Tn7 TRANSPOSITION

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

by

Stephen D. Bell

Institute of Genetics
University of Glasgow
Church St.
Glasgow

December 1992

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

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

Abbreviations iv
Acknowledgements v
Summary vi

Chapter 1 Introduction 1
1.1 General Introduction 2
1.2 Classes of Transposable Elements 3
1.3 Regulation of Transposition 5
1.4 Transposition Mechanisms 6
1.5 Replicative Transposition 7
1.6 Conservative Transposition 7
1.7 Consequences of Transposition for the Host Cell 9
1.8 *In vitro* Transposition Systems 9
1.8.1 Phage Mu 9
1.8.2 Tn10 10
1.8.3 Retro-elements 11
1.9 Transposition Insertion Sites 11
1.10 Discrimination between Intra- and Inter molecular transposition 12
1.15 Tn7 Transposition *in vitro* 17

Chapter 2 Materials and Methods 21
2.1 Bacterial strains. 22
2.2 Plasmids and Phagemids 22
2.3 Bacteriophage 22
2.4 Oligonucleotides 22
2.5 Chemicals. 22
2.6 Proteins. 22
2.7 Culture media 23
2.8 Sterilisation. 23
2.9 Buffer solutions 24
2.10 Antibiotics. 25
2.11 Growth conditions 25
2.12 Indicators. 26
2.13 Plasmid DNA isolation. 26
2.14 Chromosomal DNA isolation. 28
2.15 Transformation with Plasmid DNA. 28
2.16 Preparation of P1 Lysates. 29

2.17 P1 transduction....29
2.18 Single Colony Gel Analysis....29
2.19 Bacterial Conjugation and Mate-out Assay....30
2.20 Restriction of DNA....30
2.21 Calf Intestinal Phosphatase (CIP) Treatment....30
2.20 Ethanol Precipitation of DNA....31
2.23 Ligation of DNA Fragments....31
2.24 Gel electrophoresis of DNA....31
2.25 Purification of DNA Fragments from Agarose Gels....33
2.26 Synthesis and Purification of Oligonucleotides....34
2.27 Photography of Gels....34
2.28 Electrophoresis of Proteins in SDS-polyacrylamide Gels....34
2.29 ^{35}S Labelling of Proteins....35
2.30 Southern Blotting and Hybridisation....36
2.31 Random Priming of DNA....36
2.32 DNA Sequencing....36
2.33 End-labelling of DNA fragments....37
2.33 Preparation of Crude Cell Extracts....37
2.34 Protein Quantiation....38
2.35 Gel Binding Assays....38
2.35 Site Directed Mutagenesis....39
2.36 Attempts at Tn7 *in vitro* Transposition Reactions....39

Chapter 3 The DNA Sequence of Tn7....40
3.1 Introduction....41
3.2 Sequencing Strategy....41
3.2.1 EcoRI-HindIII Fragment....41
3.2.2 Left End....42
3.3 Analyses of Sequence....44
3.4 Restriction Map of Tn7....44
3.5 Open Reading Frames....44
3.6 TnsB:: The Tn7 Transposase ?....44
3.7 Tn7 Internal Repeats....45
3.9 TnsB Binding to Internal Repeats....46
3.10 Integrons as Transposable Elements....48
3.11 A Summary of Evidence on Tn21 Type Integrons....48
3.12 The Tn7 Integron-like Structure....50
3.13 Codon Usage in Tn7 Encoded Genes....51

Chapter 4 Tn7 Transposition *in vivo* and *in vitro* 54
4.1 Introduction 55
4.2 pSB58 Construction 56
4.3 pSB84 Construction 57
4.4 Protein Expression 57
4.5 Transposition in vivo 58
4.7 Double Strand Breaks? 61
4.8 Attempts to Establish a Tn7 *in vitro* Transposition Assay 65
4.9 Discussion 67

Chapter 5 DNA Binding Activities in pSB84 Extracts 69
5.1 Introduction 70
5.2 TnsB Activity in pSB84 extracts 70
5.3 Assays for TnsD 70
5.4 Assays of TnsD Activity in pSB84 Extracts 71
5.5 Effect of Mg++ on TnsD Binding 72
5.6 TnsD Activity in pSB84 Extracts 73
5.7 Requirement for ATP Hydrolysis to Allow Detection of TnsD Activity 74
5.8 TnsD is Interacting with a Component of TnsABC 74
5.9 Discussion 75

Chapter 6 Host Factor Binding at *attTn7* 78
6.1 Introduction 79
6.2 A Role for IHF in Tn7 Transposition ? 81
6.3 Interactions between IHF and TnsD 84
6.4 Localisation of IHF Binding Site 85
6.5 DNA Bending at *attTn7* 87
6.6 Binding to *attTn7* +9 to +55 88
6.7 Methylation at *dcm* Site 90
6.8 Consequences of IHF Binding *in vivo* 90
6.8.1 Effect of IHF on Tn7 Transposition *in vivo* 90
6.8.2 IHF and PhoS Transcription 92
6.9 Discussion 94

Chapter 7 Concluding Remarks 97

Bibliography.....101

Abbreviations

Chemicals

APS	ammonium persulphate
ATP	adenosine triphosphate
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
DTT	ditiothreitol
EDTA	ethylenediaminetetra-acetic acid (disodium salt)
MgAc	magnesium acetate
tris	tris (hydroxymethyl) amino ethane
IPTG	isopropylthio-B-D-galactoside
SDS	sodium dodecylsulphate
X-gal	5-bromo-4-chloro-3-indoyl-b-galactoside
XP	5-bromo-4-chloro-3-indoylphosphate

Measurements

bp	base pair
nt	nucleotide
kb	kilobase pair
Ci	Curie
V	Volts
A	amps
^{o}C	degrees Centigrade
µCi	micro curies
cm	centimetre
nm	nanometre
M	molar
mM	millimolar
µM	micromolar
nM	nanomolar
min	minutes
sec	seconds
hr	hours

Acknowledgements

Firstly, I would like to thank Dave Sherratt for his help and supervision throughout this project. Eternal gratitude also goes to Mary, for running the lab so well and still having time to answer my endless questions about chemicals and strains. My thanks go to members of the Tn7 group past and present, in particular Liz for help and advice and Angela, for her plasmid stocks and endless enthusiasm. I would also like to thank the rest of the sixth floor group, Mark, Marshall, Martin, Garry, Lidia, Nick, Gerhard and Dave for creating such an enjoyable and educational environment. Special thanks go to Richard for reciprocating tequila nights and generally being alive. Extra-special thanks go to Sally and Jen for sharing a lab with me and not complaining too much about the slavering bench beasts that occasionally escaped from my zoo.

Thanks also go to the prep. room ladies for providing safe and pleasant homes for my bugs.

Thanks also go to my parents for support and entusiasm for the last few years. But finally, and most of all, my thanks and love go to Jo, for putting up with my endless "I was wondering ...?" comments, supporting me with tea and ice-cream whenever it was needed and for always being there.

Summary

This project addressed a variety of aspects of the bacterial transposon Tn7. Firstly, the known sequence of the element was completed and the entire sequence compiled and presented. This was analysed with particular emphasis on the evolutionary implications of a series of directly repeated DNA motifs in the interior of Tn7 were discussed. These repeats were tested by gel retardation analysis for potential function as binding sites for Tn7 encoded proteins. These data are presented in Chapter 3.

Chapter 4 deals with the construction of plasmids designed to co-ordinately over-express the TnsA, B, C and D proteins. These plasmids were tested for their ability to support transposition *in vivo*. One plasmid gave rise to a peculiar phenomenon in these assays, , a model is proposed for its action. A number of attempts to establish a Tn7 *in vitro* reaction are also described.

Chapter 5 addresses the level of TnsD activity present in various cell extracts. An interaction between TnsD and one or more of TnsABC is detected and investigated and found be regulated by the hydrolysis of ATP. In the absence of ATP and magnesium ions a TnsABCD extract has no TnsD activity detectable by gel retardation assays, however in the presence of ATP and magnesium the same extract displays a detectable activity. Extracts containing TnsC and TnsD or TnsD on its own are not affected in this way.

Chapter 6 presents data, initially identifying an IHF binding site in the *attTn7* locus and localising the sequence determinants required for binding. IHF is demonstrated to induce a bend at the locus is demonstrated. The role of IHF *in vivo* in Tn7 transposition is investigated and it is demonstrated that transposition is reduced 6 fold in the absence of IHF. The potential role of IHF in the expression of the *phoS* gene, adjacent to *attTn7* is also addressed.

Chapter 1

Introduction

1.1 General Introduction

The study of genetic recombination has been a vital corner stone in the development of molecular biology. Recombination has been used both as a powerful tool for dissection of genetic function and equally importantly also as a topic of research in its own right. Studies of recombination processes have shed much light on processes ranging from genome evolution to the molecular intricacies of protein-DNA interactions and DNA structure. Recombination can be essentially viewed as the process whereby DNA molecules can be broken and rejoined. However, within this broad definition, it is possible to discriminate between two distinct classes of recombination reactions.

1)Homologous Recombination

This process is dependent on the presence of regions of homology in the exchanging molecules. One of the principle roles of homologous recombination is probably in the repair of DNA damage. However, homologous recombination is also important in the generation of genetic diversity in the processes of meiotic recombination in eukaryotes and Hfr generation and mating in prokaryotes (Smith, G. R., 1991).

2)Non-homologous Recombination

This class of events contains a wide and varied spectrum of rearrangement processes. The best studied non-homologous recombination reactions are those of conservative site specific recombination (CSSR) and transposition.

The simplest CSSR system comprises two DNA sites and a recombinase enzyme. The recombinase functions by bringing the two DNA sites together (synapsis) and catalyses a strand exchange reaction between the two sites. No external energy source is required, the energy of cleaved phosphodiester bonds is conserved by formation of a covalently linked recombinase-DNA intermediate (Stark et al., 1992).

Transposition is a process whereby a defined segment of DNA, the transposable element, can translocate to a novel site in the host cell's

genome. Transposable elements are a very diverse family of molecules varying in size, structure, number of encoded genes and transposition mechanism. A simplistic view of transposition involves the recognition of the ends of the transposable element by an element encoded protein, the transposase. The transposase brings about synapsis of the ends of the transposon. Thus far the process seems very similar to that of CSSR, however in transposition a third DNA site, the target site, must be recruited into the synapse prior to the strand exchange reaction. As will be discussed below the precise order of DNA cleavage and strand exchange reactions varies depending on the transposable element being studied. However, one emerging rule of transposition is that the 3'-OH groups at the ends of the element are exposed and used in a nucleophilic attack on the phosphodiester bonds in the target DNA. As with CSSR no external energy is required in the strand exchange process. In contrast with CSSR, however, in the systems where this question has been addressed the reaction appears to proceed via a one step transesterification rather than via a covalent protein-DNA intermediate (Engelman, 1991 Mizuuchi, 1991).

1.2 Classes of Transposable Elements

As stated above transposable elements are highly diverged in structure. It is possible, however, to divide bacterial elements into several broadly defined classes.

1) Insertion Sequences

Insertion sequences (IS) are structurally the most simple transposable elements. They encode only the required *cis* acting sequences and *trans* acting functions to catalyse the transposition reaction. These elements are typically 700bp to 1.6kbp in length (Galas and Chandler 1989).

2) Compound Transposons

Compound transposons are created by two IS elements mobilising the region between them. Thus, in a compound transposon, the IS encoded transposase acts at the outer ends of the flanking IS elements.

3) Complex Transposons

This is a highly diverse family of elements. They are characterised by their large size and complex architecture. The *cis* acting terminal sequences are frequently large and contain complex arrays of protein binding sites. There may also be several transposon encoded genes whose functions are required for transposition.

4) a) Conjugative transposons

These transposons differ from the elements described above in that they do not cause target site duplications on insertion. When present in a Gram positive host these elements confer the ability to act as a conjugational donor to the host. Thus the element excises from the genome of one bacterium and inserts into the genome of another. The element appears to excise by making nicks flush with either the 5' or 3' end and nicks on the opposite strand 5 bases into the flanking host sequences. These staggered ends are ligated creating a covalently closed supercoiled transposon circle, in which the ends of the element are separated by a 5 bp heteroduplex region (Caparon and Scott 1989). This region corresponds to the host extensions made in the initial excision and is termed the coupling sequence. The excision also generates a covalently closed donor backbone molecule lacking the transposon. Incision into the recipient is essentially the reverse of excision. Nicks are made at either end of the coupling sequence and a 5 base staggered cut made in the target sequence. The exposed ends are then ligated. This generates an insertion product flanked by 5 bp heteroduplexes. These heteroduplexes are apparently resolved by DNA replication. Thus the final product is the inserted transposon flanked by the initial 5 bp target sequence on one side and one of the 5 bp coupling sequences derived from the initial host (Scott 1992).

4) b) Integrons

A variety of mobile genetic elements have been shown to possess various antibiotic resistance genes in a common sequence element. This element comprises a gene with homology to lambda integrase (Mercier, Lachapelle et al. 1990), one or more antibiotic resistance genes, and a conserved

region downstream of the resistance genes (Nucken, Henschke et al. 1991, Stokes and Hall 1989). This work was recently reviewed (Amabile-Cuevas, 1992). Although the commonest genes found in integrons are for antibiotic resistance determinants other genes are also found. The resistance genes are positioned in the opposite orientation to the integrase gene and appear to be transcribed by a promoter located between the integrase and first resistance gene. Each resistance gene is associated with a conserved 59 bp element found at the 3' end of the gene (Hall, Brookes et al. 1991). The 59 bp element possesses dyad symmetry. The discovery of a family of integrons, differing only in resistance genes, led to the proposal that these elements were able to recruit resistance genes via a site specific recombination process. It has subsequently been shown that the integron encoded integrase can direct a site specific recombination reaction between 59 bp repeats. This has been demonstrated by studies of integrase mediated formation of co-integrates between two replicons containing integrons, (Martinez and De la Cruz 1988, Martinez and De la Cruz 1990), and also by the observation that over-expression of the integrase permits detection of excised, covalently closed cicular resistance gene cassettes (Collis and Hall 1992). Although the recombination site has been mapped a number of fundamental questions as to the nature of the reaction itself remain. The recombination event occurs within the sequence GTT at the 3' end of the 59 bp element. Because of the dyad symmetry of the 59bp there is a related GTT site at the 5' end of the element. However, the discoveries that inserted genes in the integrons always lie in the same orientation and that the same resistance genes in different integrons have identical 59bp repeats has demonstrated that recombination occurs exclusively at the 3' GTT. How this discrimination is mediated is unknown.
A further level of complexity to integrons is that the integron may be a transposable element itself (Bissonnette and Roy 1992). This matter as well as the integration of gene cassettes is dealt with in more detail in Chapter 3.

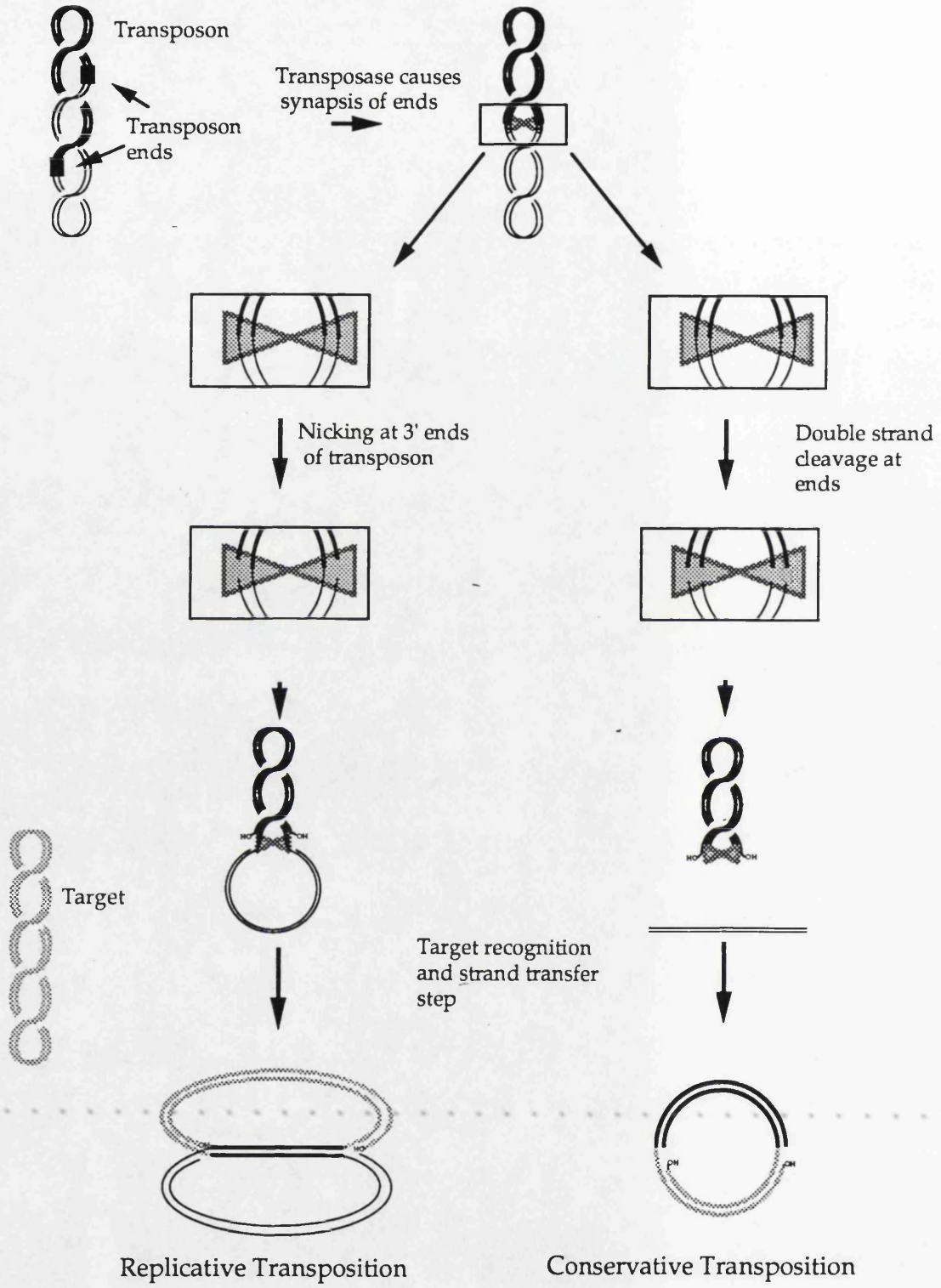
1.3 Regulation Of Transposition

Because of the potentially mutagenic effects of transposition it is important that transposable elements possess mechanisms to restrict the level of transposition. In principle transposition can be controlled at several levels.

1) By restricting the amount of transposase present in the cell.
In the case of Tn10 this is achieved by regulation of both transcription of the gene and translation of the mRNA. Both phage Mu and Tn3 also demonstrate control of expression of the transposase by repression of its transcription by other element encoded proteins (Pato 1989 ; Sherratt 1989).

2) By producing a protein which inhibits transposase activity or competes with transposase for binding to its cognate DNA site.
Perhaps the most sophisticated example of this form of control is that shown by certain IS elements. Analysis of the sequence of IS1 shows the presence of three potential open reading frames, InsA, InsB and InsB'. InsB and B' are in phase and overlap, InsB' is extended by 42 codons at the 5' end relative to InsB. The InsA protein has been detected and shown to bind the ends of IS1 (Zerbib, Jakowec et al. 1987). This protein appears to auto-regulate its expression, presumably by occlusion of its promoter in the left end of IS1. In addition to this role, InsA also exerts a negative control on transposition. InsB has never been directly detected. However both genetic and direct physical evidence have demonstrated that a fusion protein corresponding to InsAB' is produced by a -1 frameshifting event (Sekine and Ohtsubo 1989 ; Escoubas, Prere et al. 1991).
This frameshift is directed by a sequence in the IS1 mRNA which is similar to the -1 frameshifting signal used by some retroviruses in expression of *gag-pro*. Constitutive expression of an artificially engineered InsAB' fusion led to greatly elevated levels of transposition of both defective and wild type IS1 *in vivo*. Thus it appears that the InsAB' protein is the IS1 transposase.
These obsevations have led to the proposal that IS1 transposition is controlled by the relative ratios of the abundance of the InsA repressor and the InsAB' transposase. Similar frameshifting is observed with several other IS elements.

3) By coordinating transposition events with aspects of the cell cycle.
As discussed below, Tn10 perferentially undergoes transposition following DNA replication through the element.


Transposon
Transposon ends
Transposase causes synapsis of ends
Nicking at 3' ends of transposon
Double strand cleavage at ends
Target
Target recognition and strand transfer step
Replicative Transposition
Conservative Transposition

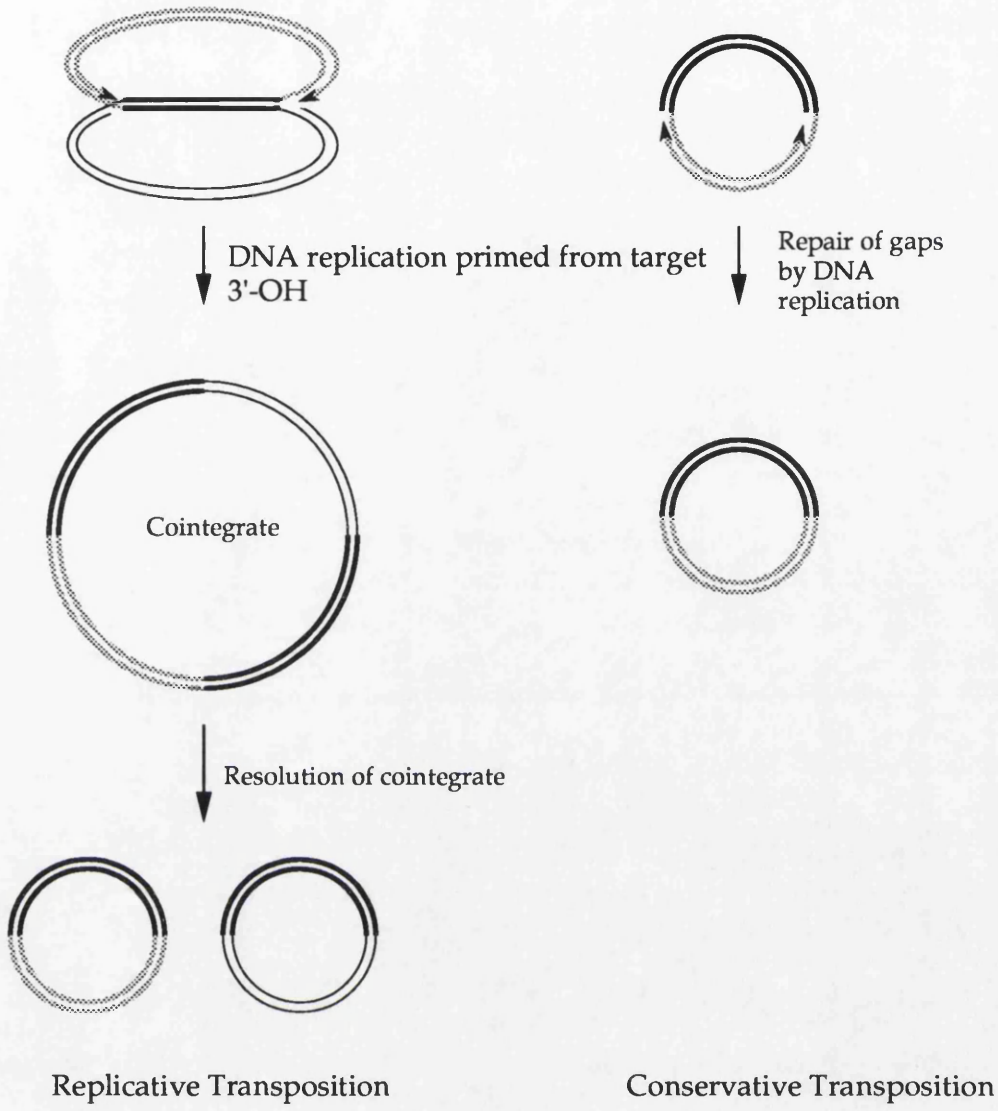


Fig. 1.1
Diagramatic Representation of Replicative and Conservative Transposition Mechanisms. As discussed in the text the precise timings and requirements for the reaction steps vary between elements.

1.4 Transposition Mechanisms

Two broad classifications of transposition mechanism were inferred from *in vivo* genetic studies of transposition. Elements can transpose by either a conservative or a replicative mechanism (see Fig. 1). Models for these transposition processes have been proposed by Shapiro, 1979, and Arthur and Sherratt, 1979. In replicative transposition DNA replication is an obligatory step in the transposition reaction. In contrast, in conservative transposition DNA replication is not an intrinsic requirement of the reaction. It may, however, be required for repair of gaps and nicks left by the conservative reaction, see Fig. 1.1.

1.5 Replicative Transposition

The best characterised elements which transpose replicatively are the members of the Tn3 family and Phage Mu (Pato 1989 ; Sherratt 1989). Tn3 encodes two functions which are required for the complete transposition reaction, transposase (*tnpA* gene) and resolvase (*tnpR* gene). Mu encodes two functions MuA and MuB which are involved in the transposition reaction. MuA is the transposase, the function of MuB is more complex but is involved in target molecule selection and also as an activator of MuA. Mu will be discussed in more detail below.

In replicative transposition the transposase catalyses the nicking of the 3' ends of the transposable element and cutting at the target site. The 3'-OH groups of the element are then joined to the exposed 5' phosphate of the cleaved target sequences. The exposed 3'-OH's of the target then prime DNA replication into the transposable element sequences. The net result of these processes is to link the donor and recipient molecule by two directly repeated copies of the transposable element. This structure is called a co-integrate. In the Tn3 family the resolvase then catalyses a CSSR reaction between sites on the two copies of the element. This resolution reaction restores the two starting molecules with the important difference that the recipient molecule now also contains a copy of the transposable element.

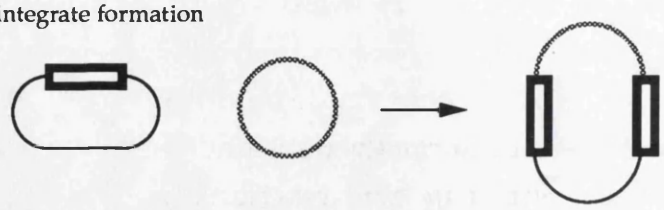
1.6 Conservative Transposition

In contrast with replicative transposition, conservative transposition results from double strand cleavage at the ends of the transposable element. The excised linear transposon is then inserted into the target site. In the systems thus far studied the exposed 3'-OH ends of the element are joined to the 5' phosphates of the cleaved recipient. Thus the net result of this is the generation of a linear donor backbone and a novel insertion product in the recipient. The insertion product will have short gaps at the 5' ends of the newly inserted transposon. These will be repaired by the DNA replication machinery of the host. The broken donor molecule will either be degraded by cellular exonucleases or repaired by the host double strand break repair mechanism. If the template used for repair is a sister chromosome which lacks a copy of the element at the analogous locus then there will be an apparent loss of the element at that locus. However, if the template has a copy of the element then the original donor will effectively be regenerated.

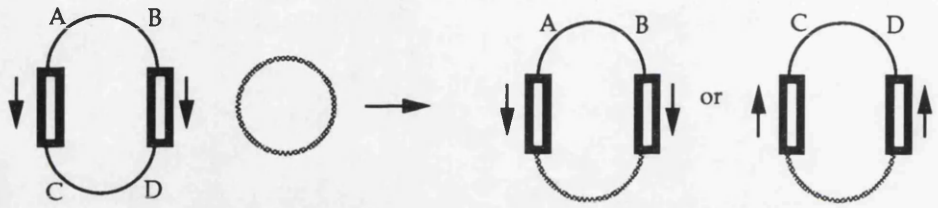
The compound transposon Tn10 has been demonstrated both *in vivo* and *in vitro* to transpose via a conservative mechanism(Morisato and Kleckner 1984 ; Bender and Kleckner 1986 ; Roberts and Kleckner 1988 ; Benjamin and Kleckner 1992). It has a mechanism whereby it increases the likelihood that the double strand gap repair will restore the donor site with a copy of Tn10. In wild type *E. coli* adenine residues within the sequence GATC are methylated at the N-6 position by the Dam methylase. However, following replication, the *dam* sites are transiently hemi-methylated. Elegant *in vivo* studies with IS10 revealed that IS10 transposition is activated by hemi-methylation (Roberts, Hoopes et al. 1985). This ensures that transposition is induced following replication. Further, one of the hemi-methylated forms is induced more strongly than the other. Combined, these effects serve to induce transposition from one of the sister molecules. As this event occurs shortly after replication the intact sister is available to act as a repair template, thus ensuring survival of the element at this locus.
Thus, although the actual transposition mechanism of Tn10 is conservative, the end result of the process is effectively replicative. This problem has dogged the study of *in vivo* transposition mechanism. Whilst it is possible by mutational analysis to determine if an element

1) Simple Insertion

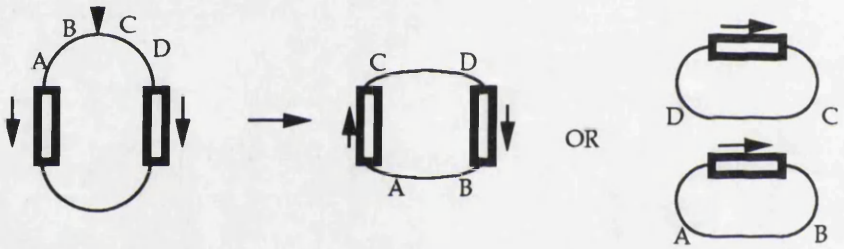
2) Cointegrate formation



3) Inverse Transposition



4) Transposition to a site within the element



5) Intramolecular replicative transposition

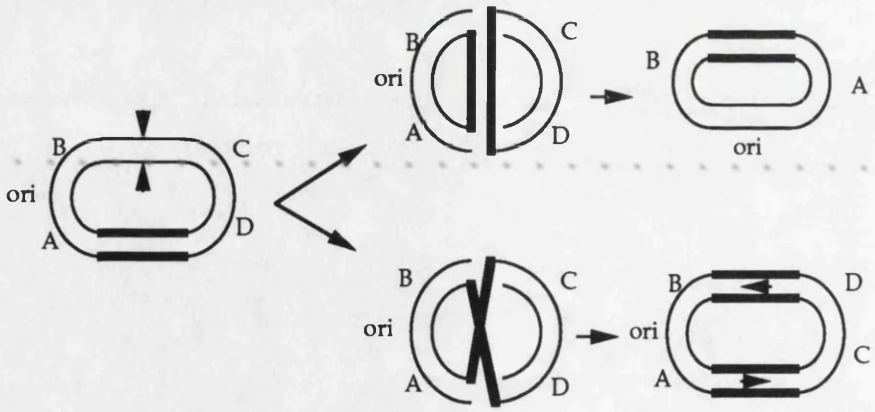


Fig1.2 Potential Transposon induced rearrangements

transposes replicatively; mutation of the resolvase gene, or its site, will stabilise co-integrates and allow their detection, it is, however, harder to rigorously prove, *in vivo*, that an element transposes conservatively.

It is necessary, therefore, to develop an *in vitro* transposition reaction in order to be able to address questions of reaction mechanism. It has also become clear from *in vitro* studies that manipulation of reaction conditions enables the stabilisation of normally transient intermediates in the reaction. Thus the development of *in vitro* reactions both addresses the overall reaction mechanism, and also enables molecular dissection of individual steps in the reaction.

1.7 Consequences of Transposition for the host cell

The most obvious effect a transposon can have is to act as an insertional mutagen, indeed it was by virtue of this property that transposable elements were first identified. However transposons can also act to cause large scale alterations to the chrormosome structure of the host. If several copies of an element exist within an organism's genome then there is the potential for homologous recombination events to occur between different copies of the element. This can, therefore, lead to either inversions, deletions or translocations of large regions of DNA. In addition to these rearrangements brought about by the host cells homologous recombination machinery, the actual transposition process itself can lead to similar large scale inversions and deletions of the host geneome. In particlular intramolecular transposition can be potentially highly damaging to the host replicon. as detailed in Kleckner, 1989 composite transposons have the largest potential to generate rearrangements. Potential transposon mediated rearrangements are shown in Fig. 1.2.

1.8 *In vitro* Transposition Systems

1.8.1 Phage Mu

Mu is a 37kb bacteriophage which replicates its genome by a replicative transposition mechanism.

Mu encodes two proteins which are required in the transposition reaction; MuA, the transposase and MuB, involved in activating strand transfer by MuA and also in target molecule selection (Mizuuchi, 1992).
MuA has been demonstrated to bind to multiple repeat motifs at the Mu termini (Craigie, 1984). The sites are based on a 22 bp consensus and are present in three copies at both ends of the element. The sites are designated L1, L2 and L3 at the left end and R1, R2 and R3 at the right end. A separate domain of MuA also binds a transpositional enhancer, the IAS (Mizuuchi, 1989). The structure and function of the IAS is discussed in Chapter 3.

Reconstruction of a Mu *in vitro* transposition reaction has allowed identification of the individual steps in the reaction. In the presence of MuA, the cellular proteins Hu and IHF, and Ca^{++} a mini-Mu substrate comprising Mu ends and the IAS forms a stable complex in which MuA has brought about synapsis of the Mu ends, this complex is termed the SSC (Surette, 1992). Addition of Mg^{++} to the reaction causes MuA to nick at the 3'-OH ends of Mu to form a Cleaved Donor Complex (CDC) (Craigie, 1987). Addition of target DNA, MuB and ATP to the CDC catalyses the strand transfer reaction. Transposition can occur in the absence of MuB, however the presence of MuB greatly enhances the rate of the strand exchange reaction (Baker, 1991).

1.8.2 Tn10

In Tn10 transposition there is only one element encoded protein required in the *in vitro* system, the transposase encoded by the IS10(right). The transposase encoded by IS10(left) is non-functional. Once again the host encoded proteins Hu and IHF play a stimulatory role in *in vitro* transposition reactions. The roles of these and other host proteins in transposition of a variety of elements is discussed in Chapter 6.

The Tn10 *in vitro* reaction produces products analogous to the Mu CDC and STC, the terminology used, however is different. To date no complex corresponding to synapsed but non-cleaved ends has been detected. Because Tn10 transposes non-replicatively the CDC analogue is an excised transposon form (ETF) of Tn10 created by double strand cleavage at the ends of the element (Haniford, 1991). The ends are synapsed by bound

transposase. Synapsis of the ends presumably occurs prior to cleavage as the excised transposon retains its supercoiling. The production of the strand transfer complex occurs primarily as an intramolecular event, producing a relaxed transposon circle as the principle product (Morisato, 1987). Advances in the Kleckner laboratory, primarily in purification of the transposase, reported at a recent meeting, have permitted the development of an efficient intermolecular *in vitro* reaction.

1.8.3 Retro-elements

Retrotransposons and retroviruses share the ability to reverse transcribe an RNA copy of their genome to double stranded DNA and integrate this DNA into the host genome. The integration event is catalysed by the retro-element encoded protein, IN (Craigie, 1990).
In the majority of retro-elements thus far studied the IN protein initially serves to trim two nucleotides from the 3' end of the element to expose the true ends of the element. This single strand cleavage can be seen as analogous to the nicking reaction carried out by MuA (Haniford, 1992). This event is not an obligatory step in the integration process, artificially generated "pre-trimmed" substrates are efficiently used *in vitro.* Furthermore, most of the yeast Ty retrotransposons have reverse-transcripts exactly the same length as the integrated element (Sandmeyer, 1992). Trimming therefore appears to function purely to expose the ends of the element.
The integration event apparently proceeds by a one step trans-esterification reaction as is found for Mu transposition (Engelman, 1991; Mizuuchi, 1991).

1.9 Transposition Insertion Sites

A feature common to most transposable elements is that they create a duplication of host sequences at the site of integration. This is ascribed to a staggered break with 5' extensions being generated in the target DNA. The 3'-OH of the transposable element are then ligated to these 5' ends. The result of this is to generate an insertion flanked by a short direct repeat of the host sequence. The length of the direct repeat varies between elements but is constant for a given element; e.g. Tn10 duplicates 9 bp, Tn7 5 bp, IS1 9 bp, Tn5 4 bp and Tn3 5bp. Whilst for some elements the choice of target

site is apparently random a growing body of evidence suggests that both sequence and structural determinants of the target DNA play important roles in determining insertion site specificity. As discussed below this is at its most extreme in the hot site transposition of Tn7, the predominant recipient site of Tn7 transposition on introduction to naive cells is a single locus in the genome, *attTn7* (Lichtenstein, 1981; Lichtenstein, 1982). This is discussed in greater detail below.
Studies of the major IS1 insertion site in the plasmid pBR322 have revealed that this locus contains several binding sites for the host encoded protein IHF (Prentki, 1987). IHF induces strong bends in this region, it is possible, therefore, that local perturbation of the DNA structure may favour insertion of IS1 at this locus.

Tn10 has a favoured target sequence, NGCTNAGCN. Sequences surrounding this consensus also influence the use of target sites (Lee, 1987; Bender, 1992). Intriguingly IHF also appears to influence target choice in Tn10 intramolecular reactions. This effect may be mediated by IHF binding to the end of the element and inducing a bend which brings internal sequences close to the synapsed ends of the ETF (unpublished data cited in Mizuuchi, 1992).

An initial analysis of favoured retroviral integration sites suggested that retroviruses prefer to insert in transcriptionally active regions and regions which demonstrate DNAseI hypersensitivity in the host genome. Recent work *in vitro* studied the integration sites of MoMLV on nucleosome associated DNA. This gave the suprising result that the presence of nucleosomes on a given stretch of DNA actually stimulated insertion in that DNA. It is probable that some structural feature of the DNA, induced by nucleosome binding, causes this selection(Pryciak, 1992; Craigie, 1992).

1.10 Discrimination between intra- and inter molecular transposition

A number of transposable elements appear to able to discriminate between target sites near to the transposon donor site and those located at a distance and to preferentially insert at the distant sites. This is the phenomenon of transpositional immunity. Another manifestation of the same process is that the presence of the transposable element in a DNA molecule blocks a second insertion event in that molecule. In the systems thus far studied

the immunity is conferred by the ends of the transposable element, specifically those sequences involved in transposase binding. In the case of Mu the presence of a single Mu end will prevent insertion of Mu within approximately 20 kb of the single end. Immunity is also observed with Tn7 and the Tn3 family. The mechanism of immunity in the Tn3 family is still mysterious, however, with the development of *in vitro* systems for Mu and Tn7 transposition some aspects of this phenomenon in these systems are becoming clear. Both Mu and Tn7 differ from Tn3 in that they encode several transposition proteins, including functions whose roles are specifically in target site selection. Thus, separate proteins are involved in transposon end binding and target site binding. The case of Mu is currently the best understood.

As described above MuA is responsible for end binding, synapsis and strand cleavage while MuB acts as a target site selector and activator of MuA catalysed strand transfer reaction. MuB is an ATP dependent DNA binding protein with little or no sequence specificity, its affinity for DNA is much greater in the presence of ATP (Adzuma, 1988). MuA protein induces the hydrolysis of ATP by MuB and thus its dissociation from DNA. Therefore, if MuA is bound to a Mu end on a DNA substrate it will effectively keep that molecule free of stably bound MuB. The net effect of these processes is to cause MuB to accumulate on DNA which does not possess Mu ends (Adzuma, 1988). A potential paradox exists as at some point MuA bound to transposon ends must come in contact with MuB bound to target DNA and perform a strand exchange reaction rather than simply cause MuB dissociation. A partial resolution of this paradox comes from the observation that MuA within the Cleaved Donor Complex is unable to cause MuB dissociation. However, once the Strand Transfer Complex is formed MuA must recover its ability to remove MuB in order to prevent the subsequent insertion of further donors into the target. Clearly several questions remain to be answered about this process.

1.11 Tn7

Tn7 is a large and complex transposon with a number of peculiar and unique properties(Craig, 1991). It encodes an unusually high number of genes whose functions are required for transposition, it transposes to a specific site, *attTn7*, at a high frequency and it can undergo two distinct transposition pathways, the site specific reaction and also a lower

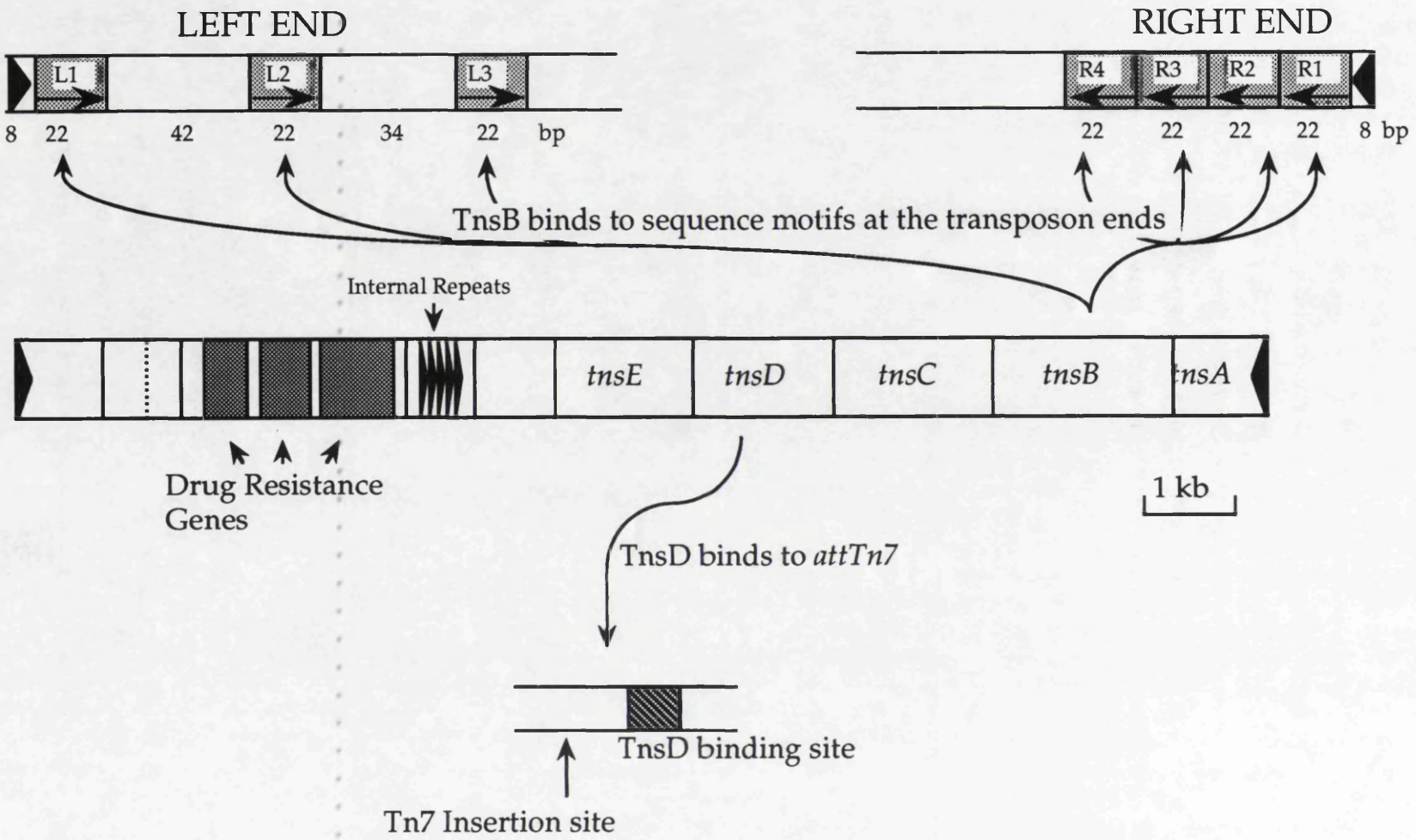


Fig. 1.3 Diagram of the structure ofTn7

frequency random event. Five Tn7 encoded genes, *tnsABCDE*, are required for transposition, overlapping subsets of these genes are required for the two pathways. The pathway directed to the "hot site", *attTn7*, requires *tnsABC* and *D* while the random site reaction requires *tnsABC* and *E*. (Rogers, 1986 ; Waddell, 1988 ; Smith, 1986)
The *tns* genes were identified by complementation assays *in vivo* and have subsequently been sequenced (Flores, 1990). They are encoded in the right end of the element and are very closely linked to each other. In a number of cases the putative coding regions overlap by a few nucleotides. This has led to the suggestion that the genes may be translationally coupled.
Tn7 also encodes three drug resistance determinants, for streptomycin/spectinomycin resistance, trimethoprim resistance and streptothricin resistence and an lambda like integrase pseudogene(Sundstrom, 1991). These genes are in a structure that resembles the integron gene recruitment system found in the Tn21 family and several broad host range plasmids(Amabile-Cuevas, 1992). The significance of this structure will be discussed in Chapter 3.

When this project was started the majority of the sequence of Tn7 had been determined. Chapter 3 describes work to complete the sequence and subsequent analysis of the data obtained.

1.12 *cis* Acting Sequences at Tn7 Ends

The ends of Tn7 possess a complex arrangement of sequence motifs. The extreme termini of Tn7 are perfect inverted repeats of an eight base pair element. The eight bp motif is adjacent to a 22bp element, thus the ends of Tn7 are effectively imperfect 30bp inverted repeats. The right end of Tn7 contains 3 more copies of the 22bp repeat, these copies are directly repeated and abut onto one another. Contrastingly, the left end of Tn7 contains a further two copies of the 22bp element, the left end motifs are also directly repeated but are more widely spaced than the right end repeats. See Fig. 1.3. These 22bp motifs are bound by TnsB (McKown, 1987; Arciszewska, 1989 ; Morrell, 1990; Arciszewska, 1991; Arciszewska, 1991; Tang, 1991). Work presented in Chapter 3 examines the sequence requirements for TnsB binding.

In addition to being structurally diverged the two ends are also functionally distinct. A mini-Tn7 element with two right ends in inverted repeat transposes at a near to wild type level, however an element with two left ends in inverted repeat fails to transpose. Intriguingly a single Tn7 right end confers transposition immunity to a 50kb plasmid, however, a left end does not have this effect(Arciszewska, 1989).

1.13 *cis* Acting Sequences at *attTn7*

The specific target site directed by the TnsABCD pathway is located at minute 84 of the *E. coli* chromosome (Lichtenstein, 1981 ; Lichtenstein, 1982). The insertion site lies within the transcriptional terminator of the *GlmS* gene (Gay, 1986). This gene is highly conserved amongst Gram negative bacteria and appears to function as a site for Tn7 insertion in these bacteria. The discrimination between site specific and random site pathways is made by the use of TnsD or TnsE respectively. It has been shown that TnsD binds specifically to DNA sequences to the right of the actual insertion site. The central nucleotide of the 5 bp sequence duplicated on Tn7 insertion is designated position 0, positions to the right, *GlmS* side, are given a positive sign, those to the left, *PstS*, side are given a negative sign. The sequence determinants required for TnsD binding lie between +28 to +53. These are the only sequences required to direct wild type levels of Tn7 insertion *in vivo*; the sequence at the actual site of insertion can be altered with no effect on its function as a target site (Qadri, 1986 ; Gringauz, 1988 ; Kubo, 1990 ; Waddell, 1989). A further feature of Tn7 insertion into *attTn7* is that it occurs in a defined orientation, the Tn7 right end is proximal to *GlmS*.

When the *attTn7* site is occupied by Tn7, although the sequences which define site activity are not altered, the site is not used as a recipient for a further insertion event. This is one manifestation of transposition immunity mediated by Tn7. When the site is occupied TnsABCD mediated transposition events still occur to related sequences in the genome, these sites are called pseudo *attTn7* sites and have sequences homologous to the +23 to +53 TnsD binding site (Kubo, 1990).
These rare pseudo *att* site events occur at approximately the same frequency as the random site transposition supported by TnsABCE, roughly 100 fold lower than the rate of *attTn7* transposition. It has been

proposed that by using the two transposition pathways Tn7 is able to exist in a safe "harbour" in the hotsite yet be able to transpose into conjugative plasmids passing through the host. This will ensure both vertical and horizontal transfer of the element.

1.14 Functions of the *tns* genes

1.14.1 *tnsA*

tnsA encodes a 31 kD protein. As with all the Tns proteins, this species has been observed both by immunoblotting, (Orle, 1991), and directly in whole cell lysates prepared from cells containing an over expression construction which have been electrophoresed on Laemmli gel followed by Coomassie staining (See Chapter 4). Although TnsA is absolutely required in both *in vivo* and *in vitro* transposition, its biochemical role is still a mystery.

1.14.2 *tnsB*

tnsB encodes an 85 kd protein. This protein has been purified and shown to bind to the 22bp repeats at the ends of Tn7. It contains a putative helix turn helix motif and has been demonstrated, by methylation protection studies, to interact with DNA primarily in the major groove. Hydroxy radical footprinting of TnsB interacting with the multiple repeats at the end of Tn7 has shown the TnsB to be present primarily on one side of the DNA double helix.

Experiments with purified TnsB have been unable to detect TnsB mediated synapsis of Tn7 ends or any cleavage or strand exchange reactions. It is possible that another *tns* protein carries out these reactions, however, by analogy with the transposases of other transposable elements it seems likely that these functions are specified by TnsB; perhaps TnsB requires an interaction with other components of the Tn7 transposition machinery to stimulate this activity (Arciszewska, 1991 ; Arciszewska, 1991; Morrell, 1990 ; Tang, 1991).

In addition to the direct role TnsB plays in transposition it also is involved in regulation of expression of the *tns* genes. The innermost 22 bp repeat

bound by Tn7 overlaps the major right end promoter of Tn7. Transcription from this promoter is reduced on TnsB binding.

1.14.3 *tnsC*

The 62kd protein encoded by *tnsC* has recently been purified and shown to be an ATP dependent non-specific DNA binding protein (Gamas and Craig 1992). Although the purified protein has not been shown to demonstrate an ATP hydrolysis activity, it is thought that the requirement for ATP hydrolysis in the Tn7 *in vitro* reaction described below is mediated by TnsC. Recent work by the Craig group has shown that in the presence of AMP-PNP, a non-hydrolysable analogue of ATP the Tn7 *in vitro* reaction can proceed with just TnsA, B and C. The transposon insertion sites are apparently chosen at random and are presumed to reflect the non-specific DNA binding activity of TnsC. Because of this observation it has been proposed that TnsC is involved in target recognition, and has been demonstrated to interact with TnsD at the *attTn7* site (R. Bainton pers comm).

1.14.4*tnsD*

The *tnsD* gene product has a predicted molecular weight of 58 kd and has been shown to bind to the *attTn7* sequences +23 to +53 (Waddell, 1989). This region of DNA contains an inverted repeat, however this inverted repeat alone does not bind TnsD. It is intriguing to note that the vast majority, >90%, of TnsABCD mediated Tn7 insertions occur assymetrically, to the left side of the TnsD recognition site. If, by analogy with other proteins which bind to inverted repeats, TnsD binds as a symmetric dimer how is this directional bias exerted? The TnsD gene has been shown to interact with TnsC at the attsite; in the presence of TnsD, TnsC binds between the TnsD binding site and the actual site of Tn7 insertion.

1.15 Tn7 Transposition *in vitro*

While the work for this thesis was being carried out details of a Tn7 *in vitro* transposition reaction were published (Bainton, 1991). This work used fully or partially purified extracts from cells overexpressing the

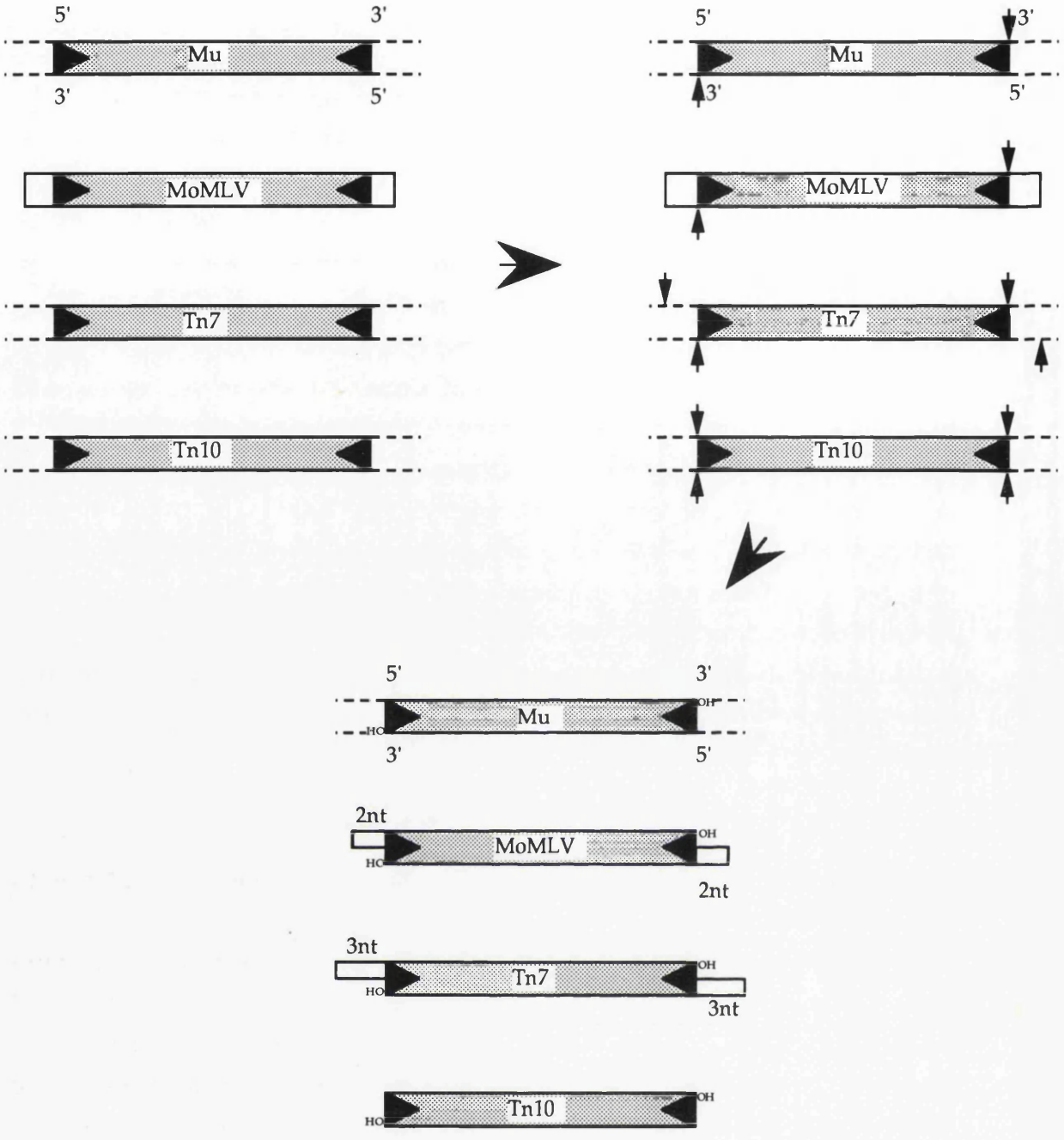


Fig 1.4 A comparison of the structures of intermediates formed in the transposition reactions of Mu, Tn10, Tn7 and retroviruses.

tnsA,B,C and *D* genes. These extracts were combined and were demonstrated to faithfully reproduce att site specific transposition using defined substrates *in vitro*. This showed that Tn7 transposes via a conservative pathway, in which an excised linear transposon (ELT) is an intermediate. Double strand breaks are produced at the ends of the element and the exposed 3' OH are joined to 5' phosphates in the cleaved recipient molecule. Several novelties are observed in the Tn7 reaction. Cleavage at the Tn7 ends is staggered, the cuts exactly expose the 3' ends of the element but create 5' extensions of donor backbone sequences, this extension is at least 3 nucleotides long. This, therefore generates an ELT that has some similarity to a trimmed retrovirus, however, with retroviruses the 5' extension is 2 nt long, see Fig. 1.4.
Unlike the case of the Mu reaction no intermediates in the Tn7 reaction are detected unless all the components of the reaction are present. Thus the Tn7 *in vitro* reaction is highly concerted, presumably requiring assembly of a large nucleoprotein complex containing all 4 proteins and both donor and target DNAs prior to any strand cleavage and exchange reactions.

The reaction requires the presence of ATP and Mg^{++}. However a preincubation prior to addition is essential. One effect of Mg^{++} is to inhibit binding to *attTn7* by TnsD (Chapter 5, this thesis).
If a non-hydrolysable analogue of ATP, AMP-PNP, is used in the *in vitro* reaction TnsD is no longer required for transposition. Thus TnsA,B and C are sufficient to promote transposition under these conditions. The target site is apparently completely random, probably reflecting the non-specific association of TnsC-AMP-PNP with DNA. Presumably TnsD interacts with TnsC and stabilises binding to *attTn7* containing DNA. Unpublished results from the Craig laboratory confirm this hypothesis, (R. Bainton pers. comm.). The phenomenon of transposition immunity is abolished in the presence of AMP-PNP. It is tempting to suggest that TnsC may play a role in target selection similar to that of MuB. An obvious difference between MuB and TnsC is that in the Mu system donor strand cleavage occurs prior to recruitment of target DNA+MuB to the synapse. As stated above, with Tn7 the reaction is highly concerted, no donor cleavage is seen in the absence of target. This may reflect the crucial difference between the replicative transposition of Mu and the conservative reaction used by Tn7. The analogous complex to the Mu CDC in Tn7 is a fully excised copy of the

element, the ELT. Generation of the ELT by definition causes the breakage and potential loss of the donor molecule. It, therefore, would make sense only to perform this highly committing reaction step when there is a suitable target site available.

It is important to note that, because of the intra-molecular nature of the published Tn10 *in vitro* reaction, it is not possible to determine if this conservative transposon also displays such a requirement for a concerted reaction.

1.16 Aims of This Project

The initial aim of this project was to compile the complete DNA sequence of Tn7. When this work was started over 12 kb of the DNA sequence of Tn7 had already been determined. Work presented in Chapter 3 describes the sequencing of the remaining segments of Tn7. This chapter also presents an analysis of the sequence and in particular examines the potential role of a series of repeated DNA motifs in the middle of the element.

A long term aim of the Tn7 project in Glasgow has been to characterise the biochemical steps in the Tn7 transposition reaction. In order to address this matter it is necessary to develop an *in vitro* Tn7 transposition assay. For a number of reasons it was felt that the best approach to take would be to use crude cell extracts containing high levels of the *tns* gene products to develop such a system. Once a functional system had been developed it could then be used as an assay for the subsequent purification of the individual Tns proteins. Chapter 4 describes the construction of plasmids designed to exploit the operon like structure of *tnsABCD* to coordinately overexpress the gene products. This chapter also describes assays of the ability of these constructs to support transposition of a mini-Tn7 element *in vivo* and finally discusses attempts to establish an *in vitro* transposition reaction.

Chapter 5 presents data on the interaction of various Tns proteins in the extracts prepared from cells containing the above constructs. This chapter also presents data on the binding of TnsD to the attsite, and in particular addresses the effect of Mg ions on binding.

Work on other bacterial transposition systems have revealed that host encoded proteins can play an accessory role in the transposition process.. The Craig group have shown that the minimal protein requirements for the reconstituted Tn7*in vitro* reaction are the purified *tns* gene products. It has been noted that addition of a boiled and spun host extract to the *in vitro* reaction does, however, stimulate the reaction(Bainton, 1991). Furthermore in the characterisation of TnsD binding to *attTn7* it was observed that a host factor also bound to this site(Waddell, 1989). Work presented in Chapter 6 attempts to identify this host factor and demonstrates a role for this factor in *in vivo* transposition of Tn7.

Chapter 2

Materials and Methods

2.1 Bacterial strains.

The derivatives of *Escherichia coli* K-12 used are listed in Table 2.1.

2.2 Plasmids and Phagemids.

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

2.3 Bacteriophage.

Phage P1 was a kind gift of Dr. M. Masters.

2.4 Oligonucleotides

The oligonucleotides listed in Table 2.3 were synthesised on an Applied Biosystems PCR Mate.

2.5 Chemicals.

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

2.6 Proteins.

Restriction and DNA modification enzymes	BRL, Boehringer Mannheim, Promega, Pharmacia.
IHF	A generous gift of G. Szatmari.

Table 2.1 Bacterial Strains

Strain	Genotype	Source
DS801 (AB1157)	*thr*1, *leu*6, *his*G4, *thi*1, *ara*14, *pro*A2, *arg*E3, *gal*K2, *sup*37, *xyl*15, *mtl*1, *tsx*33, *str*31	D. Sherratt
DS941	DS801, but *rec*F143, *sup*E44, *lacZ*ΔM15, *lacI*^{q}	D. Sherratt
CSH50	*ara*,Δ*lacpro*, *strA*, *thiA*	D. Sherratt
MR4	DS903 *attTn7*::Tn7-1	M. Rogers
	DS903 is DS801 but *recF*143	
CSH50*attTn7*::Tn7-1	Tn7-1 introduced to CSH50 by P1 transduction from MR4	Chapter 4
DS941*attTn7*::Tn7-1	Tn7-1 introduced to DS941 by P1 transduction from CSH50::Tn7-1	Chapter 4
DS940	N99 *hinA*Δ82::Tn10	D. Sherratt
DS978	DS941 but *fis::Kn*	D. Sherratt
GM320	F$^-$ *Δ(argF-lac)U*169, *ara*D139, *rpsL*150, *ptsF*25, *flbB*5381, *rbsR*, *deo*C1, *relA*1 *Φ(proU-lacZ) hyb*2 (*λplacMu*150 *osmZ* 205::Tn10	G. May
SB10	DS941 but *himA*::Tn10	Chapter 6
JM109	*F', tra D*36, *lacI*^{q}*Δ(lacZ)* M15 *proAB* */Δ(lac-proAB) thi gyrA96 endA1 hsdR*17 *(r*$_K$*-m*$_K$*+) supE*44 *relA1 ΔlacZproAB*	Promega

Table 2.1cont. Bacterial Strains

Strain	Genotype	Source
BMH 71-18 mutS	C*Δ(lac-proAB)* *[mutS*::Tn10] *[F', proAB, lacI^q*ZΔM15]	Promega
BL21	*hsd Gal*	W. Studier
BL21(DE3)	*hsd Gal (λcI ts 857 ind1 Sam7 ini*5 *lac*UV5-T7 gene-1)	W. Studier
DS916	*recA his trp spc^r rif^r λ$^+$*	D. Sherratt
ET12567	F$^-$ *dam*-13::Tn9 *dcm-6 hsdM hsdR* *rec*F143::Tn10 *galK*2 *galT*22 *ara*-14 *lacy*-1 *xyl*-5 *leuB*d *thi*-1 *tonA*31 *rpsL*136 *hisG*4 *tsx*-78 *mtl*-1 *glnV* 44	Iain Hunter

Table 2.2: Plasmids

Name of Plasmid	Marker	Origin	Description	Source
pUC18	Ap	ColE1	Cloning vactor	Yanisch-perron et al, 1985
pBend2	Ap	ColE1	Cloning vactor	Kim et al, 1990
R388	Tp	IncW	Conjugative plasmid	Datta and Hedges, 1972
pEN300	Tp	IncW	1kb *attTn7* site in R388	Rogers, 1986
pEAL1	Tc	p15A	1kb *attTn7* site in pACYC184	Lichtenstein and Brenner, 1981
pMR11	ApCm	ColE1	Tn7-1 constructed in pUC8	Rogers, 1986
pMR80	Ap	ColE1	280bp *attTn7* site in pUC8	Rogers, 1986
pMR207	Tp	ColE1	Tn7 BglII (13.1)-HindIII(10.6) in pMR78 (TnsB)	Rogers, 1986
pMR64	Ap	ColE1	Tn7 EcoRV (9.8)-ClaI(7.6) in pKK2233 (TnsD)	Rogers, 1986
pMR51	Ap	ColE1	BamH1 (11.05)-ClaI (7.6) pUC18 (TnsCD)	Rogers, 1986
pMR84	Ap	ColE1	RE-ClaI (7.6) in pUC8 (TnsABCD)	Rogers, 1986
pNE200	Ap	ColE1	203bp Tn7 RE in pUC8	E. Nimmo
pSelect-1	Tc	ColE1/f1	Mutagenesis vector	Promega
pSEL-ANT	Tc	ColE1/f1	Tn7 RE-Pst1 (13.5) in pSelect-1	Chapter 4
pSEL-MUT	TcAp	ColE1/f1	Mutagenised pSEL-ANT	Chapter 4
pT7-7	Ap	ColE1	T7 expression vector	Stan Tabor
pT7-ANT	Ap	ColE1	NdeI/PstI fragment from pSEL-MUT in pT7-7	Chapter 4
pSB58	Ap	ColE1	PstI fragment from pMR84 cloned into pT7-ANT restoring TnsABCD	Chapter 4

Table 2.2 cont.

Plasmid Name	Marker	Origin	Description	Source
pT7-5	Ap	ColE1	T7 expression vector	Stan Tabor
pSB84	Ap	ColE1	EcoRI/Cla fragment of pMR9 cloned in pT7-5	Chapter 4
pGP1-2	Kn	p15A	Source of T7 RNA polymerase	Stan Tabor
pSB70	Ap	ColE1	EcoRI/PvuII fragment of pMR84 in pT7-5	Chapter 5
pSB100	Ap	ColE1	511bp internal NaeI fragment of pUC8::Tn7 cloned into pUC18	Chapter 3
pSB110	Ap	ColE1	RsaI/HindIII fragment from pSB100 in pUC18	Chapter 3
pSB200	Ap	ColE1	Central EcoRI/HindIII fragment of Tn7 in pUC18	Chapter 3
pSB210	Ap	ColE1	Tn7 HincII(0.5kb)-HincII (2.2kb) in pUC18	Chapter 3
pSB220	Ap	ColE1	AvaI (0.6)-HincII (2.2) of pSB210 subcloned in pUC18	Chapter 3
pSB230	Ap	ColE1	PvuII-HindIII fragment of pSB200 subcloned in pUC18	Chapter 3
pSB240	Ap	ColE1	EcoRI-XbaI fragment of pSB200 sucloned in pUC18	Chapter 3

Table 2.2 cont.

Plasmid Name	Marker	Origin	Description	Source
pSB250	Ap	ColE1	AvaI-EcoRI fragment of pSB200 subcloned in pUC18	Chapter 3
pSB260	Ap	ColE1	XbaI-HindIII fragment of pSB200 sucloned in pUC18	Chapter 3
pSB500	Ap	ColE1	*attTn7* containing HincII fragment of pMR80 in pBend-1	Chapter 6
pSB400	Ap	ColE1	*attTn7* +9 to +55 oligonucleotide cloned in pUC18	Chapter 6
pAG411	Ap	ColE1	TnsA(NdeI-HincII 1.4) in pT7-7	Angela Gawthrop

Table 2.3

Oligonucleotides used in this work

Oligonucleotides used as sequencing primers

Primer LINT2R) GCGCTTTTTATGTTAGGCAT

Primer LINT2L) TAATTATTTGTGGGTGGAGG

Primer LE2R) CCTGCCGCTCAGCTTAGTAC

Oligonucleotide used in mutagenesis of pT7-ANT

CCGGACTTTCATATGGCTAAAGCA

Oligonucleotides for cloning of *attTn7* +9 to +55

CTAGGGCATCCATTTATTACTCAACCGTAACCGATTTTGCCAGGTTACG

TCGACGTAACCTGGCAAAATCGGTTACGGTTGAGTAATAAATGGATGCC

2.7 Culture media.

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

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

Iso-Sensitest Broth: 23.4g Iso-Sensitest Broth Oxoid CM473 made up to 1 l with distilled water.

Isosensitest Agar: As Iso-Sensitest broth with 15g/l agar.

4 x Davis and Mingioli minimal salts (D&M salts): 28 g K_2HPO_4, 8 g KH_2PO_4, 1 g sodium citrate, 0.4 g $MgSO_4 \cdot H_2O$, made up to 1 litre in distilled water.

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

Medium 121: A 10x stock was prepared by dissolving 4.68g NaCl, 1.5g KCl, 1.08g NH_4Cl, 0.35g Na_2SO_4, 0.2g $MgCl_2$, 0·029g $CaCl_2$, 5x10-4g $FeCl_3$, $2.7x10^{-4}$g $ZnCl_2$, $2.7x\ 10^{-4}$g Tris in 100ml of distilled H_2O. The pH was adjusted to 7.5 with HCl. Plates were prepared by adding 10 ml 10x Medium 121 to 75 ml 2% agarose in distilled water and adding glucose, amino-acids and thiamine as described above.
K_2HPO_4 was added to $8.3x10^{-4}$ M(excess Pi) or $8.3\ x10^{-5}$M (limiting Pi) and the volume made up to 100ml. XP was added to 20µg/ml and plates poured.

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

2.8 Sterilisation.

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

2.9 Buffer solutions.

DNA Electrophoresis

50 xE buffer: 1210g Tris, 410g sodium acetate, 93g $Na_2EDTA.2H_2O$, made up to 5 litres in distilled water, adjusted to pH 8.2 with glacial acetic acid.

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

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

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

4 x Horizontal agarose gel loading buffer: 25% sucrose, and 0.01% bromophenol blue in distilled water.

Protein gel running buffer: 10x stock prepared as 288.2 g Glycine, 60.6 g Tris base and 20g SDS base made up to 2 litre.

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

Buffers for DNA work

10 x restriction buffers: as supplied with restriction enzymes.

5 x ligation buffer: supplied by BRL.

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

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

EDTA: 500mM solutions of EDTA were made up and titrated to pH8.0 with 10M NaOH.

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

Phenol: All phenol used in the purification of DNA contained 0.1% 8-hydroxyquinoline and was buffered against 0.5M Tris-HCl pH8.0. The phenol was either distilled in the department or purchased pre-equilibrated from Fisons.

2.10 Antibiotics.

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

Antibiotic	Stock solution	Selective concentration
Ampicillin (Ap)	5 mg/ml (water)	50 µg/ml
Tetracycline (Tc)	1.2 mg/ml (10mM HCl)	12 µg/ml
Chloramphenicol (Cm)	2.5 mg/ml (ethanol)	25 µg/ml
Kanamycin (Km)	5 mg/ml (water)	50 µg/ml
Streptomycin (Str)	10 mg/ml(water)	100 µg/ml
Trimethoprim(Tp)	5mg/ml 50%EtOH	50µg/ml
Rifampicin(Rif)	5mg/ml MeOH	50µg/ml

All stock solutions were stored at 4°C.
Antibiotics were added to molten agar which was precooled to 55°C.

2.11 Growth conditions.

Liquid cultures for transformation, conjugation or DNA preparations were routinely grown in L broth, or Iso-sensitest when trimethoprim selection was required, at 37°C with vigorous shaking. Both L agar, isosensitest agar and minimal solid media were used. Antibiotics and supplements were used as required. Plates were generally incubated

overnight at 37°C. Any strains containing pGP1-2 were grown at 30°C to avoid induction of T7 RNA polymerase.
Bacterial strains were stored in 50% L broth, 20% glycerol and 1% peptone at -70°C or on L agar slopes at room temperature.

2.12 Indicators.

X-gal (5-bromo-4-chloro-3-indolyl-B-D-galactoside) was used in conjunction with the host strains DS941 and JM109 and the pUC and pSelect-1 vectors. Use of this indicator acts as a screen for plasmids with inserts in the polylinker region. Clones containing inserts were generally white; clones lacking inserts were blue. X-gal (40mg/ml in DMF) was stored at -20°C and added to L agar to a final concentration of 20µg/ml. IPTG (isopropyly-B-D-thiogalactoside) was added to a final concentration of 25 µg/ml to medium containing X-gal.
XP (5-bromo-4-chloro-3-indolyl phosphate) was used as described in Chapter 6 as a indicator of the level of Alkaline Phosphatase expression in cells.

2.13 Plasmid DNA isolation.

Large scale DNA preparation ((Birnboim, 1979); as modified in this laboratory).

Solutions:
I. 50mM glucose, 25mM Tris/HCl pH 8.0, 10mM EDTA.
II. 0.2M NaOH, 1% SDS; made fresh.
III. 5M potassium acetate pH 4.8; mix equal volumes of 3M CH_3COOK and 2M CH_3COOH, pH will be 4.8.

100ml cultures of stationary phase plasmid-containing cells were harvested by centrifugation (12,400g, 10 min at 4°C). The pellet was resuspended in 4 ml of solution I. 8 ml of solution II was added and the mixture was left on ice for a further 5 min. 6 ml of solution III was then added, gently mixed and the cell debris and chromosomal DNA removed by centrifugation (39,200g, 30 min at 4°C). The plasmid DNA was precipitated from the supernatant with 12 ml isopropanol for 15 minutes

at room temperature. The DNA was pelleted at 27,200g for 15 min at 20°C, rinsed with 70% ethanol and then further purified by banding on a CsCl/EtBr gradient. The DNA was resuspended in 2.1 ml of TE buffer and added to 270 µl of a 15 mg/ml ethidium bromide solution. 5 g of CsCl were dissolved in 3 ml of water and added to the DNA/EtBr solution. The gradients were centrifuged in a Beckman Ti70.1 fixed angle rotor at 200,000g for 16 hours at 20°C. Where two bands were visible, a lower supercoiled plasmid band and an upper nicked DNA band, the lower band was removed using a 1ml syringe. The ethidium bromide was removed by repeated butanol extractions and the DNA recovered by ethanol precipitation, the pellet washed in 70% ethanol and resuspended in 500µl TE.

Mini preparation of DNA

Small scale preparations of plasmid DNA were performed using the Birnboim, Doly mini-prep method

1.5 ml of a stationary culture was harvested by centrifugation in an Eppendorf tube (12,100g, 30sec) and resuspended in 100 µl of solution I (as described above for large scale prep.). 200µl of solution II was added and mixed by gentle inversion of the Eppendorf tube. Following a 5 minute incubation on ice 150µl of solution III was added and mixed by vortexing at the lowest setting of a Vortex Genie-2 (Scientific Instruments) for 10 seconds. Following a further 3 minutes on ice, the tube was centifuged (12,100g, 5 min). The supernatant was collected and extracted against 400µl of 1:1 phenol:chloroform. Following centrifugation (12,100g, 5 min) the supernatant was added to 2 volumes of ethanol. After a 5 minute incubation at room temperture and centrifugation (12,100g, 5 min) the supernatant was carefully removed and the pellet was washed with 1ml 70% ethanol, dried and resuspended in 50 µl TE buffer. 5-10 µl of this DNA was suitable for restriction enzyme digests. 1 µl of a 1 mg/ml RNase A solution was added to the DNA with the gel loading buffer.

2.14 Chromosomal DNA isolation.

Buffer A: 10mM Tris.Cl pH8.0
1mM EDTA pH8.0
0.1% Lauroyl Sarcosine
250ml NaCl
50µg/ml Proteinase K

100 ml of culture was harvested by centrifugation (12,400g, 10 min, 4°C) and resuspended in 10ml Buffer A. The suspension was frozen by immersion in liquid nitrogen and then thawed at 37°C for 45 minutes. 7ml of phenol were added and mixed well but gently, 3mls of chloroform were added, mixed and centrifuged (29,000g, 20 min). The extraction was repeated on the supernatant.
0.6 volumes isopropanol was added; after 1 hour at room temperature DNA was either spooled or recovered by centrifugation (29,000g 30 min). The pellet was washed with 70% ethanol, and resuspended in 1ml 2xTE. This DNA was suitable for restriction digestion.

2.15 Transformation with plasmid DNA.

Genetic transformation introduced plasmid DNA into different host strains. An overnight culture of the recipient strain was diluted 1 in 100 into 20 ml L broth and grown to a density of $2x10^8$ cells/ml (about 90 min). The cells were harvested (12,100g, 1 min, 4°C) and resuspended in 10 ml of 50mM $CaCl_2$. The cells were pelleted again, resuspended in 0.5 ml cold (4°C) 50mM $CaCl_2$ and kept on ice until use. 100 µl aliquots of competent cells were added to DNA in TE buffer and, after gentle mixing, were left on ice for 15 min. The cells were then heat shocked (2 min, 42°C or 5 min, 37°C) and returned to ice for 15 min. 100 µl of L broth was added to the cell suspension and incubated at 37°C for 90 min to allow expression of plasmid genes. Transformation to ampicillin resistance did not require this expression time. 100ul aliquots of the transformation mixture were spread onto selective plates.

2.16 Preparation of P1 lysates.

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

2.17 P1 transduction.

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

2.18 Single colony gel analysis.

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

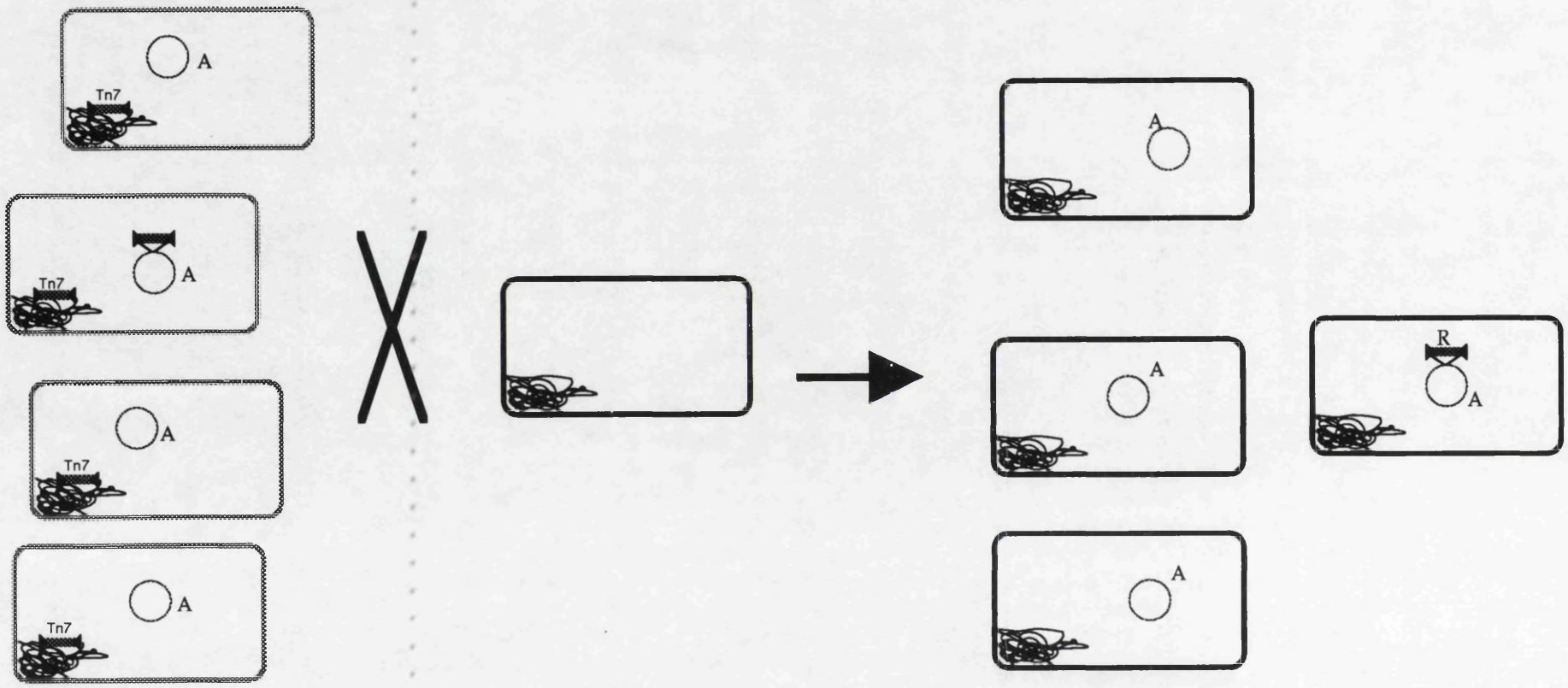


Fig. 2.1 Diagrammatic representation of the mate-out assay used to determine transposition frequency. The transposition rate is expressed as the ratio of Tn7 markers to marker A in exconjugants.

2.19 Bacterial Conjugation and Mate-out Assay

Conjugation mediated by the R388 derivative, pEN300 was used in the *in vivo* transposition assay. Cells containing a chromosomal Tn7-1 and appropriate *tns* encoding plasmid had pEN300 introduced to them by conjugation. Plate matings were carried out with 250 µl of both donor and recipient overnight culture being plated on a well dried L-agar plate. The plate was incubated at 37°C for 3-4 hours. 10 ml phage buffer were then added and the cells resuspended using a sterile glass spreader. The cell suspension was then transferred to a 40ml centrifuge tube and then centrifuged, (1200g, 5 min). The pellet was resupended in 1ml phage buffer and 10 fold serial dilutions plated on appropriate selectve plates. Following overnight selection for exconjugants, the new donor, containing Tn7-1, *tns* plasmid and pEN300 was used to innoculate a 2.5 ml culture. An analogous culture of recipient, usually DS916, was also set up. The following day another plate mating was carried out, as described above. Serial dilutions of the cell suspension, from 10^{0} to 10^{-7}, were made. 100 µl of the 10^{0} to 10^{-3} dilutions were plated on medium selecting for exconjugants containing Tn7-1, the 10^{0} to 10^{-4} were plated on medium selecting for exconjugants containing pEN300. The apparent transposition frquency is expressed as the ratio of tranfer of the Tn7-1 marker relative to the pEN300 marker (see Fig. 2.1).

2.20 Restriction of DNA.

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

2.21 Calf intestinal phosphatase (CIP) treatment.

Phosphatase was used to remove 5' phosphate groups from linearised vector to prevent recircularisation of the vector (thus increasing cloning

efficiency). 1 unit of CIP was added directly to the digest and incubated at 37°C for 15 min (or 45 min for blunt ends).

2.20 Ethanol precipitation of DNA.

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

2.23 Ligation of DNA fragments.

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

2.24 Gel electrophoresis of DNA.

a) Agarose gels

0.7-1.2% agarose gels were used.

Agarose powder was dissolved at 100°C in E buffer E and precooled to 55°C prior to use. Three types of gel apparatus were used:-

1) IBI Model QSH mini-gel kits: These kits required 25ml of molten agar. These gels were used for analysis of restriction digests.

2) IBI Model MPH midi-gel kits: Gels were 100ml of agar. These kits were used for most analysis of restriction digests, single colony gels and analyses of products of potential *in vitro* transposition reactions.

3) Large kits manufactured in the University which contained long 300ml gels. These gels were used to seperate restriction digests of genomic DNA

prior to Southern Blotting and for analysis of some attempted *in vitro* transposition reactions.

Following electrophoresis gels were stained in 0.6 μg/ml ethidium bromide. The DNA was visualised on a 254 nm wavelength UV transilluminator. If the DNA was to be isolated from the gel then it was visualised on a 365nm wavelength UV transilluminator.

b) Polyacrylamide gels
Three types were used.

1) Polyacrylamide gels for analysis of restriction products..

Vertical gel kits were used with 1.5 mm spacers. The gel apparatus was sealed with 0.6% agarose in H_2O or with a continuous length of rubber tubing compressed between the plates at the sides and bottom. An appropriate acrylamide gel mix was poured between the plates (with the insertion of a well former) and allowed to polymerise for up to 60 min.

50ml acrylamide gel consists of:-

30% acrylamide: 0.8% bisacrylamide (w/v)(Protogel)				X ml
10 x TBE	-	-	-	5 ml
10% APS (w/v)	-	-	-	200 μl
TEMED	-	-	-	20 μl

The volume was made up to 50 ml with distilled water

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

2) Non-denaturing polyacrylamide gels.

These gels were used to separate protein:DNA complexes. Two separate buffer systems were used

a) TBE based gels.

These gels were made as described above for analysis of restriction products. They were pre-run at 4°C at 200 volts for at least 45 minutes prior to loading of binding reactions. Following loading, electrophoresis was continued for 3 hours. The gel was then transferred to Whatman 3MM filter paper and dried under vacuum at 80°C. Bands were visualised by autoradiography on Kodak S1 or Fuji RX100 X-ray film for 1-3 days.

b) As detailed in the text some of the gels described in Chapter 3 were based on a Tris/Glycine buffer. This buffer, 50mM Tris/Glycine pH 9.4, 0.1mM EDTA, was found to enhance the resolution of complexes produced by TnsB binding to the ends of Tn7 (E. Morrell 1990).

These gels were pre-run and run as described above.

3) Polyacrylamide sequencing gels.

The gel kit used was an 'IBI Base Runner' kit. Gels were poured by 'sliding', as described by the manufacturers. Both straight and wedge spacers were used. Sequencing reactions were run on 8% high resolution polyacrylamide/urea gels in 1xTBE as described in the "IBI Base Runner Operator's Manual" published by IBI. Gels were pre-run for at least 45 min and run for 1-4 hours at 45W. Samples were denatured prior to loading (80°C, 2 min). After the gel run was complete, the gel was fixed in 10% acetic acid, 10% methanol for 15 min, dried under vacuum at 80°C and autoradiographed on Fuji RX100 X-ray film.

2.25 Purification of DNA fragments from Agarose Gels

Following electrophoresis, the gel was stained in distilled water with 0.6μg/ml ethidium bromide and viewed on a long wave transilluminator. The appropriate band was excised using a scalpel blade The agarose chip was then placed in a small (500 μl) Eppendorf tube over a small amount of siliconised glass wool covering a hole in the bottom of the tube. The small Eppendorf tube was then placed in a large (1.5 ml), lidless Eppendorf tube and centrifuged at 6000 rpm for 10 min. The liquid from the bottom tube was collected and the DNA was then precipated by the addition of NaOAc and ethanol. Following resuspension in distilled water the isolated DNA

could be used for ligation, further digestion or random priming as described below.

2.26 Synthesis and Purification of Oligonucleotides

Oligonucleotides were synthesised on an Applied Biosystems PCR mate. Oligonucleotides were removed fromthe column matrix and de-protected following the protocol below.
The column was opened and the glass bead packing material poured into an Eppendorf tube. 1ml of 30% aqueous ammonia was added and the tube incubated at room temperature for 1 hour. The tube was then vortexed, briefly centrifuged and the supernatant removed to a Nunc tube. A further 1ml of ammonia was added and the tube incubated at 50°C overnight.
Subsequent purification was as described in Sambrook et al (1990)

2.27 Photography of gels.

After staining in ethidium bromide, gels visualised by 254nm UV illumination were photographed using Polaroid type 667 land film or using a Pentax 35mm SRL loaded with Ilford HP5 film. Both cameras were fitted with a Kodak Wratten filter No.23A.

2.28 Electrophoresis of proteins in SDS-polyacrylamide gels.

The electrophoresis of proteins followed the procedure of (Laemmli, 1970) using 15% running gels and 4% stacking gels. Gels were poured in vertical gel kits sealed with 0.6% agarose in H_2O. The kits were either made in the University or BioRad "Mini-Protean" kits. Gel mixes for the home-made kits were made up according to the table below, mini protean kits used the same proportions of reagents but lower final quantities:

	Running Gel		Stacking Gel
	10%	15%	
30 / 0.8% acrylamide/bis (Protogel)	10	15 ml	2.0 ml
1.5M Tris-HCl pH8.8, 0.4%SDS	7.5 ml		
0.5M Tris-HCl pH6.8, 0.4%SDS			3.75 ml
10% SDS	0.3 ml		0.15 ml

H_2O	to 30 ml	to 15 ml
TEMED	20 µl	10 µl
Ammonium persulphate (10% w/v)	150 µl	150 µl

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

2.29 ^{35}S Labelling of Proteins

This procedure enabled radiolabelling of proteins encoded by genes cloned downstream of T7 promoters. Cells containing pGP1-2 and a recombinant pT7 plasmid were grown in L-broth containing ampicillin and kanamycin at 30°C. At OD_{590}=0.5, 0.2 ml of cells were centrifuged in a 1.5 ml Eppendorf tube (12,000g, 1 min). The cells were then washed in M9 medium re-centrifuged and resuspended in 1ml of M9+$MgSO_4$ supplemented with 20mg/ml thiamine (vitamin B1) and 0.01% of 18 amino acids (lacking methionine and cysteine). The cells were then grown with shaking at 30°C for one hour. Incubation temperature was then raised to 42°C for 15 min. Rifampicin was then added to 200µg/ml and incubation continued for a further 10 min. Temperature was then shifted back to 30°C for 20 min. Samples were then pulsed with 10µCi of ^{35}S-methionine for 5 min. Cells were then spun and the pellet resuspended in 50µl of protein gel loading buffer. Samples were then boiled for 3 min and loaded onto a protein gel. Following electrophoresis the gel was fixed in 30% methanol/10% acetic acid for 30 min then soaked in EN^{3}HANCE for 1 hour, transferred to soaking in water for a further 30 min, transferred onto Whatman 3MM paper and dried under vacuum at 60°C for 1.5 hours. The dried gel was then placed against an autoradiograph film and exposed at -70°C for 2 weeks.

Pulse/Chase experiment

The pulse chase experiment essentially follows the above protocol with the exception that 20μCi of ^{35}S-methionine were used, the pulse shortened to 1 minute, and immediately following the pulse unlabelled methionine added to 0.1%. Samples were then taken at various times into the chase and added to protein gel loading buffer as above.

2.30 Southern blotting and hybridisation.

Agarose gels were soaked in 0.25M HCl for 30 min, transferred to a solution of 1.5M NaCl, 0.5M NaOH for 30 min and finally neutralised in 1M Tris.Cl pH7.6, 1.5M NaCl for 30 minutes. The DNA was then transferred to a nylon membrane (Hybond N, Amersham) by pressure using a Posi-blotter (Stratagene) at 80mm Hg for 30-45 minutes using 20x SSC as transfer buffer. The blot was then UV crosslinked at 254nm for 90 seconds. Prehybridisation of the membrane was then performed for 3 hrs at 65°C in 10% dextran sulphate, 5 x Denhardt's solution, 1 x SSC, 0.5% SDS and 0.1 mg/ml denatured salmon sperm DNA. The blot was then hybridised overnight at 65°C in 10 ml of the same solution to which the denatured ^{32}P labelled probe had been added. Following hybridisation the probe was removed and the membrane washed in 2xSSC, 0.1% SDS at room temperature for 20 minutes, then twice with 1xSSC, 0.1% SDS at 65°C for 15 minutes each and finally twice with 0.1xSSC, 0.1% SDS for 15 minutes. The membrane was then dried and autoradiographed.

2.31 Random Priming of DNA.

The appropriate restriction fragment of DNA was gel isolated as described above. Following ethanol precipitation, 70% ethanol wash and drying, the pellet was resuspended in 9μl of distilled water. Random priming was carried out using a "Random Priming DNA Labeling Kit" (Boehringer Mannheim) following the manufacturers instructions. Unincorporated nucleotides were removed using a Nuctrap column (Stratagene).

2.32 DNA sequencing.

All DNA sequencing reactions were performed on double stranded plasmid DNA prepared by the large scale preparation method described

above. All sequencing reactions were carried out using the Sequenase enzyme and the reagents supplied in the Sequenase kit (United States Biochemicals). The method used was that described in the manual supplied with the kit with the exception that the initial labelling reaction was performed with the Eppendorf tubes lying on the surface of an ice bucket. This was found to reduce the incidence of sequence specific termination in all four tracks. The samples were denatured at 80°C and run on 8% polyacrylamide / urea gels as described above.

2.33 End-labelling of DNA fragments.

Purified DNA fragment were end-labelled by filling in recessed 3' ends generated by restriction digestion with the Klenow fragment of DNA polymerase I. The reaction contained 3-5µg DNA, 50nM unlabelled nucleotides (dCTP, dGTP and dTTP), 10uCi alpha ^{32}P dATP, restriction buffer (50mM Tris/HCl pH 8.2, 10mM $MgCl_2$, 50mM NaCl) and 1 unit/µg DNA Klenow enzyme in a total volume of 25µl. After incubation at room temperature for 1 hour, the reaction was stopped by heating to 70°C for 20 minutes. Unincorporated nucleotides were separated using a "NucTrap" column (Stratagene) as described in the manufacturers instructions.

2.33 Preparation of Crude Cell Extracts

Crude cell extracts of the various strains worked with were prepared as follows. 200mls of L-broth containing appropriate antibiotics were innoculated and grown overnight at 30°C, if containing pGP1-2, or otherwise at 37°C, with shaking. In the morning 200ml fresh L-broth plus antibiotics were added and cells were induced. For those cells with plac or ptac expression vectors induction was by addition of IPTG to 1mM, for those with pGP1-2 and T7 expression vectors induction was by heating to 42°C for 25 minutes then adding rifampicin to 100µg/ml prior to continuing growth at 37°C. Cultures were incubated for a further 3 hours and then harvested by centrifugation (4,420g, 15 min). The pellet was resuspended in 10 mls 1xD&M salts, transferred to a 40ml centrifuge tube and centrifuged at 12,000g, 5 min. The resultant pellet was then resuspended in 5ml of the appropriate buffer (see below) and lysozyme added to 100ng/ml. Cells were then disrupted by 6, 10 second passes of a sonicator. The lysate was then centrifuged at 49,000g for 25 minutes. The

supernatant was collected, protein concentration quantitated, and either directly used in gel binding or *in vitro* transposition reactions or was stored at -20°C following addition of glycerol to 50%.

2.34 Protein Quantiation

This was performed according to the method of Bradford, 1976

2.35 Gel binding assays.

TnsB

Binding assays with extracts containing TnsB were performed in 50mM Tris/glycine (pH 9.4), 55mM KCl, 1.1mM EDTA 1.1mM DTT. The substrate DNA was either end-labelled pSB100 (digested with EcoRI, HindIII and RsaI), pSB110 (Digested with EcoRI and HindIII) or pNE200 (digested with EcoRI and HindIII).
Binding reactions comprised:-

2 µl extract
20ng Labelled DNA
1µg Salmon Sperm DNA
18µl Binding buffer

Reactions were incubated for 10 minutes at room temperature, put on ice and loaded onto running non denaturing polyacrylamide gel.

TnsD Binding Assays

The standard TnsD binding Buffer was

50 mM Tris.Cl pH8.0
1mM EDTA pH 8.0
100mM KCl
1 mM DTT

The DNA substrate used in these experiments was end-labelled,EcoRI, HindIII digested pMR80. Other substrates were used as described in Chapter 6.
The binding reactions were as described above for TnsB.

2 µl extract
20ng Labelled DNA
1µg Salmon Sperm DNA
18µl Binding buffer

2.35 Site Directed Mutagenesis

This was performed as described in Lewis (1990) using the Altered Sites *in vitro* Mutagenesis System (Promega). See Chapter 4 for further details.

2.36 Attempts at Tn7 *in vitro* transposition reactions

These attempts essentially followed the protocol described by Bainton *et al.* (1991).
Briefly, a varying quantity, up to 500 µg, of cell extract was incubated with 0.5µg of transposon donor DNA, pMR11 and 1µg of recipient DNA, pEAL1 in 100µl reaction mixtures containing 2 mM ATP, 26 mM Hepes (pH8.0), 1.3mM Tris.Cl (pH8.0), 2.5mM KPO_4 (pH8.0), 100mM KCl, 15mM NaCl, 0.1mM EDTA, 2.2mM DTT, 100µg/ml tRNA, 50µg/ml BSA and 5% w/v PEG 8000.. The reaction was incubated for 7 minutes at 30°C then Mg acetate added to 15mM. Incubations were then continued for a further 30 minutes and DNA recovered as detailed in Bainton *et al.* (1991).

Chapter 3

The DNA Sequence of Tn7

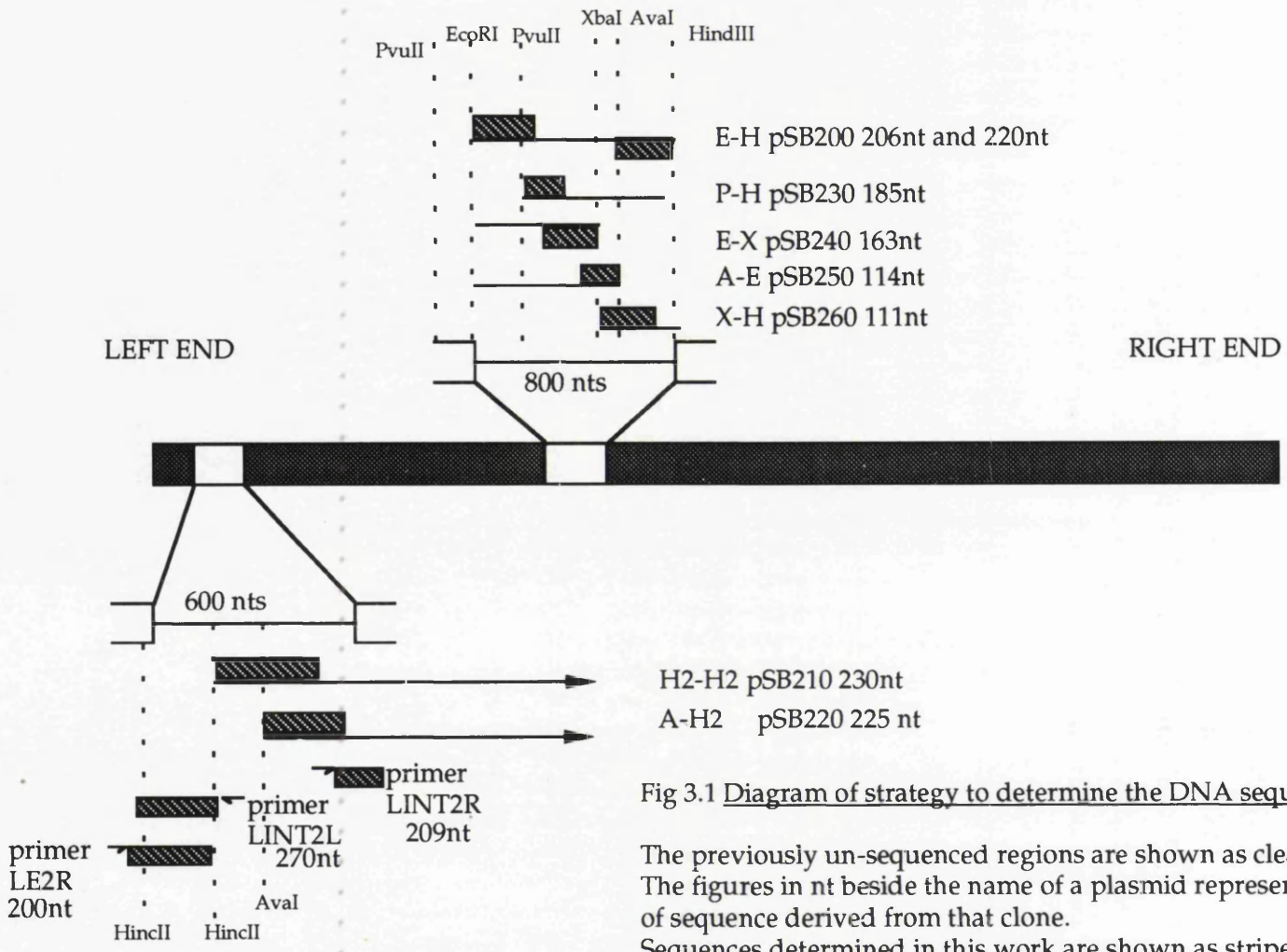


Fig 3.1 Diagram of strategy to determine the DNA sequence of Tn7

The previously un-sequenced regions are shown as clear boxes. The figures in nt beside the name of a plasmid represent the length of sequence derived from that clone.
Sequences determined in this work are shown as striped boxes. Boxes above the line of individual clones represent top strand sequence, those below, bottom strand. The symbol ← represents a synthetic oligo nucleotide primer, as described in Chapter 3.

3.1 Introduction

When this project was started the majority of the sequence of Tn7 had been determined (Lichtenstein and Brenner 1982 ; Fling and Richards 1983 ; Fling, Kopf et al. 1985 ; Flores, Qadri et al. 1990 ; Sundstrom, Roy et al. 1991). Two un-sequenced portions remained, approximately 800 nt between the unique EcoRI site and the HindIII site at the 3' end of *tnsE* gene, and approximately 800nt at the left end, the sequence of the absolute left end had already been determined.The work presented in this chapter establishes the DNA sequence of these regions. This has enabled the reconstruction of the entire Tn7 sequence. This chapter presents an analysis of the sequence. This analysis revealed the presence of several directly repeated motifs in the centre of Tn7. The potential role of these repeats is also addressed.

3.2 Sequencing Strategy

3.2.1 EcoRI-HindIII Fragment

The sequence of this fragment was determined using a series of overlapping restriction fragments, see Fig. 3.1.
This fragment was subcloned from pUC18::Tn7-1 into EcoRI/HindIII digested pUC18. This plasmid, pSB200, was sequenced as described in Chapter 2 using both forward and reverse primers. This allowed identification of internal PvuII and AvaI restriction sites. The PvuII site was 216 nt from the EcoRI site, the AvaI site 158 nt from the HindIII site. The EcoRI/AvaI and PvuII/HindIII fragments were subcloned into EcoRI/AvaI and SmaI/HindIII digested pUC 18, to create pSB250 and pSB230 respectively. The PvuII/HindIII fragment in pSB230 was sequenced using reverse primer. pSB250 was sequenced using forward primer. This sequence revealed the presence of a further internal site, XbaI, this was 101 bp from the AvaI, 259 nt from the HindIII site. The EcoRI/XbaI and HindIII/XbaI fragments were then sub-cloned into appropriately digested pUC18 to generate pSB240 and pSB260 respectively. pSB240 was then sequenced using the forward primer, pSB260 was sequenced using the reverse primer. When compiled these various sequences were shown to overlap and thereby give continuous sequence from the EcoRI site to the HindIII site. This region was found to be 790 nt in length, in good

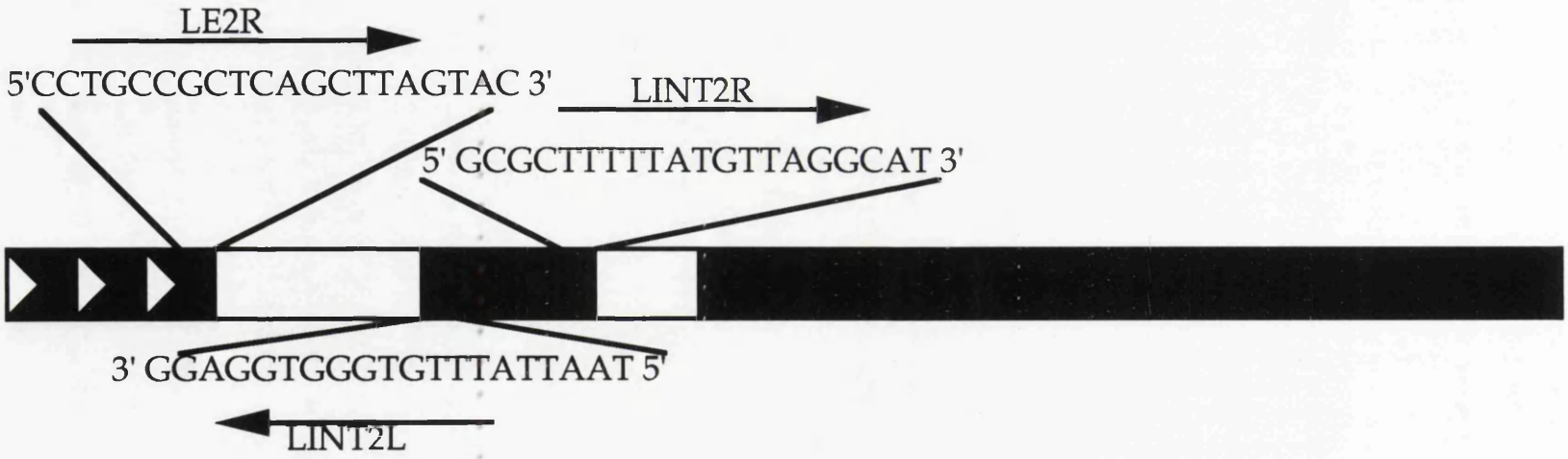


Fig. 3.2: Sequence and position of oligonucleotide primers
Regions of previously determined sequence are shown in black, un-sequenced regions are shown in white. Oligonucleotides were synthesised on an Applied Biosystems PCR mate and purified as detailed in Chapter 2.

agreement with the observed mobility of the EcoRI/HindIII fragment in a 2% agarose gel (data not shown).
From the previously published sequence of the Tn7 *aadA1* gene it was known that there was a further PvuII site 22 nt towards the left end of Tn7 from the EcoRI site (Fling, Kopf et al. 1985). The 238nt PvuII/PvuII fragment encompassing the EcoRI site was subcloned into SmaI digested pUC18. This was sequenced to confirm that the EcoRI site in Tn7 was, indeed, unique and not two sites very close together. The site was found to be unique.

3.2.2 Left end

From an examination of the restriction map of Tn7 it was apparent that fewer restriction sites had been mapped at the left end of Tn7. It was possible to subclone a 1.7kb HincII fragment, seeFig. 3.1. This fragment encompassed the right half of the unknown region and the integrase pseudogene and part of the *dhfr* open reading frame. This fragment was subcloned into Sma I digested pUC18, this made pSB210. The orientation of the fragment, confirmed by AvaI/EcoRI digestion and DNA sequence determined from the reverse primer. This sequence revealed the location of the internal AvaI site to be 79 nt to the right of the HincII site. The 1.6 kb AvaI-HindIII fragment of pSB210 was sub-cloned into AvaI-HindIII digested pUC18. The sequence of this plasmid was determined using reverse primer. No further useful restriction sites were uncovered from an analysis of this sequence, therefore the following strategy to determine the remaining sequence was adopted. Synthetic oligonucleotides were synthesised complementary to known DNA sequences at the left end of Tn7, and the left and right ends end of the sequence determined above. These oligonucleotides were designated primer LE2R, LINT2L and LINT2R respectively. Fig. 3.2 shows the sequence of these primers and the strands to which they are complementary. Primers LE2R and LINT2L were designed to hybridise to opposite strands of the denatured plasmid DNA and hopefully produce overlapping sequence. As shown in Fig. 3.1 the sequence derived from primer LINT2L completely overlapped that derived from LE2R.
A fourth primer LE2L was designed to hybridise to DNA towards the left end of Tn7 from the integrase pseudo gene. This region of un-published sequence was kindly given to me by Hilary-Kay Young of Dundee

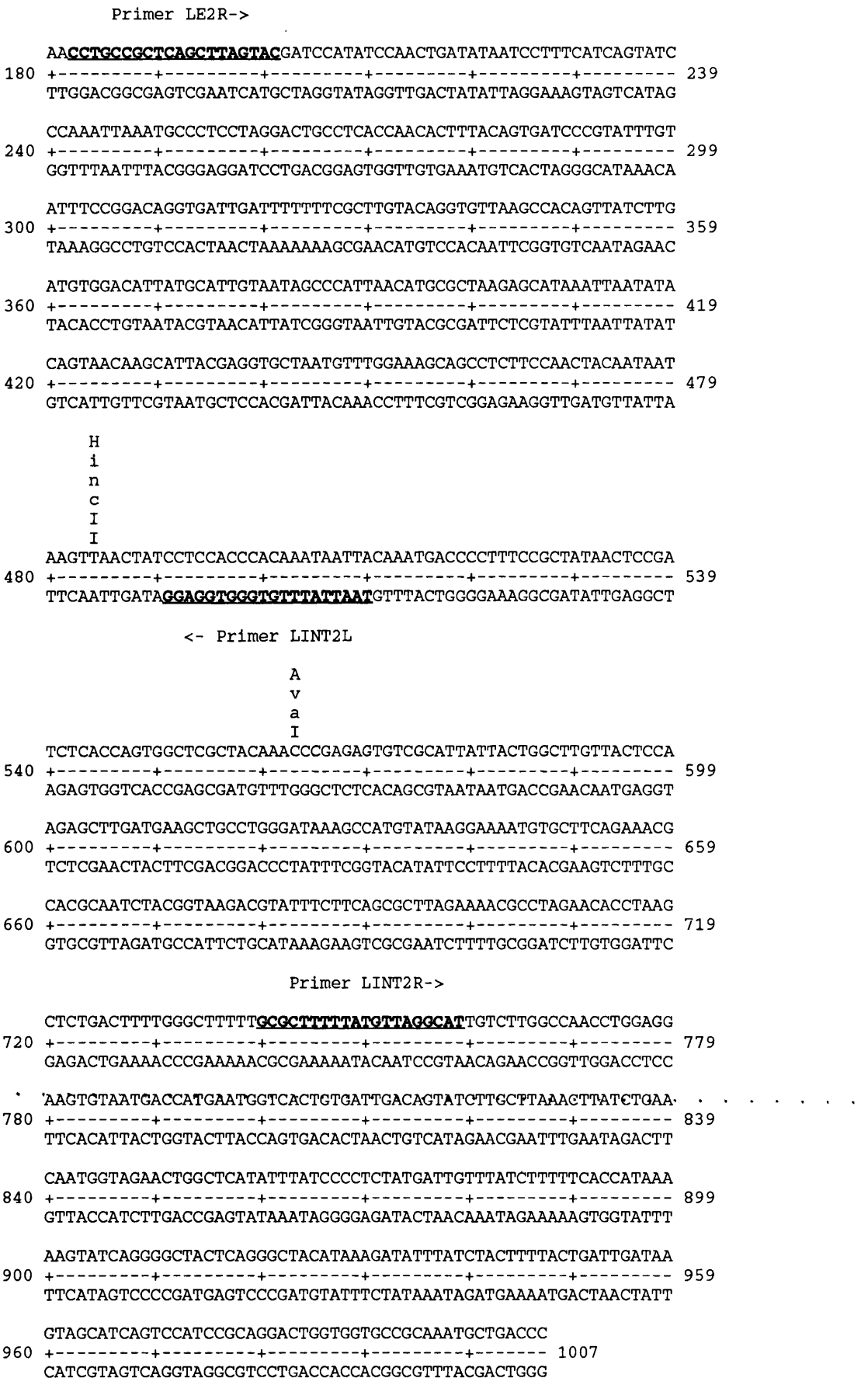


Fig. 3.3 Sequence derived from left end clones and synthetic primers, as described in the text. Sequences corresponding to the synthetic primers are indicated.

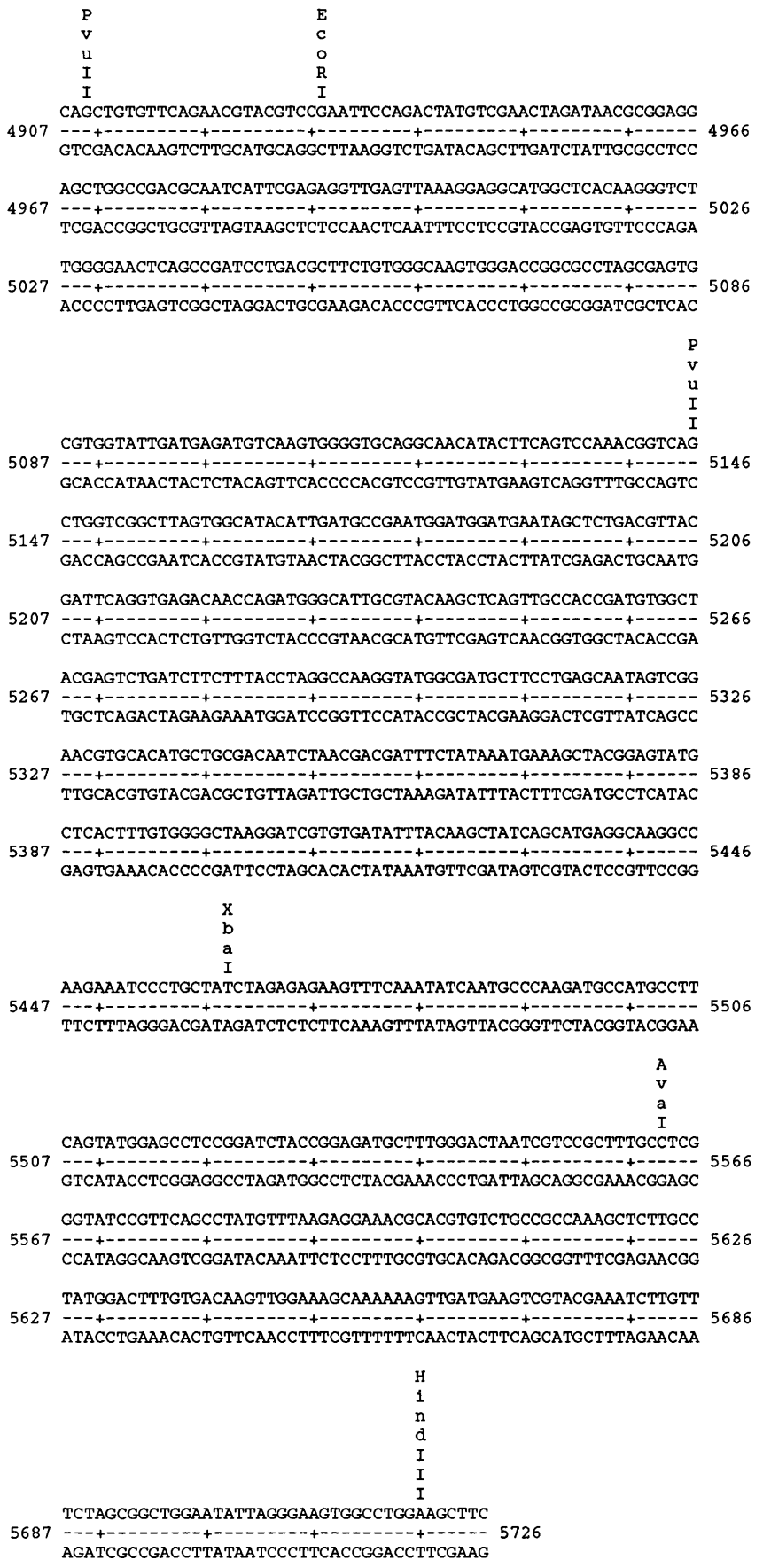


Fig. 3.4 Sequence I determined of the inner part of Tn7, see Fig. 3.1, restriction sites are highlighted

University. It was never possible to derive any sequence when this primer was used. An examination of the sequence derived from the LINT2R, which overlapped this region revealed several discrepancies between this sequence and that which I was given. This created several mismatches at the potential LE2L binding site which may have prevented primer hybridising to the denatured double stranded DNA. Regrettably there was insufficient time available for me to design, synthesise and use another primer to take the place of LE2L.

The sequences determined are shown in figures 3.3 and 3.4. These data enabled the compilation of the entire Tn7 sequence, this is listed in Fig. 3.5.

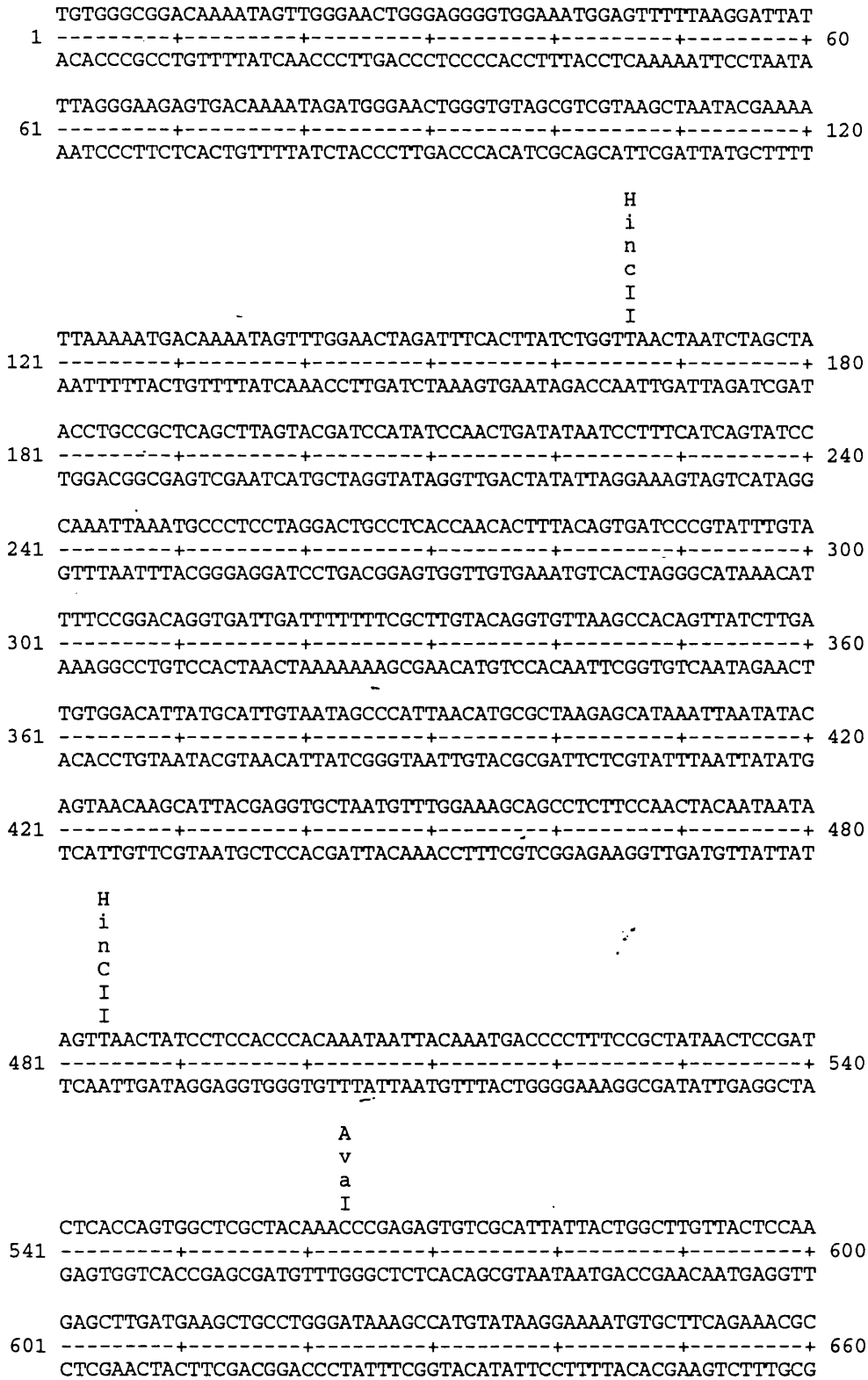


Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

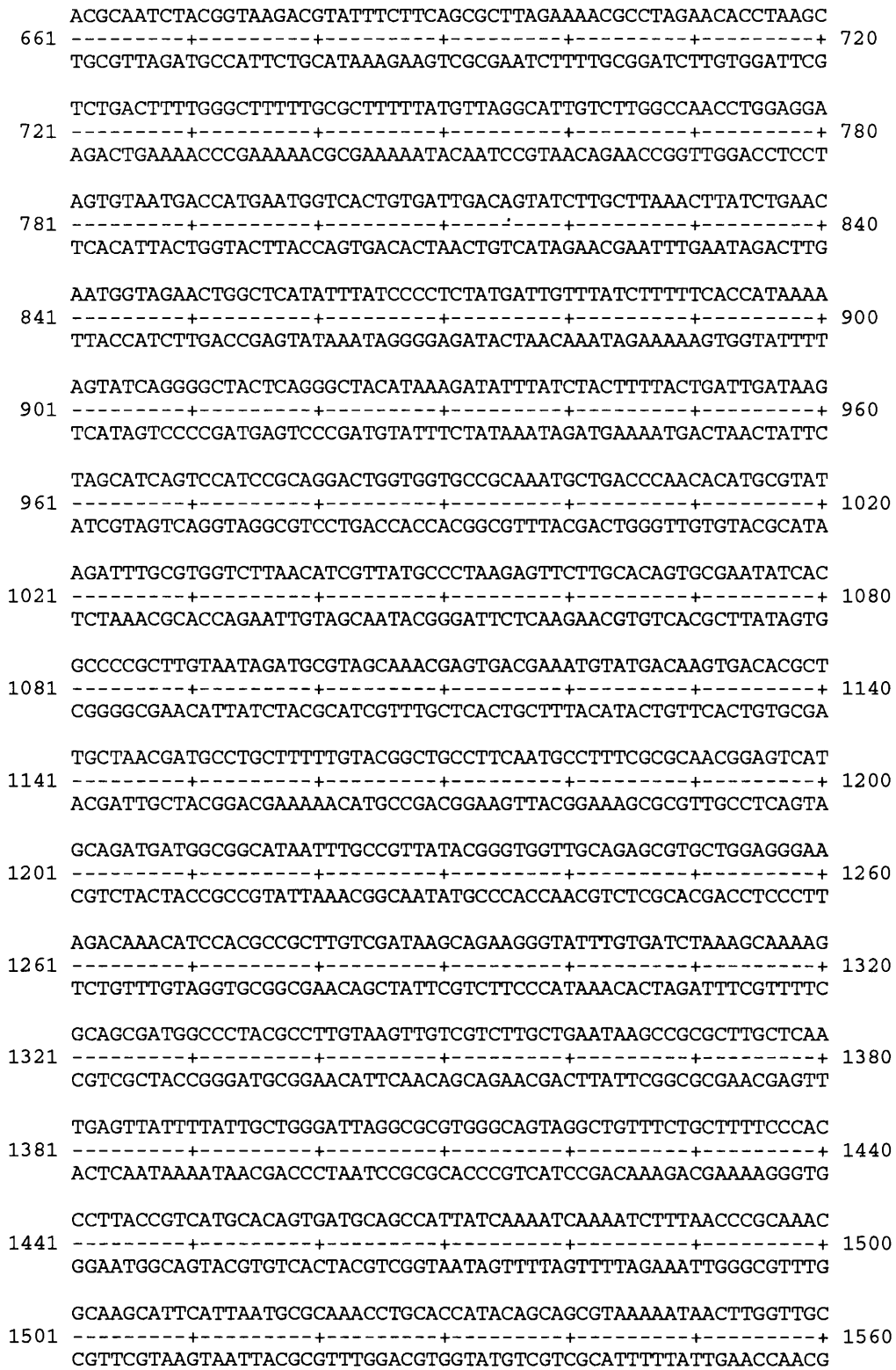


Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
     GAGTATCCATAACCTGCAAAATGCGTTGCACTTCATTTGCAGAGATAACAGAGGGTAGCC
1561 ---------+---------+---------+---------+---------+---------+ 1620
     CTCATAGGTATTGGACGTTTTACGCAACGTGAAGTAAACGTCTCTATTGTCTCCCATCGG

     GTCTAGGCTTGCTTGCAGGGATATAATCAATATCGCCCAACGGCTGTTGTAAAAACCTGT
1621 ---------+---------+---------+---------+---------+---------+ 1680
     CAGATCCGAACGAACGTCCCTATATTAGTTATAGCGGGTTGCCGACAACATTTTTGGACA

     TGTACAAAAAAGCTAGGGCATTTAAAGCGATTTTCTGCGTGTTTATGGCTACATGTCTGC
1681 ---------+---------+---------+---------+---------+---------+ 1740
     ACATGTTTTTTCGATCCCGTAAATTTCGCTAAAAGACGCACAAATACCGATGTACAGACG

     TGTTTGCTAAGCTGGATAAAAACAGCCTGACCTCTTCACTGCCCATGGTCTGAGGATGAC
1741 ---------+---------+---------+---------+---------+---------+ 1800
     ACAAACGATTCGACCTATTTTTGTCGGACTGGAGAAGTGACGGGTACCAGACTCCTACTG

     GTTTTTTGTGAAACAGAATAAAACGCTTAATCCAGTGCAGGTAAGTTTTTTCAGTTTTCA
1801 ---------+---------+---------+---------+---------+---------+ 1860
     CAAAAAACACTTTGTCTTATTTTGCGAATTAGGTCACGTCCATTCAAAAAAGTCAAAAGT

     GCGCATAACCTTTTTGTCGCATATCCGTGCGTATAGAATTTAAAAATGGACTGTTAGACA
1861 ---------+---------+---------+---------+---------+---------+ 1920
     CGCGTATTGGAAAAACAGCGTATAGGCACGCATATCTTAAATTTTTACCTGACAATCTGT

     TAAAACGCTCCTTGTCTTGCAACTGTCTGCCTATACAGCCTATTCTAGCTGGGATTTAAA
1921 ---------+---------+---------+---------+---------+---------+ 1980
     ATTTTGCGAGGAACAGAACGTTGACAGACGGATATGTCGGATAAGATCGACCCTAAATTT

     AAAGTGCCTGTTTTTTACGCCTAGAGATGCTTGTTTACCGGTAGAGTTTTAATTTAATGC
1981 ---------+---------+---------+---------+---------+---------+ 2040
     TTTCACGGACAAAAAATGCGGATCTCTACGAACAAATGGCCATCTCAAAATTAAATTACG

     TAAATAAATTAAAATGTTATGAGTTCTTTGGGTGAGATAATGTGCATCGTGCAAGCAGGA
2041 ---------+---------+---------+---------+---------+---------+ 2100
     ATTTATTTAATTTTACAATACTCAAGAAACCCACTCTATTACACGTAGCACGTTCGTCCT

     TAGACGGCATGCACGATTTGTAATAACAGAGTGTCTTGTATTTTTAAAGAAAGTCTATTT
2101 ---------+---------+---------+---------+---------+---------+ 2160
     ATCTGCCGTACGTGCTAAACATTATTGTCTCACAGAACATAAAAATTTCTTTCAGATAAA

     AATACAAGTGATTATATTAATTAACGGTAAGCATCAGCGGGTGACAAAACGAGCATGCTT
2161 ---------+---------+---------+---------+---------+---------+ 2220
     TTATGTTCACTAATATAATTAATTGCCATTCGTAGTCGCCCACTGTTTTGCTCGTACGAA

                H
                i
                n
                C
                I
                I
     ACTAATAAAATGTTAACCTCTGAGGAAGAATTGTGAAACTATCACTAATGGTAGCTATAT
2221 ---------+---------+---------+---------+---------+---------+ 2280
     TGATTATTTTACAATTGGAGACTCCTTCTTAACACTTTGATAGTGATTACCATCGATATA

     CGAAGAATGGAGTTATCGGGAATGGCCCTGATATTCCATGGAGTGCCAAAGGTGAACAGC
2281 ---------+---------+---------+---------+---------+---------+ 2340
     GCTTCTTACCTCAATAGCCCTTACCGGGACTATAAGGTACCTCACGGTTTCCACTTGTCG

     TCCTGTTTAAAGCTATTACCTATAACCAATGGCTGTTGGTTGGACGCAAGACTTTTGAAT
2341 ---------+---------+---------+---------+---------+---------+ 2400
     AGGACAAATTTCGATAATGGATATTGGTTACCGACAACCAACCTGCGTTCTGAAAACTTA
```

Fig. 3.5 Complete Sequence of Tn*7*

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

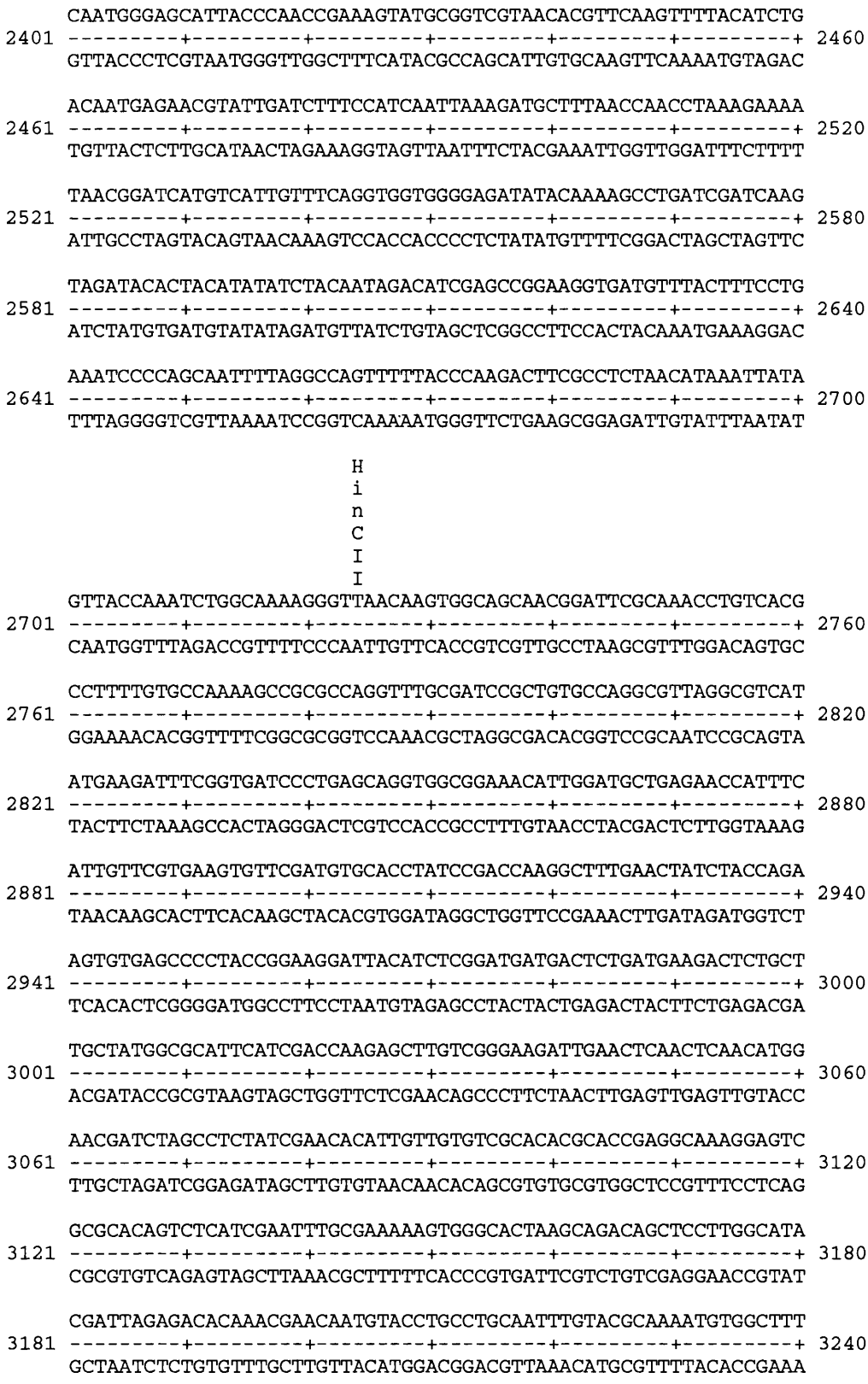


Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
     ACTCTCGGCGGCATTGACCTGTTCACGTATAAAACTAGACCTCAAGTCTCGAACGAAACA
3241 ---------+---------+---------+---------+---------+---------+ 3300
     TGAGAGCCGCCGTAACTGGACAAGTGCATATTTTGATCTGGAGTTCAGAGCTTGCTTTGT

                       A
                       v
                       a
                       I
     GCGATGTACTGGTACTGGTTCTCGGGAGCACAGGATGACGCCTAACAATTCATTCAAGCC
3301 ---------+---------+---------+---------+---------+---------+ 3360
     CGCTACATGACCATGACCAAGAGCCCTCGTGTCCTACTGCGGATTGTTAAGTAAGTTCGG

     GACACCGCTTCGCGGCGCGGCTTAATTCAGGAGTTAAACATCATGAGGGAAGCGGTGATC
3361 ---------+---------+---------+---------+---------+---------+ 3420
     CTGTGGCGAAGCGCCGCGCCGAATTAAGTCCTCAATTTGTAGTACTCCCTTCGCCACTAG

     GCCGAAGTATCGACTCAACTATCAGAGGTAGTTGGCGTCATCGAGCGCCATCTCGAACCG
3421 ---------+---------+---------+---------+---------+---------+ 3480
     CGGCTTCATAGCTGAGTTGATAGTCTCCATCAACCGCAGTAGCTCGCGGTAGAGCTTGGC

     ACGTTGCTGGCCGTACATTTGTACGGCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGT
3481 ---------+---------+---------+---------+---------+---------+ 3540
     TGCAACGACCGGCATGTAAACATGCCGAGGCGTCACCTACCGCCGGACTTCGGTGTGTCA

     GATATTGATTTGCTGGTTACGGTGACCGTAAGGCTTGATGAAACAACGCGGCGAGCTTTG
3541 ---------+---------+---------+---------+---------+---------+ 3600
     CTATAACTAAACGACCAATGCCACTGGCATTCCGAACTACTTTGTTGCGCCGCTCGAAAC

     ATCAACGACCTTTTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGTA
3601 ---------+---------+---------+---------+---------+---------+ 3660
     TAGTTGCTGGAAAACCTTTGAAGCCGAAGGGGACCTCTCTCGCTCTAAGAGGCGCGACAT

     GAAGTCACCATTGTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGCGCGAA
3661 ---------+---------+---------+---------+---------+---------+ 3720
     CTTCAGTGGTAACAACACGTGCTGCTGTAGTAAGGCACCGCAATAGGTCGATTCGCGCTT

     CTGCAATTTGGAGAATGGCAGCGCAATGACATTCTTGCAGGTATCTTCGAGCCAGCCACG
3721 ---------+---------+---------+---------+---------+---------+ 3780
     GACGTTAAACCTCTTACCGTCGCGTTACTGTAAGAACGTCCATAGAAGCTCGGTCGGTGC

     ATCGACATTGATCTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTTGCCTTGGTA
3781 ---------+---------+---------+---------+---------+---------+ 3840
     TAGCTGTAACTAGACCGATAGAACGACTGTTTTCGTTCTCTTGTATCGCAACGGAACCAT

     GGTCCAGCGGCGGAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAGGCGCTA
3841 ---------+---------+---------+---------+---------+---------+ 3900
     CCAGGTCGCCGCCTCCTTGAGAAACTAGGCCAAGGACTTGTCCTAGATAAACTCCGCGAT

     AATGAAACCTTAACGCTATGGAACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAATGTA
3901 ---------+---------+---------+---------+---------+---------+ 3960
     TTACTTTGGAATTGCGATACCTTGAGCGGCGGGCTGACCCGACCGCTACTCGCTTTACAT

     GTGCTTACGTTGTCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGAT
3961 ---------+---------+---------+---------+---------+---------+ 4020
     CACGAATGCAACAGGGCGTAAACCATGTCGCGTCATTGGCCGTTTTAGCGCGGCTTCCTA

     GTCGCTGCCGACTGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAA
4021 ---------+---------+---------+---------+---------+---------+ 4080
     CAGCGACGGCTGACCCGTTACCTCGCGGACGGCCGGGTCATAGTCGGGCAGTATGAACTT
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
     GCTAGGCAGGCTTATCTTGGACAAGAAGATCGCTTGGCCTCGCGCGCAGATCAGTTGGAA
4081 ---------+---------+---------+---------+---------+---------+ 4140
     CGATCCGTCCGAATAGAACCTGTTCTTCTAGCGAACCGGAGCGCGCGTCTAGTCAACCTT

     GAATTTGTTCACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAATGTCTAACA
4141 ---------+---------+---------+---------+---------+---------+ 4200
     CTTAAACAAGTGATGCACTTTCCGCTCTAGTGGTTCCATCAGCCGTTTATTACAGATTGT

     ATTCGTTCAAGCCGACGCCGCTTCGCGGCGCGGCTTAACTCAAGCGTTAGAGAGCTGGGG
4201 ---------+---------+---------+---------+---------+---------+ 4260
     TAAGCAAGTTCGGCTGCGGCGAAGCGCCGCGCCGAATTGAGTTCGCAATCTCTCGACCCC

     AAGACTATGCGCGATCTGTTGAAGGTGGTTCTAAGCCTCGTACTTGCGATGGCATCGGGG
4261 ---------+---------+---------+---------+---------+---------+ 4320
     TTCTGATACGCGCTAGACAACTTCCACCAAGATTCGGAGCATGAACGCTACCGTAGCCCC

     CAGGCACTTGCTGACCTGCCAATTGTTTTAGTGGATGAAGCTCGTCTTCCCTATGACTAC
4321 ---------+---------+---------+---------+---------+---------+ 4380
     GTCCGTGAACGACTGGACGGTTAACAAAATCACCTACTTCGAGCAGAAGGGATACTGATG

     TCCCCATCCAACTACGACATTTCTCCAAGCAACTACGACAACTCCATAAGCAATTACGAC
4381 ---------+---------+---------+---------+---------+---------+ 4440
     AGGGGTAGGTTGATGCTGTAAAGAGGTTCGTTGATGCTGTTGAGGTATTCGTTAATGCTG

     AATAGTCCATCAAATTACGACAACTCTGAGAGCAACTACGATAATAGTTCATCCAATTAC
4441 ---------+---------+---------+---------+---------+---------+ 4500
     TTATCAGGTAGTTTAATGCTGTTGAGACTCTCGTTGATGCTATTATCAAGTAGGTTAATG

     GACAATAGTCGCAACGGAAATCGTAGGCTTATATATAGCGCAAATGGGTCTCGCACTTTC
4501 ---------+---------+---------+---------+---------+---------+ 4560
     CTGTTATCAGCGTTGCCTTTAGCATCCGAATATATATCGCGTTTACCCAGAGCGTGAAAG

     GCCGGCTACTACGTCATTGCCAACAATGGGACAACGAACTTCTTTTCCACATCTGGCAAA
4561 ---------+---------+---------+---------+---------+---------+ 4620
     CGGCCGATGATGCAGTAACGGTTGTTACCCTGTTGCTTGAAGAAAAGGTGTAGACCGTTT

     AGGATGTTCTACACCCCAAAAGGGGGGCGCGGCGTCTATGGCGGCAAAGATGGGAGCTTC
4621 ---------+---------+---------+---------+---------+---------+ 4680
     TCCTACAAGATGTGGGGTTTTCCCCCCGCGCCGCAGATACCGCCGTTTCTACCCTCGAAG

     TGCGGGGCATTGGTCGTCATAAATGGCCAATTTTCGCTTGCCCTGACAGATAACGGCCTG
4681 ---------+---------+---------+---------+---------+---------+ 4740
     ACGCCCCGTAACCAGCAGTATTTACCGGTTAAAAGCGAACGGGACTGTCTATTGCCGGAC

     AAGATCATGTATCTAAGCAACTAGCCTGCTCTCTAATAAAATGTTAGGCCTCAACATCTA
4741 ---------+---------+---------+---------+---------+---------+ 4800
     TTCTAGTACATAGATTCGTTGATCGGACGAGAGATTATTTTACAATCCGGAGTTGTAGAT

     GTCGCAAGCTGAGGGGAACCACTAGTGTCATACGAACCTCCAAGAGACGGTTACACAAAC
4801 ---------+---------+---------+---------+---------+---------+ 4860
     CAGCGTTCGACTCCCCTTGGTGATCACAGTATGCTTGGAGGTTCTCTGCCAATGTGTTTG

     GGGTACATTGTTGATGTCATGTATGACAATCGCCCAAGTAAGTATCCAGCTGTGTTCAGA
4861 ---------+---------+---------+---------+---------+---------+ 4920
     CCCATGTAACAACTACAGTACATACTGTTAGCGGGTTCATTCATAGGTCGACACAAGTCT
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
*Tns*C nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

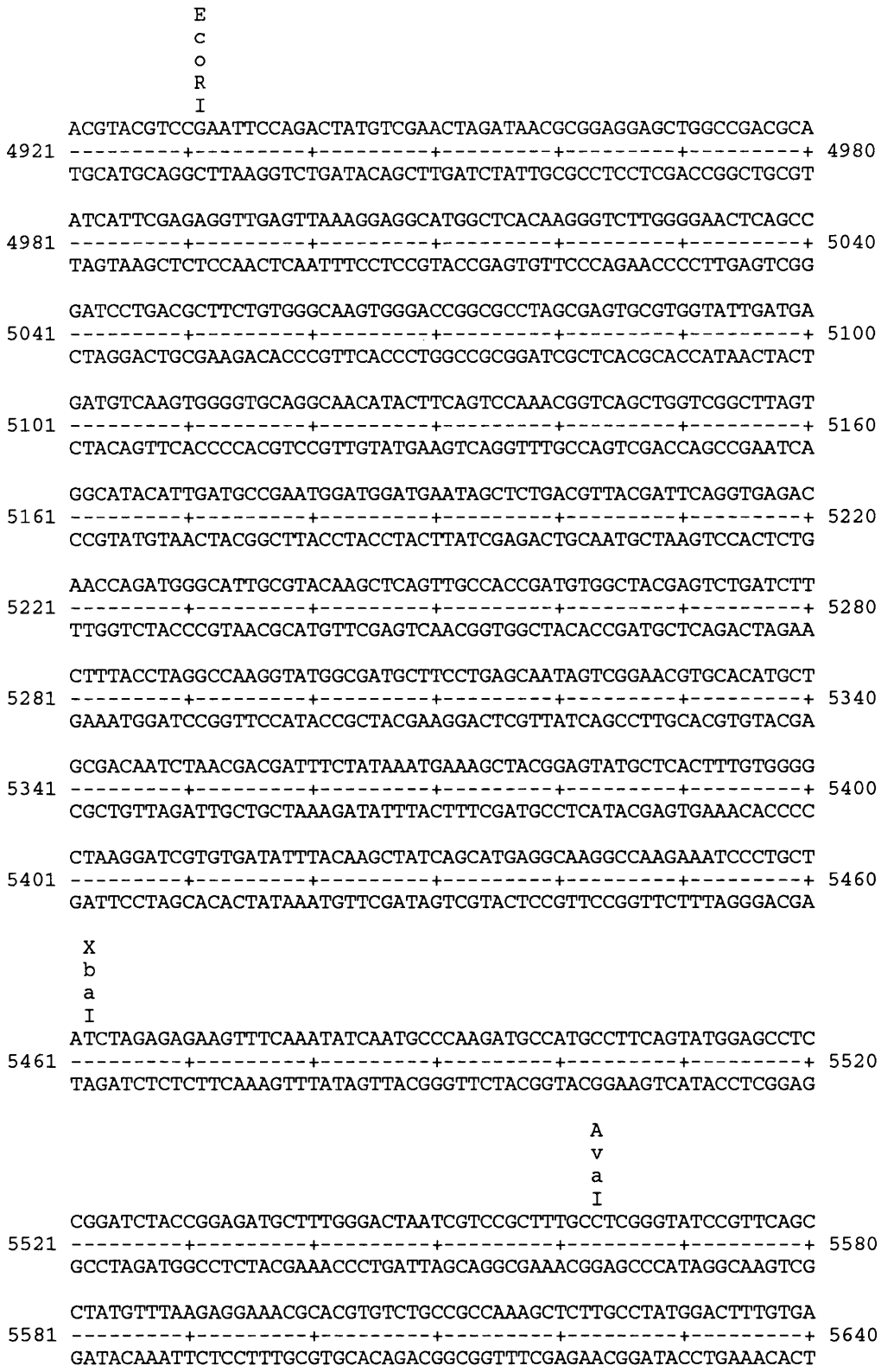


Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
     CAAGTTGGAAAGCAAAAAAGTTGATGAAGTCGTACGAAATCTTGTTTCTAGCGGCTGGAA
5641 ---------+---------+---------+---------+---------+---------+ 5700
     GTTCAACCTTTCGTTTTTTCAACTACTTCAGCATGCTTTAGAACAAAGATCGCCGACCTT

                       H
                       i
                       n
                       D
                       I
                       I
                       I
     TATTAGGGAAGTGGCCTGGAAGCTTCGTCCTTTGATCATGAAATGCCCTACTTTGACACG
5701 ---------+---------+---------+---------+---------+---------+ 5760
     ATAATCCCTTCACCGGACCTTCGAAGCAGGAAACTAGTACTTTACGGGATGAAACTGTGC


     GGGCGGGACTTCGATATTGGATTTCCTGACACGCTGTCATCCAAGATTGAGTTTTTGACC
5761 ---------+---------+---------+---------+---------+---------+ 5820
     CCCGCCCTGAAGCTATAACCTAAAGGACTGTGCGACAGTAGGTTCTAACTCAAAAACTGG

     GCGAATGGATCAATCGTCCAAGCACGTGCTCGCCTGCACTTCAAGCAGGCTCTAATTTTT
5821 ---------+---------+---------+---------+---------+---------+ 5880
     CGCTTACCTAGTTAGCAGGTTCGTGCACGAGCGGACGTGAAGTTCGTCCGAGATTAAAAA

     GCCAAGGCGTCCAAACATGTCGCCGCGCTCAAGAACTCATTGGATTACTACTATGGATCG
5881 ---------+---------+---------+---------+---------+---------+ 5940
     CGGTTCCGCAGGTTTGTACAGCGGCGCGAGTTCTTGAGTAACCTAATGATGATACCTAGC

     TCTGTAATCCCCCCTCAAAACCACGGAGGCTATTTGTAGAAAATTCCATTAATATGAACT
5941 ---------+---------+---------+---------+---------+---------+ 6000
     AGACATTAGGGGGGAGTTTTGGTGCCTCCGATAAACATCTTTTAAGGTAATTATACTTGA

     CAAGGTGTTCCCTACGGACGAATCAGCGATGAAGGTGGTGTACCTGGCTATCCAGGCGGC
6001 ---------+---------+---------+---------+---------+---------+ 6060
     GTTCCACAAGGGATGCCTGCTTAGTCGCTACTTCCACCACATGGACCGATAGGTCCGCCG

     GACCAAGAAATGGACGATGCCGATCCGCAACTGGAAGCCGGCCATGAACCGTTTTATGAT
6061 ---------+---------+---------+---------+---------+---------+ 6120
     CTGGTTCTTTACCTGCTACGGCTAGGCGTTGACCTTCGGCCGGTACTTGGCAAAATACTA

     CGAATTTGGTGACCGTTTAAATGGCCACCTTTAACTGACCGGCAAAAGCACTTACACAAA
6121 ---------+---------+---------+---------+---------+---------+ 6180
     GCTTAAACCACTGGCAAATTTACCGGTGGAAATTGACTGGCCGTTTTCGTGAATGTGTTT

     AGGGTTTACAGGCTCTTTCAGTCATTCCTTTAATGCGTAAATTGCTCTCTAACCACTCTC
6181 ---------+---------+---------+---------+---------+---------+ 6240
     TCCCAAATGTCCGAGAAAGTCAGTAAGGAAATTACGCATTTAACGAGAGATTGGTGAGAG

     TCAGCCCACCCTTCCATATCCTCTCTCGATACTTGCAGCTCCCCCAACCCCTTTGGATGA
6241 ---------+---------+---------+---------+---------+---------+ 6300
     AGTCGGGTGGGAAGGTATAGGAGAGAGCTATGAACGTCGAGGGGGTTGGGGAAACCTACT

     TTCACCCCTCTATGCCCGTCTTGTCCATATAATTGATCAAACAAACTATTTGGCCAATTC
6301 ---------+---------+---------+---------+---------+---------+ 6360
     AAGTGGGGAGATACGGGCAGAACAGGTATATTAACTAGTTTGTTTGATAAACCGGTTAAG

     AATGAGCTCTTCACCACTCCACGCCGTATCTTTTCAAAATCATTCCGCCATGTTTCGCTA
6361 ---------+---------+---------+---------+---------+---------+ 6420
     TTACTCGAGAAGTGGTGAGGTGCGGCATAGAAAAGTTTTAGTAAGGCGGTACAAAGCGAT
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
     TCAACGCCACTCAACACTTTGGTAGAAAGCATTTTCACCCCATCCGATGCATCCACCTCT
6421 ---------+---------+---------+---------+---------+---------+ 6480
     AGTTGCGGTGAGTTGTGAAACCATCTTTCGTAAAAGTGGGGTAGGCTACGTAGGTGGAGA

     AGCAATACAAATTCACTGCCATTACGTTGCACCCCAACGGCTTTAATCACTCTTGGTGAG
6481 ---------+---------+---------+---------+---------+---------+ 6540
     TCGTTATGTTTAAGTGACGGTAATGCAACGTGGGGTTGCCGAAATTAGTGAGAACCACTC

     CCATCTTTACACAGATGTAATCGGCTACGCCCAACTTTTGGCAAAACCAAGGTTTCTTCA
6541 ---------+---------+---------+---------+---------+---------+ 6600
     GGTAGAAATGTGTCTACATTAGCCGATGCGGGTTGAAAACCGTTTTGGTTCCAAAGAAGT

     AAAAGCACCCGACATGCAAATTTAGTTTTTAGAATTGAAAGTAGCTCATCAAAAGCAGCA
6601 ---------+---------+---------+---------+---------+---------+ 6660
     TTTTCGTGGGCTGTACGTTTAAATCAAAAATCTTAACTTTCATCGAGTAGTTTTCGTCGT

     AATCGATTAGCAAAAATAGAGTTGTAATTGGTTGCATCCTGCTTCCCACCCACATCTGCC
6661 ---------+---------+---------+---------+---------+---------+ 6720
     TTAGCTAATCGTTTTTATCTCAACATTAACCAACGTAGGACGAAGGGTGGGTGTAGACGG

     GCTGCTAGGACACCACCTAAGTGTGGCTCATCAGTACTAACCAAATTACTATTTTCACTT
6721 ---------+---------+---------+---------+---------+---------+ 6780
     CGACGATCCTGTGGTGGATTCACACCGAGTAGTCATGATTGGTTTAATGATAAAAGTGAA

     GTTGTTGCTTCTTCTTTTTCTAAAATAGTTTGGCTACTTTTTCTTGCCCTGCGAGATTTT
6781 ---------+---------+---------+---------+---------+---------+ 6840
     CAACAACGAAGAAGAAAAAGATTTTATCAAACCGATGAAAAAGAACGGGACGCTCTAAAA

     TCAATTCGACTAGGTCTACTAAAACTTATCCATGTCGGCTCAGCCTCTATGACTCGTGTG
6841 ---------+---------+---------+---------+---------+---------+ 6900
     AGTTAAGCTGATCCAGATGATTTTGAATAGGTACAGCCGAGTCGGAGATACTGAGCACAC


     TCTGTGTCTATATTGGCTGTTTCCTCATCGTCCAACTGTAGATGCTCATCATCAACAACT
6901 ---------+---------+---------+---------+---------+---------+ 6960
     AGACACAGATATAACCGACAAAGGAGTAGCAGGTTGACATCTACGAGTAGTAGTTGTTGA

     GACTCTGTTGAAACCGCTATGTGCTGCGTACTACCATCACCTGCCTCCTTTTCCTGAAAA
6961 ---------+---------+---------+---------+---------+---------+ 7020
     CTGAGACAACTTTGGCGATACACGACGCATGATGGTAGTGGACGGAGGAAAAGGACTTTT

     GAGGCATGGCTAATAGCTGTTGTGCTAGGAAGCATAGCGTTGATTTCTAAACCTACTATT
7021 ---------+---------+---------+---------+---------+---------+ 7080
     CTCCGTACCGATTATCGACAACACGATCCTTCGTATCGCAACTAAAGATTTGGATGATAA

     TCCTCAACTAAATAATCCTTATCCTCGTTAGAAGAACGTCCTTTTACATGTAATTTCCAA
7081 ---------+---------+---------+---------+---------+---------+ 7140
     AGGAGTTGATTTATTAGGAATAGGAGCAATCTTCTTGCAGGAAAATGTACATTAAAGGTT

     CCCTGCATGGGCGGAGGGTCAAAGCTAAAGCACCAGCTTTCAACGCCGTTCTTAATTTCT
7141 ---------+---------+---------+---------+---------+---------+ 7200
     GGGACGTACCCGCCTCCCAGTTTCGATTTCGTGGTCGAAAGTTGCGGCAAGAATTAAAGA

     CTATTTTGTTGATAGTGCCTAAAAATACTTTCATACGAATCCATAACATCTTGATCCGAA
7201 ---------+---------+---------+---------+---------+---------+ 7260
     GATAAAACAACTATCACGGATTTTTATGAAAGTATGCTTAGGTATTGTAGAACTAGGCTT

     AACAACCAAACCAAAAGCTGCACTACGGCCGACTGCTCTAACGCCCCTTTAGGAAACGAT
7261 ---------+---------+---------+---------+---------+---------+ 7320
     TTGTTGGTTTGGTTTTCGACGTGATGCCGGCTGACGAGATTGCGGGGAAATCCTTTGCTA
```

Fig. 3.5 Complete Sequence of Tn*7*

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

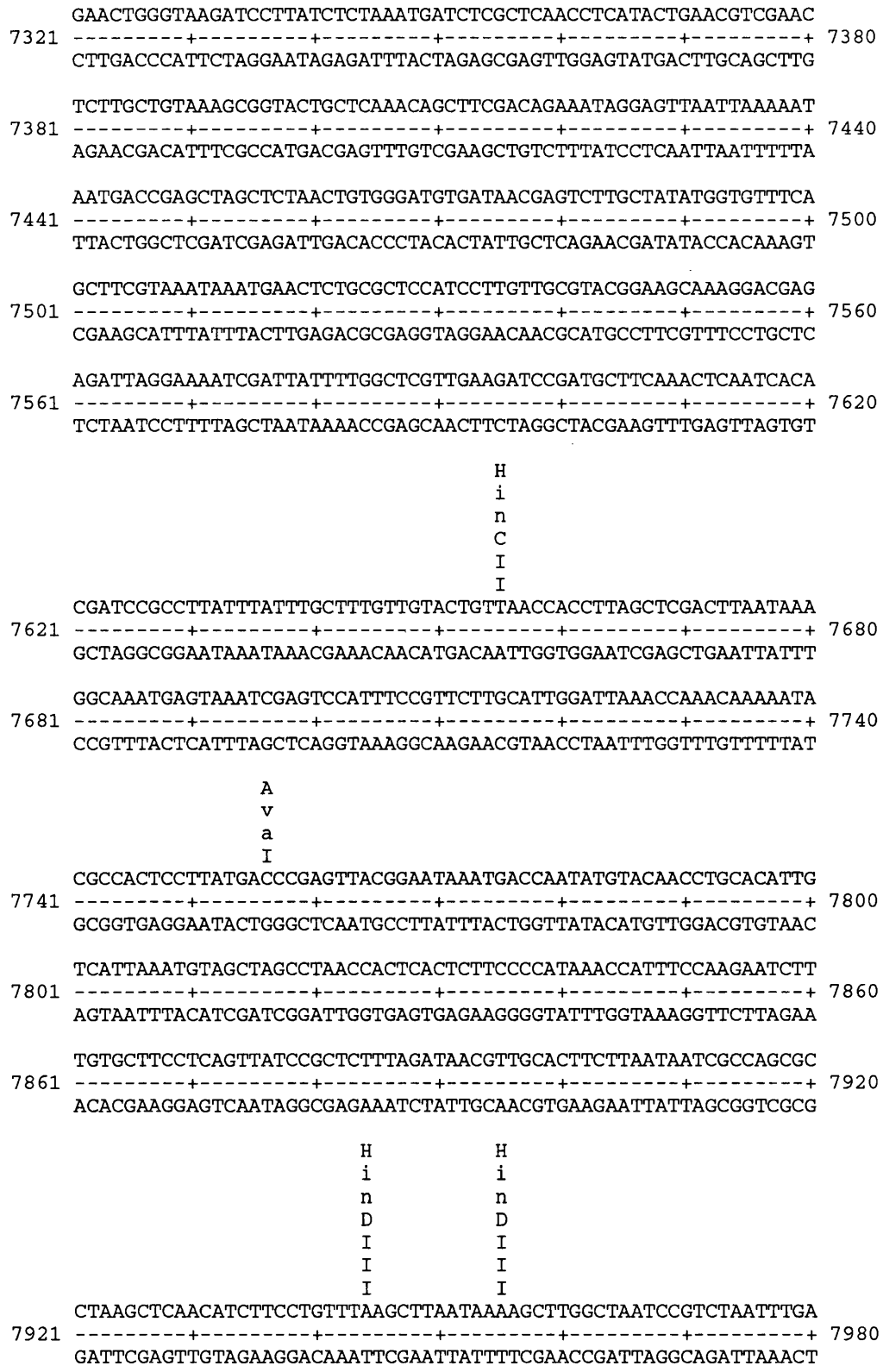


Fig. 3.5 Complete Sequence of Tn*7*

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
     TAATCTTCCACACTCTCTGAGTAACGCTTTAAGCAAAGCGCTACCAAAGGCAGTTTCTGT
7981 ---------+---------+---------+---------+---------+---------+ 8040
     ATTAGAAGGTGTGAGAGACTCATTGCGAAATTCGTTTCGCGATGGTTTCCGTCAAAGACA

     AGATTTTTTGCAAGAGAGGTTCCGTTAGGAGTTTGCTTTAACAGCCAGCTCGATGTCGCT
8041 ---------+---------+---------+---------+---------+---------+ 8100
     TCTAAAAAACGTTCTCTCCAAGGCAATCCTCAAACGAAATTGTCGGTCGAGCTACAGCGA

                          X
                          b
                          a
                          I
     CTTGGGTGATCAAGGCTACTATCTAGACGCTTTATGATTCTTAATAGTTGTCGTACAGCA
8101 ---------+---------+---------+---------+---------+---------+ 8160
     GAACCCACTAGTTCCGATGATAGATCTGCGAAATACTAAGAATTATCAACAGCATGTCGT

     ATTCGATCTCTTTGGTTCCAATCAACTCTAGGGGCGGGTGCCAGACGCTCTTGTTGATGC
8161 ---------+---------+---------+---------+---------+---------+ 8220
     TAAGCTAGAGAAACCAAGGTTAGTTGAGATCCCCGCCCACGGTCTGCGAGAACAACTACG

     TGTTGATTCCAGTGAACTAGCCAATCCCTGTCATGTCGGTAAAGCCAAGCGTATAGCACC
8221 ---------+---------+---------+---------+---------+---------+ 8280
     ACAACTAAGGTCACTTGATCGGTTAGGGACAGTACAGCCATTTCGGTTCGCATATCGTGG

     CCACCCTCTAAAGACTGTCTTGCCGCCTTAATTCCTTGGTATTTATGCACTAGTTGCTGC
8281 ---------+---------+---------+---------+---------+---------+ 8340
     GGTGGGAGATTTCTGACAGAACGGCGGAATTAAGGAACCATAAATACGTGATCAACGACG

     CAGTCTTTATGCTTAACAGATAAATCTTCAGAGTTAGGTTGCACAGACTGGCTAACAGGT
8341 ---------+---------+---------+---------+---------+---------+ 8400
     GTCAGAAATACGAATTGTCTATTTAGAAGTCTCAATCCAACGTGTCTGACCGATTGTCCA

     CTTGTCGTTATAGAGTGCTCAGTAAGAGCACTTGCCTGCTGTAGCGCTTCTATAACCGTT
8401 ---------+---------+---------+---------+---------+---------+ 8460
     GAACAGCAATATCTCACGAGTCATTCTCGTGAACGGACGACATCGCGAAGATATTGGCAA

     AGTTTTGGCAATAAGGCTTGCCACACAATACTATGCTGTAAATAACTAAAGGCTTTTCTA
8461 ---------+---------+---------+---------+---------+---------+ 8520
     TCAAAACCGTTATTCCGAACGGTGTGTTATGATACGACATTTATTGATTTCCGAAAAGAT

     TGCTTACGGAATATACTTTTCAGCCAACACGTGTCCTTGTTCTCTGCCAACTTTAAATCC
8521 ---------+---------+---------+---------+---------+---------+ 8580
     ACGAATGCCTTATATGAAAAGTCGGTTGTGCACAGGAACAAGAGACGGTTGAAATTTAGG

     AGTTTCTCTAGTGCCTCATCACTAAAAGTCTGCCTCACTCTCTCCGCCACCAAGTCATGA
8581 ---------+---------+---------+---------+---------+---------+ 8640
     TCAAAGAGATCACGGAGTAGTGATTTTCAGACGGAGTGAGAGAGGCGGTGGTTCAGTACT

     CGAATGTGCTTGCTTTTGGTTAGCCCTAGATCCTGCGCTAAGCGCTGATAAAACAGCGTC
8641 ---------+---------+---------+---------+---------+---------+ 8700
     GCTTACACGAACGAAAACCAATCGGGATCTAGGACGCGATTCGCGACTATTTTGTCGCAG

     CACTGCTCAAGGCTTGGGGAAAGCTCTTGCGCTCGTGGAGCATCTAACAGAGGGGCTATA
8701 ---------+---------+---------+---------+---------+---------+ 8760
     GTGACGAGTTCCGAACCCCTTTCGAGAACGCGAGCACCTCGTAGATTGTCTCCCCGATAT

     TAAGCAGCTAGTGCTGTTAATTGAGATAGGGAGTCTTTGGGGTAGTCTGAAAGCAGCTCA
8761 ---------+---------+---------+---------+---------+---------+ 8820
     ATTCGTCGATCACGACAATTAACTCTATCCCTCAGAAACCCCATCAGACTTTCGTCGAGT
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
     GTATGACCCAAAGCCCAAAATTGATGTCGGTGATCATCTACAGCTCTATCAAAGAAGACT
8821 ---------+---------+---------+---------+---------+---------+ 8880
     CATACTGGGTTTCGGGTTTTAACTACAGCCACTAGTAGATGTCGAGATAGTTTCTTCTGA

     AAAGCACCGTGTTTTGGACAATATGGCAAAGCGGGCAAATACCAATCTCGTTGCCAAAAG
8881 ---------+---------+---------+---------+---------+---------+ 8940
     TTTCGTGGCACAAAACCTGTTATACCGTTTCGCCCGTTTATGGTTAGAGCAACGGTTTTC

     GCTTCCCCATACCTATTTAGCTGAAGAGCAACGCAATCAGGGCAGTATCTAAAGCGGTTA
8941 ---------+---------+---------+---------+---------+---------+ 9000
     CGAAGGGGTATGGATAAATCGACTTCTCGTTGCGTTAGTCCCGTCATAGATTTCGCCAAT

              X
              b
              a
              I
     TCGCTCTTAACTCTAGAAGCAGCGACTCCTAGCATTAAATGCACCGCACCTTGCGCTTGG
9001 ---------+---------+---------+---------+---------+---------+ 9060
     AGCGAGAATTGAGATCTTCGTCGCTGAGGATCGTAATTTACGTGGCGTGGAACGCGAACC

     TACTCCATTAACCGAATAGCTTCGTCTCGGCGCTCCTTGCCTACAAACGGAGCATATAAA
9061 ---------+---------+---------+---------+---------+---------+ 9120
     ATGAGGTAATTGGCTTATCGAAGCAGAGCCGCGAGGAACGGATGTTTGCCTCGTATATTT

     GGGAATAAGGTATGCTCATAAATAAGCTGCTGAACAGCGTAACGTCCTGTTTGATGTAGA
9121 ---------+---------+---------+---------+---------+---------+ 9180
     CCCTTATTCCATACGAGTATTTATTCGACGACTTGTCGCATTGCAGGACAAACTACATCT

     TGTCTTGCTATCACACCTAAATGCGAGGGCAGACCTAAGGTAGCGACCACCTTGCGGTTG
9181 ---------+---------+---------+---------+---------+---------+ 9240
     ACAGAACGATAGTGTGGATTTACGCTCCCGTCTGGATTCCATCGCTGGTGGAACGCCAAC

     CCATACACCTCATCCAACAGCTGCTTAGGACTAACAATCCCTTGATAAACGCCTGCCCGT
9241 ---------+---------+---------+---------+---------+---------+ 9300
     GGTATGTGGAGTAGGTTGTCGACGAATCCTGATTGTTAGGGAACTATTTGCGGACGGGCA

     GCAATAGTGCTATAAATCAGCTCATTCGAGTACGGAACAGGAAAGTTTCTCATGGCTAAC
9301 ---------+---------+---------+---------+---------+---------+ 9360
     CGTTATCACGATATTTAGTCGAGTAAGCTCATGCCTTGTCCTTTCAAAGAGTACCGATTG

     CTGCTTGTTTAAATAAGCTCGCCATATCCACTATTACCCCTTTCCCTTTTAACCGTTCAT
9361 ---------+---------+---------+---------+---------+---------+ 9420
     GACGAACAAATTTATTCGAGCGGTATAGGTGATAATGGGGAAAGGGAAAATTGGCAAGTA

     GCATGGTTTTTTCAGGTTGGCGTTGTGAATAGATATAACGTAAATCCGTATCAGGCAAGC
9421 ---------+---------+---------+---------+---------+---------+ 9480
     CGTACCAAAAAAGTCCAACCGCAACACTTATCTATATTGCATTTAGGCATAGTCCGTTCG

     TATCCCAGTCGCTGGGCTTGACTACCTTTATAGCCGAAACCTTTTTACTCTTGGAGGGCT
9481 ---------+---------+---------+---------+---------+---------+ 9540
     ATAGGGTCAGCGACCCGAACTGATGGAAATATCGGCTTTGGAAAAATGAGAACCTCCCGA

     TTTCTAGTTCTGACACTACCGTTTCGCCTTCCATCAACCACTGCAAAACAAGAGGCAGTA
9541 ---------+---------+---------+---------+---------+---------+ 9600
     AAAGATCAAGACTGTGATGGCAAAGCGGAAGGTAGTTGGTGACGTTTTGTTCTCCGTCAT

     ACTTTTGTCTTGTCATCGTTGGATTCTGGCTAAACGCTTTTTTAATAGTGGGAATTAACA
9601 ---------+---------+---------+---------+---------+---------+ 9660
     TGAAAACAGAACAGTAGCAACCTAAGACCGATTTGCGAAAAAATTATCACCCTTAATTGT
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

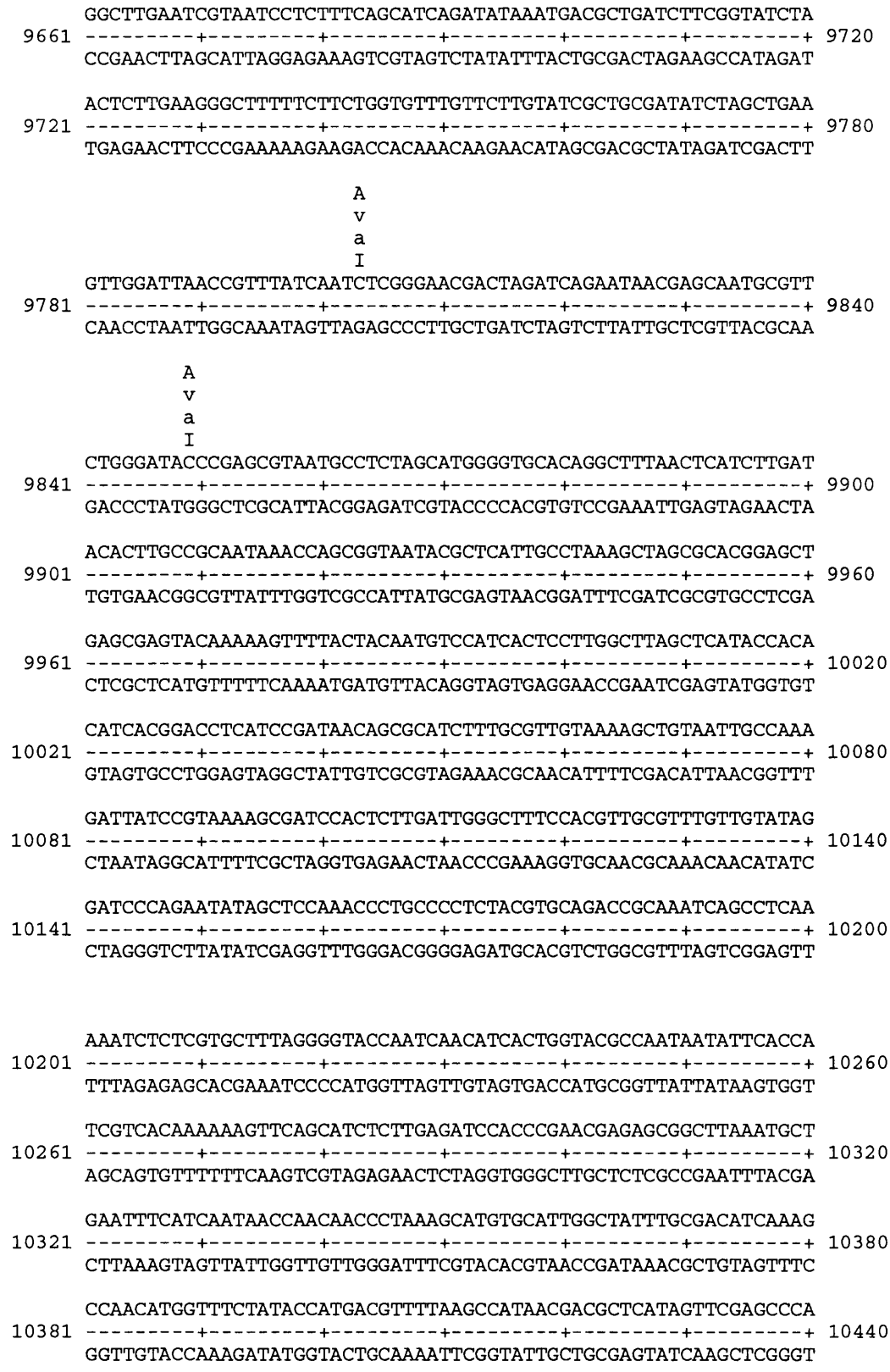


Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
      AGGCTCGATCCAACGCTCTGAAAAAATTCAAGCAGATTTCTTTTAGCGAACCATTATGCG
10441 ---------+---------+---------+---------+---------+---------+ 10500
      TCCGAGCTAGGTTGCGAGACTTTTTTAAGTTCGTCTAAAGAAAATCGCTTGGTAATACGC

      AGCAGTCTATTTTCAAATACACCACCTGCTCTACATTGAGTTCACGATGGTAAATCACCT
10501 ---------+---------+---------+---------+---------+---------+ 10560
      TCGTCAGATAAAAGTTTATGTGGTGGACGAGATGTAACTCAAGTGCTACCATTTAGTGGA

      GAGGATACGTGGCTAGAATACGATGAAGAGAGGTCGTCTTCCCACTACCAGAACAACCAA
10561 ---------+---------+---------+---------+---------+---------+ 10620
      CTCCTATGCACCGATCTTATGCTACTTCTCTCCAGCAGAAGGGTGATGGTCTTGTTGGTT

              H
              i
              n
              D
              I
              I
              I
      TTAACAATAAGCTTTGTGCCGTAGATCGTGCCTCCTCAAAGCGAAATGTCTCCAACTCTC
10621 ---------+---------+---------+---------+---------+---------+ 10680
      AATTGTTATTCGAAACACGGCATCTAGCACGGAGGAGTTTCGCTTTACAGAGGTTGAGAG

      CCGTTTGAACACGCTCATAACCATTTTGTAAATGCTTTTGTAAATCTCCTGTTTTAGGAT
10681 ---------+---------+---------+---------+---------+---------+ 10740
      GGCAAACTTGTGCGAGTATTGGTAAAACATTTACGAAAACATTTAGAGGACAAAATCCTA

      TTCTGCCTACGTAGCCACCTCGAATCATGACCGAAATACGCTCACTTAGTAGCAAATGCG
10741 ---------+---------+---------+---------+---------+---------+ 10800
      AAGACGGATGCATCGGTGGAGCTTAGTACTGGCTTTATGCGAGTGAATCATCGTTTACGC

      TACCTAATGGCTGAAAATAGTCATCTGGAATACGACAAATGGTATGAGCTCTGATAACAC
10801 ---------+---------+---------+---------+---------+---------+ 10860
      ATGGATTACCGACTTTTATCAGTAGACCTTATGCTGTTTACCATACTCGAGACTATTGTG

      GGGACTTTTGCAAGTCAGAGGAAGTAAGCTGTAAAGAGGATTTCAGTGATGCAGCACTAT
10861 ---------+---------+---------+---------+---------+---------+ 10920
      CCCTGAAAACGTTCAGTCTCCTTCATTCGACATTTCTCCTAAAGTCACTACGTCGTGATA

      TCACTGACTCTTGTAATGGTGGTAAGGCCTCGATAAAAGGATTATCACGATAAGCCTCTA
10921 ---------+---------+---------+---------+---------+---------+ 10980
      AGTGACTGAGAACATTACCACCATTCCGGAGCTATTTTCCTAATAGTGCTATTCGGAGAT

      CCCCCGTATCACGATAAACTGCTTGAATCCGGGTAGCACTCATGACTCATCCTTTTCTGG
10981 ---------+---------+---------+---------+---------+---------+ 11040
      GGGGGCATAGTGCTATTTGACGAACTTAGGCCCATCGTGAGTACTGAGTAGGAAAAGACC

      TGGATCCTGAAATAATTCAGGCACGTATGTGGGTAGGCTGTAATCTTCTTGATCATCCGC
11041 ---------+---------+---------+---------+---------+---------+ 11100
      ACCTAGGACTTTATTAAGTCCGTGCATACACCCATCCGACATTAGAAGAACTAGTAGGCG

      TTCCACTGCGTTGAAAGGAATAACTTTAGCCTCATCACCTGAAGAGCTTGGCTTCAAATG
11101 ---------+---------+---------+---------+---------+---------+ 11160
      AAGGTGACGCAACTTTCCTTATTGAAATCGGAGTAGTGGACTTCTCGAACCGAAGTTTAC

      CTCCGCACGTTTTTTACGCTCCGAGGTCACGGCTTCTTTTTTATTAGTTTTAATCTGCTT
11161 ---------+---------+---------+---------+---------+---------+ 11220
      GAGGCGTGCAAAAAATGCGAGGCTCCAGTGCCGAAGAAAAAATAATCAAAATTAGACGAA
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

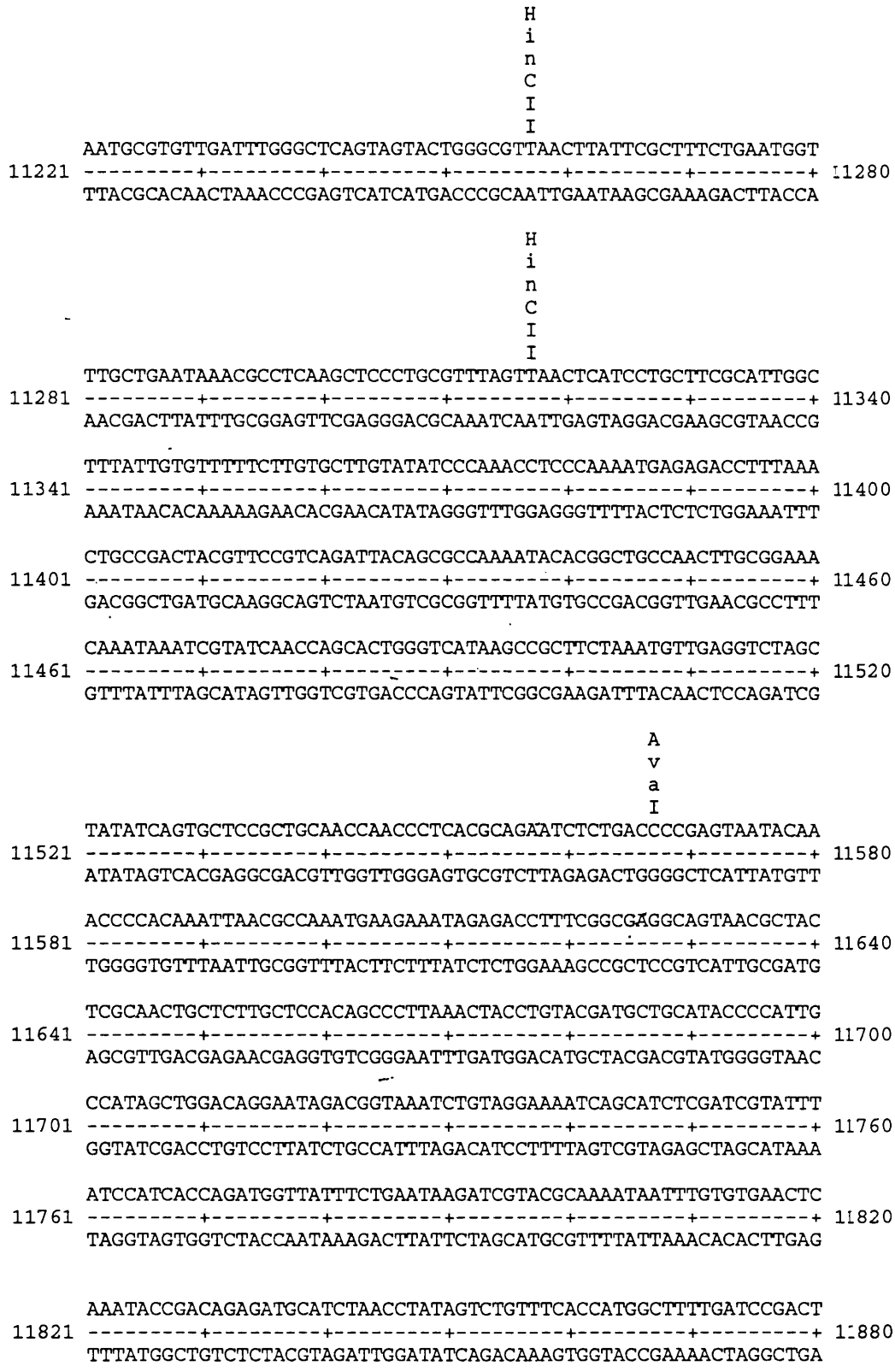


Fig. 3.5 Complete Sequence of Tn*7*

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
      GCCCTCTACAATGCCAGGTGCAAAGGACTTAAACTCGGCTTGTAGTGTTCTAAAAGTGCT
11881 ---------+---------+---------+---------+---------+---------+ 11940
      CGGGAGATGTTACGGTCCACGTTTCCTGAATTTGAGCCGAACATCACAAGATTTTCACGA

      TTCCACTATGCCTTTAGCATCGCCACGTCTAGGTGGAGCACTTTCCACTCGCACATTAAA
11941 ---------+---------+---------+---------+---------+---------+ 12000
      AAGGTGATACGGAAATCGTAGCGGTGCAGATCCACCTCGTGAAAGGTGAGCGTGTAATTT

      ACTAGAAACTAAGGCTTCGACCTGATGACTCATTAATTCGCCACGGTCCGCTAGCAACAC
12001 ---------+---------+---------+---------+---------+---------+ 12060
      TGATCTTTGATTCCGAAGCTGGACTACTGAGTAATTAAGCGGTGCCAGGCGATCGTTGTG

      ATCTGGCAAACCTACACACGGCCAGTCTGAGCTACTAATCTCAATATCATGCTGGGCACA
12061 ---------+---------+---------+---------+---------+---------+ 12120
      TAGACCGTTTGGATGTGTGCCGGTCAGACTCGATGATTAGAGTTATAGTACGACCCGTGT

      AATGGCCGTTTTGTCAGAGCAAGCATTTACAAAAGCCTGCATCGCCACCACATAAGACGG
12121 ---------+---------+---------+---------+---------+---------+ 12180
      TTACCGGCAAAACAGTCTCGTTCGTAAATGTTTTCGGACGTAGCGGTGGTGTATTCTGCC

      ATTTTCAAAGCCGATATAAAAGCCCGTGATCATCCGACTAAACACATCAATCACAATGTA
12181 ---------+---------+---------+---------+---------+---------+ 12240
      TAAAAGTTTCGGCTATATTTTCGGGCACTAGTAGGCTGATTTGTGTAGTTAGTGTTACAT

      AAGCGTTGGTCTTCCTATGATTTTTTGGCGATCATGATGATCCACTAAATAAATATCAGC
12241 ---------+---------+---------+---------+---------+---------+ 12300
      TTCGCAACCAGAAGGATACTAAAAAACCGCTAGTACTACTAGGTGATTTATTTATAGTCG

      AATCGTGGCATCAATCTCATAACGACTCCCAGGGCCTAACGCCTGAGAAGTGGCTGTACT
12301 ---------+---------+---------+---------+---------+---------+ 12360
      TTAGCACCGTAGTTAGAGTATTGCTGAGGGTCCCGGATTGCGGACTCTTCACCGACATGA

                                                X
                                                b
                                                a
                                                I
      ACTTAAGGGTCGTACGTCTTTTTTATATACCCCTGCTTTAACTCTAGACTTTAAGCGCTG
12361 ---------+---------+---------+---------+---------+---------+ 12420
      TGAATTCCCAGCATGCAGAAAAAATATATGGGGACGAAATTGAGATCTGAAATTCGCGAC

      AGCTTTAGGGTATTCTCGATCATAAAAATAACGAAACTGACGTAGTGTTGGGTAATCCTC
12421 ---------+---------+---------+---------+---------+---------+ 12480
      TCGAAATCCCATAAGAGCTAGTATTTTTATTGCTTTGACTGCATCACAACCCATTAGGAG

      TTGGGGAATGCGAGGAAAATACTGAGCAAACAAGTCCACAAATCGTCTATAGGCAACGGT
12481 ---------+---------+---------+---------+---------+---------+ 12540
      AACCCCTTACGCTCCTTTTATGACTCGTTTGTTCAGGTGTTTAGCAGATATCCGTTGCCA

                                                H
                                                i
                                                n
                                                c
                                                I
                                                I
      GGTCTTTGTACCTTTTTGATTTAACAGGTGCTTTTCTATGGTCAACCTAAAAAGGCGTTC
12541 ---------+---------+---------+---------+---------+---------+ 12600
      CCAGAAACATGGAAAAACTAAATTGTCCACGAAAAGATACCAGTTGGATTTTTCCGCAAG
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

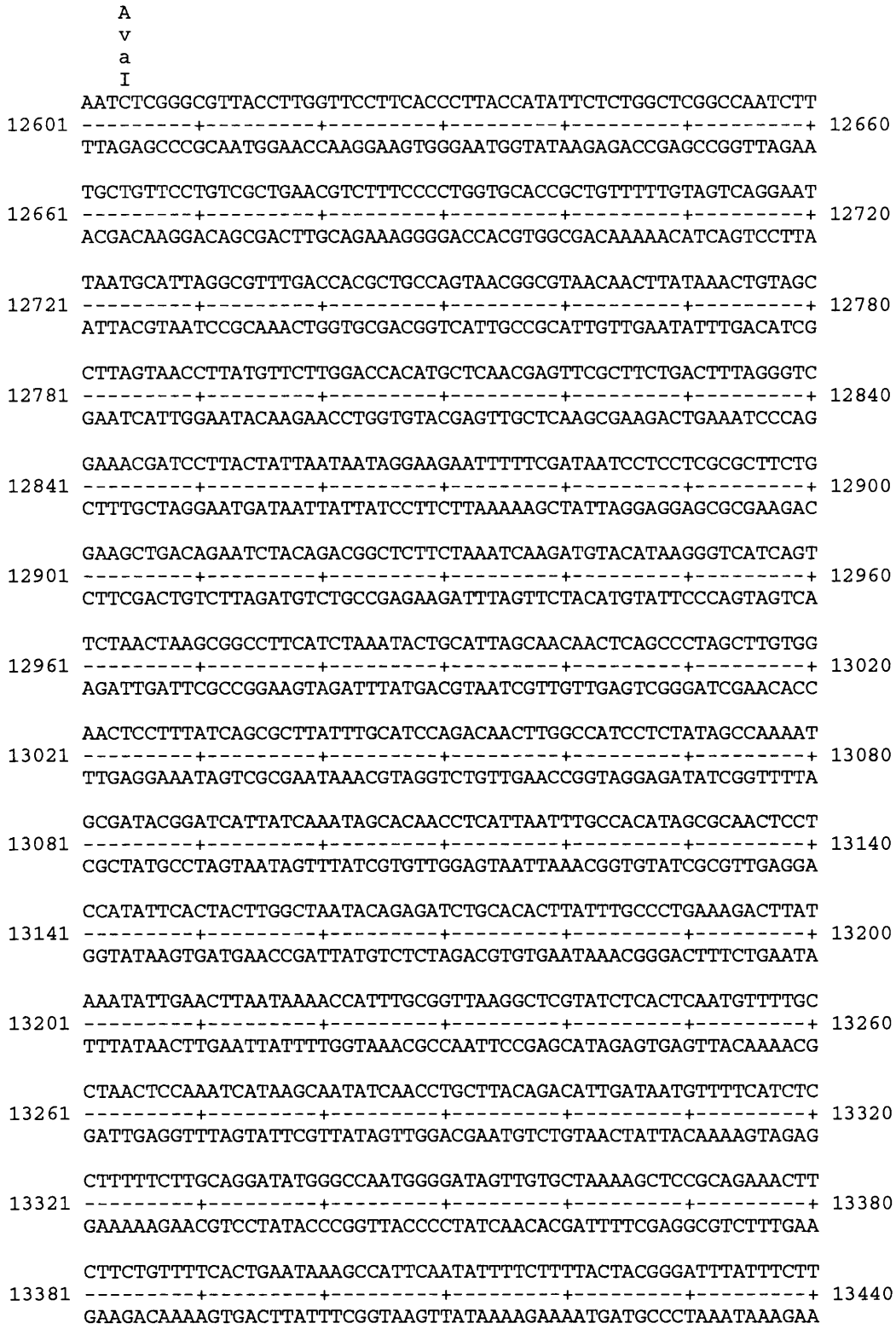


Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

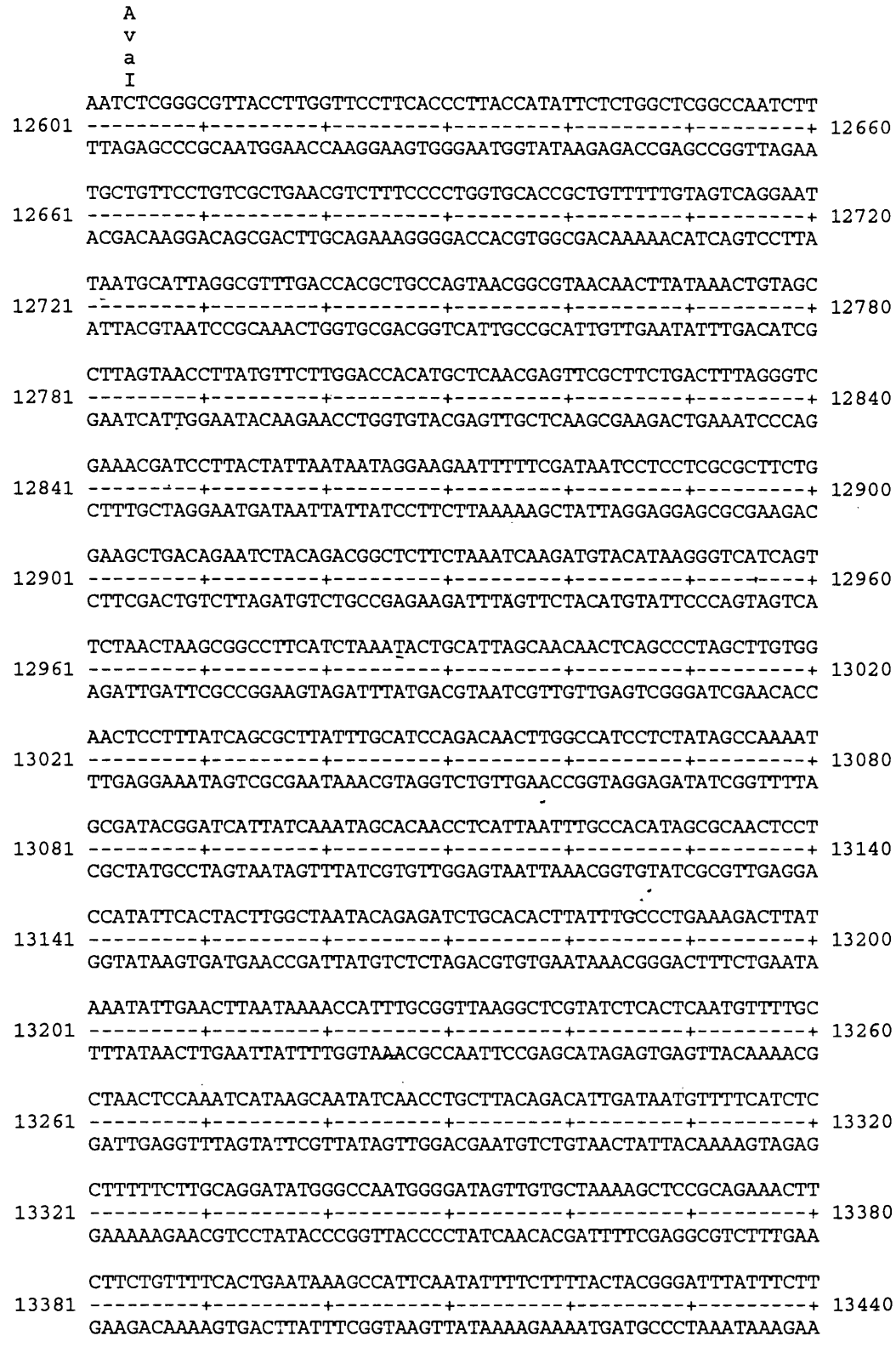


Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

```
      TATCAGTAAAAATGAACCAAGGAATTTGCTTTTGCTGCCAATAGCGACGCTCTAGTTCTA
13441 ---------+---------+---------+---------+---------+---------+ 13500
      ATAGTCATTTTTACTTGGTTCCTTAAACGAAAACGACGGTTATCGCTGCGAGATCAAGAT

                                              P
                                              s
                                              t
                                              I
      GTTTTTCTAAGGTACGCTCGTCTTGTAAGGCTGCTGCAGGTTTGACTTGAATAGCAAACT
13501 ---------+---------+---------+---------+---------+---------+ 13560
      CAAAAAGATTCCATGCGAGCAGAACATTCCGACGACGTCCAAACTGAACTTATCGTTTGA

      GCTCAAAAGGACCATCTTTGCAGTCCACTAAAAAATCAGTAGACATAACCTGATCTACAC
13561 ---------+---------+---------+---------+---------+---------+ 13620
      CGAGTTTTCCTGGTAGAAACGTCAGGTGATTTTTTAGTCATCTGTATTGGACTAGATGTG

      CACGAATAACAGGATGCTTAATACCACTATCTATTGCAATCTGCCTGGTATCACTAGGTA
13621 ---------+---------+---------+---------+---------+---------+ 13680
      GTGCTTATTGTCCTACGAATTATGGTGATAGATAACGTTAGACGGACCATAGTGATCCAT

      ATAAGGGGAACTGCTCGCGTATATCTAGCACGCTGCTCTCCCACTCAAGACTGAGAAAAA
13681 ---------+---------+---------+---------+---------+---------+ 13740
      TATTCCCCTTGACGAGCGCATATAGATCGTGCGACGAGAGGGTGAGTTCTGACTCTTTTT

      CAGCAAGCTCTAAGTCAGATAGCAAATGATGGACTCGTCCCGTCTTATGAGAATAAATAC
13741 ---------+---------+---------+---------+---------+---------+ 13800
      GTCGTTCGAGATTCAGTCTATCGTTTACTACCTGAGCAGGGCAGAATACTCTTATTTATG

      GGTGGGAACGACCTGAAGAAGGAACTTCTTGTACTGTTAGCCATGGAATATAGTCTTTAC
13801 ---------+---------+---------+---------+---------+---------+ 13860
      CCACCCTTGCTGGACTTCTTCCTTGAAGAACATGACAATCGGTACCTTATATCAGAAATG

      CATGCCCTTGGCCACGCCCCTCTTTAATACGACGGGCAATTTGCACTTCAGAAAATGAAG
13861 ---------+---------+---------+---------+---------+---------+ 13920
      GTACGGGAACCGGTGCGGGGAGAAATTATGCTGCCCGTTAAACGTGAAGTCTTTTACTTC

      AGTTTGCTTTAGCCATAACAAAAGTCCAGTATGCTTTTTCACAGCATAACTGGACTGATT
13921 ---------+---------+---------+---------+---------+---------+ 13980
      TCAAACGAAATCGGTATTGTTTTCAGGTCATACGAAAAAGTGTCGTATTGACCTGACTAA

      TCAGTTTACAACTATTCTGTCTAGTTTAAGACTTTATTGTCATAGTTTAGATCTATTTTG
13981 ---------+---------+---------+---------+---------+---------+ 14040
      AGTCAAATGTTGATAAGACAGATCAAATTCTGAAATAACAGTATCAAATCTAGATAAAAC

      TTCAGTTTAAGACTTTATTGTCCGCCCACA
14041 ---------+---------+---------+ 14070
      AAGTCAAATTCTGAAATAACAGGCGGGTGT
```

Fig. 3.5 Complete Sequence of Tn7

Integrase pseudogene nt. 1921-905
dhfr nt. 2268-2727
sat nt. 2821-3345
aadA nt. 3403-4192
Internal repeats nt 4389-4513

TnsE nt. 7828-6202
TnsD nt. 7827-9349
TnsC nt. 11023-9353
TnsB nt. 13128-11024
TnsA nt. 13936-13114

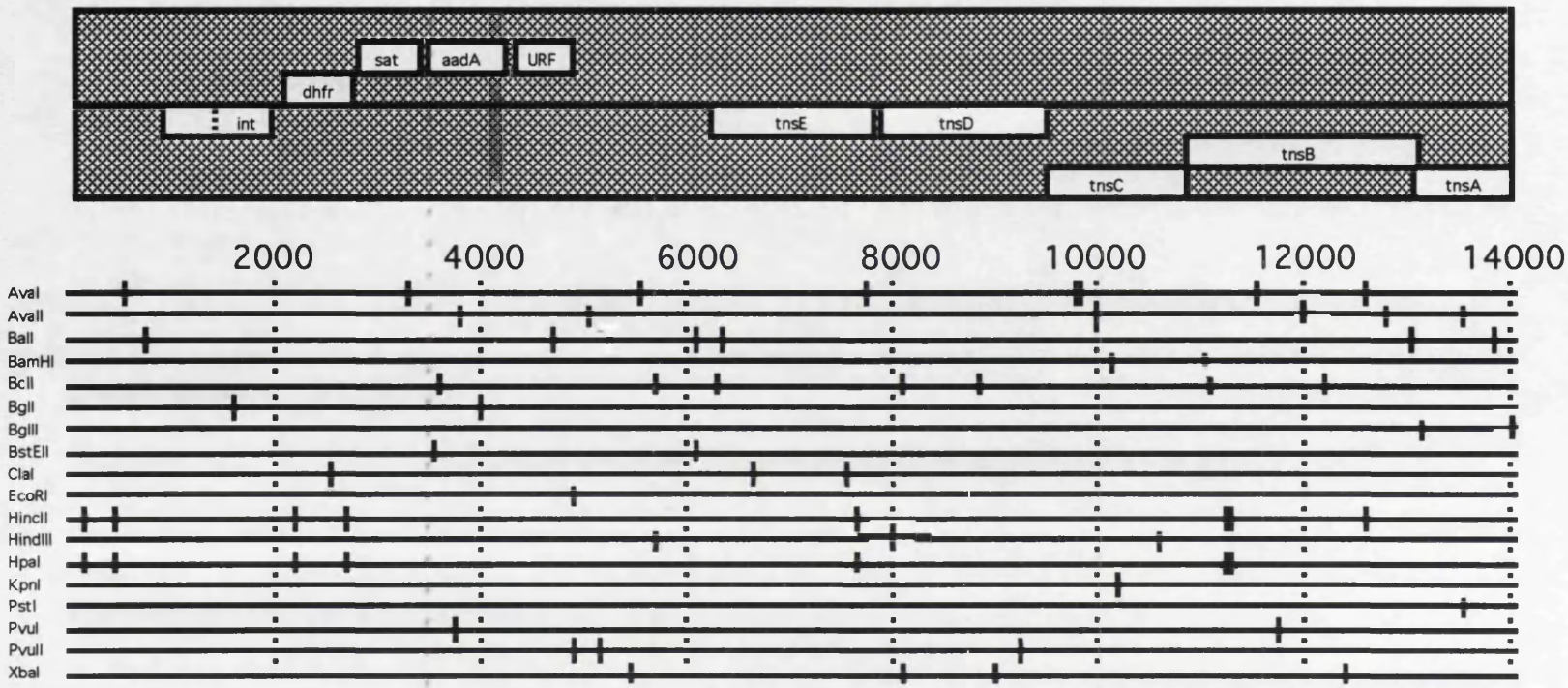


Fig. 3.6 Open reading frames and selected restriction sites in the Tn7 sequence
Coding regions are shown as white boxes in the upper panel. The dotted line in the integrase reading frame represents the stop codon, as discussed in the text.
The restriction pattern was derived using the GCG program PLOTMAP , vertical lines represent recognition sites of the listed enzymes.

```
        TGAGGGGAACCACTAGTGTCATACGAACCTCCAAGAGACGGTTACACAAACGGGTACATT
    4810 +---------+---------+---------+---------+---------+--------- 4869
        ACTCCCCTTGGTGATCACAGTATGCTTGGAGGTTCTCTGCCAATGTGTTTGCCCATGTAA

a                       V  S  Y  E  P  P  R  D  G  Y  T  N  G  Y  I

                                                P
                                                v
                                                u
                                                I
                                                I
        GTTGATGTCATGTATGACAATCGCCCAAGTAAGTATCCAGCTGTGTTCAGAACGTACGTC
    4870 +---------+---------+---------+---------+---------+--------- 4929
        CAACTACAGTACATACTGTTAGCGGGTTCATTCATAGGTCGACACAAGTCTTGCATGCAG

a       V  D  V  M  Y  D  N  R  P  S  K  Y  P  A  V  F  R  T  Y  V

         E
         c
         o
         R
         I
        CGAATTCCAGACTATGTCGAACTAGATAACGCGGAGGAGCTGGCCGACGCAATCATTCGA
    4930 +---------+---------+---------+---------+---------+--------- 4989
        GCTTAAGGTCTGATACAGCTTGATCTATTGCGCCTCCTCGACCGGCTGCGTTAGTAAGCT

a       R  I  P  D  Y  V  E  L  D  N  A  E  E  L  A  D  A  I  I  R

        GAGGTTGAGTTAAAGGAGGCATGGCTCACAAGGGTCTTGGGGAACTCAGCCGATCCTGAC
    4990 +---------+---------+---------+---------+---------+--------- 5049
        CTCCAACTCAATTTCCTCCGTACCGAGTGTTCCCAGAACCCCTTGAGTCGGCTAGGACTG

a       E  V  E  L  K  E  A  W  L  T  R  V  L  G  N  S  A  D  P  D

        GCTTCTGTGGGCAAGTGGGACCGGCGCCTAGCGAGTGCGTGGTATTGATGA
    5050 +---------+---------+---------+---------+---------+ 5100
        CGAAGACACCCGTTCACCCTGGCCGCGGATCGCTCACGCACCATAACTACT

A       A  S  V  G  K  W  D  R  R  L  A  S  A  W  Y  *  *  -
```

Fig. 3.7 Sequence of short open reading frame downstream from URF1.

3.3 Analyses of sequence

The complete, compiled sequence of Tn7 is 14070 base pairs in length. It's composition is 55.4% A/T and 44.6% G/C. The nucleotide composition of the Tns genes (nts 5720-14070) had been previously determined as 56.5% A/T and 43.5% G/C (Flores, Quadri et al. 1990). Therefore the left end of the element (nts1-5179, 53.9% A/T), is marinally less A/T rich than the right but still considerably higher than *E.coli*, 48.3% A/T.

3.4 Restriction Map of Tn7

A restriction map of Tn7 has been previously published (Gosti-Testu, Norris et al. 1983). When the restriction map of Tn7, predicted from the compiled sequence, is compared with this work a good agreement is observed. (See Fig. 3.6)
One discrepancy appears in the left end of the element. The published work places a BclI site between the two HpaI sites in the first 500 nt of the left end. This site, TGATCA, does not appear in the predicted restriction map of the compiled Tn7 sequence. Examination of this region of the sequence reveals, however, a highly related sequence, TGATCC at coordinates 285-291. This difference presumably due to a polymorphism in the different isolates of Tn7 used in the different laboratories. The only other discrepancy is in the presence of an additional BalI site in the sequence of the 3'end of the *tnsE* gene determined by Flores et al.

3.5 Open reading frames

No major open reading frames were found in addition to the previously characterised *tns* genes, drug resistance and integrase pseudo genes in the compiled Tn7 sequence. An open reading frame has been found downstream of the *aadA* gene (Fling, Kopf et al. 1985), this putative ORF is shown as URF1 in Fig. 3.6. The sequence of the putative product of this gene was used in a TFASTA search, no significant homologies were found. Another short open frame starting with a GTG codon is found downstream of URF1, the potential product of this region is only 90 amino acids long, (Fig. 3.7), and again shows no significant homologies on a TFASTA search against the GenBank and EMBL databases.

```
   Visna  SDNGPAFVAE STQLLMKYLG IEHTTGIPWN PQSQALVERT HQTLKNTLEK
    HIV1  TDNGSNFTST TVKAACWWAG IKQEFGIPYN PQSQGVIESM NKELKKIIGQ
     BLV  TDQGANYTSK TFVRFCQQFG VSLSHHVPYN PTSSGLDERT NGLLKLLLS
   HTLV1  TDNGPAYISQ DFLNMCTSLA IRHTTHVPYN PTSSGLVERS NGILKTLLY
     RSV  TDNGSCFTSK STREWLARWG IAHTTGIPGN SQGQAMVERA NRLLKDRIRV
    MMTV  TDNAPAYVSR SIQEFLARWK ISHVTGIPYN PQGQAIVERT HQNIKAQLNK
   MoMLV  TDNGPAFVSK VSQTVADLLG IDWKLHCAYR PQSSGQVERM NRTIKETLTK
   Copia  IDNGREYLSN EMRQFCVKKG ISYHLTVPHT PQLNGVSERM IRTITEKART
 552orfa  TDHGSDFTSH HMEQVAIDLK INLMFSKVGV PRGRGKIERF FQTVNQTFLE
   IS629  SDKGSQYVSL AYTQRLKEAG LLASTGSTGD SYDNAMAESI NGLYKAKVIH
   IS911  SDQGSHYTSR QFRQLLWRYQ IRQSMSRRGN CWDNSPMERF FRSLKNEWIP
 IS3orfB  TDRGGQYCSA DYQAQLKRHN LRGSMSAKGC CYDNACVESF FHSLKVECIH
     IS2  TDNGSCYRAN ETRQFARMLG LEPKNTAVRS PESNGIAESF VKTIKRDYIS
    Tnsb  ADRG.ELMSH QVEALVSSFN VRVESAPPRR GDAKGIVEST FRTLQAEFKS
           * *    *             *                *   *
```

3.8 Aligned sequences of various retroelement IN proteins, IS3 family InsB protein sequences, Tn552 transposase (552orfA) and TnsB(aa 360- 428). Sequences were aligned using the GCG program PILEUP.

3.6 TnsB:: The Tn7 transposase ?

As discussed in the introduction the transposase of a given element is a protein which generally binds to sequence motifs at the transposon termini and catalyses the strand exchange reaction. Despite the similarities in reaction mechanism catalysed by these proteins there is often very little obvious homology between them. One striking recent observation however, was of a strong similarity between the transposases of the IS3 family of bacterial insertion sequences and the IN protein of retroviruses (Fayet, Ramond et al. 1990). This similarity was also found in p480, the transposase of Tn552, a *Staphyloccocus aureus* transposon (Rowland and Dyke 1990). The strongest similarity lies between 2 absolutely conserved amino acids, an aspartate and a glutamate residue. The sequence between these residues also shows considerable sequence conservation and the 35 amino acid spacing is extremely highly conserved. Recent work has shown that mutation of the conserved aspartate or glutamate vastly reduces the ability of the HIV IN protein 3' processing, strand exchange and disintegration reactions (Engleman and Craigie, 1992).
Another feature shared by elements with transposases containing this D-35-E motif is that their 3' terminal nucleotides are 5' CA 3'. These nucleotides are generally not required for transposase binding but may be important for cleavage. Because Tn7 also has 3' termini of CA it was of interest to see if any of the Tn7 proteins possed a D-35-E motif, the best candidate for a Tn7 transposase is TnsB, the protein shown to bind the terminal 22bp motifs of Tn7. The amino acid sequences of this protein along with various members of the IS3 and retrovirus families were compiled and aligned using the GCG program PILEUP (Devereux, Haeberli et al. 1984). The aligned sequences are shown in Fig 3.8. A similar region to the D-35-E exists in TnsB, with the important difference that the spacing is 34 and not 35 amino acids between the conserved aspartate and glutamate. However, the most highly conserved residues of the D-35-E are also found inthe TnsB sequence. It is possible therefore that TnsB is the Tn7 transposase. Perhaps, the altered spacing in the motif in TnsB is indicative of the fact that additional Tn7 proteins are involved in the transposition reaction.

GACAAWAWAGTYBKRAACTRRR CONSENSUS (Mizuuchi, 1992)

CAACTAC	
GACATTTCTCCAAGCAACTAC	W=AT B=GCT K=GT
GACAACTCCATAAGAAATTAC	
GACAATAGTCCATCCAATTAC	INTERNAL
GACAACTCTGAGAGCAACTAC	
GATAATAGTTCATCCAATTAC	
GACAATAGTCGCAA	
GACAATAAAGTCTTAAACTGA	
ACAAAATAGATCTAAACTAT	RIGHT END
GACAATAAAGTCTTAAACTA	
GACAGAATAGTTGTAAACTG	
GACAAAATAGTTGGGAACTGGG	
GACAAAATAGATGGGAACTGGG	LEFT END
GACAAAATAGTTTGGAACTAGA	

Fig. 3.9 Comparison of the sequence of the internal repeats (Tn*7* nt. 4389-4513) with the sequence of the terminal repeats. The sequence listed for both internal and right end repeats is continuous sequence, the left end repeats are discontinuous. Matches to the consensus are underlined.

3.7 Tn7 Internal Repeats

An unusual feature of the Tn7 sequence is the presence of a number of directly repeated motifs downstream of the *dhfrI* gene. These repeats bear considerable similarity to the terminal 22 bp repeats. The similarity is strongest at the ends of each repeats, the inner portion is more divergent (See Fig. 3.9). The terminal repeats have been demonstrated to be the binding sites for TnsB (Morrell 1990 ; Arciszewska, Mckown et al. 1991 ; Tang, Lichtenstein et al. 1991). The following work attempted to address whether the Internal Repeats might also function as binding sites for TnsB.

Studies of phage Mu have revealed the importance of a series of internal repeats, the internal activating signal (IAS), in the transposition reaction. MuA which, like TnsB, binds the terminal repeats of the element also binds the IAS. In the case of MuA the IAS sequences are distinct from the terminal motifs and separate domains of the protein are required for binding to the two sets of repeats. Recent work has demonstrated that the net effect of this series of interactions is to facilitate synapsis of the ends of Mu. (Mizuuchi and Mizuuchi 1989; Surette and Chaconas 1992) Deletion of the Mu IAS has shown it is not essential for transposition, indeed transposition is only reduced 5-10 fold. Analysis of Tn7 has shown that deletion of a fragment containing the internal repeats has no effect on the apparent transposition frequency; however, as discussed previously, this may simply reflect the insensitivity of the transposition assay itself.

The similarity between the terminal and internal repeat sequences suggests that the internal repeats may function as binding sites for TnsB. If they do bind TnsB then it is possible, by analogy with Mu, that these repeats may be a Tn7 transpositional enhancer.

3.9 TnsB binding to Internal Repeats

The internal repeats were subcloned from pMR12::Tn7-1 as a 511bp NaeI fragment and ligated into SmaI digested pUC18. Fig. 3.10 shows a series of gel retardation assays carried out using labelled HindIII-EcoRI pEN200, Tn7 right end, and labelled HindIII RsaI digested pSB100 as DNA substrates. The source of TnsB was a crude extract of DS941 pMR207. 1µg of Salmon

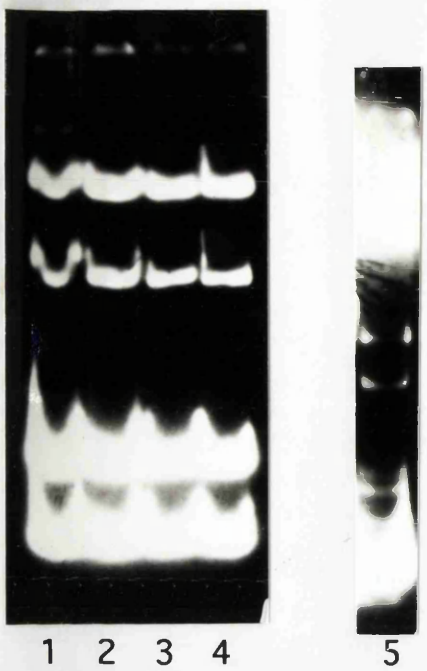


Fig. 3.10 Binding to internal repeats

Lanes 1-4 End-labelled pSB100 (HindIII/RsaI) digest
Lane 5 labelled EcoRI/HindIII digested pNE200
Incubated with:-

Lane 1. 2 mg DS941 pMR207 extract
Lane 2. 1 mg DS941 pMR207 extract
Lane 3. 0.5 mg DS941 pMR207 extract
Lane 4. No extract
Lane 5. 0.5 mg DS941 pMR207 extract

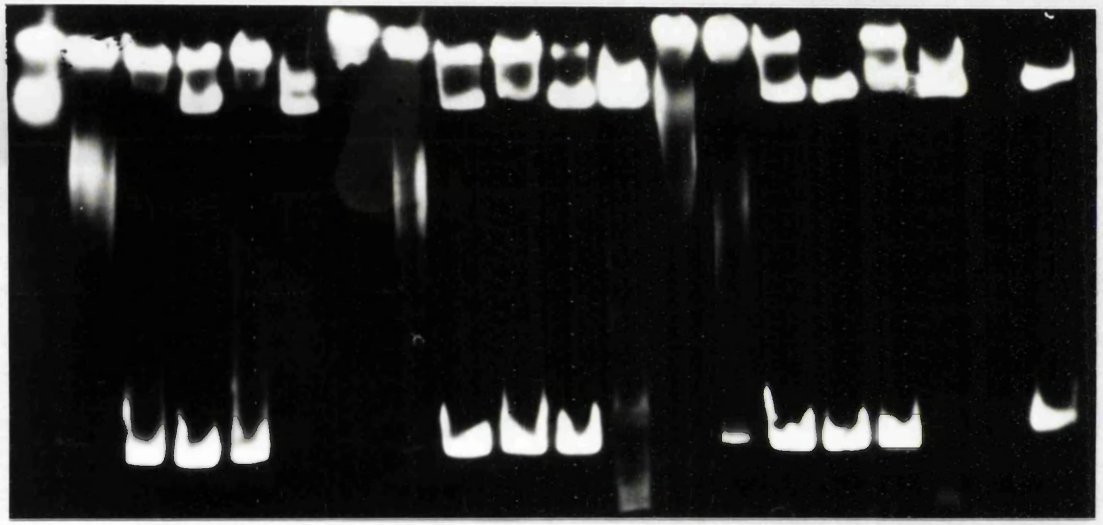


Fig. 3.11 Effect of varying salt concentrations on binding to internal repeats

Reactions were carried out under standard TnsB conditions with the exception of varying salt concentrations. Amounts of DS941 pMR207 extract added are as shown. DNA used was as detailed below.

Lanes 1-5, 7-11 and 13-17 EcoRI/HindIII digested pSB110
Lanes 6,12 and 18 EcoRI/HindIII digested pNE200

Lane 19 No extract

Sperm DNA was included as competitor. It is clear from this figure that the right end motifs are bound by the extract, the internal repeats however are apparently not bound at the extract concentrations used.
In subsequent experiments the source of internal repeat DNA was pSB110. This plasmid was made by HindIII RsaI digestion of pSB100, gel isolation of the appropriate 297bp fragment and ligation to HindIII SmaI cut pUC18. The insert in pSB110 was confirmed by DNA sequencing.

The experiment above demonstrated that under standard binding conditions no TnsB binding activity directed to the internal repeats could be detected. The experiment was therefore repeated with binding carried out at a variety of salt concentrations. The result of this is shown in Fig. 3.11, once again binding to the right-end controls was detected . At high concentrations of TnsB the internal repeat DNA was retarded. However, the DNA was smeared up the gel, furthermore, at these extract concentrations the pUC backbone sequences are also observed to be retarded. It seems likely, therefore, that the retardation observed is a non-TnsB-specific interaction.

Final possibilities are that TnsB may only transiently interact with the internal repeats or that interactions with the internal repeats may be dependent on binding to the terminal repeats.

A series of TnsB retardations were set up using standard conditions with labelled EcoRI-HindIII digested pNE200. A 2 fold dilution series of pSB84 extract (TnsABCD) was added . One set of assays contained 500ng of Salmon Sperm DNA as competitor, the other 500ng of EcoRI-HindIII digested pSB110. The latter set of reactions therefore contained unlabelled internal repeat DNA at approximately 100 times the concentration of the labelled Right end sequences. If there is any competition between Internal and terminal repeat DNA, or formation of sandwich complexes between the two fragments, distinct patterns of retardation will be observed for the reactions with specific and non-specific competitor. However, as can be seen in Fig. 3.12, the retardation patterns are essentially identical for both sets.
An analogous experiment used radiolabelled internal repeat DNA with various TnsB concentrations. Competitor DNA in these experiments was either 100,1 or 0.01ng of EcoRI-HindIII digested pMR11 (pUC18::Tn7-1) or

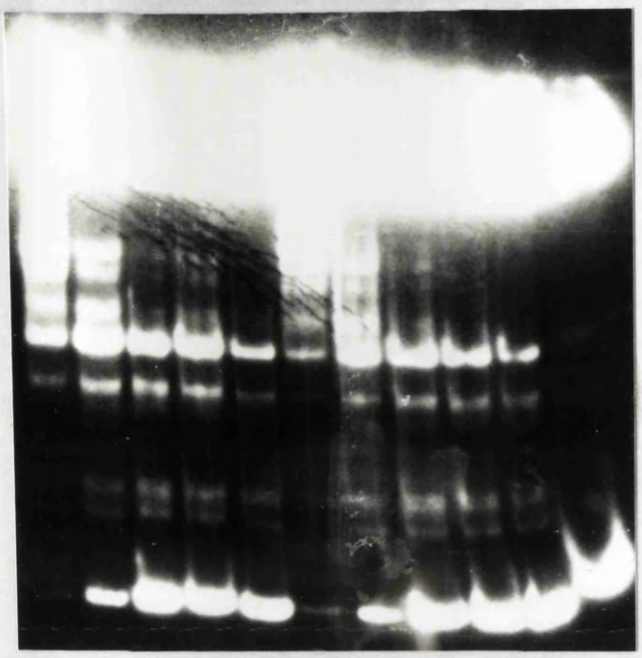


Fig. 3.12 Do the internal repeats compete for binding to the right end ?

Lanes 1-5 : 500ng Salmon sperm DNA as competitor
Lanes 6-10: 500ng pSB110 DNA as competitor

DNA substrate was end labelled, EcoRI/HindIII digested pNE200

Lanes 1&5: 2µg CSH50 pGP1-2 pSB84 (TnsABCD) extract
Lanes 2&6: 1µg CSH50 pGP1-2 pSB84 (TnsABCD) extract
Lanes 3&7: 0.5µg CSH50 pGP1-2 pSB84 (TnsABCD) extract
Lanes 4&8: 0.25µg CSH50 pGP1-2 pSB84 (TnsABCD) extract
Lanes 5&9: 0.13µg CSH50 pGP1-2 pSB84 (TnsABCD) extract
Lanes 6&10: 0.06µg CSH50 pGP1-2 pSB84 (TnsABCD) extract
Lane 11: no extract

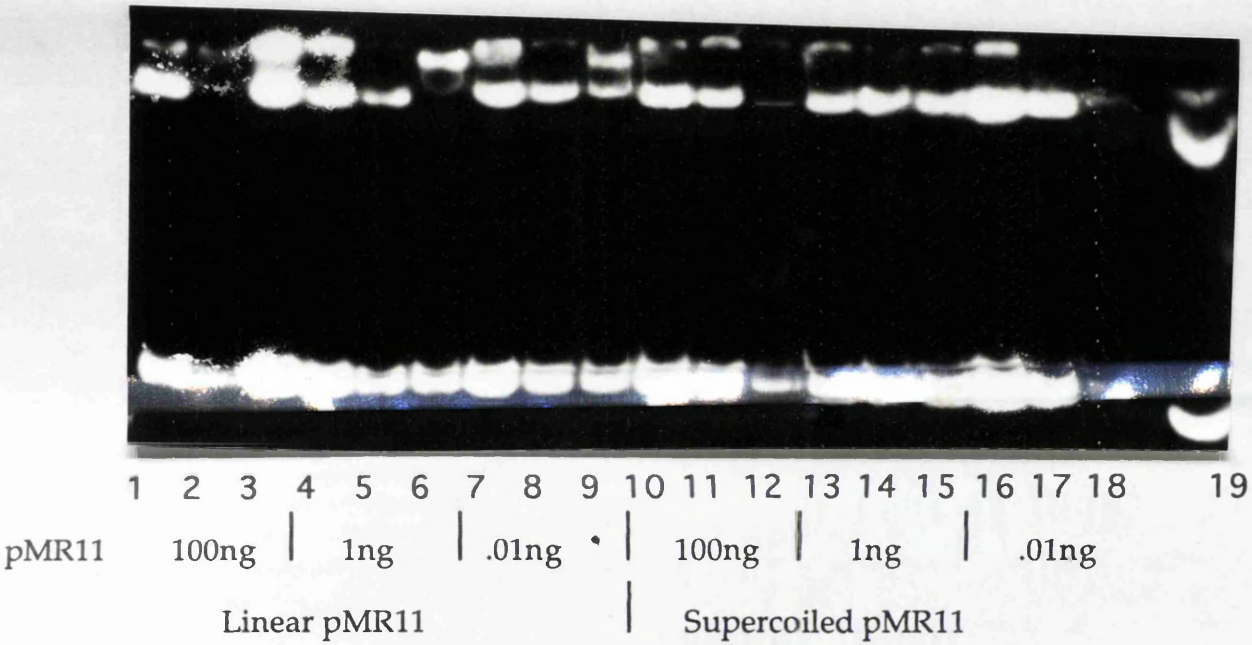


Fig. 3.13 Effect of pMR11 in attempts to detect binding to internal repeats.

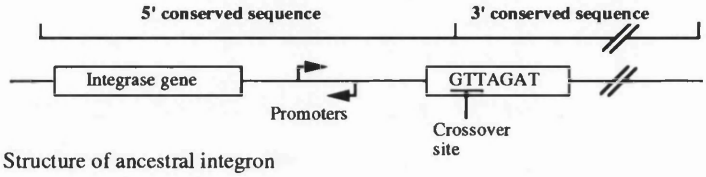
Binding reactions were carried out under standard TnsB conditions, with the addition of unlabelled pMR11 as detailed above. 1µg of DS941 pMR207 extract was used in the reactions, the labelled DNA was an EcoRI/HindIII digest of pSB110.
Lane 19 has no protein added.

100, 1 or 0.01ng of supercoiled pMR11. The aim of this experiment was to detect potential interactions between end and internal repeats; none were observed (See Fig. 3.13).

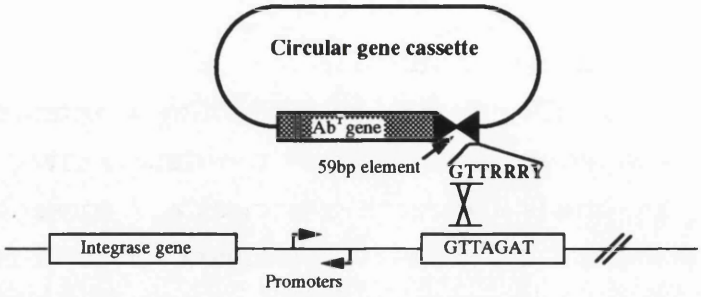
Taken together with the published data showing no loss of transposition activity on deletion of this region, the experiments described above do not support the idea that the internal repeats are functionally related to the Mu IAS. Precisely what role, if any, these repeats play is unclear. In view of the strong similarity between this sets of motifs and the sequence of the terminal 22 bp repeats it seems highly unlikely that the similarity is just co-incidental. An obvious feature of the internal repeats is that they lie in the same orientation as the left end repeats, yet the directly abutting nature of the repeats is more reminiscent of the organisation of the right end. It is possible, therefore, that these motifs represent a redundant left end, isolated in the evolutionary history of Tn7. If an insertion event took place in the left end of a primordial Tn7 within the terminal repeats , then the repeats to the right of the insertion event would be internalised.. Because these repeats would no longer be adjacent to the terminal 8bp of Tn7 they could not be used in transposition events. There would, therefore, be no selective pressures for them to retain a TnsB binding activity. Indeed, if binding activity was retained the motifs might serve as a "sink" for TnsB and could even bring about incorrect synapsis of the right end with the centre, thus inhibiting transposition. It is interesting to note that whilst an artificial mini-Tn7 with two right ends can transpose, one with two left ends cannot. This lack of functional redundancy may manifest itself in defining the unique polarity of insertion in the *attTn7* site. It may also reflect the recent aquisition of the left end.

3.10 Integrons as Transposable Elements

If the internal repeats in Tn7 do represent a redundant left end then what could have occured to isolate them during the development of the present day Tn7 ? As mentioned in Chapter 1 the drug resistance determinants in Tn7, which lie between the left end and the internal repeats, lie within a structure which has considerable similarity to the integron structure found in the Tn21 family of transposons (Sundstrom, Roy et al. 1991). A recent theory has suggested that the integron itself may be, or have been, a transposable element (Bissonnette and Roy 1992). Therefore insertion of



Structure of ancestral integron



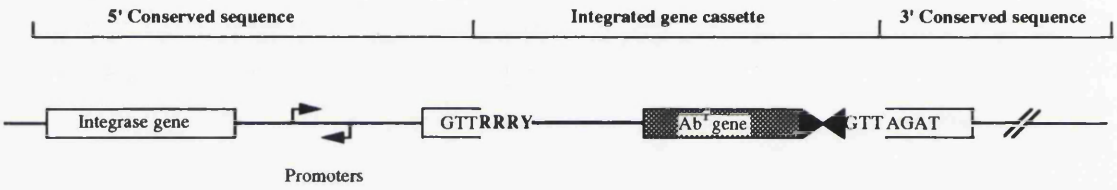


Fig. 3.14: A model for gene cassette integration in integrons.

the integron may have been the event which internalised the internal repeat.

3.11 A Summary of Evidence on Tn21 type Integrons

As discussed in Chapter 1 integrons are genetic elements which mediate the integration of discrete gene cassettes into a common structure. The following discussion will summarise the data derived from the Tn21 family on integrons.The integron comprises a 5′ conserved segment, the inserted gene(s) and a 3′ conserved segment. The 5′ conserved segment contains an open reading frame with homology to lambda type integrases, a promoter for the integrase gene and a promoter on the opposite strand which transcribes the inserted genes. The inserted gene cassettes all have a similar structures, they generally comprise the open-reading frame and a conserved element with considerable dyad symmetry at the 3′ end of the cassette. A consensus sequence of 60 bp has been derived for this element, however,for historical reasons the element is is known as the 59bp element (Hall, Brookes et al. 1991). There have been several instances where the same gene has been found in different integrons, in all these cases the same 59bp element is found with the gene, thus it appears that the gene plus 3′ 59 bp element is the mobile unit. The 3′ conserved segment comprises an open reading frame *orf4*, a sulphonamide resistance determinant, and a 2.7 kb fragment, in a number of examples a further sequence is inserted between the sulphonamide resistance determinant and the 2.7 kb fragment (Bissonnette and Roy 1992). This region is discussed in greater detail below.

Studies of recA independant, integrase dependant co-integrate structures formed by integrase mediated recombination between replicons containing two different 59 bp elements showed that recombination occurs within the sequence GTT at the 3′ end of the 59 bp element (Martinez and de la Cruz 1988 ; Martinez and de la Cruz 1990). This GTT is contained within a larger consensus, GTTRRRY. Intriguingly a dyad related site is found at the 5′ end of the 59bp element. This site, however, is apparently never used as recombination site. A GTTRRRY sequence is found at the 3′ end of the 5′ conserved segment (See Fig. 3.14 for clarification). In addition to the 59bp element, this site too can serve as a recombination site. Thus

all inserted genes in an integron are flanked by direct repeats of potential recombination site.
Excision of gene cassetes from an integron has been observed and the sequence of the novel junction in the integron was compatible with the proposal that a recombination event had occurred between the directly repeated GTTRRRY motifs flanking the excised gene cassette (Collis and Hall 1992).
Recent work has also shown that over-expression of the integrase gene can lead to excision of covalently closed circular molecules of the correct size to be the products of recombination between the directly repeated GTTRRRY flanking one or more integrated gene cassettes. It has not yet been shown that the recombination event which produced these molecules has occured between these sites or that these circular products do represent true intermediates in the dissemination of gene cassettes (Collis and Hall 1992).

The current model for the integration of gene cassettes proposese that an ancestral integron existed, this integron contained a single GTTRRRY site , not associated with a 59bp repeat. Circular gene cassettes corresponding to gene+59 bp element, from an unknown source, then underwent an int mediated recombination reaction in integrated into the ancestral integron.
Recent work has identified such a putative ancestral integron, In0 from*Pseudomonas aeruginosa*, This element comprises of the predicted 5' conserved region and 3' conserved region separated by a single GTTRRRY site. The authors of the paper identifying this structure propose this is an ancestor of more complex integrons. However, the discovery of the precise, *int* mediated excision of gene cassettes from such complex integrons makes the validity of this proposal doubtful.
This work also demonstrated that the Tn21 family integrons are delimited by 29 bp inverted repeats known as Brown's repeats. It was also stated that the 3' conserved region of these integrons contains an open reading frame with homology to the transposase of Tn552 and the *tnsB* gene of Tn7. These facts coupled with the observation that one integron, Tn402, has been observed to transpose to two different sites in bacteriophage lambda led to the propsoal that the Tn21 family integrons may themselves be transposable elements (Bissonnette and Roy 1992).

```
Tn7.int1 .........MSNSPFLNSIRTDMRQKGYALKTEKTYLHWIKRFILFHKK 40
                  :.. ..|: :|. :|   |.|:||..|:||:: || ||
Tn21 int1MKTATAPLPPLRSVKVLDQLRERIRYLHYSLRTEQAYVHWVRAFIRFHGV 50

                .         .         .         .         .
Tn7    41 RHPQTMGSEEVRLFLSSLANSRHVAINTQKIALNALAFLYNRFLQQPLGD 90
          |||.|:||.||  |||.|||.|.|.:.|:: ||.|| |:|.:.|  .|.
Tn21   51 RHPATLGSSEVEAFLSWLANERKVSVSTHRQALAALLFFYGKVLCTDLPW 100

                .         .         .         .         .
Tn7    91 IDYIPASKP.RRLPSVISANEVQRILQVMDTRNQVIFTLLYGAGLRINEC 139
          :: |. .:| |||| |:.::|| |||..::. :.::  ||||.|:||.|.
Tn21  101 LQEIGRPRPSRRLPVVLTPDEVVRILGFLEGEHRLFAQLLYGTGMRISEG 150

                .         .         .         .         .
Tn7   140 LRLRVKDFDFDNGCITVHDGKGGKSRNSLLPTRLIPAIK*LIEQARLIQQ 189
          |.|||||:|||:|.|.|::|||:|.|. :||..| |.:: :..||
Tn21  151 LQLRVKDLDFDHGTIIVREGKGSKDRALMLPESLAPSLREHVSRARAWWL 200

                .         .         .         .         .
Tn7   190 DDNLQG.VGPSLPFALDHKYPSAYRQAAWMFVFPSSTLCNHPYNGKLCRH 238
          .|. :|  |..|| ||::|||.| :  :|::||:  | :..| .| : ||
Tn21  201 KDQAEGRSGVALPDALERKYPRAGHSWPWFWVFAQHTHSTDPRSGVVRRH 250

                .         .         .         .         .
Tn7   239 HLHDSVARKALKAAVQKAGIVSKRVTCHTFRHSFATHLLQAGRDIRTVQE 288
          |:.| . .:|:| ||:.||| .|..|.||:|||||| ||..| ||||||:
Tn21  251 HMYDQTFQRAFKRAVEQAGI.TKPATPHTLRHSFATALLRSGYDIRTVQD 299

                .         .         .         .         .
Tn7   289 LLGHNDVKTTQIYTHVLGQHLRHHQSCGWTDATYQSVKVDKYLYVALSSP 338
          ||||.||.|| ||||||              ... .:|: .   ..:|.|.
Tn21  300 LLGHSDVSTTMIYTHVL.............KVGGAGVRSPLDALPPLTSE 336

Tn7   339 * 340

Tn21  337 R* 338
```

Fig. 3.15 Comparison of integrase protein sequence from Tn7 and Tn21

3.12 The Tn7 Integron-like Structure

The *aadA1* and *sat* drug resistance determinants in Tn7 both possess 59bp elements, the *dhfrI* gene also has an imperfect inverted repeat at its 3′ end (Fling and Richards 1983 ; Fling, Kopf et al. 1985). This element contains GTTRRRY sequences at its ends but the interior is not homologous to the 59bp element. Both the *aadA1* and the *dhfrI* genes have been found with their associated inverted repeats in several Tn21 like integrons (Sundstrom and Skold 1990). Two relatives of Tn7, Tn1825 and Tn1826, have been identified which possess identical restriction maps to Tn7 apart from within the drug resistance coding regions. Characterisation of these elements has revealed that they possess different resistance determinants each with their own associated 59 bp element (Tietze, Brevet et al. 1987). Thus it would appear that an integron is present within Tn7.
However, Tn7 does not possess the conserved 5′ and 3′ segments associated with the Tn21 type integrons. There is a sequence with considerable homology to the Tn21 lambda type integrase found in Tn7's analogous 5′ segment, however the open reading frame is interupted by a stop codon. If the stop codon is ignored and the putative full length reading frame is aligned with that of the Tn21 family integrases, the alignment reveals 44% amino acid sequence identity between the two sequences, Fig. 3.15. Intriguingly the same stop codon is found in the sequence of the integrase in Tn1826. This suggests that this mutation was present prior to the divergence of these two members of the Tn7 family. If this is the case was the insertion of the respective, different resistance cassettes brought about by the integrase of a second integron from which the cassettes excised or was it due to suppression of the stop codon in the ancestral Tn7 integrase ?

3.13 Codon Usage in Tn7 encoded genes

It has been demonstrated that different organisms use different subsets of synonymous codons, thus, by comparing the differential use of codons between different genes it is possible to determine the relatedness of the genes. Two important caveats must be stressed prior to this form of analysis. Firstly the different amino-acid composition of proteins, particularly in small proteins, may bias the initial determination of codon usage. Secondly, within a given organism, different genes have been

	Tn7int	*dhfr*	*sat*	*aadA*	*tnsE*	*tnsD*	*tnsC*	*tnsB*	*tnsA*	E. coli(high	E. coli(low)	Tn21 sul	Tn21 int
Tn7 int													
dhfr	**1.7**												
sat	3.09	4.63											
aadA	**2.02**	**2.49**	**2.73**										
tnsE	**2.25**	**1.41**	4.34	**2.3**									
tnsD	**1.85**	**1.64**	**2.71**	4.22	**1.11**								
tnsC	**1.82**	**1.71**	4.76	**2.51**	**1.13**	**1.46**							
tnsB	**2.06**	**1.58**	4.07	**1.83**	**0.58**	**1.2**	**0.71**						
tnsA	**2.79**	**2.14**	4.97	3.3	**1.18**	**1.84**	**0.93**	**0.86**					
E. coli (high	5.42	6.63	5.08	4.23	7.62	6.1	8.28	7.25	7.93				
E. coli (low)	**1.36**	**1.91**	**2.63**	**0.73**	**1.81**	**1.38**	**1.75**	**1.29**	**2.06**	3.39			
Tn21 Sul	5.37	4.41	**2.48**	3.98	5.63	5.83	5.22	5.24	5.97	4.23	**2.36**		
Tn21 int	5.3	4.97	3.47	4.29	6.34	5.85	6.05	5.92	7.01	3.6	**2.55**	**1.8**	

Fig. 3.16 Comparison of Codon Usage in Tn7 genes
This table lists D-square values calculated for codon usage in Tn7 genes. Values were determined using the GCG program, CORRESPOND. EcoHigh and Ecolow represent the codon frequency tables in the GCG database for high and low expression *E. coli* genes. Boxed values suggest a statistically significant similarity in codon usage.

shown to use different subsets of codons depending on the level of expression of the gene.

A previous work compared the codon usage in the integron conserved regions and the inserted gene cassettes. The methodology used was to determine the frequency of use of codons in the various reading frames by using the GCG program CODONFREQUENCY. These were then compared using the GCG program CORRESPOND (Devereux, Haeberli et al. 1984). D-square values are calculated by this program, values lower than 3 are taken to suggest a common origin for the genes. The conclusion of this work on the In0 integron was that the integrase and *sulI* open reading frames were derived from the same or similar organisms, the inserted genes however displayed different codon biases suggesting they were derived from diverse origins (Bissonnette and Roy 1992).

A similar analysis was attempted for the Tn7 open reading frames, codon usage in these genes were compared with each other and also the *E. coli* low expression and high expression codon frequency tables, and the codon usage table for the Tn21 family integrase and sulI genes. The D-square values determined in this study are shown in Fig. 3.16.

These results show similar codon usages in all 5 *tns* genes, in addition to this the tns genes also have similar codon biases to *dhfr* and *aadA* genes, the Tn7 integrase pseudogene and the table of codon usages of low expression level *E. coli* genes. The Tn7 *sat* gene had similarity to the *E. coli* low expression genes, Tn21 *sul* gene and the *aadA* gene. The codon bias of the *sat* gene also had similarity to that of *tnsD* but not the other *tns* genes. Bearing in mind the caveats mentioned above these results suggest the *sat* gene originated from a different organism from the other Tn7 genes.
The most intriguing result from this analysis is in the comparison between the codon usage of the Tn21 integrase gene and the Tn7 integrase pseudogene. The D-squared value of 5.30 suggests that these genes arose from different organisms. The Tn7 integrase gene is interrupted by a stop codon and therefore presumably non-functional. It is possible that the difference in codon usage reflects the accumulation of mutations in the Tn7 pseudogene. It is interesting to note that the Tn1826 integrase pseudogene has complete identity with the Tn7 pseudogene. Thus, unless

Tn7 and Tn1826 diverged very recently it might be expected that different random mutations would have arisen in the two elements.
It is possible that the integron like structure in Tn7 either evolved independently from the Tn21 family or that the two structures diverged early in the evolution of integrons.

No homologies between Tn7 and the non-coding portion of the Tn21 integron family 5' conserved region exist. In addition, in the Tn21 family integrons the GTTRRRY recombination site is approximately 140 nt from the start of the integrase open reading frame, in Tn7 the *dhfr* gene is integrated at a GTTRRRY site 340 nt from the start of the integrase pseudogene. Therefore, in addition to possessing a different integrase gene the structure of the Tn7 integron is distinct from the Tn21 family. However, until more members of the Tn7 family are identified and characterised it is impossible to draw firm conclusions about the relationships between the Tn7 and Tn21 integrons.

It is clear, however, that the Tn7 integron does not lie within the inverted repeat structures known as Brown's repeats. Therefore if the Tn7 integron did insert in a primordial Tn7', isolating the internal repeats in the process, either considerable further re-arrangements have occurred to delete portions of the initial integron, or the integron inserted by another mechanism.

Chapter 4

Tn7 Transposition *in vivo* and *in vitro*

4.1 Introduction

In order to address questions regarding the biochemical mechanisms of the Tn7 transposition reaction it is neccessary to develop an *in vitro* transposition assay. Two approaches to developing such an*in vitro* system suggest themselves. Firstly, fractions containing individual Tns proteins could be isolated and subsequently used to partially purify the various Tns proteins. Such fractions could then be combined and used in assays. Alternatively it may be possible to exploit the operon-like structure of *tnsA-E* to co-ordinately overexpress all of the proteins. This latter approach would have the advantage that only one extract would be required. It is also possible that this method will ensure that the various proteins will be present in the appropriate relative amounts. Previous work using p*lac* and p*tac* transcriptional fusions has shown that, of the Tns proteins, only TnsB can be over-expressed to a level such that it is visible on Coomassie blue stained Laemmli gels. It was deemed important to over-express the proteins to a readily visible level as no biochemical assay existed for TnsA and C. The only way purification of these proteins could be monitored would be to follow their fate on Laemmli gels.

The system of choice to achieve such high expression is that developed by S. Tabor, using bacteriophage T7 RNA polymerase and the T7 gene 10 promoter.
Induction of expression of a gene cloned downstream of the gene 10 promoter is modulated by controlling the level of T7 RNA polymerase present in the cell. Theoretically, as there is no cross reaction between *E. coli* RNA polymerase and T7 promoters, if there is no T7 pol in the cell then there should be no expression of the target gene(s). As detailed below this was found not to be the case.
Tabor has developed a two plasmid system to allow inducible expression of cloned genes (Tabor and Richardson 1985). The gene of choice is cloned into a ColE1 based plasmid which has the T7 gene 10 promoter upstream of a polylinker. The T7 RNA pol gene is present on a second, compatible, plasmid, pGP1-2, under the control of the lambda pL promoter. pGP1-2 also encodes the temperature sensitive cI_{857} allele of the lambda repressor, under the control of placUV5. Thus, induction of the target gene is obtained by shifting the growth temperature of the culture from 30°C to 42°C.

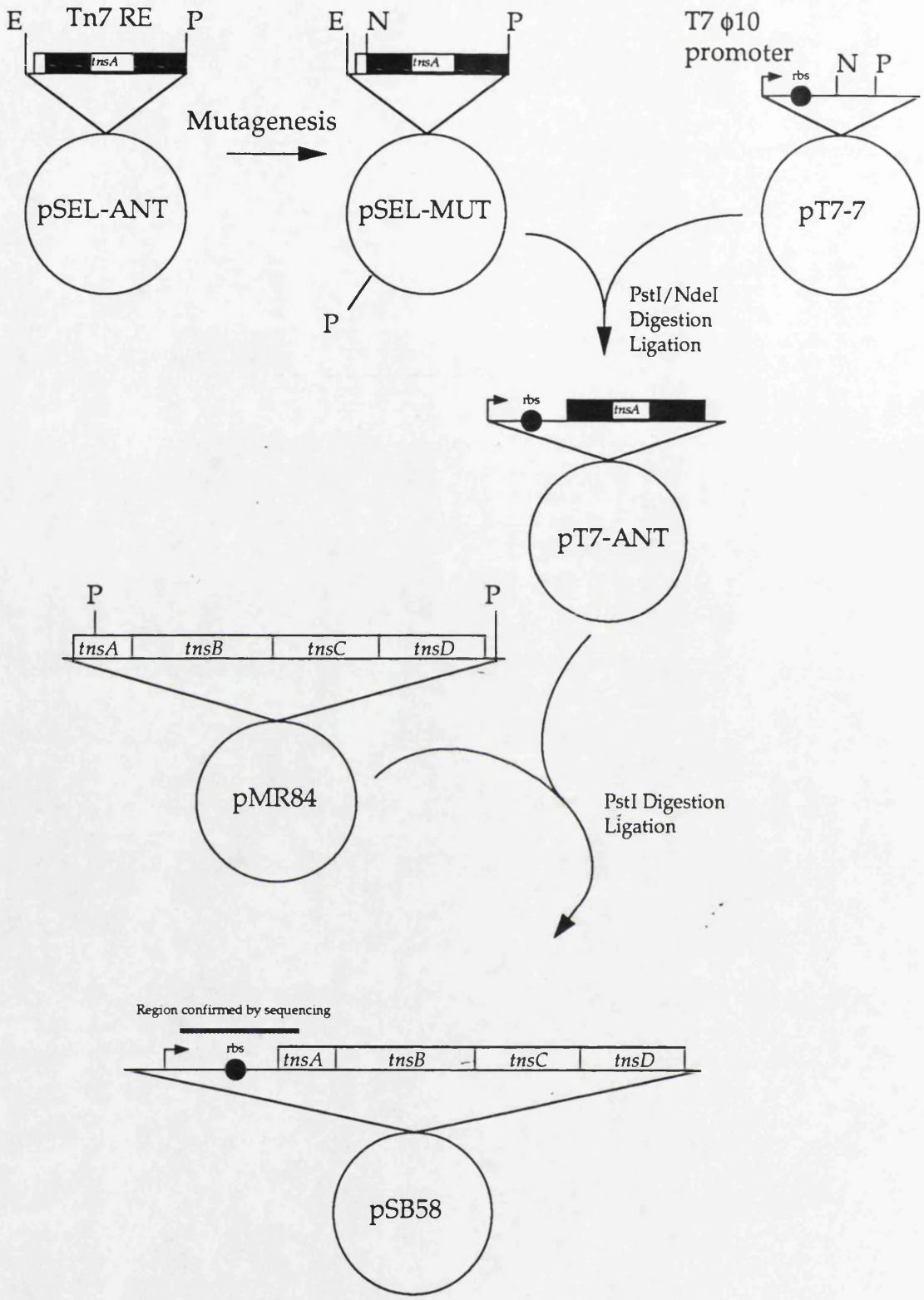
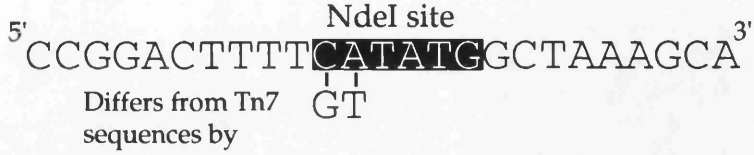


Figure 4.1

Construction of pSB58 as described in text.
E=EcoRI site
N=NdeI site
P=PstI site
rbs=Ribosome Binding Site

Fig. 4.2 Sequence of oligonucleotide used in creation of pSEL-MUT.



Tabor has developed expression vectors for transcriptional fusion, pT7-3,4,5,6, and translational fusion, pT7-7. pT7-7 has a polylinker with an NdeI site (CATATG) optimally placed adjacent to a strong Shine-Dalgarno ribosome binding site.

Two strategies were adopted to over-express *tnsABCD*, using both transcriptional and translational fusions.

4.2 pSB58 Construction

The translational fusion was constructed as shown in Fig. 4.1.
Site directed mutagenesis was used to introduce a novel NdeI site at the TnsA start codon, the mutagenesis method was that developed by Promega for use with their Altered Sites kit (Lewis and Thompson 1990). The 5' half of the *tnsA* gene was subcloned as an 600nt EcoRI-PstI fragment into the phagemid pSelect-1 to give pSel-ant. This plasmid has a tetracycline resistance gene and an ampicillin resistance gene which contains a frame-shift mutation. Single stranded DNA of the recombinant phagemid was prepared as detailed in Materials and Methods. An oligo-nucleotide (Fig. 4.2.) complementary to the region surrounding the presumptive *tnsA* start codon, but introducing an NdeI site was synthesised on an Applied Biosystems PCR mate DNA synthesiser. The oligo-nucleotide was phosphorylated and annealed to the single stranded pSel-ant, as was the ampicillin repair oligo-nucleotide. T4 DNA polymerase was used to extend from the 3'-OH of the oligo-nucleotides and DNA ligase added to covalently close the nascent strand. This was transformed into a mut-S strain of *E. coli* which was then subjected to selection in ampicillin containing medium. After overnight growth plasmid DNA was isolated by the alkaline lysis mini-prep method and used to transform competent JM109. Transformants were plated on L-agar containing ampicillin and tetracycline. Individual colonies were used to inoculate cultures and DNA was mini-prepped. Plasmids which contained the required mutation were identified by restriction with NdeI. Eight out of ten colonies tested contained the extra NdeI site. The mutant plasmid was designated pSel-mut.
The plasmid was then digested with NdeI and PstI, the novel ~500nt fragment isolated from agarose gel and ligated to NdeI-PstI digested pT7-7, this gave pT7-ant. pT7-ant was then cut with PstI and ligated to the ~6kb

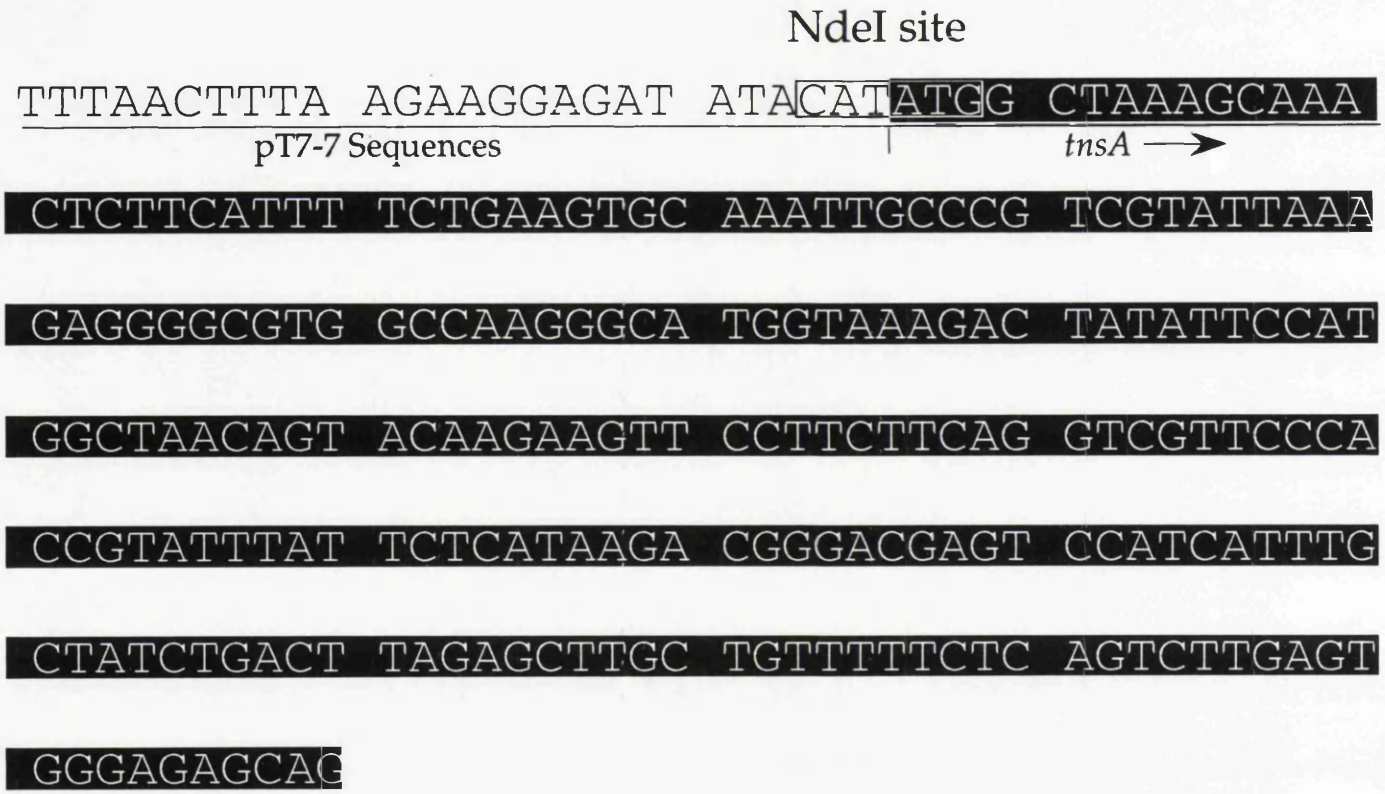
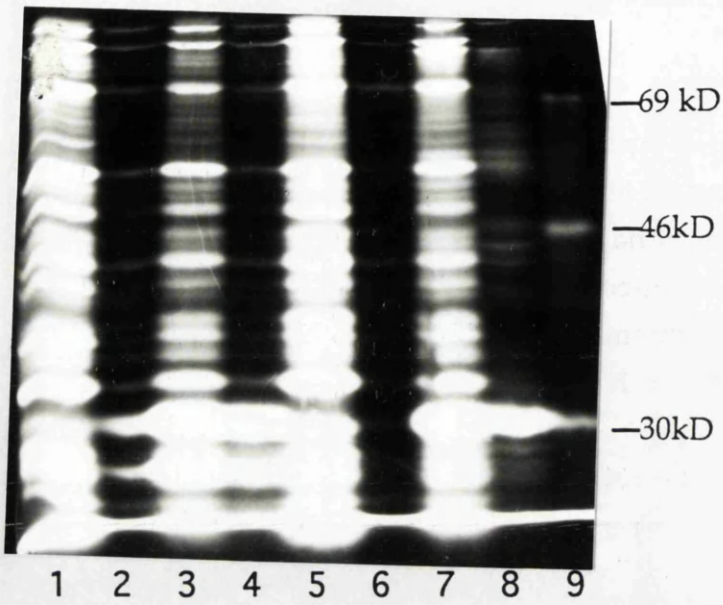


Fig. 4.3 Sequence derived from pSB58 clone as described in the text.
pT7-7 vector sequence in normal type, *tnsA* sequence in inverse type.



Lanes 1&2: BL21 pGP1-2 pT7-5

Lanes 3&4: BL21 pGP1-2 pAG411

Lanes 5&6: BL21 pGP1-2 pT7-5

Lanes 7&8: BL21 pGP1-2 pSB58

Lane 9: Markers

Lanes 2,4,6,8 Rifampicin treated cultures, as described in Materials and Methods. Lanes 1,3,5,7 labelled without rifampicin treatment.

Fig. 4.4 Autoradiograph of a Laemmli gel of whole cell extracts prepared following protocol for ^{35}S labelling of cells, as described in the text.

tnsBCD containing fragment of PstI digested pMR84. The orientation of the PstI fragment in pT7-ant was checked by restriction with HindIII and BamHI (Data not shown). The plasmid which restored the *tnsA* open reading frame was named pSB58.

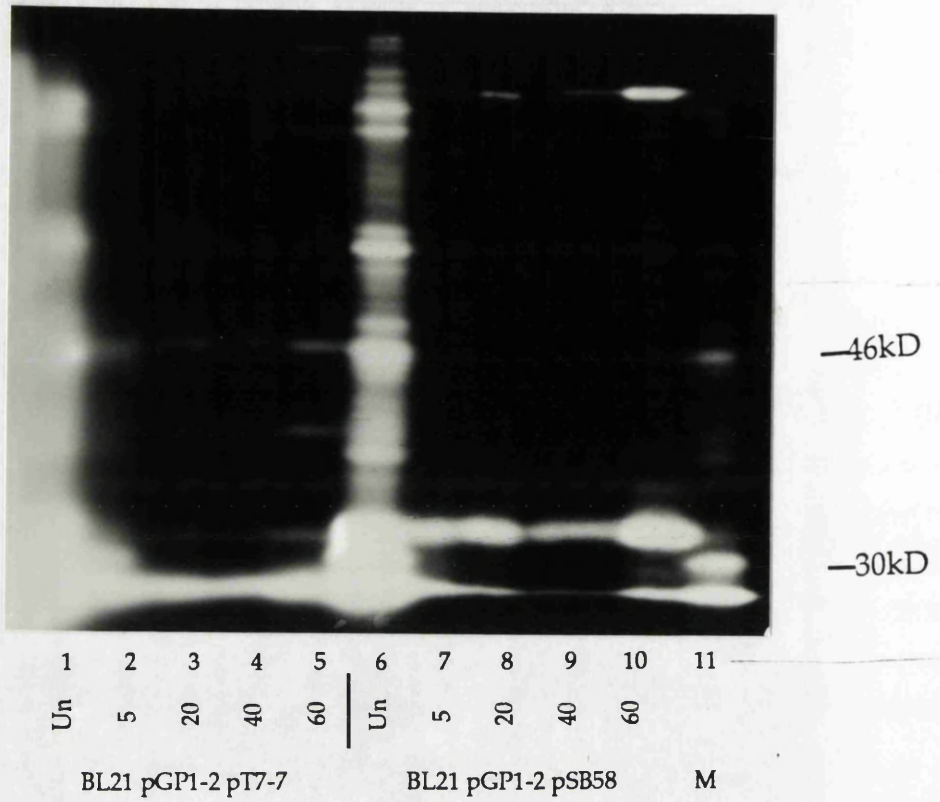
pSB58 DNA was prepared and sequenced using a primer complementary to the T7 gene10 promoter. This primer was a kind gift from Sally Rowland. This step was carried out to ensure that the mutagenesis procedure had only introduced the required mutation. The resultant sequence encompassed 250nt from the primer, corresponding to pT7-7 vector sequence and positions 134 to 358 in wild type Tn7 (Fig. 4.3). No mutations other than those creating the NdeI site were found. Although the sequence did not cover the entire fragment cloned in pSel-mut it was felt it was reasonable to assume there were no second site mutations created during the mutagenesis.

4.3 pSB84 Construction

The transcriptional fusion pSB84 was made by cloning a EcoRI-ClaI fragment containing *tnsABCD*, from pMR9, into EcoRI-ClaI cut pT7-5. This construct contains all the Right end 22 base pair motifs.

4.4 Protein Expression

A major advantage of the T7 based expression systems is that, unlike the *E coli* RNA polymerase, the T7 RNA polymerase is insensitive to the antibiotic rifampicin. By inducing expression of T7 RNA polymerase for 15 minutes then supplementing the growth medium with Rifampicin this will ensure that the only transcription occuring in the cell is by T7 RNA polymerase. Bacterial mRNA's generally have a short half life, thus a short time after exposure to rifampicin the only gene being expressed will be that transcribed by T7 RNA polymerase. This has the advantage of reducing competition with endogenous mRNA's for ribosomes and charged tRNA's. This may be particularly important if the target gene possesses a large number of rare codons. A second advantage of this system is that it enables unique radiolabelling of nascent target proteins using ^{35}S Methionine. This method is, of course, dependant on the presence of methionine in the target protein. Tns A has 3 methionines,



Lanes 1-5: BL21 pGP1-2 pT7-7
Lanes 6-10: BL21 pGP1-2 pSB58
Lane 11: Markers

Lanes 1 and 6: Uninduced
Lanes 2 and 7: 5 minutes chase
Lanes 3 and 8: 20 minutes chase
Lanes 4 and 9: 40 minutes chase
Lanes 5 and 10: 60 minutes chase

Fig 4.5 :
Pulse/chase experiment with BL21 pGP1-2 pT7-7and BL21 pGP1-2 pSB58.
This experiment was performed as descibed in materials and methods.
The whole cell extracts were electrophoresed on a 10 % gel.

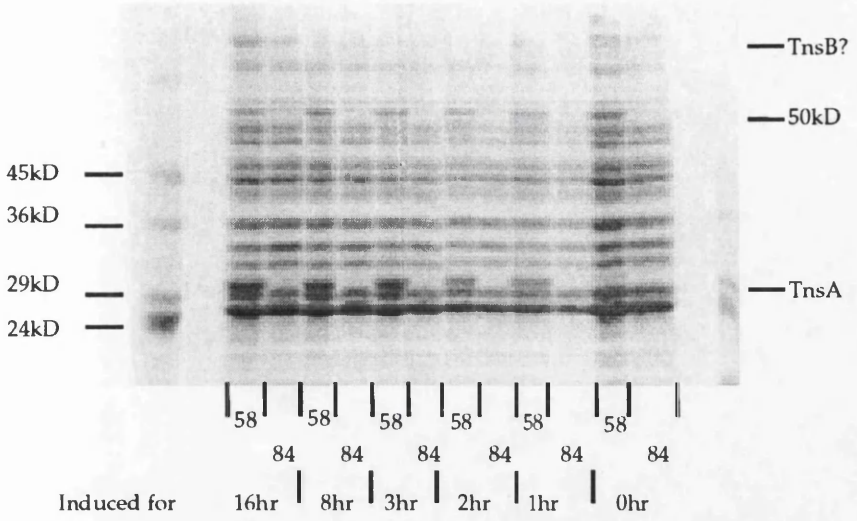


Fig. 4.6 Coomassie stained 10% Laemmli gel of whole cells of induced CSH50 pGP1-2 pSB84 and CSH50 pGP1-2 pSB58.

TnsB has 8, TnsC has 14 and TnsD has 4. Figure 4.4 shows the results of radiolabelling experiment using CSH50 pSB58 pGP1-2. This figure shows a high level of expression of a ~33 kD species, TnsA, a ~80kD species, Tns B, and several smaller species. It is not possible to absolutely ascribe any of the novel 50-65 kD species to *tnsC* or *D* gene products because of the extensive proteolysis undergone by TnsB. Nevertheless, it is very clear that a hitherto unattainable level of TnsA expression has been achieved. Fig. 4.5 shows a pulse chase experiment using the strains BL21 pGP1-2 pT7-7 and BL21 pGP1-2 pSB58. In the latter strain both TnsA and TnsB appear to be stable over the time course of the assay, however in the final time point a number of novel, but faint bands appear, these species range in size from 40-60 kD

The rifampicin/^{35}S Methionine method is useful for identifying the size of target proteins. However, because it preferentially labels these species it is not possible to draw conclusions as to the level of expression of these proteins. Thus in order to determine the expression of the *tns* gene products relative to endogenous cellular proteins the following experiment was carried out.

Cells containing either pSB58, pSB84 or pT7-5 in addition to pGP1-2 were induced and 400µl samples collected at 1 hour intervals. These cells were lysed in Laemmli loading buffer and electrophoresed on a 10% Laemmli gel. After electrophoresis the gel was stained with Coomassie. The result is shown in Fig. 4.6.

Growth curves were determined for CSH50 pGP1-2 pSB58 and 84. They clearly demonstrate a rapid slowing of the growth rate on induction of the constructs. No drop in the OD600 of the cultures, indicative of cell death, is observed (data not shown).

4.5 Transposition *in vivo*

In order to confirm that the two constructs supported transposition of a Tn7-1 *in vivo*, mate-out assays were carried out between CSH50::Tn7-1 (pEN300, pGP1-2, pSB58/84) and DS916. The results are shown in Fig. 4.7. The figures stated are the average of 5 independent assays for each plasmid.

Strain	Apparent Transposition Frequency Following Induction for 0hrs	Apparent Transposition Frequency Following Induction for 2hrs	Conjugation measured as exconjugants recovered per 1ml initial donor culture Following Induction for 0hrs	Conjugation measured as exconjugants recovered per 1ml initial donor culture Following Induction for 2hrs
1) CSH50::Tn7-1 pEN300 pGP1-2 pSB58	1.7×10^{-4}	3.0×10^{-6}	4.8×10^{6}	1.2×10^{8}
2) CSH50::Tn7-1 pEN300 pSB58	1.8×10^{-4}	1.3×10^{-4}	2.0×10^{7}	3.3×10^{8}
3) CSH50::Tn7-1 pEN300 pGP1-2 pSB84	3.0×10^{-2}	8.0×10^{-2}	5.6×10^{6}	4.4×10^{8}
4) CSH50::Tn7-1 pEN300 pSB84	2.0×10^{-2}	2.5×10^{-2}	3.1×10^{7}	1.6×10^{9}

Fig. 4.7 Apparent transposition frequencies measured by mate-out assay of the above strains, as detailed in Chapter 2, the recipient strain was DS916. Induction was attained by heating the culture to 42°C for 25 mins. then incubating at 37 °C.

The figures stated for conjugation are for exconjugants per 1ml of initial donor and recipient strains.

It is important to stress that in the absence of viable cell counts for the donor strain these figures do not represent rates of conjugation and thus it is not possible to draw conclusions about the relative rates of conjugation in the various strains.

Assays were also carried out for CSH50::Tn7-1 pSB58 or 84, in the absence of pGP1-2, and thus T7 RNApol. Transposition was supported in these cells. This presumably reflects the expression of *tnsABCD* from the endogenous Tn7 promoter, P1, in the Tn7 right end, however pSB58 has no associated promoter other than the T7 promoter. Presumably the basal expression of this construct is caused by read-through from promoters at the plasmid origin of replication.

While this work was in progress a similar approach to over-express *tnsABCD* was being employed by Conrad Lichtenstein's group in London (Flores, Cotterill et al. 1992). The pET3c vector (Studier, Rosenberg et al. 1990) developed by Studier was used to create translational fusions to the start codons of *tnsA, B, C and D*. The source of T7 RNA polymerase in their experiments was from the DE3 prophage containing the gene for T7 RNA pol under control of plac UV5. Expression of the target gene is, therefore, IPTG inducible. An additional level of control can be exerted by the presence of T7 lysozyme, encoded on pACYC184 derivatives; pLysS, expressing low levels of T7 lysozyme and pLysE, expressing higher levels. T7 lysozyme binds to T7 RNA polymerase and inhibits transcription, thus any basal activity of T7 RNA polymerase is reduced in cells containing either of these plasmids. It should be noted that it was found impossible to transform BL21 (DE3) pLysE with pSB58, presumably even the very low basal expression of this construct was highly toxic to these cells. It was however found to be possible to transform BL21 with pSB58 and then subsequently introduce pGP1-2.

The results of Flores *et al* show over-expression of the individual *tns* genes on induction of T7 RNA polymerase. In order to confirm that these proteins are functional a novel construct was made which contained all of the four genes each with its own T7 promoter and ribosome binding site. The order of genes in the construct was thus:-

T7 promoter rbs *tnsA* T7 promoter rbs *tnsB* T7 promoter rbs *tnsC* T7 promoter rbs *tnsD* T7 terminator.

Flores *et al* assayed this construct for its ability to complement transposition of a mini-Tn7 element from a plasmid to the chromosomal *attTn7* site. The transposition event was assayed by Southern blot analysis

of restricted total DNA isolated from cells containg the appropriate plasmids. Transposition was observed to be supported by the TnsATnsBTnsCTnsD construct. Intriguingly despite the fact that no Tns gene products are visible prior to induction it was observed that "transposition appears more or less as efficient in an uninduced culture as in cultures where TnsA, B, C and D expression were induced." It was also noted that the nature of this assay did not permit accurate comparison of transposition frequencies. Therefore this group also detected transposition even in the absence of induction.

The results I obtained for pSB84 suggest that as the expression of the *tnsABCD* genes is induced by T7 RNA polymerase there is a rapid increase in the apparent rate of transposition. This presumably correlates with an increase in expression of the Tns proteins.

However, the results of the assays of pSB58 supported transposition are not as readily explained. These indicate a reduction in the level of recovery of pEN300::Tn7-1 over the time course of the assay. It should be stressed that the reduction of recovery of pEN300::Tn7-1 could be due to effects in either the donor or recipient cells. Due to the absence of Tns functions in the recipient and because such effects are not seen with other TnsABCD plasmids in the donor, the following work is based on the assumption that the phenomenon occurs in the donor cell.
The drop in recovery of pEN300::Tn7-1could be explained by the selective loss of pEN300::Tn7-1 from donor cells or by the selective blockage of conjugative transfer of these molecules. It should be noted that there is no effect on the transfer rate of the pEN300 plasmids. A number of potential mechanisms whereby the recombinant pEN300::Tn7-1 could be lost or sequestered are possible.

The over expression of the *tns* gene products could be leading to very high levels of transposition, transposition could occur more rapidly than division of the host replicon. As Tn7 transposes conservatively this would lead to the loss of donor molecules from the donor cells. Another related possibility is that nicks are produced at the ends of the Tn7-1, if nicks are made on opposite strands at either end of the molecule it will be impossible to replicate pEN300::Tn7-1, thus preventing transfer of these molecules, see Fig. 4.8. A problem with this interpretation is that in the

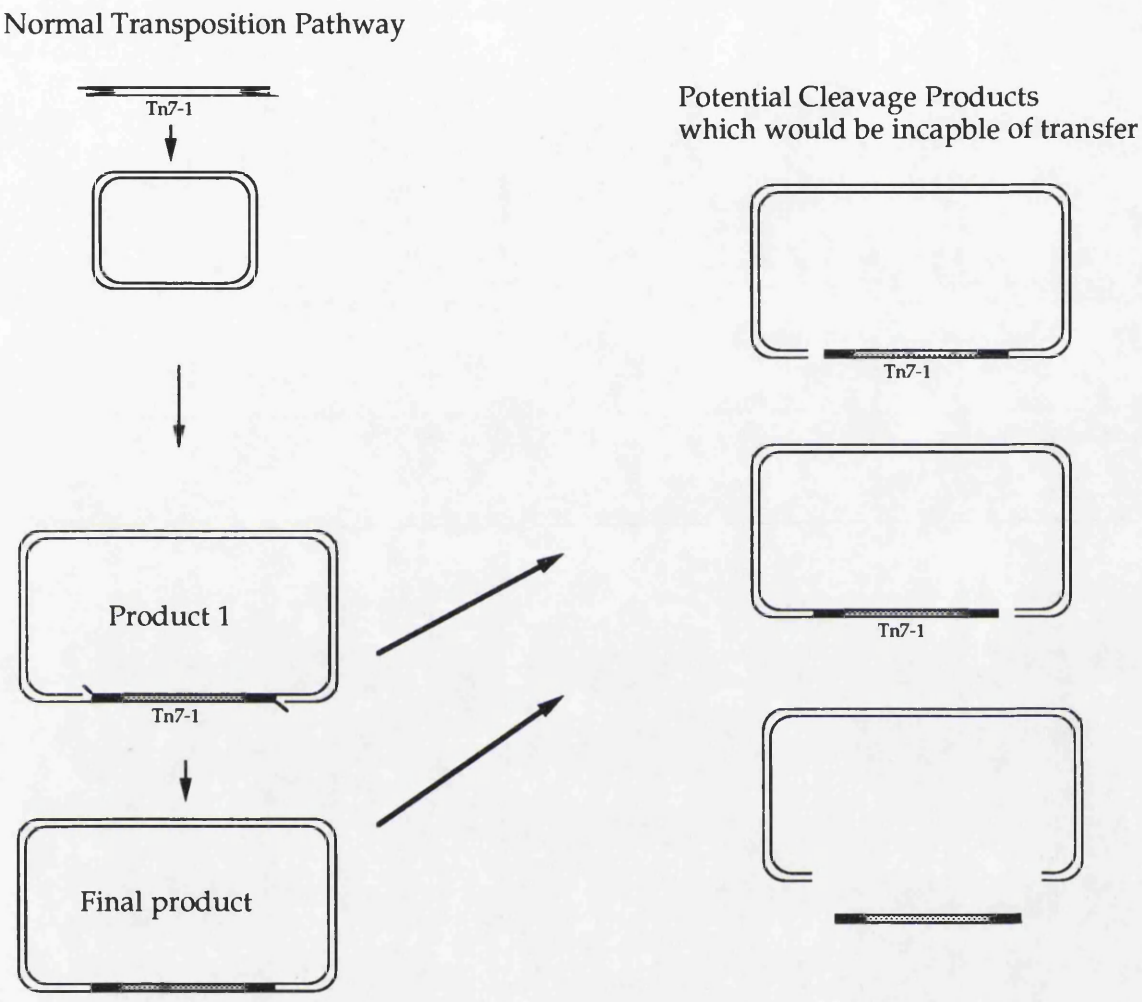


Fig.4.8 Potential points in the transposition pathway where conjugative transfer of pEN300::Tn7 could be prevented. In addition to the production of the broken molecules shown on the right half of the diagram the initial product, Product 1, will be incapable of transfer as both strands are nicked.

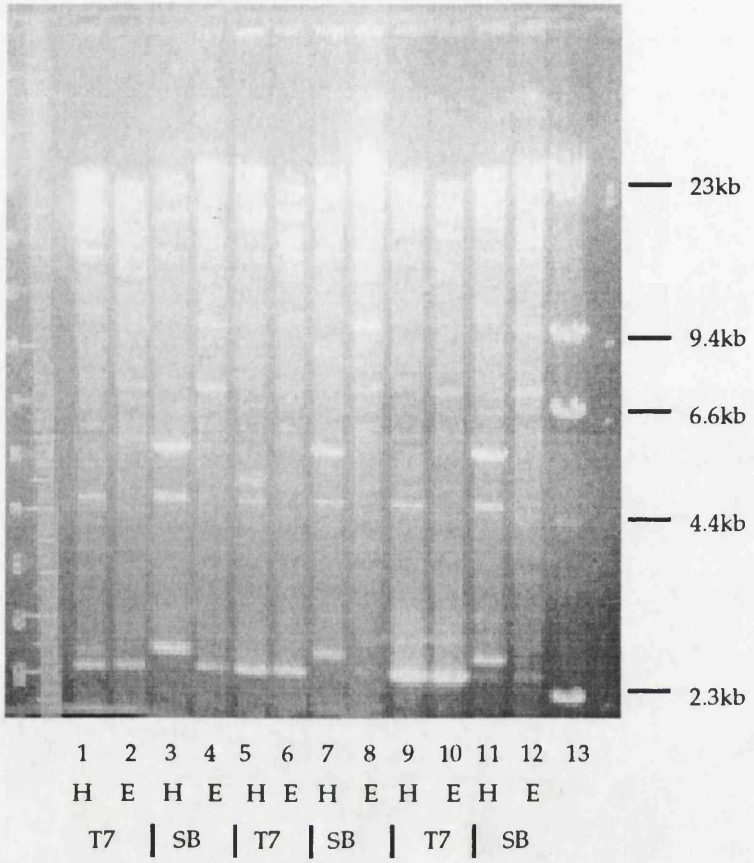


Fig. 4.9a Ethidium bromide stained 0.7% agarose gel.Lanes are alternately HindIII and EcoRI digests of chromosomal DNA prepared from the following strains.

Lanes 1&2: Un-induced CSH50 pGP1-2 pT7-7

Lanes 3&4: Un-induced CSH50 pGP1-2 pSB58

Lanes 5&6: CSH50 pGP1-2 pT7-7 induced for 30 mins.

Lanes 7&8: CSH50 pGP1-2 pSB58 induced for 30 mins.

Lanes 9&10: CSH50 pGP1-2 pT7-7 induced for 2 hrs.

Lanes 11&12: CSH50 pGP1-2 pSB58 induced for 2 hrs.

Lane 13: Lambda HindIII digest

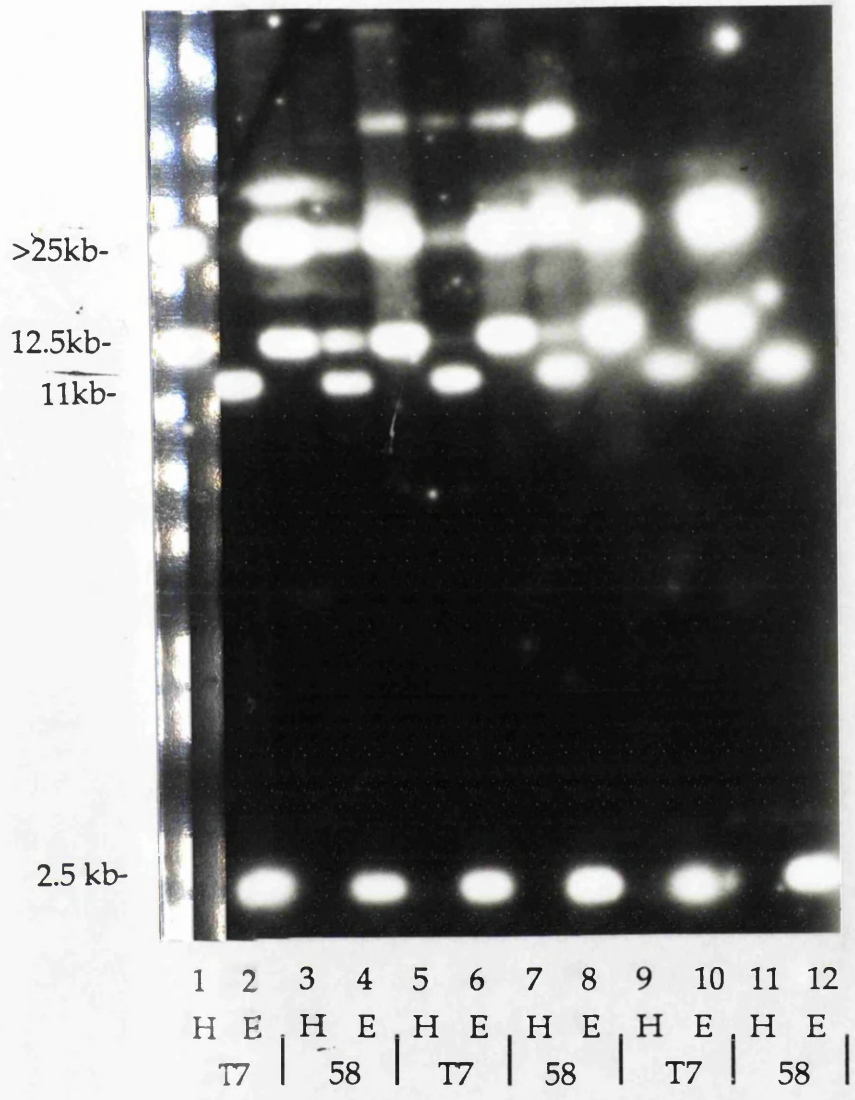


Fig. 4.9b Southern blot of agarose gel in Fig. 9a.
Probed with probe specific for *attTn7* sequences.
Lanes are alternately HindIII and EcoRI digests of
chromosomal DNA prepared from the following strains.

Lanes 1&2: Un-induced CSH50 pGP1-2 pT7-7
Lanes 3&4: Un-induced CSH50 pGP1-2 pSB58
Lanes 5&6: CSH50 pGP1-2 pT7-7 induced for 30 mins.
Lanes 7&8: CSH50 pGP1-2 pSB58 induced for 30 mins.
Lanes 9&10: CSH50 pGP1-2 pT7-7 induced for 2 hrs.
Lanes 11&12: CSH50 pGP1-2 pSB58 induced for 2 hrs.
Lane 13: Lambda HindIII digest

donor cells the chromosomal *attTn7* site is occupied. If Tn7-1 is transposing into pEN300 then undergoing a second transposition event prior to plasmid replication, then where is it moving to? The *in vitro* studies of Bainton et al have revealed that Tn7 transposition is dependant on the presence of target site. As pSB58 only encodes *tnsABCD* the most likely explanation is that pseudo att sites in the bacterial chromosome are being used. However, insertion into pseudo att sites is only observed at very low frequencies.
An alternative explanation could be that the abnormal relative abundancies of TnsA,B,C and D, combined with their gross over-expression, allow an un-coupling of transposon excision from insertion.

A more mundane explanation could be that the overexpression of the proteins is so high that they are forming insoluble aggregates within the cell. This could lead to a physical sequestration of pEN300::Tn7-1 if TnsB retains its DNA binding activity.

4.7 Double Strand Breaks?

In order to test the first hypothesis, that excessively high levels of transposition are occurring, an attempt was made to identify the proposed double strand breaks created by transposition. Studies of both Tn10 transposition and double strand break mediated meiotic recombination in yeast have allowed the detection of such breaks (Haniford, Chelouche et al. 1989 ; Cao, Alani et al. 1990).
The detection method in the above works was restriction digestion of total DNA, followed by Southern blotting

A similar analysis was attempted for pSB58 mediated Tn7-1 transposition by preparing total DNA from induced and non-induced CSH50::Tn7-1 (pEN300, pGP1-2, pSB58). The DNA was restricted with HindIII or EcoRI and electrophoresed on a 0.7% agarose gel. The DNA was transferred to Hybond-N membrane by "Posi-Blotting" (Stratagene). The membrane was pre-hybridised and then hybridised with probe prepared by Random Priming (Boeringer) of the 1kb *attsite* containing EcoRI fragment of pMR12.
In the southern blot shown in Fig. 4.9 , genomic DNA from the un-induced and induced CSH50::Tn7-1 pEN300 pGP1-2 pSB58 and

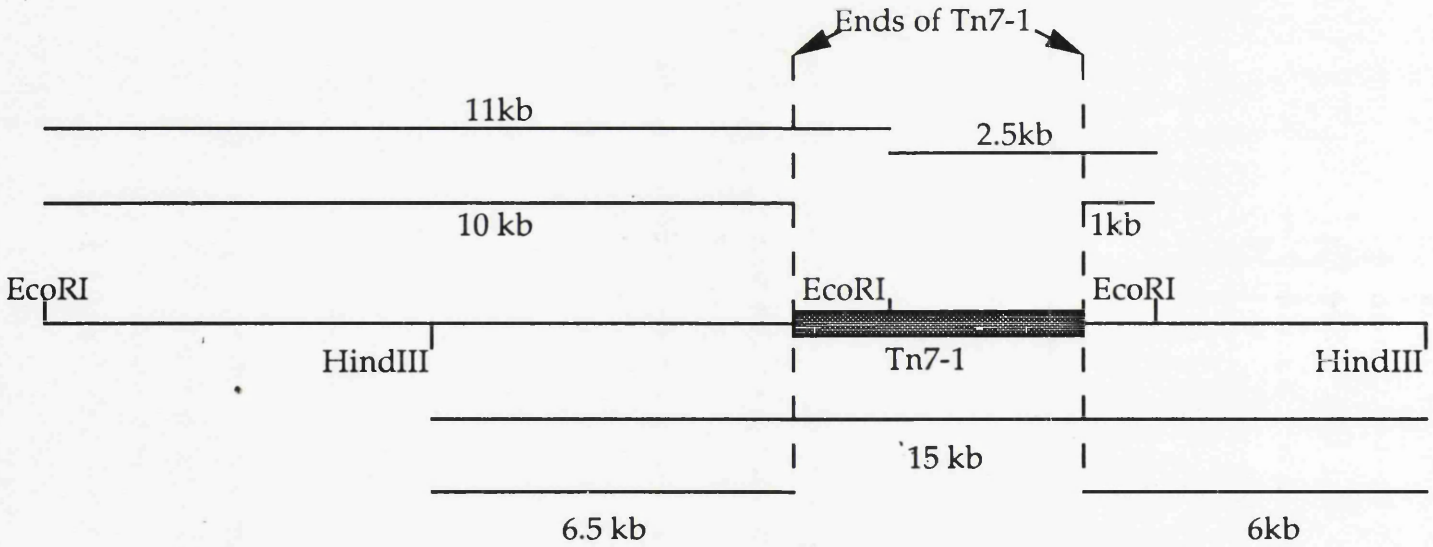


Fig 4.10: Structure of region surrounding the *attTn7* site in the *E. coli* chromosome

CSH50::Tn7-1 pEN300 pGP1-2 pT7-7 cells was prepared as detailed in Materials and Methods. The DNA was digested with EcoRI or HindIII and electrophoresed and transferred to Hybond-N as above. The membrane was then hybridised with the attsite probe. The predicted hybridisation signals are at 35 and 37 kb for HindIII digested pEN300 and pEN300::Tn7-1 respectively, the chromosomal attTn7::Tn7 locus gives a 15 kb HindIII fragment. If double strand breaks were produced at the chromosomal site the predicted HindIII-end fragments would be of approximately 6.5kb and 6kb. The EcoRI fragments would be 1kb and 10kb (See Fig. 4.10). The results of this and several other blottings did not show any novel bands appearing following induction of the pSB58 construct. On the basis of this assay there does not appear to be a sufficient level of transposition to explain the "loss" of pEN300::Tn7-1.

A second prediction which could be made, assuming high levels of transposition, is that the generation of double strand breaks would be deleterious to the host cells. This may manifest itself in a comparison of the growth curves following induction of pSB58 in isogenic cells, differing only in the presence or absence of Tn7-1. This experiment was carried out using CSH50(+/- ::Tn7-1) pGP1-2 pSB58 pEN300. It can be seen from Fig. 4.11 that following induction the growth curves of the various strains demonstrate a rapid drop in the growth rate of the cultures. However, the effect is apparently the same regardless of the presence of Tn7-1 in the strains. Thus, a Tn7-1 containing strain is not more sensitive to induction of pSB58 than a strain lacking the element.

A major problem with the interpretation of data derived from the mate-out assay is that the assay is unable to discriminate between a single transposition events which occurs early in culture growth and multiple events which occur late in growth. Thus the assay samples a population average.

Presumably the apparent transposition frequencies measured from pSB84 and pSB58 containing cells reflect this combination of events. It is apparent that the measured transposition frequency in non-induced pSB58 containing cells is at least 10 fold lower than in the non-induced pSB84 containing cells.

This lower apparent rate may simply reflect a decreased transposition rate throughout culture growth. It could alternatively be brought about by

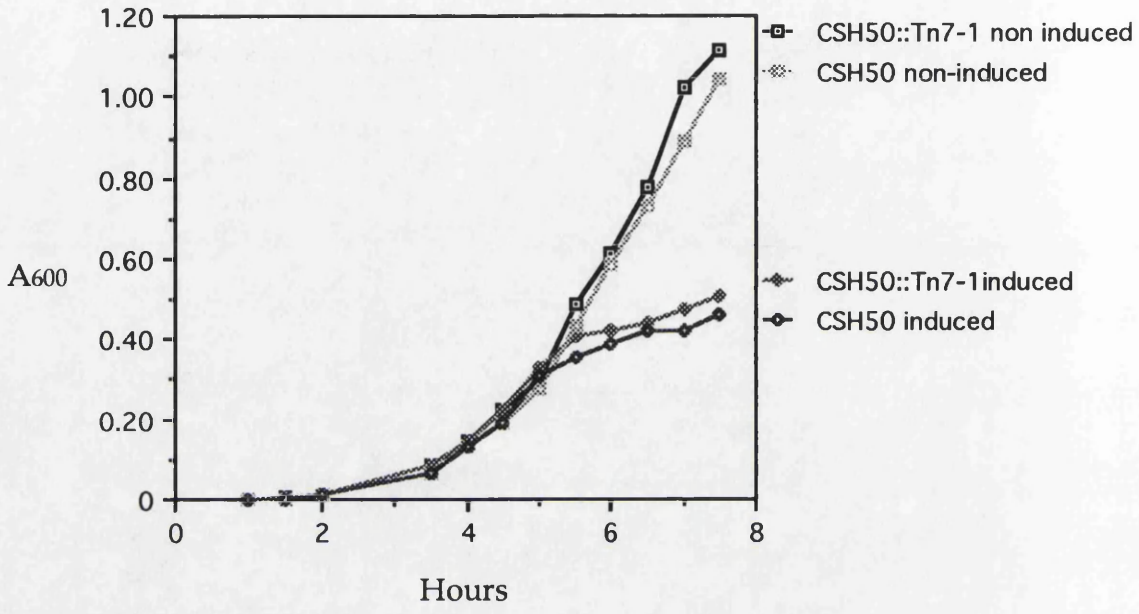


Fig. 4.11 A comparison of the growth curves of induced and non-induced CSH50::Tn7-1 pEN300 pGP1-2 pSB58 and CSH50 pEN300 pGP1-2 pSB58

500µl aliquots of overnight cultures of both strains were used to innoculate 50mls L-broth. These cultures were grown at 30° C for 5 hours then divided into 2 equal volumes one aliquot of each strain was induced for 25 minutes at 42° C, then grown at 37° C. The absorbance at 600 nm of 1ml samples of the various cultures was measured a the indicated time points.
For clarity the graph was plotted with a linear vertical scale.

Donor Strain	Exconjugants/ml initial cultures		
	Non-induced	Induced	Induced / Noninduced
CSH50::Tn7-1 pEN300::Tn7-1 pGP1-2 pSB58	6.8×10^7	3.2×10^8	4.7
CSH50::Tn7-1 pEN300 pGP1-2 pSB58	2.0×10^8	5.6×10^8	2.4

Fig. 4.12 Congugation supported by induced and non induced CSH50::Tn7-1 pEN300::Tn7-1 pGP1-2 pSB58 and CSH50::Tn7-1 pEN300 pGP1-2 pSB58.

Liquid cultures were grown at 30° C over night then 2, 10ml aliquots of each strain were separately added to 10 ml fresh pre-warmed L-broth. One aliquot of each strain was then induced 42° C, 25 mins, 37° C 2 hrs, the other aliquot was grown at 30 °C. Conjugation was then carried out with DS916 as recipient, matings with non-induced cultures were performed at 30°C, matings with induced cultures were performed at 37° C. The figures represent the average of two experiments.

It is important to stress that in the absence of viable cell counts for the donor strain these figures do not represent rates of conjugation and thus it is not possible to draw conclusions about the relative rates of conjugation in the various strains.

instability of pSB58 mediated Tn7-1 insertion into pEN300. If the pSB58 generated pEN300::Tn7-1 are unstable then it is possible that the only pEN300::Tn7-1 detected in the mate out assay are those which arise during or just prior to conjugation. Such instability could be mediated by excision of the element, as described above. The studies of Tn7 transposition *in vitro* have shown that the initial insertion product has only the 3' ends of the element covalently linked to the recipient molecule DNA, short gaps exist at the 5' ends of the element (Bainton, Gamas et al. 1991). These gaps are then presumably repaired by host DNA polymerase. Thus the initial product would require a single strand cleavage, the reverse of the integration, to release the inserted Tn7-1. The later, repaired product would require a double strand cut. It is possible that the abnormal ratios of TnsA,B,C and D may favour excision of either of these products. The induction of the construct may further enhance this process.
Alternatively, pSB58 may stabilise the gapped intermediate. Such a doubly gapped plasmid would be unable to replicate. Thus, both its propagation and conjugation would be inhibited.

To test if pSB58 can mediate the apparent retension of fully covalently closed pEN300::Tn7-1 the following experiment was performed.

pEN300::Tn7-1 was transferred from DS916 pEN300::Tn7-1 to CSH50::Tn7-1 pSB58 pGP1-2 and CSH50 pSB58 pGP1-2. This generated the two novel donor strains CSH50::Tn7-1 pSB58 pGP1-2 pEN300::Tn7-1 and CSH50 pSB58 pGP1-2 pEN300::Tn7-1. In the latter strain the chromosomal *attTn7* site is un-occupied. This means that Tn7-1 in pEN300 can potentially transpose at a high frequency to this chromosomal site. If the reduction in pEN300::Tn7-1 transfer initially observed is dependant on the on-going transposition of Tn7-1 from pEN300::Tn7-1 then it would be predicted that conjugation of pEN300::Tn7-1 from this strain should be dramatically reduced on pSB58 induction.
Overnight cultures were grown at 30°C. 10ml aliquots of overnight were added to 10ml of fresh L-broth and either kept at 30°C or induced (42°C for 25 min., then 37°C for 2 hours). Matings were then performed using DS916 as recipient. The results are shown in Fig. 4.12. Contrary to prediction the level of conjugation appears to increase on induction, however a similar increase is observed with the control pEN300

conjugation. Thus, the only conclusion that can be drawn from this experiment is that pEN300 conjugation is more efficient at 37°C.

Clearly pSB58 induction is not causing the expected 100 fold decrease in detection of the pEN300::Tn7-1 in the exconjugants. Thus it appears that the apparent prevention of transfer caused by pSB58 induction is somehow dependent on the recent transposition of Tn7-1 into pEN300. It is possible therefore that the bizarre effects of pSB58 are exerted on the gapped molecules that are the first Tn7 insertion product. This is still an untested hypothesis. It should be noted, however, that Nancy Craig's group have recently shown (Nancy Craig, personal communication to Dave Sherratt) that a synthetic structure analogous to a single end insertion can excise. This *in vitro* reaction requires TnsA,B and C. It may be significant that the most highly overexpressed pSB58 species are TnsA and B.

Because of the number of peculiar features of the transposition events supported by the pSB58 construct it was deemed unsuitable for use in attempts to develop an *in vitro* transposition assay.
The attempts to establish an *in vitro* assay therefore used pSB84 extracts.

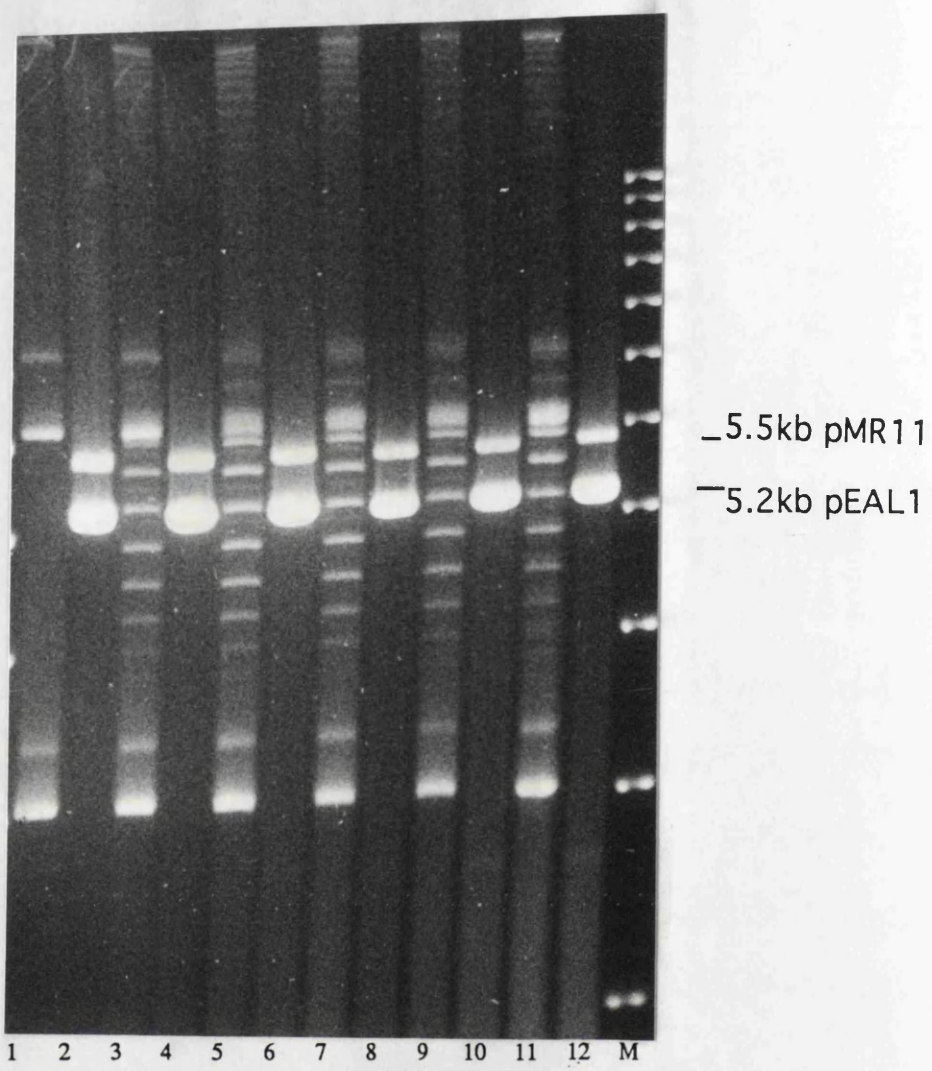


Fig. 4.13a Ethidium bromide stained gel of attempted *in vitro* transposition assay.

Reactions were carried out as detailed in Materials and Methods.

The lanes in the gel alternate uncut and HindIII digested reaction products. The protein extract used was freshly prepared CSH50 pGP1-2 pSB84 extract, prepared as detailed in Materials and Methods.

Lanes 1&2: Reactions with no protein extract.
Lanes 3&4: 5 µl (55µg) extract
Lanes 5&6: 10 µl (110µg) extract
Lanes 7&8: 20 µl (220µg) extract
Lanes 9&10: 30 µl (330µg) extract
Lanes 11&12: 40 µl (440µg) extract
M= Markers 1 kb ladder (BRL)

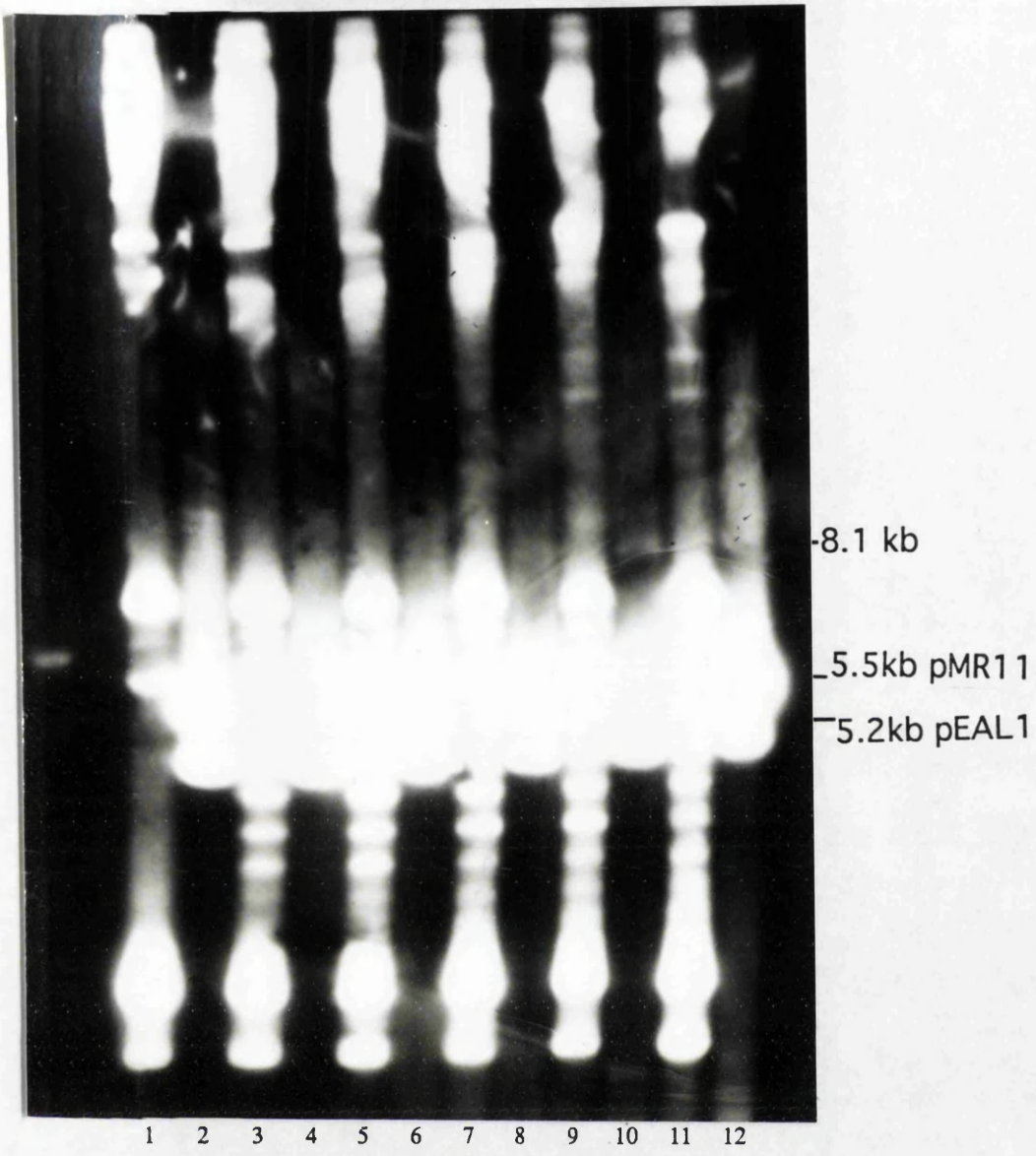


Fig. 4.13b Southern blot of the gel in Fig. 4.13a

The Southern blot was performed as detailed in Materials and Methods. The membrane was hybridised with a pMR11 specific probe generated by random priming on gel isolated pMR11 0.6kb PstI fragment.

Lanes 1&2: Reactions with no protein extract.
Lanes 3&4: 5 µl (55µg) extract
Lanes 5&6: 10 µl (110µg) extract
Lanes 7&8: 20 µl (220µg) extract
Lanes 9&10: 30 µl (330µg) extract
Lanes 11&12: 40 µl (440µg) extract

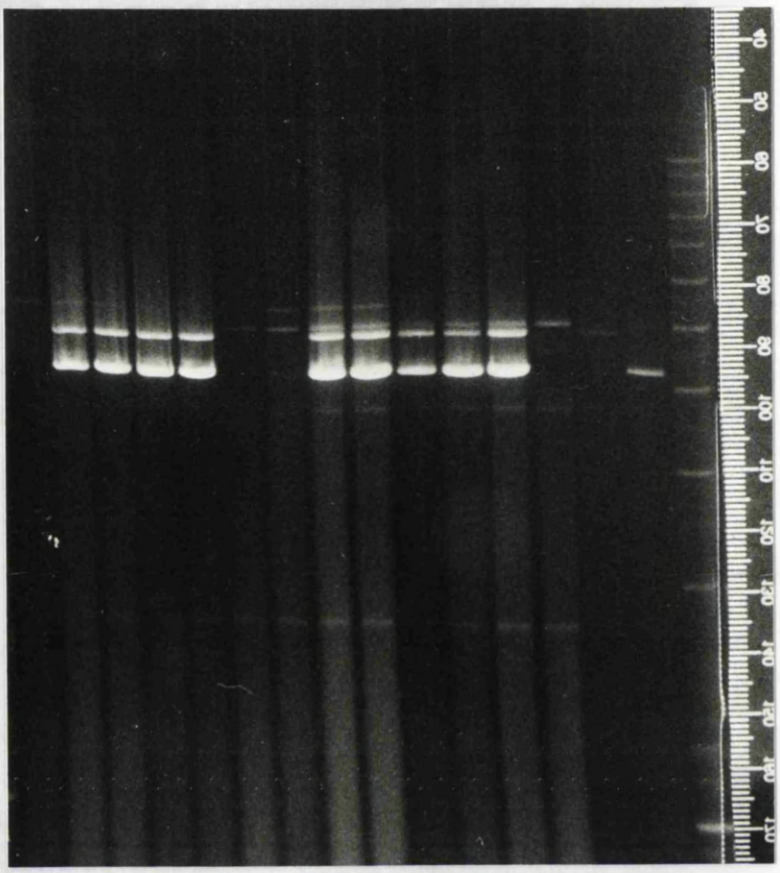
4.8 Attempts to establish a Tn7 *in vitro* transposition assay

The conditions for a Tn7 *in vitro* assay were published just as this section of the work was started. All the attempts described below use the basic reaction conditions described in Bainton, 1991.

Initial experiments were carried out using a variable amount of crude cell extract prepared from induced CSH50 pGP1-2 pSB84. The transposition donor molecule used in these assays was pMR11, (pUC8 containing Tn7-1) and the recipient was pEAL1. An initial experiment was carried out using 0,55,110,220,330 and 440 micrograms of protein extract, amounts of protein were determined by Bradford's Assay. Reactions were incubated at 30°C for 7 minutes then MgAcetate added to 15mM and incubation continued at 30°C for a further 30 minutes. DNA was recovered from the reaction by phenol extraction and ethanol precipitation. The DNA was resuspended in 10µl of TE and a 3µl aliquots digested by HindIII. The digest was then run on a 0.7% agarose gel. After ethidium staining and photography, the contents of the gel were transferred to Hybond-N by Posiblotter. The transferred DNA was linked to the membrane using a Stratalinker. After pre-hybridisation the membrane was hybridised with a probe made by random priming of a 600nt PstI fragment of pMR11. Thus both the donor pMR11 and any transposition product should be detected by this probe.

The predicted size of the HindIII restricted transposition product pEAL1::Tn7-1 is 8.1 kb. No species of this size was observed on ethidium staining of the gel. Similarly, an overnight exposure of the Southern Blot derived from this gel showed no apparent product (Fig 4.13). However, a gross over-exposure of the blot did reveal the presence of a very weak signal at approximately 8kb species. weak signals are also seen at approximately 2.6kb and 2.8 kb, the size of Tn7-1. The level of these signals were dependent on the amount of protein extract added. However, it is also apparent at this level of exposure that the probe is cross hybridising with pEAL1 sequences. It is highly possible therefore that these potential product bands may be artefactual.

A very obvious feature of the reaction is the extensive topoisomerization of the substrate plasmids. This phenomenon appeared to be independent



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
pMR11	-	+	+	+	+	-	-	+	+	+	+	+	-	+	-
pEAL1	-	+	+	+	+	-	-	+	+	+	+	+	-	-	+
pMR64 extract	10	10	5	1	-	-	10	10	5	1	-	-	-	-	-

figures represent µl of a 5mg/ml extract, as assayed by Bradfords assay

Fig. 4.14: Ethidium bromide stained gel of HindIII digested products of attempted *in vitro* assays with additional TnsD extract.

Assays 1-5 used 30µl of the extract used in the experiment shown in Fig. 4.13, stored at -70 C. Assays 6-15 used 30 µl of a fresh 8mg/ml CSH50 pGP1-2 pSB84 extract.

Assays were performed with additional, varying amounts of a 5mg/ml TnsD extract (from DS941 pMR64) as detailed above.

Other aspects of reaction conditions were as detailed in Materials and Methods.

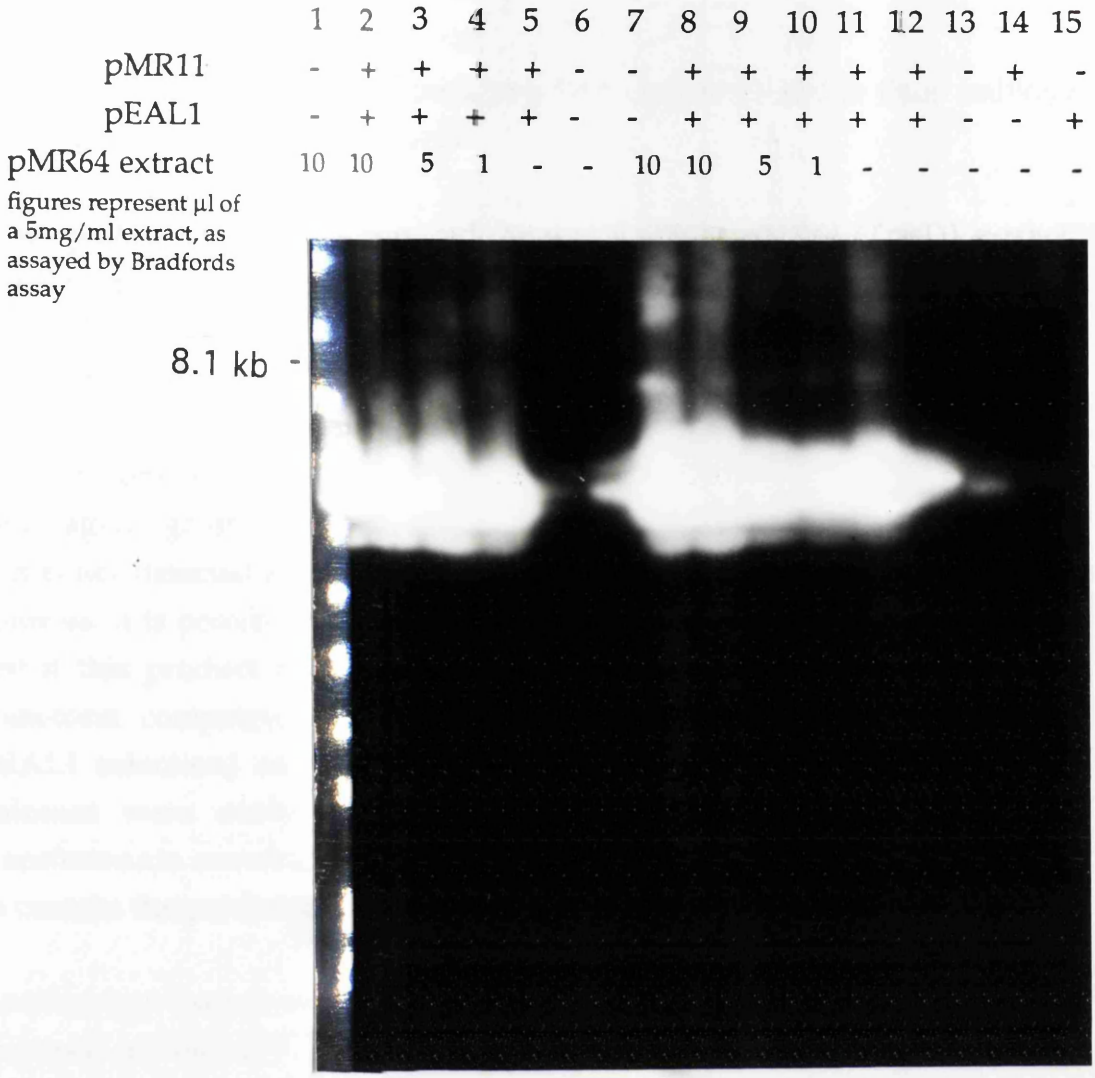


Fig. 4.14b: Southern blot of gel in Fig. 4.14a

The Southern blot was carried out as described in the Legend to Fig. 4.13b

of whether the plasmid contained any Tn7 sequences and is thus unlikely to be significant (data not shown).

Fresh pSB84 extract was prepared, as was a DS941 pMR64 (TnsD) extract. The Craig lab had found that the *in vitro* reaction was highly dependent on the level of TnsD. Reactions were carried with 300µg of pSB84 extract supplemented with 0,5,25 or 50 µg of pMR64 extract. Assays were also performed using the initial pSB84 extract, which had been stored at -70°C. The reactions were carried out as before, the results are shown in Fig. 4.14. Once again, gross over-exposure of the blot reveals a faint signal at ~8kb, this is not detected in the pMR11+extract, pEAL1+extract and extract alone controls. It is possible that this ~8kb band is a genuine reaction product. To test if this product could be isolated, 1µl of reaction I DNA was used to transform competent E. coli. The cells were plated on Tetracycline (for pEAL1 selection) and chloramphenicol (Tn7-1) containing medium. 50 colonies were analysed on Single Colony Gels, all 50 were double transformants containing both pEAL1 and pMR11, no colonies were found to contain the predicted 8.1kb product (data not shown).

Another key feature of the *in vitro* assay developed in San Francisco is the requirement for ATP. All the assays described above contain 2mM ATP. However the use of a crude cell extract in these experiments adds a potential drain for ATP in the form of endogenous host ATPases present in the extract. In order to prevent the potential depletion of ATP an ATP regeneration system was added to a set of "*in vitro*" assays. 5 units of Pyruvate kinase and 20mM Phosphoenolpyruvate. Pyruvate kinase, an enzyme in the glycolytic pathway catalyses the conversion of Phosphoenolpyruvate to pyruvate with the concomitant conversion of ADP to ATP. The assays shown in Fig 4.15 were performed. The addition of ATP regeneration did not affect any potential reaction, no product bands were observed in any of the lanes. Numerous other attempts were made to develop an *in vitro* reaction; reactions were carried out at both pH7 and 8 and also at varying salt concentrations (data not shown), it was never possible to get reproducible production of the 8.1kb potential product.

A similar lack of success has been experienced by Conrad Lichtenstein's group in London. As described above, this group has made translational fusions of TnsA,B,C and D. This has enabled them to purify TnsA,B and C

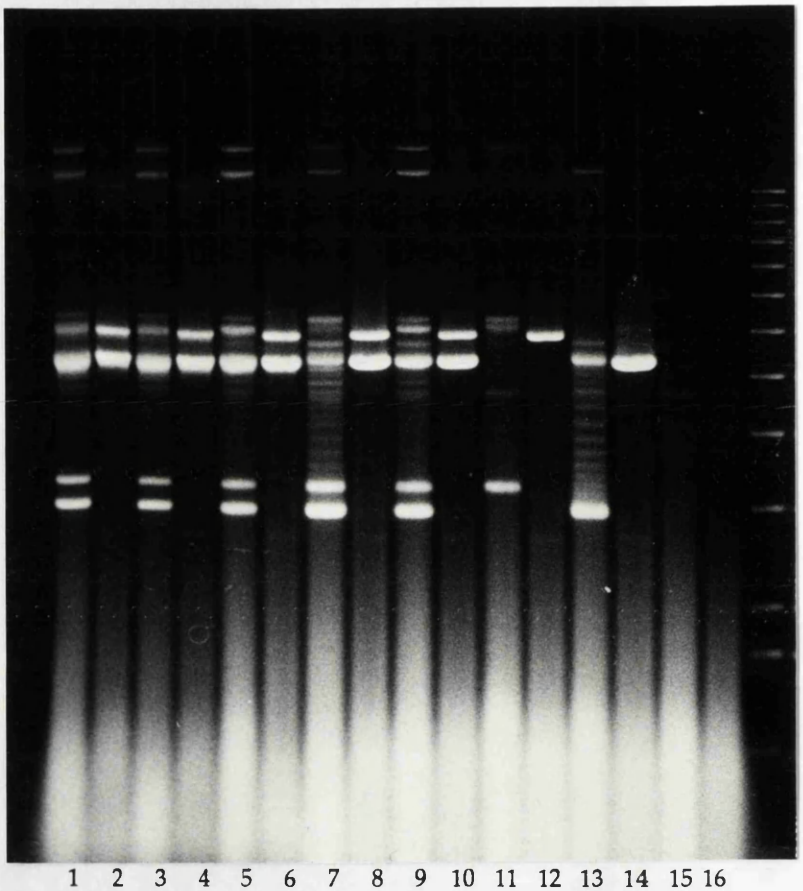


Fig. 4.15a Effect of ATP regeneration on *in vitro* reactions.

The lanes in the gel were loaded alternately with undigested and HindIII digested reaction product. The reactions were carried out as previously described, with 30µl of a 8 mg/ml pSB84 extract and 10 µl of a 5 mg/ml DS941pMR64 extract.

The reactions shown in lanes 7-16 differed from the standard by inclusion of 20mM PEP and 5 units pyruvate kinase.

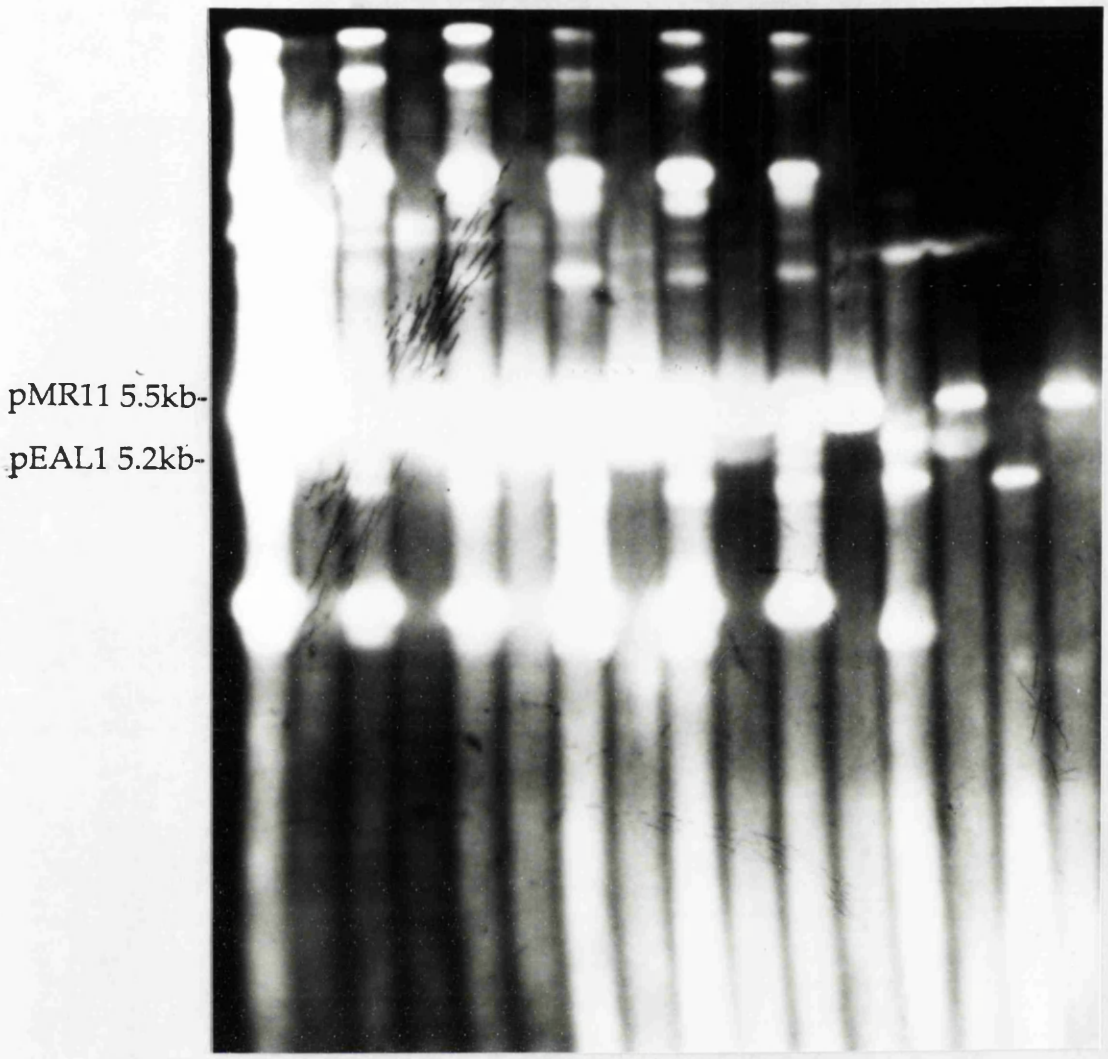


Fig. 4.15b Effect of ATP regeneration on *in vitro* reactions.

The Southern blot was performed as detailed in the legend of Fig. 4.13

The reactions shown in lanes 7-16 differed from the standard by inclusion of 20mM PEP and 5 units pyruvate kinase.

to homogeneity and substantially purify TnsD. Using both Southern blotting and PCR detection methods they have failed to reconstitute a reproducible Tn7 *in vitro* reaction from these highly purified components. (C. Lichtenstein pers. comm.)

4.9 Discussion

This chapter has described the construction of two constructs, pSB58 and pSB84, designed to express high levels of the Tns proteins. The levels of protein expressed from these constructs was examined. The ability of these constructs to support transposition of a Tn7-1 element was tested. This gave the surprising discovery that induction of pSB58 apparently prevented the conjugative transfer of pEN300::Tn7-1. Tests were carried out to determine if this apparent loss was caused by excessive levels of transposition from the pEN300::Tn7-1 by attempting to detect the proposed double strand breaks that would be generated by such reactions. No candidate species were found by Southern blotting of restricted genomic DNA. This could simply be a feature of the instability of such broken molecules, endogenous cellular nucleases will rapidly degrade molecules with exposed ends. It is also possible that the detection method employed was not sensitive enough to permit detection of these species. Thus firm conclusions cannot be drawn as to the presence or absence of such breaks. Work in Nancy Craigs lab supports the theory that double strand breaks are produced by *in vivo* transposition of Tn7. Work presented by Ann Hagerman at the FASEB conference (August. 1991) showed that homologous recombination is stimulated at the Tn7 transposition donor sites. Direct detection of the breaks has not been achieved.

Further *in vivo* experiments tentatively suggest that the pSB58 mediated effect may not be exerted on fully covalently closed pEN300::Tn7-1 molecules. The potential relevance of to the Tn7 transposition has been discussed. it should be stressed that these conjectures are completely un tested. It would be of great interest to know if manipulation of protein concentrations in the in vitro reaction established by the Craig group could reproduce these proposed effects. In particular does a large excess of TnsA stabilise the intermediate in the reaction, and would an excess of TnsA stimulate the dis-integration assays?

The attempts to generate an *in vitro* transposition assay, described in this chapter are similarly inconclusive. Candidate product molecules were occasionally detected, but never reproducibly, and always at the extreme limit of detection by Southern blotting. With the benefit of hindsight it would probably have been advantageous to couple the Southern blotting to a PCR based assay, using one primer specific to Tn7-1 and the other to the recipient, thus only transposition product would be amplified. PCR would have permitted a qualitative assay and the Southern blotting a more quantitative assay. Such an approach has been used by Conrad Lichtenstein's group, yet, even using highly purified Tns proteins, they too have failed to generate a reproducible *in vitro* reaction (C. Lichtenstein pers. com.).

A feature common to the attempts described and those of Lichtenstein's group is that the TnsB activity used in the assays generate a large number, ≥6, of complexes with the Tn7 right end in gel retardation assays (See Chapter 3 Fig. 3.12). By way of contrast the TnsB activity used by the Craig group produces a maximum of 4 complexes in such assays. Work in our group has demonstrated that the additional complexes are generated by proteolytic cleavage products of TnsB which retain DNA binding activity. Initial results from this lab suggest that over-expression of these fragments leads to some inhibition of transposition of Tn7 *in vivo*. Thus, the quality of TnsB used in the reactions may be of key importance.

Chapter 5

DNA Binding Activities in pSB84 Extracts

5.1 Introduction

The previous chapter described unsuccessful attempts to establish a Tn7 *in vitro* transposition reaction. Extracts from pSB84 containing cells were used in these attempts. This chapter demonstrates that TnsB and TnsD activities are present in the extract, and also presents circumstantial evidence for the presence of TnsC. Previous works have demonstrated specific DNA binding activities for both TnsB and TnsD. as discussed in the introduction (Waddell and Craig 1989 ; Morrell 1990). TnsB binds to the terminal 22bp repeats of Tn7 while TnsD binds to the *attTn7* site. The technique used to detect the activities of these proteins was that of gel retardation assays. The presence of these proteins in pSB84 extracts was determined using this methodology.

5.2 TnsB Activity in pSB84 extracts

TnsB activity in the pSB84 extract was detected by gel retardation assays, using pNE200 DNA, pUC8 containing the terminal right end motifs of Tn7, as substrate. This data is presented in Chapter 3, Fig. 3.12.

5.3 Assays for TnsD

Tn7 inserts at a specific location in the *attTn7*, between the *phoS* and *glmS* genes at minute 84 of the *E. coli*chromosome (Lichtenstein and Brenner 1981 ; Lichtenstein and Brenner 1982). Analyses have revealed that the architecture of the *attTn7* site is quite complex. Studies using various deletions and point mutations in the site showed the sequences defining *attTn7* activity lie to the right, *glmS*, side of the Tn7 insertion point (McKown, Orle et al. 1988 ; Qadri, Flores et al. 1989). The sequence at the actual insertion site can be varied with little or no effect on the efficiency of the site as a transposition recipient. In the nomenclature used in the literature the middle base pair of the five base pair sequence duplicated on Tn7 insertion is designated 0. Nucleotide positions to the right are given a positive sign, those to the left negative. Initial studies showed that a DNA molecule corresponding to +7 to +64 conferred near to wild-type activity (Waddell and Craig 1989).

A further feature of the TnsABCD directed pathway is that if the *attTn7* site is blocked then secondary sites can be used at a lower frequency,

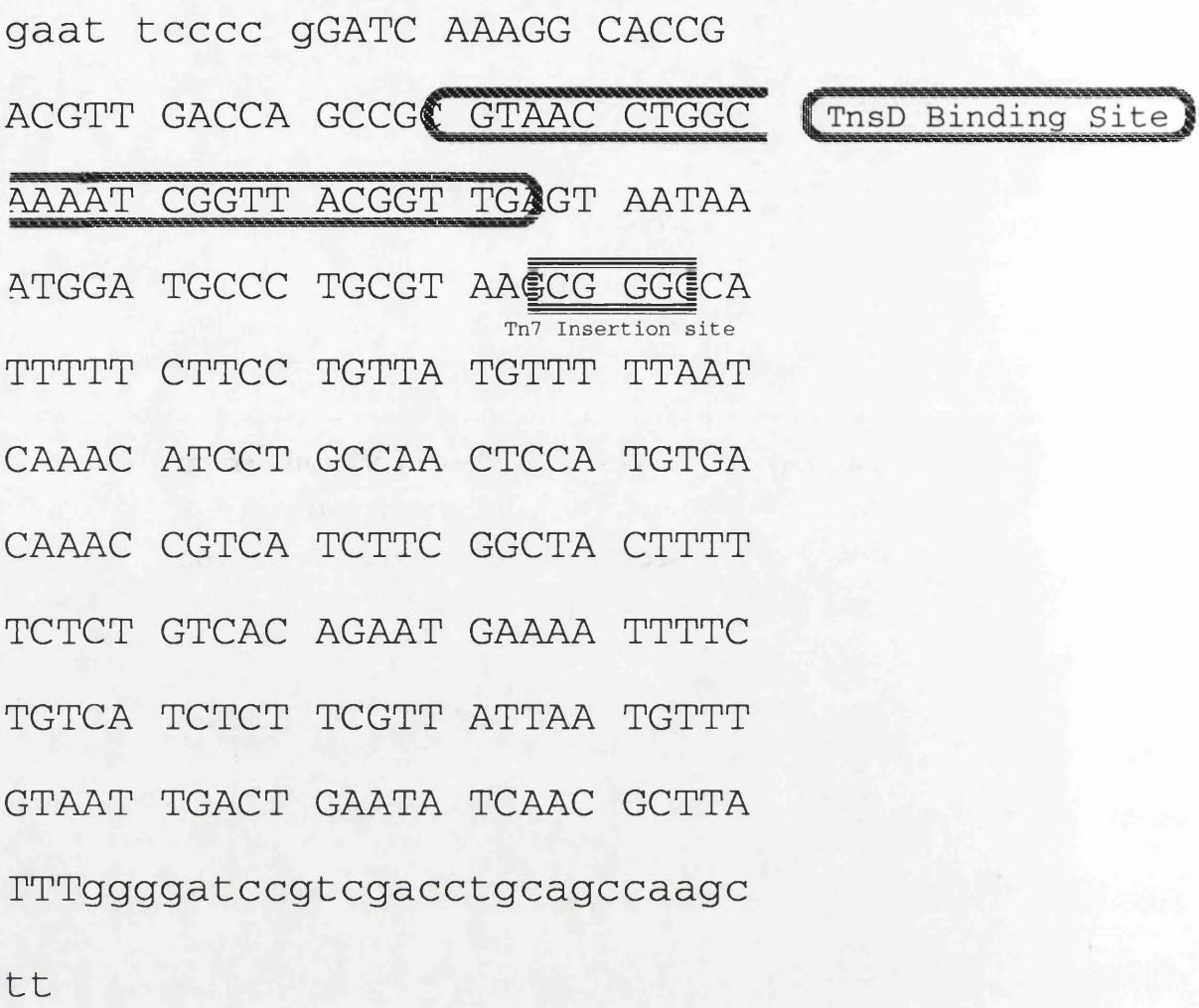


Fig.5.1
Sequence of the *attTn7* insert in pMR80.

Chromosomal sequence are in capitals, pUC18 polylinker in lowercase.
The TnsD binding site detected by Waddell and Craig, 1989 is boxed

approximately 1000-10,000 fold lower, than hot site events. Sequence analysis of a number of such sites revealed that the nucleotides at the insertion site are completely diverged. However, a block of conserved sequence was observed corresponding to *attTn7* +28 to +51 (Kubo and Craig 1990).

The most recent work to be published on the dissection of the hot site demonstrated that TnsD has a sequence specific DNA binding activity. An activity only present in *tnsD*$^+$ cells was detected which bound to DNA containing *attTn7* sequences. Analyses of various fragments of *attTn7* found the minimal region which could be bound corresponded to *attTn7* +28 to +55. Within this sequence is a region of dyad symmetry, from +32 to +55. This region, however, did not bind TnsD, nor did it function as a *attTn7* site *in vivo* (Waddell and Craig 1989).

It was also observed that an unidentified host factor bound to larger *atttn7* fragments. This activity was localised to +10 to +27, i.e. between the TnsD binding site and the insertion site. Intriguingly, although the +28 to +55 directed strong TnsD binding, the same region did not confer wild type *attTn7* activity. Instead a mini-Tn7 element transposed to this site at a 100 to 1000 fold lower frequency than to a site comprising -25 to +64.

This reduced frequency could be due to a combination of the imposition of a novel insertion point, the presence of unusual sequence between the TnsD binding site and the insertion point and the loss of the host factor binding site. Work presented in Chapter 6 attempts to establish the identity and role of the host factor.

5.4 Assays of TnsD activity in pSB84 extracts

Crude cell extracts were prepared from strains containing either pSB84 or pSB58, in the hope of using them in attempts to establish an *in vitro* assay. In order to confirm the presence of the TnsD protein in these preparations, I employed the gel retardation assay established by Waddell and Craig (1989). The DNA substrate used was an end-labelled EcoRI-HindIII fragment of pMR80. This 280 bp sequence contains the *attTn7* DNA shown in Fig 5.1. It should be noted this molecule is considerably larger than that used by Waddell and Craig.

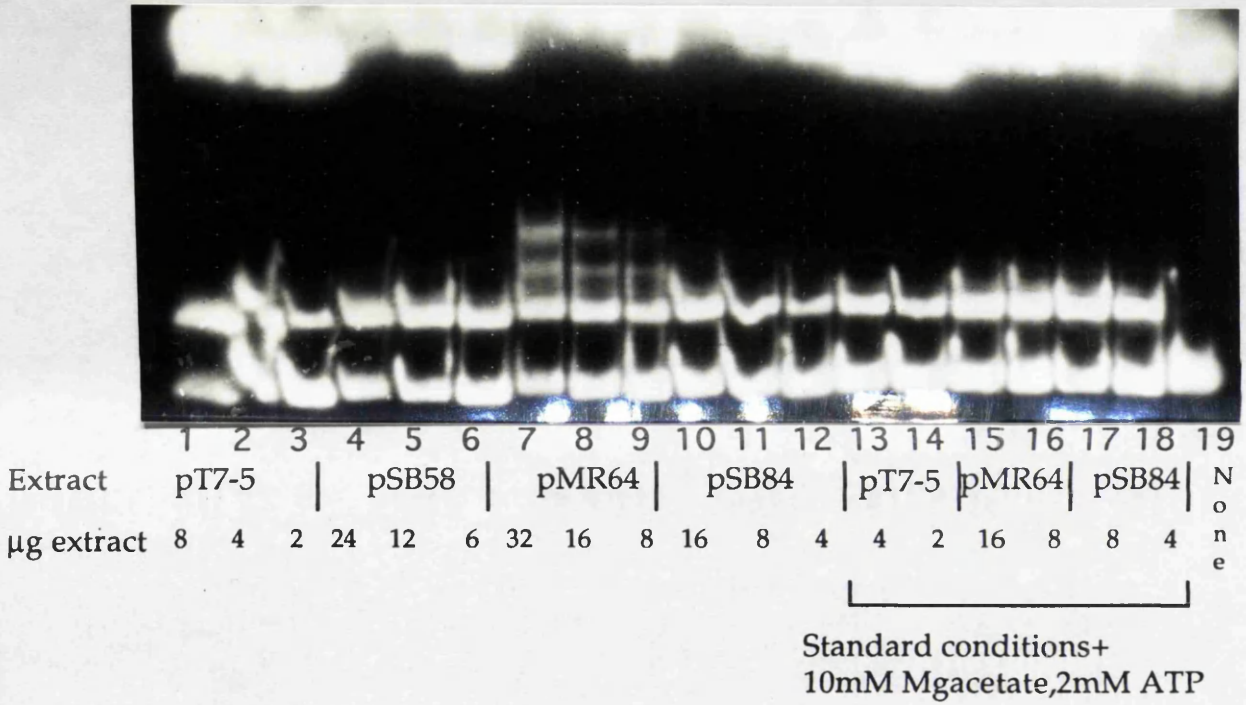


Fig. 5.2: *attTn7* Retardations with various protein extracts.

Extracts of CSH50 pGP1-2 pT7-5, CSH50 pGP1-2 pSB84, CSH50 pGP1-2 pSB58 and DS941 pMR64 were assayed for *attTn7* binding activity.

Reaction conditions were standard TnsD conditions with the exception that the assays in lanes 13-18 contain 10mM MgAcetate and 2mM ATP

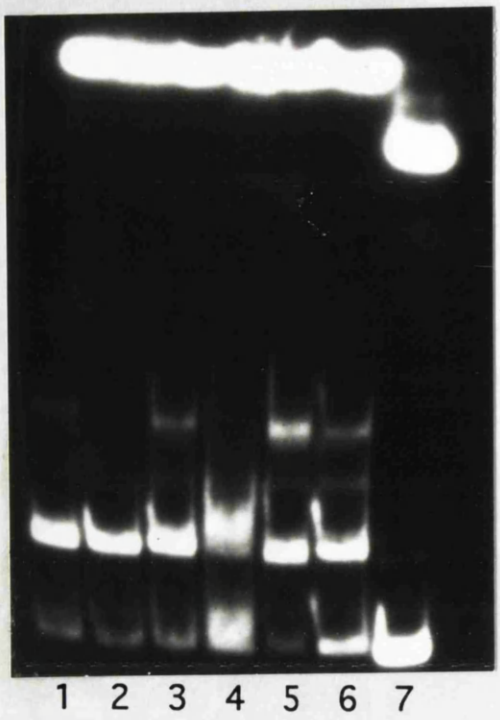
Extracts of CSH50 pGP1-2 pT7-5 and DS941 pMR64 were also assayed as negative and positive controls respectively. pMR64 contains the *tnsD* gene cloned downstream of the ptac promoter.
As shown in Fig. 5.2, in good agreement with the published results, a species of retarded mobility was observed in all assays, independent of the presence of TnsD. In this figure this appears to be a single band, however, when electrophoresis is carried out for a further half hour this band resolves into a doublet. This is in contrast to the published results of Waddell and Craig and may reflect my use of a larger *attTn7* fragment.

Two more highly retarded complexes are observed when pMR64 extracts are used in the assay. The lower complex was not always observed when different preparations of the pMR64 extract were assayed. This appeared to correspond to differing levels of host binding activity in different preparations. When high host factor binding is observed only the upper TnsD band is seen, if lower host factor binding is observed then both TnsD bands can be detected.
When pSB84 and pSB58 extracts are tested for TnsD activity none is observed under standard binding conditions.

Bainton *et al* showed ATP and Mg ions are required for the *in vitro* assay. However, the addition of Mg++ has a highly deleterious effect on the reaction unless a 7 minute preincubation is carried out prior to its addition. In order to determine if the addition of ATP and Mg has an effect on TnsD activity in pMR64, pSB84 and pSB58 extracts retardations were performed under a variety of conditions. Intriguingly, when 2mM ATP and 10mM Mg-Acetate are added to binding reactions the TnsD complexes seen with pMR64 extracts disappear, the host factor complex is untouched. A number of faint higher complexes appear in the pSB84 and pSB58 extracts.

5.5 Effect of Mg^{++} on TnsD binding

The loss of TnsD binding activity can be shown to be due to Mg++ as shown in Fig. 5.3. When ATP alone is added to pMR64 binding reactions no alteration to binding is observed. As can be seen from Fig. 5.2 and 5.4 10mM Mg prevents binding, as does 5mM Mg, albeit to a lesser extent. The



All binding reactions use 10µg DS941pMR64 extract under standard TnsD binding conditions except for the following additions

Lane 1: 5mM Mg Acetate
Lane 2: 5mM Mg Cl_2
Lane 3: 5mM $CaCl_2$
Lane 4: 5mM Spermine
Lane 5: 5mM MgAc for min then
15mM EDTA for 5 min
Lane 6: No addition
Lane 7: No protein

Fig. 5.3 Effect of cations on TnsD binding to *attTn7*

effect is caused both by $MgCl_2$ and MgAcetate but not by $CaCl_2$. 10 mM Spermine also prevents binding.

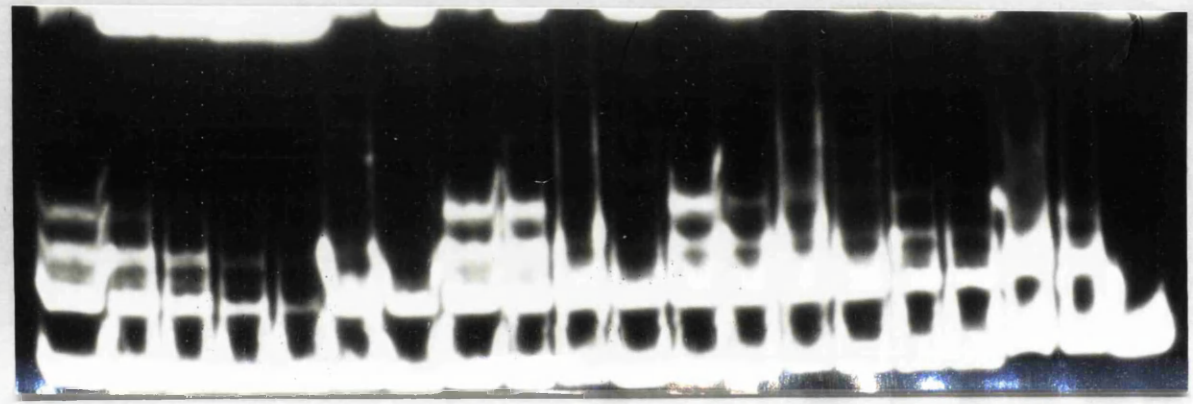
The loss of binding activity could be caused by prevention of TnsD binding, possibly by either altering the conformation of the protein or alternatively by activating a metalloproteinase which degrades the TnsD protein. In order to test these hypotheses, two series of TnsD binding reactions were carried out using pMR64 extracts. In each series two sets of conditions were used, "standard" and "standard"+10mM MgAc. One series of reactions was incubated for 20 minutes, the other for 10 minutes prior to addition of EDTA to 20mM.

The results shown in Fig. 5.3 indicate that TnsD binding can be restored by the addition of EDTA. EDTA also appears to augment the binding activity under standard conditions, this may be due to removal of endogenous Mg in the cell extract.

In the published protocol for Tn7 *in vitro* reactions the importance of allowing a 7 minute pre-incubation prior to Mg addition is stressed. When a 10 minute pMR64 binding reaction is performed with such a pre-incubation, binding is increased relative to a reaction with Mg added at the start. This is probably due to the short exposure of the presumptive DNA-protein complexes to Mg prior to electrophoresis. If the reaction is continued for a further 10 minutes after Mg addition then full inhibition is observed. Therefore it appears that Mg can not only prevent initial binding by TnsD but also cause dissociation of TnsD-DNA complexes. This effect may be caused by Mg directly disrupting complexes. Alternatively, if there is a equilibrium set up with TnsD constantly binding and dissociating, then Mg could act by preventing free TnsD from interacting with DNA.

5.6 TnsD Activity in pSB84 Extracts

The initial assays described above could detect no TnsD activity in pSB84 extracts. Following the discovery of the inhibitory effects of 10mM Mg, the assays were repeated using 5mM Mg, added after a 7 minute preincubation in "standard" conditions + 2mM ATP.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
ATP (mM)	0	0	0	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2
5mM Mg added at												7	7	7	7	0	0	0	0
µg of extract	8	4	2	1	0.5	8	4	8	4	8	4	8	4	8	4	8	4	8	4
Extract		pMR64				pSB84		pMR64		pSB84		pMR64		pSB84		pMR64		pSB84	

Fig. 5.4: Autoradiograph of gel retardations using endlabelled EcoRI/HindIII digested pMR80 as DNA susbstrate.

Extracts were prepared from CSH50 pGP1-2 pSB84 and DS941 pMR64. Reactions conditions were "standard" TnsD binding conditions as detailed in Materials and Methods with the addition of 5mM MgAcetate and/or 2mM ATP as shown above.

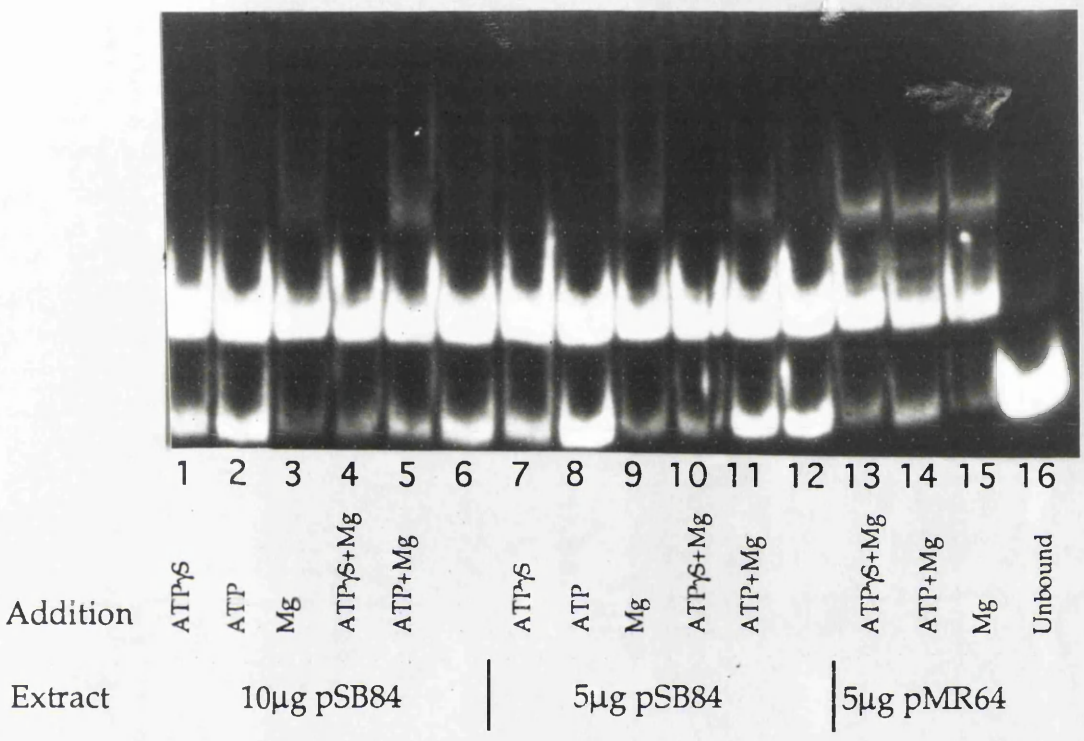


Fig. 5.5 Requirement for ATP hydrolysis

Either 10 or 5μg of a CSH50 pGP1-2 pSB84 extract or 5μg of a DS941 pMR64 extract was assayed for binding to *attTn7* by gel retardation. Binding was carried out under standard conditions or with the addition of ATP or ATPγS in the presence or absence of Mg.

Lanes 1&7: standard + 2mM ATPγS
Lanes 2&8: standard + 2mM ATP
Lanes 3,9&15: standard + 5mM Mg acetate
Lanes 4,10&13: standard + 2mM ATPγS + 5mM Mg acetate
Lanes 5,11&14: standard + 2mM ATP+ 5mM Mg acetate
Lanes 6&12: standard

Lane 16: no extract

As can be seen in Fig. 5.4 in the control reactions using pMR64 extract a slight decrease in D binding can be observed compared to that seen under standard conditions. However, when a pSB84 extract was used an entirely different result was seen. In the absence of ATP and Mg no TnsD activity can be detected; by contrast when both ATP and Mg are added to the reaction a TnsD binding activity is seen.

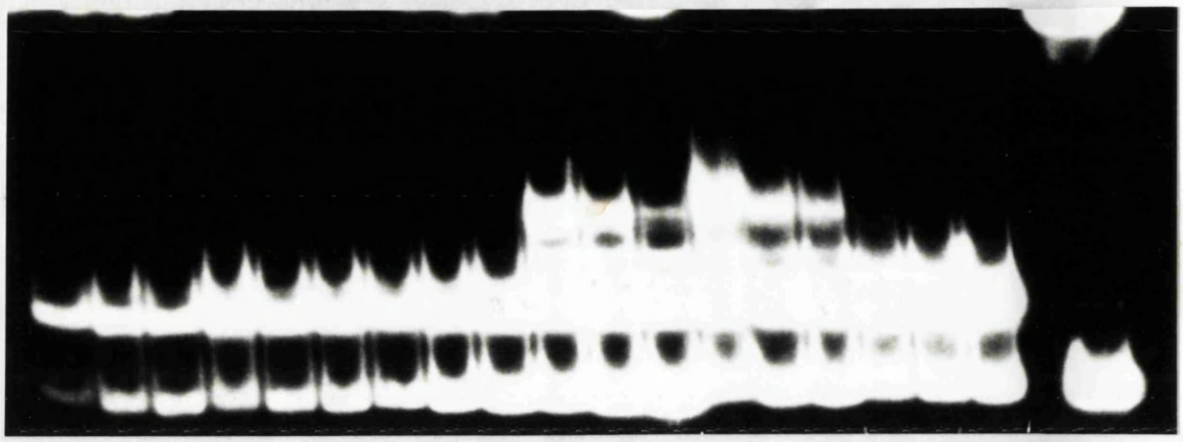
5.7 Requirement for ATP Hydrolysis to allow detection of TnsD activity

The requirement for ATP and Mg suggests that ATP hydrolysis may be necessary to permit the appearance of the TnsD activity.
This was tested by using a non-hydrolysable analogue of ATP, ATPγS in the reaction. As can be seen in Fig. 5.5 the only strong TnsD activity is seen when both ATP and Mg are added, no activity is seen when ATPγS is substituted for ATP.
TnsD activity also appears when 5mM Mg^{++} alone is added to the reaction. The Mg may act to permit hydrolysis of ATP present in the extract. The fact that no activity is seen when ATPγS and Mg are added suggests that the ATPγS may be competing with the low levels of endogenous ATP in the crude extract thus preventing TnsD "release".

No such difference is seen when a pMR64 extract is used under the same variety of conditions. This suggests that in the pSB84, extract, TnsD is interacting in some way with one or more of TnsA,B and C.

5.8 TnsD is interacting with a component of TnsABC

This TnsD sequestration and ATP dependent release can be re-constructed by addition of a TnsABC extract to a TnsD extract. Fig 5.6 a and b show the effect of pre-mixing a pMR64 extract with a CSH50 pGP1-2 pSB70 (*tnsABC*) extract. A 1:1 mix of the two extracts was made and preincubated for 10 minutes prior to addition to a binding reaction. 40ug of either individual or combined extract was added to each reaction as detailed in Fig. 5.6. Strong TnsD activity could be observed when pMR64 alone or pMR64 mixed with CSH50 pGP1-2 pT7-5 (vector) extract was assayed in the absence of ATP/Mg. However no activity was seen in pSB84 or pMR64+pSB70 extracts under these conditions. Contrastingly, these latter 2 mixed extracts



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

No Tns | TnsABC | TnsABCD | TnsD | No Tns + TnsD | TnsABC + TnsD

Lane 1: 20 µg CSH50 pGP1-2 pT7-5 extract
Lane 2: 10 µg CSH50 pGP1-2 pT7-5 extract
Lane 3: 5µg CSH50 pGP1-2 pT7-5 extract

Lane 4: 20 µg CSH50 pGP1-2 pSB70 extract
Lane 5: 10 µg CSH50 pGP1-2 pSB70 extract
Lane 6: 5 µg CSH50 pGP1-2 pSB70 extract

Lane 7: 20 µg CSH50 pGP1-2 pSB84 extract
Lane 8: 10 µg CSH50 pGP1-2 pSB84 extract
Lane 9: 5 µg CSH50 pGP1-2 pSB84 extract

Lane 10: 10 µg DS941 pMR64 extract
Lane 11: 5 µg DS941 pMR64 extract
Lane 12: 2.5 µg DS941 pMR64 extract

Lane 13: 10 µg CSH50 pGP1-2 pT7-5 extract + 10 µg DS941 pMR64 extract

Lane 14: 5 µg CSH50 pGP1-2 pT7-5 extract + 5 µg DS941 pMR64 extract

Lane 15: 2.5µg CSH50 pGP1-2 pT7-5 extract + 2.5 µg DS941 pMR64 extract

Lane 16: 10 µg CSH50 pGP1-2 pSB70 extract + 10 µg DS941 pMR64 extract

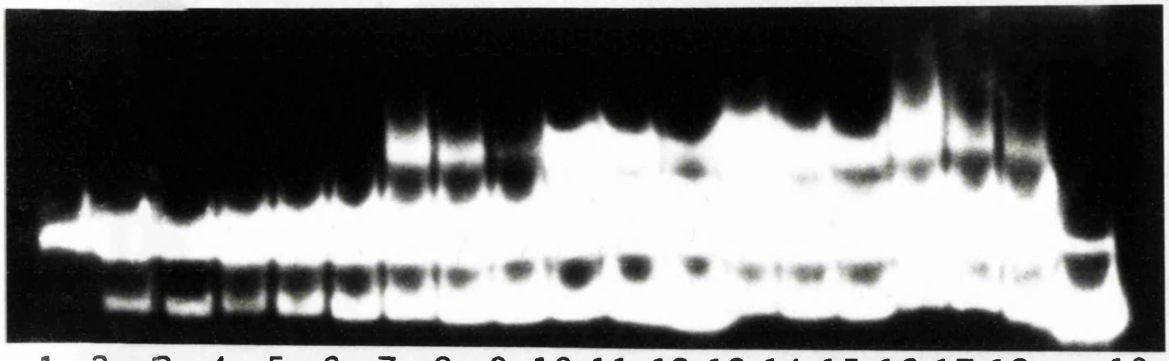
Lane 17: 5 µg CSH50 pGP1-2 pSB70 extract + 5 µg DS941 pMR64 extract

Lane 18: 2.5 µg CSH50 pGP1-2 pSB70 extract + 2.5 µg DS941 pMR64 extract

Lane 19: No protein

Binding Conditions "Standard" TnsD conditions as described in Materials and Methods

Fig. 5.6a TnsD is interacting with a component of TnsABC



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

No Tns | TnsABC | TnsABCD | TnsD | No Tns + TnsD | TnsABC + TnsD

Lane 1: 20 µg CSH50 pGP1-2 pT7-5 extract
Lane 2: 10 µg CSH50 pGP1-2 pT7-5 extract
Lane 3: 5µg CSH50 pGP1-2 pT7-5 extract

Lane 4: 20 µg CSH50 pGP1-2 pSB70 extract
Lane 5: 10 µg CSH50 pGP1-2 pSB70 extract
Lane 6: 5 µg CSH50 pGP1-2 pSB70 extract

Lane 7: 20 µg CSH50 pGP1-2 pSB84 extract
Lane 8: 10 µg CSH50 pGP1-2 pSB84 extract
Lane 9: 5 µg CSH50 pGP1-2 pSB84 extract

Lane 10: 10 µg DS941 pMR64 extract
Lane 11: 5 µg DS941 pMR64 extract
Lane 12: 2.5 µg DS941 pMR64 extract

Lane 13: 10 µg CSH50 pGP1-2 pT7-5 extract + 10 µg DS941 pMR64 extract

Lane 14: 5 µg CSH50 pGP1-2 pT7-5 extract + 5 µg DS941 pMR64 extract

Lane 15: 2.5µg CSH50 pGP1-2 pT7-5 extract + 2.5 µg DS941 pMR64 extract

Lane 16: 10 µg CSH50 pGP1-2 pSB70 extract + 10 µg DS941 pMR64 extract

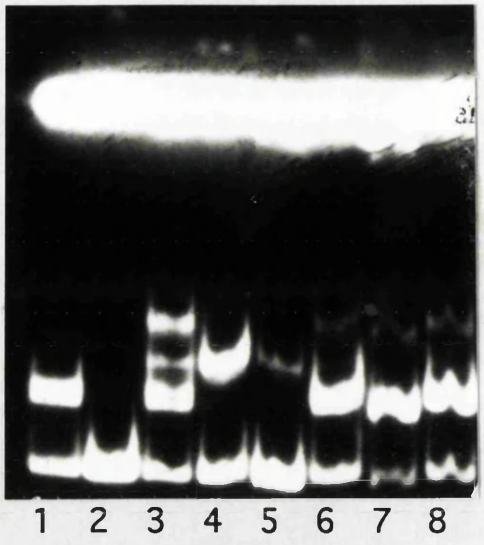
Lane 17: 5 µg CSH50 pGP1-2 pSB70 extract + 5 µg DS941 pMR64 extract

Lane 18: 2.5 µg CSH50 pGP1-2 pSB70 extract + 2.5 µg DS941 pMR64 extract

Lane 19: No protein

Binding Conditions "Standard" TnsD conditions + 2mM ATP. Reactions were incubated for 7 minutes then MgAc was added to 5mM. Incubations were continued for 5 further minutes then loaded onto a running gel.

Fig. 5.6b TnsD is interacting with a component of TnsABC



Lane 1: DS941
Lane 2: SB10
Lane 3: DS941 pMR64
Lane 4: SB10 pMR64
Lane 5: SB10 pMR51
Lane 6: DS941 pMR51
Lane 7: DS941 pMR51+2mM ATP, 7 min preincubation then 3 min with 5mM MgAc
Lane 8: DS941 pMR51 no ATP, 7 min preincubation then 3 min with 5mM MgAc

Lanes 1-8 all contain 20μg of crude extract prepared as described in Materials and Methods.
Retardation reactions performed as described in Materials and Methods.
Binding conditions in Lanes 1-6 are "standard" TnsD conditions.

Fig. 5.7 TnsD activity in a TnsCD extract

showed strong activity if ATP and Mg were added. Thus a component of the pSB70 (*tnsABC*) extract is capable of sequestering the TnsD activity. The requirement for ATP hydrolysis implicates TnsC, an ATP binding protein, as the key molecule in this process.

Assays were therefore carried out using extracts prepared from cells harbouring pMR51, demonstrated *in vivo* to be *tnsCD*$^+$. Thus, if TnsC alone was interacting with TnsD to prevent binding then a similar ATP/Mg dependent effect to that observed with pSB84 extracts should be detected. The result shown in Fig. 5.7 shows that TnsD binding activity in the pMR51 extract is ATP/Mg independent. Therefore, it appears that either TnsA and/or TnsB are required for sequestration of TnsD or, alternatively, the relative ratios of TnsC and TnsD differ between pSB84 and pMR51 extracts.

This was tested by combining a pMR51 extract with extracts prepared from cells overexpressing TnsA or TnsB. The combined extracts were assayed for TnsD binding activity in the presence and absence of ATP/Mg. This data is not shown. No ATP/Mg dependance on TnsD activity is observed with TnsA + TnsCD extract. The presence of the TnsB extract caused a smearing of the radiolabelled DNA on the gel. regretably insufficient time was available to repeat this experiment at lower concentrations of TnsB to determine if this effect was significant.

5.9 Discussion

The initial observation presented in this chapter is that Mg^{++} ions have a deleterious effect on TnsD binding. such an effect has also been observed by the Craig group (R. Bainton, pers. comm.) The binding activity of TnsD can be rescued by the addition of EDTA following Mg^{++}. This demonstrates that Mg is not bringing about any permanent alteration to either DNA or protein in the extract. It is unclear whether Mg^{++} exerts its effect on the TnsD protein or on the DNA substrate. By interacting with the charged phosphate groups on the DNA backbone, Mg^{++} can cause alterations to the conformation of the DNA. Clearly no major DNA changes can be occurring as the host factor binding is unaltered in the presence of the metal ion. It is intriguing that Mg^{++} bshould be inhibitory to formation of *attTn7*-TnsD complexes as a number of proteins have been

shown to require Mg ions for their metabolic function e.g. restriction endonucleases, DNaseI and ligase. Other proteins have also been shown to have a metal dependance for DNA binding, most notably the zinc finger proteins.

The data presented in this chapter also suggest an interaction between TnsD and components of TnsABC. No TnsD activity is detected by gel binding assays when a TnsABCD extract is assayed in the absence of ATP and Mg^{++}. In the presence of ATP and Mg^{++} TnsD activity is restored. It was also demonstrated that ATPγS does not substitute for ATP in this effect. It is unclear which of the Tns proteins interacts with TnsD.
On the basis that TnsC has been demonstrated to be an ATP binding protein it is very tempting to suggest that this protein plays a key role in the effects described. However, it is apparent that the effect is not due to an interaction between only TnsD and TnsC, as TnsD activity in an extract prepared from cells containg pMR51 does not show an ATP/Mg dependance.
The Craig group have demonstrated that purified TnsC can interact with TnsD at the att site, this effect is dependent on ATP; in agreement with my initial observation they too have shown that Mg^{++} concentration in excess of 10mM prevent TnsD binding.
In my work I have never observed additional higher complexes in TnsD gel binding assays in the presence of TnsC. This may reflect my use of crude cell extracts, in particular a strong host factor complex is observed in these retardation assays, it is possible that binding by this factor inhibits TnsC interactions with TnsD in these crude cell extracts.
It would appear that the interaction between the Tns proteins may involve all four proteins. Because of the presence of chromosomal DNA in the crude extracts used it is not possible to determine if the interaction occurs on DNA or in solution; presumably both Tn7 right end, TnsB binding, motifs from pSB84 and *attTn7* DNA, from sheared chromosomal DNA are present. There is no similar ATP dependance on TnsB binding observed in pSB84 extracts (data not shown), this, however, does not preclude a role for TnsB in this process. On the basis of western blotting (Orle and Craig 1991), and analysis of transcriptional and translational fusions to Tn7 genes (Ekaterinaki 1987) TnsB appears to be the most highly expressed levels of the various *tns* genes. It is possible, therefore that only a small fraction of TnsB in the pSB84 containing cells is involved in this

putative interaction. In comparison, expression of TnsD is very low and it is possible that the higher levels of the other Tns proteins allow all the TnsD present in the cell to be sequestered in this way.

The observation of the Mg^{++} dependance of TnsD binding coupled with the requirement for ATP hydrolysis to detect its binding in pSB84 extracts suggest a reason for the inability to reconstruct a Tn7 *in vitro* system.

In the initial preincubation of Tns proteins, ATP and DNA substrates, the TnsD in the extract will be sequestered, on addition of Mg^{++} to 15mM, ATP hydrolysis will occur and TnsD will be released, however at these high concentrations of Mg^{++} TnsD does not bind to *attTn7* and therefore no rection can occur. Unfortunately there was insufficient time available to test this hypothesis, by performing an in vitro reaction under the conditions similar to those used in the gel retardation assays. It is possible that the inclusion of ATP and 5mM Mg^{++} in the reaction from the start may allow TnsD to bind to *attTn7* and allow transposition to occur.

Chapter 6

Host Factor Binding at *attTn7*

6.1 Introduction

Although the nicking and strand exchange reactions in transposition of an element are carried out by a transposon encoded protein, the transposase, a growing body of evidence suggests that several small host encoded proteins can play a stimulatory role in these processes. Bacteria possess a number of related small DNA binding proteins which, on binding, influence the local structure of DNA and play roles in diverse cellular processes. The best characterised of these proteins are Integration Host Factor (IHF), HU, FIS and H-NS. FIS and IHF show sequence specificity in their DNA binding (Craig and Nash 1984 ; Finkel and Johnson 1992), whereas HU and H-NS do not. Recent work has suggested that while H-NS may not show a sequence specificity it does preferentially associate with intrinsically bent DNA (Owen-Hughes, Pavitt et al. 1992). All of these proteins have been demonstrated to play roles in a variety of recombination reactions and, in several cases, have been shown to play structural rather than enzymatic roles.

Integration Host Factor (IHF) is a heterodimeric protein of approximately 20kd, composed of an alpha and a beta subunit. It was, as its name suggests, first identified as an essential host encoded component in the *in vitro* lambda integration reaction. Lambda is unable to establish lysogens in cells mutant in either α or β subunit, this enabled the identification of the two genes encoding the subunit proteins, *himA* and *hip*.(Friedman 1988)

IHF belongs, along with HU, to a class of small "histone-like" proteins found in the bacterial nucleoid. IHF shows considerable sequence homology to HU. However, where HU is known to bind DNA non-specifically, IHF has been demonstrated to be a sequence specific DNA binding protein. A consensus sequence for binding has been derived (Craig and Nash 1984).
The similarity between IHF and HU has made possible the use of molecular modelling approaches based on the known HU crystallographic structure to suggest a structure for IHF. This proposed structure along with DNA footprinting data has led to the proposal that IHF makes contacts with DNA in the minor groove. In particular, studies of interference due to alkylation of bases have been highly informative.

Alkylation of nearly all adenine in an IHF binding site, which occurs in the minor groove, is highly deterimental to IHF binding. By contrast alkylation of guanine bases, a major groove effect, shows only a few alkylated guanines interfere with binding (Yang and Nash 1989).
By analogy with the HU crystal structure contact with the minor groove may be mediated by a two stranded beta ribbon, on both subunits of the IHF. Some recent work, however, has revealed considerable differences in the interactions of HU and IHF with DNA. A single HU heterodimer interacts with only 9 bp of DNA, whereas IHF protects about 40bp from nuclease attack (Bonnefoy and Rouviere-Yaniv 1991). Analysis of the stoiciometry of IHF-DNA complexes have revealed that a single IHF heterodimer is responsible for this large region of protection (Yang and Nash 1989). It appears therefore, that IHF may make additional contacts with DNA outwith the conserved WATCAANNNNNTTR region.

Whatever the mode of interaction of IHF with DNA, it is clear IHF is important in many other varied processes in addition to lambda integration. These processes include control of plasmid pSC101 DNA replication, F factor transfer, regulation of gene expression and transposition of a variety of elements (Friedman 1988 ; Freundlich, Ramani et al. 1992).

In vitro studies have shown IHF binding to preferred transposition target sites for both IS1 and γδ (Tn1000) (Gamas, Chandler et al. 1987 ; Prentki, Chandler et al. 1987 ; Wiater and Grindley 1988). Also, IHF binds to the inverted terminal repeats of IS1 and also the ends of IS10 and γδ. For γδ it has also been demonstrated that there is mutual co-operativity in the binding of binding of transposase and IHF to the ends of the element (Wiater and Grindley 1990).

Despite these well characterised *in vitro* phenomena *himA* or *hip* mutations appear to have no effect on the level of transposition of IS1 *in vivo* (Galas and Chandler 1989).
Intriguingly, while transposition of a γδ derivative which lakcks the terminal IHF binding sites has been shown to be reduced about three fold in IHF^- cells, transposition of the wild type transposon, with its terminal IHF sites, is actually stimulated about three fold in IHF^- cells. How these effects are mediated is unclear. The principle effect that IHF has *in vivo*

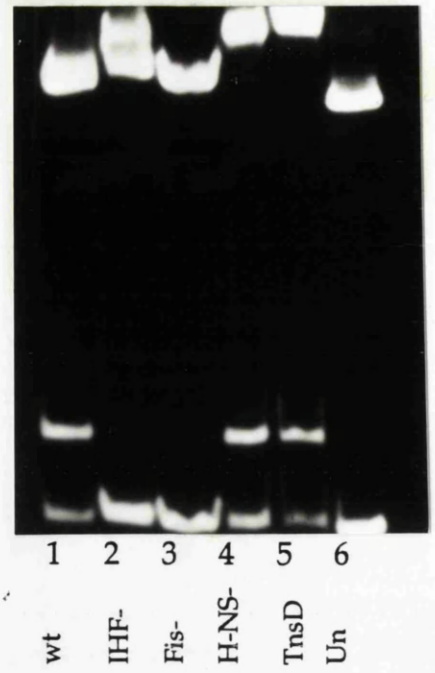
appears to be to stimulate the level of transposition immunity conferred by γδ ends. This effect is presumed to be due to the stimulation of γδ transposase binding (Wiater and Grindley 1990).
Finally, it has been shown that IS10-induced rearrangements occur at a 5 to 9-fold reduced frequency in *hip* or *himA* mutants relative to wild type cells (Kleckner 1989).

The relatively small loss of activity *in vivo* in IHF mutants may be because HU appears to be able to compensate for loss of IHF activity to a limited extent. This is reflected in the discovery that while IHF$^-$ cells and HU$^-$ cells are viable, mutations in both IHF and HU are conditionally lethal (Kano and Imamoto 1990). This partial functional redundancy is also true *in vitro* where it has been observed that HU can replace IHF, albeit relatively inefficiently, in a variety of reactions.

It initially appears puzzling that such a small protein can influence such a wide range of biological activities. However, the observation that IHF induces a sharp bend in DNA on binding suggests that its role in the above processes may be structural rather than enzymatic. In the case of lambda site specific recombination it has been demonstrated that an intrinsically bent stretch of DNA can functionally replace an IHF binding site (Goodman and Nash 1989). Whilst the effect that IHF has on altering the local DNA conformation may have direct effects on alignment of DNA sites; it is clear that IHF can also influence the binding of other proteins to DNA. It is possible that this effect is a consequence of the alteration of DNA structure, IHF binding may induce a conformation more easily recognised by a second protein. as stated above, positive cooperative binding interactions have been observed between IHF and the γδ transposase at the ends of γδ and also between lambda integrase and IHF at *attL* and *attR*. (Lee, 1992).

6.2 A role for IHF in Tn7 Transposition ?

Waddell and Craig, in their characterisation of TnsD binding to the hot-site, proposed that an unidentified host-factor bound specifically to *attTn7* sequences between the TnsD binding site and the *Tn7* insertion site (positions +10 to +27, see Fig. 5.1) (Waddell and Craig 1989). No attempt was made to identify this factor, although it was noted to be heat-stable.



Lane 1: 10µg DS941 extract
Lane 2: 10µg DS978 extract
Lane 3: 10µg DS940 extract
Lane 4: 10µg GM230 extract
Lane 5: 10µg DS941 pMR64 extract

Reactions were carried out under standard TnsD binding conditions

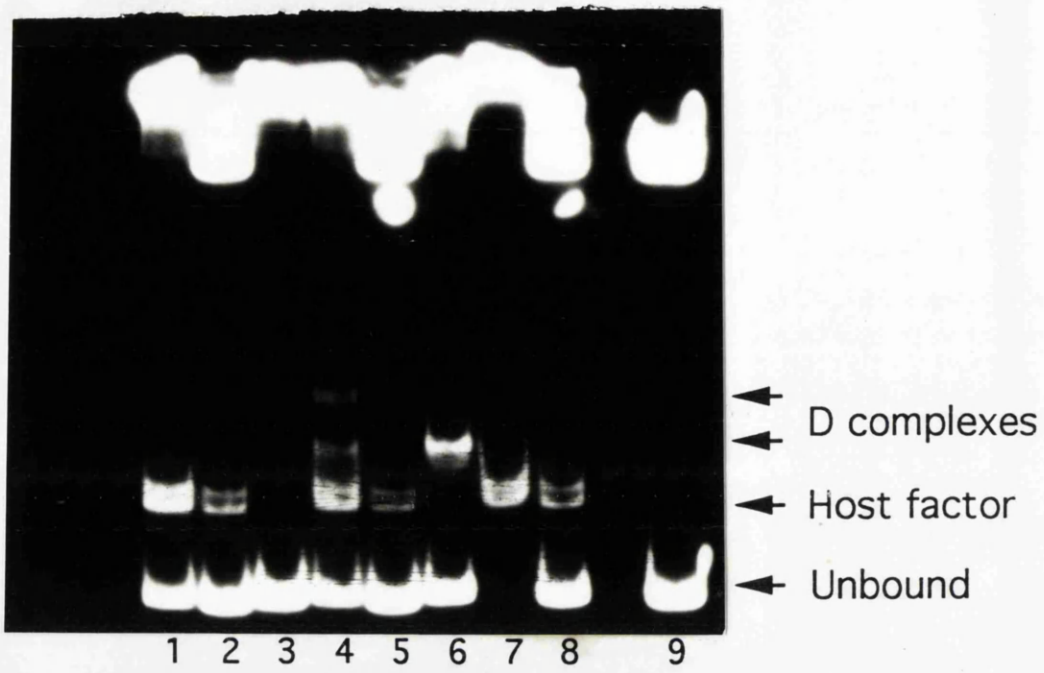
Fig. 6.1 Identification of Host Factor binding to *attTn7*

As discussed in the previous chapter, I observe a doublet host factor dependent band in gel retardation assays using crude extracts from a TnsD overexpressing strain and *attTn7* , -180 to +90. The additional retarded species is presumably due to my use of a larger DNA molecule containing an extra, unknown, binding site. At this stage in my studies I assumed that one of these bands was due to the interaction observed by Waddell and Craig, the other being due to an as yet uncharacterised binding activity. It should be stressed, however, that these complexes represent a specific binding; all binding reactions were carried out in the presence of 2µg of Salmon Sperm DNA as competitor.

In order to try to identify the host factor(s) responsible for forming these complexes, gel retardation assays were performed using extracts prepared from DS941 (positive control), DS978, DS940 and GM230· These latter 3 strains lacked FIS, IHF and HNS respectively. As can be seen from Fig 6.1 the host-factor activity was unaffected in *osmZ*$^-$, reduced in *fis*$^-$ cells and completely absent in *himA*$^-$ cells. This suggests that IHF may be the host factor responsible for both retarded complexes. The reduction in host factor binding in the *fis*$^-$ strain may be a result of a direct involvement of FIS in expression of *himA* and/or *hip*.. There is, however, no published evidence to support this hypothesis. It is also possible that there is an indirect effect. IHF levels have been shown to increase as *E. coli* enters stationary phase. The growth rate of *fis*$^-$ cells is considerably reduced relative to DS941 (J. Roberts pers. comm.), this may account for the low level of host factor binding seen.

The various strains tested had diverse genotypes; therefore I introduced the *himA::Tn10* allele into DS941 by P1 transduction, thus generating otherwise isogenic IHF$^-$ and IHF$^+$ strains. The DS941 (*himA::Tn10*) strain was designated SB10. The IHF$^-$ phenotype was checked by preparing competent DS941 and SB10 and transforming with the spectinomycin resistant pSC101 derivative, pLA108. Plasmids with a pSC101 origin cannot be maintained in IHF$^-$ strains (Stenzel, Patel et al. 1987). The competent cells were separately transformed with pUC18 as a control. The inability to recover SB10 pLA108 transformants confirm that SB10 is IHF$^-$ (see Fig. 6.2)

SB10 and DS941 were then transformed with pMR64 (*tnsD*$^+$). Extracts were prepared from the various strains and again assayed by gel retardation (Fig. 6.3). The result agrees with the previous finding that the



Lane 1: DS941
Lane 2: Heat treated DS941
Lane 3: DS941 (*himA*)
Lane 4: DS941 pMR64
Lane 5: Heat treated DS941 pMR64
Lane 6: DS941 (*himA*) pMR64
Lane 7: 20 ngPurified IHF
Lane 8: 2ng of IHF
Lane 9: No extract

Lanes 1-6 all contain 20μg of crude extract prepared as described in Materials and Methods.
Retardation reactions performed as described in Materials and Methods.
Binding conditions are "standard" TnsD conditions

Fig. 6.3 Host Factor Binding

Strain / Plasmid	DS941	SB10
pUC18	2.1×10^5	4.6×10^5
pLA108	1.9×10^5	0

Fig. 6.2 Transformation of DS941 and SB10 by pUC18 or pLA108. Figures in transformants per μg of DNA

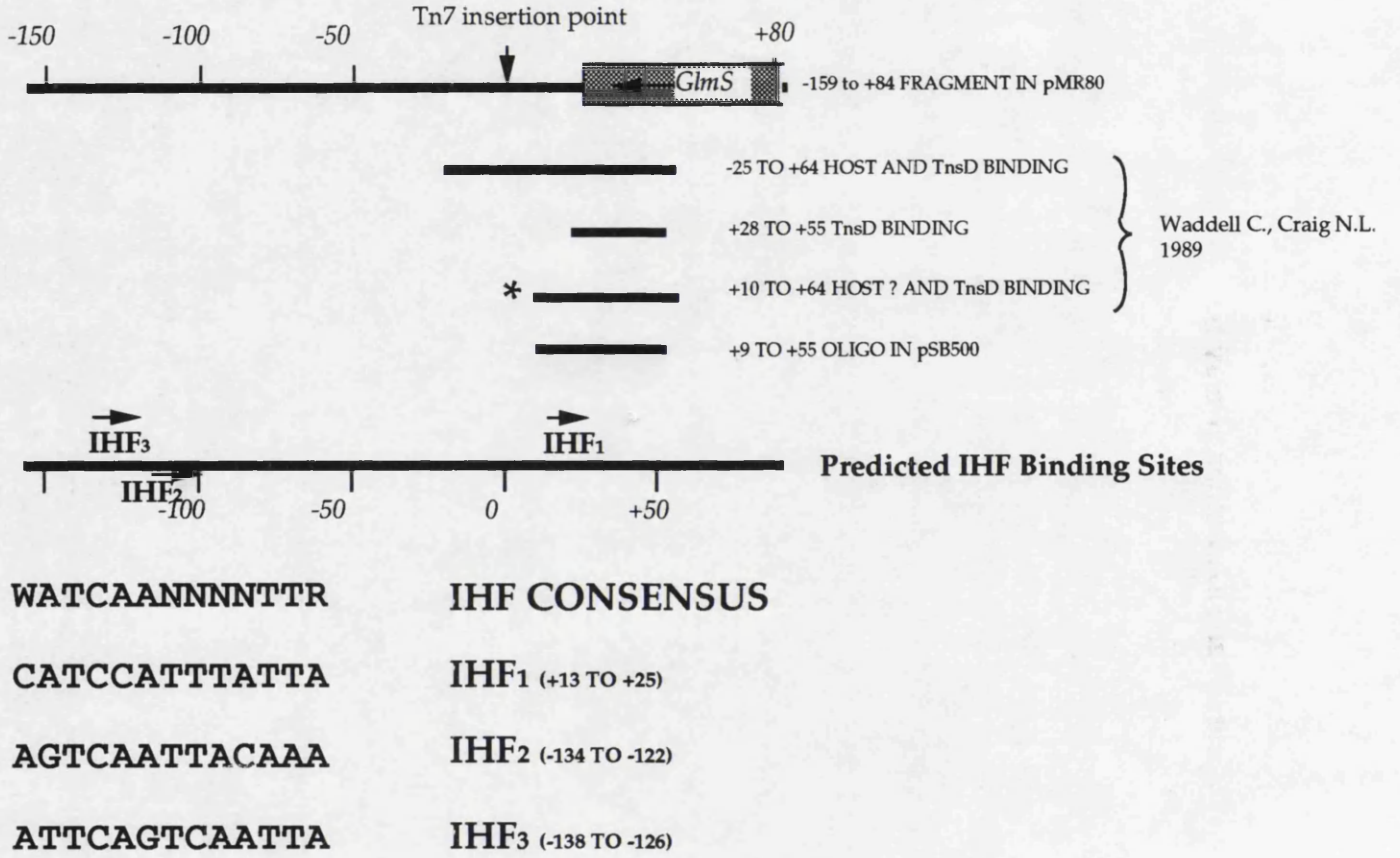


Fig. 6.4:
Summary of DNA binding data at*attTn7* .
Location of host factor and TnsD binding sites are shown
An oligo of the fragment marked * binds a host factor weakly
but does not compete for host binding to -25 to +55 (See Waddell & Craig,1989 Fig.3 Lane 8)
Potential IHF binding sites are also indicated.

host factor binding is not present in *himA* mutants. In addition to this it can be seen that whilst two TnsD specific complexes are detected with DS941 pMR64 extract, only the lower of these is observed with DS941(*himA*) pMR64 extract. This suggests that the lower D-specific complex may be TnsD interacting with naked DNA whereas the upper complex may be due to both TnsD and host factor binding to DNA.

IHF has a wide range of activities, therefore it is possible that IHF itself may not be the factor binding to *attTn7*. Rather, it may be required for the expression of a separate factor which is responsible for this host binding activity.
Strong evidence for a direct role in IHF binding is shown in figure 6.3. IHF has the unusual property of being highly heat resistant, retaining activity even after 10 minutes at 100°C, this feature is shared by several small DNA binding proteins including Fis. DS941 extract retained host binding activity after a 10 minute immersion in boiling water. Furthermore, a partially purified IHF preparation showed a high level of binding activity. Finally, on the basis of sequence analysis, there are several potential IHF binding sites in the pMR80 sequence (Fig 6.4). This is discussed in more detail below. Put together these data strongly support the idea that IHF itself binds *attTn7*.

As stated above, it is observed that the host factor band resolves as a doublet on extended electrophoresis. It should be noted however, that this is not always the case(eg see Fig. 6.5) However, it is clear from Fig 6.3 that both these complexes are dependent on IHF, they are not observed when SB10 extracts are assayed. Furthermore, they are formed when partially purified IHF is assayed. Thus, it seems likely that both complexes are caused by IHF binding, an alternative explanation would require that one complex is due to IHF, the other being created by binding of a factor that is dependent on IHF for its expression and which also co-purifies with IHF. This seems unlikely. However, as discussed below, problems exist with the first interpretation , that both complexes are caused by IHF binding.
Two potential explanations exist to explain the formation of the two complexes.

1) The complexes correspond to IHF bound to different sites in the molecule, thus inducing distinct bend centres. Molecules with differently located bends would migrate differently on electrophoresis.
2) The lower and upper complexes could correspond to 1 and 2 IHF molecules bound to the DNA.

The first hypothesis is unlikely because studies on the mobility of fragments carrying multiple IHF sites have shown that IHF binding is a dynamic process. IHF appears able to transfer from site to site on a given fragment. The overall effect is to cause a singly bound multi-site molecule to migrate at a single, averaged mobility (Prentki, Chandler et al. 1987). Further data against the first hypothesis lies in the observation that even at high IHF concentrations, where no unbound DNA remains, no third, upper complex corresponding to a molecule with both presumptive sites occupied is detected.

The second hypothesis initially also seems unlikely. A naive expectation would be that the shift in mobility from unbound to singly bound would be similar to the shift from singly to doubly bound. However, this interpretation does not take the bending induced by IHF into account. Work by Snyder et al has shown that the mobility shift from singly to doubly bound molecules is dependant on the helical phasing between the IHF sites. Sites which are in phase result in an augmented bend and thus a large mobility shift. Out of phase sites effectively cancel out the respective bends and can even cause an increase in mobility of doubly bound DNA relative to singly bound (Snyder, Thompson et al. 1989).

6.3 Interactions between IHF and TnsD

The proposed IHF site IHF_1 lies immediately adjacent to the sequences implicated in TnsD binding. It might, therefore, be predicted that IHF may influence TnsD binding to the site. In the case of the lambda integration and excision reactions it has been demonstrated that there is co-operative binding of IHF and Int at *attP* and IHF, Int and Xis at *attR*. Studies of IHF binding to the termini of γδ have also shown co-operative interactions between the γδ transposase and IHF (Wiater and Grindley 1990). In order to test if similar co-operative interactions occur at *attTn7* a series of

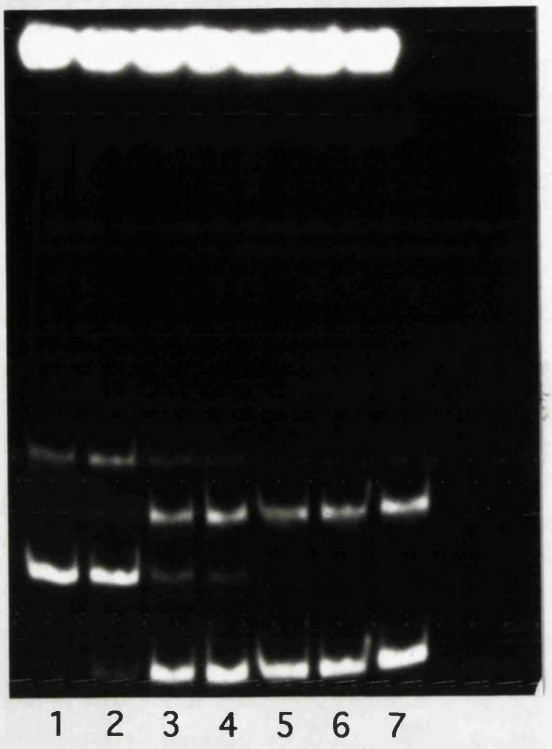


Fig 6.5: Autoradiograph of gel used to examine co-operativity in TnsD and IHF binding.

All reactions performed in standard TnsD conditions.
All reactions have 10μg of a SB10 pMR64 extract and varying amounts of IHF.

Lane 1: 20ng IHF
Lane 2: 10ng IHF
Lane 3: 5ng IHF
Lane 4: 2.5ng IHF
Lane 5: 1.3ng IHF
Lane 6: 0.6 ng IHF
Lane 7: no IHF

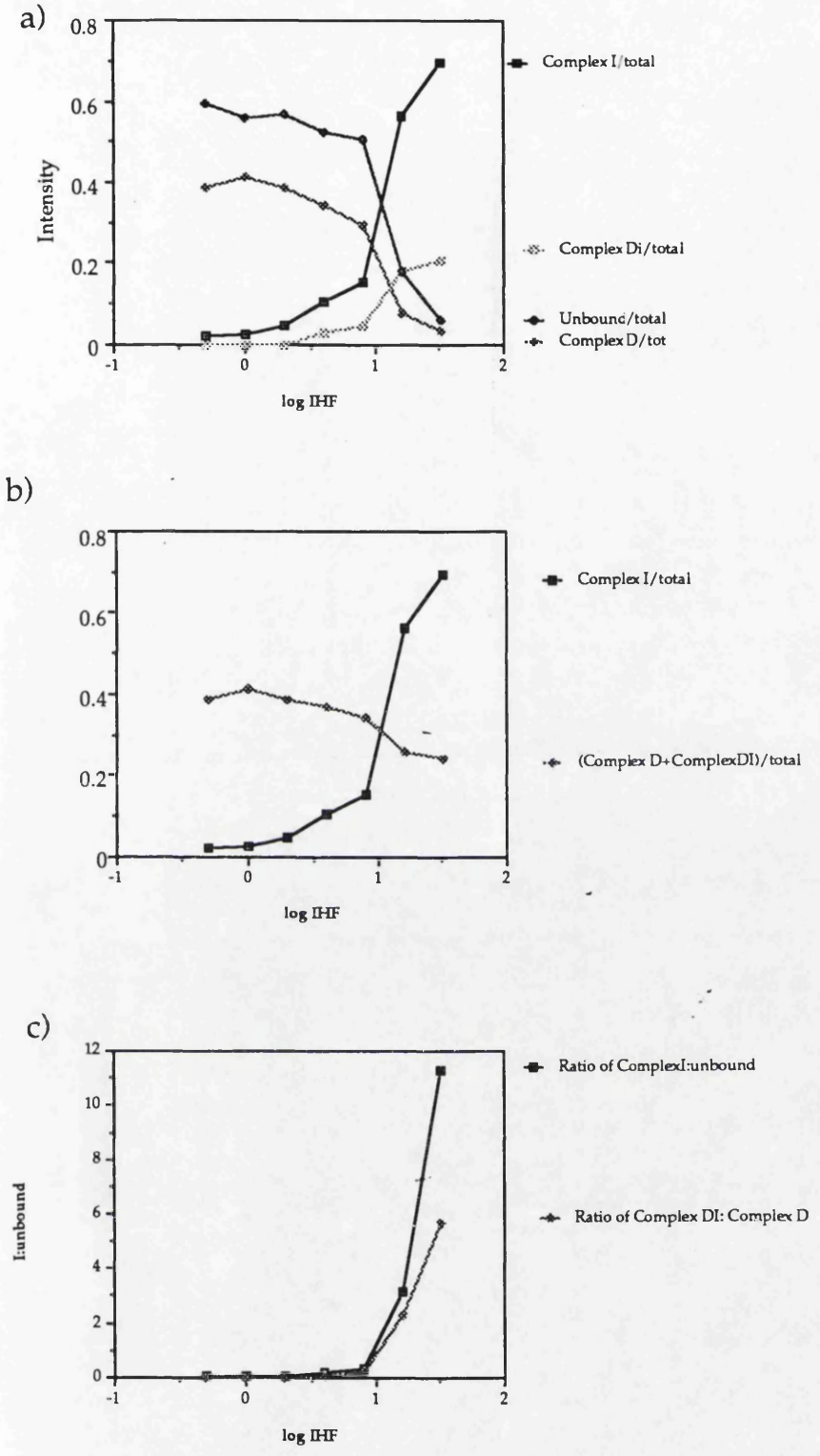


Fig. 6.6 Plot of relative intensities of TnsD and IHF complexes from Fig. 6.5
a) The intensity of individual complexes standardised against the sum of intensities at varying IHF concentratioins
b) The intensity of the IHF complexes and the sum of the intensities of complex D and DI against IHF concentration
c) The ratios of intensity of complex I/unbound and complex DI/D at varying IHF concentration.
Because the absolute concentration of IHF is not known the x-axis is in arbitary values

retardations were performed using end labelled EcoRI/HindIII digested pMR80.
This DNA substrate was incubated with a sub-saturating amount of SB10 pMR64 extract such that approximately one third of the *attTn7* containing fragment was bound by TnsD. Aliquots of this binding reaction were added to two fold serial dilutions of partially purified IHF. Binding was continued for ten minutes and then the reactions were loaded onto a 6% poly-acrylamide gel and electrophoresed. An autoradiograph of the gel is shown in Fig. 6.5. The film was pre-flashed prior to exposure to the dried gel. This was done to increase the film's sensitivity and also to attempt to obtain a linear response to signal intensity. The following discussion of the result of this experiment assumes that the film has responded in such a linear fashion.

It is apparent from the autoradiograph that IHF binds to *attTn7* independently of TnsD. If this was the case it would be predicted that the ratio of complex I to unbound DNA would be the same as the ratio of complex DI to complex D. If IHF and TnsD interact positvely then the ratio of DI to D would be greater than Complex I to unbound. If they interact negatively then DI:D would be less than I:unbound.

The autoradiograph was scanned into a digital image in a Macintosh computer using a Apple "OneScanner". The resultant image was analysed using the Image 1.41 package. The apparent intensities of the various bands of the gel were quantified and the resultant values of intensity plotted against log(IHF). In addition the ratios of complex I to unbound and complex DI to complex D intensities were calculated and plotted (Fig. 6.6). This more quantitative analysis suggests that IHF binding to the *attTn7* sequences does not facilitate TnsD binding and may actually be slightly inhibitory.

6.4 Localisation of IHF binding Site

An initial search using the core IHF binding consensus as "probe" identified the putative sites IHF_1 IHF_2 and IHF $_3$. These potential sites differ by at least one base from the consensus. In the case of IHF_1 the A-C deviation from the consensus is likely to have a highly deleterious effect on IHF binding. The relative contributions of bases within the consensus

site was examined by mutational analysis of the lambda H' site. Mutant IHF sites with decreased ability to bind IHF were identified by a phage challenge assay. A mutant site with an A-C change at this position caused a several thousand fold drop in lysogenisation frequency relative to wild type (Lee, 1991). It appears unlikely therefore that the putative IHF_1 site is a functional IHF binding site.

A feature of IHF binding sites is that a large (approximately 48nt) region is protected from nuclease digestion on IHF binding. Recent work has compiled data from 27 known IHF binding sites and proposed an expanded binding site consensus (Goodrich, Schwartz et al. 1990).

These workers complied the base scores for each position of a 48 base pair region surrounding and including each of the 27 IHF consensus sites. A similarity score can be calculated by the equation shown below

SS= 100x(Sum of base scores for candidate site-baseline score)/(maximum score-baseline score)

The similarity scores calculated for the 27 bona fide IHF sites ranged in value from 46.2 to 76.6. When the similarity score for IHF_1 was calculated it was found to be 14.2, thus this region of the *attTn7* locus is unlikely to be the binding site for IHF.

When the similarity score for IHF_3 was determined it was found that the eight base-pairs at the 5' end of the 48bp sequence belonged to pUC8 polylinker. Similarity scores were therefore calculated for both the cloned site and the native chromosomal site. These values were 47.6 and 52.4 respectively. On the basis of this analysis it appears that IHF_3 may be the binding site for IHF detected in the above retardation assays.
The similarity score was also calculated for the cloned IHF_2 site as 40.7, lower than the lowest score for a genuine site (46.2).

In order to determine which sequence elements within the 280 nt *attTn7* sequence are recognised by the various proteins various end-labelled restriction fragments were generated. pMR80 was digested with either EcoRI or HindIII and end-labelled. The resultant DNA's were purified and then digested with MboII and MseI to create the fragments shown in Fig

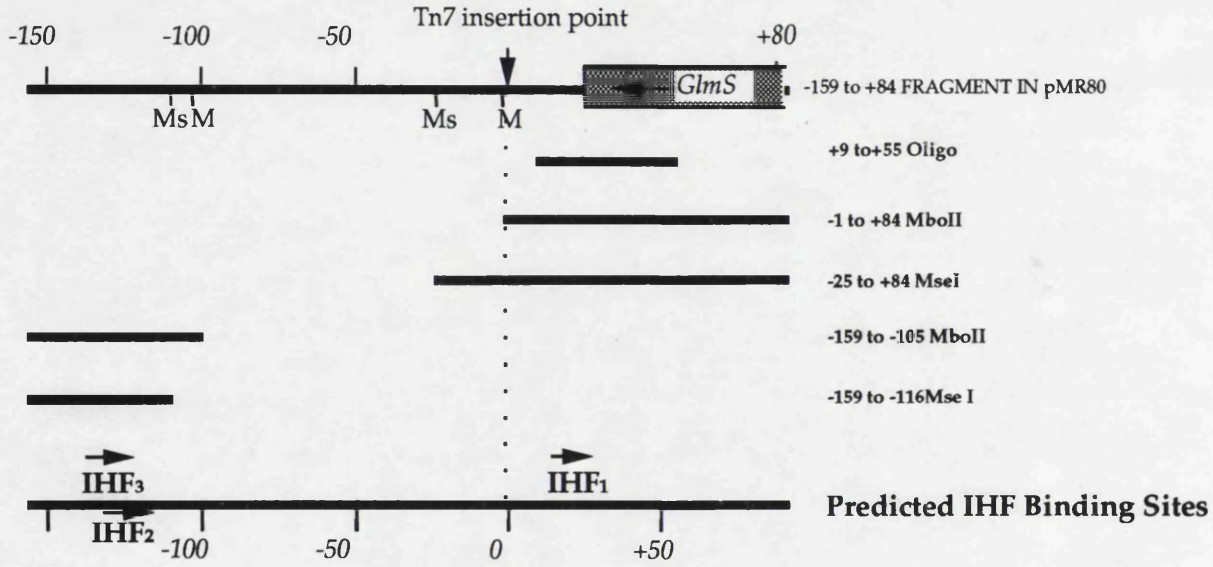


Figure 6.7:
Restriction fragments used to examine DNA binding at *attTn7* .
Potential IHF binding sites are indicated.
Ms=MseI site, M=MboII site.
Fragments were made by restricting with either EcoRI or HindIII, endlabelling, isolating the DNA then digesting with either MseI or MboII.

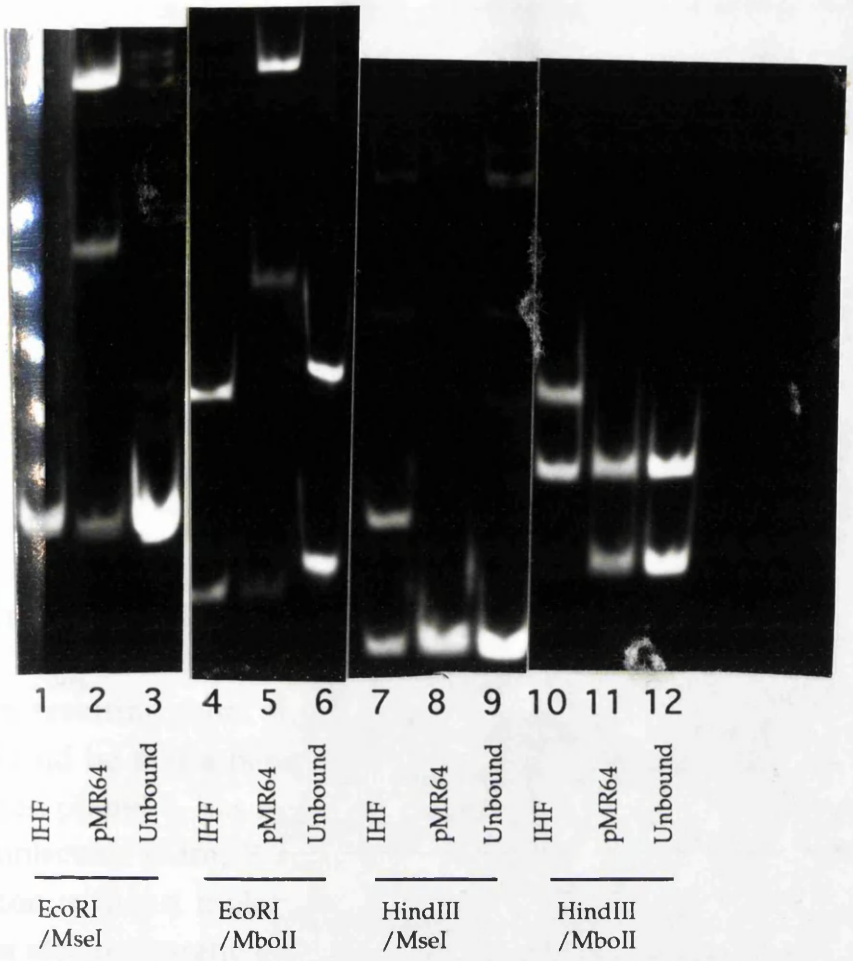


Fig.6.8 Localisation of IHF binding.
The restriction fragments in described in Fig. 6.7 were used in retardation assays under standard TnsD conditions with either 20ng IHF or 10µg DS941 pMR64 extract.
The reactions loaded in each lane are as below:

Lanes 1-3: EcoRI-MseI fragments
Lanes 4-6: EcoRI-MboII fragments
Lanes 7-9: HindIII-MseI fragments
Lanes 10-12: HindIII-MboII fragments

Lanes 1, 4,7,10 IHF
Lanes 2,5,8,11 DS941 pMR64 extract
Lanes 3,6,9,12 Unbound

6.7. Because of the scarcity of unique restriction sites in the *attTn7* sequence it was not possible to probe for binding directed to the central portion of the pMR80 insert.

Gel retardation assays were performed using purified IHF and DS941 pMR64 (IHF^+, $TnsD^+$) extract with each of the various substrate DNA's. The autoradiograph is shown in Fig. 6.8.

This result shows that, as previously demonstrated, TnsD binds to the left end of the pMR64 insert. In addition it reveals that IHF binds to the left end, presumably to the putative IHF_3 motif. No IHF binding is detected to the IHF_1 motif adjacent to the TnsD site.

6.5 DNA Bending at *attTn7*

A prediction resulting from the observation of IHF binding to *attTn7* sequences would be that a bend should be induced in the bound DNA. A plasmid vector, pBend-2, has been constructed for use as a tool in detecting such bent molecules (Kim, Zweib et al. 1989). The theoretical basis for bend detection is that a molecule with a central bend will migrate more slowly on an electrophoretic gel than an identically sized molecule with a bend nearer its end (Wu and Crothers 1984). pBend-2 exploits this observation by having a long, directly repeated polylinker with two unique cloning sites between the direct repeats. The sequence of interest is cloned into pBend-2 using these unique sites. Restriction of the recombinant plasmid by enzymes which cleave within the direct repeats generates a series of identical sized fragments with the insert DNA a varying distance from the ends of the fragments . Electrophoresis of these fragments will then allow detection of differences in mobility indicative of bending.

The 250 nt HincII fragment of pMR80 was gel isolated and ligated to blunt ended Klenow treated SalI digested pBend-2. The orientation of the insert in a recombinant clone, pSB500, was confirmed by digestion with BamHI.

1µg aliquots of pSB500 were digested with MluI, StyI and PvuII. The resultant fragments were end-labelled using Klenow, dATP, dTTP, dGTP, α-^{32}PdCTP. It should be noted that the PvuII digest was incubated with Klenow for 15 minutes prior to addition of dNTPs. This was to enable the 3′ to 5′ exonuclease activity of the enzyme to generate suitable ends for filling in. A consequence of this action was that the PvuII fragment was

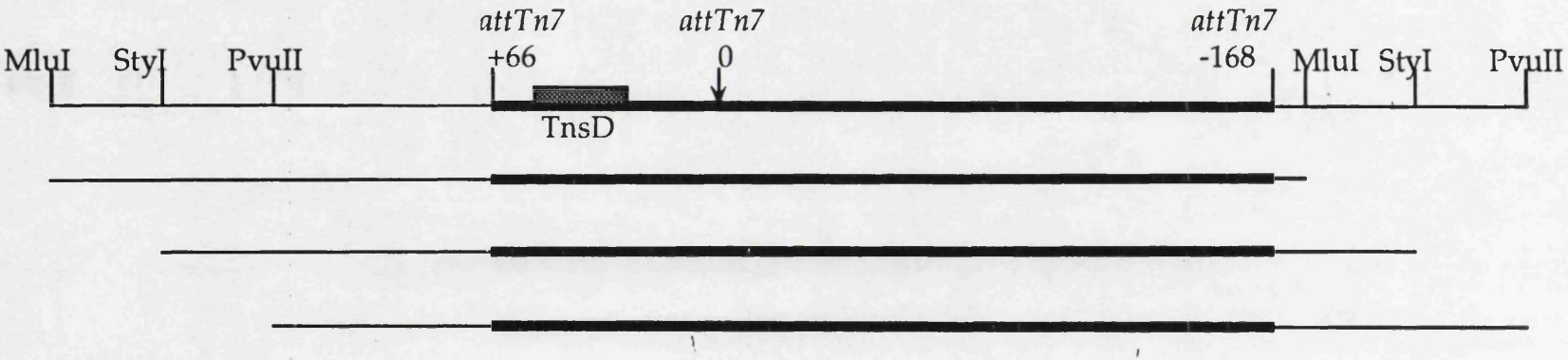


Fig 6.9a Diagram of the structure of the insert in pSB500.
Restriction with MluI, StyI or PvuII will release circularly permuted but identically sized fragments .

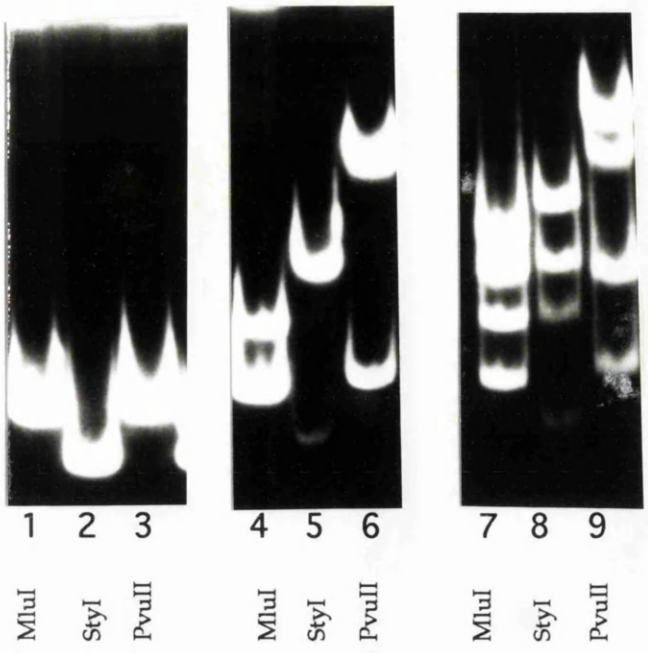


Fig. 6.9 DNA Bending at *attTn7*

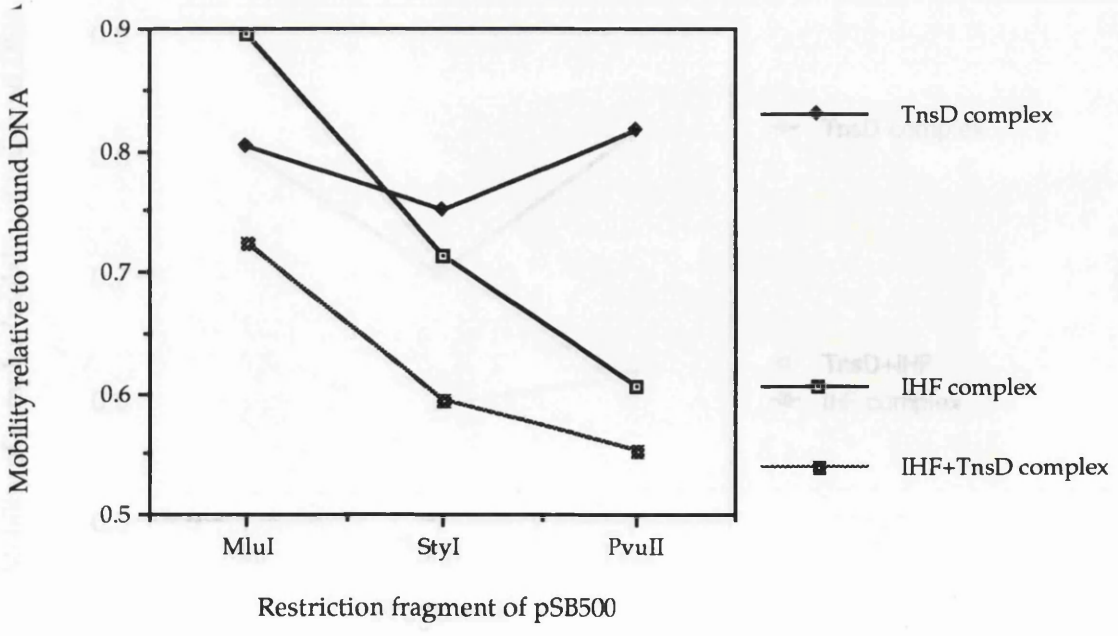
Lanes 1,2 & 3: No extract

Lanes 4,5 & 6: 20ng IHF

Lanes 7,8 & 9: 10µg DS941 pMR64 extract

Lanes 1,4 & 7: MluI digested pSB500
Lanes 2,5 & 8: StyI digested pSB500
lanes 3,6,& 9: PvuII digested pSB500

Fig. 6.9b Gel retardation of pSB500 fragments with IHF or TnsD exreacts
The vertical axis on the plot below is of mobility of complex/ mobility of unbound DNA.



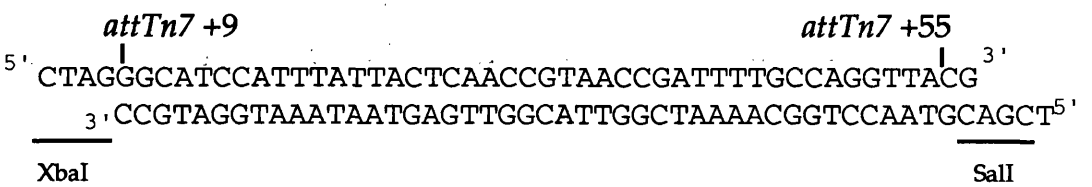


Fig. 6.10
attTn7 oligonucleotides used to make pSB400

2nt shorter than the other att-site containing molecules. This difference does not affect interpretation of the results of this experiment.
The endlabelled fragments were used in gel retardation assays with partially purified IHF and a DS941 pMR64 extract, (IHF^+, $TnsD^+$). Following binding, the complexes were electrophoresed on a 4% polyacrylamide gel which was dried and autoradiographed. The result is shown in Fig. 6.9.
It is clear that the mobility of the IHF complex varies depending on the location of the att-site relative to the end of the molecule. This indicates that, as predicted, IHF induces a bend upon binding. An examination of the data suggests that the induced bend centre is near the left end of the att-site sequence, some distance from, and on the opposite side of the Tn7 insertion site to, the TnsD binding site. This is in good agreement with the findings of the experiment to localise the IHF binding described above.
The data also suggest that TnsD binding itself also induces a bend upon binding, however this bend is considerably less than that induced by IHF. There also apears to be an intrinsic bend within the *attTn7* sequences, the centre of this bend is situated towards the right end of the molecule.

6.6 Binding to *attTn7* +9 to +55

The results of the localisation experiments above show only TnsD binding to the EcoRI MseI fragment. However, Waddell and Craig show a host factor binding activity directed to this region. They further purport to have localised this activity to be directed to *attTn7* +10 to +27
In order to address this matter I synthesised two oligonucleotides corresponding to both strands of *attTn7* sequences +9 to +55, those regions shown by Waddell and Craig to bind both TnsD and the host factor. The insertion site proximal, left, and insertion site distal, right, ends of the annealed oligonucleotides had 5' extensions compatible with XbaI and SalI restricted DNA respectively (See Fig. 6.10). The oligonucleotides were deprotected and gel purified as detailed in Materials and Methods. They were then annealed and ligated into XbaI/SalI digested pUC18 to generate pSB400. The insert in this plasmid was confirmed by DNA sequencing.

The DNA substrate used for the initial retardations in this section was an end-labelled EcoRI/HindIII digest of pSB400. The *attTn7* containing

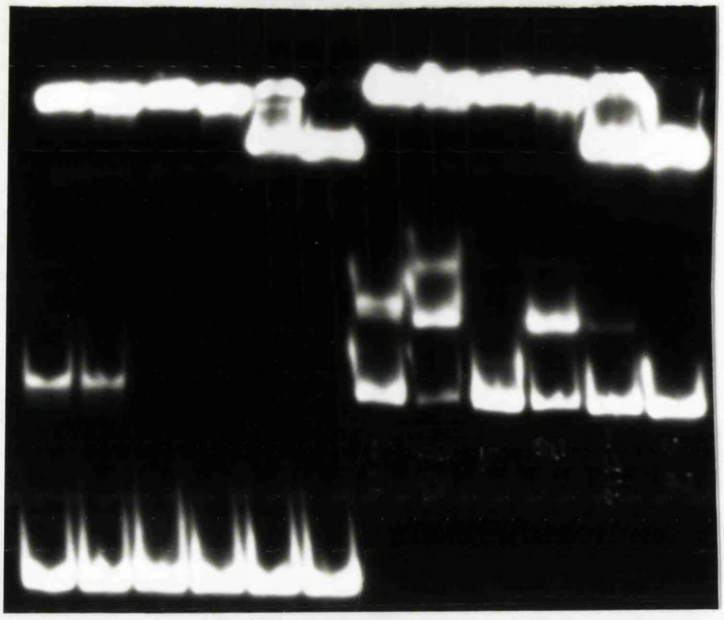


Fig. 6.11: Binding to cloned +9-+55 oligonucleotide

Reactions were in standard TnsD conditions.

Lanes 1&7: 10 µg SB10 pMR64 extract
Lanes 2&8: 10 µg DS941 pMR64 extract
Lanes 3&9: 10 µg SB10 extract
Lanes 4&10: 10 µg DS941 extract
Lanees5&11: 20ng IHF

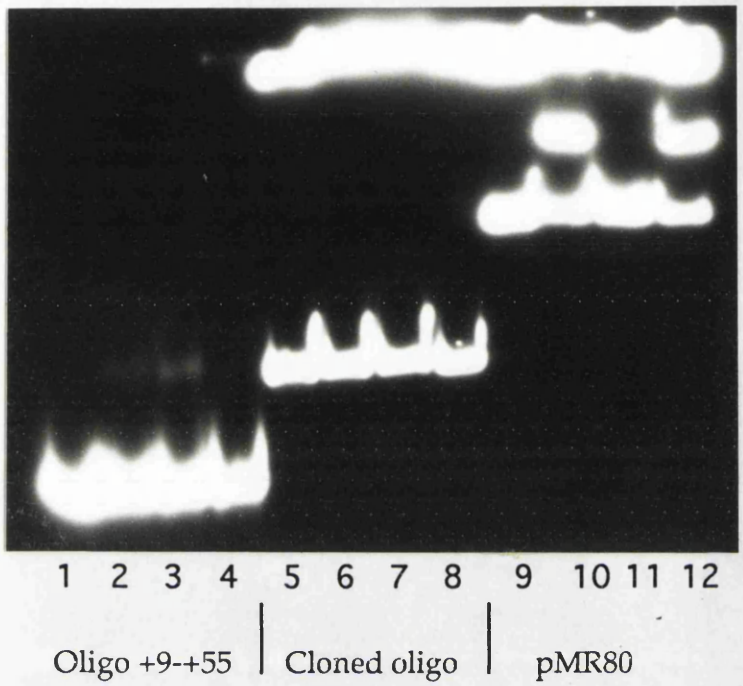


Fig. 6.12 Binding to various substrates by TnsD and IHF.
Reaction conditions standard TnsD

Lanes 1-4 Labelled substrate is annealed oligonucleotide +9- +55
Lanes 5-8 The substrate is a labelled EcoRI/HindIIIdigest of pSB400
Lanes 9-11 The substrate is labelled EcoRI/HindIIIdigest of pMR80

Lanes 1,5&9: No extract
Lanes 2&6: 10 µg DS941 extract
Lanes 3,7&11: 10 µg SB10 extract
Lanes 4,8&12: 20ng IHF

fragment therefore comprised the cloned oligonucleotide flanked by pUC18 polylinker sequences. DNA binding assays were carried out using equal quantities, 20 μg, of DS941, SB10 (*himA*$^-$), DS941 pMR64 and SB10 pMR64 extracts, assays were also performed using 20ng of partially purified IHF. The results are shown in figure 6.11. A single retarded species was seen with both TnsD containing extracts, surprisingly no retarded species are observed with DS941 extract or IHF. Binding reactions with the 280nt fragment of pMR80 were also done as controls for binding activities in the extracts.

This result initially appears at variance with that of Waddell and Craig, that *attTn7* +10 to +27 confers a host binding activity. In their experiment an annealed and endlabelled oligonucleotide of +10 to +64 is shown to bind a host activity. Thus their DNA substrate differs from the one I used in three ways:-

1) It possesses *attTn7* +55 to +64.
2) It is not flanked by pUC18 polylinker sequences.
3) It is synthetic DNA, it will neither be methylated nor 5' phosphorylated.

The latter difference is dependent on the DNA possessing a methylation signal. There is a *dcm* methylation site centred at *attTn7* +48.

The experiment was repeated, this time including the use of end-labelled annealed oligonucleotide *attTn7* +9 to +55 as substrate (Fig. 6.12). This molecule was observed to bind an activity present in both DS941 and SB10 but not to bind partially purified IHF. Thus it would appear that either some unknown feature imposed by the pUC polylinker, or methylation differences prevents binding by this host factor to digested pSB400. However, this is clearly an IHF independent phenomenon, unlike the binding activities described above.

A close examination of the data presented by Waddell and Craig allows re-interpretation of their results. They see a host binding activity directed to the synthetic +10 to +64 and also to cloned *attTn7* -25 to +64. Their assumption is that these activities are the same. However, the synthetic fragment, whilst able to compete effectively for TnsD binding, does not

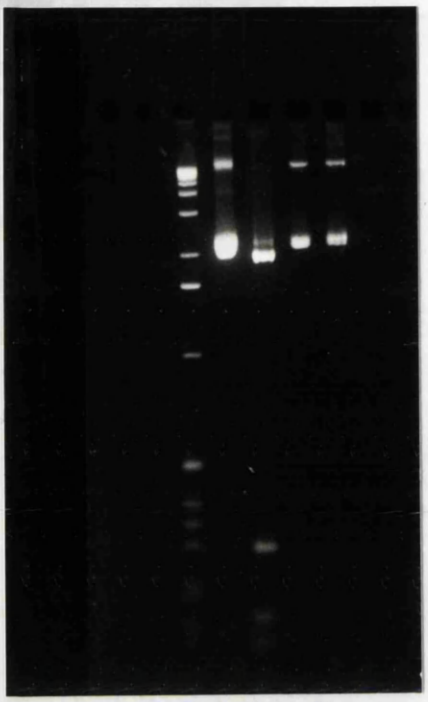


Fig. 6.13 Confirmation of *dcm* phenotype.
Agarose gel of pSB400 DNA prepared from DS941 and ET12567 electrophoresed uncut or restricted with EcoRII prior to loading.

Lane 1 BRL1 kb ladder
Lane 2 Un-cut pSB400 (ex ET12567)
Lane 3 EcoRII digested pSB400 (ex ET12567)
Lane 4 Un-cut pSB400 (ex DS941)
Lane 5 EcoRII digested pSB400 (ex DS941)

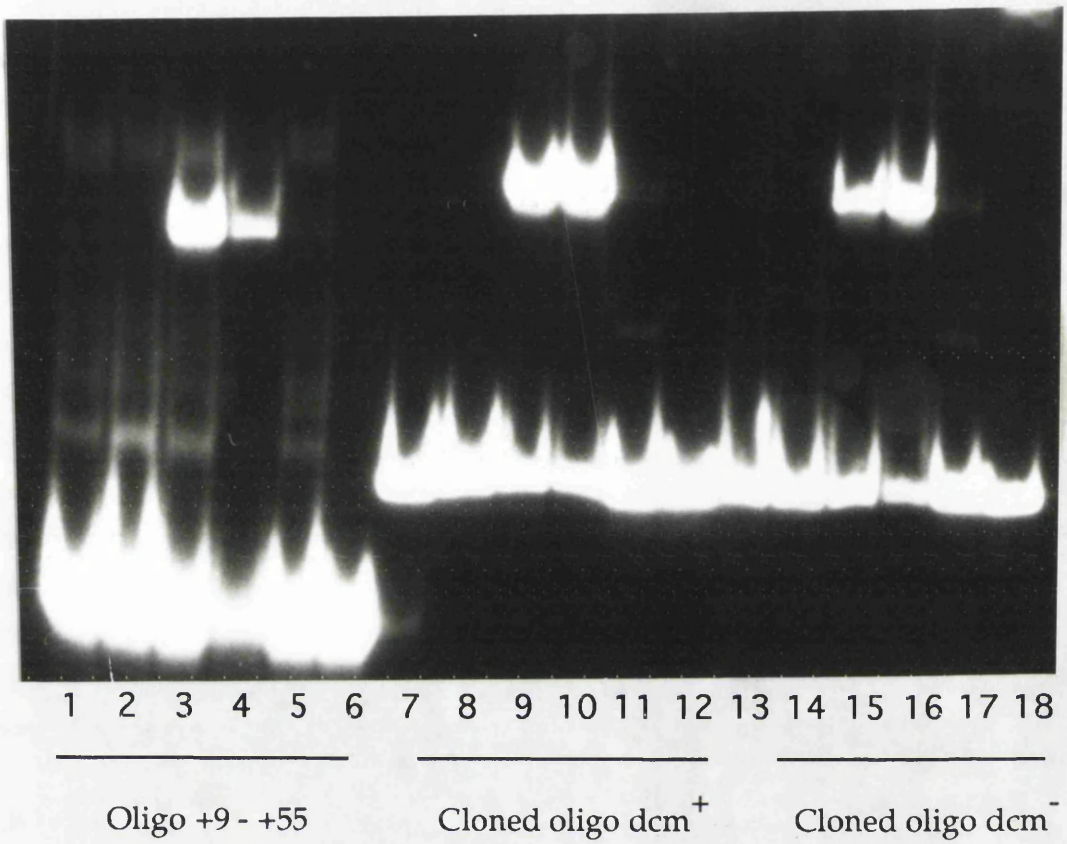


Fig. 6.14: Effect of dcm methylation on binding to cloned *attTn7* +9- +55.

Lanes 1-6: DNA substrate is annealed oligonucleotide *attTn7* +9- +55
Lanes 7-12: DNA substrate is labelled EcoRI/HindIII digest of pSB400, prepared from DS941
Lanes 13-18:DNA substrate is labelled EcoRI/HindIII digest of pSB400, prepared from ET12567 i.e no dcm methylation

Extracts

Lanes 1,7,13: 10µg DS941 extract
Lanes 2,8,14: 10µg SB10 extract
Lanes 3,9,15: 10µg DS941 pMR64 extract
Lanes 4,10,16: 10µg SB10 pMR64 extract
Lanes 5,11,17: 10µg ET12567 extract
Lanes 6,12,18: no protein extract

compete for host factor binding to *attTn7* -25 to +64.(Waddell and Craig (1989) Fig.2 Lanes 8 and 9) This argues that two separate activities are being observed.

6.7 Methylation at dcm site

In order to determine whether the binding activity directed to the oligonucleotide is due to the lack of methylation a *dcm*$^-$ strain was obtained, a kind gift of Dr. Iain Hunter. This strain was transformed with pSB400. L-broth was inoculated with this strain and allowed to grow for approximately 40 generations, thereby reducing the initial methylated DNA to an insignificant level in the population. Plasmid DNA was prepared by large scale alkaline lysis method as described in Materials and Methods. The non-methylated status of this DNA was confirmed by restriction digestion with EcoRII, this enzyme cuts DNA only at non-methylated *dcm* sites (Fig. 6.13). Having confirmed the absence of methylation 1μg of pSB400 (*dcm*$^-$) was digested with EcoRI and HindIII and end-labelled. This DNA was then used in the retardation assays shown in Fig. 6.14.
This figure indicates that the methylation status of the pSB500 insert does not influence binding by a host factor.

6.8 Consequences of IHF binding *in vivo*

6.8.1 Effect of IHF on Tn7 transposition *in vivo*

Previous work in this laboratory had suggested that there was a ~6 fold decrease in apparent hot-site transposition frequencies in IHF$^-$ cells relative to IHF$^+$ cells (N. Ekaterinaki, unpublished results). These data were derived from 3 sets of mate out assays using non-isogenic IHF$^+$ and IHF$^-$ cells as donor. Two IHF mutants were used,*himA::Tn10* and *hip* ::Cm, in which a chloramphenicol resistance cassette is inserted in the *hip* gene.
In order to determine the effect of lack of IHF in otherwise iso-genic strains, Tn7-1 was introduced into SB10 by P1 transduction. The P1 lysate as made from DS941::Tn7-1. Mate out assays were carried out as detailed in Materials and Methods, using DS941::Tn7.1 pSB84 pEN300 and SB10::Tn7.1 pSB84 pEN300 as donor and DS916 as recipient. The assays

Donor strain	DS941	SB10
	8.0	1.6
	2.3	1.1
Apparent	6.6	0.4
Transposition	9.6	0.76
Frequencies	3.9	0.81
(% of	5.6	0.62
exconjugants	5.7	0.46
that are Cm	2.1	1.0
resistant)	2.2	0.86
	4.0	4.3
Average	5.0	1.2

Fig. 6.15 Table of apparent transposition frequencies of IHF+ and IHF- cells. Mate-out assays were performed as detailed in the text.

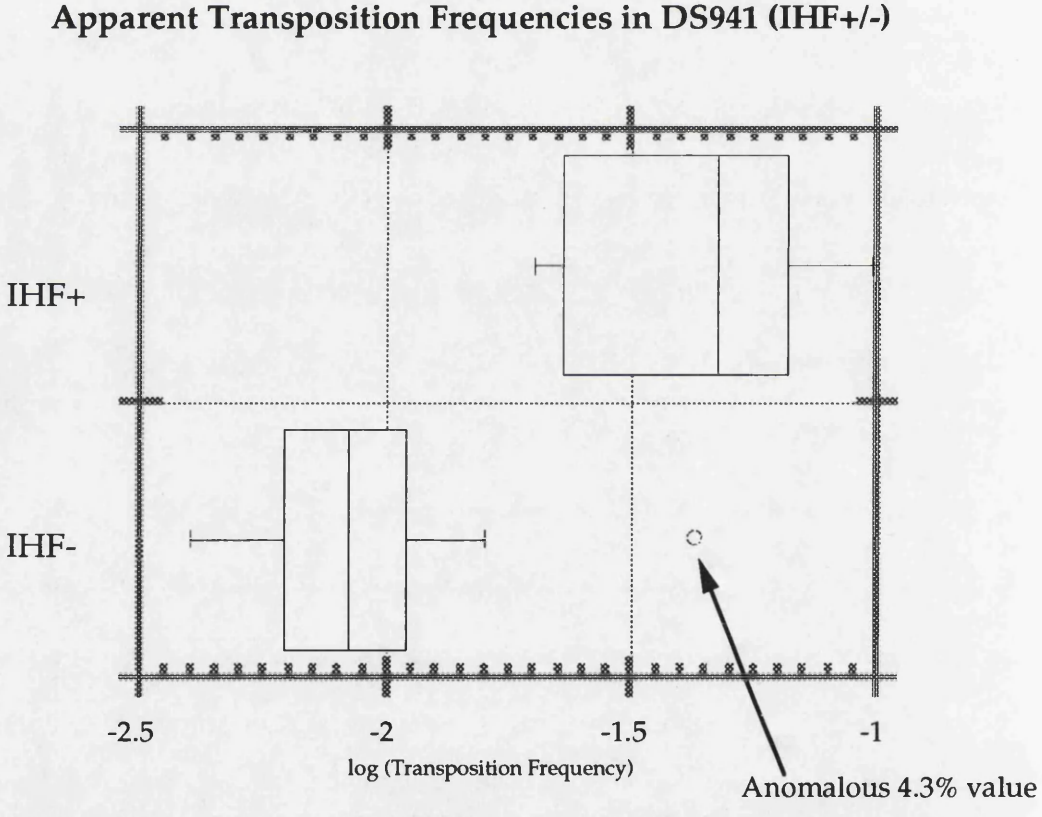


Fig. 6.16
Box plot of values of log(apparent transposition frequency) for IHF+ and IHF- donor cells
Boxes represent the range of values with the mean indicated as a vertical line. Error bars of 2x Standard Error from mean are also shown.

were performed on three separate occasions each time using freshly prepared donor strains. In total 10 colonies of both donor strains were assayed. The results obtained are shown in Figs 6.15 and 6.16. These results suggest a lower level of apparent transposition in the IHF$^-$ donor strain. One data point, 4.3%, in the IHF$^-$ set appears anomalous.

A *t*-Test was employed to test the statistical significance of the two result sets. If the 4.3% value of the IHF$^-$ data set was included the test gave a value of t=5.43, $p<0.001$, indicating a statistically significant difference between sets. The means of the two sets differ by a factor of 2.5. If the 4.3% value is discounted then the difference between means rises to 6 fold.

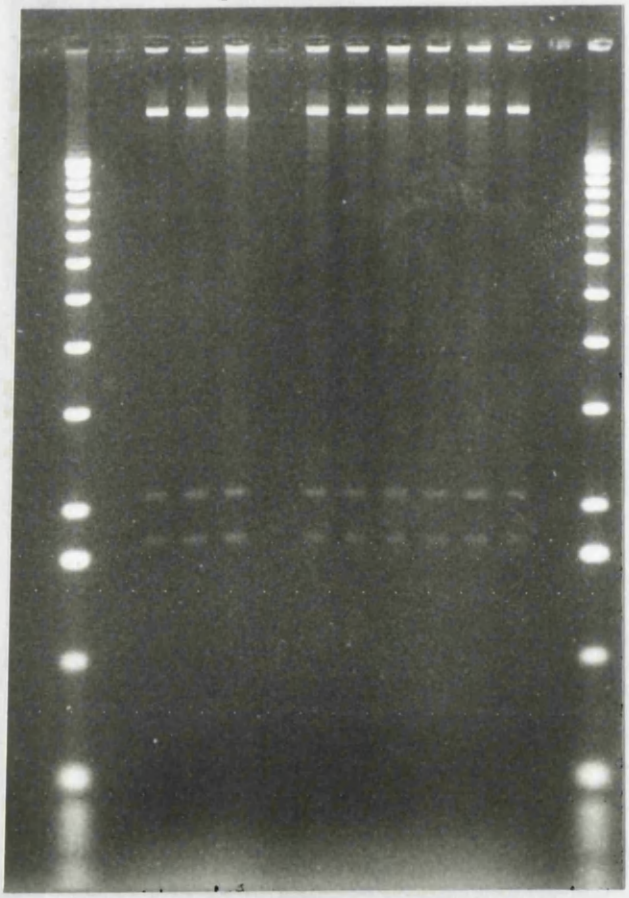
The results suggest a small (2.5-6 fold) but significant difference between the transposition levels assayed in the IHF$^+$ and IHF$^-$ cells. The donor strain used in the above assays were isogenic apart from the *himA::Tn10* allele. Thus the reduction in apparent transposition may caused by some unknown form of interference between Tn7 and Tn10. However, when these data are added to those of N. Etkaterinaki, who used a two different IHF mutations, it appears likely that the lack of IHF is mediating this reduction. With such an insensitive assay it is not possible to determine at which stage in the transposition process this effect is exerted. However, the IHF dependent *attTn7* binding activity suggests it may be a direct effect. The finding by the Craig group that a boiled and spun host extract stimulates the *in vitro* Tn7 transposition assay adds further circumstantial evidence to this proposal (Bainton, Gamas et al. 1991, R. Bainton *pers. comm.*).

The levels of conjugation were essentially identical in each set of assays, thus there is no obvious loss of ability to transfer pEN300 in the IHF mutant. Thus, unlike the case of the F-factor and R100-1, R388 conjugative transfer is unaltered in the *himA* strain (Dempsey 1987).

The location of the insertion event was determined by EcoRI digestion of plasmid DNA prepared from ex-conjugants. The restricted DNA was then electrophoresed on a 0.8% agarose gel (Fig. 6.17). The insertion sites appear to be the same in both assay sets. A further 10 exconjugants were tested, one had an insertion at a novel location in the 1kb *attTn7* containing EcoRI fragment of pEN300, data not shown. This is not

Fig. 6.17 Agarose gel of digested pEN300::Tn7-1.

The markers are BRL 1 kb ladder, lanes contain EcoRI digested mini-preps of plasmid DNA from exconjugants, as described in the text. The bands migrating at 1.7 and 2.1 kb are indicative of Tn7-1 insertion at *attTn7* position 0.



significant as approximately 10% of all Tn7-1 insertions in pEN300 occur at such sites in wild type strains. (M. Rogers, Ph.D Thesis, Glasgow)
Thus, if the lack of IHF is altering choice of the site of insertion, it is doing so at a level below the sensitivity of detection of this assay.

6.8.2 IHF and PhoS Transcription

The data presented above show IHF binding to a site on the opposite side of the Tn7 insertion point from the TnsD binding region. The site, IHF_3 in figure 6.1, lies a short distance upstream of the promoter regions of the *PhoS* gene. This gene, also known as *PstS*, encodes a 35 kD periplasmic phosphate binding protein (Magota, K. et al., 1984; Surin, B. et al. 1984)). The PhoS protein plays an important role in the Pst system for the uptake of inorganic phosphate from the growth medium. Induction of the Pst system occurs on depriving *E. coli* of phosphate. The Pst system also exerts negative control over the synthesis of alkaline phosphatase. Thus, a phenotype conferred by *PhoS* mutants is constitutive expression of alkaline phosphatase.
Both *in vivo* and *in vitro* studies have mapped the transcription initiation site of the *PhoS* mRNA and promoter elements have been characterised (Makino, K et al., 1988). The transcription initiation site lies approximately 100 nts downstream of the IHF_3 site. It should be stressed that none of the promoter elements lie in the *attTn7* containing insert in pMR80.

As discussed in the introduction, IHF is implicated in the expression of a number of genes in *E. coli* (Freundlich, Ramani et al. 1992).
Expression of the *ilvGMEDA*, *ilvBN*, *aceB1*, *pfl* and *psp* operons have shown to be reduced in strains lacking IHF. However expression of several other operons e.g. *ompC*, *ompF* and *pyrB* is increased in IHF mutants. Thus, IHF can both activate or repress transcription. A recent review has also demonstrated the presence of IHF binding sites in the proximity of the promoters of a number of genes . However, a role for IHF in the expression of these genes has not yet been demonstrated.
The location of the IHF binding site relative to the start of transcription of these various genes varies from -179 to +131. There is no apparent correlation between location of binding site and either stimulatory or inhibitory effect of IHF.

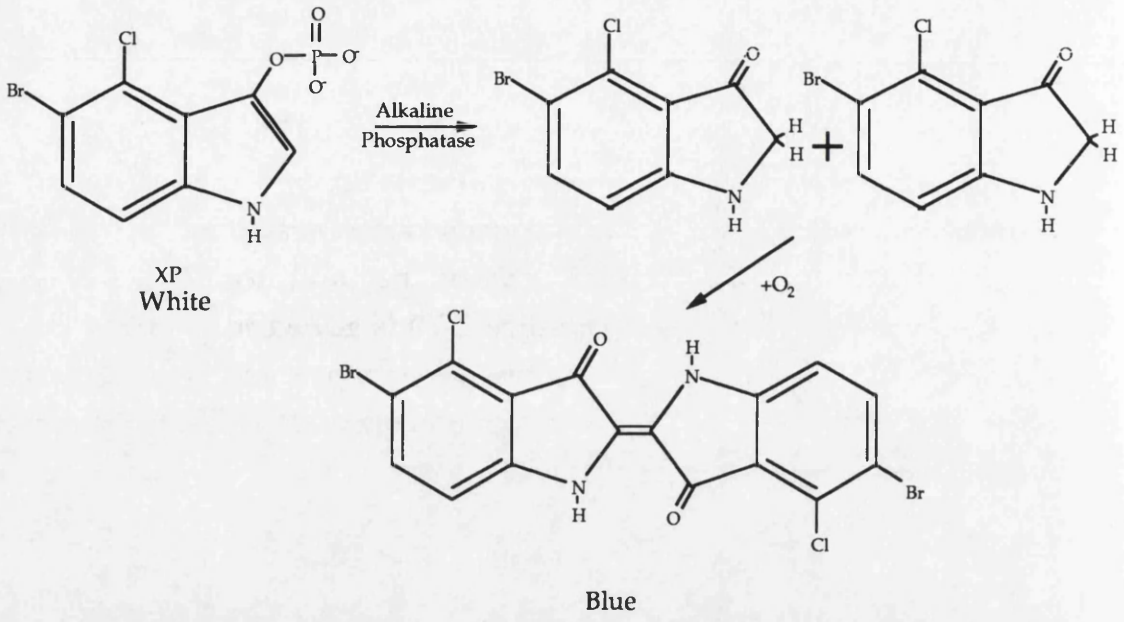


Fig. 6.18
Chemical basis of blue/white Alkaline Phosphatase Test

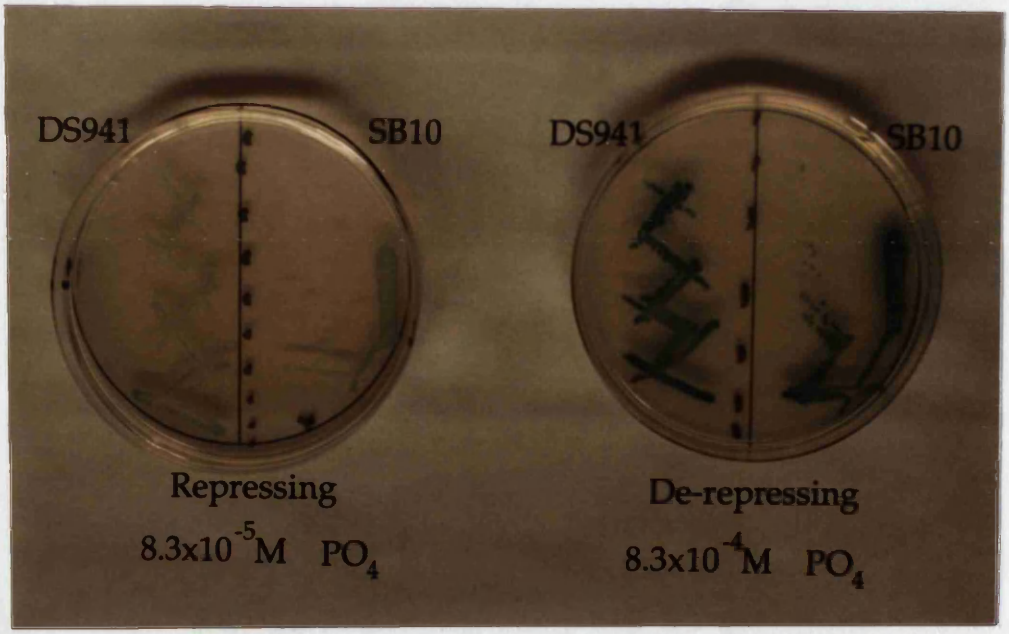


Fig. 6.19 Photograph of DS941 and SB10 grown on Medium 121 plates in repressing or de-repressing conditions in the presence of XP.

By analogy with the above examples it is possible that IHF binding to the IHF3 site adjacent to *PhoS* may influence expression of the *PhoS* gene.
As stated above, a loss of *PhoS* expression leads to constitutive synthesis of alkaline phosphatase; the inverse has also been demonstrated, increased *PhoS* expression leads to decreased alkaline phosphatase activity, even under de-repressing conditions (Amemura, 1982).
A blue/white plate based assay analogous to the B-galactosidase/X-Gal test exists for alkaline phosphatase activity.The indicator dye, 5-bromo-4-chloro-3-indolylphosphate-*p*-toluidine (XP) is added to minimal medium 121 plates containing either de-repressing or repressing concentrations of phosphate.(Berg P.E., 1981) Colonies producing alkaline phosphatase stain blue, those not producing stain white(Fig. 6.18).

DS941 and SB10 were plated on Medium 121 containing 40 μg/ml XP and either $8.3x10^{-4}M$ K_2HPO_4 (excess phosphate) or $8.3x10T^{-5}M$ K_2HPO_4 (limiting phosphate). After overnight growth at 37°C the colour of the resultant colonies was observed.
As can be seen from Fig. 6.19 SB10 and DS941 produced equally intense blue colonies on the low phosphate plates and white colonies on the high phosphate plates. Thus as measured by this assay there is no detectable difference in levels of alkaline phosphatase in the presence or absence of IHF. However, the sensitivity of this assay is such that it will only detect gross alterations in the level of alkaline phosphatase in the cells. An important caveat to the interpretation of the results of this indirect assay is that IHF may influence transcription of the alkaline phosphatase gene itself, thus masking any effects of *PhoS* expression levels.

It remains possible that IHF may modulate the expression of *PhoS* to a limited extent, below the level of detection of this indirect assay. A more accurate determination of alkaline phosphatase activity, or preferably, direct quantitation of the *PhoS* gene product or use of a promoter probe construct to detect promoter activity in the IHF^+ and IHF^- strains will be necessary to detect any more subtle effects of IHF on transcription of the *PhoS* gene.

6.9 Discussion

The data presented in this chapter suggest a complex arrangement of protein binding sites within the *attTn7* locus. The retardation assays using the 280nt pMR80 fragment show, in addition to the TnsD specific complexes, two closely migrating host-factor complexes. On following criteria both these complexes appear to be caused by IHF binding to two sites in this molecule.

1) The activity is absent in *himA*$^-$ cell
2) The complexes are formed even when the protein extract has been submitted to 10 minutes at 100^{o}C
3) Binding is strongly enhanced when a partially purified IHF extract is assayed.
4) One strong match to the IHF binding consensus site is found in the pMR80 insert, at the position designated IHF3 in Fig. 6.4. Two further poorer matches to the consensus are also detected.
5) Binding of the host factor induces a sharp bend in the DNA.

Whilst individually none of the above observations prove that the host activity is IHF, the combination of observations strongly support this theory.

With the exception of IHF3 the matches to the IHF consensus are not particularly strong. Because of the large region of protection caused by IHF binding to other substrates it appears that sequences beyond the WATCAANNNNTTR consensus are important for IHF binding. Indeed, some potential IHF sites with strong matches to the core consensus bind IHF very poorly. Therefore, one of the complexes is probably due to binding to IHF3, this is supported by the localisation of binding to the HindIII-MseI and HindIII-MboII restriction fragments of pMR80, and by the apparent centre of the bend induced by IHF binding. As noted above the second IHF dependent band is not always present, this may reflect a weak interaction with one of the poorer IHF sites. It has been demonstrated that IHF will bind tightly but non-sequence specifically to DNA containing an intrinsic bend. Thus, in addition to the requirement for the consensus sequence, local structure probably strongly influences the binding of the protein (Bonnefoy and Rouviere-Yaniv 1991).

It has been observed that binding of HU to DNA can influence the affinity of binding of IHF to the same molecule (Bonnefoy and Rouviere-Yaniv 1992). Thus it is possible that sites with weak affinites for IHF *in vitro* may become better binding sites *in vivo* . Although HU may be present in the crude extracts used in binding reactions, the presence of excess Salmon Sperm DNA will prevent its binding to the pMR80 DNA.

Further evidence for a role for IHF in Tn7 transposition comes from the observation that the apparent transposition of Tn7-1 is reduced in a *himA*$^-$ strain. Obviously this slight effect could be caused at any stage in the transposition process. However, it is important to note that it has been demonstrated, in this laboratory, that IHF does not bind to the transposon's termini (Morrell 1990). In light of the above binding data, it is tempting to suggest that IHF may favourably influence the conformation of the att-site, either for transposon excision from the chromosomal *attTn7::Tn7-1* donor site or incision of the Tn7-1 element in the recipient pEN300, or both processes. Work by Mark Rogers, presented in his PhD Thesis, has shown that the transposition rate of Tn7 is apparently not influenced by the flanking donor host sequences. Thus if IHF binding to the att-site is influencing transposition it is more likely to be at the level of insertion.

The work of Waddell and Craig agrees in part with the observations described above. In addition to the TnsD specific complex they also detected a complex caused by a heat stable host factor. Only one host complex was observed, presumably reflecting their use of shorter DNA substrates in the binding reactions. They claim to have localised important sequence determinants for this binding to *attTn7* +10 to +27. As I discussed earlier, I feel that their data was incorrectly interpreted, since +10 to +27 does not compete for binding of a protein to -25 to +64. This suggests that different binding activities are being detected on these over-lapping fragments. The cause of the weak binding to the +9 to +55 oligonucleotide observed in my work may be the same factor that binds to the +10 to +27 in the published work . This binding is possibly due to recognition of some feature of the synthetic DNA. It is clearly not due to the absence of methylation at the *dcm* site at +48.

Although the data presented in this chapter identifies an IHF site in the *attTn7* locus, it has not allowed identification of the host factor responsible for the binding detected by Waddell and Craig. It is clear that in the absence of IHF no host binding is observed using the full length pMR80 insert. It is possible that the binding observed by Waddell and Craig corresponds to the weak, and sometimes irreproducible, second band produced by IHF binding to the pMR80 insert.

Chapter 7

Concluding Remarks

This project addressed various aspects of the transposon Tn7. The initial work, presented in Chapter 3 completes the known DNA sequence of the element. A entire DNA sequence is compiled and presented in this chapter. An intriguing finding is that a considerable proportion of the element appears to be non-coding; in addition two potential genes have little or no function *in vivo*, a lambda integrase in the left end of the Tn7 is interupted by a stop codon, furthermore the *sat* gene is only expressed at very low levels , conferring resistance to only very low levels of streptothricin. The low level of expression of this gene is proposed to be due to insertion of the upstream *dhfr* gene cassette in the ribosome binding site of the *sat* gene.

It is suprising that an element such as Tn7 , which could be viewed as parasitic DNA should have so much non-coding DNA. It is possible that, by analogy with the IAS of phage Mu, some of these regions have a structural role in Tn7 transpostion process. The potential role of a series of repeated motifs in the centre of the transposon was also addressed in this chapter. It is clear from previous studies of Tn7 transposition that the only *cis* acting sequences required for Tn7 transposition are those at the termini of the element. However with the relative insensitivity of the *in vivo* transposition assay it remains a possibility that internal elements may play a role in transposition. Future workers could address this question by using short restriction fragments of the inner regions of Tn7 in gel retardation assays with extracts prepared from cells overexpressing the Tns proteins in the hope of detecting specific binding to these regions.

Perhaps the nature of the Tn7 transposition pathways has permitted the accumulation of large regions of apparently non-functional DNA with little detriment to the transposon. Studies indicate that Tn7 is generally found at a single site in the chromosome of several species of Gram negative bacteria. In *E coli* it has been demonstrated that apart from the acquisition of drug resistance this insertion is phenotypically silent, i.e. insertion does not disrupt any cellular gene expression. By spending most of its time in the chromosome and only occasionally transposing to broad host range plasmids perhaps there is little selective pressure for loss of these non-coding regions of Tn7.

Chapter 4 presented data on the construction of plasmids to coordinately over express TsnABCD, one construct was observed to have bizarre effects

on the detection of transposition products *in vivo*. A model was proposed for this effect and attempts were made to detect double strand breaks, none were observed but this may reflect a limitation of the detection method. For future workers it would be useful to repeat this analysis using the pSB84 construct , which supported high levels of *in vivo* transposition.
This chapter also described attempts to generate a Tn7 *in vitro* system, again, this was unsuccessful, potential transposition products were detected, but never reproducibly. Chapter 5 presented data that suggests that in pSB84 extracts TnsD is prevented from recognising the *attTn7* site by an interaction with one or more of TnsABC. The presence of ATP and low levels of Mg ions enabled detection TnsD binding. These data may explain why the attempts to develop an *in vitro* system were largely unsuccessful. Any future work to develop a Tn7 *in vitro* system must take these observations into account. Presumably in the Craig group *in vitro* system condition this effect is overcome by the relative ratios of the Tns proteins. In light of the data presented in Chapter 4 on the interference of Mg^{++} with TnsD binding, it may be possible to alter the conditions of the *in vitro* reaction to have low Mg^{++} and ATP present throughout the entire time course of the reaction.
Finally, work presented in Chapter 6 identified an IHF binding site in the *attTn7* locus. Only a prelimnary characterisation of the site is presented, further work should examine the protection from cleavage by DNaseI conferred by binding to this site and use other DNA footprinting techniques. A potential role for IHF in the transcription of the *phoS (pstS)* gene is also suggested, to examine this question transcriptional fusions of a reporter gene to this locus should be constructed and levels of transcription assayed in isogenic IHF^+ and IHF^- cells.

In conclusion, it is clear that there are many aspects of the Tn7 transposition reaction that are poorly understood, with the development of an *in vitro* reaction by the Craig group, and hopefully here in Glasgow, more understanding will be gained on the complex interactions between the Tns proteins and DNA sites involved in Tn7 transposition.

Bibliography

Adzuma, K. Mizuuchi, K. (1988) Target immunity of Mu-transposition reflects a differential distribution of Mu-B-protein. Cell **53**: 257-266

Adzuma, K. Mizuuchi, K. (1989) Interaction of proteins located at a distance along DNA - mechanism of target immunity in the Mu DNA strand-transfer reaction. Cell **57**: 41-47

Amabile-Cuevas, C. F. Chicurel, M. E. (1992) Bacterial Plasmids and Gene flux. Cell **70**: 189-199

Amemura, M., Shinagawa, H., Makino, K., Otsuji, N., Nakata, A. (1982) Cloning of and Complementation tests with Alkaline Phosphatase Regulatory genes (*phoS* and *phoT*) of *E. coli* Journal of Bacteriology **152**: 692-701

Arciszewska, L. K., Mckown, R. L. and Craig, N. L. (1991). Purification of TnsB, a transposition protein that binds to the ends of Tn7. Journal Of Biological Chemistry **266**: 21736-21744.

Arciszewska, L. K. Drake, D. Craig, N. L. (1989) Transposon Tn7 cis-acting sequences in transposition and transposition immunity. Journal Of Molecular Biology **207**: 36-52

Arciszewska, L. K. Craig, N. L. (1991) Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon Nucleic Acids Research **19**: 5021-5029

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

Bainton, R., Gamas, P. and Craig, N. L. (1991). Tn7 transposition *invitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. Cell **65**: 805-816.

Bender, J. and Kleckner, N. (1986). Genetic-evidence that Tn10 transposes by a non-replicative mechanism. Cell **45**: 801-815.

Bender, J. and Kleckner, N. (1992) Tn10 insertion specificity is strongly dependent upon sequences immediately adjacent to the target-site consensus sequence. Proceedings Of The National Academy Of Sciences Of The United States Of America. **89**: 7996-8000

Benjamin, H. W. and Kleckner, N. (1989) Intramolecular transposition by Tn10. Cell **53**: 373-383

Benjamin, H. W. and Kleckner, N. (1992). Excision of Tn10 from the donor site during transposition occurs by flush double-strand cleavages at the transposon termini. Proceedings Of The National Academy Of Sciences Of The United States Of America **89**: 4648-4652.

Berg, P.E. (1981) Cloning and Characterisation of the *E. coli* Gene coding for Alkaline Phosphatase. Journal of Bacteriology **146**: 660-667

Birnboim, M.M. and Doly, J. (1979) A rapid alkaline extraction method for screening recombinant plasmid DNA. Nucleic Acids Research **9**: 1513-1523

Bissonnette, L. and Roy, P. H. (1992). Characterisation of InO of *P. aeruginosa* plasmid pVS1, an anscestor of integrons of multireistance plasmids and transposons of gram negative bacteria. Journal of Bacteriology **174**: 1248-1257.

Bonnefoy, E. and Rouviere-Yaniv, J. (1991). HU and IHF, 2 homologous histone-like proteins of *Escherichia-coli*, form different protein DNA complexes with short DNA fragments. EMBO Journal **10**: 687-696.

Bonnefoy, E. and Rouviere-Yaniv, J. (1992). HU, the major histone-like protein of escherichia-coli, modulates the binding of IHF to oriC. EMBO Journal **11**: 4489-4496.

Cao, L., Alani, E. and Kleckner, N. (1990). A pathway for generation and processing of double-strand breaks during meiotic recombination in saccharomyces-cerevisiae. Cell **61**: 1089-1101.

Caparon, M. G. and Scott, J. R. (1989). Excision and insertion of the Conjugative Transposon Tn916 involves a novel recombination mechanism. Cell **59**: 1027-1034.

Collis, C. M. and Hall, R. M. (1992). Gene cassettes from the insert region of integrons are excised as covalently closed circles. Molecular Microbiology **6**: 2875-2885.

Collis, C. M. and Hall, R. M. (1992). Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. Journal Of Bacteriology **174**: 1574-1585.

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

Craig, N. L. (1991) Tn7 - a target site-specific transposon Molecular Microbiology **5**: 2569-2573

Craigie, R., Mizuuchi, M., Mizuuchi, K. (1984) Site specific recognition of Mu ends by the MuA protein Cell **39**: 387-394

Craigie, R., Mizuuchi, K. (1987) Transposition of Mu-DNA - joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. Cell **51**: 493-501

Craigie, R. Fujiwara, T. Bushman, F. (1990) The IN protein of Moloney Murine Leukaemia Virus processes the Viral DNA ends and accomplishes their integration *in vitro* Cell **62**: 829-837

Craigie, R. (1992) Hotspots and warm spots - integration specificity of retroelements. Trends in Genetics **8**: 187-190

Datta, N., Hedges, R.W. (1972) Trimethoprim resistance conferred by W plasmids in Enterobacteriacae Journal of General Microbiology **72**:349-355

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

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence-analysis programs for the Vax. Nucleic Acids Research **12**: 387-395.

Ekaterinaki, N. (1987). Expression and function of Tn7 transposition proteins. PhD Thesis, Glasgow University :

Engelman, A. Mizuuchi, K. Craigie, R. 1991 HIV-1 DNA integration - mechanism of viral-DNA cleavage and DNA strand transfer. Cell **67**: 1211-1221

Engelman, A. Craigie, R. (1992) Identification of conserved amino-acid-residues critical for human-immunodeficiency-virus type-1 integrase function-*in vitro* Journal Of Virology **66**: 6361-6369

Escoubas, J. M., Prere, M. F., Fayet, O., Salvignol, I., Galas, D., Zerbib, D. and Chandler, M. (1991). Translational control of transposition activity of the bacterial insertion-sequence IS1. EMBO Journal **10**: 705-712.

Fayet, O., Ramond, P., Polard, P., Prère, M. F. and Chandler, M. (1990). Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? Molecular Microbiology **4**: 1771-1777.

Finkel, S. E. and Johnson, R. C. (1992). The Fis protein: it's not just for DNA inversion anymore. Molecular Microbiology **6**: 3257-3265.

Fling, M. E., Kopf, J. and Richards, C. (1985). Nucleotide sequence of the transposon Tn7 gene encoding an amino-glycoside modifying enzyme. Nucleic Acids Research **13**: 7095-7106.

Fling, M. E. and Richards, C. (1983). The nucleotide-sequence of the trimethoprim-resistant dihydrofolate-reductase gene harbored by Tn7. Nucleic Acids Research **11**: 5147-5158.

Flores, C., Qadri, M. I. and Lichtenstein, C. (1990). DNA-sequence analysis of 5 genes - TnsA, TnsB, TnsC, TnsD and TnsE, required for Tn7 transposition. Nucleic Acids Research **18**: 901-911.

Flores, C. C., Cotterill, S. and Lichtenstein, C. P. (1992). Overproduction of 4 functionally active protein-TnsA, protein-B, protein-C, and protein-D, required for Tn7 transposition to its attachment site, *atttn7*. Plasmid **28**: 80-85.

Freundlich, M., Ramani, N., Mathew, E., Sirko, A. and Tsui, P. (1992). The role of integration host factor in gene-expression in Escherichia-coli. Molecular Microbiology **6**: 2557-2563.

Friedman, D. I. (1988). Integration host factor - a protein for all reasons. Cell **55**: 545-554.

Galas, D. and Chandler, M. (1989). Bacterial Insertion Sequences. Mobile DNA published by ASM : 109.

Gamas, P., Chandler, M. G., Prentki, P. and Galas, D. J. (1987). Escherichia-coli integration host factor binds specifically to the ends of the insertion-sequence IS1 and to its major insertion hot-spot in pBR322. Journal Of Molecular Biology **195**: 261-272.

Gamas, P. and Craig, N. L. (1992). Purification and characterization of TnsC, a Tn7 transposition protein that binds ATP and DNA. Nucleic Acids Research **20**: 2525-2532.

Gay, N. J. Tybulewicz, V.L.J. Walker, J. E. (1986) Insertion of transposon Tn7 into the Escherichia-coli-Glms transcriptional terminator. Biochemical Journal **234**: 111-117

Goodman, S. D. and Nash, H. A. (1989). Functional replacement of a protein-induced bend in a DNA recombination site. Nature **341**: 251-254

Goodrich, J. A., Schwartz, M. L. and Mcclure, W. R. (1990). Searching for and predicting the activity of sites for DNA-binding proteins - compilation and analysis of the binding-sites for *Escherichia-coli* integration host factor (IHF). Nucleic Acids Research **18**: 4993-5000.

Gosti-Testu, F., Norris, V. and Brevet, J. (1983). Restriction Map of Tn7. Plasmid **10**: 96-99.

Gringauz, E. Orle, K. A. Waddell, C. S. Craig, N. L. (1988) Recognition of Escherichia-coli atttn7 by transposon Tn7 - lack of specific sequence requirements at the point of Tn7. insertion Journal Of Bacteriology **170**: 2832-2840

Hall, R. M., Brookes, D. E. and Stokes, H. W. (1991). Site-specific insertion of genes into integrons - role of the 59-base element and determination of the recombination cross-over point. Molecular Microbiology **5**: 1941-1959.

Haniford, D. B., Chelouche, A. R. and Kleckner, N. (1989). A specific class of IS10 transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. Cell **59**: 385-394.

Haniford, D. B. Benjamin, H. W. Kleckner, N. (1991) Kinetic and structural-analysis of a cleaved donor intermediate and a strand transfer intermediate in Tn10 transposition. Cell **64**: 171-179

Haniford, D.B. Chaconas, G. (1992) Mechanistic aspects of DNA transposition Mechanistic aspects of DNA transposition **2**: 698-704

Higgins, C.F., Dowman, C.J., Stirling, D.L., Waddel, L., Booth, I.R., May, G., Brewer, E. (1988) a physiological role for DNA supercoiling in the osmoregulation of gene expression of *S. typhimurium* and *E. coli*. Cell **52**: 569-584

Kano, Y. and Imamoto, F. (1990). Requirement of integration host factor (IHF) for growth of *E. coli* deficient in HU protein. Gene **89**: 133-137.

Kim, J., Zweib, C., Wu, C. and Adhya, S. (1989). Bending of DNA by gene regulatory proteins: construction and use of a DNA bending vector. Gene **85**: 15-23.

Kleckner, N. (1989). Transposon Tn10. Mobile DNA, published by ASM : 227-269.

Kubo, K. M. and Craig, N. L. (1990). Bacterial transposon-Tn7 utilizes 2 different classes of target sites. Journal Of Bacteriology **172**: 2774-2778.

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

Lavoie, B. D. Chaconas, G. (1990) Immunoelectron microscopic analysis of the A, B, and HU protein-content of bacteriophage-Mu transpososomes. Journal Of Biological Chemistry. **265**: 1623-1627

Lee, S. Y. Butler, D. Kleckner, N. (1987) Efficient Tn10 transposition into a DNA insertion hot spot *invivo* requires the 5-methyl groups of symmetrically disposed thymines within the hot-spot consensus sequence. Proceedings Of The National Academy Of Sciences Of The United States Of America **84**: 7876-7880

Lee, E. C.Macwilliams, M. P. Gumport, R. I. Gardner, J. F. Genetic-analysis of *Escherichia-coli* integration host factor interactions with its bacteriophage-lambda H' recognition site. Journal Of Bacteriology **173**: 609-617

Lewis, M. K. and Thompson, D. V. (1990). Efficient site directed *in vitro* mutagenesis using ampicillin selection. Nucleic Acids Research **18**: 3439-3443.

Lichtenstein, C. and Brenner, S. (1981). Site specific proerties of Tn7 transposition in the *E. coli* chromosome. Molecular and General Genetics **183**: 380-387.

Lichtenstein, C. and Brenner, S. (1982). Unique insertion site of Tn7 in the *E. coli* chromosome. Nature **297**: 601-603.

Magota, K.Otsuji, N.Miki, T.Horiuchi, T.Tsunasawa, S.Kondo, J.Sakiyama, F.Amemura, M., Morita, T., Shinagawa, H., Nakata, A. (1984) Nucleotide-sequence of the phoS-gene, the structural gene for the phosphate-binding protein of *Escherichia-coli* Journal of Bacteriology **157**; 909-917

Makino, K., Shinagawa, H., Amemura, M., Kimura, S., Nakata, A., Ishihama, A. (1988) Regulation of the phosphate regulon of *Escherichia-coli* - activation of pstS transcription by phoB protein *invitro* Journal of Molecular Biology **203**: 85-95

Martinez, E. and De la Cruz, F. (1988). Transposon Tn21 encodes a recA-independent site-specific integration system. Molecular & General Genetics **211**: 320-325.

Martinez, E. and De la Cruz, F. (1990). Genetic elements involved in Tn21 site-specific integration, a novel mechanism for the dissemination of antibiotic-resistance genes. EMBO Journal **9**: 1275-1281.

McKown, R. L. Waddell, C. S. Arciszewska, L. K. Craig, N. L. (1987) Identification of a transposon Tn7-dependent DNA-binding activity that recognizes the ends of Tn7 Proceedings Of The National Academy Of Sciences Of The United States Of America **84**: 7807-7811

McKown, R. L., Orle, K. A., Chen, T. and Craig, N. L. (1988). Sequence requirements of Escherichia-coli atttn7, a specific site of transposon Tn7 insertion. Journal Of Bacteriology **170**: 352-358.

Mercier, J., Lachapelle, J., Couture, F., Lafond, M., Vezina, G., Boissinot, M. and Levesque, R. C. (1990). Structural and functional-characterization of tnpI, a recombinase locus in Tn21 and related beta-lactamase transposons. Journal Of Bacteriology **172**: 3745-3757.

Mizuuchi, K. (1992). Transpositional recombination - mechanistic insights from studies of Mu and other elements. Annual Review Of Biochemistry **61**: 1011-1051.

Mizuuchi, K. Adzuma, K. (1991) Inversion of the phosphate chirality at the target site of Mu-DNA strand transfer - evidence for a one-step transesterification mechanism Cell **66**: 129-140

Mizuuchi, M. and Mizuuchi, K. (1989). Efficient Mu-transposition requires interaction of transposase with a DNA-sequence at the Mu-operator - implications for regulation. Cell **58**: 399-408.

Mizuuchi, M. Baker, T. A. Mizuuchi, K. (1991) DNAse protection analysis of the stable synaptic complexes involved in Mu transposition Proceedings Of The National Academy Of Sciences Of The United States Of America **88**: 9031-9035

Morisato, D. and Kleckner, N. (1984). Transposase promotes double strand breaks and single strand joints at Tn10 termini *in vivo*. Cell **39**: 181-190.

Morisato, D. Kleckner, N. (1987) Tn10 transposition and circle formation invitro. Cell **51**: 101-111
Morrell, E. (1990). Transposition of Tn7. PhD Thesis (Glasgow) :

Nucken, E. J., Henschke, R. B. and Schmidt, F. R. J. (1991). Site-specific integration of genes into hot spots for recombination flanking *aadA* in Tn21 transposons. Molecular and General Genetics **229**: 137-146.

Orle, K. A. and Craig, N. L. (1991). Identification of transposition proteins encoded by the bacterial transposon Tn7. Gene **104**: 125-131.

Owen-Hughes, T. A., Pavitt, G. D., Santos, D. S., Sidebotham, J. M., Hulton, C. S. J., Hinton, J. C. D. and Higgins, C. F. (1992). The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene-expression. Cell **71**: 255-265.

Pato, M. L. (1989). Bacteriophage Mu. Mobile DNA : 23-53.

Prentki, P., Chandler, M. and Galas, D. J. (1987). Escherichia-coli integration host factor bends the DNA at the ends of IS1 and in an insertion hotspot with multiple IHF binding-sites. EMBO Journal **6**: 2479-2487.

Pryciak, P. M. Varmus, H. E. (1992) Nucleosomes, DNA-binding proteins, and DNA-sequence modulate retroviral integration target site selection. Cell **69**: 769-780

Qadri, M. I., Flores, C. C., Davis, A. J. and Lichtenstein, C. P. (1989). Genetic-analysis of *atttn7*, the transposon Tn7 attachment site in escherichia-coli, using a novel M13-based transduction assay. Journal Of Molecular Biology **207**: 85-98.

Roberts, D. and Kleckner, N. (1988). Tn10 transposition promotes recA-dependent induction of a lambda-prophage. Proceedings Of The National Academy Of Sciences Of The United States Of America **85**: 6037-6041.

Roberts, D. E., Hoopes, B. C., McClure, W. R. and Kleckner, N. (1985). IS10 transposition is regulated by DNA adenine methylation. Cell **43**: 117-130.
Rogers, M. Ekaterinaki, N. Nimmo, E. Sherratt, D. (1986) Analysis of Tn7 transposition. Molecular & General Genetics. **205**: 550-556

Rogers, M.S. (1986) Transpositional properties and organisation of Tn7. PhD Thesis, Glasgow University

Rowland, S. J. and Dyke, K. G. H. (1990). Tn552, a novel transposable element from *Staphyloccocus aureus*. Molecular Microbiology **4**: 961-975.

Sambrook, J. Fritsch, E.F. Maniatis, T. Molecular cloning: a laboratory manual (second edition) Cold Spring Harbour Laboratory Press, New York

Sandmeyer, S. Yeast retrotransposons. Current Opinion in Genetics and Development. **2**: 705-711

Scott, J. R. (1992). Sex and the single circle. Journal of Bacteriology **174**: 6005-6010.

Sekine, Y. and Ohtsubo, E. (1989). Frameshifting is required for production of the transposase encoded by insertion sequence 1. Proceedings Of The National Academy Of Sciences Of The United States Of America **86**: 4609-4613.

Shapiro, J.A. (1979). Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proceedings Of The National Academy Of Sciences Of The United States Of America **76**: 1933-1937

Sherratt, D. J. (1989). Transposon Tn3 and related elements: Site specific recombination and transposition. Mobile DNA published by ASM : 163-185.

Smith, G. M. Jones, P. (1986) Tn7 transposition - a multigene process - identification of a regulatory gene-product Nucleic Acids Research **14**: 7915-7927

Smith, G.R. (1991) Conugational recombination in *E. coli*: Myths and Mechanisms Cell **64**: 19-27

Snyder, U. K., Thompson, J. F. and Landy, A. (1989). Phasing of protein induced DNA bends in a recombination complex. Nature **341**: 255-257.

Stark, W.M., Boocock, M.R., Sherratt, D.J., (1992) Catalysis by site specific recombinases. Trends in Genetics **8**: 432-438

Stenzel, T. T., Patel, P. and Bastia, D. (1987). The integration host factor of *Escherichia-coli* binds to bent DNA at the origin of replication of the plasmid-pSC101. Cell **49**: 709-717.

Stokes, H. W. and Hall, R. M. (1989). A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions - integrons. Molecular Microbiology **3**: 1669-1683.

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

Sundstrom, L., Roy, P. H. and Skold, O. (1991). Site-specific insertion of 3 structural gene cassettes in transposon Tn7. Journal Of Bacteriology **173**: 3025-3028.

Sundstrom, L. and Skold, O. (1990). The dhfrI trimethoprim resistance gene of Tn7 can be found at specific sites in other genetic surroundings. Antimicrobial Agents And Chemotherapy **34**: 642-650.

Surette, M. G. and Chaconas, G. (1992). The Mu-transpositional enhancer can function in trans - requirement of the enhancer for synapsis but not strand cleavage. Cell **68**: 1101-1108.

Surin, B. P., Jans, D. A., Fimmel, A. L., Shaw, D. C., Cox, G. B., Rosenberg, H., (1984) Structural gene for the phosphate-repressible phosphate-binding protein of *Escherichia-coli* has its own promoter - complete nucleotide-sequence of the phoS gene, Journal of Bacteriology,**157**: 772-778

Tabor, S. and Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled expression of specific genes. Proceedings Of The National Academy Of Sciences Of The United States Of America **82**: 1074-1078.

Tang, Y., Lichtenstein, C. and Cotterill, S. (1991). Purification and characterization of the TnsB protein of Tn7 - a transposition protein that binds to the ends of Tn7. Nucleic Acids Research **19**: 3395-3402.

Tietze, E., Brevet, J. and Tschape, H. (1987). Relationships among the streptothricin resistance transposons Tn1825 and Tn1826 and the trimethoprim resistance transposon Tn7. Plasmid **18**: 246-249.

Waddell, C. S. and Craig, N. L. (1989). Tn7 transposition - recognition of the atttn7 target sequence. Proceedings Of The National Academy Of Sciences Of The United States Of America **86**: 3958-3962.

Waddell, C. S. Craig, N. L. (1988) Tn7 transposition - 2 transposition pathways directed by 5 Tn7-encoded genes. Genes & Development **2**:137-149.

Wiater, L. A. and Grindley, N. D. F. (1988). Gamma-delta transposase and integration host factor bind cooperatively at both ends of gamma-delta. EMBO Journal **7**: 1907-1911.

Wiater, L. A. and Grindley, N. D. F. (1990). Integration host factor increases the transpositional immunity conferred by gamma-delta ends. Journal of Bacteriology **172**: 4951-4958.

Wu, C. and Crothers, D. M. (1984). The locus of sequence directed and protein induced DNA bending. Nature **308**: 509-513.

Yang, C. C. and Nash, H. A. (1989). The interaction of *Escherichia-coli* IHF protein with its specific binding-sites. Cell **57**: 869-880.

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

Zerbib, D., Jakowec, M., Prentki, P., Galas, D. J. and Chandler, M. (1987). Expression of proteins essential for IS1 transposition - specific binding of InsA to the ends of IS1. EMBO Journal **6**: 3163-3169.