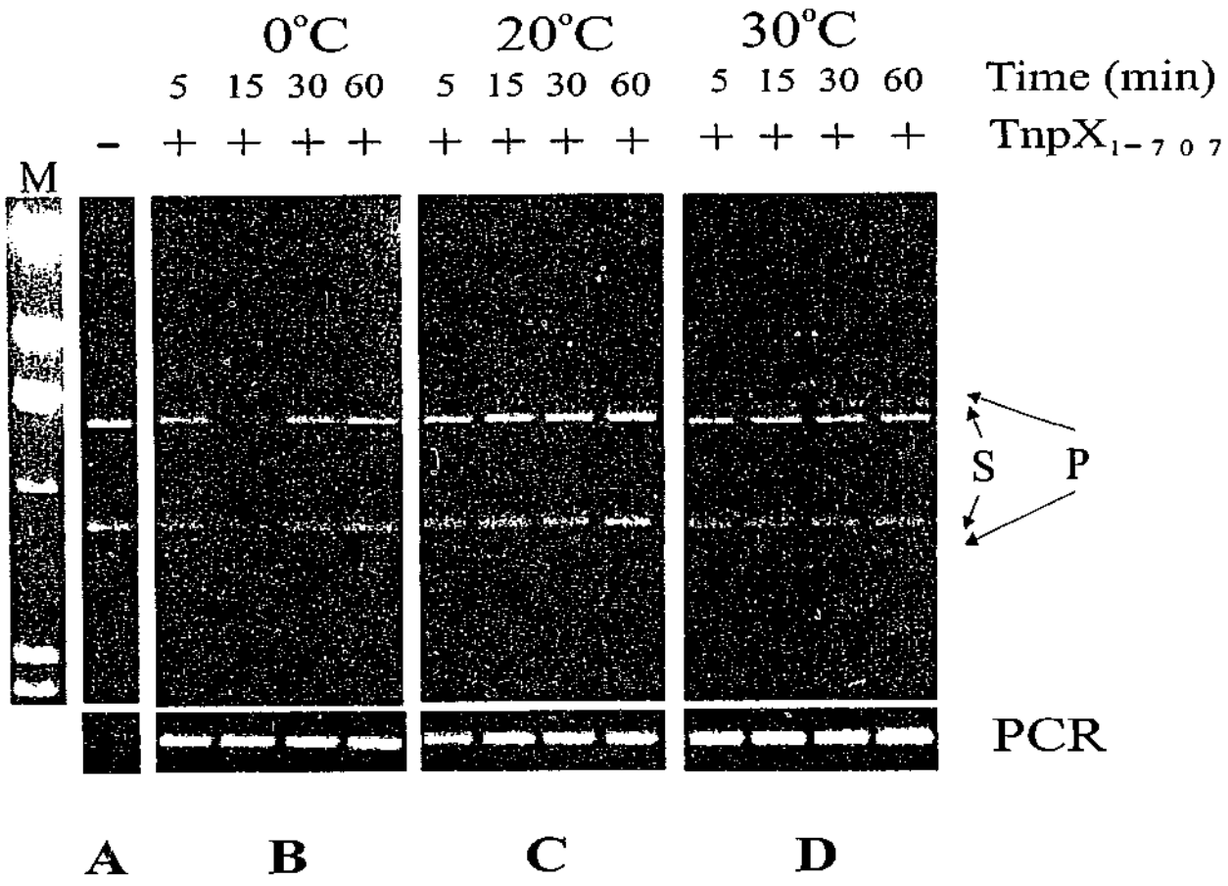

[TnpX1-707] (μM)
M 0 0.03 0.1 0.2 0.4
M 0 0.2 0.4 0.6 0.8 1.0 1.2
S
P
PCR
A
B

Figure 3.8 Optimization of the *in vitro* excision assay – effect of temperature

Excision reactions were performed as described in Fig 3.5 with 10 mM $CaCl_2$ and a $TnpX_{1-707}$ concentration of 0.2 μM. Time course experiments were performed as described in Fig 3.6A at three different temperatures, 0°C (B), 20°C (C) and 30°C (D). Samples were taken at each time point and transferred immediately to 80°C for heat inactivation. The starting substrates (S) and recombination products (P) are indicated. PCR reactions were performed using oligonucleotide primers 8187 and 15894. The size of PCR products was ~400 bp as expected for the joint of the circular intermediate (405 bp). Molecular weight markers (labeled M) were *Hin*dIII digested λ DNA. PCR markers (Promega) were used when PCR reactions were subjected to agarose gel electrophoresis. The no protein control lane is labelled A.




0°C
20°C
30°C
5 15 30 60 5 15 30 60 5 15 30 60
Time (min)
– + + + + + + + + + + + +
TnpX1–707
M
S
P
PCR
A
B
C
D

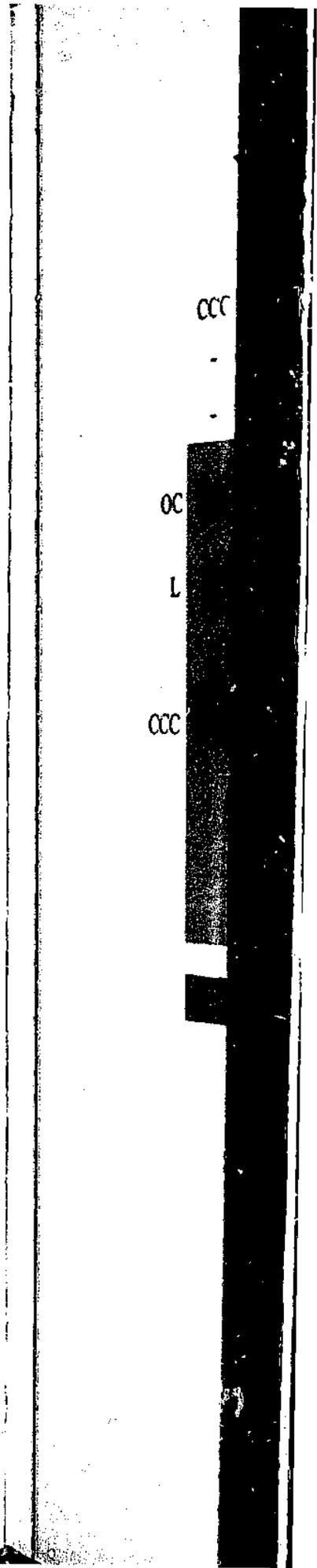
open circular species was difficult, as the gel purification procedure was optimised for linear DNA, not circular DNA. Standard reaction conditions were used for all reactions, consisting of 0.3 μM of $TnpX_{1-707}$ and 10 mM $CaCl_2$, incubated at 37°C for 30 minutes. Southern hybridization was performed on the resultant gel to detect small amounts of excision products that may have been produced (Fig 3.9). Both the supercoiled and open circular DNA were somewhat damaged during the gel purification procedure as both showed the presence of linear DNA and open circular DNA was also evident in the supercoiled DNA preparation (Fig 3.9, lanes 1 and 3). However, Southern hybridization analysis clearly showed that excision occurred in the supercoiled preparation (presence of a 6.2 kb band representing the CI joint) but that only minimal activity was evident with either the linear or open circular substrates (PCR amplification of the CI joint was positive for all substrates when $TnpX_{1-707}$ was present). It was not considered that the difference in the amount of circular intermediate produced using the different substrates was due to a difference in DNA concentration as the linear preparation was the most concentrated. It appeared that the addition of $TnpX_{1-707}$ resulted in the loss of linear DNA species regardless of which DNA substrate was used (Fig 3.9, lanes 5, 7 and 9). It is postulated that the loss of this partially digested DNA is due to non-productive complexes being formed with $TnpX_{1-707}$, since the loss of the linear species did not correspond to an increase in the 6.2 kb CI DNA fragment in the linear and open circular DNA substrate lanes (Fig 3.9).

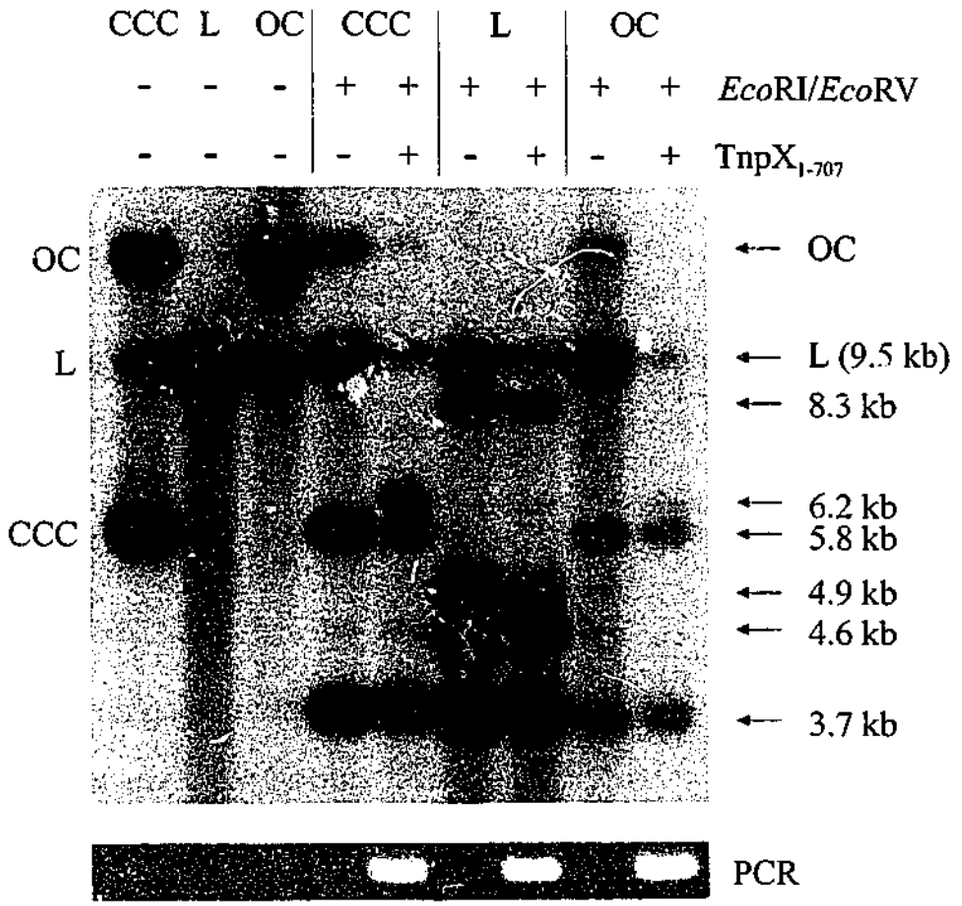
d) *In Vitro* Insertion.

The development of an *in vitro* insertion assay was also attempted using either plasmid molecules or linear PCR products as substrates. Recombination was attempted between two different sized plasmids both containing the cloned joint of the circular intermediate, on the basis that higher levels of recombination may occur between two circular

Figure 3.9 Effect of substrate topology on *in vitro* excision

Lanes 1-3 contain an example of the starting DNA substrate before the reactions were performed. The supercoiled (CCC) sample (lane 1) contains all three species tested, however, neither of the other preparations contain supercoiled DNA. The open circular (OC) DNA also contains a proportion of linear (L) DNA. Extraction of the linear DNA resulted in only linear species. Aliquots of these three DNA species were then used in excision reactions with and without $TnpX_{1-707}$ (0.3 μM), incubated at 37°C, heat inactivated and digested with *Eco*RI and *Eco*RV as indicated. The reactions were run on a 0.7% gel and transferred to nylon filters for Southern analysis. This was performed using a circular intermediate probe, generated using oligonucleotide primers 8187 and 15894, consequently only DNA fragments containing transposon DNA were detected. PCR reactions were performed as described in Fig 3.5. The starting substrates (black arrow, 5.8, 4.6 and 3.7 kb) and recombination product, circular intermediate, (green arrow, 6.2 kb) are indicated. Digestion with *Eco*RI and *Eco*RV did not proceed to completion in all reactions and partial digestion products are indicated by red arrows. In some instances the DNA was not digested at all, resulting in the presence of some linear and open circular DNA species (top two red arrows), while partial digestion also occurred (bottom two red arrows: 4.9 kb when no digestion by *Eco*RI and 8.3 kb when no digestion by *Eco*RV).




CCC L OC CCC L OC
- - - + + + + + + EcoRI/EcoRV
- - - - + - + - + TnpX1-707
OC
L
CCC
OC
L (9.5 kb)
8.3 kb
6.2 kb
5.8 kb
4.9 kb
4.6 kb
3.7 kb
PCR

intermediate joints. Recombination was also attempted with a cloned circular intermediate joint and a fragment containing a known target site. All recombination sites contained the TA dinucleotide, as was used for the *in vitro* excision assays. It was thought that as the same catalytic mechanism was required for both excision and insertion, the TA dinucleotides may also constitute a preferred insertion substrate.

Plasmid based insertion assays involved the use of pJIR2159, a 7.8 kb plasmid containing a 1014 bp PCR-generated fragment that encompassed the circular intermediate joint (Fig 3.10). The second plasmid, pJIR2292, was a 4.7 kb plasmid that also contained the joint of the CI, although this fragment was generated using different oligonucleotide primers, resulting in a smaller PCR product (909 bp) (Fig 3.10). The third plasmid used in these assays was pJIR2501, a 3.3 kb plasmid derived from pJIR1452 (Fig 3.10). Note that TnpX-mediated fusion of pJIR2159 and pJIR2501 represents the equivalent reaction observed upon the integration of Tn*4451* into its target site.

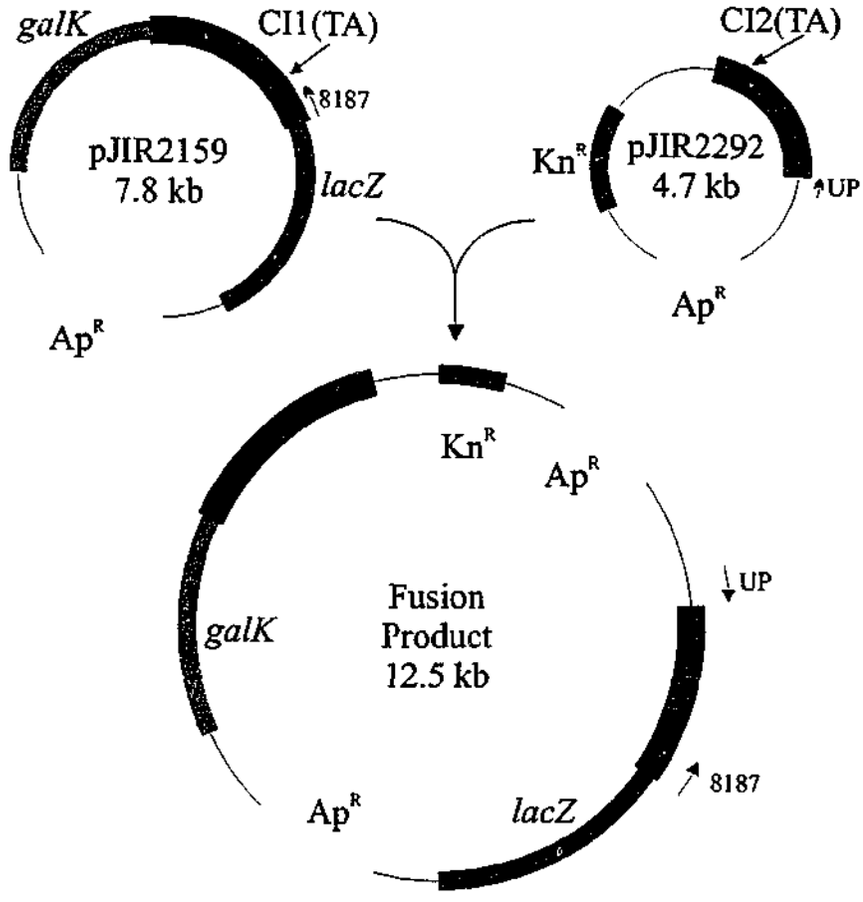
Plasmid DNA was again purified through CsCl gradients and dialyzed extensively. Several conditions were varied including [$TnpX_{1-707}$], [Ca^{2+}] and incubation time (Fig 3.11). No fusion products could be identified although treatment of the plasmid DNA substrates with $TnpX_{1-707}$ resulted in the appearance of laddered DNA species (Fig 3.11) that were the result of $TnpX_{1-707}$ activity since they were absent from the no TnpX control. An increase in incubation time had little discernible effect other than to increase the amount of linear and open circular plasmid species present and production of the laddered DNA, except for the 24 hour sample, where the DNA had been somewhat degraded (Fig 3.11). Increasing the calcium concentration also had little effect, while increasing the $TnpX_{1-707}$ concentration gave a similar result to increasing the incubation time (Fig 3.11). PCR designed to detect the fusion of pJIR2159 and pJIR2292 gave the most product

Figure 3.10 ***In vitro*** **insertion assay plasmid substrates and possible products**

A Reactions between two circular intermediate joints. Plasmid pJIR2159 contains the joint of the circular intermediate (blue arc) generated by PCR using oligonucleotide primers 4675 and 8187 (1014 bp product) and includes a TA dinucleotide at the joint of the two transposon ends and also encodes *galK* (light purple arc), *lacZ* (magenta arc) and ampicillin resistance (yellow arc). The plasmid pJIR2292 contains the joint of the circular intermediate (green arc) generated using the oligonucleotide primers 4675 and 4676 (909 bp product) and also includes a TA dinucleotide (red arrows) at the joint of the circular intermediate. This plasmid also carries kanamycin (red arc) and ampicillin (yellow arc) resistance genes. TnpX-mediated fusion of these plasmids would yield the fusion product indicated by the blue arrows. PCR amplification (1.1 kb product) should only occur between oligonucleotide primers 8187 and UP in the fusion conformation.

B Reactions between a circular intermediate joint and a target site. Plasmid pJIR2501 is derived from pJIR1452 by TnpX-mediated deletion of Tn*4451ΔtnpX1*. It also contains a TA dinucleotide (red arrow) in the middle of the deletion site (dark purple). The fusion plasmid expected from recombination between pJIR2159 and pJIR2501 is indicated by the blue arrow.

A


galK
CI1(TA)
8187
pJIR2159
7.8 kb
lacZ
Ap^R
CI2(TA)
Kn^R
pJIR2292
4.7 kb
UP
Ap^R
Kn^R
Ap^R
UP
galK
Fusion
Product
12.5 kb
Ap^R
lacZ
8187

B

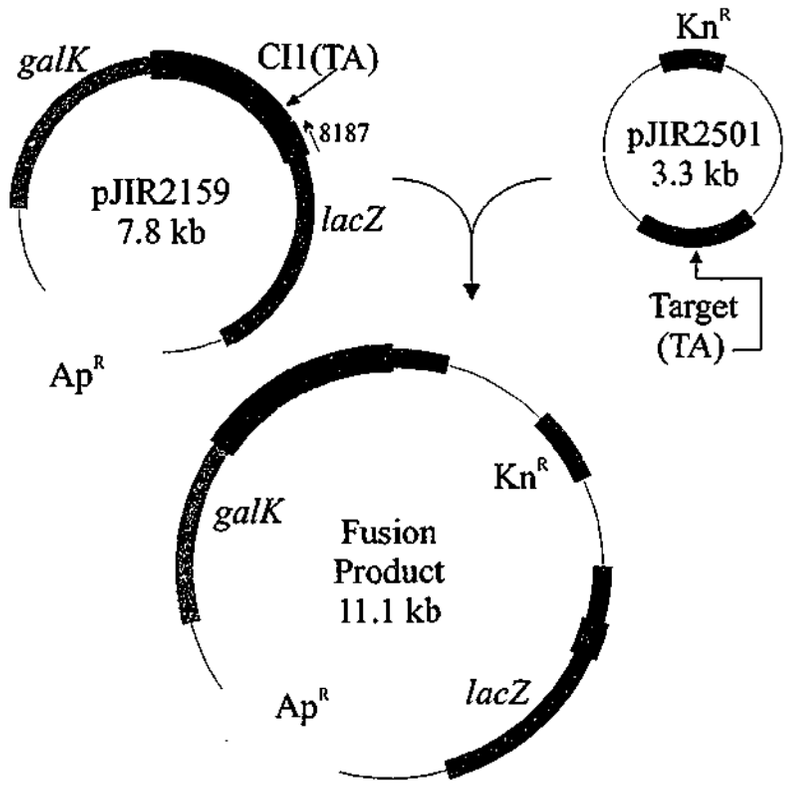
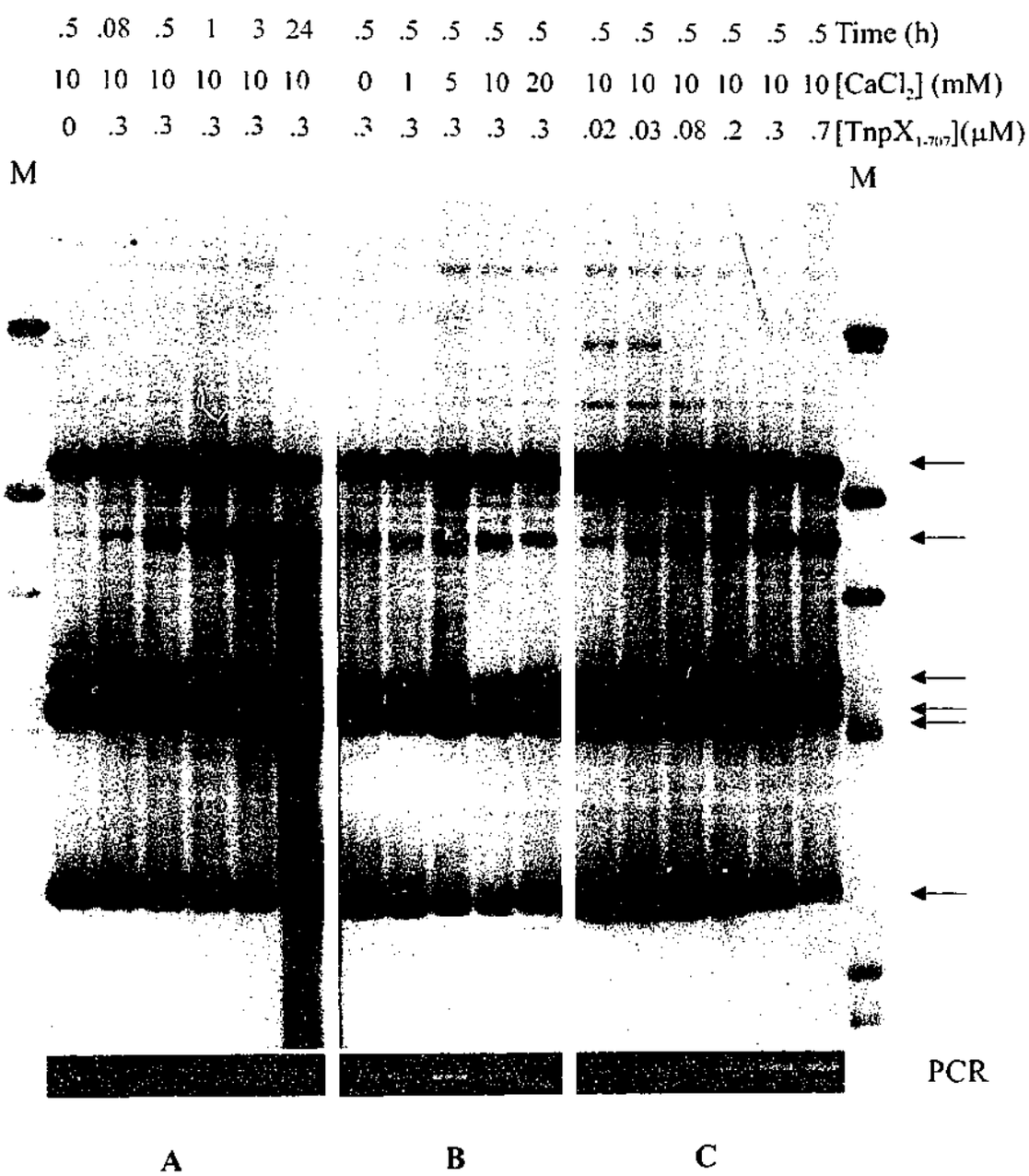

galK
CI1(TA)
8187
pJIR2159
7.8 kb
lacZ
Ap^R
Kn^R
pJIR2501
3.3 kb
Target
(TA)
Kn^R
galK
Fusion
Product
11.1 kb
Ap^R
lacZ

Figure 3.11 *In vitro* insertion assays using plasmid DNA substrates

Supercoiled plasmids were used in insertion reactions containing the indicated amounts of $CaCl_2$, $TnpX_{1-707}$ and the specified incubation time. Insertion reactions were performed in a similar manner to the excision reactions described in Fig 3.5, however, in these reactions two different plasmids, pJIR2159 and pJIR2292, were incubated together in the presence of $TnpX_{1-707}$, heat inactivated and subjected to agarose gel electrophoresis through a 0.8% gel. The DNA was transferred to a nylon membrane and Southern Hybridization was carried out. The probe used to visualize the DNA was a 405 bp PCR product derived from the circular intermediate joint.

An insertion assay time course (panel A) was performed at 37°C, similar to that described in Fig 3.6. Determination of the optimal calcium concentration (panel B) was also performed by varying the amount of $CaCl_2$ added to the reactions. Determination of the optimal $TnpX_{1-707}$ concentration was also attempted (panel C). PCR amplification of the circular intermediate joint is shown below the autoradiograph and was performed using the oligonucleotide primers UP and 8187. The binding site for UP was located on the plasmid pJIR2292, while the binding site for 8187 was only present on plasmid pJIR2159 (Fig 3.10). Consequently amplification would only occur if the two plasmids became fused in such a way that a normal circular intermediate joint was reconstituted (that is the left end of the transposon juxtaposed to the right end, not two left ends or two right ends joined together). The various forms of the substrate plasmids are indicated by arrows: plasmid pJIR2159 forms (red arrows), open circular plasmid at the top, linear in the middle and supercoiled at the bottom, similarly the plasmid pJIR2292 (blue arrows) forms are indicated. The molecular size markers (M) are DIG labelled *Hin*dIII digested λ markers.


.5 .08 .5 1 3 24 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 .5 Time (h)
10 10 10 10 10 10 0 1 5 10 20 10 10 10 10 10 10 [CaCl2] (mM)
0 .3 .3 .3 .3 .3 .3 .3 .3 .3 .3 .02 .03 .08 .2 .3 .7 [TnpX1-707](μM)
M
M
PCR
A
B
C

between 1-24 hours and between 0.2-0.7 μM of $TnpX_{1-707}$, and more PCR product was produced at 5 and 10 mM $CaCl_2$ (Fig 3.11).

The plasmid based assays appeared to show an orientation bias, which was detected by PCR (data not shown), resulting in the reconstruction of an intact circular intermediate joint when two CI joints were reacted together (rather than two left or right ends of the transposon being joined together). Detection of this product was difficult, since if insertion occurred between the same two plasmids the resultant insertion plasmid would be indistinguishable from the parental plasmid by restriction digestion or PCR (Fig 3.10). Assays were also attempted using only a single plasmid, pJIR2159, and electrophoresing the DNA without digestion to try and identify the fusion product as a DNA species with a lower electrophorectic mobility although no products of the expected size could be identified (data not shown).

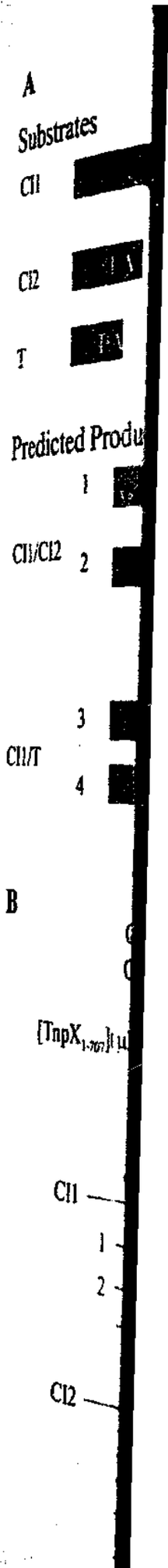
The most notable result observed with these insertion assays was the appearance of DNA laddering in the presence of $TnpX_{1-707}$ (Fig 3.11). This effect may have been indicative of insertion events occurring at multiple different sites rather than recombination between one joint and another. Alternatively, it may have been a reflection of TnpX-mediated topoisomerase activity. Insertion assay PCR results were variable due the unfortunate similarity in sequence surrounding the cloned CI joints as it seems that some mis-priming was occurring. However, it would seem that insertion was occurring at a very low frequency (Fig 3.11). Insertion of the CI joint into the target site occurred at an even lower frequency so this combination was not pursued further using the plasmid system (data not shown).

To develop a system that would allow easier detection of insertion activity the use of different sized linear PCR products as substrates was attempted (Fig 3.12A). This

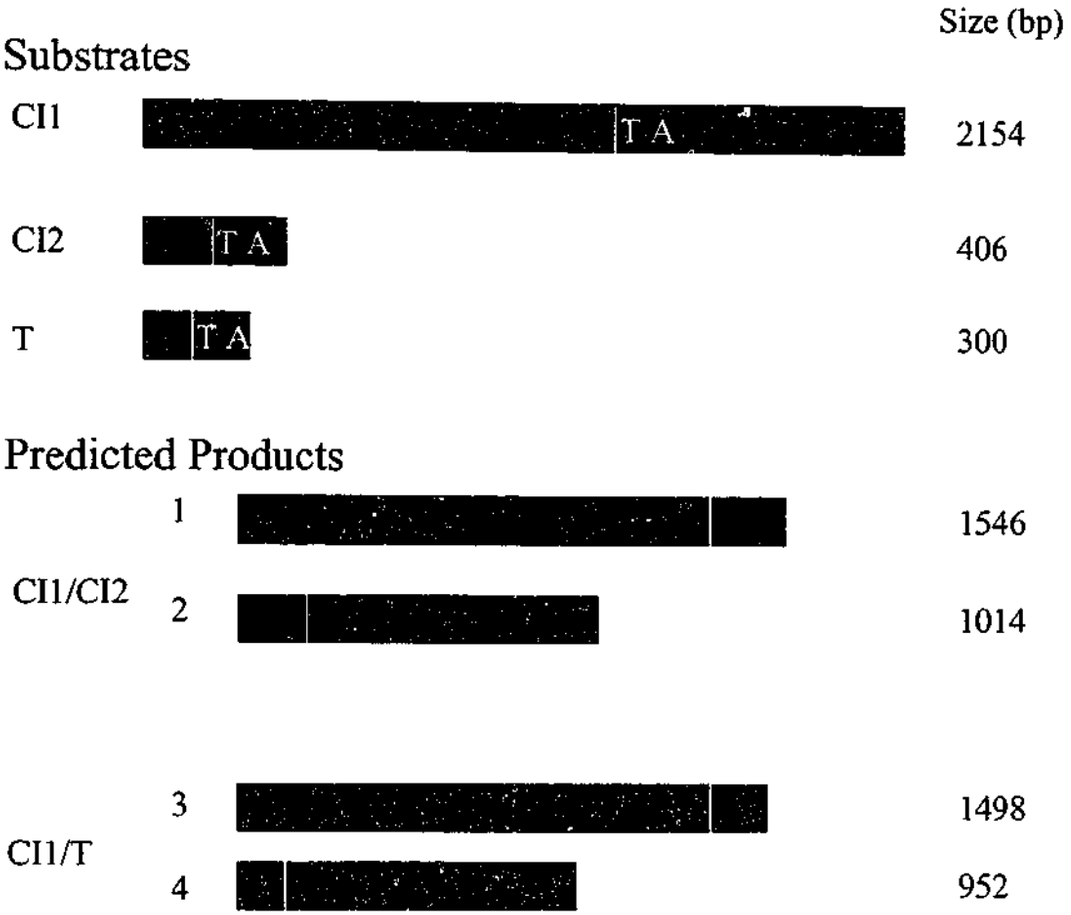
Figure 3.12 *In vitro* insertion assays using linear PCR products as substrates

A PCR products used. The PCR products used (CI1, CI2 and T, all TA derivatives) and the expected reaction products from TnpX-mediated insertion events are shown. Vertical lines on the reactants signify the position of the TA dinucleotides at the joint.

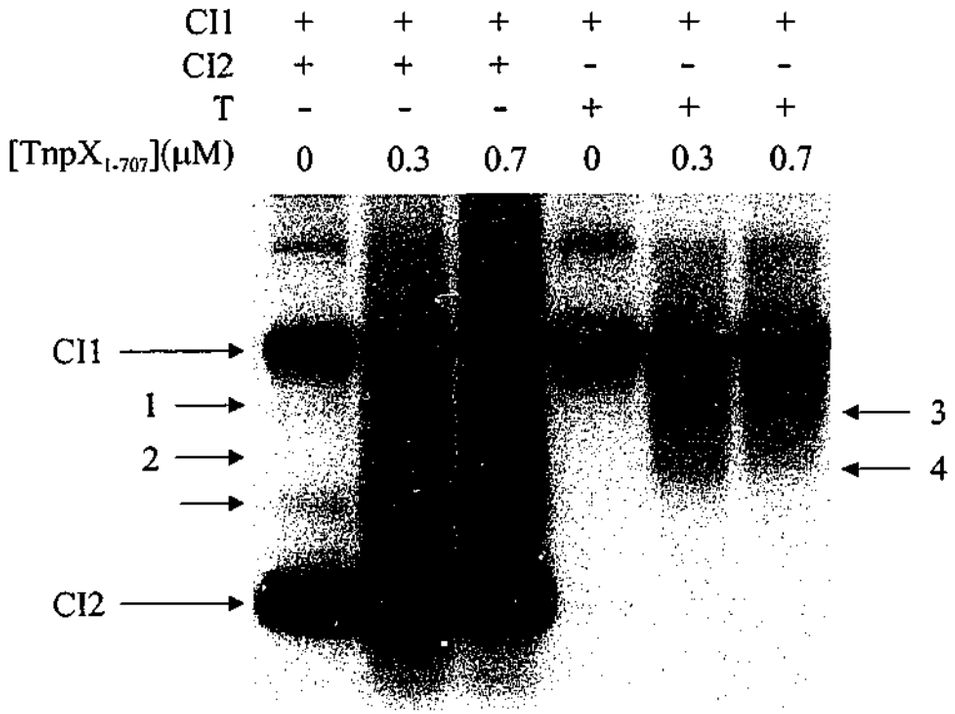
B Two circular intermediate PCR products of different size were mixed together (CI1 2154 bp and CI2 406 bp, generated using oligonucleotide primer pairs 169/4675 and 8187/15894 respectively), incubated with and without TnpX for 4.5 hours then heat inactivated (lanes 1-3). CI1 and a target fragment (T, 300 bp, generated using oligonucleotide primers 8186 and 16208) were incubated in the same way (lanes 4-6). After heat inactivation the reactions were electrophoresed through a 1% agarose gel. The DNA was transferred to a nylon membrane and the DNA detected using a DIG-labelled probe derived from the joint of the circular intermediate, as a consequence the DNA fragment T did not hybridize to the labelled probe. The recombinant products obtained through insertion of one PCR product into another of a different size are indicated by numbers which correspond to the numbers beside the predicted products (panel A). The red arrow indicates a DNA fragment that corresponds in size to a fragment of CI1 cleaved at the TA residue.



A


Size (bp)
Substrates
CI1
T A
2154
CI2
T A
406
T
T A
300
Predicted Products
1
1546
CI1/CI2
2
1014
3
1498
CI1/T
4
952

B


CI1 + + + + + +
CI2 + + + - - -
T - - - + + +
[TnpX1-707](μM) 0 0.3 0.7 0 0.3 0.7
CI1
1
2
CI2
3
4

approach was somewhat more successful. Since linear fragments were used there could only be two combinations of reactions, those of one CI joint recombining into an identical CI, a reaction that was again undetectable, and the second event whereby one CI joint recombined with a CI joint (or target) of a different size, forming a hybrid molecule different in size to both of the original fragments. This second event could then be detected directly by agarose gel electrophoresis or by Southern hybridization analysis. Such hybrid fragments can be seen in Figure 3.12B where the reaction of CI1 (2154 bp) and CI2 (406 bp) produced recombinant products of 1.55 and 1.0 kb. Likewise when CI1 and the target (T) (300 bp) were combined, fragments of 1.50 and 0.95 kb in size were obtained. Neither recombinant combination could be clearly seen by agarose gel electrophoresis and it was necessary to perform Southern hybridization to confirm that they were present (Fig 3.12B). This result illustrates that insertion *in vitro*, under the conditions used, occurs at a low frequency. However, since this assay was clearly not sufficiently robust for routine use no further studies were carried out on *in vitro* insertion.

Discussion

Purified $TnpX_{1-707}$ is functional, being capable of catalyzing excision *in vitro*. Combined with *in vivo* studies carried out in this laboratory (D.Lyras and J. I. Rood, unpublished results) the results presented in this chapter provide convincing evidence that TnpX is the only Tn*4451*-encoded protein required for transposon excision, and most probably for transposon insertion. The *in vivo* studies involved the mutagenesis of each of the genes encoded on Tn*4453a* and the subsequent genetic analysis of the mutants. The results showed that mutagenesis of the genes other than *tnpX* did not affect the excision or insertion reactions.

The *in vitro* excision reaction requires the presence of 10 mM $CaCl_2$ for efficient recombination, a level considerably higher than intracellular calcium concentrations, which

are normally in the nM range and are tightly controlled (Michiels *et al.*, 2002). Since recombination still occurs in the absence of calcium ions (i.e. in the presence of EDTA) it is concluded that Ca^{2+} is not involved in catalysis, which would be inconsistent with a resolvase-mediated mechanism of catalysis (Hallet and Sherratt, 1997; Mizuuchi, 1997). It is postulated that calcium ions stabilize the synaptic complex or TnpX structure *in vitro*. It has been shown that some calcium binding domains form a dimerization domain between two protein subunits, as for *E. coli* phospholipase A. Alternatively, calcium binding may lead to the proper folding of protein domains within the one monomeric unit (Michiels *et al.*, 2002). The TnpX amino acid sequence was examined for the presence of a Ca^{2+} binding domain but none could be found. Furthermore, in the absence of exogenous $CaCl_2$, TnpX was shown to exist as a dimer in solution, although the protein may have retained endogenous Ca^{2+} through the protein purification procedure. These data suggest that high concentrations of Ca^{2+} are not required for dimerization. However, Ca^{2+} could be required for intramolecular folding of the TnpX DNA binding domain. For example, if distal regions of the TnpX amino acid sequence combine to form the three dimensional DNA binding domain, Ca^{2+} may be required to stablize this structure and may form a novel Ca^{2+} binding domain. Preliminary results from flame ionization experiments have suggested that both Ca^{2+} and Zn^{2+} are associated with purified $TnpX_{1-707}$, although these experiments need to be repeated (F. Tynan and I. Lucet, unpublished results). Alternatively, the positive charge provided by the Ca^{2+} ions may facilitate interaction with the negatively charged DNA. If so, it would be expected that a similar affect would be observed in the presence of Mg^{2+} or Mn^{2+}, which had no effect on excision (Fig 3.5).

The excision reaction reached a point of saturation after which no further recombination occurred. Saturation may have been due to the presence of open circular and linear species of plasmid DNA in the starting DNA substrate, since it was demonstrated that these molecules are non-preferred substrates for excision (Fig 3.9). However, the proportion of

substrate DNA found to be supercoiled compared to the linear and open-circular species does not account for the approximately 50% maximal recombination. Another explanation is that the excision reaction may be in equilibrium with the insertion reaction. Given the low rate of insertion seen *in vitro* this hypothesis seems unlikely. The reaction product topology has not been determined, although standard resolvase-catalyzed excision produces 2-noded catenanes as almost the exclusive resolution product (Benjamin *et al.*, 1996; Nash, 1996; Stark and Boocock, 1995). The preference for supercoiled substrate DNA exhibited by TnpX is consistent with the standard resolvase reaction (Stark and Boocock, 1995).

Excessive protein concentrations resulted in loss of a proportion of the substrate DNA, which was trapped in the wells. Such complexes may have been the result of aborted recombination reactions, where the synaptic complex was dissociated prior to rejoining of the recombinant DNA strands. Why this occurred only at high protein concentrations was not clear but may be related to a deregulation of the process in the presence of excess TnpX. This situation would be very rare within the cell due to the tight regulation of TnpX expression. The presence of excess TnpX in the *in vitro* excision reaction may result in the release of cleaved, linear DNA still covalently attached to TnpX molecules. This DNA population would consist of high molecular weight material, even after heat inactivation of the reactions to remove any protein not covalently attached to the DNA, and would be a product of the artificial conditions used in the assay. The presence of the high molecular weight complexes could be alleviated by treatment with proteinase K prior to electrophoresis allowing higher concentrations of protein to be used.

The *in vitro* insertion assay did not yield easily interpretable results (Fig 3.11). Reaction of supercoiled plasmid DNA containing the joint of the circular intermediate and purified TnpX resulted in the production of new DNA bands in a laddered arrangement (Fig 3.11).

There are two possible explanations for this result. Laddering may have resulted from the insertion of different plasmid combinations, each with differing amounts of supercoiling, or from the insertion of linear DNA into linear or supercoiled species. These different combinations would result in bands with different DNA mobilities upon agarose gel electrophoresis. The other explanation is that insertion was not occurring and that laddering was due to topoisomerase activity exhibited by TnpX. If so, the ladders would represent plasmid species with different degrees of supercoiling or plasmid topologies as some resolvase reactions can result in the production of knotted structures (Stark *et al.*, 1991). The most likely explanation, given that an insertion-specific PCR gave a positive result, is that both possibilities were occurring. The laddering pattern was regular and it would be expected that if the insertion substrates were random then the reaction products would also be random, and subsequently would not show a uniform banding pattern.

In vitro insertion assays using PCR products as substrates were more conclusive and showed small amounts of the predicted recombinant reaction products. This experiment illustrated that $TnpX_{1\text{-}707}$ was capable of catalyzing *in vitro* insertion but that the level of insertion was not high. Note that the assay system did not show all the insertion activity that would be occurring, since the product obtained from the insertion of one PCR product into a different but identical PCR product would be indistinguishable from the original molecules. Irrespective, the insertion frequency observed *in vitro* was expected to be considerably higher given the observed *in vitro* excision and *in vivo* insertion frequencies (Chapters 4 and 5). Clearly, the reaction conditions for the *in vitro* insertion assay were not optimal. Other factors, ranging from a more appropriate starting substrate (the TA derivatives may not be optimal for insertion), to buffer conditions and possibly even other cellular co-factors, may be required for efficient *in vitro* insertion. For example, the IS*10* transposase only requires IHF *in vitro* to facilitate bending of the short, linear DNA fragments used for *in vitro* reaction assays (Allingham and Haniford, 2002).

Other large resolvase proteins such as Int from ϕC31 have been shown to catalyze insertion or integration *in vitro* (Thorpe and Smith, 1998). Recombination between the ϕC31 *attB* and *attP* sites has been demonstrated at 30°C but 16 hours incubation is required to produce 50% integration (Thorpe and Smith, 1998). The addition of either 10 mM $MgCl_2$ or 10 mM EDTA has no effect on the integration frequency and both linear and supercoiled DNA molecules serve as efficient substrates for integration (Thorpe and Smith, 1998). The Int protein from ϕTP901-1 is also integration proficient *in vitro* (Stoll *et al.*, 2002). Like ϕC31 Int, integration activity reaches a maximum after approximately 15 hours and proceeds to approximately 90% completion (Stoll *et al.*, 2002). However, neither of these proteins are capable of catalyzing efficient excision, either *in vitro* or *in vivo*, when only the Int protein is supplied. Both enzymes require an additional phage-encoded protein to catalyze high levels of excision (Breüner *et al.*, 1999; Stoll *et al.*, 2002; Thorpe and Smith, 1998; Thorpe *et al.*, 2000).

The most closely related large resolvase to TnpX is the Tn*5397*-encoded TndX protein. *In vivo*, TndX has been shown to be the only transposon encoded protein required for both excision and insertion (Wang and Mullany, 2000). The results presented in this chapter have shown that *in vitro* TnpX behaves in a similar fashion with respect to Tn*4451/3a*. In contrast, several of the phage-encoded large resolvase proteins, such as the Int proteins from ϕC31 and ϕTP901-1, Mjl from ϕFC1 and Sre from ϕR4, can catalyze integration or insertion without other proteins but efficient excision of the phage genome requires the presence of another phage-encoded protein (Breüner *et al.*, 1999; Chritiansen *et al.*, 1996; Matsuura *et al.*, 1996; Thorpe and Smith, 1998; Yang *et al.*, 2002). The recent review by Smith and Thorpe (2002) illustrated that the large resolvase proteins are clearly distinct from other resolvase proteins but that the large resolvase group constitutes a very diverse

subgroup. Functional studies suggest that the transposon-associated large resolvases, TndX and TnpX, have evolved independently from the phage related large resolvases.

Chapter 4

Identification of Functional Domains of TnpX

Introduction

The large serine recombinases are two to three times larger than traditional resolvase proteins (Smith and Thorpe, 2002). Several members of the resolvase family, such as the resolvase from IS*Xc5*, encode a C-terminal extension that is not required for function (Liu *et al.*, 1998; Smith and Thorpe, 2002). However, it is likely that the group IV, large serine recombinases do require the C-terminal extension as removal of this region from TndX and other large resolvase proteins results in a non-functional protein (Matsuura *et al.*, 1996; Smith and Thorpe, 2002; Wang and Mullany, 2000).

To facilitate excision and insertion TnpX must be capable of DNA binding, dimerization, multimerization and catalysis. Since TnpX is the only protein required for excision (Chapter 3) it must be capable of recognizing the two ends of the transposon, bringing the ends together and catalyzing strand exchange to form the circular intermediate and to reform the insertion site. To delineate which region of TnpX facilitates these functions, truncated derivatives of TnpX were constructed based upon computer analysis of the TnpX amino acid sequence.

Results

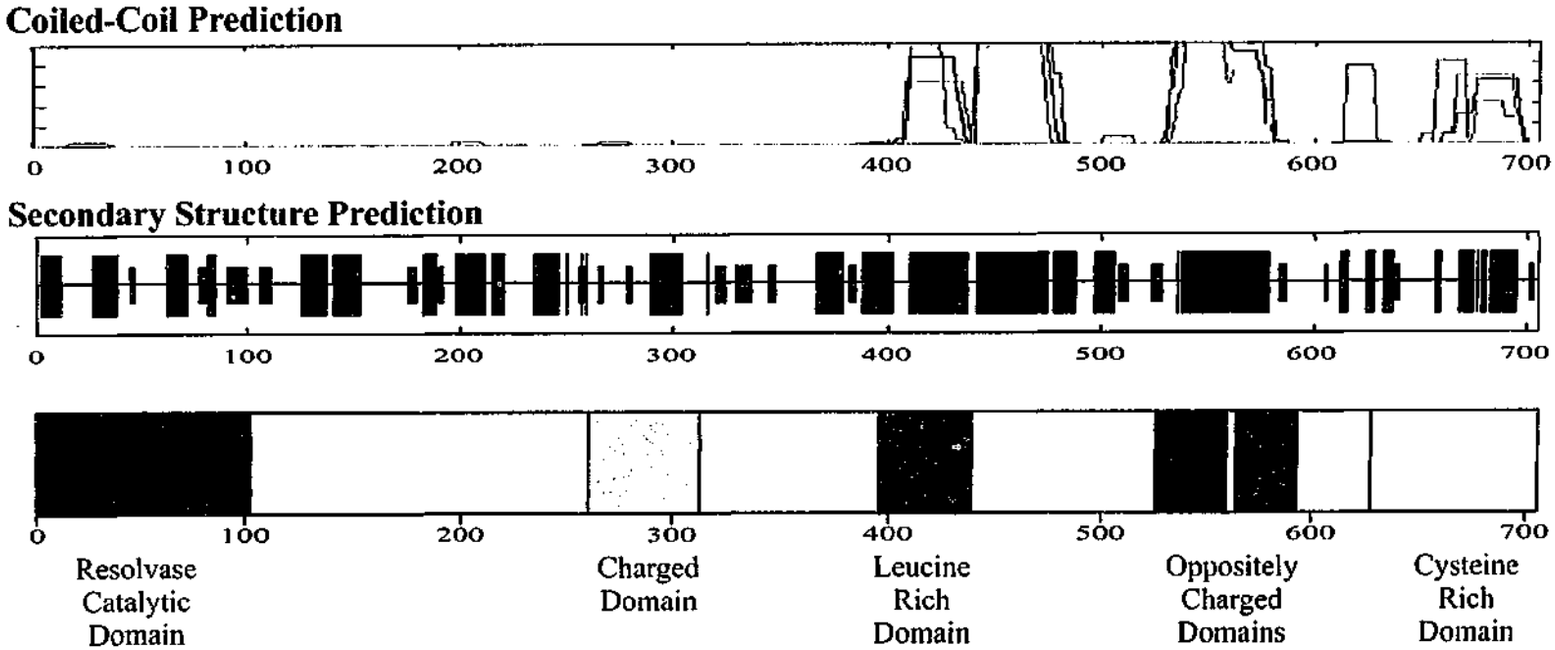
a) Computer Analysis of TnpX and Development of Truncated *tnpX* Derivatives

Initially the *tnpX* sequence was divided randomly into smaller sections in an effort to identify individual domains or motifs. The sequence fragments were submitted to the NCBI Blast server (http://www.ncbi.nlm.nih.gov/) and numerous significant matches to resolvases (large and small) were identified. Many of the matches without a significant score or e value were DNA binding proteins or multimeric proteins, which may indicate that such proteins have similar structural characteristics.

In addition to this analysis, the amino acid sequence of TnpX was submitted to various secondary structure prediction programs from the Network Protein Sequence @nalysis protein structure server (http://npsa-pbil.ibcp.fr/). These programs use statistical algorithims to predict the secondary structure of proteins. A consensus program collects and collates results from many different algorithims to give a predicted structural consensus. From this analysis it was evident that TnpX was likely to be highly α helical in structure and may contain three or more coiled-coil domains towards the C-terminal end of the protein (Fig 4.1). Coiled-coil structures are formed by the interaction of two or three α-helical strands in parallel, intertwining at an angle of 20°, as typified by the protein myosin (Lupas *et al.*, 1991). Many other proteins possess a coiled-coil structure, including leucine zipper proteins and other DNA binding proteins (Lupas *et al.*, 1991). Coiled-coil structures are often associated with protein-protein and protein-DNA interactions (Lupas *et al.*, 1991).

Figure 4.1 Computational analysis of TnpX structure

Computer analysis of the TnpX amino acid sequence involved the use of secondary structure prediction programs from the Network Protein Sequence @nalysis site (http://npsa-pbil.ibcp.fr/). Analysis included prediction of coiled-coil structures (Lupas et al., 1991). The output from this program is a graph showing the probability (0-1, Y-axis) that a particular region may form a coiled-coil structure with a maximum probability of 1. The pink, blue and green traces are calculated using window sizes of 28, 21 and 14 amino acids respectively. Secondary structure prediction was performed using the consensus program from the Network Protein Sequence @nalysis site. The output indicates regions of putative α-helical structure (blue bars), β-sheets (smaller pink bars) and random coil (black line). Finally, comparative sequence analysis carried out using the NCBI Blastp server (http://www.ncbi.nlm.nih.gov/) lead to the identification of regions within the TnpX sequence that showed sequence similarity to resolvase proteins (dark purple rectangle). Examination of the amino acid sequence of TnpX lead to the identification of other areas of interest (light purple, green, blue, red and yellow rectangles) labeled below the figure. The length of the truncated and full length TnpX derivatives is indicated by the black lines below the sequence analysis figures.



$TnpX_{1-356}$

$TnpX_{1-492}$

$TnpX_{1-597}$

$TnpX_{1-707}$

The amino acid sequence of TnpX was also examined using Blast searches (Altschul *et al.*, 1997) and several regions of interest were identified. The first ~100 amino acids had significant similarity to the catalytic domain of resolvase proteins. The region between amino acids 250 and 300 contained a reasonably large number of charged and polar residues (16 charged, 34 charged and polar amino acids). The region between 373 to 449 contained 14 leucine residues but did not form the heptad repeat characteristic of leucine zipper proteins (Haren *et al.*, 2000) or the motif found in leucine rich repeat proteins (Buchanan and Gay, 1996). From amino acids 536-577 there were two regions of interest, both highly charged in nature. The first, from amino acids 536-555, contained nine acidic residues in a total of 20 amino acids, while the second region (residues 558-577) contained 11 basic residues in a total of 20 amino acids. Clearly, this was not a random arrangement. The last 110 amino acids of TnpX contained 10 of the 17 cysteine residues in the protein and was consequently designated as the cysteine rich domain. Both the leucine rich region and the oppositely charged regions corresponded to predicted α–helical structures and coiled-coils (Fig 4.1).

This analysis was performed to guide the construction of deletion derivatives of TnpX lacking particular domains. Based on these data three truncated derivatives of TnpX were produced (Fig 4.1). The longest derivative, $TnpX_{1\text{-}597}$, would consist of a deletion of the last of the predicted coiled-coil structures and the cysteine rich domain. The truncation leading to $TnpX_{1\text{-}492}$ would serve to remove the oppositely charged domains and their corresponding coiled-coil structure. The shortest derivative $TnpX_{1\text{-}356}$, would no longer contain the leucine rich domain or any of the coiled-coil structures. All of the truncated derivatives would contain the resolvase domain and therefore the catalytic serine residue, S15 (Crellin and Rood, 1997).

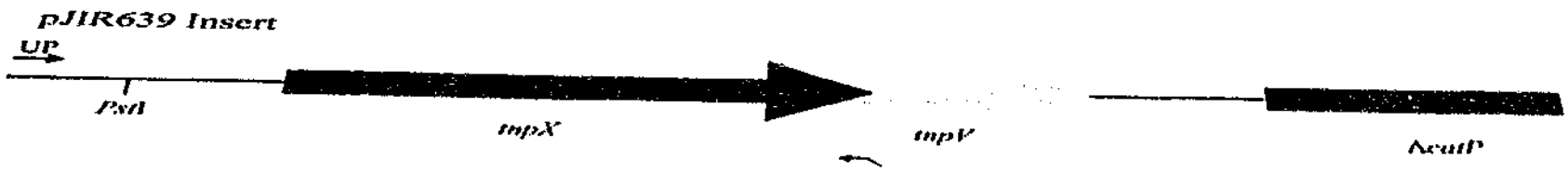
The truncated genes were generated by PCR using the same oligonucleotide primer for the 5' end (UP) and a different oligonucleotide primer for the 3' end of each gene. An in frame stop codon was introduced in each of the different 3' oligonucleotide primers, directly followed by the recognition site for the restriction enzyme *Eco*RI (Fig 4.2). The $tnpX_{1\text{-}356}$ gene was generated by introducing a stop codon downstream of the codon for amino acid R356 in the nucleotide sequence. Stop codons were also introduced in the appropriate positions to construct $tnpX_{1\text{-}492}$ and $tnpX_{1\text{-}597}$. Finally, a full length version of the *tnpX* gene was generated in a similar fashion, after incorporation of an *Eco*RI restriction enzyme site directly after the existing *tnpX* stop codon.

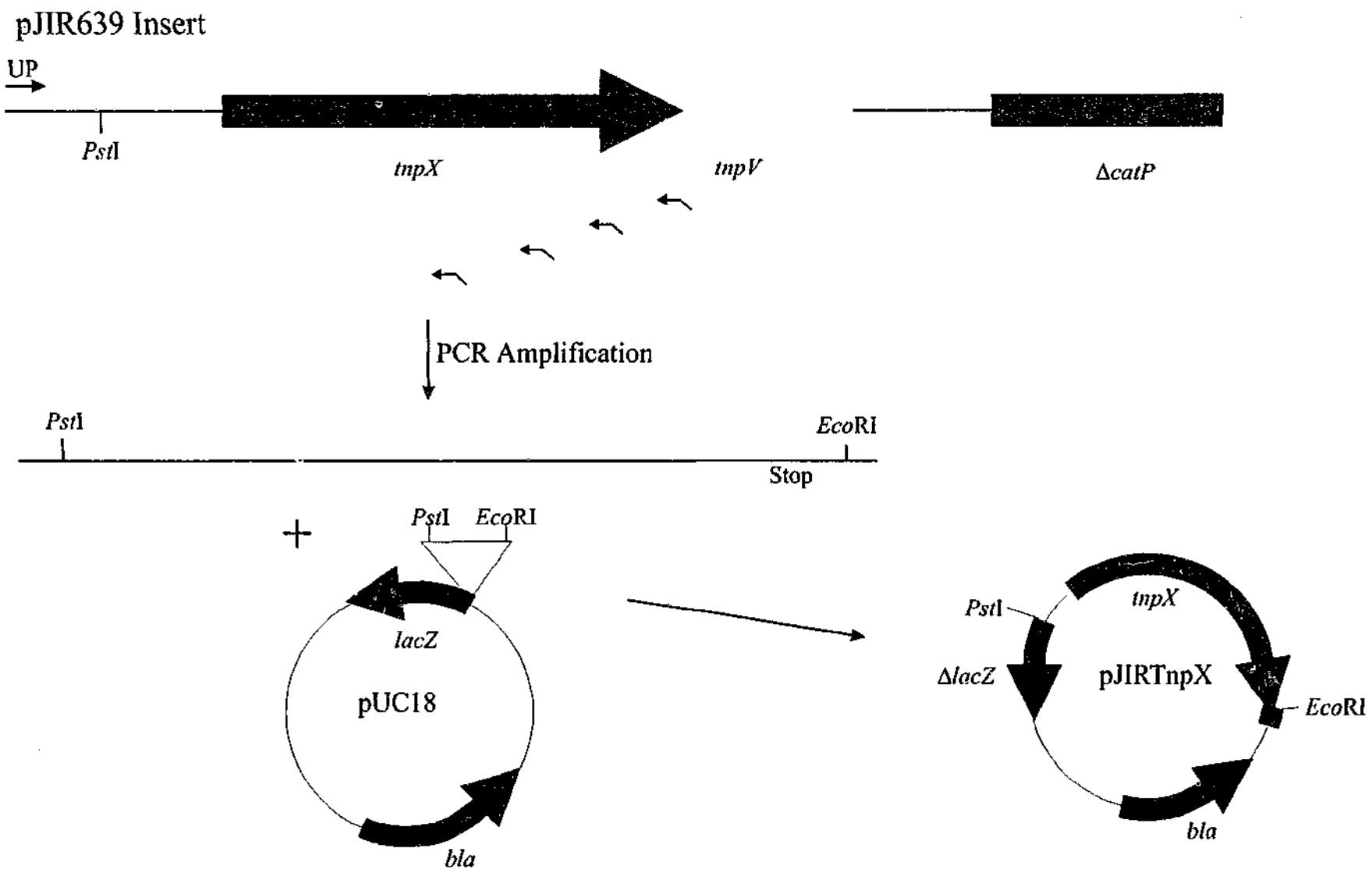
Initial attempts to PCR amplify the *tnpX* gene using *Taq* DNA polymerase (Roche) resulted in the introduction of errors within the sequence. As a consequence PCR reactions aimed at the generation of the truncated *tnpX* derivatives were performed using the Expand high fidelity enzyme (Roche). The template used for each PCR reaction was pJIR639 (Bannam *et al.*, 1995), a pBLUESCRIPT based plasmid containing the *tnpX* gene from Tn*4451* (Fig 4.2). The oligonucleotide primer, UP, binds approximately 500 bp upstream of the start of the *tnpX* gene. The resultant PCR products containing truncated *tnpX* genes were subsequently digested with *Pst*I and *Eco*RI and cloned into pUC18. The inserts in the recombinant plasmids were sequenced to ensure that no errors had been introduced into the *tnpX* gene region during PCR amplification. The full length *tnpX* construct contained a nucleotide substitution that resulted in a silent mutation but was used in subsequent experiments since no alteration was made to the derived amino acid sequence.

The plasmids carrying the truncated *tnpX* genes were designated as pJIR1860 ($tnpX_{1\text{-}356}$), pJIR1861 ($tnpX_{1\text{-}492}$), and pJIR1862 ($tnpX_{1\text{-}597}$). The control plasmid, pJIR1898, carried the full length *tnpX* gene. These derivatives were cloned in such a way that

Figure 4.2 Cloning strategy for production of truncated *tnpX* derivatives

The truncated derivatives were produced by PCR amplification using the plasmid pJIR639 as the template DNA. The pJIR639 insert is shown at the top of the figure and contains DNA usptream of Tn*4451* (downstream of the *Pst*I site), the *tnpX* gene (green arrow), which overlaps *tnpV* (yellow arrow), and the 5' end of the *catP* gene (magenta rectangle). Each PCR involved the use of the forward oligonucleotide primer UP and one of the reverse oligonucleotide primers indicated by the black arrows with a red tail. The reverse oligonucleotide primers were complementary to the *tnpX* sequence (black arrow) but also contained a region not complementary (red tail) that encoded a stop codon and the recognition sequence for the restriction endonuclease *Eco*RI. The resultant PCR products (black, green and red line) were digested with *Pst*I and *Eco*RI to facilitate introduction into the same sites on pUC18 (containing the *lacZ'* gene, large black arrow, and the *bla* gene, purple arrow) to construct the truncated *tnpX* derivatives.





transcription of the cloned genes was not driven by the *lac* promoter. Expression of *tnpX* from a strong promoter had previously been found to be lethal to *E. coli* cells and consequently the *tnpX* genes were deliberately cloned in the opposite orientation to the *lac* promoter (D. Lyras and J. I. Rood, unpublished).

b) Functional Analysis of Truncated Derivatives

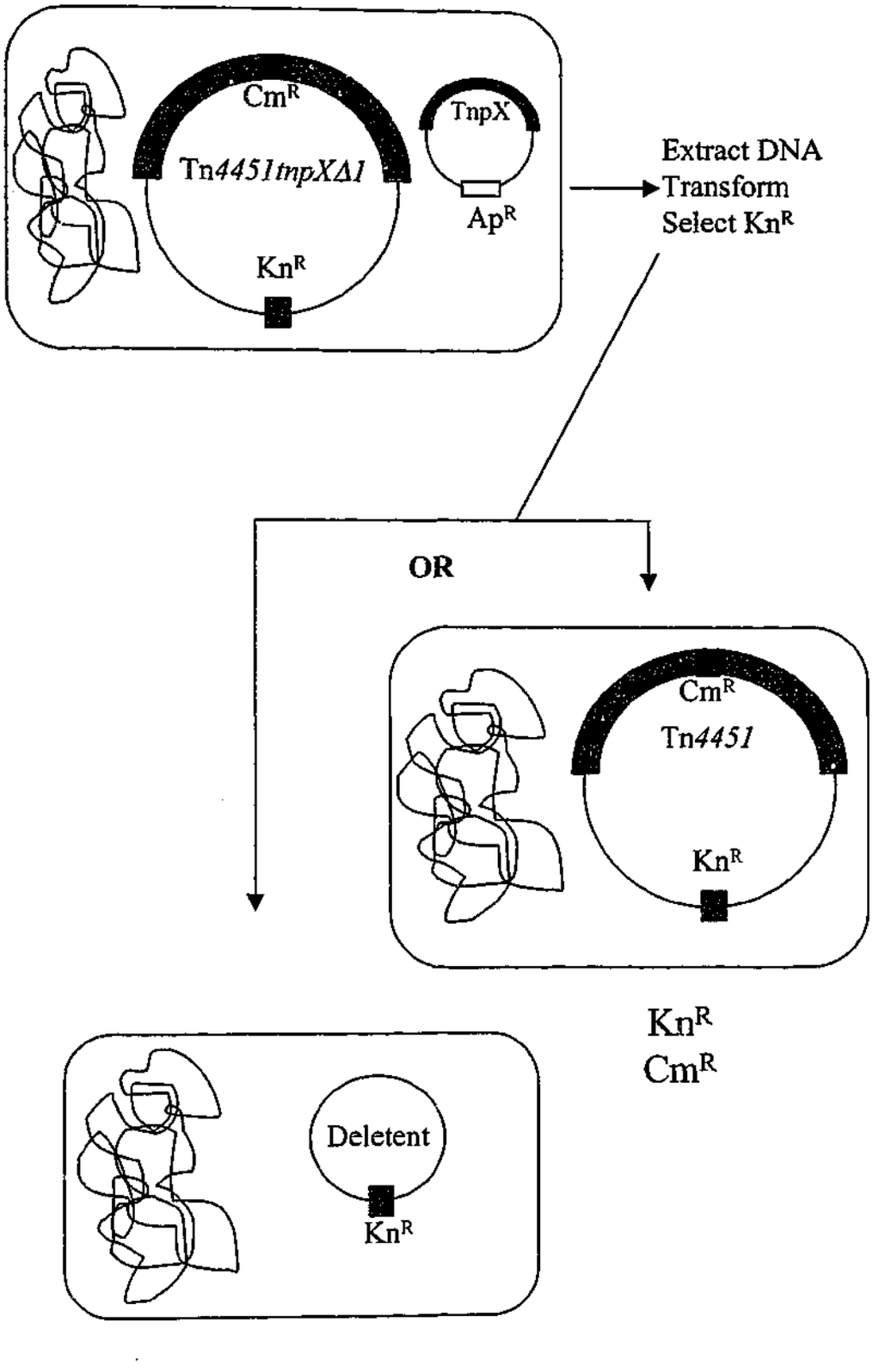
Three *in vivo* assays are commonly used in this laboratory for characterizing TnpX activity. They are a stability assay that measures excision (Crellin and Rood, 1997), a transposition assay (Lyras *et al.*, 1998) and the recently developed transposon insertion assay (D. Lyras and J. I. Rood, unpublished results). The truncated *tnpX* derivatives and the full length *tnpX* gene ($tnpX_{1\text{-}707}$) were tested for biological activity using the *in vivo* assays, which are described below.

c) Stability Assays Show that $TnpX_{1\text{-}597}$ is Functional

The stability assay monitors the excision of a defective Tn*4451* transposon. This transposon, Tn*4451tnpXΔ1* (Bannam *et al.*, 1995), contains a deletion of 185 bp within the *tnpX* gene, rendering the transposon completely stable. In these experiments, Tn*4451tnpXΔ1* carries a functional *catP* gene and is located on the kanamycin resistance plasmid, pSU39 (Fig 4.3). Excision of the defective transposon can be accomplished by introducing a wild-type *tnpX* gene on a compatible plasmid and monitoring the loss of chloramphenicol resistance (Fig 4.3). Transposon stability is determined after patching colonies from kanamycin onto chloramphenicol medium and determining the percentage of chloramphenicol resistant colonies that have retained the transposon. Therefore, a high stability level denotes a low TnpX-mediated excision frequency.

Figure 4.3 The Tn*4451* excision assay

The excision assay involves the use of an *E. coli* strain containing Tn*4451tnpXΔ1* (green) on a plasmid. This transposon is unable to excise from the plasmid and is therefore completely stable. A functional *tnpX* gene (mauve) is present on another plasmid supplying functional TnpX that excises the transposon *in trans*. This strain is incubated overnight, the plasmid DNA extracted and used to transform *E. coli* to kanamycin resistance. Those bacteria receiving the plasmid with the transposon still integrated will be resistant to chloramphenicol, those which have lost the transposon will be susceptible.


Cm^R
Tn4451tnpXΔ1
TnpX
Ap^R
Kn^R
Extract DNA
Transform
Select Kn^R
OR
Cm^R
Tn4451
Kn^R
Kn^R
Cm^R
Deletent
Kn^R
Kn^R
Cm^S

The truncated *tnpX* derivatives and the full length *tnpX* construct were introduced separately into a strain containing pJIR773 (pSU39ΩTn*4451tnpXΔ1*) (Bannam *et al.*, 1995) and stability assays carried out as before (Crellin and Rood, 1997). Each strain was assayed at least three times to obtain statistically significant data. To confirm that excision of the transposon resulted in the production of the circular intermediate, PCR was used to detect the presence of this molecule (Bannam *et al.*, 1995), using plasmid DNA extracted as part of the excision assay as the PCR template.

The results showed that neither $TnpX_{1\text{-}356}$ nor $TnpX_{1\text{-}492}$ were capable of excision. Both yielded 100% transposon stability and were PCR negative for the circular intermediate (Fig 4.4, Table 4.1). However, $TnpX_{1\text{-}597}$ was capable of excision, although it was not as efficient as the wild-type $TnpX_{1\text{-}707}$ protein. From these data it was concluded that the last 110 aa were not essential for *in vivo* excision. Further truncation of TnpX resulted in the inability to catalyze excision (Fig 4.4).

d) Transposition Assays of Truncated *tnpX* Derivatives

The truncated *tnpX* derivatives were subsequently tested for their ability to catalyze transposition. Since the transposition assay (Lyras *et al.*, 1998) tests both the excision and the insertion processes a defect in only one process cannot be distinguished. The benefit of this assay is that it involves the entire transposition process and other parameters can be investigated, such as TnpX-TnpX interactions. The assay involves the use of an *E. coli* strain that has an intact chromosomal copy of Tn*4453a*. This transposon copy shows a higher transposition frequency than a similar copy of Tn*4451* and was used to increase the sensitivity of the assay (Lyras *et al.*, 1998). Provision of additional TnpX molecules from a multicopy $tnpX^+$ plasmid significantly increases the

Table 4.1 Summary of *in vivo* assay results

TnpX Derivative	% Stability*	CI-PCR	Transposition Frequency†	Insertion Frequency‡
None	100%	-	$3.6\pm4.3 \times 10^{-6}$	$4.8\pm5.1 \times 10^{-8}$
$TnpX_{1-356}$	100%	-	$2.0\pm1.3 \times 10^{-6}$	$2.2\pm1.2 \times 10^{-6}$
$TnpX_{1-492}$	100%	-	$1.0\pm1.6 \times 10^{-7}$	$6.2\pm5.4 \times 10^{-7}$
$TnpX_{1-597}$	55.7±18.5%	+	$2.4\pm1.7 \times 10^{-4}$	$1.6\pm0.3 \times 10^{-2}$
$TnpX_{1-707}$	10.6±9.3%	+	$9.8\pm8.9 \times 10^{-5}$	$5.8\pm5.7 \times 10^{-3}$

Statistical analysis was performed using a students t-test performed using the statistical software Minitab.

* Statistical analysis of stability assay data indicates that the negative control is not significantly different from $TnpX_{1-356}$ and $TnpX_{1-492}$ ($p\geq0.88$) and is significantly different from $TnpX_{1-597}$ ($p=0.0001$) and $TnpX_{1-707}$ ($p=0.0001$). Analysis also showed that $TnpX_{1-597}$ is significantly different from $TnpX_{1-707}$ ($p=0.0018$).

† Statistical analysis indicates that $TnpX_{1-707}$, $TnpX_{1-597}$ and $TnpX_{1-492}$ are significantly different from the negative control ($p\leq0.03$) and that $TnpX_{1-356}$ is not significantly different from the negative control ($p=1$). It was calculated that TnpX1-597 and TnpX1-707 were not significantly different from each other ($p=1$).

‡ Statistical analysis indicated that all data sets are significantly different from all other data sets ($p\leq0.04$)

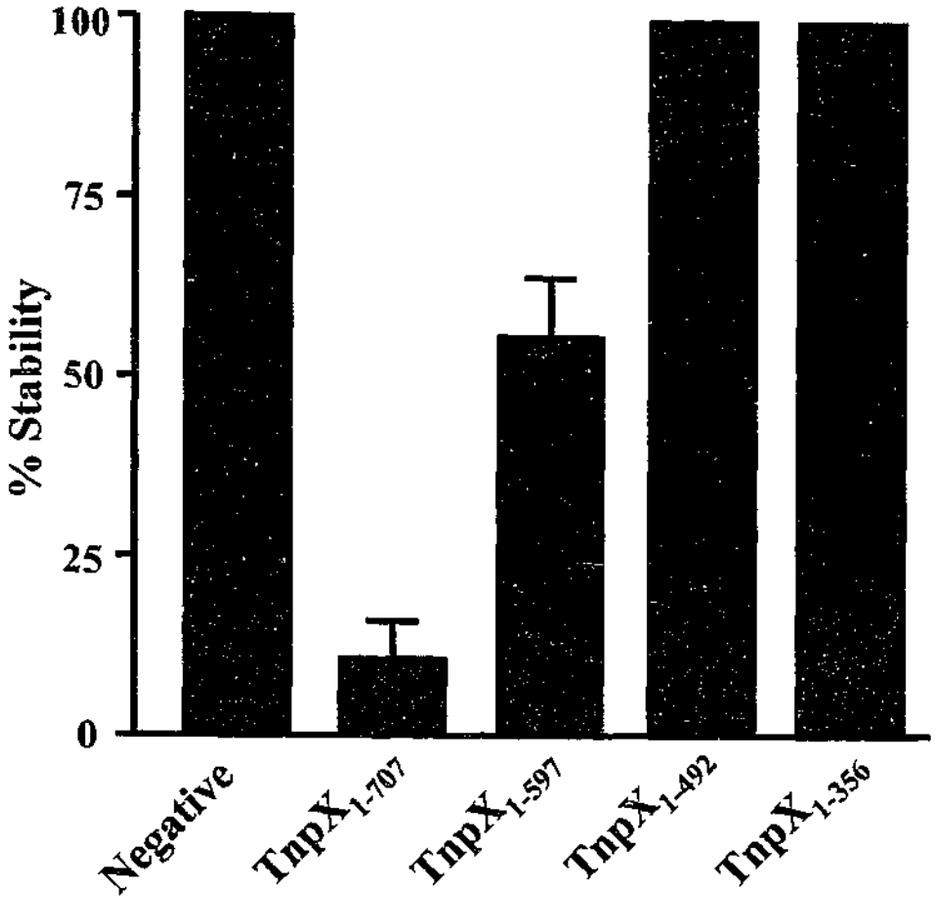


Figure 4.4 Excision activity of the truncated TnpX derivatives assessed using an *in vivo* stability assay

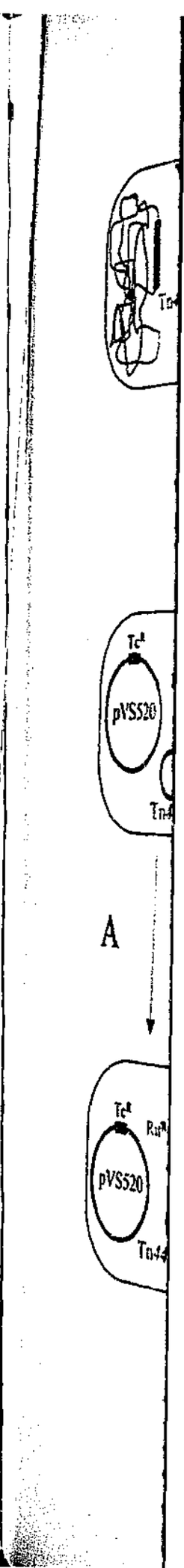
The stability assay measures the ability of a TnpX derivative to excise a defective transposon *in vivo* (Bannam *et al.*, 1995). Excision is detected by the loss of chloramphenicol resistance and the stability is calculated as the percentage of chloramphenicol resistant colonies. Results are an average of at least three independent assays. PCR was performed to detect the presence of the circular intermediate (Table 4.1). The negative control strain contains pUC18.

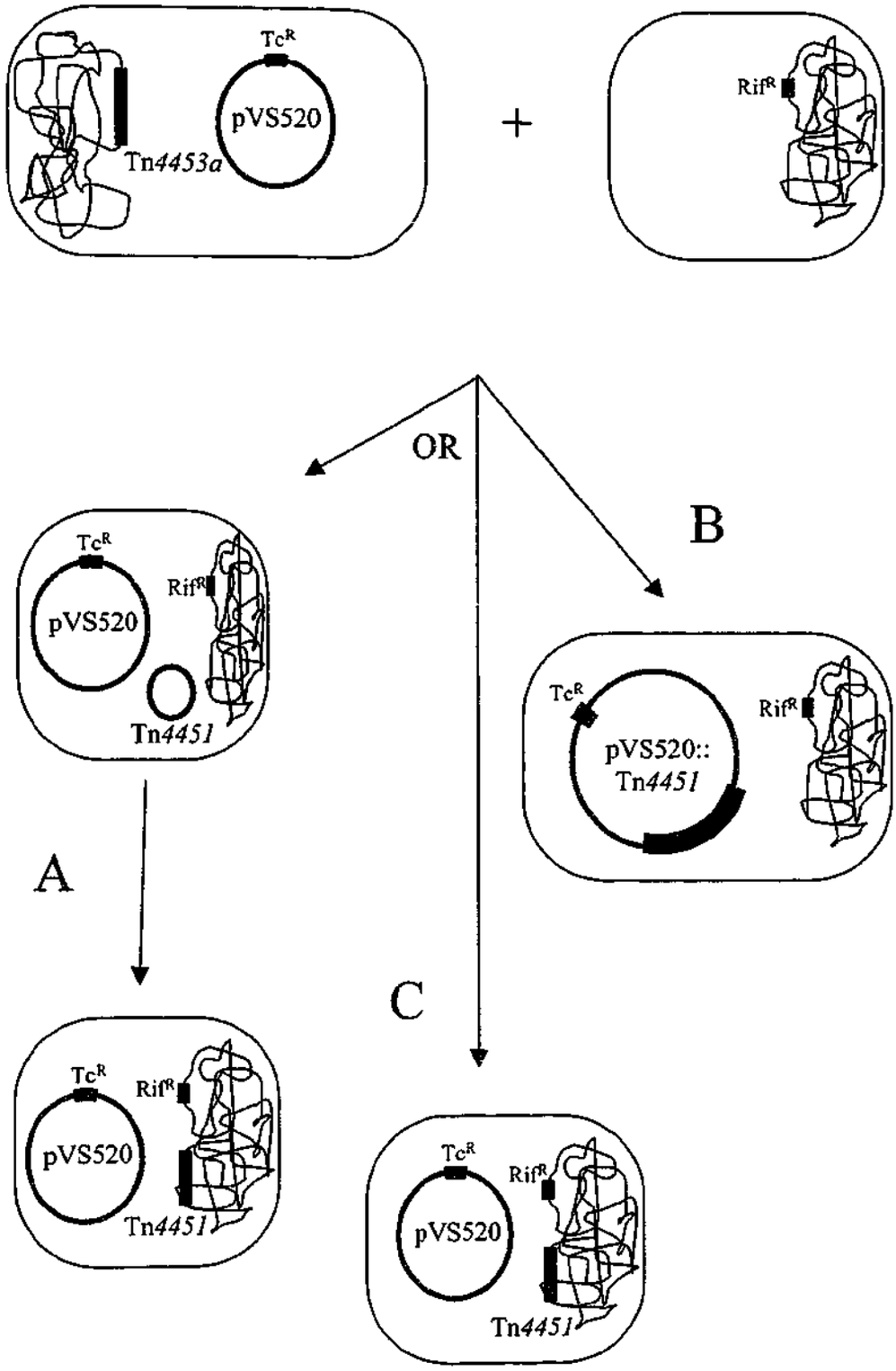
level of transposition (Lyras *et al.*, 1998). In this assay system the detection of transposition relies on the presence of the tetracycline resistant conjugative plasmid, pVS520, a derivative of RP4, and involves a filter mating between the donor strain containing Tn*4453a* and pVS520 and the recipient strain LT101, a rifampicin resistant derivative of HB101. Excision of Tn*4453a* during or prior to mating results in transfer of the circular intermediate into the recipient cell either through the action of the transposon-encoded mobilization functions, followed by insertion into the recipient chromosome or by direct insertion of the transposon into pVS520, which is subsequently transferred to the recipient (Fig 4.5). Either method requires transposition, i.e. excision from the chromosome and insertion into a new target site. Transposition is quantitated by determining the frequency of chloramphenicol resistant recipient cells detected after the matings and is internally controlled by determining the frequency of transfer of pVS520 in each experiment. The transposition frequency is calculated by dividing the number of chloramphenicol resistant recipients by the number of tetracycline resistant transconjugants.

The truncated *tnpX* constructs were introduced into the donor transposition strain and transposition assays were conducted. Since the Tn*4453a* transposon used in this assay encodes a functional *tnpX* gene there is a basal level of transposition activity evident in the control strain carrying pUC18 (Fig 4.6). The frequency obtained for full length $TnpX_{1\text{-}707}$ was approximately one log higher than the basal level while the frequency for $TnpX_{1\text{-}597}$ was consistently half a log higher again. $TnpX_{1\text{-}356}$ showed no increase from basal levels indicating that this protein did not participate in transposition or interfere in the activity of the wild-type TnpX protein encoded on the chromosome. Note that $TnpX_{1\text{-}492}$ showed a significant decrease (Table 4.1) in transposition frequency below the basal level, indicating that the $TnpX_{1\text{-}492}$ protein was inhibiting transposition of

Figure 4.5 Tn*4453* transposition assay

The transposition assay involves the use of an *E. coli* strain containing Tn*4453a* (blue) integrated onto the chromosome, as well as a conjugative plasmid encoding tetracycline resistance, pVS520 (green). This strain is mated with a rifampicin resistant (Rif^R) recipient. Tn*4453a* can be mobilised into the recipient cell either by itself, using TnpZ and the RS_A site (A), or by transposing onto the conjugative plasmid (B). Tn*4453a* may also move into the same cell as pVS520 but do so independently (C).




Tc^R
pVS520
Tn4453a
+
Rif^R
OR
B
Tc^R
Rif^R
pVS520
Tn4451
pVS520::
Tn4451
A
C
Tc^R
Rif^R
pVS520
Tn4451
Tc^R
Rif^R
pVS520
Tn4451

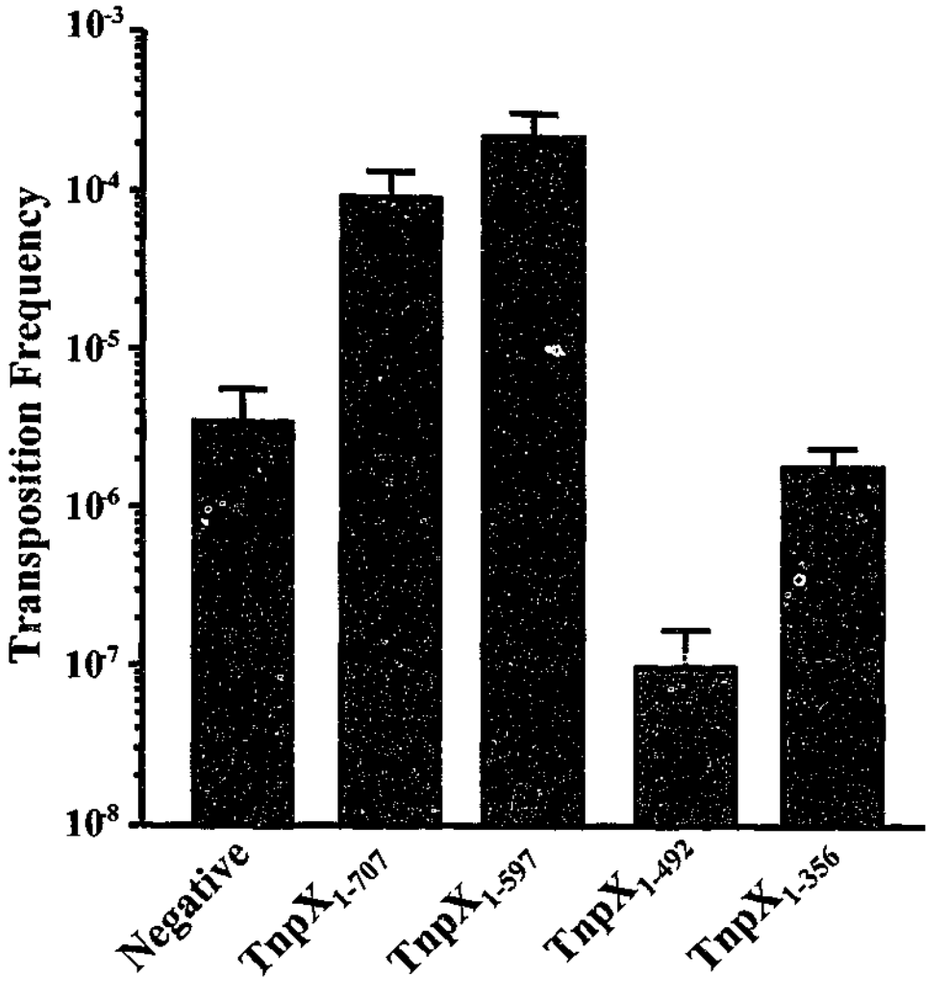


Figure 4.6 Activity of the truncated TnpX derivatives assessed using an *in vivo* transposition assay

The transposition assay involves a filter mating between a donor cell containing an intact copy of Tn*4453a* on the chromosome and the conjugative plasmid pVS520, and an additional TnpX derivative encoded by another plasmid, and a rifampicin resistant recipient. The mating mixtures were plated onto two different media; rifampicin and tetracycline, to select for recipient cells into which pVS520 has transferred, and rifampicin and chloramphenicol to detect transfer of Tn*4453a*. The transposition frequency was determined by dividing the Tn*4453a* transfer frequency by the pVS520 transfer frequency (Lyras and Rood, 2000). The negative control strain carried pUC18.

Tn*4453a* catalyzed by the wild-type TnpX protein. This phenotype was referred to as negative dominance.

e) Insertion Assays of Truncated *tnpX* Derivatives

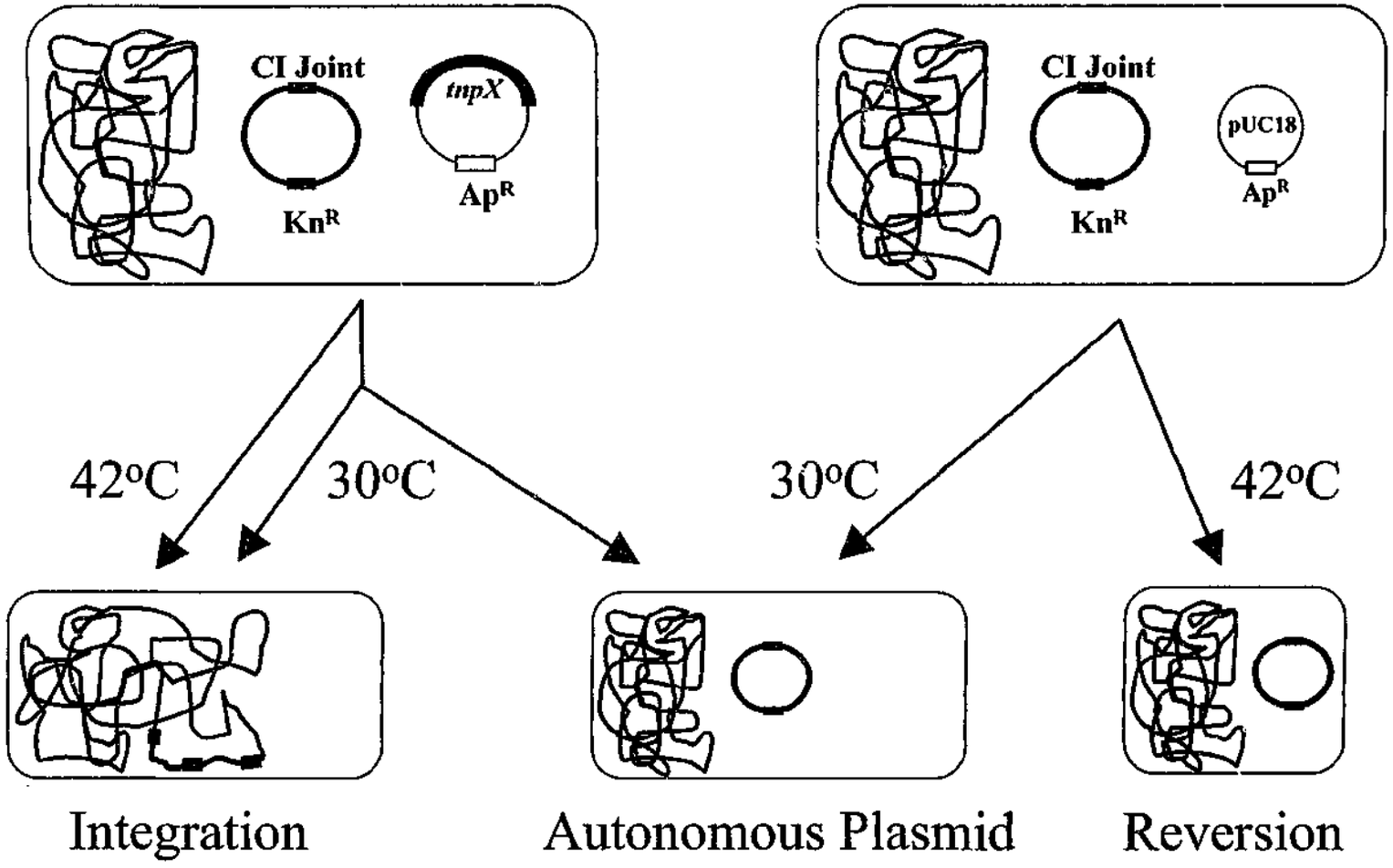
The final assay performed on the truncated *tnpX* derivatives was the insertion assay (D. Lyras and J. I. Rood, unpublished results; Fig 4.7). This assay relied on the use of a temperature sensitive derivative of the low copy number plasmid pTSK29 (Phillips, 1999). A 0.25 kb fragment that contained the joint of the circular intermediate was cloned into pTSK29 to construct pJIR2376. Due to the conditional mutation in the replication region of pJIR2376, an *E. coli* strain carrying this plasmid will only replicate when grown at 30°C on selective media. The plasmid is unable to replicate, and is subsequently lost, from a culture grown at 42°C. Introduction of a $tnpX^+$ plasmid into a strain of *E. coli* containing pJIR2376 enables the temperature sensitive plasmid to integrate into the bacterial chromosome, through the activity of TnpX acting on the joint of the circular intermediate. Consequently, when the incubation temperature is elevated these cells retain kanamycin resistance.

The truncated *tnpX* constructs were introduced into the host strain carrying pJIR2376 and grown at 30°C for several hours in 2 x YT broth containing ampicillin and kanamycin. Viable counts were then determined for each strain at both 30°C and 42°C. The frequency of insertion was determined by calculating the ratio of the number of colony forming units at 42°C and 30°C. Note that there was a basal level of growth at 42°C due to low level reversion of the temperature sensitive mutation (Fig 4.7), this reversion frequency was determined from the control strain containing pUC18 (Fig 4.8). The insertion frequency for $TnpX_{1\text{-}707}$ was found to be four orders of magnitude higher than the reversion frequency. The $TnpX_{1\text{-}597}$ derivative inserted at almost five orders of

Figure 4.7 Schematic of insertion assay

The figure shows two insertion test strains. The positive control (top left) containing a functional *tnpX* gene and a negative control (top right) containing no *tnpX* gene (pUC18). Each bacterial strain is grown overnight at 30°C, diluted and grown to mid-log phase and plated onto duplicate kanamycin plus ampicillin plates. One set of plates is incubated at 30°C and the second at 42°C. The kanamycin resistance (blue rectangle) plasmid (red circle) contains the joint of the circular intermediate (green rectangle) and contains a temperature sensitive replicon. The joint of the circular intermediate can be used to introduce the temperature sensitive plasmid (red circle) into the bacterial chromosome in the presence of a functional *tnpX* gene. The figure shows the different bacterial populations present at the different temperatures for both the positive control (red lines) and the negative control (blue lines). In the cell in the bottom left corner the temperature sensitive plasmid has been integrated into the bacterial chromosome. The negative control strain will only grow on kanamycin containing plates at 42°C if the mutation causing temperature sensitivity has reverted (purple circle).

sensitivity has reverted (purple circle).



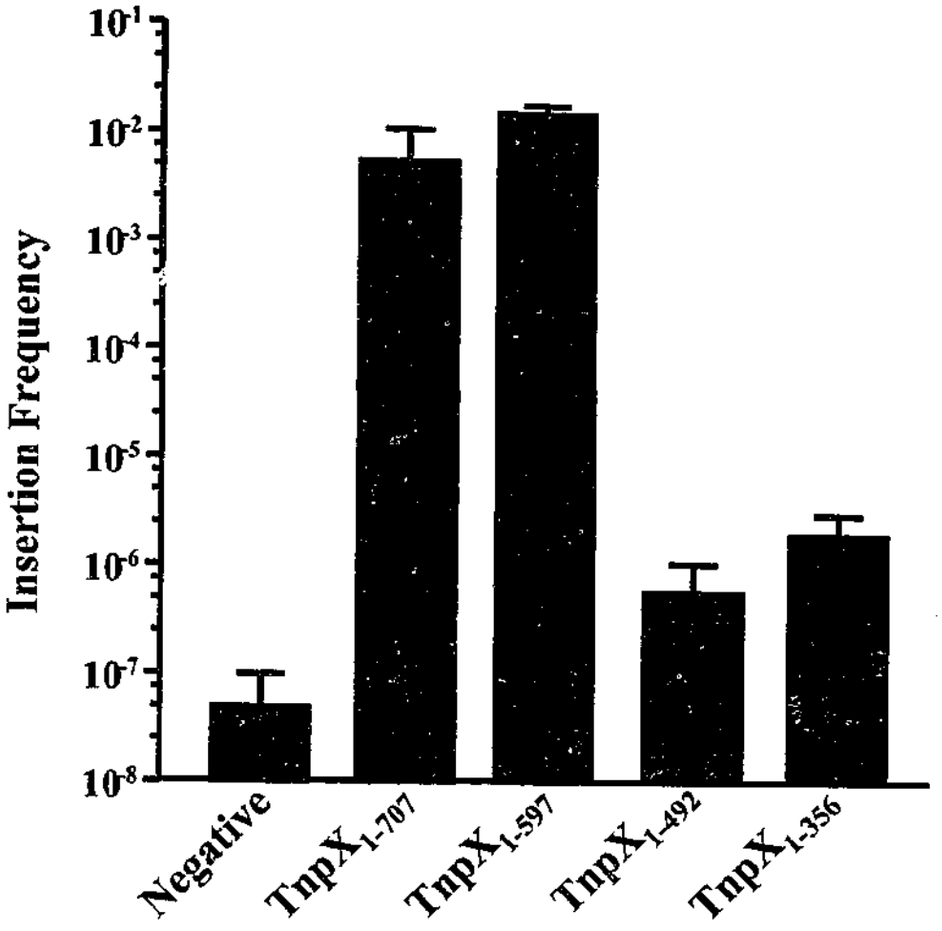


Figure 4.8 Insertion activity of the truncated TnpX derivatives assessed using an *in vivo* insertion assay

The insertion strains used for the insertion assay contained the plasmid pJIR2376, a temperature sensitive plasmid encoding kanamycin resistance and containing the cloned joint of the circular intermediate. The replication region of this plasmid did not function at 42°C. The insertion strains also contained a plasmid encoding *tnpX* or a *tnpX* derivative. The insertion assay involved duplicate plating of each insertion strain and incubating one set of plates at 42°C and the second at 30°C. Viable counts were determined for both temperatures and the insertion frequency calculated as the viable count at 42°C divided by the viable count at 30°C. The negative control contained pUC18.

magnitude higher than the reversion frequency and five times higher than the full length TnpX derivative (Fig 4.8). This result was similar to the slight but consistently higher transposition frequency obtained for $TnpX_{1-597}$ compared to the full length derivative and suggests that the increase in transposition frequency is due to the increased insertion acitivity exhibited by the $TnpX_{1-597}$ enzyme. Surprisingly, $TnpX_{1-356}$ showed an insertion frequency 100-fold higher than the reversion frequency indicating that this derivative has some insertion activity although activity was still two orders of magnitude less than the full length construct (Fig 4.8). $TnpX_{1-492}$ also showed an increase in viable colonies at 42°C although for this strain the increase in frequency was only 10-fold (Fig 4.8).

Statistical analysis indicated that each result was significantly different ($p<0.042$) to the other insertion results (Table 4.1). This analysis indicated that both $TnpX_{1-356}$ and $TnpX_{1-492}$ were still capable of catalyzing insertion of the circular intermediate joint into the *E. coli* chromosome, although at a greatly reduced level. In addition, $TnpX_{1-597}$ was capable of a significantly higher level of insertion, suggesting that the last 110 aa may actually hinder the insertion reaction (Table 4.1, Fig 4.8).

f) Expression and Purification of the Truncated TnpX Proteins

Purification of the truncated TnpX proteins was accomplished as for the full length TnpX protein, through the use of a hexa-His tag. Oligonucleotide primers were synthesized that were complementary to the ends of the truncated *tnpX* derivatives, but which removed the stop codon and replaced it with the recognition sequence for the restriction enzyme *Xho*I. Different reverse oligonucleotide primers were used for each mutant, all of which incorporated a *Xho*I site in place of the stop codon of each derivative. The oligonucleotide primer 12882 was used as the forward primer, resulting

in the incorporation of an *Nde*I site at the 5' ends of the PCR products. The PCR products were subsequently cloned into the expression vector pET22b+, as for the full length *tnpX* gene (Chapter 3). As before the sequence of each gene was determined to ensure that no errors were introduced during PCR amplification. The truncated expression constructs were introduced into the expression strain C43(DE3), expressed and purified as described for the full length protein.

Expression levels of the recombinant $TnpX_{1-356}$ and $TnpX_{1-597}$ proteins were higher than that of the full length recombinant protein, probably due to the loss of the cysteine rich domain and possibly also due to the reduction in size, from 83 kDa to 42 kDa and 70 kDa, respectively. The $TnpX_{1-492}$ protein was expressed poorly, perhaps because the truncation was within a domain and resulted in partially folded protein that was readily degraded within the cell (I. Lucet and J. I. Rood, unpublished results). The proteins were purified by metal affinity chromatography using a Talon column (Clontech). As a consequence of their higher expression levels the $TnpX_{1-356}$ and $TnpX_{1-597}$ proteins were easier to purify than $TnpX_{1-707}$ or $TnpX_{1-492}$ (Fig 4.9). $TnpX_{1-707}$ was particularly prone to degradation (Fig 4.9) and protease inhibitors were used during purification. The apparent molecular size of the recombinant proteins were very similar to the calculated molecular sizes of 83 kDa ($TnpX_{1-707}$), 70 kDa ($TnpX_{1-597}$), 58 kDa ($TnpX_{1-492}$) and 42 kDa ($TnpX_{1-356}$) (Fig 4.9).

The recombinant proteins were subjected to gel filtration using a Superdex 200 16/60 column (Pharmacia) to determine if the proteins formed multimers in solution (Table 4.2). Previously it was suggested that $TnpX_{1-707}$ is a dimer in solution and the asymmetry of the elution peak led to a higher apparent molecular size than was predicted for a dimer (Chapter 3 and I. Lucet and J. I. Rood, unpublished results). $TnpX_{1-597}$ eluted from the column in a similar fashion to $TnpX_{1-707}$ and at a volume

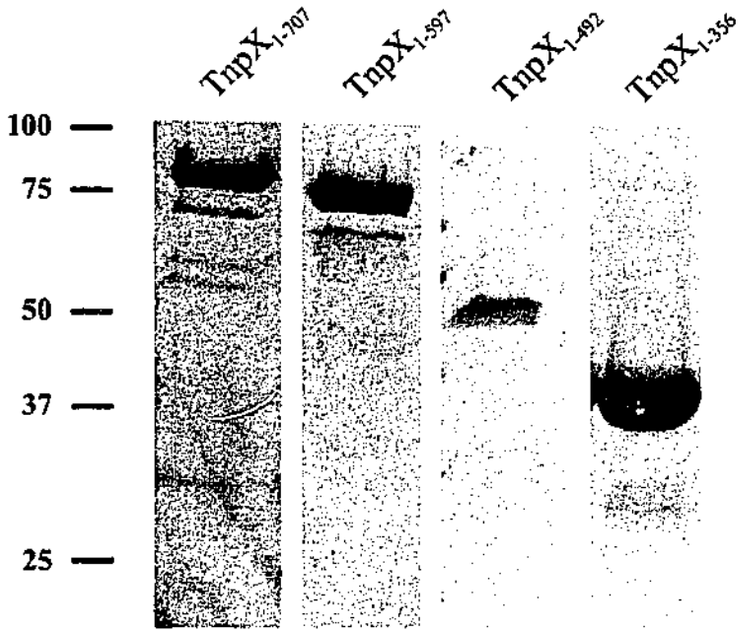


Figure 4.9 Purified TnpX and Truncated Derivatives

Coomassie stained SDS-PAGE of the Talon purified proteins used in this study. Molecular size markers are indicated on the left in kDa.

Table 4.2 Observed molecular sizes of truncated TnpX derivatives

The calculated molecular weight is shown for each TnpX derivative and constitutes the His_6-tagged monomer. The elution volume (elution vol.) was used to calculate the apparent molecular size of the TnpX derivatives in solution by comparison to protein standards of known size, namely thyroglobulin (670 kDa), gamma globulin (158 kDa), ovalbumin (44 kDa), myoglobulin (17 kDa) and vitamin B12 (1.35 kDa) (Biorad). Details of calculations are given in Chapter 2.

Table 4.2 Gel filtration data for truncated TnpX derivatives

TnpX Derivative	Calculated Molecular Weight (kDa)	Peak no.	Elution vol. (mL)	Apparent Molecular Size (kDa)
$TnpX_{1-356}$	42	1	85.3*	40
		2	77.1**	82
$TnpX_{1-492}$	58	1	75	98
$TnpX_{1-597}$	70	1	69.7	155
$TnpX_{1-707}$	83	1	67	196

* $TnpX_{1-356}$ monomer peak

** $TnpX_{1-356}$ dimer peak

suggestive of a size of 155 kDa, which corresponded to more than two molecules from the estimated monomeric size but less than three. Similar to $TnpX_{1-707}$, the $TnpX_{1-597}$ peak was asymmetrical, leading to the slightly larger than expected apparent molecular size (Fig 4.10). Recombinant $TnpX_{1-492}$ eluted at a size suggestive of a dimer as the peak was more symmetrical, while recombinant $TnpX_{1-356}$ had two peaks, one corresponding to a monomer peak and the other a dimer, these peaks were also more symmetrical than those of $TnpX_{1-707}$ and $TnpX_{1-597}$ (Fig 4.10). From these data it was concluded that like full length TnpX, the truncated TnpX derivatives were capable of dimerization, which suggests that the dimerization site is likely to be located prior, but close to the $TnpX_{1-356}$ truncation site. The asymmetry noted in the elution peaks of $TnpX_{1-707}$ and $TnpX_{1-597}$ may be due to the presence of the coiled-coil domain between amino acids 370-500 giving these proteins a less than globular shape. This structure is absent in $TnpX_{1-356}$, and likely to be destabilized in $TnpX_{1-492}$, hence both of these proteins show more symmetrical elution peaks, supporting this conclusion.

g) DNA Binding Analysis of TnpX

Gel mobility shift assays were employed to determine the DNA binding sites of the TnpX protein. It was expected that purified TnpX would bind to the left and right ends of the integrated transposon, to the joint of the circular intermediate and to the sites derived from transposon deletion and insertion events (Fig 4.11). Each DNA fragment tested was produced by PCR amplification, labelled with DIG at the 3' end of the DNA and was detected using a standard chemiluminescent DIG detection method using CSPD as the substrate.

The left end DNA fragment encompassed a region extending 163 bp upstream of the GA dinucleotide (flanking DNA) and 195 bp downstream of the GA dinucleotide

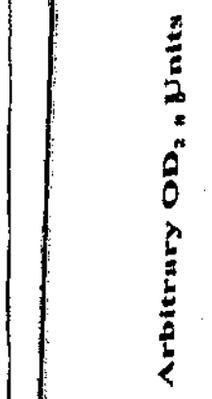


Figure 4.10 Determination of native molecular weight of truncated derivatives

Purified $TnpX_{1-707}$ and truncated derivatives were applied to a Superdex 200 16/60 gel filtration column (Pharmacia) and eluted. Fractions were collected and analyzed by SDS-PAGE. The elution volume was determined from the UV absorbance peaks shown by SDS-PAGE to correspond to each protein, $TnpX_{1-707}$ (red trace), $TnpX_{1-597}$ (blue trace), $TnpX_{1-492}$ (green trace) and $TnpX_{1-356}$ (pink trace).

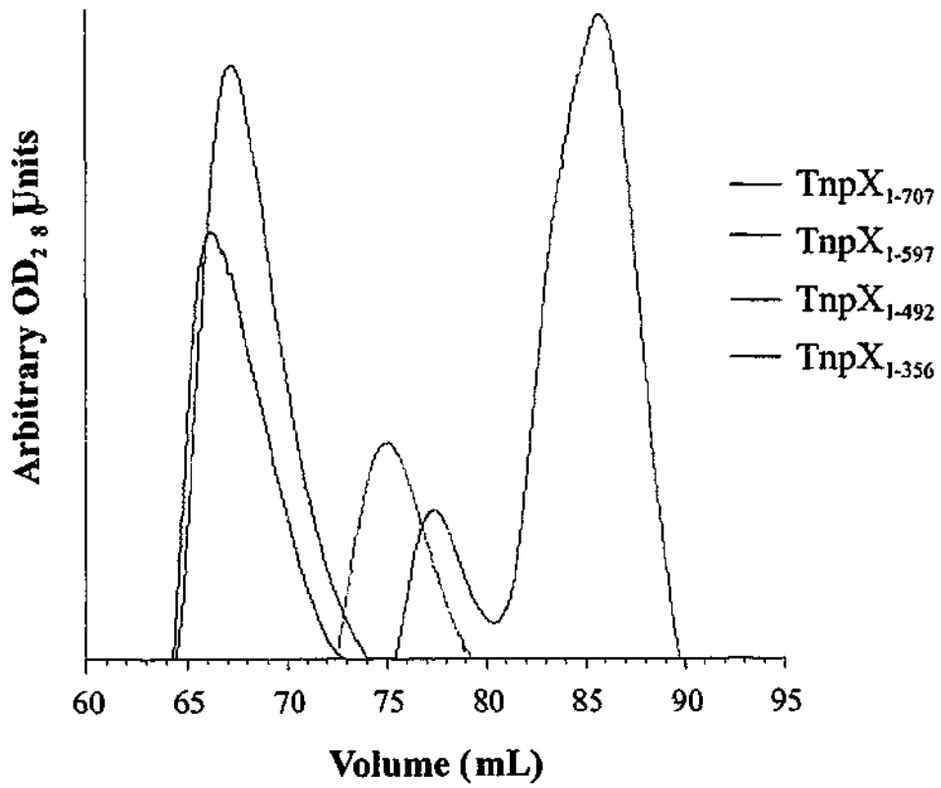

Arbitrary OD_{280} Units
60
65
70
75
80
85
90
95
Volume (mL)
$TnpX_{1-707}$
$TnpX_{1-597}$
$TnpX_{1-492}$
$TnpX_{1-356}$

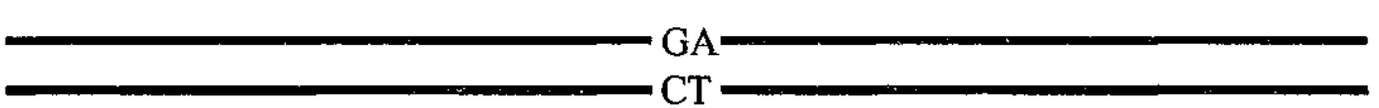
Figure 4.11 Location of DNA fragments used for TnpX binding studies

DNA target fragments used for gel mobility shift analysis and DNaseI protection assays. The top diagram shows the integrated or linear transposon. The *tnpX* (blue block arrow and blue block arc) end of the transposon is designated as the left end and the left end fragment (orange line) encompasses 163 bp of flanking DNA, the GA dinucleotide and 195 bp of the left end of the transposon. The *tnpW* (red arrows) end of the transposon is designated as the right end and the right end fragment includes 208 bp of the right end of the transposon, the GA dinucleotide and 135 bp of flanking DNA. The joint of the circular intermediate (CI) results from the juxtaposition of the left and right ends and the CI fragment comprises 195 bp of the left end of the transposon, the GA dinucleotide and 208 bp of the right end of the transposon. A deletion site is a site from which the transposon has been excised (Abraham and Rood, 1987). It contains the GA dinucleotide, found approximately in the centre of a 369 bp fragment. Insertion sites are regions into which transposon insertion has been detected (Crellin and Rood, 1997), both sites contain a GA dinucleotide.

Left End

Right End

Circular Intermediate



Deletion/Insertion Sites

(transposon DNA), a total of 360 bp (Fig 4.11). The DNA was mixed with varying amounts of purified TnpX, incubated at room temperature for 15 minutes then loaded onto a native polyacrylamide gel for electrophoresis. Incubation with full length TnpX protein led to the concentration dependent formation of two shifted complexes (Fig 4.12). Complex I was the less predominant of the two complexes and when sufficient quantities of TnpX were added complex II was almost the only complex formed (Fig 4.12). Competitive gel mobility shift assays were performed to determine if binding of TnpX to the left end fragment was specific. Two species of unlabeled competitor DNA were added to the normal binding reaction, either specific competitor (the same DNA fragment) or non-specific competitor (a 270 bp fragment derived from upstream of the unrelated *pfoA* gene from *C. perfringens*). If binding was specific for the DNA fragment tested, the addition of excess specific competitor would result in the loss of DNA binding whereas addition of excess non-specific DNA would have no effect on DNA binding. The results showed that $TnpX_{1-707}$ binding to the left end of Tn*4451* was sequence specific (Fig 4.13).

The binding experiments were repeated using PCR fragments derived from the right end and the joint of the circular intermediate (Fig 4.11). The circular intermediate fragment was 405 bp in length encompassing 195 bp of the *tnpX* end of the transposon and 208 bp of the *tnpW* side (Fig 4.11). The right end fragment was 343 bp in length including 208 bp of transposon sequence and 135 bp of flanking DNA (Fig 4.11). TnpX was found to bind to both of these fragments (Fig 4.14). Due to the larger size of the circular intermediate fragment (405 bp) the gel mobility shift assays carried out on this DNA target were analyzed by agarose gel electrophoresis since the DNA-protein complex was too large to enter the polyacrylamide gel. Binding of TnpX to the right end fragment showed a binding pattern very similar to that of the left end fragment (Fig 4.12 and 4.14). Binding of TnpX to the circular intermediate was slightly different,

Figure 4.12 Gel mobility shift analysis of the left end of Tn*4451*

The DNA fragment was generated by PCR and labelled with DIG. The labelled DNA fragment was incubated with the indicated amounts of purified $TnpX_{1-707}$ for 15 minutes followed by non-denaturing PAGE to separate the unbound DNA from the protein-DNA complexes. The DNA was subsequently transferred to a nylon membrane and the DIG-labelled DNA detected using an anti-DIG antibody and the CSPD substrate, followed by exposure to and development of the X-ray film. The figure shows the developed autoradiograph labelled at the bottom with the amount of $TnpX_{1-707}$ (ng) incubated with the DNA. The positions of the unbound DNA and the two shifted complexes produced are indicated to the right.

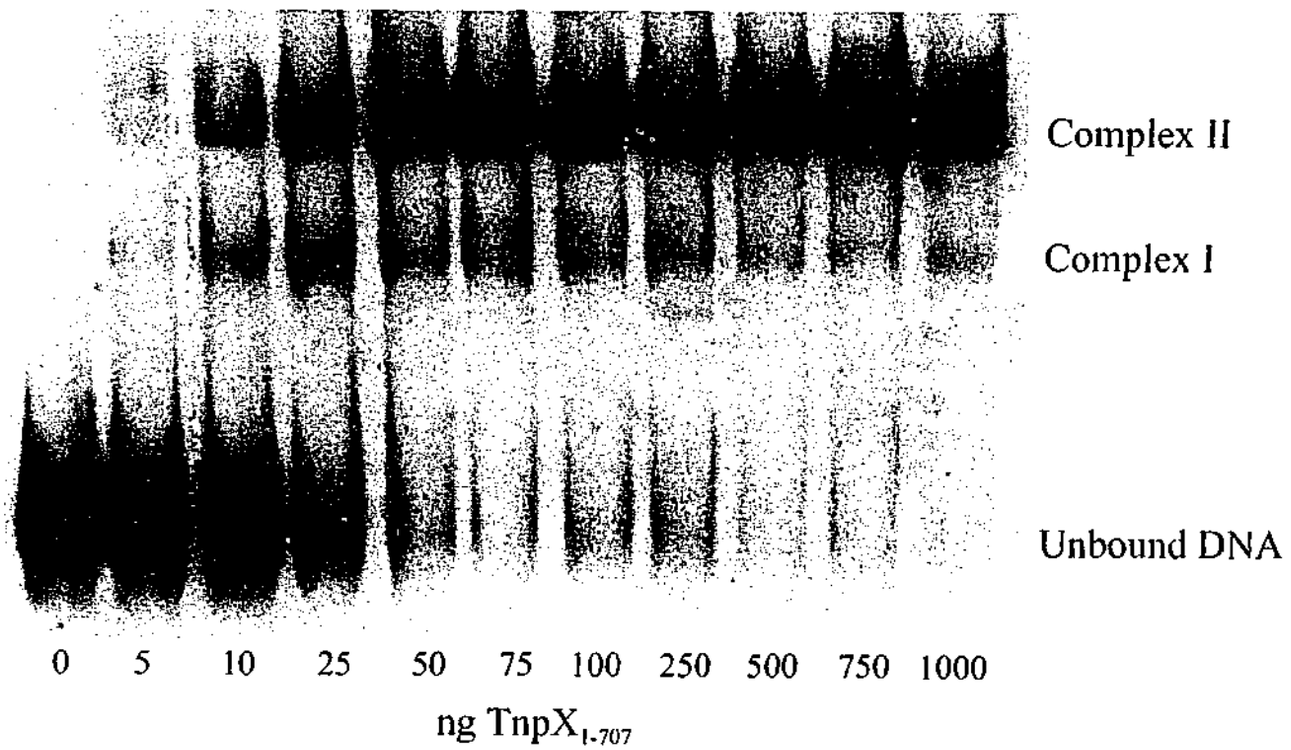
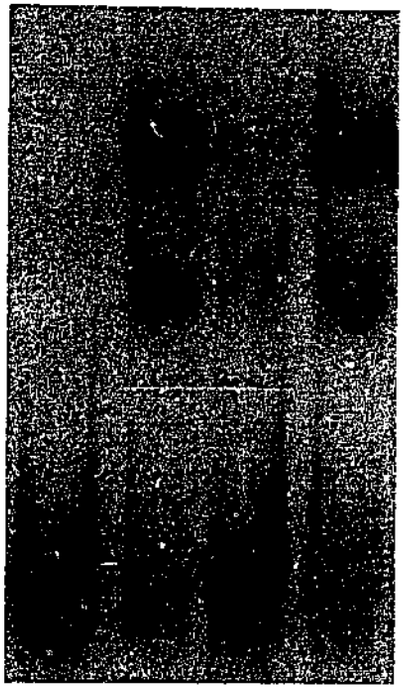

Complex II
Complex I
Unbound DNA
0 5 10 25 50 75 100 250 500 750 1000
ng TnpX1-707

Figure 4.13 Determination of the binding specificity of TnpX for the left end fragment

A gel mobility shift assay was performed as described in Figure 4.12. The amount of DIG-labelled left end fragment added to the reaction was 15 fmol. $TnpX_{1-707}$ (50 ng) was added as shown. To the second reaction (lane 2) only labelled left end DNA and $TnpX_{1-707}$ were added. Reaction three (lane 3) also contained 500 fmol of unlabelled left end DNA (Sp DNA). Similarly reaction four (lane 4) also contained 500 fmol of non-specific competitor DNA (Non-Sp DNA) derived from the upstream region of the *C. perfringens pfoA* gene.

$TnpX_{1-707}$	-	+	+	+
Non-Sp DNA	-	-	-	+
Sp DNA	-	-	+	-



Lane Number	1	2	3	4

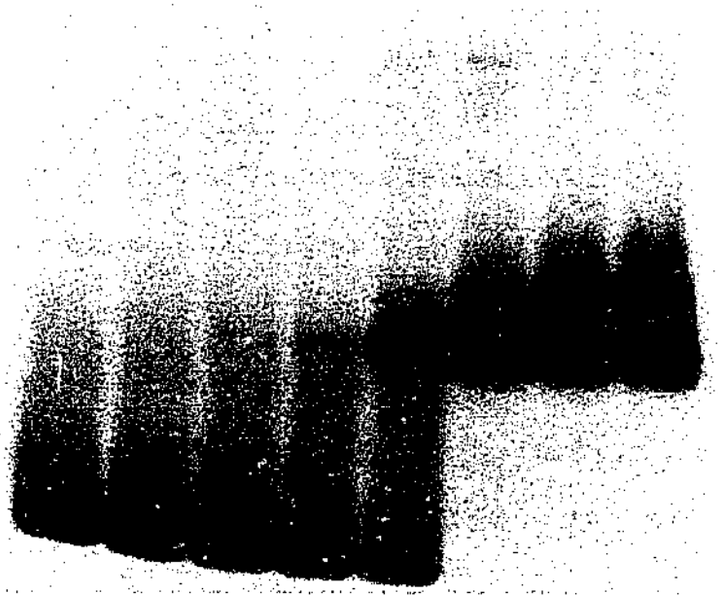
Figure 4.14 Gel mobility shift analysis of the circular intermediate and the right end of Tn*4451*

(A) The circular intermediate DNA fragment was generated by PCR and labelled with DIG. The labelled DNA fragment was incubated with the indicated amounts of purified $TnpX_{1\text{-}707}$ for 15 minutes followed by electrophoresis through a 2% agarose gel. The DNA was subsequently transferred to a nylon membrane and the DIG-labelled DNA was detected as before. The positions of the unbound DNA and the bound complexes are indicated on the right.

(B) The right end DNA fragment was generated by PCR and treated similarly to the circular intermediate DNA fragment. However, the right end-protein reactions were subjected to non-denaturing PAGE. The positions of the unbound DNA and the two shifted complexes are indicated on the right.

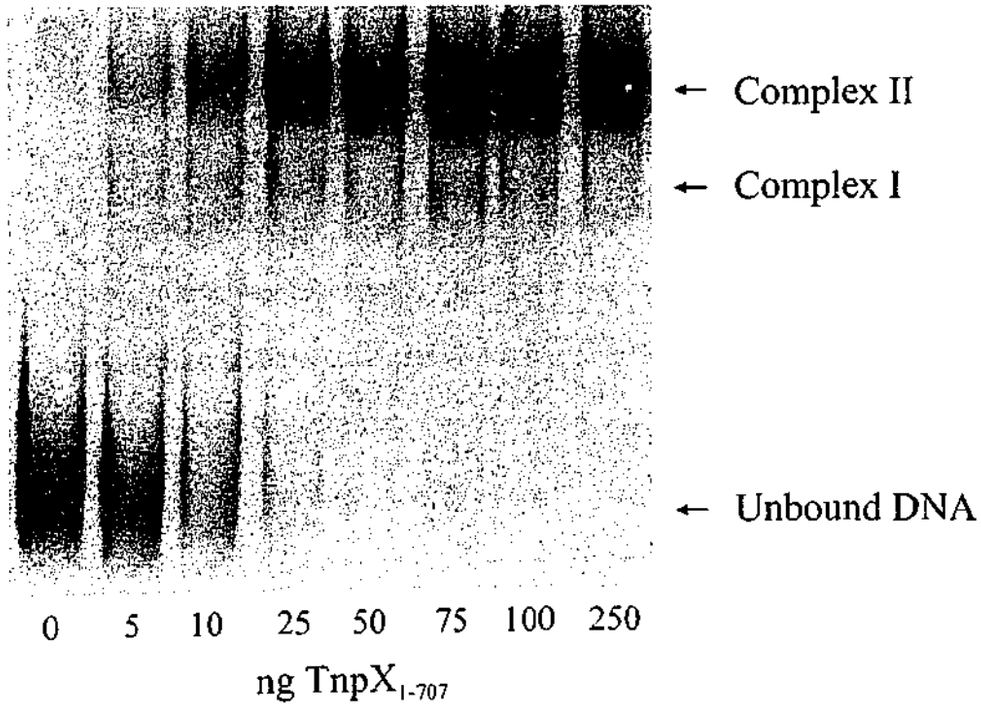
A

Circular Intermediate



B

Right End



with the suggestion that there were three concentration dependent bands present rather than two (Fig 4.14). Binding specificity experiments were repeated using the right end and the circular intermediate fragments and binding was shown to be specific (data not shown).

The ability of purified TnpX to bind to deletion and insertion sites was investigated by PCR amplifying DNA fragments of a known Tn*4451* deletion site from *C. perfringens* (Abraham and Rood, 1987) and two *E. coli* chromosomal insertion sites (Abraham and Rood, 1987). These target sites were designated D1, I1 and I2, respectively, and were 369 bp in size for the deletion site and approximately 320 bp and 280 bp for the two insertion sites. $TnpX_{1-707}$ was shown to bind to these sites (Fig 4.15) but only at much higher concentrations of protein. However, binding was still found to be specific (data not shown).

h) Quantitative DNA Binding Studies

To determine the binding affinity of TnpX for the different binding sites, quantitative gel mobility shift assays using [^{32}P] labelled DNA were conducted. Protein titrations using the different DNA fragments were performed and the DNA binding was quantitated. This was achieved by detecting the radiolabelled DNA by exposure to a phosphor screen, which was subsequently scanned and the percentage of DNA bound to protein plotted against protein concentration. The results showed that TnpX binds with the same affinity to the left end and right end fragments and with equal or slightly better affinity to the circular intermediate joint (Fig 4.16A). However, the most interesting result was the difference in binding of $TnpX_{1-707}$ to the deletion and insertion sites compared to the high affinity binding $TnpX_{1-707}$ exhibits to the transposon ends and the joint of the circular intermediate. The binding affinity of $TnpX_{1-707}$ for the deletion and

Figure 4.15 Gel mobility shift analysis of deletion and insertion sites

The DNA fragments were generated by PCR and labelled with DIG. The labelled DNA fragments were incubated with the indicated amounts of purified $TnpX_{1-707}$ for 15 minutes followed by non-denaturing PAGE and analysed as before. The panel labelled D1 was performed using DNA derived from a known *C. perfringens* deletion site. The second and third panels labelled I1 and I2 were performed using two different *E. coli* insertion sites. The positions of the unbound DNA and the shifted complexes are indicated on the right.

D1 I1 I2

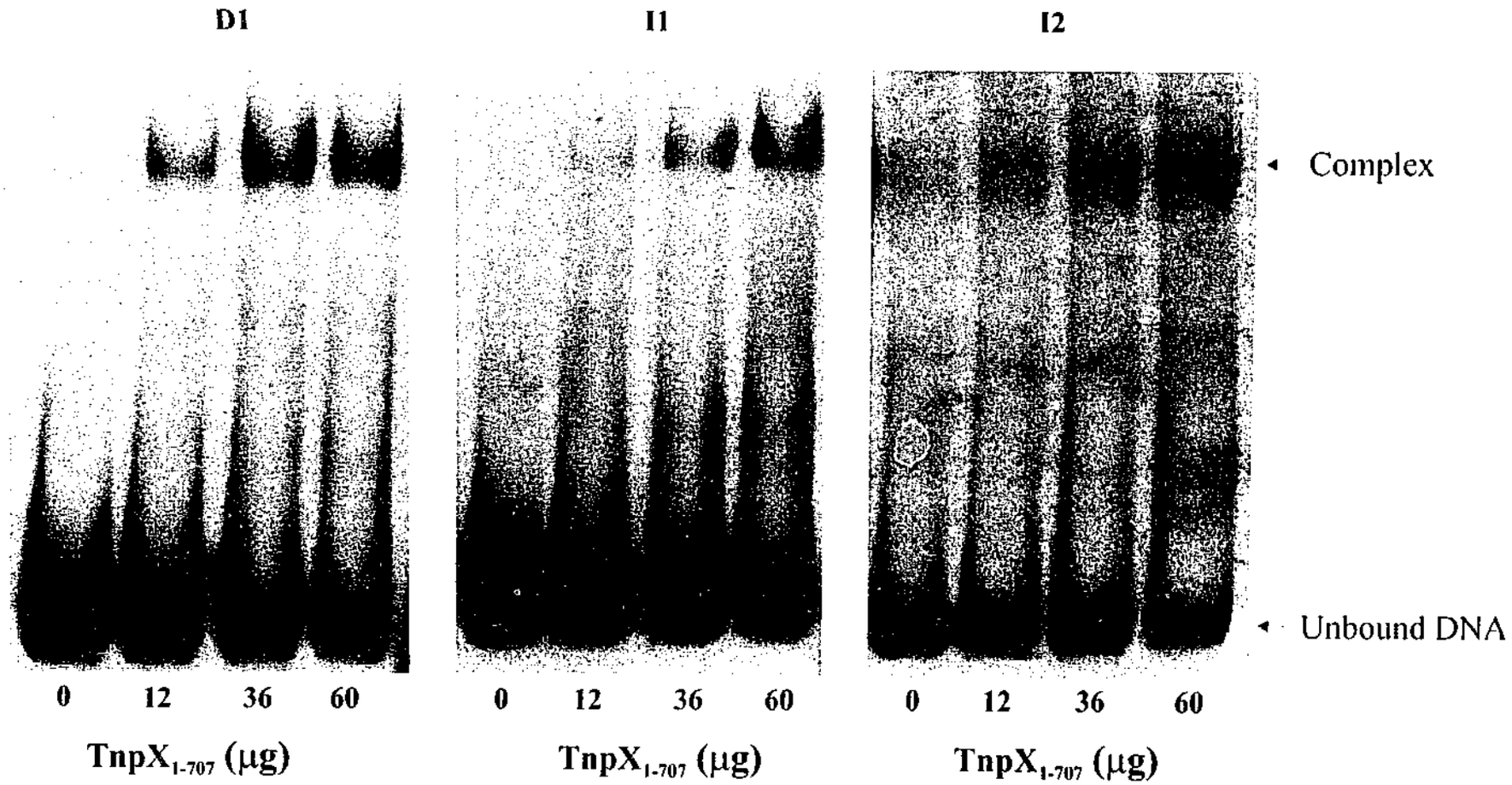
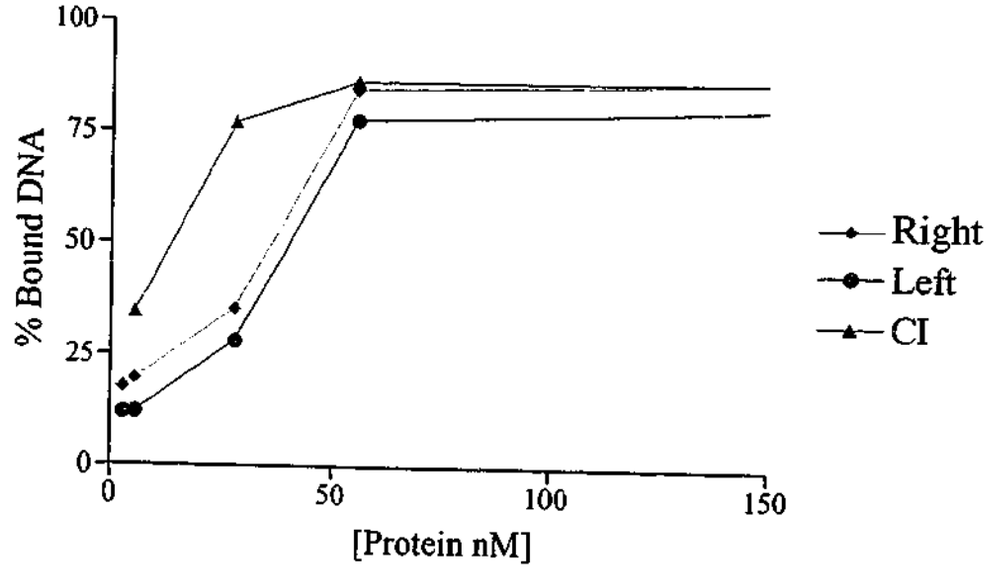

D1
I1
I2
Complex
Unbound DNA
0
12
36
60
TnpX1-707 (μg)
0
12
36
60
TnpX1-707 (μg)
0
12
36
60
TnpX1-707 (μg)

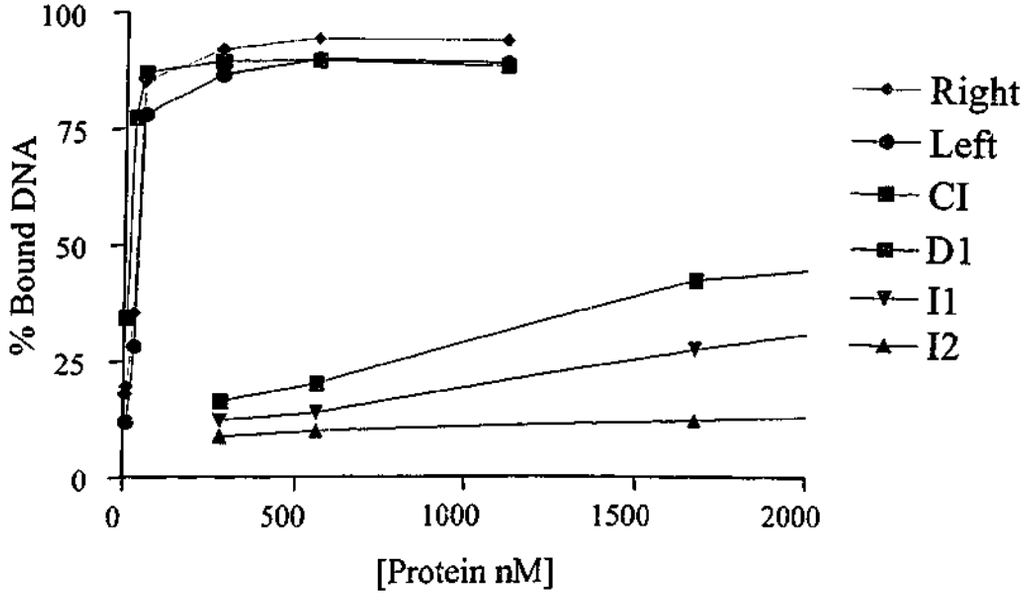
Figure 4.16 Quantitative binding of TnpX to DNA from the left end, right end, circular intermediate, deletion and insertion sites

Quantitative gel mobility shift assays were performed using the same DNA fragments as those used in the previous gel mobility shift analyses. The DNA fragments were 5' end labelled with [^{32}P]-ATP and the fragments were incubated with increasing amounts of protein. The binding reactions using the left end, right end, deletion and insertion site DNA fragments were separated by native PAGE or agarose gel electrophoresis. The gels (acrylamide and agarose) were dried using a vacuum dryer and exposed to a phosphor screen overnight. The phosphor screen was scanned and the number of pixels, representing labelled DNA, were determined for both the bound and unbound bands. The % bound DNA was calculated and plotted against the protein concentration. The binding affinity of $TnpX_{1\text{-}707}$ for the right end (green trace), left end (red trace) and circular intermediate (blue trace) was compared (Panel A) using a protein concentration scale of 0-150 nM. $TnpX_{1\text{-}707}$ only bound the deletion and insertion fragments (purple traces, Panel B) at a much higher concentration.

A


100
75
50
25
0
% Bound DNA
0
50
100
150
[Protein nM]
Right
Left
CI

B


100
75
50
25
0
% Bound DNA
0
500
1000
1500
2000
[Protein nM]
Right
Left
CI
D1
I1
I2

insertion sites was lower by almost two orders of magnitude (Fig 4.16B). There was also a difference in the binding affinities between the different deletion and insertion binding sites (Fig 4.16B). Of the deletion/insertion sites the highest binding affinity was for the deletion site while I2 had the lowest binding affinity, not even reaching 50% binding (Fig 4.16B). The results indicate that $TnpX_{1-707}$ has high affinity binding sites (left end, right end and CI joint) and low affinity binding sites (deletion and insertion sites). In addition, there is also a difference between the affinities for the different deletion and insertion sites.

A previous study (Crellin and Rood, 1997) had produced several mutants in which the GA dinucleotides at both the left and right end of Tn*4451* had been altered. Subsequent stability assays performed using the mutant transposons had shown that excision frequencies were affected, generally decreasing the ability of TnpX to excise the transposon. The exception was when both GA dinucleotides were changed to TA, which resulted in increased excision of Tn*4451*. It was postulated that the observed differences in excision frequency may be due to an effect on the ability of TnpX to bind to the mutated transposon ends (Crellin and Rood, 1997). To ascertain whether this hypothesis was valid, mutated derivatives from the left and right ends of the transposon were PCR amplified, labelled with [^{32}P] and used in quantitative gel mobility shift assays. The results showed that there were no differences in TnpX binding affinities (Fig 4.17A and B). Clearly, the difference in the stability scores between the mutants was not due to a difference in TnpX binding as previously suggested (Crellin and Rood, 1997).

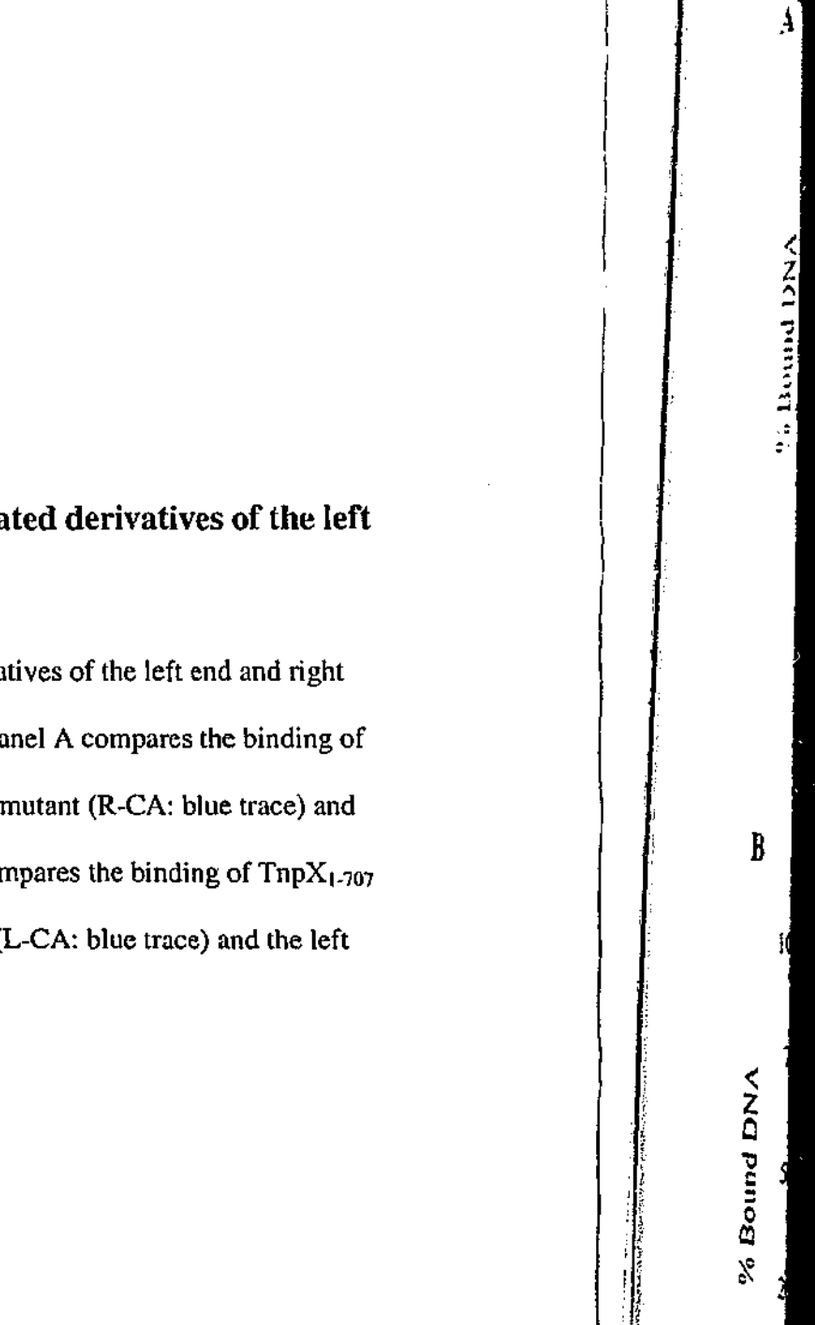
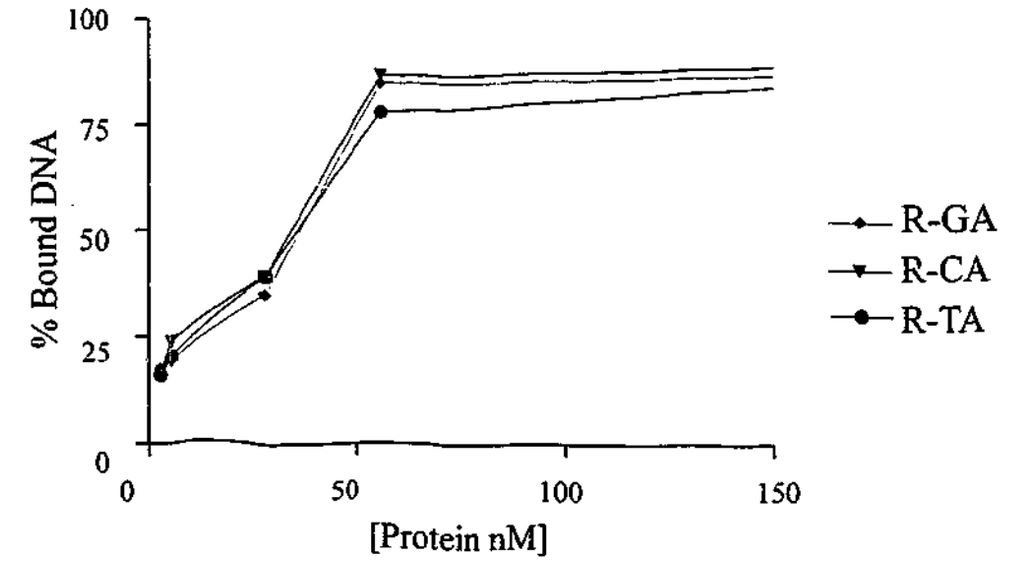


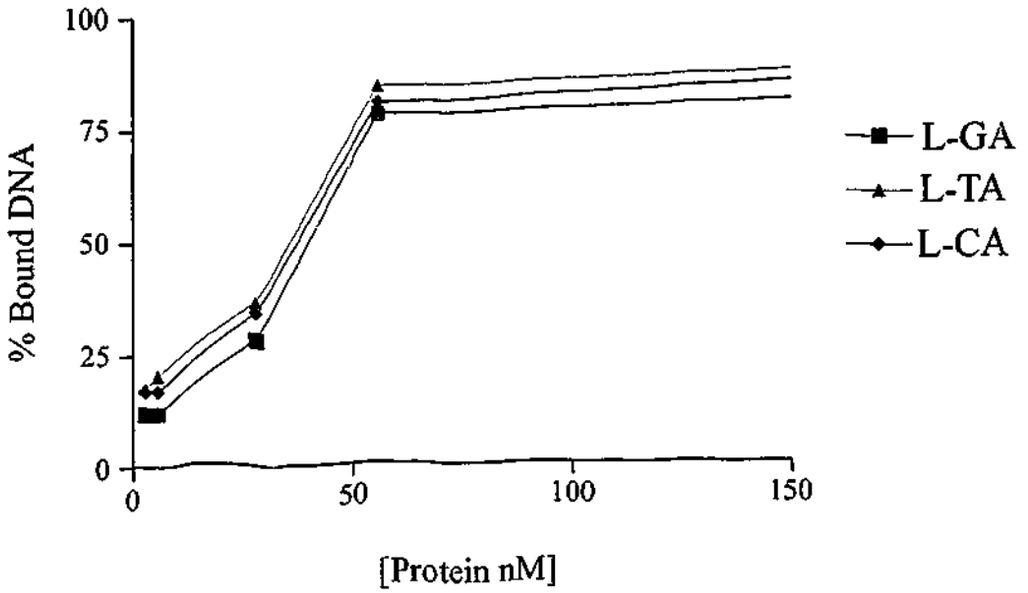
Figure 4.17 Ability of $TnpX_{1-707}$ to bind to mutated derivatives of the left and right ends of Tn*4451*

The binding affinity of TnpX to the different mutant derivatives of the left end and right end DNA fragments was determined as for Figure 4.16. Panel A compares the binding of $TnpX_{1-707}$ for the right end (green trace), the right end CA mutant (R-CA: blue trace) and the right end TA mutant (R-TA: purple trace). Panel B compares the binding of $TnpX_{1-707}$ for the left end (L-GA: red trace), the left end CA mutant (L-CA: blue trace) and the left end TA mutant (L-TA: green trace).

A


100
75
50
25
0
% Bound DNA
0
50
100
150
[Protein nM]
R-GA
R-CA
R-TA

B


100
75
50
25
0
% Bound DNA
0
50
100
150
[Protein nM]
L-GA
L-TA
L-CA

i) Determination of the DNA Content of TnpX Bound Complexes

The multiple complexes exhibiting reduced mobility that were observed in the gel mobility shift assays could have been produced in several different ways. Complex I may constitute a single DNA fragment bound to either a TnpX monomer or dimer. Complex II may consist of a single DNA fragment bound to a dimer of TnpX or may represent two DNA fragments bound either to a monomer, dimer or tetramer. Alternatively, complex II may represent the same constituents as complex I but with an altered conformation, such that the complex is retarded significantly more during electrophoresis.

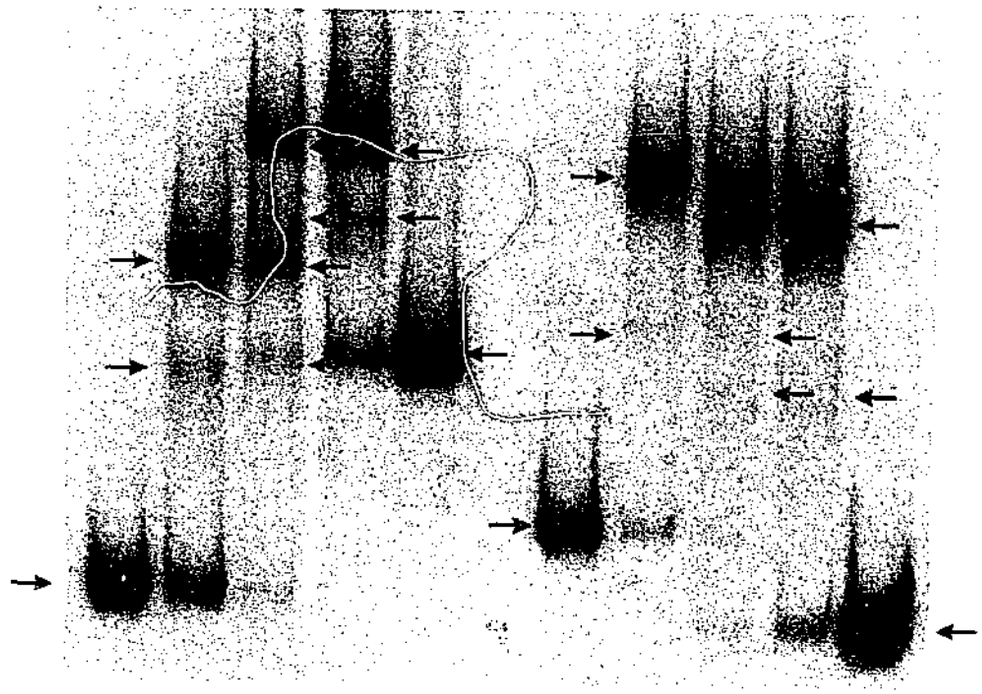
To determine if two DNA fragments were present in complex II, a gel mobility shift assay was performed using DNA fragments of different size. Two different sized left end fragments were mixed and incubated with $TnpX_{1-707}$, control reactions consisted of the same fragments incubated separately with $TnpX_{1-707}$. If two DNA fragments were present in either complex I or II, then hybrid complex I and II bands would be present when two DNA fragments of different sizes were reacted together. The results showed that both complex I and II consisted of a single DNA fragment (Fig 4.18), since no hybrid bands were observed. Similar results were obtained when right end fragments of different sizes were used (Fig 4.18).

j) Analysis of the Ability of TnpX Derivatives to Bind to DNA

To determine the DNA binding ability of the truncated TnpX derivatives, gel mobility shift assays were performed using the purified truncated TnpX proteins. The results (Fig 4.19) showed that $TnpX_{1-356}$ and $TnpX_{1-492}$ had a greatly reduced ability to bind to the left end of the transposon, although a small degree of binding was sometimes evident at very high protein concentrations (5-10 μg/reaction). In contrast, $TnpX_{1-597}$

Figure 4.18 Are multiple DNA fragments involved in formation of complexes I and II?

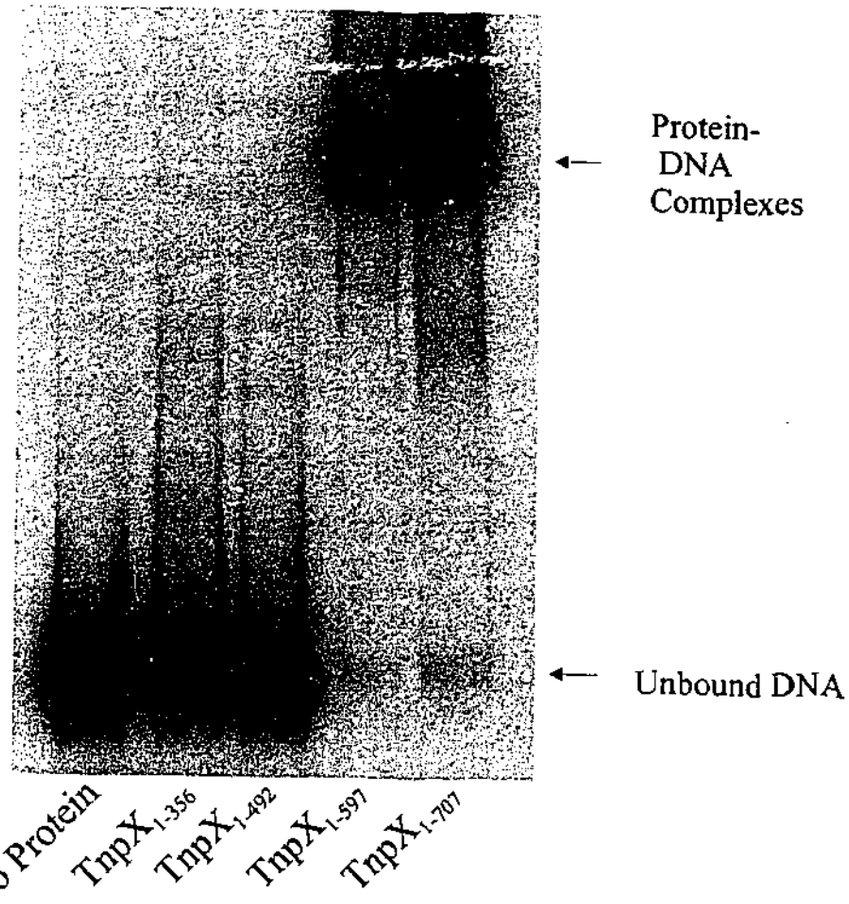
Gel mobility shift assays were performed using four different DNA fragments; two derived from the left end (the original 360 bp left end fragment, and a larger 563 bp fragment) and two from the right end (the original 343 bp right end fragment, and a larger 414 bp fragment). Reactions comprised of either a single DNA fragment with or without 1 μg of $TnpX_{1-707}$ or both the left end fragments or both the right end fragments, in the presence of $TnpX_{1-707}$. The red arrows indicate free DNA, the blue arrows indicate complexes I and II in lanes containing a single DNA fragment and the green arrows indicate bands present in lanes containing two different DNA species.



	$TnpX_{1-707}$	-	+	+	+	-	-	+	+	+	-
	360	+	+	+	-	-	-	-	-	-	-
DNA	563	-	-	+	+	+	-	-	-	-	-
Fragments (bp)	343	-	-	-	-	-	+	+	+	-	-
	414	-	-	-	-	-	-	-	+	+	+

Figure 4.19 Determination of DNA binding characteristics of the truncated TnpX derivatives

Gel mobility shift assays were performed using the left end DNA fragment incubated with the three truncated TnpX derivatives or $TnpX_{1-707}$, as described in Fig 4.12. Twelve pmol of each truncated TnpX derivative (equivalent to 1 μg of $TnpX_{1-707}$) was incubated with the labelled DNA. Unbound DNA and protein-DNA complexes are indicated.



ed with

e pmol

with the

showed a DNA shift very similar to that of the $TnpX_{1-707}$ protein (Fig 4.19). Quantitative gel mobility shift assays indicated that for each of the DNA binding fragments (left end, right end and circular intermediate), the binding affinities of $TnpX_{1-597}$ and $TnpX_{1-707}$ were not significantly different (Fig 4.20).

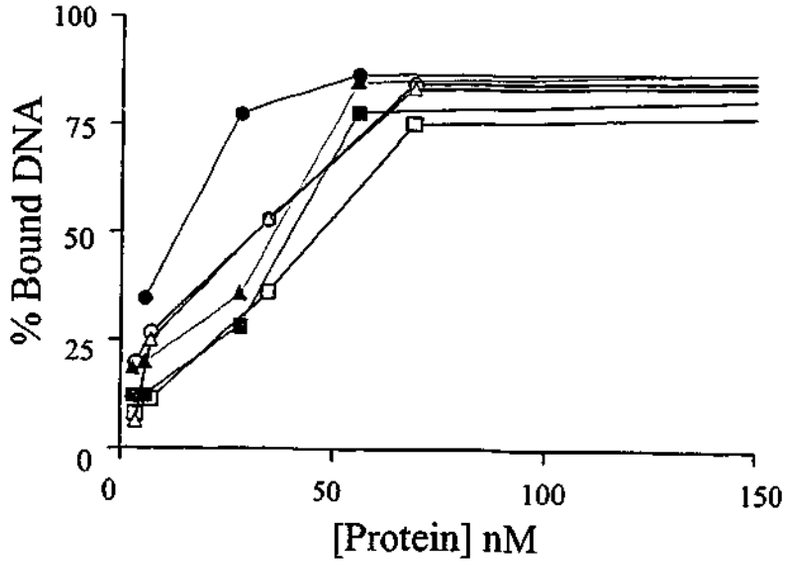
k) Determination of the Extent of the TnpX Binding Site

The extent of the TnpX DNA binding region at the left and right ends of the transposon, and at the joint of the circular intermediate, was determined using DNaseI protection assays. The DNA fragments used were identical to those used previously for gel mobility shift analysis but were labelled on one strand only, so that binding to the individual DNA strands could be examined. The DNA fragments were incubated with purified protein and then DNaseI was added. The resultant partially digested DNA fragments were separated by electrophoresis through a polyacrylamide sequencing gel. The banding patterns for reactions containing either $TnpX_{1-707}$ or $TnpX_{1-597}$ were compared to control reactions containing BSA.

The results showed that $TnpX_{1-707}$ and $TnpX_{1-597}$ both bound predominantly to the transposon side of the left (Fig 4.21) and right end (Fig 4.22) fragments, although there was some protection of the flanking DNA. Binding to the circular intermediate was found to be relatively symmetrical and to centre around the GA dinucleotide (Fig 4.23). Binding on both DNA strands was found to be overlapping and the protected transposon sequences from the left and right end correlated well with the protection of the same sequences on the circular intermediate (Fig 4.24). Only one area of protection was identified on each fragment tested, suggesting that there is only one binding site present on each fragment. The protected regions differed slightly between $TnpX_{1-707}$ and $TnpX_{1-597}$, with the region being more extensive for $TnpX_{1-707}$.

Figure 4.20 Comparison of $TnpX_{1\text{-}707}$ and $TnpX_{1\text{-}597}$ binding

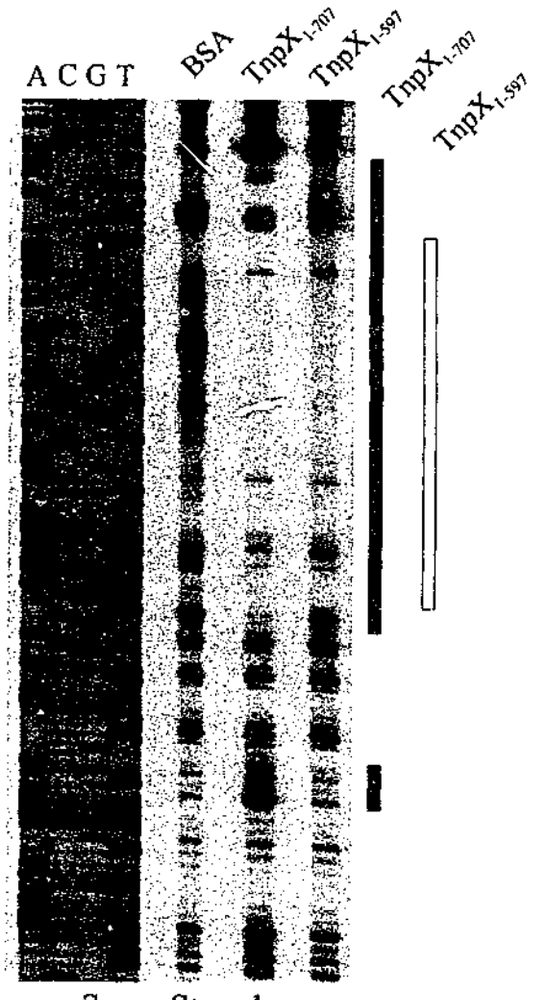
Quantitative gel mobility shift assays were performed as described in Fig 4.16. All three preferred DNA binding fragments were tested (left end, right end and circular intermediate (CI)) with either $TnpX_{1\text{-}707}$ (coloured traces, closed symbols) or $TnpX_{1\text{-}597}$ (black traces, open symbols).



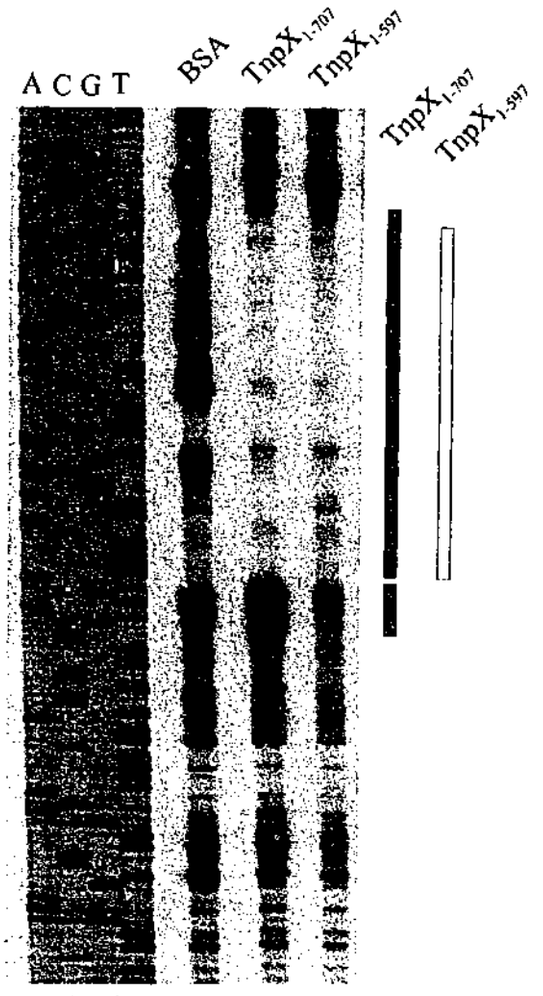
- Left end $TnpX_{1-707}$
- Left end $TnpX_{1-597}$
- Right end $TnpX_{1-707}$
- Right end $TnpX_{1-597}$
- CI $TnpX_{1-707}$
- CI $TnpX_{1-597}$

Figure 4.21 DNaseI footprinting of the sense and antisense strands of the left end DNA fragment

The DNA fragment was labelled with [^{32}P]-dATP on either the sense or antisense strand. The labelled DNA was incubated with BSA (negative control), $TnpX_{1-707}$ or $TnpX_{1-597}$ for 15 minutes at room temperature before partial digestion by treatment with DNaseI for 2 minutes at room temperature. The digested DNA was then loaded onto an acrylamide sequencing gel to separate the fragments. A sequencing ladder prepared from the same primers was used to identify regions protected from DNaseI digestion. Protection afforded by $TnpX_{1-707}$ is indicated by red rectangles, that by $TnpX_{1-597}$ using yellow rectangles. Areas showing an altered banding pattern are indicated by blue rectangles and regions of DNaseI hypersensitivity by green rectangles. The sequencing ladder is to the left of the digested reactions.



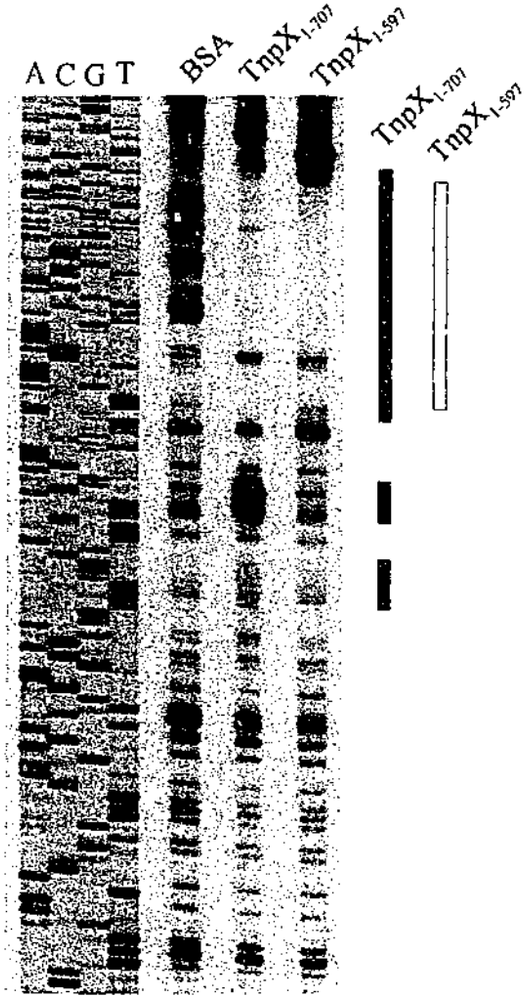
Sense Strand



Antisense Strand

Figure 4.22 DNaseI footprinting of the sense and antisense strands of the right end DNA fragment

The right end DNA fragment was labelled and analysed as described in Figure 4.21. Protection afforded by $TnpX_{1-707}$ is indicated by red rectangles, that by $TnpX_{1-597}$ using yellow rectangles. Areas showing an altered banding pattern are indicated by blue rectangles and regions of DNaseI hypersensitivity by green rectangles. The sequencing ladder is to the left of the digested reactions.

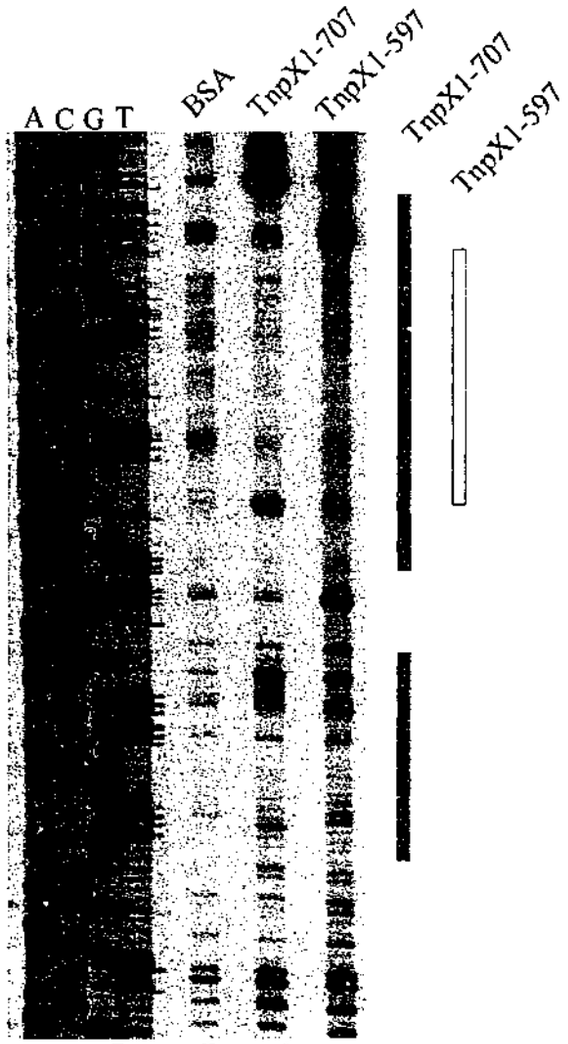


Sense Strand

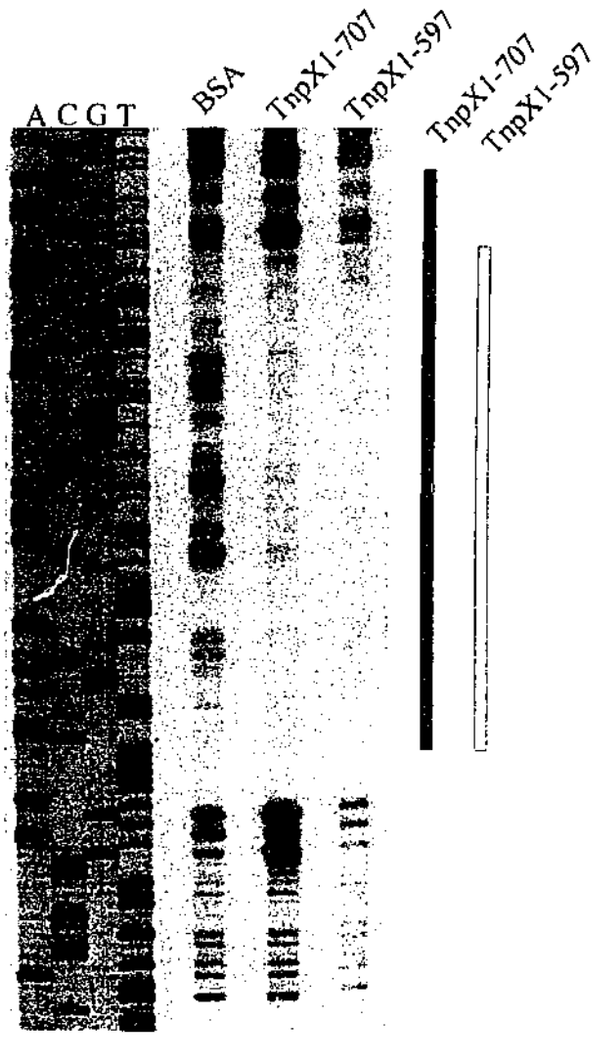
Antisense Strand

Figure 4.23 DNaseI footprinting of the sense and antisense strands of the circular intermediate DNA fragment

The circular intermediate DNA fragment was labelled and analysed as described in Figure 4.21 and 4.22. Protection afforded by $TnpX_{1-707}$ is indicated by red rectangles, that by $TnpX_{1-597}$ using yellow rectangles. Areas showing an altered banding pattern are indicated by blue rectangles and regions of DNaseI hypersensitivity by green rectangles. The sequencing ladder is to the left of the digested reactions.



Sense Strand



Anti-sense Strand

Figure 4.24 Extent of TnpX-mediated footprint

Only the protected sequences are shown. The double stranded protected sequence at the right end is shown at the top, flanking DNA is printed in black text, the GA dinucleotide in red and transposon-derived sequences in green text. Similarly, the middle sequence shows the circular intermediate, and the left end sequences are at the bottom.

CGTTTAATTTAGAAAATCCACAAAAAGTATCTTGACAAACTCGACCCCGAAGGATATGAGCTTACAACTGAAAC
TTAAATCTTTTAGGTGTTTTTCATAGAACTGTTTGAGCTGGGGCTTCCTATACTCGAATGTTGACTTT

Right End

AATTTAGAAAATCCACAAAAAGTATCTTGACAAACTCGACCCCGAGGGCTATACTTTAATAGGACAAAAAAATTAACTGTCCTAAATTGAAAAAAT
GCAAATTAAATCTTTTAGGTGTTTTTCATAGAACTGTTTGAGCTGGGGCTCCCGATATGAAATTATCCTGTTTTTTTAATTGACAGGATTTAACTTTTTTA

Circular Intermediate

GAGAATACAAAATCTCAGAAACTCAAGCTCCAGAGGGCTATACTTTAATAGGACAAAAAAATTAACTGTCCTAAATTGAAAAAAT
TGTTTTAGAGTCTTTGAGTTCGAGGTCTCCCGATATGAAATTATCCTGTTTTTTTAATTGACAGGATTTAACTTTTTTA

Left End

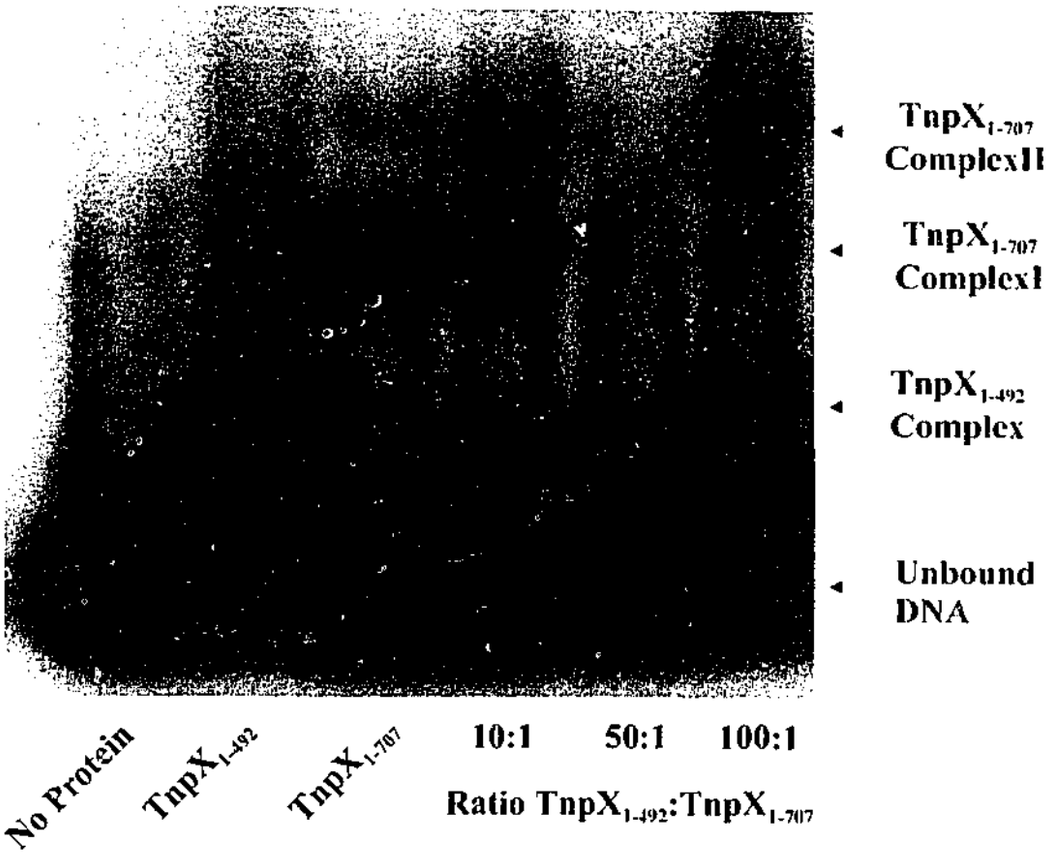
l) Competitive Gel Mobility Shift Assays of $TnpX_{1-707}$ and $TnpX_{1-492}$

As previously described, the *tnpX*$_{1-492}$ gene encoded a negative dominant phenotype in the transposition assay. To determine if this result was due to competitive inhibition of wild-type TnpX binding, a competition gel mobility shift assay was performed. In this experiment the left end fragment was radiolabelled for increased sensitivity. The purified proteins were mixed in a normal gel mobility shift reaction prior to the addition of DNA. Due to the poor yield of $TnpX_{1-492}$ the concentration of this protein preparation was low. As a consequence, only a small amount of $TnpX_{1-707}$ was added to the reactions, just enough to barely illicit a DNA shift, as it was anticipated that an excess of $TnpX_{1-492}$ would be required to show an inhibitory effect. These conditions would also serve to mimic the transposition assay, where the levels of the plasmid-encoded truncated TnpX derivative would have exceeded the levels of wild-type TnpX produced from the chromosomal copy of the gene.

The two proteins, $TnpX_{1-492}$ and $TnpX_{1-707}$, were mixed in molar ratios with $TnpX_{1-492}$ in excess by 10 fold, 50 fold and 100 fold. Control reactions containing $TnpX_{1-492}$ and $TnpX_{1-707}$ alone were included in addition to the no protein control. The results suggested that $TnpX_{1-492}$ was capable of a low level of DNA binding. This binding is very poor as it was only observed at very high concentrations of $TnpX_{1-492}$ (1.4 μg) and even at these concentrations the proportion of DNA bound was much less than 50% (Fig 4.25). The results illustrate that the presence of excess $TnpX_{1-492}$ abolished DNA binding by $TnpX_{1-707}$ (Fig 4.25), however, whether this was due to displacement of $TnpX_{1-707}$ from the DNA by competitive binding or whether this is due to the formation of heterodimeric proteins unable to bind DNA productively is unclear.

Figure 4.25 Competition gel mobility shift assay between $TnpX_{1-707}$ and $TnpX_{1-492}$

A gel mobility shift assay was performed essentially as described in Figure 4.16. The first reaction (lane 1) contains no protein, the second reaction (lane 2) contains 1.4 μg of $TnpX_{1-492}$, reaction three (lane 3) contains $TnpX_{1-707}$ (0.02 μg). The next three reactions (lanes 4, 5 and 6) contain a mixture of $TnpX_{1-707}$ (0.02 μg) and increasing amounts of $TnpX_{1-492}$ in the molar ratios indicated to a maximum of 1.4 μg of $TnpX_{1-492}$:0.02 μg of $TnpX_{1-707}$.


TnpX1-707 ComplexII
TnpX1-707 ComplexI
TnpX1-492 Complex
Unbound DNA
No Protein
TnpX1-492
TnpX1-707
10:1
50:1
100:1
Ratio TnpX1-492:TnpX1-707

m) Do the Truncated TnpX Derivatives Catalyze Excision *In Vitro*?

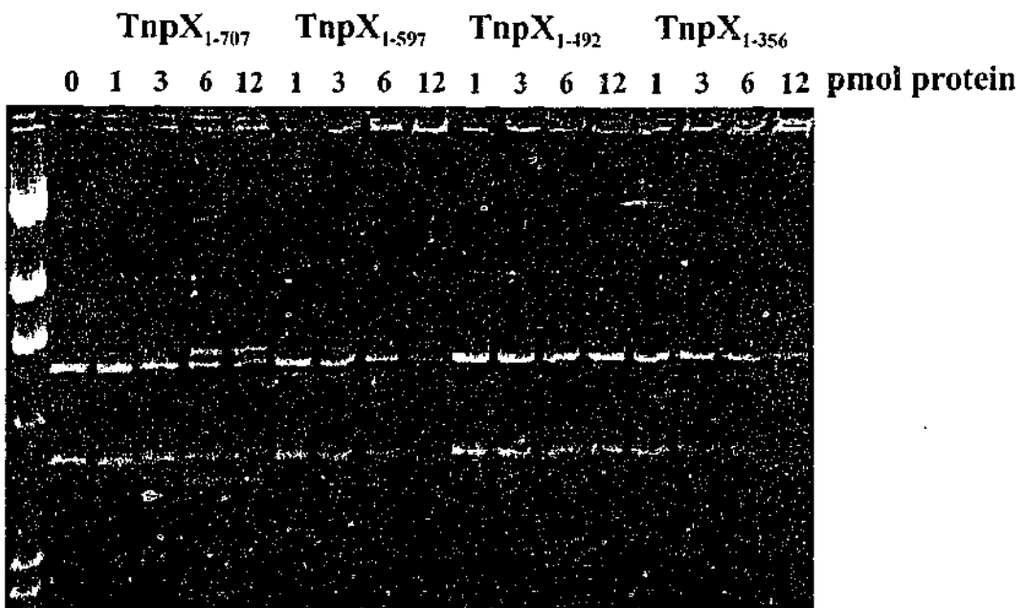
The purified full length TnpX and truncated TnpX proteins were tested for their ability to excise Tn*4451* using the *in vitro* excision assay (Chapter 3). Excision reactions were incubated using conditions previously described, heat inactivated and subjected to restriction enzyme digestion and electrophoresis (Fig 4.26). As expected, full length TnpX and $TnpX_{1-597}$ were capable of catalyzing excision, although a significant reduction in activity between $TnpX_{1-707}$ and $TnpX_{1-597}$ was evident. Additionally, $TnpX_{1-356}$ and $TnpX_{1-492}$ did not catalyze excision (Fig 4.26). These data were consistent with *in vivo* studies and indicated that $TnpX_{1-597}$, like $TnpX_{1-707}$, was capable of excision in the absence of other cellular proteins, at a lower level.

Discussion

Comparative analysis of the TnpX amino acid sequence revealed several regions of interest. The first of these regions contained a resolvase catalytic domain, which was previously demonstrated to be essential for TnpX function since site-directed mutants of the conserved residues, S15 and R89, are unable to catalyse excision (Crellin and Rood, 1997) or transposition (Lyras *et al.*, 1998). Based upon this evidence it was concluded that TnpX acts using a resolvase-like mechanism, as expected for a protein that has a significant amino acid sequence identity to resolvase proteins within the first 100 amino acids. In addition to the N-terminal catalytic domain, members of the small resolvase family also contain dimerization and DNA binding domains. The DNA binding domain often consists of a helix-turn-helix motif (Smith and Thorpe, 2002). However, this kind of DNA binding domain has not been identified in any of the large resolvase proteins, including TnpX (Smith and Thorpe, 2002). TnpX diverges structurally from the small resolvase proteins after approximately 100 aa. After this point computer analysis revealed a putative zinc binding domain (residues 324-360), a leucine rich region

Figure 4.26 *In vitro* excision assays with truncated TnpX derivatives

The excision reactions were performed using varying amounts of protein as indicated. Reactions were performed as described in Chapters 2 and 3 and subjected to restriction enzyme analysis. PCR amplification was performed on a sample of each reaction to detect the presence of circular intermediate DNA (bottom panel).


TnpX1-707
TnpX1-597
TnpX1-492
TnpX1-356
0 1 3 6 12 1 3 6 12 1 3 6 12 1 3 6 12 pmol protein
PCR

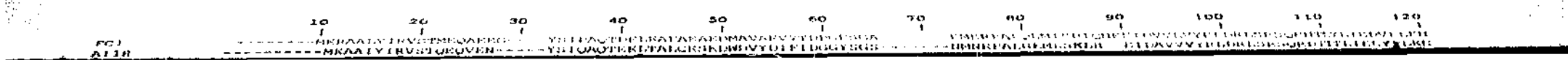
(residues 373-449), a charged domain (aa 536-577) and a cysteine rich domain (last 110 aa).

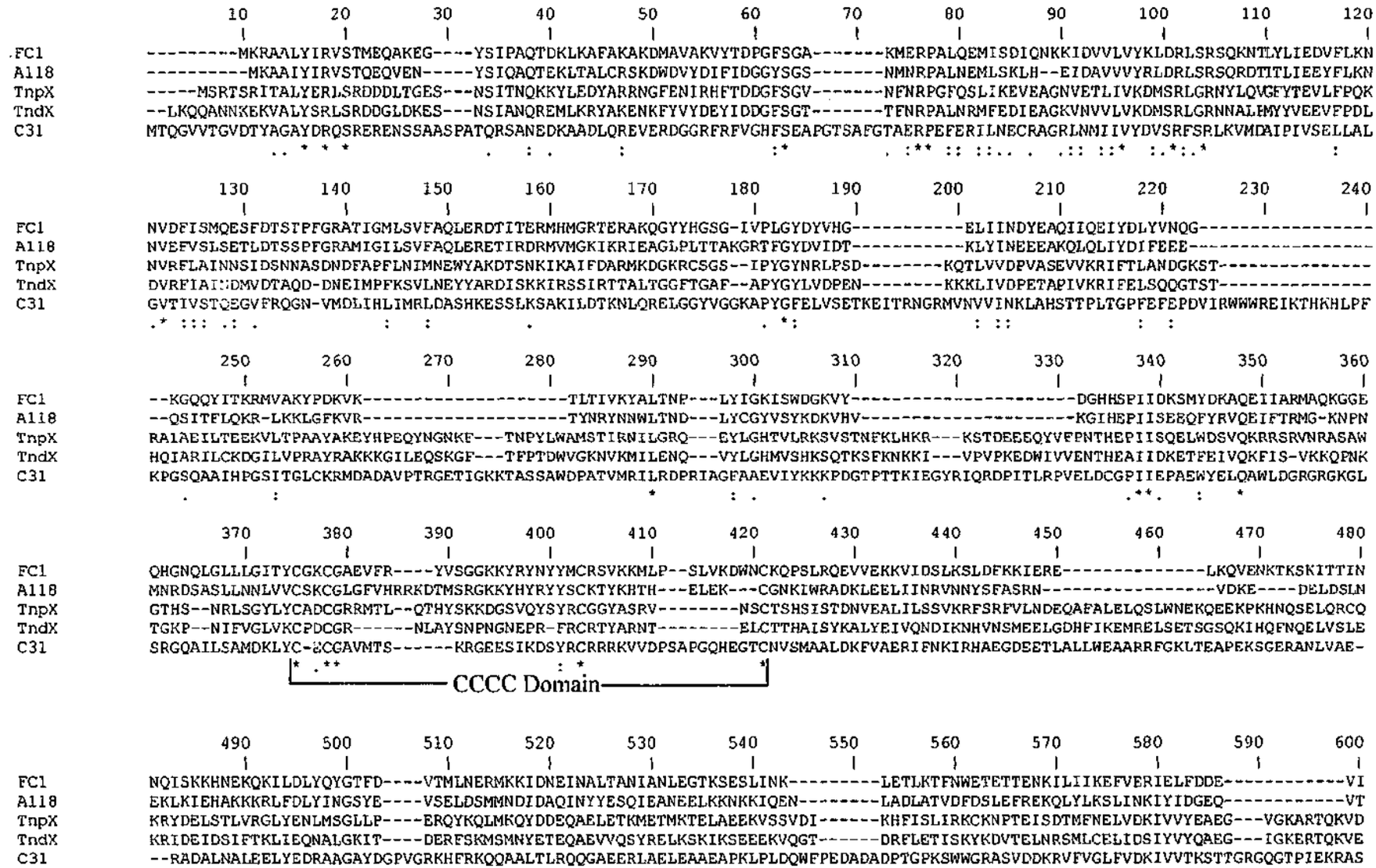
The putative zinc binding domain is conserved between other members of the large resolvase family and consists of four conserved cysteine residues (Fig 4.27). In TnpX this region shows a predicted secondary structure consistent with zinc binding proteins but does not contain a canonical C_2H_2 zinc binding domain, instead having a CCCC domain found in other atypical zinc binding proteins (Laity *et al.*, 2001). Most zinc binding proteins such as TFIIIA, a transcription factor from *Xenopus*, bind to their DNA targets through the co-ordination of zinc (Laity *et al.*, 2001). However, recently a number of zinc binding proteins have been identified that dimerize through interaction between two zinc binding domains (Laity *et al.*, 2001). It is therefore possible that the putative zinc binding domain of TnpX is responsible for mediating TnpX dimerization.

Another putative structure that may be capable of facilitating dimerization is the leucine rich domain (aa 373-449). The spacing of the leucine residues in this region did not follow the heptad repeat structure characteristic of leucine zipper proteins (Haren *et al.*, 2000) and did not have similarity to the leucine rich repeat family (Buchanan and Gay, 1996). Both of these motifs have been implicated as dimerization domains (Buchanan and Gay, 1996; Haren *et al.*, 2000) that facilitate DNA binding (Haren *et al.*, 2000). The loss of this region, as in the $TnpX_{1-356}$ protein, resulted in a protein still capable of dimerization although the dimer that was produced showed reduced stability compared to that of $TnpX_{1-707}$ dimer. This result suggests that the leucine rich region is not responsible for dimerization. $TnpX_{1-356}$ still contains almost all of the putative zinc binding domain, with the exception of the last cysteine residue, which may explain the significant reduction in dimer stability exhibited by this protein.

Figure 4.27 Alignment of large resolvase proteins

A clustal W (http://npsa-pbil.ibcp.fr/) alignment of five large resolvase proteins is shown. The proteins are the integrase proteins from the phages ϕFC1, ϕA118 and ϕC31 and the clostridial large resolvase proteins TndX from Tn*5397* and TnpX from Tn*4451*. Their accession numbers are as follows: ϕFC1 (AAD26564), ϕA118 (CAB53817), TnpX (AAB51419.1), TndX(AAF35174.1) and ϕC31(NP_047974.1). Residues that are completely conserved are indicated in red and by an asterisk below the alignment, residues showing a high degree of similarity are in green with two dots below the alignment and residues with a lower level of similarity are indicated in blue with one dot below the alignment. The CCCC domain is also indicated.




FC1
A118
TnpX
TndX
C31
CCCC Domain

In addition to a dimerization site there is likely to be a multimerization site for synapse formation, unless TnpX binds to DNA targets as a monomer. The latter is unlikely since TnpX exists as a dimer in solution and a minimum of two molecules of TnpX would be required at each cleavage site for two DNA strands to be broken via covalent attachment to the catalytic serine residue, then rotated and re-joined in a characteristic resolvase reaction (Hallet and Sherratt, 1997). It is possible that if the leucine rich region is not required for dimer formation it might be involved in the protein-protein interactions required for synapse formation. Alternatively, the leucine rich region may not have any functional significance.

Between residues 536 and 555 there was a region with a large proportion of acidic residues (9 of 20 amino acids are aspartic or glutamic acid). Adjacent to this region, from amino acids 558 to 577, there was a region consisting of 11 basic residues. This oppositely charged domain (amino acids 536-577) also corresponded to a predicted coiled-coil structure (Fig 4.1). Analysis of the DNA binding ability of the truncated TnpX derivatives showed that both $TnpX_{1-707}$ and $TnpX_{1-597}$ were capable of binding to DNA. However, neither $TnpX_{1-356}$ nor $TnpX_{1-492}$ were capable of DNA binding. The results suggest that there is a DNA binding domain between amino acids 493 and 597. Given this information, it seems feasible that the charged, putative coiled-coil domain between amino acids 536 and 577 constitutes a DNA binding domain.

The last 110 amino acids of TnpX were not essential for activity since the truncated derivative $TnpX_{1-597}$ was functional for excision, insertion, transposition, dimerization and DNA binding. However, $TnpX_{1-707}$ and $TnpX_{1-597}$ did not have the same activity profile. The *in vivo* and *in vitro* excision activity for $TnpX_{1-597}$ is reduced compared to that of the full length protein. These results suggest that the last 110 amino acids may contribute in some way to the excision reaction. In addition, DNaseI protection assays

indicated that the area of protection mediated by $TnpX_{1-597}$ was not as extensive as that of $TnpX_{1-707}$. This result was not unexpected as the area of protection will to some extent be dependent on the size of the DNA binding protein affording protection. As $TnpX_{1-597}$ is smaller than $TnpX_{1-707}$, DNaseI may have access to DNA residues that are protected by the extra bulk of $TnpX_{1-707}$. Whether this additional protection is due to the last 110 amino acids contacting the DNA and constituting a second DNA binding domain is unknown. If this 110 amino acid region also binds to DNA it may serve to stabilize the synaptic complex formed during transposon excision, which may explain the decrease in excision observed with $TnpX_{1-597}$. Regardless, excision was possible without the last 110 amino acids of the protein.

Studies conducted on the DNA binding characteristics of other large resolvase proteins have shown that the minimum DNA fragment length still resulting in recombination is quite small. Deletion analysis has indicated that the minimum functional binding site for Int from phage TP901-1 is approximately 40-50 bp (Breüner *et al.*, 2001; Stoll *et al.*, 2002). Similarly, DNaseI protection studies conducted on the binding sites of Int from phage C31 revealed protected areas of 50-60 bp (Thorpe *et al.*, 2000). In comparison, the typical small resolvase binding site, *res*, is considerably larger (Sherratt, 1989). The *res* site consists of three subsites, binding six resolvase proteins in total, a dimer at each subsite (Craigie, 1995). The *res* site spans approximately 120 bp, considerably larger than the minimum sites shown to be active for large resolvase activity (Breüner *et al.*, 2001; Sherratt, 1989; Stoll *et al.*, 2002; Thorpe *et al.*, 2000). DNaseI protection assays indicated that the binding of $TnpX_{1-707}$ to its DNA binding site resulted in approximately 50 bp of DNA being protected, similar to the Int proteins of the phages TP901-1 and C31 (Breüner *et al.*, 2001; Stoll *et al.*, 2002; Thorpe *et al.*, 2000). No other regions protected from DNaseI digestion were found on the DNA fragments tested, which ranged in size from 564 bp (larger left end transposon fragment (GA still

approximately in the middle), data not shown) to 343 bp (right end transposon fragment).

Subsites II and III of the small resolvase *res* site are postulated to be involved in orientating and facilitating synapse formation and it is thought that the resolvase dimers bound to these sites are responsible for activating the catalytic resolvase molecules bound to subsite I (Arnold *et al.*, 1999). This complicated arrangement, involving multiple sites and proteins, serves to restrict resolvase activity to excision of DNA segments from a single, supercoiled DNA molecule containing *res* sites that are directly repeated (Nash, 1996; Stark and Boocock, 1995). The large resolvase proteins are not restricted in this manner as they are capable of catalyzing recombination between two sites on separate DNA molecules (Breüner *et al.*, 2001; Stoll *et al.*, 2002; Thorpe and Smith, 1998; Wang and Mullany, 2000). DNaseI protection studies suggested that there was only a single DNA binding site for TnpX, as for Int from ϕC31 and TP901-1 (Breüner *et al.*, 2001; Nash, 1996; Stoll *et al.*, 2002; Thorpe and Smith, 1998).

Mutants of the Tn*3* small resolvase that can catalyze excision in the absence of subsites II and III have been isolated. They result from two amino acid substitutions (Arnold *et al.*, 1999) and are capable of catalyzing intramolecular recombination between two subsite I sequences located on different DNA molecules. These data suggest that the restrictions observed for resolvase reactions are imposed by the need for the accessory binding sites to activate the catalytic dimer bound to subsite I, in turn leading to the requirement for a defined and restrictive synaptic complex (Arnold *et al.*, 1999).

Reactions catalyzed by the large resolvase proteins may bypass the need for accessory binding sites by providing the internal scaffolding within the larger protein molecules. This structural difference may explain why $TnpX_{1\text{-}507}$, although still able to catalyse

excision, shows reduced excision activity. Perhaps the last 110 amino acids are required for more efficient synaptic complex formation either via protein-protein interactions or further protein-DNA contact.

The converse appears to be the case for the insertion reaction, the reaction that differentiates large resolvase proteins from small resolvases. Insertion or integration has been demonstrated both *in vivo* and *in vitro* for Int from ϕC31, in the absence of any other phage or cellular proteins (Thorpe and Smith, 1998). Similarly, it has been demonstrated that TnpX requires no other transposon-encoded proteins for insertion (D. Lyras and J. I. Rood, unpublished results). $TnpX_{1-597}$ is able to catalyse insertion of the joint of the circular intermediate into the host chromosome at a higher frequency than the wild-type protein. Therefore, either the last 110 amino acids inhibit the insertion reaction, in direct contrast to the situation for excision, or the level of $TnpX_{1-597}$ expression or the protein stability (less prone to degradation) is higher than that of $TnpX_{1-707}$ *in vivo*.

A significant difference between the excision reaction and the insertion reaction was shown by the differential binding affinity evident between TnpX and the reaction substrates. The DNA substrates involved in the insertion reaction are the circular intermediate joint, to which TnpX binds with a high affinity, and the potential insertion sites, to which TnpX binds very poorly. In contrast, TnpX binds to the substrates of the excision reaction, the left and right ends of the transposon, with equally high affinity. It has been shown that $TnpX_{1-597}$ is capable of a high level of insertion activity while excision activity is decreased. These results provide evidence that there are fundamental differences between the insertion and excision reactions. The insertion reaction may require accessory cellular proteins or there may be a substantial difference in synaptic complex formation or in the conformation of the complex.

Knowledge of the composition of the two complexes observed by gel mobility shift analysis would contribute significantly to an understanding of the excision and insertion reactions. The more mobile complex I may have comprised a single DNA fragment bound to a single molecule of TnpX, which would explain why this complex was most obvious at low protein concentrations. The less mobile complex II may have comprised a single DNA fragment to which a TnpX dimer is bound. Alternatively, both complex I and II may consist of a TnpX dimer bound to a single DNA molecule but differences in the conformation of the two complexes, ie DNA bending, may result in complexes with differing mobility. Binding of TnpX to the circular intermediate is more complicated as higher order complexes appeared to be present at higher protein concentrations. These more retarded bands may have resulted from binding of a second dimer to the initial dimer-DNA complex, forming a tetramer, or may reflect further complexity of the interactions.

The excision reaction may be facilitated by binding of two TnpX dimers to the ends of the integrated transposon, which are then brought together for synaptic complex formation. This process would lead to the juxtaposition of the ends of the transposon, bringing the two TnpX dimers into close proximity, allowing catalysis and strand exchange to occur. By contrast, the insertion reaction may proceed by binding of a TnpX dimer to the joint of the circular intermediate, followed by binding by a second TnpX dimer. This putative tetramer-DNA complex may then have a higher affinity for a potential insertion site, may bind other cellular or transposon-encoded proteins, or may be dependent on the local topology of the DNA, which is not present in short, linear DNA fragments used in the binding assays. The resolvase-invertase proteins, which are capable of excision or inversion depending on the orientation of the binding sites, require an accessory protein bound to the recognition site (Alonso *et al.*, 1996;

Janière *et al.*, 1996). The TnpX-circular intermediate joint complex, having recruited the target site, would then be capable of undergoing recombination. If this hypothesis is correct it would seem that there is a significant difference in synaptic complex formation between the excision reaction and the insertion reaction.

Chapter 5

Functional Analysis of TnpX

Introduction

A significant amount of structural data is available for the small resolvase derived from the Tn*3*-like transposon, γδ, including crystal structures (Rice and Steitz, 1994; Sanderson *et al.*, 1990; Yang and Steitz, 1995) and an NMR solution structure (Pan *et al.*, 2001). Most of the data describes the catalytic domain since in the absence of DNA the DNA binding domain is disordered in crystal structures of the intact protein (Rice and Steitz, 1994). However, the crystal structure of a γδ resolvase dimer complexed to a 34 bp DNA fragment has been solved (Yang and Steitz, 1995). The small resolvase structure consists of three regions, the catalytic domain (residues 1-105 in γδ resolvase), a dimerization helix (the E helix) and a DNA recognition region, a helix-turn-helix structure (Sanderson *et al.*, 1990; Yang and Steitz, 1995). The structure of the catalytic domain consists of three α helices bordering four parallel β strands and one antiparallel β strand (Sanderson *et al.*, 1990). The catalytic serine residue is positioned on a mobile loop (Pan *et al.*, 2001). The data obtained from studies of resolvase crystal structure has given insights into the structure of the catalytic domain and has indicated why particular residues, such as arginine-8, serine-10 and various other well conserved resolvase residues located towards the N-terminus, are important for function. Mutation of these residues has produced mutants affected in catalysis but not DNA binding. The location of these residues within the crystal (Sanderson *et al.*, 1990) or co-crystal structure (γδ complexed with DNA) (Yang and Steitz, 1995) has indicated that they form part of a catalytic pocket.

The first approximately 100 amino acids that constitute the catalytic domain of the resolvase/invertase family of proteins, including the large resolvase subgroup, is highly conserved (Smith and Thorpe, 2002; Yang and Steitz, 1995). To identify amino acids of TnpX that were important for structure or function random mutagenesis of the *tnpX* gene was performed. Mutants of TnpX were subsequently characterized using the *in vivo* assays described previously (Chapter 4) and selected mutants were overexpressed and purified for further analysis.

Results

a) Development of a Strategy for the Detection of TnpX Mutants

To detect non-functional *tnpX* mutants, it was necessary to design an efficient and rapid screening procedure. The method that was used was based on the excision assay, where the ability of the potentially mutated TnpX proteins to excise a stable copy of Tn*4451 in trans* was assessed. Two new plasmids were constructed for this assay. The first plasmid, pJIR1872 (Fig 5.1), contained the *Bam*HI/*Asp*718 *tnpX* fragment from pJIR639 cloned into the *Bam*HI/*Asp*718 sites of the vector pSU39. This vector is a medium copy number vector that contains a kanamycin resistance gene and the *lacZα* gene downstream of the *lac* promoter (Bartolomé *et al.*, 1991). The *tnpX* insert in pJIR1872 contains the wild-type *tnpX* and *tnpV* genes and a 3' truncation of the *catP* gene (Fig 5.1). In this plasmid the *tnpX* gene is located downstream of the *lac* promoter, although there is approximately 600 bp between the promoter and the *tnpX* start codon.

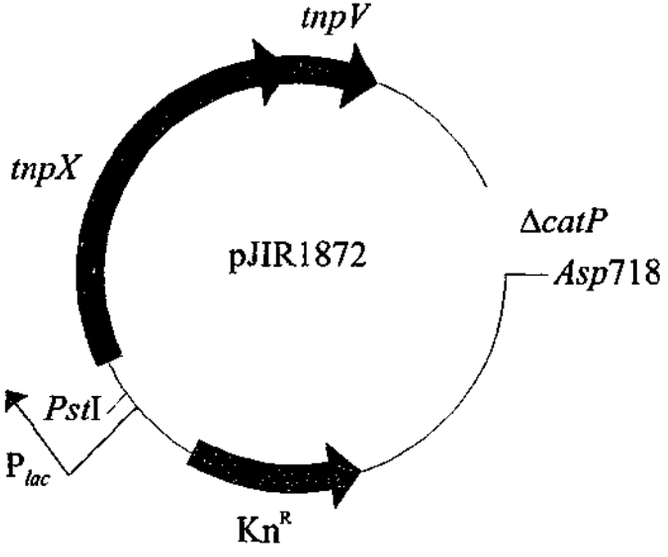
In vivo the *tnpX*$^+$ plasmid was found to confer a very high level of excision activity, approaching 100% (data not shown). To increase this level further the target plasmid, pJIR1880 (Fig 5.1B), contained a defective copy of Tn*4451* bounded not by GA

Figure 5.1 Plasmids used for random mutagenesis and subsequent screening

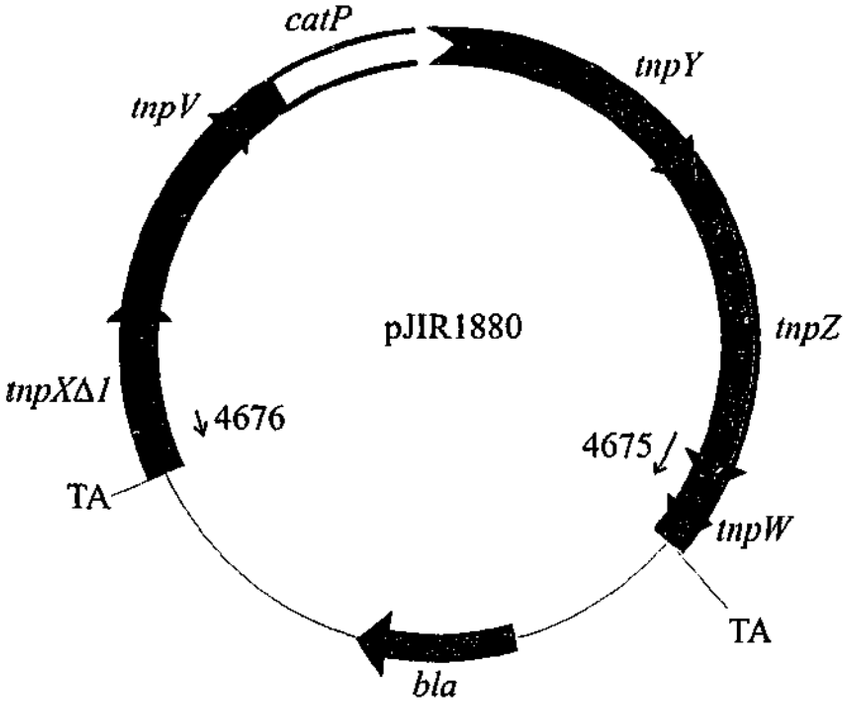
A Plasmid pJIR1872 encodes the wild-type *tnpX* gene (blue arrow), an intact *tnpV* gene (green arrow), a truncated *catP* gene (yellow block) and a kanamycin resistance gene (pink arrow). The *lac* promoter is indicated by the bent arrow. The *Pst*I and *Asp*718 restriction sites were used to subclone the *tnpX* fragment from pJIR639 (Bannam et al., 1995) into pSU39.

B Plasmid pJIR1880 contains the transposon Tn*4451tnpXΔ1*. This plasmid encodes *tnpW*, *tnpZ* and *tnpY* (red arrows), *catP* (yellow arrow) and *tnpV* (green arrow). The *tnpXΔ1* gene (blue) has an internal 185 bp deletion located towards the 5' end of the gene; the resultant protein is non-functional. This plasmid also encodes ampicillin resistance (pink arrow). The thin black arrows indicate the approximate positions of the oligonucleotide primers (4675 and 4676) used for the circular intermediate PCR. As shown, the ends of the transposon boundary (black block) contain TA residues not GA residues.

A


tnpV
tnpX
pJIR1872
ΔcatP
Asp718
PstI
P lac
Kn R

B


catP
tnpY
tnpV
pJIR1880
tnpZ
tnpXΔ1
4676
4675
TA
tnpW
TA
bla

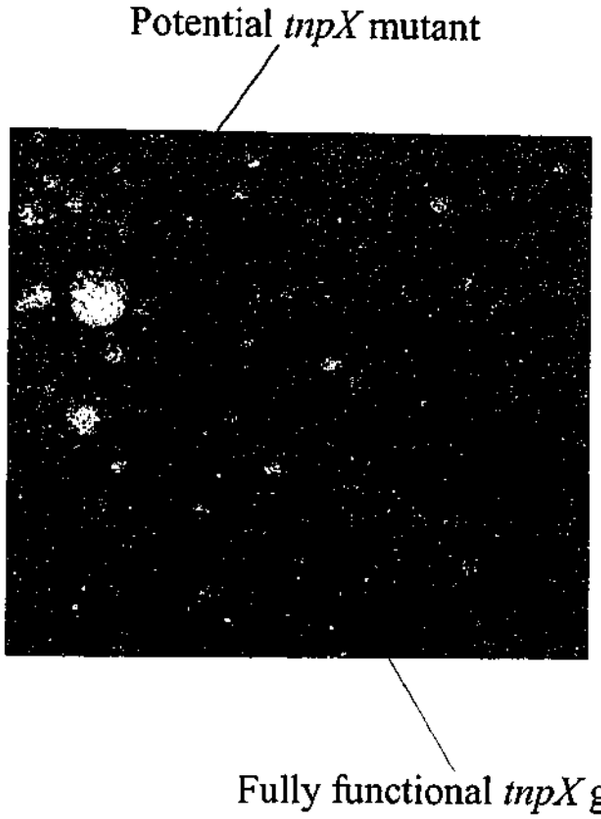
dinucleotides but by TA dinucleotides, which were shown previously to give a higher excision frequency (Crellin and Rood, 1997). Additionally, the defective Tn*4451tnpXΔ1*-TA derivative was cloned into the low copy vector pWSK29 to form pJIR1880 (Fig 5.1B). When *E. coli* strain DH12S, containing pJIR1872 and pJIR1880, was grown on medium containing high levels of chloramphenicol (40 μg/mL) the bacterial colonies observed after 18 hrs were extremely small compared to the negative control colonies containing pJIR1880 and pSU39 (Fig 5.2). The colony size difference between those cells containing a functional *tnpX* gene and a non-functional *tnpX* gene formed the basis of the screening procedure (Fig 5.2). The colony size difference was due to the loss of the *catP* gene, and consequently chloramphenicol resistance, by TnpX-mediated excision of the Tn*4451* transposon on pJIR1880.

b) Random Mutagenesis and Mutant Screening

The plasmid pJIR1872 was mutagenized by passage through the mutator cell line XL1-Red (Stratagene). The XL1-Red strain contains mutations within three DNA mismatch repair genes, *mutS*, *mutD* and *mutT*, consequently the mutation rate of this strain is 5000-fold greater than the parent strain (Instruction manual, XL1-Red competent cells, Stratagene). The passaged DNA was subsequently used to electroporate DH12S (pJIR1880). The cells were subcultured onto medium containing kanamycin (20 μg/mL) and chloramphenicol (40 μg/mL) and incubated overnight. The largest colonies were selected (i.e. those colonies most likely to contain a defective *tnpX* gene) and used to extract plasmid DNA, which was subsequently used as the PCR template to detect the presence of the circular intermediate. PCR amplification using oligonucleotide primers 4676 and 4675 could potentially yield two different DNA fragments (Fig 5.1). A larger fragment of 6.6 kb would be obtained if the reaction had resulted in polymerization around the plasmid, comprising predominantly pWSK29 DNA and lacking most of the

Figure 5.2 Differentiation of *tnpX*$^+$/*tnpX* genes by examination of the colony size

Cells containing potentially mutated pJIR1872 DNA and the indicator plasmid pJIR1880 were grown on medium supplemented with kanamycin (20 μg/mL) and chloramphenicol (40 μg/mL). Large (*tnpX* mutant) and small (*tnpX*$^+$) colonies are indicated.


Potential *tnpX* mutant
Fully functional *tnpX* gene

defective Tn*4451* transposon DNA. A 0.9 kb DNA fragment would be generated by amplification across the joint of the circular intermediate, if this molecule was present. The 0.9 kb DNA fragment would be preferentially amplified as it is much smaller. Consequently, even if only a small amount of circular intermediate DNA was present it would be detected, indicated by the presence of both PCR products.

Four aliquots of pJIR1872 DNA were introduced separately into the XL1-Red strain for mutagenesis and the plasmid DNA reisolated. The reisolated DNA was stored in small aliquots to be subsequently used to transform DH12S(pJIR1880) to kanamycin resistance. Approximately 20 such transformations were conducted using aliquoted DNA that had been passaged through XL1-Red. Bacteria from the subsequent transformations were selected on medium containing kanamycin and a high concentration of chloramphenicol, leading to the appearance of approximately 40 large bacterial colonies from each transformation. The DNA extracted from the large colonies was tested by circular intermediate PCR. If very little or none of the 0.9 kb fragment was present then the potentially mutated pJIR1872 plasmid was separated from pJIR1880 by transformation into DH5α and patching onto kanamycin plates, screening for ampicillin sensitivity. The purified DNA was subjected to restriction endonuclease analysis to ensure that no large deletions or insertions had occurred. Determination of the colony size phenotype and the circular intermediate PCR was repeated and then the *tnpX* gene of each potential mutant was sequenced. Twenty-two mutated *tnpX* genes containing single amino acid substitutions were isolated and retained for further analysis (Table 5.1).

Table 5.1 List of random TnpX mutants

Plasmid Number	Nucleotide Change	Amino Acid Change
pJIR2085	C44T	S15L
pJIR2063	C77T	S26F
pJIR2094	C100T	L34F
pJIR2067	G167A	G56D
pJIR2100	G182T	R61L
pJIR1922	T196C	S66P
pJIR1947	A248G	D83G
pJIR1923	G252A	M84I
pJIR2011	G254A	S85N
pJIR2087	G262A	G88R
pJIR2049	G266A	R89Q
pJIR1948	A392G	N131S
pJIR1945	C411T	A138V
pJIR1921	G412A	A138T
pJIR2040	G418A	D140N
pJIR1919	A469G	K157E
pJIR1946	G481C	G161S
pJIR2041	G751A	E251K
pJIR1920	A778G	K260E
pJIR1984	T781C	S261P
pJIR1918	C854T	T285I
pJIR2068	T970C	C324R

c) *In Vivo* Testing of Random *tnpX* Mutants

The 22 *tnpX* mutants were introduced into suitable strains so that stability, insertion and transposition assays could be performed. The strains used for these assays were slightly different than those described previously (Chapter 4) to allow for different plasmid compatibilities and antibiotic resistance phenotypes.

Stability assays were performed using the plasmid pJIR683. This plasmid is almost identical to pJIR773 except that it is a pBLUESCRIPT based plasmid encoding ampicillin resistance. The plasmid pJIR683 contains Tn*4451tnpXΔ1* and is stable (it does not spontaneously lose chloramphenicol resistance) in the absence of an additional, functional *tnpX* gene. Both pJIR683 and each pJIR1872-derived mutant were introduced into DH5α to perform the stability assays. Stability was determined by patching onto medium containing ampicillin or chloramphenicol and determining the percentage of ampicillin resistant colonies that were chloramphenicol resistant.

In the presence of the pSU39 negative control, pJIR683 was found to be 100% stable as expected, while the positive control, pJIR1872, exhibited a very low stability of 0.7%. The mutants generally conferred high stability (Table 5.2, Fig 5.3), which was expected since the detection process was based on an inability to excise Tn*4451tnpXΔ1*. Several of the mutants, such as $TnpX_{M84I}$, $TnpX_{A138V}$, $TnpX_{K157E}$ and $TnpX_{T285I}$, did catalyze a significant amount of excision. These mutants were among the first isolated, reflecting the fact that the selection procedure became more stringent as the experiments proceeded.

The insertion assay was also performed using an alternative insertion vector. Previously, a kanamycin resistant temperature sensitive plasmid was used, however, this

Table 5.2 Random TnpX mutant *in vivo* assay data

Plasmid/ Mutation	% Stability [a]	Insertion Frequency [b]	Transposition Frequency [c]
pJIR1872	0.7 ± 1.1	0.5 ± 0.5	$(1.3 \pm 1.8) \times 10^{-4}$
pSU39	100	$(9.5 \pm 7.0) \times 10^{-5}$	$(1.8 \pm 3.6) \times 10^{-5}$
S15L	100	$(4.2 \pm 8.9) \times 10^{-5}$	$(7.3 \pm 9.3) \times 10^{-7}$
S26F	100	$(3.7 \pm 2.8) \times 10^{-5}$	$(2.3 \pm 2.6) \times 10^{-7}$
L34F	99.7 ± 0.5	$(4.4 \pm 3.0) \times 10^{-5}$	$(1.5 \pm 2.0) \times 10^{-6}$
G56D	100	$(4.1 \pm 1.2) \times 10^{-5}$	$(6.9 \pm 10.5) \times 10^{-7}$
R61L	100	$(4.1 \pm 2.4) \times 10^{-5}$	$(9.6 \pm 5.0) \times 10^{-8}$
S66P	98.3 ± 2.2	0.6 ± 0.8	$(3.0 \pm 1.0) \times 10^{-7}$
D83G	100	$(4.2 \pm 2.9) \times 10^{-5}$	$(1.5 \pm 0.6) \times 10^{-8}$
M84I	19.1 ± 0.1	0.6 ± 0.6	$(1.8 \pm 2.8) \times 10^{-7}$
S85N	100	$(6.2 \pm 3.4) \times 10^{-5}$	$(2.3 \pm 0.9) \times 10^{-8}$
G88R	100	$(4.8 \pm 2.4) \times 10^{-5}$	$(2.7 \pm 2.2) \times 10^{-7}$
R89Q	99.7 ± 0.5	$(3.3 \pm 2.0) \times 10^{-5}$	$(8.1 \pm 11.2) \times 10^{-8}$
N131S	90.5 ± 6.2	$(8.1 \pm 3.6) \times 10^{-5}$	$(5.1 \pm 7.6) \times 10^{-7}$
A138V	2.8 ± 1.8	0.4 ± 0.5	$(3.3 \pm 4.2) \times 10^{-6}$
A138T	61.6 ± 10.4	$(2.8 \pm 1.7) \times 10^{-3}$	$(1.4 \pm 1.5) \times 10^{-7}$
D140N	100	$(2.2 \pm 1.5) \times 10^{-4}$	$(4.0 \pm 4.1) \times 10^{-8}$
K157E	8.8 ± 1.3	$(5.3 \pm 5.8) \times 10^{-3}$	$(5.3 \pm 4.1) \times 10^{-8}$
G161S	91.6 ± 1.7	$(2.9 \pm 1.2) \times 10^{-4}$	$(1.2 \pm 1.5) \times 10^{-7}$
E251K	100	$(6.2 \pm 5.9) \times 10^{-5}$	$(4.1 \pm 5.1) \times 10^{-8}$
K260E	100	$(7.8 \pm 3.0) \times 10^{-5}$	$(5.5 \pm 4.2) \times 10^{-8}$
S261P	100	$(5.7 \pm 1.2) \times 10^{-5}$	$(1.8 \pm 1.7) \times 10^{-8}$
T285I	1.5 ± 1.3	$(1.1 \pm 0.4) \times 10^{-2}$	$(9.0 \pm 7.0) \times 10^{-7}$
C324R	100	$(4.2 \pm 1.9) \times 10^{-5}$	$(3.1 \pm 5.3) \times 10^{-7}$

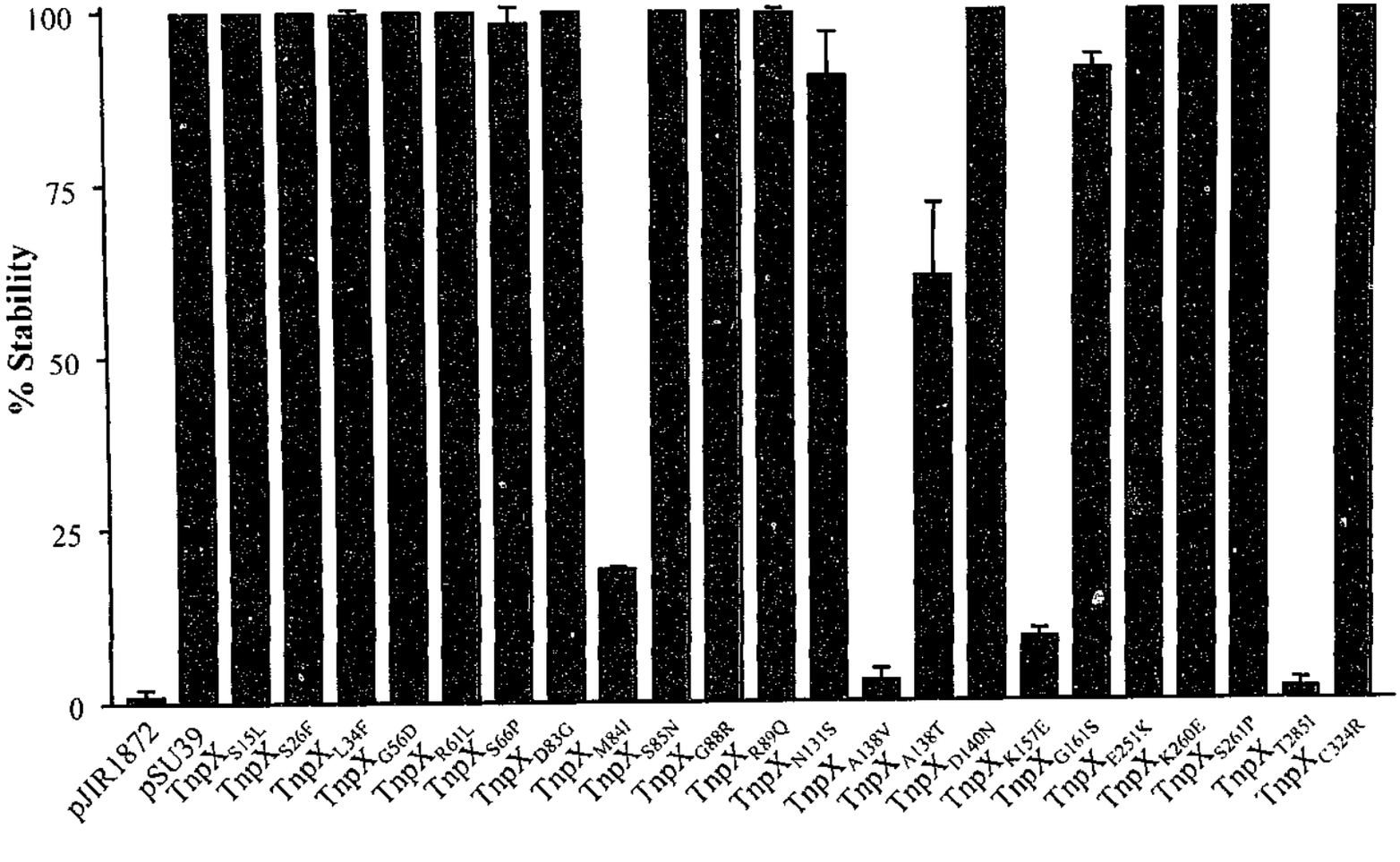
[a] The % stability was calculated as the percentage of ampicillin resistant colonies that were chloramphenicol resitant.

[b] The insertion frequency was calculated as the viable count at 42°C divided by the viable count at 30°C.

[c] The transposition frequency was calculated as the number of transconjugants conferring chloramphenicol resistance (Tn*4453a*) divided by the number of transconjugants conferring tetracycline resistance (pVS520).

Figure 5.3 Analysis of the ability of the TnpX mutants to excise Tn*4451tnpXΔ1*

Plasmid DNA was isolated from each strain and used to transform DH5α to ampicillin resistance. Subsequent colonies were cross patched onto medium containing ampicillin or chloramphenicol and the percentage of chloramphenicol resistant colonies determined. This figure shows a graph of the percentage stability calculated from at least three independent assays for each mutant and the positive (pJIR1872) and negative (pSU39) controls.


% Stability
100
75
50
25
0
pJIR1872
pSU39
TnpX S15L
TnpX S26F
TnpX L34F
TnpX G56D
TnpX R61L
TnpX S66P
TnpX D83G
TnpX M84I
TnpX S85N
TnpX G88R
TnpX R89Q
TnpX N131S
TnpX A138V
TnpX A138T
TnpX D140N
TnpX K157E
TnpX G161S
TnpX E251K
TnpX K260E
TnpX S261P
TnpX T285I
TnpX C324R

plasmid was not suitable for testing the pJIR1872 plasmids, which also conferred kanamycin resistance. The alternative plasmid, pJIR2421, encodes chloramphenicol resistance and is otherwise identical to the original insertion plasmid, being temperature sensitive and containing the joint of the circular intermediate. Each mutant, positive control (pJIR1872) and negative control (pSU39) plasmid was introduced separately into DH5α(pJIR2421) and the insertion assay was performed as described (Chapter 4). The insertion frequency for each mutant was then calculated (Table 5.2, Fig 5.4). Most of the mutants exhibited no insertion activity above the background reversion frequency indicated by the pSU39 results. However, $TnpX_{A138T}$, $TnpX_{D140N}$, $TnpX_{K157E}$, $TnpX_{G161S}$ and $TnpX_{T285I}$ showed a 5 to 50-fold increase and $TnpX_{S66P}$, $TnpX_{M84I}$ and $TnpX_{A138V}$ showed wild-type levels of insertion (Fig 5.4).

The random *tnpX* mutants were also tested using the transposition assay, with only a slight alteration being made to the assay procedure. Previously, the *tnpX* encoding plasmid also encoded ampicillin resistance, which was used for selection. Instead, in these experiments kanamycin resistance was used for strain selection. The transposition strains containing the random *tnpX* mutants were constructed as described previously and the assay was performed as before (Chapter 4). The transposition frequencies were calculated and it was found that every mutant displayed a negative dominant phenotype, although there was some variation in transposition frequency (Table 5.2, Fig 5.5).

d) Purification and DNA Binding Analysis of Selected TnpX Mutants

At the commencement of these studies it was anticipated that random mutations would be obtained outside of the N-terminal 200 aa, thereby identifying residues important to structure or function that were outside the resolvase catalytic domain. Five such mutants were detected, $TnpX_{E251K}$, $TnpX_{K260E}$, $TnpX_{S261P}$, $TnpX_{T285I}$ and $TnpX_{C324R}$.

Figure 5.4 Analysis of the ability of the TnpX mutants to catalyze insertion

The insertion assay strains were incubated at 30°C in the presence of appropriate antibiotics prior to duplicate plating for incubation at either 30°C or 42°C. The viable count of each strain at each temperature was determined to calculate the insertion frequency. The graph shows the average insertion frequencies obtained from at least three independent assays for the mutant TnpX proteins and the positive (pJIR1872) and negative (pSU39) controls.

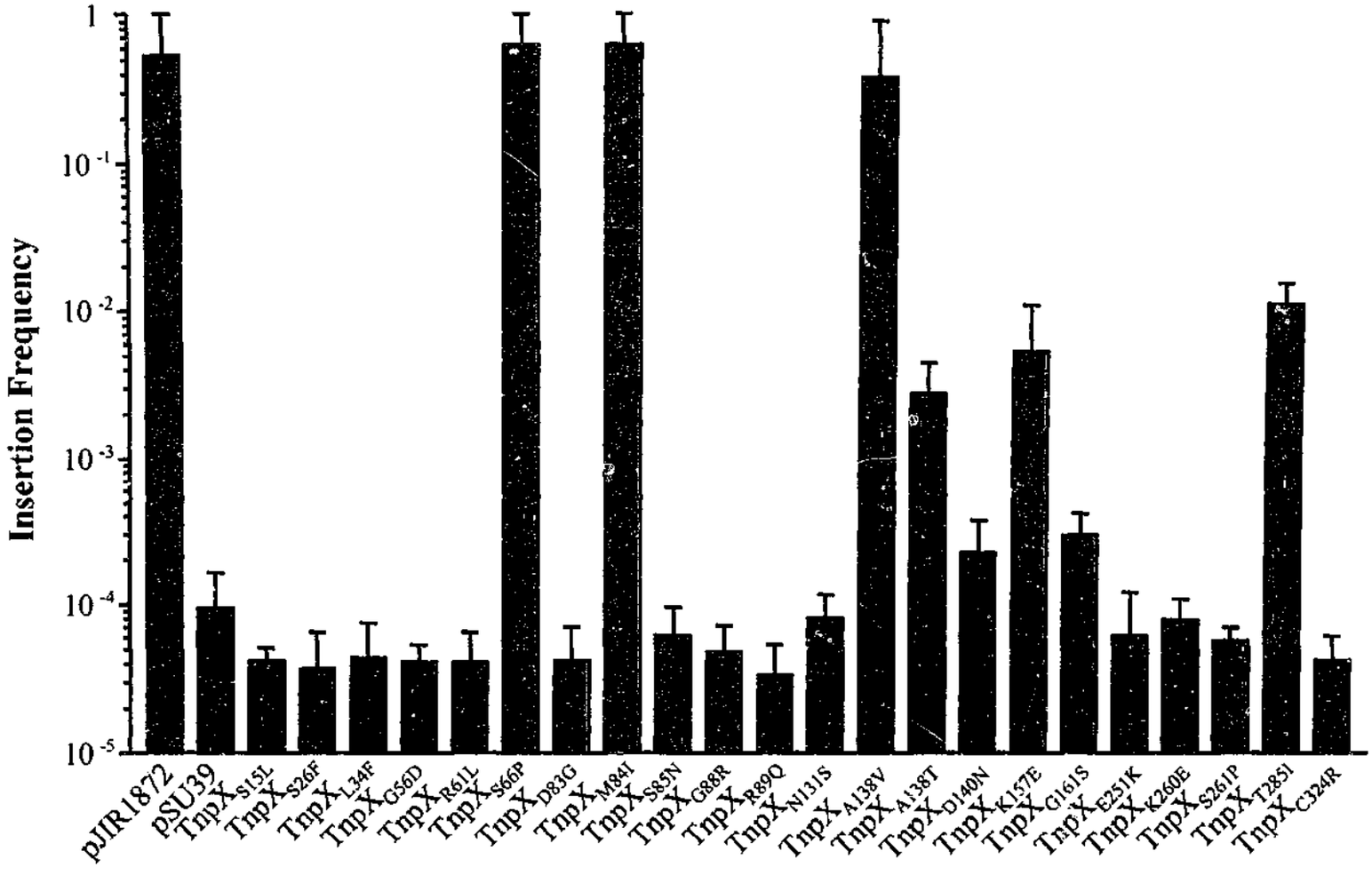
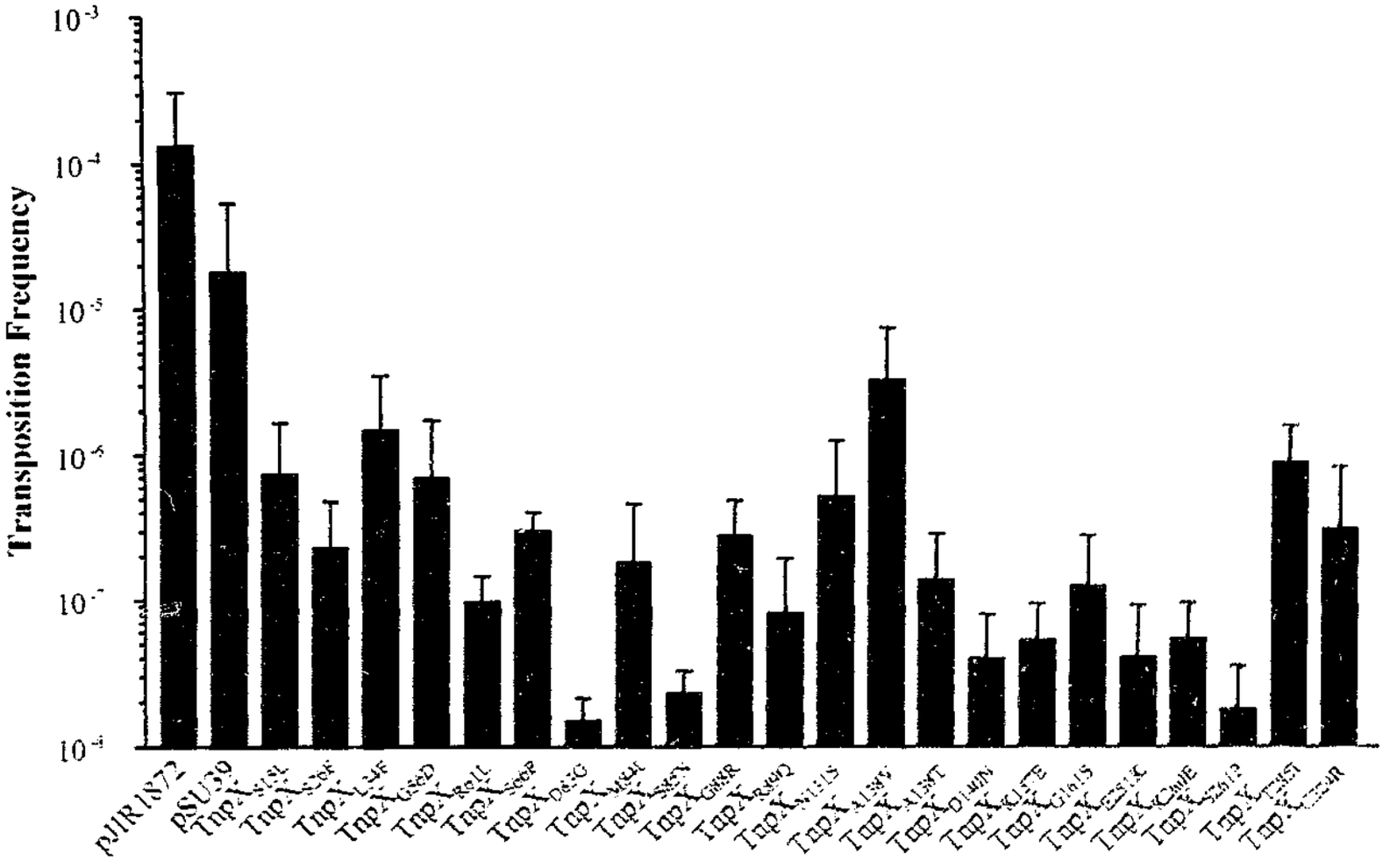

Insertion Frequency
1
10 -1
10 -2
10 -3
10 -4
10 -5
pJIR1872
pSU39
TnpX S15L
TnpX S26F
TnpX L34F
TnpX G56D
TnpX R61L
TnpX S66P
TnpX D83G
TnpX M84I
TnpX S85N
TnpX G88R
TnpX R89Q
TnpX N131S
TnpX A138V
TnpX A138T
TnpX D140N
TnpX K157E
TnpX G161S
TnpX E251K
TnpX K260E
TnpX S261P
TnpX T285I
TnpX C324R

Figure 5.5 Analysis of the ability of the TnpX mutants to catalyze transposition

The transposition donor strains were used in matings with the rifampicin resistant recipient strain, LT101. Transconjugants that had received the transposon were detected by selection on rifampicin and chloramphenicol and transconjugants that had received the conjugative plasmid pVS520 were selected on medium containing rifampicin and tetracycline. The transposition frequency was subsequently determined from at least three independent assays. The graph shows the average transposition frequencies of each TnpX mutant and the positive (pJIR1872) and negative (pSU39) control.


Transposition Frequency
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
pJIR1872
pSU39

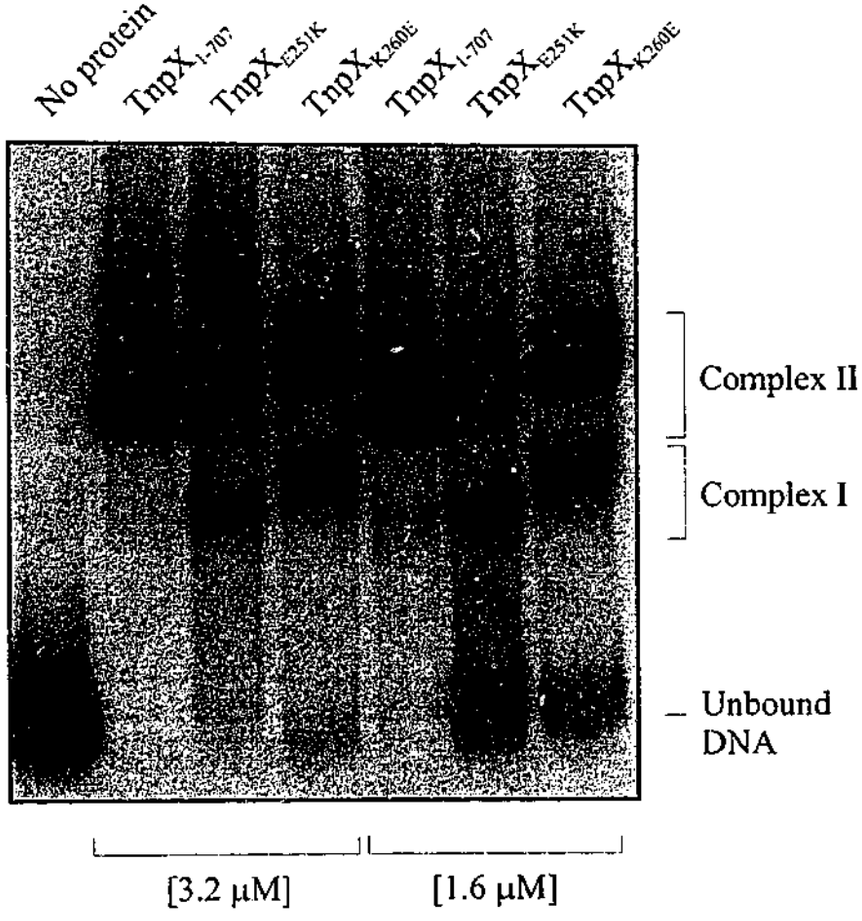
Phenotypic analysis indicated that $TnpX_{T285I}$ was not significantly reduced in function and it was thought that $TnpX_{S261P}$ was a radical amino acid sequence change that was likely to affect the folding of the protein. Consequently, purification of the $TnpX_{S261P}$ and $TnpX_{T285I}$ mutant proteins was not attempted. The genes encoding the $TnpX_{E251K}$, $TnpX_{K260E}$ and $TnpX_{C324R}$ mutations were introduced into the expression plasmid pET22b in the same manner as for the wild-type *tnpX* gene. The resultant constructs were sequenced to confirm that no errors had been incorporated during PCR amplification and the His-tagged recombinant proteins were over-expressed and purified as for the $TnpX_{1-707}$ protein described previously (Chapter 3).

Purification of both $TnpX_{E251K}$ and $TnpX_{K260E}$ was performed and yielded protein preparations similar in quality to those obtained for $TnpX_{1-707}$. $TnpX_{E251K}$ and $TnpX_{K260E}$ were subjected to gel filtration to determine their molecular size in solution. It was found that both behaved in a similar manner to $TnpX_{1-707}$, eluting at the same volume and showing the same asymmetrical elution peak as $TnpX_{1-707}$ (Chapter 3). The elution volumes corresponded to molecular sizes of 190 kDa (data not shown). However, the $TnpX_{C324R}$ protein could not be purified as the protein became insoluble during the purification process.

Gel mobility shift analysis was carried out to determine whether the two mutant proteins were capable of binding DNA. The $TnpX_{1-707}$ protein produces two complexes, I and II, when incubated with appropriate target DNA with the lower mobility complex, complex II constituting the main complex. Complex I, although being barely discernable at high protein concentrations, being visualized at lower protein concentrations. Both $TnpX_{E251K}$ and $TnpX_{K260E}$ had aberrant banding patterns when incubated with the left end Tn*4451* fragment compared to that of $TnpX_{1-707}$ (Fig 5.6). The banding pattern exhibited by $TnpX_{E251K}$ was closer to the $TnpX_{1-707}$ pattern than that of $TnpX_{K260E}$ but

Figure 5.6 Gel mobility shift analysis of purified $TnpX_{E251K}$ and $TnpX_{K260E}$ proteins

Each of the mutant proteins and $TnpX_{1-707}$ were mixed separately with radiolabelled left end DNA. The binding mixture was subjected to electrophoresis through a native polyacrylamide gel followed by autoradiography. The location of complexes I and II are indicated on the right-hand side of the autoradiograph.


No protein
TnpX1-707
TnpXE251K
TnpXK260E
TnpX1-707
TnpXE251K
TnpXK260E
Complex II
Complex I
Unbound DNA
[3.2 μM]
[1.6 μM]
TnpX Concentration

still differed significantly. The $TnpX_{E251K}$ complexes had significantly more DNA migrating at approximately the same position as complex I of $TnpX_{1-707}$, the proportion being approximately 50% at lower protein concentrations and slightly less at higher protein concentrations. Both $TnpX_{E251K}$ complexes migrated at a very similar position to the $TnpX_{1-707}$ complexes although there maybe a slight decrease in mobility.

The $TnpX_{K260E}$ complexes produced were significantly altered when compared to the banding pattern of $TnpX_{1-707}$ binding, although again two complexes were formed. Similar to $TnpX_{E251K}$ the proportion of DNA located in each of the complexes was approximately 50% at lower protein concentrations however, unlike $TnpX_{E251K}$ this did not change when the protein concentration was increased. The mobility of the $TnpX_{K260E}$ complexes was also affected, as the complexes had a lower electrophorectic mobility than the $TnpX_{1-707}$ or $TnpX_{E251K}$ complexes.

Discussion

It has been shown by the analysis of partial chymotryptic digests that the small resolvase protein, γδ resolvase, consists of two domians. The first 140 aa contains the catalytic and protein-protein interaction domains and the last 43 aa constitutes the DNA binding domain (Abdel-Meguid *et al.*, 1984). More recently, several structures have been solved for γδ resolvase (Abdel-Meguid *et al.*, 1984; Sanderson *et al.*, 1990; Yang and Steitz, 1995).

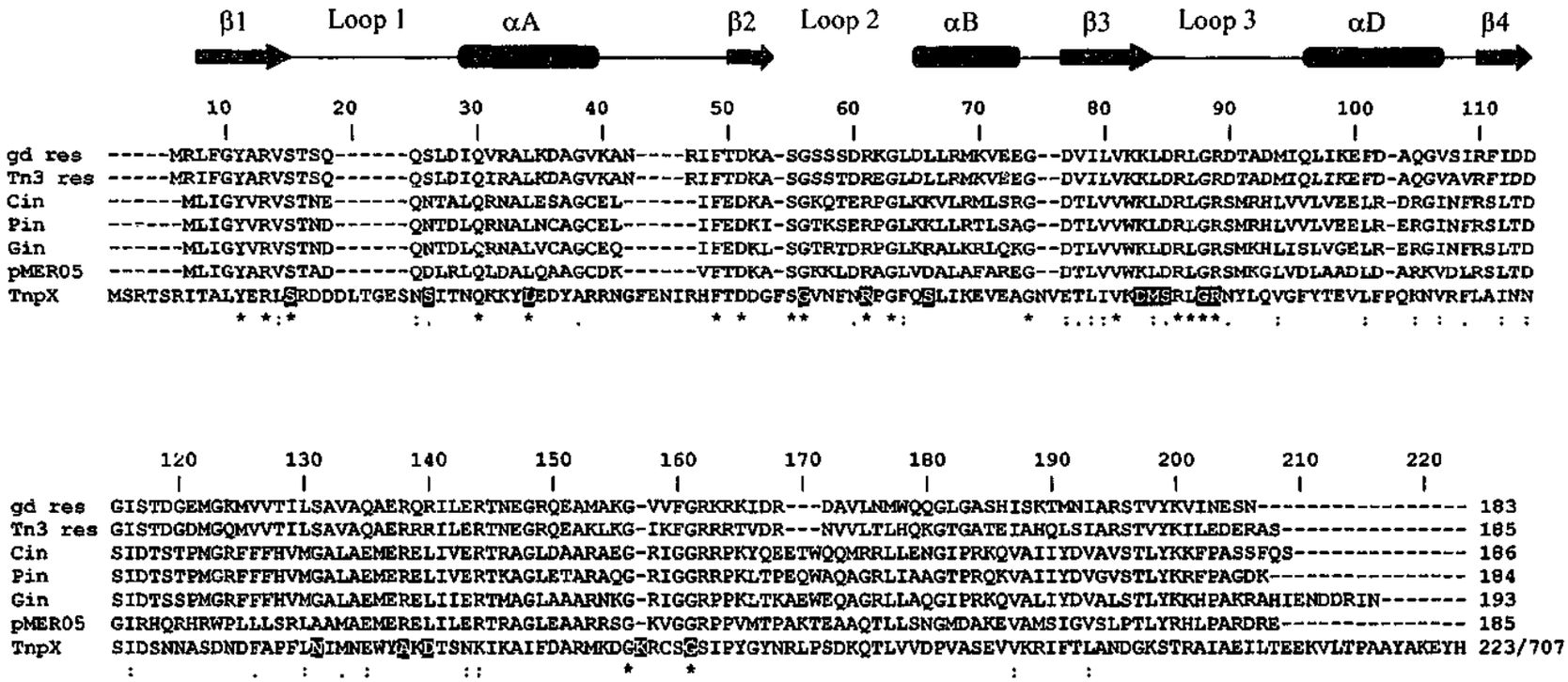
Within the catalytic domain of resolvase proteins there are many residues that are highly conserved within the resolvase family (Fig 5.7) (Pan *et al.*, 2001; Smith and Thorpe, 2002; Yang and Steitz, 1995). Several of the TnpX random mutants shown here to be important for function were located at residues that were conserved within the resolvase

Figure 5.7 Alignment of conserved resolvase domain amino acid sequences

The sequences aligned using the clustal W algorithm at the NPS@ protein analysis site (Thompson et al., 1994) and include: gd res (γδ resolvase accession number (acc. no.) P03012), Tn3res (resolvase from Tn*3*, acc. no. P03011), Cin (Cin invertase, acc. no. P21703), Pin (Pin invertase, acc. no. NP 700400), Gin (Gin invertase, acc. no. NP050655), pMER05 (resolvase from pMER05, acc. no. AAB02643) and TnpX (acc. no. S78538). Complete sequence conservation is indicated by red text and a red star below the sequences. Strongly conserved residues are indicated by green text and two dots below the alignment. Low level conservation is indicated by blue text and a single dot below the text. The position of the TnpX random mutants is indicated by black boxes around the mutated residue. Numbers at the ends of the aligned sequences give the total protein length except for TnpX, where only the N-terminal 223 amino acids are aligned. Above the alignment the secondary structure of γδ resolvase derived from the known crystal structure is shown, based on a figure from Pan *et al* (2001). The four β–sheets (β1-4) are indicated by pink block arrows. The three α-helical structures (αA, αB and αD) are represented by blue barrels. The mobile loops Loop1 (green line) containing the catalytic serine residue, Loop 2 (yellow line) and Loop 3 (red line) are also indicated.

arrows. The three α-helical structures (αA, αB and αD) are represented by blue barrels.

catalytic serine residue, Loop 2 (yellow line) and Loop 3 (red line) are also indicated.



family. These residues were located within the catalytic domain and include S15, L34, G56, R61, G88 and R89 (Fig 5.7). Several of these residues have either been mutated and found to be important to γδ resolvase function or are thought to play an important role in structure or function of the catalytic site given their position within the co-crystal structure (Hughes *et al.*, 1990; Newman and Grindley, 1984; Pan *et al.*, 2001; Yang and Steitz, 1995). The equivalent γδ residue to the TnpX mutant $TnpX_{R89Q}$ is thought to be responsible for contacting the DNA backbone close to the cleavage position (Yang and Steitz, 1995), while the γδ residue equivalent to $TnpX_{S85N}$, although not a serine, contributes to the stabilization of the catalytic pocket (Yang and Steitz, 1995). Catalytic serine mutants of γδ resolvase ($TnpX_{S15L}$ in TnpX) are capable of DNA binding but incapable of recombination *in vivo* (Newman and Grindley, 1984). Two serine γδ mutants were obtained, S10L and S10C. Interestingly, the cysteine change resulted in a protein capable of wild-type DNA binding while the leucine mutant showed altered binding affinities for the different subsites within the recombination site, *res* (Newman and Grindley, 1984).

Mutagenesis of the resolvase γδ has lead to the identification of several other mutants such as R8Q, R68H, L69P and R71H, which are recombination deficient but proficient for DNA binding (Hughes *et al.*, 1990; Newman and Grindley, 1984). From mutagenic studies, and from their location within the crystal structure, these residues are thought to be important for stabilization of scissile bonds during catalysis or stabilization of the catalytic pocket (Pan *et al.*, 2001; Yang and Steitz, 1995). Several TnpX mutations resulting in loss of function lie very close to the equivalent position of γδ R68-R71, including $TnpX_{D83G}$, $TnpX_{M84I}$, $TnpX_{S85N}$, $TnpX_{G88R}$ and $TnpX_{R89Q}$. This region within the γδ resolvase structure is located on a flexible loop, Loop 3, which is capable of coming into close proximity of Loop 1, containing the catalytic serine residue. It is

postulated that residues located within this region may participate in catalysis by coordinating the phosphate moiety during the cleavage reaction (Pan *et al.*, 2001). The TnpX mutant, $TnpX_{S26F}$, would be predicted to be located within the Loop 1 region and this residue, although not completely conserved within the resolvase family, does generally contain a relatively small polar or charged residue (Fig 5.7), which in $TnpX_{S26F}$ has been replaced by an aromatic ring. This change may have resulted in inflexibility or disruption of the local protein structure.

Other γδ resolvase mutants such as G129D, A133T and I164M are incapable of both recombination and DNA binding (Hughes *et al.*, 1990). Several TnpX mutants were located close to this area (Fig 5.7) including $TnpX_{N131S}$, $TnpX_{A138V}$, $TnpX_{A138T}$, $TnpX_{D140N}$, $TnpX_{K157E}$ and $TnpX_{G161S}$, although the similarity of TnpX to small resolvase protein begins to diminish after the first 100 aa of TnpX. Regardless, the TnpX mutants $TnpX_{N131S}$, $TnpX_{D140N}$ and $TnpX_{G161S}$ were also found to be recombination deficient, although their ability to bind to target DNA was not tested.

Most of the non-functional TnpX mutants were located within the catalytic domain and were incapable of catalyzing both excision and insertion. As shown previously, TnpX-mediated excision is dependent on an intact resolvase catalytic domain (Crellin and Rood, 1997) and for the first time it has been shown that key resolvase catalytic residues are also essential for the insertion process.

One of the most interesting mutants was $TnpX_{S66P}$. This mutant was severely reduced in excision activity, although the insertion frequency was the same as for the wild-type. This residue did not coincide with a conserved resolvase residue (Fig 5.7) and the non-conservative substitution to a proline residue suggests that there may have been a significant alteration to the protein structure. However, the protein must still be folded

correctly since its ability to catalyze the insertion process remains unaffected. These data suggest that the complexes involved in the excision and insertion reactions differ, therefore a structural change can affect one reaction but not the other. These data provide the first evidence that different regions or residues of TnpX are required for excision and insertion.

All of the 22 random *tnpX* mutants exhibited a negative dominant phenotype in the transposition assay. It is postulated that this phenotype is attributable to interference by the mutant TnpX protein with the wild-type TnpX protein expressed from the intact copy of the transposon that was located on the chromosome of the donor cell. The non-functional proteins may form a non-productive dimer with the wild-type protein; this dimer may be unable to bind DNA, or to form a productive synapse. Alternatively, the mutant proteins may bind to the target DNA but not dimerize, preventing access of the productive wild-type TnpX dimers to the DNA binding site. Several of the mutants were at best reduced in function rather than non-functional. These mutants included $TnpX_{M84I}$, $TnpX_{A138V}$, to a lesser extent $TnpX_{A138T}$, $TnpX_{K157E}$ and $TnpX_{T285I}$. These mutants still demonstrated a negative dominant phenotype in the transposition assay. This result suggests that although some of the mutants were almost fully functional it is likely that they could not form productive dimers with the wild-type protein.

Only four non-functional mutants were located outside of the resolvase catalytic domain. Clearly, the catalytic domain is critical for resolvase function. However, since mutants for some of the conserved resolvase family residues were not detected in the random mutagenesis screening it was likely that the procedure had not reached saturation point. The screening process was stopped after obtaining the key catalysis mutant $TnpX_{S15L}$, which was obtained late in the procedure after multiples of other mutants were isolated. Lack of saturation could explain why few residues outside of the

catalytic domain were found to be mutated. Alternatively, alteration of a single amino acid residue may be insufficient to completely abolish function within the C-terminal domain of the protein.

The non-functional mutants located furthest from the N-terminus included $TnpX_{E251K}$, $TnpX_{K260E}$, $TnpX_{S261P}$ and $TnpX_{C324R}$. None of these mutants could catalyze excision. Time constraints dictated that not all of the random mutants could be overexpressed and purified and it was decided that the mutants within TnpX-specific regions, outside the resolvase domain, were of greatest interest. Purification of $TnpX_{S261P}$ was not attempted, for reasons already discussed. Consequently, purification of $TnpX_{E251K}$, $TnpX_{K260E}$ and $TnpX_{C324R}$ was attempted. Unfortunately, $TnpX_{C324R}$ could not be purified as the protein was insoluble and prone to degradation. The inability to purify $TnpX_{C324R}$ was unfortunate as the C324 residue is conserved within the large resolvase family as part of a putative zinc finger motif that may be involved in Zn^{2+} binding (I. Lucet and J. I. Rood, unpublished results). The purification data suggested that the structure of TnpX was affected by the alteration of the cysteine residue, indicating that this residue may be of structural importance.

The purified mutant proteins were capable of dimerization and DNA binding, although the binding patterns varied from that of wild-type TnpX. The amino acid substitutions in $TnpX_{E251K}$ and $TnpX_{K260E}$ both resulted in a charge reversal compared to the wild-type protein, which may have affected electrophoretic mobility in the native gel used in the binding analysis. However, the shift in electrophoretic mobility is more likely to result from a difference in the protein-DNA complex that leads to greater retardation within the gel matrix. There is a precedent for this model. It has been shown that when γδ resolvase, is bound to site I it induces a DNA bend, whereas the mutant resolvase protein, E118K, shows an altered DNA binding profile, is incapable of recombination

and does not induce a DNA bend at site I. Consequently, the electrophoretic mobility of DNA bound to the mutant resolvase protein differs from that of the wild-type. The stability of DNA binding is also affected, leading to dissociation during electrophoresis (Hatfull *et al.*, 1987).

The second difference between the wild-type and mutant TnpX binding patterns involved the proportions of the two complexes. The wild-type complex I constituted a very minor complex, complex II was almost the exclusive product at high protein concentrations. However, the mutant proteins produced substantially more complex I (Fig 5.6). It is unfortunate that the exact constituents of complexes I and II are not yet known since their determination would provide additional information about the aberrant binding patterns.

In summary, results described in this chapter have identified a key region of TnpX outside of the catalytic domain that is of functional significance (amino acids 250-261). Residues within the resolvase catalytic domain that are important to function have also been identified, some of these residues are conserved while others are not. Finally, and most importantly, was the identification of the mutant $TnpX_{S66P}$. The characteristics of this mutant suggest that there are regions of TnpX that differ in their structural or functional importance between the excision and insertion reactions. In future studies this mutant could be used to study the excision and insertion reactions in isolation.

Chapter 6

Discussion and Future Directions

The Tn*4451/53* group of elements constitute a family of clostridial mobilizable transposons whose transposition is mediated by the site-specific recombinase TnpX (Lyras *et al.*, 1998). It was previously demonstrated that TnpX utilized a resolvase catalytic mechanism to mediate transposon excision as a circular molecule (Bannam *et al.*, 1995; Crellin and Rood, 1997) and the circular molecule was subsequently shown to be the transposition intermediate (Lyras and Rood, 2000b). It was not known whether other transposon-encoded proteins or cellular proteins were required for transposon excision, or whether TnpX was involved in the insertion process. Additionally, nothing was known about the structure or function of the TnpX protein apart from the resolvase domain. Finally, the nature and position of the TnpX DNA binding sites was unknown, although they were assumed to be adjacent to the GA dinucleotides where DNA cleavage occurs.

Structure of TnpX

The native TnpX protein is 707 aa in length, although the purified TnpX protein used in this study contains a hexa-His tag and is 715 aa long. Analysis of the amino acid sequence of TnpX had led to the identification of several regions of interest. The region between amino acids 324-360 contains four cysteine residues that are conserved within members of the large resolvase family. It is thought that this region may form a zinc binding motif as the predicted secondary structure of TnpX in this region also shows the characteristic structure of this motif, namely the presence of two β-sheets and an α-helix (Laity *et al.*, 2001). The next region of note is the leucine rich region between amino acids 404-450. This region is predicted to be highly α-helical and likely to form

a coiled-coil structure. Between amino acids 541-557 is a region containing a large number of glutamic acid residues, followed by residues 558-574 which contain a large number of arginine and lysine residues. The formation of a coiled-coil is also predicted in this region. The final ~100 amino acids of TnpX includes ten of the seventeen cysteine residues found within the protein. These cysteine residues include a repeated sequence, $CXXCX_{12}FCSXXC$ (I. Lucet and J. I. Rood, unpublished results). This sequence appears twice between amino acids 591-660 but its significance is unknown.

The results presented in this thesis have led to the identification of regions of TnpX that may serve a specific function. Testing of the truncated derivatives of TnpX using gel mobility shift analysis indicated that the loss of the region between amino acids 493-597 resulted in the loss of the ability to bind DNA. These data suggest that a DNA binding domain exists within this region. The presence of a coiled-coil structure is predicted in this oppositely charged region so it is postulated that this domain is involved in DNA binding. This hypothesis is supported by the fact that this region is present in the binding proficient derivative, $TnpX_{1-597}$, but absent in $TnpX_{1-492}$, which is unable to bind DNA very effectively. Recent studies (I. Lucet and J. I. Rood, unpublished results) have shown that the region of TnpX between amino acids 533-583 can be overexpressed and purified and is capable of binding the same DNA binding sites as $TnpX_{1-707}$.

$TnpX_{1-492}$ is capable of binding DNA only when present at very high protein concentrations. If the region between amino acids 493-597 was the only DNA binding region then no DNA binding should be evident with $TnpX_{1-492}$. These results suggest that there may be another region of TnpX that can either bind DNA or that contributes to DNA binding. Another such region was identified by random mutagenesis of TnpX. Two random TnpX mutants were overexpressed and purified, $TnpX_{E251K}$ and

$TnpX_{K260E}$. These proteins did not show a normal DNA binding pattern, which suggests that this region is also involved in DNA binding. The binding demonstrated by both $TnpX_{E251K}$ and $TnpX_{K260E}$ is dependent on the presence of the last 110 amino acids. That is, if these residues are removed these proteins are no longer capable of DNA binding (I. Lucet and J. I. Rood, unpublished results). Additional mutants in the same region as $TnpX_{E251K}$ and $TnpX_{K260E}$ have now been produced using site-directed mutagenesis and have been shown to exhibit the same aberrant binding pattern as the original random mutants (M. Chiarezza, I. Lucet, D. Lyras and J. I. Rood, unpublished results), providing further support for the hypothesis that the region is involved in DNA binding.

Another important function of TnpX is its ability to dimerize. The truncated TnpX derivative $TnpX_{1-356}$ was not capable of dimerizing to the same extent as the other TnpX derivatives. This result suggested that the dimerization site was altered by this truncation. The putative zinc binding motif is located between amino acids 324-360. The last cysteine is C360, consequently $TnpX_{1-356}$ is missing this residue, which may explain why the equilibrium of this protein lies towards a monomeric rather than the dimeric form. Perhaps the last cysteine residue is important for dimer stability. The production and testing of two other truncated mutants, one with the 324-360 region intact and one lacking this region, would allow further characterization of this region and confirmation as to whether it is in fact the site of dimerization.

The function of the last domain of TnpX, from residues 598 to 707, is unknown. The domain is not essential for TnpX activity but the presence of this region did increase the excision frequency substantially. In contrast, in the absence of this domain there was an increase in insertion frequency. Furthermore, both $TnpX_{1-356}$ and $TnpX_{1-492}$ were capable of catalyzing a low level of insertion *in vivo*. During the purification of the

truncated TnpX derivatives it was noted that $TnpX_{1-597}$ purified more readily than the full length protein, suggesting that $TnpX_{1-597}$ was a more stable derivative than $TnpX_{1-707}$. If so this could explain the difference in insertion activity evident in the *in vivo* insertion assay, since the levels of functional $TnpX_{1-597}$ within the cell may be higher than the wild-type levels, thereby appearing to give a higher level of insertion activity. It is possible that the last domain, between amino acids 598-707, may be involved in stabilization of the excision reaction complex. Indeed, recent studies have indicated that the amino acid region 583-707, when overexpressed and purified, is capable of DNA binding (I. Lucet and J. I. Rood, unpublished results), suggesting that a stabilization role may be accomplished by further protein-DNA contact.

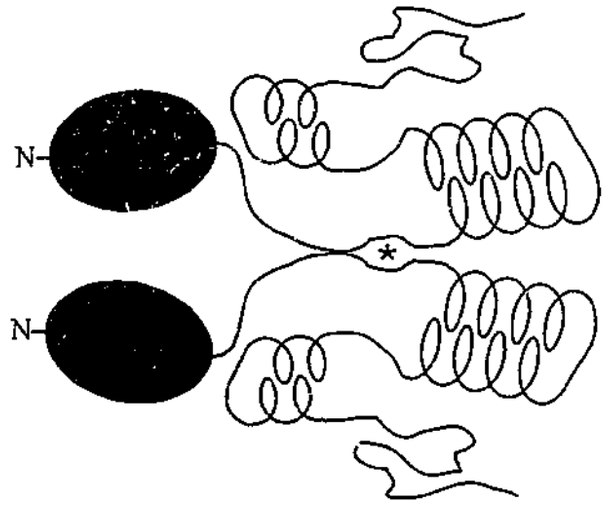
The gel filtration data presented in this thesis suggests that TnpX is not a completely globular protein since the elution peaks were asymmetrical, presumably due to the non-uniform movement of the TnpX dimers through the gel matrix. The progressive loss of the C-terminal regions of TnpX results in the production of gel filtration peaks with a more symmetrical appearance. This result suggests that the last few domains of TnpX are less globular than the preceding domains.

The results enable the construction of a model that describes the structure of TnpX (Fig 6.1). It is postulated that outside the resolvase catalytic domain there is the first region of a bipartite DNA binding domain (the 250-260 amino acid region), the second region being the oppositely charged coiled-coil located between residues 541-574. These two regions presumably interact to form the intact DNA binding domain. Other work conducted in this laboratory supports this hypothesis (M. Chiarezza, I.Lucet, D. Lyras and J. I. Rood, unpublished results). The region just downstream of the first DNA binding site, between amino acids 324-360, is thought to be the dimerization site and also may be a putative zinc binding domain. The leucine rich region between

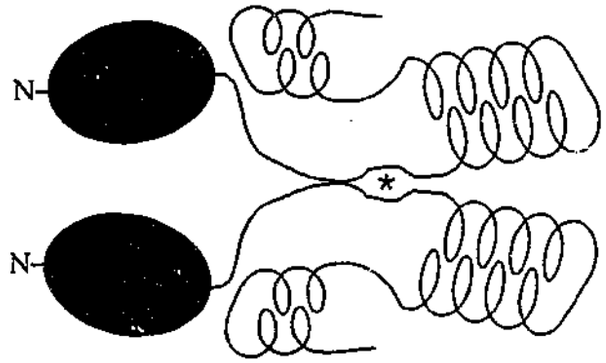
Figure 6.1 Structural model of a TnpX dimer

The resolvase domain (purple oval) is at the N-terminus of the protein (N), followed by the conserved charged domain (purple line), with the dimerization site indicated by a red star. The multimerization domain is shown by the green spiral, indicating the presence of a putative coiled-coil structure. The second coiled-coil structure (red and blue spirals) indicates the oppositely charged coiled-coil thought to be part of the major DNA binding domain. The final cysteine-rich domain (black line) is located at the carboxy terminus. All four TnpX derivatives used in this study are shown. $TnpX_{1-597}$ lacks the cysteine rich domain, $TnpX_{1-492}$ lacks part of the leucine rich coiled-coil and the oppositely charged coiled coil and $TnpX_{1-356}$ lacks all of the leucine rich coiled-coil and part of the conserved charged domain.

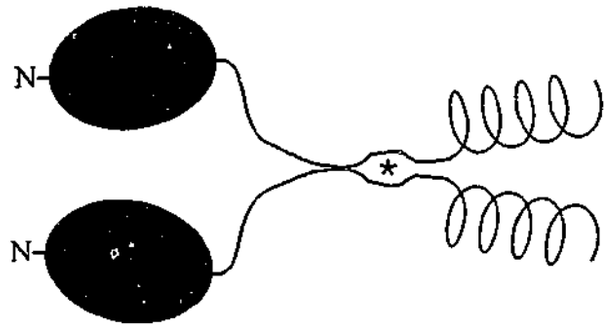
TnpX$_{1-707}$



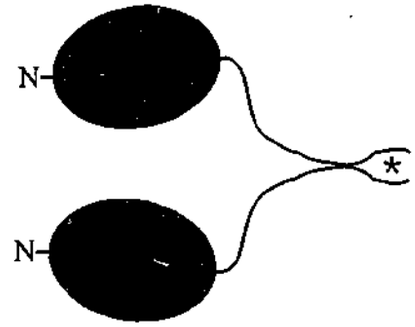
TnpX$_{1-597}$



TnpX$_{1-492}$



TnpX$_{1-356}$



amino acids 400-500 contains a high degree of helical structure also postulated to form a coiled-coil domain. It is primarily the leucine rich coiled-coil domain that is thought to form the most non-globular part of the protein, and is also thought to be the site of multimerization. There is no supporting data for this hypothesis, but it is presumed that TnpX dimers are able to multimerize to facilitate synaptic complex formation. The structure of the final domain and the role of the cysteine repeats in this region are unknown. However, given the preliminary data it is suggested that this domain is another DNA binding domain important for the excision reaction rather than the insertion reaction.

Elucidation of the structure of TnpX would most easily be accomplished by solving the TnpX crystal structure. Attempts to produce TnpX crystals have so far proved refractory as expression levels of $TnpX_{1-707}$ are not very high and $TnpX_{1-707}$ is not particularly soluble in buffers conducive to crystallization. However, given the higher stability of the truncated derivative $TnpX_{1-597}$, an attempt to crystallize this derivative may prove to be more feasible. In addition, it would be worthwhile trying to solve a TnpX crystal structure that also contained DNA. The addition of DNA to the crystallization trays, if not producing co-crystals, may also serve to stablize the TnpX protein, allowing higher protein concentrations to be obtained before the protein becomes insoluble (J. Rossjohn, personal communication).

Other studies could include the production of additional truncated TnpX proteins, including protein fragments such as the 230-260 amino acid region in the presence and absence of the 541-574 amino acid region. The objective would be to look at interactions between these two regions as well as assessing DNA binding capacity. The region between amino acids 324-360 could be assessed for dimerization as mentioned earlier. Similarly, the leucine rich domain could be investigated using site-directed

mutagenesis to mutate particular amino acid residues. To determine if this region is responsible for multimerization the resultant proteins could then be tested for the production of higher order complexes when TnpX binds to the circular intermediate.. One way to assess all of the above interactions would be through the use of surface plasmon resonance. This system involves detection of differences in the refractive index of an anchored ligand with the addition of a second ligand that interacts with the first (McDonnell, 2001). For example, this technique has been used to investigate the interaction kinetics between the DNA binding protein Cmb1, and its DNA substrate (Sassoon *et al.*, 2001). However, the ability of TnpX derivatives containing amino acid regions 230-260 and 541-574, both alone and when mixed together, to bind DNA could initially be assessed by gel mobility shift analysis. Surface plasmon resonance could also be used to determine if these two amino acid regions also interact and whether the amino acid region 324-360 is capable of dimerization, although gel filtration could also be used (McDonnell, 2001).

Transposition Mechanism

The results (Chapter 3) show that TnpX is the only protein required *in vitro* for transposon excision. Additional studies involving *in vitro* insertion assays were less conclusive but indicated that a small amount of insertion can be catalyzed by TnpX in the absence of other proteins. In support of these results, it has been shown recently that none of the other proteins encoded by Tn*4451* are essential for the transposition process in *E. coli* (D. Lyras and J. I. Rood, unpublished results). In particular, a *tnpY* mutant was wild-type for excision, insertion and transposition. The situation may be different in the native host, since *in vitro* studies using purified TnpY have shown that this protein is capable of interacting both with TnpX and with its DNA targets (I. Lucet, D. Lyras and J. I. Rood, unpublished results). TnpY has sequence similarity to DNA

repair proteins and contains a putative Walker A box (Bannam *et al.*, 1995)(D. Lyras and J. I. Rood, unpublished results). TnpY may have a regulatory function that is too subtle to be detected when tested using *in vivo* assay conditions. Alternatively, the function of TnpY may be redundant in *E. coli*, while being essential in *C. perfringens*.

The studies presented here provide further evidence that TnpX catalyzes excision and insertion by a resolvase mechanism. The random mutant study (Chapter 5) confirmed that the resolvase catalytic domain was essential for TnpX function. It was previously shown that if R13, S15, G56 or R89 are mutated, there is a loss of TnpX-mediated excision activity (Crellin, 1997). Several of these residues were also found to be mutated during the random mutagenesis study, specifically S15, G56 and R89. These residues were again found to be essential for TnpX-mediated excision and were also essential for TnpX-mediated insertion. However, within the resolvase catalytic domain, another random mutant of TnpX was isolated, $TnpX_{S66P}$. This mutant although almost incapable of catalysing excision, exhibited an insertion frequency equivalent to that of the wild-type (Chapter 5), indicating that there are some TnpX residues more important for one recombination event than the other and, most critically implying that the two events are distinct reactions. This conclusion can be drawn despite the fact that both reactions involve the same resolvase-mediated catalytic process.

$TnpX_{S66P}$, will be a valuable mutant for examining the insertion process. It was possible that the inability to detect significant activity in the *in vitro* insertion assay was due to the excision and insertion reactions being in equilibrium, with more excision occurring *in vitro* than insertion. These reactions could be uncoupled with the use of a purified $TnpX_{S66P}$ protein. Since $TnpX_{S66P}$ is essentially only capable of insertion, this protein could be used to optimize the *in vitro* insertion reaction in the knowledge that no appreciable excision would be catalyzed.

The DNA substrates for excision are the left and right ends of a supercoiled linear transposon. It has been shown that TnpX binds to these sites with very similar affinity. However, the DNA substrates for the insertion reaction are the joint of the circular intermediate and the intended insertion site. TnpX binds to the circular intermediate joint with high affinity but binding to the insertion sites was shown to be lower by several orders of magnitude. In addition, the binding of TnpX to different target sites was also shown to be variable. This result was not completely unexpected as the sequences surrounding the GA dinucleotide are different (Crellin and Rood, 1997) (Fig 6.2). Furthermore, it would be expected that the excision frequencies from the different transposon insertion sites would also vary due to the different levels of TnpX expression, as it has been shown that when the transposon is integrated the formation of a promoter for *tnpX* is dependent on the upstream sequences (Lyras and Rood, 2000b).

Another difference between the two reactions was evident during the *in vitro* assays. The *in vitro* excision assay was relatively efficient with a maximal conversion of approximately 50% of the starting substrate to recombination product within 15 minutes. In contrast, the *in vitro* insertion assay worked very poorly, despite the use of conditions that were found to be optimal for the excision reaction (Chapter 3). These results imply that the optimal conditions for the two reactions are different. It is considered that the insertion and excision reactions must occur at a comparable efficiency within the bacterial cell as TnpX-mediated insertion of the temperature sensitive plasmid occurs in almost 100% of cells, which is comparable to that of the *in vivo* excision reaction (Chapter 5).

Analysis of the truncated derivatives also indicated that the insertion reaction was different to the excision reaction. The two smallest truncated derivatives, $TnpX_{1-356}$ and

```
GCTCCAGAAGGATATGAGCTTAC          D1

GACCCCGAACGGTAAAAGCGCAT          I1

GACCCCGAAGATTACGCGTATCA          I2

        GAGGGCTATACTTTAAT        Tn4451 Left

GACCCCGA                         Tn4451 Right
```

Figure 6.2 Alignment of deletion and insertion sites and the ends of Tn*4451*

The GA dinucleotide is shown in red. Regions of sequence identity are indicated in green. Only transposon derived sequences are shown for the transposon ends (Tn*4451* left and right). The deletion site from *C. perfringens* is shown (D1) and two insertion sites from *E. coli* (I1 and I2).

$TnpX_{1-492}$, were unable to catalyze any transposon excision. In contrast, they were able to catalyze a small amount of insertion, although the latter may be indicative of the greater sensitivity of the insertion assay compared to the excision assay. Additionally, the truncated derivative, $TnpX_{1-597}$, was able to catalyze insertion at a significantly higher rate than the wild-type TnpX protein, with the observed excision activity catalyzed by $TnpX_{1-597}$ being approximately 50% of the wild-type levels (Chapter 4). Even taking into account the possibility of stability differences between the two proteins, $TnpX_{1-597}$ does not catalyze excision as well as the wild-type protein. Clearly, the requirement for particular TnpX regions and therefore the function of the TnpX domains is different for the excision reaction and the insertion reaction. It may be that the excision reaction requires a larger protein to provide protein scaffolding for the excision mechanism or for activation of the synaptic complex and induction of strand cleavage. The DNA binding site of TnpX appears to contain a single TnpX site unlike the small resolvase binding site, *res*. The additional subsites contained within the *res* site allow binding of additional resolvase dimers that contribute to the structure of the synaptic complex (Arnold *et al.*, 1999). It has been shown that γδ resolvase can be rendered capable of catalysing resolution in the absence of sites II and III of the *res* site by the introduction of two mutations that are thought to result in activated resolvase proteins that, to catalyze strand exchange, do not require activation by the resolvase dimers bound to sites II and III during synapsis (Arnold *et al.*, 1999). In a comparable manner it is possible that the last domain allows stabilization and or activation of the TnpX proteins during synapse.

The excision model postulated by Crellin and Rood (1997) is supported by the data presented in this thesis. This model proposes that the excision synapse is assembled as would be expected for a resolvase-type reaction, by the binding of a TnpX dimer at each end of the transposon. A recombination synapse would then be formed by interactions

between the two TnpX dimers bound at the transposon ends. Strand exchange would occur and the transposon excised from the surrounding DNA as a circular molecule, with the deletion site being reformed in the same state that it was in before transposon insertion (Fig 6.3). It was previously shown that the transposon is excised as a circular molecule (Crellin and Rood, 1997) and the data presented in this thesis have shown that TnpX binds to both the left and right ends of the transposon. Whether the TnpX protein(s) that are bound are dimeric remains to be determined, although TnpX appears to dimerize in solution and resolvase proteins normally function as dimers (Murley and Grindley, 1998). The identification of the exact nature of the TnpX protein(s) complexed to the left and right ends of the transposon DNA will allow determination of whether a protein monomer or dimer is present at each transposon end. Characterization of the protein-DNA complexes could be accomplished using either chemical cross-linking or analytical ultracentrifugation using fluorescently labelled DNA (Lee *et al.*, 2001).

Based on the results described in this thesis, models for the insertion reaction can be constructed, accepting the premise that the synaptic complex formed during the insertion reaction must be different to the excision reaction synapse. Given the differential binding affinity found between TnpX and the circular intermediate joint and the potential insertion sites, it is suggested that TnpX binds as a tetramer to the joint of the circular intermediate. Binding of the second dimer to the circular intermediate joint may facilitate a conformational change that increases the affinity of the bound TnpX for the insertion site DNA (Fig 6.4), leading to the positioning of the four TnpX proteins required for cleavage of four DNA strands and leaving a DNA binding domain free to bind to the insertion site. This structure may then effectively search for a suitable insertion site. Alternatively, interaction of the circular intermediate-TnpX complex with an insertion site may be dependent on the presence of other non-transposon encoded

Figure 6.3 Model for TnpX-mediated excision

The linear transposon (A, blue rectangle) has a TnpX dimer bound to each end (red and green shapes). The ends of the element are brought together in a synapse (B), strand exchange occurs and the circular intermediate (blue circle) and donor DNA (black lines) are released with a resulting exchange of cognate dimer partners (C), consistent with resolvase models of co-integrate resolution (Hallet and Sherratt, 1997).

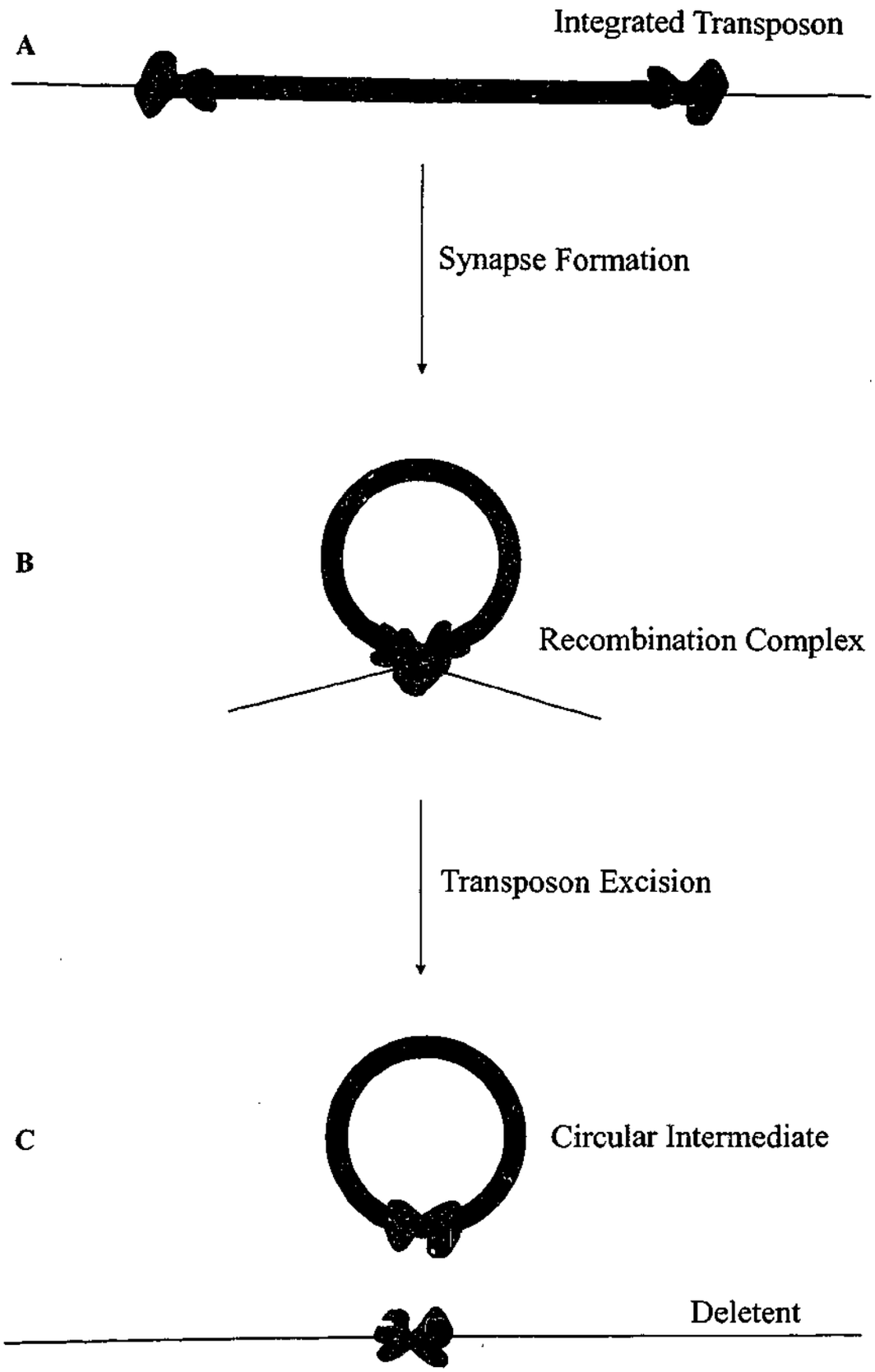

A
Integrated Transposon
Synapse Formation
B
Recombination Complex
Transposon Excision
C
Circular Intermediate
Deletent

Figure 6.4 Model for TnpX-mediated insertion

The circular intermediate (blue circle) has one TnpX dimer (red) bound to the joint (A). A second TnpX dimer (green) is then recruited (B). This complex encounters suitable target DNA (black line) either due to binding of host DNA binding proteins (purple) resulting in an alteration in the conformation of the DNA, by interaction with host DNA binding proteins (pink) directly or by direct interaction with the target DNA. Synapsis (C), followed by strand exchange then occurs, resulting in integration of the circular intermediate into the target DNA and the formation of the linear transposon form (D).

DNA binding proteins. These proteins may serve to increase the affinity of the circular intermediate-TnpX complex for the insertion site DNA. The increase in affinity may be dependent on host protein-mediated DNA bending or there may be a direct interaction between the host protein and the TnpX complex. Synapsis would occur after binding of the circular intermediate joint-TnpX complex to the insertion site DNA, followed by strand exchange and insertion of the transposon into the host DNA (Fig 6.4).

Although there is some evidence to support this model, a great deal of work remains to be done before the insertion reaction is properly understood. The complexes formed when TnpX was incubated with circular intermediate DNA were different to those formed when either the left or right end fragments were used (Chapter 4). The exact nature of these complexes is unknown and it is not known if the circular intermediate binds four molecules of TnpX. However, DNaseI footprinting studies showed that TnpX protected mainly transposon DNA at both the left and right ends. Since the circular intermediate joint contains both of these TnpX binding sites it is possible that there are two binding sites within this region. Determination of the constituents of the bound complexes produced during gel mobility shift analysis of the circular intermediate DNA fragment may provide evidence for the concentration dependent formation of a DNA-TnpX tetramer complex.

There is a fundamental difference between the TnpX-mediated excision reaction and normal resolvase reactions. The binding site of small resolvase proteins, such as those encoded by the transposons Tn*3* and γδ, is asymmetrical (Alonso *et al.*, 1996). Similarly, DNaseI protection studies of the left and right ends of Tn*4451* have indicated that the TnpX binding site is also asymmetrical with respect to the GA dinucleotide where strand cleavage occurs. However, the binding sites for the Tn*3* and γδ resolvases are directly repeated within the same DNA molecule. The binding sites of TnpX are in

an inverted arrangement, as binding predominantly occurs within the transposon sequences. DNaseI protection studies carried out using the Int protein from ϕC31 have indicated that the binding sites are more symmetrical with respect to the dinucleotide cleavage site (Thorpe *et al.*, 2000). However, the ϕTP901-1 large resolvase system also appears to have asymmetrical minimal DNA sites that can still be used for recombination, although DNA binding studies have not been conducted on this system (Breüner *et al.*, 2001). It is not unexpected that the ϕC31 system utilizes symmetrical recombination sites as the phage genome is integrated into a specific chromosomal location unless this site is not available (Combes *et al.*, 2002), unlike the Tn*4451* system where integration is more random (Crellin and Rood, 1997). The mechanistic implications for the difference in binding site orientation are unclear but may reflect a difference in synapse formation between the two systems. There is some evidence for a difference between the transposon and phage-related systems as many of the phage-related large resolvase proteins have been shown to require additional phage-encoded proteins for efficient excision *in vivo* (Bibb and Hatfull, 2002; Breüner *et al.*, 1999).

The most important questions raised by this study are: (1) what are the constituents of the complexes formed when TnpX is bound to either the left or right ends of the transposon or the circular intermediate joint?; and (2) how can the affinity of TnpX for potential insertion sites be increased? The question of what multimeric form of TnpX is bound to each DNA target could be addressed by analytical ultracentrifugation in the presence of a circular intermediate joint DNA fragment that had been labelled using a fluorophore to allow distinction between absorbance derived from the protein and that of the DNA (Lee *et al.*, 2001). Alternatively, surface plasmon resonance could be used to determine the stoichiometry of the binding interaction (Sassoon *et al.*, 2001). The dependence of the insertion reaction on host-encoded DNA binding proteins could be

examined with the use of mutant *E. coli* strains to determine any effect on insertion frequency. Alternatively, purified proteins such as IHF or HU, which are known DNA binding proteins involved in many different kinds of recombination events, could be obtained for *in vitro* insertion assay optimization studies, which could include the excision deficient TnpX mutant $TnpX_{S66P}$. The role, if any, of TnpY could also be investigated in this system.

In summary, there are three major findings reported in this thesis. Firstly, it has been shown that TnpX is the only protein required to catalyze excision *in vitro*. Secondly, deletion studies and random mutation of the TnpX amino acid sequence has led to the assignment of particular TnpX functions to different regions of the TnpX structure, including a putative dimerization site and a proposed bipartite DNA binding domain. This analysis has led to the construction of a model for the structure of TnpX. Finally, the TnpX DNA binding sites have been identified and mapped. Binding studies have shown that the affinity of TnpX for the ends of the transposon was high whereas the affinity for the flanking DNA, i.e. the potential insertion sites, was very low. This information supports the TnpX excision model and has led to the conclusion that there is a significant mechanistic difference between the excision and insertion reactions that are catalyzed by TnpX.

REFERENCES

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

Abraham, L.J. and Rood, J.I. (1987) Identification of Tn*4451* and Tn*4452*, Chloramphenicol Resistnace Transposons from *Clostridium perfringens*. *J. Bacteriol.*, **169**, 1579-1584.

Abraham, L.J. and Rood, J.I. (1988) The *Clostridium perfringens* Chloramphenicol Resistance Transposon Tn*4451* Excises Precisely in *Escherichia coli*. *Plasmid*, **19**, 164-168.

Abraham, L.J., Wales, A.J. and Rood, J.I. (1985) Worldwide Distribution of the Conjugative *Clostridium perfringens* Tetracycline Resistant Plasmid pCW3. *Plasmid*, **14**, 37-46.

Allingham, J.S. and Haniford, D.B. (2002) Mechanisms of metal ion action in Tn*10* transposition. *J. Mol. Biol.*, **319**, 53-65.

Allingham, J.S., Pribil, P.A. and Haniford, D.B. (1999) All Three Residues of the Tn*10* Transposase DDE Catalytic Triad Function in Divalent Metal Ion Binding. *J. Mol. Biol.*, **289**, 1195-1206.

Alonso, J.C., Ayora, S., Canosa, I., Weise, F. and Rojo, F. (1995) Site-specific recombination in Gram-positive theta-replicating plasmids. *FEMS Microbiol. Lett.*, **142**, 1-10.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.

Aravind, L., Makarova, K.S. and Koonin, E.V. (2000) Survey and Summary Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417-3432.

Arnold, P.H., Blake, D.G., Grindley, N.D.F., Boocock, M.R. and Stark, W.M. (1999) Mutants of Tn*3* resolvase which do not require accessory binding sites for recombination activity. *EMBO J.*, **18**, 1407-1414.

Ausubel, F.M., Brent, R., Kingston, R.E. and Moore, D.D. (eds.) (1994) *Preparation of genomic DNA from bacteria*. Greene Publishing Associates and John Wiley and Sons, New York.

Bannam, T.L., Crellin, P.K. and Rood, J.I. (1995) Molecular genetics of the chloramphenicol-resistance transposon Tn*4451* from *Clostridium perfringens*: the TnpX site-specific recombinase excises a circular transposon molecule. *Mol. Microbiol.*, **16**, 535-551.

Bartolomé, B., Jubete, Y., Martínez, E. and Cruz, F.d.l. (1991) Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. *Gene*, **102**, 75-78.

Bass, K.A. and Hecht, D.W. (2002) Isolation and characterization of cLV25, a *Bacteroides fragilis* chromosomal transfer factor resembling multiple *Bacteroides* sp. mobilizable transposons. *J. Bacteriol.*, **184**, 1895-1904.

Benevides, J.M., Kukolj, G., Autexier, C., Aubrey, K.L., DuBow, M.S. and Jr., G.J.T. (1994) Secondary structure and integration of phage D108 Ner repressor with a 61 bp operator: evidence for altered protein and DNA structures in the complex. *Biochemistry*, **33**, 10701-10710.

Benjamin, H.W. and Cozzarelli, N.R. (1988) Isolation and characterization of the Tn*3* resolvase synaptic intermediate. *EMBO J.*, **7**, 1897-1905.

Benjamin, K.R., Abola, A.P., Kanaar, R. and Cozzarelli, N.R. (1996) Contributions of Supercoiling to Tn*3* Resolvase and Phage Mu Gin Site-Specific Recombination. *J. Mol. Biol.*, **256**, 50-65.

Bennett, P.M. (1999) Integrons and gene cassettes: a genetic construction kit for bacteria. *J. Antimicrob. Chemother.*, **43**, 1-4.

Berg, D.E. (1989) Transposon Tn*5*. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 185-210.

Berryman, D.I., Lyristis, M. and Rood, J.I. (1994) Cloning and sequence analysis of *ermQ*, the predominant macrolide-lincosamide-streptogramin B resistance gene in *Clostridium perfringens*. *Antimicrob. Agents Chemother.*, **38**, 1041-1046.

Berryman, D.I. and Rood, J.I. (1995) The closely related *ermB/AM* genes from *Clostridium perfringens*, *Enterococcus faecalis* (pAMβ1), and *Streptococcus agalactiae* (pIP501), are flanked by variants of a directly repeated sequence. *Antimicrob. Agents Chemother.*, **39**, 1830-1834.

Bhasin, A., Goryshin, I.Y. and Reznikoff, W.S. (1999) Hairpin Formation in Tn*5* Transposition. *J. Biol. Chem.*, **274**, 37021-37029.

Bibb, L.A. and Hatfull, G.F. (2002) Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, ϕRv1. *Mol. Microbiol.*, **45**, 1515-1526.

Blakely, G.W., Davidson, A.O. and Sherratt, D.J. (1997) Binding and Cleavage of Nicked Substrates by Site-specific Recombinases XerC and XerD. *J. Mol. Biol.*, **265**, 30-39.

Brefort, G., Magot, M., Ionesco, H. and Sebald, M. (1977) Characterization and transferability of *Clostridium perfringens* plasmids. *Plasmid*, **1**, 52-66.

Breüner, A., Brøndsted, L. and Hammer, K. (1999) Novel Organization of Genes Involved in Prophage Excision Identified in the Temperate Lactococcal Bacteriophage TP901-1. *J. Bacteriol.*, **181**, 7291-7297.

Breüner, A., Brøndsted, L. and Hammer, K. (2001) Resolvase-like recombination performed by the TP901-1 integrase. *Microbiology*, **147**, 2051-2063.

Brown, H.J., Stokes, H.W. and Hall, R.M. (1996) The Integrons In0, In2 and In5 Are Defective Transposon Derivatives. *J. Bacteriol.*, **178**, 4429-4437.

Brynestad, S. and Granum, P.E. (1999) Evidence that Tn*5565*, which includes the enterotoxin gene in *Clostridium perfringens*, can have a circular form which may be a transposition intermediate. *FEMS Microbiol. Lett.*, **170**, 281-286.

Brynestad, S., Synstad, B. and Granum, P.E. (1997) The *Clostridium perfringens* enterotoxin gene is on a transposable element in type A human food poisoning strains. *Microbiology*, **143**, 2109-2115.

Buchanan, S.G.S.C. and Gay, N.J. (1996) Structural and functional diversity in the leucine-rich repeat family of proteins. *Prog. Biophys. Mol. Biol.*, **65**, 1-44.

Canosa, I., Ayora, S., Rojo, F. and Alonso, J.C. (1997) Mutational analysis of a site-specific recombinase: characterization of the catalytic and dimerization domains of the β recombinase of pSM19035. *Mol. Gen. Genet.*, **255**, 467-476.

Carrasco, C.D., Ramaswamy, K.S., Ramasubramanian, T.S. and Golden, J.W. (1994) *Anabaena xisF* gene encodes a developmentally regulated site-specific recombinase. *Genes Dev.*, **8**, 74-83.

Caspers, P., Dalrymple, B., Iida, S. and Arber, W. (1984) IS*30*, a new insertion sequence of *Escherichia coli* K12. *Mol. Gen. Genet.*, **196**, 68-73.

Chaconas, G., Lavoie, B.D. and Watson, M.A. (1996) DNA transposition: jumping gene machine, some assembly required. *Curr. Biol.*, **6**, 817-820.

Cheng, Q., Paszkiet, B.J., Shoemaker, N.B., Gardner, J.F. and Salyers, A.A. (2000) Integration of a *Bacteroides* Conjugative Transposon CTnDOT. *J. Bacteriol.*, **182**, 4035-4043.

Cho, E.H., Jr, R.A., Gumport, R.I. and Gardner, J.F. (2000) Characterization of Bacteriophage Lambda Excisionase Mutants Defective in DNA Binding. *J. Bacteriol.*, **182**, 5807-5812.

Chritiansen, B., Brøndsted, L., Vogensen, F.K. and Hammer, K. (1996) A Resolvase-Like Protein is Required for the Site-Specific Integration of the Temperate Lactococcal Bacteriophage TP901-1. *J. Bacteriol.*, **178**, 5164-5173.

Clewell, D.B., Susan E, F. and Jaworski, D.D. (1995) Unconstrained bacterial promiscuity: the Tn*916*-Tn*1545* family of conjugative transposons. *Trends Microbiol.*, **3**, 229-236.

Combes, P., Till, R., Bee, S. and Smith, M.C.M. (2002) The *Streptomyces* genome contains multiple pseudo-*attB* sites for the ϕC31-encoded site-specific recombination system. *J. Bacteriol.*, **184**, 5746-5752.

Craig, N.L. (1989) Transposon Tn*7*. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 211-226.

Craig, N.L. (1996) Transposition. In Neidhardt, F.C., III, R.C., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (eds.), *Escherichia coli and Salmonella Cellular and Molecular Biology*. ASM Press, Washington DC, Vol. 2, pp. 2339-2362.

Craigie, R. (1995) Resolving a Resolvase. *Struct. Biol.*, **2**, 607-609.

Crellin, P.K. (1997) Functional analysis of the clostridial transposon Tn*4451*. *Microbiology*. Monash University, Melbourne.

Crellin, P.K. and Rood, J.I. (1997) The Resolvase/Invertase Domain of the Site-Specific Recombinase TnpX Is Functional and Recognizes a Target Sequence That Resembles the Junction of the Circular Form of the *Clostridium perfringens* Transposon Tn*4451*. *J. Bacteriol.*, **179**, 5148-5156.

Crellin, P.K. and Rood, J.I. (1998) Tn*4451* from *Clostridium perfringens* is a mobilizable transposon that encodes the functional Mob protein, TnpZ. *Mol. Microbiol.*, **27**, 631-642.

Crisona, N.J., Weinberg, R.L., Peter, B.J., Sumners, D.W. and Cozzarelli, N.R. (1999) The Topological Mechanism of Phage λ Integrase. *J. Mol. Biol.*, **289**, 747-775.

Ellemor, D.M., Baird, R.N., Awad, M.M., Boyd, R.L., Rood, J.I. and Emmins, J.J. (1999) Use of genetically manipulated strains of *Clostridium perfringens* reveals that both alpha-toxin and theta-toxin are required for vascular leukostasis to occur in experimental gas gangrene. *Infect. Immun.*, **67**, 4902-4907.

Farrow, K.A., Lyras, D. and Rood, J.I. (2000) The Macrolide-Lincosamide-Streptogramin B resistance determinant from *Clostrdium difficile* 630 contains two *erm*(B) genes. *Antimicrob. Agents Chemother.*, **44**, 411-413.

Farrow, K.A., Lyras, D. and Rood, J.I. (2001) Genomic analysis of the erythromycin resistance element Tn*5398* from *Clostridium difficile*. *Microbiology*, **147**, 2717-2728.

Garnier, T. and Cole, S.T. (1986) Characterization of a bacteriocinogenic plasmid from *Clostridium perfringens* and molecular genetic analysis of the bacteriocin-encoding gene. *J. Bacteriol.*, **168**, 1189-1196.

Glasgow, A.C., Hughes, K.T. and Simon, M.I. (1989) Bacterial DNA inversion systems. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington D. C., pp. 637-659.

Golden, J.W., Carrasco, C.D., Mulligan, M.E., Schneider, G.J. and Haselkorn, R. (1988) Deletion of a 55-Kilobase-Pair DNA Element from the Chromosome during Heterocyst Differentiation of *Anabaena* sp. Strain PCC 7120. *J. Bacteriol.*, **170**, 5034-5041.

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

Gopaul, D.N. and Duyne, G.D.V. (1999) Structure and mechanism in site-specific recombination. *Curr. Opin. Struct. Biol.*, **9**, 14-20.

Grainge, I. and Jayaram, M. (1999) The integrase family of recombinases: organization and function of the active site. *Mol. Microbiol.*, **33**, 449-456.

Grindley, N.D.F. and Leschziner, A.E. (1995) DNA Transposition: From Black Box to a Colour Monitor. *Cell*, **83**, 1063-1066.

Groth, A.C., Olivares, E.C., Thyagarajan, B. and Calos, M.P. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 5995-6000.

Halford, S.E., Jordan, S.L. and Kirkbride, E.A. (1985) The resolvase protein from the transposon Tn*21*. *Mol. Gen. Genet.*, **200**, 169-175.

Hall, R.M. and Collis, C.M. (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.*, **15**, 593-600.

Hallet, B. and Sherratt, D.J. (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.*, **21**, 157-178.

Haren, L., Bétermier, M., Pollard, P. and Chandler, M. (1997) IS*911*-mediated intramolecular transposition is naturally temperature sensitive. *Mol. Microbiol.*, **25**, 531-540.

Haren, L., Normand, C., Polard, P., Alazard, R. and Chandler, M. (2000) IS*911* Transposition is Regulated by Protein-Protein Interactions via a Leucine Zipper Motif. *J. Mol. Biol.*, **296**, 757-768.

Haren, L., Ton-Hoang, B. and Chandler, M. (1999) Integrating DNA: Transposases and Retroviral Integrases. *Annu. Rev. Microbiol.*, **53**, 245-281.

Hatfull, G.F., Noble, S.M. and Grindley, N.D.F. (1987) The γδ resolvase induces an unusual DNA structure at the recombinational crossover point. *Cell*, **49**, 103-110.

Haykinson, M.J., Johnson, L.M., Soong, J. and Johnson, R.C. (1996) The Hin dimer interface is critical for Fis-mediated activation of the catalytic steps of site-specific DNA inversion. *Curr. Biol.*, **6**, 163-177.

Henderson, D.J., Lydiate, D.J. and Hopwood, D.A. (1989) Structural and functional analysis of the mini-circle, a transposable element of *Streptomyces coelicolor* A3(2). *Mol. Microbiol.*, **3**, 1307-1318.

Hughes, R.E., Hatfull, G.F., Rice, P., Steitz, T.A. and Grindley, N.D.F. (1990) Cooperativity mutants of the γδ resolvase identify an essential interdimer interaction. *Cell*, **63**, 1331-1338.

Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23-28.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C. and Hiramatsu, K. (2001) Structural comparison of three types of Staphylococcal Cassette Chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, **45**, 1323-1336.

Ito, T., Katayama, Y. and Hiramatsu, K. (1999) Cloning and Nucleotide Sequence Determination of the Entire *mec* DNA of Pre-Methicillin-Resistant *Staphylococcus aureus* N315. *Antimicrob. Agents Chemother.*, **43**, 1449-1458.

Janière, L., McGovern, S., Pujol, C., Petit, M.-A. and Ehrlich, S.D. (1996) *In vivo* analysis of the plasmid pAMβ1 resolution system. *Nucleic Acids Res.*, **24**, 3431-3436.

Jia, Y. and Churchward, G. (1999) Interactions of the Integrase Protein of the Conjugative Transposon Tn*916* with Its Specific DNA Binding Sites. *J. Bacteriol.*, **181**, 6114-6123.

Johanesen, P.A., Lyras, D. and Rood, J.I. (2001) Induction of pCW3-encoded tetracycline resistance in *Clostridium perfringens* involves a host factor. *Plasmid*, **46**, 229-232.

Johnson, S.R. and Romig, W.R. (1981) *Vibrio cholerae* conjugative plasmid pSJ15 contains transposable prophage dVcA1. *J. Bacteriol.*, **146**, 632-638.

Katayama, Y., Ito, T. and Hiramatsu, K. (2000) A New Class of Genetic Element, Cassette Chromosome *mec*, Encodes Methicillin Resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, **44**, 1549-1555.

Kilbride, E., Boocock, M.R. and Stark, W.M. (1999) Topological Selectivity of a Hybrid Site-specific Recombination System with Elements from Tn*3 res*/ resolvase and Bacteriophage P1 *loxP*/Cre. *J. Mol. Biol.*, **289**, 1219-1230.

Kiss, J. and Olasz, F. (1999) Formation and transposition of the covalently closed IS*30* circle: the relation between tandem dimers and monomeric circles. *Mol. Microbiol.*, **34**, 37-52.

Kleckner, N. (1989) Transposon Tn*10*. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 227-268.

Kuhstoss, S. and Rao, R.N. (1991) Analysis of the Integration Function of the Streptomycete Bacteriophage ϕC31. *J. Mol. Biol.*, **222**, 897-908.

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

Laity, J.H., Lee, B.M. and Wright, P.E. (2001) Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.*, **11**, 39-46.

Lawrence, G.W. (1997) The pathogenesis of enteritis necroticans. In Rood, J.I., McClane, B.A., Songer, J.G. and Titball, R.W. (eds.), *The Clostridia Molecular Biology and Pathogenesis*. Academic Press, California, pp. 197-210.

Lee, S.P., Fuior, E., Lewis, M.S. and Han, M.K. (2001) Analytical ultracentrifugation studies of Translin: analysis of protein-DNA interactions using a single-stranded fluorogenic oligonucleotide. *Biochemistry*, **40**, 14081-14088.

Liebert, C.A., Hall, R.M. and Summers, A.O. (1999) Transposon Tn*21*, Flagship of the Floating Genome. *Microbiol. Mol. Biol. Rev.*, **63**, 507-522.

Liu, C.-C., Hühne, R., Tu, J., Lorbach, E. and Dröge, P. (1998) The resolvase encoded by the *Xanthomonas campestris* transposable element IS*Xc5* constitutes a new subfamily closely related to DNA invertases. *Genes Cells*, **3**, 221-233.

Lupas, A., Dyke, M.V. and Stock, J. (1991) Predicting coiled-coils from protein sequences. *Science*, **252**, 1162-1164.

Lyras, D. and Rood, J.I. (1997) Transposable Genetic Elements and Antibiotic Resistance Determinants from *Clostridium perfringens* and *Clostridium difficile*. In Rood, J.I., McClane, B.A., Songer, J.G. and Titball, R.W. (eds.), *The Clostridia: Molecular Biology and Pathogenesis*. Academic Press, San Diego, California, pp. 73-92.

Lyras, D. and Rood, J.I. (2000a) Clostridial Genetics. In Fischetti, V.A., Novick, R.P., Ferretti, J.J., Portnoy, D.A. and Rood, J.I. (eds.), *Gram-Positive Pathogens*. ASM Press, Washington DC, pp. 529-539.

Lyras, D. and Rood, J.I. (2000b) Transposition of Tn*4451* and Tn*4453* involves a circular intermediate that forms a promoter for the large resolvase, TnpX. *Mol. Microbiol.*, **38**, 588-601.

Lyras, D., Storie, C., Huggins, A.S., Crellin, P.K., Bannam, T.L. and Rood, J.I. (1998) Chloramphenicol Resistance in *Clostridium difficile* Is Encoded on Tn*4453*

Transposons That Are Closely Related to Tn*4451* from *Clostridium perfringens*. *Antimicrob. Agents Chemother.*, **42**, 1563-1567.

Mahillon, J. and Chandler, M. (1998) Insertion Sequences. *Microbiol. Mol. Biol. Rev.*, **62**, 725-774.

Mahillon, J., Leonard, C. and Chandler, M. (1999) IS elements as constituents of bacterial genomes. *Res. Microbiol.*, **150**, 675-687.

Matsuura, M., Noguchi, T., Aida, T., Asayama, M., Takahashi, H. and Shirai, M. (1995) A gene essential for the site-specific excision of actinophage R4 prophage genome from the chromosome of a lysogen. *J. Gen. Appl. Microbiol.*, **41**, 53-61.

Matsuura, M., Noguchi, T., Yamaguchi, D., Aida, T., Asayama, M., Takahashi, H. and Shirai, M. (1996) The *sre* Gene (ORF469) Encodes a Site-Specific Recombinase Responsible for Integration of the R4 Phage Genome. *J. Bacteriol.*, **178**, 3374-3376.

McClane, B.A., Lyerly, D., Moncrief, J.S. and Wilkins, T.D. (2000) Enterotoxic Clostridia: *Clostridium perfringens* Type A and *Clostridium difficile*. In Fischetti, V.A., Novick, R.P., Ferretti, J.J., Portnoy, D.A. and Rood, J.I. (eds.), *Gram-Positive Pathogens*. ASM Press, Washington D. C., pp. 551-562.

McDonnell, J.M. (2001) Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Curr. Opin. Chem. Biol.*, **5**, 572-577.

Mendiola, M.V., Bernales, I. and Cruz, F.d.l. (1994) Differential roles of the transposon termini in IS*91* transposition. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 1922-1926.

Meselson, M. and Yuan, R. (1968) DNA restriction enzyme from *E. coli*. *Nature*, **217**, 1110-1114.

Michiels, J., Xi, C., Verhaert, J. and Vanderleyden, J. (2002) The functions of Ca^{2+} in bacteria: a role for EF-hand proteins? *Trends Microbiol.*, **10**, 87-93.

Miroux, B. and Walker, J.E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.*, **260**, 289-298.

Mizuuchi, K. (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells*, **2**, 1-12.

Morelle, G. (1989) A plasmid extraction procedure on a miniprep scale. *Focus*, **11**, 7-8.

Morgan, G.J., Hatfull, G.F., Casjens, S. and Hendrix, R.W. (2002) Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus, Neisseria* and *Deinococcus. J. Mol. Biol.*, **317**, 337-359.

Mullany, P., Wilks, M., Lamb, I., Clayton, C., Wren, B. and Tabaqchali, S. (1990) Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis. J. Gen. Microbiol.*, **136**, 1343-1349.

Murley, L.L. and Grindley, N.D.F. (1998) Architecture of the γδ Resolvase Synaptosome: Oriented Heterodimers Identify Interactions Essential for Synapsis and Recombination. *Cell*, **95**, 553-562.

Murphy, E. (1989) Transposable Elements in Gram-Positive Bacteria. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 269-288.

Murphy, E., Reinheimer, E. and Huwyler, L. (1991) Mutational analysis of *att554*, the target of the site-specific transposon Tn*554*. *Plasmid*, **26**, 20-29.

Nanassy, O.Z. and Hughes, K.T. (1998) *In Vivo* Identification of Intermediate Stages of the DNA Inversion Reaction Catalysed by the Salmonella Hin Recombinase. *Genetics*, **149**.

Nanassy, O.Z. and Hughes, K.T. (2001) Hin Recombinase Mutants Functionally Disrupted in Interactions with Fis. *J. Bacteriol.*, **183**, 28-35.

Nash, H.A. (1996) Site-Specific Recombination: Integration, Excision, Resolution, and Inversion of Defined DNA Segments. In Neidhardt, F.C., III, R.C., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (eds.), *Escherichia coli and Salmonella Cellular and Molecular Biology*. ASM Press, Washington DC, Vol. 2, pp. 2363-2376.

Newman, B.J. and Grindley, N.D.F. (1984) Mutants of the γδ resolvase: a genetic analysis of the recombination function. *Cell*, **38**, 463-469.

Nunes-Düby, S.E., Kwon, H.J., Tirumalai, R.S., Ellenberger, T. and Landy, A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.*, **26**, 391-406.

Olasz, F., Stalder, R. and Arber, W. (1993) Formation of the tandem repeat (IS*30*)$_2$ and its role in IS*30*-mediated transpositional DNA rearrangements. *Mol. Gen. Genet.*, **239**, 177-187.

Oram, M., Szczelkun, M.D. and Halford, S.E. (1995) Pieces of the site-specific recombination puzzle. *Curr. Biol.*, **5**, 1106-1109.

Palombo, E.A., Yusoff, K., Stanisich, V.A., Krishnapillai, V. and Willetts, N.S. (1989) Cloning and genetic analysis of *tra* cistrons of the Tra2/Tra3 region of plasmid RP1. *Plasmid*, **22**, 59-69.

Pan, B., Maciejewski, M.W., Marintchev, A. and Mullen, G.P. (2001) Solution structure of the catalytic domain of γδ resolvase. Implications for mechanism of catalysis. *J. Mol. Biol.*, **310**, 1089-1107.

Pato, M.L. (1989) Bacteriophage Mu. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 23-52.

Perkins-Balding, D., Duval-Valentin, G. and Glasgow, A.C. (1999) Excision of IS*492* Requires Flanking Target Sequences and Results in Circle Formation in *Pseudoalteromonas atlantica*. *J. Bacteriol.*, **181**, 4937-4948.

Peters, J.E. and Craig, N.L. (2001a) Tn*7* recognizes transposition target structures associated with the DNA-binding protein TnsE. *Genes Dev.*, **15**, 737-747.

Peters, J.E. and Craig, N.L. (2001b) Tn*7*: smarter than we thought. *Nat Rev Mol Cell Biol*, **2**, 806-814.

Petit, M.-A., Ehrlich, D. and Jannière, L. (1995) pAMβ1 resolvase has an atypical recombination site and requires the histone-like protein HU. *Mol. Microbiol.*, **18**, 271-282.

Phillips, G.J. (1999) New cloning vectors with temperature-sensitive replication. *Plasmid*, **41**, 78-81.

Polard, P., Seroude, L., Fayet, O., Prère, M.-F. and Chandler, M. (1994) One-Ended Insertion of IS*911*. *J. Bacteriol.*, **176**, 1192-1196.

Pribil, P.A. and Haniford, D.B. (2000) Substrate Recognition and Induced DNA Deformation by Transposase at the Target-capture Stage of Tn*10* Transposition. *J. Mol. Biol.*, **303**, 145-159.

Ramaswamy, K.S., Carrasco, C.D., Fatma, T. and Golden, J.W. (1997) Cell-type specificity of the *Anbaena fdxN*-element rearrangement requires *xisH* and *xisI*. *Mol. Microbiol.*, **23**, 1241-1249.

Rao, J.E. and Craig, N.L. (2001) Selective recognition of pyrimidine motif triplexes by a protein encoded by the bacterial transposon Tn*7*. *J. Mol. Biol.*, **307**, 1161-1170.

Recchia, G.D. and Hall, R.M. (1995) Gene cassettes: a new class of mobile element. *Microbiology*, **141**, 3015-3027.

Reimmann, C. and Haas, D. (1990) The *isiA* gene of insertion sequence IS*21* is essential for cleavage at the inner 3' ends of tandemly repeated IS*21* elements *in vitro*. *EMBO J.*, **9**, 4055-4063.

Reimmann, C., Moore, R., Little, S., Savioz, A., Willets, N.S. and Haas, D. (1989) Genetic structure, function and regulation of the transposable element IS*21*. *Mol. Gen. Genet.*, **215**, 416-424.

Reznikoff, W.S., Bhasin, A., Davies, D.R., Goryshin, I.Y., Mahnke, L.A., Naumann, T., Rayment, I., Steiniger-White, M. and Twining, S.S. (1999) Tn*5*: A Molecular Window on Transposition. *Biochem. Biophys. Res. Commun.*, **266**, 729-734.

Rice, L.B. (1998) Tn*916* Family Conjugative Transposons and Dissemination of Antimicrobial Resistance Determinants. *Antimicrob. Agents Chemother.*, **42**, 1871-1877.

Rice, P.A. and Steitz, T.A. (1994) Model for a DNA-mediated synaptic complex suggested by crystal packing of γδ resolvase subunits. *EMBO J.*, **13**, 1514-1524.

Roberts, A.P., Braun, V., Eichel-Streiber, C.v. and Mullany, P. (2001a) Demonstration that the Group II Intron from the Clostridial Conjugative Transposon Tn*5397* Undergoes Splicing In Vivo. *J. Bacteriol.*, **183**, 1296-1299.

Roberts, A.P., Johanesen, P.A., Lyras, D., Mullany, P. and Rood, J.I. (2001b) Comparison of Tn*5397* from *Clostridium difficile*, Tn*916* from *Enterococcus faecalis* and the CW459*tet*(M) element from *Clostridium perfringens* shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology*, **147**, 1243-1251.

Roberts, A.P., Pratten, J., Wilson, M. and Mullany, P. (1999) Transfer of a conjugative transposon, Tn*5397* in a model oral biofilm. *FEMS Microbiol. Lett.*, **177**, 63-66.

Rood, J.I. (1998) Virulence Genes of *Clostridium perfringens*. *Annu. Rev. Microbiol.*, **52**, 333-360.

Rowe-Magnus, D.A., Guérout, A.-M. and Mazel, D. (1999) Super-integrons. *Res. Microbiol.*, **150**, 641-651.

Rudy, C., Taylor, K.L., Hinerfeld, D., Scott, J.R. and Churchward, G. (1997) Excision of a conjugative transposon *in vitro* by the Int and Xis proteins of Tn*916*. *Nucleic Acids Res.*, **25**, 4061-4066.

Salyers, A.A., Shoemaker, N.B. and Li, L.-Y. (1995a) In the Driver's seat: The *Bacteroides* Conjugative Transposons and the Elements they Mobilize. *J. Bacteriol.*, **177**, 5727-5731.

Salyers, A.A., Shoemaker, N.B., Stevens, A.M. and Li, L.-Y. (1995b) Conjugative Transposons: an Unusual and Diverse Set of Integrated Gene Transfer Elements. *Microbiol. Rev.*, **59**, 579-590.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a laboratory manual.* Cold Spring Harbour Laboratory Press, Cold Springs Habour, New York.

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

Sassoon, J., Lilie, H., Baumann, U. and Kohli, J. (2001) Biochemical characterization of the structure-specific DNA-binding protein Cmb1 from *Schizosaccharomyces pombe*. *J. Mol. Biol.*, **309**, 1101-1115.

Sato, T., Harada, K. and Kobayashi, Y. (1996) Analysis of Suppressor Mutations of *spoIVCA* Mutations: Occurrence of DNA Rearrangements in the Absence of a Site-Specific DNA Recombinase SpoIVA in *Bacillus subtilis*. *J. Bacteriol.*, **178**, 3380-3383.

Sato, T., Harada, K., Ohta, Y. and Kobayashi, Y. (1994) Expression of the *Bacillus subtilis spoIVA* Gene, Which Encodes a Site-Specific Recombinase, Depends on the *spoIIGB* Product. *J. Bacteriol.*, **176**, 935-937.

Sato, T., Samori, Y. and Kobayashi, Y. (1990) The *cisA* Cistron of *Bacillus subtilis* Sporulation Gene *spoIVC* Encodes a Protein Homologous to a Site-Specific Recombinase. *J. Bacteriol.*, **172**, 1092-1098.

Schmid, S., Seitz, T. and Haas, D. (1998) Cointegrase a Naturally Occurring, Truncated Form of IS*21* Transposase, Catalyses Replicon Fusion rather than Simple Insertion of IS*21*. *J. Mol. Biol.*, **282**, 571-583.

Schneider, K. and Beck, C.F. (1986) Promoter-probe vectors for the analysis of divergently arranged promoters. *Gene*, **42**, 37-48.

Scott, J.R. and Churchward, G.G. (1995) Conjugative Transposition. *Annu. Rev. Microbiol.*, **49**.

Sekine, Y., Eisaki, N. and Ohtsubo, E. (1996) Identification and Characterization of the Linear IS*3* Molecules Generated by Staggered Breaks. *J. Biol. Chem.*, **271**, 197-202.

Sherratt, D. (1989) Tn*3* and Related Transposable Elements: Site-Specific Recombination and Transposition. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington DC, pp. 163-184.

Shiga, Y., Sekine, Y. and Ohtsubo, E. (1999) Transposition of IS*1* circles. *Genes Cells*, **4**, 551-561.

Shoemaker, N.B., Wang, G.-R. and Salyers, A.A. (2000) Multiple Gene Products and Sequences Required for Excision of the Mobilizable Integrated *Bacteroides* Element NBU1. *J. Bacteriol.*, **182**, 928-936.

Sloan, J., McMurry, L.M., Lyras, D., Levy, S.B. and Rood, J.I. (1994) The *Clostridium perfringens* Tet P determinant comprises two overlapping genes: *tetA*(P), which mediates active tetracycline efflux, and *tetB*(P), which is related to the ribosomal protection family of tetracycline-resistance determinants. *Mol. Microbiol.*, **11**, 403-415.

Smith, C.J., Tribble, G.D. and Bayley, D.P. (1998) Genetic Elements of *Bacteroides* Species: A [illegible]ving Story. *Plasmid*, **40**, 12-29.

Smith, M., Jessee, [illegible]anders, T. and Jordan, J. (1990) High efficiency bacterial electroporation: 1 x 10^{10} *E. coli* transformants/mg. *Focus*, **12**, 38-40.

Smith, M.C.M., Burns, R.N., Wilson, S.E. and Gregory, M.A. (1999) The complete genome sequence of the *Streptomyces* temperate phage ϕC31: evolutionary relationships to other viruses. *Nucleic Acids Res.*, **27**, 2145-2155.

Smith, M.C.M. and Thorpe, H.M. (2002) Diversity in the Serine Recombinases. *Mol. Microbiol.*, **44**, 299-307.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments seperated by gel electrophoresis. *J. Mol. Biol.*, **98**, 503-517.

Stackebrandt, E. and Rainey, F.A. (1997) Phylogenetic Relationships. In Rood, J.I., McClane, B.A., Songer, J.G. and Titball, R.W. (eds.), *The Clostridia Molecular Biology and Pathogenesis*. Academic Press, California, pp. 3-20.

Stark, W.M. and Boocock, M.R. (1995) Topological selectivity in site-specific recombination. In Sherratt, D.J. (ed.) *Mobile Genetic Elements*. Oxford University Press, Oxford, pp. 101-129.

Stark, W.M., Boocock, M.R. and Sherratt, D.J. (1992) Catalysis by site-specific recombinases. *Trends Genet.*, **8**, 432-439.

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

Stellwagen, A.E. and Craig, N.L. (2001) Analysis of gain-of-function mutants of an ATP-dependent regulator of Tn7 transposition. *J. Mol. Biol.*, **305**.

Stoll, S.M., Ginsburg, D.S. and Calos, M.P. (2002) Phage TP901-1 site-specific integrase functions in human cells. *J. Bacteriol.*, **184**, 3657-3663.

Stragier, P., Kunkel, B., Kroos, L. and Losick, R. (1989) Chromosomal rearrangement generating a composite gene for a developmental transcription factor. *Science*, **243**, 507-512.

Takemaru, K.-i., Mizumo, M., Sato, T., Takeuchi, M. and Kobayashi, Y. (1995) Complete nucleotide sequence of a *skin* element excised by DNA rearrangement during sporulation in *Bacillus subtilis*. *Microbiology*, **141**, 323-327.

Tan, H.-M. (1999) Bacterial catabolic transposons. *Appl. Microbiol. Biotechnol.*, **51**, 1-12.

Tavakoli, N., Comanducci, A., Dodd, H.M., Lett, M.-C., Albiger, B. and Bennett, P. (2000) IS1294, a DNA Element That Transposes by RC Transposition. *Plasmid*, **44**, 66-84.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.

Thorpe, H.M. and Smith, M.C.M. (1998) *In vitro* site-specific integration of bacteriophage DNA catalysed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 5505-5510.

Thorpe, H.M., Wilson, S.E. and Smith, M.C.M. (2000) Control of directionality in the site-specific recombination system of the *Streptomyces* phage ϕC31. *Mol. Microbiol.*, **38**, 232-241.

Tobiason, D.M., Buchner, J.M., Thiel, W.H., Gernert, K.M. and Karls, A.C.G. (2001) Conserved amino acid motifs from the novel Piv/MooV family of transposases and site-specific recombinases are required for catalysis of DNA inversion. *Mol. Microbiol.*, **39**, 641-651.

Ton-Hoang, B., Bétemier, M., Polard, P. and Chandler, M. (1997) Assembly of a strong promoter following IS*911* circularization and the role of circles in transposition. *EMBO J.*, **16**, 3357-3371.

Ton-Hoang, B., Polard, P., Haren, L., Turlan, C. and Chandler, M. (1999) IS*911* transposon circles give rise to linear forms that can undergo integration *in vitro*. *Mol. Microbiol.*, **32**, 617-627.

Tribble, G.D., Parker, A.C. and Smith, C.J. (1997) The *Bacteroides* Mobilizable Transposon Tn*4555* Integrates by a Site-Specific Recombination Mechanism Similar to That of the Gram-Positive Bacterial Element Tn*916*. *J. Bacteriol.*, **179**, 2731-2739.

Tribble, G.D., Parker, A.C. and Smith, C.J. (1999a) Genetic Structure and Transcriptional Analysis of a Mobilizable, Antibiotic Resistance Transposon from *Bacteroides*. *Plasmid*, **42**, 1-12.

Tribble, G.D., Parker, A.C. and Smith, C.J. (1999b) Transposition genes of the *Bacteroides* mobilizable transposon Tn*4555*: role of a novel targeting gene. *Mol. Microbiol.*, **34**, 385-394.

Turlan, C. and Chandler, M. (1995) IS*1*-mediated intramolecular rearrangements: formation of excised transposon circles and replicative deletions. *EMBO J.*, **14**, 5410-5421.

Turlan, C. and Chandler, M. (2000) Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.*, **8**, 268-274.

Turlan, C., Ton-Hoang, B. and Chandler, M. (2000) The role of tandem IS dimers in IS*911* transposition. *Mol. Microbiol.*, **35**, 1312-1325.

Vendantam, G., Novicki, T.J. and Hecht, D.W. (1999) *Bacteroides fragilis* Transfer Factor Tn*5520*: the smallest Bacterial Mobilizable Transposon Containing

Single Integrase and Mobilization Genes That Function in *Escherichia coli*. *J. Bacteriol.*, **181**, 2564-2571.

Wang, H. and Mullany, P. (2000) The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn*5397*. *J. Bacteriol.*, **182**, 6577-6583.

Wang, H., Roberts, A.P., Lyras, D., Rood, J.I., Wilks, M. and Mullany, P. (2000a) Characterization of the Ends and Target Sites of the Novel Conjugative Transposon Tn*5397* from *Clostridium difficile*: Excision and Circularization Is Mediated by the Large Resolvase, TndX. *J. Bacteriol.*, **182**, 3775-3783.

Wang, J., Shoemaker, N.B., Wang, G.-R. and Salyers, A.A. (2000b) Characterization of a *Bacteroides* Mobilizable Transposon, NBU2, Which Carries a Functional Lincomycin Resistance Gene. *J. Bacteriol.*, **182**, 3559-3571.

Wang, R.F. and Kushner, S.R. (1991) Construction of versatile low-copy number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene*, **100**, 195-199.

Waters, V.L. (1999) Conjugative Transfer in the Dissemination of Beta-Lactam and Aminoglycoside Resistance. *Front. Biosci.*, **4**, 416-439.

Watson, M.A., Boocock, M.R. and Stark, W.M. (1996) Rate and Selectivity of Synapsis of *res* Recombination Sites by Tn*3* Resolvase. *J. Mol. Biol.*, **257**, 317-329.

Yang, H.-Y., Kim, Y.-W. and Chang, H.-I. (2002) Construction of an integration-proficient vector based on the site-specific recombination mechanism of enterococcal temperate phage ϕFC1. *J. Bacteriol.*, **184**, 1859-1864.

Yang, W. and Steitz, T.A. (1995) Crystal Structure of the site-specific recombinase γδ resolvase complexed with a 34 bp cleavage site. *Cell*, **82**, 193-207.

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