Transposition of ISY100

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

by

Xiaofeng Feng

Division of Molecular Genetics

Institute of Biomedical and Life Science

University of Glasgow

August 2006

The research presented here, except where otherwise stated, is my own original work, and has not been submitted for any other degree.

Acknowledgements

Firstly I would like to thank Sean, my supervisor, for guidance, patience and support throughout these years, as well as the critical reading and the correction of this manuscript. Thanks to Amy and Nadia for their tolerance of my occupying so much of Sean's time that belongs to them. Thanks to Marshall, Mary, Martin and Sally from the Sixth floor for help on methods, advice, plasmids, strains, reagents and use of equipment. Thanks to Jiuya and his family for support and happy times. Thanks to Sarah, Elizabeth and Stephen for practical help. Thanks to Drs. M. Jackson and R. Sutcliffe, my assessors, for their comments and suggestions. Thanks to everyone on the first and seventh floors for letting me use PCR machines. Thanks to the Wellcome Trust for funding, and the Graduate School of IBLS at University of Glasgow for financial support.

Thanks to Drs. J. Clutterbuck, C. Mullineaux, B. Kan and F. Hayes for their kind gifts of plasmids and genomic DNA.

Thanks to my wife Yangjie, who gave up her career and familiar life in China and came here to support me, none of this would have been possible without her support.

This thesis is dedicated with all my love to my parents, my wife and my son.

Abstract

The insertion sequence ISY100 is a member of the IS630/Tc1/*mariner* superfamily of transposable elements and was first found in the cyanobacterium *Synechocystis* sp. PCC6803. ISY100 is active in *E. coli*, transposing into TA dinucleotides like other members of the IS630/Tc1/*mariner* superfamily (Urasaki, *et al.*, 2002).

An *in vivo* transposition assay was set up in this study, confirming the TA target preference of ISY100 and showing that 30 bp from each end of ISY100 is sufficient for efficient transposition. Single mutations in the proposed catalytic residues (D136, D218 and E253) in the C-terminal domain inactivated ISY100 transposase *in vivo*.

Adding 6 histidines to the C-terminus of the transposase did not change its transposition activity *in vivo*. His-tagged transposase was purified and catalysed efficient transposition *in vitro*. Purified transposase was the only protein required for transposition.

The *in vitro* system was used to study the different steps in the transposition reaction. Purified His-tagged transposase bound the transposon ends, protecting approximately 26 bp from cleavage by DNase I at each end. Two helix-turn-helix DNA binding motifs linked by an 'AT-hook'-like sequence were predicted in the N-terminal domain of ISY100 transposase. Deletion derivatives containing just these helix-turn-helix motifs bound efficiently to the transposon ends.

Analysis of *in vivo* cleavage products suggested that ISY100 transposase cleaves exactly at the transposon ends to produce 3' OH ends and two nucleotides inside ISY100 on the other strand to produce 5' phosphates (Urasaki *et al.*, 2002). In this study, supercoiled plasmid containing ISY100 ends, and synthetic linear transposon ends were tested for cleavage by transposase *in vitro*. Cleavage products were observed and the cleavage sites were mapped. Supercoiled plasmid substrates were cleaved *in vitro* exactly as found *in vivo* by Urasaki *et al.* (2002). However, linear DNA fragments containing single ISY100 ends were cleaved mainly one nucleotide inside the transposon end to produce a 3' OH and one nucleotide outside the transposon end to produce a 5' phosphate.

Changes in the flanking TA dinucleotides at either one end or both ends of ISY100, reduced the efficiency of transposition *in vivo*. These changes also reduced the efficiency of cleavage *in vitro*. Changes at only one end inhibited cleavage at both ends implying communication between the two transposon ends.

Synthetic pre-cut transposon ends were tested in an *in vitro* integration assay, and transposase catalysed the insertion of transposon 3'-OH ends into a target plasmid.

Transposase mediated efficient integration of a mini-ISY100, pre-excised by transposase or restriction enzyme, into TA targets *in vitro*, confirming that excised transposon fragments are intermediates in the reaction.

Target sequences of ISY100 from published data and this study were analyzed, yielding the consensus target sequence ADWTAWHT (W = A or T, D = not C and H = not G), in which the central TA is the duplicated target dinucleotide.

Transposase was fused with the DNA-binding domain of Zif268, separated by different-length linkers. Transposition activities of these chimeric proteins were tested *in vivo* with a series of target plasmids carrying TA target sequences adjacent to a Zif268 binding site. Results with chimeric transposases showed reduced transposition frequencies. However, when the Zif268 DNA-binding domain was linked to ISY100 transposase via the DNA-binding domain of Tn3 resolvase, transposition occurred into TA dinucleotides to one side of the Zif268 binding site with elevated frequency. This could be developed into a genetic tool for target-specific integration.

Abbreviations

Units

m	10^{-3}
μ	10^{-6}
n	10^{-9}
p	10^{-12}

Chemicals/Reagents

A	adenine
Amp	ampicillin
ATP	adenosine triphosphate
BSA	bovine serum albumin
C	cytosine
CIP	Alkaline Phosphatase, Calf Intestinal
Cm	chloramphenicol
DMSO	dimethyl sulfoxide
DNase I	deoxyribonuclease I
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EtBr	ethidium bromide
G	guanine
IPTG	isopropyl ß-D-thiogalactopyranoside
Kan	kanamycin
SDS	sodium dodecyl sulphate
T	thymidine
TEMED	N,N,N',N'-tetramethyl-1,2-diaminoethane
TdT	Terminal Transferase
Tris	tris -(hydroxymethyl)aminoethane

Other Terms

bp	Base pair(s)
cps	Counts per second
kb	Kilobase pairs
kDa	Kilo Dalton
nt	Nucleotide(s)
ori	Origin of replication
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
X^r	Resistance to X
X^s	Sensitivity to X
L	Litre

Contents

Chapter I
Introduction

1.1 Introduction

Transposons are defined mobile genetic elements that can jump, or transpose, from one site to another in DNA. This transposition is a form of genetic recombination and is catalysed by specialised enzymes called transposases. Genetic transposition was first observed in maize by Barbara McClintock half century ago (McClintock, 1951; Comfort, 2001). Since then, transposons have been identified widespread in nature, and are found in virtually all organisms. Large-scale sequencing results have revealed that a large fraction of DNA in many eukaryotic genomes consists of transposable elements and their remnants. For example, about 40% of both human and mouse genomes are derived from transposable elements (Lander *et al.*, 2001; Waterston *et al.*, 2002). Transposition leads to a variety of genomic rearrangements such as insertions, deletions and translocations (for examples see Craig *et al.*, 2002), is important in the spread of antibiotic resistance between bacteria (reviewed in Gomez-Lus, 1998), and has contributed to the evolution of the genomic DNA of species from all three kingdom (Bennett, 2004; Kazazian, 2004).

1.2 Transposition in brief

1.2.1 Class I transposons

Based on whether transposition occurs via an RNA intermediate, transposons have been categorised into two classes (reviewed in Finnegan, 1992). Class I transposons (or retrotransposons) move via an RNA intermediate and can be further divided into long terminal repeat (LTR) and non-LTR retrotransposons. To perform transposition, LTR retrotransposons, including retroviruses and retrovirus-like transposons, are first transcribed into a RNA copy from the integrated DNA copy. This RNA is then reverse-transcribed into a full length linear double strand cDNA in a process involving template switching between the LTRs (reviewed in Wilhelm and Wilhelm, 2001). Finally the cDNA copy of the retrovirus is integrated into genomic DNA in a reaction catalysed by the integrase protein (IN) (reviewed in Lewinski and Bushman, 2005; Craigie, 2002).

Non-LTR retrotransposons do not have long terminal repeats but usually have a 3' poly(A) tail or a similar sequence, so are also called poly(A) retrotransposons. Based on their length, non-LTR retrotransposons can be further divided into long and short interspersed nuclear elements (LINEs and SINEs) (reviewed in Deininger and Batzer, 2002). SINEs are nonautonomous forms of LINEs and depend on proteins encoded by the

LINEs for their transposition. To transpose, an mRNA copy is first transcribed from the DNA of the integrated element and translated to generate functional proteins with endonuclease and reverse transcriptase activities. Endonuclease cleaves the target DNA revealing a poly(T) stretch which base pairs with the poly(A) sequence of the LINE or SINE RNA. Target site primed reverse transcription then synthesizes the first cDNA strand. The RNA is degraded, the second strand is synthesized and DNA joining and repair completes the integration of the non-LTR element into a new location in the genome (reviewed in Ostertag and Kazazian, 2001).

1.2.2 Class II transposons

In contrast to class I transposons, class II transposable elements transpose directly at the DNA level and can be found in all species from Archaea and Eubacteria to Eukarya. According to their mechanisms of transposition, class II transposons can be subdivided into two categories, non-replicative transposons and replicative transposons. During transposition, non-replicative transposons are cut on both strands at both ends to produce a fully excised element which then inserts into a new target site (Figure 1.1). This mechanism has also been called cut-and-paste transposition. Replicative transposons are cut on only one strand at each transposon end. The transposition products can then be processed by DNA replication into a cointegrate molecule containing donor and target replicons separated by two copies of the transposon. The cointegrate molecule can then be resolved by site-specific recombination into the original donor plasmid and the target plasmid with a new transposon insertion (Figure 1.1).

One type of class II transposons common in bacteria are called "insertion sequences" (IS). IS are short DNA mobile elements with a simple genetic structure and no selectable phenotype (Figure 1.2). A number of eukaryotic class II transposons (such as *mariner* and Tc1) have a similar organisation and structure to these simple bacterial insertion sequences.

In bacteria, two ISs can mobilise a segment of DNA (often carrying phenotypic markers) in between them, forming a composite transposons (Figure 1.2). Two examples of this are Tn10, formed by two copies of IS10 flanking a tetracycline resistance gene, and Tn5, formed by two copies of IS50 flanking a kanamycin resistance gene. Other bacterial transposons, such as Tn7 and Tn3, have more complex genetic structures. These may contain multiple genes required for transposition as well as antibiotic resistance genes, all flanked by transposon end sequences which are recognised by the transposase proteins.

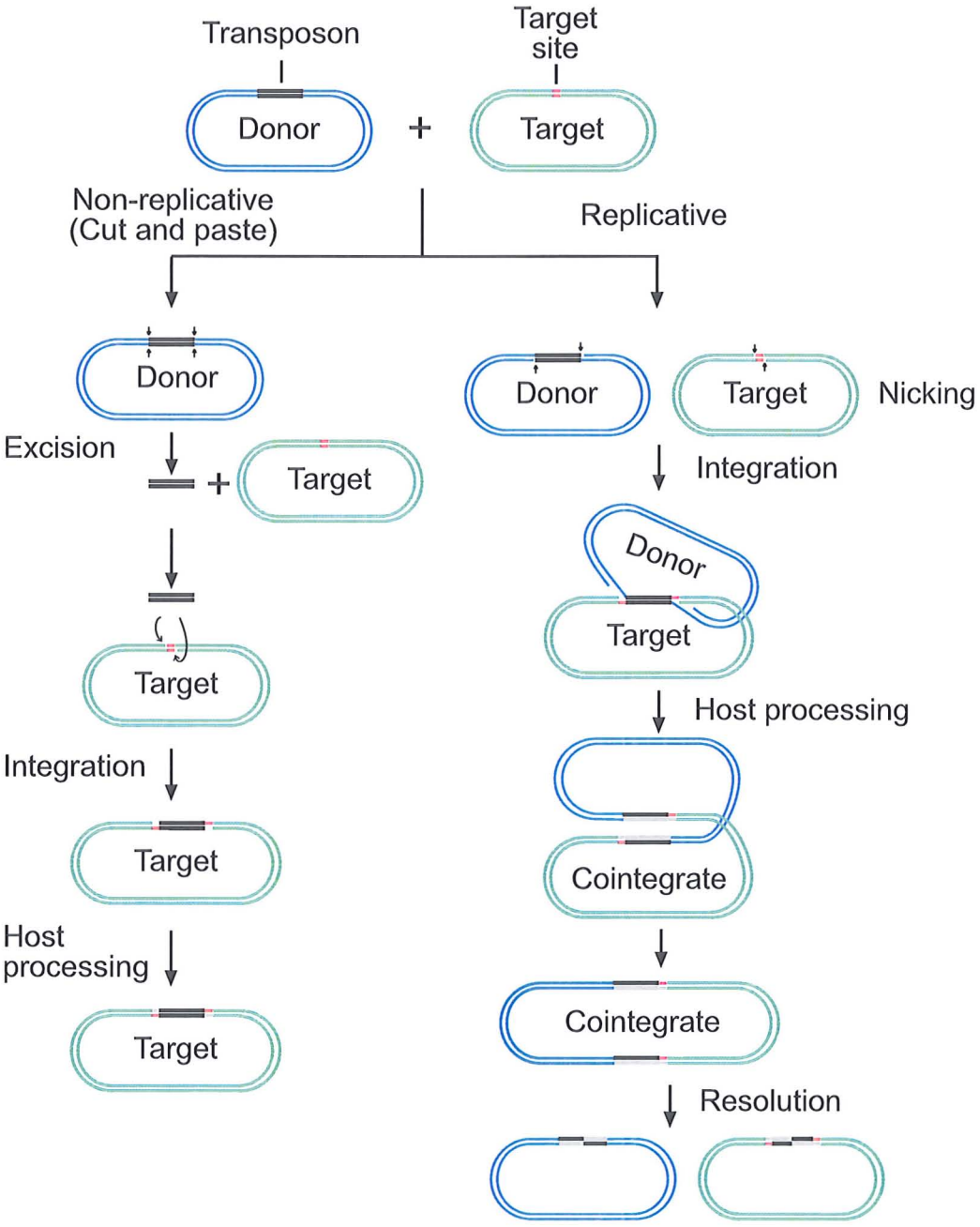


Figure 1.1 Replicative and non-replicative transposition. Left: non-replicative (cut-and-paste) transposition, in which the transposon is fully excised from the donor molecule and then inserted into a new target. Right: replicative transposition, in which the transposon is cleaved only at its 3' ends. Integration of these 3' ends into a target molecule produces two replication fork-like structures. Replication produces a cointegrate which can be resolved by site-specific recombination into donor and target molecules, each carrying a single copy of the transposon. The diagram is coloured as follows: transposon, thick black lines; donor replicon, blue; target replicon, green; target site, red; newly replicated DNA, grey.

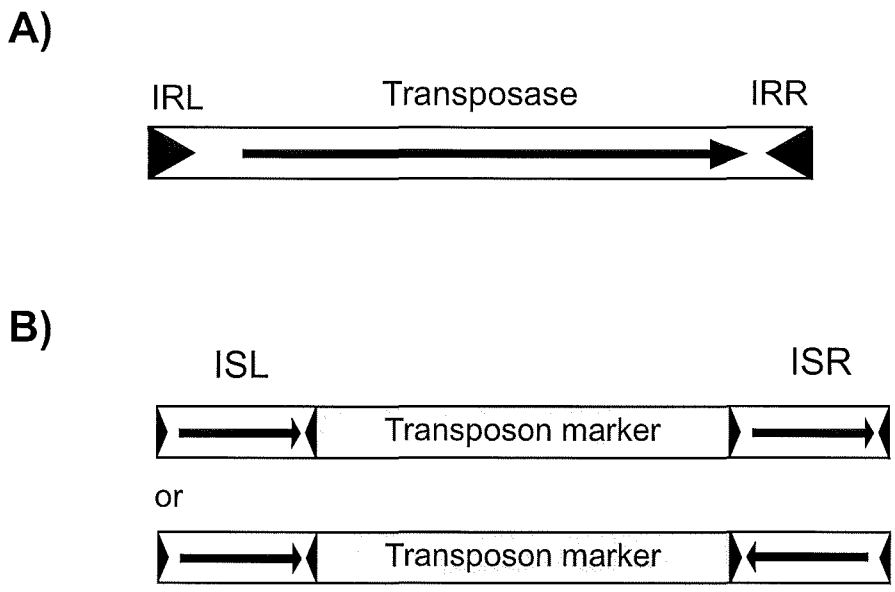


Figure 1.2 Structure of typical insertion sequences and composite transposons. A) Insertion sequences contain single open reading frames encoding the transposase (shown as arrow) flanked by the terminal inverted repeats IRL and IRR (shown as black triangles). **B)** Composite transposons contain transposon marker genes flanked by two insertion sequences, which can be in direct or inverted repeat.

1.2.3 Examples of class II transposons

A large number of class II transposons have been studied in great detail, facilitated by *in vitro* systems that recreate the full transposition reaction with purified proteins and DNA substrates. Some examples are briefly introduced below.

The bacterial composite transposon Tn5 jumps in a cut-and-paste reaction that requires only the transposase protein. Transposase binds to the IRs at both ends of the transposon, brings the ends together, cuts both strands at both ends and inserts the transposon at a new location (Goryshin and Reznikoff, 1998). Tn10 transposes by a similar mechanism, but requires the host protein IHF (Integration Host Factor) for efficient transposition (Chalmers *et al.*, 1998).

The bacterial transposon Tn7 carries at least three antibiotic resistance genes in between transposon ends containing multiple binding sites for the transposase protein TnsB (Arciszewska and Craig, 1991). Transposition also requires the other transposon-encoded proteins TnsA, TnsC and either TnsD or TnsE. TnsA and TnsB cooperate to excise the transposon from its original location, and TnsB catalyses insertion at a new location (Sarnovsky *et al.*, 1996). TnsC, D and E are involved in directing Tn7 to its new target (Bainton *et al.*, 1993; Peters and Craig, 2001).

Tc1 and Tc3 from *Caenorhabditis elegans*, *Mos1* from *Drosophila mauritiana* and other *mariner* elements, which are found widespread in eukaryotes (reviewed in Plasterk *et al.*, 1999), as well as *Hermes* from the *hAT* superfamily jump by cut-and-paste mechanisms (Zhou *et al.*, 2004), and have also been studied *in vitro*.

IS911, an insertion sequence first isolated from *Shigella dysenteriae*, uses a different strategy to produce the linear excised transposon via a circular intermediate. This linear excised transposon then inserts at new locations in DNA much like other class II transposons (Polard and Chandler, 1995; Polard *et al.*, 1996; Ton-Hoang *et al.*, 1999).

The best studied elements that jump by a replicative mechanism are Tn3 and Mu. Mu is a transposon that has adopted the lifestyle of a bacteriophage by acquisition of genes required for the lytic and lysogenic phases of the bacteriophage lifecycle, and uses transposition to replicate its DNA in the lytic phase (reviewed in Chaconas and Harshey, 2002). Tn3 was one of the earliest transposons to be discovered because of its ability to confer resistance to ampicillin (reviewed in Grindley, 2002). Transposition of Tn3 produces a cointegrate molecule that is resolved by the well studied Tn3 resolvase protein (reviewed in Grindley, 2002).

Retroviral integration reactions, for instance that catalysed by HIV integrase, are closely related to the transposition reaction catalysed by the transposases of class II transposons mentioned above, and have also been well studied (reviewed in Craigie, 2001).

1.2.4 Transposons as genetics tools

Transposons have been widely used as genetic tools. They are most commonly used to create tagged mutations in the study of model organisms ranging from bacteria to fungi, insects and mammalian cells in tissue culture. Transposons are potent mutagens, and transposition occurs at random or semi-random sites. Methods have been developed to isolate large numbers of independent transposition events. Transposition may take place *in vivo* in the target organism, or may be carried out *in vitro* into isolated target DNA.

Once a transposon tagged mutant has been selected, the known transposon sequence can be used to isolate the flanking DNA and identify the gene of interest. Transposons can also be used as mobile primer sites for sequencing projects or PCR-based strategies (reviewed in Hayes, 2003). Transposons can also be used to deliver single copy integrated transgenes to a wide variety of organisms. Any DNA can be inserted into the genome, so long as it is flanked by transposon end sequences and the transposase protein is supplied, usually *in trans*.

In *Drosophila melanogaster*, the P-element transposon has been developed as a genetic tool for about 20 years (Rubin and Spradling 1982; Spradling and Rubin 1982). P elements are the method of choice for producing transgenic *D. melanogaster* and transposon-based systems are being developed for other important insects such as mosquitoes. More recently transposons from the Tc1/*mariner* family and *hAT* superfamilies have been used or are being developed as genetic tools in vertebrate organisms (reviewed in Miskey *et al.*, 2005). The purposes of using transposon in multicellular organisms are similar to those in microorganisms, including characterization of functional genes identified by genome sequencing projects, transgenesis and insertional mutagenesis (Adelman *et al.*, 2002; Carlson *et al.*, 2003; reviewed in Hamer *et al.*, 2001). New applications such as transposon-based RNAi for generating gene knockdowns in mammalian cells have also been developed (Heggestad *et al.*, 2004).

1.3 Mechanism of transposition

The transposons described in section 1.2.3 are those that have been most intensively studied, and whose mechanisms of transposition are understood best. The following

discussion of transposition mechanisms is therefore based largely on results from these elements.

1.3.1 Strand cleavage and transfer reactions in transposition

Transposases use two basic chemical steps in their transposition reactions: DNA strand cleavage or hydrolysis to produce a 5' phosphate and a 3' OH, and transesterification reactions to join this 3' OH terminus to another DNA strand (see section 1.3.2). For many transposons, including Tn5 and Tn10, both cleavage and strand transfer reactions are catalysed by a single active site in the transposase protein (Bolland and Kleckner, 1996; Davies *et al.*, 1999; Kennedy *et al.*, 2000).

Non-replicative transposons use a variety of strategies to cleave the transferred and non-transferred strands of the transposon ends. Double strand cleavages at Tn5 and Tn10 ends occur via hairpin intermediates (Bhasin *et al.*, 1999; Kennedy *et al.*, 1998). One strand at each end, the transferred strand, is first cleaved in a hydrolysis reaction that liberates the transposon 3' OH end (Bolland and Kleckner, 1995). This then attacks the opposite strand, and becomes covalently joined to it forming a hairpin structure at the transposon end (Figure 1.3A; Bhasin *et al.*, 1999; Kennedy *et al.*, 1998). The hairpin is opened by transposase in another hydrolysis reaction to produce transposon 3' OH and 5' phosphate ends (Figure 1.3A).

The eukaryotic transposon *Hermes* from the *hAT* family also jumps via a hairpin intermediate. However, in contrast to Tn5 and Tn10, the hairpin is on the sequences flanking the transposon rather than on the transposon ends. The donor molecule is first cleaved one nucleotide outside the transposon on the non-transferred strand. The liberated 3' OH attacks the opposite strand, forming a hairpin structure on the transposon flanking sequences, and a double-stranded break is produced at the transposon ends (Zhou *et al.*, 2004). This same mechanism is also used during V(D)J recombination (reviewed in Oettinger, 1999).

Elements such as IS911 and IS21 are excised from the donor site via a circular DNA intermediate (reviewed in Berger and Haas, 2001). Hydrolysis first occurs on the transferred strand at just one end (Polard and Chandler, 1995). The liberated transposon 3' OH attacks the same strand at the other transposon end, forming a covalently closed single stranded circle. Replication yields a double stranded circular transposon, with the two ends joined together. Transposase then cleaves both strands at the junction between the two

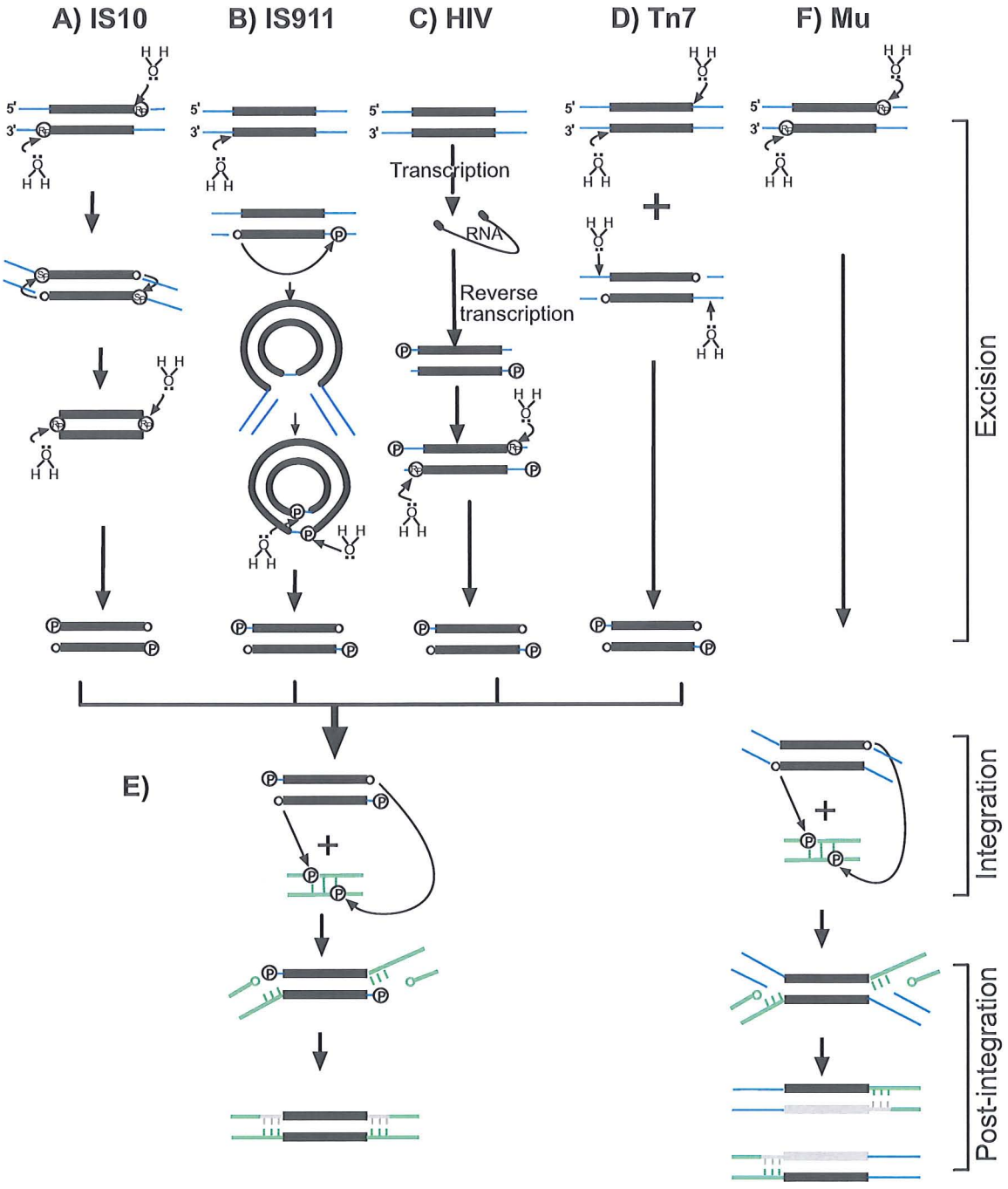


Figure 1.3 Different transposons use different transposition pathways. A) Double strand cleavage at IS10 and IS50 ends takes place via a hairpin intermediate. **B)** Pathway of IS911 excision. Transposition occurs via a circular intermediate. **C)** Retroviral integrases cleave two nucleotides from the 3' ends of the full length cDNA. **D)** The 3' ends of Tn7 are cleaved by TnsB, while the 5' ends are cleaved by TnsA. **E)** After excision, the transposon 3' ends liberated in steps A-D are inserted into staggered positions on the target DNA. Gaps and unpaired strands are repaired by host machinery to produce the target duplication (grey). **F)** For Mu and Tn3 only the transferred-strands are cleaved, producing 3' OH ends which are transferred to the target DNA. The strand transfer product is replicated to form a cointegrate. The diagram is coloured as follows: Transposon DNA, thick black lines; flanking donor DNA, blue; target DNA, green. Phosphates are shown as circles with a ' P ' or where known the preferred phosphorothioate stereoisomer Sp or Rp. The exposed 3' OH groups involved in subsequent reactions are shown as small open circles. Positions of strand cleavage are indicated by short arrows. Nucleophilic attack on the DNA phosphodiester bond by water or 3' hydroxyls are also indicated by arrows. (Adapted from Haren *et al.,* 1999; Turlan and Chandler, 2000).

ends, forming a linear excised transposon which can be inserted into a new target site (Figure 1.3B; Polard and Chandler, 1995; Polard *et al.*, 1996; Ton-Hoang *et al.*, 1999).

Full-length double stranded retroviral cDNAs are only a few bp longer than the integrated form of the virus. The extra base pairs are cleaved from the cDNA 3' ends by the integrase protein in a reaction similar to the strand cleavage reactions catalysed by transposases (Figure 1.3C).

Unlike other transposons, excision of Tn7 from the donor DNA is carried out by a heteromultimeric complex of two proteins, TnsB and TnsA. The active site in TnsB cleaves the transferred strands and catalyses strand transfer reactions, while TnsA cleaves the non-tranferred strands of Tn7 (Figure 1.3D; Sarnovsky *et al.*, 1996). While TnsB is homologous to other transposase proteins, TnsA is structurally related to type II restriction endonucleases (Hickman *et al.*, 2000).

Once the transposon has been cleaved at the 3' ends, either with or without cleavage of the other strand, the transposon 3' OH termini are transferred to phosphates on opposite strands of the target DNA (Figure 1.3E). This yields a transposition product with nicks that can be repaired by the host repair machinery. The two ends of the transposon are usually transferred to staggered positions on the target DNA, resulting after repair in a short direct target repeat flanking the transposon insertion. The length of this target repeat is characteristic of a given transposon.

1.3.2 Chemistry of the strand-cleavage and transfer reactions

The two different types of chemical reaction catalysed by transposase, strand cleavage and strand transfer, share a similar mechanism and are usually catalysed by a single active site in the transposase protein. In the DNA cleavage reaction, a water molecule acts as a nucleophile to attack a phosphate at the cleavage site, generating a 5' phosphate and a 3'OH at the cleaved ends. In the strand transfer reaction, a 3' OH (generated by a previous cleavage reaction) acts as a nucleophile to attack a DNA phosphate. The attacking 3' end becomes covalently linked to the target phosphate displacing the original target 3' end. For Mu transposase and HIV integrase, it has been shown that both DNA cleavage and strand transfer occur by a one-step in-line substitution mechanism (Engelman *et al.*, 1991; Mizuuchi and Adzuma, 1991). A nucleophile attacks the activated phosphate from one side and a 3' OH leaves the phosphate on the other side of the phosphate. Evidence for this one step mechanism came from the inversion of chirality observed on phosphorothioate-containing substrates (Figure 1.4; Mizuuchi *et al.*, 1999). This contrasts with the

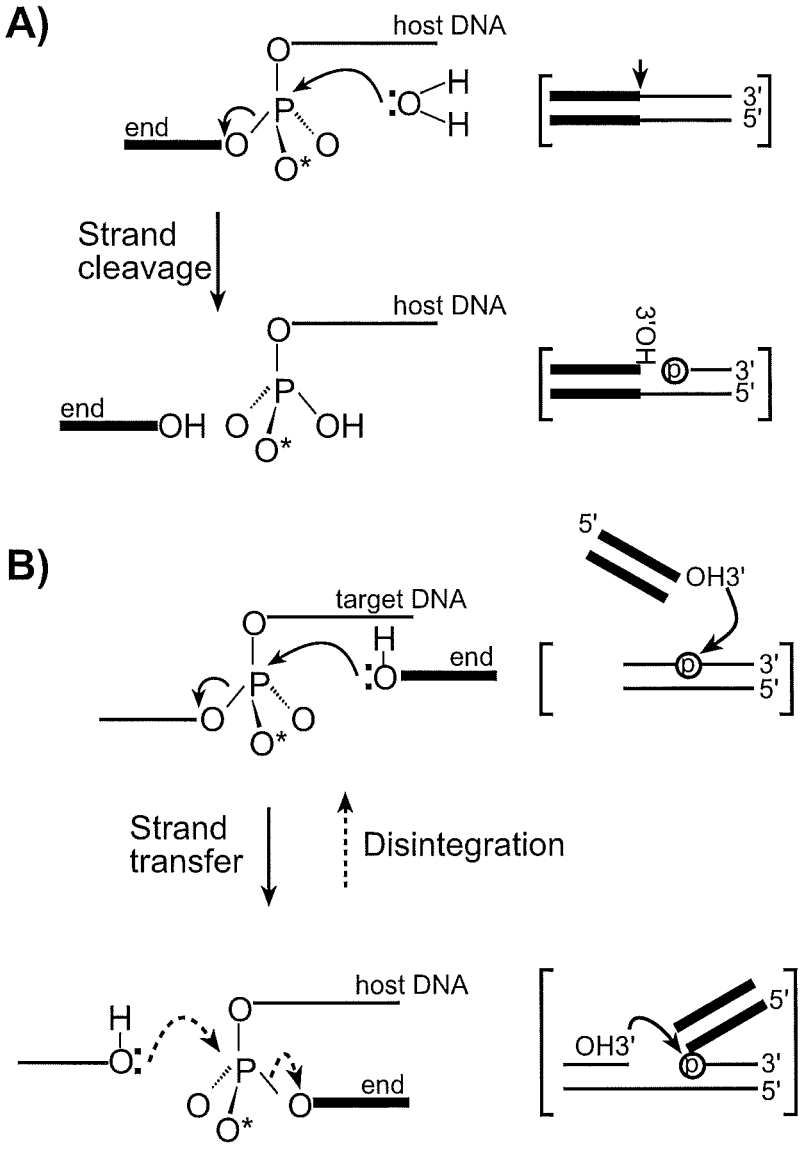


Figure 1.4 The chemistry of strand cleavage and transfer. A) Strand cleavage. **B)** Strand transfer and disintegration.The transposon DNA is represented as a bold line. Only one of the relevant DNA strands is shown. Nucleophilic attack (curved arrows) is carried out either by H_2O (A), or a 3' hydroxyl from the exposed transposon end for strand transfer, or from the exposed 3' hydroxyl in the target DNA for disintegration (B). The phosphodiester bond that undergoes cleavage in these reactions is shown as a chiral form, in which a non-bridging oxygen is represented as O*. The hydrolysis (A) and transesterification (B) reactions shown here result in an inversion of chirality. The corresponding steps involving double-strand donor and target DNA together with their polarity are presented on the right of this figure for clarity. (Adapted from Haren *et al.*, 1999).

conservation of chirality seen in the two-step transesterification reaction mechanisms adopted by site-specific recombinases, which cut and rejoin via covalent protein-DNA intermediates (reviewed in Hallet and Sherratt, 1997).

1.3.3 An N-terminal DNA-binding domain

Transposition of class II transposons is initiated by the transposase specifically bound to the transposon ends. The N-terminal domains of many different transposases have been shown to bind specifically to sequences found at both transposon ends (Colloms *et al.*, 1994; Schukkink and Plasterk, 1990; Tavakoli *et al.*, 1997; Weinreich *et al.*, 1994; Zhang *et al.*, 2001). Helix-turn-helix DNA-binding motifs have been recognised by sequence homology in the N-terminal domains of several different transposases (Nagy *et al.*, 2004; Pietrokovski and Henikoff, 1997; Ton-Hoang *et al.*, 2004) and in two cases the presence of these helix-turn-helix DNA-binding domains has been verified by X-ray crystallography (Clubb *et al.*, 1994; Watkins *et al.*, 2004).

In bacterial transposons, the N-terminal DNA-binding domain may fold as it is translated and bind to the transposon end before the C-terminal catalytic domain of the transposase has been synthesised. This property could account for the *cis* preference of some transposons, where transposase acts preferentially on the transposon copy encoding it (Derbyshire and Grindley, 1996; Morisato *et al.*, 1983; Zhou and Reznikoff, 1997).

1.3.4 A conserved catalytic DDE domain

A conserved catalytic domain has been recognised by sequence homology in a large variety of DNA transposases as well as the retroviral integrases (reviewed in Haren *et al.*, 1999). This domain contains three highly conserved acidic residues, two aspartates and a glutamate, which are essential for catalytic activity (Junop and Haniford, 1997; Kulkosky *et al.*, 1992). Most of the sequence similarity clusters around these three acidic residues (Figure 1.5). This transposase catalytic domain has been termed the DDE domain, and proteins that contain it are referred to as DDE transposases. X-ray crystal structures have been solved for several different DDE domains, all of which share a structural fold related to that of RNase H (Bolland and Kleckner, 1996; Lins *et al.*, 2000; Richardson *et al.*, 2006; Steiniger-White *et al.*, 2004; reviewed in Rice and Baker, 2001). The three acidic residues coordinate a pair of divalent metal ions (Mg^{++} or Mn^{++}), which are thought to catalyse the strand cleavage and transfer reactions by stabilising the pentavalent phosphorous

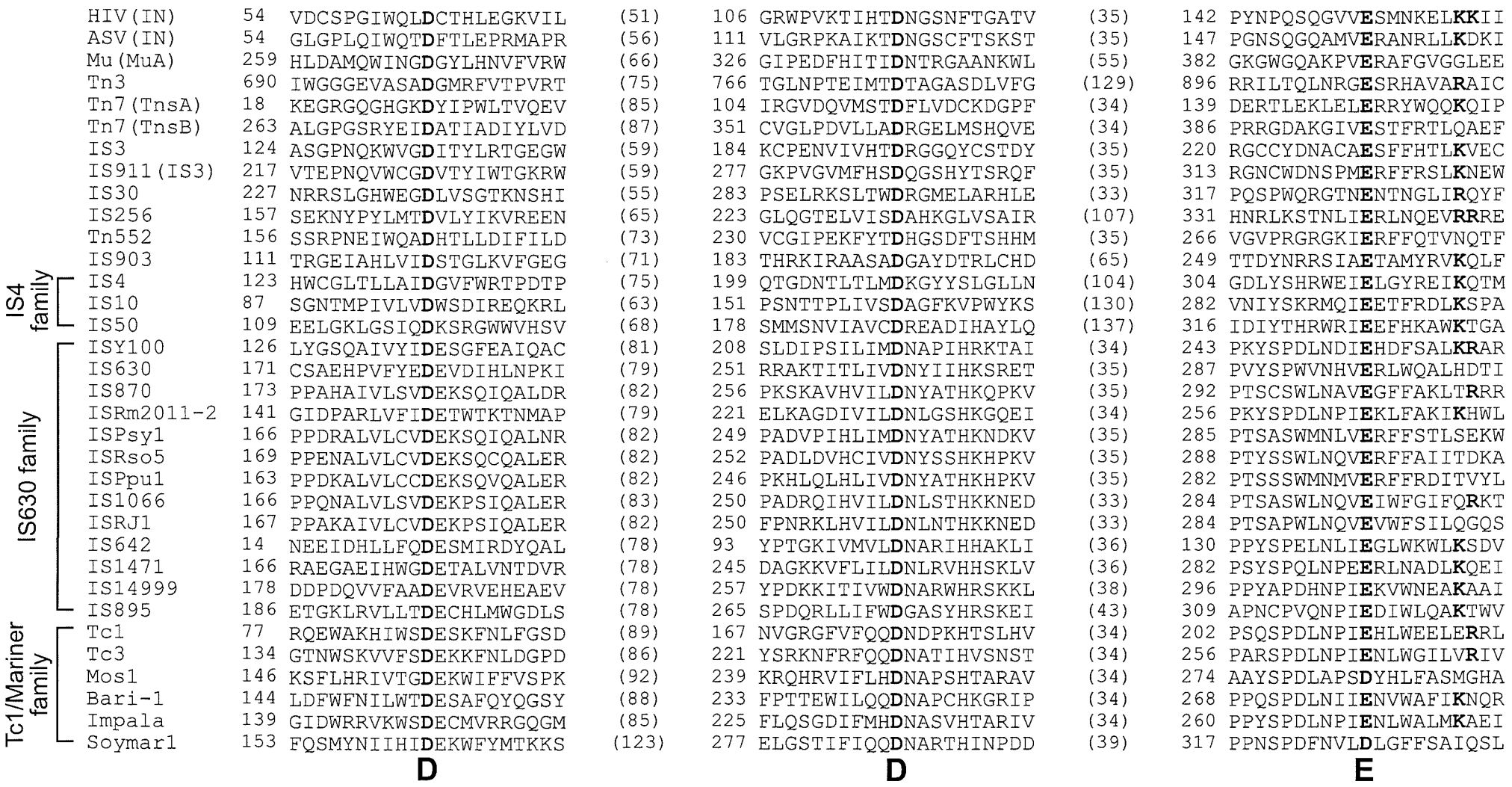


Figure 1.5 Alignment of DDE domains from a selection of transposases. Conserved DDE residues are shown in bold. The positions of the first residue in each segment in the proteins are indicated and the distances between the catalytic acidic residues are shown in brackets.

intermediate, activating the nucleophile and stabilising the negative charge on the 3' OH leaving group (Figure 1.6; reviewed in Yang *et al.*, 2006).

Not all DNA transposases fall into the DDE family (reviewed in Mahillon and Chandler, 1998). Some elements that have been classified as insertion sequences use enzymes that share homology and mechanistic similarities with enzymes involved in rolling circle replication, or site-specific recombinases. The mechanism of transposition of these elements may be more similar to site-specific excision and integration carried out by bacteriophage such as λ (Lenich and Glasgow, 1994; Serre *et al.*, 1995). Transposases of some other insertion sequences have active sites that have yet to be identified.

1.4 Insertion sequences and the IS630/Tc1/*mariner* superfamily

The work described in this thesis concerns an insertion sequence belonging to the IS630/Tc1/*mariner* family of transposons. The sections below give a brief introduction to bacterial insertion sequences and then go on to describe the properties of the IS630/Tc1/*mariner* family in more detail.

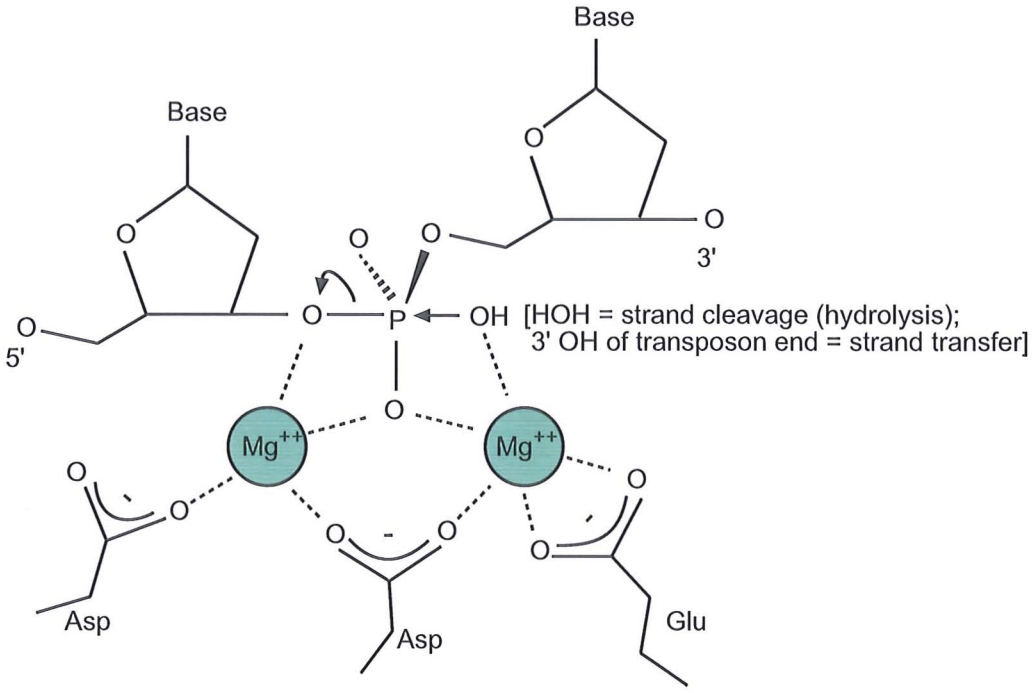
Insertion sequences are small genetically compact DNA transposons found in bacteria (reviewed in Mahillon and Chandler, 1998). They are one of the major causes of mutation in bacteria and were first discovered for their ability to cause polar mutations blocking transcription of downstream genes in operons (Fiandt *et al.*, 1972). They also provide repeated sequences in bacterial genomes that allow DNA rearrangements by homologous recombination (for example see Parkhill *et al.*, 2003).

In general, ISs have short terminal inverted repeats marking the transposon ends, flanking a single gene that encodes a functional transposase (Figure 1.2). The IR upstream of the transposase gene is referred to as IRL and the one downstream of the transposase gene is IRR.

So far, more than 500 different IS elements have been found, and they are classified into various families or groups based on their structures and the degree of homology between their nucleotide and protein sequences (http://www-is.biotoul.fr). One family of ISs is the prokaryotic IS630 family, closely related to the eukaryotic Tc1/*mariner* family of transposons. Together these elements form the IS630/Tc1/*mariner* superfamily (Doak *et al.*, 1994).

More than 50 different IS630 family insertion sequences have been identified in eubacteria and archaebacteria (http://www-is.biotoul.fr). Members of the bacterial IS630

A)



B)

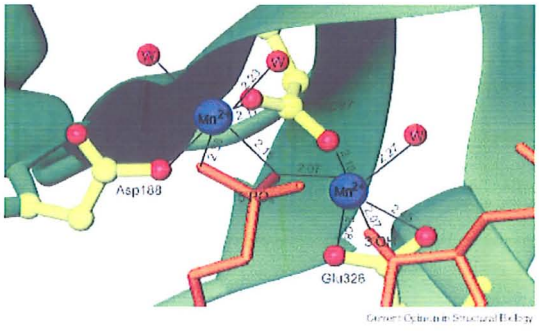


Figure 1.6 Coordination of divalent metal ions by the DDE catalytic residues. A) Proposed mechanism of two-metal-ion catalysis by DDE transposases (Steitz and Steitz, 1993; reviewed in Yang *et al.*, 2006). The two metal ions are coordinated by carbonyl oxygens of the three acidic residues. A non-bridging oxygen on the scissile phosphate is directly between the two metal ions. One metal ion assists nucleophilic attack by a 3' OH or a water molecule and the other stabilises the negative charge on the 3' OH leaving group. **B)** Crystal structure of the active site of Tn5 transposase, showing the coordination of Mn^{++} by the DDE residues (From Steiniger-White, *et al.*, 2004).

family of ISs include IS630 from *Shigella sonnei* (Matsutani et al., 1987), ISRm2011-2 from *Rhizobium meliloti* (Selbitschka *et al.*, 1995) and IS1066 from *Pseudomonas* sp. strain P51 (van der Meer *et al.*, 1991) (See Table 1.1, Figure 1.7).

The eukaryotic transposon *Mos1 mariner* was first discovered in *D. mauritiana* 20 years ago (Hartl, 2001; Jacobson *et al.*, 1986). Members of the *mariner* family have now been found in a wide diversity of species including insects, other arthropods, fish and humans (Feschotte and Wessler, 2002; Oosumi *et al.*, 1995; Robertson, 1993; Robertson and Lampe, 1995; Wicker *et al.*, 2005). Most members of the *mariner* family are inactive mutants, relics of past active transposons that have been inactivated by mutation.

Two active *mariner* elements, *Mos1* and *Himar1* have been isolated in different ways. The *Mos1* element was discovered as an active transposon in *D. mauritiana* (Jacobson *et al.*, 1986), whereas *Himar1* was reconstructed from the consensus sequence of multiple inactive transposon copies in the horn fly *Haematobia irritans* (Lampe *et al.*, 1996; Robertson and Lampe, 1995).

The first members of the Tc1 family to be isolated were found in the nematode *C. elegans*. Tc1 itself was discovered as an active transposon in *C. elegans*, at about the same time that *mariner* was discovered in the fruit fly (Emmons *et al.*, 1983). Another naturally active Tc1 family member, Tc3 was found later, also in *C. elegans*. Other Tc1-like elements, such as *Sleeping Beauty (SB)* were reconstructed from inactive elements found in the genomes of fish or other organisms (Ivics *et al.*, 1997).

1.4.1 Structural analysis of the IS630/Tc1/*mariner* superfamily

Generally, members from this superfamily have a typical IS structure (Figure 1.2). All elements are 900-2400 bp in length and, apart from a few exceptions, carry a single ORF encoding the transposase protein. A distinguishing feature of this family is that all members integrate into the sequence TA, and insertions are flanked by TA on both sides. The presence of the palindromic sequence TA at both the original target site and flanking the new insertion makes the position of the transposon ends and the length of the target duplication ambiguous. The TA target site might be duplicated during transposition, or one nucleotide at each end might be part of the transposon. Results obtained with Tc3, IS630 and *Mos1 mariner,* either by changing the sequences flanking the donor transposon, or by precise mapping of the excised transposon ends, showed that the TA is a target duplication and is not part of the transposon (Dawson and Finnegan, 2003; Tenzen *et al.*, 1990; van Luenen *et al.*, 1994).

Table 1.1 Transposable elements of IS630/Tc1/*mariner* superfamily

A) IS630 family

Name	Origin	Accession No.		Length (bp)
		Nucleotide	Protein	
IS14999	*Corynebacterium glutamicum* ATCC 14999	AB186419	BAD89377	1149
IS1471	*Burkholderia cepacia*	U67938	AAB38861	1113
IS642	*Bacillus halodurans* C-125	AP001515	BAB06239	1142
IS630	*Shigella sonnei*	X05955	CAA29389	1153
IS630-Spn1	*Streptococcus pneumoniae*	AF026471	AAD10186	895
IS870	*Agrobacterium vitis* 2657	Z18270	CAA79150	1146
IS895*	*Anabaena sp. PCC7120*	M67475	AAA98138 AAA98139	1192
ISAr1	*Agrobacterium rhizogenes* A4	K03313	-	1146
ISPpu1	*Pseudomonas putida*	AJ245436	CAB54057	1164
ISPpu2	*Pseudomonas putida*	AJ233397	CAB69082	1131
ISPpu3	*Pseudomonas putida*	AJ233397	-	1135
ISPsy1	*Pseudomonas syringae pv. glycinea*	AF169828	AAD50908	2009
ISRf1 (ISRfr1)	*Rhizobium fredii* (Sinorhizobium fredii) USDA257	M73698	-	1136
ISRso5	*Ralstonia solanacearum*	NC_003295	CAD16890	1167
ISRm10	*Sinorhizobium meliloti* BE31CC	AF143444	AAD37358	1047
ISRm10-1	*Sinorhizobium meliloti* 2011	AJ242573	CAB43594	1042
ISRm2011-2**	*Rhizobium meliloti* 2011	U22370	AAC43490	1053
ISTcSa	*Synechocystis* sp. PCC 6803	U38915	AAB72124	947
ISY100	*Synechocysis* sp. PCC 6803	BA000022	BAA16620	947
IS1066	*Pseudomonas* sp. P51	U15298	-	1137
ISRj1	*Bradyrhizobium japonicum* USDA 110	X02581	CAA26419	1122

B) Tc1/*mariner* family

Name	Origin	Accession No.		Length (bp)
		Nucleotide	Protein	
Tc1	*Caenorhabditis elegans*	X01005	CAA25498	1610
Tc3	*Caenorhabditis elegans*	AF025458	P34257	2335
Impala	*Fusarium oxysporum f. sp. melonis*	AF282722	AAB33090.2	1281
Bari1	*Drosophila melanogaster*	X67681	CAA47913	1728
Mariner	*Drosophila mauritiana*	M14653	AAA28678	1286
Soymar1	*Glycine max*	AF078934	AAC28384	3491
Mos1	*Drosophila mauritiana*	X78906	-	1293
Minos	*Drosophila hydei*	X61695	S26856	1775

*: Contains two ORFs

**: Putative transposase from frame-shifting translation.

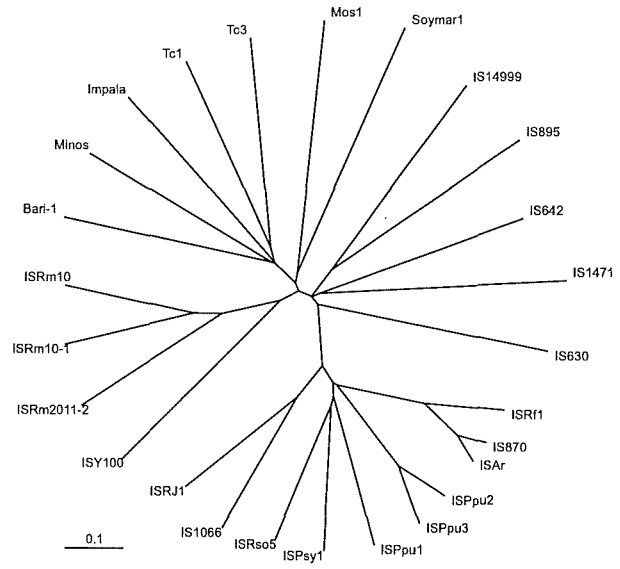


Figure 1.7 Phylogenetic analysis of the IS630/Tc1/*mariner* superfamily of transposases. The tree was generated by the program clustalX 1.83 based on amino acid sequences of a selection of the transposases shown in Table 1.1.

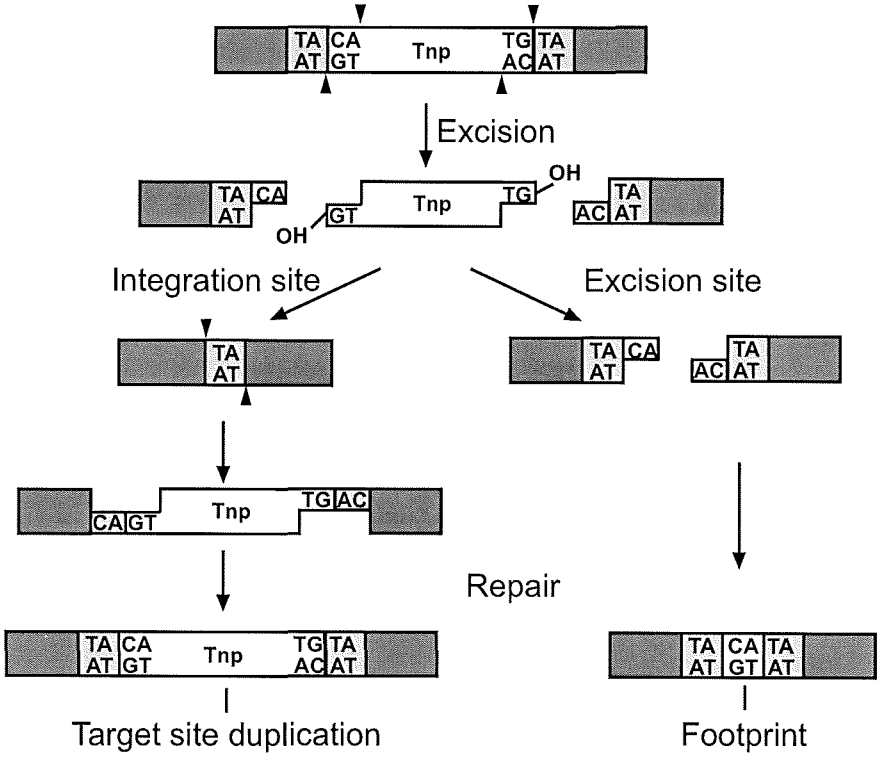


Figure 1.8 Proposed mechanism of transposition of IS630/Tc1/*mariner* superfamily. The cut-and-paste transposition model is based on studies of the Tc3 element. Tc3 is first excised from the donor molecule by double-stranded cleavage at the ends of the inverted repeats. These transposase-mediated cleavages cut the Tc3 element exactly at 3' ends, and two nucleotides inside the transposon at the 5' ends. The excised Tc3 inserts into staggered positions on a TA dinucleotide target site by transposase-mediated nucleophilic attack of transposon 3' OH ends. Host processing results in a TA target duplication flanking the transposon and a footprint at the donor site. (Adapted from Plasterk *et al.*, 1999).

The transposase genes of the bacterial IS630 family elements take up almost the entire length of the transposon and extend into the right inverted repeat. A few elements contain frameshift mutations and premature stop codons; these could simply be inactivating mutations in the transposase. Alternatively they might be involved in the regulation of transposase expression, as has been found for other bacterial ISs (reviewed in Mahillon and Chandler, 1998). The eukaryotic transposase genes, such as those from Tc1 and Tc3 contain introns, accounting in part for the greater length of these elements.

The transposase proteins encoded by IS630/Tc1/*mariner* family elements have an N-terminal DNA-binding domain that specifically recognises the transposon ends (IRs) and a C-terminal catalytic domain with homology to other DDE transposases (reviewed in Plasterk *et al.*, 1999). The distance between the last two acidic residues in the catalytic domain is 35 for the prokaryotic IS630 family, and 34 for the Tc1 and *mariner* families. The last acidic residue in the triad is D in the *mariner* elements giving them a DD(34)D motif (Capy *et al.*, 1997; Doak *et al.*, 1994). In recent years, other transposons have been identified in this family with catalytic domains classified as DD(37)E motifs (Shao and Tu, 2001; Turcotte and Bureau, 2002).

The length of terminal inverted repeats in members of the IS630/Tc1/*mariner* family ranges from 8 bp to greater than 400 bp. The family can be divided into three groups based on the inverted repeat length (Tu and Shao, 2002). Transposons in the first group, including Tc1 (54 bp IRs), *Mos1* (28 bp IRs) and IS630-like elements (8-29 bp IRs) have terminal inverted repeats shorter than 100 bp, with single transposase binding sites at the transposon termini. The second group of transposons, represented by *Sleeping Beauty* and *Minos*, have 200-250 bp long terminal inverted repeats and two binding sites in each inverted repeat. The third group, for example Tc3 with 462 bp IRs, carry long IRs (> 400 bp) that extend inward significantly beyond the second target binding site (reviewed in Plasterk *et al.*, 1999). However, inverted repeat length does not necessarily correlate with transposase sequence. For instance, not all Tc3-like transposons have inverted repeats as long as Tc3. Some have IRs in the second class, for example Tc3-CbIIa and Tc3-CbIIb from *C. briggsae*, while MsqTc3 from *A. aegypti* contains short (36 bp) terminal inverted repeats (Shao *et al.*, 2001; Tu and Shao, 2002). While extra transposase binding sites are important for transposition of *Sleeping Beauty* (Cui *et al.*, 2002), they are not required for transposition of Tc3 (Fischer *et al.*, 1999). It is not clear how these long, often almost perfect inverted repeats arise in Tc3, nor what if any is their function in transposition.

1.4.2 Cut-and-paste transposition of IS630/Tc1/*mariner* superfamily

The mechanism of cut-and-paste transposition in this family was first elucidated by studies on Tc3 transposition (Figure 1.8; van Luenen *et al.*, 1994).

Transposition of Tc3 contains several defined steps. Transposition is initiated by the binding of the Tc3 transposase to the ends of the inverted repeats (van Luenen *et al.*, 1993). The N-terminal domain of Tc3 transposase binds specifically to Tc3 ends (Colloms *et al.*, 1994). The structure of this domain bound to DNA has been solved by X-ray crystallography and contains two paired-like helix-turn-helix DNA-binding domains (van Pouderoyen *et al.*, 1997; Watkins *et al.*, 2004). The second step of transposition is double-strand cleavage at the Tc3 ends, to excise Tc3 from the donor site. The excised element was produced upon induction of transposase in *C. elegans* and contains the complete transposon sequence at the 3' ends but lacks the terminal 2 nucleotides at the 5' ends. The third step is presumed to be target capture by the excised transposon transposase complex, followed by integration of the excised transposon into the target site. The whole process is mediated by Tc3 transposase, and the three acidic residues in the DDE motif are essential for the catalytic activity of transposase (van Luenen *et al.*, 1994). Finally, it is thought that after transposition, the cellular machinery repairs the gap at each end of the newly inserted transposon, and so produces a full length Tc3 flanked by a duplicated TA sequence. As a result of transposon excision, two nucleotides from each transposon end remain behind in the donor site as a two nucleotide protruding 3' single stranded tail. This is frequently repaired producing a 'footprint' TACATA sequence, in which the two TA dinucleotides are from the duplication during the original transposition event (van Luenen *et al.*, 1994).

Later *in vitro* transposition studies on Tc1 (Vos *et al.*, 1996) and the *mariner* element *Himar1* (Lampe *et al.*, 1996) have shown that they all work by a similar mechanism to Tc3. Furthermore, the transposon encoded transposase protein is the only protein required for transposition. Similar mechanisms have since been confirmed for other members of the family, such as *Sleeping Beauty*, Frog Prince and *Mos1* (Dawson and Finnegan, 2003; Ivics *et al.*, 1997; Miskey *et al.*, 2003).

Mos1 is now one of the best studied members of the Tc1/*mariner* family. Transposition has been reconstituted *in vitro* with purified transposase (Tosi and Beverley, 2000). The two strands of the transposon are cleaved with a defined order. The two ends of the transposon are brought together in a paired end complex only after cleavage of the non-transferred strands, and the transferred strands are cleaved in this complex. A 3' OH from

the first strand cleavage is not required for second strand cleavage and cleavage does not occur via a hairpin intermediate (Dawson and Finnegan, 2003). More recently, a crystal structure has been solved for the catalytic domain of *Mos1* transposase, showing the two catalytic divalent metal ions, and their coordination by the three conserved acidic residues in the DD(34)D motif (Richardson *et al.*, 2006).

Studies on the bacterial IS630 family are less far advanced than those on the eukaryotic Tc1/*mariner* family. At the start of this project, no *in vitro* transposition systems had been established, so that most of our knowledge came from *in vivo* transposition studies and bioinformatics. The first prokaryote member to be discovered was IS630, isolated from *Shigella sonnei* (Matsutani *et al.*, 1987). Transposition of IS630 can take place in *E. coli* when the transposase protein is appropriately expressed (Tenzen *et al.*, 1990). IS630 does not mediate cointegration between different replicons, suggesting non-replicative transposition. IS630 preferentially inserts into TA dinucleotides in the tetranucleotide 5'-CTAG-3', and inserts are flanked by TA at both ends (Tenzen *et al.*, 1990; Tenzen and Ohtsubo, 1991). When the TA sequences were changed at both ends of IS630, the new sequences were not carried with the transposon upon transposition. Thus the TAs are a target duplication and not part of the transposon (Tenzen *et al.*, 1990).

The next best studied bacterial IS630 family element comes from the cyanobacterium *Synechocystis* sp. strain PCC6803, and was originally named IS1987 (Cassier-Chauvat *et al.*, 1997). It has also been called ISTcSa, indicating that it is a Tc1-like IS from *Synechocystis* (http://www-is.biotoul.fr). However, this element is referred to as ISY100 by the *Synechocystis* genome sequencing project (http://www.kazusa.or.jp/cyano) and by Urasaki *et al.* (2002) who were the first to study its transposition in *E. coli*. The name ISY100 will be used throughout the rest of this thesis. Urasaki *et al.* (2002) demonstrated that ISY100 undergoes double strand cleavage at the transposon ends, and transposes to new TA targets in *E. coli* on induction of the ISY100-encoded transposase.

1.5 Advantages of using IS630/Tc1/*mariner* superfamily as genetic tools

Transposons from IS630/Tc1/*mariner* superfamily hold great promise and have many advantageous traits for use as genetic tools. To date, several transposons of the IS630/Tc1/*mariner* family have been used in genome mapping (Akerley *et al.*, 2002; Wong and Mekalanos, 2000) and producing transgenic organisms (Adelman *et al.*, 2002;

Akerley *et al.*, 1998; Dupuy *et al.*, 2005; Matsuoka *et al.*, 2004; Rubin *et al.*, 1999). Some of the advantages of these elements are outlined below.

1.5.1 No host co-factors are required for transposition

Many transposases require host factors to perform transposition. For example, Tn10 needs IHF, a site-specific DNA binding protein, to assemble the transpososome (Morisato and Kleckner, 1987). In the replicative bacteriophage Mu transposition pathway, accessory proteins HU and IHF help in the formation of a stable three-site synapse (Maxwell *et al.*, 1987). This requirement for host-specific proteins makes it difficult to use these transposons outside their natural hosts. Other transposons, such as Tn7 and Tn3 require multiple transposon-encoded proteins for transposition, again making it more difficult to set up transposition system in heterologous hosts. *In vitro* transposition studies on members of the IS630/Tc1/*mariner* superfamily have demonstrated that transposase is the only protein required to perform transposition (Lampe *et al.*, 1996; Tosi and Beverley, 2000; Vos *et al.*, 1996). It is therefore easy to set up transposition systems in a wide range of host species – from bacteria to *Xenopus* and from flies to mice – with transposons from these families (reviewed in Plasterk *et al.*, 1999).

1.5.2 The transposon ends required to bring about transposition are relatively small

Apart from transposase, the only other requirement for a transposition system is the inverted repeats flanking the transgene to be mobilised. IRs of IS630/Tc1/*mariner* superfamily are often short and the length of DNA required for efficient transposition can be smaller than the inverted repeat size. For instance, only 26 bp out of the 54 bp Tc1 IRs are required for the transposition *in vitro* (Vos *et al.*, 1996). The terminal 100 bp of the 462 bp Tc3 IRs are sufficient to maintain transposition at wild-type level (Fischer *et al.*, 1999). In other cases, for instance *Sleeping Beauty*, longer sequences are needed at both ends. This requirement for relatively short transposon-derived sequences may allow larger payloads to be carried by artificial transposons. For example, GFP and phleomycin resistance reporter genes have been introduced into mini-*Mos1* elements to make detection of transposition easier (Goyard *et al.*, 2001).

1.5.3 Efficient and detectable transposition

Many wild-type transposases from other families show low levels of transposition activity. To be useful as a genetics tool, high levels of transposition activity are often required. In some cases this has been achieved by the selection of hyperactive transposase mutants (Goryshin and Reznikoff, 1998; Lampe *et al.*, 1999; Pledger and Coates, 2005). The large number of inactive remnant copies of transposons from the IS630/Tc1/*mariner* family is a treasure-trove for construction of new and hyperactive transposase variants. Good examples are provided by *SB* and *Himar1*, which were reconstructed from the inactive remains of once active transposons (Ivics *et al.*, 1997; Lampe *et al.*, 1996).

1.5.4 Broad heterologous host range

IS630/Tc1/*mariner* transposons have been found in a wide range of species and there is evidence for horizontal transfer between species, indicating a broad host range. The fact that transposons of this family can horizontally transfer to a broad spectrum of species broadens their potential use as genetic tools. In the laboratory, many genetic transformation experiments using Tc1/*mariner* transposons have been carried out in heterologous species (reviewed in Plasterk *et al.*, 1999). For example, *Mos1* has been used as a transformation vector for *D. melanogaster* (Lidholm *et al.*, 1993), *Leishmania* (Gueiros-Filho and Beverley, 1997), the yellow fever mosquito *Aedes aegypti* (Coates *et al.*, 1998), zebrafish (Fadool *et al.*, 1998) and chicken (Sherman *et al.*, 1998) with varying degrees of success.

As for other transposons, there are some challenges for using members from this superfamily as gene delivery tools. One is that their insertion occurs at essentially random sites in the genome. Insertions may therefore knock out, or inappropriately activate important host genes. Also, insertion into heterochromatin can lead to poor expression. It would therefore be useful to be able to control the precise location of transposon integration in the host genome.

1.6 Project aims

The project described here is on the transposition of ISY100. The aims of this project were

i) To characterise and determine the mechanism of transposition of the bacterial insertion sequence ISY100 *in vivo* and *in vitro*.

ii) To investigate the potential of ISY100 as a site-specific integration vector for gene delivery applications.

Chapter II

Materials and Methods

2.1 Bacterial strains

The derivatives of *E. coli* K-12 or *E. coli* B used are listed in Table 2.1.

Table 2.1 Bacterial Strains

Strain	Genotype	Source
BL21 <DE3>	*E. coli* B F^- *dcm ompT hsdS*($r_B^-m_B^-$) *gal* λ(DE3)	Novagen
BL21 <DE3> pLysS	*E. coli* B F^- *dcm ompT hsdS*($r_B^-m_B^-$) *gal* λ(DE3) [pLysS Cm^r].	Novagen
DH5α	F^- Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA1 endA1 hsdR17*(r_K^-, m_K^+) *phoA supE44 thi-1 gyrA96 relA1*	Invitrogen
DS941	F^- *thr-1 araC14 leuB6* Δ(*gpt-proA62*) *lacY1 tsx-33 glnV44 galK2 hisG4 rfbC1 mgl-51 rpoS396 rpsL31*(Str^R) *kdgK51 xylA5 mtl-1 argE3 thi-1 recF143 lacZ*ΔM15 *lacI*^{q}	D. J. Sherratt
DS964	*thr-1 araC14 leuB6* Δ(*gpt-proA62*) *lacY1 tsx-33 glnV44 galK2 hisG4 rfbC1 mgl-51 recA13 rpoS396 rpsL31*(Str^R) *kdgK51 xylA5 mtl-1 argE3 thi-1* $λ^+$ lysogen	D. J. Sherratt
GM2163	*ara-14 leuB6 fhuA31 lacY1 tsx-78 glnV44 galK2 galT22 mcrA dcm-6, hisG4, rfbD1, R*(*zgb*210::*Tn10*)Tet^S *rpsL136 dam13*::*Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	NEB
TOP10	F^- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80 *lacZ*ΔM15 Δ*lacX74 recA1 araD139* Δ(*ara-leu*)*7697 galU galK rpsL* (Str^R) *endA1 nupG*	Invitrogen

Table 2.2 Oligonucleotides

Name	Sequence	Purpose
AcuI-TA-L Top	5′-AATTCCTGAAGCTAAGAGCGCAGTATAGTCATTTCAATTAACGATGAGAGAATTTG-3′	pXF153
AcuI-TA-L Bot	5′-GATCCAAATTCTCTCATCGTTAATTGAAATGACTATACTGCGCTCTTAGCTTCAGG-3′	pXF153
AcuI-TA-R Top	5′-AGCTTCTGAAGCTAAGAGCGCAGTATAGTAGTTTCAAATAAAGCTGAGACGCTAAG-3′	pXF153
AcuI-TA-R Bot	5′-GATCCTTAGCGTCTCAGCTTTATTTGAAACTACTATACTGCGCTCTTAGCTTCAGA-3′	pXF153
BseRI-TG Top	5′-AATTCGAGGAGGATATCTGTACATCTAGA-3′	TG-ISY100
BseRI-TG Bot	5′-AGCTTCTAGATGTACAGATATCCTCCTCG-3′	TG-ISY100
BseRI-TC Top	5′-AATTCGAGGAGGATATCTCTAGATCTAGA-3′	TC-ISY100
BseRI-TC Bot	5′-AGCTTCTAGATCTAGAGATATCCTCCTCG-3′	TC-ISY100
BseRI-TT Top	5′-AATTCGAGGAGGATATCTTTAAATCTAGA-3′	TT-ISY100
BseRI-TT Bot	5′-AGCTTCTAGATTTAAAGATATCCTCCTCG-3′	TT-ISY100
BseRI-AA Top	5′-AATTCGAGGAGGATATCAATATTTCTAGA-3′	AA-ISY100
BseRI-AA Bot	5′-AGCTTCTAGAAATATTGATATCCTCCTCG-3′	AA-ISY100
BseRI-CA Top	5′-AATTCGAGGAGGATATCCATATGTCTAGA-3′	CA-ISY100
BseRI-CA Bot	5′-AGCTTCTAGACATATGGATATCCTCCTCG-3′	CA-ISY100
BseRI-TA Top	5′-AATTCGAGGAGGATATCTATATATCTAGA-3′	TA-ISY100
BseRI-TA Bot	5′-AGCTTCTAGATATATAGATATCCTCCTCG-3′	TA-ISY100
YIRR58 Top	5′-ACGAATTCCTGAAGAGGAGGATCAGCTATAGTAGTTTCAAATAAAGCTGAGACGCTAA-3′	IRR-58
YIRR58 Bot	5′-TTAGCGTCTCAGCTTTATTTGAAACTACTATAGCTGATCCTCCTCTTCAGGAATTCGT-3′	IRR-58
YIRL58 Top	5′-ACGAATTCCTGAAGAGGAGGATCAGCTATAGTCATTTCAATTAACGATGAGAGAATTT-3′	IRL-58
YIRL58 Bot	5′-AAATTCTCTCATCGTTAATTGAAATGACTATAGCTGATCCTCCTCTTCAGGAATTCGT-3′	IRL-58
YIRL79 Top	5′-TAGTTTCTAAAGTAAAAATTAAGAGGTATAGTCATTTCAATTAACGATGAGAGAATTTAATGTAAAATTATGGAGTGTA-3′	IRL-79

YIRL79 Bot	5′- TACACTCCATAATTTTACATTAAATTCTCTCATCGTTAATTGAAATGACTATACCTCTTAATT TTTACTTTAGAAACTA-3′	IRL-79
YIRR79 Top	5′- CTCATCCGTATAATGCTTTCTCGTATTATAGTAGTTTCAAATAAAGCTGAGACGCTAAACGCC ACAGTAAGAACGGATA-3′	IRR-79
YIRR79 Bot	5′- TATCCGTTCTTACTGTGGCGTTTAGCGTCTCAGCTTTATTTGAAACTACTATAATACGAGAAA GCATTATACGGATGAG-3′	IRR-79
YIRL-52/51-Top	5′-ATAGTCATTTCAATTAACGATGAGAGAATTTAATGTAAAATTATGGAGTGTA-3′	Integration assay
YIRL-49/51-Top	5′-GTCATTTCAATTAACGATGAGAGAATTTAATGTAAAATTATGGAGTGTA-3′	Integration assay
YIRL-51/51-Bot	5′-TACACTCCATAATTTTACATTAAATTCTCTCATCGTTAATTGAAATGACTA-3′	Integration assay
YIRL-50/51-Bot	5′-TACACTCCATAATTTTACATTAAATTCTCTCATCGTTAATTGAAATGACT-3′	Integration assay
RI-Z1-TANN Top	5′-AATTCGCGTGGGCGTTAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z+1)
RI-Z1-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTAACGCCCACGCG-3′	TA array (Z+1)
RI-Z2-TANN Top	5′-AATTCGCGTGGGCGTCTAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z+2)
RI-Z2-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTAGACGCCCACGCG-3′	TA array (Z+2)
RI-Z3-TANN Top	5′-AATTCGCGTGGGCGTCATAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z+3)
RI-Z3-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTATGACGCCCACGCG-3′	TA array (Z+3)
RI-Z4-TANN Top	5′-AATTCGCGTGGGCGTCAGTAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z+4)
RI-Z4-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTACTGACGCCCACGCG-3′	TA array (Z+4)
RI-rZ1-TANN Top	5′-AATTCACGCCCACGCTAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z-1)
RI-rZ1-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTAGCGTGGGCGTG-3′	TA array (Z-1)
RI-rZ2-TANN Top	5′-AATTCACGCCCACGCCTAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z-2)
RI-rZ2-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTAGGCGTGGGCGTG-3′	TA array (Z-2)
RI-rZ3-TANN Top	5′-AATTCACGCCCACGCCATAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z-3)
RI-rZ3-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTATGGCGTGGGCGTG-3′	TA array (Z-3)
RI-rZ4-TANN Top	5′-AATTCACGCCCACGCCAGTAGATATCTAAGTAAATACGTATGTATGTATATAGAA-3′	TA array (Z-4)
RI-rZ4-TANN Bot	5′-AGCTTTCTATATACATACATACGTATTTACTTAGATATCTACTGGCGTGGGCGTG-3′	TA array (Z-4)

Table 2.3 Primers

Name of primer	Sequence	Purpose
6803_Top	5′-ggggaattcATATGTCCGGACTTCGCTATAGTTTCTAAAG-3′	pXF106
6803_Bot	5′-cccggatccATGGTCTAGATGTCCTCATCCGTATAATGC-3′	pXF106
y100-51_Top	5′-gggctcgaGGATCCTACACTCCATAATTTTAC-3′	ISY100-51
y100-51_Bot	5′-cccagatctGGATCCTCCGTTCTTACTGTGGC-3′	ISY100-51
y100-30_Top	5′-gggctcgaGGATCCAAATTCTCTCATCGTTAATTG-3′	ISY100-30
y100-30_Bot	5′-cccagatctGGATCCTTAGCGTCTCAGCTTTATTTG-3′	ISY100-30
Tnp100_Top	5′-ggaattccatATGGCTTACAGTTTAGA-3′	Tnp
Tnp100_Bot	5′-cccgcatgcggatccCTAAACGCCACAGTAAGAACGGAT-3′	Wild-type Tnp
Tnp100_Chis	5′-cccggatccCTAGTGGTGGTGGTGGTGGTGAACGCCACAGTAAGAACGGAT-3′	Tnp-His_6
EagI-BamHI	5′-AAGGATCCGGTACCCTAACGGCCGTAAACGCCACAGTAAGAACGG-3′	Z-transposase
AgeI-Hi6-BamHI	5′-GAACCGGTCACCACCACCACCACCACTAGGGATCCTCTAGAG-3′	Truncated Tnp
ISY100N37	5′-GAACCGGTTCTGTATATCGAGGCTTTTCC-3′	Tnp_{1-37}
ISY100N38	5′-GAACCGGTCCATCTGTATATCGAGGCTTTTCC-3′	Tnp_{1-38}
ISY100N46	5′-GAACCGGTTGGGCTTAAATCTACTCTATTTAACC-3′	Tnp_{1-46}
ISY100N77	5′-GAACCGGTGGCTCTGTCTATCAATCTTGC-3′	Tnp_{1-77}
ISY100N95	5′-GAACCGGTTTTCATTTTCTTTAATGCG-3′	Tnp_{1-95}
ISY100N57	5′-GAACCGGTGTCTAATTTCCTATGGCGACGC-3′	Tnp_{1-57}
ISY100N68	5′-GAACCGGTATTTTCTTCTACGTCTTTTTTTAG-3′	Tnp_{1-68}
ISY100N110	5′-GAACCGGTCCGGTTTCTTTCTCGATAACG-3′	Tnp_{1-110}
ISY100 D136A-Top	5′-GTCAAGCTATAGTTTACATAGCAGAATCTGGATTCGAAGCAATCC-3′	Tnp D136A
ISY100D136A-Bot	5′-GGATTGCTTCGAATCCAGATTCTGCTATGTAAACTATAGCTTGAC-3′	Tnp D136A
ISY100 D218A-Top	5′-GACATTCCATCAATATTAATAATGGCCAATGCTCCTATTCATCGTAAAACTG-3′	Tnp D218A
ISY100 D218A-Bot	5′-CAGTTTTACGATGAATAGGAGCATTGGCCATTATTAATATTGATGGAATGTC-3′	Tnp D218A
ISY100 E253A-Top	5′-CTCCTGATTTAAATGATATTGCGCATGACTTTAGTGCCTTG-3′	Tnp E253A
ISY100 E253A-Bot	5′-CAAGGCACTAAAGTCATGCGCAATATCATTTAAATCAGGAG-3′	Tnp E253A
ISY100 K260A-Top	5′-GCATGACTTTAGTGCCTTGGCTCGAGCTAGAATGTACGCTC-3′	Tnp K260A
ISY100 K260A-Bot	5′-GAGCGTACATTCTAGCTCGAGCCAAGGCACTAAAGTCATGC-3′	Tnp K260A
ISY100 R261A-Top	5′-GAGCATGACTTTAGTGCCTTGAAGGCAGCTAGAATGTACGCTCCTATTG-3′	Tnp R261A
ISY100 R261A-Bot	5′-CAATAGGAGCGTACATTCTAGCTGCCTTCAAGGCACTAAAGTCATGCTC-3′	Tnp R261A

ISY100 KR260AA-Top	5′-GATATTGAGCATGACTTTAGTGCCTTGGCTGCAGCTAGAATGTACGCTCCTATTGACAC-3′	Tnp K260A, R261A
ISY100 KR260AA-Bot	5′-GTGTCAATAGGAGCGTACATTCTAGCTGCAGCCAAGGCACTAAAGTCATGCTCAATATC-3′	Tnp K260A, R261A
Kan-L	5′-AAGACGTTTCCCGTTGAATATGG-3′	Sequencing
Kan-R	5′-ATTGCAGTTTCATTTGATGCTCG-3′	Sequencing
polyG	5′-cccccatccatatgaagcttGGGGGGGGGGGG-3′	Mapping cleavage site
nest-polyG	5′-cccccatccatatgaagcttG-3′	Mapping cleavage site
Uni-MluI	5′-TCAGGACGCGTCAGCGGGTGTTG-3′	Mapping cleavage site
Rev-ClaI	5′-TGAGCATCGATTTTTGTGATGCTC-3′	Mapping cleavage site
Uni_103	5′-GCTATTACGCCAGCTGGCGAAAG-3′	PCR and sequencing
Uni_203	5′-AAATACCGCACAGATGCGTAAGG-3′	PCR and sequencing
Rev 108	5′-CATTAGGCACCCCAGGCTTTACAC-3′	PCR and sequencing
Rev 308	5′-CCTTTGAGTGAGCTGATACCGCTC-3′	PCR and sequencing
pCB104-sphI_Top	5′-ctgttccgctgggcGtgccaggacaacttc-3′	Remove one SphI site on pCB104
pCB104-sphI_Bot	5′-gaagttgtcctggcaCgcccagcggaacag-3′	Remove one SphI site on pCB104

Table 2.4 Plasmid list

Plasmid	Description (antibiotic resistance)	Source/ Reference
pUC18	general purpose, pBR322 derived vector (Amp)	Yanisch-Perron *et al.*, 1985
pUC19	general purpose, pBR322 derived vector (Amp)	Yanisch-Perron *et al.*, 1985
pUC4K	Kanamycin resistance cassette, pBR322 derived vector (Amp, Kan)	Vieira and Messing, 1982
pCB104	λ-dv derived vector with pUC18 polylinker (Cm)	Boyd and Sherratt, 1995
pH2	pIC20 containing a 2.6 kb TA-rich sequence from *Aspergillus nidulans* (Amp)	Aleksenko and Clutterbuck, 1996
pTrc99a	Expression plasmid, carries a strong hybrid trp/lac promoter (Amp)	Amann *et al.*, 1988
pKET3a	pET3a derived expression vector (Kan)	Studier *et al.*, 1990; S.J.Rowland unpublished
pXF80	pCB104 derived, one SphI site was eliminated (Cm)	Chapter III
pXF106	pUC18 containing ISY100a (Amp)	Chapter III
pXF101	Expression vector pKET3a + ISY100 transposase gene (Kan)	Chapter IV
pXF102	Expression vector pKET3a + C-terminal His_6-tagged ISY100 transposase gene (Kan)	Chapter IV
pXF104	Expression vector pTrc99a + ISY100 transposase gene (Amp)	Chapter III & Chapter IV
pXF105	Expression vector pTrc99a + C-terminal His_6-tagged ISY100 transposase gene (Amp)	Chapter III & Chapter IV
pXF109	pUC18 + ISY100-51-Kan (Amp)	Chapter III & Chapter IV
pXF113	pXF75 + ISY100-51-Kan. (Kan, Cm)	Chapter III
pXF114	pXF75 + ISY100-30-Kan. (Kan, Cm)	Chapter III
pXF115	PXF113 + inducible wild-type transposase gene (Kan, Cm)	Chapter III
pXF116	PXF114 + inducible wild-type transposase gene (Kan, Cm)	Chapter III
pXF117	PXF113 + inducible Tnp-His_6 transposase gene (Kan, Cm)	Chapter III
pXF118	PXF114 + inducible Tnp-His_6 transposase gene (Kan, Cm)	Chapter III

pXF119	pUC18 + 51 bp ISY100 IRL (Amp)	Chapter III & Chapter IV
pXF120	pUC18 + 30 bp ISY100 IRL (Amp)	Chapter III
pXF121	pUC18 + 49 bp ISY100 IRR (Amp)	Chapter III & Chapter IV
pXF122	pUC18 + 30 bp ISY100 IRR (Amp)	Chapter III
pAN2	resolvase-Zif268 fusion plasmid	Akopian *et al.*, 2003
pAN14	resolvase-$(GSG)_3$-Zif268 fusion plasmid	Akopian *et al.*, 2003
pAN15	resolvase-$(GSG)_4$-Zif268 fusion plasmid	Akopian *et al.*, 2003
pXF123	pUC18 + wild-type transposase gene with KpnI-EagI-BamHI polylinker upstream of the stop codon. (Amp)	Chapter V
pXF124	pXF123 + KpnI-EagI fragment from pAN15, Tnp-$(GSG)_4$-Zif268 DBD (Amp)	Chapter V
pXF125	pXF123 + KpnI-EagI fragment from pAN14, Tnp-$(GSG)_3$-Zif268 DBD (Amp)	Chapter V
pXF127	pXF123 + KpnI-EagI fragment from pAN2, Tnp-Tn3 DBD-Zif268 DBD (Amp)	Chapter V
pXF128-132	pTrc99a + His_6-tagged Tnp_{1-37}, Tnp_{1-38}, Tnp_{1-46}, Tnp_{1-77} and Tnp_{1-95} (Amp)	Chapter IV
pXF163, pXF133-134, pXF136	pTrc99a + transposase derivatives from pXF123-pXF125 and pXF127. (Amp)	Chapter V
pXF141-145	pKET3a + His_6-tagged Tnp_{1-37}, Tnp_{1-38}, Tnp_{1-46}, Tnp_{1-77} and Tnp_{1-95} (Kan)	Chapter IV
pXF147	pUC18 + 30 bp ISY100 IRL (Amp)	Chapter III
pXF149	pUC18 + 30 bp ISY100 IRR (Amp)	Chapter III
pXF153	pUC18 + ISY100-30-Kan (Amp, Kan)	Chapter III
pXF164-166,168	pXF114 + Tnp-Zif268 expression cassettes from pXF163, pXF133,pXF134 and pXF136 (Kan, Cm)	Chapter V
pXF169-174	Expression vector pTrc99a + transposase mutants D136A, D218A, E253A, K260A, R261A or KR260AA (Amp)	Chapter III
pXF195	pUC19 + ISY100-30-Kan (Amp)	Chapter III
pXF198	pUC-based plasmid + mini-ISY100 with 30 bp IRR at both ends (Amp)	Chapter III
pXF199	pUC-based plasmid + mini-ISY100 with 30 bp IRL at both ends (Amp)	Chapter III
pXF200-207	pUC18 + Z+1 to Z+4 and Z-1 to Z-4 target $(TANN)_9$ array sequences (Amp)	Chapter V
pXF208-213	pUC18 + HindIII-BseRI-EcoRI polylinkers with different dinucleotides flanking BseRI cut sites. (Amp)	Chapter III

pXF214-219	pXF208-pXF213 + ISY100-30-Kan (Amp,Kan)	Chapter III, IV
pXF220	pUC18 + mini-ISY100 with 30 bp IRR at both ends (Kan, Amp)	Chapter III, IV
pXF226	pUC18 + mini-ISY100 with 30 bp IRL at both ends (Kan, Amp)	Chapter III, IV
pZ+1 to pZ+4, pZ-1 to pZ-4	pH2 + Z+1 to Z+4 and Z-1 to Z-4 target sequences. (Amp)	Chapter V
pXF240-242	pTrc99a-based expression plasmids for His_6-tagged Tnp_{1-57}, Tnp_{1-68} and Tnp_{1-110} (Amp)	Chapter III
pXF243-245	pKET3a-based expression plasmids for His_6-tagged Tnp_{1-57}, Tnp_{1-68} and Tnp_{1-110} (Kan)	Chapter IV
pXF246-251	pXF114 + inducible mutated transposase gene for D136A, D218A, E253A, K260A, R261A and KR260AA (Kan, Cm)	Chapter III
pXF252-257	pXF116-like plasmids. Contain ISY100-30-Kan flanked by dinucleotides TA, TG, TC, TT, AA and CA at both ends. (Kan, Cm)	Chapter III
pXF258	pXF116-like plasmid. mini-ISY100 has 30 bp IRR at both ends (Kan, Cm)	Chapter III
pXF259	pXF116-like plasmid. mini-ISY100 has 30 bp IRL at both ends (Kan, Cm)	Chapter III
pXF262-264	pUC18 + mini-ISY100 with 30 bp IRR at both ends. IRRs at right end are flanked by TA and left IRRs are flanked by TT, AA and CA. (Amp)	Chapter III, IV
pXF265	pXF116-like plasmid. TG at IRL, TA at IRR of the ISY100-30-Kan. (Kan, Cm)	Chapter III
pXF266	pXF116-like plasmid. TC at IRL, TA at IRR of ISY100-30-Kan. (Kan, Cm)	Chapter III
pXF267-269	pXF116-like plasmids. Mini-ISY100 has two 30 bp IRR at both ends. IRRs at right end are flanked by TA and left IRRs are flanked by dinucleotides TT, AA and CA. (Kan, Cm)	Chapter III

2.2 Oligonucleotides

Oligonucleotides used for construction of plasmids and *in vitro* transposition assays are listed in Table 2.2. They were synthesized by MWG Biotech.

2.3 Primers

Non-commercial primers used for sequencing and primers for PCR or PCR-based site-directed mutagenesis are listed in Table 2.3. They were synthesized by MWG Biotech. Commercial primers for sequencing were supplied by MWG Biotech.

2.4 Plasmids

The main plasmids used are listed in Table 2.4.

2.5 Chemicals and buffer solutions

2.5.1 Chemicals

Sources of general chemicals are listed below.

Chemical	Source
General chemicals, Biochemicals, Organic solvents	Sigma-Aldrich, Bio-Rad, VWR, New England Biolabs, Promega and Invitrogen
Media	Difco, Oxoid
Agarose	BRL/Invitrogen, Cambrex Bio Science Rockland
37:1 or 19:1 acrylamide: bisacrylamide solutions	VWR Acrylogel, Sigma
Radiochemicals	Amersham Pharmacia Biotech

2.5.2 Buffer solutions

All solutions were made using deionised or double distilled water. All solutions for protein purification were made using double distilled water. Standard solutions are listed below and will be shown as their abbreviations in the following text. All percentages represent w/v measurements unless otherwise stated.

Acrylamide gel elution buffer	0.5 M ammonium acetate, 1 mM EDTA (pH 8.0)
Agarose gel loading dye (5×)	30% glycerol, 3 mM Tris, 0.3 mM EDTA, 0.5% SDS and 0.25% bromophenol blue. Proteinase K was added to 0.2 mg/ml if necessary.
Alkaline agarose gel loading buffer (6×)	300 mM NaOH, 6 mM EDTA, 18% Ficoll (Type 400, Pharmacia), 0.15% bromocresol green, 0.25% xylene cyanol
Alkaline gel electrophoresis buffer (10×)	500 mM NaOH, 10 mM EDTA
Ammonium persulphate (APS)	10% ammonium persulphate in water made fresh before each use.
Buffer A for purification of ISY100 transposase	50 mM Tris, 14.3 mM β-mercaptoethanol, 1 M NaCl, 0.2% triton X-100
Commercial enzyme buffers	As supplied with enzymes
Doly I	50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA
Doly II	0.2 M NaOH, 1% SDS
Doly III	3 M potassium acetate, 2 M acetic acid
Formamide loading buffer (2×)	20 mM EDTA (30 mM EDTA for footprinting assays), 0.1% Xylene cyanol, 0.1% bromophenol blue in 99% deionised formamide
L-Broth (Luria broth)	1% Bacto tryptone, 0.5% Bacto yeast extract, 0.5% NaCl, adjust to pH 7.5 with NaOH.
Neutralizing solution for alkaline agarose gels	1 M Tris-HCl (pH 7.6), 1.5 M NaCl
Single colony gel loading buffer	2.5% Ficoll, 1.25% SDS, 0.1% Bromophenol Blue, 1× TAE
SDS-PAGE loading buffer (2×)	62.5 mM Tris-HCl (pH 6.8), 25% glycerol, 2% SDS (w/v), 0.1% Bromophenol blue (w/v) and 0.5% β-mercaptoethanol

SDS-PAGE running buffer (10×)	0.25 M Tris, 1.92 M Glycine and 0.1% SDS
SDS-PAGE staining buffer	50% methanol, 10% acetic acid, 0.1% Coomassie brilliant blue
SDS-PAGE destaining buffer	10% methanol, 10% acetic acid
SOC	2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.
TAE (50×)	2 M Tris base, 1 M acetic acid, 0.05 M EDTA (pH 8.0) and 1 M sodium acetate
TBE (5×)	0.445 M Tris base, 0.445 M boric acid, 10 mM EDTA
TTE (10×)	0.89 M Tris base, 5 mM EDTA, 0.29 M taurine.
TE	10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)
1/10 TE	1 mM Tris-HCl (pH 8.0), 0.1 mM EDTA (pH 8.0)
Tris-tricine gel buffer (3×)	3.0 M Tris-HCl (pH 8.45), 0.3% SDS
Tricine gel cathode buffer (5×)	0.5 M Tris base, 0.5 M Tricine, 0.5% SDS
Tricine gel anode buffer (5×)	1.0 M Tris-HCl (pH 8.9)
Tris-tricine gel loading buffer (2×)	100 mM Tris-HCl (pH6.8), 4% SDS, 10% glycerol, 5% β-mercaptoethanol

2.6 Kits

QIAprep Spin Miniprep Kit, QIAquick Gel Extraction Kit, MinElute Gel Extraction Kit and QIAquick PCR Purification Kit were obtained from Qiagen. The TOPO TA cloning kit was supplied by Invitrogen.

2.7 Bacterial growth media

E. coli strains were routinely grown in liquid L-Broth or on solid L-Broth agar (L-Broth with 15 g/L agar). Media for bacterial growth were sterilised at 121°C, 15 pounds/inch^2 for 15 minutes.

2.8 Bacterial growth conditions

Bacteria in L-Broth normally grew with shaking at 37°C. For expression of transposase, cells were grown with shaking at 30°C. Bacteria on LB Agar plates were incubated at 37°C overnight. For long term storage of bacterial strains, 1 ml of overnight culture was added to 0.7 ml of 40% glycerol, 2% peptone solution and kept at -70°C.

2.9 Competent *E. coli* cells

2.9.1 Commercial chemical competent cells

Commercial chemical competent cells were obtained from Invitrogen Ltd. Commercial competent cells were transformed following protocols provided by the manufacturer.

2.9.2 $CaCl_2$ competent cells

A single colony of *E. coli* was inoculated into 20 ml L-Broth with appropriate antibiotics and incubated at 37°C with shaking overnight. 200 µl of this overnight culture was used to inoculate 20 ml fresh L-Broth containing appropriate antibiotics. This was incubated in a shaking incubator at 37°C and 200 rpm for 90 minutes. The cells were harvested by centrifugation at 6,400 g for 1 minute, the supernatant was discarded and the cell pellet was resuspended with 10 ml precooled 50 mM $CaCl_2$. This was incubated on ice for 1 hour and the cells were again centrifuged at 6,400 g for 1 minute, and resuspended in 1 ml precooled 50 mM $CaCl_2$. This yielded 1 ml competent cells, which were stored on ice until use for transformation as described in section 2.10.1.

2.9.3 Electrocompetent cells

Electrocompetent cells for the electroporation were prepared according to the protocol provided by Bio-Rad Laboratories Ltd. (Hertfordshire, UK). Briefly, 500 ml of cells grown shaking at 37°C to an OD_{600} at 0.5-0.7 was kept on ice for about 20 minutes and then was centrifuged at 4,000 g for 15 minutes at 4°C. The cell pellet was washed at least three times with cold sterile 10% glycerol. Cells were resuspended in a final volume of 1-2 ml of ice-cold sterile 10% glycerol to give 1-3 $\times$ 10^{10} cells per ml and were transformed as described in section 2.10.2.

2.10 Transformation

2.10.1 Transformation of $CaCl_2$ competent cells

Plasmid DNA or ligation product was transformed into 50-100 µl of $CaCl_2$ chemically competent cells. The DNA sample and competent cells were mixed gently in an eppendorf tube and the mixture was incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 90 seconds, and then were incubated on ice for a further 5 minutes. 200 µl of L-Broth was added and the cells were incubated at 37°C for 1 hour to allow expression of the antibiotic resistance genes carried on the plasmid. Aliquots of the culture were spread onto selective L-Broth agar plates.

2.10.2 Electroporation

DNA samples were treated as follows before use in electroporation: For plasmid DNA purified using the QIAprep Spin Miniprep Kit, 1/10 TE instead of buffer EB was used to elute DNA from the spin column in the last step. Ligation or *in vitro* transposition reaction products were precipitated with ethanol after the addition of 1 µg of yeast tRNA. The DNA pellet was then resuspended in 10 µl of 1/10 TE. 1-2 µl of this DNA was added to 40 µl of electrocompetent cells in a 0.2 cm cuvette (Bio-Rad) and left on ice for 5 minutes. The cells were then pulsed at 2.5 kV using a Gene Pulser (Bio-Rad) set on E2 following the manufacturer's instructions. 1 ml of SOC was added to the cuvette immediately and mixed gently. The culture was then transferred into a 15 ml centrifuge tube and was incubated in a shaking incubator (200 rpm) at 37°C for 1 hour. Aliquots of the culture were spread onto selective L-Broth agar plates.

2.11 Antibiotics

Antibiotics used were supplied by Sigma and are listed below.

Antibiotic (abbreviation)	Stock solution	Working concentration
Ampicillin (Amp)	100 mg/ml in H_2O	100 µg/ml
Kanamycin (Kan)	25 mg/ml in H_2O	25 µg/ml
Chloramphenicol (Cm)	25 mg/ml in ethanol	25 µg/ml

2.12 Preparation of DNA

2.12.1 Large-scale preparation of plasmid DNA

Plasmid DNA was prepared following a modified version of the method of Birnboim and Doly (1979). 200 ml of overnight culture was centrifuged at 4,000 g (5 minutes at 4°C) and the cell pellet was resuspended in 4 ml of Doly I buffer on ice. 8 ml of freshly made Doly II buffer was added and mixed in by inverting the tube gently. The tube was held on ice for 4 minutes to allow full lysis. The mixture was neutralised by the addition of 6 ml of cold Doly III, and then mixed gently. The samples were centrifuged at 39,200 g, 4°C for 30 minutes. The supernatant was transferred into new tubes and plasmid DNA was precipitated by adding 12 ml of isopropanol at room temperature for 15 minutes. The precipitated DNA was recovered by centrifugation at 39,200 g, 20°C for 30 minutes. The DNA pellets were washed with 2 ml of 70% of ethanol to remove the remaining salts and then were resuspended in 2 ml of TE buffer and incubated for 1 hour at 37°C. The insoluble debris was removed by centrifugation (17,400 g, 5 minutes, 4°C) and the soluble plasmid DNA solution was ready for CsCl equilibrium density gradient centrifugation.

Ethidium bromide CsCl equilibrium density gradient centrifugation

DNA in TE (2 ml) was transferred into a Beckman quick-seal ultra-centrifuge tube (16 × 76 mm). 4.324 ml of CsCl solution (5 g CsCl plus 3 ml H_2O) and 0.27 ml of an ethidium bromide solution (10 mg/ml) were then added. The tubes were filled with liquid paraffin and then sealed. The CsCl density gradient was achieved by centrifugation in a Beckman Ti70 fixed angle rotor at 45,000 rpm for 16 hours at 18°C (with no brake). The supercoiled plasmid DNA was visualised on a long-wave UV source (365 nm) and was recovered by piercing the tube immediately below the desired band with a 21 G hypodermic needle attached to a 1 ml syringe. About 0.5 ml of sample was recovered and the ethidium bromide was removed by extraction with water-saturated n-butanol. CsCl was removed from samples by dialysis against 400 volumes of TE three times at 4°C. The DNA samples were stored at -20°C.

2.12.2 Small-scale preparation of plasmid DNA

Plasmid DNA was purified from 3 ml of culture using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instruction. The purified plasmid DNA was normally eluted from the column with 50 μl of the supplied EB buffer or 30 μl of 1/10 TE for use in electroporation.

2.13 Restriction enzyme digestion of DNA

Restriction enzyme digestions were carried out according to the suppliers' instructions. Partial digestions were accomplished by reducing the enzyme concentration and/or reducing the reaction time.

2.14 Electrophoresis

2.14.1 Agarose gel electrophoresis

Native agarose gels (1.0% or 1.2%) were prepared by mixing appropriate amounts of agarose and TAE buffer and dissolving by heating in a microwave oven. The solution was allowed to cool to ~ 60°C and poured into a gel former with an appropriate comb. After the gel was set, the comb was removed and the gel was run at 40 V for 16 hours or 100 V for 4 hours in a gel tank containing 1× TAE buffer just covering the gel surface. The gel was stained and then visualized as described in section 2.15.1.

Alkaline agarose gel electrophoresis was performed to separate single stranded DNA. To make up an alkaline 1.33% (w/v) agarose gel, agarose was dissolved in H_2O in a microwave oven. After cooling, $1/9^{th}$ volume of 10× alkaline agarose gel electrophoresis buffer was mixed in and the gel was poured immediately. After the gel was set, it was transferred to a gel tank containing freshly made 1× alkaline electrophoresis buffer just covering the gel. The gel was run at <3.5 V/cm. After the bromocresol green had migrated approximately 0.5-1 cm into the gel, the power was turned off and a glass plate was placed on top of the gel. Electrophoresis was then continued until the bromocresol green had migrated approximately 2/3 of the length of the gel. The gel was soaked in neutralizing buffer for 45 minutes at room temperature before staining with ethidium bromide and visualisation as described in section 2.15.1.

2.14.2 Non-denaturing polyacrylamide gel electrophoresis

Transposase electrophoretic mobility shift assays (EMSA) were performed using non-denaturing polyacrylamide gel electrophoresis. A 30 ml solution containing 6% acrylamide: bis-acrylamide in a 37:1 ratio and 1/3× TBE was prepared. Polymerisation was initiated by the addition of 18 μl TEMED and 360 μl 10% ammonium persulphate (APS). The comb was removed and the gel was assembled in a vertical gel tank with 1/3× TBE in top and bottom reservoirs. The gel was pre-run for 45 min at 200 V and 4°C. After loading the samples, the gel was run at a constant voltage (200 V) for 2-4 hours at 4°C. Non-denaturing polyacrylamide gel electrophoresis in 1× TBE was also used to separate DNA fragments, running at 200 V for 2 hours at room temperature.

2.14.3 Denaturing polyacrylamide gel electrophoresis

DNase I footprinting and mapping the cleavage sites on the transposon ends were carried out on denaturing polyacrylamide gels using the BRL Sequencing System, Model S2. 80 ml of gel mixture containing 8% acrylamide: bisacrylamide (19:1), 7.5 M urea in 1× TTE was prepared. Polymerisation was initiated by the addition of 800 μl of 10% APS and 21.6 μl of TEMED. The gel was poured using 0.1 mm spacers, allowed to polymerise at room temperature, and was then set up according to the manufacturer's instructions. The gel was pre-run at 80 W for 45 minutes with 1× TTE buffer. An equal volume of formamide loading buffer (2×) was added to the samples and they were heated at 80°C for 5 minutes prior to loading. The gel was run for 1.5-3 hours at 80 W. The gel was dried under vacuum and autoradiography was carried out as described in Section 2.32.

Purification of synthetic oligonucleotides was carried out on denaturing polyacrylamide gels using CBS Scientific Co. Adjustable Slab gel kits (model # ASG-250). 40 ml of gel mixture was made containing 10% acrylamide: bisacrylamide (19:1), 7.5 M urea and 1× TBE. Polymerisation was initiated by the addition of 400 μl of 10% fresh APS and 12.5 μl of TEMED. The gel kit was set up according to the manufacturer's instructions. The gel was pre-run at 400 V. An equal volume of formamide loading buffer (2×) was added to the samples and they were heated to 70°C, 5 min prior to loading. The gel was run at 400 V for 2 hours. The desired oligonucleotides were recovered from the gel as described in section 2.26.

2.14.4 Discontinuous SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

A discontinuous SDS-PAGE system (Laemmli, 1970) was used to analyse protein samples. Gels were prepared in two steps. First, 30 ml of resolving gel containing 10% acrylamide: bisacrylamide (37:1), 0.375 M Tris-HCl (pH 8.8) and 0.1% SDS was prepared. Polymerisation was initiated by addition of 160 µl 10% APS and 16 µl of TEMED. The gel was poured and overlaid with 1 ml of isopropanol and allowed to polymerise for 45 minutes at room temperature. After polymerisation, 9 ml of stacking gel containing 4% acrylamide: bisacrylamide (37:1), 0.125 M Tris-HCl (pH 6.8), 0.1% SDS was prepared. Polymerisation was initiated by addition of 60 µl of freshly made 10% APS and 6 µl of TEMED. The stacking gel was poured on top of the resolving gel after removing the isopropanol overlay, and the comb was added immediately. After polymerisation, the comb was removed and the wells were rinsed with 1× SDS – PAGE running buffer.

An equal volume of 2× SDS-PAGE loading buffer was added to the protein samples and they were boiled for 2 minutes before loading. The gel was run at 120-150 V and stained as described in section 2.15.2.

2.14.5 Tricine SDS-PAGE

Cell extracts containing N-terminal transposase derivatives were analysed using a discontinuous tricine SDS-PAGE (Schagger and von Jagow, 1987). 30 ml of separating gel mixture containing 16.5% acrylamide: bisacrylamide (19:1), 10.7% glycerol in 1× Tricine gel buffer and 9 ml of stacking gel mixture containing 4% acrylamide: bisacrylamide (37:1) in 1× Tricine gel buffer were prepared. Polymerisation was initiated by addition of 150 µl 10% APS and 15 µl TEMED to the separating gel mix and 50 µl 10% APS and 10 µl TEMED to the stacking gel mix. The separation gel was poured first, the stacking gel was poured immediately onto the top of the separating gel, and the comb was then inserted. After polymerisation, the comb was removed and the tricine gel was set up with 1× cathode buffer in the top tank and 1× anode buffer in the bottom tank. Protein samples were mixed with an equal volume of 2× Tricine gel sample buffer and were boiled for 2 minutes before loading on the gel. The gel was run at 30 mA for 16 hours and stained as described in section 2.15.2.

2.15 Visualising DNA and protein

2.15.1 Visualising DNA

After electrophoresis, agarose or polyacrylamide gels were stained with an ethidium bromide solution (0.5 μg/ml) for 30-60 minutes, and then were destained with deionised water for 60 minutes. DNA bands on the gels were visualised on a 254 nm UV short wavelength transilluminator, or on a 365 nm UV long wavelength transilluminator for preparative work. Gels were photographed as described in 2.32.

2.15.2 Visualising Proteins

Proteins on Laemmli gels were visualised by incubating gels in the SDS-PAGE staining buffer for 60 minutes with shaking. Gels were then destained in several changes of SDS-PAGE destaining buffer.

2.16. Ligation of DNA fragments

Ligation of insert DNA fragment into linear vector DNA was performed with a molar ratio of 10:1 insert to vector. Ligation of double stranded oligonucleotides into vector was performed with a molar ratio of 3:1 insert to vector. A standard DNA ligation was carried out in 1× ligation buffer (NEB) with 33 units of T4 DNA ligase (NEB) in a volume of 10 μl. The mixture was incubated at room temperature for 2-16 hours and then used to transform competent cells as described in section 2.10.

2.17 Over-expression of ISY100 transposase and its derivatives

To express ISY100 transposase and its derivatives in *E. coli*, 4 ml of overnight culture of the expression strain BL21 (DE3) pLysS [or BL21 (DE3)] carrying the appropriate transposase expression plasmid was inoculated into pre-warmed (37°C or 30°C) L-broth (400 ml) supplemented with kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml) [using kanamycin only for the strain BL21 (DE3)] in a 2 L flask. The cells were grown with vigorous shaking (200 rpm) at 37°C (or 30°C) to an OD_{600} of 0.4-0.6. Transposase expression was induced by the addition of 0.5 mM IPTG and the cells were grown for a further 2 hours. Cells were harvested by centrifugation at 4,000 g for 5 minutes at 4°C in a precooled JA14 rotor. The supernatant was discarded and the cell pellet was resuspended on ice in 100 ml of TME buffer (20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 0.1 mM EDTA).

The cells were pelleted (4,000 g, 5 minutes, 4°C) and the supernatant was removed. Cell pellets were stored at –70°C.

2.18 Purification of ISY100 transposase and its derivatives

All the purification steps were carried out at 4°C or on ice unless otherwise indicated. One gram of frozen induced cells (from section 2.17) was resuspended in 10 ml TBP buffer (50 mM Tris-HCl pH 7.5, 14.3 mM β-mercaptoethanol, 1 mM PMSF). Cells were sonicated three times for 15 seconds using a Vibra-cell VC100 sonicator at 30% amplitude with a button probe 2. The cells were allowed to cool on ice between the bursts. An equal volume of TB buffer (50 mM Tris-HCl pH 7.5, 14.3 mM β-mercaptoethanol) was added and the solution was made up to 10 mM $MgCl_2$ and gently vortexed. After 5 minutes on ice, the mixture was centrifuged at 28,300 g for 15 minutes and the supernatant was transferred to a new tube. Over a period of more than 5 minutes, 200 mM spermine was added drop by drop to a final concentration of 2.5 mM with stirring. Stirring was continued for a further 10 minutes and the mixture was then centrifuged at 16,000 g for 5 minutes. The pellet was resuspended in 10 ml TB buffer, and an equal volume of TB + 2 M NaCl + 0.4% triton X-100 was added drop by drop with gentle vortexing. The mixture was kept on ice for 1 hour and was centrifuged at 16,000 g for 15 minutes. The supernatant, containing soluble transposase, was applied onto a Ni-NTA (Qiagen) column, pre-charged with $NiSO_4$ according to the manufacturer's instructions, for further purification.

All steps of column chromatography were monitored by UV absorbance at 260 nm and 280 nm. After loading the sample, the column was first washed with 10 column volumes of Buffer A (50 mM Tris-HCl pH 7.5, 14.3 mM β-mercaptoethanol, 1 M NaCl, 0.2% triton X-100), and then a further 3 volume columns of Buffer A plus 20 mM imidazole, and 3 volumes of Buffer A plus 50 mM imidazole. The purified transposase was eluted from the column with Buffer A plus 200 mM imidazole. Fractions corresponding to absorbance peaks were analysed on a Laemmli gel to check the purity and estimate protein concentrations. The purified transposase fractions were dialysed against dialysis buffer (50 mM Tris-HCl pH 7.5, 1 M NaCl, 0.1% triton X-100 and 0.1 mM DTT) at 4°C overnight to remove imidazole. The sample was then centrifuged in an Eppendorf microcentrifuge at 14,000 rpm for 5 minutes at 4°C. The supernatant, containing purified transposase, was transferred to new tubes, made up to 50% glycerol and kept at -20°C or -70°C for further use.

His_6-tagged transposase was also purified under denaturing conditions. One gram of frozen cells (see section 2.17) was resuspended in 10 ml denaturing buffer (50 mM Tris, 0.5 M NaCl, 6 M urea and 1 mM PMSF) and was then sonicated as described above. The sonicate was centrifuged at 28,300 g for 15 minutes and the supernatant was loaded onto a pre-charged Ni-NTA column. The column was washed by 10 column volumes of denaturing buffer B (50 mM Tris, 0.5 M NaCl and 6 M urea) and then by denaturing buffer B plus 50 mM imidazole to remove contaminating proteins. The purified transposase was eluted from the column with denaturing buffer B plus 200 mM imidazole. The fractions corresponding to absorbance peaks were analysed on a Laemmli gel. The purified transposase fractions were dialysed against dialysis buffer (50 mM Tris-HCl pH 7.5, 1 M NaCl, 20% glycerol, 0.1 mM EDTA and 0.1 mM DTT) at 4°C overnight to remove urea and imidazole. The sample was centrifuged in an Eppendorf microcentrifuge at 14,000 rpm for 5 minutes at 4°C. The supernatant, which contained purified transposase, was transferred to new tubes and an equal volume of 80% glycerol was added. Purified transposase was kept at -20°C or -70°C for further use.

2.19 *In vivo* transposition assay

A suicide-vector *in vivo* transposition assay was established and used to detect transposition *in vivo*. A pBR322-derived target plasmid (Amp^r) and a pCB104-based donor λ-dv plasmid which carries a transposase gene and a kanamycin resistance marked mini-ISY100 were introduced into a *recA E. coli* strain DH5α in succession by transformation. A transformant (or a portion of a mixture of several transformants equal to one transformant) was grown at 37°C with shaking (200 rpm) for 18 hours in 5 ml L-Broth containing, ampicillin, kanamycin and chloramphenicol. 5 μl of the culture was then inoculated into 5 ml L-Broth containing ampicillin, kanamycin and chloramphenicol and was then incubated at 37°C with shaking for another 18 hours. Plasmid DNA was purified from 3 ml culture using the Qiagen QIAprep Spin Miniprep Kit as described in section 2.12.2 and eluted in 30 μl 1/10 TE. 100 ng of this DNA was electroporated into 40 μl of λ-lysogen *E. coli* strain DS964 by electroporation as described in section 2.10.2. After electroporation, the transformants were suspended in 1 ml SOC and incubated at 37°C with shaking (200 rpm) for 1 hour. 100 μl of a 10^{-3} dilution was spread on an L-Broth agar plate containing ampicillin and the rest was plated on an L-Broth agar plate containing kanamycin, followed by incubation at 37°C overnight. The total number of the Amp^r

colonies and Kan^r colonies in the electroporation reaction were then calculated. The total number of the Kan^r Cm^s colonies was then calculated from the chloramphenicol resistance of 20 randomly picked Kan^r single colonies. The transposition frequency was calculated by dividing the total number of Kan^r Cm^s transformants by the total number of Amp^r transformants.

2.20. *In vitro* transposition assay

2.20.1 *In vitro* transposition assay using supercoiled donor and target plasmids

A standard *in vitro* transposition assay was carried out by incubating 0.12 pmol supercoiled donor plasmid or 0.37 pmol of purified pre-cut mini-ISY100, together with target plasmid and purified transposase in 20 µl of transposase assay buffer (10 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA, 50 mM Tris-HCl pH 7.5, and 50 mM NaCl) at 30°C for 3-4 hours. The reaction was stopped by heating the samples at 75°C for 10 minutes. The samples were analysed by agarose gel electrophoresis or electroporated into DS964 as described in section 2.10.2. The transposition frequency was obtained by dividing the total number of the Kan^r colonies by the total number of Amp^r colonies.

2.20.2 Cleavage assay

DNA fragments carrying mini-ISY100 or a single end of ISY100 was incubated with 1 µl of purified transposase in 20 µl of transposase assay buffer (10 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA, 50 mM Tris-HCl pH 7.5, and 50 mM NaCl) at 30°C for 3-4 hours. There is 50 mM NaCl in the reaction from the purified transposase. DMSO was included at a final concentration of 2% when double stranded oligonucleotides were used as substrate. The reaction was stopped by heating at 75°C for 10 minutes.

2.20.3 Integration assay

DNA fragment carrying transposon end(s) pre-cut at appropriate positions were incubated with target plasmid DNA together with purified transposase in 20 µl of transposase assay buffer (10 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml BSA, 50 mM Tris-HCl pH 7.5 and 50 mM NaCl) at 30°C for 3-4 hours. The reaction was stopped by heating at 75°C for 10 minutes and was analysed by electrophoresis.

2.21 Electrophoretic mobility shift assay (EMSA)

An appropriate amount of end-labelled DNA (~20-50 cps) was mixed with 1 μg of poly dIdC in 10 μl of the binding buffer (10 mM Tris-HCl pH 7.5, 0.1 mg/ml BSA, 0.1 mM EDTA, 10% glycerol and 1 mM DTT). Transposase dilutions were added (1/10 volume) to the mixture. The samples were mixed and then were incubated at 30°C for 10 minutes. Samples were separated by non-denaturing gel electrophoresis on 6% polyacrylamide gels in 1/3× TBE as described in section 2.14.2. Gels were dried and autoradiography was performed as described in section 2.32.

2.22 DNase I footprinting

A standard 8 μl DNA binding assay was set up in DNA binding buffer (see section 2.21.1) with fragment (1,000 cps) containing ISY100 IRL or IRR which was 3'- or 5'- end labelled at one end. Purified transposase or cell extract containing transposase was then added. 1 μl of DNase I (4 μg/ml) in 50 mM $CaCl_2$ and 100 mM $MgCl_2$ was mixed in and the mixture was incubated at room temperature for one minute. The reaction was stopped by adding 6 μl of formamide loading dye containing 30 mM EDTA. The samples together with size markers such as G+A ladder were separated on 6% or 8% denaturing polyacrylamide gels in 1× TTE buffer as described in section 2.14.3.

2.23 PCR

Standard PCR reactions used either *Pfu Turbo* (Stratagene) or *Taq* (Promega) DNA polymerase and were carried out in 50 μl of 1× buffer supplied by the manufacturer. Reactions contained 0.2 mM dNTPs, appropriate amount of template DNA and 2 μM of each primer. For PCR reactions using *Taq* polymerase, $MgCl_2$ was added to a final concentration of 2.5 mM. A typical PCR reaction was denatured at 94°C for 1 minute and then went through 30 cycles of one minute at 94°C, one minute at 55°C and two minutes at 72°C. Finally, the reaction was incubated at 72°C for 10 minutes.

Colony PCR was carried out as follow: A single colony was touched with a toothpick and the cells were suspended in 40 μl double distilled water in a 0.5 ml Eppendorf tube. The tube was heated at 100 °C for 3 minutes and then was centrifuged at 16,000 g for 1 minute. 0.5 μl of the supernatant was used as template in a 25 μl standard PCR.

The 3' termini of donor cleavage sites were determined using a two-step PCR reaction. pXF153 was cleaved with ISY100 transposase in a standard cleavage reaction and the donor backbone was purified from an agarose gel. The purified donor backbone (10 ng) was then treated at 37°C for 15 minutes with 25 units of recombinant terminal deoxynucleotidyl transferase (Invitrogen) and 2 mM dCTP in TdT buffer containing 2 mM $CoCl_2$ (dC-tailing reaction). PCR was carried out for 4 cycles with 1 ng of the dC-tailed donor backbone as template, using primers Rev-ClaI and polyG to amplify the left donor flank, or Uni-MluI and polyG to amplify the right donor flank. The parameters for PCR were 30 seconds at 94°C, 30 seconds at 55°C and 70 seconds at 72°C. This PCR reaction was diluted 1 µl into 25 µl as template for a second PCR employing primers Rev-ClaI and nest-polyG for the left donor flank, or primers Uni-MluI and nest-polyG for the right donor flank. PCR was carried out using the same reaction conditions for a further 25 cycles. The PCR products were cloned by TOPO TA cloning (Invitrogen) following the manufacturer's instructions, and then sequenced.

2.24 Site-directed mutagenesis

ISY100 transposase gene was mutagenized using the QuikChange site-directed mutagenesis Kit (Stratagene) according to the manufacturer's instructions. Mutagenesis was also used to eliminate an extra SphI site on the plasmid pCB104 to facilitate construction of donor plasmids for the *in vivo* transposition assay.

2.25 Sequencing of plasmid DNA or PCR products

DNA samples were purified using the Qiagen Miniprep Spin Kit and then sequenced by MWG Biotech or MBSU DNA Sequencing Service, University of Glasgow.

2.26 Purification of synthetic oligonucleotides

Full-length oligonucleotides were separated from prematurely terminated oligonucleotides by denaturing polyacrylamide gel electrophoresis (section 2.14.3). After electrophoresis, the oligonucleotides were visualised by staining with "Stains-all" (1-Ethyl-2-[3-(1-ethylnaphtho [1,2-d]thiazolin-2-ylidene)-2-methylpropenyl] naphth [1,2-d] thiazolium bromide; Sigma). The gel was first stained in 100 ml of staining solution [80 ml H_2O, 20 ml isopropanol and 10 ml of 0.1% solution of "Stains-all" (w/v) in formamide]. After

destaining in water, the full-length oligonucleotide band was cut from the gel, and the DNA was recovered as described in section 2.27. The DNA pellet was dissolved in 1/10 TE buffer. The purified oligonucleotides were stored at -20°C.

2.27 Purification of DNA fragments from polyacrylamide gels

Following polyacrylamide gel electrophoresis, the gel was stained with ethidium bromide (EtBr) as described in section 2.15.1. DNA fragments were visualised on a long wavelength transilluminator (365 nm). The desired DNA band was cut out and transferred to a microcentrifuge tube. The gel slice was crushed against the tube wall using a pipette tip and then soaked into 500 μl acrylamide gel elution buffer or TE if T4 polynucleotide kinase treatment was to follow. The gel slurry was incubated at 37°C on a rotating wheel for 4-12 hours. The polyacrylamide was removed by spinning at maximum speed for 1 minute and then passing the supernatant through a Costar Spin-X filter (0.22 μm) for 3 minutes at 2,700 rpm in an Eppendorf microcentrifuge. The DNA was precipitated with ethanol and the pellet was dissolved in the desired volume of TE or H_2O for further use.

2.28 Purification of DNA from agarose gels

Following agarose gel electrophoresis, DNA was visualised on a long wavelength transilluminator (365 nm) as described in section 2.15.1. The desired DNA band was excised and transferred to a fresh tube. Depending on the size of the DNA fragment, it was purified from the agarose slice using QIAquick Gel Extraction Kit (Qiagen) or MinElute Gel Extraction Kit (Qiagen) following instructions provided by the manufacturer.

DNA was purified from SeaPlaque GTG agarose gel using phenol extraction. The excised band was transferred to a 1.5 ml microfuge tube and melted at 67°C for 10 minutes. 1 volume of pre-warmed (67°C) TE buffer (20 mM Tris-HCl, 1 mM EDTA, pH 7.8) was added to the melted gel and mixed. This diluted SeaPlaque GTG agarose was extracted twice with an equal volume of phenol saturated with TE buffer, once with phenol/chloroform and then once with chloroform alone. The DNA was precipitated with ethanol and dissolved in the desire amount of buffer for further use.

2.29 Preparation of radiolabelled DNA

2.29.1 Labelling 3′ termini of double-stranded DNA using the Klenow fragment of *E. coli* DNA polymerase

Double-stranded DNA was dissolved in 20 µl of 1× EcoPol Buffer (NEB) or NEBuffer 2 and mixed with 20 µCi [α-^{32}P]dNTP and 10 µM of other three dNTPs. One unit Klenow per µg DNA was added to the mixture and the mixture was then incubated at room temperature for 15 minutes. 2.2 µl of 2 mM solution of all 4 dNTPs was added to the reaction and the reaction was incubated for 5 minutes at room temperature. The reaction was stopped by adding 2.5 µl of 250 mM EDTA (pH 8.0).

2.29.2. Phosphorylation of DNA molecules with protruding 5'-hydroxyl termini

Double-stranded DNA restriction fragment was first dephosphorylated by calf intestinal phosphatase (CIP) as described in section 2.30. In a 20 µl reaction, dephosphorylated DNA (10-50 pmoles) was incubated with [γ-^{32}P] ATP (10 µCi) and T4 kinase (1 µl, 10 U) in the supplied T4 kinase buffer for 30 minutes at 37°C. Unlabelled ATP was then added to final concentration of 1 mM and the reaction was incubated at 37°C for a further 15 minutes. The reaction was stopped by heating at 70°C for 5 minutes.

2.29.3 Obtaining fragments labelled at only one end

First double-stranded DNA fragment was labelled at both ends as described in 2.29.1 or 2.29.2. Unincorporated labelled nucleotides were removed by MicroSpin G-50 column following the manufacturer's instructions (Amersham Biosciences). Proteins were removed by extraction once with phenol and once with chloroform. The labelled DNA fragment was precipitated with ethanol and dissolved in the restriction enzyme buffer. The labelled fragment was cleaved with an appropriate restriction enzyme as described in section 2.13. 10 µg of tRNA was added to the reaction and the DNA was precipitated with ethanol and dissolved in 20 µl of TE. The DNA was run on a polyacrylamide gel (as described in 2.14.2) and the desired band was cut and recovered from the gel as described in section 2.27. Labelled DNA fragments were dissolved in TE buffer and stored at -20°C for further use.

2.29.4 5' end-labelling of oligonucleotides

Synthetic oligonucleotides lack phosphate groups at their 5' ends so can be labelled by phosphorylation directly. In a 10 µl reaction, 10 pmoles of purified oligonucleotide was incubated with [γ-^{32}P] ATP (10 µCi) and T4 kinase (1 µl, 10 U) in the supplied T4 kinase buffer for 30 minutes at 37°C. The reaction was stopped by heating at 68°C for 10 minutes. Unincorporated labelled nucleotides were removed by MicroSpin G-50 column following the manufacturer's instruction (Amersham Biosciences). Proteins were removed by extraction once with phenol and once with chloroform. Double-stranded oligonucleotides were obtained by annealing the labelled oligonucleotide with its complementary oligonucleotide as described in section 2.31. Labelled double-stranded oligonucleotides were separated from single-stranded oligonucleotides by polyacrylamide gel electrophoresis as described and stored at -20°C for further use.

2.29.5 Preparation of G+A ladder

A 10 µl reaction was set up in a 2 ml gas-tight microcentrifuge tube with 2,000 cps end-labelled ds DNA and 1 µg poly dIdC in H_2O. 1 µl of 1:26 diluted formic acid was added and the mixture was incubated at 37°C for 25 minutes. 150 µl of 1:10 diluted piperidine was added and the mixture was incubated at 90°C for 30 minutes. 1.8 ml of butanol was added, mixed, and the tube was centrifuged at the maximum speed for 1 minute in a desktop centrifuge. The supernatant was discarded and 150 µl of 1% SDS was added to resuspend the pellet. Another 1.8 ml of butanol was added for one more precipitation as described above. The DNA pellet was dissolved in TE and an equal volume of 2× formamide loading dye was added. The sample was stored at -20°C for future use, and heated for 5 minutes at 80°C prior to loading on urea denaturing polyacrylamide gels.

2.29.6 Labelling of supercoiled plasmid.

5 µg of plasmid was linearised by restriction digestion in 40 µl of appropriate buffer. 2.5 units of CIP was added 10 minutes from the end of the reaction. The reaction product was extracted three times with phenol and once with chloroform to inactive the enzyme. The DNA was precipitated with ethanol and redissolved in T4 kinase buffer (20 µl). [γ-^{32}P] ATP (10 µCi) and T4 kinase (1 µl, 10 U) was added and the mixture was incubated for 30 minutes at 37°C. ATP was added to the mixture to final concentration of 1 mM and the mixture was incubated for another 15 minutes. The reaction was stopped by heating at

70°C for 5 minutes. 100 µl of 5× ligase buffer (Invitrogen), 370 µl of water, and 10 µl of 100 µg/ml ethidium bromide was added to the reaction and mixed well. 2 units of T4 ligase was added and the mixture was incubated overnight at room temperature (16 h at 20°C). 10 µg (1 µl) of tRNA was added to the reaction and the DNA was recovered by ethanol precipitation. The pellet was dissolved in 20 µl of TE and separated on a 1% Seaplaque agarose gel. Supercoiled DNA was cut from the gel and purified as described in 2.28.

2.29.7 5' end-labelling by the phosphate exchange reaction

The radioactive 1 kb DNA ladder used as DNA molecular marker on agarose gels was labelled by the phosphate exchange reaction following instructions provided by the manufacturer (Invitrogen).

2.30 Dephosphorylation of DNA fragments with alkaline phosphatase

5'-phosphate groups were removed from DNA with calf intestinal alkaline phosphatase (CIP) following instructions provided by the manufacturer (New England Biolabs). The product was extracted three times with phenol and once with chloroform to remove the CIP. The DNA was precipitated with ethanol and was dissolved in the desired amount of TE for further use.

2.31 Annealing of oligonucleotides

A pair of purified complementary oligonucleotides were annealed by mixing equal amounts (1-10 pmoles) of each oligonucleotide in 1 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 100 mM NaCl and heating to 80°C for 5 minutes. The mixture was then cooled to room temperature over a period of 5 hours. Annealed oligonucleotides were stored at -20°C.

2.32 Photography, autoradiography and phosphorimagery

Agarose and polyacrylamide gels were photographed using either a Polaroid camera and Polaroid 667 film, or a Canon EOS-D30 digital camera with a 480 nm band pass filter (Peca products).

For autoradiography, gels were dried under vacuum in a Bio-Rad slab gel dryer, and exposed to a sheet of Kodak X-ray film. The film was developed in an X-OMAT automated processor (Kodak).

Phosphorimaging was carried out by drying the gel as described above, and then exposing it for 6-16 hours to a phosphorimaging screen. The screen was imaged in a Fuji BAS-1500 phosphorimaging system and the image was analysed with the supplied software.

2.33 Quantification of DNA and protein

The concentration of double-stranded DNA was calculated from the OD_{260} of the DNA solution using the formula [DNA] = $A_{260} \times 50$ μg/ml. Oligonucleotide concentrations were determined from the absorbance at 260 nm using the extinction coefficient calculated from the DNA sequence. Protein concentration was measured by Bio-Rad protein assay according to the supplied protocol.

Chapter III

In vivo transposition of ISY100

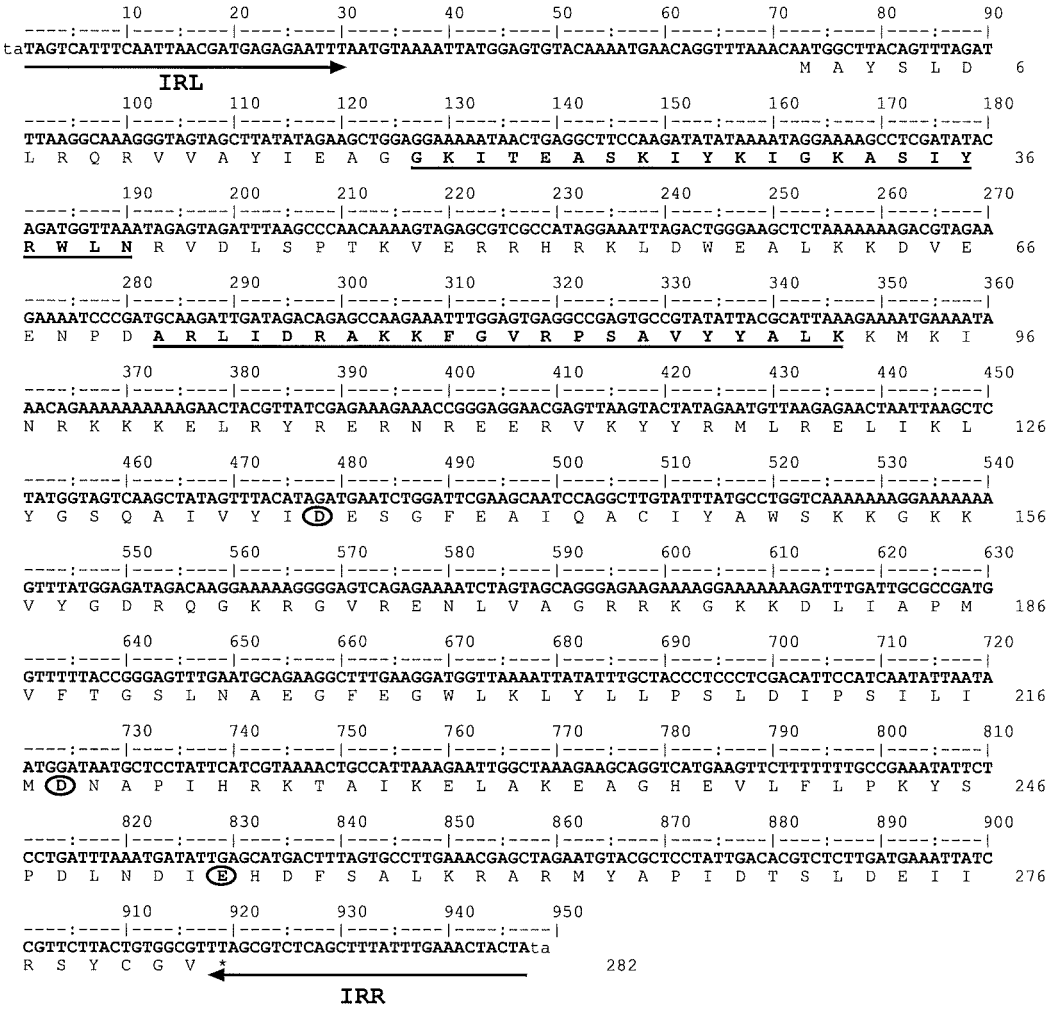
3.1 Introduction

ISY100 was first found in the cyanobacterium *Synechocystis* sp. strain PCC6803 (Cassier-Chauvat *et al.*, 1997). There are 22 copies of ISY100 in the genome of PCC6803 and at least three copies have been found in plasmids from this strain (http://www.kazusa.or.jp/cyano/Synechocystis/). A typical ISY100 is 947 bp in length and contains one open reading frame which is proposed to encode the transposase protein (Figure 3.1 A). In some copies of ISY100, the transposase gene contains a frame-shift mutation which will lead to the production of a truncated protein. ISY100 has 24 bp imperfect inverted repeats (IRs) at its termini (Figure 3.1 B). Insertions are flanked by TA dinucleotides, which are presumed to be target duplications. Based on sequence homology of its putative transposase, ISY100 belongs to the IS630/Tc1/*mariner* superfamily of transposons.

The structure of ISY100 suggests that it is an insertion sequence, in other words a simple DNA transposon containing only the genes needed for its own transposition. ISY100 is found in multiple copies in the genome of *Synechocystis* sp. strain PCC6803 and one copy was found in a putative IS4-like transposase gene suggesting recent transposition activity (Cassier-Chauvat *et al.*, 1997). ISY100 was confirmed to be active by Urasaki *et al.* (2002), who showed that expression of ISY100 transposase leads to excision and transposition of ISY100 in *E. coli.* Mapping of the cleavage products revealed that they were cut exactly at the transposon boundaries at the 3' ends but recessed by 2 nucleotides at the 5' ends. ISY100 jumped exclusively into TA dinucleotides in *E. coli*, and these TA dinucleotides appear to be duplicated after transposition. These properties are all characteristics of the IS630/Tc1/*mariner* superfamily, consistent with ISY100 being a member of this group.

This chapter describes my work on the transposition of ISY100 *in vivo*. My ultimate aim was to set up an *in vitro* system using purified transposase and DNA. However, I started this work by setting up an *in vivo* assay for several reasons. First I wanted to replicate some of the results of Urasaki *et al.* (2002). This would allow me to ensure I had isolated an active transposase gene and to determine the *cis* requirements for transposition. The system was also used to study the effects of mutations in the transposase gene, the IRs and the flanking sequences. Furthermore, the plasmids constructed for the *in vivo*

A)



B)

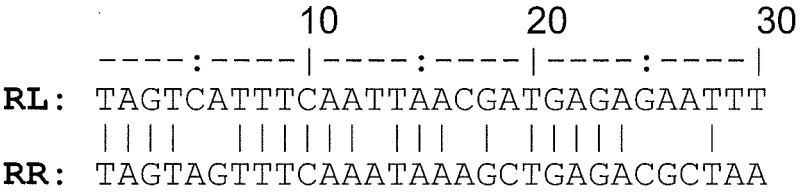


Figure 3.1 Structure of ISY100a **A)** Terminal inverted repeats are shown by arrows. The lowercase TA sequences outside the ISY100 sequence show the target site duplication. The amino acid sequence of the transposase is shown below the nucleotide sequence. The positions of nucleotide residues are shown by the ruler; the position of the last amino acid residue in each line is indicated at the right. Potential helix-turn-helix domains are underlined. D136, D218, and E253 of the predicted DDE motif are circled. **B)** Alignment of IRL and IRR sequences of ISY100.

transposition system could later be used to set up a sensitive genetic screen for transposition *in vitro*.

3.2 Establishment of an *in vivo* transposition assay

3.2.1 *In vivo* assays for transposition

Various assays have been used to study transposition in bacteria. Many transposons have been discovered by phenotypic changes they bring about in their hosts. These changes can be caused by genes carried on the transposon itself, as in the case of antibiotics resistance conferred by the transposon, or by insertional mutagenesis or other effects on neighbouring genes.

Two main methods have been developed to study transposition quantitatively in bacteria: mating-out assays and assays based on suicide vectors. In both assays a marked copy of the transposon together with the transposase gene and a target DNA are introduced into the same cell. The two assays then use different strategies to distinguish between the transposon donor and the target DNA, so that the transposition frequency can be calculated without interference from the donor. Mating-out assays use a conjugative plasmid, such as the F factor-derived pOX38 which lacks all known transposable elements, as target. After allowing time for transposition to happen, conjugation is used to transfer the target plasmid to a recipient cell. The transposition frequency is calculated by comparing the number of transposon-containing exconjugants to the total number of exconjugants (Groenen *et al.*, 1985; Johnson and Reznikoff, 1984).

In a typical suicide-vector transposition assay, the transposon is carried on a conditionally replicating vector. After allowing time for transposition to occur, the conditions are changed so that the transposon donor cannot replicate but the target can. Target molecules containing insertions are then selected using the antibiotic resistance conferred by the marked transposon. The transposition frequency is calculated as the number of colonies carrying the target DNA with inserted transposon (for instance see Tenzen and Ohtsubo, 1991).

Both mating-out and suicide-vector assays allow the transposase gene to be *in cis* or *in trans* to the transposon ends, and mutations can be easily introduced into the transposase gene and the inverted repeat sequences.

3.2.2 Previous studies on transposition of ISY100 *in vivo*

In the study of Urasaki *et al.* (2002), transposition of ISY100 was studied in *E. coli* using a modified suicide-vector transposition assay. This used a pBR322-derived donor plasmid carrying an IPTG-inducible transposase gene and an artificial mini-ISY100, consisting of 51 bp ISY100 IRL and IRR sequences, flanking a chloramphenicol (Cm) resistance gene. The target plasmid used was a spectinomycin resistant (Sp^r) R100 derivative. The assay was set up by transforming the donor plasmid into an *E. coli* K12 strain which already harboured the target plasmid. To allow transposition to take place, a transformant was incubated overnight, diluted into a fresh culture and incubated overnight again. IPTG was added to the second overnight incubation to induce expression of the transposase. Plasmid DNA was isolated and the plasmid mixture was then electroporated into the *polA E. coli* strain JE6638. DNA polymerase I, encoded by the *polA* gene, is required for replication of pBR322-based plasmids, but not for replication of R100 (Grindley and Kelley, 1976). Therefore, the donor plasmid cannot replicate in this strain but the target plasmid can. Transformants carrying the target plasmid with a mini-ISY100 insertion were distinguished from those carrying the original target plasmid by the mini-transposon-encoded chloramphenicol resistance. The transposition frequency was calculated as the ratio of the number of Sp^r Cm^r transformants to the total number of Sp^r transformants and represents the fraction of target plasmids having a transposon insertion.

3.2.3 Choice of transposition assay

When setting out to develop an *in vivo* transposition assay for studies of ISY100, I had to decide whether to set up a mating out or a suicide-vector based system. One of the main problems with mating-out transposition assays is the low copy number of conjugative plasmids such as F and POX38. This makes it difficult to map the junction sequences between the transposon and the target directly. In the mating-out assay established for transposition of *S. sonnei* IS630 (my unpublished data), an extra step of inverse PCR was required to identify the IS630 target sites on the F plasmid. To avoid this extra step, it was decided to establish an *in vivo* transposition assay for ISY100 based on the strategy used by Urasaki *et al.* (2002).

Results from mating-out experiments suggested that *S. sonnei* IS630 transposase acts preferentially *in cis* rather than *in trans* (my unpublished data). I thus decided to construct a suicide-vector transposition assay for ISY100 in which the transposon ends and transposase gene are carried on the same donor plasmid.

3.2.4 Construction of plasmids for *in vivo* transposition assay

The initial donor plasmid contained the same basic components as the donor plasmid used by Urasaki *et al.* (2002).

First, the same copy of ISY100 used by Urasaki *et al.* (ISY100a) together with short flanking sequences was isolated from *Synechocystis* sp. strain PCC6803 genomic DNA (donated by C. Mullineaux) by PCR with primers 6803_Top and 6803_Bot (Figure 3.2). The PCR fragment was then cleaved at the EcoRI and XbaI sites incorporated into the primer sequences, and inserted into pUC18, yielding pXF106.

A fragment containing 51 bp from ISY100 left end, 49 bp from ISY100 right end and the pUC18 vector sequences, was obtained by PCR with two outward facing primers within ISY100 (y100-51_Top and y100-51_Bot) using pXF106 as template. This PCR product was cleaved at BamHI sites within the primer sequences, and ligated to a 1.3 kb BamHI Kan^r fragment from pUC4K, giving pXF109 (Figure 3.2). pXF109 therefore contains a kanamycin resistance gene flanked by 51 bp ISY100 IRL and 49 bp ISY100 IRR sequences in the natural genomic context of ISY100a. The Kan^r marked mini-ISY100 element contained within pXF109 is here referred to as ISY100-51-Kan.

The transposase gene of ISY100a was isolated from pXF106 by PCR with primers Tnp100_Top and Tnp100_Bot, which introduce an NdeI site at the start codon and a BamHI site just after the stop codon of the transposase open reading frame. The transposase gene was then inserted into the NdeI-BamHI-cleaved expression vector pTrc99a to produce pXF104.

To produce the finished suicide transposon donor plasmid, ISY100-51-Kan was inserted into the λ-dv plasmid pXF80 to give pXF113 (Figure 3.2). Then the transposase expression cassette from pXF104 was inserted into pXF113 to give pXF115 (Figure 3.2). λ-dv-based plasmids can replicate normally in most *E. coli* hosts, but are unable to replicate in λ lysogens, or any other hosts that express the λ repressor.

An initial *in vivo* transposition assay was carried out using pUC18 as target (Figure 3.3). The donor plasmid pXF115 was transformed into the *recA E. coli* strain DH5α already containing pUC18 as target plasmid. A parallel control experiment was set up using pXF113, which contains ISY100-51-Kan but not the transposase gene, as donor plasmid. Because the donor and target plasmids have different origins, they are compatible and can co-exist in the same cell. The resulting strains were incubated in L-Broth overnight to allow transposition to occur. In this initial experiment, IPTG was added when the OD_{600}

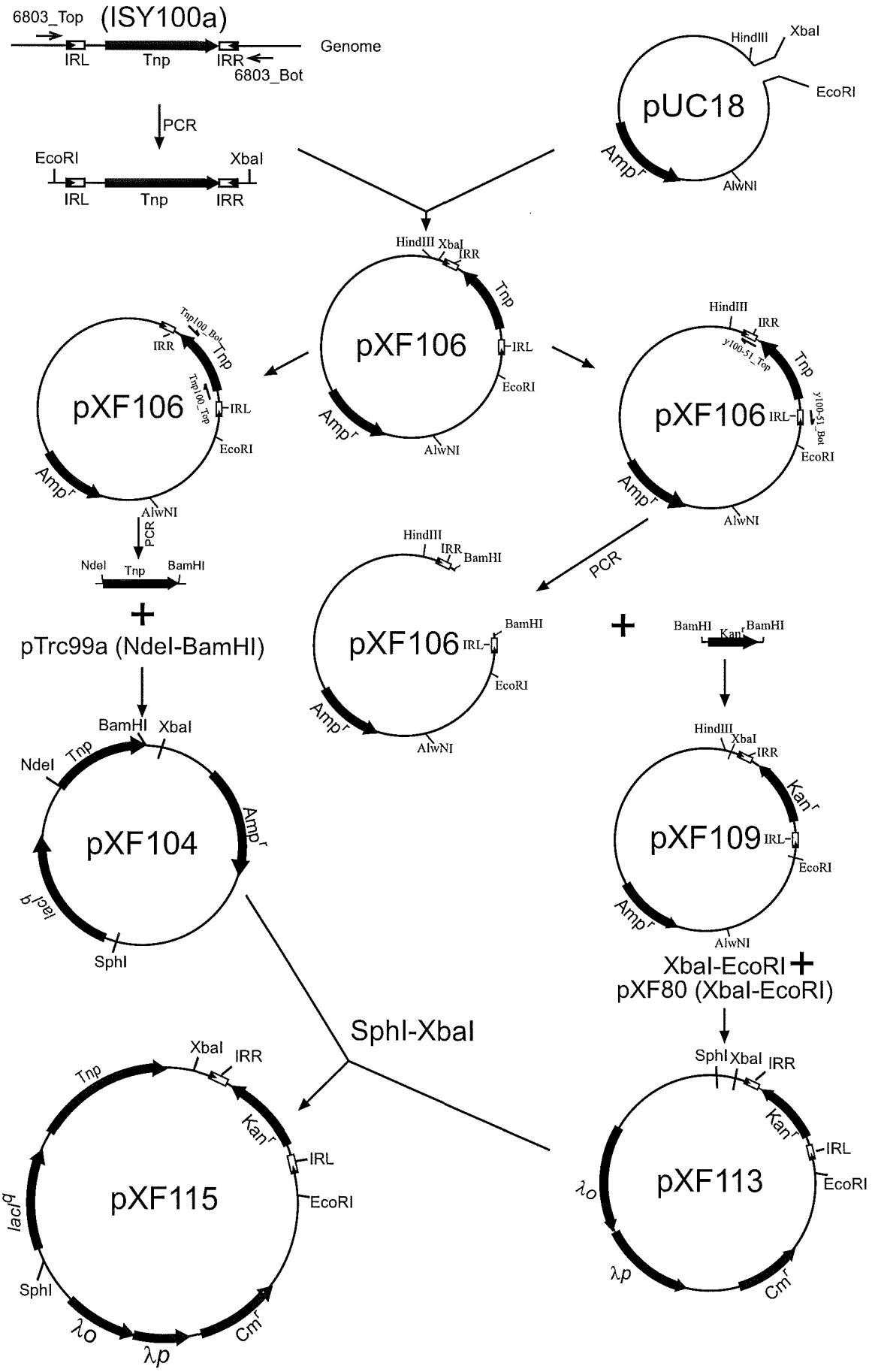

6803_Top
(ISY100a)
IRL
Tnp
IRR
Genome
6803_Bot
PCR
EcoRI
XbaI
IRL
Tnp
IRR
pUC18
HindIII
XbaI
EcoRI
AlwNI
Amp^r
pXF106
HindIII
XbaI
IRR
Tnp
IRL
EcoRI
AlwNI
Amp^r
pXF106
Tnp100_Bot
Tnp100_Top
IRR
Tnp
IRL
EcoRI
AlwNI
Amp^r
PCR
NdeI
Tnp
BamHI
pXF106
HindIII
IRR
y100-51_Top
Tnp
IRL
y100-51_Bot
EcoRI
AlwNI
Amp^r
PCR
pXF106
HindIII
IRR
BamHI
IRL
BamHI
EcoRI
AlwNI
Amp^r
pTrc99a (NdeI-BamHI)
BamHI
Kan^r
BamHI
pXF104
BamHI
XbaI
NdeI
Tnp
Amp^r
lacI^q
SphI
pXF109
HindIII
XbaI
IRR
Kan^r
IRL
EcoRI
AlwNI
Amp^r
XbaI-EcoRI
pXF80 (XbaI-EcoRI)
SphI-XbaI
pXF115
XbaI
IRR
Tnp
Kan^r
IRL
EcoRI
lacI^q
SphI
λo
λp
Cm^r
pXF113
SphI
XbaI
IRR
Kan^r
IRL
EcoRI
λo
λp
Cm^r

Figure 3.2 Construction of donor plasmid for *in vivo* transposition. ISY100a and its flanking sequence was first isolated from *Synechocystis* sp. strain PCC6803 genomic DNA by PCR and then inserted into pUC18 to construct pXF106. The λ-dv based donor plasmid pXF115 and the control plasmid lacking transposase (pXF113) were constructed using the steps shown. The donor plasmid pXF116 was constructed in the same way as pXF115 but carries a C-terminal His-tagged transposase obtained from pXF106 by PCR using a different bottom primer. The donor pXF114 was made in the same way as pXF113 but carries a Kan^r marked mini-ISY100 with 30 bp IRs produced from pXF106 by PCR using primers y100-30_Top and y100-30_Bot. The donor plasmids pXF117 and pXF118 were constructed from pXF114 in the same way as pXF115 and pXF116.

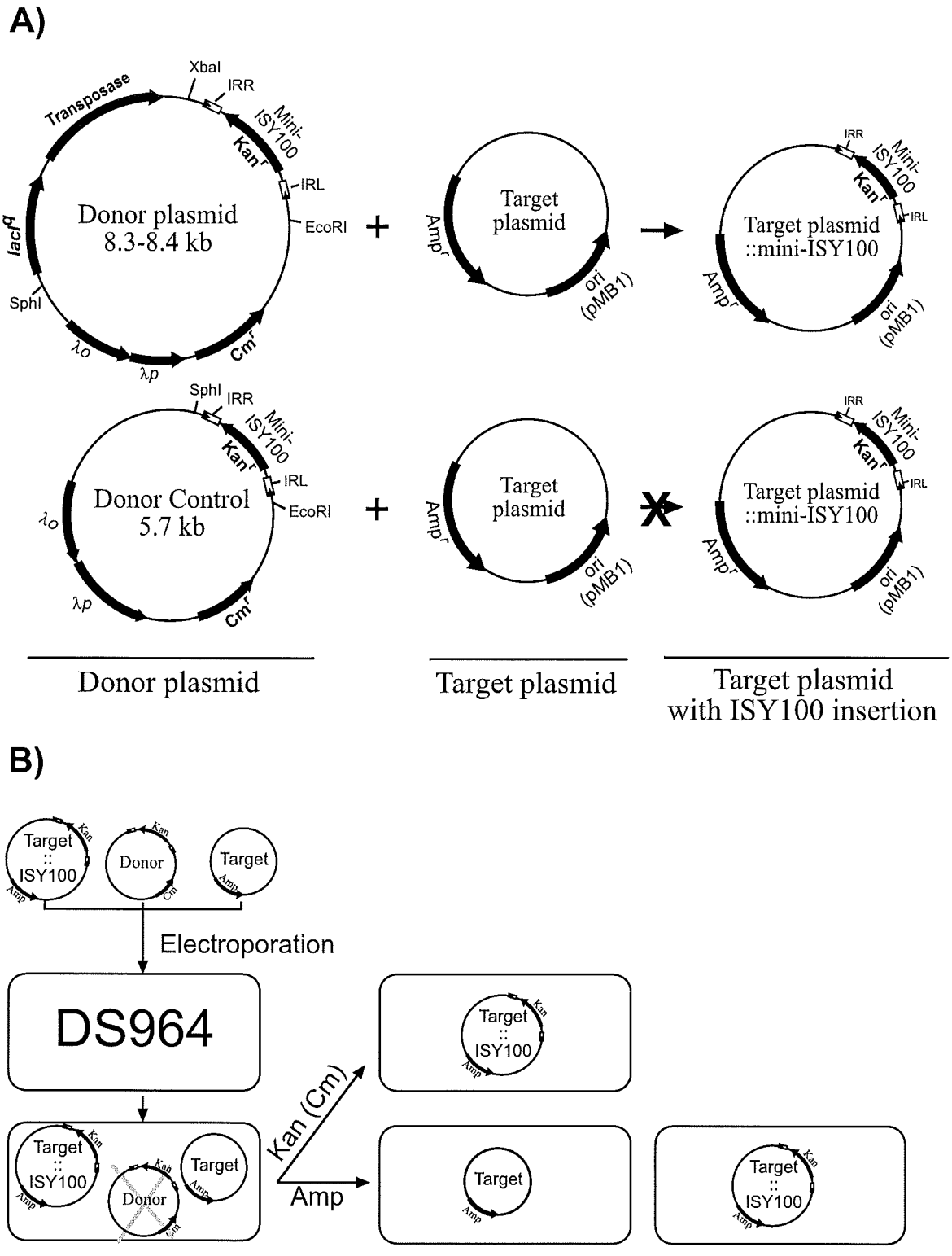


Figure 3.3 *In vivo* transposition assay. A) *In vivo* transposition. Donor and target plasmids were co-transformed into the *recA E. coli* strain DH5α and incubated to allow transposition to occur (see chapter II for details). Expression of the transposase was expected to mediate transposition of the mini-ISY100 from donor plasmid to the target plasmid. In the absence of transposase, no transposition should occur. **B)** Detection of transposition events. Plasmid DNA was purified from the DH5α after the incubation, and was then electroporated into a λ-lysogen *E. coli* strain (DS964), in which the λ-dv donor or donor control plasmid cannot replicate. 10 µl of a 10^{-2} dilution of the transformation was spread onto an Amp plate and the rest was spread onto a Kan plate. The total number of target transformants was calculated from the number of single colonies on the Amp plate. The transposition frequency was calculated by dividing the total number of the Kan^r Cm^s colonies by the total number of Amp^r colonies.

reached 0.3 to induce expression of transposase. After overnight growth, plasmid DNA was purified and was then electroporated into a *recA* mutant λ-lysogen *E. coli* strain (DS964), in which the donor λ-dv plasmid cannot replicate. Transformants were selected on plates containing ampicillin and kanamycin, to select for target plasmids containing ISY100-51-Kan. Such Amp^r Kan^r transformants were obtained from the pXF115 transposition reaction, but not from the pXF113 reaction (data not shown). Therefore, transposition seems to be taking place and is dependent on the presence of the transposase gene on the donor plasmid. A single colony gel was used to analyse the DNA from the transformants, demonstrating the presence of plasmids of the correct size to be the target pUC18 containing ISY100-51-Kan insertions (data not shown).

3.3 Optimisation of the transposition assay and requirements for transposition

A quantitative transposition assay was then developed and used to study the requirements for transposition. Since the mini-ISY100 can potentially insert into the Amp^r gene, resistance to kanamycin rather than ampicillin and kanamycin was used to identify target plasmids containing mini-ISY100 insertions. The transposition frequency was then calculated by dividing the number of Kan^r transformants by the number of Amp^r transformants. This *in vivo* transposition assay was then tested and improved as described below using transposition frequency as the main indicator of transposition efficiency.

In earlier work, I set up a similar assay to study transposition of *S. sonnei* IS630 (my unpublished data). In those assays, and in pilot assays for ISY100, a small number of Kan^r transformants containing only the unmodified target plasmid were sometimes obtained from electroporation into DS964. These transformants were found even in the absence of transposase and could be distinguished from the genuine transposition events because they were resistant to chloramphenicol as well as kanamycin. This suggests that the donor suicide plasmid was still present. The way in which the "suicide" donor plasmid survives in the λ-lysogen cell is unclear. λ may be lost from DS964 at low frequency, allowing replication of the low copy-number donor plasmid. However, the donor plasmid was never observed on single colony gels, so it may have integrated into the chromosome even though DS964 is *recA*$^-$. Although Kan^r Cm^s transformants were rare, it was decided that their number should be estimated and that this number should be subtracted from the number of the Kan^r colonies before calculating the transposition frequency. Normally, 20

Kan^r transformants were randomly picked and checked separately for resistance to kanamycin and chloramphenicol, and the predicted total number of $Kan^r Cm^s$ transformants was then calculated accordingly.

3.3.1 IPTG is not needed for efficient transposition

In the study of Urasaki *et al.* (2002), IPTG at a final concentration of 0.5 mM was added when the OD_{600} of the cell-culture reached 0.3 in the second overnight incubation. In the experiments above, IPTG was used to keep the conditions as close as possible to those of Urasaki *et al.* (2002). To test whether IPTG made any difference to the transposition frequency, transposition assays were carried out with pXF115 and pXF113 as donors and pUC18 as target, with and without IPTG. For each experiment, two individual transformant colonies containing donor and target plasmid were picked and incubated overnight for 18 hours in L-broth. The cultures were diluted 1 in 1000 and incubated for another 18 hours overnight, with or without the addition of IPTG when the culture reached an OD_{600} of 0.3. There was no transposition in the absence of transposase (pXF113; Table 3.1). In the presence of transposase, there was a modest increase in the transposition frequency when IPTG was added (pXF115; Table 3.1), although it is not clear whether this increase was due to the presence of IPTG or was statistical variation in the assay (see section 3.3.5).

Transposition in the absence of IPTG suggests that the P_{trc} promoter on pXF115 is not tightly regulated. The leakiness of the P_{trc} promoter has also been observed in this lab in experiments with the *E. coli pepA* gene (S. Colloms, personal communication). The presence of *lac* operators on the high copy number pUC18 target plasmid might also weaken repression by the *lac* repressor encoded on the donor plasmid. Although adding IPTG gave higher transposition frequencies, to simplify the assay and increase its reproducibility, IPTG was not generally used in later experiments.

3.3.2 30 bp from IRR and IRL are sufficient for transposition

ISY100 contains imperfect 24 bp inverted repeats (Figure 3.1). This suggests that 24 bp from each transposon end might be sufficient for transposition. To test whether the 51 bp ISY100 ends in pXF115 are required for transposition, a donor plasmid with only 30 bp from each end was constructed. ISY100-30-Kan, carrying 30 bp IRL and IRR flanking the kanamycin resistance gene, was constructed in a similar way to ISY100-51-Kan and inserted into donor plasmids with and without the transposase gene to give pXF116 and

Table 3.1 Frequency of simple insertion mediated by mini-ISY100

Reaction			Amp^r ($\times 10^6$)	$Amp^r Kan^r$	Cm^s	$Amp^r Kan^r Cm^s$	Transposition frequency ($\times 10^{-5}$)	
No.	pUC18 +	IPTG[1] (mM)					Individual	Average
#1	pXF113 (ISY100-51-Kan, Tnp^-)	–	3.6	31	0/31	0	$< 2.8 \times 10^{-7}$	$< 6.6 \times 10^{-8}$
#2			11.6	14	0/14	0	$< 8.6 \times 10^{-6}$	
#1	pXF113 (ISY100-51-Kan, Tnp^-)	0.1	8.5	91	0/91	0	$< 1.2 \times 10^{-6}$	$< 5.8 \times 10^{-8}$
#2			8.7	14	0/14	0	$< 1.1 \times 10^{-6}$	
#1	pXF115 (ISY100-51-Kan, Tnp^{wt})	–	8.5	175	15/20	131	1.5	1.5 ± 0.0
#2			10.7	172	19/20	163	1.5	
#1	pXF115 (ISY100-51-Kan, Tnp^{wt})	0.1	6.4	141	20/20	141	2.2	2.4 ± 0.2
#2			5.0	135	19/20	128	2.5	
#1	pXF114 (ISY100-30-Kan, Tnp^-)	–	13.9	25	0/25	0	$< 7.2 \times 10^{-8}$	$< 3.3 \times 10^{-8}$
#2			16.0	13	0/13	0	$< 6.3 \times 10^{-8}$	
#1	pXF116 (ISY100-30-Kan, Tnp^{wt})	0.1	2.8	369	20/20	369	13.0	8.1 ± 6.9
#2			4.2	151	18/20	136	3.2	
#1	pXF117 (ISY100-51-Kan, Tnp^{His6})	0.1	3.5	240	20/20	240	6.8	7.1 ± 0.4
#2			3.1	227	20/20	227	7.4	
#1	pXF118 (ISY100-30-Kan, Tnp^{His6})	0.1	4.8	169	18/20	152	3.2	4.6 ± 2.0
#2			3.2	193	20/20	192	6.0	

(1): IPTG was added at $OD_{600} = 0.3$ during the second 18 hours incubation.

DH5α containing pUC18 together with the indicated plasmid was incubated overnight for 18 hours, then diluted 1:1000 and incubated for a further 18 hours with or without IPTG as indicated. DNA was then isolated and electroporated into the λ-lysogen strain DS964. A dilution was plated onto Amp plates and the remaining cells were plated onto Amp + Kan plates. At least 20 single colonies were picked from the Amp + Kan plate to test their resistance to chloramphenicol to assess the total number of Amp^r Kan^r Cm^s colonies. The transposition frequency was calculated by dividing the number of Amp^r Kan^r Cm^s colonies by the number of Amp^r colonies. Transposition frequencies are shown as the average of all repeats of the assay ± the standard deviation (σ_{n-1}).

pXF114 respectively (Figure 3.2). These donor plasmids were used in an *in vivo* transposition assay with pUC18 as target. No Kan^r Cm^s transformants were obtained in the absence of transposase (pXF114; Table 3.1). In the presence of transposase, transposition was observed at an even higher frequency than that obtained with ISY100-51-Kan (pXF116; Table 3.1). Although this increase in transposition frequency may not be statistically significant, it seems safe to conclude that 30 bp from the ends of ISY100 are sufficient to mediate efficient transposition. A transposon with only 24 bp from each end has not yet been tested, but it seems likely that it would work.

3.3.3 Transposase carrying a C-terminal His_6-tag is active in transposition

A PCR product encoding ISY100 transposase with a C-terminal 6-His tag (Tnp-His_6) was obtained from pXF106 with primers Tnp100_Top and Tnp100_Chis. Donor plasmids carrying this Tnp-His_6 gene together with either ISY100-51-Kan or ISY100-30-Kan were constructed as described for pXF115, giving pXF117 and pXF118 respectively. The transposition frequency was then measured as described in the previous section giving frequencies similar to those obtained for equivalent plasmids with the wild-type transposase (Table 3.1).

In a similar *in vivo* assay for IS630 transposition, the addition of a His_6-tag at C-terminus of IS630 transposase increased the transposition frequency more than 3 fold whereas the addition of a His_6-tag at the N-terminus of IS630 transposase reduced the transposition frequency more than 20 fold (My unpublished data).

The fact that ISY100 transposase with a C-terminal His_6-tag was active *in vivo* was later exploited in the purification of transposase for *in vitro* transposition.

3.3.4 The time of incubation affects the transposition frequency

Transposition occurs during the period when the cells containing both donor and target plasmid are incubated. A longer incubation time might lead to an increase in transposition frequency. To test this, the effect of changing the incubation time was investigated (Table 3.2). In this experiment, the donor plasmid pXF116 containing only 30 bp from each transposon ends and wild-type transposase was used. pXF116 was transformed into DH5α already containing pUC18 as a target plasmid. Three single transformant colonies were picked and processed in parallel, to assess the variability of the assay. DNA was isolated from samples taken at 14, 16 and 18 hours after inoculation. After 18 hours, the culture was diluted 1000 fold into fresh L-broth and samples were again taken at 14, 16 and 18

Table 3.2 Incubation time affects the *in vivo* transposition frequency

Incubation time	Reaction No.	Amp^r ($\times 10^4$)	Kan^r	Transposition frequency ($\times 10^{-6}$)	
				Individual(1)	Average
14 hours	1	891	29	3.25	2.96 ± 1.10
	2	861	15	1.74	
	3	695	27	3.88	
16 hours	1	740	48	6.49	6.91 ± 1.20
	2	769	46	5.98	
	3	980	81	8.27	
18 hours	1	843	43	5.10	7.56 ± 2.29
	2	644	62	9.63	
	3	804	64	7.96	
32 hours	1	1047	68	6.49	10.46 ± 3.85
	2	1093	117	10.70	
	3	825	117	14.18	
34 hours	1	1023	169	16.52	22.52 ± 5.92
	2	908	206	22.69	
	3	963	273	28.35	
36 hours	1	1068	163	15.26	16.52 ± 1.10
	2	1087	186	17.11	
	3	1151	198	17.20	

(1) Not adjusted for the number of Cm^r colonies.

In vivo transposition was performed by incubating 3 single DH5α colonies carrying pXF116 and pUC18 for 14 hours, 16 hours and 18 hours. After 18 hours the cells were diluted 1:1000 into fresh media and incubated for a further 14 hours, 16 hours and 18 hours (32 hours, 34 hours and 36 hours total incubation time). The plasmids were purified and 100 ng of the DNA from each individual assay was electroporated into DS964. One ten thousandth of each electroporation was plated onto an ampicillin plate and the rest was plated on a kanamycin plate. The transposition frequency was calculated by dividing the total number of the Kan^r colonies by total number of Amp^r colonies.

hours. The DNA was electroporated into DS964 and the transposition frequency was calculated from the numbers of Kan^r and Amp^r colonies. The transposition frequencies from the different colonies varied by less than a factor of 2.5 at each time-point. The results at each time point were averaged, giving an equal weighting to the results from each starting colony and the standard deviation was calculated to give an idea of statistical variation in the assay. After the first overnight incubation, the transposition frequency increased from $3.0 \pm 1.1 \times 10^{-6}$ at 14 hours to $7.6 \pm 2.3 \times 10^{-6}$ at 18 hours incubation. After a second overnight incubation, the transposition frequency increased further to $10.5 \pm 3.9 \times 10^{-6}$ at 14 hours and $22.5 \pm 5.9 \times 10^{-6}$ at 16 hours incubation, but then decreased to $16.5 \pm 1.1 \times 10^{-6}$ at 18 hours. Although some of this variation may be due to random variation in the assay, there is an upward trend in the transposition frequency over time (Figure 3.4). The results suggest that longer incubation times and the second overnight culture lead to higher transposition frequencies. This presumably reflects the accumulation of transposition events into the target plasmid over time. However, competition between the target plasmid with and without mini-transposon insertions might complicate matters. The results suggested that it was important to fix the incubation time to make results comparable between different assays. For convenience, it was decided to use two overnight incubations of 18 hours in the standard assay.

3.3.5 Variability in the transposition assay

It was noticed in the time course and other experiments described above that there was some variation in transposition frequency obtained from independent repeats of the same experiments. Transposition is an inherently random event, and depending on whether transposition has occurred early or late in the colony that is picked, a higher or lower transposition frequency might be obtained.

One way to attempt to smooth out this variation is to pool multiple transformant colonies and to prepare DNA from cultures started from these pools. To investigate the effects of this procedure, a transposition assay was carried out with pXF116 as donor and pUC18 as target. DNA was prepared from three individual transformants and from three pools of three colonies, then electroporated into DS964. Transposition frequencies from the three single colonies ranged from 1.94×10^{-5} to 9.45×10^{-5}, a difference of nearly 5 fold (Table 3.3). As expected, taking pools of three colonies seemed to smooth this variation somewhat, giving frequencies from 3.98×10^{-5} to 5.46×10^{-5} in the three

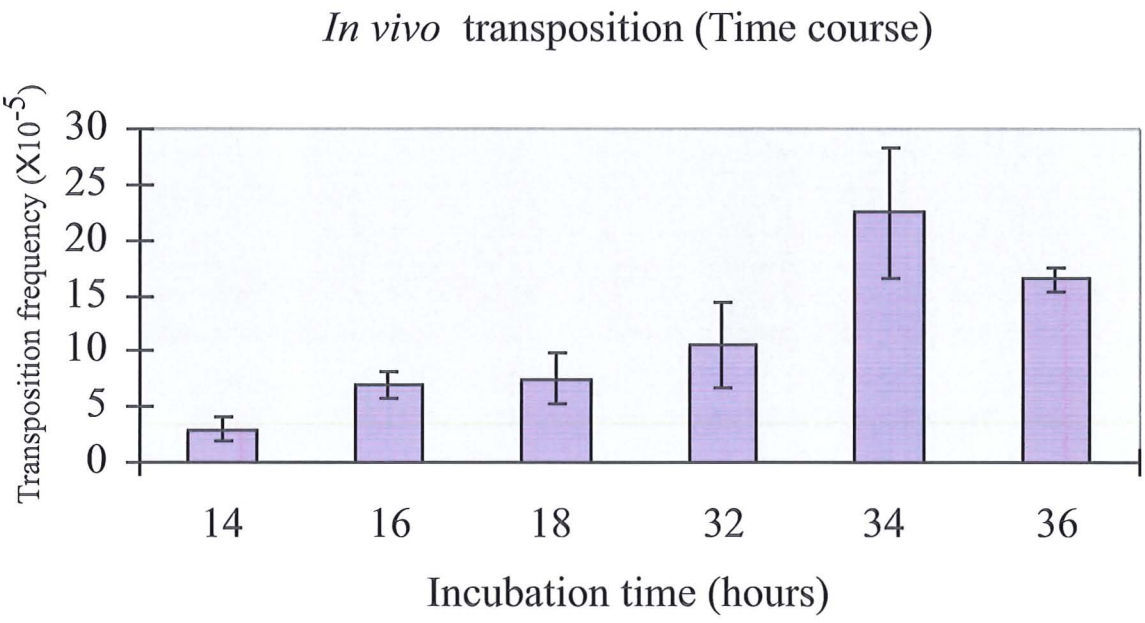


Figure 3.4 Incubation time affects the *in vivo* transposition frequency. *In vivo* transposition assays to test the impact of incubation time on the transposition frequency were carried out as described in Table 3.2. The average transposition frequency after different incubation times is shown. The error bars represent the standard deviation σ_{n-1}.

Table 3.3 Comparison of transposition frequencies obtained from assays started with multiple colonies and single colonies

Starting colonies	Reaction No.	Amp^r ($\times 10^4$)	Kan^r	Transposition frequency ($\times 10^{-5}$)	
				Individual[1]	Average
Single colonies	1	504	330	6.55	5.98 ± 3.79
	2	904	175	1.94	
	3	513	485	9.45	
Multiple colonies	1	485	265	5.46	3.98 ± 1.49
	2	1610	640	3.98	
	3	1204	300	2.49	

(1) Not adjusted for the number of Cm^r colonies.

In vivo transposition assays employing pXF116 as donor and pUC18 as target were carried out using standard conditions. Cultures were started from 3 single colonies or from 3 mixtures of 3 single colonies.

Table 3.4 The amount of DNA electroporated affects the calculated transposition frequency

DNA used	Kan^r colonies	Amp^r colonies ($\times 10^6$)	Transposition efficiency ($\times 10^{-4}$)
50ng	151	1.06	1.42
100ng	277	1.34	2.07
200ng	342	1.26	2.71
300ng	375	2.43	1.54
500ng	257	1.34	1.94

Different amounts of DNA from a single *in vivo* transposition assay were electroporated into same amount of competent cell, and transposition frequencies were calculated using the standard method.

Table 3.5 *In vivo* transposition employing pH2 as target plasmid

pH2 +	Amp^r ($\times 10^4$)	Kan^r Cm^s	Transposition frequency ($\times 10^{-4}$)	
			Individual	Average
pXF114 (ISY100-30-Kan, Tnp^-)	23	0	$< 4.3 \times 10^{-6}$	$< 2.4 \times 10^{-6}$
	18	0	$< 5.6 \times 10^{-6}$	
pXF116 (ISY100-30-Kan, Tnp^{wt})	4	18	4.50	2.21 ± 1.53
	15	24	1.60	
	510	1128	2.21	
pXF118 (ISY100-30-Kan, Tnp^{His})	6	35	5.83	4.84 ± 3.75
	6	48	8.00	
	540	379	0.70	

replicates carried out (Table 3.3). It was decided it might be best to pool multiple colonies in future assays as done here, however, most of the assays reported in this chapter were carried out by taking three single colonies.

Another factor that may affect the apparent transposition frequency is that the electroporation results may not reflect the true ratio of the plasmid molecules present. Specifically, the total amount of DNA used in the electroporation may affect the observed ratio. To investigate this, different amounts of the same DNA isolated from a transposition reaction with pXF116 and pH2 (see later) were electroporated into DS964. The number of Amp^r and Kan^r colonies depended on the amount of DNA, resulting in apparent transposition frequencies varying by nearly a factor of two between the highest and lowest (Table 3.4). It seems likely that different batches of competent cells, with different levels of competence might give similar variation to that seen here with different amounts of DNA. Thus the electroporation and plating procedures are another source of variability in these assays.

Mixtures of known ratios of target plasmid and the transposition product could be used to determine whether the frequencies obtained by electroporation reflect the true ratio of DNA molecules present, however this was not done.

The variation in transposition frequency obtained from independent repeats of the same experiment should be kept in mind when interpreting all of the results from *in vivo* transposition assays in this chapter and chapter V.

3.3.6 The target plasmid affects the transposition frequency

In their *in vivo* assays, the transposition frequency obtained by Urasaki *et al.* (2002) was 3.3×10^{-4} Cm^r Sp^r colonies per Sp^r colony in the presence of IPTG. However, the average frequency obtained here with pUC18 and pXF115, which carries a transposon with the same ends used by Urasaki *et al.* (2002), was approximately 2.4×10^{-5} Kan^r colonies per Amp^r colony in the presence of IPTG and 1.5×10^{-5} in absence of IPTG (Table 3.1). pXF116, with only 30 bp ISY100 ends, gave a higher frequency (8.11×10^{-5}) than pXF115 in the presence of IPTG (Table 3.1), but this is still 4-fold less than the frequency obtained by Urasaki *et al.* (2002). One difference between the assays described here so far, and that of Urasaki *et al.* (2002) is that they use a larger target plasmid (6.6 kb), with more potential TA target sites. It was therefore decided to test the effect of changing the target plasmid in the assay developed here. An alternative target plasmid called pH2 was used. This plasmid contains a 2.6 kb TA rich DNA fragment from *Aspergillus nidulans* in the pUC18-like

cloning vector pIC20. IS630 from *S. sonnei* transposes almost exclusively into the target sequence CTAG (Tenzen and Ohtsubo, 1991), a rare tetranucleotide in pUC18 and most other bacterial sequences (Burge *et al.*, 1992). pH2 was originally used as a target in an *in vivo* transposition assay for IS630, where it gave higher transposition frequencies than pUC18 because it contains a large number of CTAG tetranucleotides (my unpublished results).

pH2 was therefore tested as a target for ISY100 transposition, and the results are shown in Table 3.5. The transposition frequency obtained with pH2 was generally greater than 10^{-4} in absence of IPTG (average of 3 experiments with pH2 and pXF116 = $2.77 \pm 1.53 \times 10^{-4}$), whereas frequencies obtained with pUC18 as target were generally lower than 10^{-4} (average transposition frequency for 6 experiments with pXF116 and pUC18 = $3.81 \pm 3.37 \times 10^{-5}$; Table 3.2, 3.3 absence of IPTG).

This increase in transposition frequency might be caused simply by the increased size of the target plasmid and/or the increased number of available TA target sequences. Assuming that ISY100 transposes only into TA target sites, the number of potential targets in pUC18 and pH2 can be calculated. Insertions into the plasmid replication origin are likely to inactivate the origin and should not be recovered in the transposition assay. Therefore, the 589 bp region between the -35 box of the RNAII promoter and the point where DNA synthesis starts from the RNAII primer was excluded from this analysis. pH2 contains 343 potential TA target sites outside the origin region, whereas pUC18 contains only 107. This three-fold difference in the number of target sites might therefore account for a large proportion of the difference in the observed transposition frequency between these two targets. Transposition of ISY100 is not random and there are some sites that are used preferentially as targets (see chapter V). It is possible that some of the difference in transposition frequency between pH2 and pUC18 could be accounted for by a difference in the number of preferred target sites on these two plasmids.

3.3.7 Kan^r Cm^s colonies represent genuine transposition events

Plasmid DNA was purified from 10 Kan^r transformant colonies obtained after *in vivo* transposition with pXF116 and pH2 (here referred to as T1-T10). A further 10 plasmids (here referred to as T11-T20) were purified after transposition with pXF118 and pH2. All of them, except one from the pXF118 pH2 reaction, were chloramphenicol sensitive, and were therefore thought to be the products of transposition. Plasmid DNA isolated from these 20 transformants was run uncut and cut with BamHI on an agarose gel (Figure 3.5).

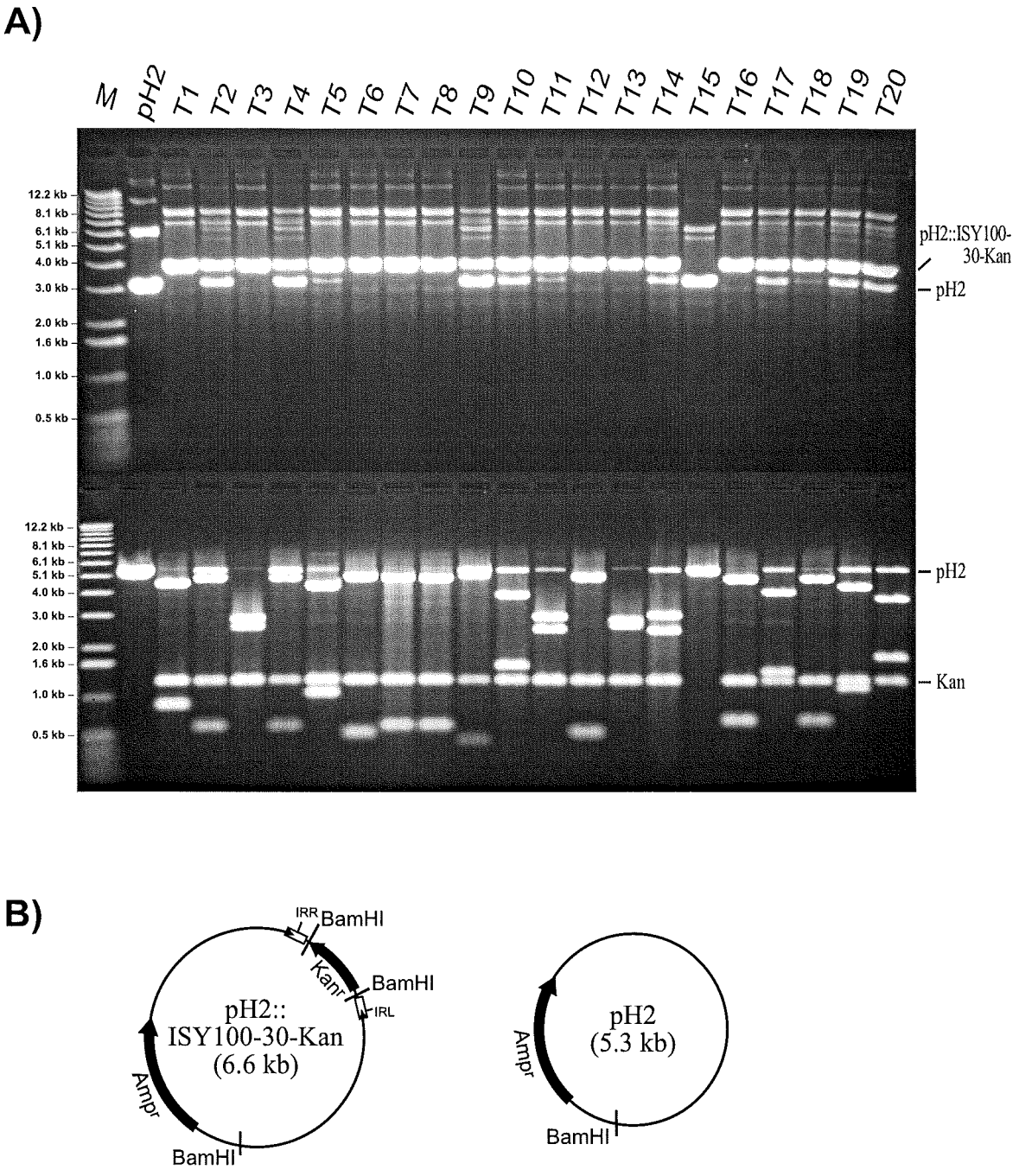


Figure 3.5 Restriction analysis of purified plasmids from *in vivo* transposition assays. **A)** Lanes T1 to T10 are plasmids purified from Kan^r Cm^s colonies from the *in vivo* transposition assay which employed pH2 as target plasmid and pXF116 as donor. Lanes T11 to T20 are plasmids purified from Kan^r Cm^s (Kan^r Cm^r for T15) colonies from the *in vivo* transposition assay which employed pH2 as target plasmid and pXF118 as donor. Top gel: Uncut plasmid from each colony. Bottom gel**:** BamHI digestion of each purified plasmid. **B)** pH2 is 5.3 kb and contains a single BamHI site. The transposition product pH2::ISY100-30-Kan is 6.6 kb and contains three BamHI sites. Cleavage of transposition products is predicted to give a Kan^r fragment of 1.3 kb (Kan) and two fragments of different sizes (adding up to 5.3 kb) depending on the site of insertion.

All of the Kan^r Cm^s colonies contained a plasmid of the same size, about 1.3 kb larger than pH2 consistent with an insertion of ISY100-30-Kan into pH2 (Figure 3.5A). Some of them also carried a minor extra plasmid the same size as pH2. This probably comes from co-transformation of DS964 with two plasmid molecules, one molecule of pH2 and one of pH2::ISY100-30-Kan. Although these two plasmids contain the same origin and should be incompatible, it may take many generations for the Kan^s pH2 to segregate out.

pH2 contains a single BamHI site, and ISY100-30-Kan contains two BamHI sites flanking the 1.3 kb kanamycin resistance gene. Therefore, pH2 with a ISY100-30-Kan insertion is expected to give a 1.3 kb Kan^r fragment, and two other fragments that add up to the size of pH2 (5.3 kbp) when cleaved with BamHI. The exact sizes of these two fragments will depend on the position of the insertion. Cleavage with BamHI gave different restriction patterns for the different colonies, consistent with single insertions of the mini-ISY100 into different positions in pH2. Those colonies containing a pH2-sized plasmid in the uncut sample also gave an extra band of the correct size to be linear pH2 in the BamHI cleaved samples.

The Kan^r Cm^r colony from the pXF118 + pH2 transposition reaction contained only a plasmid of the correct size to be pH2 (T15 in Figure 3.5A). This supports the hypothesis that this class of colony comes from integration of the donor plasmid into the DS964 chromosome, and reinforces the need to check that transformants are Cm^s.

All 19 transposition products were sequenced using outward facing primers from the left and right ends of the ISY100-30-Kan (primers Kan_L and Kan_R, Table 2.3). The sequencing results confirmed that all 19 plasmids contained insertions into TAs in pH2. The insertions occurred precisely at the boundaries of the ISY100 ends, and the TA target sequences appear to be duplicated. All of these properties are those expected for legitimate transposition events.

Four of the 19 insertions sequenced were into the same site in the pIC20 vector sequence, and another two shared a site in the TA-rich *A. nidulans* sequence in pH2 (Figure 5.4). It is not clear if these are independent transposition events or sibling plasmids. 11 out of the 15 different target sites (73%) were in the *A. nidulans* sequence, and the other four (27%) were in the pIC20 vector sequences, reflecting the fraction of TA dinucleotides in these regions (69% and 31% respectively). The insertions sites used are analysed in more detail in chapter V.

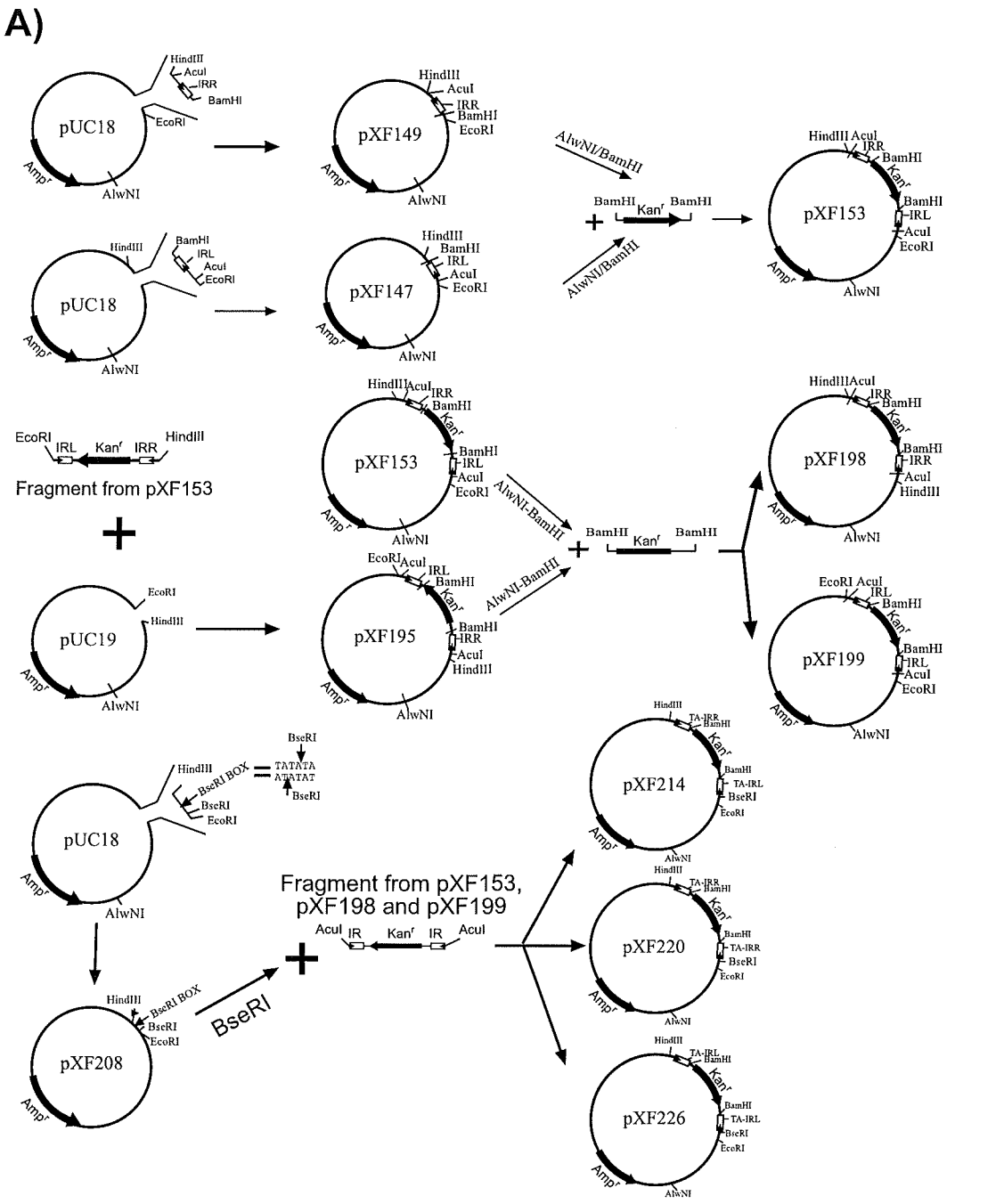
3.4 The two ends of ISY100 are not equally proficient for transposition

The *in vivo* transposition assay described in the previous sections can be used to study the requirements for transposition and the steps involved in the transposition reaction. In this section, differences between the left and right ends of ISY100 are explored by using mini-transposons containing either two ISY100 left ends or two ISY100 right ends.

Like many transposons, ISY100 is bounded by imperfect inverted repeats. The terminal inverted repeats of ISY100 are 24 bp long, not counting the presumed TA target duplication. 19 out of these 24 bp are identical between the left and right ends, and there are thus 5 mismatches (Figure 3.1). It is possible that one or both ends of ISY100 are suboptimal for transposition. As a first step to investigate this possibility, mini-transposons containing either two left or two right ends were constructed and tested for transposition.

A strategy was devised to construct Kan^r mini-ISY100 elements with different combinations of left and right ends, and with different flanking dinucleotides (Figure 3.6). Double stranded oligonucleotides representing 30 bp from the left or right end of ISY100 were inserted into pUC18 between EcoRI and BamHI or HindIII and BamHI sites respectively. Fragments from these plasmids were then combined with the BamHI Kan^r fragment from pUC4K to construct pXF153, carrying a marked mini-ISY100 element with one left and one right end of ISY100 (Figure 3.6A). This mini-ISY100 was then moved into pUC19 to create pXF195, and fragments from pXF153 and pXF195 were combined to make pUC-based plasmids carrying mini-ISY100s with two right ends (pXF198; Figure 3.6A) and two left ends (pXF199; Figure 3.6A). These three plasmids were designed so that the mini-ISY100s could be excised without flanking TA dinucleotides by digestion with AcuI, and inserted into the BseRI site of pXF208 (Figure 3.6A). This yielded a set of plasmids carrying marked mini-ISY100 elements with one left and one right end (pXF214, TA-IRL/TA-IRR), two right ends (pXF220, TA-IRR/TA-IRR) or two left ends (pXF226, TA-IRL/TA-IRL). These three mini-ISY100s, with identical flanking sequences and differing only in the sequences of their inverted repeats, were used to make λ-dv based donor plasmids similar to pXF116 (pXF252, pXF258 and pXF259; Figure 3.6 D). It should be noted that the sequences flanking these artificial mini-transposons are different from those found in pXF116, which carried a mini-ISY100 in its natural *Synechocystis* ISY100a genomic context.

The new donor plasmids were used in an *in vivo* transposition assay with pH2 as target. The transposition frequency for pXF252 (TA-IRL/TA-IRR) was $2.7 \pm 1.5 \times 10^{-4}$ (Table 3.6A), similar to the frequency found for pXF116 using pH2 as a target ($2.2 \pm 1.5 \times$


A)
pUC18
pXF149
pXF147
pXF153
pXF195
pXF198
pXF199
pUC19
pXF214
pXF220
pXF226
pXF208
Fragment from pXF153
Fragment from pXF153, pXF198 and pXF199
BseRI
AlwNI/BamHI
AlwNI-BamHI
AcuI
B)
pUC18
pXF209-pXF213
pXF153
pXF215-pXF219
pXF218-2 pXF219-2
AcuI
BseRI
pXF215: TG-L/TG-R
pXF216: TC-L/TC-R
pXF217: TT-L/TT-R
pXF218: AA-L/AA-R
pXF219: CA-L/CA-R

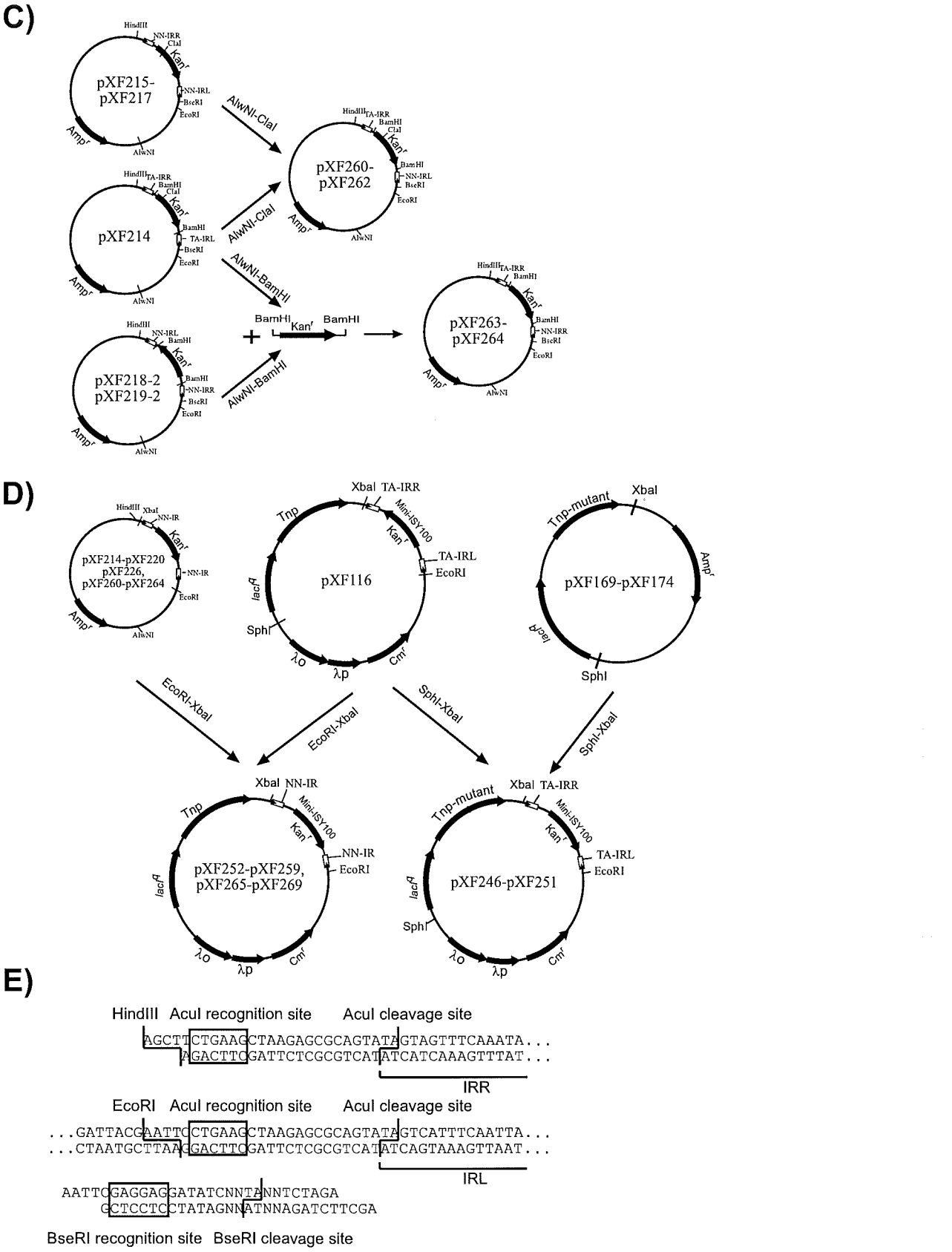


Figure 3.6 Construction of donor plasmids carrying transposase mutants, changes at the IRs of mini-ISY100 or the flanking sequence of the mini-ISY100. A) Construction of pUC18-based plasmids carrying mini-ISY100 with IRL and IRR, two IRRs and two IRLs. **B)** Construction of pUC18-based plasmids carrying mini-ISY100 with identical flanking sequences apart from TA dinucleotides. **C)** Construction of pUC18-based plasmids carrying mini-ISY100 flanked by TA at one end and other dinucleotides at the other end. **D)** Construction of λ-dv donor plasmids from pUC18-based plasmids carrying mini-ISY100 with different ends or flanking dinucleotides, or pTrc99a-based plasmids carrying mutated transposase genes. **E)** Sequences of ds oligonucleotides inserted into pUC18 to make pXF149, pXF147 and pXF208-213.

Table 3.6 Mutations at the transposase, transposon ends and flanking dinucleotides affect *in vivo* transposition frequency

A) Inverted repeats sequences affect *in vivo* transposition frequency

pH2 +	Mini-ISY100 and Flanking sequence	Reac-tion No.(1)	Amp^r ($\times 10^4$)	Kan^r	Cm^r	Kan^r Cm^s	Transposition frequency ($\times 10^{-6}$)	
							Individual	Average/ (relative)
pXF252	TA-L/TA-R (Wild-type)	16	156	400	1/20	380	244	273 ± 152 (100%)
		17	101	442	0/20	442	438	
		18	134	186	0/20	186	139	
pXF258	TA-R/TA-R	34	153	103	0/20	103	67.3	65.1 ± 11.4 (23.8%)
		35	145	109	0/20	109	75.2	
		36	201	112	1/20	106	52.7	
pXF259	TA-L/TA-L	37	98	548	0/20	548	559	563 ± 27 (206%)
		38	122	656	0/20	656	538	
		39	117	692	0/20	692	591	

B) Mutations in DDE motif inactivate ISY100 transposase

pH2 +	Mutation in Transposase	Reac-tion No.(1)	Amp^r ($\times 10^4$)	Kan^r	Cm^r	Kan^r Cm^s	Transposition frequency ($\times 10^{-6}$)	
							Individual	Average / (relative)
pXF252	Wild-type	16	156	400	1/20	380	244	273 ± 152 (100%)
		17	101	442	0/20	442	438	
		18	134	186	0/20	186	139	
pXF246	D136A	1*	185	5	0/5	5	2.70	3.75 ± 1.19 (1.37%)
		2*	99	6	1/6	5	5.05	
		3*	143	9	4/9	5	3.50	
pXF247	D218A	4*	155	7	1/7	6	3.87	3.03 ± 0.96 (1.11%)
		5*	151	3	0/3	3	1.99	
		6*	155	7	2/7	5	3.23	
pXF248	E253A	7*	146	9	1/9	8	5.48	3.10 ± 2.14 (1.13%)
		8*	149	2	0/2	2	1.34	
		9*	162	4	0/4	4	2.47	
pXF249	K260A	40*	162	2	0/2	2	1.23	1.40 ± 0.97 (0.51%)
		41*	123	4	1/4	3	2.44	
		42*	189	1	0/1	1	0.53	
pXF250	R261A	10*	126	7	1/7	6	4.76	3.39 ± 1.52 (1.24%)
		11*	164	8	2/8	6	3.66	
		12*	170	7	4/7	3	1.76	
pXF251	KR260AA	13*	182	8	4/8	4	2.20	2.60 ± 0.84 (0.95%)
		14*	147	5	2/5	3	2.04	
		15*	140	7	2/7	5	3.57	

C) Mutations in flanking dinucleotides affect *in vivo* transposition frequency

pH2+	Mini-ISY100 and Flanking sequence	Reac-tion No.(1)	Amp^r ($\times 10^4$)	Kan^r	Cm^r	Kan^r Cm^s	Transposition frequency ($\times 10^{-6}$)	
							Individual	Average/ (relative)
pXF253	TG-L/TG-R	19*	72	5	1/5	4	5.56	16.2 ± 18.1 (5.92%)
		20	97	42	3/20	36	37.1	
		21*	68	7	3/7	4	5.88	
pXF254	TC-L/TC-R	22*	105	3	1/3	2	1.90	1.55 ± 0.55 (0.57%)
		23*	109	3	2/3	1	0.92	
		24*	163	6	3/6	3	1.84	
pXF255	TT-L/TT-R	25*	186	4	3/4	1	0.54	1.32 ± 0.70 (0.48%)
		26*	157	4	1/4	3	1.91	
		27*	200	6	3/6	3	1.50	
pXF256	AA-L/AA-R	28*	192	3	2/3	1	0.52	2.37 ± 2.28 (0.87%)
		29*	300	7	2/7	5	1.67	
		30	183	12	3/12	9	4.92	
pXF257	CA-L/CA-R	31	235	201	0/20	201	85.5	105 ± 73 (38.5%)
		32	189	352	0/20	352	186	
		33	184	85	1/20	81	44.0	
pXF265	TG-L/TA-R	43	131	518	0/20	518	395	335 ± 59 (123%)
		44	192	535	0/20	535	279	
		45	262	868	0/20	868	331	
pXF266	TC-L/TA-R	46	159	11	4/11	7	4.40	10.4 ± 6.0 (3.82%)
		47	163	22	5/22	17	10.4	
		48	164	34	4/20	27	16.5	
pXF267	TT-R/TA-R	49	167	96	2/20	86	51.5	49.5 ± 10.1 (76.1%)(2)
		50	178	104	0/20	104	58.4	
		51	210	81	0/20	81	38.6	
pXF268	AA-R/TA-R	52	149	260	1/20	247	166	174 ± 11 (267%)(2)
		53	167	312	0/20	312	187	
		54	162	288	1/20	274	169	
pXF269	CA-R/TA-R	55	126	59	2/20	53	42.1	89.0 ± 40.7 (137%)(2)
		56	176	214	2/20	193	109	
		57	177	204	0/20	204	115	

(1): Reaction number corresponds to the lane number in Figure 3.8
(2): Compared with the donor plasmid carrying mini-ISY100 TA-R/TA-R.
*: No DNA was recovered from single Kan^r Cm^s colonies picked from these reactions.

Standard *in vivo* transposition assays employing pH2 as target plasmid were carried out as described in section 2.19. Donor plasmids containing mutants in the transposase, or changes in the mini-ISY100 at the IRs or flanking dinucleotides were tested. For each donor plasmid, three individual assays were performed. The transposition frequency is given as the ratio of Kan^r Cm^s to Amp^r colonies and also as a percentage of the frequency obtained with the appropriate unchanged donor plasmid.

10^{-4}; Table 3.5). Thus there seems to be little difference in transposition frequency due to the differences in flanking sequences in the new plasmids. The mini-ISY100 with two left ends (pXF259; TA-IRL/TA-IRL) transposed 2.1-fold more efficiently than the transposon with one left and one right end, giving a transposition frequency of $5.6 \pm 0.3 \times 10^{-4}$ (Table 3.6A). In contrast, the transposon with two right ends (pXF258; TA-IRR/TA-IRR) transposed 4.2-fold less efficiently than the transposon with one left and one right end, giving a transposition frequency of $6.5 \pm 1.1 \times 10^{-5}$ (Table 3.6A). It seems that the sequence of the left end is in some way more optimal for transposition than the right end. The step in the transposition reaction affected by these sequence differences is studied in more detail in chapter IV.

3.5 The effects of changing the flanking TA dinucleotides

As well as the transposon ends, the flanking sequences of some transposons may affect their transposition (Ason and Reznikoff, 2004; Jaworski and Clewell, 1994). Although no difference was observed in transposition frequency between a donor plasmid containing ISY100 in its natural context (pXF116, Table 3.5) and one containing ISY100 in an artificial context (pXF252, Table 3.6A), both of these donor transposons are flanked by the normal TA dinucleotide. To see if this sequence is important for transposition, it was decided to change it at one or both transposon ends. Another reason to change the sequence of the TA dinucleotides flanking ISY100 is to obtain information about the boundaries of the transposon (see section 1.4.1).

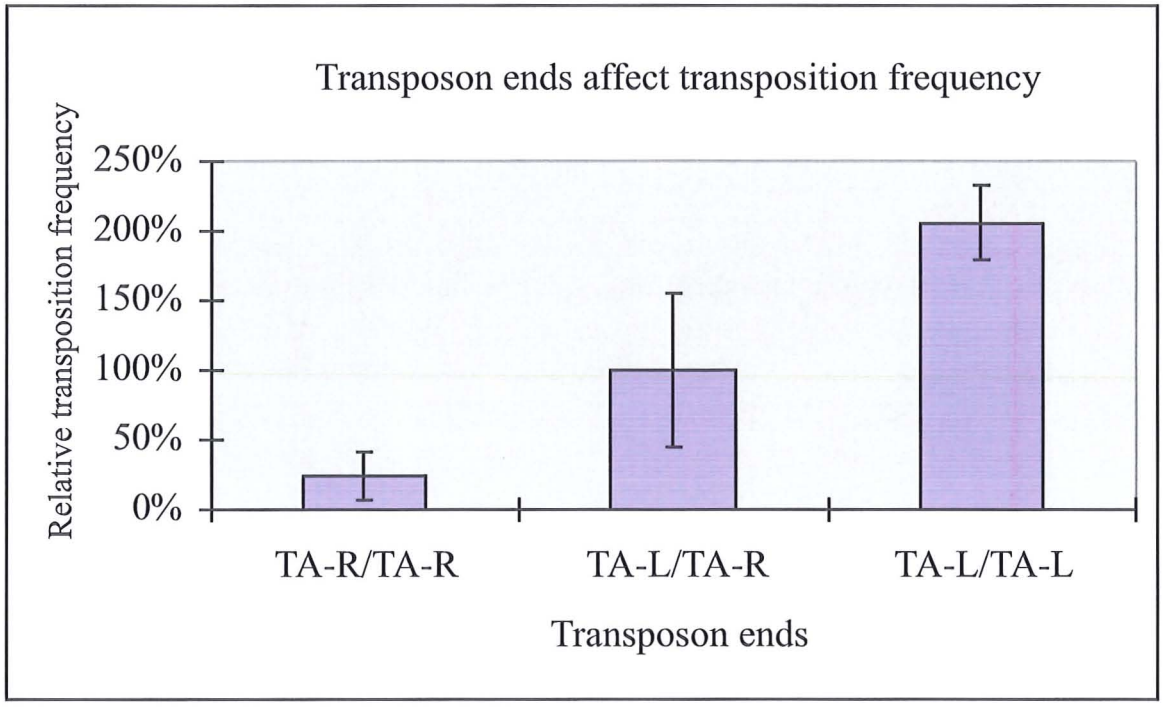
The AcuI fragment from pXF153, carrying the mini-ISY100 without its flanking TA dinucleotides, was inserted into a series of double stranded oligonucleotides cloned in pUC18. This produced a set of plasmids carrying mini-transposons with altered flanking dinucleotides (Figure 3.6B, E). These mini-transposons had one left inverted end (IRL), one right inverted repeat (IRR) and identical changes in the flanking dinucleotides at both ends. These mini-transposons are referred to here as NN-L/NN-R, where the sequence of the flanking dinucleotide is given at both ends reading 5' to 3' from the outside towards the transposon inverted repeat. The A at position –1, just outside the proposed transposon end, was changed to all three possible nucleotides, producing TG-L/TG-R, TC-L/TC-R and TT-L/TT-R. The T at position –2 was changed to two of the three other possible nucleotides, giving CA-L/CA-R and AA-L/AA-R.

λ-dv based donor plasmids carrying the wild-type transposase cassette and these altered mini-transposons were constructed (Figure 3.6 D) and tested in a standard *in vivo* transposition assay (Table 3.6C). Changing the A at the –1 position to a G at both ends led to a reduction in transposition frequency to 5.9% of the wild-type level (Figure 3.7). Changing this A to C or T at both ends led to a further decrease in transposition frequency, to 0.6% or 0.5% of the wild-type level (Figure 3.7). Changing the T at the –2 position to C at both ends resulted in a decrease in transposition frequency to 38.5% of the wild-type level, while changing it to an A decreased the transposition frequency to 0.9% of wild-type.

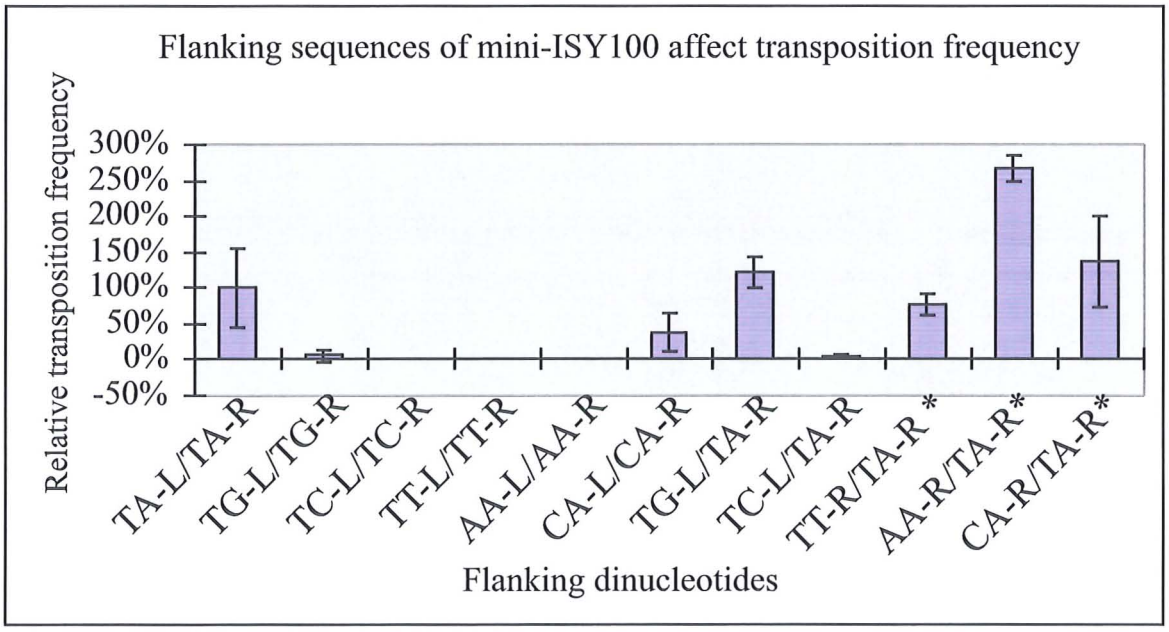
Donor plasmids were also constructed with changes to the flanking TA at only one end (Figure 3.6C). Two plasmids were constructed containing IRR with the normal flanking TA, and IRL with the A at position –1 changed to either G or C (TG-L/TA-R and TC-L/TA-R). Another three plasmids were constructed, which unintentionally contained two copies of IRR, with TA at one end and TT, AA or CA at the other end (TT-R/TA-R, AA-R/TA-R and CA-R/TA-R) (Figure 3.6C). In the constructs with one IRR and one IRL, changing the A adjacent to IRL at position -1 to G gave a small increase in the transposition frequency, whereas changing this A to C decreased the frequency to 3.8% of the wild-type transposition level (Table 3.6C; Figure 3.7). Changing the A at position –1 to T at one end of the IRR/IRR mini-ISY100 decreased the transposition frequency slightly, to 76% of the level seen for TA-R/TA-R. Changing the T at position –2 to A, at one end of the IRR/IRR mini-ISY100, increased the transposition 2.67-fold, while changing this T to a C increased the transposition frequency 1.37-fold relative to TA-R/TA-R. It is hard to explain why changes at one end appear to increase the transposition frequency of the mini-transposon with two right ends, and some of these results might be due to random variation in the assay. However some general conclusions can be drawn. The effect of changing the TA at only one end is much less severe than the effect of the equivalent change at both ends (Figure 3.7). At position –1 the order of preference appears to be A>>G>T≅C, so that purines are preferred over pyrimidines, but a T is better than a C, while the base at the –2 position seems to be less important than the base at position –1.

Plasmid DNA was purified from a single Kan^r Cm^s colony for each transposition reaction (3 for each donor plasmid), and a BamHI digestion was carried out to check that transposition had taken place. Surprisingly, a number of colonies failed to give any plasmid DNA at all (Figure 3.8). These colonies were all from reactions giving low

A)



B)



* The relative transposition frequencies was calculated relative to TA-R/TA-R

Figure 3.7 A) Transposon ends affect the *in vivo* transposition frequency. **B)** Flanking dinucleotides of the mini-ISY100 affect the *in vivo* transposition frequency. Data are from Table 3.6.

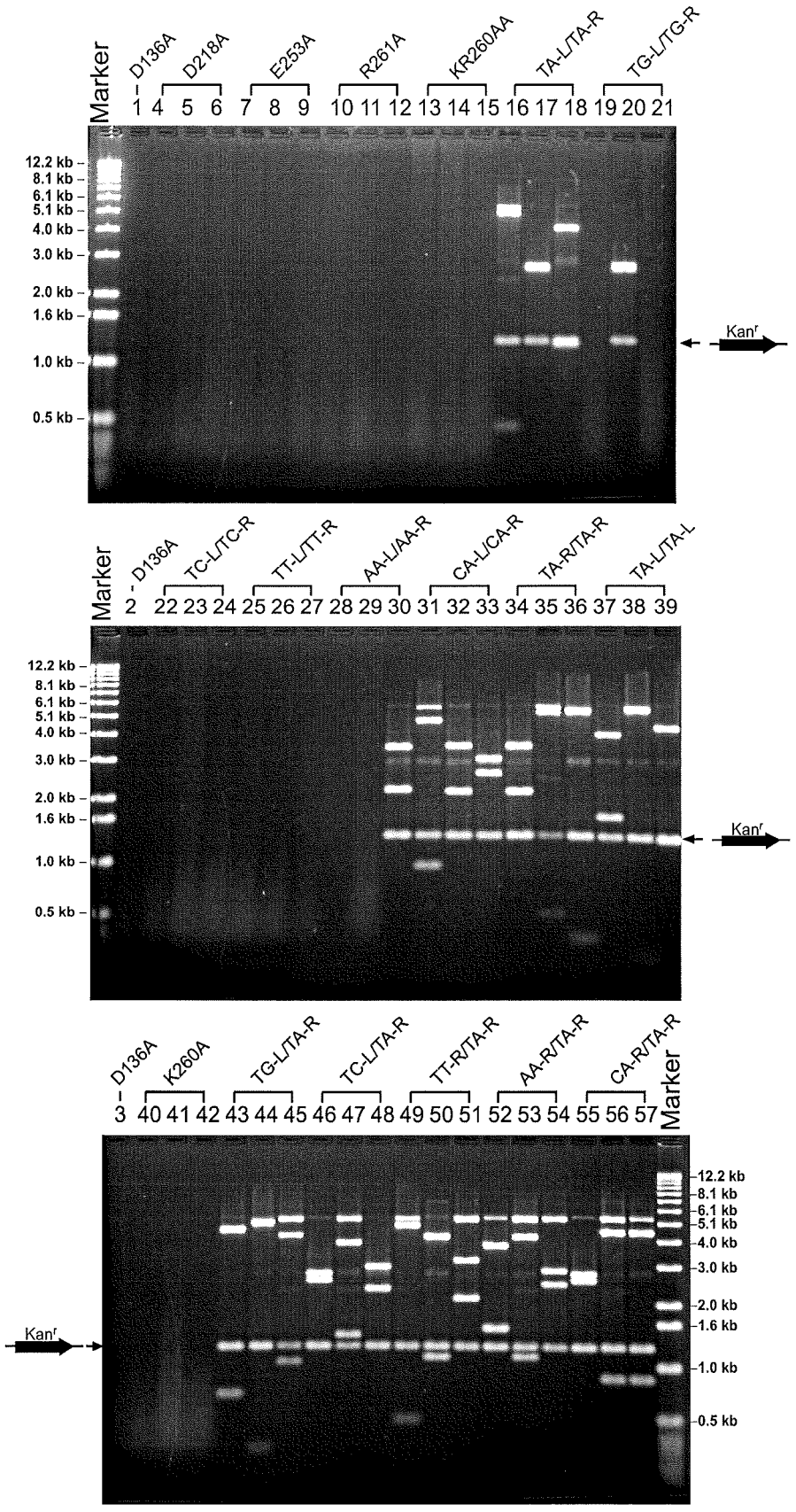


Figure 3.8 Restriction analysis of purified plasmids from *in vivo* transposition assays. *In vivo* transposition assays employing target plasmid pH2 and donor plasmids carrying transposase mutants, changes at the IRs of mini-ISY100 or the flanking sequence of the mini-ISY100 were carried out as described in Table 3.6. Plasmid was purified from one randomly selected single Kan^r Cm^s colony for each individual assay and subjected to 1.2% agarose gel electrophoresis after cleavage with BamHI. The number of each digestion corresponds to the number of the assay showed in Table 3.6. A brief description of the assay is given above each set of 3 lanes.

transposition frequencies (< 1% of wild-type level, or a total number of Kan^r Cm^s colonies < 7) and might be explained in several ways. One possibility is that transposition occurs after electroporation into DS964, and is from the donor plasmid onto the chromosome of DS964. This will give a Kan^r Cm^s colony carrying an insertion of marked mini-ISY100 on the chromosome of DS964, but containing no plasmid. Another possibility is that there was a low level of contamination of the DS964 electrocompetent cells during this experiment. Unfortunately, a negative control electroporation was not carried out to test this possibility during these experiments. It seems likely that many or all of the colonies seen at low frequency in these assays are the result of some background process that may not involve transposition, so that the transposition frequencies reported in Table 3.6 should be taken as maximum values.

All of the colonies that yielded plasmid DNA gave restriction patterns consistent with insertion of mini-ISY100 into different sites in pH2 (Figure 3.8). This included one plasmid from a donor with the A at position -1 changed to a G at both ends and 9 plasmids from donors with the A at position -1 changed to G, C or T at only one end. All of the resulting plasmids were sequenced with outward facing mini-ISY100 primers. All of the mini-ISY100 had inserted into TA targets in pH2 and were flanked by TA dinucleotides at both ends, regardless of the flanking sequence in the donor plasmid (Figure 5.4). Therefore the TA dinucleotides at the insertion sites are target duplications and are not part of the transposon, consistent with the sequence of the excised ISY100 reported by Urasaki *et al.* (2002) and the cleavage sites mapped *in vitro* in chapter 4.

3.6 Mutations in the DDE catalytic motif abolish transposition *in vivo*

To test the requirement for the three acidic residues of the proposed DDE motif of ISY100 transposase, the codons for D136, D218 and E253 (Figure 3.1) were individually changed to alanine codons by site-directed mutagenesis. The mutant transposase genes were then used to replace the wild-type transposase gene in pXF116 to make a new set of donor plasmid (Figure 3.6E), and the tested in triplicate in a standard *in vivo* transposition assay (Table 3.6B). The transposition frequencies from all three mutant transposases were less than 1.5% of the wild-type frequency, and no more than 6 Kan^r Cm^s colonies were observed in any transposition assay. A single Kan^r Cm^s colony was picked from each assay and used to isolate plasmid DNA. However, no plasmid DNA was recovered from any of these colonies (Figure 3.8). These assays were carried out at the same time as the assays

reported in section 3.5, and it appears that the same level of background events are giving rise to Kan^r Cm^s colonies that do not represent transposition events. Nevertheless, it is safe to conclude that mutations in the DDE motif substantially reduce the transposition frequency.

A number of transposases from various transposon families contain a conserved basic residue (lysine or arginine) 6-7 amino acids downstream from the glutamate residue in the DDE motif. In the IS4 family, this conserved lysine is a part of Y(2)R(3)E(6)K signature, in which the E is the last acidic amino acid of the DDE motif acting in the catalytic activity. This motif is involved in binding hairpins at the termini of these elements. Studies on the IS4 members Tn5 (Naumann and Reznikoff, 2002) and Tn10 (Bolland and Kleckner, 1996) demonstrated that mutations in the Y or R residues of this YREK signature reduced the overall catalytic activity of transposases whereas mutation at the K residue reduced the ability of Tn5 transposase to form protein-DNA complexes (Naumann and Reznikoff, 2002). Although no conserved Y and R residues are found in the correct position relative to the DDE motif in members of IS630/Tc1/*mariner* superfamily, some active transposases like Tc1, *Sleeping Beauty* and ISY100 have basic residues (K and/or R) in the correct position. ISY100 transposase has two basic residues (K260 and R261) 6 and 7 residues downstream from E253. To see if these residues are important for catalysis of transposition, they were changed to alanine, and tested for their effects on transposition *in vivo*. Single mutants K260A and R261A and the double mutant transposase gave transposition frequencies of <1.5% of wild-type transposase, and again no plasmid DNA could be recovered from the Kan^r Cm^s colonies representing apparent transposition events (Figure 3.7, Table 3.6B).

3.7 Discussion

In this chapter, a standardised *in vivo* transposition assay was set up for ISY100 transposition. This assay was used to study the effects of changing the sequences of the terminal inverted repeats of ISY100 and their flanking sequences. Mutations in proposed catalytic residues of the transposase protein were also studied.

Transposons containing two left ends transposed at a higher frequency than transposons containing two right ends, so it appears that the left end is in some way better for transposition. This situation seems to be quite common in transposons. For instance, a similar study on the *Mos1 mariner* element revealed that a transposon with two right ends

jumped more efficiently in bacteria than one with one left and one right end (Auge-Gouillou *et al.*, 2001). There are five differences between the sequences of the left and right inverted repeats of ISY100. The sequences responsible for the different efficiencies of transposition could be ascertained by making mosaic ends, combining sequences from both left and right ends as has been done for Tn5 (Zhou et al., 1998). However, this has not yet been done.

Changes in the TA dinucleotides that normally flank ISY100 were also studied. Changes in these sequences had a surprisingly large effect on the transposition frequency. Changes at position –1 had a stronger effect than changes at position –2, transitions (purine to purine and pyrimidine to pyrimidine) were tolerated better than transversions and changes at a single end had a much smaller effect than changes at both ends.

Sequencing the products of transposition from donors with alterations in the flanking TAs revealed important information about the exact boundaries of the transposon. Two strategies have been used to determine these boundaries and whether the target TA is duplicated during transposition. One method is to analyse the cleavage points in the excised transposon. For ISY100 the ends of the excised transposon produced *in vivo* have been mapped precisely by Urasaki *et al.* (2002). Their study indicated that the TA is not part of the excised transposon, and these results are consistent with my findings *in vitro* (Chapter IV). The other strategy is to change the sequence of the dinucleotides flanking the donor transposon and to ascertain whether these changes are moved with the transposon to new insertion sites. When changes were made at the –1 positions at both ends of the transposon, transposition was rare. However the one event recovered and sequenced had inserted into a TA and had the normal TA flanking sequences. When changes at -1 position were made at only one end, transposition was more efficient and a total of 9 insertions were recovered and sequenced at both ends. All of these insertions contained TAs at both ends, again consistent with duplication of the target TA (Figure 5.4). It therefore appears that the flanking TAs are target duplication and not part of the transposon, consistent with findings for all other members of the IS630/Tc1/*mariner* family of transposons.

In previous experiment on IS630 and Tc3, the flanking TAs were changed to GA or GC (Tenzen *et al.*, 1990; van Luenen *et al.*, 1994). These changes had a much smaller effect on transposition frequency than those reported here. However, all insertions were into TA targets and these TAs were duplicated, consistent with the result shown here.

Finally, the *in vivo* transposition assay was used to study the effects of mutations in the DDE motif of ISY100 transposase. These three residues coordinate one or two divalent

metal ions (Mg^{2+} or Mn^{2+}) in the crystal structures of a number of transposase proteins and retroviral integrases (Bujacz *et al.*, 1996; Davies *et al.*, 2000; Hickman *et al.*, 2000; Maignan *et al.*, 1998; Richardson *et al.*, 2006) and are essential for catalysis. Consistent with this finding, mutations in any of these three residues in ISY100 transposase reduced transposition *in vivo* to the background level of the assay. Mutations in two basic residues just downstream from the E253 had a similar effect on transposition, suggesting these residues may play a catalytic role in transposition. However, the significance of this result is less clear because these residues are not conserved in other members of the IS630/Tc1/*mariner* family.

CHAPTER IV

Transposition of ISY100 *in vitro*

4.1 Introduction

The previous chapter reports my studies of ISY100 transposition *in vivo* in *E. coli*. Although these studies gave us important information about ISY100 transposition, an *in vitro* system is required to determine the detailed mechanism of ISY100 transposition. The transposition reaction can be broken down into a number of different steps which include: binding of transposase to IR sequences, synapsis of transposon ends, cleavage of transferred and non-transferred strands, target binding and strand transfer. With an *in vitro* transposition system using purified transposase, each of these steps can be studied individually. For example, electrophoretic mobility shift and footprinting assays can be used to study the protein and DNA requirements for DNA binding and synapsis. Cleavage and transposition assays with simple oligonucleotide substrates can provide important information about the biochemical mechanism of the later steps in the reaction. In addition, if the entire transposition reaction can be made efficient, the *in vitro* system could be used for a number of practical applications. These include the generation of insertion libraries to characterise plasmid-borne genes and provide primer sites for sequencing projects.

In this chapter, the purification of a His-tagged version of ISY100 transposase is described. It is demonstrated that this protein has transposition activity *in vitro*. Various steps in the transposition mechanism, including DNA binding, cleavage of transferred and non-transferred strands, and the strand transfer reaction are then studied in more detail.

4.2 Purification of His-tagged ISY100 transposase by metal chelate affinity chromatography

To study transposition *in vitro*, the first challenge is to obtain soluble and active transposase protein. Purification removes any contaminating activities such as nucleases and proteases, which might interfere with transposition and other assays. Furthermore, purification will allow the identification of any host factors that are necessary for transposition. Many transposase proteins are poorly soluble and over-expression in bacterial expression systems often leads to the production of insoluble inclusion bodies. A number of different strategies have been used to overcome this problem including: expressing the transposase as a fusion protein, expressing it at relatively low temperatures (25° - 30°C) to reduce the expression rate and allow chaperone proteins to aid proper

protein folding, using detergents, high ionic strength and other unusual buffer conditions to increase protein solubility, and using mutant transposases with increased solubility.

The solubility of ISY100 transposase, and how it was affected by induction and cell lysis conditions was first investigated. ISY100 transposase with a C-terminal 6-His-tag has approximately the same activity as wild-type transposase *in vivo* (see chapter III). Since proteins containing a His-tag can be purified relatively simply using metal affinity chromatography, it was decided to use the His-tagged ISY100 transposase in these studies. A DNA fragment containing the His-tagged transposase gene was excised from pXF105 (Chapter III) and inserted into the bacterial expression vector pKET3a as a NdeI-BamHI fragment to make pXF102 (Figure 4.1A). This places the transposase gene under the control of the strong T7 promoter. pXF102 was transformed into the *E. coli* B strain BL21 <DE3> with or without pLysS. Upon induction with IPTG, BL21<DE3> expresses a chromosomal copy of the T7 RNA polymerase gene from the lacUV5 promoter. pLysS encodes the T7 lysozyme gene, a natural inhibitor of the T7 RNA polymerase, and therefore cuts down on the levels of gene expression, both in the absence and the presence of IPTG, which might be helpful in the case of toxic proteins. BL21 also lacks the Lon and OmpT proteases, which may help to prevent proteolytic degradation of the over-expressed transposase protein.

BL21<DE3> pXF102 and BL21 <DE3> pLysS pXF102 were grown in L-broth to an OD_{600} of 0.5, and the expression of transposase was induced by adding IPTG to 0.5 mM. Cells were grown for a further 2 hours at 37°C and cells were then harvested and analysed by SDS-PAGE. A protein of the correct size to be ISY100 transposase (33.6 kDa) was present in the induced cultures, but absent in the uninduced control cultures, with and without pLysS (Figure 4.1B). However, BL21 <DE3> pLysS gave a slightly higher level of expression. It was noticed that the strain without pLysS grew slower than that containing pLysS, even in the absence of IPTG, suggesting that there might be a low level of transposase expression in the absence of pLysS, and that this is harmful to the host cell. This might lead to selection for cells that have lost the expression plasmid, or plasmid mutants that do not express transposase, and give a lower level of expression. It was therefore decided to use BL21<DE3> pLysS pXF102 in subsequent expression experiments.

Since many transposase proteins are insoluble after over-expression in *E. coli*, finding cell lysis conditions to maximise the amount of soluble transposase protein was the

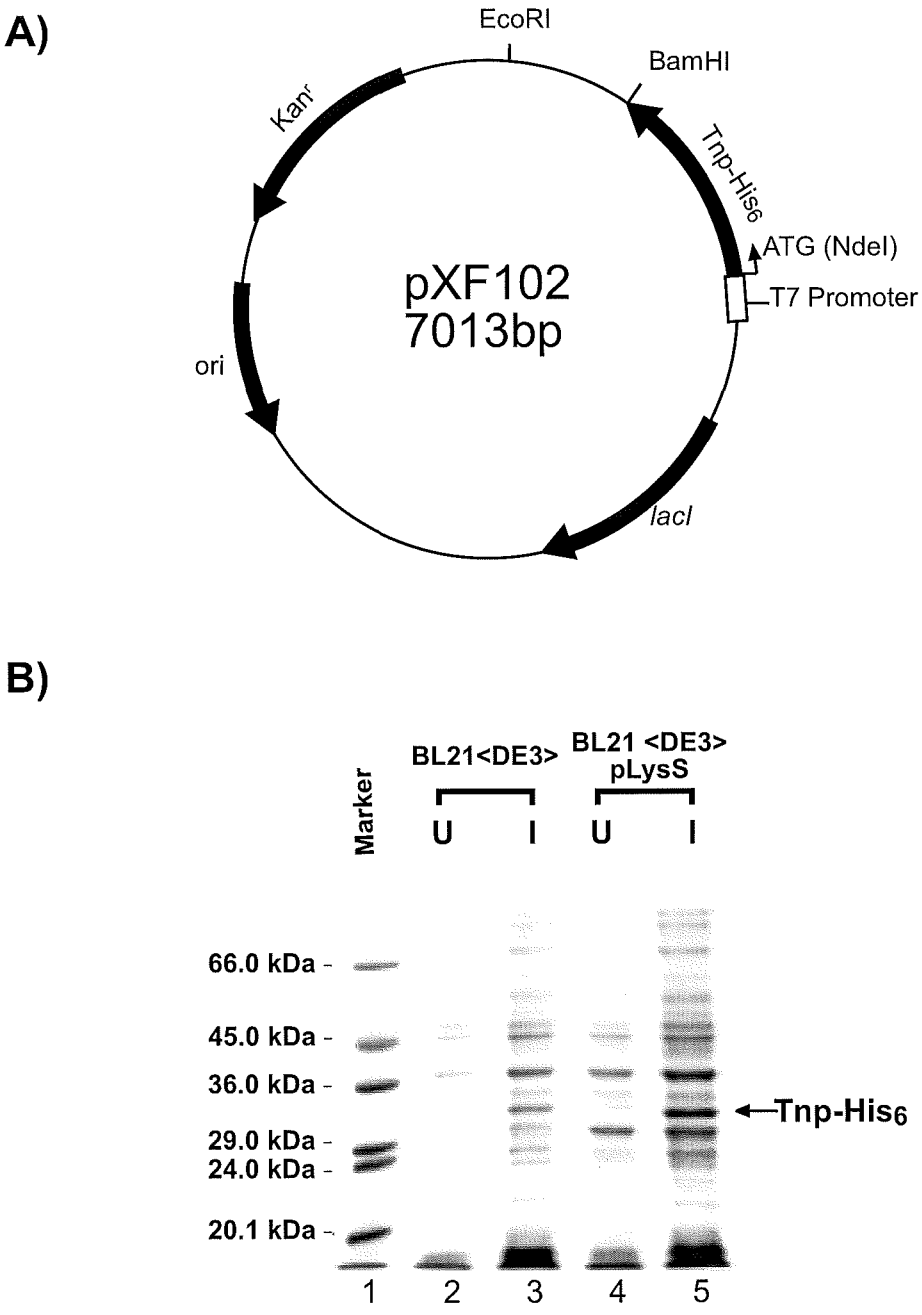


Figure 4.1 Expression of His-tagged ISY100 transposase in BL21 <DE3> and BL21 <DE3> pLysS. **A)** Map of the expression plasmid pXF102. ISY100 transposase with a C-terminal His_6 tag was expressed from the T7 promoter in pKET3a, a kanamycin resistant version of pET3a (Studier *et al.,* 1990; S.J. Rowland unpublished) **B)** His-tagged ISY100 transposase was expressed from pXF102 in BL21 <DE3> (lanes 2 and 3) or BL21 <DE3> pLysS (lanes 4 and 5). Plasmid-carrying strains were either uninduced (U) or induced (I) with 0.5 mM IPTG. Lane 1 contains a protein standard marker. Transposase is indicated by an arrow.

first priority. The solubility of transposase in various cell lysis buffers was therefore investigated. Lysis buffers containing 50 mM Tris-HCl (pH 7.6), 0.1 mM EDTA and different concentrations of NaCl were first tried. It was found that the majority of transposase was soluble at all of the NaCl concentrations tested, from 0 mM to 1 M. The addition of 0.2% triton X-100 increased the solubility of transposase, especially at high NaCl concentrations.

When the cells were lysed at 0 mM NaCl, the transposase could be co-precipitated with DNA, by the addition of 2.5 mM spermine. It appears that much of the transposase is associated with cellular DNA at low ionic strength. Transposase could then be recovered from the spermine pellet in a buffer containing 1 M NaCl and 0.2% triton X-100. The His-tagged protein could then be purified further using metal chelate affinity chromatography on a Ni-NTA column. Transposase bound well to the column in this buffer, and was eluted by the addition of 200 mM imidazole. One slight disadvantage of using triton X-100 is that its absorbance peaks at 254 nm and 278 nm made it difficult to monitor the protein purification by ultraviolet absorption at 260 nm and 280 nm.

The steps outlined above were used to develop a non-denaturing purification of ISY100 transposase as outlined in Figure 4.2 and described in detail in chapter II. The purification was followed by coomassie-stained SDS-PAGE and the results of a typical purification are shown in Figure 4.3. Comparison of lanes 3 and 4 shows that the majority of the transposase was soluble in the lysis buffer, although some was insoluble. Precipitation with spermine is a very good step and removes many of the contaminating cellular proteins (Figure 4.3; lanes 5-6). Recovery in buffer A (containing 1 M NaCl and 0.2% triton X-100) was also very efficient (Figure 4.3; lanes 7-8). Comparison of lanes 8 and 9 shows that more than 50% of the transposase bound to the Ni-NTA column while most of the contaminating proteins flowed through. It is possible that the His-tag is not available for binding to the column on a proportion of the protein, perhaps because the protein is misfolded or maybe because the His-tag is bound by some other factors. Low concentrations of imidazole (20 – 50 mM) were used to elute proteins which were not tightly bound to the column (Figure 4.3; lanes 11, 12) and transposase eluted from the column substantially pure at an imidazole concentration of 200 mM (Figure 4.3; lane 13-16). Imidazole was removed by dialysis, an equal volume of glycerol was added to the dialysed transposase, and it was stored at -20°C for future use. Typical transposase

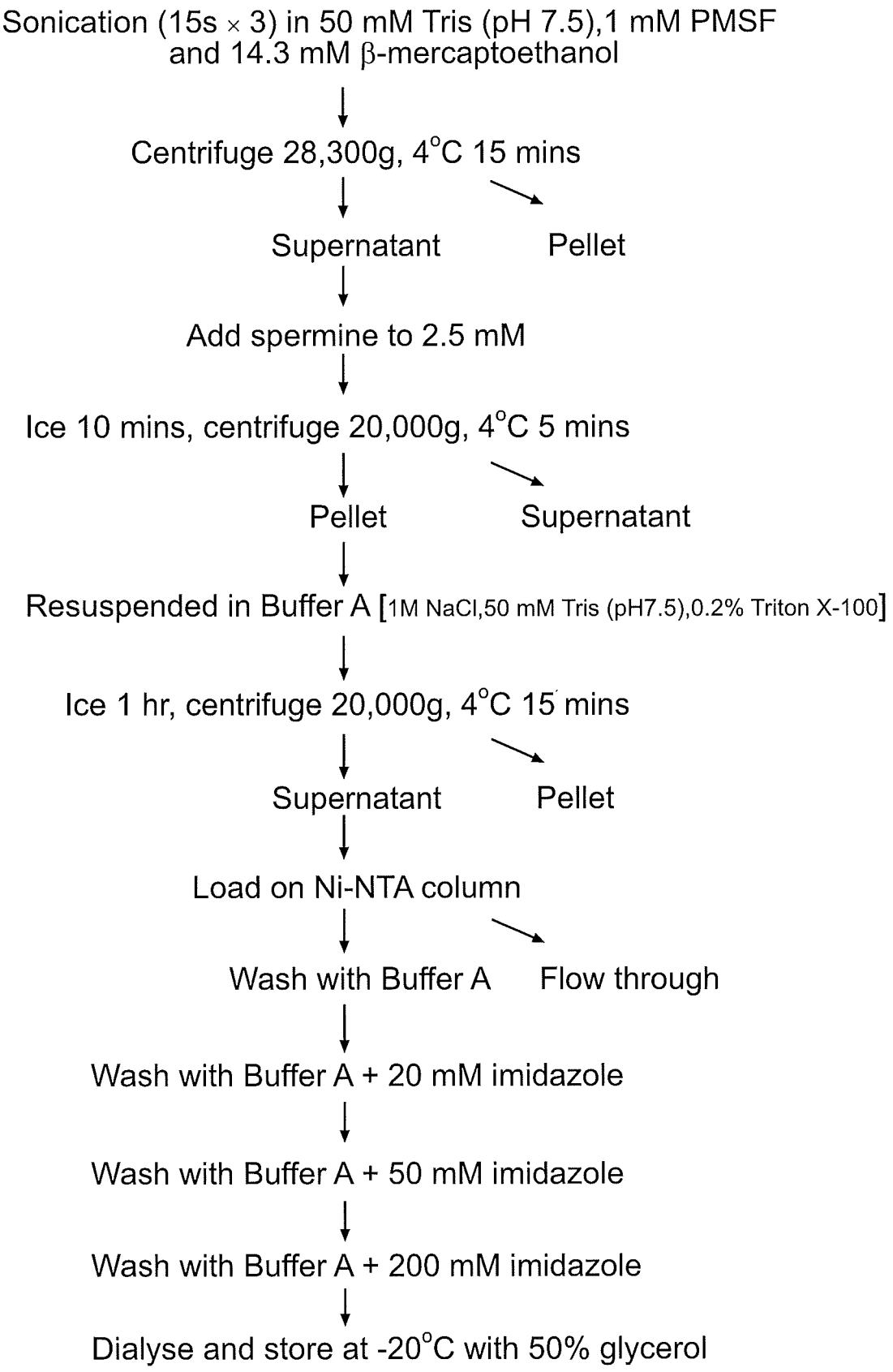


Figure 4.2 Flowchart for the purification of ISY100 transposase.

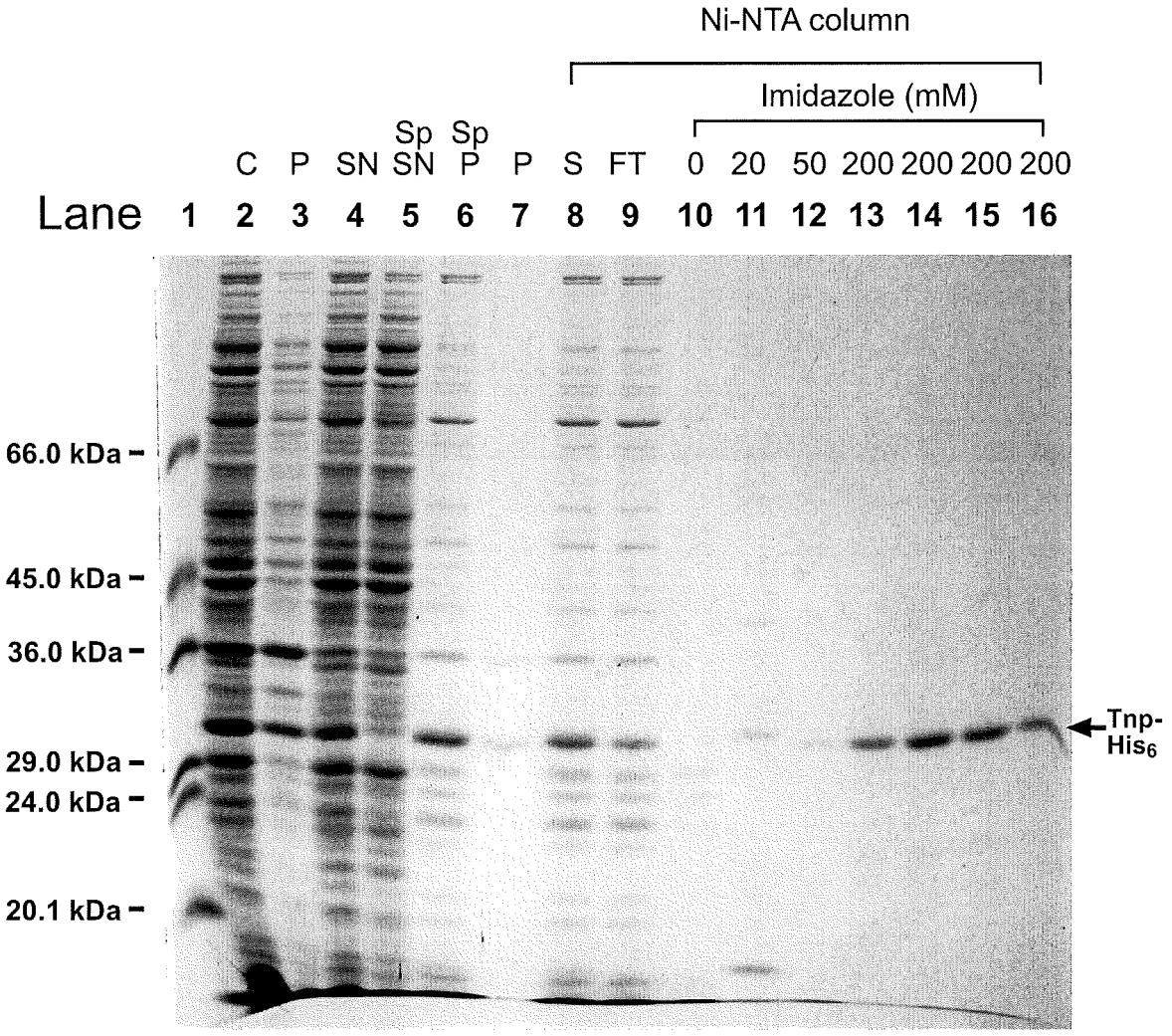


Figure 4.3 SDS-polyacrylamide gel electrophoresis of fractions obtained during purification of ISY100 transposase by Ni-NTA column chromatography. Approximately equivalent proportions of the total material were loaded at each step.

Lane 1: Protein standard marker
Lane 2: IPTG induced BL21 <DE3> pLysS pXF102 whole cells (C)
Lane 3: Sonication pellet (P)
Lane 4: Sonication supernatant (SN)
Lane 5: Spermine supernatant (Sp SN)
Lane 6: Spermine pellet (Sp P)
Lane 7: 1 M NaCl/triton X-100 pellet (P)
Lane 8: 1 M NaCl/triton X-100 supernatant (load) (S)
Lane 9: Flow through (FT)
Lane 10: Buffer A wash
Lane 11: Buffer A plus 20 mM imidazole wash fraction #5
Lane 12: Buffer A plus 50 mM imidazole wash fraction #5
Lane 13-16: Buffer A plus 200 mM imidazole elution fractions 3 - 6

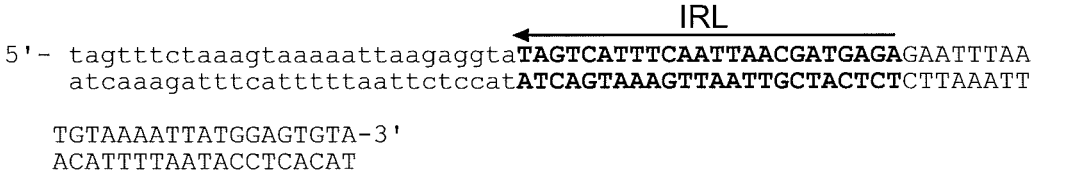
concentrations were 0.09 - 0.17 mg/ml. Transposase stored at –20° remained stable and active for at least one year in these conditions.

In other experiments, the transposase was purified under denaturing condition. Cells expressing transposase were broken in a buffer containing 6 M urea and centrifuged at 28,300 g to remove insoluble material. The supernatant was then directly applied onto a Ni-NTA column. The column was washed with low concentrations of imidazole in a buffer containing 6 M urea, and transposase was then eluted in the same buffer containing 200 mM imidazole. The purified transposase was then renatured by dialysis against a buffer containing no urea (see chapter II for details). The transposase purified in this way showed the same activity *in vitro* as that purified under native condition (data not shown).

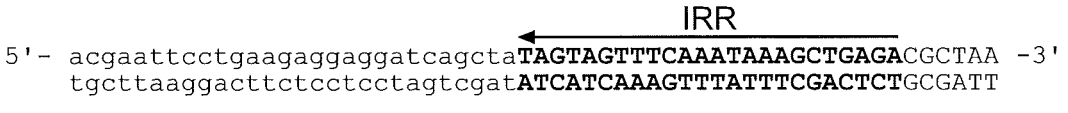
4.3 Purified transposase binds to transposon ends

The simplest assay for transposase activity is for binding to transposon end sequences. Electrophoretic mobility shift assays (EMSA) were therefore used to investigate binding of fractions from the transposase purification to a DNA fragment IRL 29, which contains 29 bp from the left end of ISY100 (Figure 4.4). The induced BL21<DE3> plysS pXF102 cell extract bound to this DNA fragment to give one major retarded complex (Figure 4.5, lane 2, complex I). This binding appears to be dependent on expression of ISY100 transposase since an extract from cells containing just the empty expression vector did not bind to a 79 bp labelled DNA fragment IRL79 (Figure 4.10B, lane 2), containing 51 bp from the left end of ISY100 (Figure 4.4). The DNA binding activity was still present in the spermine precipitated material (Figure 4.5, lane 3), though at least two other slower migrating complexes (II and III) were also visible. The majority of the DNA-binding activity producing complexes I and III did not bind to the nickel column, coming out in the flow through and the buffer A wash (Figure 4.5, lanes 4 and 5). In contrast, the purified full-length protein eluting at 200 mM imidazole specifically produced complex II. One possible explanation of this is that proteolytic fragments of transposase with DNA-binding activity were present in the cell extract. Since transposase has a C-terminal His-tag, N-terminal fragments will flow through the nickel column, whereas full-length transposase, and any C-terminal fragments, will bind to the column and elute at 200 mM imidazole. Consistent with this idea, N-terminal fragments of transposase bound to transposon end sequences producing complexes with similar mobility to complex I in Figure 4.5 (see section 4.3.2).

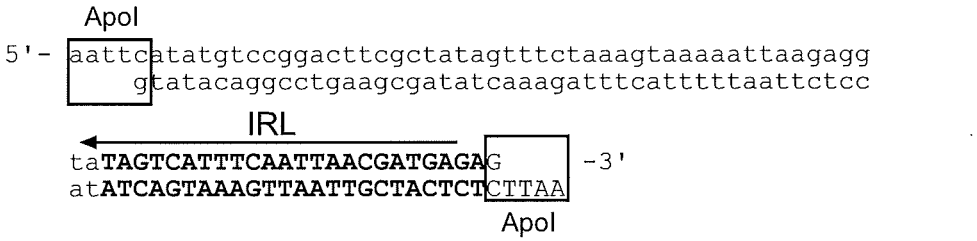
IRL79:



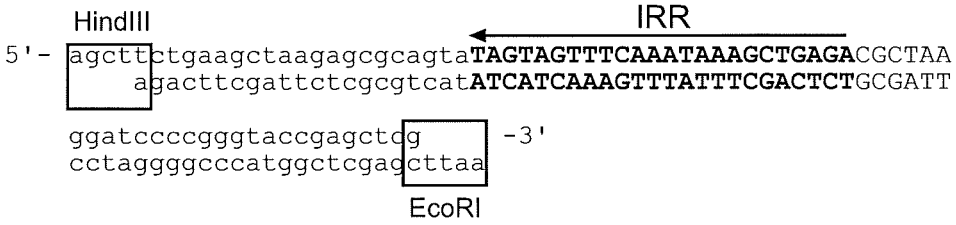
IRR58:



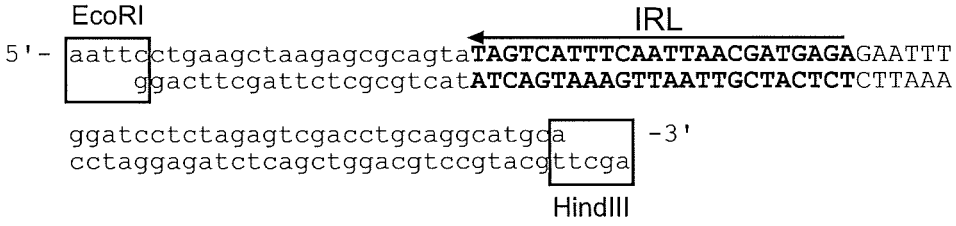
IRL29 (ApoI fragment from pXF110):



IRR30 (EcoRI-HindIII fragment of pXF149):



IRL30 (EcoRI-HindIII fragment of pXF147):



IRR49 (HindIII-EcoRI or HindIII-BamHI fragment of pXF121)

Figure 4.4 Sequences of fragments containing ISY100 end sequences used in binding and footprinting reactions. Transposon sequences are shown in uppercase, the 24 bp inverted repeats are shown in bold. IRL79 and IRR58 were made by annealing single stranded oligonucleotides. The other fragments were obtained from plasmids by restriction digestion as indicated.

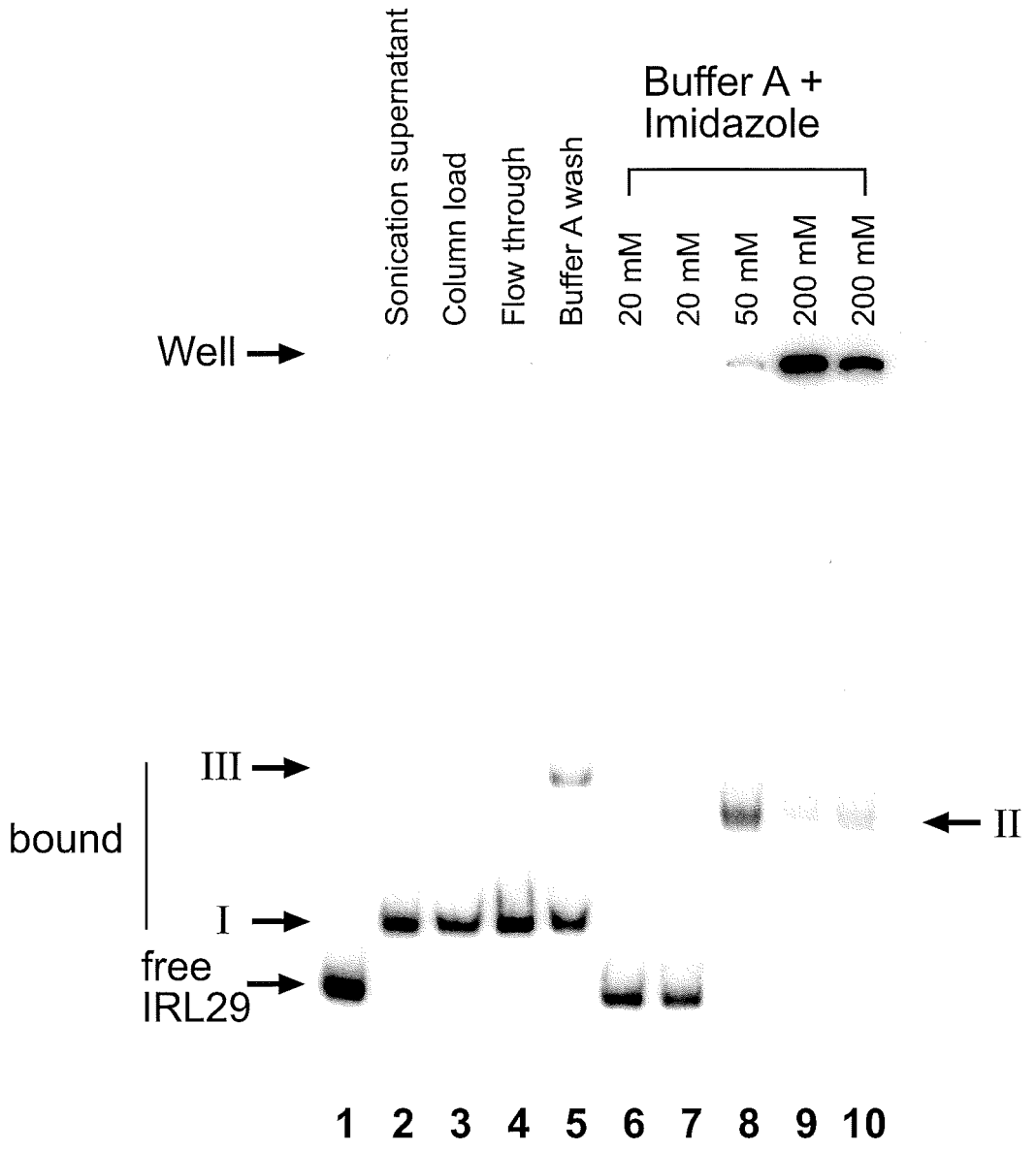


Figure 4.5 DNA-binding activities of fractions from the purification. Indicated fractions from the purification of ISY100 transposase (Figure 4.3) were incubated with 3' end-labelled IRL29, in a standard binding reaction for 5 minutes at 30°C. Bound and unbound DNA were separated on a 6% native polyacrylamide gel.

Lane 1: Protein dilution buffer
Lane 2: BL21 <DE3> pLysS pXF102 sonication supernatant
Lane 3: Spermine pellet resuspended in 1M NaCl and 0.2% triton X-100 (column load)
Lane 4: Ni-NTA flow through
Lane 5: Buffer A wash
Lane 6: Buffer A plus 20 mM imidazole wash fraction #6
Lane 7: Buffer A plus 20 mM imidazole wash fraction #7
Lane 8: Buffer A plus 50 mM imidazole wash fraction #5
Lane 9: Buffer A plus 200 mM imidazole elution fraction #3 - #6 mixture
Lane 10: Buffer A plus 200 mM imidazole elution fraction #7 - #8 mixture

Further binding assays were carried out with the purified full-length transposase. To measure the affinity of binding, transposase dilutions were incubated with a DNA fragment IRL30 or IRR30 (Figure 4.6), containing 30 bp from either the left or the right end of ISY100 (Figure 4.4). Transposase bound to both DNA fragments with similar affinity, giving 50% binding somewhere between the 2^{-0} and the 2^{-1} dilutions, corresponding to an apparent K_d of approximately 130 nM. Both left and right end fragments gave single retarded complexes with similar mobility to each other (relative mobility Rf = 0.70 and 0.71 respectively) and to complex II in Figure 4.6 (Rf = 0.71). The fact that 30 bp of ISY100 end sequences were sufficient for DNA binding is consistent with the fact that these same end sequences were functional for transposition *in vivo* (see chapter III).

4.3.1 Analysis of transposase binding by DNase I footprinting

The experiments above suggest that purified transposase can bind to ISY100 ends. Binding reactions were carried out in the presence of a large excess of poly dIdC so this binding appears to be specific. To find out what regions of ISY100 ends are bound by transposase, DNase I footprinting experiments were carried out. DNase I footprinting is one of the simplest methods for mapping protein binding sites on DNA (Galas and Schmitz, 1978), and has been used to study transposase-DNA interactions for many different transposons for example see Arciszewska et al., (1991) and Mizuuchi et al., (1991). DNase I cuts double-stranded DNA relatively non-specifically, leaving 5′-phosphate and 3′-hydroxyl ends. Where a protein is bound to the DNA, it prevents access by DNase I, producing a "footprint" of uncleaved positions in the DNA molecule.

DNA restriction fragments containing transposon right end sequences were 3' end labelled at just one end on either the top or the bottom strand. Work on other transposons has shown that the 3'-OH at the transposon end is transferred to the target DNA (Bainton *et al.*, 1991; Bhasin *et al.*, 1999; Mizuuchi, 1984). This is later shown also to be the case for ISY100 (see section 4.7). The bottom and top strands as shown in Figure 4.4 are hereafter referred to throughout this chapter as transferred and non-transferred strands according to this result. Labelled DNA fragments were incubated in binding buffer with or without transposase, and then subjected to mild cleavage by DNase I, to cleave the DNA on average once per molecule. The cleavage products were then separated on a denaturing polyacrylamide gel adjacent to a Maxam and Gilbert G plus A sequencing ladder (Figure 4.7A) (Maxam and Gilbert, 1980). Comparison of the ladders produced in the presence and absence of transposase reveals the sequences that were protected by transposase, and the

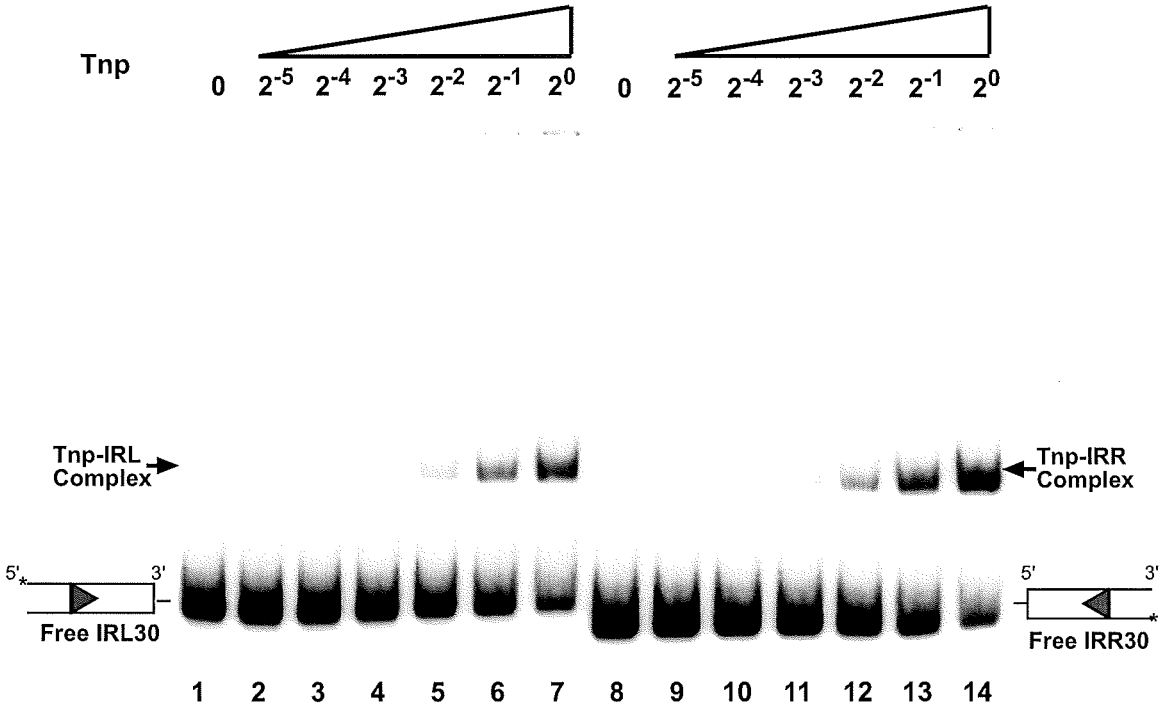
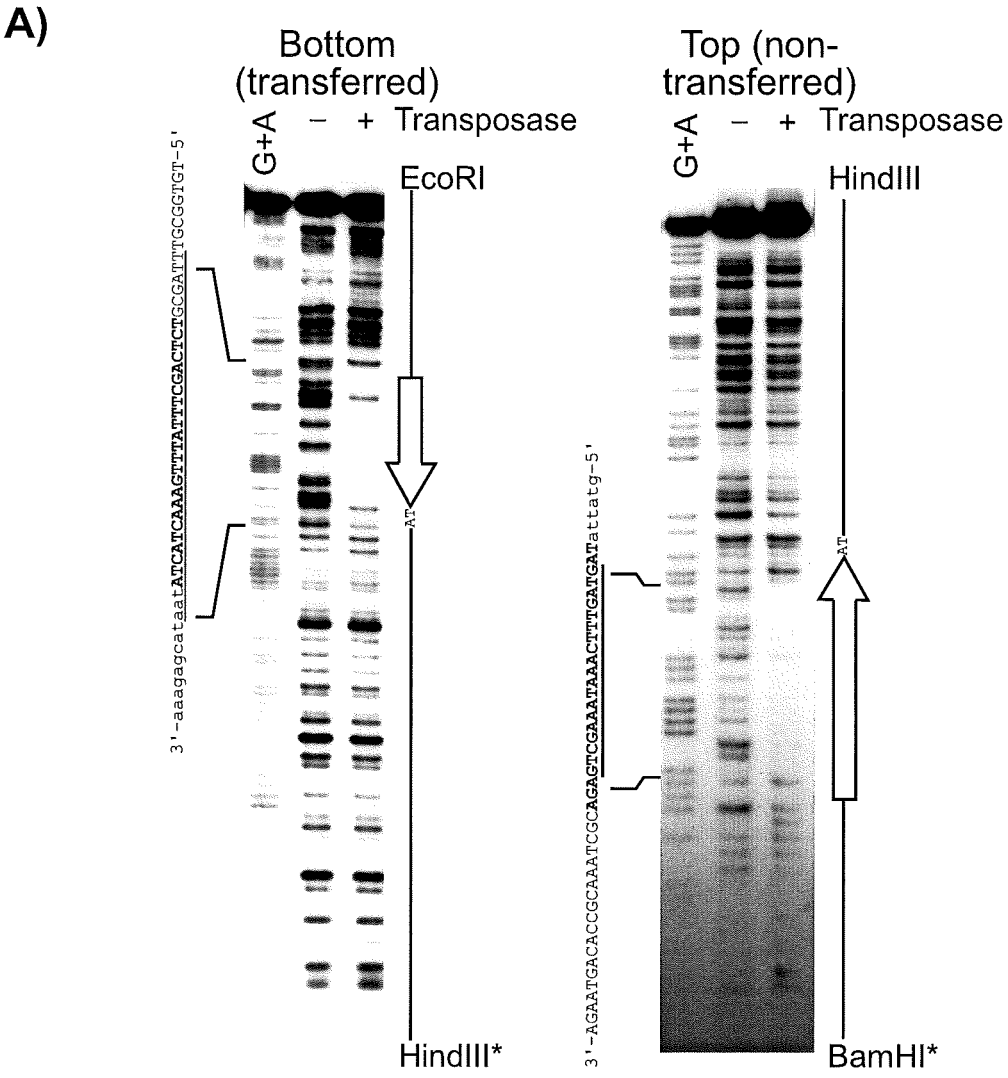


Figure 4.6 Gel retardation of IRL and IRR DNA fragments using purified His-tagged ISY100 transposase. DNA binding assays were set up using purified ISY100 transposase and 5' end-labelled IRL30 (lanes 1-7) or IRR30 (lanes 8-14). The binding reaction was carried out for 10 minutes at 30°C and complexes were separated by electrophoresis on a 6% native polyacrylamide gel. Positions of radiolabelling are indicated by asterisks. Arrows indicate nucleoprotein complexes. The relative protein concentration in each assay is given above the corresponding lane. Lane 1 and lane 8 are negative controls containing protein dilution buffer. The transposase concentration in the undiluted lane (2^0) is approximately 130 nM.



B)

IRR49 pXF121 HindIII*-EcoRI and HindIII-BamHI*

HindIII

5-' agcttgcatgcctgcaggtcgactctagatgtcctcatccgtataatgct
tcgaacgtacggacgtccagctgagatctacaggagtaggcatattacga

IRR

ttctcgtattaTAGTAGTTTCAAATAAAGCTGAGACGCTAAACGCCACAG
aagagcataatATCATCAAAGTTTATTTCGACTCTGCGATTTGCGGTGTC

TAAGAACGGAggatccccgggtaccgagctcgaatt -3'
ATTCTTGCCTcctaggggcccatggctcgagcttaa

BamHI EcoRI

— Top (non-transferred) strand
— Bottom (transferred) strand

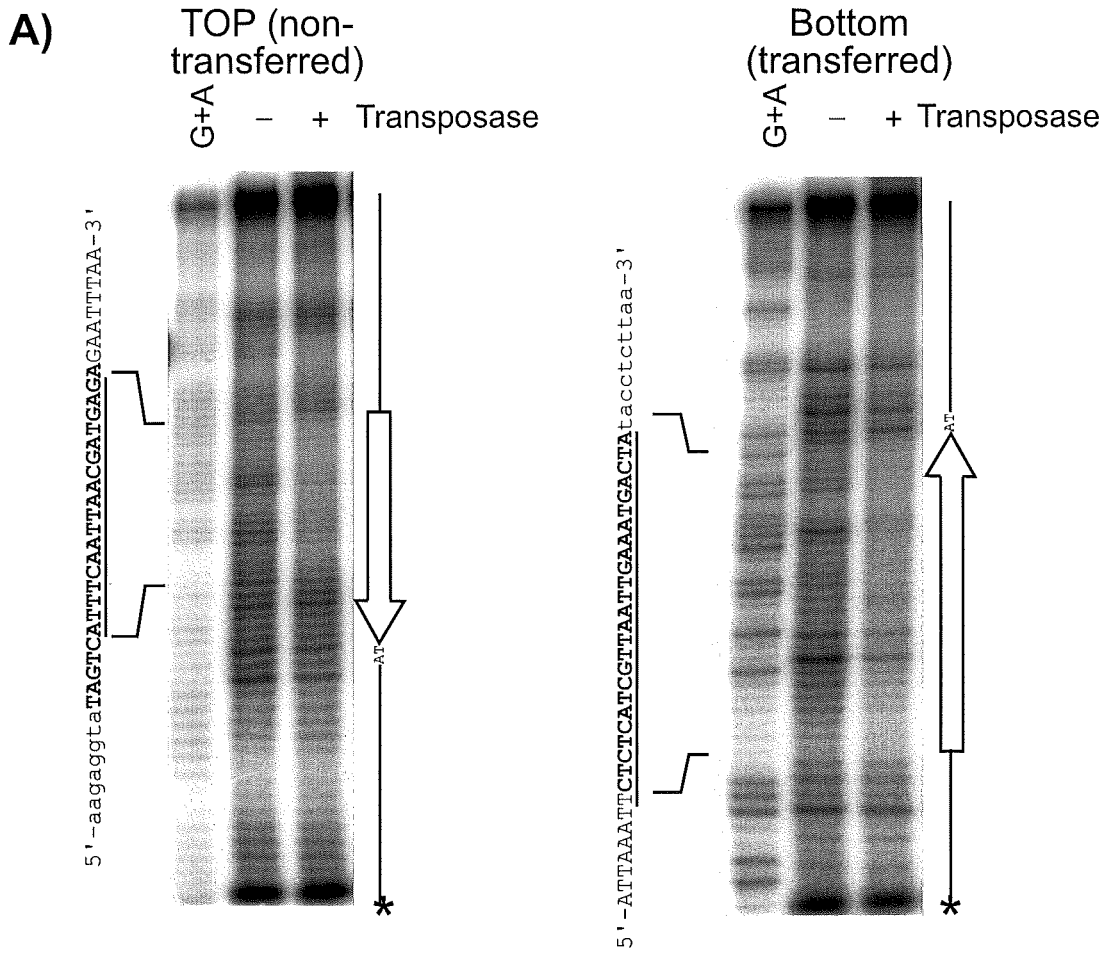
Figure 4.7 DNase I footprinting of ISY100 transposase IRR complexes. **A)** HindIII- or BamHI-cleaved pXF121 were 3' end-labelled by Klenow and then cut by EcoRI and HindIII respectively. The end-labelled fragments containing 49 bp of IRR were gel purified. Binding reactions were set up with purified transposase and incubated at 30°C for 10 minutes. The mixtures were then treated with DNase I as described in chapter II. Reactions were separated on a 8% denaturing polyacrylamide gel together with a G+A ladder. **B)** DNA sequence of the fragments used and summary of protection. Transposon end sequences are shown in capitals and flanking sequence in lowercase, the 24 bp of IRR is in bold. The protected sequence is underlined.

results are summarised in Figure 4.7B. To footprint the left end, top and bottom strands of IRL79 oligonucleotides (Figure 4.4) containing 51 bp from ISY100 IRL were 5' end-labelled with kinase and annealed to unlabelled complementary strands. These labelled oligonucleotides were footprinted in a similar way and the results are shown in Figure 4.8A. The protected regions are quite similar at both ends with majority of the 24 bp inverted repeat sequences protected on both strands. At both IRR and IRL, protection extended further on the bottom (transferred) strand than on the top (non-transferred) strand, extending inward beyond the start of the inverted repeat, and outward to cover the transposon end and, in the case of IRR, the flanking target TA.

4.3.2 ISY100 transposase contains an N-terminal DNA binding motif

Many different DDE transposases have been shown to contain N-terminal DNA binding-domains that specifically recognise the terminal inverted repeats of their own transposon. There is little sequence conservation between the terminal inverted repeats of different transposons, and it appears that transposases co-evolve with their terminal inverted repeats, so as to act on their own ends, but not those of other elements. The DNA binding domains of many DDE transposases, including all members of the IS630/Tc1/*mariner* family, appear to contain a pair of helix-turn-helix (HTH) DNA-binding motifs within their N-terminal domains. The structure of one such DNA-binding domain, that of the *C. elegans* Tc3 transposase, has been determined by X-ray crystallography in complex with DNA (Watkins *et al* 2004). The N-terminal domain of Tc3 transposase folds into two independent helix-turn-helix domains separated by a linker that interacts with the minor groove of DNA. This linker contains an 'AT hook'-like motif, which is also present in other bipartite DNA-binding domains belonging to the paired domain family (Xu *et al.*, 1999; Xu *et al.*, 1995). Each helix-turn-helix motif consists of three α-helices, with the third helix inserted into the DNA major groove in each case. In Tc3 transposase, the N-terminal helix-turn-helix binds to more transposon internal sequences, while the second helix-turn-helix binds to sequences closer to the transposon end.

To see if ISY100 transposase contains any potential helix-turn-helix motifs, the entire protein sequence of the ISY100 transposase was analysed using the program *helixturnhelix* from the EMBOSS package (Dodd and Egan, 1987, 1990; Rice *et al.*, 2000). Two potential HTH structures were predicted by this program with high probability (Figure 4.9A). Both predicted HTH motifs were located N-terminally, one between residues 19-40 and the other between residues 71-92. To validate this method, the same program was also used on Tc3



B)

IRL79:

IRL

5'-*tagtttctaaagtaaaaattaagaggtaTAGTCATTTCAATTAACGATGAGAGAATTTAA
atcaaagatttcatttttaattctccatATCAGTAAAGTTAATTGCTACTCTCTTAAATT

TGTAAAATTATGGAGTGTA-3' — Top (non-transferred) strand
ACATTTTAATACCTCACAT* — Bottom (transferred) strand

Figure 4.8 DNase I footprinting of ISY100 transposase IRL complex. **A)** double-stranded oligonucleotide IRL79 containing 51 bp of IRL was 5' end-labelled by kinase on either the top or bottom strand. Binding reactions were set up with purified transposase and incubated at 30°C for 10 minutes. The mixtures were then treated with DNase I as described in chapter II. Reactions were separated on an 8% denaturing polyacrylamide gel together with a G+A ladder. **B)** DNA sequence of IRL79 and summary of protection. Transposon end sequences are shown in capitals and flanking sequence in lowercase, the 24 bp of IRL is in bold. The protected sequence is underlined.

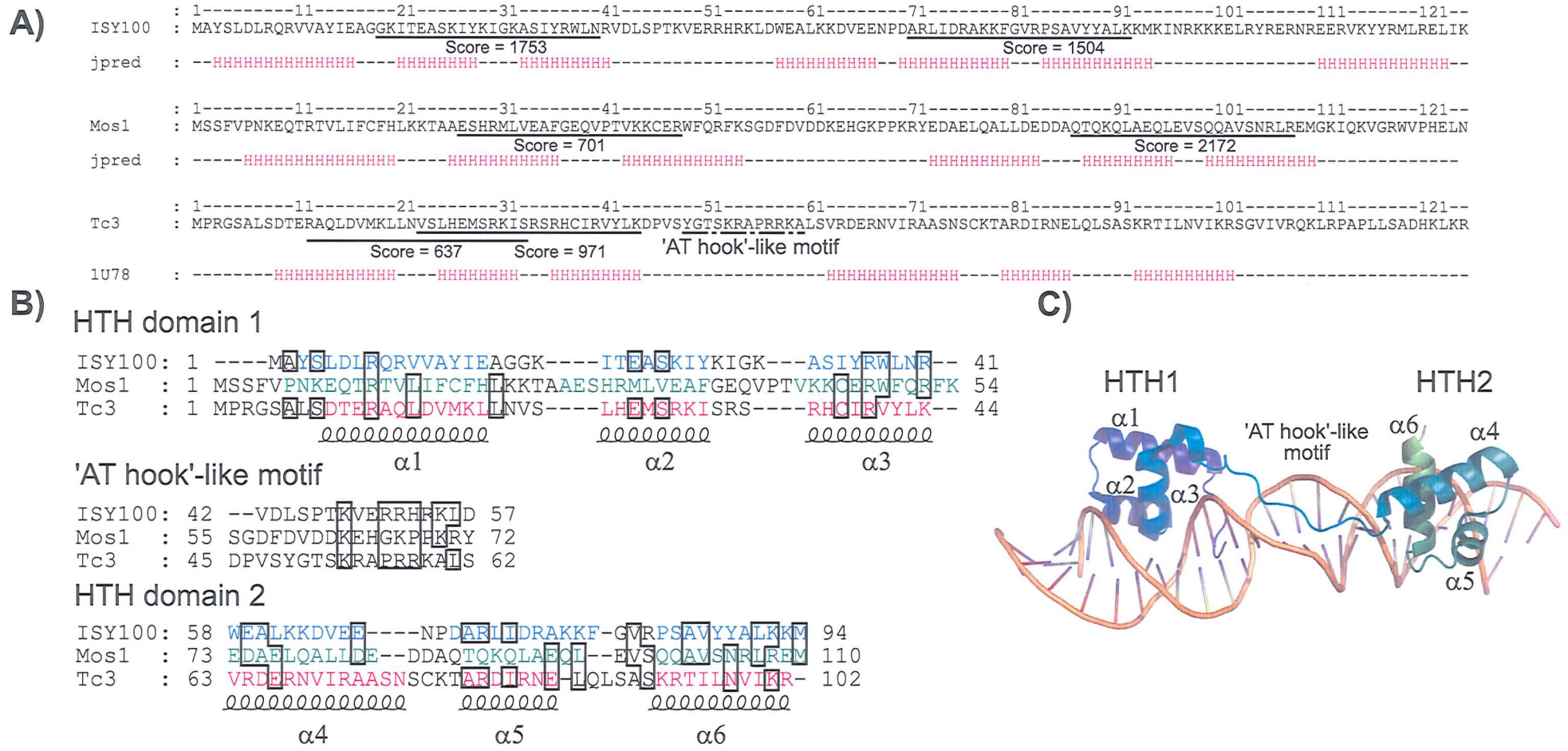
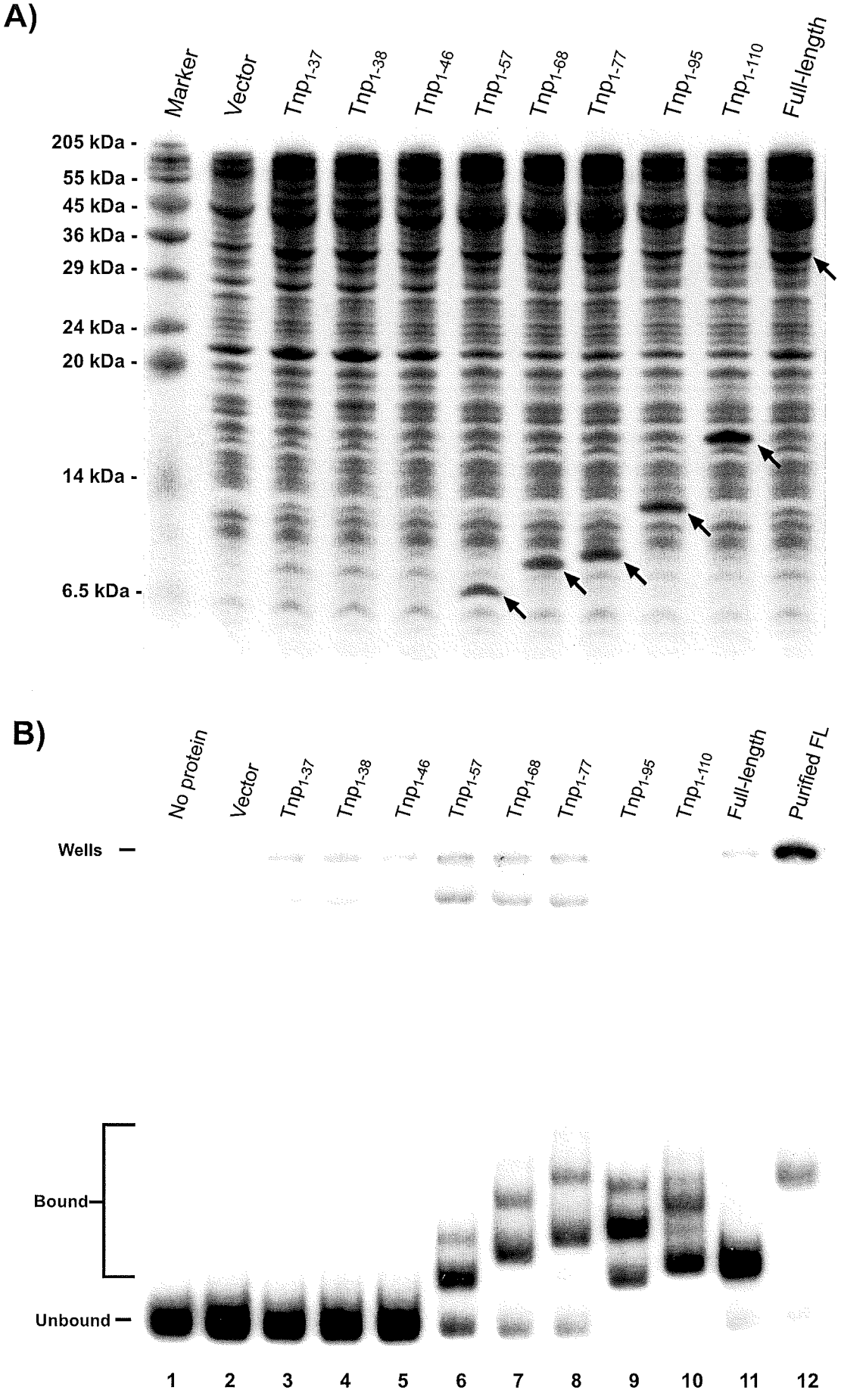


Figure 4.9 The N-terminal sequence of ISY100 transposase has two potential HTH structures. A) The sequences of the N-terminal 123 amino acids of ISY100 (BAA16620), Mos1 (AAA28678) and Tc3 (BAA16620) transposases are shown with predicted helix turn helix motifs (Dodd and Egan, 1990) shown underlined. The secondary structures, predicted by jpred (Cuff and Barton, 2000) are shown below each sequence (H = helix, E = sheet). **B)** Secondary structure guided sequence alignments of ISY100, Mos1 and Tc3 transposase HTH1, HTH2 and 'AT hook'-like domains. **C)** The crystal structure of the Tc3 DNA-binding domain bound to DNA (PDB accession code 1U78).

and *Mos1* transposases, locating two potential HTH motifs in each case. However, the two predicted HTH motifs of Tc3 transposase overlap. The highest scoring motif corresponds to α-helix 2 and 3 of the first HTH in the crystal structure, while the other corresponds to α-helix 1 and 2 of the same motif (Figure 4.9A). The second Tc3 HTH known from the X-ray structure was not identified by this program. Finally JPRED was used to predict the secondary structure of ISY100 transposase (Figure 4.9A). This program uses PSI-blast to make a multiple alignment of related proteins and uses this alignment to predict the secondary structure of the query protein with quite high accuracy (Cuff and Barton, 2000). JPRED predicted six α-helices in the N-terminal 95 amino acids of ISY100 transposase, consistent with the presence of two three-helix HTH motifs (residues 3-41 and 58-92) in this domain. A similar prediction was made for *Mos1*. The secondary structure prediction and the *helixturnhelix* results were used to guide the alignment of ISY100 transposase with the N-terminal domains of Tc3 and *Mos1* transposases (Figure 4.9B). The results show that the N-terminal domain of ISY100 transposase is likely to have a similar structure to Tc3 transposase, with two three-helix HTH motifs separated by a linker containing the basic sequence RRHRK (residues 51-55) which may bind in the DNA minor groove like the 'AT hook'-like motif found in Tc3 transposase.

To verify that these predicted HTH structures bind specifically to ISY100 ends, the DNA-binding activity of a set of truncated transposase derivatives, containing zero, one or two predicted HTH motifs, was investigated. The 5' end of the ISY100 transposase gene was amplified using a forward primer that introduced an NdeI site at the transposase start codon, and reverse primers that introduced C-terminal His_6 tags at various positions within the first 110 codons of the transposase gene. These PCR fragments were inserted into the expression plasmid pKET3a under the control of the T7 promoter. In this way, plasmids encoding the N-terminal 37, 38, 46, 57, 68, 77, 95 and 110 amino acids of ISY100 transposase (Tnp_{1-37}, Tnp_{1-38}, Tnp_{1-46}, Tnp_{1-57}, Tnp_{1-68}, Tnp_{1-77}, Tnp_{1-95} and Tnp_{1-110}) were constructed (Figure 4.10C). These plasmids were introduced into the strain BL21 <DE3> pLysS, and N-terminal transposase derivatives were induced by the addition of IPTG. pXF102, expressing C-terminal His-tagged full length transposase, and pKET3a, the empty expression vector, were used as controls. Extracts were produced from the induced cells by sonication followed by centrifugation at 28,300 g. These cell extracts were then separated by electrophoresis on a tricine-SDS gel, which can resolve proteins of much lower molecular weight than standard Tris-glycine gels, and stained with coomassie blue.


A)
Marker
Vector
Tnp1-37
Tnp1-38
Tnp1-46
Tnp1-57
Tnp1-68
Tnp1-77
Tnp1-95
Tnp1-110
Full-length
205 kDa -
55 kDa -
45 kDa -
36 kDa -
29 kDa -
24 kDa -
20 kDa -
14 kDa -
6.5 kDa -
B)
No protein
Vector
Tnp1-37
Tnp1-38
Tnp1-46
Tnp1-57
Tnp1-68
Tnp1-77
Tnp1-95
Tnp1-110
Full-length
Purified FL
Wells
Bound
Unbound
1
2
3
4
5
6
7
8
9
10
11
12

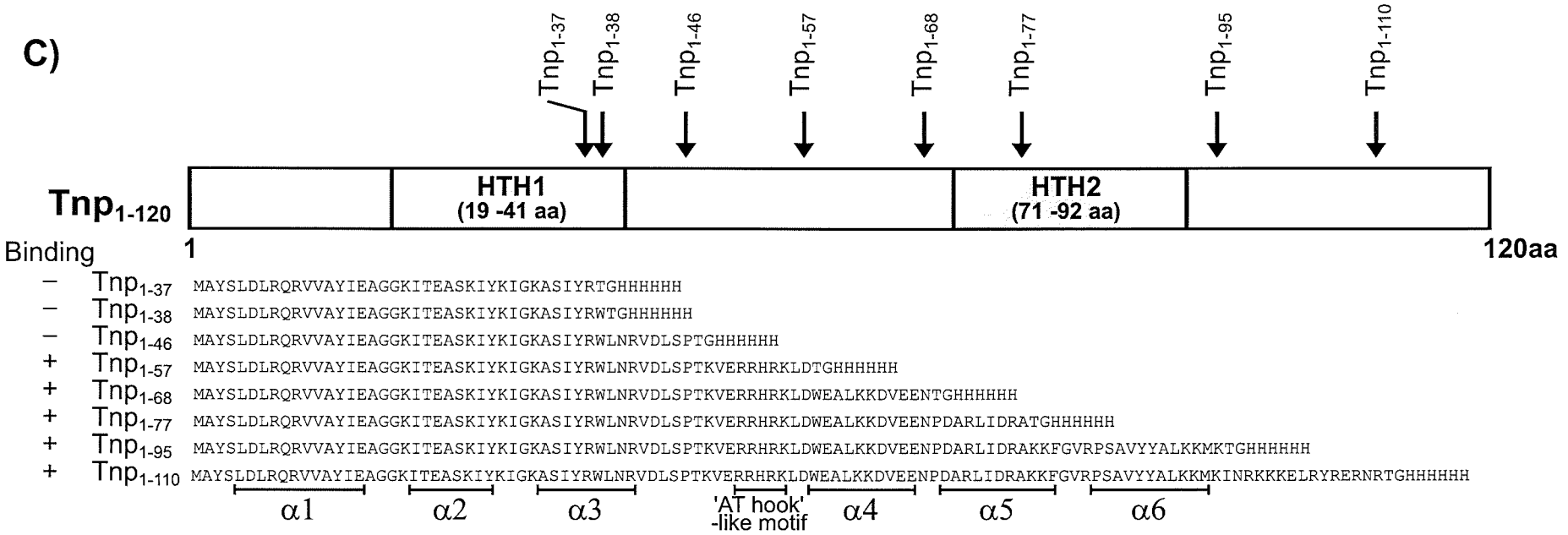


Figure 4.10 N-terminal derivatives of ISY100 transposase. A) Tricine-SDS gel electrophoresis of whole cell extracts from strains expressing truncated transposase derivatives. DNA fragments encoding N-terminal fragments of ISY100 transposase with C-terminal 6 × His tags were inserted into pKET3a and expressed in BL21 <DE3> pLysS. Cells were grown at 37°C to an A_{600} of 0.5 and induced for 2 hours with 0.5 mM IPTG. Cells were sonicated in 50 mM Tris-Cl (pH 7.5) and 10 mM β-mercaptoethanol and centrifuged at 19,000 rpm (JA20) for 15 minutes at 4°C, and supernatants were analysed on a 15% Tricine-SDS polyacrylamide gel. Cell extracts from induced BL21 <DE3> pLysS containing pKET3a only (lane 2) or pXF102 carrying full-length transposase (lane 11) were run on the same gel. Induced transposase fragments are indicated with arrows. **B)** The cell extracts shown in A were used in a gel shift assay with labelled IRL79. Purified full length Tnp-His_6 was used as a control (lane 12). **C)** Structure of N- terminal domain of ISY100 transposase, showing the positions of the predicted HTH motifs and sequences of the truncated transposases.

Induced proteins of the correct size were observed for Tnp_{1-57}, Tnp_{1-68}, Tnp_{1-77}, Tnp_{1-95} and Tnp_{1-110} and the full length transposase (Figure 4.10A). However, no induced bands were visible for Tnp_{1-37}, Tnp_{1-38}, Tnp_{1-46}. It may be that these proteins are too small to be well separated on this gel system. Alternatively, these smaller proteins may not be efficiently expressed, or they may be missing from the cleared lysates because they are insoluble or degraded for some reason.

Cell extracts containing transposase N-terminal derivatives were used in a gel shift DNA binding assay (Figure 4.10B) with IRL79, a 79 bp DNA fragment containing 51 bp from the left end of ISY100 (Figure 4.4). The vector alone extract gave two faint retarded bands that were presumably due to binding by cellular proteins. Tnp_{1-37}, Tnp_{1-38}, Tnp_{1-46} gave no additional retarded bands, either because these derivatives do not bind the transposon end, or because no transposase derived protein is present in these extracts. At the protein concentration used, the extract containing Tnp_{1-57} gave one major retarded band, and a slightly less prominent slower migrating band. Tnp_{1-68} and Tnp_{1-77} both gave a similar pattern of bands, but as the size of the transposase derivative was increased, the complexes migrated more slowly. Tnp_{1-95} and Tnp_{1-110} also bound to the ISY100 left end fragment, but the fastest migrating complex in both cases ran quicker than the fastest migrating complex produced by Tnp_{1-77}. This faster migration could be due to a different conformation of the complexes formed by the proteins containing two HTH motifs, or could be caused by the presence of shorter protein fragments in the cell extract.

From these results, it appears that the N-terminal HTH motif and an extra 16 amino acids containing the 'AT hook'-like motif (Tnp_{1-57}), is sufficient to bind to an ISY100 end. It is not clear whether the single HTH motif without the 'AT hook'-like motif found in Tnp_{1-46}, is unable to bind to the transposon end, or whether this protein is not present in the extract. The fact that Tnp_{1-95} and Tnp_{1-110} leave no unbound DNA (Figure 4.10B, lanes 9 and 10) suggest that they might bind more tightly, consistent with the presence of two HTH-motifs in these proteins. Confirmation of this result awaits more quantitative experiments with purified proteins.

In the Tc3 crystal structure, the first HTH motif binds to the inner end of the IR, while the second binds the IR closer to the transposon end. To see whether the same is true for ISY100 transposase, footprinting experiments were carried out with cell extracts containing N-terminal derivatives of ISY100 transposase. It was thought that Tnp_{1-57}, Tnp_{1-68} and Tnp_{1-77}, containing just one HTH, might protect only the inner parts of the IR, while Tnp_{1-95} and Tnp_{1-110} might protect all or most of the IR. DNase I footprinting assays were

carried out with IRL79 labelled at the 5' end on either the top or bottom strand. Tnp_{1-95} and Tnp_{1-110} protected almost all of the IR sequence, consistent with the presence of two HTH motifs binding to two different regions of the IR. Somewhat surprisingly, this protection was indistinguishable from the protection by the full length transposase, suggesting that the catalytic domain does not interact with the DNA in the footprinting reactions with the full length protein. The three truncated transposases with only a single HTH failed to protect IRL in this assay (Figure 4.11). Tnp_{1-57} and Tnp_{1-77} were purified on a Ni-NTA column and tested, but neither of them gave any detectable protection in the footprinting assay although they did produce retarded bands on a gel shift assay (data not shown). The lack of a footprint might reflect the reduced specificity or reduced affinity of binding of transposase derivatives containing only one HTH motif.

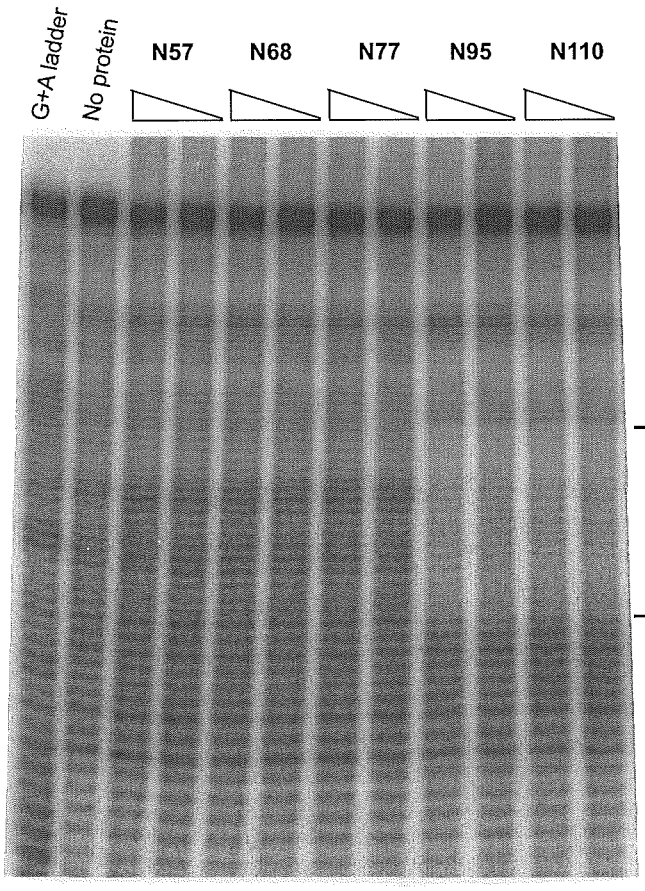
4.4 Attempts to observe paired end complexes

After DNA binding, the transposase must bring two transposon ends together to form a synaptic complex, also known as a paired end complex (PEC). For many transposons, including Mu, Tn5, Tn10 and IS911, the ends are cleaved only after the formation of a PEC (Bhasin *et al.*, 2000; Normand *et al.*, 2001; Sakai *et al.*, 1995; Savilahti *et al.*, 1995). However, for *Mos1*, it is thought that cleavage of the non-transferred strand precedes PEC formation (Dawson and Finnegan, 2003). The simplest way to detect a PEC is to carry out a DNA binding assay, using a mixture of two different length DNA fragments. If a PEC is produced in the binding reaction, three different complexes will be seen on the binding gel: one complex containing two long DNA fragments, one containing two short DNA fragments and one containing one short and one long DNA fragment.

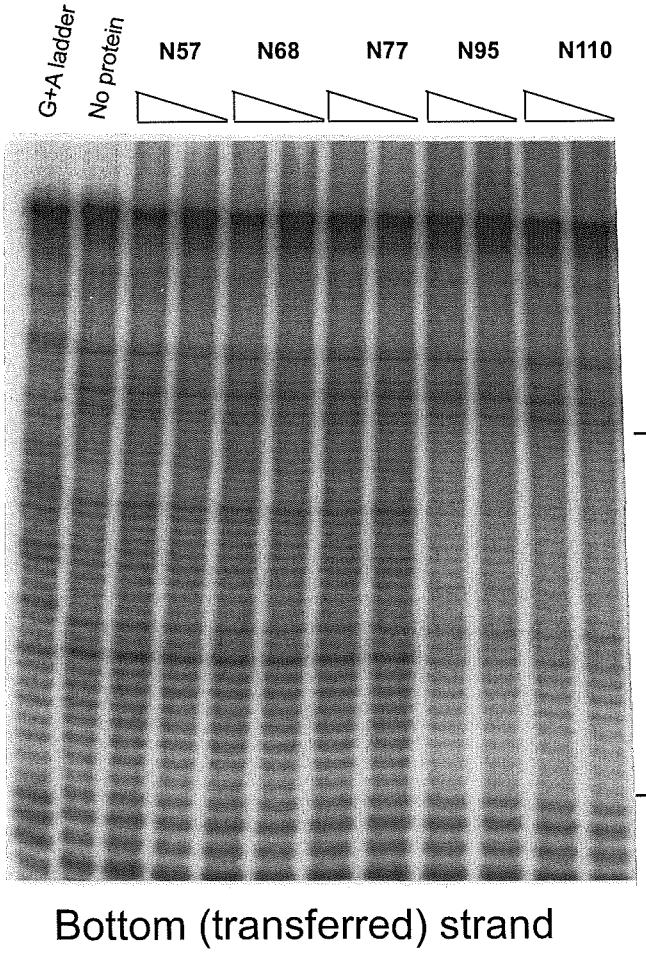
Long and short radiolabelled oligonucleotide substrates containing ISY100 inverted repeat sequences (IRL79 and IRR58, Figure 4.4) were incubated with purified His-tagged transposase either singly or in an equimolar mixture. Mg^{2+} and DMSO were added to allow catalysis of strand transfer in case this was required for PEC formation. IRL79 and IRR58 both produced one major retarded band, and at least one other more retarded minor band. However no additional bands, diagnostic for PEC formation, were observed in the presence of both DNA fragments in any of the conditions tested (Figure 4.12), strongly suggesting that the observed complexes contain only one molecule of DNA.

For *Mos1*, cleavage of the non-transferred strand is required for PEC formation. For this reason, PECs can only be detected in conditions that allow strand cleavage (Dawson

A)



Top (non-transferred) strand



Bottom (transferred) strand

B)

Figure 4.11 Footprinting of N-terminal transposase derivatives to the ISY100 left end. A) IRL79 5' end-labelled on either the top or bottom strand was incubated with cell extracts containing the indicated protein and subjected to a standard DNase I footprinting reaction. Reactions were separated on an 8% sequencing gel adjacent to a G+A ladder generated from the same DNA fragment. Binding reactions contained either 1 µl or 0.5 µl of the indicated cell extract. **B)** Summary of footprinting results. The 24 bp IR is shown in bold, transposon end sequences are shown in uppercase and regions protected by $Tnp_{1\text{-}95}$ and $Tnp_{1\text{-}110}$ are underlined.

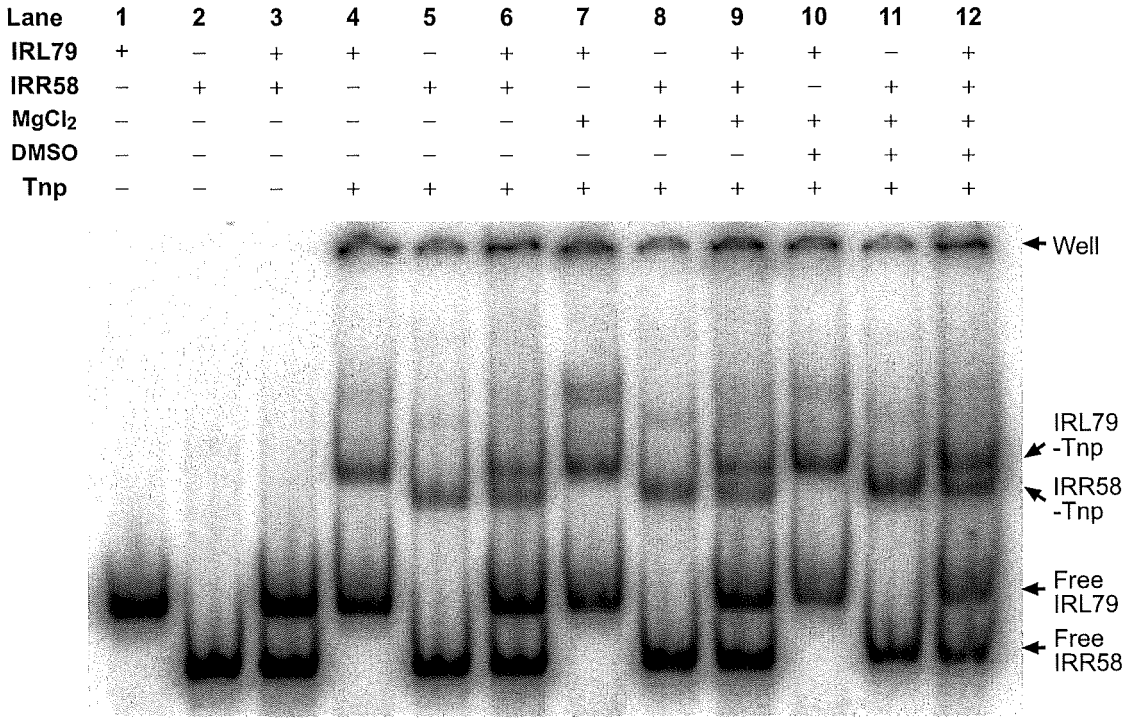


Figure 4.12 Binding reactions with DNA fragments of different lengths to detect PECs. Labelled DNA fragment IRL79 and IRR58 were incubated singly or in combination, with purified transposase in standard binding conditions (lanes 4-6) or with the addition of 10 mM $MgCl_2$ (lanes 7-9), or 10 mM $MgCl_2$ and 20% DMSO (lanes 10-12). Reactions were carried out for 10 minutes at 25°C and loaded directly onto a native polyacrylamide gel. Complexes formed by IRL79 or IRR58 are indicated.

and Finnegan, 2003). *Mos1* transposase requires DMSO and divalent cations for catalysis, and for this reason DMSO and Mg^{2+} were added to one set of binding reactions shown in Figure 4.12. However, the reactions were not run on strand separating gels to check whether any strand cleavage had occurred. In later experiments, it was shown that oligonucleotide substrates are cleaved inefficiently even in the presence of Mg^{2+} and DMSO. The binding reactions shown here were only allowed to proceed for 10 minutes, so it is possible that strand cleavage is required for PEC formation and not enough time was allowed for this to happen. Another possibility is that synaptic complexes form, but are not stable enough to survive on the gel system used here.

4.5 Cleavage by ISY100 transposase

After synapsis, the next step in transposition is strand cleavage, to excise the transposon from its donor site.

An *in vitro* cleavage assay is required to study how ISY100 transposase catalyses strand cleavage and whether this is similar to *Mos1* and other members of the Tc1/*mariner* family of transposons (see chapter I). This section describes work to set up assays for strand cleavage on supercoiled plasmid substrates and on short oligonucleotide substrates containing transposon terminal inverted repeat sequences. The conditions for cleavage were optimised, and the positions of strand cleavage were mapped.

4.5.1 ISY100 transposase cleaves at ISY100 terminal inverted repeats *in vitro*

To test whether transposase can carry out the excision reaction *in vitro*, pXF109, containing 51 bp IRL and 49 bp IRR sequences flanking a 1.3 kb kanamycin resistance gene in pUC18, was incubated with ISY100 transposase. Supercoiled pXF109 was incubated for four hours at 30°C with transposase in a simple buffer containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 0.1 mg/ml BSA and 0.1 mM DTT. The reactions also contained 50 mM NaCl and 0.01% triton X-100 from the transposase protein or dilution buffer. The reactions were stopped by heating at 75°C for 10 minutes and then loading buffer containing SDS and proteinase K was added before electrophoresis on an agarose gel (Figure 4.13). The products of excision were detected after staining with ethidium bromide. Two bands were produced of the correct size to be the 1.3 kb excised transposon and the 2.7 kb vector backbone. Thus transposase appears to produce double strand breaks specifically at the transposon ends. Linearised plasmid, probably from cleavage at only one

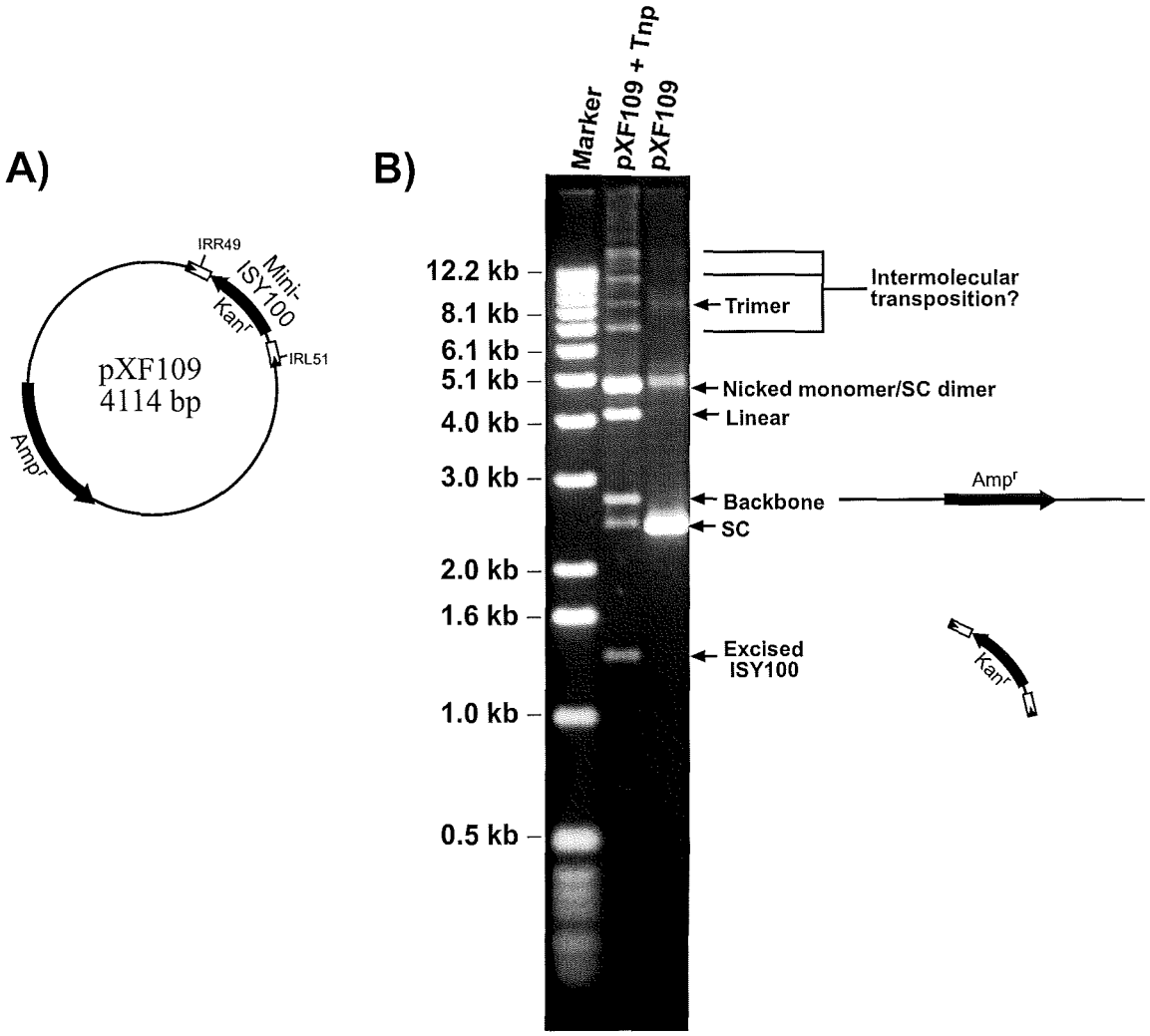


Figure 4.13 *In vitro* cleavage catalysed by ISY100 transposase. A) Map of donor plasmid pXF109. **B)** Purified transposase cleaves supercoiled donor plasmid *in vitro*. pXF109, carrying a 1.3 kb mini-ISY100, was incubated with purified His-tagged transposase in a buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 0.1 mg/ml BSA and 0.1 mM DTT for 4 hours at 30°C. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. Excised mini-ISY100, supercoiled substrate (SC), linear substrate, and backbone of the donor plasmid left after excision of the mini-ISY100 are indicated.

transposon end in pXF109, was also produced (see section 4.5.4). Another major product was open circular DNA that has lost all supercoiling because it contains a nick on at least one strand. Recent work has demonstrated that this DNA contains nicks specifically on the non-transferred strand at the transposon ends (Sean Colloms and Victoria Moffat, personal communications). Other products migrating slower than open circle substrate are also visible on the gel. Some of these are probably higher molecular weight cleavage products coming from the supercoiled dimer plasmid present in the substrate DNA. It is later shown that at least one of these higher bands is the product of one ended intermolecular transposition into the substrate plasmid.

4.5.2 Optimisation of cleavage by ISY100 transposase

Next, the reaction conditions for cleavage by ISY100 transposase were optimized and the requirements for cleavage were defined. First the requirement for divalent metal ions was investigated. The side chains of the catalytic DDE residues have been shown to coordinate divalent metal ions in a number of different transposases (Junop and Haniford, 1997; Richardson *et al.*, 2006; Steiniger-White *et al.*, 2004). Magnesium ions are thought to be the physiologically relevant cation, but manganese often functions as well or better *in vitro*. In some cases, manganese causes a relaxation in specificity, allowing reactions to proceed that do not work in the presence of magnesium (Allingham *et al.*, 1999; Kim *et al.*, 1995; Rowland *et al.*, 1995). Cleavage reactions were carried out at a range of Mg^{2+} and Mn^{2+} concentrations (Figures 4.14 and 4.15), using pXF153 carrying 30 bp from each transposon end, flanking a Kan^r gene. Excision was detectable at Mg^{2+} concentrations from 4 - 20 mM, while Mn^{2+} gave excision from 0.2-10 mM. The highest level of cleavage was obtained at around 10 mM $MgCl_2$ or 2 mM $MnCl_2$ (Figures 4.14 and 4.15). A range of concentrations of $NiSO_4$, $ZnSO_4$, $CuSO_4$, $CoCl_2$, and $CaCl_2$ from 0.1 mM to 15 mM were also tested in the cleavage assay but gave no detectable excision (data not shown).

Mn^{2+} gave more open circular product and a lower level of cleavage than magnesium. Mn^{2+} also gave much lower levels of the presumed intermolecular transposition products than did Mg^{2+}. The use of Mn^{2+} at higher concentrations led to nicking even in the absence of protein, probably through the production of reactive oxygen species such as hydroxyl radicals. For these reasons, Mg^{2+} (10 mM $MgCl_2$) was used for subsequent cleavage experiments.

To investigate the effect of NaCl concentration on the cleavage reaction, the transposase stock solution, containing 500 mM NaCl, was diluted 8 fold into a buffer

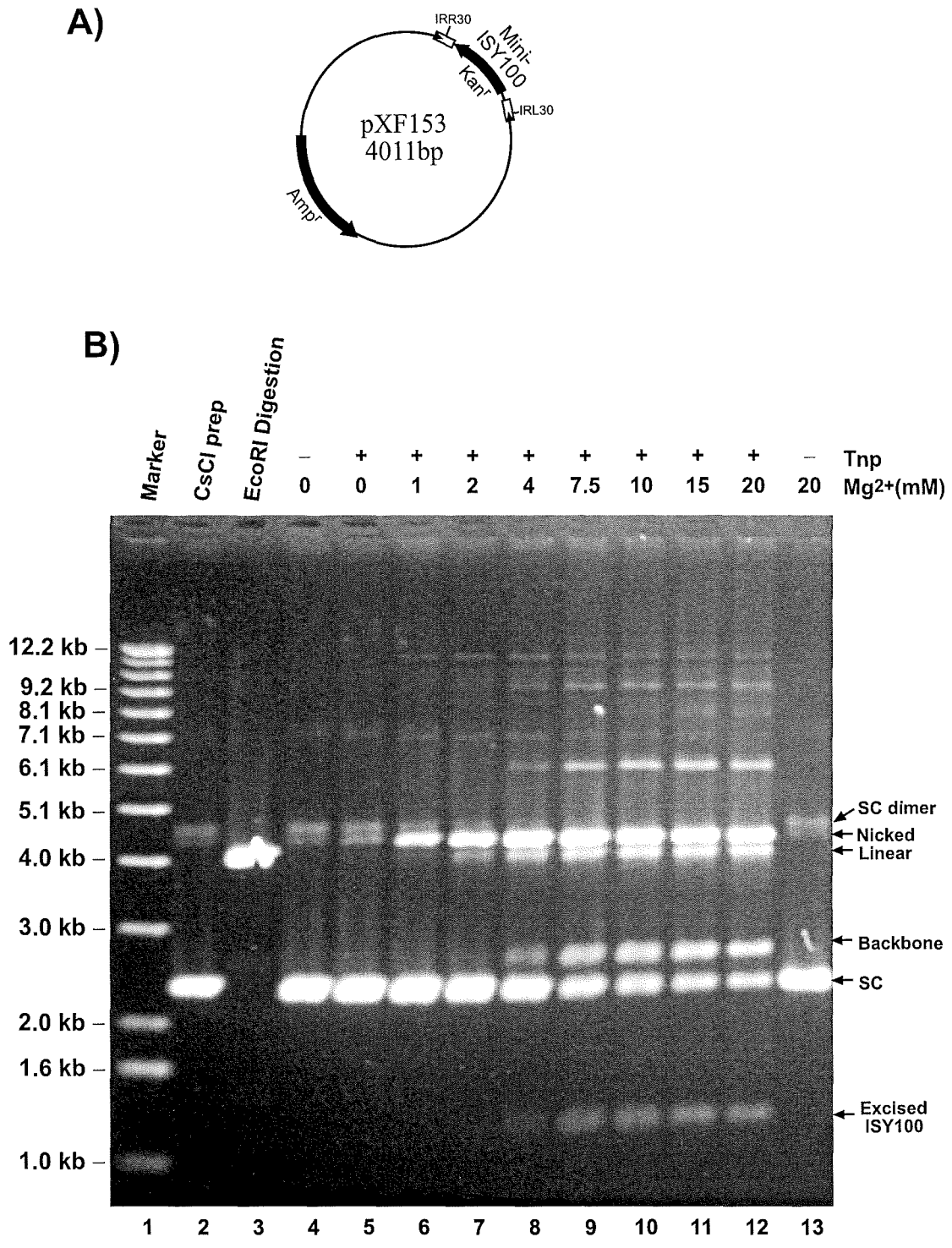


Figure 4.14 Effect of Mg^{2+} concentration on the cleavage activity of ISY100 transposase *in vitro*. **A)** Map of donor plasmid pXF153. **B)** Cleavage assays were performed by incubating 0.15 pmole supercoiled pXF153 together with 2.6 pmoles purified His-tagged transposase in a standard cleavage reaction buffer with the indicated concentration of $MgCl_2$ for 4 hours at 30°C. The products were separated by electrophoresis on a 1% agarose gel and stained with ethidium bromide. Lanes 2 and 3 show supercoiled and EcoRI linearised pXF153 respectively. Excised mini-ISY100, supercoiled (SC), linear and nicked substrate, and backbone of the donor plasmid are indicated.

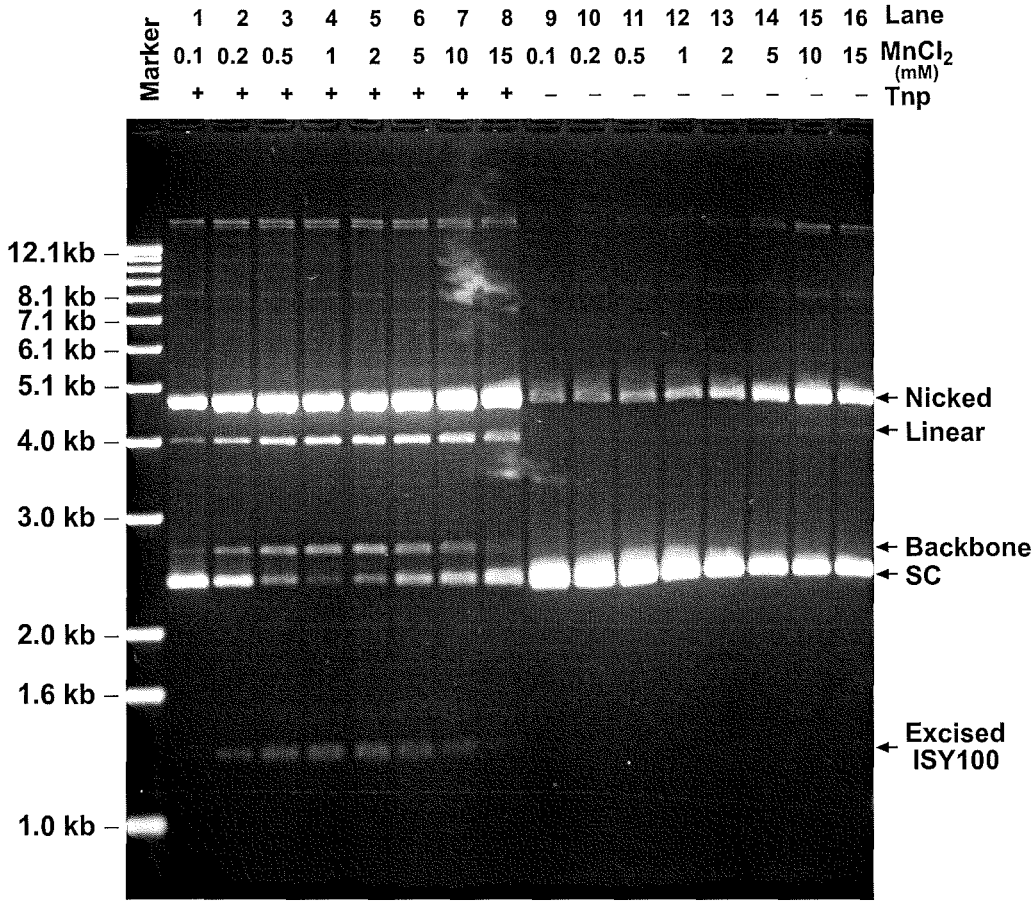


Figure 4.15 Effect of Mn^{2+} concentration on the cleavage activity of ISY100 transposase *in vitro*. Cleavage assays were performed by incubating 0.15 pmole supercoiled pXF153 plasmid with 2.6 pmoles purified His-tagged transposase in a standard cleavage reaction buffer (see Chapter II) without $MgCl_2$ but containing the indicated concentrations of $MnCl_2$, for 4 hours at 30°C. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. Bands are indicated as in Figure 4.14.

without NaCl. This dilution was then used in cleavage reactions with additional NaCl in the individual reactions to give final NaCl concentrations from 6.25 mM to 500 mM. Excision could be detected from 6.25 to 150 mM NaCl, and was most efficient at 40 to 80 mM NaCl (Figure 4.16). As the NaCl concentration was increased, the amount of single end cutting started to increase, and then above 80 mM NaCl the amount of nicking also increased. At 200 mM NaCl almost all of the substrate was nicked and there was very little double strand cleavage, and at 350 mM NaCl and above there was almost no activity at all. After this, 50 mM NaCl, coming from a one in ten dilution of the transposase dilution buffer, was used in the standard reaction conditions.

Next the effect of changing the reaction temperature was investigated (Figure 4.17). The reaction worked best between 20 and 30°C. There was almost no cleavage at 16°C and below, and temperatures of 37°C and above led to a marked reduction in the level of double strand cleavage. 30°C was adopted as the standard reaction temperature for subsequent experiments.

Then the effect of varying the pH was investigated (Figure 4.18). Reactions were performed in buffers ranging in pH from 7.0-10.7. Excision was only efficient in the buffers with pH 7.0, 7.5, 8.0 and 8.5. Some nicking and linearisation of the substrate were observed at higher pH, but further work is needed to determine whether cleavage is occurring at the transposon ends. It should be noted that different buffers were used to obtain the different pH values: Tris-HCl was used for pH 7.0 to 8.5, CHES was used for pH 9.0 to 9.5, CAPS was used for pH 9.7 to 10.5 and K_2HPO_4 was used for pH 10.7. Excision was only efficient in Tris buffers, so it is possible that the buffer as well as the pH is having an effect on the efficiency of cleavage.

Cleavage at ISY100 ends in pXF153 was then investigated in a time course experiment (Figure 4.19). Cleavage products were first detected at about 60-90 minutes and continued to accumulate over time. Cleavage continued throughout the time course and the maximum amount of cleavage was observed at the final 16-hour time point. The nicked substrate also accumulated over the full time course, appearing more quickly than the double strand cleavage products. There may be slightly less nicked substrate at 16 hours than at 4 hours, suggestion that the nicked substrate may be converted to double strand cleavage products by cutting of the second strand. For convenience, a four-hour incubation period was used for subsequent cleavage assays.

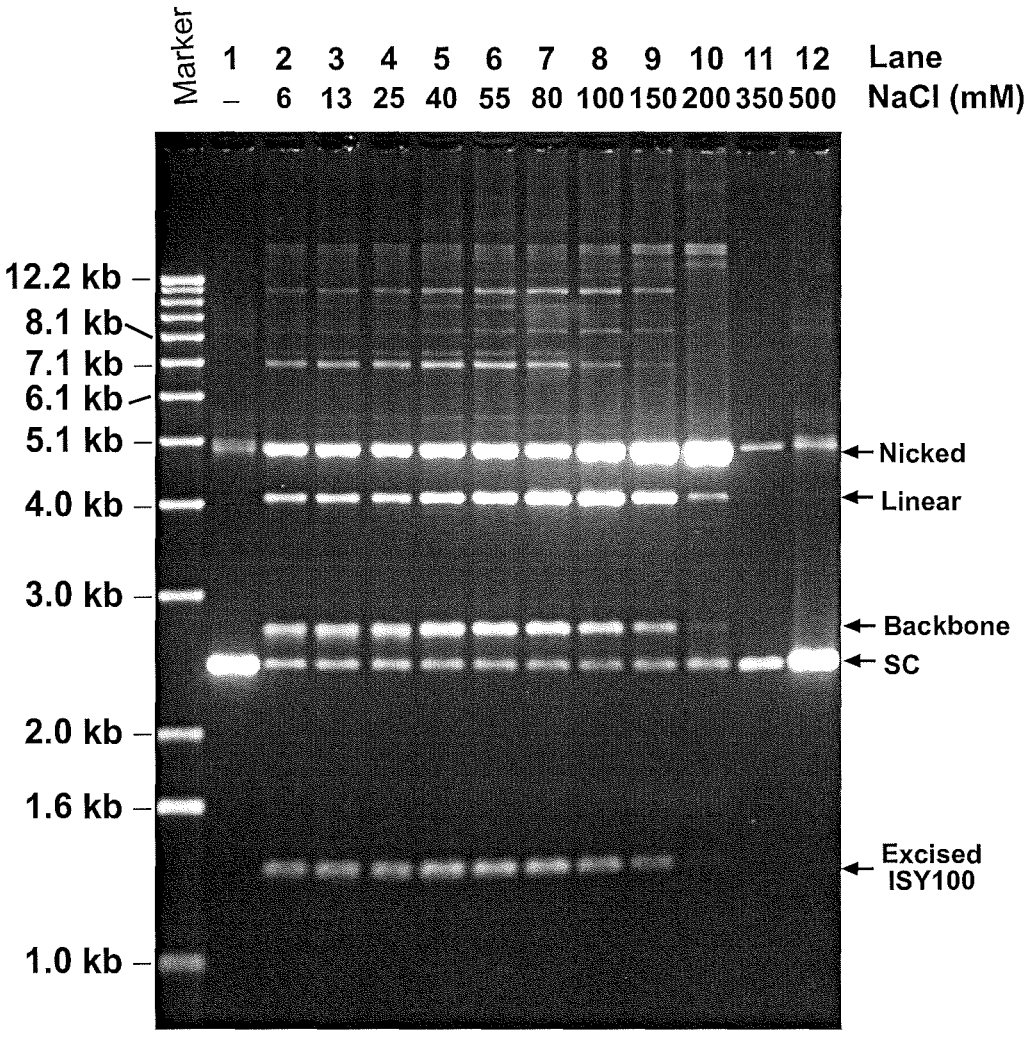


Figure 4.16 Effect of changing NaCl concentration on the cleavage activity of ISY100 transposase *in vitro*. Cleavage assays were performed by incubating 0.1 pmole supercoiled pXF153 together with 2.6 pmoles purified His-tagged transposase in a standard cleavage reaction buffer in the presence of 10 mM $MgCl_2$ and the indicated concentration of NaCl for 4 hours at 30°C (lanes 2 - 12). The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. Lane 1 shows untreated substrate. Bands are indicated as in Figure 4.14.

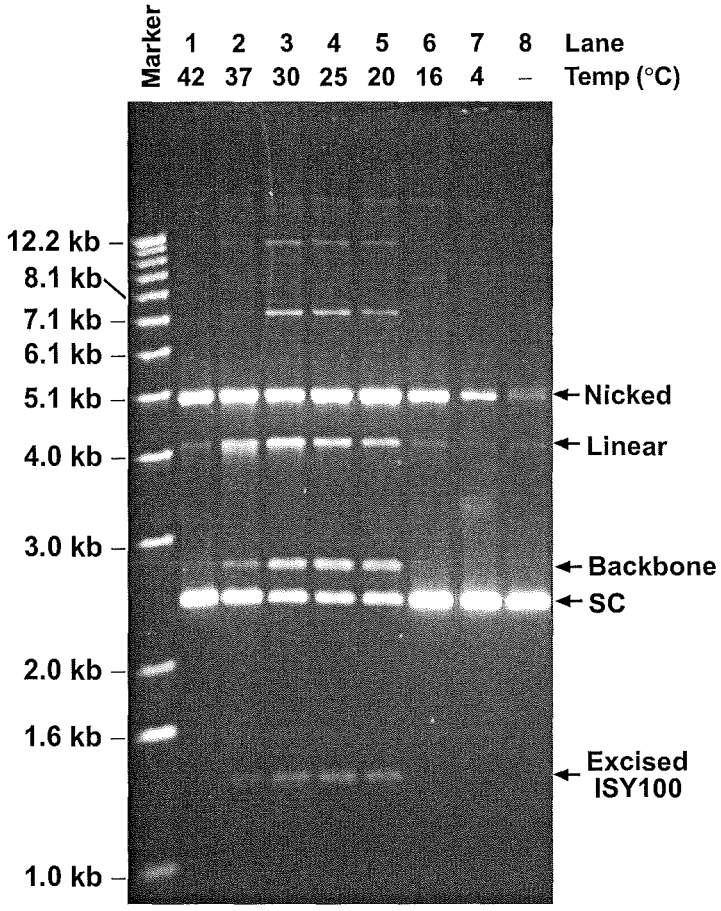


Figure 4.17 Effect of reaction temperature on the cleavage activity of ISY100 transposase *in vitro*. Cleavage assays were performed by incubating 0.10 pmole supercoiled pXF153 plasmid together with 2.6 pmoles purified His-tagged transposase in a standard cleavage reaction buffer for 4 hours at the indicated temperatures. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. Bands are indicated as in Figure 4.14.

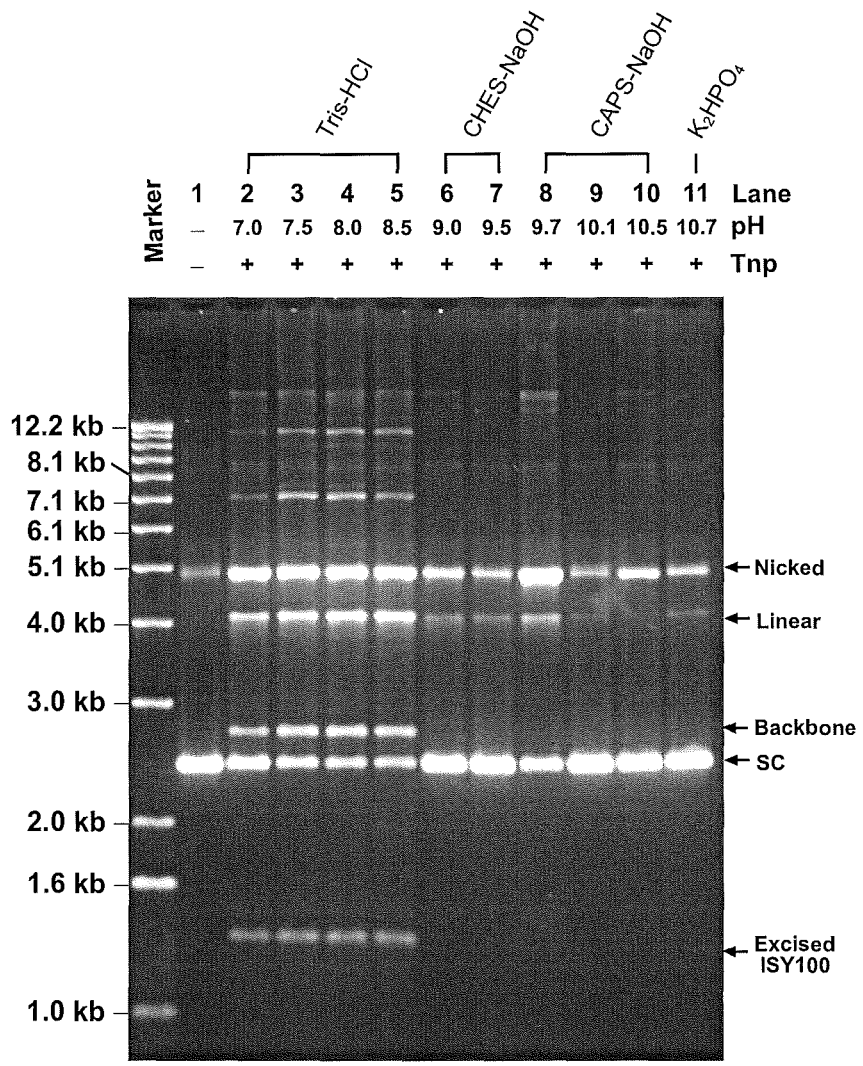


Figure 4.18 Effect of pH on the cleavage activity of ISY100 transposase *in vitro*. Cleavage assays were performed by incubating 0.10 pmole supercoiled plasmid pXF153 together with 2.6 pmoles purified His-tagged transposase in a standard cleavage reaction buffer at different pH values. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. pH values and buffers used are indicated on the top of each lane. Bands are indicated as in Figure 4.14.

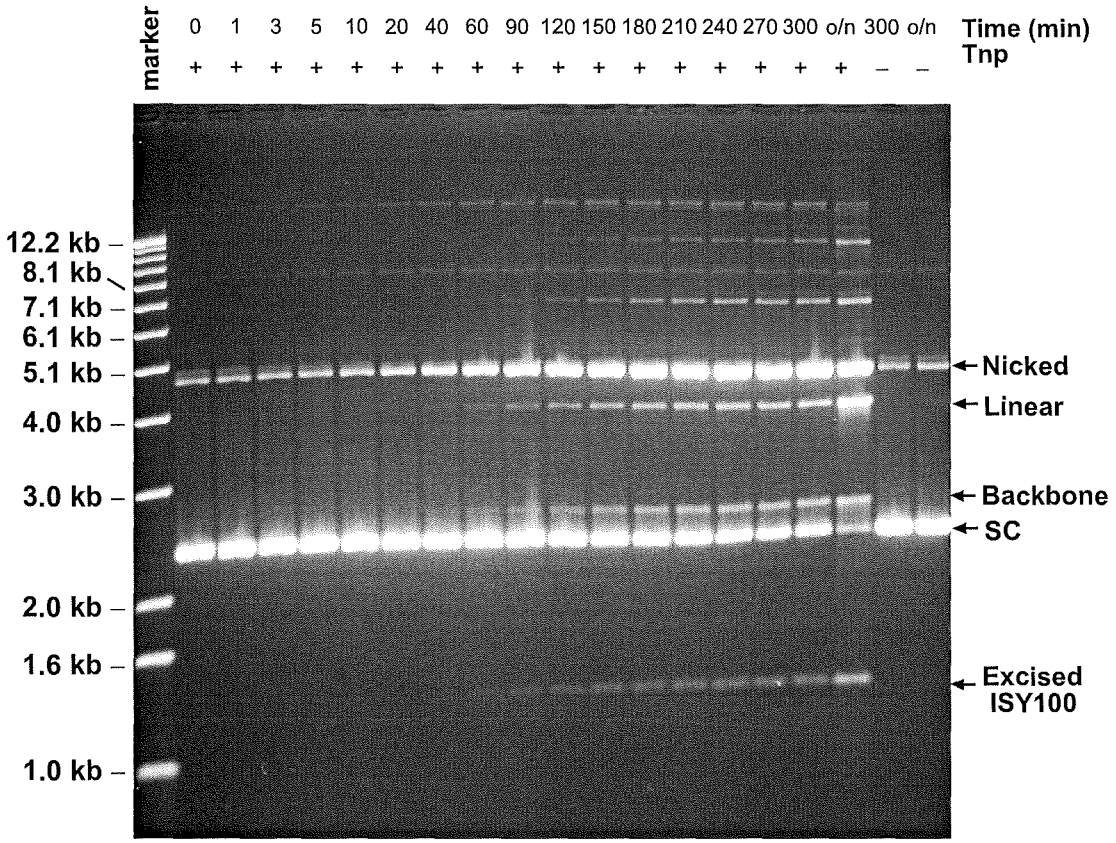


Figure 4.19 Effect of incubation time on the cleavage activity of ISY100 transposase *in vitro*. Cleavage assays were setup in a single 340 µl reaction containing 0.12 pmoles supercoiled donor plasmid pXF153 together with 2.6 pmoles purified His-tagged transposase for each time point. 20 µl samples were withdrawn at the indicated time points and stopped by heating at 75°C for 10 minutes. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. The incubation time is indicated on the top of each lane. Bands are indicated as in Figure 4.14. o/n indicates overnight for 16 hours.

To see if the ratio of transposase to DNA is important for the amount of DNA cleavage, reactions were carried out at a range of DNA concentrations at two different protein concentrations (Figure 4.20). At a given transposase concentration, increasing the amount of DNA present increased the amount of excised transposon produced up to a certain point. As the DNA concentration increased beyond this point, the total amount of excised transposon did not increase any more, and even started to decrease slightly. When the protein concentration was doubled, the DNA concentration giving maximum excision also doubled. In other words, having too much or too little protein per DNA molecule inhibited the cleavage reaction and it was the ratio of protein to DNA that determined the extent of the reaction. Too much protein per DNA molecule might lead to multimerisation of transposase on the transposon ends or binding at incorrect sites, inhibiting cleavage. Too little protein per DNA molecule might produce plasmid molecules that have transposase bound at only one end and prevent synapsis and concerted cleavage. An excess of 5 -10 fold of protein over DNA was required to nick approximately 50% of the substrate in 4 hours. This could be because a proportion of the transposase is inactive. However it appears that each transposase does not catalyze multiple events.

All of the experiments reported above were carried out using a plasmid with two ISY100 ends. To see whether two ends are required for cleavage, cleavage assays were carried out with supercoiled substrates carrying only one IR from ISY100 (Figure 4.21). Supercoiled pXF119 and pXF121, carrying 51 bp and 49 bp from the left and right ends of ISY100 respectively, were incubated for four hours with dilutions of transposase under standard reaction conditions. Nearly half of the substrate DNA was nicked in this time at the highest transposase concentration, and a small amount of linearised plasmid was also produced. However, the main conclusion from this experiment is that substrates with one transposon end are cleaved much less efficiently than substrates with two transposon ends (compare the pXF109 control in lane 1 with the equivalent pXF119 and pXF121 reactions; Figure 4.21 lanes 1, 3 and 9). This result suggests that communication between the two ends on the same DNA molecule is required for efficient cleavage.

4.5.3 Cleavage on linear DNA substrates

The experiments reported in the previous section all used supercoiled DNA substrates, mimicking the state of DNA *in vivo*. To test the requirement for DNA supercoiling, cleavage assays were carried out on linearised substrates carrying one or two ISY100 ends (Figure 4.22). In standard conditions, cleavage of linearised pXF109, pXF119 and pXF121

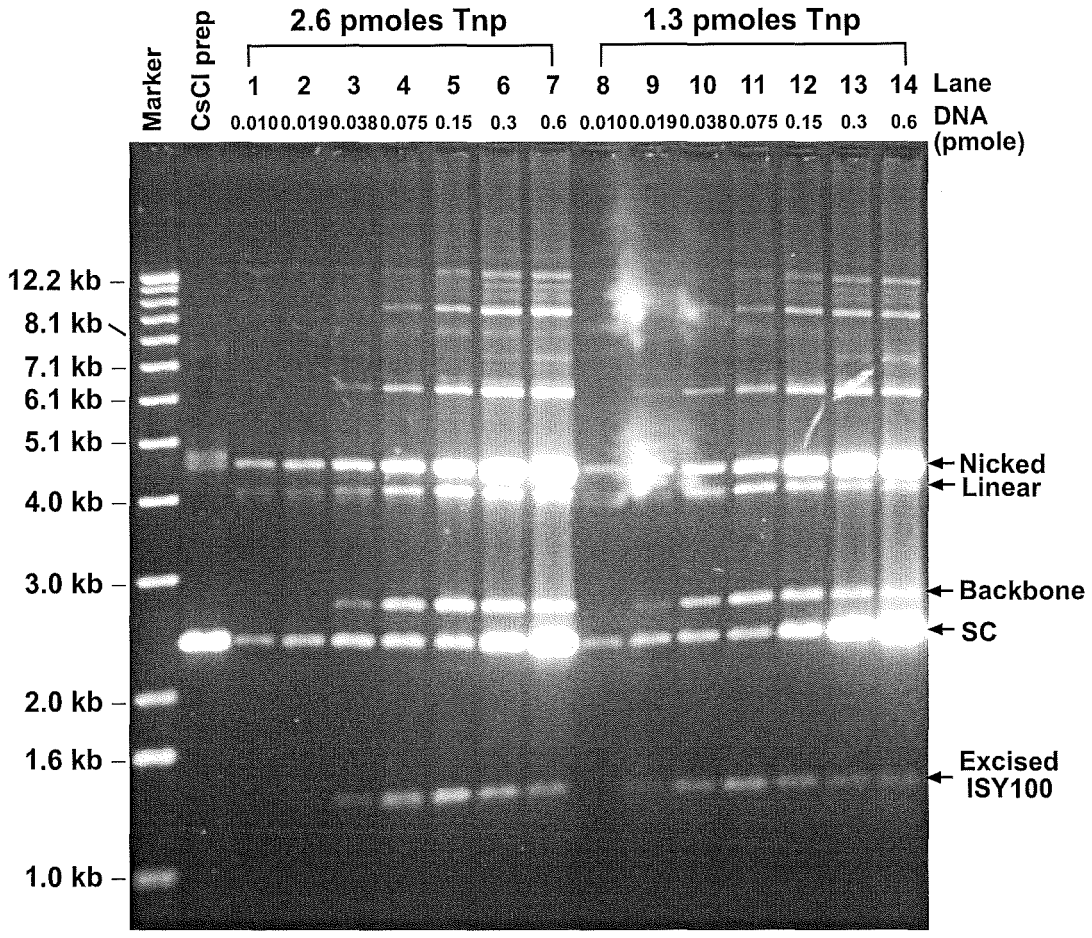


Figure 4.20 Effect of transposase / substrate ratio on the cleavage activity of ISY100 transposase *in vitro*. Cleavage assays were performed at different DNA concentrations in the presence of two different concentrations of transposase in a standard cleavage reaction at 30°C for 4 hours. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. The amounts of protein and substrate used in the reactions are indicated on the top of each lane. Bands are indicated as in Figure 4.14.

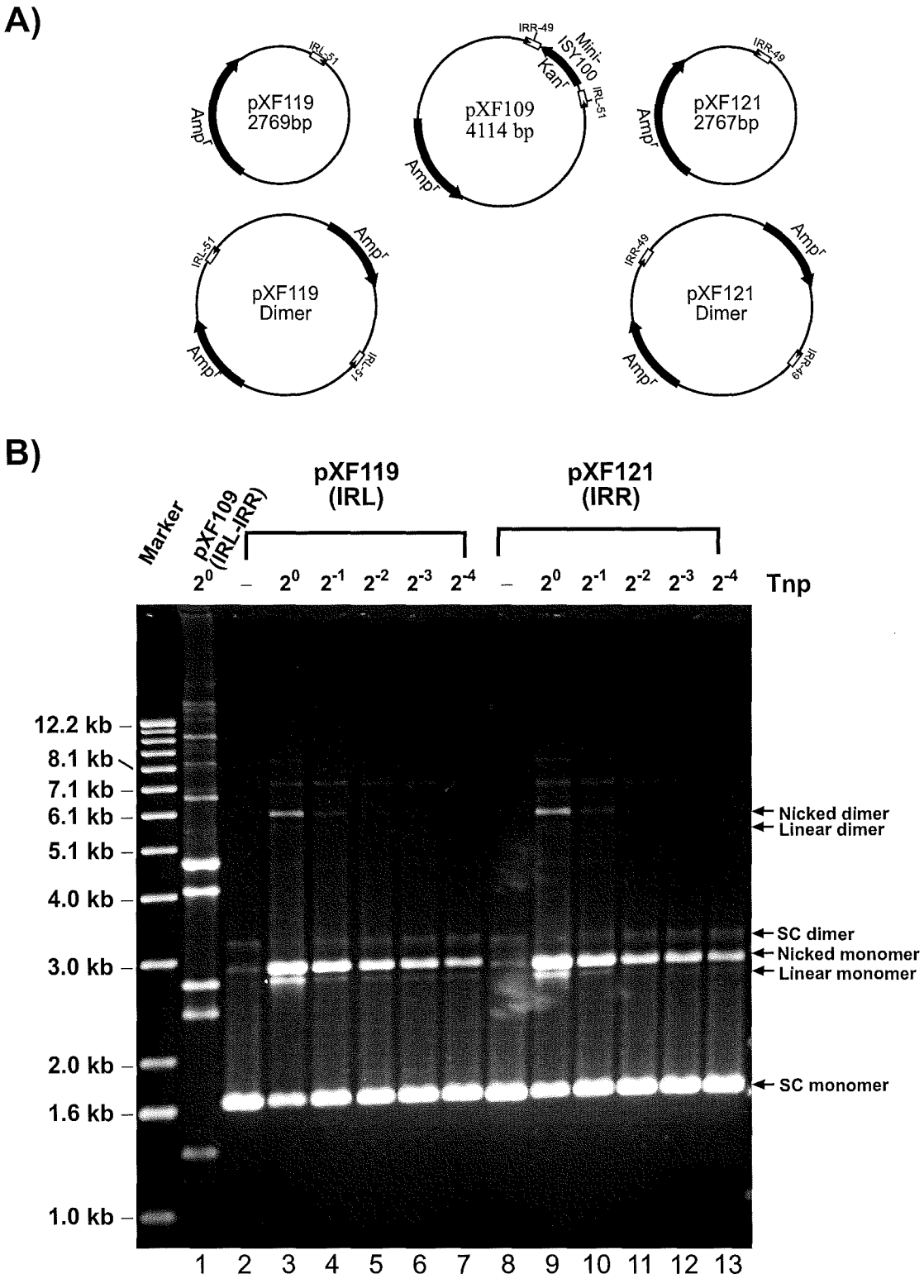


Figure 4.21 ***In vitro*** **cleavage activity of ISY100 transposases on substrates containing one end of ISY100. A)** Structure of plasmids used in this assay. pXF119 and pXF121 contain 51 bp and 49 bp from the left end and right end of ISY100 respectively in pUC18. Small amounts of supercoiled dimer were present in the plasmid substrates. **B)** Cleavage assays were performed with the indicated relative concentrations of purified His-tagged transposase in a standard cleavage reaction buffer at 30°C for 4 hours. 2^0 corresponds to a 2.6 pmoles protein per 20 µl reaction. Nicked and linearised monomer and dimer bands are indicated.

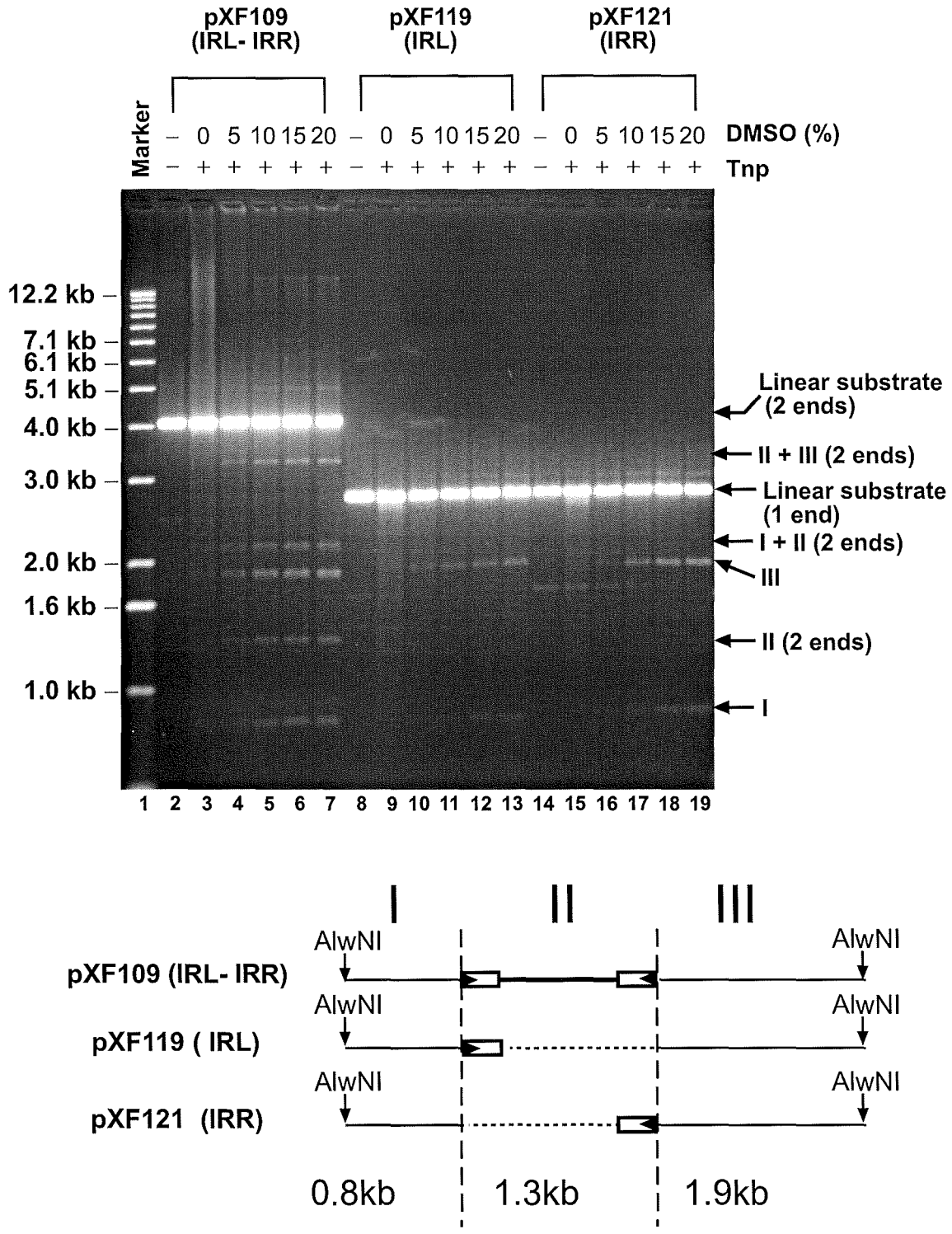


Figure 4.22 DMSO enhances the cleavage activity of ISY100 transposase on linear substrates. pXF109, pXF119 and pXF121 containing IRL + IRR, or just IRL or IRR of ISY100 were linearised with AlwNI and then treated with purified transposase in a standard cleavage assay with various concentration of DMSO at 30°C for 4 hours. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. The concentrations of DMSO are shown above each lane, and the predicted fragments from cleavage at transposon ends are shown to the right of the gel.

(with IRL and IRR, just IRL or just IRR respectively) was almost undetectable. The addition of DMSO has been reported to allow transposition on linear DNA substrates in other systems (Baker and Mizuuchi, 1992; Dawson and Finnegan, 2003), so the effect of DMSO on cleavage of linearised pXF109, pXF119 and pXF121 was tested (Figure 4.22). When 5%-20% DMSO was added to the reactions, the amount of cleavage on all three substrates increased, although cleavage was still relatively inefficient. The efficiency of cleavage at 10%, 15% and 20% DMSO was approximately equal, and the three higher concentrations appear to give better cleavage than 5% DMSO. The plasmid with two ends was not cleaved much better than the plasmids with one end, and cleavage on pXF109 does not appear to be highly concerted at the two ends.

The effect of DMSO was also tested on a supercoiled substrate containing two ISY100 ends, with a slightly different result (Figure 4.23). DMSO was not required for excision and has a smaller effect. As the concentration of DMSO was increased from 0 to 10%, the efficiency of first strand cleavage increased so that the amount of supercoiled substrate decreased. At 10% DMSO and above, no supercoiled substrate remained. However DMSO appeared to have very little effect on the extent of double strand cleavage. Increasing the DMSO concentration from 0-10% also reduced the yield of presumed intermolecular transposition product.

4.5.4 The sequence of the flanking dinucleotide affects cleavage efficiency

The results reported in chapter III show that the dinucleotide sequences flanking ISY100 affect the transposition frequency *in vivo*. The flanking sequences are removed at the cleavage step so it seems most likely that their effect on transposition is before or during this step. To test this, supercoiled substrates carrying mini-ISY100s with different flanking dinucleotides were tested in the *in vitro* cleavage assay (Figure 4.24).

The results of standard cleavage assays on plasmids containing identical changes at both ends of mini-ISY100 elements are shown in Figure 4.24A (lanes 1-12). Changing the base adjacent to the transposon (position -1) from A (reading 5' to 3' into ISY100) to any of the other three nucleotides almost completely abolished double strand cleavage. Transposase still nicked these substrates efficiently, although it has not been determined whether this nicking occurs at the transposon ends. When the base at position -2 was changed at both ends from T to A or C, double strand cleavage was greatly reduced, giving just a small amount of linear full length cleavage product, but almost no detectable excised transposon fragment (ETF) or vector fragment. Again, these changes did not seem to have

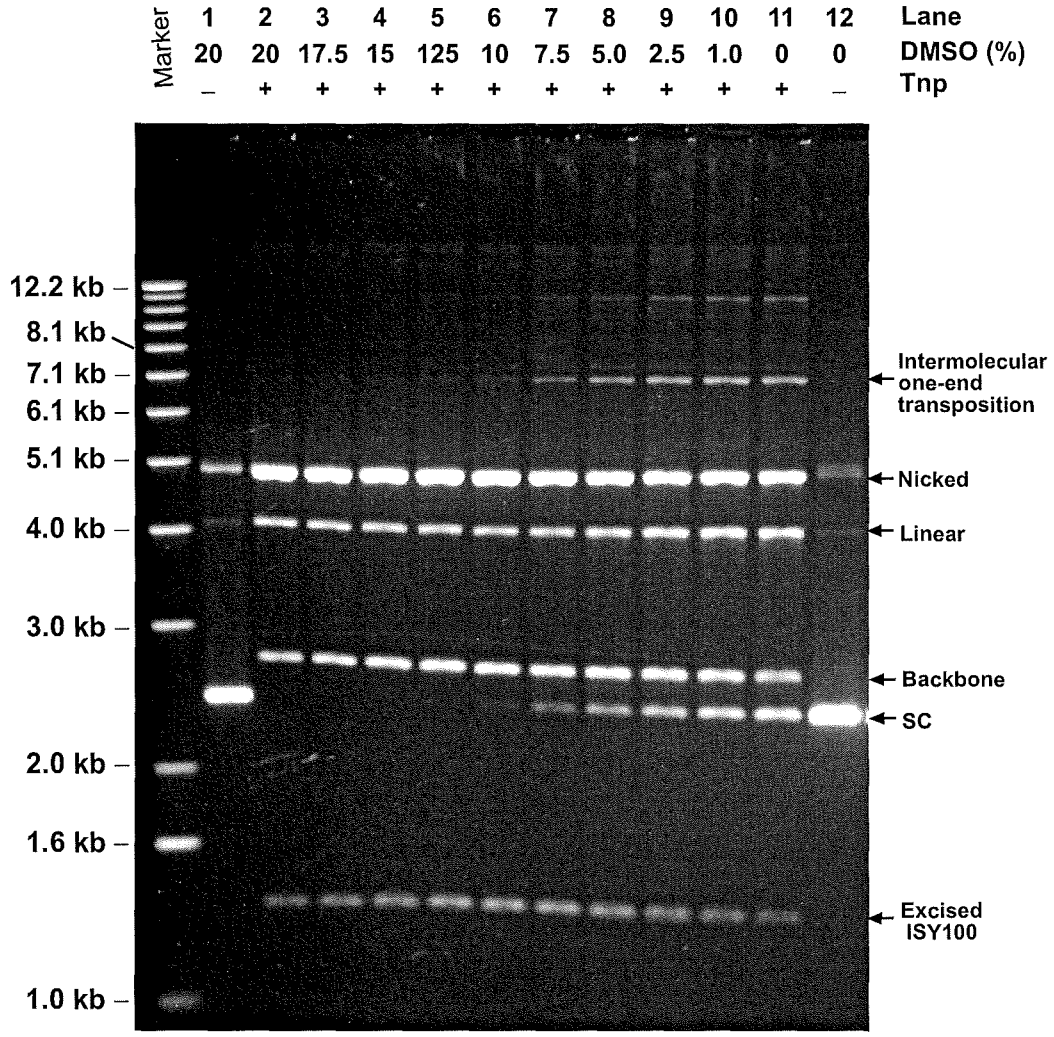


Figure 4.23 Effect of DMSO concentration on the cleavage activity of ISY100 transposase *in vitro*. Cleavage assays were performed by incubating 0.10 pmole supercoiled plasmid pXF153 together with 2.6 pmoles purified His-tagged transposase in a standard cleavage reaction with various concentrations of DMSO at 30°C for 4 hours. The products were separated by electrophoresis on a 1.2% agarose gel and stained with ethidium bromide. Concentrations of DMSO are indicated above each lane. Excised mini-ISY100 and backbone of the donor plasmid are indicated.

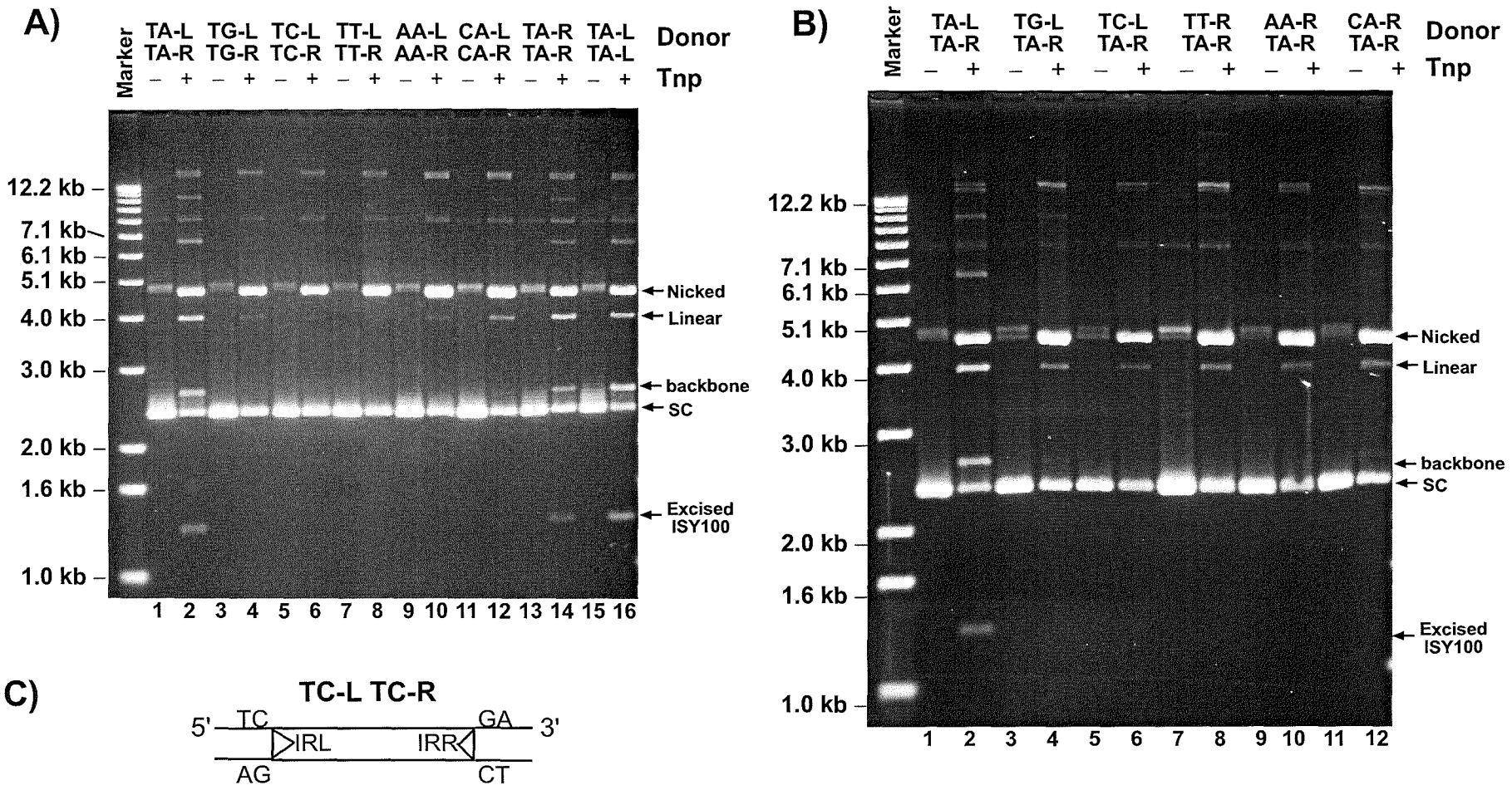


Figure 4.24 Flanking dinucleotides affect the cleavage activity of ISY100 transposase *in vitro*. Standard cleavage assays were performed on supercoiled substrates carrying mini-ISY100s with alternative flanking dinucleotides. Transposons with two left IRs or two right IRs were also tested. The products were separated by electrophoresis on 1.2% agarose gels and stained with ethidium bromide. **A)** Mini-transposons with symmetrical changes in the flanking dinucleotides at both ends and transposons with two left inverted repeats or two right inverted repeats as indicated. **B)** Mini-transposons with changes at only one end. **C)** Mini-transposons are referred to according to the sequences of the flanking dinucleotides reading 5' - 3' towards the transposon at both ends.

much of an effect on first strand cleavage, as large amounts of nicked substrate were produced. These same changes were also introduced at just one end of the mini-transposon, abolishing production of ETF and donor backbone, but also greatly reducing the yield of linearised plasmid (Figure 4.24B). It appears that changes at one transposon end have an effect on double strand cleavage at the other end. However, there was no detectable effect on first strand cleavage, as all of these substrates were nicked efficiently by transposase.

To investigate whether the linear product from these reactions was cleaved at one of the transposon ends and, if so, at which end, the full-length linear cleavage product was gel purified from reactions with TA-L/TA-R, CA-L/CA-R and TC-L/TA-R. This DNA was then cut with AlwNI and run on an 1.2% agarose gel (Figure 4.25). This analysis revealed that the linear product from TA-L/TA-R had been cleaved by transposase either at IRL or at IRR, with more cleavage at IRL than at IRR (Figure 4.25; lane 5). Altering the T at position -2 to C at both ends (CA-L/CA-R) reduced the amount of linear product so that the AlwNI products were very hard to detect by ethidium staining. Nevertheless, cleavage at both left and right ends was just detectable (Figure 4.25; lane 7). When a change was made at just one end (TC-L/TA-R), although double strand cleavage was greatly reduced, the cleavage that did take place occurred only at the unaltered end (Figure 4.25; lane 9).

The sequence of the inverted repeats affected the transposition frequency *in vivo*, such that a mini-transposon with two copies of IRR transposed less efficiently than one with two copies of IRL (see chapter III). This effect could be caused by a difference either before or after the cleavage steps. Substrates containing two IRR sequences or two IRL sequences with normal flanking TAs were therefore tested in the cleavage assay. Under the same reaction conditions, more ETF and donor backbone were obtained from the substrate containing two copies of IRL in inverted repeat than from the substrate containing two copies of IRR (Figure 4.24 A; lanes 13-16), suggesting this substrate is affected at the cleavage step of transposition.

4.5.5 Mapping the cleavage site on the transposon ends

Different transposons employ different mechanisms to cleave the transposon ends prior to the strand transfer reactions. The replicative transposons cleave only the transferred strand at each end, while non-replicative transposons cleave both strands. The mechanism for cleaving both strands can involve a hairpin intermediate, which can be on the transposon end as in Tn5 and Tn10 (Bhasin *et al.*, 1999; Kennedy *et al.*, 1998) or on the flanking sequence as in *Hermes* (Zhou *et al.*, 2004). In other cases, such as *mariner* and Tn7, there

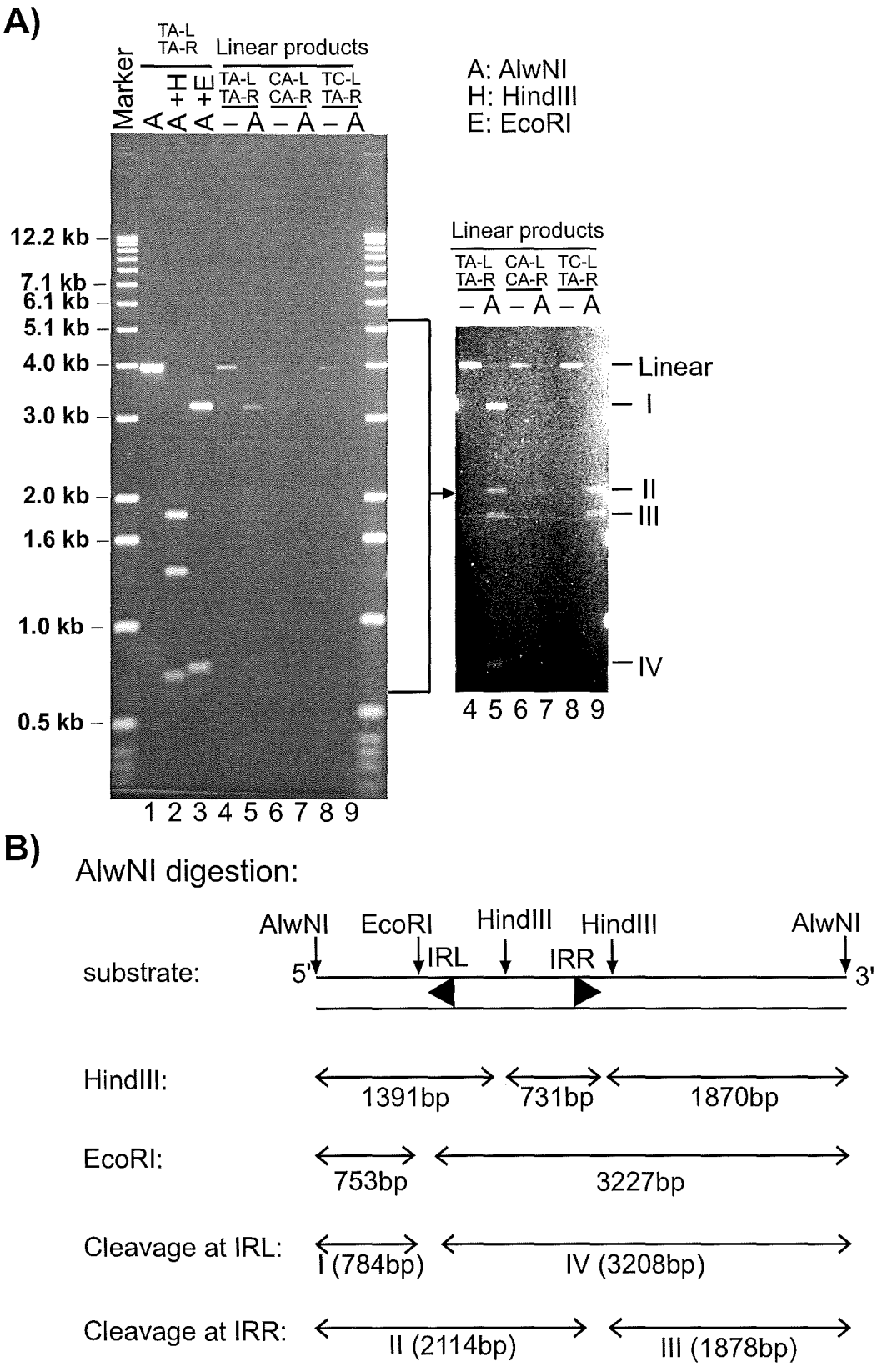


Figure 4.25 Analysis of linear products from cleavage assays on supercoiled substrates with different flanking dinucleotides. A) 1.2% agarose gel electrophoresis of linear products from cleavage reactions with supercoiled substrates with the indicated flanking dinucleotides (see Figure 4.24). Lanes 1 - 3 contain digestion products of supercoiled substrate as size markers. Lanes 4, 6 and 8 contain gel purified linear products from the cleavage reactions and lanes 5, 7 and 9 contain the AlwNI digestion products. Lanes 4 - 9 are shown at the right with the contrast adjusted to enhance faint bands. **B)** Map of substrate showing restriction sites and predicted transposase cleavage products.

is no hairpin intermediate. In all DDE transposons studied so far, the transferred strands are cleaved exactly at the transposon boundaries to produce 3'OH ends that attack the target DNA in the strand transfer reaction. The non-transferred strand can be cleaved outside the transposon, exactly at the transposon end, or within the transposon, giving protruding 5' ends, blunt ends, or recessed 5' ends respectively. In all cases studied to date, the non-transferred strand carries a 5' phosphate group. Studies on the positions of cleavage by transposase, and intermediates in the cleavage reaction, provide information on the mechanism of the transposition reaction and the organisation of the transposase at the transposon ends.

The cleavage sites on ISY100 were first mapped by Urasaki and colleagues (Urasaki *et al.*, 2002). In that study, excised transposon fragments were produced *in vivo* from a mini-ISY100 donor plasmid by over-expression of transposase in *E. coli*. The excised transposon fragment was purified electrophoretically and then tailed with dGTP using terminal deoxynucleotide transferase (TdT). The poly-G tailed transposon end was amplified by PCR using a poly-C primer and an end specific primer, and the PCR product was then sequenced, demonstrating that the transposon was cleaved exactly at the transposon boundary at its 3' ends. The fact that the 3' ends could be extended with terminal transferase implies that a 3' OH was present at these ends. The 5' ends of the transposons were mapped by a more indirect method. The transposon ends were first made blunt with the Klenow fragment of DNA polymerase I in the presence of dNTPs. This should extend the 3' ends if protruding 5' ends are present, or shorten the 3' ends if the 5' ends are recessed. After this treatment, the transposon 3' ends were determined using the same dG-tailing PCR based method described above. The results of this experiment implied that the 5' ends of the excised ISY100 transposon were recessed by two nucleotides. The experimental techniques used by Urasaki *et al.* (2002) to determine the transposase cleavage sites have some disadvantages: 1) The excised transposon was produced *in vivo*, and may have been processed by other enzymes after cleavage by transposase. 2) The method only reveals the cleavage points of those molecules that have a 3' OH. Any molecules that cannot be extended by TdT will be missing from the analysis. 3) Klenow may not efficiently produce blunt ends, so the 5' cleavage site deduced may not be accurate. 4) The sequencing reaction on the PCR products will miss any minor cleavage products. For these reasons, it was decided to determine the positions of cleavage by purified transposase *in vitro*.

The first approach taken was to use T4 DNA ligase to circularise the vector fragment produced in an *in vitro* transposition reaction. The circularised vector fragment was transformed into DH5α giving a large number of colonies, suggesting that both ends of the fragment contained compatible sticky or blunt ends. Plasmid DNA was purified from two single colonies and sequenced. Both plasmids had identical sequences, with both flanking TA dinucleotides and an extra TA dinucleotide at the empty donor site (Figure 4.26A). This result is consistent with the results of Urasaki *et al.* (2002), cleavage at the exact 3' ends of transposon ends with 2 nucleotide recessed 5' ends. However, it is also consistent with 5' ends exactly at the transposon ends and two nucleotides recessed 3' ends, or with blunt cleavage 1 bp inside the transposon ends (Figure 4.26B). The latter two suggestions seem unlikely as they would leave truncated transposon 3' ends. A drawback of this method is that it only gives information about the cleavage sites in those molecules that can self-ligate, that is those with 5' phosphates, 3' hydroxyls and compatible overhanging or blunt ends.

Another approach was therefore taken to determine the cleavage sites directly. An experiment was first carried out to map the cleavage position on the non-transferred strand of IRL in the presence and absence of DMSO, to see if DMSO is required for cleavage on short linear substrates and to see if DMSO has any effect on the cleavage position. A DNA fragment (IRL30; Figure 4.4) containing 30 bp from ISY100 IRL was labelled at the 3' end at the HindIII site. This DNA fragment was incubated with transposase and the cleavage products were separated on a sequencing gel adjacent to a G+A Maxam and Gilbert sequencing ladder. To provide an additional marker, the DNA fragment was also cut with AcuI, which cleaves the top (non-transferred) strand 2 nucleotides inside the transposon end on this fragment, exactly at the transposase cleavage position predicted from the results of Urasaki *et al.* (2002). As expected, cleavage was much more efficient in the presence of DMSO, although a long exposure revealed that some cutting occurred in the absence of DMSO (Figure 4.27A). One major cleavage site, one nucleotide outside the transposon, and a minor cleavage site, two nucleotides outside the transposon, were observed on the non-transferred strand. DMSO did not alter the sites of cleavage (Figure 4.27). There was no evidence of cleavage at the expected site, two nucleotides inside the transposon. The cleavage products were treated with phosphatase to see if cleavage left a phosphate group on the transposon 5' end. Removing the 5' phosphate should decrease the charge and hence the mobility of the cleavage product. This result was not so clear, but

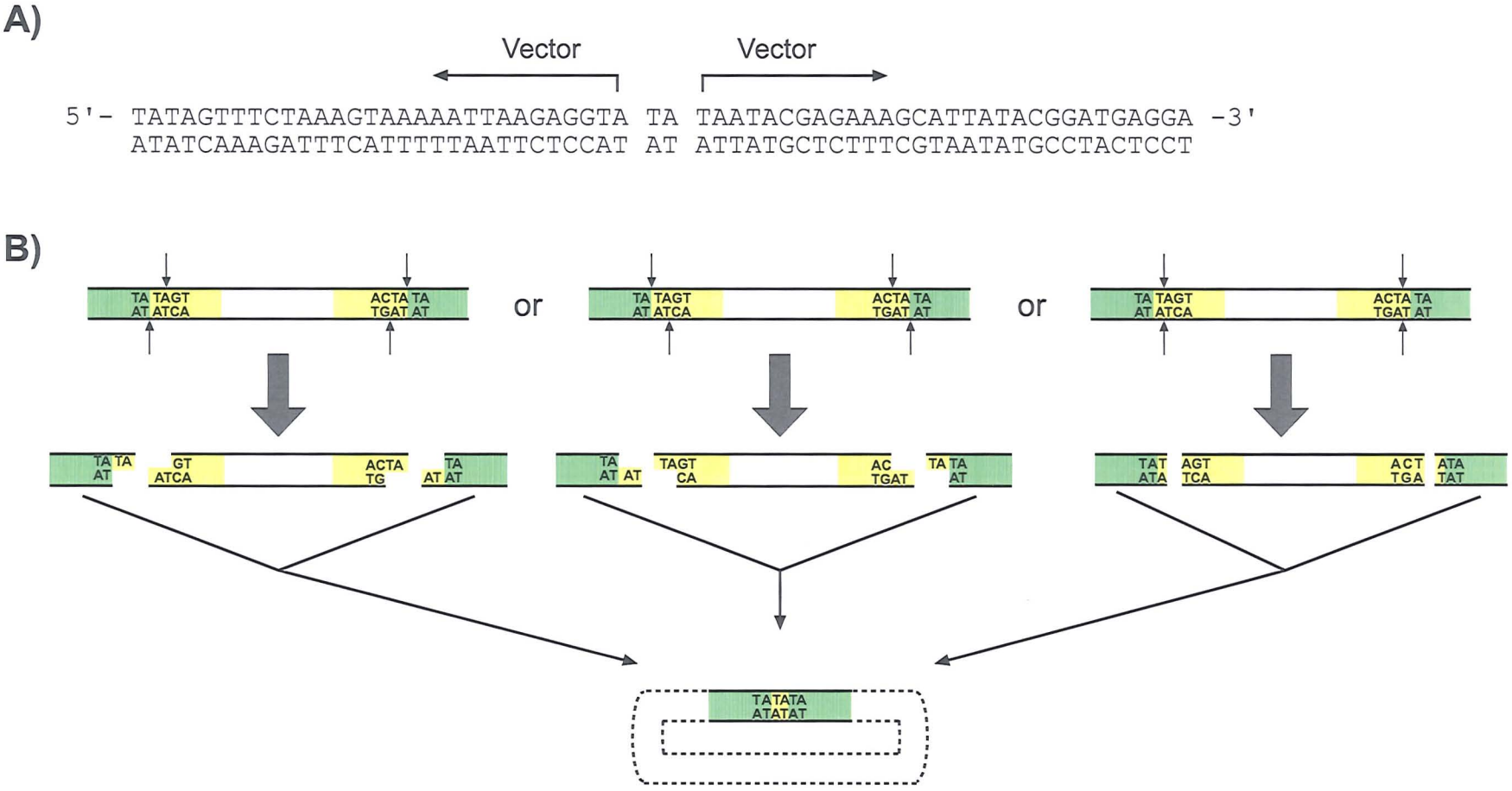


Figure 4.26 Junction sequence of self-ligated donor backbone. A) Two plasmids produced by circularising the backbone fragment from an *in vitro* excision reaction were recovered by transformation into DH5α and sequenced using the M13 uni-43 and rev-49 primers. The junction sequence is shown. **B)** Models for cleavage at ISY100 ends. Yellow represents transposon end sequences and green represents vector sequence.

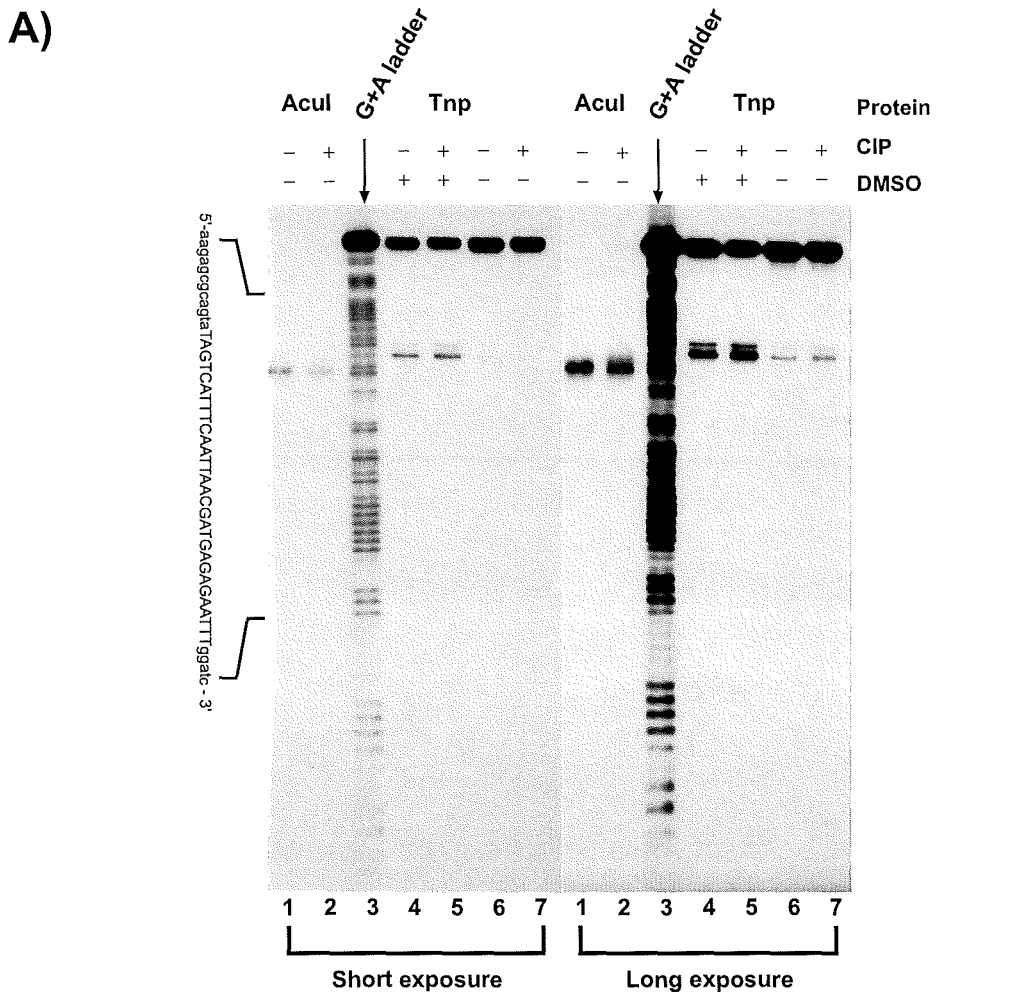


Figure 4.27 DMSO enhances the cleavage activity of ISY100 transposase on a radiolabelled IRL DNA fragment *in vitro*. IRL30 was 3' end-labelled at the HindIII site with [α-^{32}P] dATP and the Klenow fragment of DNA polymerase. This was incubated with purified His-tagged transposase in a standard cleavage assay at 30°C for 2.5 hours with or without 20% DMSO. Reactions were stopped by heating at 75°C for 10 minutes and then were split into 2 fractions, one of which was treated with calf intestinal alkaline phosphatase (CIP) for 30 minutes at 37°C. The labelled fragment was also digested by AcuI for 30 minutes and half of the digestion product was then dephosphorylated by CIP. Samples were separated on an 8% denaturing polyacrylamide gel together with a G+A ladder and visualized on X-ray films by autoradiography. **A)** Results from 24 (left) and 48 (right) hours exposure. **B)** Sequence of IRL30 showing the AcuI recognition site, and predicted cleavage site, and the position of transposase cleavage relative to the AcuI cleavage site.

there was some evidence for the appearance of slower migrating bands behind the cleavage products, suggesting that there is a 5' phosphate and that the phosphatase reaction was incomplete.

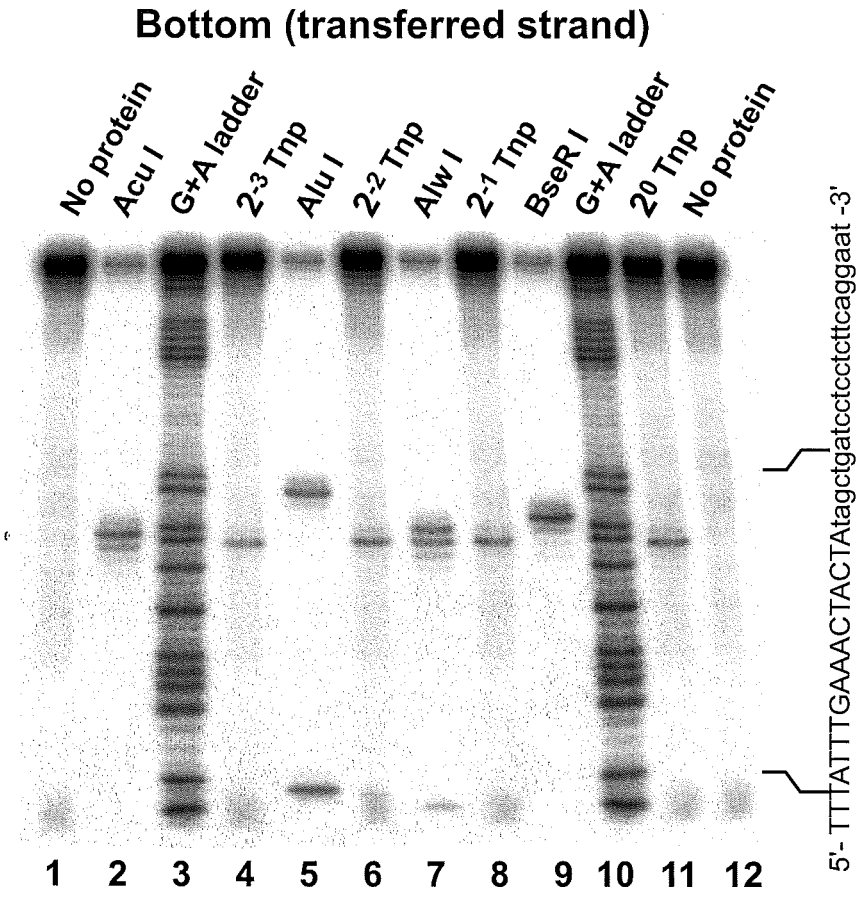
In the experiments reported above, the cleavage sites were determined by comparing to a G + A ladder and AcuI cleaved fragment. Interpretation of these sequencing ladders is complicated by the different ends left by chemical and enzymatic degradation of DNA. To check that the sequencing ladder was being interpreted correctly, another set of cleavage reactions were carried out using a double stranded oligonucleotide containing IRR adjacent to sites for a variety of restriction enzymes that cut at different distances from their recognition sites. The oligonucleotide was designed so that cleavage with these restriction enzymes would provide markers near to the predicted transposase cleavage sites. For some unknown reason the restriction markers ran 1½ nucleotides different from the G+A sequencing ladder instead of the ½ nucleotide difference predicted. However using the restriction markers it can be seen that the major cleavage site on the presumed transferred strand was one nucleotide inside the transposon, with a minor cleavage site just at the transposon end (Figure 4.28A). On the non-transferred strand, there was a cut at the expected position, two nucleotides inside the transposon, and then five more cut sites, at the transposon boundary and 1, 2, 3 and 4 nucleotides outside the transposon (Figure 4.28B). It is not clear why the non-transferred strand of IRR is cleaved with less sequence specificity here than IRL was in the previous assay, though the different sequence of the substrate may be responsible.

In these experiments, a hairpin structure on either the flanking or transposon DNA would be observed as a single stranded DNA longer than the original labelled DNA substrate. No such products were observed on any of the labelled cleavage gels run (data not shown).

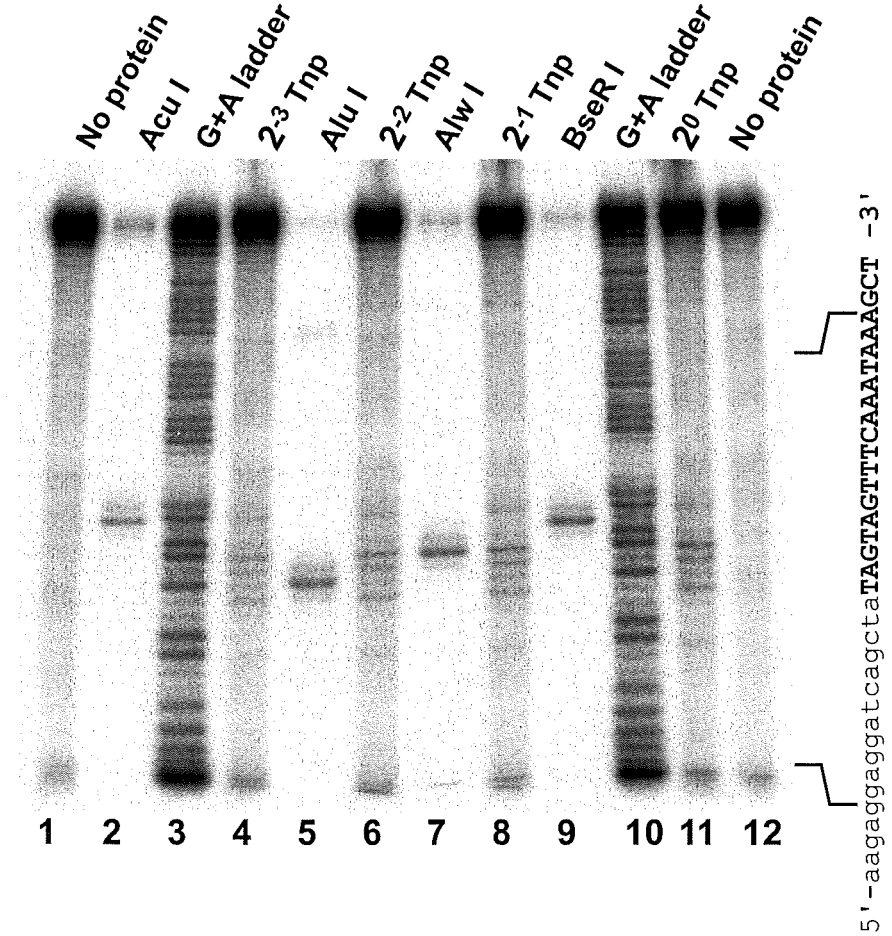
Cleavage of the transferred strand one nucleotide inside the transposon end would lead to loss of transposon sequences. Furthermore, the major cleavage sites are not compatible with the sequence of the ligated vector reported above. If the cleavage sites mapped on linear substrates are correct, it seems that the majority of the cleavage product will not be suitable for subsequent steps of the transposition reaction. Alternatively, the cleavage sites mapped on short linear substrates may not accurately reflect the cleavage sites on supercoiled plasmid substrates with two transposon ends.

Therefore, the cleavage site on a supercoiled plasmid containing a mini-ISY100 was investigated. An *in vitro* transposase cleavage reaction was carried out on pXF153, and the

A)


Bottom (transferred strand)
No protein
Acu I
G+A ladder
2-3 Tnp
Alu I
2-2 Tnp
Alw I
2-1 Tnp
BseR I
G+A ladder
20 Tnp
No protein
1 2 3 4 5 6 7 8 9 10 11 12
5'- TTTATTTGAAACTACTAtagctgatcctcctcttcaggaat -3'

B)


Top (non-transferred strand)
No protein
Acu I
G+A ladder
2-3 Tnp
Alu I
2-2 Tnp
Alw I
2-1 Tnp
BseR I
G+A ladder
20 Tnp
No protein
1 2 3 4 5 6 7 8 9 10 11 12
5'-aagaggaggatcagctaTAGTAGTTTCAAATAAAGCT -3'

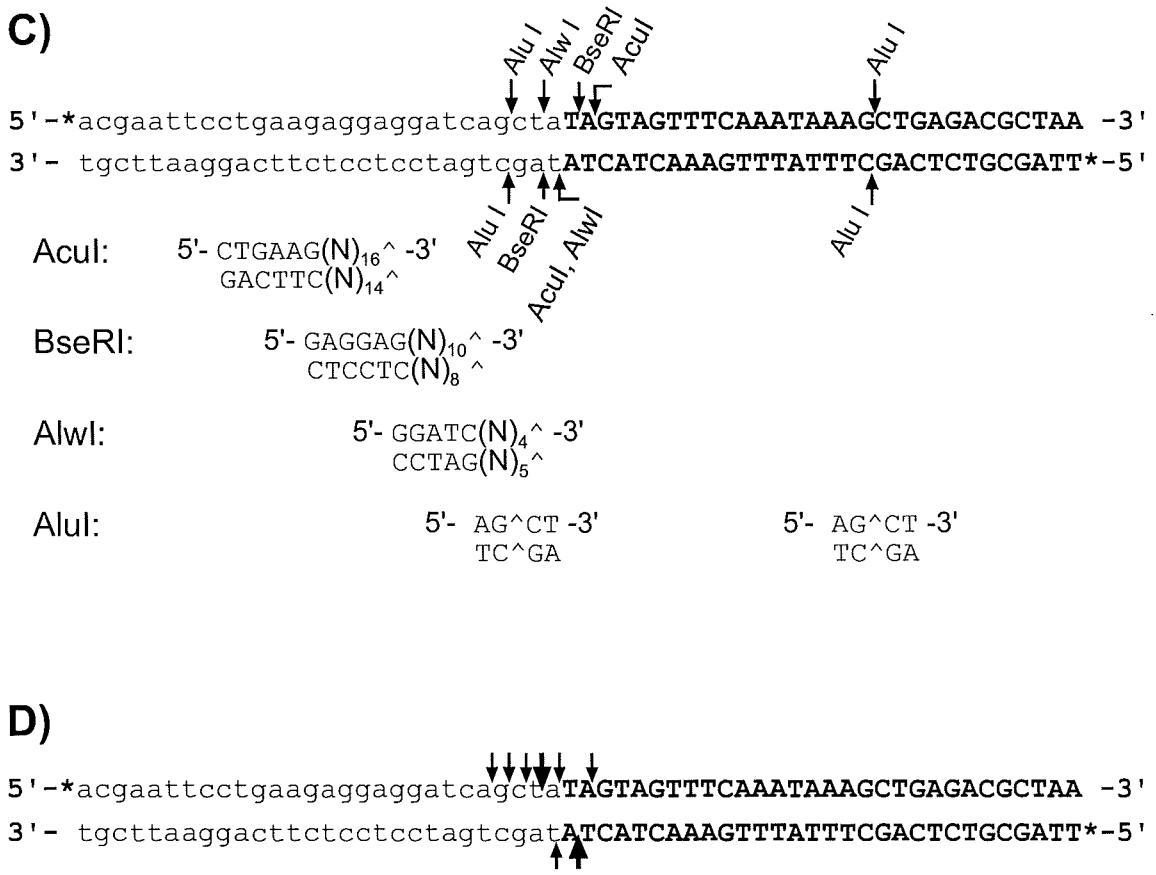


Figure 4.28 Mapping the cleavage sites on IRR *in vitro*. IRR58, an oligonucleotide carrying 30 bp of IRR and multiple restriction sites, was labelled at the 5' end of either the top or bottom strand. Labelled IRR58 was incubated with the indicated concentration of purified His-tagged ISY100 transposase in cleavage buffer together with 20% DMSO for 2 hours at 30°C. The cleavage products were separated on 8% denaturing polyacrylamide gels. Labelled IRR58 was also digested by AcuI (lane 2), AluI (lane 5), AlwI (lane 7) and BseRI (lane 9) and the digests were run as markers. **A)** Bottom strand cleavage. **B)** Top strand cleavage. **C)** Recognition sites and predicted cleavage positions of restriction enzymes. The transposon end sequence is shown in bold uppercase. **D)** Positions of transposase cleavage calculated relative to restriction enzyme cleavage sites. Thick arrows indicate preferred cleavage sites.

DNA fragments representing excised transposon and donor backbone were recovered from an agarose gel. These fragments were tailed using terminal transferase with dGTP or dCTP to add poly-G or poly-C tails to the 3' ends. PCR reactions were then performed using specific vector or outward facing transposon primers and either poly-C or poly-G primers as appropriate. The PCR products were ligated into the TOPO TA cloning vector pCR2.1 and then sequenced to determine the excised transposon fragment and donor backbone 3' ends. Unfortunately, due to a poly-G tract near the ends of the kanamycin cassette, the PCR reactions with the excised transposon fragment failed to give the correct product. Five junctions from the left end of the donor backbone, and nine from the right end were sequenced. These junction sequences reveal the sequence of the 3' end of the donor backbone, which presumably reflects the sequence of the 5' end of the transposon. The majority (4 out of 5 for IRL, 7 out of 9 for IRR) of the inferred cut sites on the non-transferred strands were two nucleotides inside the transposon. The remaining junctions indicated cleavage sites only one nucleotide inside transposon and at the exact transposon end (Figure 4.29).

Another approach was attempted to determine the cleavage site at the 3' end of the transposon. The gel-purified excised mini-ISY100 was tailed with dGTP and terminal transferase and annealed to pUC18 cleaved with PstI and dC-tailed with terminal transferase. After transformation into *E. coli*, 5 colonies carrying insert were selected for sequencing but only one of them gave a readable sequence, showing an intact mini-ISY100 flanked by poly-G (Figure 4.29D). This result suggests that the transferred strands of the transposon were cleaved at the exact transposon end.

These results suggest that the form of the substrates may affect the cleavage specificity of transposase. When supercoiled donor plasmid was used as substrate, the results supported the proposed cleavage sites from the *in vivo* studies of Urasaki *et al.*, (2002): cleavage at the transposon boundaries at the 3' ends, with 2 nucleotide recessed 5' ends. Only a small fraction of molecules were cleaved elsewhere. However, when linear oligonucleotide substrates were used, cleavage at the 3' end was mainly one nucleotide inside the transposon and cleavage at the 5' ends occurred mainly one nucleotide outside the transposon.

4.6 Strand cleavage occurs with a defined order

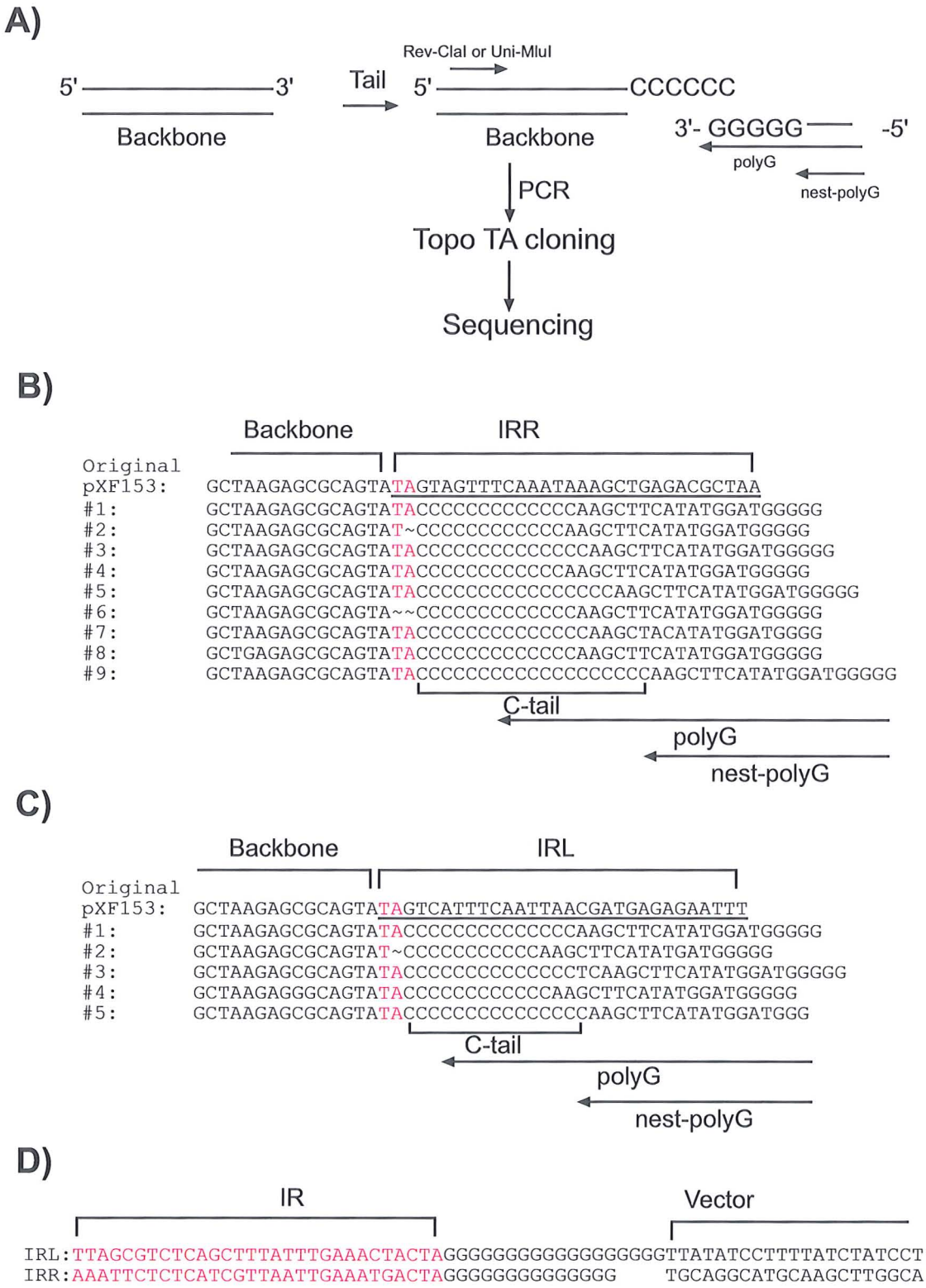


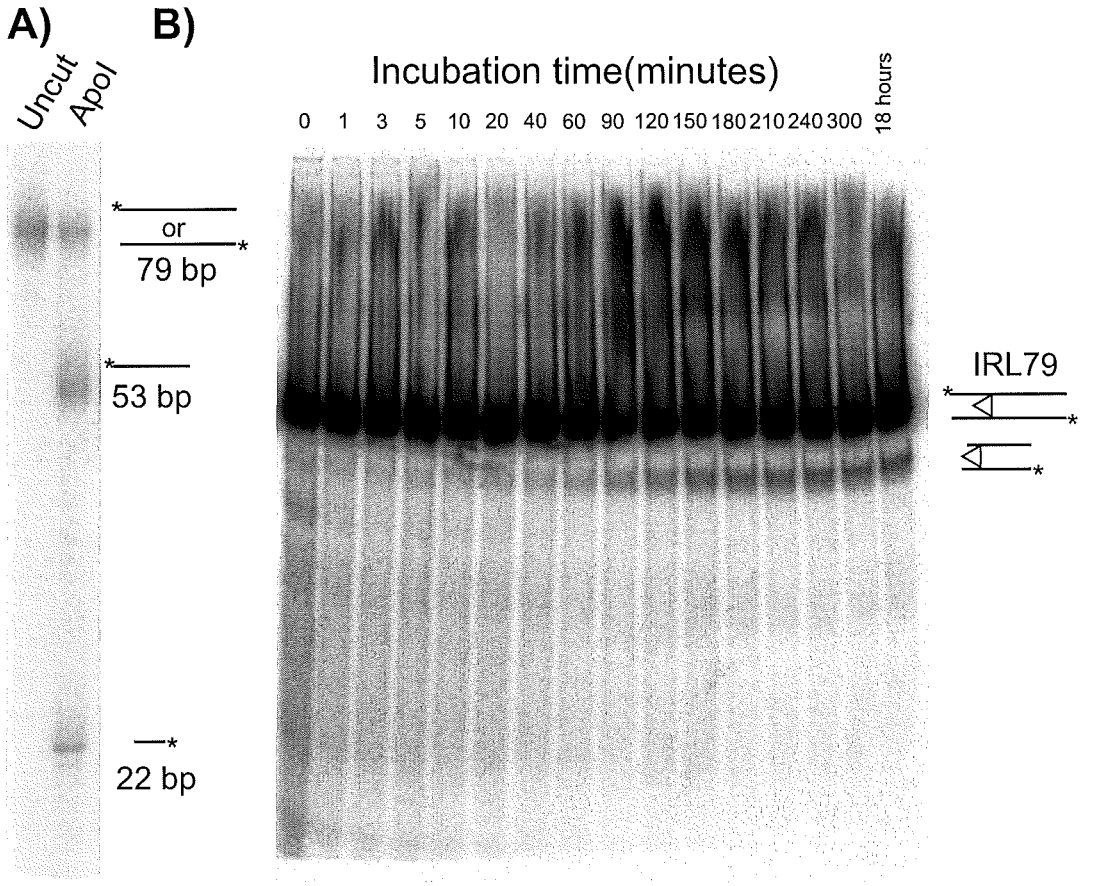
Figure 4.29 Mapping of cleavage sites by C-tailing and PCR. **A)** The donor backbone fragment was tailed with TdT and polyC and then amplified by PCR with a backbone specific primer and a polyG primer (see chapter II). PCR products were cloned by TOPO TA cloning and then sequenced. **B)** 3' end of excised backbone sequence flanking IRR. **C)** 3' end of excised backbone flanking IRL **D)** Junction between a tailed excised mini-ISY100 and C-tailed PstI cleaved pUC18.

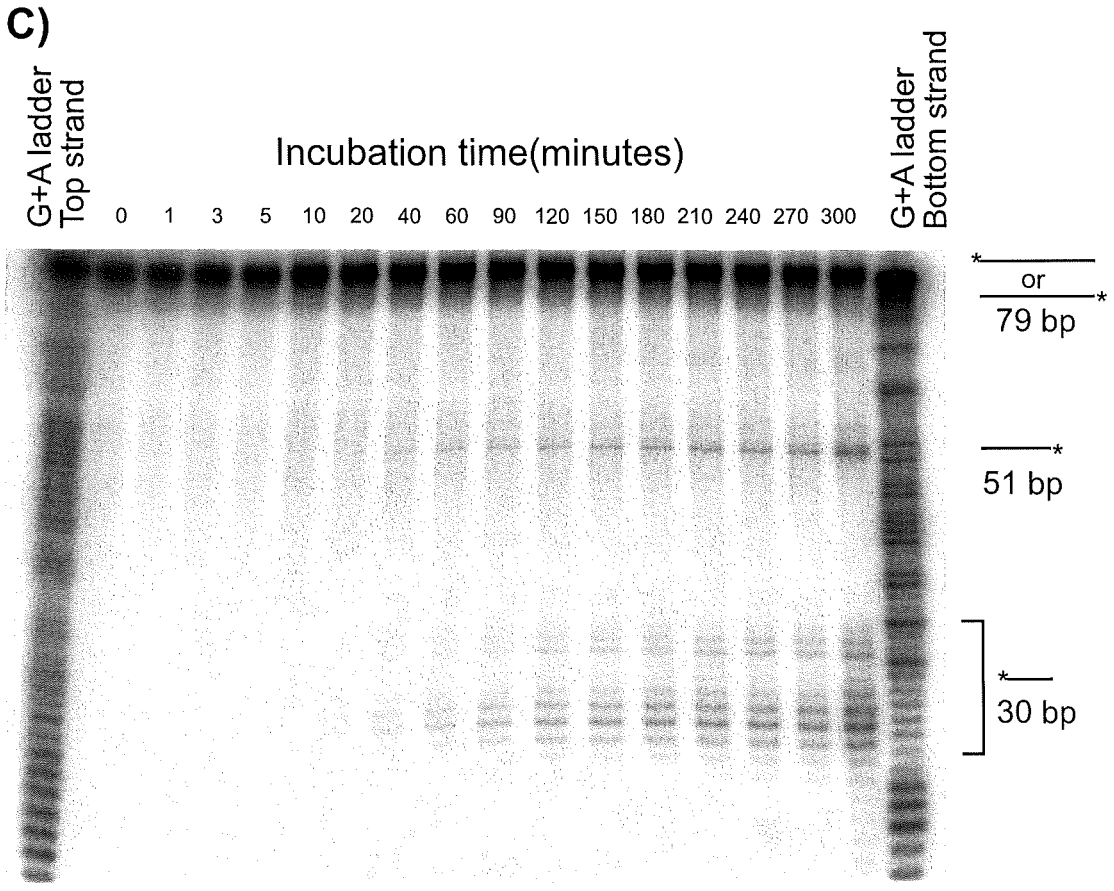
Dawson and Finnegan (2003) reported that cleavage of the two strands at the transposon end by *Mos1* transposase occurs with a defined order. The non-transferred strand is cleaved first to produce a transposon 5' phosphate end, and only then is the transferred strand cleaved to produce a 3' OH. To investigate whether this is also the case for ISY100, the kinetics of top and bottom strand cleavage on IRL79 were measured. IRL79 was 5' end-labelled at both ends, cut with Apo I and run on a denaturing gel acrylamide gel to assess the efficiency of labelling on top and bottom strands. IRL79 was then incubated with transposase in a time course experiment and run on denaturing and non-denaturing gels. Cleavage of the top strand yields a labelled single stranded fragment of approximately 30 nucleotides, whereas cleavage of the bottom strand yields a labelled single stranded fragment of 51 nucleotides that can be seen on the denaturing gel. Cleavage of both strands yields double-strand cleavage products that can be seen on the non-denaturing gel. The gels were quantitated using a phosphoimager, and the percentages of top-, bottom-, and double-strand cleavage were calculated at each time-point (figure 4.30D). The top (non-transferred) strand was cleaved much faster than the bottom strand. Furthermore, the kinetics of double-strand cleavage exactly followed the kinetics of bottom strand cleavage. This shows that bottom strand cleavage only occurs on molecules where the top strand has already been cleaved, and that the non-transferred strand is always cleaved first.

4.7 Integration catalysed by ISY100 transposase.

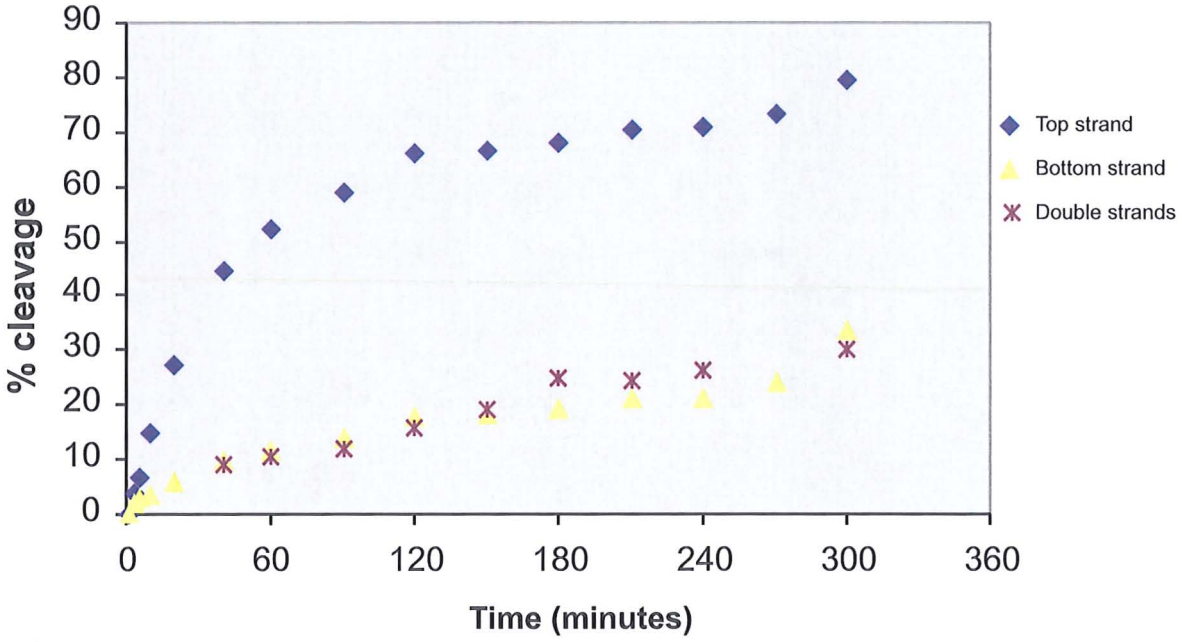
Results presented in the previous section showed that transposase can cleave ISY100 from a donor plasmid. The next step in the transposition reaction is integration of the excised transposon into a new target. This step was studied with double stranded oligonucleotide substrates designed to mimic excised transposon ends. Because contradictory results had been obtained about the exact cleavage sites at ISY100 ends, substrates were designed with a number of different 5' and 3' transposon ends.

The first experiment used an IRL substrate with a 3'OH precisely at the predicted transposon end on the bottom (transferred) strand. The other strand had either a 1 nucleotide protruding 5' end, coinciding with the major cleavage point on oligonucleotide substrates, or a 2 nucleotide recessed 5' end coinciding with the cleavage point on supercoiled substrates *in vivo* and *in vitro* (Figure 4.31A). Double stranded oligonucleotides, 5' end-labelled on either the top or bottom strand, were incubated with pH2 target DNA and transposase in a standard transposition reaction. The products were


A)
B)
Uncut
ApoI
Incubation time(minutes)
0 1 3 5 10 20 40 60 90 120 150 180 210 240 300 18 hours
or
79 bp
53 bp
22 bp
IRL79


C)
G+A ladder
Top strand
Incubation time(minutes)
0 1 3 5 10 20 40 60 90 120 150 180 210 240 270 300
G+A ladder
Bottom strand
or
79 bp
51 bp
30 bp

D)



E)

IRL79:

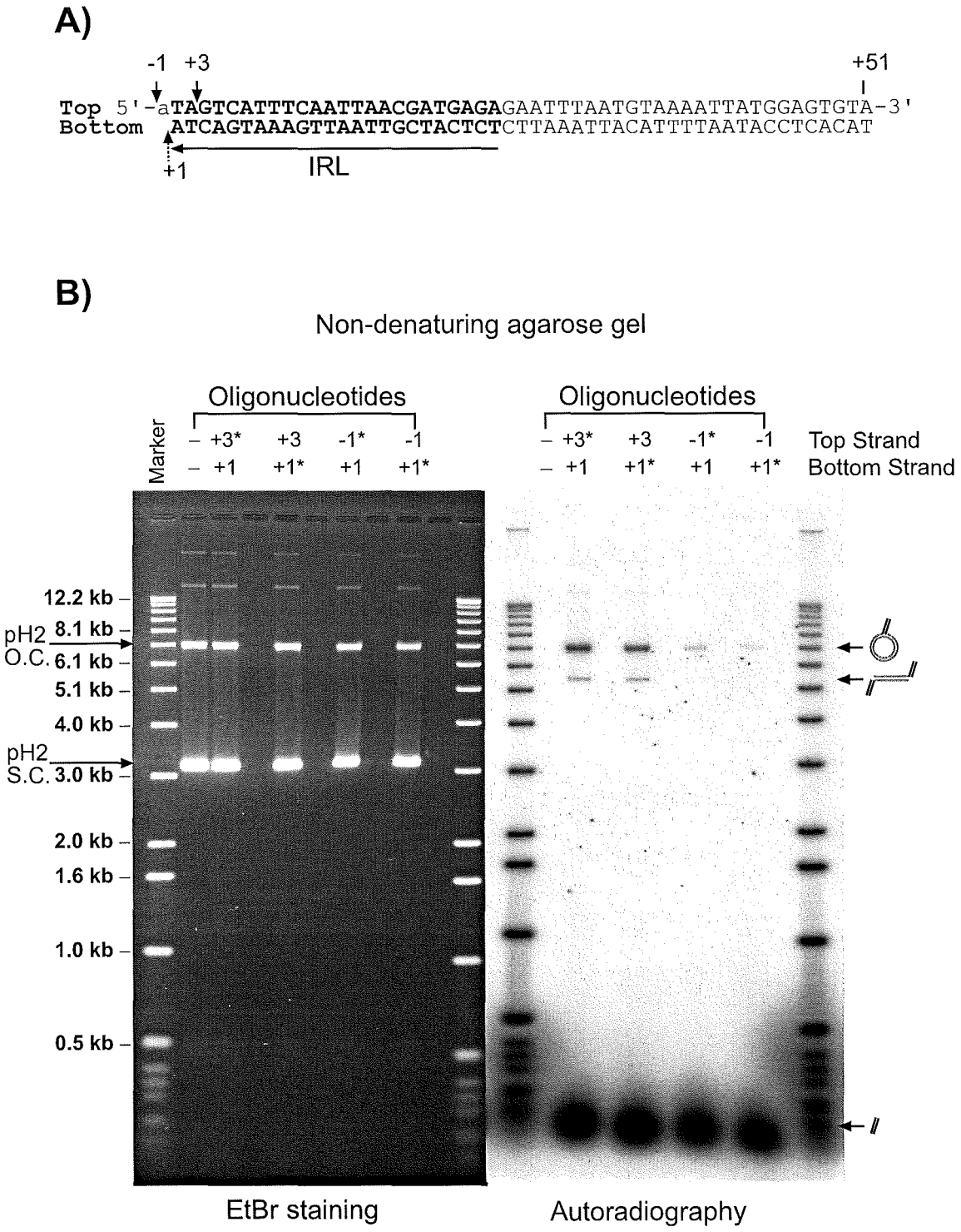
IRL

ApoI

5'-*tagtttctaaagtaaaaattaagaggta**TAGTCATTTCAATTAACGATGAGA**GAATTTAA
atcaaagatttcatttttaattctccat**ATCAGTAAAGTTAATTGCTACTCT**CTTAAATT

TGTAAAATTATGGAGTGTA-3'
ACATTTTAATACCTCACAT*

Figure 4.30 Kinetics of IRL79 cleavage by purified transposase *in vitro*. IRL79 was 5' end-labelled on both strands and was used in a cleavage assay. Samples were withdrawn at the indicated time points and stopped by adding either agarose loading dye or formamide loading dye containing EDTA in excess over the Mg^{++} concentration in the sample. **A)** IRL79 was cleaved with ApoI and separated on an 8% denaturing gel to determine the relative efficiency of labelling of both strands. **B)** Cleavage reactions stopped at the indicated time points were separated on a 10% non-denaturing polyacrylamide TBE gel to measure the extent of double strand cleavage. **C)** Cleavage reactions stopped at the indicated time points were separated on a 8% denaturing polyacrylamide TTE gel to measure the extent of single strand cleavage. **D)** Graph showing the extent of top-strand, bottom-strand and double strand cleavage of IRL79 over time. The phosphorimages shown in B and C were quantitated using the peak area analysis function of the Fuji Image gauge software. Values shown are the percentage of initial substrate cleaved. Top and bottom strand cleavage were corrected for the relative efficiency of labelling on top and bottom strands. **E)** DNA sequence of IRL79. Transposon end sequences are shown in capitals and flanking sequence in lowercase, the 24 bp IRL is in bold. The ApoI recognition site is indicated.


A)
-1 +3
+51
Top 5'-aTAGTCATTTCAATTAACGATGAGAGAATTTAATGTAAAATTATGGAGTGTA-3'
Bottom ATCAGTAAAGTTAATTGCTACTCTCTTAAATTACATTTTAATACCTCACAT
+1
IRL
B)
Non-denaturing agarose gel
Oligonucleotides
Oligonucleotides
Marker
– +3* +3 -1* -1
– +1 +1* +1 +1*
Top Strand
Bottom Strand
12.2 kb
8.1 kb
pH2 O.C.
6.1 kb
5.1 kb
4.0 kb
pH2 S.C.
3.0 kb
2.0 kb
1.6 kb
1.0 kb
0.5 kb
EtBr staining
Autoradiography

C)

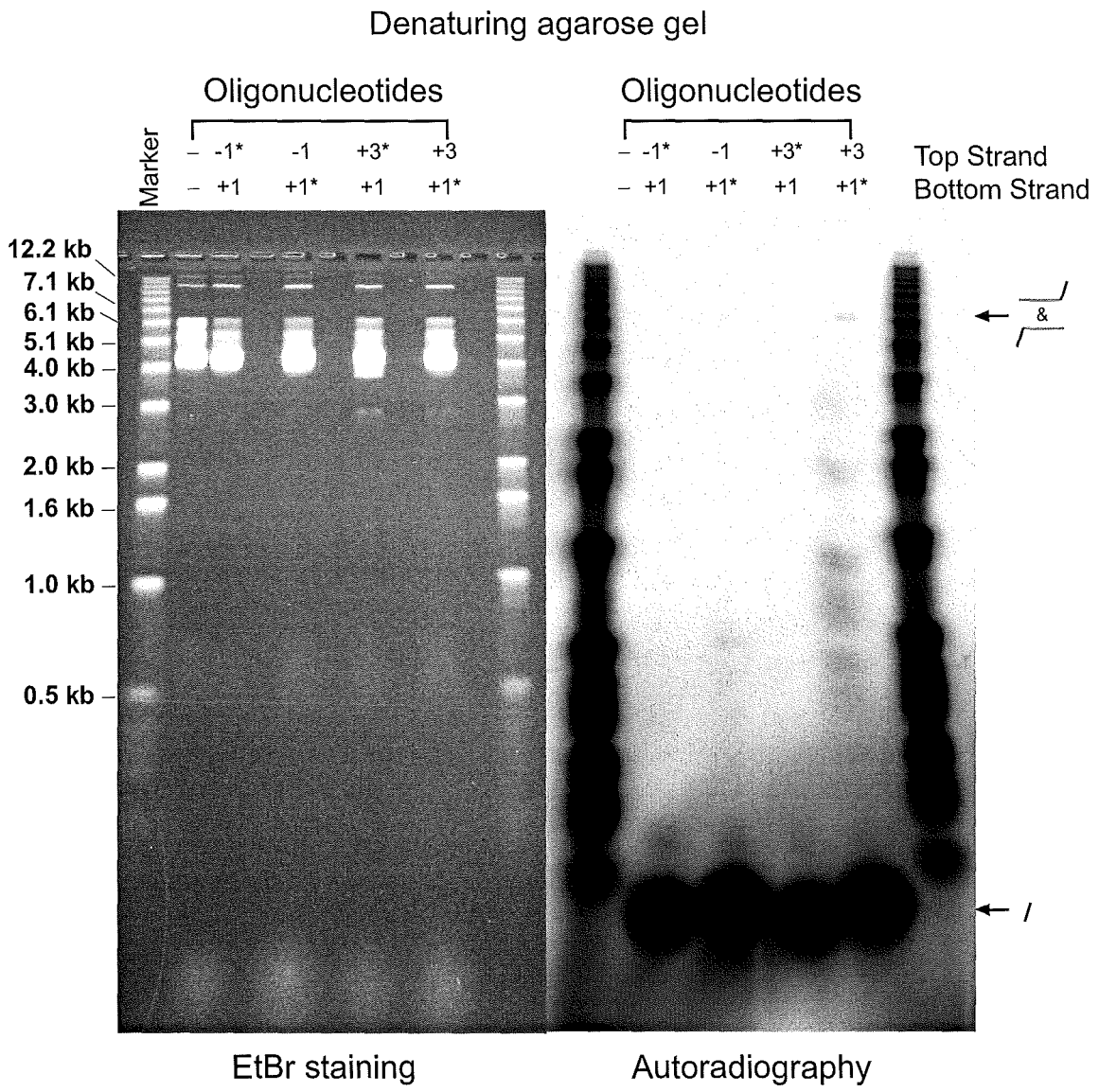


Figure 4.31 Strand transfer by ISY100 transposase. A) Sequence of synthetic oligonucleotides. ISY100 sequence is shown in uppercase and the 24 bp IRL is shown in bold. A bottom strand oligonucleotide consisting of nucleotides +1 to +51 was annealed to a top strand oligonucleotide consisting of either nucleotides -1 to +51 or +3 to +51. **B** and **C)** Integration reaction. Double strand oligonucleotides (+3 top / +1 bottom or -1 top / +1 bottom) were 5' end-labelled on only one strand, as indicated by the asterisk (*), and used in an integration assay with pH2 as target (see chapter II). Reaction products were separated on a 1.2% non-denaturing agarose gel (B) and a 1.2% denaturing alkaline agarose gel (C). The gels were stained by EtBr and then visualised by UV transillumination (left panel), and then dried and visualised by autoradiography (right panel).

then run on a TAE agarose gel and visualized by both ethidium bromide staining and autoradiography (Figure 4.31B). The substrate with a recessed 5' end produced a major product co-migrating with the open circular form of pH2, and a minor product migrating at approximately the same size as linear pH2. The major band is most likely to be pH2 with a single copy of the oligonucleotide transposon end integrated at a TA dinucleotide. Integration will introduce a single nick, yielding a non-supercoiled circular product with a small protruding tail that will migrate at almost exactly the same position as open circle pH2. The minor product is thought to come from concerted integration of two transposon ends on both strands at approximately the same position, yielding a linear molecule with a labelled oligonucleotide covalently attached at both ends. The oligonucleotide substrate with a protruding 5' end gave much less product, and only the one ended integration product was detectable (Figure 4.31B). The position of the label, on either the top or bottom strand, made no difference to the products observed on this non-denaturing gel system.

The same reaction products were then run out on a denaturing gel to see if the labelled oligonucleotides became covalently attached to the pH2 target DNA (Figure 4.31C). A radioactively labelled product of the correct size to be the oligonucleotide covalently linked to a single strand of pH2 was observed only when the bottom strand of the oligonucleotide was 5' end-labelled. This result therefore shows that the transposon 3' end becomes covalently linked to the target DNA, as has been found previously for all other DDE transposons.

A major cleavage was observed on oligonucleotide substrates one nucleotide inside the transposon on the transferred strand (Figure 4.28). To see it transposon ends cleaved at this position can integrate into a target, or are dead-end products, oligonucleotide substrates were made to mimic these cleavage products. Top and bottom strand-labelled substrates were made with 3' ends 1 nucleotide inside the transposon end, and 5' ends either 1 nucleotide outside or 2 nucleotides inside the transposon end. When these were incubated together with transposase and pH2, no strand transfer products were seen (Figure 4.32). Thus it seems that cleavage products missing the 3' terminal nucleotide of ISY100 are dead-end products that cannot be integrated into a target plasmid.

The divalent ion requirements of the strand transfer reaction were also tested (Figure 4.33). Integration was dependent on the presence of divalent cations, and had similar requirements to the cleavage reaction, occurring at 5-15 mM Mg^{2+} and 0.5-2.0 mM Mn^{2+}.

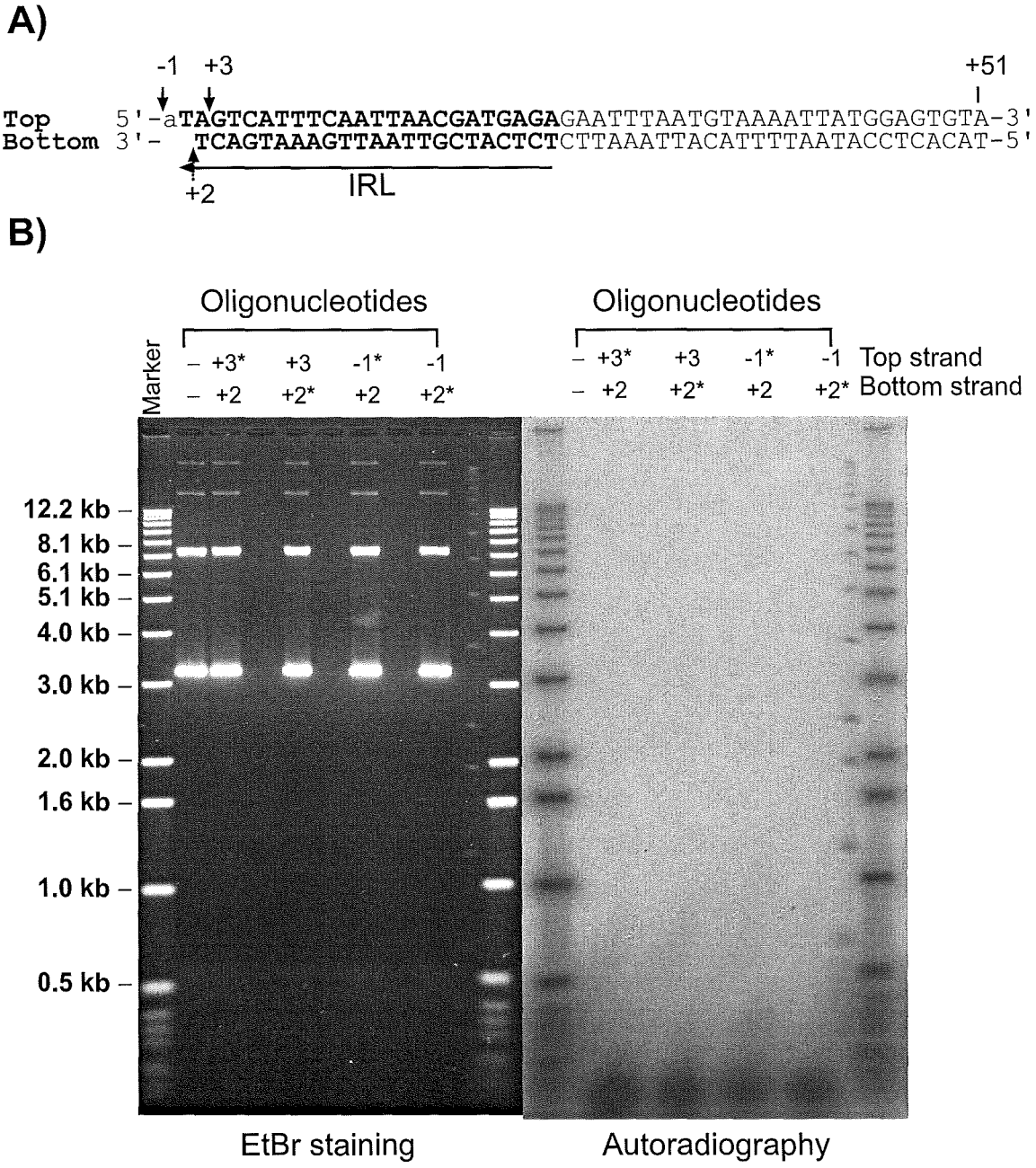
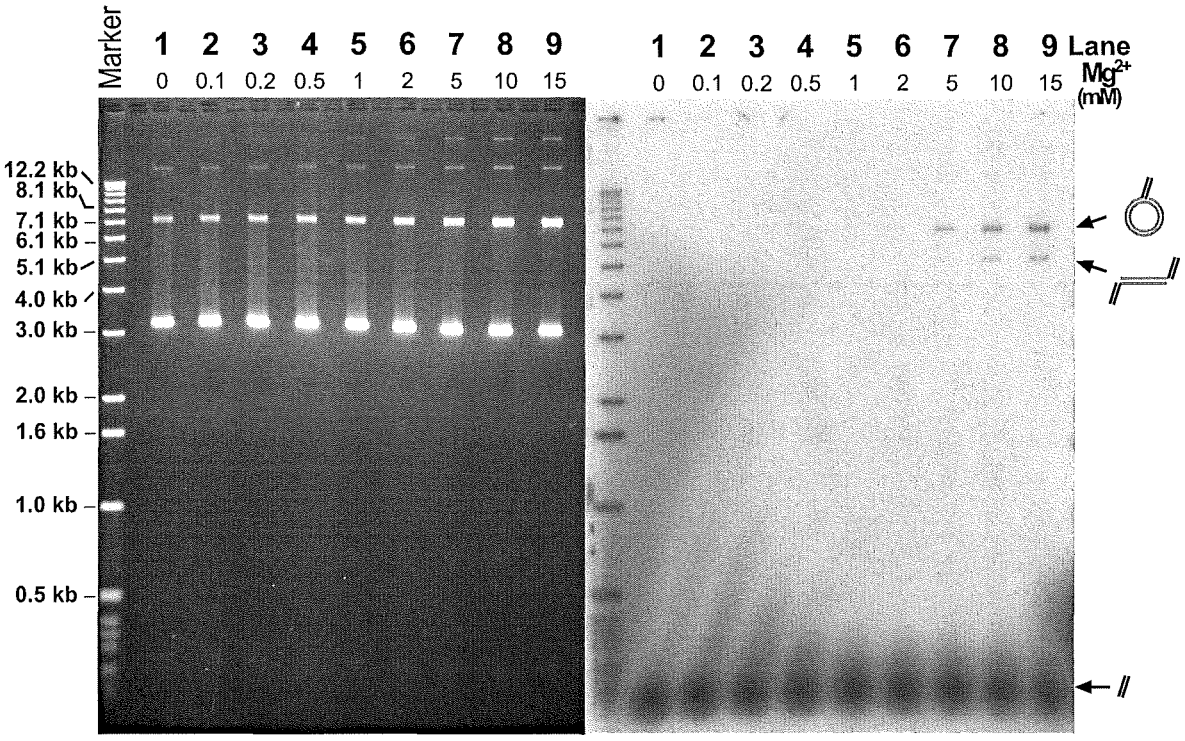


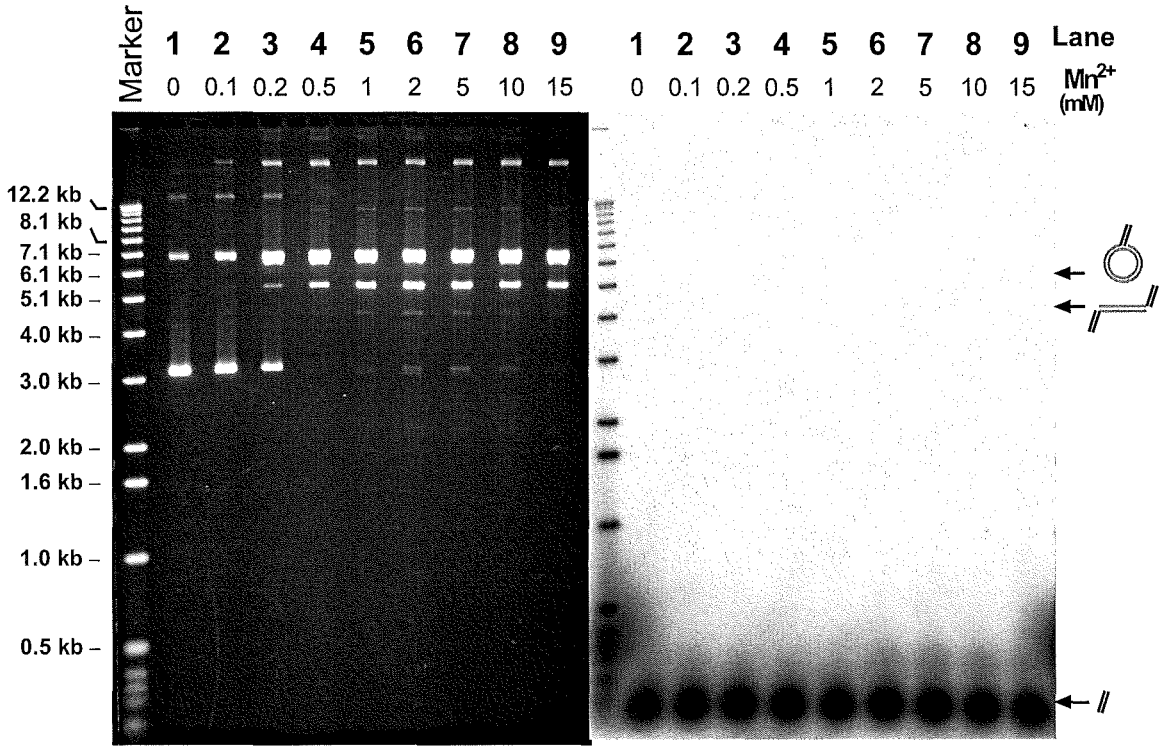
Figure 4.32 Integration assays with truncated bottom strand substrates. A) Sequence of synthetic oligonucleotides. ISY100 sequence is shown in uppercase and the 24 bp IRL is shown in bold. A bottom strand oligonucleotide consisting of nucleotides +2 to +51 of the sequence shown was annealed to a top strand consisting of either nucleotides -1 to +51 or +3 to +51. **B)** Integration reaction. Double strand oligonucleotides were 5' end-labelled on only one strand as indicated by asterisk (*) and used in an integration assay with pH2 as target (see chapter II). Reaction products were separated on a 1.2% non-denaturing agarose gel. The gel was stained by EtBr and then visualised by UV transillumination (left panel), and then dried and visualised by autoradiography (right panel).

A)



EtBr staining Autoradiography

B)



EtBr staining Autoradiography

C)

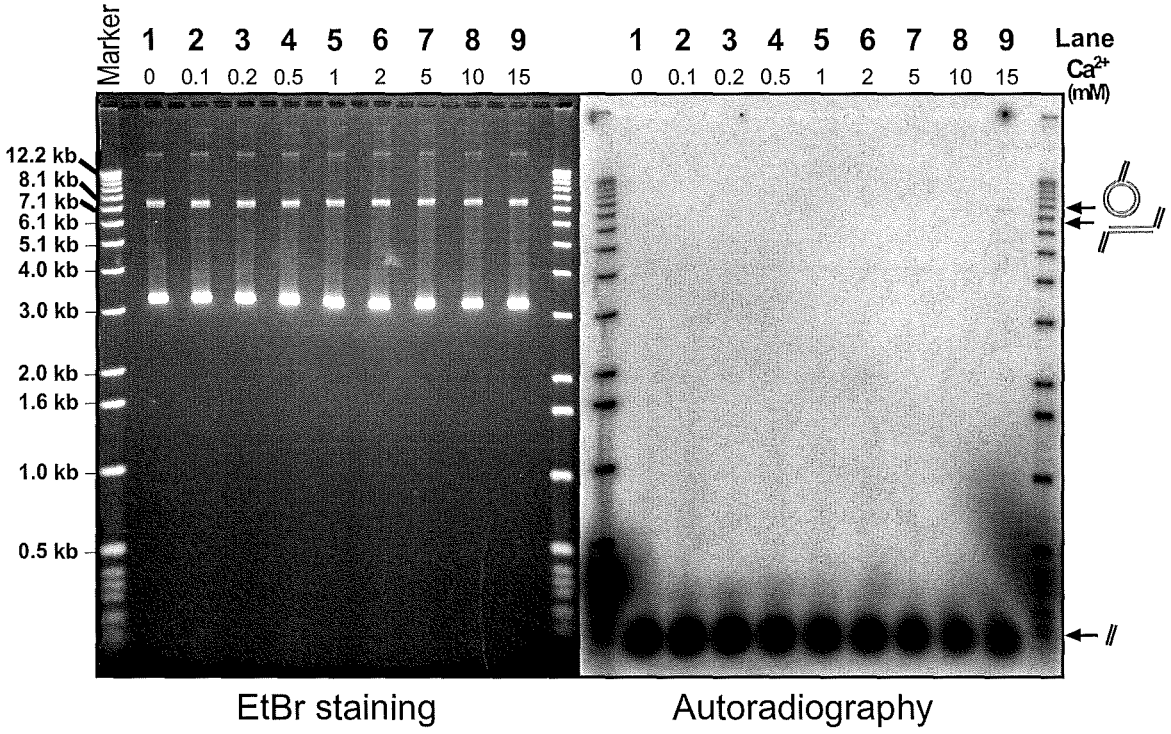


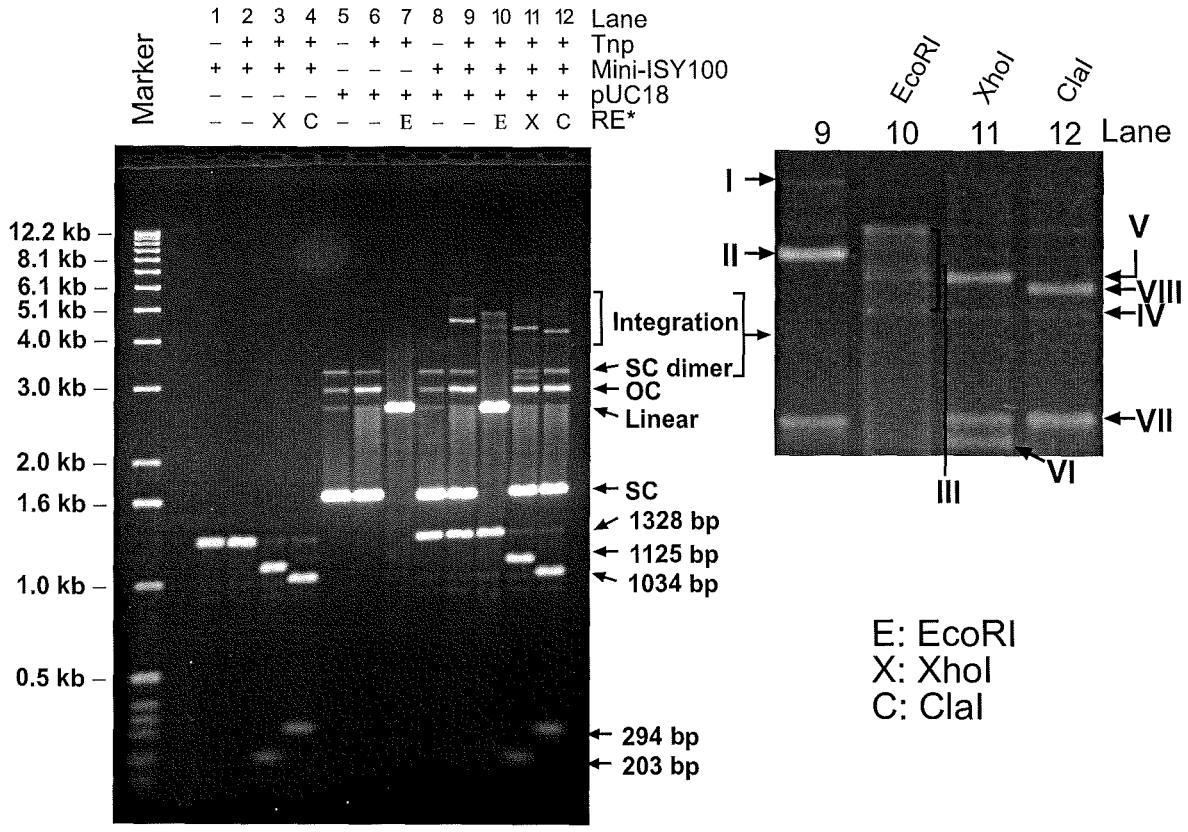
Figure 4.33 Influence of divalent metal ions on the integration. 20 pmoles of radiolabelled double stranded IRL oligonucleotide, consisting of nucleotides +1 to +51 on the bottom (transferred) strand and nucleotides +3 to +51 on the top (non-transferred) strand (see Figure 4.31A), was used in an integration assay (see chapter II) in the presence of various concentrations of Mg^{2+}, Mn^{2+} or Ca^{2+}. 0.15 pmole of supercoiled plasmid pH2 was used as target in each reaction. Reaction products were separated on 1.2% non-denaturing agarose gels. The gels were stained by EtBr and then visualised by UV transillumination (left panel) and then dried and visualised by autoradiography (right panel). Concentrations of divalent ions used are indicated.

However, unlike the cleavage reaction, a small amount of integration could be detected in the presence of 15 mM Ca^{2+}.

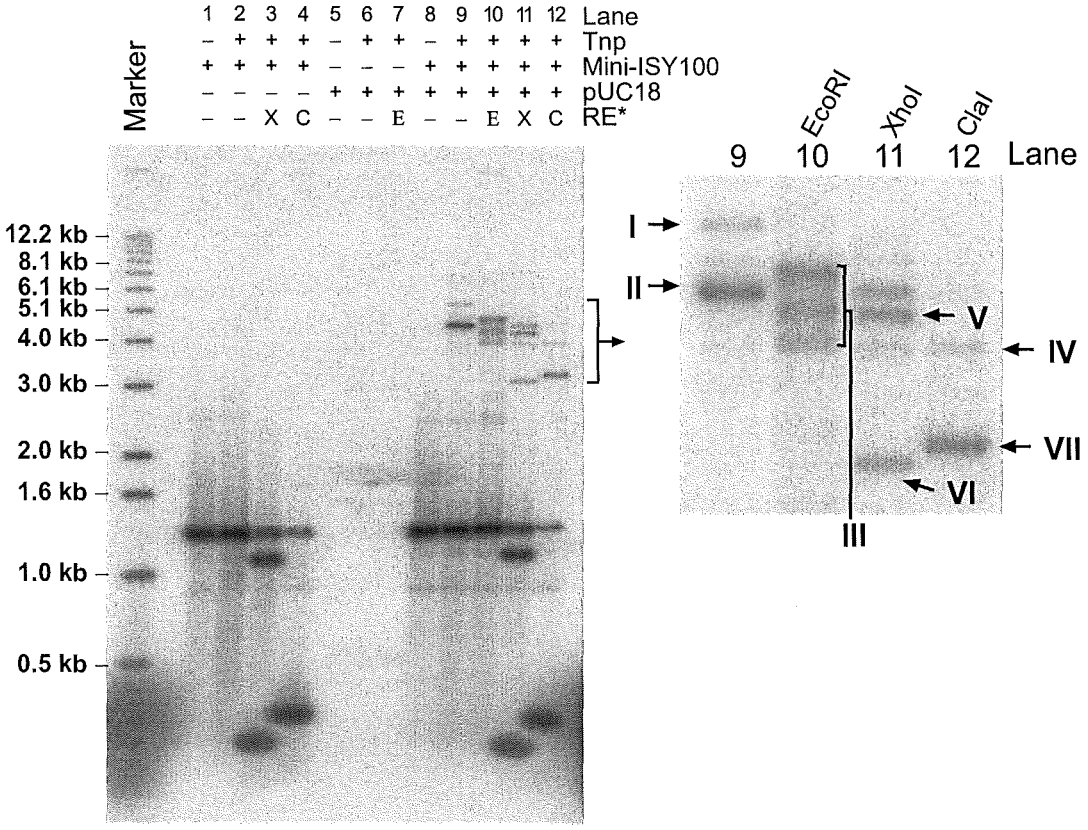
A slightly different integration assay was carried out using a linear 1.3 kb excised transposon fragment with ISY100 left and right ends. This fragment was radioactively labelled at an internal XhoI site, and then cut out from pXF153 using AcuI, so that it had the exact same ends produced by transposase: 3'OH ends precisely at the transposon ends and 2 nucleotide recessed 5' phosphates. This fragment was incubated with purified transposase and pUC18 plasmid DNA as target. This produced one new major radioactive product (Figure 4.34C, band II), running significantly slower than nicked pUC18, and a less abundant even slower product (band I). When cleaved with XhoI, this produced two labelled products, one migrating close to pUC18 open circle (band VI) and the other (band V) migrating closer to band II. When cleaved with ClaI, there were two major bands (VII and VIII) only one of which was radioactive. This is fully consistent with integration of only one end of the transposon into pUC18, with the left and right ends inserted at roughly equal frequency (Figure 4.34C). When this product was cleaved with EcoRI, which cleaves once in pUC18 but not in the transposon fragment, a series of bands was produced (III), presumably containing one-ended insertions at different sites in pUC18. A small amount of product was seen migrating at the 4.0 kb position (band IV) in the EcoRI, ClaI and XhoI digestions. This is consistent with concerted integration of both ends of ISY100 into a single site in pUC18 to produce a nicked circular insertion product that can be linearised with a restriction enzyme that cuts once either in pUC18 or in the transposon fragment.

If this is the case, then these products could be recovered as Kan^r Amp^r colonies by transformation into *E. coli*. To test this idea, various amounts of the purified pXF153 AcuI excised mini-transposon fragment were incubated with transposase and supercoiled pH2 as target, and the resulting products were run on an agarose gel (Figure 4.35 B). A band that could be one-end integration was seen in increasing amounts as the amount of mini-transposon fragment was increased. When the products were electroporated into DS964, kan^r colonies were obtained with increasing frequency as the amount of mini-transposon fragment was increased, and over 1 per 1000 of the target plasmids had received a transposon insertion at the highest mini-transposon concentration. Individual kan^r colonies were picked for plasmid isolation, restriction mapping and DNA sequencing, demonstrating that they represented genuine transposon insertions into TA dinucleotides in pH2 (see chapter V for more details).

A)



EtBr staining



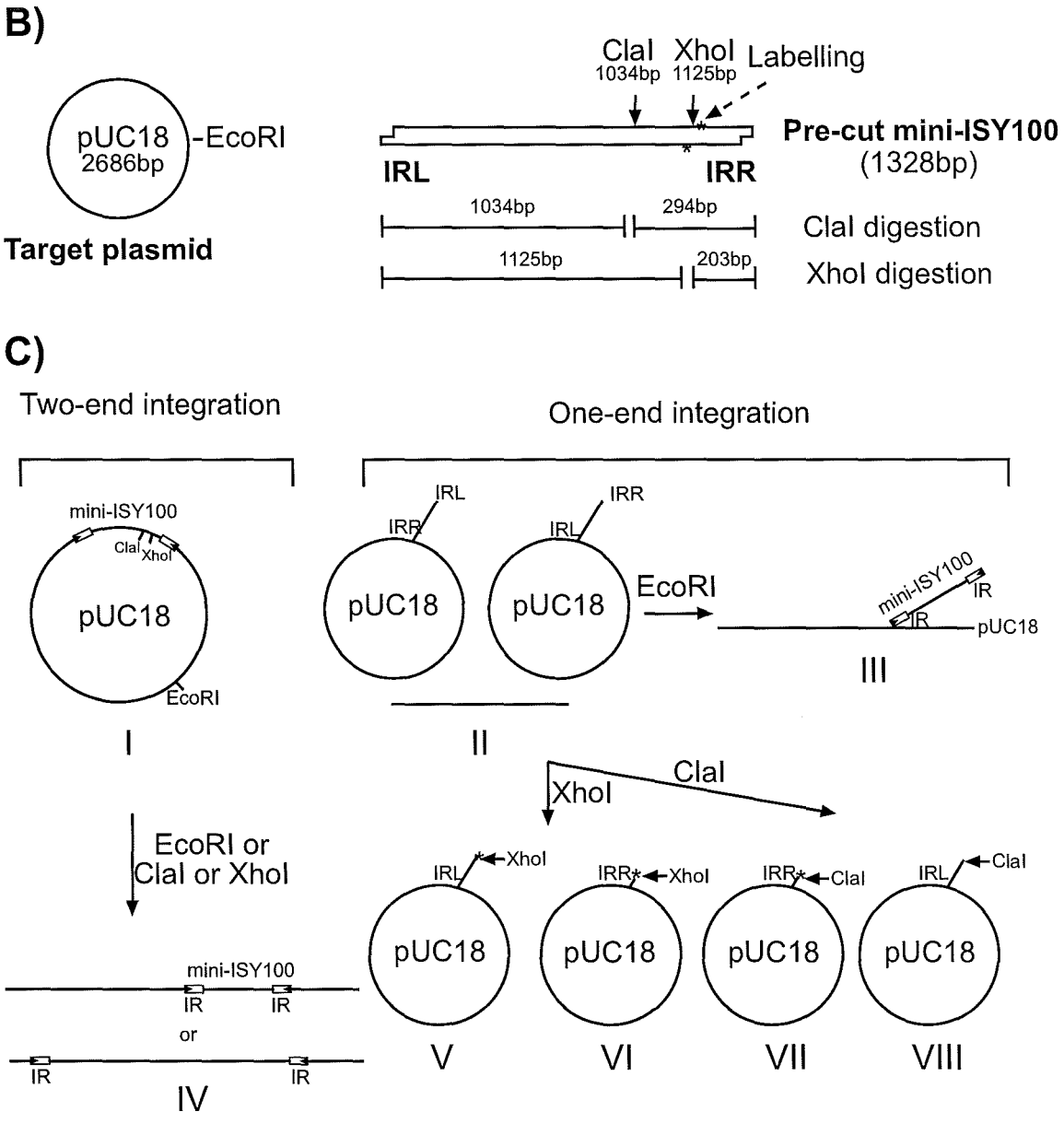


Figure 4.34 Integration of pre-cut mini-ISY100 into target DNA. pXF153 was cleaved with XhoI, end-labelled with ^{32}P using T4 polynucleotide kinase and then re-ligated. This was then cleaved with AcuI to release a 1.3 kb mini-ISY100, cleaved precisely at the transposase cleavage sites, internally labelled at the XhoI site. 0.37 pmole of labelled precut mini-ISY100 was incubated with 0.92 pmole of pUC18 together with 1.31 pmoles purified His-tagged transposase in a standard transposition buffer for 3 hours at 30°C. The reaction was stopped, proteins were removed and the buffer was changed using a Qiagen gel extraction kit. The products were then digested with either EcoRI, XhoI or ClaI. **A)** The products were separated on a 1.2% agarose gel and the gel was stained with EtBr and visualised by UV transillumination (top gel), and then dried and visualised by autoradiography (bottom gel). The region of the gel containing integration products is shown expanded to the right of each gel. Bands of interest are indicated with roman numerals (I - VIII). **B)** Diagram of substrate DNAs (precut mini-ISY100 and pUC18 target) used in the assay. **C)** Diagram showing the products of integration and restriction digestion. Products are indicated with roman numerals as in A).

A)

Reaction	Pre-cut mini-ISY100 (pmole)	pH2 (pmole)	Tnp (pmole)	Amp^r ($\times 10^6$)	Kan^r ($\times 10^6$)	Transposition frequency ($\times 10^{-4}$)
#1	0.60	–	–	0	5	0
#2	–	0.31	1.31	2.77	0	0
#3	0.075				1,618	5.84
#4	0.15				2,508	9.05
#5	0.30				3,442	12.43
#6	0.60				>4,000	>14.44

B)

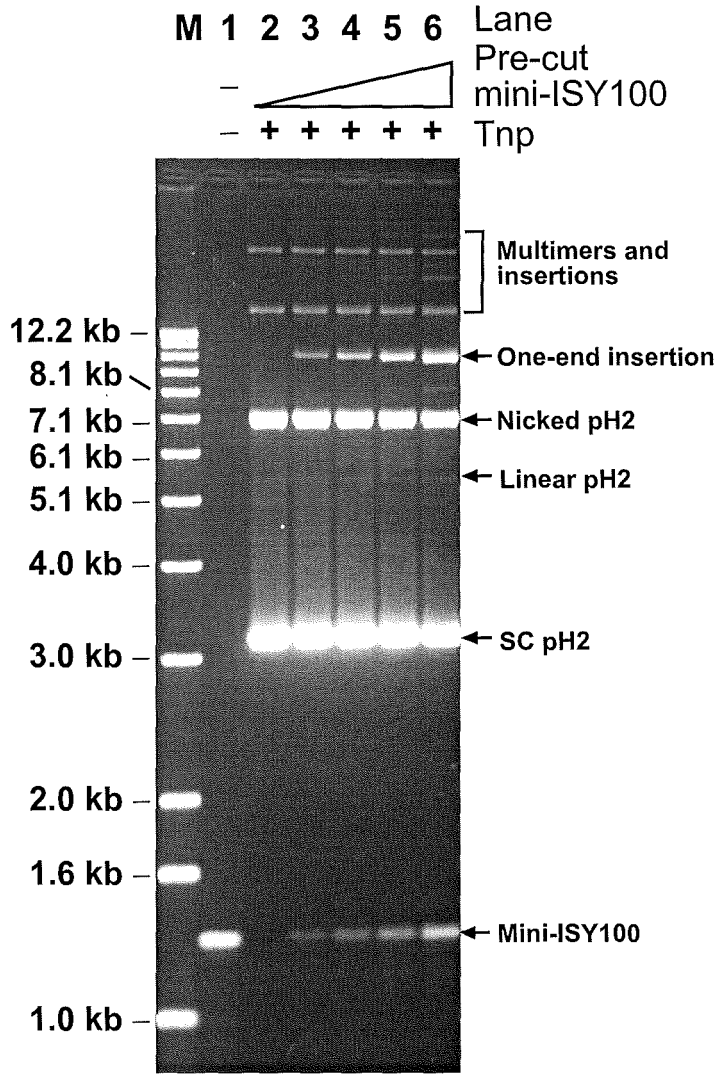
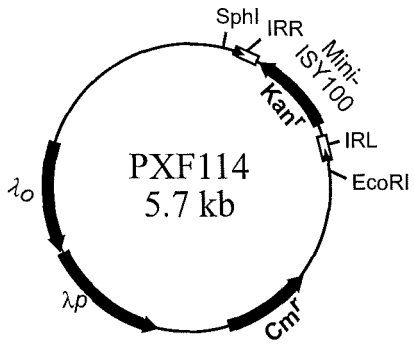
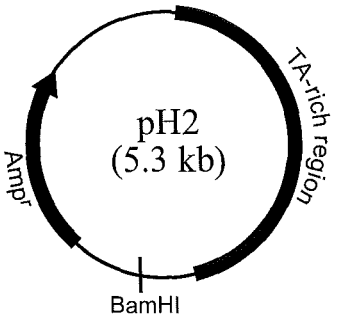


Figure 4.35 *In vitro* transposition using pre-cut mini-ISY100 as donor. *In vitro* transposition was performed by incubating various amount of 1.3 kb pXF153/AcuI digestion product, a mini-ISY100 with 2 nucleotide recessed 5' ends, with supercoiled target plasmid pH2 and purified transposase in a standard transposition buffer for 4 hours at 30°C. **A)** Transposition frequencies measured by electroporation into DS964. **B)** Analysis of products from the *in vitro* transposition on a 1.2% agarose gel.

A)



B)



C)

Reaction	Amp^r ($\times 10^6$)	Kan^r Cm^s	Transposition frequency ($\times 10^{-3}$)	
			Individual	Average
#1	1.21	1700	1.49	1.72±0.33
#2	1.30	2530	1.95	

D)

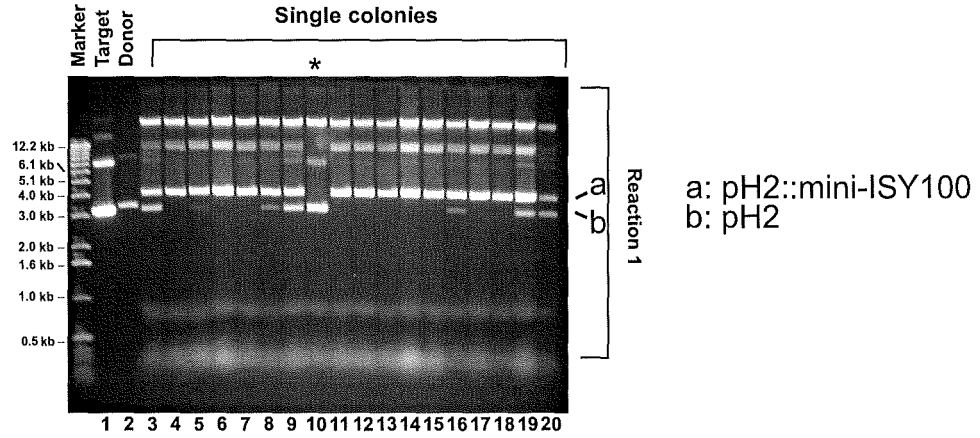


Figure 4.36 ***In vitro*** **transposition assay. A)** Map of λ-dv donor plasmid pXF114. **B)** Map of target plasmid pH2. **C)** Transposition frequency of two independent *in vitro* transposition assays employing purified transposase, pXF114 and pH2 in a buffer containing 10 mM Mg_2Cl (see Chapter II for details) **D)** Single colony gel analysis of kanamycin resistant colonies from *in vitro* transposition 1. 18 DS964 kanamycin resistant transformant colonies, of which one was also chloramphenicol resistant (labelled with *), were picked from *in vitro* transposition reaction 1. Plasmid DNA was analysed on a 1.2% agarose single colony gel. 1μg of target plasmid pH2 (lane 1) and 0.28 μg of donor plasmid pXF114 (lane 2) were loaded on the gel as markers.

Finally, transposase was tested for its ability to catalyse a complete transposition reaction. Purified transposase was mixed with a supercoiled donor plasmid pXF114 (Figure 4.36A) and a supercoiled target plasmid (pH2) (Figure 4.36B) in a standard transposition assay. The DNA was then ethanol precipitated and electroporated into a λ-lysogen strain (DS964). The donor plasmid carries a kanamycin resistance gene flanked by 30 bp ISY100 IRL and IRR sequences, on a λ-dv vector. If transposition occurs, the mini-ISY100 will jump from the donor plasmid into the target plasmid, giving DNA molecules that can be detected as Kan^r transformant colonies in DS964. The donor plasmid cannot replicate in DS964 and so there should be no other way of obtaining Kan^r colonies.

As shown in Figure 4.36C, such DNA molecules were detected with a frequency of approximately 1.5 - 2 Kan^r colonies per 10^3 Amp^r colonies. The majority of Kan^r colonies obtained were sensitive to chloramphenicol, consistent with loss of the donor backbone, as expected for a proper transposition event. Gel electrophoresis of plasmids obtained from Kan^r Cm^s colonies was consistent with insertion of the mini-transposon into the target plasmid pH2 (Figure 4.36D). Subsequent sequencing results showed that transposition occurred exactly at the ISY100 ends with duplication of the target TA (see chapter V for details).

4.8 Discussion and future work

This chapter reports the purification of full-length His-tagged transposase from the ISY100 insertion sequence of *Synechocystis* sp. PCC6803. The purified transposase was active *in vitro* and catalysed all of the steps in a "cut and paste" transposition reaction. The fact that purified transposase was active *in vitro* demonstrates that no host factors are required for ISY100 transposition (although it is possible that host factors could contribute to the efficiency of transposition). The fact that ISY100 transposase is active *in vitro* without host factors suggests that it could be active in cultured mammalian cells, or in model organisms. ISY100 might therefore have applications in biotechnology as has already been shown for other members of the Tc1/*mariner* family of transposons. The question of whether the bacterial ISY100 transposase can localise to the nucleus of a mammalian cell, where it would be required for activity, has yet to be addressed.

A number of steps in the transposition reaction were studied in detail. An early step in transposition is binding of transposase to the transposon ends. It was shown that the N-terminal 95 amino acids of ISY100 transposase, containing two helix-turn-helix motifs,

binds specifically to the left and right ISY100 inverted repeats. N-terminal fragments of transposase carrying only one helix-turn-helix bound to transposon inverted repeats, but the binding appeared to have lower sequence specificity than protein fragments with two helix-turn-helix motifs. A ladder of complexes was found as the protein concentration increased (data not shown), suggesting binding to multiple site in the DNA, and no footprint was observed.

The footprints obtained with Tnp_{1-95} and Tnp_{1-110} were identical to the footprint obtained with full-length transposase. This suggests that the catalytic domain of transposase is not interacting with the transposon end in the complexes obtained. Perhaps the catalytic domain only interacts stably with the transposon end in a paired end complex, which was not observed in the conditions used here. Alternatively, it is possible that the complexes observed with full-length transposase come from N-terminal proteolytic transposase fragments that may be present in the purified protein. This seems unlikely, since transposase was purified with a C-terminal His-tag. However it could be tested by cutting out the complex with full-length transposase from a binding gel, separating it by SDS-PAGE, doing a western-blot and looking for full-length transposase in the complexes with an antibody directed against the C-terminal His-tag.

The findings that Tnp_{1-95} is sufficient for recognition and binding transposon ends might have implications for the regulation of ISY100 transposition in the cyanobacterium *Synechocystis* sp. PCC6803. 25 copies of ISY100 were reported on the genome and plasmids of PCC6803 (http://www.kazusa.or.jp/cyano/Synechocystis/). Ten copies on the genome encode the full length ISY100 transposase, with only minor sequence variation. However, nine copies on the genome and three copies on the plasmids contain frameshift mutations that will lead to the production of proteins containing only the first 101 amino acids of transposase in a protein of 119 amino acids. A further copy contains a frame shift that will produce a 94 amino acid transposase derivative. These truncated transposase proteins are predicted to compete for binding with the full-length transposase and act as negative regulators of transposition.

Although there was no evidence from footprinting experiments for interaction between the catalytic domain of transposase and the transposon ends, the catalytic domain must interact with the transposon ends in subsequent cleavage and strand transfer reactions. On supercoiled substrates, cleavage was exactly at the transposon end on one strand to produce a 3'-OH, and two nucleotides inside on the other strand to produce a 5' phosphate. On linear substrate, the specificity of cleavage was reduced and cleavage

occurred at a number of nearby positions on both strands. Correctly cleaved 3'-OH ends could be covalently attached to the target DNA in the strand transfer reaction. Although apparent concerted insertion of two transposon ends into opposite strands of a single target site could be observed, and could be recovered as transposon insertions after electroporation in to *E. coli*, single end insertions were much more common. Further studies are needed to investigate whether the C-terminal His-tag interferes with the integration reaction, or whether a host factor helps bring about concerted integration.

Some of the most interesting findings presented in this chapter concern the order of events in the strand cleavage reaction. To excise fully from its donor site, both strands of the transposon must be cleaved at both ends. On a small linear substrate, the non-transferred strand was cleaved much faster than the transferred strand (Figure 4.30). Furthermore, the transferred strand was only cleaved on molecules that had already been cleaved on the non-transferred strand (Figure 4.30). Thus it appears that there is a strict order of strand cleavage at the transposon ends. Consistent with this, nicked molecules accumulated from supercoiled substrates much more quickly than double strand cleavage products, and later work done by Sean Colloms and Victoria Moffat has shown that these nicked molecules contain nicks only on the non-transferred strands.

There is also evidence suggesting that cleavage of the transferred strands requires synapsis or some other communication between the two transposon ends, whereas cleavage of the non-transferred strands does not. Plasmids containing just one transposon end were nicked efficiently, but double strand cleavage was much less efficient than on plasmids with two ends. Intermolecular synapsis is much more likely than intramolecular synapsis at low DNA concentrations, and may be favoured by DNA supercoiling, so this result suggests that synapsis is required for second strand cleavage.

Mutations in the flanking TA at just one end blocked double-strand cleavage at the unchanged end as well as at the mutated end, but nicking was not affected. This again supports the idea that synapsis is not required for cleavage of the first (non-transferred) strand but is required for cleavage of the second (transferred) strand.

A final piece of evidence that second strand cleavage occurs in a synapse is that double strand cleavage appears to be concerted at the two ends on supercoiled substrates. Even when there was only a small amount of double strand cleavage, there was always at least as much ETF and donor backbone as there was linear cleaved plasmid. If cleavage at the two ends were independent, much more full-length linear than ETF and donor backbone would be produced when the overall level of cleavage was low. This concerted

cleavage at both ends was lost on linear substrates, suggesting that supercoiling is required for efficient synapsis.

Thus the current working model for transposition of ISY100 is that transposase binds to the two inverted repeats and cleaves the non-transferred strands without synapsis. After this, the two ends are brought together and the transferred strands are cleaved. The excised transposon with bound transposase then locates a target TA and the 3'-OH ends are transferred to the 5' phosphates of the T residues in a single transesterification step. A similar model for the order of strand cleavage events has previously been proposed for the *Mos1* transposon (Dawson and Finnegan, 2003), where assembly of a paired end complex appears to be required prior to cleavage of the transferred strand.

Chapter V

Target-specific transposition catalysed by chimeric transposases

5.1 Introduction

Transposition has been used for a wide variety of genetic and biotechnology applications. Traditionally transposons have been used to generate large numbers of different insertion mutations in a wide variety of organisms, including bacteria, insects and mammals. The recent development of *in vitro* transposition systems has allowed development of new applications for transposition such as delivery of primer sites for DNA sequencing projects, and *in vitro* mutagenesis of large DNA fragments carried on cosmids, Bacs or other vectors (Chatterjee and Coren, 1997; Chatterjee *et al.*, 2004; Holtman *et al.*, 2005).

Transposons can also be used as gene delivery vectors for insertional transgenesis, because they generate stable single copy integrations into chromosomal DNA. For instance, P elements are routinely used to insert DNA into the model organism *D. melanogaster,* and retrovirus-based vectors have been used in gene therapy trials. Although size limitations might be a problem, essentially any DNA sequence can be inserted as long as it is flanked by transposon end sequences. There has been much interest in use of Tc1/*mariner* family elements for gene delivery (reviewed in Plasterk *et al.,* 1999).

One of the chief disadvantages in the use of transposons and retrovirus-based vectors for gene delivery is their random integration into the host genome. This can have serious unwanted side effects. One problem is that insertions can occur into or adjacent to an important host gene, knocking it out or activating it inappropriately. This led to serious problems in gene therapy trials using retroviral vectors to replace the defective gene in severe combined immunodeficiency (SCID). In two patients, the retroviral vector integrated adjacent to a proto-oncogene in some cells, leading to activation of this oncogene and development of cancer (Hacein-Bey-Abina *et al.*, 2003). Another consequence of insertion into random sites is that insertion into areas of heterochromatin can lead to poor expression of the introduced gene (reviewed in Ellis, 2005).

We hoped that ISY100 could be used to develop a targeted transposition system that could deliver a transgene to a chosen site in a complex genome, removing the risk of insertional mutagenesis. The method chosen for this was to fuse a zinc finger DNA-binding domain (DBD) to ISY100 transposase, in the hope that it would direct transposition into sites adjacent to its binding site. Similar experiments have previously been carried out with retroviral integrase (IN) proteins fused to the zinc finger DNA-

binding domains from Sp1, Zif268 and E2C (Bushman and Miller, 1997; Peng *et al.*, 2002; Tan *et al.*, 2006; Tan *et al.*, 2004).

In the experiments of Tan *et al.* (2004, 2006), the synthetic polydactyl zinc finger protein E2C was fused to either the N- or the C- terminus of HIV integrase, giving proteins called IN-E2C and E2C-IN respectively. Integration catalysed by IN-E2C was clustered around the E2C binding site with an asymmetric distribution (Tan *et al.*, 2004).

Other DNA-binding domains, such as the lambda repressor and *E. coli* LexA repressor have also been fused to IN proteins (Bujacz *et al.*, 1996; Bushman, 1994; Goulaouic and Chow, 1996). These DNA-binding domains increased HIV integration into sites adjacent to their binding sites (Bushman, 1994; Goulaouic and Chow, 1996).

In these experiments, transposition reactions into defined DNA molecules were normally carried out *in vitro*. The distribution of target sites was then determined by PCR using a target-specific primer and a retrovirus-specific primer, and the results were generally displayed on a denaturing polyacrylamide gel (Bushman, 1994; Bushman and Miller, 1997; Tan *et al.*, 2006; Tan *et al.*, 2004). Thus the pattern of insertions was only monitored over a region of about 600 - 1,000 bp of DNA. Transposition generally increased close to the DBD-binding site, and was reduced in a region of DNA protected by the DBD. However the level of integration into sites distant from the specific binding site was unchanged. Therefore all of these systems still carry out random integration at a high level, and none of them can deliver a transposon to a precisely defined target sequence.

It was hoped that the existing TA target specificity of IS630/Tc1/*mariner* superfamily elements and the broad host range of these elements might be a useful starting point to develop targeted transposition. This would be done by fusing a specific DNA-binding domain to the ISY100 transposase.

Although other DNA-binding domains have been used for some of these studies, we decided to use the Zif268 DNA-binding domain (Zif268-DBD) because techniques have been developed to produce Zif268 derivatives that bind specifically to virtually any chosen DNA sequence. Furthermore the Zif268-DBD was readily available from A. Akopian and W. M. Stark. Zinc fingers are one of the most common types of eukaryotic DNA-binding domains and are found in a wide variety of transcription factors. Zinc finger proteins are normally made up from 3 or more repeated zinc-binding motifs termed zinc fingers. Each zinc finger binds to a zinc ion using a total of four histidine and cysteine residues, and folds into an independent DNA-binding motif. There are several families of zinc finger proteins, classified according to protein sequence. Zif268 belongs to the Cys_2His_2 family of

zinc fingers, which contain two cysteines and two histidines in its consensus sequence (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X_5-ψ-X_2-H-X_{3-5}-H, where X represents any amino acid and ψ is a hydrophobic residue (Wolfe *et al.*, 2000). Zif268 is a three zinc fingers domain from a murine transcription factor that specifically recognizes the 10 bp sequence 5′-GCGTGGGCGT – 3′ (Christy *et al.*, 1988). The structure of the Zif268-DNA complex has been solved by X-ray crystallography (Elrod-Erickson *et al.*, 1996). Each finger of Zif268 folds into a ββα architecture to form a compact domain stabilized by zinc. A single zinc ion is coordinated between two cysteines from the N-terminal β-sheet and two histidines from the C-terminal α-helix. In the structure, the C-terminal α-helix of each finger makes sequence-specific contacts with three basepairs in the major groove. Although each finger makes some interaction to the first base-pair of the next triplet (Figure 5.1), each finger binds relatively independently to 3 bp of sequence. The DNA-binding specificity of each finger can be changed by mutations in amino acids that make base-specific contacts with the DNA. Using this structural information, strategies have been developed to select Zif268 derivatives that bind tightly and specifically to a large number of chosen 9-10 bp sequences (Greisman and Pabo, 1997; Rebar and Pabo, 1994; Wolfe *et al.*, 2000).

Several other strategies have been proposed to integrate a gene to a specific target sequence. One widely used strategy is to use a site-specific recombinase, for example Cre or ϕC31 integrase. Cre from bacteriophage P1 can integrate a circular DNA molecule containing a *loxP* sequence into an existing genomic *loxP* site. The Cre-*loxP* system has been widely used and commercialized for this and other purposes such as conditional gene inactivation or inversion. The *Streptomyces* phage ϕC31 integrase catalyzes recombination between the bacterial *attB* and phage *attP* attachment sites. The irreversible and stable nature of the integration makes ϕC31 integrase a favorable transgenesis tool, and it has been successfully applied in many species other than bacteria such as human cells, flies and mouse *in vivo* (Groth *et al.*, 2004; Groth *et al.*, 2000; Olivares *et al.*, 2002). The advantage of using site-specific recombinases as mentioned above for the transgenesis is their simplicity and efficiency. However the requirement for pre-existing sites for integration on the target chromosome prevents them from being used more widely.

Since zinc finger domains can specifically recognize virtually any chosen sequence, they are an attractive option for changing the target specificity of various enzymes that act on DNA. As well as the example of HIV integrase mentioned above, other proteins have

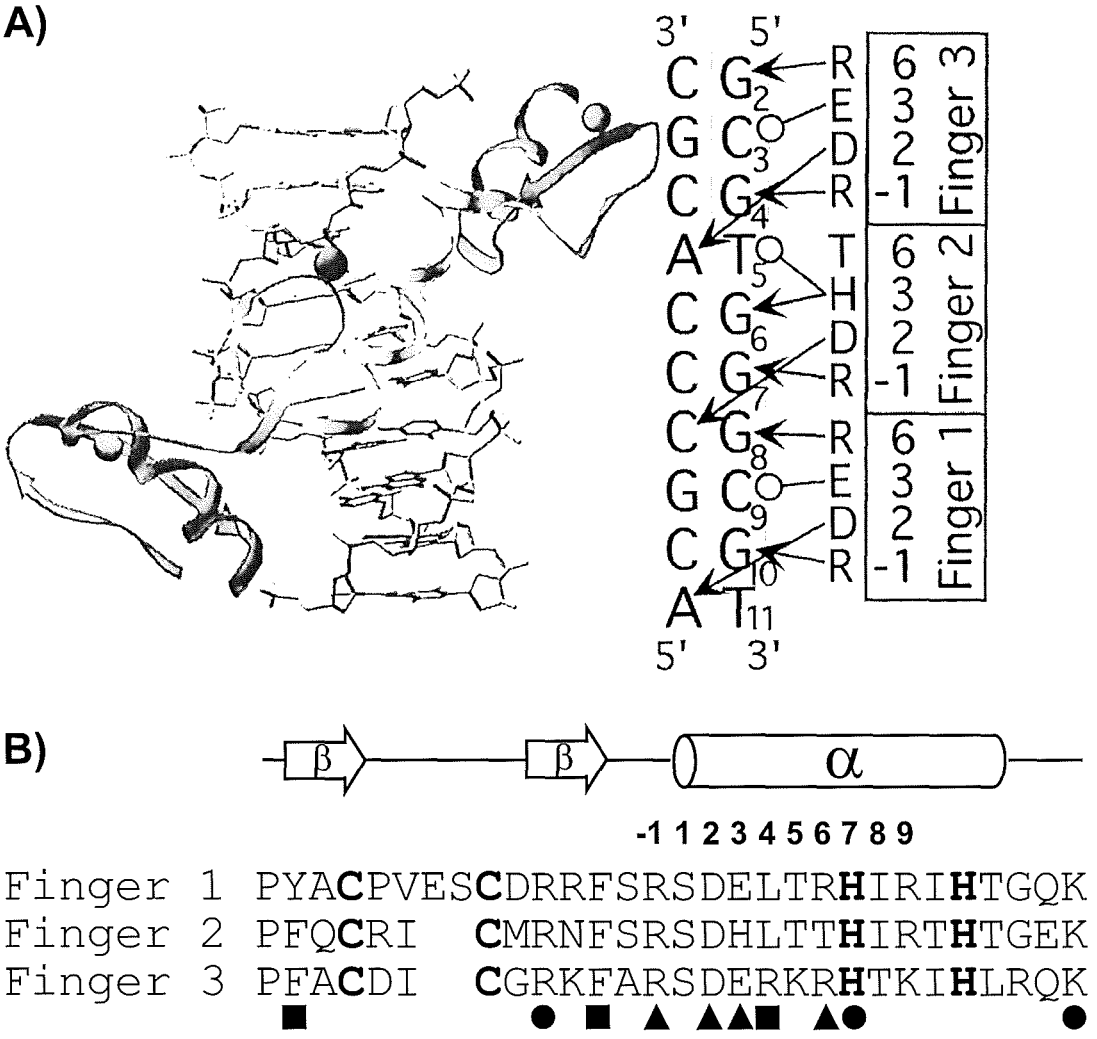


Figure 5.1 Structure of the three fingers of Zif268 bound to DNA. A) Base contacts made from positions -1, 2, 3, and 6 of each α-helix are indicated schematically to the right of the structure. Arrows indicate contacts mediated by hydrogen bonds; open circles indicate hydrophobic interactions. **B)** The sequence of the three fingers of Zif268. Diagram shows the positions of the β-sheets and α-helix above the corresponding residues. The cysteines and histidines involved in zinc coordination are indicated in bold. Filled squares below the sequence indicate the position of the conserved hydrophobic residues. Filled circles and triangles indicate residue positions that are involved in phosphate and base contacts respectively in most of the fingers. (Taken from Wolfe *et al.*, 2000).

also been modified by the addition of zinc finger domains to improve or change their target specificity. For instance, the Zif268 DNA-binding domain has been fused to the N-terminal catalytic domain of Tn3 resolvase. This chimeric recombinase can catalyze efficient recombination between recombination sites consisting of the central region of Tn3 *res* site I flanked by two Zif268 binding sequences in inverted repeat in *E. Coli* (Akopian *et al.*, 2003). If this chimeric recombinase is functional in eukaryotic cells, it may have applications in insertional transgenesis as well as controlled deletion of unwanted DNA sequences.

Another strategy that has been used is to fuse a (Cys_2His_2) zinc finger DBD to the nonspecific DNA cleavage domain of the restriction enzyme FokI. This yields a protein that will cleave specifically at a site consisting of two adjacent zinc finger binding sites in inverted repeat (Smith *et al.*, 1999). In this strategy, the specific double strand break can then invoke homologous recombination between the chromosome and a donor plasmid to integrate a foreign DNA sequence into the site (Smith *et al.*, 2000). This zinc finger fusion nuclease works in human cells and plant cells at a reasonable efficiency, and may be a powerful tool for gene modification (Urnov *et al.*, 2005; Wright *et al.*, 2005).

This chapter reports my attempts to modify the target specificity of ISY100 transposase by fusing the DNA-binding domains of Zif268 to ISY100 transposase.

5.2 Target sequence preference of wild-type transposase

Before attempting to modify the target specificity of ISY100 transposase, I first decided to investigate the target choice of wild-type and His_6-tagged ISY100 transposase.

Different transposases show different levels of target selectivity, preferring or avoiding insertions into certain sequences, regions or structures (Craig, 1997). For instance, Tn7 transposes at high frequency into the *attTn7* site, a single attachment site found in many bacterial genomes. This target-specific transposition uses transposon-encoded proteins TnsA, TnsB, TnsC and TnsD. TnsD binds to the attachment site and introduces DNA distortions which are recognized by TnsC, which somehow directs TnsAB-mediated transposition to *attTn7* (Kuduvalli *et al.*, 2001; Waddell and Craig, 1989). In another transposition pathway of Tn7, TnsA, TnsB, TnsC and TnsE bring about lower frequency transposition to sites which were initially thought to be random but were later found to be regions where chromosomal DNA replication terminates, and sites proximal to DNA double-strand breaks (Peters and Craig, 2000). Other transposons show

varying levels of preference for specific DNA sequences encompassing the duplicated target sequence and the surrounding region. For instance Tn5 prefers the sequences 5'-AGNT$^{T}/_{C}{}^{A}/_{T}{}^{A}/_{G}$ANCT-3' (N = any base). The sequence 5'- GATC$^{A}/_{T}$GATC -3' was a very good target site for *in vivo* transposition, especially when it was made into in a tandem array (Goryshin *et al.*, 1998). Tn10 prefers a 9 bp consensus target sequence 5' NGCTNAGCN (Bender and Kleckner, 1992; Halling and Kleckner, 1982), and the catalytic DDE domain is implicated in target choice (Junop and Haniford, 1997). As previously discussed, members of IS630/Tc1 family always jump into the dinucleotide TA. However, preferences for specific sequences outside the duplicated TA have also been reported for Tc1, Tc3, IS630 and *Sleeping Beauty* (Liu *et al.*, 2005; Preclin *et al.*, 2003; Tenzen and Ohtsubo, 1991).

Tc1 inserts into widely distributed sites on the *C. elegans* genome. A commonly studied target is *unc-22* because insertion and excision events in this gene are associated with an easily observed (uncoordinated) phenotype. An alignment of 12 insertion target sites, including seven in *unc-22*, revealed a consensus sequence 5'- GA$^{G}/_{T}$A**TA**TGT -3' in which the TA (in bold) was duplicated after insertion (Mori *et al.*, 1988). To investigate this further, target sequences of 204 independent Tc1 insertions and 166 Tc3 insertions in a 1 kb region of *gpa-2* gene were isolated by PCR and analyzed in more detail by DNA sequencing (van Luenen and Plasterk, 1994). No consensus target sequence was found for Tc1 or Tc3 in this study. However the different pattern of inserts for Tc1 and Tc3 suggested that the hotspots may depend on the primary DNA sequences, and/or sequence dependent DNA structure, but not the local chromatin structure. In another study, sequencing 378 alleles of Tc1 left flanks and 340 alleles of right flanks in strains carrying high copy numbers of Tc1 revealed a weak consensus target sequence of 5'-CA$^{C}/_{T}$A**TA**T$^{G}/_{A}$TG -3', with the TA in the middle (Korswagen *et al.*, 1996). Recently, 588 Tc1 insertions and 223 Tc3 insertions into a wider region of the genome were analysed and the result suggested that Tc1 prefers a target sequence 5'-A$^{C}/_{T}$A**TA**T$^{G}/_{A}$T -3' and Tc3 prefers 5'- ATA**TA**TTT -3' (Preclin *et al.*, 2003). The observation that Tc3 does have a preferred target sequence suggests that the failure of the previous study to identify a consensus target site (van Luenen and Plasterk, 1994) might be due to the small sample size.

To investigate the target preference of ISY100, the sequences flanking insertions from various sources were analyzed as follows. The original target sequences were first reconstructed *in silico* by precise removal of the ISY100 sequence and one copy of the TA

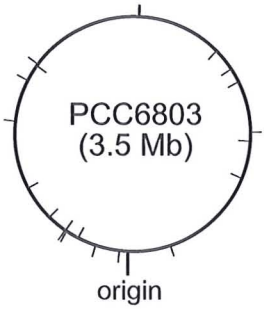
target duplication. The flanking sequences were aligned from left to right in the direction of the insertion, that is with IRL on the left and IRR on the right. The base frequencies were then calculated at every position for 20 bp on either side of the target TA. The alignments were also analyzed using the program WebLogo, which generates graphical sequence logos showing the sequence similarity of multiple sequence alignments providing more information than simple consensus sequences (Crooks *et al.*, 2004). The sequence logo is comprised of stacks of letters, one stack for each position in the sequence. The total height of each stack indicates the sequence conservation, or information content at that position, and the height of each letter within the stack represents the relative frequency of that nucleotide. The information content is measured in bits (calculated according to the formula shown in Figure 5.2C), where 0 indicates no deviation from random sequence and 2 represents 100% DNA sequence conservation. Note that the smaller the sample size, the higher the probability of producing a high information content by chance, as indicated by the error bars on the diagrams.

The target choice of the 24 ISY100 insertions in the genomic sequence of *Synechocystis* sp. strain PCC6803 and its plasmids pSYSG and pSYSA were analyzed first. Out of the 21 genomic ISY100 insertions, 11 are in one orientation and 10 are in the other. Both orientations are spread randomly throughout the genome, with no obvious correlation between the orientation of insertion and the presumed direction of DNA replication (Figure 5.2 A). All of the insertions are flanked by TA target duplications on both sides, except one chromosomal copy that has TA on one side and CA on the other. This could be due to a mutation that occurred after insertion of this copy of ISY100. The alignment of these 24 insertions, and the WebLogo produced from this alignment, suggests that there is a tendency for three base-pairs on either side of the target TA to be A-T rich with weak consensus Ts at positions -1, +1, +2 and +3 as well as an A at position -3 (Figure 5.2D,E).

Urasaki *et al.* reported the sequence of 21 mini-ISY100 insertions into plasmids pSEK80 and pAU5 generated in their *in vivo* transposition system (Urasaki *et al.*, 2002). The sequences of these insertions were analyzed in a similar way (Figure 5.3). Again there is a tendency for at least three base-pairs on either side to be A-T rich, although there was no strong consensus sequence outside the target TA.

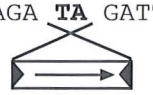
Next, the target sequences of 48 insertions generated *in vivo* in this work (Chapter III) were analyzed. These insertions were generated in a variety of assays, with donor plasmids containing different combinations of ISY100 left and right ends, or different

A)



B)

```
5'- CGTATAATGCTTTCTCGTAT TA CCTCTTAATTTTTACTTTAG -3'
    AATAGCGTAAATTTTCCACT TA CTTTGGGATCAGGGTTGTCA
    CTGGAAGTATTCCCAGCTTT TA TGTGTCTAAATCAGGTTCAT
    CTCAAGGAAAGTTAGAAAAG TA ACGACCACAAAAATGCTAAT
    CCGGTTTGAATTACAAAGCA TA GTTCAAAAATTAACCGAAGA
    AGAGGTTGAGGCTTTGTTTT TA TTATCGGAACAACATTAATC*
    GTAAAGGAGCAAGTTTTAGT TA TATTCCGATCCCTATCCTCA
    TATATGAAAATAGTCAAATC TA TTTTCATAAAACTAAATTCA
    TTACCCACCTCTTTATATTT TA CTCCTGATACTTTTGTCATA
    TTGGGATAACAATTCGGGAA TA TTGTCTCGAATAATGCTCCA
    CTTCTATTGTTTCATTCCCA TA TTTATTATCATGACCTTGTC
    TGTTACCCGCTTTGAAGATT TA TTTGCTGAATTTTGCGGAAC
    ACAGTAAATATCAACGATTA TA TTTTTAAAATTTCTAGAAAC
    TTTTAGGCAAAATTTAGATT TA AAAATCAAACTTGCAGAGCT
    CATAGGTTTCTCCCAAAAAA TA ATGTATGAAGCTCTCTCTTA
    TCTTAATACATGATATTTGT TA TTCTCCTTTTGTAATCTCTT
    ACAATGCCAACTATGGGTTG TA TTTCACTTTCTGGGATAAAA
    ATTATCGGTCAACAAATCAA TA CTAATACGGCGCAAATGTTA
    AAGCAATCCAGGCTTGTATT TA TGCCTGGTCAAAAAAAGGAA
    AAAAGAGCTTGGTTTCAATA TA AAGGTAGTATTAGCATATAG
    GTTAACGGGCAAAATGTTGA TA CAGTCGAACAGAGACTTTTT
    CGAGCCTGATCGACGACGAT TA TATGGGATCTATGACCGCCG ]- pSYSG
    CCACATGAGACTCACTTACA TA ATTTTAGTGATGAACTAAAC ]
    ATCCTGGGGCGGATTCAAGA TA GATTGCACCACCGCATTCAA — pSYSA
```

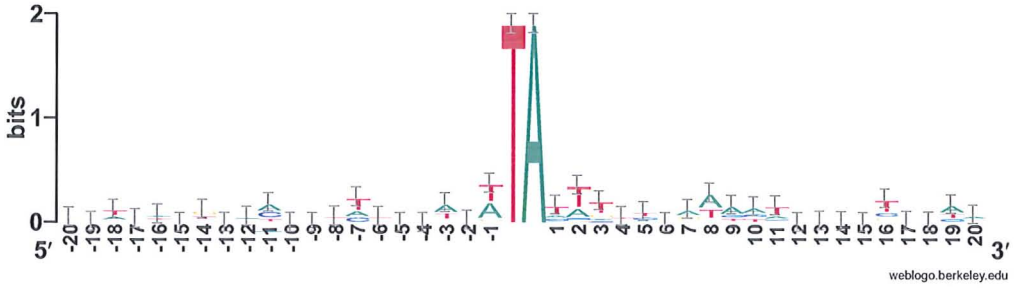


C)

$$R_{seq} = S_{max} - S_{obs} = \log_2 N - (-\sum_{n=1}^{N} p_n \log_2 p_n)$$

R_{seq} = information content of a single position
S_{max} = maximum possible entropy
S_{obs} = the entropy of the observed symbol distribution at that position
N = number of possible symbols per position (=4 for DNA)
p_n = observed frequency of symbol n at that position

D)

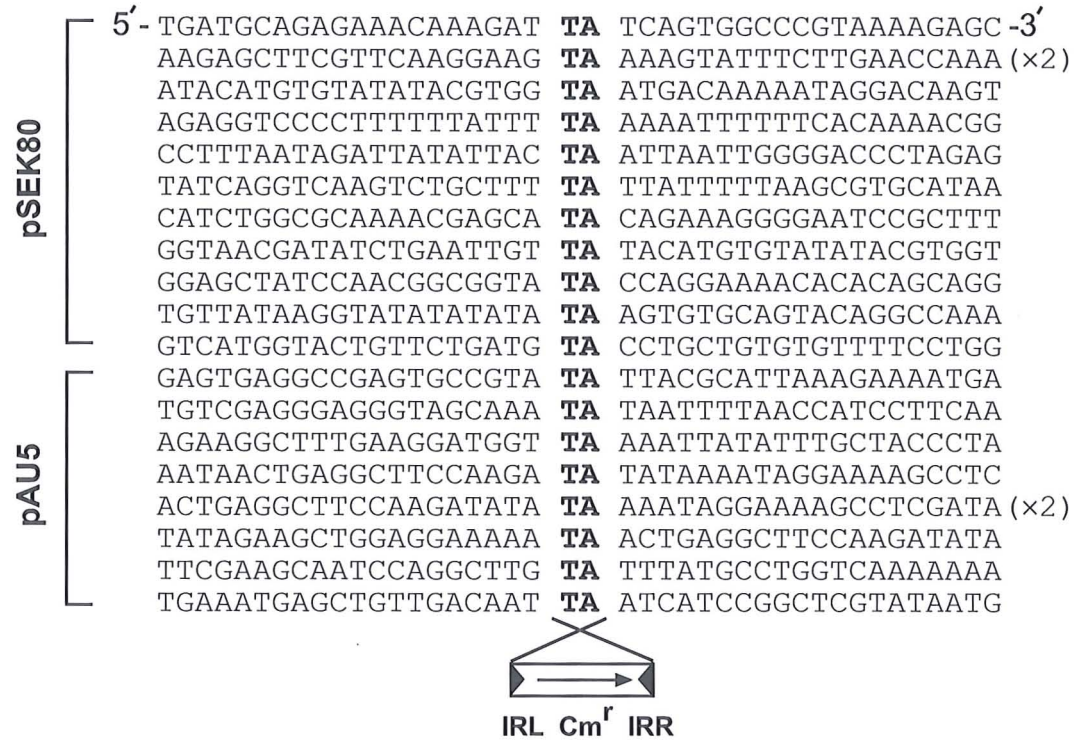


E)

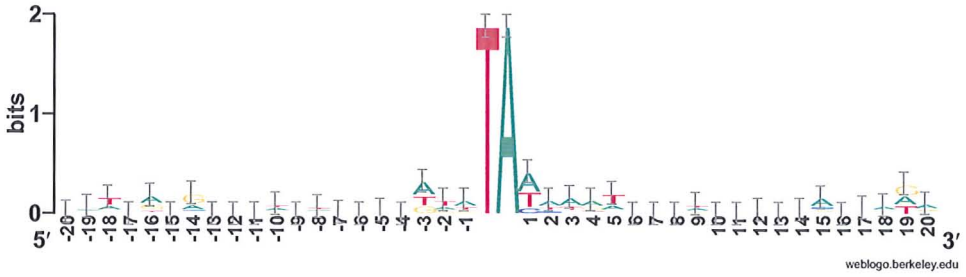
	-5	-4	-3	-2	-1	**T**	**A**	+1	+2	+3	+4	+5
A	8	8	11	6	10	0	24	5	6	3	4	3
C	4	4	2	4	1	0	0	5	2	3	5	8
G	7	5	3	4	2	0	0	2	2	5	4	3
T	5	7	8	10	11	24	0	12	14	13	11	10

Figure 5.2 Analysis of target sequences of insertion of ISY100 in the genome of *Synechocystis* sp. strain PCC6803. A) Distribution of ISY100 insertions on the chromosome of *Synechocystis* sp. strain PCC6803. Map of the PCC6803 genome showing the locations of 21 insertions. Lines on the outside of the circle indicate insertions in a clockwise orientation, those on the inside are anticlockwise. The presumed origin of replication deduced by the skewed oligomers method (Salzberg *et al.,* 1998) is also shown. **B)** Alignment of flanking sequences of ISY100 insertions on the genome and plasmids of *Synechocystis* sp. strain PCC6803. All the sequences are shown oriented in the direction of ISY100 transposase. The bold TA in the middle of the sequence was duplicated after the insertion of the ISY100. The insertion marked with an asterisk (*) was flanked by TA at one side and CA on the other. The last three target sequences are from the plasmids whose names are shown at the end of the sequence. **C)** The formula used by Weblogo to calculate the information content of a sequence alignment (Crooks *et al.,* 2004). **D)** Weblogo of the sequence alignment shown in B). **E)** Base frequencies of 5 nucleotides on either side of the target TA.

A)



B)



C)

	-5	-4	-3	-2	-1	T	A	1	2	3	4	5
A	8	5	8	6	8	0	19	9	8	9	8	5
C	4	5	0	1	1	0	0	3	4	2	1	2
G	5	4	5	4	4	0	0	0	1	2	6	2
T	2	5	6	8	6	19	0	7	6	6	4	10

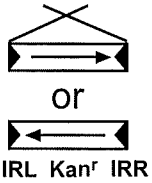
Figure 5.3 Analysis of sequences flanking ISY100 insertions isolated *in vivo* by Urasaki *et al.*, 2002. A) Alignment of sequences flanking the duplicated target TA (in bold). **B)** Weblogo of the aligned target sequences. **C)** Base frequencies of 5 nucleotides on either side of the target TA.

flanking dinucleotides. All of the target sequences from these assays are shown aligned in Figure 5.4. Because only a small number of insertions were sequenced for some donor plasmids, it is not possible to see if the differences between the donors affected the target choice. However, because ISY100 is likely to excise fully from the donor plasmid before the target is located, it seems unlikely that the donor flanking sequences will affect the target choice. The targets of *in vivo* transposition were therefore analyzed in three groups according to the combination of left and right ends in the mini-transposon. 33 insertions of ISY100-IRL/IRR, 12 insertions of ISY100-IRR/IRR, and 3 insertions of ISY100-IRL/IRL were analyzed separately. Although some individual target sites were used several times in these assays, each sequence was counted only once in any alignment, so as to avoid introducing an apparent preference for sequences that in fact have no effect on target choice. For ISY100-IRL/IRR, the strongest target preference outside the central TA was for an A at position –3 and a T at position +3 (Figure 5.4B). The number of insertions of the other two classes of elements was small, but the target sites used do not seem significantly different from those obtained with ISY100-IRL/IRR. Thus there seems to be no strong preference for left ends to insert adjacent to one type of sequence and right ends adjacent to another type. Adding the target sites of IRL/IRL and IRR/IRR insertions to the target sites of IRL/IRR insertions in either orientation did not significantly change the results obtained from the latter class alone.

Finally, the insertions obtained from *in vitro* transposition reactions were analyzed in a similar way. *In vitro* transposition reactions used three different sources of transposon DNA: supercoiled donor plasmid, transposase-excised mini-ISY100, or AcuI-excised mini-ISY100. Alignments of target sequences from these three groups gave similar Weblogo patterns and were then analyzed together (Figure 5.5). In the combined data set, there is a strong preference for A (with a weaker preference for T) at position –3, and a symmetrical preference for T (or A) at position +3. There is also a weaker preference for either A or T at positions ±1, A, G or T at position -2 and A, C or T at position +2. This preference coincides quite well with the preference deduced for ISY100-IRR/IRL in the *in vivo* assays. Combining the target sites of ISY100-IRL/IRR from *in vivo* and *in vitro* assays in this study gives a consensus sequence ARW**TA**WYT (W = A or T, R = A or G, Y = C or T; Figure 5.6). When all of the ISY100-IRL/IRR insertions from this study and study of Urasaki *et al.* (2002) and the genome of PCC6803, are analyzed together, the consensus target sequence ADW**TA**WHT (W = A or T, D = not C, H = not G; Figure 5.6) can be

A)

	No.	Flanking sequences			Orientation of insertion	Position of insertion	
TA-L, TA-R, Tnp-His$_6$	T11	TGACGTATAAACGTTGCTTG	**TA**	TTTATTATTACAATTAAAGG	(+)	186 bp	
	T12	GAGTGAGACAAAGAGCGAAA	**TA**	AGAGCTTCAAAGGCTATCTC	(+)	2182bp	
	T13	CCACGGTATCTTGTTAAAAA	**TA**	TATATTTGTGGCGGGCTTAC	(+)	64 bp	
	T14	GCAAACCTGAGATTAGTTTC	**TA**	ACTAGTCTCTAGGTTTTCTA	(-)	232 bp	
	T16	CAGCGTAAGCTCCCCGTCCT	**TA**	ACTGCCTCGAGATAGTGACA	(+)	2023bp	
	T17	CCTAGGTACGGGGCTGAATG	**TA**	ACTTGCCTTTCCTTTCTCAT	(+)	1245bp	
	T18	ACGGAAATGTTGAATACTCA	**TA**	CTCTTCCTTTTTCAATATTA	(-)	3274bp	
	T19	TTGGACTTGGTGCTGCTATA	**TA**	TTAGCTAATACACTAGTTAG	(+)	1539bp	
	T20	AGACTCAGACAAAGACCAAC	**TA**	GTTATTGACCAGTCTTCTAG	(-)	937 bp	
TA-L, TA-R	T1	TCCCTCTGCCAACAGAAACA	**TA**	TTGGTCAAGCACTTGAAGGA	(-)	1794bp	
	T2	TTGTCTCATGAGCGGATACA	**TA**	TTTGAATGTATTTAGAAAAA	(-)	3211bp	(=T4/7/8)
	T3	TTGTGTGTTTAGTTTATAAT	**TA**	GTCTCTTATTAATTTGATGT	(-)	108 bp	
	T5	TGATAACACTGCGGCCAACT	**TA**	CTTCTGACAACGATCGGAGG	(+)	3678bp	
	T6	GAGTGAGACAAAGAGCGAAA	**TA**	AGAGCTTCAAAGGCTATCTC	(+)	2182bp	
	T9	TAAAAGAGATTTCGACTAGT	**TA**	TGTACCTGGAGTTTGGATAT	(+)	2265bp	
	T10	TTAAATTAAAAATGAAGTTT	**TA**	AATCAATCTAAAGTATATAT	(-)	4176bp	
	T16'	TAAAAGAGATTTCGACTAGT	**TA**	TGTACCTGGAGTTTGGATAT	(+)	2266bp	
	T17'	TTCGTGCTATCACAAAACAG	**TA**	TACAAAAAATAAGCTTTCGC	(-)	16 bp	
	T18'	AGAGTAAGTAGTTCGCCAGT	**TA**	ATAGTTTGCGCAACGTTGTT	(-)	3849bp	
TA-L, TA-L	T37	GATCTTCACCTAGATCCTTT	**TA**	AATTAAAAATGAAGTTTTAA	(-)	4196bp	
	T38	GCCTTTCTGGTTCTACGAAG	**TA**	TCCAGTCCCGCCTGCAACGC	(-)	2545bp	
	T39	CAAACTATTAACTGGCGAAC	**TA**	CTTACTCTAGCTTCCCGGCA	(+)	3861bp	
TA-R, TA-R	T34	TGATGGCAGACATGGTAGGT	**TA**	ATACCTTGTTAGTTATTGCT	(-)	635 bp	
	T35	AGAGATTTCGACTAGTTATG	**TA**	CCTGGAGTTTGGATATAGGA	(+)	2269bp	
	T36	GCACCATATGCGGTGTGAAA	**TA**	CCGCACAGATGCGTAAGGAG	(-)	2876bp	
TG-L, TG-R	T20	ACCGTGGAAAAACTTCGTGC	**TA**	TCACAAAACAGTATACAAAA	(-)	29 bp	
AA-L, AA-R	T30	CAAGGTGTTCAGAGACTAGT	**TA**	TTGACTAGTCAGTGACTAGC	(+)	596 bp	
CA-L, CA-R	T31	GAGCACTTTTAAAGTTCTGC	**TA**	TGTGGCGCGGTATTATCCCG	(+)	3496bp	
	T32	TGGCAGACATGGTAGGTTAA	**TA**	CCTTGTTAGTTATTGCTAGT	(-)	631 bp	
	T33	CCTTTAATTGTAATAATAAA	**TA**	CAAGCAACGTTTATACGTCA	(-)	186 bp	
TG-L, TA-R	T43	GCTGCCGGATTTCAAGATAG	**TA**	AGCCGACCCGACTAGCTTTA	(-)	1985bp	
	T44	CGTTCCTCGGTGGTCGCAGA	**TA**	ACTCCCTAAATTCCCTGAGC	(-)	2432bp	
	T45	TGATAACACTGCGGCCAACT	**TA**	CTTCTGACAACGATCGGAGG	(+)	3678bp	
TC-L, TA-R	T46	GTAGGTAAGCCCGCCACAAA	**TA**	TATATTTTTAACAAGATACC	(-)	68 bp	
	T47	CCTAGGTACGGGGCTGAATG	**TA**	ACTTGCCTTTCCTTTCTCAT	(+)	1245bp	
	T48	GATTCTGTGGATAACCGTAT	**TA**	CCGCCTTTGAGTGAGCTGAT	(+)	4996bp	
TT-R, TA-R	T49	AAATCTAGTCAAGAACTAGA	**TA**	CTTAACCTTCATACTGCTTA	(-)	2224bp	
	T50	TGCGTAAGTTTCTGTGAGTC	**TA**	ACTAGTGTATTAGCTAATAT	(-)	1558bp	
	T51	ATATATACTTTAGATTGATT	**TA**	AAACTTCATTTTTAATTTAA	(+)	4176bp	
AA-R, TA-R	T52	AGTGGTGAGGATAATAGTAG	**TA**	AGTTATTCTAGCTTCAGAGT	(-)	527 bp	
	T53	TGGTGCTGCTATATATTAGC	**TA**	ATACACTAGTTAGACTCACA	(+)	1546bp	
	T54	TAATAAATACAAGCAACGTT	**TA**	TACGTCAAAAGCATTTATAA	(-)	173 bp	
CA-R, TA-R	T55	GCACCCCAGGCTTTACACTT	**TA**	TGCTTCCGGCTCGTATGTTG	(+)	5227bp	
	T56	CGGGATAATACCGCGCCACA	**TA**	GCAGAACTTTAAAAGTGCTC	(-)	3495bp	(=T57)



B)

ISY100 IRL/IRR

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	9	6	15	10	10	0	26	8	5	5	8	4
C	9	7	2	5	4	0	0	5	7	4	6	9
G	7	4	0	6	4	0	0	2	4	3	8	5
T	1	9	9	5	8	26	0	11	10	14	4	8

ISY100 IRR/IRR

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	2	3	6	2	3	0	11	5	2	4	2	5
C	3	2	1	1	2	0	0	3	4	2	4	1
G	1	3	3	3	2	0	0	1	2	1	3	2
T	5	3	1	5	4	11	0	2	3	4	2	3

ISY100 IRL/IRL

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	0	0	2	2	0	0	3	1	1	0	2	1
C	3	1	0	0	1	0	0	1	1	1	0	1
G	0	2	0	0	1	0	0	0	0	0	0	1
T	0	0	1	1	1	3	0	1	1	2	1	0

C)

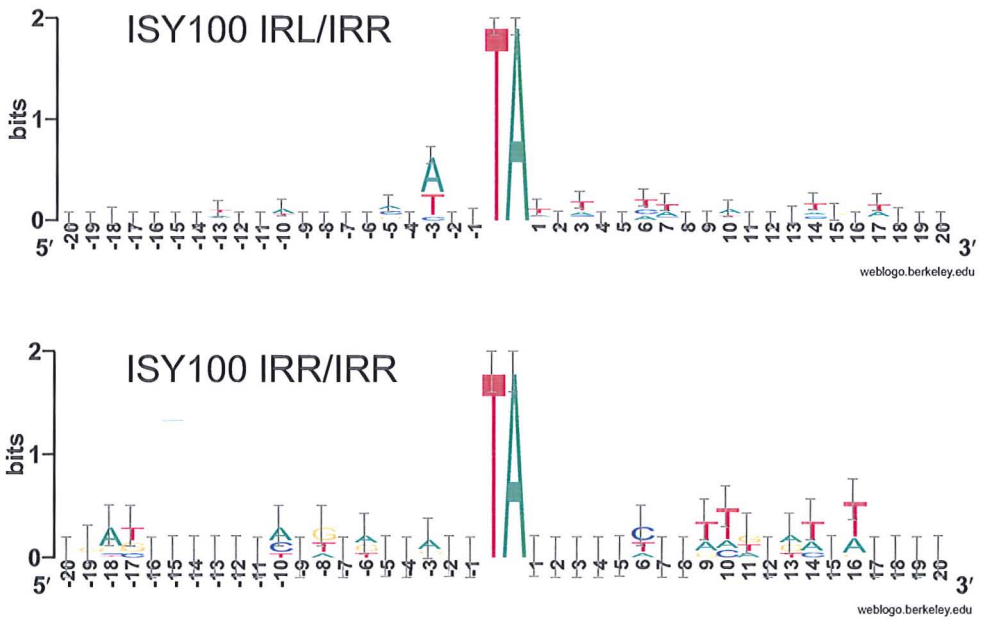
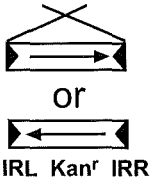


Figure 5.4 Analysis of sequences flanking ISY100 insertions isolated *in vivo* in this study. A) Alignment of sequences flanking the duplicated target TA (in bold) from individual insertions. Multiple insertions at the same site are listed in the right hand column. All the assays used pH2 as target plasmid and the positions of the insertions are given. The orientation of the insertions were labelled as + or –, indicating the orientation from IRL to IRR of insertions (or the orientation of kanamycin resistance gene on mini-ISY100 with two identical ends) relative to the ampicillin resistance gene on pH2. The first group of insertions were obtained with His_6-tagged transposase. All of others were obtained with wild-type transposase. **B)** Base frequencies of 5 nucleotides on either side of the target TA. Duplicated target sequences were removed so that each target sequence contributed only once to the analysis in any direction. **C)** Weblogo of the aligned independent target sequences.

A)

	No.		Flanking sequences	Orientation of the insertion	Position of the insertion	
I Supercoiled plasmid donor	T2		TGTCAGACCAAGTTTACTCA **TA** TATACTTTAGATTGATTTAA	(+)	4157bp	
	T3		AACTAGTGTATTAGCTAATA **TA** TAGCAGCACCAAGTCCAAGG	(−)	1537bp	
	T5		AGAAAAAGTCGGTCAAGAGG **TA** ACTCTAAGTCGGCCATTCCT	(+)	2345bp	
	T6		AAAAAAACAAGACAGCCACA **TA** TTAGGGAGACTAGTTAGAAG	(+)	348 bp	
	T7		ACTAGTTTCTGACTAGTTAT **TA** ACTAGTTGACAGGTTCCCTC	(−)	1830bp	
	T9		CTGACTCCCCGTCGTGTAGA **TA** ACTACGATACGGGAGGGCTT	(−)	4048bp	(=T12/
	T11		GCACAACATGGGGGATCATG **TA** ACTCGCCTTGATCGTTGGGA	(+)	3744bp	13/14)
	T15		TTGCAACTAATACTAGTTAG **TA** TCTAGTCTCCTATAACTCTG	(+)	491 bp	
	T16		AAATCTAGTCAAGAACTAGA **TA** CTTAACCTTCATACTGCTTA	(−)	2224bp	
II Tnp-excised mini-ISY100	T1		GGCAGTGAGCGCAACGCAAT **TA** ATGTGAGTTAGCTCACTCAT	(+)	5182bp	
	T3		CCTGCAGATTTTGACCTAGT **TA** ACAAGGCACATCATGCTATC	(−)	2620bp	
	T4		TTGGACTTGGTGCTGCTATA **TA** TTAGCTAATACACTAGTTAG	(+)	1539bp	(=T7)
	T5		CTAACTAGTGTATTAGCTAA **TA** TATAGCAGCACCAAGTCCAA	(−)	1539bp	(=T14/15/
	T6		AGGTAAGCCCGCCACAAATA **TA** TATTTTTAACAAGATACCGT	(−)	66bp	16/20)
	T8		AAGCCCTCCCGTATCGTAGT **TA** TCTACACGACGGGGAGTCAG	(+)	4048bp	
	T9		AATTAATAAGAGACTAATTA **TA** AACTAAACACACAAGCAAGC	(+)	114bp	
	T10	*	CTAACTAGTGTATTAGCTAA **TA** TATAGCAGCACCAAGTCCAA	(−)	1539bp	
	T11		CGGGCTTACCTACATCAAAT **TA** ATAAGAGACTAATTATAAAC	(+)	97bp	
	T12		AACTGATCTTCAGCATCTTT **TA** CTTTCACCAGCGTTTCTGGG	(−)	3366bp	
	T13		ATAGGAGACTAGATACTAAC **TA** GTATTAGTTGCAACTAACCT	(−)	484bp	
	T17		GATAGCATGATGTGCCTTGT **TA** ACTAGGTCAAAATCTGCAGG	(+)	2620bp	
	T18		CAGTCTAGTTGACAACTAGA **TA** GATTCTAGATTGAGGCCAAA	(+)	429bp	
	T19		AGTTATAGGAGACTAGATAC **TA** ACTAGTATTAGTTGCAACTA	(−)	488bp	
III AcuI-excised mini-ISY100	T1		TTTTTCTAAATACATTCAAA **TA** TGTATCCGCTCATGAGACAA	(+)	3211bp	
	T3		CTAACTAGTGTATTAGCTAA **TA** TATAGCAGCACCAAGTCCAA	(−)	1539bp	(=T6)
	T4		AAGCCCTCCCGTATCGTAGT **TA** TCTACACGACGGGGAGTCAG	(+)	4048bp	
	T7		TATTTGTTTATTTTTCTAAA **TA** CATTCAAATATGTATCCGCT	(+)	3201bp	
	T8		AGCGGATACATATTTGAATG **TA** TTTAGAAAAATAAACAAATA	(−)	3201bp	
	T10		ATATCCAAACTCCAGGTACA **TA** ACTAGTCGAAATCTCTTTTA	(−)	2265bp	
	T11		CCTGCAGATTTTGACCTAGT **TA** ACAAGGCACATCATGCTATC	(−)	2620bp	
	T12		GACTAGTCAGTGACTAGCAA **TA** ACTAACAAGGTATTAACCTA	(+)	619bp	
	T13		TCCCAACGATCAAGGCGAGT **TA** CATGATCCCCCATGTTGTGC	(−)	3744bp	
	T14		GCTCAGGGAATTTAGGGAGT **TA** TCTGCGACCACCGAGGAACG	(+)	2432bp	
	T17		TAAAAGAGATTTCGACTAGT **TA** TGTACCTGGAGTTTGGATAT	(+)	2265bp	
	T18		CTATGCGGCATCAGAGCAGA **TT** GTACTGAGAGTGCACCATAT	(−)	2908bp	
	T19	*	GTAAGTAGTTCGCCAGTTAA **TA** GTTTGCGCAACGTTGTTGCC	(−)	3846bp	(=T24)
	T20		CGACTAGCTTTAAGCCTAGT **TA** CTGACTAGTCCACAGGACGA	(−)	1955bp	
	T21		CCTGTCGTGCCAGCTGCATT **AA** TGAATCGGCCAACGCGCGGG	(−)	5122bp	
	T22		TGTCAGACCAAGTTTACTCA **TA** TATACTTTAGATTGATTTAA	(+)	4157bp	
	T25		TTGGACTTGGTGCTGCTATA **TA** TTAGCTAATACACTAGTTAG	(+)	1539bp	
	T26		TTGTCTCATGAGCGGATACA **TA** TTTGAATGTATTTAGAAAAA	(−)	3211bp	



B)

Plasmid donor

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	2	1	6	2	5	0	9	4	2	1	5	2
C	2	3	0	2	0	0	0	1	5	0	3	2
G	3	1	0	3	3	0	0	0	0	1	1	4
T	2	4	3	2	1	9	0	4	2	7	0	1

Tnp-excised mini-ISY100

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	2	4	8	5	5	0	13	6	4	4	6	1
C	6	3	0	0	2	0	0	1	4	1	0	4
G	4	0	0	4	0	0	0	2	0	1	1	6
T	1	6	5	4	6	13	0	4	5	7	6	2

AcuI-excised mini-ISY100

	-5	-4	-3	-2	-1	T	A	1	2	3	4	5
A	3	1	14	5	10	1	17	3	4	4	11	3
C	6	5	1	3	0	0	0	3	5	0	1	7
G	8	3	0	7	1	0	0	2	3	1	4	5
T	1	9	3	3	7	17	1	10	6	13	2	3

All

	-5	-4	-3	-2	-1	T	A	1	2	3	4	5
A	6	6	25	11	17	1	34	12	8	7	18	6
C	12	9	1	4	2	0	0	5	12	1	4	10
G	13	4	0	12	4	0	0	4	3	3	5	13
T	4	16	9	8	12	34	1	14	12	24	8	6

C)

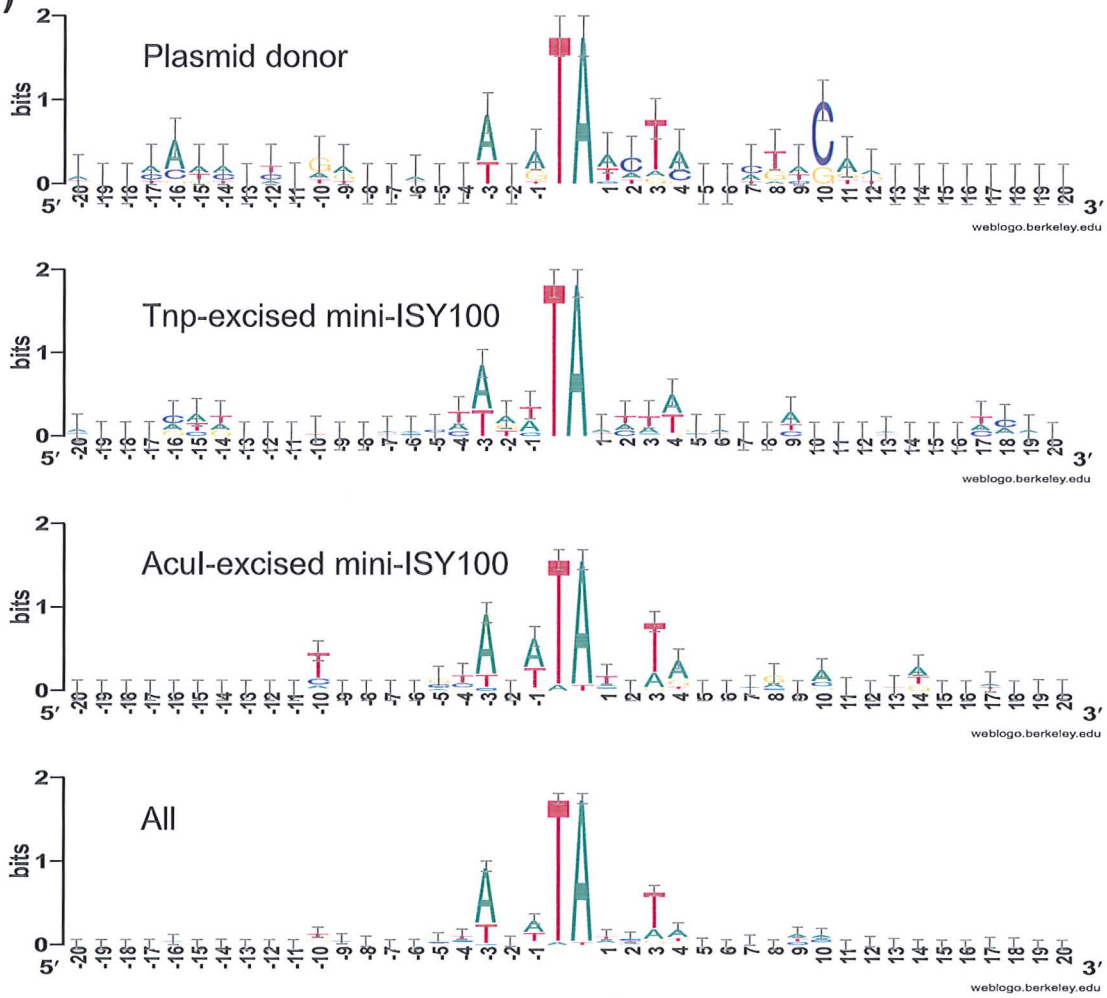


Figure 5.5 Analysis of sequences flanking ISY100 insertions isolated *in vitro* in this study. A) Alignment of sequences flanking target TA dinucleotides. All the assays used pH2 as target plasmid. Group I results are from *in vitro* transposition assay using supercoiled plasmid carrying mini-ISY100 as donor, group II and III results are from *in vitro* transposition assays using mini-ISY100 excised from pXF153 by ISY100 transposase (group II) or by AcuI (group III). The orientation of the insertions were labelled as "+" or "–", indicating the orientation from IRL to IRR of insertions relative to the ampicillin resistance gene on pH2. Sequences marked with an asterisk (*) were flanked by TA at one side but only A at the other. **B)** Base frequencies of 5 nucleotides on either side of the target TA. Duplicated target sequences were removed so that each target sequence contributed only once to the analysis in any direction. **C)** Weblogo of independent target sequences from A).

A)

In vivo (from PCC6803 + Urasaki *et al.*, 2002)

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	16	13	19	12	18	0	43	14	14	12	12	8
C	8	9	2	5	2	0	0	8	6	5	6	10
G	12	9	8	8	6	0	0	2	3	7	10	5
T	7	12	14	18	17	43	0	19	20	19	15	20

In vivo + *in vitro* (IRL/IRR transposon - this work)

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	14	12	37	21	25	1	57	20	13	11	25	9
C	19	16	3	8	6	0	0	10	19	5	10	17
G	20	8	0	17	8	0	0	6	6	6	11	18
T	5	22	18	12	19	57	1	22	20	36	12	14

All

	-5	-4	-3	-2	-1	**T**	**A**	1	2	3	4	5
A	30	25	56	33	43	1	100	34	27	23	37	17
C	27	25	5	13	8	0	0	18	25	10	16	27
G	32	17	8	25	14	0	0	8	9	13	21	23
T	12	34	32	30	36	100	1	41	40	55	27	34

B)

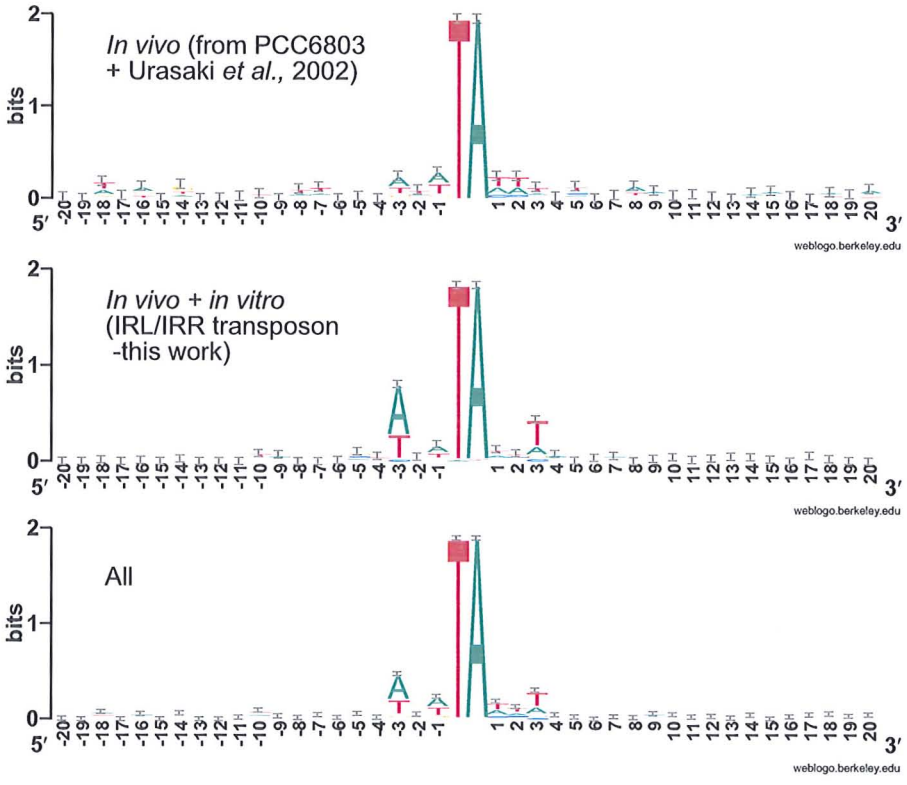


Figure 5.6 Analysis of target sequences from independent ISY100 insertion sites. Independent target sequences of mini-ISY100 with IRL and IRR from *in vivo* and *in vitro* transposition assays in this study and target sequences obtained from *in vivo* transposition assay done by Urasaki *et al.*, 2002 together with the deduced target sequences of ISY100 on the genomic DNA were analysed separately and in combination. **A)** Base frequency distribution of 5 nucleotides on either side of the target TA. **B)** Weblogo of the alignment of the independent target sequences.

deduced. At position –3, A is present in 55% of targets and T is present in 32%, whereas at position +3, T is present in 55% of targets and A is present in 23%. At positions –1 and +1, either T or A is present in 74-78% of insertions and at positions –2 and +2, either T or A is present in 62-66% (Table 5.1). At position -2, G is present in 25% of target sites, whereas C is unfavourable. Conversely at position +2, C is present at 25% of target sites and G is unfavourable.

The deduced consensus target sites of IRL/IRR insertions from this study as well as that deduced from all the published data, is palindromic. This is the expected result if there is no orientation preference for insertion at any site, even if the preferred sequence is not palindromic. This is because the targets are aligned according to the direction of ISY100 insertions and a non-palindromic sequence, when averaged in both orientations will give a palindromic sequence. Multiple insertions were observed at several different sites, and both orientations were observed at these sites in approximately equal numbers, agreeing with the idea that there is no orientation preference for insertion (Table 5.2, Figure 5.7). This suggests that the left and right ends of the transposon are equivalent in the target search, as would be expected if the strand transfer complex is functionally 2-fold symmetric. Nevertheless, it is still possible that palindromic targets are preferred over non-palindromic sequences. This question can be addressed by asking whether palindromic bases are more common than expected at mirror positions. In other words, if an A is present at position –3, is there an increased probability of a T at position +3? Analysis of the frequency of palindromic bases at mirror positions (Table 5.3) shows that palindromic bases do not occur at higher than expected frequency at mirror positions in the 3 base pairs flanking the target TA. Thus there does not seem to be any inherent preference for palindromic targets.

ANN**TA**NNT (N = any base) was the simplest consensus target sequence deduced for ISY100 (IRL/IRR) insertions. 31 out of 101 (30.7%) different insertion sites in this study and in PCC6803 genome and in the results of Urasaki *et al.* (2002) match this consensus sequence. Among these 31 sites, only 2 are fully palindromic over the 8 bp around the central TA, supporting the lack of preference for palindromic target sites. A total of 33 out of the 84 sequenced ISY100-IRL/IRR insertions into pH2 (39.3%), or 21 out of the 58 different target sites (36.2%) match up to the consensus sequence ANN**TA**NNT (Table 5.2, Figure 5.4A, 5.5A), and 15 out of the 36 available ANNTANNT sequences in pH2 (41.7%) have at least one insertion. Two out of the 14 ISY100-IRL/IRL and ISY100-IRR/IRR insertions were also into sites matching this consensus sequence

Table 5.1 Number and percentage occurrence of bases flanking duplicated target dinucleotides

A) Independent target sequences used by IRL-IRR transposons *in vivo* and *in vitro* in this study (Figure 5.6A).

Base	positions relative to target dinucleotides					
	-3	-2	-1	+1	+2	+3
A	37 (63.8%)	21 (36.2%)	25 (43.1%)	20 (34.5%)	13 (22.4%)	11 (19.0%)
C	3 (5.2%)	8 (13.8%)	6 (10.3%)	10 (17.2%)	19 (32.8%)	5 (8.6%)
G	0 (0%)	17 (29.3%)	8 (13.8%)	6 (10.3%)	6 (10.3%)	6 (10.3%)
T	18 (31.0%)	12 (20.7%)	19 (32.8%)	22 (37.9%)	20 (34.5%)	36 (62.1%)

B) Independent target sequences used by IRL-IRR transposons in this study, and the study of Urasaki *et al.* (2002) and in the PCC6803 genome (Figure 5.6A).

Base	positions relative to target dinucleotides					
	-3	-2	-1	+1	+2	+3
A	56 (55.4%)	33 (32.7%)	43 (42.6%)	34 (33.7%)	27 (26.7%)	23 (22.8%)
C	5 (5%)	13 (12.9%)	8 (7.9%)	18 (17.8%)	25 (24.8%)	10 (9.9%)
G	8 (7.9%)	25 (24.8%)	14 (13.9%)	8 (7.9%)	9 (8.9%)	13 (12.9%)
T	32 (31.7%)	30 (29.7%)	36 (35.6%)	41 (40.6%)	40 (39.6%)	55 (54.5%)

Table 5.2 Distribution of ISY100 insertions in pH2 from this study

A) Insertion of ISY100 IRL/IRR

Position (bp)	ANNTANNT	Orientation 1 (+)[1]		Orientation 2 (–)[1]		Total
		in vivo	*in vitro*	*in vivo*	*in vitro*	
16				1		1
29				1		1
64	Y	1				1
66	Y				1	1
68	Y			1		1
97			1			1
108				1		1
114			1			1
186		1		1		2
232				1		1
348			1			1
429	Y		1			1
484					1	1
488					1	1
491			1			1
595		1				1
619			1			1
631				1		1
937	Y			1		1
1245	Y	2				2
1537					1	1
1539		1	3		8	12
1794				1		1
1830					1	1
1955					1	1
1985				1		1
2023		1				1
2182		2				2
2224	Y				1	1
2265	Y	2	1		1	4
2345	Y		1			1
2432	Y		1	1		2
2620			1		2	3
2908*					1	1
3201	Y		1		1	2
3211	Y		1	4	1	6
3274				1		1
3366					1	1
3495		1				1
3678	Y	2				2
3744	Y		1		1	2
3846					2	2
3849				1		1
4048	Y		2		4	6
4157			2			2
4176				1		1
4996		1				1
5122*					1	1
5182			1			1
Total		**15**	**21**	**18**	**30**	**84**

B) Insertion of ISY100 IRL/IRL (*in vivo*)

Position (bp)	ANNTANNT	Orientation 1 (+)[1]	Orientation 2 (–)[1]
2545			1
3861	Y	1	
4196			1
Total		**1**	**2**

C) Insertion of ISY100 IRR/IRR (*in vivo*):

Position (bp)	ANNTANNT	Orientation 1 (+)[1]	Orientation 2 (–)[1]
173			1
527			1
635			1
1546		1	
1558			1
2224			1
2269	Y	1	
2876			1
3495			2
4176		1	
5227		1	
Total		**4**	**8**

(1) "+" indicates that the insertion is in the same direction (from IRL to IRR) as the Amp^r gene in pH2. For insertions with two identical IR sequences, "+" indicates that the Kan^r gene in the mini-ISY100 is in the same orientation as the Amp^r gene of pH2. "–" indicates insertions in the opposite orientation. All insertions are into TA targets on pH2 except those labelled with an asterisk (*), which inserted into TT/AA dinucleotides instead.

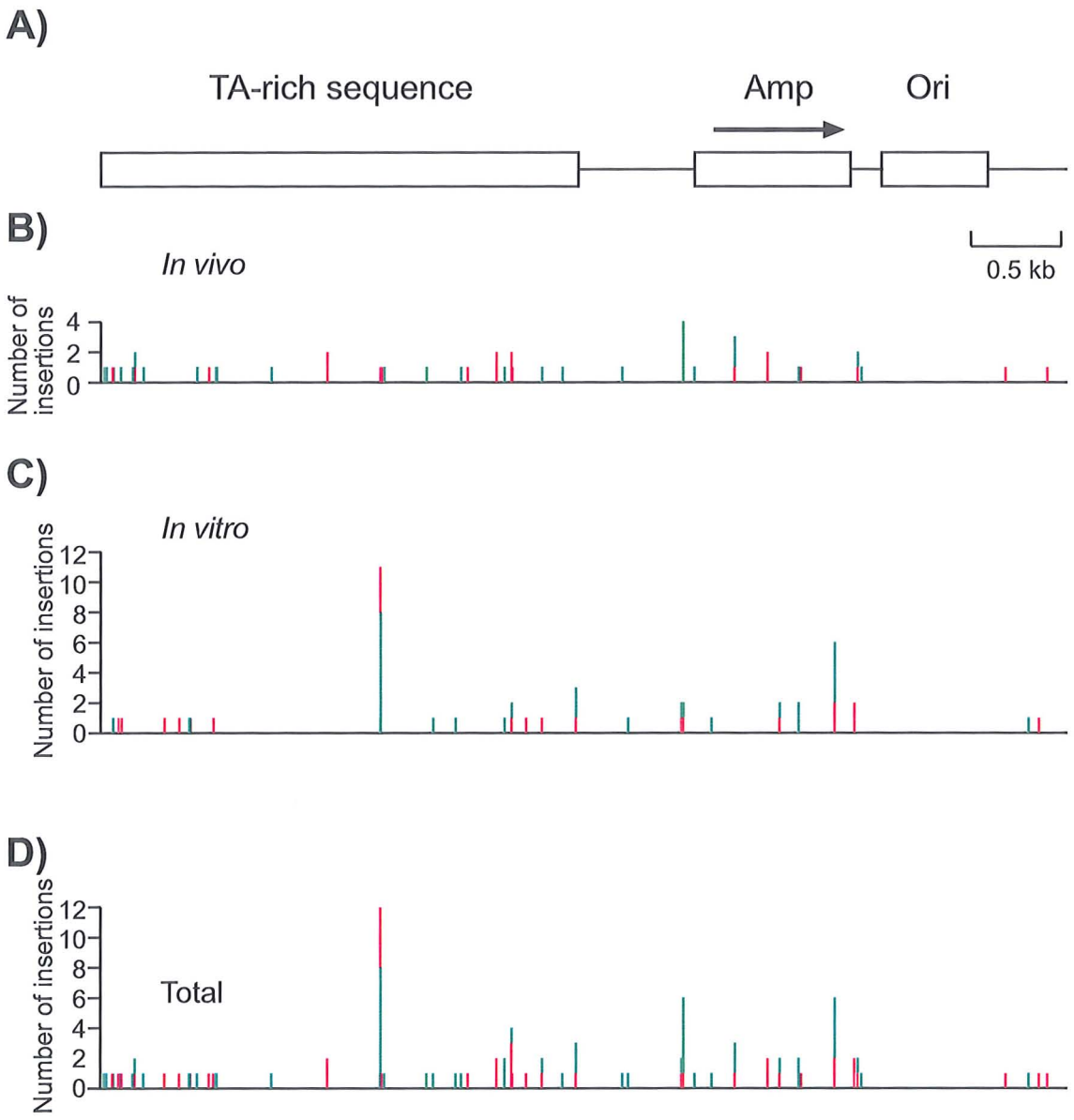


Figure 5.7 Distribution of insertions of mini-ISY100 on pH2. A) Structure of pH2. The length of the genes are drawn to scale. **B), C)** and **D)** show the distribution of insertions on pH2 from *in vivo* transposition assays, *in vitro* transposition assays and both assays. The position of insertion is shown along the X-axis and the number of insertions at each position is indicated by the height of the bar. Insertions in the "+" orientation are shown in red, those in the "–" insertion are shown in green (details in Table 5.1).

Table 5.3 Absence of bias toward a palindromic sequence for ISY100 insertion sites

A) Independent target sequences used by IRL-IRR transposons *in vivo* and *in vitro* in this study.

position in sequence	+1		+2		+3	
	T	A	T	C	T	A
Percentage (and number) of occurrence	37.9% (22/58)	34.5% (20/58)	34.5% (20/58)	32.8% (19/58)	62.1% (36/58)	19.0% (11/58)
Percentage (and number) of base occurrence at the mirror position[a]						
A	50.0% (11/22)		30.0 % (6/20)		58.3% (21/36)	
T		40% (8/20)				9.1% (1/11)
G				42.1% (8/19)		

B) Independent target sequences used by IRL-IRR transposons in this study, Urasaki *et al.* (2002) and in the PCC6803 genome.

position in sequence	+1		+2		+3	
	T	A	T	A	T	A
Percentage (and number) of occurrence	40.6% (41/101)	33.7% (34/101)	39.6% (40/101)	26.7% (27/101)	54.5% (55/101)	22.8% (23/101)
Percentage (and number) of base occurrence at the mirror position[a]						
A	41.5% (17/41)		27.5 % (11/40)		56.4% (31/55)	
T		35.3% (12/34)		29.6% (8/27)		17.4% (4/23)

[a] The mirror positions for positions +1, +2, +3 are positions -1, -2, -3 respectively.

(Table 5.2). Thus the consensus sequence may have some use in predicting which sites will be used by ISY100.

The impact of DNA structure, including DNA bending and flexibility, has been shown to be important for the target selection for several transposons (Kuduvalli *et al.*, 2001; Liu *et al.*, 2005; Manna *et al.*, 2004; Pribil and Haniford, 2003; Vigdal *et al.*, 2002). In this study, ISY100 preferentically inserted into some 'hot' target sites. The target that was used most frequently (at position 1539 bp on pH2) with 9 insertions in one orientation and 3 insertions in the other, had the sequence ATA**TA**TTA, which does not fit the consensus sequence. Perhaps the very AT rich nature of this sequence gives this sequence structural features that make up for the lack of a T at position +3. However, more work is needed to investigate whether the local DNA structure has an effect on the target preference of ISY100 insertion.

For all of the transposition events where the donor was supercoiled plasmid DNA, whether the reaction was *in vivo* or *in vitro*, TA was used as target and duplicated in the reaction. However, when transposition was carried out *in vitro* using AcuI-cleaved mini-ISY100, two insertions were isolated where TT (or AA) was used as a target instead of TA. In these cases the TT (AA) target sequence was duplicated after insertion (Figure 5.5 A-III). In two other cases, TA was used as a target but while the normal TA was found adjacent to the left end of the transposon, only A was found at the right end (Figure 5.5 A-II, III). It appears in these cases that the strand transfer reaction occurred at a TA but at positions staggered by 1 bp instead of the normal 2 bp. It seems that transposition following assembly of transposase onto the pre-cleaved transposon is in some way less precise than the normal reaction. Perhaps the excision reaction ensures that transposase assembles properly onto the transposon ends.

5.3 Targeted transposition catalysed by transposase-zinc finger fusion proteins

To find out whether ISY100 could be used for targeted transposition, the zinc finger DNA-binding domain of the mouse transcription factor Zif268 was fused to the C-terminus of transposase to create Z-transposase. Transposition catalysed by Z-transposases was then studied *in vivo* using a modification of the transposition assay described in chapter III.

5.3.1 Designs of chimeric Zif268-transposases and constructions of donor plasmids for *in vivo* transposition

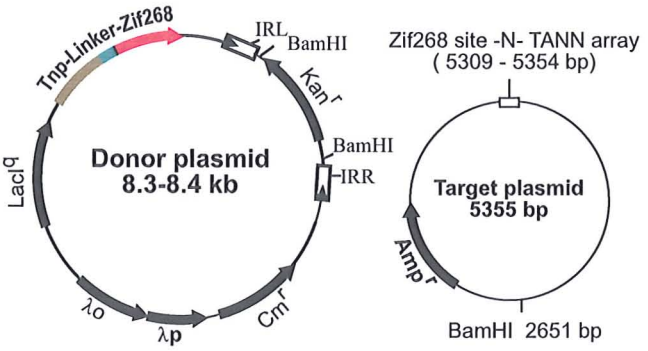
In experiments to alter the substrate specificity of Tn3 resolvase, Akopian *et al.* fused the DNA-binding domain of Zif268 to the C-terminus of the intact Tn3 resolvase, or to the C-terminus of the N-terminal catalytic domain of resolvase (Akopian *et al.*, 2003). The genes for the Zif268 DNA-binding domain with various N-terminal linkers were readily available from Aram Akopian as an EagI-KpnI restriction fragment. A transposon expression vector was therefore constructed with an EagI restriction site in the correct frame at the 3' end of the transposase gene and a KpnI site after the new transposase stop codon (Figure 5.8B). This expression vector encodes a protein with three extra amino acids (YGR) after the normal transposase C-terminus. Three different EagI-KpnI fragments were then inserted into this plasmid to make plasmids expressing ISY100 transposase fused to the Zif268 DNA-binding domain by three different protein linkers (Figure 5.8B). Two linkers were short repetitive glycine- and serine-rich peptides designed to be flexible ($[GSG]_3$ and $[GSG]_4$) (Figure 5.8B). The third linker consisted of the entire C-terminal DNA-binding domain (residues 141-185) of Tn3 resolvase and a few extra amino acids (Figure 5.8B). The genes for these three transposase-Zif268 fusion proteins ("Z-transposase": Tnp-$[GSG]_3$-Zif268, Tnp-$[GSG]_4$-Zif268, and Tnp-Tn3-Zif268) as well as the gene encoding transposase with three extra amino acids at the C-terminus (Tnp-YGR), were used to replace the wild-type transposase gene in the λ-dv based transposon donor plasmid pXF116 (Figure 5.8A). These plasmids were then used for *in vivo* transposition assays with the target plasmids described below.

5.3.2 Design of target plasmids for Z-transposase

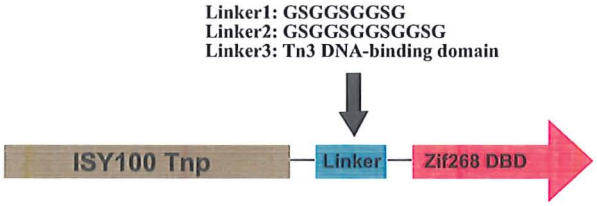
Zif268 binds asymmetrically to its binding site, with finger-1 bound to the 3' end and finger-3 bound to the 5' end of the Zif268 binding site 5'- GCGTGGGCGT -3'. Z-transposase bound to a target site by a C-terminal Zif268 DNA-binding domain will be tethered with a specific orientation, and might catalyse insertions only to one side of the Zif268 binding site. Furthermore, the transposase might be constrained by Zif268 such that it can only catalyse insertions at a fixed distance from the Zif268 binding site, or on only one face of the DNA helix relative to the Zif268 binding site. Transposition into sites at the wrong distance, or on the wrong face of the DNA helix, would require unfavourable distortion of either the DNA or the protein.

Target sites were therefore designed with potential TA target sites at all possible distances (from 1 bp to 36 bp) from both possible orientations of the Zif268 binding site.

A)



B)



Tnp-YGR

NdeI EagI KpnI BamHI

catATGGCTTACAGTTTAGATTTAAGG.....TACTGTGGCGTTTACGGCCGTtagggtaccggatcc

M A Y S L D L R .....Y C G V Y G R *

Tnp

Linker 1 $(GSG)_3$

EagI SpeI

cggccgCAGGCGTGGCTCTGGCGGTTCCGGCGGCTCTGGTACTAGT

G R R R **G S G G S G G S G** T S

Linker 2 $(GSG)_4$

EagI SpeI

cggccgCAGGCGTGGCTCTGGCGGTTCCGGCGGCTCTGGTGGCAGTGGTACTAGT

G R R R **G S G G S G G S G G S G** T S

Linker 3 Tn3 DBD

EagI

cggccgCAGGCGTACCGTGGACAGGAACGTCGTGCTGACGCTTCATCAGAAGGGCACTGGT

G R R R T V D R N V V L T L H Q K G T G

GCAACGGAAATTGCTCATCAGCTCAGTATTGCCCGCTCCACGGTTTATAAAATTCTTGAA

A T E I A H Q L S I A R S T V Y K I L E

SpeI

GACGAAAGGGCGAGCTCAGACCCAACTTCACAGACTAGT

D E R A S S D P T S Q T S

Zif268 -DBD

SpeI

actagtGAACGTCCGTATGCTTGTCCGGTTGAATCCTGTGACCGTCGTTTCTCGAGATCA

T S **E R P Y A C P V E S C D R R F S R S**

GACGAACTGACCCGTCACATCCGTATCCACACCGGTCAGAAACCGTTCCAGTGCCGTATA

D E L T R H I R I H T G Q K P F Q C R I

TGCATGAGGAACTTCTCCAGATCTGACCACCTGACCACCCACATCCGTACGCACACTGGC

C M R N F S R S D H L T T H I R T H T G

GAAAAACCGTTCGCATGCGATATCTGCGGTCGTAAATTCGCGCGCTCTGATGAACGTAAA

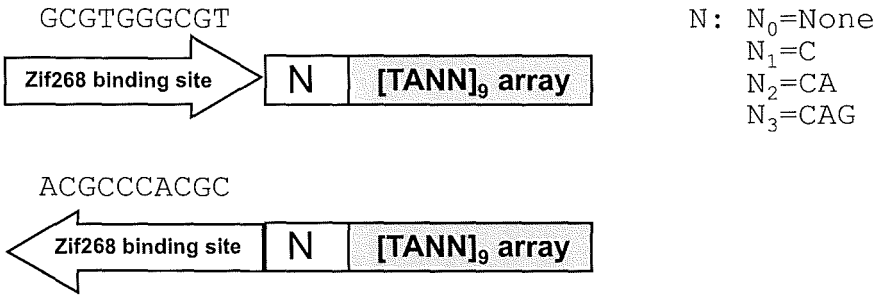
E K P F A C D I C G R K F A R S D E R K

KpnI BamHI

CGTCACACCAAAATCCACCTGCGTCAGAAAGATTCGAGCTCAtgagggtaccggatcc

R H T K I H L R Q K D S S S *

C)



Zif268 site -N- TANN array

```
Zif268 binding site:    5'-GCGTGGGCGT-3'
                              10        20        30
                        ----:----|----:----|----:----|----:-
TANN-array:   5'-TAGATATCTAAGTAAATACGTATGTATGTATATAGA-3'
                  1   2   3   4   5   6   7   8   9
```

Figure 5.8 Donor and target plasmids used to assay for *in vivo* transposition catalysed by ISY100 transposase-Zif268 fusion proteins. A) Map of the donor and target plasmids. Donor plasmids are pXF116-like plasmids, encoding transposase-Zif268 fusion proteins. Target plasmids are pH2 with target sequences, consisting of the Zif268 recognition sequence adjacent to a $[TANN]_9$ array. **B)** Sequences of transposase-Zif268 fusion proteins. Restriction sites are indicated above the DNA sequences and the amino acid sequences of serine-rich glycine-rich linkers, Tn3 resolvase DNA-binding domain, and the Zif268 DNA-binding domain are shown in bold and underlined. **C)** Structures and sequences of the Zif268- $[TANN]_9$ target sites inserted into pH2.

This was done by placing a repeated array of 9 copies of the sequence TANN (where N was a randomly chosen base) adjacent to the Zif268 binding site, at four different spacings (0, 1, 2 and 3 bp) (Figure 5.8C). The sequence $[TANN]_9$ has the added advantage that all but the first and last TA dinucleotides have an A at the -3 and a T at the +3 position and therefore conform to the consensus preferred target site. A total of 8 different target sequences, with both possible orientations of the Zif268 binding site and four different spacings between the Zif268 binding site and the TA array, were inserted as double stranded oligos into pH2 to construct Amp^r target plasmids (Figure 5.8A). These plasmids are referred to here as pZ±X where + or – indicates the orientation of the Zif268 binding site and X = 0, 1, 2, or 3 indicates the length of the spacer. All of these plasmids were based on pH2 so that transposition could either go into the TA array adjacent to the Zif268 binding site, or the large TA-rich region of *A. nidulans* DNA in pH2.

5.3.3 Z-transposases catalyse transposition to one side of the Zif268 binding site.

In the first experiment (Experiment A, Table 5.4), 32 separate *in vivo* transposition assays were set up by co-transforming one λ-dv-based donor plasmid and one target plasmid into DH5α. Eight different target plasmids were used, with both orientations of the Zif268 binding site and 4 different spacings between the Zif268 binding site and the TA array. Four donor plasmids were used, differing only in the transposase gene. Three donor plasmids carried the three different Z-transposase genes, while the fourth carried transposase with three additional amino acids (Tnp-YGR), but no Zif268 DNA-binding domain. A standard *in vivo* transposition assay was then carried out. 10 single colonies containing donor and target plasmids were pooled and grown for 18 hours in liquid culture, diluted 1 in 1000 and grown a further 18 hours. DNA was isolated from these cells and plasmid molecules in which the mini-ISY100 had jumped into the target plasmid were isolated by transformation into the λ lysogen strain DS964 and selection for kanamycin resistance. Twenty colonies were picked from each transformation and checked for chloramphenicol resistance, and the transposition frequency was calculated in the usual way. Transposition catalysed by the control transposase (Tnp-YGR) occurred with a frequency ranging from 0.42- 4.00×10^{-3} (mean ± 1 standard deviation= $2.11 \pm 1.13 \times 10^{-3}$) into the 8 different substrates (Table 5.4A). Transposition catalysed by Tnp-$(GSG)_3$-Zif268 and Tnp-$(GSG)_4$-Zif268 occurred with a reduced frequency ($0.70 \pm 0.38 \times 10^{-3}$ and $0.76 \pm 0.35 \times 10^{-3}$ respectively), while the transposition frequency obtained with Tnp-Tn3-Zif268 ($0.47 \pm 0.35 \times 10^{-3}$) was approximately 4-fold lower than that obtained with Tnp-YGR.

Table 5.4 *In vivo* transposition frequency ($\times 10^{-3}$) of Z-transposase

Experiment A

Target plasmid	Transposases			
	Tnp-YGR	Tnp-$(GSG)_3$-Zif268	Tnp-$(GSG)_4$-Zif268	Tnp-Tn3-Zif268
pZ+1	1.87	0.27	0.38	0.58
pZ+2	0.42	0.11	0.80	0.13
pZ+3	1.70	0.54	1.28	0.84
pZ+4	2.17	0.85	1.08	1.13
pZ-1	4.00	0.82	0.35	0.33
pZ-2	2.98	1.22	0.98	0.33
pZ-3	2.71	1.07	0.81	0.22
pZ-4	1.03	0.69	0.41	0.19
Average (± 1 SD)	2.11 ± 1.13	0.70 ± 0.38	0.76 ± 0.35	0.47 ± 0.35

Experiment B

Target plasmid	Transposases			
	Wild-type Tnp	Tnp-$(GSG)_3$-Zif268	Tnp-$(GSG)_4$-Zif268	Tnp-Tn3-Zif268
pZ+1	0.80	0.31	0.50	0.09
pZ+2	1.08	0.34	0.76	0.06
pZ+3	1.55	0.19	0.28	0.09
pZ+4	1.46	0.27	1.34	0.06
pZ-1	1.91	0.32	0.60	0.05
pZ-2	5.77	0.34	0.61	0.09
pZ-3	4.20	0.37	0.56	0.06
pZ-4	1.75	0.27	0.48	0.04
pH2	2.81	0.10	0.24	0.04
Average (± 1 SD)	2.37 ± 1.63	0.28 ± 0.09	0.60 ± 0.32	0.06 ± 0.02

In vivo transposition assays to test the transposition activity of chimeric transposases. Both experiments A and B were started from multiple (10) single colonies, carrying both donor and target plasmid and were carried out as described in chapter II. In experiment A 1 µl of plasmid DNA was electroporated into DS964, whereas in experiment B the DNA used in the electroporation was standardised to 100 ng.

Thus fusing the Zif268 DNA-binding domain to ISY100 transposase does not abolish transposition, although the linker consisting of the resolvase C-terminal domain substantially reduced the transposition frequency.

To get a reasonable estimate of the frequency of insertions into the designed target adjacent to the Zif268 binding site, it was decided that 20 colonies should be examined for each transposition assay. The approximate location of mini-ISY100 insertions catalysed by Z-transposases could be determined by restriction digestion of plasmids purified from individual DS964 transformants. However, to avoid purifying plasmid DNA from 640 individual colonies, an initial PCR-based screen was carried out on transformant colonies. Twenty colonies from each transposition assay were picked and used directly as template for PCR as described in Chapter II. Two combinations of primers were used, to detect both possible orientations of the mini-ISY100 in the TANN array of the target plasmids (Figure 5.9A). PCR products of approximately 720 bp with primers Rev_308 and KanR, and approximately 530 bp with primers Rev_308 and KanL indicated insertions close to the Zif268 binding site. This process identified 143 plasmids out of the 640 screened, containing insertions close to the Zif268 binding site. DNA was prepared from these 143 colonies, digested with BamHI and run on an agarose gel to verify the PCR results. The mini-ISY100 used in this assay contains two BamHI sites flanking the Kan^r gene, 30 bp in from each end. The 5.4 kb target plasmids all contain a single BamHI site almost exactly 2.7 kb away from the Zif268 binding site. Therefore plasmids with insertions close to the Zif268 binding site will give a 1.3 kb Kan^r fragment and two co-migrating fragments of 2.7 kb (Figure 5.10). Those plasmids giving this restriction pattern were sequenced using both KanL and KanR primers. Sequencing results showed that all the insertions were into TA targets and were flanked by TA at both sides. 121 out of the 143 plasmids, identified by PCR and BamHI digestion as carrying insertions close to the Zif268 binding site, carried insertions in the $[TANN]_9$ array. The other 22 carried insertions nearby, but outside the $[TANN]_9$ array.

The number of insertions into the TANN array differed for the different combinations of donor and target plasmids. For the control transposase (Tnp-YGR), between 1 and 3 out of 20 insertions were into the $[TANN]_9$ target (Table 5.5, 5.6). Thus the $[TANN]_9$ array is used as target approximately 9% of the time, even in the absence of a Zif268 DNA-binding domain. Tnp-$(GSG)_3$-Zif268 and Tnp-$(GSG)_4$-Zif268 gave similar results, except that Tnp-$(GSG)_4$-Zif268 gave 10 out of 20 inserts into the TA target array of one substrate, pZ+1 (Table 5.5, 5.6). Tnp-Tn3-Zif268 gave much more promising results. For the "–"

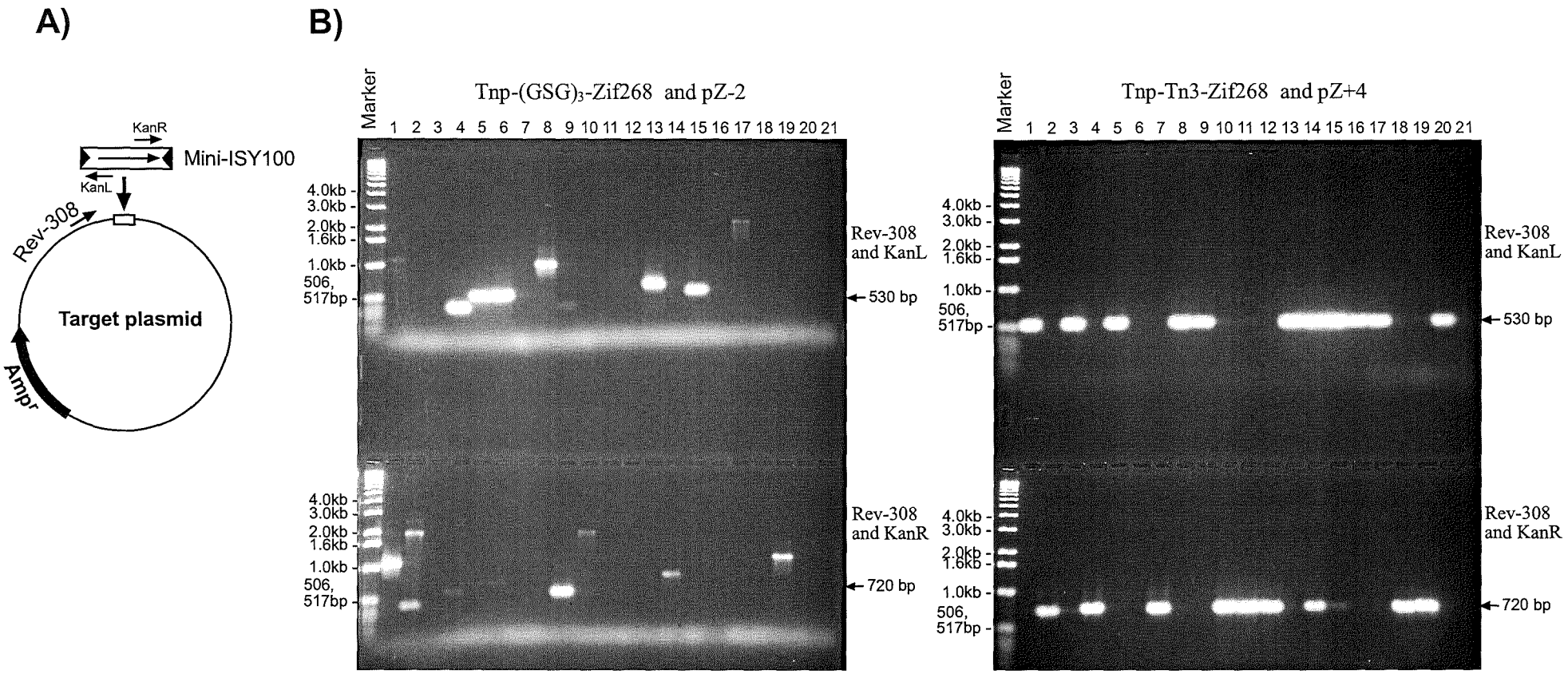


Figure 5.9 Colony PCR to screen transposition events for insertions close to the Zif268 binding site. A) Diagram showing positions of primers on the target plasmid and mini-ISY100. Primer Rev-308 is located upstream of the Zif268 binding site and TA array (white box). Primers KanL and KanR are located close to the ends of the mini-ISY100. **B)** 1.2% agarose gel electrophoresis of PCR products obtained from single DS964 transformant colonies. Lanes 1-20 of each gel show PCR products using 20 kanamycin resistant single colonies from *in vivo* transposition assays. The donor plasmid and target plasmid used are indicated above each gel. The PCR primers used are indicated to the side of each gel. Lane 21 shows a negative control PCR reaction.

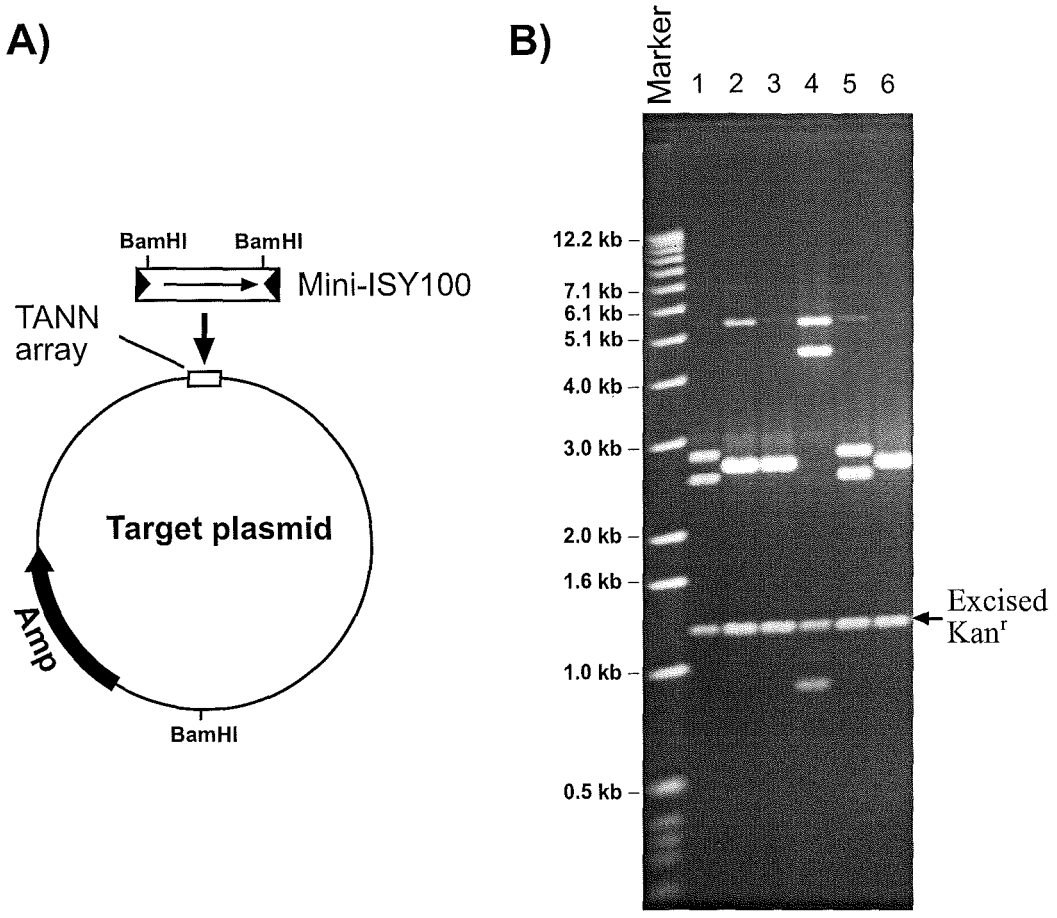


Figure 5.10 BamHI digestion to screen potential target specific transposition. A) Diagram shows insertion of a mini-ISY100 into the $[TANN]_9$ array on the target plasmid. The relative positions of BamHI sites on the mini-ISY100 and the target plasmid are shown. **B)** 1.2% agarose gel electrophoresis of the BamHI digestion products of candidate plasmids carrying an insertion of mini-ISY100 from an *in vivo* transposition assay with Tnp-$(GSG)_3$-Zif268 and pZ-2.These plasmids were purified from Kan^r Cm^s single colonies which gave PCR products close to the desired size. Lanes 1 - 6 correspond to lanes 4, 5, 6, 7, 9 and 15 respectively on Figure 5.9.

Table 5.5 Target selection in transposition by Z-transposase (Experiment A)

Target plasmid	Number of insertions into TANN array			
	Tnp-YGR	Tnp-$(GSG)_3$-Zif268	Tnp-$(GSG)_4$-Zif268	Tnp-Tn3-Zif268
pZ+1	1/20	3/20	10/20 (P<0.05)	11/20 (P<0.05)
pZ+2	3/20	2/20	2/20	7/20
pZ+3	2/20	1/20	0/20	18/20 (P<0.05)
pZ+4	2/20	2/20	1/20	19/20 (P<0.05)
pZ-1	3/20	2/20	4/20	4/20
pZ-2	2/20	3/20	0/20	1/20
pZ-3	1/20	2/20	4/20	3/20
pZ-4	1/20	3/20	3/20	1/20

A *chi-squared* test was performed to compare the number of insertions into the $[TANN]_9$ array mediated by chimeric proteins with the number mediated by Tnp-YGR. Those giving P<0.05 are indicated in the table.

Table 5.6 Details of the *in vivo* transposition of Z-transposases (Experiment A)

Tnp	Target plasmid	Amp^r colonies ($\times 10^4$)	Kan^r colonies	Transposition frequency ($\times 10^{-3}$)	No. of insertions into TA array	
					Total No. in the 20 general insertions	No. of insertions into specific TA[(1)] (No.: position of TA)
Tnp-YGR	pZ+1	177	3308	1.87	1/20	1 : 2nd TA
	pZ+2	314	1320	0.42	3/20	2 : 7th TA 1 : 7th TA (+)
	pZ+3	187	3184	1.70	2/20	1 : 4th TA 1 : 8th TA
	pZ+4	230	5000	2.17	2/20	1 : 6th TA (+) 1 : 8th TA
	pZ-1	110	4400	4.00	3/20	1 : 3rd TA 1 : 4th TA 1 : 7th TA (+)
	pZ-2	83	2476	2.98	2/20	2 : 7th TA (+)
	pZ-3	79	2140	2.71	1/20	1 : 9th TA
	pZ-4	102	1052	1.03	1/20	1 : 4th TA
Tnp-$(GSG)_3$-Zif268	pZ+1	240	647	0.27	3	3 : 4th TA
	pZ+2	139	156	0.11	2/20	1 : 2nd TA 1 : 6th TA
	pZ+3	171	921	0.54	1/20	1 : 7th TA
	pZ+4	101	863	0.85	2/20	2 : 2nd TA
	pZ-1	87	717	0.82	2/20	2 : 7th TA
	pZ-2	80	973	1.22	3/20	1 : 2nd TA 2 : 5th TA
	pZ-3	84	895	1.07	2/20	1 : 7th TA (+) 1 : 7th TA
	pZ-4	69	474	0.69	3/20	1 : 4th TA (+) 1 : 6th TA 1 : 7th TA (+)

(Continued on next page)

Tnp	Target plasmid	Amp^r colonies ($\times 10^4$)	Kan^r colonies	Transposition frequency ($\times 10^{-3}$)	No. of insertions into TA array	
					Total No. in the 20 general insertions	No. of insertions into specific TA[(1)] (No.: position of TA)
Tnp-$(GSG)_4$-Zif268	pZ+1	132	500	0.38	10/20	1 : 2nd TA 3 : 4th TA 5 : 7th TA 1 : 7th TA (+)
	pZ+2	92	738	0.80	2/20	1 : 2nd TA 1 : 2nd TA (+)
	pZ+3	68	870	1.28	0/20	–
	pZ+4	122	1315	1.08	1/20	1 : 2nd TA
	pZ-1	114	394	0.35	4/20	3 : 2nd TA 1 : 8th TA
	pZ-2	98	959	0.98	0/20	–
	pZ-3	74	601	0.81	4/20	2 : 2nd TA 1 : 4th TA (+) 1 : 7th TA (+)
	pZ-4	37	151	0.41	3/20	1 : 5th TA (+) 1 : 6th TA 1 : 7th TA
Tnp-Tn3-Zif268	pZ+1	59	343	0.58	11/20	1 : 3rd TA (+) 2 : 4th TA 2 : 6th TA (+) 5 : 7th TA 1 : 7th TA (+)
	pZ+2	75	97	0.13	7/20	5 : 5th TA 2 : 7th TA
	pZ+3	92	770	0.84	18/20	11 :2nd TA 4 : 2nd TA (+) 3 : 7th TA
	pZ+4	69	780	1.13	19/20	10 : 2nd TA 7 : 2nd TA (+) 2 : 7th TA
	pZ-1	61	204	0.33	4/20	3 : 2nd TA (+) 1 : 7th TA
	pZ-2	28	91	0.33	1/20	1 : 7th TA
	pZ-3	72	160	0.22	3/20	3 : 7th TA(+)
	pZ-4	216	410	0.19	1/20	1 : 7th TA

(1): (+) indicates that the mini-ISY100 is inserted with the Kan^r gene in the same orientation as the Amp^r gene on the target plasmid. All other insertions have the opposite orientation.

orientation of the Zif268 binding site, the $[TANN]_9$ target was used approximately 11% of the time, roughly the same frequency as in the absence of the Zif268 DNA-binding domain. However, for the "+" orientation, 18 out of 20 insertions into pZ+3 and 19 out of 20 insertions into pZ+4 were in the $[TANN]_9$, while 11 out of 20 and 7 out of 20 insertions into pZ+1 and pZ+2 were in the $[TANN]_9$ array. Thus it appears that Tnp-Tn3-Zif268 fusion protein catalyses insertions with increased frequency only at one side of the Zif268 binding site (Table 5.5, 5.6).

Next, the exact locations of insertions into the $[TANN]_9$ arrays were analysed. The locations of insertions catalysed by the four different transposases into the eight different target sites are shown in Table 5.6 and summarised graphically in Figure 5.11. Analysing the 15 different insertions catalysed by Tnp-YGR, the TA closest to the Zif268 binding site was never used, and nor was the fifth, while the seventh TA was used six times. If all of the insertions where there was no increase in insertion into the $[TANN]_9$ array are added into this analysis, it can be seen that the second, fourth and seventh TA from the Zif268 binding site are preferred sites of integration. Perhaps the flanking sequences of these TAs have sequence or structural features that make them preferred targets for normal transposase. The first TA is never used as a target, perhaps because it does not have an A at the –3 position and/or because it is too close to the GC-rich Zif268 binding site. Z-transposase might also block integration into this site by binding to the Zif268 binding site. When insertions catalysed by Tnp-Tn3-Zif268 into targets adjacent to the Zif268 binding site in the "+" orientation are examined, it appears that different TAs are favoured in the different substrates. The second TA was used for 15/18 and 17/19 insertions for pZ+3 and pZ+4 respectively, with the seventh TA used for the other insertions. The fifth TA in pZ+2 and the seventh TA in pZ+1 were also preferred. This suggests that insertions might be favoured at certain fixed distances from the Zif268 binding site possibly on one face of the DNA helix, though this needs to be verified by further repeats of this experiment.

To avoid the large number of PCRs and plasmid purifications described above, an alternative method was used to display the distribution of insertions catalysed by Z-transposases. After transformation into DS964, all of the colonies should contain different ISY100 insertions into the target plasmid. Colonies were therefore washed off the transformation plates and plasmid DNA was purified from these pooled colonies. This DNA was cleaved with BamHI and run on an agarose gel. All of the plasmid molecules in these pools should give a 1.3 kb Kan^r fragment from the mini-ISY100. Plasmids with an insertion adjacent to the Zif268 binding site will give two bands of the same size at 2.7 kb,

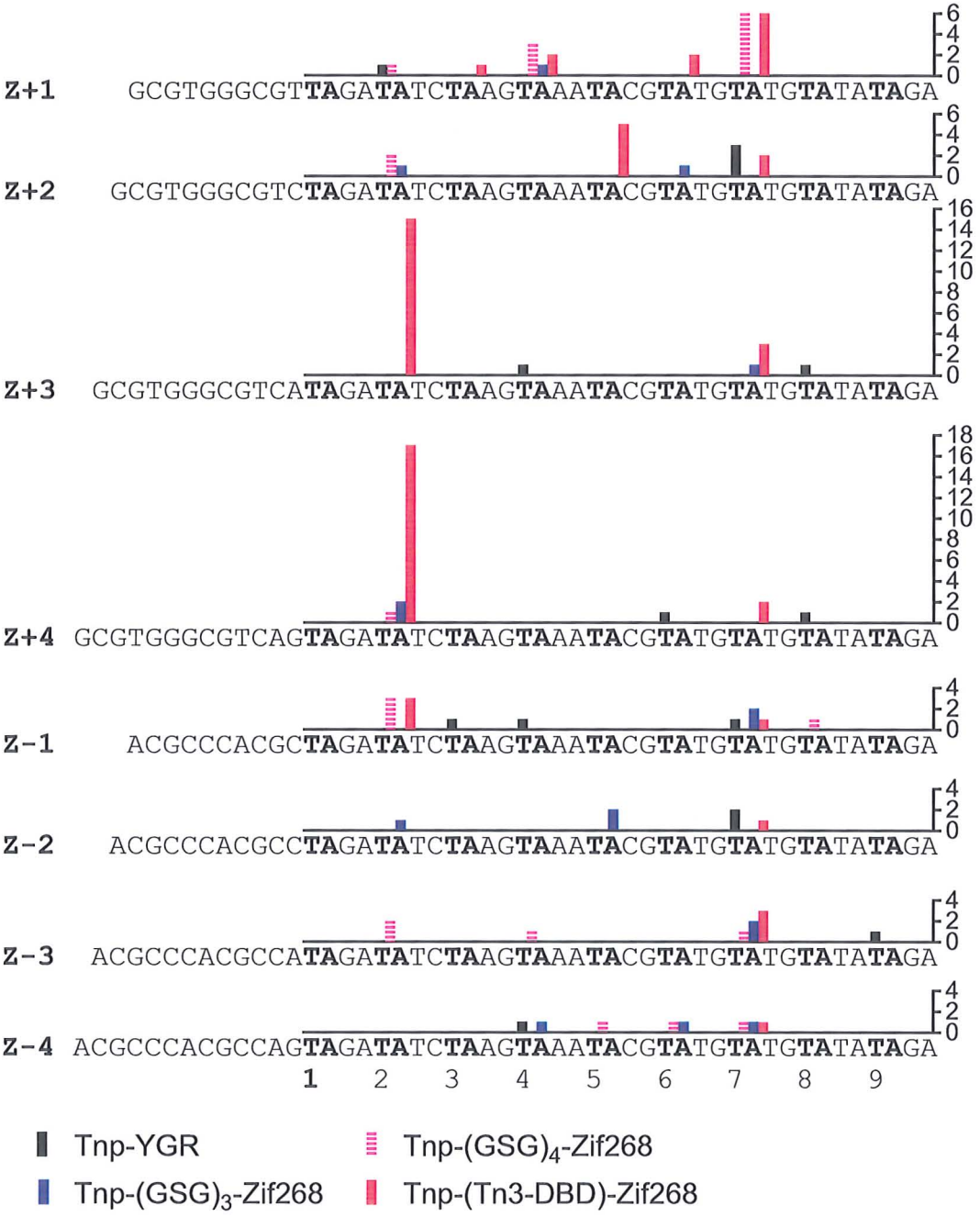


Figure 5.11 Distribution of insertions in the [TANN]$_9$ array from *in vivo* Z-transposition assays. Histogram to show number of insertions into each TA in the target sites.

while all of the other plasmids will give two bands, one larger than 2.7 kb and the other smaller. The mixture will therefore give a ladder of different sized fragments, with the relative intensity of the band at 2.7 kb position indicating the proportion of insertions in, or close to, the TANN array. The results of this analysis from the same *in vivo* transposition assays described above (Experiment A) are shown in Figure 5.12. As expected, a ladder of bands representing insertions at different locations is seen in all of the lanes. Tnp-YGR, Tnp-$(GSG)_3$-Zif268 and Tnp-$(GSG)_4$-Zif268 give complex ladders of bands, with the 2.7 kb band making up approximately 10-20% of this ladder. Thus only 10-20% of the plasmid molecules in these mixtures contain insertions in or close to the TANN designed target. Tnp-Tn3-Zif268 gave less complex ladders of bands, perhaps correlating with the low transposition frequency and a smaller total number of transformant colonies in the pools. For the targets with the Zif268 binding site in the "+" orientation, but not for those with this site in the "–" orientation, the 2.7 kb band makes up a large proportion of the total amount of DNA, consistent with the results shown in Table 5.6. This technique therefore gives a good idea of the proportion of insertions into the designed target, and the results are generally consistent with those obtained by purifying large numbers of individual plasmids.

Another thing that can be noted in Figure 5.12 is that a high proportion of molecules can sometimes contain insertions into a single site. For instance in the DNA purified from the transposition reaction with Tnp-$(GSG)_4$-Zif268 and pZ+2 (Figure 5.12A, lane 12), two very bright bands of approximately 2.3 and 3.0 kb, probably representing a single insertion into one site on pZ+2, are present. This result presumably reflects an early transposition event into this site in the original strain carrying both target and donor plasmid in DH5α.

The best way to distinguish between a genuine target preference and an early fortuitous transposition event giving multiple copies of the same insertion, is to carry out multiple independent repeats of the transposition assay. As a first step toward confirming the target specificity of the different Z-transposases, a second set of *in vivo* transposition assays (Experiment B) was carried out. This experiment used the same three Z-transposase donor plasmids and a control plasmid carrying the wild-type transposase (pXF116). As a control, pH2 was also used as a target for all 4 donor plasmids, so that a total of 36 assays were carried out. A standard *in vivo* transposition assay was carried out as before, giving similar transposition frequencies (Table 5.7). DNA was then prepared from pooled kanamycin resistant colonies, cut with BamHI and run on an agarose gel (Figure 5.13). The

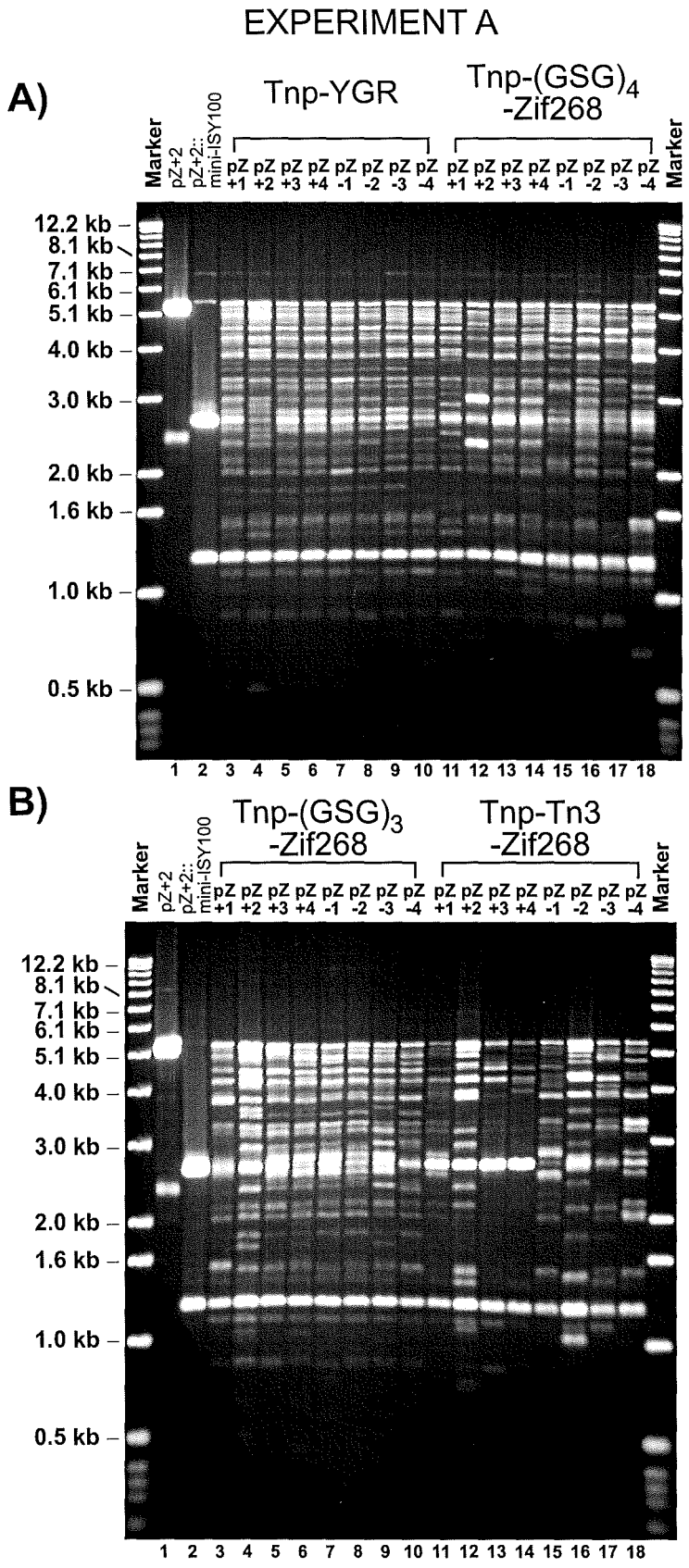


Figure 5.12 Electrophoretic display of insertions catalysed by Z-transposases. Plasmid DNA was purified from pooled Kan^r colonies from each individual transposition assay in experiment A. DNA was cleaved with BamHI and separated on a 1.2% agarose/TAE gel. The transposase and target plasmid used are indicated above each lane. Lane 1 of each gel contains BamHI-digested pZ+2. Lane 2 of each gel contains a transposition product with a single insertion in the $[TANN]_9$ array.

Table 5.7 Details of the *in vivo* transposition of Z-transposases (Experiment B)

		Target plasmids	Amp^r ($\times 10^4$)	Kan^r	Transposition frequency ($\times 10^{-3}$)	
Wild-type Tnp	1	pZ+1	169	1360	0.80	Mean ± 1SD: 2.37 ± 1.63
	2	pZ+2	145	1560	1.08	
	3	pZ+3	110	1710	1.55	
	4	pZ+4	110	1610	1.46	
	5	pZ-1	117	2240	1.91	
	6	pZ-2	137	7910	5.77	
	7	pZ-3	41	1720	4.20	
	8	pZ-4	68	1190	1.75	
	C_0	pH2	77	2167	2.81	
Tnp-$(GSG)_3$-Zif268	17	pZ+1	130	401	0.31	Mean ± 1SD: 0.28 ± 0.09
	18	pZ+2	188	646	0.34	
	19	pZ+3	203	391	0.19	
	20	pZ+4	187	508	0.27	
	21	pZ-1	228	732	0.32	
	22	pZ-2	147	494	0.34	
	23	pZ-3	267	976	0.37	
	24	pZ-4	300	815	0.27	
	C_2	pH2	160	154	0.10	
Tnp-$(GSG)_4$-Zif268	9	pZ+1	118	586	0.50	Mean ± 1SD: 0.60 ± 0.32
	10	pZ+2	108	817	0.76	
	11	pZ+3	88	248	0.28	
	12	pZ+4	72	963	1.34	
	13	pZ-1	94	566	0.60	
	14	pZ-2	85	515	0.61	
	15	pZ-3	152	850	0.56	
	16	pZ-4	111	535	0.48	
	C_1	pH2	143	349	0.24	
Tnp-Tn3-Zif268	25	pZ+1	208	182	0.09	Mean ± 1SD: 0.06 ± 0.02
	26	pZ+2	402	229	0.06	
	27	pZ+3	251	232	0.09	
	28	pZ+4	323	190	0.06	
	29	pZ-1	309	159	0.05	
	30	pZ-2	272	256	0.09	
	31	pZ-3	141	88	0.06	
	32	pZ-4	221	88	0.04	
	C_3	pH2	172	71	0.04	

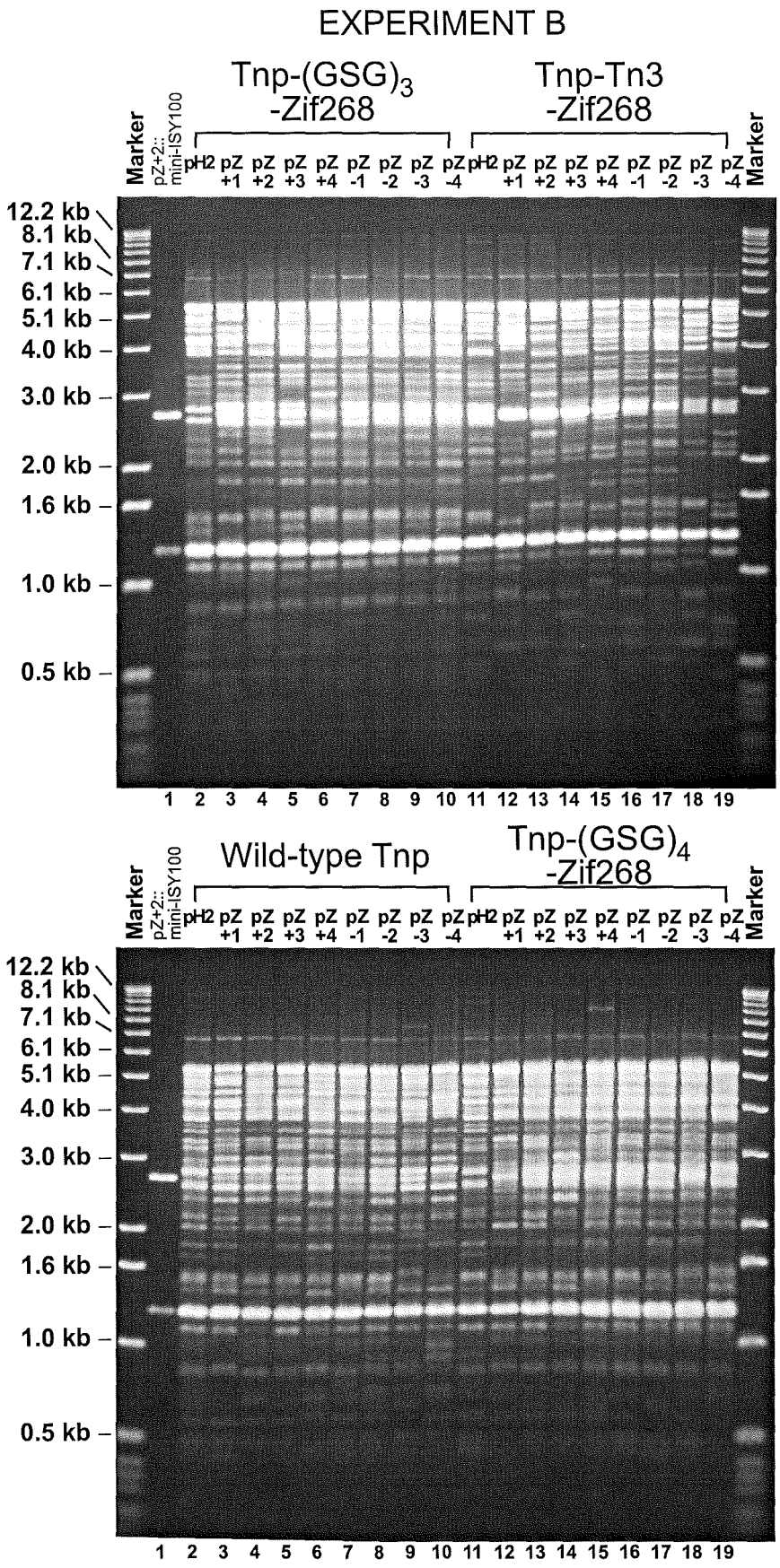


Figure 5.13 Electrophoretic display of insertions catalysed by Z-transposases. Plasmid DNA was purified from pooled Kan^r colonies from each individual transposition assay in experiment B.DNA was cleaved with BamHI and separated on a 1.2% agarose/TAE gel. The transposase and target plasmid used are indicated above each lane. Lane 1 of each gel contains BamHI-digested pZ+2. Lane 2 of each gel contains a transposition product with a single insertion in the $[TANN]_9$ array.

sizes of these pools range from 1190-7910 colonies for wild-type transposase, to 71-256 colonies for Tnp-Tn3-Zif268. With the wild-type transposase, it can be seen that adding the TANN array increases transposition to this area of all eight target plasmids. However, there is clear evidence for targeted transposition by Tnp-Tn3-Zif268, with a much brighter band at 2.7 kb for all of the substrates with the Zif268 binding site in the "+" orientation compared to the wild-type transposase. BamHI digestion of a randomly selected 20 Kan^r single colonies from the assays employing Tnp-Tn3-Zif268 together with target plasmid pZ+3 or pZ+4 suggested more than half (60% and 55% respectively) of the insertions were in the designed target array adjacent to the Zif268 binding site (Figure 5.14), although the exact target sites need to be further identified by sequencing. There was no evidence for a high level of targeting by either Tnp-$(GSG)_3$-Zif268 or Tnp-$(GSG)_4$-Zif268 in these assays.

5.4 Discussion

In this chapter, the target preference of ISY100 outside the duplicated target TA was first investigated based on the deduced target sites used by ISY100 *in vivo* and *in vitro* in this study, in the study of Urasaki *et al.* (2002) and the insertions in the sequenced genome of *Synechocystis* sp. PC6803. The consensus target sequences ADW**TA**WHT (where W = A or T, D = not C and H = not G) or more generally ANN**TA**NNT (where N = any base), with the duplicated target TA at the centre, was deduced from these data. The strongest preference was for T and A at +/-3, with a weaker preference for AT-rich sequence at positions ±1 and ±2 relative to the central TA.

Although the insertion sites might be biased by selection against insertions into essential genes in the genome, or by the relatively small number of available targets in the plasmids used, the single biggest limitation in this study was probably the small number of insertions studied. Only about 100 different insertion sites were analysed in this study, in contrast to more than 800 different sites for Tc1 and Tc3 in the studies of Preclin *et al.* (2003).

The deduced consensus target sequence (ANN**TA**NNT) was adapted and used in the design of $[TANN]_9$ target sites that could be placed next to the binding site for Zif268 to investigate the target specificity of ISY100 transposase-Zif268 fusion proteins. These arrays were placed at 4 different positions so that transposition into TA target sites could be studied at all possible distances from 1 to 36 bp from the Zif268 binding site.

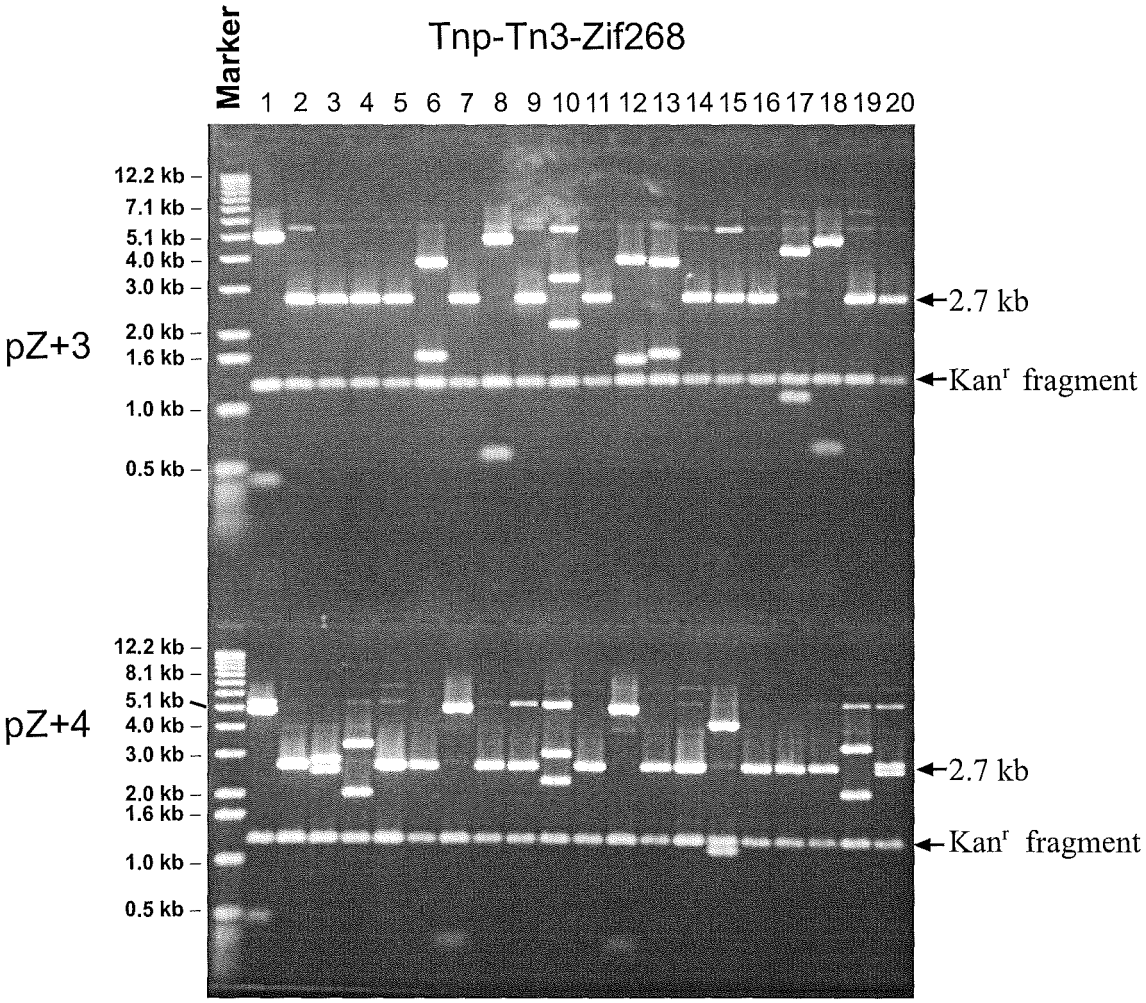


Figure 5.14 BamHI digests of transposition products catalysed by Tnp-Tn3-Zif268 into pZ+3 and pZ+4. DNA from 20 randomly selected individual Kan^r transformants of DS964 from experiment B with Tnp-Tn3-Zif268 and either pZ+3 or pZ+4 was cleaved with BamHI and run on a 1.2% agarose/TAE gel.

Transposition was studied *in vivo* rather than *in vitro* because the *in vivo* assay is easy to carry out and there is no need to purify the Z-transposase. The *in vivo* assay set up could also be adapted to screen for mutants of Z-transposase that have increased activity or target specificity.

It was decided to fuse the Zif268 DNA-binding domain (Zif268-DBD) to the C-terminus of ISY100 transposase partly due to the ease with which expression plasmids could be constructed. Another consideration was that addition of a His-tag to the N-terminus of IS630 transposase inactivated the transposase *in vivo*, whereas a C-terminal His-tag led to an increase in activity (my unpublished data). It was therefore thought that a C-terminal fusion was less likely to inactivate ISY100 transposase than an N-terminal fusion. In similar experiments carried out in other labs, the zinc finger domains of Zif268 and E2C have been fused to both the N-terminus and the C-terminus of HIV integrase without loss of activity (Bushman and Miller, 1997; Tan *et al.*, 2004). A recent *in vivo* study on the IS630/Tc1/*mariner* superfamily transposase *Sleeping Beauty* suggested that attaching another Cys_2His_2 zinc finger protein (Sp1) to the C-terminus of *Sleeping Beauty* transposase abolished transposase activity, while the addition of Sp1 or zinc fingers 3-8 of the human zinc finger protein ZNF202 to the N-terminus of SB did not (Wilson *et al.*, 2005). However, no alteration of target selectivity was observed in this study. It is therefore possible that N-terminal fusions of zinc finger DNA-binding domains to ISY100 transposase might also be active.

Preliminary evidence strongly suggested that fusing the Zif268 DNA-binding domain to ISY100 transposase via the C-terminal DNA-binding domain of Tn3 resolvase led to targeted transposition on just one side of the Zif268 recognition site. Whether the preference for TAs at certain distances from the Zif268 binding site is due to tethering of transposase in a fixed configuration with respect to the Zif268 binding site, or is caused by the different flanking sequences of the TAs in the TANN array, requires further investigation. The role of the Tn3 resolvase DNA-binding domain is also unclear. It may function simply as a non-specific linker protein to allow the Zif268 DNA-binding domain to reach the target site while the transposase is bound in the integration complex. Alternatively, it may play some more specific role, for instance by binding to the DNA between the Zif268 binding site and the target site. However, there is no evidence that it is influencing target choice by binding to *res* site-related sequences.

The Zif268 binding site is 9-10 bp long, with each "finger" recognising 3-4 bp. Together with the TA target specificity of ISY100 transposase, the chimeric transposase

can potentially recognise a specific 11 - 12 bp sequence. In theory this sequence will occur only once in every 4 Mbp of random sequence, equivalent to one site in the genome of *E. coli*. To increase this specificity, more zinc finger domains can be added onto the zinc finger DNA-binding domain (Beerli and Barbas, 2002). Another possibility is to construct a transposase that is targeted to a TA flanked by two Zif268 binding sites in indirect repeat. Because of the expected 2-fold symmetry of the integration complex, this should efficiently target transposition to the central TA. A preliminary experiment to set up such a system has been carried out by Marko Prorocic, with some limited success. The distance between the inverted repeat Zif268 recognition sites was chosen based on the distance between a single Zif268 binding site and the preferred target TA found in my studies. Such a transposase could potentially recognise a sequence of 22 bp in length, which should occur once in 10^{12} bp, and is long enough to specify a unique site in a mammalian genome.

Other systems have been set up that take advantage of the ability of two zinc finger DNA-binding domains to specify a unique site in a large genome. Zinc finger nucleases, based on the restriction enzyme FokI, require two zinc finger target sites in inverted repeat (Mani *et al.*, 2005). These nucleases have been engineered to recognise several different sites in the human genome, and have been used to initiate homologous gene replacement of medically important genes in human cells (Durai *et al.*, 2005). Tn3 resolvase has been fused to Zif268 to produce Z-resolvases. Recombination requires two Zif268 recognition sites flanking the *res* site I central region in inverted repeat (Akopian *et al.*, 2003).

More work is needed, first to confirm the target-specific transposition reported in this chapter, and then to increase the specificity and efficiency of transposition. One of the problems with the targeted transposition reported here is that there is still a relatively high level of transposition into random sites. It seems possible that mutant transposases could be isolated that cannot bind tightly enough to target DNA to allow transposition. These mutants might be rescued by the zinc finger DNA-binding domain, in which case transposition would only occur adjacent to the zinc finger binding site. The *in vivo* transposition assays developed in this work could perhaps be modified to allow selection of such mutants, and also of hyperactive Z-transposase mutants that bring about transposition with a higher frequency.

Chapter VI
Concluding remarks

The main objectives of this project were to increase our understanding of the transposition mechanism of ISY100, and to investigate the potential to develop ISY100 into a target-specific transposition tool. The results presented here demonstrate that ISY100 transposes via a cut-and-paste mechanism, and that ISY100 transposase is the only protein required for transposition. Transposase binds transposon ends via its N-terminal helix-turn-helix domains, and then cleaves the transposon exactly at its 3' ends, and two nucleotides inside the transposon at its 5' ends. Transposase integrates the excised transposon into target TA sites, which are duplicated after transposition.

Double-stranded breaks at the ISY100 ends occur in a defined order. The non-transferred strand is always cleaved first, and the transferred strand can only be cleaved once the non-transferred strand has been cleaved. This result is consistent with previous observations on the cleavage of *Mos1 mariner*, a eukaryotic member from the same superfamily (Dawson and Finnegan, 2003). For some transposons, such as Tn5 and Tn10, transposase cleaves the second strand via a hairpin intermediate, and this accounts for the defined order of strand cleavage for these elements. However, all of the evidence suggests that *Mos1* does not cleave via a hairpin intermediate. Instead, it seems that the non-transferred strands are cleaved independently at each end. Only when first strands have been cleaved can the two ends come together to form a paired end complex (PEC). Second strand cleavage occurs within the PEC, possibly by a domain swapping mechanism in which the transposase bound to one end cleaves the transferred strand of the other end. No hairpin structures were seen at ISY100 ends, consistent with the results for *Mos1*. Although there was no direct evidence for PEC formation for ISY100, second strand cleavage appeared to be coordinated at the two ends, suggesting formation of some sort of synaptic complex. More experiments are needed to identify the mechanism responsible for the defined order of cleavage by ISY100 transposase and the nature of any PEC. Further work is also needed to show whether the lack of observable PECs might be due to an unstable synaptic complex and whether this might account for the low level of coordinated insertion of transposon ends in the *in vitro* system.

Fusion of the Zif268 DNA-binding domain to the C-terminus of ISY100 transposase, with the Tn3 DNA-binding domain as a spacer, resulted in a chimeric Z-transposase that mediated transposition mostly into target sites adjacent to the Zif268 recognition site. This suggested that Z-transposase could be developed into a useful genetic tool. However, many more experiments need to be done to investigate and develop this system. For instance, transposition assays with different sequences adjacent to the Zif268 recognition sites could

be carried out to verify whether targeting of specific TAs adjacent to the Zif268 site was a consequence of the local DNA sequence surrounding the TA targets, or tethering of transposase to the Zif268 site such that transposition could only occur at fixed distances from the Zif268 recognition site. Similar fusion proteins containing shorter linkers between transposase and Zif268 failed to give site-specific transposition. No obvious Tn3 resolvase recognition sites were present close to the Zif268 recognition site. Experiments employing protein spacers with different lengths and sequence are required to investigate whether the Tn3 resolvase DNA-binding domain contributes in some structure-specific manner to targeted transposition, or is simply providing a protein linker of the correct length and geometry. To investigate whether this system could be used to target transposition to different target sites, the transposase could be fused to zinc fingers with different DNA-binding specificities. Experiments employing chimeric proteins with more zinc finger motifs, or with binding sites flanking the target TA in inverted repeat, are needed to see if transposition could be targeted to longer sites, expected to occur only once in a large eukaryotic genome.

Results from this study suggested that improvements in specificity might be obtained by decreasing the transposition frequency into random sites without reducing transposition into specific target sites. A protein that had lower target binding affinity, and was therefore dependent on the zinc finger DNA binding domain for target binding, would have the desired properties. A papillation assay has been established in *E. coli* in this lab since work on this thesis was completed (S. Colloms, personal communication), to facilitate screening for hyperactive transposase mutants. Further assays are being set up to screen for mutants with increased target specificity.

ISY100 transposase is the only protein required for transposition *in vitro*, suggesting ISY100 might be functional in wide-range hosts. Zinc finger proteins are common eukaryotic transcription factors, and should work well in eukaryotic hosts. However, wild-type ISY100 transposase and ISY100-based Z-transposases have yet to be tested in eukaryotic cells, and this is a priority for future development of this system. One possible problem is the lack of a nuclear localisation signal (NLS) in the prokaryotic ISY100 transposase. A region between the N-terminal DNA binding domain and the catalytic DDE domain, consisting of two basic residues followed by a 10 amino acid spacer and a cluster of 5 amino acids, of which three are basic, has been identified as a NLS in many transposases of the Tc1/*mariner* family (Ivics *et al.*, 1996). ISY100 transposase has a

similar basic region that might also function as a NLS in eukaryotes, but this has yet to be tested.

The *in vivo* reversibility of transposition can be problematic when it is desirable to obtain stable mutations or insertions. This problem can be minimized in several ways. The main strategy that has been used is to control expression of the transposase protein, so that it is only supplied transiently. For instance, the transposase gene can be carried on a vector that cannot replicate in the host cell, or the transposase can be supplied as mRNA which is quickly degraded. Another possibility is to introduce pre-assembled transpososomes, catalytically competent synaptic complexes of transposase and excised transposon, into the cell (Goryshin *et al.*, 2000). Recent experiments have demonstrated that *E. coli* can be electroporated with transpososomes assembled from ISY100 transposase and pre-excised transposon (my unpublished results). This yielded Kan^r colonies that presumably reflect transposase-catalysed integration of the transposon into *E. coli* chromosomal DNA. This may function, and be useful, in other cell types.

Results from this study showed that the cleavage activity of ISY100 transposase was affected by the sequence of the dinucleotides flanking the donor transposon. The maximum cleavage activity was obtained with the wild-type TA dinucleotide, and the activity was dramatically reduced in the presence of alternative flanking dinucleotides. The work presented in this thesis demonstrated that the dinucleotides flanking new insertions are produced by duplication of the target site. Furthermore, pre-excised transposon DNA is an efficient substrate for integration. Therefore, ISY100 transposase mutants with changes in preference for either the donor cleavage site or the target site (but not both) could possibly be useful for preventing further events after the initial transposon insertion. Finding transposase mutants with alternative target site specificity may also be necessary for extending the range of sequences available for targeting by Z-transposases.

As yet little is known about when and how members of the IS630/Tc1/*mariner* family capture target DNA during transposition. Transposons Tn7 and Mu have separate proteins responsible for target binding and strand-transfer, and target site selection is done before cleavage. For Tn10, where transposition is catalysed by a single transposon-encoded protein, target capture occurs only after transposon excision (Sakai and Kleckner, 1997). Since excision and strand transfer reactions are carried out by a single catalytic site, it seems likely that the target site is recognized by same part of the transposase complex that binds the donor flank in the excision reaction. In this study, transposase mediated transposition of pre-excised ISY100 *in vitro* into TT (AA) dinucleotide instead of TA at an

elevated frequency. This suggests that target site selection may depend on the conformation of the complex formed by transposase and the two transposon ends. Studies of target capture by ISY100 transposase may be facilitated by Z-transposase, which can mediate transposition to a specific target site in a DNA molecule. For instance, it might be possible to obtain a crystal structure (or failing that, other structural information) of Z-transposase bound to transposon ends and target DNA in a target capture complex. This sort of study could also aid in the development and refinement of Z-transposase as a genetic tool.

Although much work remains to be done, ISY100 transposase and Z-transposase hold much promise, and may form a useful tool in the armoury presently being developed in laboratories throughout the world to manipulate complex genomes in a controlled fashion.

References

Adelman, Z. N., N. Jasinskiene and A. A. James (2002). Development and applications of transgenesis in the yellow fever mosquito, *Aedes aegypti*. Mol. Biochem. Parasitol. **121**, 1-10.

Akerley, B. J., E. J. Rubin, A. Camilli, D. J. Lampe, H. M. Robertson and J. J. Mekalanos (1998). Systematic identification of essential genes by *in vitro mariner* mutagenesis. Proc. Natl. Acad. Sci. U. S. A. **95**, 8927-32.

Akerley, B. J., E. J. Rubin, V. L. Novick, K. Amaya, N. Judson and J. J. Mekalanos (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. Proc. Natl. Acad. Sci. U. S. A. **99**, 966-71.

Akopian, A., J. He, M. R. Boocock and W. M. Stark (2003). Chimeric recombinases with designed DNA sequence recognition. Proc. Natl. Acad. Sci. U. S. A. **100**, 8688-91.

Aleksenko, A. and A. J. Clutterbuck (1996). The plasmid replicator AMA1 in *Aspergillus nidulans* is an inverted duplication of a low-copy-number dispersed genomic repeat. Mol. Microbiol. **19**, 565-74.

Allingham, J. S., P. A. Pribil and D. B. Haniford (1999). All three residues of the Tn 10 transposase DDE catalytic triad function in divalent metal ion binding. J. Mol. Biol. **289**, 1195-206.

Amann, E., B. Ochs and K. J. Abel (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. Gene **69**, 301-15.

Arciszewska, L. K. and N. L. Craig (1991). Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. Nucleic Acids Res. **19**, 5021-9.

Arciszewska, L. K., R. L. McKown and N. L. Craig (1991). Purification of TnsB, a transposition protein that binds to the ends of Tn7. J. Biol. Chem. **266**, 21736-44.

Ason, B. and W. S. Reznikoff (2004). DNA sequence bias during Tn5 transposition. J. Mol. Biol. **335**, 1213-25.

Auge-Gouillou, C., M. H. Hamelin, M. V. Demattei, M. Periquet and Y. Bigot (2001). The wild-type conformation of the *Mos-1* inverted terminal repeats is suboptimal for transposition in bacteria. Mol. Genet. Genomics **265**, 51-7.

Bainton, R., P. Gamas and N. L. Craig (1991). Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell **65**, 805-16.

Bainton, R. J., K. M. Kubo, J. N. Feng and N. L. Craig (1993). Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified *in vitro* system. Cell **72**, 931-43.

Baker, T. A. and K. Mizuuchi (1992). DNA-promoted assembly of the active tetramer of the Mu transposase. Genes Dev. **6**, 2221-32.

Beerli, R. R. and C. F. Barbas, 3rd (2002). Engineering polydactyl zinc-finger transcription factors. Nat. Biotechnol. **20**, 135-41.

Bender, J. and N. Kleckner (1992). Tn10 insertion specificity is strongly dependent upon sequences immediately adjacent to the target-site consensus sequence. Proc. Natl. Acad. Sci. U. S. A. **89**, 7996-8000.

Bennett, P. M. (2004). Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangement. Methods Mol. Biol. **266**, 71-113.

Berger, B. and D. Haas (2001). Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS21 and related elements. Cell. Mol. Life Sci. **58**, 403-19.

Bhasin, A., I. Y. Goryshin and W. S. Reznikoff (1999). Hairpin formation in Tn5 transposition. J. Biol. Chem. **274**, 37021-9.

Bhasin, A., I. Y. Goryshin, M. Steiniger-White, D. York and W. S. Reznikoff (2000). Characterization of a Tn5 pre-cleavage synaptic complex. J. Mol. Biol. **302**, 49-63.

Birnboim, H. C. and J. Doly (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**, 1513-23.

Bolland, S. and N. Kleckner (1995). The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. Proc. Natl. Acad. Sci. U. S. A. **92**, 7814-8.

Bolland, S. and N. Kleckner (1996). The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. Cell **84**, 223-33.

Boyd, A. C. and D. J. Sherratt (1995). The pCLIP plasmids: versatile cloning vectors based on the bacteriophage lambda origin of replication. Gene **153**, 57-62.

Bujacz, G., M. Jaskolski, J. Alexandratos, A. Wlodawer, G. Merkel, R. A. Katz and A. M. Skalka (1996). The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. Structure **4**, 89-96.

Burge, C., A. M. Campbell and S. Karlin (1992). Over- and under-representation of short oligonucleotides in DNA sequences. Proc. Natl. Acad. Sci. U. S. A. **89**, 1358-62.

Bushman, F. D. (1994). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. U. S. A. **91**, 9233-7.

Bushman, F. D. and M. D. Miller (1997). Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. J. Virol. **71**, 458-64.

Capy, P., T. Langin, D. Higuet, P. Maurer and C. Bazin (1997). Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? Genetica **100**, 63-72.

Carlson, C. M., A. J. Dupuy, S. Fritz, K. J. Roberg-Perez, C. F. Fletcher and D. A. Largaespada (2003). Transposon mutagenesis of the mouse germline. Genetics **165**, 243-56.

Cassier-Chauvat, C., M. Poncelet and F. Chauvat (1997). Three insertion sequences from the cyanobacterium *Synechocystis* PCC6803 support the occurrence of horizontal DNA transfer among bacteria. Gene **195**, 257-66.

Chaconas, G. and Harshey, R. M. (2002). Chapter 17, Transposition of Phage Mu DNA, p. 384-402. In Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M. (eds.), Mobile DNA II. American Society for Microbiology, Washington, D. C.

Chalmers, R., A. Guhathakurta, H. Benjamin and N. Kleckner (1998). IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. Cell **93**, 897-908.

Chatterjee, P. K. and J. S. Coren (1997). Isolating large nested deletions in bacterial and P1 artificial chromosomes by *in vivo* P1 packaging of products of Cre-catalysed recombination between the endogenous and a transposed *loxP* site. Nucleic Acids Res.. **25**, 2205-12.

Chatterjee, P. K., S. Mukherjee, L. A. Shakes, W. Wilson, 3rd, J. S. Coren, K. R. Harewood and G. Byrd (2004). Selecting transpositions using phage P1 headful packaging: new markerless transposons for functionally mapping long-range regulatory sequences in bacterial artificial chromosomes and P1-derived artificial chromosomes. Anal. Biochem. **335**, 305-15.

Christy, B. A., L. F. Lau and D. Nathans (1988). A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences. Proc. Natl. Acad. Sci. U. S. A. **85**, 7857-61.

Clubb, R. T., J. G. Omichinski, H. Savilahti, K. Mizuuchi, A. M. Gronenborn and G. M. Clore (1994). A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase. Structure **2**, 1041-8.

Coates, C. J., N. Jasinskiene, L. Miyashiro and A. A. James (1998). *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. Proc. Natl. Acad. Sci. U. S. A. **95**, 3748-51.

Colloms, S. D., H. G. van Luenen and R. H. Plasterk (1994). DNA binding activities of the *Caenorhabditis elegans* Tc3 transposase. Nucleic Acids Res.. **22**, 5548-54.

Comfort, N. C. (2001). From controlling elements to transposons: Barbara McClintock and the Nobel Prize. Trends Genet. **17**, 475-8.

Craig, N. L. (1997). Target site selection in transposition. Annu. Rev. Biochem. **66**, 437-74.

Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M. (eds.) (2002). Mobile DNA II. American Society for Microbiology, Washington, D.C.

Craigie, R. (2001). HIV integrase, a brief overview from chemistry to therapeutics. J. Biol. Chem. **276**, 23213-6.

Craigie, R. (2002). Chapter 25, Retroviral DNA Integration, p. 613-630. In Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M. (eds.), Mobile DNA II. American Society for Microbiology, Washington, D. C.

Crooks, G. E., G. Hon, J. M. Chandonia and S. E. Brenner (2004). WebLogo: a sequence logo generator. Genome Res. **14**, 1188-90.

Cuff, J. A. and G. J. Barton (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins **40**, 502-11.

Cui, Z., A. M. Geurts, G. Liu, C. D. Kaufman and P. B. Hackett (2002). Structure-function analysis of the inverted terminal repeats of the *Sleeping Beauty* transposon. J. Mol. Biol. **318**, 1221-35.

Davies, D. R., I. Y. Goryshin, W. S. Reznikoff and I. Rayment (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. Science **289**, 77-85.

Davies, D. R., L. Mahnke Braam, W. S. Reznikoff and I. Rayment (1999). The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-A resolution. J. Biol. Chem. **274**, 11904-13.

Dawson, A. and D. J. Finnegan (2003). Excision of the *Drosophila mariner* transposon *Mos1*. Comparison with bacterial transposition and V(D)J recombination. Mol. Cell **11**, 225-35.

Deininger, P. L. and M. A. Batzer (2002). Mammalian retroelements. Genome Res. **12**, 1455-65.

Derbyshire, K. M. and N. D. Grindley (1996). *Cis* preference of the IS903 transposase is mediated by a combination of transposase instability and inefficient translation. Mol. Microbiol. **21**, 1261-72.

Doak, T. G., F. P. Doerder, C. L. Jahn and G. Herrick (1994). A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. U. S. A. **91**, 942-6.

Dodd, I. B. and J. B. Egan (1987). Systematic method for the detection of potential lambda Cro-like DNA-binding regions in proteins. J. Mol. Biol. **194**, 557-64.

Dodd, I. B. and J. B. Egan (1990). Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res.. **18**, 5019-26.

Dupuy, A. J., K. Akagi, D. A. Largaespada, N. G. Copeland and N. A. Jenkins (2005). Mammalian mutagenesis using a highly mobile somatic *Sleeping Beauty* transposon system. Nature **436**, 221-6.

Durai, S., M. Mani, K. Kandavelou, J. Wu, M. H. Porteus and S. Chandrasegaran (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res. **33**, 5978-90.

Ellis, J. (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. Hum. Gene Ther. **16**, 1241-6.

Elrod-Erickson, M., M. A. Rould, L. Nekludova and C. O. Pabo (1996). Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. Structure **4**, 1171-80.

Emmons, S. W., L. Yesner, K. S. Ruan and D. Katzenberg (1983). Evidence for a transposon in *Caenorhabditis elegans*. Cell **32**, 55-65.

Engelman, A., K. Mizuuchi and R. Craigie (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell **67**, 1211-21.

Fadool, J. M., D. L. Hartl and J. E. Dowling (1998). Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. Proc. Natl. Acad. Sci. U. S. A. **95**, 5182-6.

Feschotte, C. and S. R. Wessler (2002). *Mariner*-like transposases are widespread and diverse in flowering plants. Proc. Natl. Acad. Sci. U. S. A. **99**, 280-5.

Fiandt, M., W. Szybalski and M. H. Malamy (1972). Polar mutations in lac, gal and phage lambda consist of a few IS-DNA sequences inserted with either orientation. Mol. Gen. Genet. **119**, 223-31.

Finnegan, D. J. (1992). Transposable elements. Curr. Opin. Genet. Dev. **2**, 861-7.

Fischer, S. E., H. G. van Luenen and R. H. Plasterk (1999). *Cis* requirements for transposition of Tc1-like transposons in *C. elegans*. Mol. Gen. Genet. **262**, 268-74.

Galas, D. J. and A. Schmitz (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5**, 3157-70.

Gomez-Lus, R. (1998). Evolution of bacterial resistance to antibiotics during the last three decades. Int Microbiol **1**, 279-84.

Goryshin, I. Y., J. Jendrisak, L. M. Hoffman, R. Meis and W. S. Reznikoff (2000). Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat Biotechnol **18**, 97-100.

Goryshin, I. Y., J. A. Miller, Y. V. Kil, V. A. Lanzov and W. S. Reznikoff (1998). Tn5/IS50 target recognition. Proc. Natl. Acad. Sci. U. S. A. **95**, 10716-21.

Goryshin, I. Y. and W. S. Reznikoff (1998). Tn5 *in vitro* transposition. J. Biol. Chem. **273**, 7367-74.

Goulaouic, H. and S. A. Chow (1996). Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and *Escherichia coli* LexA protein. J. Virol. **70**, 37-46.

Goyard, S., L. R. Tosi, J. Gouzova, J. Majors and S. M. Beverley (2001). New *Mos1 mariner* transposons suitable for the recovery of gene fusions *in vivo* and *in vitro*. Gene **280**, 97-105.

Greisman, H. A. and C. O. Pabo (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. Science **275**, 657-61.

Grindley, N. D. F. (2002). Chapter 14, The Movement of Tn3-Like Elements: Transposition and Cointegrate Resolution, p. 272-302. In Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M. (eds.), Mobile DNA II. American Society for Microbiology, Washington, D.C.

Grindley, N. D. and W. S. Kelley (1976). Effects of different alleles of the *E. coli* K12 pol A gene on the replication of non-transferring plasmids. Mol. Gen. Genet. **143**, 311-8.

Groenen, M. A., E. Timmers and P. van de Putte (1985). DNA sequences at the ends of the genome of bacteriophage Mu essential for transposition. Proc. Natl. Acad. Sci. U. S. A. **82**, 2087-91.

Groth, A. C., M. Fish, R. Nusse and M. P. Calos (2004). Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. Genetics **166**, 1775-82.

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

Gueiros-Filho, F. J. and S. M. Beverley (1997). Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. Science **276**, 1716-9.

Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science **302**, 415-9.

Hallet, B. and D. J. Sherratt (1997). Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. FEMS Microbiol. Rev. **21**, 157-78.

Halling, S. M. and N. Kleckner (1982). A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. Cell **28**, 155-63.

Hamer, L., T. M. DeZwaan, M. V. Montenegro-Chamorro, S. A. Frank and J. E. Hamer (2001). Recent advances in large-scale transposon mutagenesis. Curr. Opin. Chem. Biol. **5**, 67-73.

Haren, L., B. Ton-Hoang and M. Chandler (1999). Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. **53**, 245-81.

Hartl, D. (2001). Discovery of the transposable element *mariner*. Genetics **157**, 471-6.

Hayes, F. (2003). Transposon-based strategies for microbial functional genomics and proteomics. Annu. Rev. Genet. **37**, 3-29.

Heggestad, A. D., L. Notterpek and B. S. Fletcher (2004). Transposon-based RNAi delivery system for generating knockdown cell lines. Biochem. Biophys. Res. Commun. **316**, 643-50.

Hickman, A. B., Y. Li, S. V. Mathew, E. W. May, N. L. Craig and F. Dyda (2000). Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. Mol. Cell **5**, 1025-34.

Holtman, C. K., Y. Chen, P. Sandoval, A. Gonzales, M. S. Nalty, T. L. Thomas, P. Youderian and S. S. Golden (2005). High-Throughput Functional Analysis of the Synechococcus elongatus PCC 7942 Genome. DNA Res. **12**, 103-15.

Ivics, Z., P. B. Hackett, R. H. Plasterk and Z. Izsvak (1997). Molecular reconstruction of *Sleeping Beauty*, a Tc1-like transposon from fish, and its transposition in human cells. Cell **91**, 501-10.

Ivics, Z., Z. Izsvak, A. Minter and P. B. Hackett (1996). Identification of functional domains and evolution of Tc1-like transposable elements. Proc Natl Acad Sci U S A **93**, 5008-13.

Jacobson, J. W., M. M. Medhora and D. L. Hartl (1986). Molecular structure of a somatically unstable transposable element in *Drosophila*. Proc. Natl. Acad. Sci. U. S. A. **83**, 8684-8.

Jaworski, D. D. and D. B. Clewell (1994). Evidence that coupling sequences play a frequency-determining role in conjugative transposition of Tn916 in *Enterococcus faecalis*. J. Bacteriol. **176**, 3328-35.

Johnson, R. C. and W. S. Reznikoff (1984). Copy number control of Tn5 transposition. Genetics **107**, 9-18.

Junop, M. S. and D. B. Haniford (1997). Factors responsible for target site selection in Tn10 transposition: a role for the DDE motif in target DNA capture. EMBO J. **16**, 2646-55.

Katz, R. A., G. Merkel and A. M. Skalka (1996). Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: *in vitro* activities and incorporation of a fusion protein into viral particles. Virology **217**, 178-90.

Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. Science **303**, 1626-32.

Kennedy, A. K., A. Guhathakurta, N. Kleckner and D. B. Haniford (1998). Tn10 transposition via a DNA hairpin intermediate. Cell **95**, 125-34.

Kennedy, A. K., D. B. Haniford and K. Mizuuchi (2000). Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. Cell **101**, 295-305.

Kim, K., S. Y. Namgoong, M. Jayaram and R. M. Harshey (1995). Step-arrest mutants of phage Mu transposase. Implications in DNA-protein assembly, Mu end cleavage, and strand transfer. J. Biol. Chem. **270**, 1472-9.

Korswagen, H. C., R. M. Durbin, M. T. Smits and R. H. Plasterk (1996). Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. Proc. Natl. Acad. Sci. U. S. A. **93**, 14680-5.

Kuduvalli, P. N., J. E. Rao and N. L. Craig (2001). Target DNA structure plays a critical role in Tn7 transposition. EMBO J. **20**, 924-32.

Kulkosky, J., K. S. Jones, R. A. Katz, J. P. Mack and A. M. Skalka (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12**, 2331-8.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**, 680-5.

Lampe, D. J., B. J. Akerley, E. J. Rubin, J. J. Mekalanos and H. M. Robertson (1999). Hyperactive transposase mutants of the *Himar1 mariner* transposon. Proc. Natl. Acad. Sci. U. S. A. **96**, 11428-33.

Lampe, D. J., M. E. Churchill and H. M. Robertson (1996). A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. EMBO J. **15**, 5470-9.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, *et al.* (2001). Initial sequencing and analysis of the human genome. Nature **409**, 860-921.

Lenich, A. G. and A. C. Glasgow (1994). Amino acid sequence homology between Piv, an essential protein in site-specific DNA inversion in *Moraxella lacunata*, and transposases of an unusual family of insertion elements. J. Bacteriol. **176**, 4160-4.

Lewinski, M. K. and F. D. Bushman (2005). Retroviral DNA integration--mechanism and consequences. Adv. Genet. **55**, 147-81.

Lidholm, D. A., A. R. Lohe and D. L. Hartl (1993). The transposable element *mariner* mediates germline transformation in *Drosophila melanogaster*. Genetics **134**, 859-68.

Lins, R. D., T. P. Straatsma and J. M. Briggs (2000). Similarities in the HIV-1 and ASV integrase active sites upon metal cofactor binding. Biopolymers **53**, 308-15.

Liu, G., A. M. Geurts, K. Yae, A. R. Srinivasan, S. C. Fahrenkrug, D. A. Largaespada, J. Takeda, K. Horie, W. K. Olson and P. B. Hackett (2005). Target-site preferences of *Sleeping Beauty* transposons. J. Mol. Biol. **346**, 161-73.

Mahillon, J. and M. Chandler (1998). Insertion sequences. Microbiol. Mol. Biol. Rev. **62**, 725-74.

Maignan, S., J. P. Guilloteau, Q. Zhou-Liu, C. Clement-Mella and V. Mikol (1998). Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. J. Mol. Biol. **282**, 359-68.

Mani, M., J. Smith, K. Kandavelou, J. M. Berg and S. Chandrasegaran (2005). Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. Biochem. Biophys. Res. Commun. **334**, 1191-7.

Manna, D., A. M. Breier and N. P. Higgins (2004). Microarray analysis of transposition targets in *Escherichia coli*: the impact of transcription. Proc. Natl. Acad. Sci. U. S. A. **101**, 9780-5.

Matsuoka, T., S. Awazu, N. Satoh and Y. Sasakura (2004). *Minos* transposon causes germline transgenesis of the ascidian *Ciona savignyi*. Dev. Growth Differ. **46**, 249-55.

Matsutani, S., H. Ohtsubo, Y. Maeda and E. Ohtsubo (1987). Isolation and characterization of IS elements repeated in the bacterial chromosome. J. Mol. Biol. **196**, 445-55.

Maxam, A. M. and W. Gilbert (1980). Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**, 499-560.

Maxwell, A., R. Craigie and K. Mizuuchi (1987). B protein of bacteriophage mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. Proc. Natl. Acad. Sci. U. S. A. **84**, 699-703.

McClintock, B. (1951). Chromosome organization and genic expression. Cold Spring Harb. Symp. Quant. Biol. **16**, 13-47.

Miskey, C., Z. Izsvak, K. Kawakami and Z. Ivics (2005). DNA transposons in vertebrate functional genomics. Cell. Mol. Life Sci. **62**, 629-41.

Miskey, C., Z. Izsvak, R. H. Plasterk and Z. Ivics (2003). The Frog Prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. Nucleic Acids Res. **31**, 6873-81.

Mizuuchi, K. (1984). Mechanism of transposition of bacteriophage Mu: polarity of the strand transfer reaction at the initiation of transposition. Cell **39**, 395-404.

Mizuuchi, K. and K. Adzuma (1991). Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. Cell **66**, 129-40.

Mizuuchi, K., T. J. Nobbs, S. E. Halford, K. Adzuma and J. Qin (1999). A new method for determining the stereochemistry of DNA cleavage reactions: application to the SfiI and HpaII restriction endonucleases and to the MuA transposase. Biochemistry (Mosc). **38**, 4640-8.

Mizuuchi, M., T. A. Baker and K. Mizuuchi (1991). DNase protection analysis of the stable synaptic complexes involved in Mu transposition. Proc. Natl. Acad. Sci. U. S. A. **88**, 9031-5.

Mori, I., G. M. Benian, D. G. Moerman and R. H. Waterston (1988). Transposable element Tc1 of *Caenorhabditis elegans* recognizes specific target sequences for integration. Proc. Natl. Acad. Sci. U. S. A. **85**, 861-4.

Morisato, D. and N. Kleckner (1987). Tn10 transposition and circle formation *in vitro*. Cell **51**, 101-11.

Morisato, D., J. C. Way, H. J. Kim and N. Kleckner (1983). Tn10 transposase acts preferentially on nearby transposon ends *in vivo*. Cell **32**, 799-807.

Nagy, Z., M. Szabo, M. Chandler and F. Olasz (2004). Analysis of the N-terminal DNA binding domain of the IS30 transposase. Mol. Microbiol. **54**, 478-88.

Naumann, T. A. and W. S. Reznikoff (2002). Tn5 transposase active site mutants. J. Biol. Chem. **277**, 17623-9.

Normand, C., G. Duval-Valentin, L. Haren and M. Chandler (2001). The terminal inverted repeats of IS911: requirements for synaptic complex assembly and activity. J. Mol. Biol. **308**, 853-71.

Oettinger, M. A. (1999). V(D)J recombination: on the cutting edge. Curr. Opin. Cell Biol. **11**, 325-9.

Olivares, E. C., R. P. Hollis, T. W. Chalberg, L. Meuse, M. A. Kay and M. P. Calos (2002). Site-specific genomic integration produces therapeutic Factor IX levels in mice. Nat. Biotechnol. **20**, 1124-8.

Oosumi, T., W. R. Belknap and B. Garlick (1995). *Mariner* transposons in humans. Nature **378**, 672.

Ostertag, E. M. and H. H. Kazazian, Jr. (2001). Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. **35**, 501-38.

Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, *et al.* (2003). Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat. Genet. **35**, 32-40.

Peng, W. J., C. M. Chang and T. H. Lin (2002). Target integration by a chimeric Sp1 zinc finger domain-Moloney murine leukemia virus integrase *in vivo*. J. Biomed. Sci. **9**, 171-84.

Peters, J. E. and N. L. Craig (2000). Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. Mol. Cell **6**, 573-82.

Peters, J. E. and N. L. Craig (2001). Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. Genes Dev. **15**, 737-47.

Pietrokovski, S. and S. Henikoff (1997). A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. Mol. Gen. Genet. **254**, 689-95.

Plasterk, R. H., Z. Izsvak and Z. Ivics (1999). Resident aliens: the Tc1/*mariner* superfamily of transposable elements. Trends Genet. **15**, 326-32.

Pledger, D. W. and C. J. Coates (2005). Mutant *Mos1 mariner* transposons are hyperactive in *Aedes aegypti*. Insect Biochem. Mol. Biol. **35**, 1199-207.

Polard, P. and M. Chandler (1995). An *in vivo* transposase-catalyzed single-stranded DNA circularization reaction. Genes Dev. **9**, 2846-58.

Polard, P., B. Ton-Hoang, L. Haren, M. Betermier, R. Walczak and M. Chandler (1996). IS911-mediated transpositional recombination *in vitro*. J. Mol. Biol. **264**, 68-81.

Preclin, V., E. Martin and L. Segalat (2003). Target sequences of Tc1, Tc3 and Tc5 transposons of *Caenorhabditis elegans*. Genet. Res. **82**, 85-8.

Pribil, P. A. and D. B. Haniford (2003). Target DNA bending is an important specificity determinant in target site selection in Tn10 transposition. J. Mol. Biol. **330**, 247-59.

Rebar, E. J. and C. O. Pabo (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science **263**, 671-3.

Reznikoff, W. S. (2002). Chapter 18, Tn5 Transposition, p. 403-422. In Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M. (eds.), Mobile DNA II. American Society for Microbiology, Washington, D.C.

Rice, P., I. Longden and A. Bleasby (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16**, 276-7.

Rice, P. A. and T. A. Baker (2001). Comparative architecture of transposase and integrase complexes. Nat. Struct. Biol. **8**, 302-7.

Richardson, J. M., A. Dawson, N. O'Hagan, P. Taylor, D. J. Finnegan and M. D. Walkinshaw (2006). Mechanism of *Mos1* transposition: insights from structural analysis. EMBO J. **25**, 1324-34.

Robertson, H. M. (1993). The *mariner* transposable element is widespread in insects. Nature **362**, 241-5.

Robertson, H. M. and D. J. Lampe (1995). Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. Mol. Biol. Evol. **12**, 850-62.

Rowland, S. J., D. J. Sherratt, W. M. Stark and M. R. Boocock (1995). Tn552 transposase purification and *in vitro* activities. EMBO J. **14**, 196-205.

Rubin, E. J., B. J. Akerley, V. N. Novik, D. J. Lampe, R. N. Husson and J. J. Mekalanos (1999). *In vivo* transposition of *mariner*-based elements in enteric bacteria and mycobacteria. Proc. Natl. Acad. Sci. U. S. A. **96**, 1645-50.

Rubin, G. M. and A. C. Spradling (1982). Genetic transformation of Drosophila with transposable element vectors. Science **218**, 348-53.

Sakai, J., R. M. Chalmers and N. Kleckner (1995). Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn10 transposition. EMBO J. **14**, 4374-83.

Sakai, J. and N. Kleckner (1997). The Tn10 synaptic complex can capture a target DNA only after transposon excision. Cell **89**, 205-14

Salzberg, S. L., A. J. Salzberg, A. R. Kerlavage and J. F. Tomb (1998). Skewed oligomers and origins of replication. Gene **217**, 57-67.

Sarnovsky, R. J., E. W. May and N. L. Craig (1996). The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. EMBO J. **15**, 6348-61.

Savilahti, H., P. A. Rice and K. Mizuuchi (1995). The phage Mu transpososome core: DNA requirements for assembly and function. EMBO J. **14**, 4893-903.

Schagger, H. and G. von Jagow (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. **166**, 368-79.

Schukkink, R. F. and R. H. Plasterk (1990). TcA, the putative transposase of the *C. elegans* Tc1 transposon, has an N-terminal DNA binding domain. Nucleic Acids Res. **18**, 895-900.

Selbitschka, W., W. Arnold, D. Jording, B. Kosier, N. Toro and A. Puhler (1995). The insertion sequence element ISRm2011-2 belongs to the IS630-Tc1 family of transposable elements and is abundant in *Rhizobium meliloti*. Gene **163**, 59-64.

Serre, M. C., C. Turlan, M. Bortolin and M. Chandler (1995). Mutagenesis of the IS1 transposase: importance of a His-Arg-Tyr triad for activity. J. Bacteriol. **177**, 5070-7.

Shao, H., Y. Qi and Z. Tu (2001). MsqTc3, a Tc3-like transposon in the yellow fever mosquito *Aedes aegypti*. Insect Mol. Biol. **10**, 421-5.

Shao, H. and Z. Tu (2001). Expanding the diversity of the IS630-Tc1-*mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics **159**, 1103-15.

Sherman, A., A. Dawson, C. Mather, H. Gilhooley, Y. Li, R. Mitchell, D. Finnegan and H. Sang (1998). Transposition of the *Drosophila* element *mariner* into the chicken germ line. Nat. Biotechnol. **16**, 1050-3.

Smith, J., J. M. Berg and S. Chandrasegaran (1999). A detailed study of the substrate specificity of a chimeric restriction enzyme. Nucleic Acids Res. **27**, 674-81.

Smith, J., M. Bibikova, F. G. Whitby, A. R. Reddy, S. Chandrasegaran and D. Carroll (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res. **28**, 3361-9.

Spradling, A. C. and G. M. Rubin (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. Science **218**, 341-7.

Steiniger-White, M., I. Rayment and W. S. Reznikoff (2004). Structure/function insights into Tn5 transposition. Curr. Opin. Struct. Biol. **14**, 50-7.

Steitz, T. A. and J. A. Steitz (1993). A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. U. S. A. **90**, 6498-502.

Studier, F. W., A. H. Rosenberg, J. J. Dunn and J. W. Dubendorff (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185**, 60-89.

Tan, W., Z. Dong, T. A. Wilkinson, C. F. Barbas, 3rd and S. A. Chow (2006). Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. J. Virol. **80**, 1939-48.

Tan, W., K. Zhu, D. J. Segal, C. F. Barbas, 3rd and S. A. Chow (2004). Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. J. Virol. **78**, 1301-13.

Tavakoli, N. P., J. DeVost and K. M. Derbyshire (1997). Defining functional regions of the IS903 transposase. J. Mol. Biol. **274**, 491-504.

Tenzen, T., S. Matsutani and E. Ohtsubo (1990). Site-specific transposition of insertion sequence IS630. J. Bacteriol. **172**, 3830-6.

Tenzen, T. and E. Ohtsubo (1991). Preferential transposition of an IS630-associated composite transposon to TA in the 5'-CTAG-3' sequence. J. Bacteriol. **173**, 6207-12.

Ton-Hoang, B., P. Polard, L. Haren, C. Turlan and M. Chandler (1999). IS911 transposon circles give rise to linear forms that can undergo integration *in vitro*. Mol. Microbiol. **32**, 617-27.

Ton-Hoang, B., C. Turlan and M. Chandler (2004). Functional domains of the IS1 transposase: analysis *in vivo* and *in vitro*. Mol. Microbiol. **53**, 1529-43.

Tosi, L. R. and S. M. Beverley (2000). *cis* and *trans* factors affecting *Mos1 mariner* evolution and transposition *in vitro*, and its potential for functional genomics. Nucleic Acids Res. **28**, 784-90.

Tu, Z. and H. Shao (2002). Intra- and inter-specific diversity of Tc3-like transposons in nematodes and insects and implications for their evolution and transposition. Gene **282**, 133-42.

Turcotte, K. and T. Bureau (2002). Phylogenetic analysis reveals *stowaway*-like elements may represent a fourth family of the IS630-Tc1-*mariner* superfamily. Genome **45**, 82-90.

Turlan, C. and M. Chandler (2000). Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. Trends Microbiol. **8**, 268-74.

Urasaki, A., Y. Sekine and E. Ohtsubo (2002). Transposition of cyanobacterium insertion element ISY100 in *Escherichia coli*. J. Bacteriol. **184**, 5104-12.

Urnov, F. D., J. C. Miller, Y. L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory and M. C. Holmes (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature **435**, 646-51.

van der Meer, J. R., A. J. Zehnder and W. M. de Vos (1991). Identification of a novel composite transposable element, *Tn5280*, carrying chlorobenzene dioxygenase genes of *Pseudomonas* sp. strain P51. J. Bacteriol. **173**, 7077-83.

van Luenen, H. G., S. D. Colloms and R. H. Plasterk (1993). Mobilization of quiet, endogenous Tc3 transposons of *Caenorhabditis elegans* by forced expression of Tc3 transposase. EMBO J. **12**, 2513-20.

van Luenen, H. G., S. D. Colloms and R. H. Plasterk (1994). The mechanism of transposition of Tc3 in *C. elegans*. Cell **79**, 293-301.

van Luenen, H. G. and R. H. Plasterk (1994). Target site choice of the related transposable elements Tc1 and Tc3 of *Caenorhabditis elegans*. Nucleic Acids Res. **22**, 262-9.

van Pouderoyen, G., R. F. Ketting, A. Perrakis, R. H. Plasterk and T. K. Sixma (1997). Crystal structure of the specific DNA-binding domain of Tc3 transposase of C.elegans in complex with transposon DNA. EMBO J. **16**, 6044-54.

Vieira, J. and J. Messing (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**, 259-68.

Vigdal, T. J., C. D. Kaufman, Z. Izsvak, D. F. Voytas and Z. Ivics (2002). Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/*mariner* transposable elements. J. Mol. Biol. **323**, 441-52.

Vos, J. C., I. De Baere and R. H. Plasterk (1996). Transposase is the only nematode protein required for *in vitro* transposition of Tc1. Genes Dev. **10**, 755-61.

Waddell, C. S. and N. L. Craig (1989). Tn7 transposition: recognition of the attTn7 target sequence. Proc. Natl. Acad. Sci. U. S. A. **86**, 3958-62.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature **420**, 520-62.

Watkins, S., G. van Pouderoyen and T. K. Sixma (2004). Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. Nucleic Acids Res. **32**, 4306-12.

Weinreich, M. D., L. Mahnke-Braam and W. S. Reznikoff (1994). A functional analysis of the Tn5 transposase. Identification of domains required for DNA binding and multimerization. J. Mol. Biol. **241**, 166-77.

Wicker, T., J. S. Robertson, S. R. Schulze, F. A. Feltus, V. Magrini, J. A. Morrison, E. R. Mardis, R. K. Wilson, D. G. Peterson, A. H. Paterson, *et al.* (2005). The repetitive landscape of the chicken genome. Genome Res. **15**, 126-36.

Wilhelm, M. and F. X. Wilhelm (2001). Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. **58**, 1246-62.

Wilson, M. H., J. M. Kaminski and A. L. George, Jr. (2005). Functional zinc finger/*Sleeping Beauty* transposase chimeras exhibit attenuated overproduction inhibition. FEBS Lett. **579**, 6205-9.

Wolfe, S. A., L. Nekludova and C. O. Pabo (2000). DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. **29**, 183-212.

Wolfe, S. A., E. I. Ramm and C. O. Pabo (2000). Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. Structure **8**, 739-50.

Wong, S. M. and J. J. Mekalanos (2000). Genetic footprinting with *mariner*-based transposition in *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. U. S. A. **97**, 10191-6.

Wright, D. A., J. A. Townsend, R. J. Winfrey, Jr., P. A. Irwin, J. Rajagopal, P. M. Lonosky, B. D. Hall, M. D. Jondle and D. F. Voytas (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J. **44**, 693-705.

Xu, P. X., X. Zhang, S. Heaney, A. Yoon, A. M. Michelson and R. L. Maas (1999). Regulation of Pax6 expression is conserved between mice and flies. Development **126**, 383-95.

Xu, W., M. A. Rould, S. Jun, C. Desplan and C. O. Pabo (1995). Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. Cell **80**, 639-50.

Yang, W., J. Y. Lee and M. Nowotny (2006). Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol. Cell **22**, 5-13.

Yanisch-Perron, C., J. Vieira and J. Messing (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**, 103-19.

Zhang, L., A. Dawson and D. J. Finnegan (2001). DNA-binding activity and subunit interaction of the *mariner* transposase. Nucleic Acids Res. **29**, 3566-75.

Zhou, L., R. Mitra, P. W. Atkinson, A. B. Hickman, F. Dyda and N. L. Craig (2004). Transposition of *hAT* elements links transposable elements and V(D)J recombination. Nature **432**, 995-1001.

Zhou, M., A. Bhasin and W. S. Reznikoff (1998). Molecular genetic analysis of transposase-end DNA sequence recognition: cooperativity of three adjacent base-pairs in specific interaction with a mutant Tn5 transposase. J. Mol. Biol. **276**, 913-25.

Zhou, M. and W. S. Reznikoff (1997). Tn5 transposase mutants that alter DNA binding specificity. J. Mol. Biol. **271**, 362-73.