Studies on the *tnsA* and *tnsB* genes of transposon Tn*7*

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

by

Angela Barbara Gawthrop

Institute of Genetics
University of Glasgow
Church Street
Glasgow
UK

March, 1993

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

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

SUMMARY

Tn*7* is a 14kb transposable element found in a variety of enteric bacteria. It encodes five genes, *tnsA*, *tnsB*, *tnsC*, *tnsD* and *tnsE* which are required for transposition of the element. Two of these genes, *tnsA* and *tnsB*, are the subject of this study.

The *tnsA* gene is poorly expressed from the Tn*7* *tns* promoter and ribosome binding site; the possible reasons for this were investigated, through both examination of the DNA sequence and experimentation. Different expression systems for *tnsA* were utilised, and use of the pT7-7 translational fusion expression vector was found to yield high concentrations of the TnsA protein. This protein was, however highly insoluble. Various strategies for solubilising the protein were tested, and conditions under which partially purified TnsA could be maintained in a soluble form were found.

Regulation of transposition of a number of prokaryotic elements is brought about by the binding to the terminal repeat sequences of an inactive truncated form of the transposase protein. TnsB was previously known to bind to repeated 22bp motifs at the Tn*7* ends; peptides believed to be proteolytic fragments of TnsB are produced *in vivo* particularly under certain growth conditions, and it was hypothesised that these might be involved in regulation of Tn*7* transposition. A series of plasmids encoding TnsB peptides truncated at the amino or carboxyl end was therefore constructed. These plasmids were used in three different lines of investigation: *in vitro* binding of protein extracts derived from cells expressing these truncated *tnsB* genes to DNA fragments from the left and right ends of Tn*7*; galactokinase assays to measure the degree of repression of the *tns* promoter (which overlaps the terminal repeat at the Tn*7* right end) in the presence of the TnsB peptides; and *in vivo* transposition frequencies of Tn*7* in the presence of a high concentration of TnsB peptide, relative to that of the wildtype TnsB. Results suggest that, although binding is observed which is specific to TnsB peptides including a region near the amino end, these peptides do not significantly repress transcription of the *tns* genes or reduce the observed *in vivo* transposition frequencies. During the course of this work, a weak promoter was identified in the *tnsA* coding sequence and reading in the opposite

direction to the *tns* genes; it is thought unlikely that this plays a significant rôle in regulation of transposition.

A sequence bearing similarity to that of a helix-turn-helix DNA binding motif had been previously identified towards the amino end of TnsB. All the TnsB peptides shown here to exhibit binding activity to the right or left end motifs of Tn*7* contained this sequence. Site-directed mutagenesis of the putative recognition helix of TnsB did not, however, lead to loss of binding activity, and so the existence of the helix-turn-helix remains in doubt.

Agarose and polyacrylamide binding gels were used to search for synapsis of the Tn*7* ends mediated by TnsB; such a synapsis is likely to be a prerequisite for transposition. Results from polyacrylamide gel electrophoresis were consistent with the formation of DNA-protein-DNA complexes involving linear DNA fragments. TnsB-specific retardation was also observed on agarose gels with circular molecules containing a mini-Tn*7*, but there was no evidence that this was due to synapsis of the Tn*7* ends and not simply to the added mass of the binding protein molecules. The clarity of the stained gels would be improved by the use of pure TnsB protein. Strategies for improving on the previous partial purification procedure were therefore investigated but no significant improvement was made; insolubility of TnsB in low salt concentrations and a predilection for proteolysis (the latter of which might be solved by carrying out the whole procedure at low temperature) were the main problems encountered.

ACKNOWLEDGEMENTS

I wish to thank past and present members of the Institute of Genetics for help with matters academic, technical and ancillary, secretarial and social. My supervisor was David Sherratt. This work was carried out with financial assistance from the Crowther Fund of the Open University.

ABBREVIATIONS

Abbreviation	Full name
3-12	N-dodecyl-N,N' dimethyl 3-ammonio-1-propansulphonate
Ap	ampicillin
ATP	adenosine triphosphate
bp	base pairs
cAMP	cyclic adenosine monophosphate
CDC	cleaved donor complex
cDNA	complementary DNA
Cm	chloramphenicol
cpm	counts (of radioactive decay) per minute
CTAB	mixed alkyltrimethylammonium bromide
DEAE	diethylaminoethyl
DMSO	dimethyl sulphoxide
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid disodium salt
FIS	factor for inversion stimulation
GCG	Genetics Computer Group of the University of Wisconsin
GST	glutathione *S*-transferase
Gu	guanidinium
HEPES	*N*-[2-hydroxyethyl]piperazine-*N'*-[2-ethanesulphonic acid]
HIV	human immunodeficiency virus
IAS	internal activating sequence
IHF	integration host factor
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase pairs
Km	kanamycin monosulphate
LE	left end of Tn*7*
Nal	nalidixic acid sodium salt
OD	optical density
ONPG	*o*-nitrophenyl-β -D-galactoside
ORF	open reading frame
ori	origin of replication
PAGE	polyacrylamide gel electrophoresis
PMSF	phenylmethylsulphonyl fluoride
RBS	ribosome binding site
RE	right end of Tn*7*
Rif	rifampicin
RNase	ribonuclease
SDS	sodium dodecylsulphate
Sm	streptomycin sesquisulphate
s/n	supernatant
Sp	spectinomycin dihydrochloride
SSC	stable synaptic complex
STC	strand transfer complex
Su	sulphanilamide
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
Tc	tetracycline
TE	Tris EDTA
TEMED	*NNN'N'*-tetramethyl-1,2-diaminoethane
Tp	trimethoprim
Tris	tris(hydroxymethyl)aminoethane
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CONTENTS

CHAPTER 1

INTRODUCTION TO Tn*7* AND RELATED MOBILE DNA ELEMENTS

1.1 A short survey of mobile genetic elements

Mobile genetic elements, now known to be ubiquitous, were first recognised in maize over forty years ago (McClintock, 1948, 1951) by the altered phenotype their insertion brought about. Genome rearrangements as a result of transposition include translocations, duplications, deletions and inversions, and thus phenotypic variation is a common consequence of transposition. This thesis is concerned with one prokaryotic transposable element, Tn*7*; the prokaryotic transposable elements are generally characterised by a DNA genome flanked by inverted repeats typically of about 6-30bp and encoding one or more genes for a transposition function. Transposition is through a DNA intermediate, and may be replicative, as in the case of Tn*3*, or conservative (non-replicative), as with Tn*10*. Even where non-replicative transposition occurs, the transposon copy number may increase if transposition occurs to a site in front of a replication fork from one behind it, or if the gap left by an excised element is repaired by gene conversion from a homologous chromosome.

Many eukaryotic transposable elements, such as the P element of *Drosophila*, Tc*1* of *Caenorhabditis elegans* and the *Ac-Ds* elements of maize originally studied by McClintock, have similar characteristics to those of prokaryotic elements, but several other kinds of transposable element are also found in eukaryotes. Most of this latter category of mobile elements are believed to transpose through an RNA intermediate, and encode a putative reverse transcriptase gene. LINEs (Long Interspersed Nuclear Elements) and the *Drosophila* Jockey element are examples of a class of mobile element which typically has no terminal inverted repeat sequences or long direct terminal repeat sequences, but is demarcated at one end by a poly-A.T tail; members of this class generally contain two open reading frames, bearing resemblance to the *gag* and *pol* genes of retroviruses. Retrotransposons, such as the yeast Ty elements, are yet more complex, containing not only *gag*-like and *pol* genes, but also long terminal repeats and the PBS (tRNA primer binding site) and PRS (purine rich priming

sequence) typical of retroviruses; the packaging of retrotransposon RNA may be an essential step in transposition of these elements.

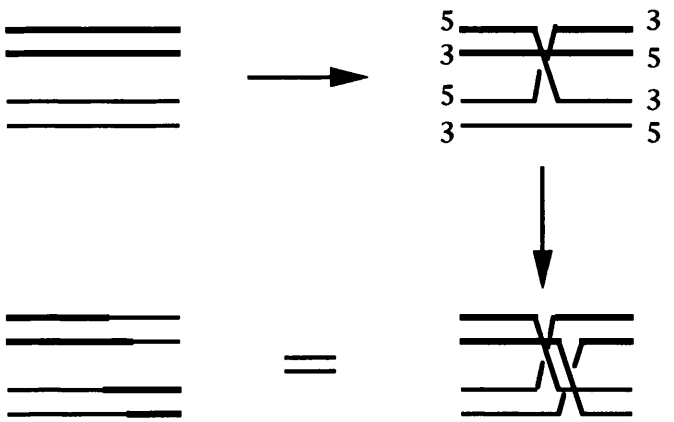
Other DNA elements encoding their own mobility include mobile introns (Cech *et al.*, 1990) and integrons (Stokes and Hall, 1989), the latter of which are further discussed in Section 1.4.4.

1.2 Transposition and site-specific recombination in prokaryotes

Transposition and site-specific recombination share some properties, but also have several differences. In site-specific recombination, two sites are involved, four covalent phosphodiester bonds are broken and four new bonds are made, catalysed by the recombinase enzyme, without the synthesis of any new DNA; in transposition, three sites are involved: two transposon ends and the target insertion site (although one-ended transposition is occasionally observed with, for example Tn*3*), either four or six phosphodiester bonds are broken, and probably the formation of only two bonds is catalysed by the transposase, with host enzymes being responsible for repair of other breaks at the transposon/target DNA junctions. There is a requirement in site-specific recombination for sequence homology between the cross-over regions of the two sites involved in such a reaction, whereas no sequence homology is required between the ends of a transposable element and the DNA at the site of insertion of the element.

The transient covalent linkage of a serine or tyrosine residue of the catalytic protein to the cleaved DNA has been seen only in site-specific recombination, and is believed not to occur in transposition; the experiment described below supports this view. The phosphate groups of the DNA backbone participating in the bond cleavages of site-specific recombination and transposition are not themselves chiral, but can be made chiral by the substitution of a sulphur atom for one of the non-bonding oxygen atoms. Using these mutagenised substrates, it has been demonstrated that the strand transfer reaction in bacteriophage Mu transposition (Mizuuchi and Adzuma, 1991) and HIV integration (Engelman *et al.*, 1991) occurs with the inversion of chirality at the reacting phosphate group which would be expected with a one step S_N1 reaction in the absence of a covalently linked protein-DNA intermediate; while Lambda site-specific recombination takes place with the retention of chirality

(a) integrases



(b) resolvases and invertases

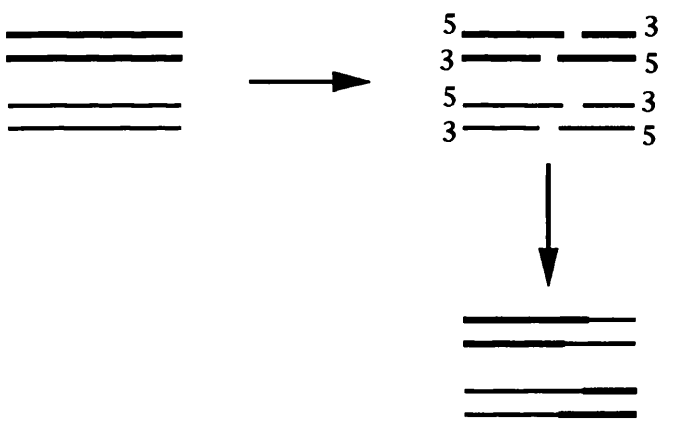


Figure 1.1. The two basic mechanisms of site-specific recombination in prokaryotes. Systems in the integrase family recombine one strand at a time through a branch-migrating Holliday junction intermediate; resolvases and invertases exchange through a double strand break. The recombinase proteins, which are covalently bound to the cleaved DNA ends in intermediate reaction steps, are not shown here. It should be noted that structure and cellular function do not necessarily coincide: cointegrate resolution of Tn*4430* (Mahillon and Lereclus, 1988), for example, is catalysed by a protein of the integrase family.

typical of a covalently bound protein-DNA intermediate (Mizuuchi and Adzuma, 1991).

Site-specific recombination reactions, as depicted in Figure 1.1, may be subdivided into those which take place through initial single stranded nicks at the reacting sites and a Holliday junction intermediate, as is seen for example with bacteriophage Lambda insertion into and excision from the *E. coli* chromosome, or through a double stranded DNA break, as is believed to occur in the cointegrate resolution reaction following the transposition of Tn*3* and in the mechanistically similar DNA inversion reactions. In both cases, staggered cuts at the reacting ends match exactly, and there is thus no requirement for a gap repair mechanism.

Transposition may proceed via single stranded nicks at the 3' transposon termini, but sometimes the bonds at the transposon 5' ends are also broken at the same time, leading to an excised transposon; there is no requirement in transposition for target DNA cleavage to occur concurrently with these former reactions, as has been demonstrated with Tn*7* (Bainton *et al.*, 1991). The target DNA may have some structural or functional specificity, this depending on the particular system. There is normally a staggered cut in the target DNA (Benjamin and Kleckner, 1989; Bainton *et al.*, 1991) with the result that the 5' overhang on the target DNA does not match the overhang (if any) on the ends of the transposable element; the transposition reaction must therefore generally be followed by repair of the gaps at the junctions, leaving a repeat of several base-pairs of target site DNA each side of the inserted transposon.

1.3 Mechanisms of transposition and its regulation in prokaryotes

Prokaryotic transposable elements have traditionally been classified according to their structure (for example Craig and Kleckner, 1987); see Figure 1.2. Class I elements included insertion sequences, which encode only transposase protein(s), flanked by terminal inverted repeat sequences, and composite transposons (which are formed of two copies of an insertion sequence between which is one or more genes for selectable markers), though Tn*4727* (Dumans *et al.*, 1989) contains virtually the whole of Tn*7* between two copies of IS*1*; Class II embraced elements like Tn*3* and Tn*21*, which encode a cointegrate resolution function and a *res* site at which this resolution takes place, in addition to the transposase; and Class

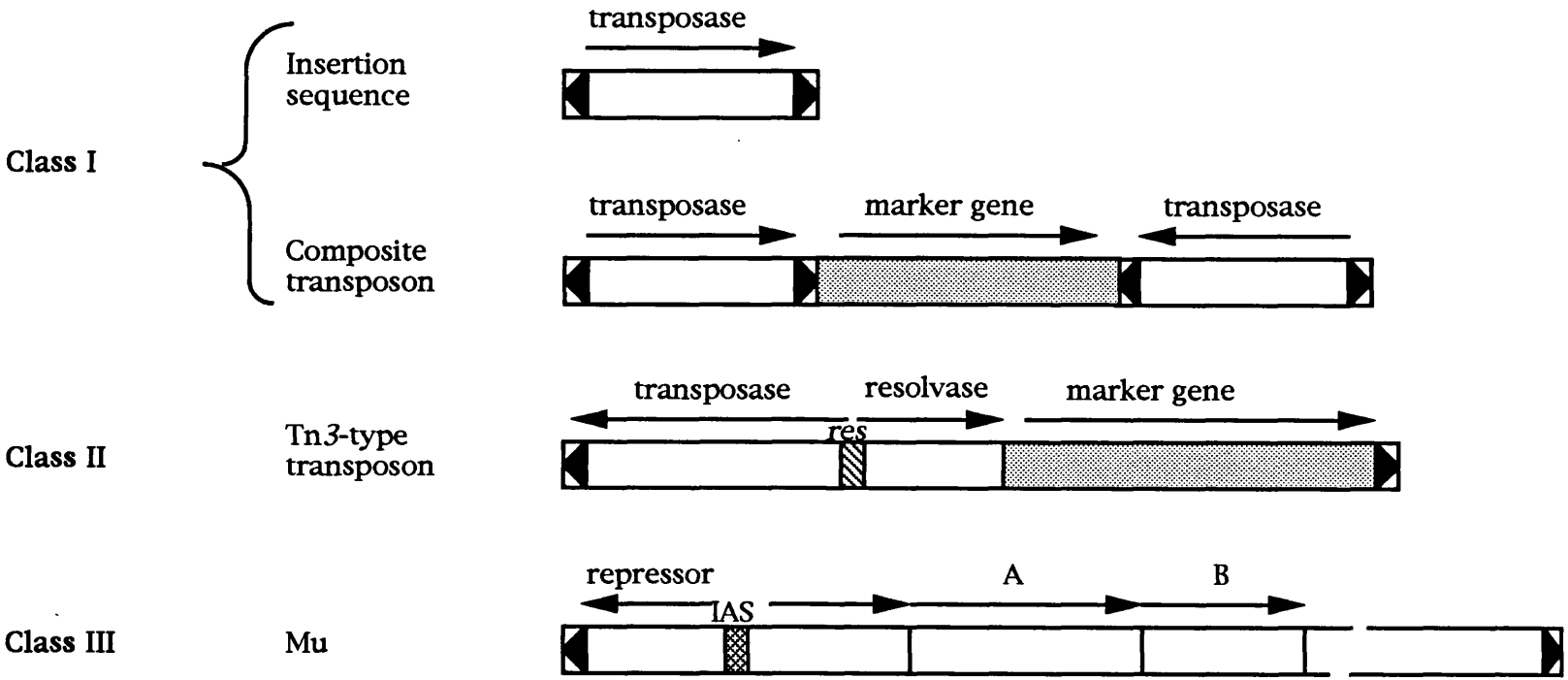


Figure 1.2 The traditional classification of prokaryotic transposable elements.
(IAS=internal activating sequence;
◥◤ denotes terminal inverted repeat)

III covered Bacteriophage Mu and the closely related D108 of *E. coli* and several transposing bacteriophages from *Pseudomonas* strains. Tn*7* does not fall into this classification at all.

It is perhaps more meaningful to classify elements according to their mechanism of transposition. The Class I member Tn*10*, for instance is known to transpose by a non-replicative "cut and paste" mechanism *in vivo* (Bender and Kleckner, 1986) and *in vitro* (Haniford *et al.*, 1991), while IS*6*, also in Class I, is believed to transpose replicatively (Galas and Chandler, 1989); Tn*3* transposes replicatively to form a cointegrate of donor and target replicons where transposition is intermolecular, and a deletion where intramolecular transposition occurs; and Mu adopts both replicative and non-replicative mechanisms, both likely to be via a Shapiro intermediate (Figure 1.3), rather than via an excised transposon, as explained below. Clearly the original classification of prokaryotic transposable elements does not distinguish between these various mechanisms. Confusion about whether transposition of a particular element is conservative or replicative can occur because homologous recombination may also take place *in vivo* between two copies of an element on the same or different replicons; and also, the gap left by the excision of a transposon may be filled by another copy of the element through a process of gene conversion (Engels *et al.*, 1990).

Transposition in Gram-negative bacteria tends to take place by one of a small number of mechanisms, described further below. Other prokaryotes, however, employ different strategies, some of which are outlined in Section 1.3.6.

The common mechanisms are illustrated in Figure 1.3. The basic features depicted are: (i) recognition of and binding to the ends of the element by a transposon-encoded protein, transposase-catalysed nucleophilic attack by a molecule, probably usually of water, on the 3' transposon ends and cleavage at the transposon ends (either double strand cleavage as in Tn*10* (Haniford *et al.*, 1991) and Tn*7* (Bainton *et al.*, 1991) or single strand nicking as is believed to occur with Tn*3* and Mu); (ii) nucleophilic attack by the resultant exposed 3'-OH groups of the transposon ends on the target DNA, resulting in concerted double strand cleavage of the target DNA and transfer of 3' transposon ends to 5' ends of the target DNA with the formation of new phosphodiester bonds, catalysed by one or more transposon-encoded proteins; and (iii) repair of the consequent

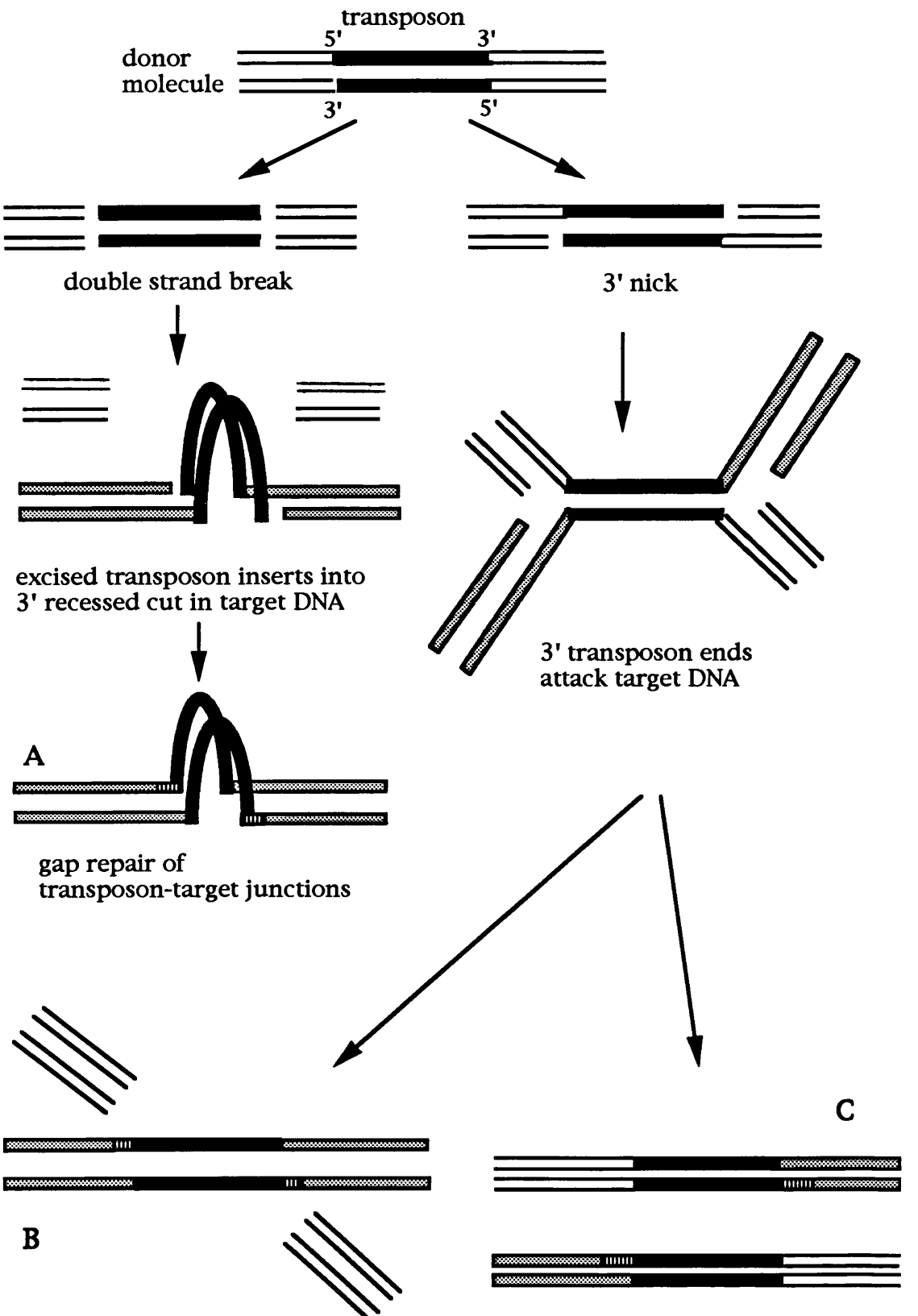

transposon
5'
3'
donor
molecule
3'
5'
double strand break
3' nick
excised transposon inserts into
3' recessed cut in target DNA
3' transposon ends
attack target DNA
A
gap repair of
transposon-target junctions
C
B
cleavage of 5' transposon ends;
gap repair of transposon-target
junctions
replication of transposon DNA and
gap repair, resulting in cointegrate
formation

Figure 1.3 The more common transposition mechanisms of prokaryotic elements.

The two different outcomes are conservative (or non-replicative) transposition of the element, as depicted in **A** and **B**, and replicative transposition, shown in **C**. The two strands of the transposon DNA are shown black, the donor molecule white, and the target DNA (which can be the transposon or the donor molecule in the case of intramolecular transposition) stippled.

The initial cleavage of the transposon ends can be on both strands of the same end simultaneously, as shown in the left hand pathway and exemplified by Tn*10* and Tn7, or by single stranded nicks at the 3' end of the element, depicted in the right hand pathway.

The left hand pathway shows the "cut and paste" mechanism. An excised transposon is formed by double-stranded breaks; the 3' OH groups of the element subject the target DNA to nucleophilic attack at sites typically staggered on the two strands of the target molecule, as shown. Since the cleavage pattern of the transposon ends does not normally coincide with that of the target site, insertion of the element leaves gaps at the 5' ends which are repaired, presumably by the host cell's machinery, as indicated by vertical bars in **A**.

The donor molecule may be repaired through a gene conversion step, or may remain linear and be degraded.

Single strand nicking of the transposon ends and staggered cutting of the target DNA leads to a "Shapiro intermediate" (Shapiro, 1979) in which the 3' ends of the element are ligated to the target DNA while the 5' ends remain linked to the donor DNA. From here, two distinct reactions are possible: the 5' ends of the transposon DNA can be cleaved from the donor DNA and ligated to the free ends of the target DNA (with gap repair as described above), leaving the product **B**, which is a simple insertion identical to **A**, as is likely to happen in the initial insertion of bacteriophage Mu. Alternatively, DNA replication can be primed fom the free target DNA ends and continued through the transposon DNA; this results in semi-conservative replication of the transposon DNA, with one end of each copy of the transposon attached to donor DNA and the other end to target DNA; in intermolecular transposition, a cointegrate of donor and target molecules is formed, with two copies of the element in direct repeat (**C**). This cointegrate may be resolved back into two separate molecules, each now containing one copy of the element, either through site-specific recombination, as seen with Class II transposable elements, or by homologous recombination.

breaks in the target DNA, this latter probably catalysed by host-encoded proteins. Repair of the donor DNA is known to occur *in vivo* for replicative transposition, since intact supercoiled cointegrate molecules are seen, and sometimes at least also for non-replicative transposition as in, for example, Tn*916* (Caparon and Scott, 1989) and also the eukaryotic *Drosophila* P elements (Engels *et al.*, 1990), although there is evidence against this donor repair in the case of Tn*10* (Bender *et al.*, 1991).

The transposition properties of some individual elements, and their regulatory mechanisms, are discussed below. Where *in vitro* transposition systems have been set up, the experimental techniques used are described in Section 1.5.

1.3.1 IS*1* and the IS*3* family

Although the transposase encoded by IS*1* differs substantially in sequence from that encoded by elements of the IS*3* family, these elements are discussed together here because they have a related overall organisation of open reading frames: there are typically two open reading frames, transcribed in the same direction. The smaller upstream reading frame (*orfA*) is, however, in the +1 phase with respect to the larger downstream open reading frame (*orfB*). It has been demonstrated with some of these elements (Sekine and Ohtsubo, 1989; Vogele *et al.*, 1991; Polard *et al.*, 1991) that programmed frameshifting occurs with the result that one large protein, OrfAB, is produced; the small protein OrfA is also produced when frameshifting fails to occur and, in the case of IS*911* (a member of the IS*3* family) at least, the OrfB protein is also translated (Polard *et al.*, 1991), although it does not bind the terminal inverted repeat DNA and it has no known function as a separate protein (Polard *et al.*, 1992). In the case of IS*1*, the InsAB protein is the transposase (Lüthi *et al.*, 1990), and InsA acts as a negative regulatory element (Machida and Machida, 1989; Zerbib *et al.*, 1990): it has the capacity to bind to the terminal inverted repeats of the transposable element, but fails to catalyse transposition. IS*911* by contrast appears to require both the OrfAB and OrfA proteins for high frequency intermolecular transposition to take place, and tends to produce excised circularised transposons, containing also one of the original target DNA direct repeats, in the absence of OrfA (Polard *et al.*, 1992). It is thought that these circles are not an obligatory intermediate in IS*911* transposition (Polard *et al.*, 1992).

The translated *orfB* open reading frame of members of the IS*3* family shows similarity with the region of retroviral integrase proteins which has been shown to be essential for strand cleavage and transfer (Fayet *et al.*, 1990; Engelman and Craigie, 1992). This is discussed further in Section 1.3.7.

It is believed that IS*1* transposes both replicatively and non-replicatively (Biel and Berg, 1984), but no studies have been published on the transposition mechanism or on how the decision between the two pathways is made. IS*911*, by contrast, appears to transpose by an exclusively conservative "cut and paste" pathway (Polard *et al.*, 1992).

1.3.2 IS*50* and Tn*5*

Tn*5* is a composite transposon, consisting of two copies of the insertion sequence IS*50* (denoted IS*50*L and IS*50*R) in opposite orientations, flanking kanamycin, bleomycin and streptomycin resistance genes. Both Tn*5* and the component IS*50* elements are believed to transpose conservatively (Berg, 1989) through a "cut and paste" mechanism.

The IS*50*L sequence contains a stop codon in the transposase gene, and so only IS*50*R transposase is normally functional. The transposase gene acts preferentially in *cis*; a *trans*-acting inhibitor protein, Inh, is encoded by the same reading frame as the transposase, but with a second, apparently stronger, promoter and with translation initiating at a site internal to the transposase coding sequence (Krebs and Reznikoff, 1986). Regulation of transposition is thus brought about by excess of inhibitor over transposase combined with its greater activity in *trans*; it is not yet known whether Inh acts by binding to the DNA at the transposon ends, or by complexing with the transposase. Transposition frequency is also influenced by the *dam* methylation state of the transposon DNA (McCommas and Syvanen, 1988; Yin and Reznikoff, 1988; Dodson and Berg, 1989); there are terminal *dam* methylation sites, one of them overlapping the transposase promoter, and both transcription of the transposase (but not the inhibitor) and transposase binding are increased when the element is in the hemi-methylated state following the passage of a replication fork.

1.3.3 IS*10* and Tn*10*

Tn*10* is also a composite transposon, consisting of two copies of IS*10*, which encode a *cis*-acting transposase, in inverted repeat, flanking genes required for tetracycline resistance. It is interesting to note that changes

at the three terminal base pairs of the transposon inhibit strand transfer, but do not block the donor DNA cleavage (Huisman *et al.*, 1989); this suggests that the two reactions are not obligatorily coupled and that the transposon terminal DNA is recognised by the transposase in a different manner for the two reactions. An *in vitro* transposition system has been developed for Tn*10*, though the products seen here are generally intramolecular (Morisato and Kleckner, 1987), rather than the intermolecular ones more commonly seen *in vivo*. Tn*10* transposes non-replicatively *in vitro* through a double stranded "cut and paste" mechanism (Benjamin and Kleckner, 1992), with blunt-ended cuts, in contrast to those observed with Tn*7* and retroviruses; and it is believed that the same occurs *in vivo* (Bender and Kleckner, 1986). The cellular SOS mechanism is induced by a Tn*10* transposition event *in vivo* (Roberts and Kleckner, 1988), probably as a result of chromosomal double strand breaks occurring on excision of the element. There is limited target site specificity; GCTNAGC is preferred though not absolutely required (Halling and Kleckner, 1982). As well as the transposase, host-encoded IHF and HU proteins are involved in transposition (Morisato and Kleckner, 1987); these latter are probably involved in conformational changes in DNA which facilitate the synapsis and strand cleavage.

IS*10* and Tn*10* transposition is regulated through several different mechanisms. The transposase gene has a poor ribosome binding site (Kleckner, 1990b), and the binding of a highly stable antisense RNA to this ribosome binding site further inhibits translational initiation (Ma and Simons, 1990). *dam* methylation also plays a rôle in transpositional regulation: there are two potential methylation sites in the DNA of IS*10*, one overlapping the transposase promoter and the other within the opposite terminal transposase binding site (Kleckner, 1990b); the DNA becomes transiently hemi-methylated when a replication fork passes through, and one of these hemi-methylated daughter elements transposes at high frequency as a result of both increased transcription of transposase and more stable binding of it to the transposon termini (Kleckner, 1990b). By this means, Tn*10* and IS*10* transposition, like that of Tn*5* and IS*50*, is programmed to occur at a particular time in the cell cycle, or immediately after the transposon has entered the cell on a conjugative plasmid.

1.3.4 Class II elements

The Class II elements are represented on Figure 1.2 by Tn*3*. The characteristics of this family of elements are the presence within the transposon of a gene encoding a resolvase protein which breaks down the cointegrate molecules resulting from intermolecular replicative transposition, and an internal site, denoted *res*, at which this recombinase binds and acts through a site-specific recombination reaction, as described in Section 1.2. There are two basic types of gene organisation within this family of elements: that exemplified by Tn*3* and shown on Figure 1.2, and the alternative, as seen in Tn*21* and several others, where the resolvase gene and upstream *res* site are in the opposite orientation with respect to the transposase gene. The transposition reaction is believed to take place with a Shapiro intermediate step (Shapiro, 1979), in which the 3' end of each DNA strand is linked to recipient DNA, while the 5' end of each strand remains attached to the donor DNA, as shown in Figure 1.3. While much research has been devoted to the resolution systems of Class II elements, the transposition mechanism adopted by this family has not been well studied and few details are known.

In Tn*3* the *res* site overlaps the promoters for both the transposase and resolvase genes, and thus transposition can be negatively regulated at the transcriptional level by the binding of resolvase to the *res* site (Chou *et al.*, 1979). In transposons of the Tn*21* type, the resolvase and transposase are co-transcribed, again from a promoter within the *res* site, and negative control of transposition by the resolvase is possible.

Transposition immunity, the inhibition of transposition into a replicon already harbouring a copy of the transposon, is observed with Tn*3* and other members of this family (Robinson *et al.*, 1977).

Tn*21*, in addition to the characteristic structure described above, appears to contain an integron: a number of antibiotic resistance genes delineated by a characteristic palindromic sequence and an integrase gene likely to be responsible for insertion of these genes. A similar structure is found in Tn*7*, and integrons are discussed further in Section 1.4.4.

1.3.5 Mu

The temperate bacteriophage Mu (Figure 1.2) is probably the most studied of all transposable elements; its transposition system has been well characterised *in vitro* as well as *in vivo*. It was first identified (Taylor, 1963) through its mutagenic effect on the *E. coli* chromosome as, during

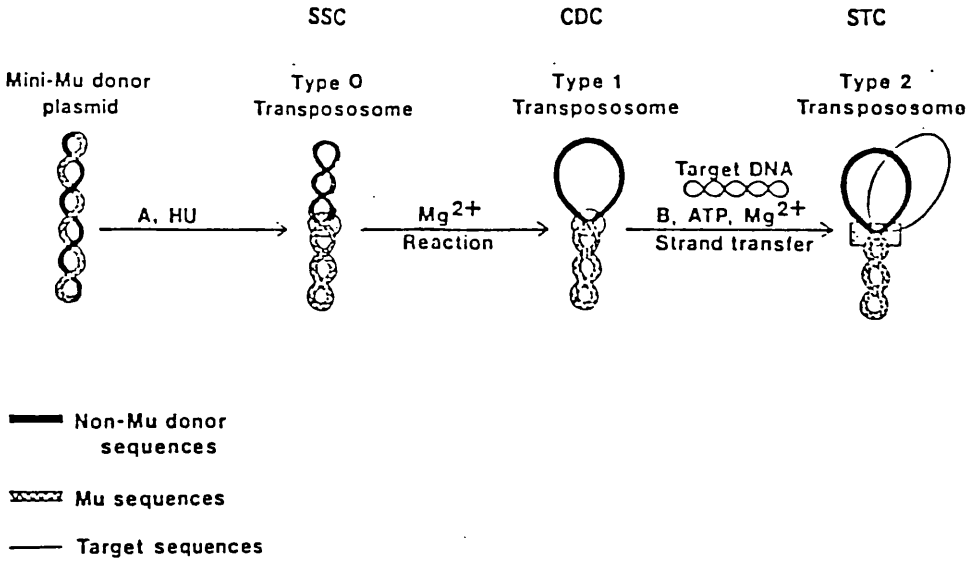


Figure 1.4 Steps in the replicative intermolecular *in vitro* transposition reaction of bacteriophage Mu, leading to a cointegrate product.

SSC = stable synaptic complex

CDC = cleaved donor complex

STC = strand transfer complex

The *in vivo* reaction may take place through the same mechanism, although this is not certain; wildtype Mu generally transposes intramolecularly within the chromosome, leading to genomic rearrangements rather than formation of a cointegrate between two molecules.

The Figure is adapted from Surette *et al.* (1987).

lysogeny, it integrated at a large number of sites (unlike bacteriophage Lambda). The complete element is 37kb long, but the only sequences required in *cis* for transposition are the correctly oriented elaborate terminal repeats and (unless in *trans* in a vast molar excess; Surette and Chaconas, 1992) the internal activating sequence (IAS); the Mu A and Mu B proteins are the only Mu-encoded proteins required for replicative transposition, and these can be supplied in *trans*. Like IS*1*, Mu transposes both conservatively and replicatively; the initial integration of DNA from the infecting bacteriophage is mechanistically a conservative transposition reaction, and subsequent transposition events in the lytic phase of the phage's lifestyle are replicative. The Mu genome is packaged in its protein coat as a double stranded DNA molecule with the ends of the element flanked by host DNA and held together by the phage-encoded N protein. The integration event involves cleavage of the Mu-host DNA junctions and transfer of both strands of the Mu ends to the target DNA of the infected bacterium. The Mu A protein is required for this reaction, and integrative transposition occurs at only very low levels in the absence of Mu B protein in addition to Mu A (Chaconas *et al.*, 1985).

Replication of the bacteriophage DNA to approximately 100 copies prior to lysis also occurs by transposition; it is known that this replication is semi-conservative, and it is believed to take place through a Shapiro intermediate (Shapiro, 1979), as with Tn*3*; there is, however, no evidence for a Mu-encoded resolvase function. Such a resolution function is probably unnecessary to the survival of Mu since the replicative transposition is generally intramolecular and immediately precedes packaging and cell lysis. The *in vitro* transposition system is well established; see Section 1.5.1 for technical details. Several intermediates of the *in vitro* replicative transposition reaction have been identified (Surette *et al.*, 1987; Mizuuchi *et al.*, 1991; Mizuuchi *et al.*, 1992), and it appears that the reaction proceeds as in Figure 1.4. Three distinct transposition intermediates have been studied: the Stable Synaptic Complex (SSC; Type 0 transpososome), the Cleaved Donor Complex (CDC; Type I transpososome), where the 3' ends of the Mu DNA have been nicked, and the Strand Transfer Complex (STC; Type II transpososome), where the hydroxyl groups of the nicked ends have attacked phosphodiester bonds of the target DNA. In each of these complexes, MuA protein is thought to be bound as a tetramer to the terminal motifs (Kuo *et al.*, 1991; Lavoie *et al.*, 1991; Mizuuchi *et al.*, 1992). The IAS, which can be bound by Mu A, is also required for the initial

synapsis of the Mu ends in the SSC, although its presence is not necessary for subsequent steps (Mizuuchi *et al.*, 1992; Surette and Chaconas, 1992). The SSC and CDC involve DNA only of the supercoiled donor replicon, the supercoiling facilitating the correct orientation of transposon ends in the synapse; target DNA is incorporated in the STC, where Mu B protein and ATP are also necessary in addition to the Mu A, Mg^{2+} ions and HU host protein required for the SSC and CDC, and the IHF required for correct conformation of the IAS in the SSC.

Mu B protein in the presence of ATP oligomerises and binds nonspecifically to the target DNA molecule and thus, by its interaction with Mu A, makes the site bound by Mu B a preferred site for Mu insertion (Adzuma and Mizuuchi, 1988); Mu B may however sometimes activate Mu A without first binding to DNA (Baker *et al.*, 1991). It appears that there is no strict sequence of events leading to the completion of Mu transposition but, rather, a network of steps (Mizuuchi, 1992). The three named complexes are of increasing stability as the reaction proceeds (Surette *et al.*, 1987), thus driving the reaction forwards.

Like Tn*3*, Mu exhibits immunity to repeated transposition into the same plasmid replicon, although clearly *in vivo* transposition is generally from one chromosomal site to another; in the case of Mu, this immunity is known to occur through interactions between Mu A and Mu B proteins. Mu B oligomers bound to DNA can be activated to dissociate and release the DNA when interacting with adjacent Mu A, which stimulates the ATPase activity of Mu B (Adzuma and Mizuuchi, 1991). In this way, the probability of transposition into a site close to the donor transposon is reduced.

The Mu repressor protein, encoded by the *c* gene at the left end of the element, regulates both the initial lysogeny/lysis decision and subsequent replicative Mu transposition. The IAS overlaps repressor binding sites in the promoters for both the repressor and the transposition genes, with the result that transcription of the *A* and *B* genes is blocked by bound repressor. The repressor also competes with Mu A for occupation of the IAS; the stable synaptic complex, a prerequisite for transposition, cannot be formed unless the IAS is held in the correct conformation by bound Mu A.

1.3.6 Tn*554*, Tn*916* and other unusual transposons

The transposons Tn*554* and Tn*916* are both found in Gram positive bacterial hosts. Tn*554* encodes three genes involved in transposition; of

these, *tnpA* and *tnpB* have sequence similarity to Lambda-like integrase genes. Tn*554* transposes at high frequency with site and orientation specificity (Bastos and Murphy, 1988); the third gene, *tnpC*, is not essential for transposition, but is required for orientation specificity. Unlike the majority of prokaryotic transposable elements, Tn*554* does not contain any terminal repeat sequence; neither does its insertion generate a duplication of the target site DNA sequence. There is some homology between its terminal sequences and those of its preferred insertion site (Murphy and Löfdahl, 1984) and it is likely that its transposition mechanism is nearer to that of Lambda integration than to those described in Section 1.2.

The regulation of transposition in Tn*554* is not well studied. A site *tnpI* at the left end of Tn*554*, which may overlap with the *tnp* promoter, is implicated in inhibition of transposition, by means as yet unknown (Murphy, 1983).

Tn*916* is 16.4kb transposon which encodes functions not only for transposition, but also for its conjugative transfer between bacterial cells (Senghas *et al.*, 1988; Scott *et al.*, 1988). It has the capacity to excise from the host DNA and circularise covalently; in this form the transposon ends, which contain short imperfect inverted repeats, are believed to be linked by a 5 or 6bp heteroduplex derived from the different donor DNA sequences at the insertion site (Caparon and Scott, 1989). This circle can either transpose to another site in the same host cell, or move to a different bacterium, sometimes of a different species. It has been proposed (Caparon and Scott, 1989; Poyart-Salmeron *et al.*, 1990) that transposition of Tn*916* and the related Tn*1545* is also more akin to Lambda excision and integration than to "traditional" transposition, although the homology between the core sequences necessary for Holliday junction branch migration and subsequent resolution in Lambda excision is not generally found with these transposons.

Tn*552* is also found in *Staphylococcus*. It has terminal inverted repeats, two genes encoding proteins pA and pB which have sequence similarity with retroviral integrases and MuB and which are believed to have transposition functions, a recombination site, *resL*, and an enzyme encoded by *binL*, the sequence of which appears more similar to that of an invertase than a resolvase and which acts on the *resL* site and on a similar site on the host plasmid (Rowland and Dyke, 1990; Rowland, pers. comm.). Transposition of Tn*552* has not yet been demonstrated.

Another unusual transposable element, this time isolated from *E. coli*, is IS*91*. The element has distinctive ends: each end contains a palindromic sequence, of 16 and 30bp, but there is no sequence similarity between these two (Mendiola *et al.*, 1992). IS*91* transposes without target duplication and has limited site-specificity. It has been proposed (Mendiola and de la Cruz, 1992) that transposition of this element may involve single strand nicking of one end of the element only, followed by single stranded rolling-circle replication of the transposon; the transposase shows sequence similarity with plasmid rolling-circle replication proteins.

1.3.7 Eukaryotic retroviruses

The eukaryotic retroviruses, such as HIV, are included in this survey because, although not traditionally classified as transposable elements, the mode of insertion of their reverse-transcribed DNA into the host chromosome appears to be the same mechanistically as that of many prokaryotic transposable elements (Craigie and Mizuuchi, 1987).

The virion consists of two identical RNA molecules bound to virus-encoded reverse transcriptase-RNaseH and integrase proteins, and packaged in a protein nucleocapsid and exterior lipid-protein coat. The nucleocapsid is injected into the cytoplasm of the host cell, and the retroviral genome is subsequently reverse transcribed into double stranded DNA within the nucleocapsid; the DNA then moves to the nucleus, where it is integrated into a random site in the host DNA with the assistance of the viral integrase protein to form a provirus.

The proviral DNA contains long terminal repeats, at the outside ends of which are short inverted repeats; the 3' terminal dinucleotides are typically 5'-CA-OH-3', a sequence which is found in the same position in Tn*552* (Rowland and Dyke, 1990), Tn*7*, IS*21*, IS*3*, IS*911* and several other insertion elements (Galas and Chandler, 1989). The 3' DNA sequence prior to integration continues for 2 or 3 bp beyond the CA dinucleotide, and is processed to leave the 3' recessed ends which are involved in nucleophilic attack on the target DNA (Craigie *et al.*, 1990). The staggered ends of the retroviral DNA have however proved not to be essential for integration, and the reaction also takes place with blunt $^{CA}/_{GT}$ 3' ends (Engelman *et al.*, 1991).

The retroviral integrase (IN) proteins contain a zinc-finger motif in the amino part of the protein, which is believed to be involved in specific binding to the terminal DNA, although it has been suggested (Woerner *et*

al., 1992) that the carboxyl half of the HIV integrase protein is involved in the DNA binding activity. The central cores of these proteins all contain a motif which is also found in the carboxyl region of several prokaryotic transposase proteins, including those of the IS*3* family (Fayet *et al.*, 1990), as shown in Figure 1.8. It has recently been shown by site-directed mutagenesis (Engelman and Craigie, 1992) that the three invariant residues D-64, D-116 and E-152 of HIV-1 integrase are absolutely required for 3' processing of the DNA ends prior to integration, and also for the strand transfer reactions of proviral integration, and it is likely that these three residues are all part of the active site for the strand cleavage and transfer reactions of retroviral integration. Other invariant or highly conserved residues in the central cores of the retroviral integrase sequences were identified by Engelman and Craigie, but were not tested by them, and it is not known whether these also play an essential rôle in the 3' end processing and strand cleavage reactions.

It is not yet known by what criteria the target DNA for retroviral insertion is selected, or how intermolecular, rather than intramolecular, integration is favoured.

1.4 Tn*7*

Tn*7*, the subject of this thesis, has an complex construction, depicted in Figure 1.5. The *cis* acting end repeats, shown on Figure 1.6, are more elaborate than those found in other transposons, excepting perhaps bacteriophage Mu; an unequalled number of five genes is involved in transposition; and the three resistance genes appear to be relics of an integron-like system. These features of the transposon are discussed further below.

1.4.1 A brief history of the study of Tn*7*

The transposon Tn*7* was first recognised as such by Barth *et al.* (1976) who were studying clinical isolates of *E. coli*, and observed transfer of linked trimethoprim and streptomycin resistance between conjugative plasmids and from plasmids to the *E. coli* chromosome, and named the responsible factor TnC. It was noted at the time that the transposon was approximately 14kb long, and that it inserted at one particular position on the *E. coli* chromosome, and also appeared to show orientational specificity and some site specificity in its insertion into the plasmid RP4 (Barth and

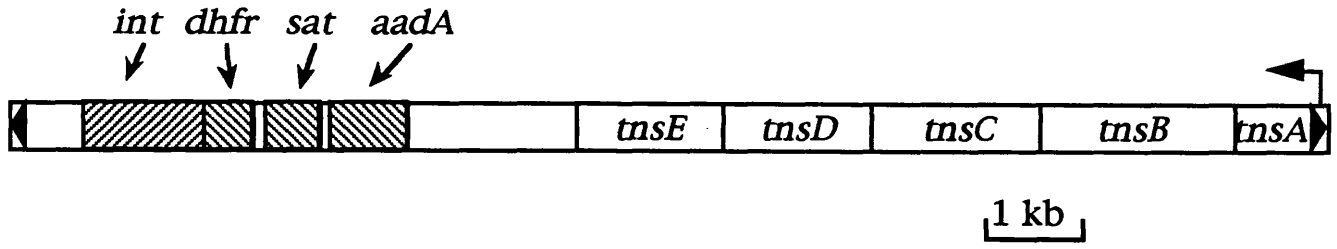


Figure 1.5 General structure of Tn*7*. The element is approximately 14kb long, and is flanked by inverted repeats, the structure of which is shown in more detail in Figure 1.6. The five transposition genes of Tn*7* are located in the right half, as drawn here and read from right to left. Only one promoter for these genes has been unambiguously identified; it overlaps the terminal motifs, and is shown arrowed. The left side of the transposon contains the three antibiotic resistance genes *dhfr*, *sat* and *aadA*, all reading from left to right and encoding resistance to trimethoprim, streptothricin and streptomycin/spectinomycin respectively, and the integrase gene *int*, which contains an internal stop codon in the isolates sequenced.

Grinter, 1977). TnC was subsequently redesignated Tn*7* (Cohen, 1976). Study of the transposition genes and transpositional insertion sites, and to a lesser extent the antibiotic resistance genes, preceded knowledge of their DNA sequence. As the sequence became known (Flores *et al.*, 1990; Fling and Richards, 1983; Fling *et al.*, 1985; H. Kay-Young, pers. comm.), comparisons with other transposable elements at the DNA level became possible, and a new gene, likely to have once been active as a lambda-type integrase, was identified. The sequence of Tn*7* has recently been completed by S. Bell (in press).

1.4.2 The *tns* genes and their transposition system

It is now known that Tn*7* encodes five genes involved in transposition (Rogers *et al.*, 1986; Flores *et al.*, 1990). The transposition functions, designated *tnsABCDE*, are located in what is conventionally called the right half of the transposon. All five genes read in the same direction from the right end inwards and they either overlap slightly or are immediately adjacent to each other. The only unambiguously identified promoter lies upstream of *tnsA*, close to the right end of the transposon (Gay *et al.*, 1986), and has been estimated (Rogers *et al.*, 1986) to have a strength approximately 1.1 times that of the *lac* promoter. A second, much weaker, promoter has been proposed upstream of *tnsB* (Rogers *et al.*, 1986), but this is disputed by Waddell and Craig (1988) who claimed that insertions in *tnsA* are polar on *tnsB*, and thus that the two genes must be co-transcribed. Waddell and Craig also believed that each of *tnsC*, *tnsD* and *tnsE* is a separate transcriptional unit; promoters for these three genes were not identified by Rogers *et al.*

It was clear from an early stage of the research on Tn*7* that there were two distinct transposition pathways: the more common was that to the preferential *E. coli* chromosomal insertion site, now denoted *attTn7*, and required *tnsA*, *tnsB*, *tnsC* and *tnsD*. Transposition at a much lower frequency to other chromosomal sites and to plasmids required *tnsA*, *tnsB*, *tnsC* and *tnsE*, but not *tnsD* (Hauer and Shapiro, 1984; Rogers *et al.*, 1986).

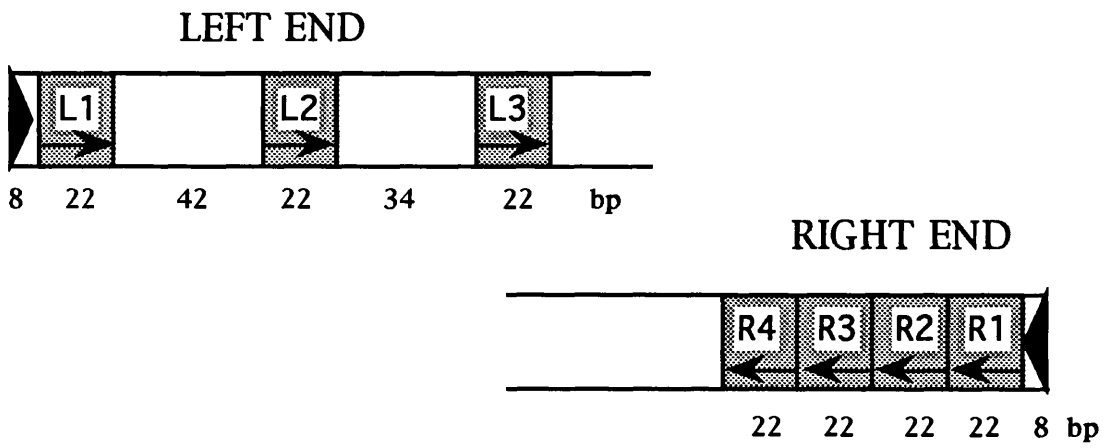
At the commencement of this work, it was known that *tnsB* encoded a protein which was associated with a specific binding activity to the Tn*7* Right End (M. Rogers, unpublished results), work which was later refined elsewhere (McKown *et al.*, 1987). There were indications that a *tnsD*-specific binding activity might be associated with *attTn7* (Ekaterinaki, 1987). Nothing was known about *tnsA*, *tnsC* or *tnsE* or their possible protein

products apart from the requirement for the genes in transposition, and the fact that these genes, like *tnsB* and *tnsD*, could act *in trans*.

During the course of this work, the complete nucleotide sequence of the *tns* genes was made available (Flores *et al*, 1990) and it became clear that each gene was associated with an open reading frame, thus making it more likely that a protein product was required for transposition. Flores *et al*. carried out limited analysis of the proposed amino acid sequences of the gene products but, apart from putative helix-turn-helix motifs (of various levels of credibility) in TnsA, TnsB, TnsD and TnsE, and homology to a nucleotide binding motif in TnsC, no strong similarity to other proteins was identified by them. Further discussion of *tnsA* and *tnsB* appears in Chapters 3 and 4 respectively. It has recently been confirmed that TnsC binds ATP and also binds DNA non-specifically (Gamas and Craig, 1992), though it is unlikely that ATP hydrolysis is energetically linked to strand transfer.

The ends of Tn*7*, in common with those of the majority of prokaryotic transposable elements, are inverted repeats. However, as can be seen from Figure 1.6, these are much more complex than is normally found: there is a terminal 30bp repeat, but the inner 22bp of this form a motif which is repeated several times at each end, always in the same orientation as the terminal copy. The four copies of the motif at the right end of Tn*7*, R1, R2, R3 and R4, overlap slightly and are aligned so that there are two turns of the (undistorted) DNA double helix for each motif; at the left end, the three motifs, L1, L2 and L3, are more widely spaced, and there does not appear to be an integral number of turns of the DNA helix separating these motifs. Substantially purified TnsB protein has been shown to bind to the motifs of both the right and left end (Arciszewska and Craig, 1991). Copies of a sequence similar to that of the terminal 22bp motifs are also found internally, between the *tns* genes and the antibiotic resistance genes and oriented as in the Tn*7* left end (Figure 1.6; S. Bell, in press), but no TnsB-specific binding activity to them has been detected. The R2 motif contains a GATC *dam* methylation site; it has been shown (Diekmann, 1987) that DNA methylation can enhance or induce inherent DNA curvature and this, or simply the presence or absence of the methyl group, could be a factor in differential TnsB binding or in adoption of the correct conformation of the Tn*7* ends for transposition. Ekaterinaki (1987) however did not detect any effect of the *dam* methylation state of R2 on *in vivo* transposition frequency, as measured by plate mating.

A



B

RIGHT END

TGTGGGCG	GACAATAAAGTCTTAAACTGAA	R1
	AACAAAATAGATCTAAACTATG	R2
	GACAATAAAGTCTTAAACTAGA	R3
	GACAGAATAGTTGTAAACTGAA	R4

LEFT END

TGTGGGCG	GACAAAATAGTTGGGAACTGGG	L1
	GACAAAATAGATGGGAACTGGG	L2
	GACAAAATAGTTTGGAACTAGA	L3

INTERNAL

	GACATTTCTCCAAGCAACTACG	I1
	GACAACTCCATAAGAAATTACG	I2
	GACAATAGTCCATCCAATTACG	I3
	GACAACTCTGAGAGCAACTACG	I4
	GATAATAGTTCATCCAATTACG	I5

Figure 1.6 A Structure of the left and right ends of Tn*7*, showing the relative position and orientation of the copies of the 22bp repeated motif.

B Sequence of one strand of the various copies of the 22bp motif found at the right end, at the left end, and internally between the antibiotic resistance genes and *tnsE*, with numbering of the motifs as found throughout this thesis. Bases which are part of two overlapping motifs are shown italicised in the first listed motif. The *dam* methylation site in R2 is indicated by underlining. The internal repeats were both sequenced and identified by S. Bell (in press).
The terminal 8bp of the right and left ends are also shown; the first 22bp motif starts immediately after these.

The transposition reaction is apparently conservative *in vivo*, although the evidence is indirect. A successful *in vitro* transposition system has been described (Bainton *et al.*, 1991) and the technical details of this are given in Section 1.5.6. The reactions observed *in vitro* have also been conservative (Bainton *et al.*, 1991), and proceeded through an intermediate stage in which the transposon was completely excised from the donor replicon by double strand breaks; although the 3' end of each DNA strand of Tn*7* was apparently cut precisely, there was an overhang of 3 nucleotides of donor DNA seen at the 5' ends, as has also been postulated in the excision of IS*911* (Polard *et al.*, 1992). These extra nucleotides were not however an essential feature of the *in vitro* transposition mechanism. The 3' end of each transposon strand was then covalently joined to a free 5' end of the cleaved target DNA. It is not certain that *in vivo* transposition proceeds by the same mechanism as that seen *in vitro*, since *in vivo* conditions are much more complex than those *in vitro*. Apart from the simple insertion product, showing a 5bp target duplication and site- and orientation-specificity as *in vivo*, several other DNA molecules were seen; these were believed to be donor molecules with a staggered double strand break at one or other Tn*7* end, excised linear transposons and complementary gapped donor molecules. The excised linear Tn*7* was shown to transpose, given all the requirements of the complete transposition reaction. The donor molecules with a double strand break had the 3' transposon end cleaved precisely and a 3bp overhang at the 5' end of the transposon; they displayed reaction kinetics typical of a transposition intermediate. The reaction products had Tn*7* 3' ends covalently linked to the target DNA; there were gaps at the 5' ends, which is not surprising since there was no gap repair system in the reaction mixes.

Recent *in vivo* results (Hagemann and Craig, 1993) indicate that Tn*7* transposition may be followed by repair of the donor molecule by homologous recombination with a sister chromosome. This would have the effect of increasing the copy number of Tn*7* within a cell.

1.4.3 Interactions between Tn*7* and IS*21*

It has been observed (Schurter and Holloway, 1987) that when Tn*7* and the prokaryotic insertion element IS*21* are present together in a cell, each can stimulate the other to transpose. IS*21* has terminal 8bp repeat sequences which have 6 out of 8 bp identical to the corresponding base pairs of the Tn*7* terminal repeats (Reimmann *et al.*, 1989). A further

similarity between the two transposable elements is that, when tandem copies of IS*21* are present in on a plasmid, double strand, staggered cleavage of the ends occurs *in vitro* between the two copies of the insertion sequence in a similar fashion to that seen with Tn*7* (Reimmann and Haas, 1990). This staggered cleavage of the IS*21*-IS*21* junction is observed only in the presence of IstA; the 3' ends are cut precisely, leaving 5' overhangs of 2 or 3 bp (which could not be distinguished by the experimental technique used), similar to those seen with retroviruses and with Tn*7* transposition. IstB was shown not to be necessary for this cleavage reaction. No insertion products were seen in the *in vitro* system, but the experiments were not designed to look for these. The above observations suggest that end recognition and strand cleavage in the Tn*7* and IS*21* systems may take place through catalysis by enzymes with very similar active sites. IS*21* encodes two genes involved in transposition; as indicated above, *istA* is known to be required for strand cleavage (Reimmann and Haas, 1990). There is, however, only very limited similarity between the amino sequences of IstA and TnsB; the D-(35)-E motif of TnsB occurs as N-(35)-E in IstA, although the occurrence of this motif may be purely fortuitous; see Section 1.6 for further discussion of this.

A further similarity between the two elements is that both Tn*7* (Hauer and Shapiro, 1984) and IS*21* (Danilevich and Kostyuchenko, 1985) exhibit immunity to repeated transposition into the same replicon molecule. Immunity is found in Tn*3* transposition; it has, however been more thoroughly studied in Mu, where immunity is mediated by the nonspecific binding of MuB to target DNA, and MuA-stimulated ATP hydrolysis catalysed by MuB which releases the protein from the DNA.

The Tn*7* immunity is found in both the *tnsD* and *tnsE*-specific pathways (Arciszewska *et al.*, 1989) and may possibly involve nonspecific binding by TnsC to the target replicon, linked to ATP hydrolysis (Gamas and Craig, 1992). It is conceivable that IS*21* transposition immunity is influenced by binding of the IstB protein to DNA; IstB, like TnsC, has a nucleotide binding motif, but there has as yet been no published evidence that IstB binds DNA. The existence of immunity in both Tn*7* and IS*21* may however be purely fortuitous, rather than indicating functional similarities.

1.4.4 Properties of *att* sites and TnsE-dependent target sites

As mentioned in Section 1.4.2, Tn*7* can transpose by two different pathways, utilising either *tnsD* or *tnsE* in addition to *tnsA*, *tnsB* and *tnsC*.

High frequency transposition is mediated by *tnsD* pathway. The preferential *E. coli* chromosomal target site for Tn*7* transposition, *attTn7*, lies at approximately minute 84, between the *glmS* and *pstS* (*phoS*) genes (Lichtenstein and Brenner, 1982), such that expression of these genes is unaffected by the presence of the transposon. Insertion here is not only site-specific, but also orientation-specific; it is always the Right End of Tn*7* (as defined on Figure 1.5) which is adjacent to the carboxyl terminus of the *glmS* gene. The insertion site is in the transcriptional terminator of *glmS* (Gay *et al.*, 1986); the recognition site does not include the actual insertion site but is located to the side of it, in a manner reminiscent of Type III restriction enzymes. The base pairs 28 to 55 from the insertion site, including part of the coding region of *glmS*, are recognised and bound by TnsD protein (Waddell and Craig, 1988) and a subset of these, not yet precisely defined, is required for transposition to *attTn7*. In common with the majority of other prokaryotic transposition systems, but unlike the situation found with site-specific recombination by resolvases/invertases and integrases, the sequence of the actual insertion site does not bear any resemblance to that of the ends of the transposon. Also, as with most prokaryotic transposable elements, a duplication of the insertion site occurs upon transposition; in the case of Tn*7*, 5bp are duplicated, *in vitro* at least by a staggered cut of the target DNA with 5' overhangs, prior to transposon insertion. This break is presumably subsequently repaired *in vivo* by the host bacterium's DNA repair machinery.

If *attTn7* is deleted from the chromosome, Tn*7* will transpose, at a lower frequency, to other sites bearing structural similarity to *attTn7* (Craig, 1989), again using the pathway involving *tnsD*. Transposition to a specific chromosomal site also occurs in other Gram negative bacterial species (Thomson *et al.*, 1981; Ely, 1982; Caruso and Shapiro, 1982; Turner *et al.*, 1984; Thomasian and Voll, 1989) and it is likely that a common structural feature of the DNA is recognised in each case.

Low frequency transposition by the *tnsE* pathway occurs regardless of the presence or otherwise of *tnsD*. The chromosomal and plasmid target sites are apparently random, and bear no similarity either to *tnsD*-dependent sites or to the Tn*7* ends (Craig, 1989).

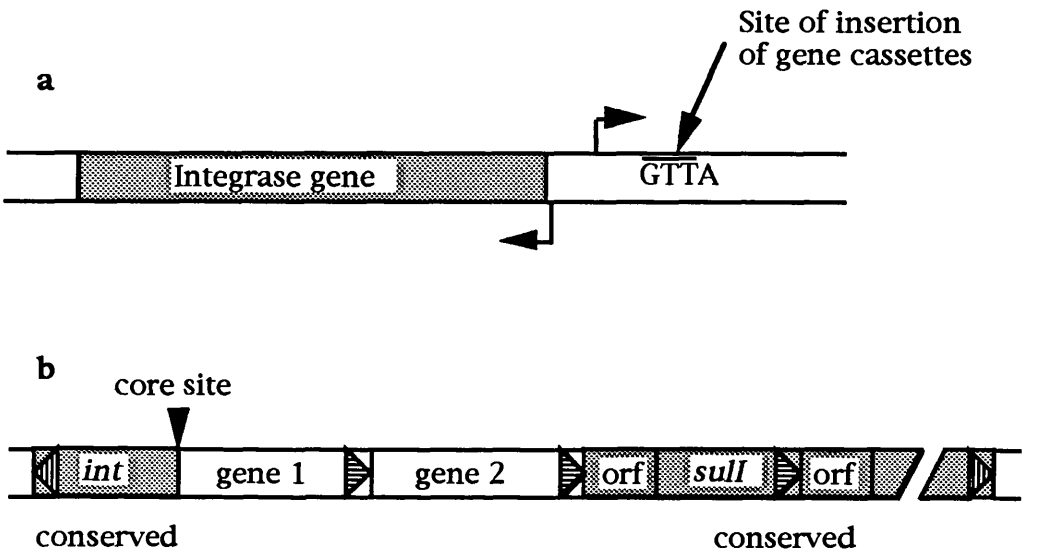


Figure 1.7 Integrons

(a) The integrase gene, divergent promoters and GTTA sequence at which cassette integration is believed to take place.
(b) Structure of a typical integron, showing conserved sequence including the integrase (*int*) and sulphonamide resistance (*sulI*) genes stippled, with terminal 25bp inverted Brown's repeats (striped vertically), and two inserted gene cassettes, each with a downstream copy of the 59bp palindromic sequence containing GTTA (striped horizontally). The 3' conserved sequence also contains an open reading frame (not shown on the Figure) the sequence of which bears similarity to those of TnsB and the A transposition gene of Tn*552*.

1.4.5 The antibiotic resistance genes and their possible derivation; relatives of Tn*7*

Tn*7* has three antibiotic resistance genes, *dhfr*, *sat* and *aadA*, encoding resistance to trimethoprim, streptothricin and streptomycin/ spectinomycin respectively. As shown in Figure 1.5, they are oriented in the same direction, and they are in fact believed to be co-transcribed from a promoter upstream of *dhfr*. Also upstream of *dhfr*, and reading in the opposite direction, is what appears to be a pseudogene bearing similarity to Lambda-type integrase genes; the sequence is interrupted by a stop codon in the isolates sequenced (Fling and Richards, 1983; Sundström *et al.*, 1991). The general arrangement of resistance genes and integrase is reminiscent of that found in integrons (Stokes and Hall, 1989).

A generalised integron structure, as postulated by Stokes and Hall (1989) and Martinez and de la Cruz (1990), and elaborated by Bissonnette and Roy (1992), is shown in Figure 1.7. The essential features are a gene encoding a site-specific integrase function, a divergent promoter, and a site at which genes, generally encoding antibiotic resistance, can be inserted (and possibly subsequently excised). Systems satisfying this criterion are found in plasmids and transposons of Gram negative bacteria (Stokes and Hall, 1989), and also in evolutionarily distant organisms such as the mycobacterial transposon Tn*610* (Martin *et al.*, 1990). Typically, each antibiotic resistance gene in an integron is demarcated at the 3' end by a copy of a 59bp imperfect inverted repeat sequence; the actual insertion site is towards the 3' end of this sequence (Hall *et al.*, 1991); this core site is also found between the 5' end of the integrase gene and the first antibiotic resistance gene (see Figure 1.7).

Mobility of the integron antibiotic encoding genes *in vivo* has recently been reported (Collis and Hall, 1992b). In the presence of the integrase gene, covalently closed DNA circles have been seen; these encoded one or more excised antibiotic genes, with no additional DNA apart from that in the 59bp motif. Insertion of these cassette circles has not yet been demonstrated.

The structure in Tn*7* associated with the three antibiotic resistance genes differs somewhat from the above description of integrons. However, there is probably sufficient similarity between the two systems to propose a common mechanism of antibiotic resistance dissemination.

Two transposable elements with a similar *tns* region to Tn*7*, but with a different arrangement of antibiotic resistance genes, are known. Tn*1825* and Tn*1826* (Tschäpe *et al.*, 1986; Tietze *et al.*, 1987) contain the *sat* and *aadA* genes of Tn*7*, but do not contain the *dhfr* gene which confers trimethoprim resistance to Tn*7* hosts. Tn*1826*, the shorter of the two elements also contains sequence coding for the putative integrase gene upstream of the *sat* gene (Tietze and Brevet, 1990); the sequenced part of the upstream DNA in Tn*1825* diverges from that of Tn*1826* (Heim *et al.*, 1989), and it is not known whether there is any integrase gene present.

1.4.6 The rest of Tn*7*

The Tn*7* sequence has recently been completed and assembled by S. Bell (in press). The transposon is now known to be 14070bp in length. Disappointingly, apart from the additional internal 22bp motifs, described in Section 1.4.2, which may or may not have evolutionary significance, no new interesting features have been identified.

1.5 *In vitro* systems for transposition

While most of the study of transposable elements has been carried out on *in vivo* systems, the means of studying transposition *in vitro* have recently been developed for a number of elements. These systems vary in the purity of their reagents, from crude cell extracts to proteins purified to homogeneity, as elucidated below. The reaction conditions and products of the *in vitro* transposition reactions are not necessarily those which are most commonly seen *in vivo*, particularly in the case of Tn*10* reaction products. Nevertheless, the *in vitro* transposition systems provide a useful indication of the reaction requirements *in vivo*, and alteration of the reaction conditions yields helpful information on reaction intermediates.

The reaction conditions used for the known *in vitro* systems are given below; the conclusions drawn from *in vitro* experiments are detailed in Sections 1.3 and 1.4.2.

1.5.1 Bacteriophage Mu

The best studied *in vitro* transposition system is that of the replicative transposition of bacteriophage Mu. (The conservative initial integration event has proved less amenable to study.) An *in vitro* reaction system was first reported by Mizuuchi in 1983; this was loosely based on the

system developed for the replication of the *E. coli* chromosomal origin (Fuller *et al.*, 1981), and involved plasmid-borne mini-Mu transposons containing the correctly oriented ends and (although this was not recognised until later) the IAS; circularised bacteriophage Lambda was added as a target molecule for transposition; this was subsequently packaged and used to infect *E. coli* cells. The Mu A and Mu B proteins, overexpressed from plasmids, were supplied in slightly purified form from crude cell extracts; RNA, deoxynucleotides (sometimes radioactively labelled), Mg^{2+}, ATP and the components of the cellular ATP regeneration system were also added. The key to the success of the reactions was considered to be the volume-excluding hydrophilic polymer polyvinyl alcohol (PVA), the presence of which enabled the reaction components to come into close proximity without being in very high overall concentration.

The *in vitro* reaction appeared to have the same characteristics as the *in vivo* reaction: requirement for correctly oriented Mu ends and for Mu A and Mu B protein; lack of target site specificity; cointegrate product molecules carrying two copies of the mini-Mu and exhibiting evidence of substantial incorporation of DNA replication precursors; and immunity to repeated transposition into the same DNA molecule.

Since these first experiments, the transposition reaction conditions have been refined; Mu A and Mu B proteins have been purified (Craigie and Mizuuchi, 1985b; Chaconas *et al.*, 1985) and the host factor requirements (for ATP and HU) have been clarified. *In vitro* transposition intermediates have been identified both by electron microscopy and by electrophoretic analysis, and the different pathways by which transpososomes may assemble have been elucidated.

1.5.2 Tn*10*

The technique of Tn*10* transposition *in vitro* was developed (Morisato and Kleckner, 1987), using as a basis the established system for bacteriophage Mu *in vitro* transposition. The two differed, however in three respects: firstly, whereas the Mu transposition products appeared the same *in vivo* and *in vitro*, the *in vitro* reaction products seen in the Tn*10* system were nearly all intramolecular, rather than the intermolecular ones detected *in vivo*. Secondly, the Tn*10* transposase proved much more difficult to purify than the Mu transposition proteins and so, until very recently, reactions have been carried out using only partially pure protein

and thus masking a possible requirement for a host protein in the reaction additional to IHF and HU; in the most recent published paper (Benjamin and Kleckner, 1992), the Tn*10* transposase was purified to 50% of total protein and stored in 2M urea to maintain activity, and so there was still urea in the transposition reaction mixture; this may possibly have affected the reactions. A more recent transposase purification procedure (R. Chalmers and N. Kleckner, unpublished) disposes of the requirement for urea; it is not known to the present author whether the reaction is sensitive to the transposase purification conditions. A third difference between the Tn*10* and Mu reactions is that the latter requires ATP, but no energy source is needed for the Tn*10* reaction. Excised linear transposons are clearly seen in the Tn*10 in vitro* reaction mixtures, but it has not been demonstrated that these are an intermediate in the transposition reaction, and they may be an artefact of the *in vitro* system.

1.5.3 Retroviral integration

Research has recently been initiated into the *in vitro* integration of a number of retroviruses. The purity of the reagents and efficiency of the reaction vary from system to system and appear to be inversely related: the Moloney murine leukaemia virus inserts into DNA moderately efficiently, using a system where the integrase protein is only partially purified (Craigie *et al.*, 1990), whereas HIV integration occurs *in vitro* with purified IN protein at much lower frequency (Mizuuchi, 1992). The most effective systems are those in which integrase-retrovirus complexes are isolated from nucleocapsids and provided with target DNA (for example, Bowerman *et al.*, 1989; Farnet and Haseltine, 1990). This suggests that the reaction requires more than simply IN protein and retroviral DNA, and that some component, either of the viral nucleocapsid or of the host cell is additionally involved. In the early experiments, selectable markers were added to the retroviral genome to simplify scoring of insertion events into a bacteriophage Lambda derivative which was subsequently packaged and used to infect *E. coli* (Bowerman *et al.*, 1989).

As implied above, integration can still take place *in vitro* if the retroviral genome is altered, providing the terminal sequences are retained. The cation specificity of the reactions varies depending on the DNA structure and on the purity of the IN protein (Mizuuchi, 1992).

1.5.4 P element

An *in vitro* system for studying transposition of the *Drosophila* P element has recently been reported (Kaufman and Rio, 1992). As with the early experiments with *in vitro* retrovirus integration, a marker selectable in *E. coli* was inserted into the mini-P element. Partially purified transposase protein was added to the donor plasmid and a recipient plasmid; transposition occurred at the same frequency regardless of whether the donor plasmid was supercoiled or linear, neither was there a requirement for supercoiling of the target plasmid. It was found that the only other requirements for the reaction were Mg^{2+} ions and GTP; non-hydrolysable analogues of GTP also satisfied the requirement, indicating that GTP hydrolysis was not linked to the strand transfer reaction, a conclusion similar to that arrived at with Tn*7* transposition (Gamas and Craig, 1992). Since the transposase was only partially pure, it is possible that other reaction requirements were contained in the protein extract used. The efficiency of the reaction was reduced in the presence of a high concentration of monovalent cation. The reaction products had the flanking 8bp target site repeat seen in *in vivo* reactions, and were consistent with the "cut and paste" mechanism proposed for the *in vivo* reaction (Engels *et al.*, 1990); transposition could proceed with pre-cleaved transposon DNA. Where the 3' end of these linearised elements was altered to a dideoxynucleotide, no transposition took place, suggesting that the polarity of strand transfer is the same, *in vitro* at least, for the P element as for other transposable elements studied.

1.5.5 IS*2 1*

This system (Reimmann and Haas, 1990) addressed only the question of donor cleavage, and no attempt was made to observe integrated transposons. It mirrored the commonly found *in vivo* situation where IS*2 1* is found in tandemly repeated copies, separated by 3bp on plasmids; the two IS*2 1* copies are able to act like a composite transposon and, by cleavage at the junction of the insertion sequences, to transpose non-replicatively with the plasmid backbone and form cointegrates. The IstA protein in the experiments of Reimmann and Haas was supplied in crude cell extracts in which the *istA* gene had been overexpressed. The $(IS2\,1)_2$ plasmid substrate was supercoiled, although it was not clear whether this was necessary; linearised DNA did not act as a substrate.

1.5.6 Tn*7*

The only functioning *in vitro* transposition system known for Tn*7* is that of Bainton *et al.* (1991). This initially used crude cell extracts as a source of the Tns proteins, but these are now purified to varying degrees. The Tn*7* used was a mini-Tn*7* on a donor plasmid; transposition was more effective when this plasmid was supercoiled. A target plasmid encoding *att Tn7* was also added; very few reaction intermediates were seen in the absence of target DNA (Bainton, 1992). Apart from DNA and TnsA, B, C and D, the *in vitro* transposition reaction required Mg^{2+} and either ATP or dATP. The Mg^{2+} was added after preincubation of the other requirements; failure to do this resulted in a substantial reduction in transposition frequency. There was no evidence for the requirement for a host-encoded protein such as IHF or HU, but this would not have been detected since the Tns proteins were only partially purified and were contaminated by host proteins.

1.6 Tn*7* transposition and its possible relationship with retroviral integration

It has been observed (Fayet *et al.*, 1990) that there are similarities between the amino acid sequences of the retroviral integrases and the transposases of the IS*3* family. Further inspection reveals that this similarity extends beyond the IS*3* family to the A protein of transposon Tn*552* found in Gram-positive bacteria (Rowland and Dyke, 1990), and to TnsB of Tn*7* and IstA of IS*21* (Fayet *et al.*, 1990). An alignment, carried out with assistance from S. Rowland (pers. comm.), summarising these similarities, is shown in Figure 1.8. The amino terminal part of these proteins is believed to contain a DNA binding domain, probably in the form of a cysteine/histidine zinc finger in the retroviral proteins (Engelman and Craigie, 1992), and as a helix-turn-helix structure in the transposases. As mentioned in Section 1.3.7, the conserved amino acids include some of those necessary for the retroviral integration process (Engelman and Craigie, 1992), although it has not been proved that these residues are actually in the active site of the enzyme, and do not simply have an important structural rôle in the enzyme. It can be seen from Figure 1.8 that some, but not all, of the residues which are highly conserved in the integrase proteins of retroviruses are also found in TnsB and in IstA of IS*21*. For example, the aspartic acid residue marked • on the Figure is

```
Visna      51 SNKRGID... HWQVDYTHYE ..........  DKIIL....V WVETNSGLIY AERVK.GETG
HIV1       49 QVDCSPG... IWQLDCTHLE ..........  GKVIL....V AVHVASGYIE AEVIP.AETG
BLV        55 TIQRGWAPNH IWQADITHYK ..........  YKQFTYALHV FVDTYSGATH ASAKR.GLTT
HTLV1      52 HIRRGLLPNH IWQGDITHFK ..........  YKNTLYRLHV WVDTFSGAIS ATQKR.KETS
RSV        49 VNPRGLGPLQ IWQTDFTLEP ..........  RMAPRSWLAV TVDTASSAIV VTQHG.RVTS
MMTV       50 VNPRGLKPRV LWQMDVTHVS ..........  EFGKLKYVHV TVDTYSHFTF ATART.GEAT
MoMLV      62 TRVRGHRPGT HWEIDFTEIK P.........  GLYGYKYLLV FIDTFSGWIE AFPTK.KETA
copia      64 THIK..RPLF VVHSDVCGPI TPV.......  TLDDKNYFVI FVDQFTHYCV TYLIKYKSDV
Tn552     123 ESSR...PNE IWQADHTLLD IYILDQKGN. ..INRPWLTI IMDDYSRAIA GYFISFDAPN
Tn402     134 PPAVT.APLE QVQIDHTVID LIVVDDRDRQ P.IGRPYLTL AIDVFTRCVL GMVVTLEAPS
TnsB      165 ATSQALGPGS RYEIDATIAD IYLVDHHDRQ KIIGRPTLYI VIDVFSRMIT GFYIGFENPS
IS21       71 RFHVFAAP.. KQDAEHTYES LVRAFR.... RYFGGCVKTV LVDNQKAA.. ..........
IS629     113 RQFVAERPDQ LWVADFTYVS ..........  TWQGFVYVAF IIDVFAGYIV GWRV.....S
IS911     107 RQFAVTEPNQ VWCGDVTYIW ..........  TGKRWAYLAV VLDLFARKPV GWAM.....S
IS3       114 QDFYASGPNQ KWAGDITYLR ..........  TDEGWLYLAV VIDLWSRAVI GWSM.....S
IS2       122 QRWCSD.... ....GFEFCC ..........  DNGERLRVTF ALDCCDREAL HWAVTTGGFN
                          •  •                              •

Visna          QEFRVQTMKW Y.AMF..... ..........  ......A.PK SLQSDNGPAF VAESTQLLMK
HIV1           QETAYFLLKL A.GRW..... ..........  ......P.VK TVHTDNGSNF TSTTVKAACW
BLV            QTTIEGLLEA I.VHL..... ..........  ......GRPK KLNTDQGANY TSKTFVRFCQ
HTLV1          SEAISSLLQA I.AHL..... ..........  ......GKPS YINTDNGPAY ISQDFLNMCT
RSV            VAVQHHWATA I.AVL..... ..........  ......GRPK AIKTDNGSCF TSKSTREWLA
MMTV           KDVLQHLAQS F.AYM..... .........  .......GIPQ KIKTDNAPAY VSRSIQEFLA
MoMLV          TVVTKKLLEE IFPRF..... ..........  ......GMPQ VLGTDNGPAF VSKVSQTVAD
copia          FSMFQDFVAK SEAHFN.... ..........  ......LKVV YLYIDNGREY LSNEMRQFCV
Tn552          AQNTALTLHQ AIWNKNN... ..........  TNWPVCGIPE KFYTDHGSDF TSHHMEQVAI
Tn402          AVSVGLCLVH VACDKRPWLE GLNVEMDWQM  SASPCCSR*
TnsB           YVVAMQAFVN ACSDKTAICA QHDIEISS..  SDWPCVGLPD VLLADRG.EL VSHQVEALVS
IS21           .......... ..........  ..........  .......... .LKNNNGKVV FNSGFLLLAD
IS629          SSMETTFVLD A.........  ...LEQALWA R.RPSGTV..  .HHSDKGSQY VSLAYTQRLK
IS911          FSPDSRLTMK A.........  ...LEMAWET RGKPGGVM..  .FHSDQGSHY TSRQFRQLLW
IS3            PRMTAQLACD A.........  ...LQMALWR RKRPRNVI..  .VHTDRGGQY CSADYQAQLK
IS2            STEVQDVMLG A.........  ...VERRFGN DLPSSPVE..  .WLTDNGSCY RANETRQFAR
                                                             •••••  •

Visna          YLGIEHTTGI PWNPQSQALV ERTHQTLKNT 118
HIV1           WAGIKQEFGI PYNPQSQGVI ESMNKELKKI 126
BLV            QFGVSLSHHV PYNPTSSGLD ERTNGLLKLL 123
HTLV1          SLAIRHTTHV PYNPTSSGLV ERSNGILKTL 130
RSV            RWGIAHTTGI PGNSQGQAMV ERANRLLKDR 157
MMTV           RWKISHVTGI PYNPQGQAIV ERTHQNIKAQ 152
MoMLV          LLGIDWKLHC AYRPQSSGQV ERMNRTIKET 179
copia          KKGISYHLTV PHTPQLNGVS ERMIRTITEK 367
Tn552          DLKINLMFSK VGVPRGRGKI ERFFQTVNQT 195
Tn402
TnsB           SFNVRVESAP PRRGDAKGIV ESTFRTLQAE 297
IS21           HYNFLPRACR PRRARTKGKV ERMVKYLKEN  61
IS629          EAGLLASTGS TGDSYDNAMA ESINGLYKAK  56
IS911          RYQIRQSMSR RGNCWDNSPM ERFFRSLKNE  50
IS3            RHNLRGSMSA KGCCYDNACV ESFFHSLKVE  48
IS2            MLGLEPKNTA VRSPESNGIA ESFVKTIKRD  58
                  •          •   • • ••    ••
```

Figure 1.8 The region of similarity in the amino acid sequences of retroviral integrases, the retrotransposon *copia* integrase and prokaryotic transposases. Highly conserved residues are indicated •. The alignment was carried out with the assistance of S.Rowland, using the GCG Pileup program. The unpublished Tn*402* sequence was obtained from P. Rådström.

HIV	Human Immunodeficiency Virus	BLV	Bovine Leukaemia Virus
HTLV	Human T-cell Leukaemia Virus	RSV	Rous Sarcoma Virus
MMTV	Mouse Mammary Tumour Virus	MoMLV	Moloney Murine Leukaemia Virus

absolutely required in HIV-1 and HIV-2 integrase, and substitution by asparagine (Engelman and Craigie, 1992) or isoleucine (van Gent *et al.*, 1992) completely inactivates the protein; yet asparagine rather than aspartic acid is found at this point in the sequence of IstA. This suggests that the structure (and possibly the function) of IstA must be sufficiently different from that of the retroviral integrases to accommodate this substitution; since the transposition machineries of Tn*7* and IS*21* can interact, it is likely that, *if* the enzymatic rôle of TnsB were homologous to that of IstA, which has not been demonstrated, then the residues marked • may not be critical for function of TnsB either.

A further similarity between the elements discussed above is that each has the same terminal DNA sequence of 5'-CA-OH-3'; it is known that a staggered cut at this point, leaving a short 5' overhang, occurs with Tn*7* (Bainton *et al.*, 1992), IS*21* (Reimmann and Haas, 1990), Moloney Murine Leukaemia Virus (Craigie *et al.*, 1990) and HIV (Engelman *et al.*, 1991), and it has been proposed with IS*911* (Polard *et al.*, 1992).

From the above, it may seem that TnsB might be homologous to the retroviral integrases. However no strand processing or strand transfer rôle akin to that of the retroviral integrases has been assigned to TnsB; also, not all the residues conserved in the integrases appear in TnsB, and the spacing between the two crucial residues marked • on Figure 1.8 is 34 amino acids in TnsB, but 35 amino acids in all the other integrases and transposases; this may or may not be significant. Furthermore, TnsB is considerably larger (81kD) than the 32-46kD of retroviral integrases, and may well code for additional functions. The integrase protein of the retrotransposon *copia* is also substantially larger than the retroviral integrase proteins, and it is possible that these two share some additional function. It is also possible that TnsB carries out the same functions as the integrases, but that the residues identified as conserved are not in fact the ones which are significant.

1.7 Aims of this project

The aims of this project were to investigate the rôles of the *tnsA* and *tnsB* genes of Tn*7* in the transposition reaction of the element, both *in vivo* and *in vitro*. While there was some pre-existing information on the protein product of the *tnsB* gene and its properties, nothing was known about the

tnsA gene apart from the fact that its presence was required either in *cis* or in *trans* for transposition to take place.

More specifically, it was hoped to overexpress the *tnsA* gene, to purify the protein product if it was required for transposition, and to determine the rôle *tnsA* played in transposition. It was also planned to investigate regulation of Tn*7* transposition by the TnsB protein and by TnsB peptides; and to ascertain whether TnsB was involved in synapsis of the two Tn*7* ends during the early stages of transposition.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Strains, plasmids, phagemids and transposons

2.1.1 *E. coli* strains

Strain	Genotype	Source
BMH71-18mutS	*thi*,*supE*,Δ(*lac proAB*),[*mutS*::Tn*10*] [F',*proAB*,*lacI*^{q}ZΔM15]	Promega
CSH26	*ara*,Δ(*lac pro*),*thi*	D.Sherratt
DS801 (AB1157)	*thr*1,*leuB*6,*hisG*4,*thi*1,*ara*14,*lacY*1,*proA*2,*argE*3, *galK*2,*supE*44,*xyl*5,*mtl*1,*tsx*33,*rpsL*31	D.Sherratt
DS889 (CB51)	CSH26, but *dam*3,*rpsL*	C.Boyd
DS902 (AB2463)	DS801, but *recA*13,*arg*$^+$	D.Sherratt
DS903	DS801, but *recF*143	D.Sherratt
DS916	JC5466 but rif^r,sp^r,*recA*,*his*,*trp*,λ$^+$	D.Sherratt
DS941	DS903, but *supE*44,*lacI*^{q}ZΔM15	D.Sherratt
JM109	*endA*1,*recA*1,*gyrA*96,*thi*,*hsdR*17($r_k^-m_k^+$), *relA*1,*supE*44,λ$^-$,Δ(*lac proAB*), [F',*traD*36,*proAB*,*lacI*^{q} ZΔM15]	Promega
LA547	$lacZ_{am}$,trp_{am},pho_{am},$supC_{ts}$,*rpsL*,$htpR_{165}$	L.Arciszewska
ML308	*lacI*$^-$	I.S.Hunter
MR1	DS916, but nal^r	M.Rogers
MR4	DS903::Tn*7*-1	M.Rogers
MR5	MR4 [F',*proAB*,*lacI*^{q}ZΔM15]	M.Rogers
DS941::Tn*7*-1	P1 transduction of Tn*7*-1 into *attTn7*	K.McCurrach
AG2	LA547::Tn*7*-1, made by P1 transduction	Chapter 4
AG3	LA547::Tn*7*, made by P1 transduction	Chapter 4

Table 2.1 Bacterial strains

Table 2.2 Plasmids and phagemids used

Plasmid	Size/kb	Marker	Origin	Comments	Source
pUC18	2.7	Ap	ColE1	p*lac* cloning vector	C. Yanisch-Perron *et al.* (1985)
pACYC184	4.2	Tc Cm	p15A	Cloning vector	A. Chang and S. Cohen (1978)
pBEND2	2.5	Ap	ColE1	Cloning vector	C. Zwieb *et al.* (1989)
R388	32.2	Tp Su	IncW	Naturally occurring conjugative plasmid	N. Datta and R. Hedges (1972)
pMR100	5.6	Km	λdv	p*tac* expression vector	M. Rogers *et al.* (1986)
pMR78	3.4	Tp	ColE1	p*tac* expression vector	M. Rogers (1986)
pGLW8	4.5	Ap	ColE1	p*tac* expression vector similar to pKK223-3	I. S. Hunter
pGEX-3X	4.1	Ap	ColE1	p*tac* expression vector for GST fusions	Pharmacia
pBluescript	3.0	Ap	ColE1/ f1	Cloning vector phagemid	Stratagene
pSELECT-1	5.7	Tc (Ap)	ColE1/ f1	Cloning vector phagemid for site-directed mutagenesis	Promega
pT7-5	2.5	Ap	ColE1	T7 expression vector for transcriptional fusions	S. Tabor (unpublished)
pT7-7	2.5	Ap	ColE1	T7 expression vector for translational fusions	S. Tabor (unpublished)
pGP1-2	7.2	Km	p15A	T7 RNA polymerase producer	S. Tabor and C. Richardson (1985)
p253	5.8	Cm	p15A	*galK* transcriptional fusion vector	M. Burke (Ref R. Slatter, 1987)
pUC18::*lacZ*	5.8	Ap	ColE1	BamH1 fragment containing *lacZ* in pUC18 polylinker	M. Burke (unpublished)
pSB58	8.8	Ap	ColE1	0.135-6.495kb from Tn*7* RE in pT7-7; encodes active *tnsA,B,C,D*	S. Bell (unpublished)
pSB84	8.8	Ap	ColE1	0.001-6.495kb from Tn*7* RE in pT7-5; encodes active *tnsA,B,C,D*	S. Bell (unpublished)
pEAL1	5.2	Tc	p15A	1kb *att*Tn*7* in pACYC184	C. Lichtenstein & S. Brenner (1981)
pEN300	33.3	Tp Su	IncW	1kb *att*Tn*7* in R388	E.Nimmo (ref. M.Rogers *et al.*, 1986)
pMR80	3.0	Ap	ColE1	280bp *att*Tn*7* in pUC18	M. Rogers (1986)
pMR86	3.7	Ap	ColE1	1kb *att* Tn*7* in pUC8	M. Rogers (1986)
pNE200	2.9	Ap	ColE1	0.001-0.203kb from Tn*7* RE in pUC8	N. Ekaterinaki (1987)
pMR9	11.7	Ap	ColE1	0.001-~9kb (EcoRI) from Tn*7* RE in pUC8	M. Rogers (1986)
pMR11	5.5	Ap Cm	ColE1	Tn*7*-1 constructed in pUC18	M. Rogers *et al.* (1986)

pMR25	5.8	Tc	p15A	0.001-1.485kb from Tn*7* RE in pACYC184	M. Rogers *et al.* (1986)
pMR106	9.0	Km	λdv	3.024-6.495kb from Tn*7* RE in pMR100; encodes active *tnsC,D*	M. Rogers *et al.* (1986)
pMR121	12.0	Km	λdv	0.001-6.495kb from Tn*7* RE in pMR100; encodes active *tnsA,B,C,D*	M. Rogers (1986)
pMR204	4.8	Tp	ColE1	0.125-1.485kb from Tn*7* RE in pMR78; encodes active *tnsA*	M. Rogers (unpublished)
pMR205	6.7	Tp	ColE1	0.125-3.437kb from Tn*7* RE in pMR78; encodes active *tnsA,B*	M. Rogers (unpublished)
pMR207	5.9	Tp	ColE1	0.899-3.437kb from Tn*7* RE in pMR78; encodes active *tnsB*	M. Rogers (unpublished)
pLA50	5.0	Ap Km	ColE1	miniTn*7* lacking R4 constructed in pUC18	L. Arciszewska *et al.* (1989).
pLA77	3.0	Ap	ColE1/ f1	L3 cloned into Bluescript-KS	L. Arciszewska and N. Craig (1991)
pLM5	2.7	Ap	ColE1	0.009-0.030kb from Tn*7* RE in pUC18	E. Morrell (1991)
pLM7	2.7	Ap	ColE1	0.001-0.030kb from Tn*7* RE in pUC18	E. Morrell (1991)
pLM8	2.6	Ap	ColE1	0.001-0.030kb from Tn*7* RE in pBEND2	E. Morrell (1991)
pAG012	5.0	Ap Cm	ColE1	pMR11 with the PstI Tn*7* RE-containing fragment deleted	Chapter 5
pAG013	5.7	Ap Cm	ColE1	pMR11 with the PstI Tn*7* RE-containing fragment reversed in orientation	Chapter 5
pAG077	4.6	Ap	ColE1	1.485-3.437kb from Tn*7* RE with RBS and in-frame start codon from pT7-7, in pUC18	Chapter 4
pAG078	3.7	Ap	ColE1	0.899-1.859kb from Tn*7* RE in pUC18	Chapter 4
pAG079	4.7	Ap	ColE1	0.125-2.209kb from Tn*7* RE in pUC with 5' *lacZ-tnsA* translational fusion;encodes active *tnsA*	Chapter 4
pAG080	4.0	Ap	ColE1	0.899-2.209kb from Tn*7* RE in pUC18	Chapter 4
pAG081	4.0	Ap	ColE1	0.125-1.485kb from Tn*7* RE in pUC with 5'*lacZ-tnsA* translational fusion; encodes active *tnsA*	Chapter 3
pAG082	5.9	Ap	ColE1	0.125-3.437kb from Tn*7* RE in pUC with 5'*lacZ-tnsA* translational fusion; encodes active *tnsA,B*	Chapter 3

pAG083	3.5	Ap	ColE1	0.125-1.010kb from Tn*7* RE in pUC with 5'*lacZ-tnsA* translational fusion; encodes active *tnsA*	Chapter 3
pAG084	4.0	Ap	ColE1	insert as pAG081, but in reverse orientation	Chapter 3
pAG085	3.2	Ap	ColE1	0.889-1.485kb from Tn*7* RE in pUC18	Chapter 4
pAG086	5.3	Ap	ColE1	0.889-3.437kb from Tn*7* RE in pUC18; encodes active *tnsB*	Chapter 4
pAG087	4.6	Ap	ColE1	1.485-3.437kb from Tn*7* RE in pUC8	Chapter 4
pAG088	4.1	Ap	ColE1	0.125-1.668kb from Tn*7* RE in pUC8	Chapter 4
pAG089	3.3	Ap	ColE1	0.899-1.668kb from Tn*7* RE in pUC8	Chapter 4
pAG091	2.9	Ap	ColE1	Two R1 motifs in the same orientation separated by 42bp, cloned in pUC18	Chapter 4
pAG095	2.7	Ap	ColE1	R4 motif cloned SalI-XbaI into pUC18	Chapter 4
pAG096	2.7	Ap	ColE1	R3+R4 motifs cloned SalI-XbaI into pUC18	Chapter 4
pAG101	12.5	Km	λdv	0.532-8.351kb from Tn*7* RE in pMR100; encodes active *tnsB,C,D,E*	Chapter 3
pAG201	3.2	Tp	ColE1	exoIII deletion of pMR204, leaving 0.125-~0.560kb *tnsA* sequence	Chapter 4
pAG202	4.2	Tp	ColE1	exoIII deletion of pMR204, leaving 0.125-~1.170kb from Tn*7* RE; encodes active *tnsA*	Chapter 4
pAG204	4.3	Tp	ColE1	0.125-1.010kb from Tn*7* RE in pMR78; encodes active *tnsA*	Chapter 3
pAG205	4.2	Tp	ColE1	0.899-1.485kb from Tn*7* RE in pMR78	Chapter 4
pAG208	5.4	Tp	ColE1	1.485-3.441kb from Tn*7* RE, with RBS and in-frame start codon from pT7-7, in pMR78	Chapter 4
pAG209	4.7	Tp	ColE1	0.899-2.209kb from Tn*7* RE in pMR78	Chapter 4
pAG210	4.1	Tp	ColE1	0.899-1.668kb from Tn*7* RE in pMR78	Chapter 4
pAG211	4.6	Tp	ColE1	0.899-1.859kb from Tn*7* RE in pMR78	Chapter 4
pAG301	6.2	Ap	ColE1	0.125-1.485kb from Tn*7* RE in pGEX-3X; encodes active GST-*tnsA* fusion	Chapter 3

pAG302	6.2	Ap	ColE1	As pAG302, but frameshift between GST and *tnsA*	Chapter 3
pAG401	7.1	Tc (Ap)	ColE1/ f1	0.125-1.485kb from Tn*7* RE in pSELECT-1; encodes active *tnsA*	Chapter 3
pAG402	7.1	Tc Ap	ColE1/ f1	As pAG401, but *bla* gene repaired, and start codon of *tnsA* destroyed and mutagenised to HpaI site	Chapter 3
pAG404	7.1	Tc Ap	ColE1/ f1	As pAG401, but *bla* gene repaired, and start codon of *tnsA* retained and mutagenised to NdeI site	Chapter 3
pAG411	3.8	Ap	ColE1	0.135-1.485kb from Tn*7* RE (cut at pAG404 NdeI site) and cloned into pT7-7; encodes active *tnsA*	Chapter 3
pAG415	3.3	Ap	ColE1	0.135-1.010kb from Tn*7* RE cloned into pT7-7; encodes active *tnsA*	Chapter 3
pAG416	3.8	Ap	ColE1	As pAG411, but *tnsA* start codon has been destroyed	Chapter 3
pAG417	4.6	Ap	ColE1	1.485-3.437kb from Tn*7* RE cloned in frame into pT7-7	Chapter 4
pAG421	7.1	Tc Ap	ColE1/ f1	As pAG401, but *bla* gene repaired, and TnsB K121→Q with introduced PvuII site	Chapter 4
pAG422	7.1	Tc Ap	ColE1/ f1	As pAG401, but *bla* gene repaired, and TnsB R124→T with introduced HpaI site	Chapter 4
pAG423	7.1	Tc Ap	ColE1/ f1	As pAG401, but *bla* gene repaired, and TnsB K116→S with introduced SpeI site	Chapter 4
pAG424	7.1	Tc Ap	ColE1/ f1	As pAG401, but *bla* gene repaired, and TnsB V119→Y with introduced SnaBI site	Chapter 4
pAG431	5.9	Ap	ColE1	pAG082 with TnsB K121→Q and PvuII site (as pAG421)	Chapter 4
pAG432	5.9	Ap	ColE1	pAG082 with TnsB R124→T and HpaI site (as pAG422)	Chapter 4
pAG433	5.9	Ap	ColE1	pAG082 with TnsB K116→S and SpeI site (as pAG423)	Chapter 4

pAG434	5.9	Ap	ColE1	pAG082 with TnsB V119→Y and SnaBI site (as pAG424)	Chapter 4
pAG501	6.2	Cm	p15A	0.037-0.468kb from Tn*7* RE cloned with *galK* and 5'*tnsA* in opposite orientations in p253	Chapter 4
pAG502	6.2	Cm	p15A	0.037-0.468kb from Tn*7* RE cloned with *galK* and 5'*tnsA* in the same orientation in p253	Chapter 4
pAG503	6.2	Cm	p15A	0.001-0.468kb from Tn*7* RE cloned with *galK* and 5'*tnsA* in the same orientation in p253	Chapter 4

2.1.2 Plasmids and phagemids

See Table 2.2

2.1.3 Transposons

Tn*7*	wildtype Tn*7* as found on R483	P. Barth
Tn*7*-1	Tn*7* RE:1-568; Cm^r; Tn*7* LE:168-1	M. Rogers

2.2 Chemicals

2.2.1 Antibiotics

Name	stock solution	selective concentration
Ampicillin	5 mg ml^{-1} in H_2O	50 μg ml^{-1}
Trimethoprim	5 mg ml^{-1} in 50% ethanol	50 μg ml^{-1}
Chloramphenicol	2.5 mg ml^{-1} in ethanol	25 μg ml^{-1}
Tetracycline	1 mg ml^{-1} in 10mM HCl	10 μg ml^{-1}
Kanamycin monosulphate	5 mg ml^{-1} in H_2O	50 μg ml^{-1}
Rifampicin	5 mg ml^{-1} in methanol	50 μg ml^{-1}
Nalidixic acid sodium salt	2 mg ml^{-1} in H_2O	20 μg ml^{-1}
Sulphanilamide	100 mg ml^{-1} in DMSO	1 mg ml^{-1}
Streptomycin sesquisulphate	30 mg ml^{-1} in H_2O	300 μg ml^{-1} (chromosomal) 5 μg ml^{-1} (Tn*7*)
Spectinomycin dihydrochloride	2.5 mg ml^{-1} in H_2O	25 μg ml^{-1}

Isosensitest broth/agar (Oxoid) was used instead of L broth/agar when selecting for trimethoprim or sulphanilamide resistance.

2.2.2 General chemicals

Other chemicals were obtained from Sigma, B.D.H., Fisons, BRL and Boehringer Mannheim.

2.2.3 Stock solutions

Standard solutions were made and stored as described by Maniatis *et al.* (1982). The pH of all solutions made was measured at room temperature (20°C).

2.3 Gels

2.3.1 Agarose gels

Horizontal submarine gels were run in Tris acetate buffer pH7.9 (Maniatis *et al.*, 1982). Standard DNA gels were run at room temperature; binding gels were run at 4°C.

2.3.2 Agarose/polyacrylamide composite gels

These were poured and run according to Dahlberg *et al.* (1969).

2.3.3 Polyacrylamide gels

Small DNA fragments were separated on 6% polyacrylamide gels, run in TBE pH8.3 (Maniatis *et al.*, 1982) at room temperature.

Sequencing gels were made as described in the BRL sequencing instruction manual with 8% polyacrylamide in TBE and run at 49W. They were soaked in 10% acetic acid, 12% methanol to remove urea before drying. Oligonucleotides were purified on short sequencing gels run at 30W; the DNA was excised from the wet gel.

Binding gels were of 6% polyacrylamide and were made and run in one of TE (Maniatis *et al.*, 1982), TBE or 50mM Tris-glycine pH9.4 (Bednarz *et al.*, 1990). They were pre-run at 200W for at least 60 minutes at 4°C with recirculation of the running buffer before loading the reaction samples. Gels containing radioactively labelled DNA were dried before exposure to film.

SDS polyacrylamide protein gels were run according to the method of Laemmli (1970).

2.3.4 Staining methods

DNA gels were stained with ethidium bromide.

Protein gels were stained with Coomassie Brilliant Blue G250.

Gels containing low protein concentrations were silver stained as described below:

Prefix 50% methanol, 10% acetic acid	30 minutes;
Prefix 5% methanol, 7% acetic acid	30 minutes;
Fix 10% glutaraldehyde	30 minutes;
Rinse and soak in distilled water	60 minutes;
Change water and soak again	30 minutes;
Soak in 0.5mg DTT in 100ml distilled water	30 minutes;

Pour off the solution, but do not rinse the gel;

Add 1ml 20% $AgNO_3$ to 199ml distilled water 30 minutes;

Rinse in a small volume of distilled water;

Rinse twice in 50µl 37% formaldehyde in 100ml 3% Na_2CO_3; soak with gentle agitation to develop;

Stop developing by agitating with added 5 ml 2.3M citric acid 10 minutes;

Soak several times in distilled water.

2.3.5 Photography of gels

Ethidium-, Coomassie blue- and silver-stained gels were photographed with Polaroid type 67 land film or with a Pentax 35mm SLR camera loaded with Ilford HP5 film. Both cameras were fitted with a Kodak Wratten filter no 23A.

Radioactively labelled proteins and DNA on gels were subjected to autoradiography, using Fuji RX100 Xray film. Laemmli gels containing ^{35}S methionine were treated with En^{3}hance (duPont) before drying.

2.3.6 Purification of DNA and proteins from gels

DNA fragments were eluted from agarose gels by spinning the excised bands through a glass wool plug in a perforated Eppendorf tube at 6000rpm for 10 minutes. DNA was isolated from polyacrylamide gels by the "crush and soak" method (Maxam and Gilbert, 1977).

Proteins were eluted from polyacrylamide gel slices using the method of Hager and Burgess (1980).

2.4 Plasmid DNA preparation

DNA was prepared by the method of Birnboim and Doly (1979). Small scale preparations were phenol extracted. Large scale preparations were spun in the presence of ethidium bromide through a CsCl gradient in a Beckman Ti70.1 rotor for 16 hours at 200,000g, with the ethidium subsequently butanol extracted.

Plasmids in cells were visualised on single colony gels using the method described by Morrell (1990).

2.5 DNA Sequencing

Plasmid sequencing was used. The two DNA strands were separated by treatment with 0.2M NaOH, 0.2mM EDTA and heated to 70°C before annealing the primer. A Sequenase kit (United States Biochemical Corporation) was used for the reactions, following the manufacturer's instructions; [α-^{35}S]thio-dATP was added for autoradiographic detection.

2.6 Restriction and modification enzymes

DNA restriction and modification enzymes were obtained from BRL, New England Biolabs or Boehringer and used according to the manufacturer's instructions. EcoRI at 4000 units μl^{-1} was a gift from S. Halford.

2.6.1 Exonuclease III deletions

These were made using the method of Henikoff (1984).

2.6.2 DNA nicking

Singly nicked DNA was obtained by treating the supercoiled substrate with 2μg ml^{-1} DNAseI in the presence of 300μg ml^{-1} ethidium bromide in 50mM Tris.Cl pH8.0, 10mM $MgCl_2$ for 30 minutes at 30°.

2.7 Southern hybridisation

The DNA on an agarose gel was transferred for approximately 24 hours to a Pall Biodyne nylon membrane in the presence of 0.4M NaOH. The membrane was washed briefly in 2xSSC (Maniatis *et al.*, 1982). Prehybridisation was carried out for 3 hours at 60°C using

2g dextran sulphate
1 ml 100xDenhardt's solution (Maniatis *et al.*, 1982)
1 ml 20xSSC
1 ml 10% SDS
17ml water
2mg denatured sheared salmon sperm DNA.

Hybridisation was carried out overnight in the same solution, with the addition of the ^{32}P labelled DNA probe, which was random primed using a kit and method supplied by Boehringer. The membrane was rinsed in 2xSSC, 0.1%SDS, then washed twice at 50°C in 0.1xSSC, 0.1% SDS before autoradiography.

2.8 Transformation with plasmid DNA

Mid-log phase cells were spun down and made competent for transformation by resuspension in ice-cold 50mM $CaCl_2$. 2mM RbCl was added for transformation of cells with the products of ligation reactions.

2.9 P1 transduction

This followed the method outlined by Colloms (1990). The P1 phage was originally obtained from M. Masters.

2.10 Transfection with M13 phage

This was performed as described by Colloms (1990).

2.11 Transposition assays

2.11.1 *in vivo* mate-out transposition assays

The donor strain was generally DS941 or LA547; Tn*7* or Tn*7*-1 was located on the chromosome or a plasmid, as specified in the text and, where appropriate, *tns* functions were supplied *in trans* under a controllable promoter. The conjugative plasmid pEN300 was used as the target DNA molecule. Inocula from overnight cultures were grown to mid-log phase, inducing the *tns* genes where possible, then were mixed with an excess of the log-phase recipient strain MR1 for plate mating for approximately 90 minutes. After the mating step, selective plates were used to identify the recipient strain and to distinguish between those isolates of the recipient strain containing pEN300 with and without a resident Tn*7*.

The transposition frequency was defined as:

$$\frac{\text{number of recipients with pEN300 plasmid carrying the transposon marker}}{\text{total number of recipients containing pEN300 plasmid}}$$

2.11.2 *in vitro* transposition assays

These were carried out as described in Section 3.9.

2.12 Oligonucleotides

Oligonucleotides were manufactured on a Applied Biosystems 391 PCR-Mate oligonucleotide synthesiser The oligonucleotides were released from the column matrix and deprotected by overnight incubation at 50°C in saturated aqueous ammonia, then ethanol precipitated. Oligonucleotides of the correct length were purified by denaturing gel electrophoresis and elution from the excised bands. The 5' ends were phosphorylated before use. The concentration was determined by UV spectrophotometry. Complementary oligonucleotides synthesised for cloning were annealed by heating to 85°C in TE with 200mM NaCl and then cooling slowly to room temperature.

2.13 Site-directed mutagenesis

This was carried out using a Promega Altered Sites mutagenesis kit and following the manufacturer's instructions. Oligonucleotides were synthesised and treated as above.

2.14 ^{35}S methionine labelling of proteins

The culture was grown in M9 minimal medium with casamino acids; the cells were spun down, rinsed in M9 salts and resuspended in Methionine Assay Medium (Difco) with induction of the appropriate gene. After 10 minutes' growth, 10μCi ^{35}S methionine was added and growth was continued to allow incorporation of the radioactivity. An excess of unlabelled methionine was then added to chase out unincorporated ^{35}S methionine from the cells before concentrating the cells by centrifugation and loading samples on to a Laemmli gel.

Where the T7 expression system was used, 250 μg ml^{-1} rifampicin could be added to the gene expression medium to ensure that the majority of the 35S methionine incorporation was into the protein produced from the T7 promoter.

2.15 Cell lysis methods for protein extraction

The standard method of lysing cells for protein extraction followed that of Kay *et al.* (1985) but with 1mM benzamidine instead of aprotinin, and with 3 freeze-thaw cycles after the incubation with lysozyme. Sonication of the sample on ice was sometimes used in addition to or instead of the lysozyme and freeze-thaw steps. Overall methods for the subsequent purification of TnsA and TnsB are described in Chapters 3 and 4 respectively; individual techniques are outlined below.

CTAB was used for the lysis of cells for β-galactosidase assays, and toluene for galactokinase assays.

2.16 Protein purification methods

Ammonium sulphate precipitation of proteins was performed at 4°C with gentle stirring, using enzyme grade ammonium sulphate.

Dialysis was carried out for at least 12 hours, generally at 4°, with two changes of buffer.

DEAE sephacel columns were used for the separation of proteins from DNA; these were made by pouring the matrix equilibrated with the appropriate buffer into 1ml syringes plugged with glass wool.

A Waters 650E Advanced Protein Purification System was used for HPLC. The elution of proteins was detected by monitoring OD_{280} and OD_{260}; samples were collected on a fraction collector at room temperature and transferred to ice as soon as possible. Pharmacia Mono S HR5/5 and Phenyl-Superose HR 5/5 columns were used with 250µl injected samples; the columns were treated according to the manufacturer's instructions.

Bradford assays (Bradford, 1976) were used for quantification of protein concentration.

Protein extracts were stored on ice during purification; long term storage was in 50% glycerol at -20°C; extracts used for binding reactions had a total protein concentration of approximately 15-20 mg ml^{-1}.

2.17 Gel binding assays

The binding reactions described in Chapter 4 generally contained approximately:

^{32}P-labelled DNA	0.5ng
carrier DNA (sheared salmon sperm)	200-800ng
protein extract (1M KCl) in 50% glycerol	10µg (or dilution, as specified in the text)

The final reaction buffer contained 10mM Tris, 1mM EDTA, 1mM DTT, 200mM KCl, 5% glycerol in a volume of 20µl. Reactions were incubated at 37°C for 10 minutes unless otherwise specified, and were then loaded immediately on to the pre-run 6% polyacrylamide gel; no additional loading buffer was used. Gels were run at 200V for 3 hours at 4°C with recirculation of the running buffer.

For the reactions in Chapter 5 using complete plasmids, the above method was modified as follows:

unlabelled specific DNA plasmid	1μg
carrier DNA (generally supercoiled pUC18)	2μg
protein extract in 50% glycerol	10μg

Variations on this, and additional steps in the reaction were as described in the text of Chapter 5.

2.17.1 ^{32}P End-labelling of DNA

DNA was cleaved to leave cohesive ends, which were filled in by the Klenow fragment of DNA polymerase I with the appropriate $\alpha^{32}P$ dNTP and other unlabelled nucleotides to leave a blunt end. The labelled DNA was generally purified by passage through a Nuctrap push-column (Stratagene).

2.18 Enzyme assays

2.18.1 β-galactosidase assays.

These were carried out according to Miller (1972) with the following variations:

(1) β-mercaptoethanol was not added to the Z buffer;

(2) ONPG was dissolved in Z buffer;

(3) 50μg CTAB dissolved in Z buffer was used instead of 1ml toluene to lyse the cells.

2.18.2 Galactokinase assays

These followed the method of McKenney *et al.* (1981) as modified by Slatter (1987), except that cells were lysed with toluene for 1 hour at 37°C, instead of using CTAB; and the scintillant used was Ecoscint.

The number of galactokinase units (nanomoles of galactose phosphorylated per minute per ml of cells at OD_{650} of 1.0) was calculated using the formula:

$$\frac{\text{cpm from test system - cpm from blank system x 5200}}{\text{mean cpm from unwashed filters - incubation time/minutes x } OD_{650}}$$

2.19 Raising antibodies

The TnsA protein was run on a 12.5% Laemmli gel, and Coomassie stained and destained in water. Wet polyacrylamide gel slices containing the denatured TnsA protein were ground up and mixed with Freund's incomplete adjuvant and injected into the peritoneal cavity of two rats by J D. Barry. Blood samples were collected before injection of the antigen, and should have been collected at intervals subsequently.

2.20 Computer programs

The following programs incorporated in the GCG package supplied by the University of Wisconsin were used:

SeqEd	GelAssemble	
Map	Mapsort	
Compare	DotPlot	
BestFit	Gap	PileUp
WordSearch	Segments	
FastA	TFastA	
Terminator	StemLoop	
Fold	CodonFrequency	
Isoelectric	Translate	
PeptideStructure	PlotStructure	

Database searching for similarities to the Tns proteins was also carried out by A. Coulson at the University of Edinburgh using the Prosrch program developed by J. Collins and A. Coulson.

CHAPTER 3

TnsA: OVEREXPRESSION AND PURIFICATION STRATEGIES

3.1 Introduction

Tn*7* encodes five genes which are involved in the transposition process. Of these, only *tnsB* had been assigned any function at the commencement of this work, although it later became clear that *tnsC* and *tnsD* also had well defined DNA binding activity, presumed to be an indication of their rôles in transposition. The function of *tnsA* was completely unknown, although it was clear that deletion of the gene (Rogers *et al.*, 1986) or disruption of it (Waddell and Craig, 1988) resulted in loss of transposition. It was therefore of great importance to the understanding of the overall transposition process to determine what part *tnsA* played in transposition.

tnsA is the gene proximal to the main (and possibly the only) Tn*7* promoter; there is an open reading frame, potentially encoding a 31kDa protein of 273 amino acids, preceded by a poor match to the consensus Shine-Dalgarno sequence (UGGACU) (Flores *et al.*, 1990). The amino acid sequence shows no obvious homology to that of any other protein in available databases, although there is a short sequence

LxRRYWQxxQxP

which is common to TnsA and TnsB and not found elsewhere; it is predicted by the GCG program PLOTSTRUCTURE to occur on the external surface of the protein at the carboxy end of an α-helical secondary structure in TnsA (see Figure 3.1), and is also immediately downstream of a putative exposed helix-turn-helix motif in TnsB (Flores *et al.*, 1990). The significance, if any, of this is unknown, but it is possible that this is a region which in both TnsA and TnsB is involved in interaction with a third party, perhaps one of the other Tns proteins.

Gay *et al.* (1986) have identified a region in TnsA which bears similarity to the amino acid sequences of the transposases of Tn*3* and IS*1*:

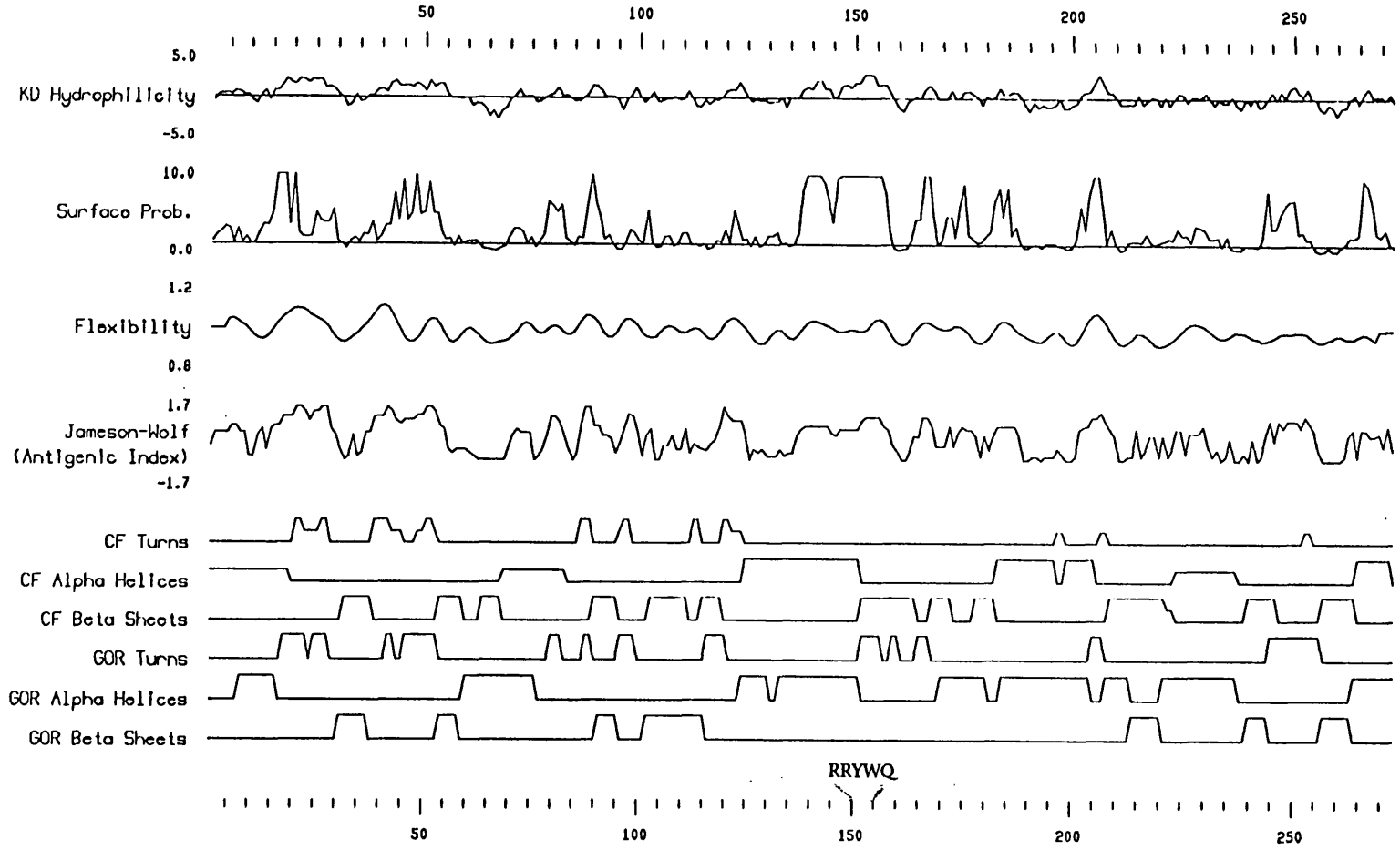


Figure 3.1 PLOTSTRUCTURE of TnsA. KD = Kyte and Doolittle; CF = Chou and Fasman; GOR = Garnier, Osguthorpe and Robson. The RRYWQ sequence common to TnsA and TnsB is shown at the bottom of the Figure.

```
Tn3 TnpA     148   ATGWLMQHEI ILPGATTLTR LISEVREKAT LRLWNKLALI
                         :  .   |      |     :|  .   |     |.: :.  .|:
Tn7 TnsA      47   IYSHKTGRVH HLLSDLELAV FLSLEWESSV LDIREQFPLL
                    .    ||:.  |||  :: : ::::. |       .:: .. ::
IS1 InsAB'   142   RTMATLGRLM SLLSPFDVVI WMTDGWPLYE SRLKGKLHVI

                   PSAEQRSQLE MLLGPTDCSR LSLLESLKKG
                   ||  :   ::   :        :  : |.
                   PSDTEQIAID SGIKHPVIRG VDQVMSTDFL
                        |       :      :       |. |
                   SKRYTQRIER HNLNLRQHLA RLGRKSLSFS
```

According to Gay *et al.*, the match between TnsA and the Tn*3* and IS*1* transposases shown above would occur by chance in 1:300 000 and 1:200 000 sequences respectively, suggesting that these conserved amino acid sequences may play an important rôle in transposition. Dissection of the structure-function relationship of the Tn*3* and IS*1* transposases is still at an early stage, and no precise activity has yet been assigned to the sequences identified above; the Tn*3* fragment is at the amino-terminal end of the protein and is contained within a region which has been associated with specific binding to the terminal inverted repeats of Tn*3* (Maekawa *et al.*, 1993), whereas the IS*1* fragment is towards the carboxy-terminus of a much smaller protein, in the InsB' region, which is probably involved in the catalysis of strand cleavage and transfer during transposition, as discussed further in Section 1.3.1.

Flores *et al.* (1990) have identified a putative helix-turn-helix DNA-binding domain in TnsA but, according to the algorithm of Dodd and Egan (1990), the score calculated is much lower than would be considered significant; this region is not identified as α-helical by the GCG PLOTSTRUCTURE program (Figure 3.1). The pI of TnsA is estimated, using the GCG ISOELECTRIC program, to be 5.96. This would make the protein considerably more acidic than is normally found in DNA-binding proteins; however acidic DNA-binding proteins are known: the *trp* repressor, for instance, has an experimentally determined pI of 5.9 (Joachimiak *et al.*, 1983), and binds through a helix-turn-helix motif (Kelley and Yanofsky, 1985). Protein binding to DNA is possible through mechanisms other than the helix-turn-helix; protein crystallography has shown that the Met repressor, encoded by *metJ*, for instance binds DNA through interaction with β strands (Somers and Phillips, 1992; He *et al.*, 1992), and that the yeast transcriptional activator GCN4 uses the same leucine zipper structure for both protein dimerisation and DNA binding (Ellenberger *et al.*, 1992). Gel

binding assays have been carried out using the ^{32}P-labelled Tn*7* right end and left end, and a crude cell extract of TnsA-producing cells. There is no evidence for any binding of TnsA to the Tn*7* ends (see also Chapter 4), but it is conceivable that the TnsA produced was unable to participate in binding due to insolubility or for other reasons.

The TnsA amino acid sequence was examined for features identified in other proteins of known function, for example the helicase D-E-A-D box (Linder *et al.*, 1989) and the D-35-E motif common to retroviral integrases and many prokaryotic transposases (Section 1.6; Fayet *et al.*, 1990), but no motifs were identified.

It is not known at present in what multimerisation state TnsA exists in solution or in Tn*7* transpososomes. Leucine-zipper domains, believed to be the site of protein-protein interactions, have been identified in some prokaryotic proteins, for example the Lac repressor (Alberti *et al.*, 1991), but none such is evident in TnsA, although it seems likely that protein-protein contact with one or more of the other Tn*7*-encoded transposition proteins is important in TnsA's rôle in transposition.

When this work was initiated, TnsA had not been visualised on Coomassie- or silver-stained Laemmli gels; it had, however, been detected by Rogers (1986) in ^{35}S-methionine labelled minicells. The low protein level could theoretically have been due to several factors: poor transcription due to a weak promoter or an abundant repressor protein; transcriptional termination within the open reading frame; unstable RNA; the existence of antisense RNA; poor translational initiation because of the inefficient ribosome-binding site; autorepression of transcription or translation by the TnsA protein; atypical codon usage or proteolysis. Each of these possibilities was examined (described in Section 3.2), as was the hypothesis that it was not the protein product of *tnsA* which was required for transposition (Section 3.3).

The aims of this section of the project were to improve the expression of *tnsA*, and thus to purify and characterise it. As will be seen, the first aim (expression) was successfully achieved, the second (purification) only partially so and, because of the problems with purification, the final step (characterisation of the protein's structure and function) was not possible in the time available.

3.2 Preliminary investigation of the *tnsA* sequence

Before attempting practical work on expression of the *tnsA* gene, the sequence was studied, both by eye and by utilising the GCG computer package, for indications of reasons for the low protein level observed.

3.2.1 Promoter strength

This was examined by Ekaterinaki (Rogers *et al.*, 1986) who determined using galactokinase assays that the Tn*7* promoter upstream of *tnsA* had an *in vivo* strength approximately 1.1 that of p_{lac}.

The sequence of the presumed promoter, as shown on Figure 3.2, is

TTGTAA..17bp..TATGCT.

This compares favourably with the consensus *E. coli* σ^{70} promoter derived by Hawley and McClure (1983):

TTGACA..17±1..TATAAT,

where underlined bases are found in more than 75% of such promoters examined.

Thus it seems that the low TnsA level observed is not due to very low promoter strength.

3.2.2 Possible transcriptional termination

The Tn*7* chromosomal *att* site is immediately downstream of the *glmS* gene; the *glmS* transcript overlaps the Tn*7* promoter, with ρ-dependent termination only of *glmS* transcription 17bp downstream of the Tn*7* transcriptional start (Gay *et al.*, 1986); initiation of transcription of *tnsA* may be reduced by occlusion of the Tn*7* promoter by overlapping *glmS* transcripts.

Ekaterinaki (1987) used galactokinase terminator probe vectors to search for transcriptional terminators in the *tnsA* region; none were located.

A computer search was carried out using the GCG TERMINATOR program, and a possible weak ρ-independent terminator was located in the centre of the *tnsA* coding region, between 590 and 620 bp from the Tn*7* right end; it is not known at present whether any transcription does terminate here.

However, where both *tnsA* and *tnsB* are expressed from an intact Tn*7* promoter, the TnsB level observed is consistently markedly higher than that of TnsA. There is either no promoter (Waddell and Craig, 1988) or only a very weak one (Ekaterinaki, 1987) immediately upstream of *tnsB*,

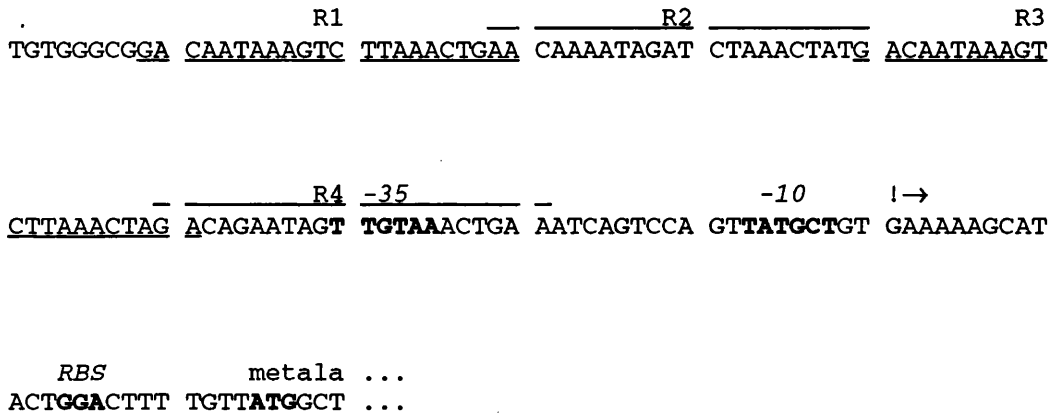


Figure 3.2 The nucleotide sequence at the right end of Tn*7*, showing the four 22bp motifs R1, R2, R3 and R4 alternately underlined and overlined, the presumed -35 and -10 boxes of the *tns* promoter, the transcriptional initiation site (!→), the likely ribosome binding site (RBS) and the translational start of *tnsA*.

suggesting that transcripts initiating at the Tn7 promoter upstream of *tnsA* continue in substantial numbers through to *tnsB*.

3.2.3 Repression of transcription

As can be seen on Figure 3.2, the innermost of the four 22bp motifs found at the Tn7 RE overlaps the *tns* promoter sequence. There is well documented evidence (Ekaterinaki, 1987; Arciszewska and Craig, 1991; this work) that TnsB binds to the -35 box of the presumed promoter and represses transcription. This repression cannot, however, account for the differential levels of TnsA and TnsB observed if, as is generally believed, both genes are expressed from the same *tns* promoter.

3.2.4 RNA stability

No published studies have been carried out with Tn7 RNA. However, as with the possibility of transcriptional termination, observed TnsB levels suggest that RNA stability is not a major factor in determining the TnsA level, unless the mRNA is being specifically cleaved upstream of *tnsB*.

3.2.5 Antisense RNA

Antisense RNA has been implicated in the control of transposition in several prokaryotic transposable elements. In the case of Tn*10*, for instance, there are divergent promoters with overlapping transcripts (Lee and Schmidt, 1985; Case *et al.*, 1990; Ma and Simon, 1990). One transcript encodes the transposase; it forms a duplex with the antisense transcript, thus preventing translation of the transposase gene.

It is possible that such a system also exists in Tn7. The galactokinase assays performed (Section 4.7) suggested that there might be either a weak antisense promoter, or a strong antisense promoter followed within 400bp by a transcriptional terminator. Examination of the Tn7 sequence revealed the sequence:

5'-TTGCAA..19bp..TAATAA-3'

located with the 5' base 417 base pairs from the right end of the transposon, and reading in the opposite direction to that of the *tns* genes, followed by a purine rich sequence. This might act as a σ^{70} promoter, although the spacing between the -35 and -10 boxes is greater than ideal, and would presumably necessitate distortion of the double helix for optimal binding of RNA polymerase. No transcriptional terminator sequence was identified in the antisense sequence cloned into the *galK* vector p253 (see Section 4.7).

The possible rôle of antisense RNA in *tnsA* expression was investigated by measuring transposition in the presence of artificially generated RNA antisense to *tnsA*. An *in vivo* mate out transposition assay was carried out using Tn*7*-1 and *tns* functions *in trans*, as described in Chapter 2. The donor cells harboured the conjugative plasmid pEN300 which contains a cloned chromosomal *att Tn7* site as a target sequence for transposition. The cells were then transformed with pAG101 as a source of *tnsB*, *C* and *D*, and one of two different plasmids encoding *tnsA*: in pAG081 the *tnsA* gene (without its upstream *tns* promoter) is cloned downstream of the pUC *lac* promoter, whereas in the alternative pAG084 the cloned insert is reversed in orientation, so that a transcript antisense to *tnsA* is produced. The transformants were grown overnight. The pEN300 was then mated out into a second strain, and the proportion of exconjugants containing a copy of Tn*7*-1 (as assayed by chloramphenicol resistance) was determined. The measured transposition rates, over the approximately 25 cell generations when the *tns* genes and pEN300 were co-resident in the cells, were:

pAG101 only	$<1.6 \times 10^{-8}$
pAG101 + pAG081	5.0×10^{-3}
pAG101 + pAG084	5.0×10^{-4}
pAG101 + pAG084 (IPTG induction)	5.0×10^{-4}

These results suggest firstly that transposition can still occur in the absence of the correctly oriented *lac* promoter, presumably because pUC plasmids also harbour a weak promoter antisense to those for the *lac* and *bla* genes, and secondly that induction of the antisense promoter by addition of IPTG to the growth medium does not block transposition. These conditions are not, of course, those found in the wildtype Tn*7* transposon, but they nevertheless suggest that regulation of *tnsA* expression by antisense RNA may not be significant.

The Tn*7* sequence was also examined for σ^{54} and σ^{28} promoter sequences in the antisense direction, but none was identified.

3.2.6 Ribosome binding site strength

The 13 nucleotides at the 3' end of 16s rRNA are

5'-GAUCACCUCCUUA$_{OH}$-3'

and these are exposed and thus available for annealing to the mRNA. The wild-type *tnsA* ribosome binding site has the sequence:

ACUGGACUU

which is a poor match to the Shine-Dalgarno consensus. The *tnsB* ribosome binding site, by comparison, is

GGAGGAGUU.

This suggests that, taking only these sequences into consideration, translational initiation is likely to be considerably more efficient for *tnsB* than for *tnsA*.

A poor ribosome binding site is frequently found upstream of bacterial transposases, for example those of IS*10* and Tn*3*, so inefficient translation is a common phenomenon with transposable elements.

The above discussion may be too simplistic: the translational initiation region is, however, often considered to extend both upstream (Schauder and McCarthy, 1989) and downstream (Ringquist *et al.*, 1992) of the ribosome binding site, with a much less defined consensus sequence in the extended region; optimal extended ribosome binding site sequences are therefore correspondingly more difficult to recognise.

3.2.7 Autorepression

Repression of transcription by TnsA might, by analogy with other known systems, be expected to involve direct binding of TnsA to the promoter. When this work was initiated, it was thought that TnsA might bind to the Tn*7* right end; see Chapter 4. Binding was examined, using gel retardation as diagnostic, using crude extracts from TnsA-overproducing strains (Figure 3.3). No evidence of TnsA binding to the standard Tn*7* right end fragment (the EcoRI-Hin*d*III fragment of pNE200) was obtained in the absence of TnsB; and the retardation pattern observed with TnsB-containing extracts was the same as that observed with TnsA and TnsB containing extracts (Figure 4.12). Attempts at detecting TnsA-dependent binding under different pH conditions were also made by S. Bell (unpublished), and again no retardation was observed. This suggests that TnsA does not bind directly to the DNA around the promoter.

An alternative, but unlikely, possibility is that TnsA is binding not to DNA but to part of the RNA polymerase complex, and thus in some way

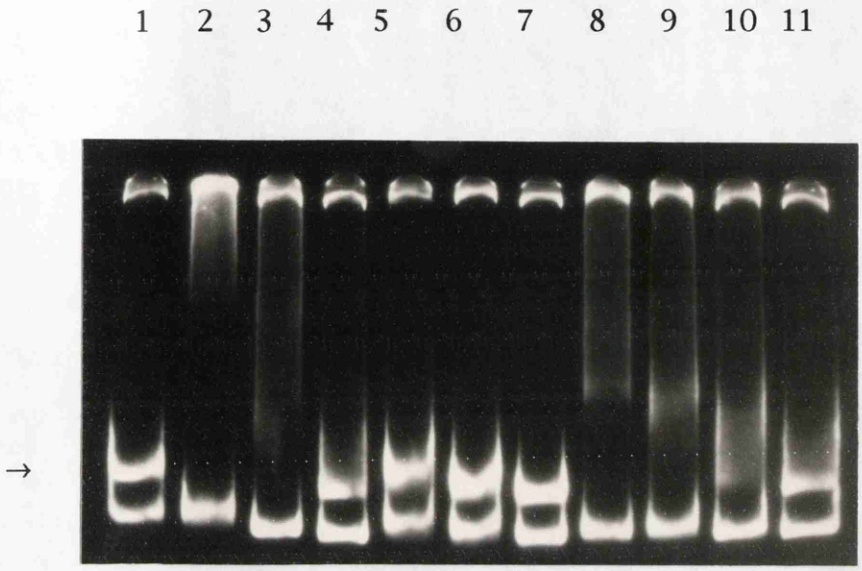


Figure 3.3 Gel retardation assay to show the specific binding activity associated with crude cell extracts of DS941 expressing *tns* genes encoded by the plasmids pMR204, pAG204, pAG205 and pMR78 (vector with no insert); these inserts are shown diagrammatically on Figure 4.4. The ^{32}P labelled DNA is pNE200 (see Figure 4.11) cut with Hin*d*III, EcoRI and NdeI; the DNA fragment containing the Tn*7* sequence is the second one up, as indicated with an arrow. The reactions were carried out, and the gel poured and run in TE, as described in Chapter 2. The protein extracts used were as listed below.

1	pMR78 extract	2^{-3} dilution of stock extract
2	pMR204 extract	2^{-3}
3	pMR204 extract	2^{-4}
4	pMR204 extract	2^{-5}
5	pAG204 extract	2^{-3}
6	pAG204 extract	2^{-4}
7	pAG204 extract	2^{-5}
8	pAG205 extract	2^{-3}
9	pAG205 extract	2^{-4}
10	pAG205 extract	2^{-5}
11	pAG205 extract	2^{-6}

impeding the action of the polymerase to transcribe from the *tns* promoter. In the absence of antibodies to TnsA, it is difficult to determine whether any such protein-protein complex is formed.

A similar arrangement of genes to that of Tn*7* is found in bacteriophage Mu (see Figure 1.2), where the *A* and *B* genes, which are involved in transposition, are adjacent and transcribed from the same promoter, upstream from *A*; the measured molar ratio of A to B proteins is approximately 1:20 (Parsons and Harshey, 1988). It has been suggested (Parsons and Harshey, 1988) that the A protein of bacteriophage Mu represses its own translation by binding specifically to its mRNA. It is possible that TnsA could likewise block translation by binding either to the *tnsA*-encoding mRNA or by binding to part of the translation machinery. Neither of these hypotheses relating to TnsA has been tested.

3.2.8 Codon usage

The codon usage of *tnsA*, *tnsB*, *tnsC*, *tnsD* and *tnsE* was examined (Table 3.1) and was compared with that of *E. coli* proteins of differing cellular abundances (de Boer and Kastelein, 1986). It has been observed (de Boer and Kastelein, 1986) that genes encoding *E. coli* proteins with a high expression level display a much greater codon bias than do those encoding rare proteins.

It can be seen from Table 3.1 that the 27 leucine codons found in *tnsA* include 6 occurrences of the rare codon CUA, and only 1 of CUG, which is normally by far the most commonly used codon. (This is a phenomenon also found in *tnsD*, which is also apparently very poorly expressed.) Again, the proline-encoding codon most common in *tnsA*, CCU, is the one which is generally rare, and the most common *E. coli* proline codon, CCG, is not found at all in *tnsA*. Further, AAC, which is by far the more common asparagine codon in *E. coli* overall is greatly under-represented in Tn*7*, with 8 of the 10 asparagine residues encoded by AAU (a phenomenon also found in *tnsB*, *tnsC* and *tnsE*). This suggests that part, at least, of the reason why TnsA is in low abundance could be that translational elongation is slowed down, possibly by saturation of use of certain tRNAs, although this mechanism has been disputed (Sørensen *et al.*, 1989).

If rare codon usage had evolved in this particular instance as a deliberate mechanism for differential gene expression, it might be expected that attenuation of *tnsA* translation would occur near the N-terminal amino acids of the protein, thus minimising the number of amino

Amino acid	Codon	*tnsA*	*tnsB*	*tnsC*	*tnsD*	*tnsE*
leu	CUG	4	12	10	16	4
	CUU	15	9	11	17	13
	CUC	7	4	4	0	2
	UUG	19	26	25	13	28
	UUA	33	37	36	30	37
	CUA	22	12	14	23	15
Total	number	27	59	72	69	46
arg	CGU	53	28	50	14	22
	CGC	20	21	15	12	6
	CGA	13	16	10	19	25
	AGA	0	19	13	37	22
	CGG	0	7	13	9	14
	AGG	13	9	0	9	11
Total	number	15	57	40	43	36
gly	GGU	50	38	38	33	33
	GGC	17	38	22	17	15
	GGG	17	12	22	28	37
	GGA	17	12	19	22	15
Total	number	12	34	32	18	27
ile	AUC	9	12	22	5	15
	AUU	73	71	56	60	50
	AUA	18	17	22	35	35
Total	number	22	41	36	20	26
tyr	UAC	0	32	18	41	20
	UAU	100	68	82	59	80
Total	number	7	28	17	22	10
his	CAC	14	14	8	30	25
	CAU	86	86	92	70	75
Total	number	7	14	12	20	16
ser	UCU	40	25	22	27	11
	UCC	10	2	10	6	6
	AGC	15	23	20	24	20
	UCG	0	6	22	9	17
	UCA	15	17	13	6	15
	AGU	20	27	13	27	31
Total	number	20	48	40	33	54
val	GUU	30	29	21	42	43
	GUA	25	22	29	13	16
	GUG	30	35	38	35	36
	GUC	15	14	12	10	5
Total	number	20	51	34	31	44

pro		CCG	0	16	0	9	11
		CCA	20	26	35	27	44
		CCU	60	42	39	45	33
		CCC	20	16	26	18	11
	Total	number	10	38	23	22	18
thr		ACU	22	24	8	32	23
		ACC	33	16	31	21	15
		ACG	11	26	46	26	8
		ACA	33	34	15	21	54
	Total	number	9	38	26	19	26
ala		GCU	22	37	39	41	41
		GCA	50	17	31	27	28
		GCG	6	24	17	20	16
		GCC	22	22	14	11	16
	Total	number	18	46	36	44	33
gln		CAG	37	51	30	45	48
		CAA	63	49	70	55	52
	Total	number	16	39	37	38	25
asn		AAC	20	17	13	64	33
		AAU	80	83	87	36	67
	Total	number	10	18	15	11	24
lys		AAA	61	59	62	48	54
		AAG	39	41	38	52	46
	Total	number	18	44	29	27	28
asp		GAC	27	22	17	27	22
		GAU	73	78	83	73	78
	Total	number	15	46	29	26	32
glu		GAA	46	47	39	31	49
		GAG	54	53	61	69	51
	Total	number	24	51	38	29	45
cys		UGC	25	25	50	60	20
		UGU	75	75	50	40	80
	Total	number	4	4	4	5	5
phe		UUC	36	17	13	29	11
		UUU	64	83	87	71	89
	Total	number	11	29	15	14	28
trp		UGG					
	Total	number	5	11	6	13	10
met		AUG					
	Total	number	3	8	14	4	6

Table 3.1. Percentage use of each codon for the amino acids in *tnsA*, *tnsB*, *tnsC*, *tnsD* and *tnsE*. For each amino acid, codons are listed in order of decreasing frequency in *E. coli* according to de Boer and Kastelein (1986).

Amino acid	Codon	*int*	*dhfr*	*sat*	*aadA*
leu	CUG	26	25	7	23
	CUU	3	0	14	19
	CUC	8	8	29	10
	UUG	24	17	14	32
	UUA	26	17	7	3
	CUA	13	33	29	13
Total	number	38	12	14	31
arg	CGU	24	25	14	7
	CGC	28	25	0	53
	CGA	12	25	29	13
	AGA	12	0	43	7
	CGG	16	0	14	7
	AGG	8	25	0	13
Total	number	25	4	7	15
gly	GGU	38	45	0	15
	GGC	38	9	70	62
	GGG	25	18	10	0
	GGA	0	27	20	23
Total	number	16	11	10	13
ile	AUC	37	38	50	53
	AUU	47	25	40	41
	AUA	16	38	10	6
Total	number	19	16	10	17
tyr	UAC	33	50	71	50
	UAU	67	50	29	50
Total	number	9	6	7	6
his	CAC	31	0	83	50
	CAU	69	100	17	50
Total	number	16	2	6	6
ser	UCU	26	21	29	0
	UCC	16	0	7	25
	AGC	37	14	14	25
	UCG	11	7	29	33
	UCA	0	36	7	8
	AGU	11	21	14	8
Total	number	19	14	14	12
val	GUU	39	42	17	23
	GUA	17	33	8	35
	GUG	22	8	50	23
	GUC	22	17	25	19
Total	number	18	12	12	26
pro	CCG	19	13	0	43
	CCA	31	38	0	29
	CCU	31	25	75	14
	CCC	19	25	25	14
Total	number	16	8	4	14

thr		ACU	35	11	20	13
		ACC	12	33	10	33
		ACG	47	11	30	40
		ACA	6	44	40	13
	Total	number	17	9	10	15
ala		GCU	22	43	15	35
		GCA	39	14	31	23
		GCG	26	14	31	15
		GCC	13	29	23	27
	Total	number	23	7	13	26
gln		CAG	33	17	43	67
		CAA	67	83	57	33
	Total	number	18	6	7	9
asn		AAC	67	50	71	40
		AAU	33	50	29	60
	Total	number	15	10	7	5
lys		AAA	63	55	57	50
		AAG	37	45	43	50
	Total	number	19	11	7	8
asp		GAC	29	33	43	39
		GAU	71	67	57	61
	Total	number	14	9	14	18
glu		GAA	71	57	67	60
		GAG	29	43	33	40
	Total	number	7	7	12	25
cys		UGC	80	0	67	0
		UGU	20	0	33	0
	Total	number	5	0	3	0
phe		UUC	0	13	63	20
		UUU	100	88	38	80
	Total	number	14	8	8	5
trp		UGG				
	Total	number	2	1	4	6
met		AUG				
	Total	number	6	2	1	1
stop	Total	number	1	0	0	0

Table 3.2. Percentage use of each codon for the amino acids in the full-length putative integrase and the three antibiotic resistance genes of Tn*7*. For each amino acid, codons are listed in order of decreasing frequency in *E. coli* according to de Boer and Kastelein (1986).

acids sequestered in abortive peptides. The rare codons of leucine, proline and asparagine in TnsA are in fact distributed throughout the coding sequence for the protein, suggesting that translational attenuation probably does not occur to a significant extent.

Alternatively, it is possible that this pattern of codon usage indicates an origin for Tn*7* in a bacterium other than *E. coli*; the codon distribution among the five *tns* genes is similar, suggesting a common origin.

It is however interesting to compare the codon usage of the *tns* genes with that of the putative integrase and the three antibiotic resistance genes found in the left half of Tn*7*, as shown in Table 3.2. These latter genes generally conform better to the *E. coli* codon distribution of de Boer and Kastelein (1986) than do those of the *tns* genes, suggesting that the two sets of genes may possibly have different ancestry, with the "integron"-associated genes coming from *E. coli* or a closely related species, and the *tns* genes derived from a more distantly related bacterium. The general distribution of codons has not unfortunately been studied in a wide enough range of bacteria for the hypothesis of a bacterial source other than *E. coli* to be examined critically. Tn*7* is found in a wide range of Gram-negative bacteria, but its paths of transmission between species are not known.

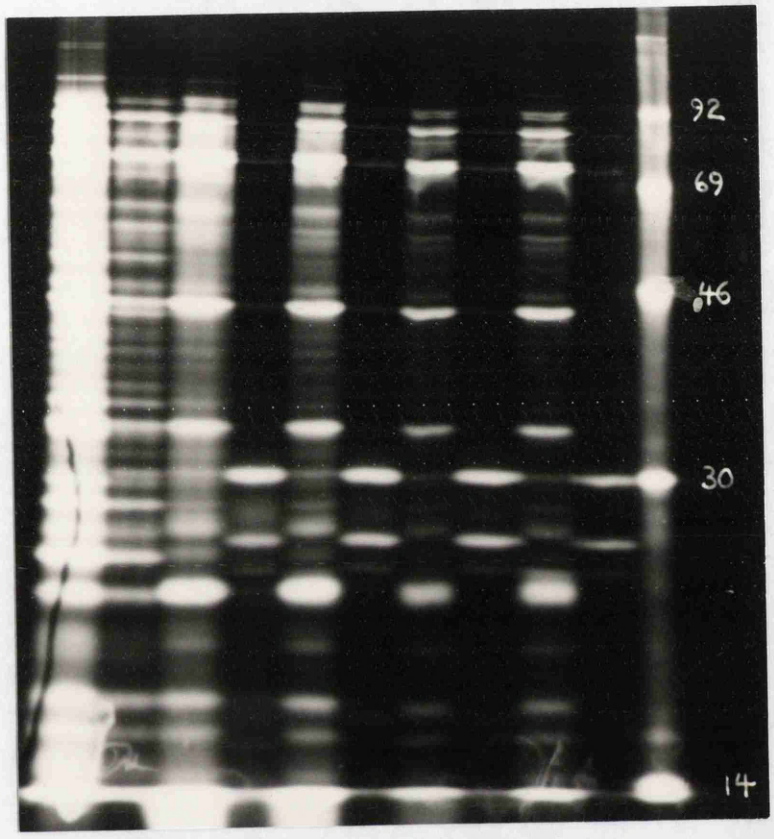
3.2.9 Possible proteolysis

It is known that several transposases, for example that of IS*903* (Derbyshire *et al.*, 1990), are susceptible to proteolysis and hence have a short half-life, and it was considered possible that the same might be true in the case of TnsA. This was examined, using the plasmid pAG411 with *tnsA* under the control of the T7 promoter, in a pulse-chase experiment (see Fig. 3.4). The protein appears to be stable for at least the 60 minutes during which samples were collected from the culture, though its long-term stability was not ascertained by this method.

It is believed that both the amino-terminal and carboxy-terminal residues of a protein play a part in determining whether that protein is the subject of proteolysis.

Varshavsky (Tobias *et al.*, 1991; Varshavsky, 1992) has put forward the hypothesis that the *in vivo* half-life of an *E. coli* protein is related to the identity of its N-terminal residue, with proteins possessing arginine or lysine as the N-terminal amino acid being susceptible to cleavage by the Clp protease. *E. coli* proteins are usually translated with methionine as the

1 2 3 4 5 6 7 8 9 10 11



TnsA—

Figure 3.4 Autoradiograph of 12.5% polyacrylamide Laemmli gel showing expression in whole cells of ^{35}S methionine-labelled *tnsA* from the plasmid pAG411, the construction of which is described in Section 3.7.1; pT7-7 is the expression vector into which the *tnsA* sequence was cloned. The T7 RNA polymerase in the plasmid pGP1-2 is under the control of Lambda cI_{857}, and is therefore heat-inducible. A pulse-chase experiment was carried out as described in Chapter 2; the chase times with unlabelled methionine are shown below.

1	pT7-7	no heat shock; no rifampicin
2	pAG411	no heat shock; no rifampicin
3	pT7-7	heat shock; rifampicin; no chase
4	pAG411	heat shock; rifampicin; no chase
5	pT7-7	heat shock; rifampicin; 5 minutes chase
6	pAG411	heat shock; rifampicin; 5 minutes chase
7	pT7-7	heat shock; rifampicin; 15 minutes chase
8	pAG411	heat shock; rifampicin; 15 minutes chase
9	pT7-7	heat shock; rifampicin; 60 minutes chase
10	pAG411	heat shock; rifampicin; 60 minutes chase
11	molecular weight markers (sizes in kDa)	

first amino acid, but this residue is frequently removed during or immediately after translation. Hirel *et al.* (1989) have proposed that in *E. coli* the retention of this initial methionine is correlated with the length of the side chain of the next amino acid: if the second amino acid is glycine, alanine, proline, serine, threonine, valine or cysteine, then the methionine is usually removed; otherwise it is usually retained. According to these rules, an *E. coli* protein is very unlikely to have arginine or lysine as the amino-terminal residue unless it has already been subject to proteolysis by some other mechanism. The first two amino acids of the Tns proteins are believed to be:

TnsA:	Met	Ala
TnsB:	Met	Tyr
TnsC:	Met	Ser
TnsD:	Met	Arg
TnsE:	Met	Val

so TnsD is the only one of the five proteins which could possibly be subject to Clp-dependent proteolysis according to the above rules; in particular, there is no reason to believe that Clp might be acting on TnsA.

It has been suggested (Parsell *et al.*, 1990) that, as well as *lon* dependent thermal stability and the amino-terminal dependent stability discussed above, the degree to which proteins are degraded in *E. coli* is to some extent determined by the carboxy-terminal amino acids: nonpolar residues at the 5 carboxy-terminal residue positions destabilise a protein, whereas charged residues, particularly at the terminal position, have the opposite effect. The five carboxy-terminal residues of the Tn*7* transposition proteins are:

TnsA	**RYVAN**
TnsB	**EKDES**
TnsC	**FKQAG**
TnsD	**VYGEE**
TnsE	**EQFTH**

where stabilising residues according to Parsell *et al.* are shown in bold type. It can be seen that, while four of the last five amino acids of TnsB are charged, only two of the carboxy-terminal residues of TnsA are likely to

Wild-type Tn*7* sequence

TTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCATACTGGA CTTTTGTT ATG

In pMR78 (pMR204/pAG204); *tac* promoter

TTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGAATTCCCGGA CTTTTGTT ATG

In pSELECT (before mutagenesis - pAG401); T7 promoter

TGTAATACGACTCACTATAGGGCGAATTCCCGGA CTTTTGTT ATG

In pSELECT after mutagenesis with start codon removed (pAG402); T7 promoter

TGTAATACGACTCACTATAGGGCGAATTCCCGGA CTTTTGTT AAC

In pSELECT after mutagenesis with start codon retained (pAG404); T7 promoter

TGTAATACGACTCACTATAGGGCGAATTCCCGGA CTTTTCAT ATG

In pT7-7 with start codon retained (pAG411/pAG415); T7 promoter

ATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAG ATATACAT ATG

In pT7-7 with start codon removed (pAG416); T7 promoter

ATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAG ATATACATA AAC

Figure 3.5 Sequences upstream of the *tnsA* coding region.
Underlined bases indicate the position of the promoter, ribosome binding site(s) and translation start codon. The transcriptional start is indicated in bold type.

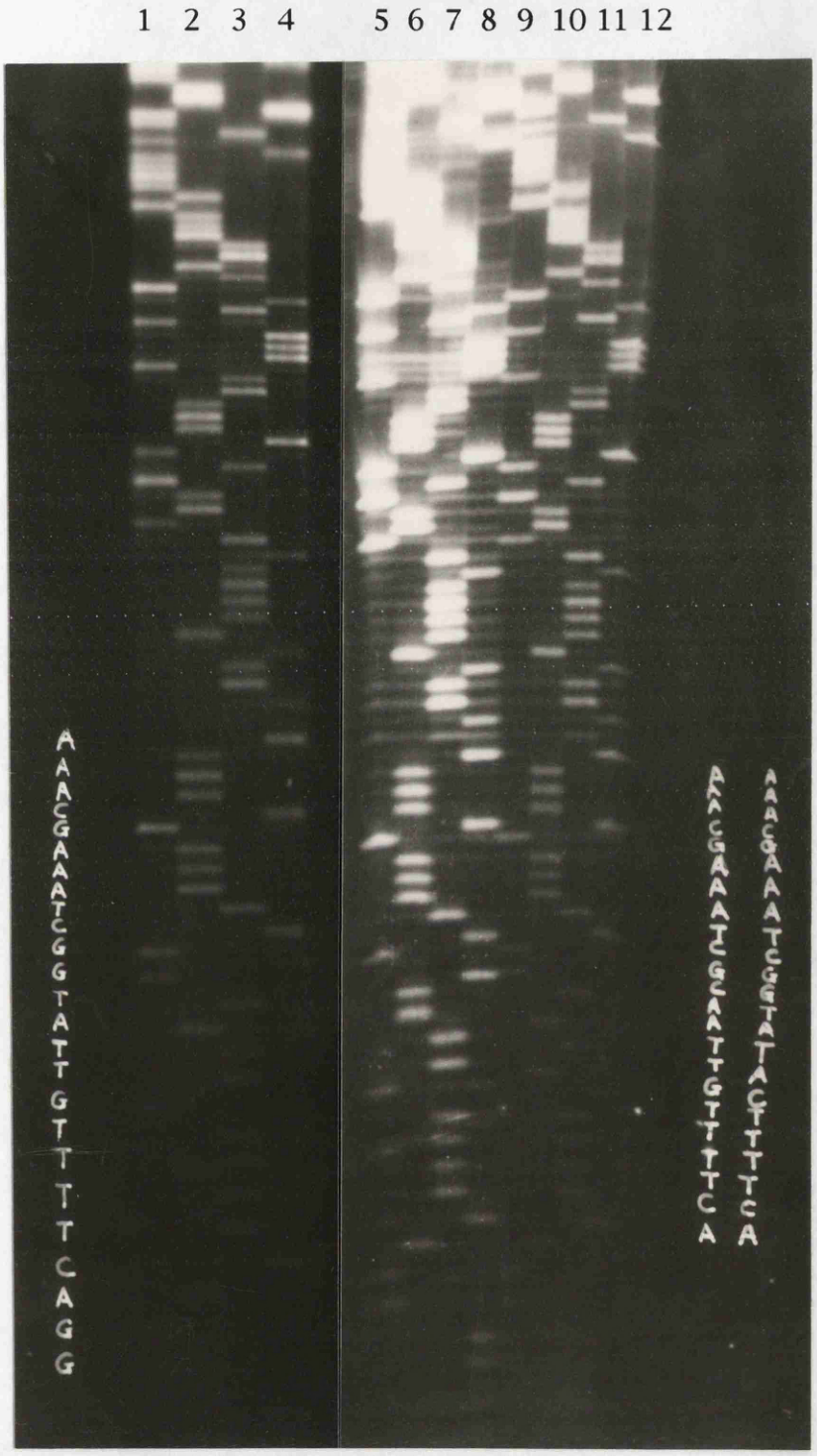


Figure 3.6 Dideoxynucleotide sequencing gel confirming the mutagenised sequences at the 5' end of *tnsA* in the plasmids pAG402 and pAG404. (The sequence is adjacent to that of the primer and hence the bands are weak.)

1	pAG401	G	5	pAG402	G	9	pAG404	G
2	pAG401	A	6	pAG402	A	10	pAG404	A
3	pAG401	T	7	pAG402	T	11	pAG404	T
4	pAG401	C	8	pAG402	C	12	pAG404	C

confer a substantial stabilising effect. The differential observed concentration of these two proteins could therefore be partly explained by their carboxy-terminal sequence.

3.3 Requirement for the protein product

The formal possibility existed that it was not the protein product of *tnsA* which was required for transposition, since the *in vitro* transposition system developed by Bainton *et al.* (1991) used extracts from *tnsA*-expressing cells, rather than the pure protein. To test this hypothesis, two plasmids were constructed, in each of which the start codon of *tnsA* translation had been mutated to AAC, for which there is no known record of its ability to initiate translation.

The initial site-directed mutagenesis was carried out on pAG401, in which the EcoRI / Hin*d*III *tnsA*-encoding fragment of pMR204 was cloned into pSELECT, and the procedure outlined in Section 2.13 and Section 3.7.1 was used to remove the start codon, and simultaneously to introduce a diagnostic HpaI site. This plasmid was pAG402.

The mutated gene was then isolated from pAG402 as a HpaI / Hin*d*III fragment and inserted into pT7-7 cut with NdeI / Hin*d*III and with the NdeI site filled in using the Klenow fragment of DNA polymerase I, to give the plasmid pAG416. The sequences of the 5' junctions of *tnsA* with the vectors are shown in Figure 3.5. They were confirmed using direct plasmid sequencing, as shown in Figure 3.6.

The plasmids pAG402 and pAG416, and their start codon retaining analogues pAG404 and pAG411, were used in standard plate-mating transposition assays, with pAG101 to supply *tnsB-E* and DS941::Tn*7*-1 as the donor strain. T7 RNA polymerase was not supplied as it had previously been shown (Section 3.7.2 and S. Bell, pers. comm.) that the measured transposition rate was not different in the absence, compared with the presence, of T7 RNA polymerase. The average transposition rate with each *tnsA* supplier was as follows:

pSELECT (vector)	$<6.5 \times 10^{-8}$
pAG402	1.9×10^{-6}
pAG404	6.0×10^{-3}
pT7-7 (vector)	$<8.9 \times 10^{-9}$
pAG411	3.1×10^{-2}
pAG416	8.5×10^{-3}

It can be seen from the above results that Tn*7*-1 transposition rates in the presence of pAG411 and pAG416 were very similar. This might suggest that the translational start codon is not necessary for transposition. However, closer study of the pAG416 sequence on Figure 3.5 reveals that there is an in-frame AUA upstream of the AAC codon. There is precedent (Singer *et al.*, 1981) for the use of AUA as an *E. coli* start codon. Comparison of the measured transposition rates in the presence of pAG402 and pAG404 shows that there is a clear difference between the properties of these two plasmid-encoded genes, manifested in their ability to participate in transposition. Unlike pAG416, pAG402 has no recognisable start codon between the promoter and the site of the wildtype start codon; there is a GUG codon 30 bp downstream from the normal start position, but this is not preceded by a ribosome binding site, and so is unlikely to be a major site of translational initiation; it is not known whether the residual translation initiates upstream or downstream of the normal start site, and whether the low level of transposition is due to low TnsA concentration or production of a protein species which is suboptimal for transpositional function.

It therefore appears that translation of *tnsA* is indeed necessary for transposition to occur, although the precise translational start point does not appear to be critical.

3.4 Strategies for overexpressing *tnsA*

A wide range of strategies was adopted for attempting to overexpress *tnsA*, using different promoters and ribosome binding sites; some methods involved the construction of fusion proteins. These various methods are described in the following sections. Three main tactics were employed: fusions to the 5' end of *lacZ* (Section 3.5), glutathione-*S*-transferase fusions

(Section 3.6) and use of the bacteriophage T7 gene 10 promoter and ribosome binding site (Section 3.7).

Unless stated specifically, all experiments involving expression of the *tnsA* gene were carried out in the *lacI*^{q} strain DS941. The *lacI*^{q} allele is a *lacI* promoter up-mutation, leading to over-expression of the *lac* repressor protein, and thus enhancing repression of a gene under the control of a *lac* or *tac* promoter before induction with IPTG.

The plasmids examined initially for their ability to express *tnsA* to a high level were the pMR78-based pMR204 (see Figure 4.3), constructed by M. Rogers (unpublished), and its derivative pAG204. In both plasmids the *tns* promoter has been deleted and *tnsA* is under the control of the *tac* promoter while retaining its Tn*7* ribosome binding site; pMR204 also encodes 542 bp of the 5' end of *tnsB*, while pAG204 encodes only 67 bp of *tnsB*. Following induction with 1mM IPTG, no protein corresponding in size to TnsA can be seen on a Laemmli gel although it can be shown that under these conditions, when *tnsB, C* and *D* are provided on a compatible plasmid, pAG101, a mini-Tn*7* can transpose *in vivo* with the normal observed frequency for plate-mating transposition assays, thus indicating that protein is being produced, albeit at a very low level.

3.5 Construction and investigation of *lacZ-tnsA* fusion proteins

Having taken into account the various theoretical considerations outlined in Section 3.2, it was thought likely that the poor *tnsA* ribosome binding site was responsible for the lack of detectable protein on Laemmli gels in the early experiments described above. To counteract this, the plasmids pAG081, pAG082 and pAG083 were constructed; in each of these, the *tnsA* gene is fused to the 5' end of the *lacZα* gene of pUC18; the sequence of the junction as confirmed by plasmid sequencing is:

```
 [RBS]          lacZ→           [ EcoRI ]
CAG GAA ACA GCT ATG ACC ATG ATT ACG AAT TCC CGG ACT TTT GTT ATG GCT
                met thr met ile thr asn ser arg thr phe val met ala
```

where the Tn*7* wildtype ribosome binding site and presumed start codon are shown underlined. The sequence of the *lacZ* ribosome binding site appears to be slightly nearer the consensus than that of the Tn*7* site, and thus it is likely that translation initiates at the first codon of *lacZ*, producing a fusion

protein of the N-terminal 7 amino acids of LacZ, 4 amino acids derived from *tnsA* upstream sequence, and TnsA itself. The *lac* promoter of the pUC plasmids is not as strong as the *tac* promoter of pMR78-based plasmids, but the former can be repressed to a much greater degree, and so it was hoped that any protein production subsequent to IPTG induction would be easily detected on Laemmli gels without resorting to radioactive labelling. The plasmids pAG081 and pAG082 were first tested for their ability to encode *tnsA* and *tnsA+tnsB* respectively by using a plate mating transposition assay, in which pAG101 supplied *tnsBCD* and pMR106 supplied *tnsCD*. In both cases, the measured transposition frequency was $\sim 10^{-2}$ both in the presence and in the absence of 1mM IPTG. This suggests that only a very low level of gene expression is required for transposition, and that the repression of the *lac* promoter, even in a *lacI^q* strain is not complete.

It was decided to examine directly the TnsA and TnsB proteins on a Laemmli gel, in order to ascertain the levels of expression under conditions of induction and repression. 200ml of either rich broth or minimal medium was inoculated with DS941 harbouring the appropriate plasmid (pAG081 or pAG082) and grown with ampicillin to OD_{600} = 0.4. The culture was then induced with 1mM IPTG, a further 200ml broth with ampicillin added and growth continued for 3 hours. The cells were lysed by sonication in the presence of 1mM PMSF and spun; the pellet was resuspended in 1M KCl. Samples of uninduced and induced whole cells, the supernatants and the resuspended pellets were run on a 12½% Laemmli gel; production of TnsA was not visible. Under these conditions, TnsB could not be detected by differential expression in the extracts from induced pAG082-containing cells either, and so it was thought that a clearer distinction between repressed and induced cells should be made.

The *lac* promoter of pUC contains the upstream CAP (catabolite activator protein) binding sites found in the chromosomal *lac* operon, and so transcription should be greatly reduced by the addition of glucose or gluconate to the growth medium; this addition would have the effect of reducing intracellular cAMP concentration and thus inactivating CAP. To this end repressed cells were grown in broth with the addition of glucose to various concentrations from 0.2% to 4.0%; there was no difference in growth rate with added glucose. Induction was carried out by adding 0.2% fructose with IPTG. Again no TnsA could be seen on Laemmli gels. A ^{35}S-methionine pulse-labelling experiment was carried out with cells grown

Sugar in growth medium	IPTG induction?	Measured transposition rate
Glucose	No	6.5×10^{-4}
Glucose	Yes	1.3×10^{-2}
Fructose	No	1.0×10^{-2}
Fructose	Yes	1.0×10^{-2}

Table 3.3 The effect of different sugars in the growth medium on the measured *in vivo* transposition rate, using the mate-out assay described in Chapter 2. Cells were grown in minimal medium with Davis and Mingioli salts, casamino acids, thiamine and either glucose or fructose to 1.3% w/v. The donor *E. coli* strain was DS941::Tn*7*-1(pEN300); pAG081 (high copy number; *lac* promoter) was used as a source of TnsA; pAG101 (medium copy number; *tac* promoter) provided TnsB, TnsC and TnsD. 1mM IPTG was used where indicated to induce the *lac* and *tac* promoters for 2 hours prior to mating out pEN300 into the recipient strain, MR1.

initially in 1% glucose, then induced with IPTG 15 minutes before the labelled methionine was added; no TnsA could be seen on the autoradiograph. The effect of glucose addition on *in vivo* transposition was investigated. In one experiment it was found that 1.3% (72mM) added glucose, as the sole carbon source, reduced transposition rates ~100fold, an effect which was negated by supplying 1mM IPTG in addition. There was no difference in transposition frequency when 1.3% fructose was added (see Table 3.3). Conjugation frequency of the donor and recipient strains in the mate-out assay was unaffected by the addition of glucose, either in the liquid growth medium or on the plates on which conjugation took place, and conjugation took place at the same frequency on rich and minimal medium plates. It therefore appeared that the variation in measured transposition rates was likely to be due to differential expression from one or both of the plasmids pAG081 and pAG101 which encoded the *tns* genes, although the degree of glucose repression was lower than might have been expected. It was not certain, however, that TnsA production was in any way affected by the identity of the sugar present in the growth medium; it was equally possible that the concentration of one of TnsB, TnsC or TnsD was the limiting factor in transposition.

Attempts were made to control the *lac* promoter better by using 20mM sodium gluconate in addition to glucose in the growth medium for maximum repression in the absence of IPTG, but again the catabolite repression of transcription of the *tns* genes was not sufficient to block transposition; this still took place, at a measured frequency of $\sim 1.4 \times 10^{-3}$, when 220mM glucose and 220mM sodium gluconate were added together to the growth medium.

The ability to repress the *lac* promoter under these conditions was tested by using in DS941, instead of pAG081 / pAG082, the plasmid pUC18::*lacZ*, in which the complete *lacZ* gene is under the control of the *lac* promoter. Cells were streaked on minimal plates containing X-gal and various concentrations of glucose and sodium gluconate, ranging from 20mM glucose to 400mM of each. It was found that with up to 60mM glucose + 60mM sodium gluconate the colonies were blue, while with at least 100mM glucose + 100mM sodium gluconate the colonies were white. When a single chromosomal copy of *lacZ* is repressed, colonies are *just* blue; induction of the chromosomal *lacZ* yields ~1000 times more β-galactosidase. pUC is a high copy-number plasmid and so approximately 100 times more β-galactosidase

might be expected to be produced than from the single-copy chromosomal gene. Hence repression of the *lac* promoter in pUC18::*lacZ* to white colony status must involve repression of the order of 10^5-fold. The fact that Tn*7* transposition still occurs, albeit at a 100-fold lower frequency at 60-100mM glucose suggests that only very low expression of the *tnsA* gene is required for transposition.

In order to investigate more quantitatively the effect of glucose and gluconate concentrations on the *lac* promoter, β-galactosidase assays were carried out. The method followed that of Miller (1972), with modifications as described in Chapter 2. The strains used were:

1. DS941, which is $lacI^q$ and has no complete *lacZ* gene, so should display no β-galactosidase activity;
2. ML308, a $lacI^-$ strain;
3. MR5 (pEN300, pAG101, pUC::*lacZ*), which was chosen to mimic as closely as possible the conditions used for the *in vivo* transposition assays using pAG081 or pAG083.

The molar concentrations of glucose and sodium gluconate used were equal and were as indicated in Table 3.4 below.

Number of Miller units β-galactosidase activity

[glucose] [gluconate] mM	DS941 - IPTG	DS941 + IPTG	ML308 - IPTG	ML308 + IPTG	MR5 (pEN300, pAG101,pUC18::*lacZ*) - IPTG	MR5 (pEN300, pAG101,pUC18::*lacZ*) + IPTG
10	0	0	$3.9x10^3$	$3.9x10^3$	$4.1x10^3$	$3.7x10^3$
50	0	0	$1.8x10^3$	$1.5x10^3$	$8.4x10^2$	$9.6x10^2$
100	0	0	$1.5x10^3$	$1.6x10^3$	$1.0x10^3$	$1.2x10^3$
200	0	0	$1.7x10^3$	$1.6x10^3$	$2.9x10^3$	$2.2x10^3$
300	0	0	$1.8x10^3$	$1.5x10^3$	$6.8x10^3$	$6.1x10^3$
400	0	0	$2.0x10^3$	$1.5x10^3$	$5.2x10^3$	$6.1x10^3$

Table 3.4 Expression of the *lacZ* gene, measured as β-galactosidase activity, from three different *E. coli* strains in the absence and presence of induction by 1mM IPTG.

There appeared to be no evidence for glucose/gluconate repression of the *lac* promoter, either chromosomal in ML308, or plasmid-encoded, in MR5 (pEN300,pAG101,pUC18::*lacZ*). The level of expression from the multicopy pUC18::*lacZ* was not significantly higher than that from the single copy in ML308. The reason for this is not obvious, but it is possible that an insufficient concentration of the ONPG substrate for the assays was provided or that it was unable to enter the cells freely. It is also clear from the above results that neither single copy nor multicopy p*lac* is repressed by high concentrations of glucose and gluconate.

It was decided that further refinement of the pAG081/083 expression system would not be fruitful in view of the above results on promoter control, and also because it could not be established whether it was the amino-terminal LacZ-TnsA fusion protein or the wildtype TnsA protein which was responsible for the *in vivo* transposition observed.

3.6 Glutathione-*S*-transferase fusions

Attempts to visualise TnsA from expression on the plasmids pAG204 (wildtype protein, *tac* promoter) and pAG081/083 (N-terminal LacZ fusion, *lac* promoter) were not successful. It was therefore decided to make use of the pGEX system of making fusion proteins (Smith and Johnson, 1988).

The structure of the plasmid pGEX-3X is shown below in Figure 3.7. The glutathione *S*-transferase (GST) encoding gene was obtained from a cDNA of *Schistosoma japonicum*, and is under the control of the *tac* promoter. At the 3' end of the gene is a sequence which will, when translated, give the cleavage recognition site, Ile Glu Gly Arg, of the serine protease, Factor X_a; this is followed by BamHI, SmaI and EcoRI restriction sites for the cloning in of the gene to be co-expressed and stop codons in all three reading frames. The plasmid also contains the *lacI*^{q} gene for enhanced control of the *tac* promoter.

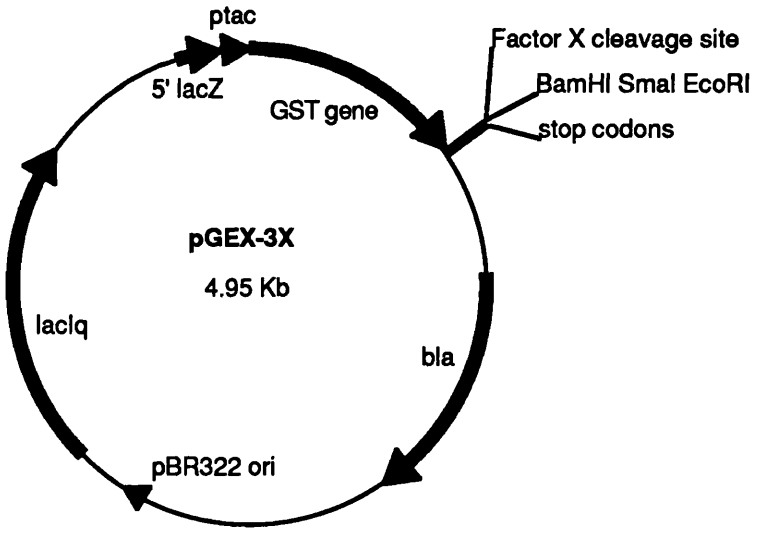


Figure 3.7 Structure of the expression vector pGEX-3X

The sequence of the region around the cleavage site is:

```
                    ⌈  Factor Xa  ⌉↓
Pro Lys Ser Asp Leu Ile Glu Gly Arg Gly Ile Pro Gly Asn Ser Ser ***
CCA AAA TCG GAT CTG ATC GAA GGT CGT GGG ATC CCC GGG AAT TCA TCG TGA
                                    ⌊ BamHI ⌈⌋SmaI⌊⌉EcoRI ⌋

CTG ACT GAC GAT CTG
```

where underlining indicates the position of a stop codon.

Upon IPTG induction, a fusion protein with the desired protein downstream of glutathione *S*-transferase is produced. In theory, the fusion protein can be purified through its affinity for glutathione, which may be immobilised on beads of agarose or the closely related sepharose. The fusion protein can then be cleaved, on the beads, with Factor X_a, the activated form of Factor X, to release the desired protein.

The plasmid pAG301 was constructed by excising the Tn*7* fragment from pAG081 with EcoRI and Hinc*II*; pGEX3-X was cut with EcoRI; and the Tn*7* fragment, encoding *tnsA* and the 5' 0.5Kb of *tnsB*, was inserted, retaining the EcoRI site and the reading frame. The sequence at the junction was thus:

```
... Pro Gly Asn Ser Arg Thr Phe Val Met Ala ...
... CCC GGG AAT TCC CGG ACT TTT GTT ATG GCT ...
          EcoRI
```

where the normal start codon of *tnsA* is indicated by bold type. The fusion protein was predicted, using GCG programs, to have a molecular weight of 59kDa and a pI of 6.2. Cleavage of this fusion protein with Factor X_a would yield a TnsA protein with 10 additional amino acids at the amino-terminal end; it was hoped that this would not make any difference to the activity of the protein.

The plasmid pAG302 was also constructed; it differed from pAG301 in that the EcoRI site found in pAG301 between the GST and *tnsA* coding regions had been filled in with the Klenow fragment of DNA polymerase I, with a resultant loss of reading frame, and hence, as is evident from Figure 3.8, no fusion protein was detectable. Wildtype TnsA protein could, however still be produced from pAG302 as the original ribosome binding site and translational start codon were retained.

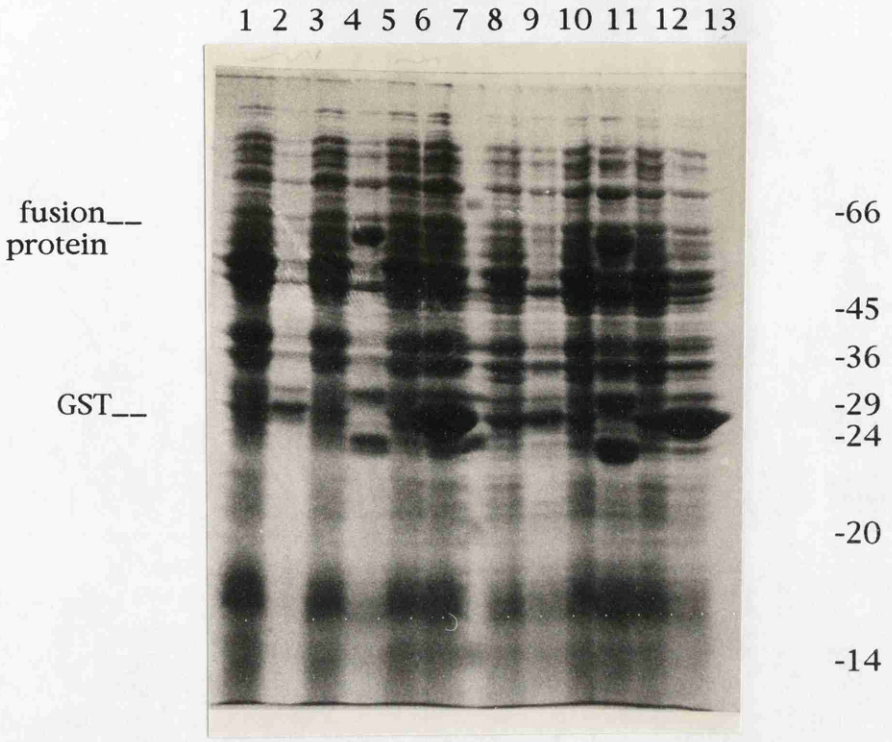


Figure 3.8 Coomassie-blue stained 12.5% polyacrylamide Laemmli gel of whole cells, showing production of the GST-TnsA fusion protein in DS941, and also of the two apparent cleavage products of this fusion protein. The three plasmids used were pGEX-3X (the expression vector), pAG301, which encoded the fusion protein, and pAG302, in which there was a frameshift. Induction was with 1mM IPTG for the times shown below; uninduced control cultures were grown for the same time as the induced cells.

1	pGEX-3X	no IPTG	4.5 hours
2	pGEX-3X	IPTG	4.5 hours induction
3	pAG301	no IPTG	4.5 hours
4	pAG301	IPTG	4.5 hours induction
5	pAG302	no IPTG	4.5 hours
6	pAG302	IPTG	4.5 hours induction
7	molecular weight markers (sizes in kDa)		
8	pGEX-3X	no IPTG	7 hours
9	pGEX-3X	IPTG	7 hours induction
10	pAG301	no IPTG	7 hours
11	pAG301	IPTG	7 hours induction
12	pAG302	no IPTG	7 hours
13	pAG302	IPTG	7 hours induction

Plate mating transposition assays were carried out to determine whether the fusion protein encoded by pAG301 could substitute for TnsA in transposition of the mini-Tn7 in DS941::Tn7-1 to the *att* site of pEN300. pAG101 was used to supply TnsB, C and D; and either pAG081, pGEX-3X, pAG301 or pAG302 was also transformed into the donor cells. After mating into MR1, the mean measured transposition rates (which do not vary substantially with IPTG induction) were:

with pAG081	1.3×10^{-2}
with pGEX-3X	$<2.5 \times 10^{-7}$
with pAG301	5.5×10^{-4}
with pAG302	7.0×10^{-3}

The transposition rate in cells encoding the GST fusion protein was markedly lower than that normally observed (as seen with pAG081 or pAG302-containing cells) and suggests that a fusion protein was indeed being produced, and that it was interfering with the transposition process in some way; there might for example be both the GST-TnsA fusion protein and a low concentration of TnsA present in the cells since the *tnsA* ribosome binding site and start codon were not removed during the cloning, with the possible formation of heterodimers (or other oligomers) which sequestered a large proportion of the TnsA in an inactive form; or there might simply be competition between the wildtype (active) and the fusion (inactive) proteins.

It is noticeable that the measured transposition rate in the presence of pAG302, which contains the frameshift in the fusion gene, is similar to that normally observed with wildtype *tns* genes under the control of their own or a *lac* or *tac* promoter. It is presumed that active wildtype TnsA is being produced under these conditions, since the ribosome binding site and translational start codon are unaltered by the frameshift cloning. This implies that, in the presence of pAG301, sufficient wildtype TnsA for transposition is available, and supports the proposal that the fusion protein may play no positive rôle in transposition.

Transposition assays were also carried out with pMR121 instead of pAG101; wildtype TnsA is therefore known to be present in the donor cells as well as any fusion protein. in this case; the observed transposition frequencies were:

No added plasmid	1.5×10^{-2}
+ pGEX-3X	1.7×10^{-2}
+ pAG301	6.0×10^{-3}
+ pAG302	2.0×10^{-2}

In GST-TnsA fusion protein producing cultures, transposition appears to occur at a higher frequency in the presence of wildtype TnsA. There was a slight tendency for transposition to be enhanced in IPTG induced cells, but this was never more than 6-fold.

Prior to attempts at purification, growth of cells containing pGEX-3X or one of its derivatives was investigated. Duplicate sets of 200ml L broth with ampicillin in a 250ml flask were inoculated with an overnight culture of DS941 containing pGEX-3X, pAG301 or pAG302; after growth to early logarithmic phase, one of each pair of flasks was induced with 1mM IPTG and the OD_{600} was monitored over a period of 7 hours. The results are shown in Figure 3.9. IPTG induction has a marked effect on the growth rate of cells containing either pGEX-3X or pAG301 but, strangely, those containing pAG302 did not show any decrease in growth; it is possible that they had acquired a mutation in the GST gene or the promoter region. It was clear from the continued growth that induction of the fusion gene in pAG301 was not lethal to the cells and so it was decided to continue with maximisation of the production and purification of the GST-TnsA fusion protein.

Samples of cells were harvested $4^1/2$ and 7 hours after induction and were run on a Laemmli gel; see Figure 3.8. Glutathione *S*-transferase is clearly visible after induction of the *tac* promoter in pGEX-3X and pAG302-containing cells. In pAG301-containing cells, three polypeptides are present in high concentration, together probably constituting greater than 50% of the total cell protein, judged by Coomassie staining, as can be seen on Figure 3.8. The largest, of ~58kDa, is presumed to be the GST-TnsA fusion protein. Two smaller proteins run with mobilities corresponding to 24kDa and 30kDa, and the sum of their sizes suggests that they may be natural cleavage products of the GST-TnsA fusion. However, unless the size difference between the two is a gel artefact, cleavage is not occurring at the desired position, so that purification by affinity chromatography would become inefficient. The spontaneous cleavage of GST fusion proteins has

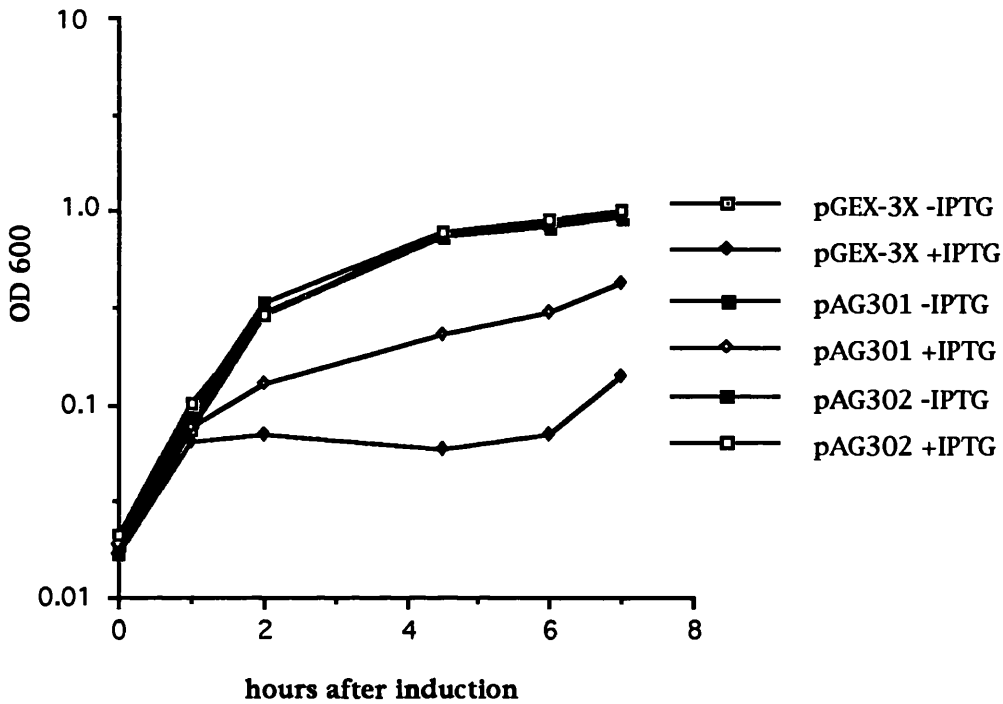


Figure 3.9 Growth of DS941 harbouring one of the GST-encoding plasmids, as indicated above, following induction with 1mM IPTG.

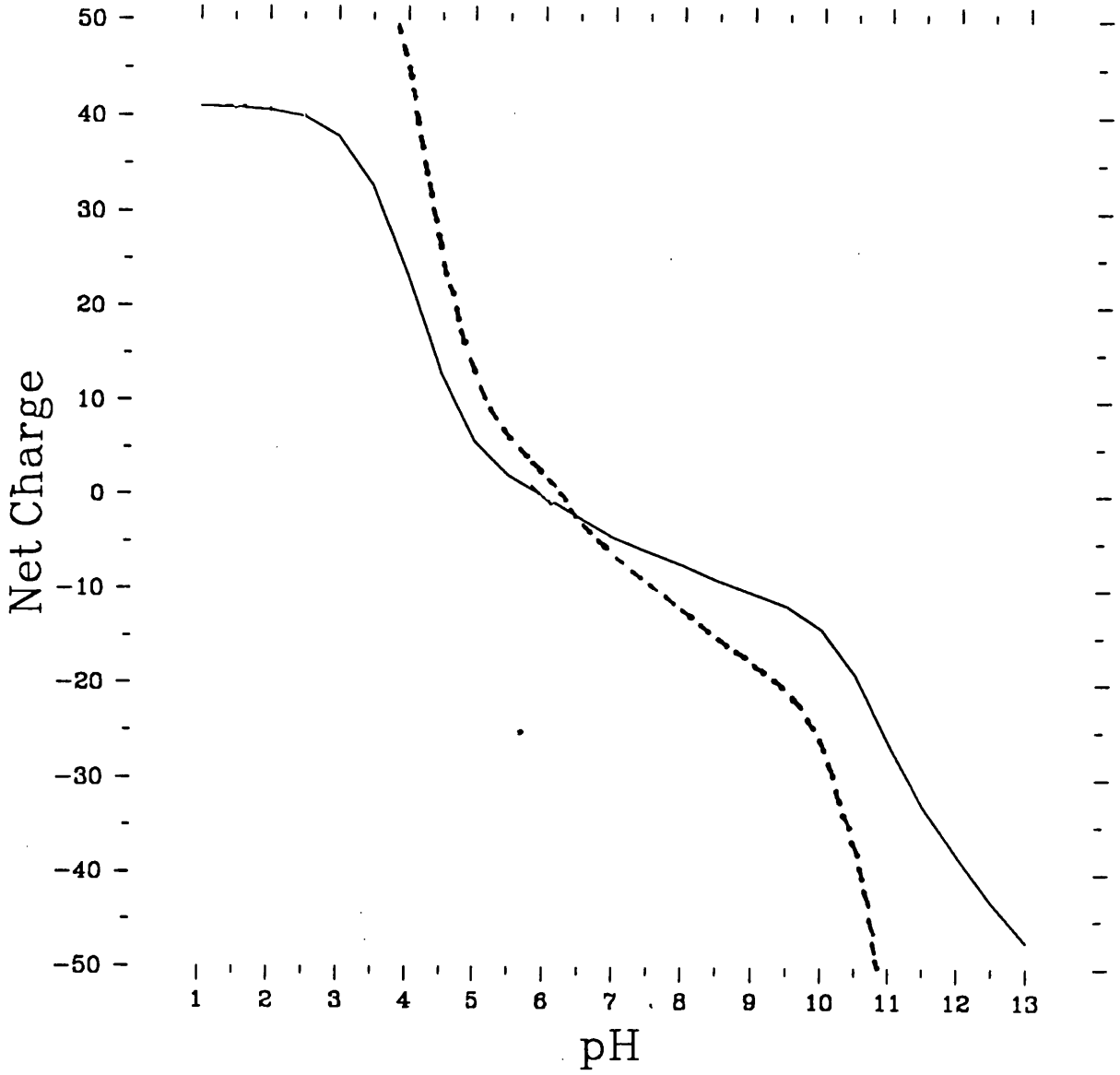


Figure 3.10 The variation of charge with pH in TnsA and in the glutathione-*S*-transferase-TnsA fusion protein encoded by pAG301.

——— TnsA

- - - - - - - - fusion protein

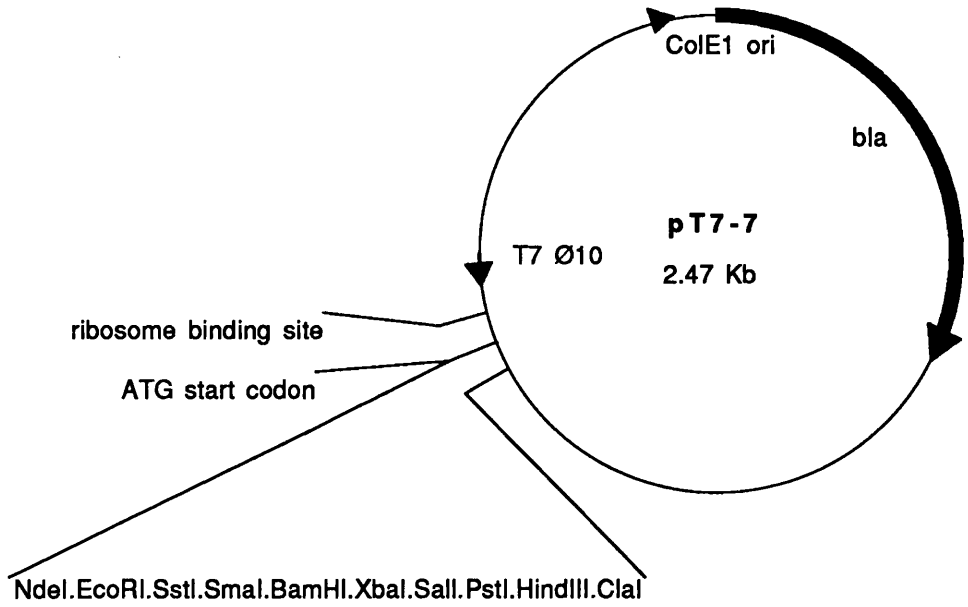
been recognised in other situations (N. Spibey, pers. comm.). Use of a protease-deficient strain might have alleviated the observed GST-TnsA cleavage, but unfortunately no such strain was available at the time this work was carried out.

Cells were lysed with lysozyme in a 100mM KCl buffer with PMSF at pH8.0, at which the fusion protein molecule would be expected to have a charge of ~-12 (see Figure 3.10): GST in pGEX-3X containing cells was soluble; the fusion protein and the putative cleavage products were all insoluble at this salt concentration. The cell pellet was resuspended in 1M KCl buffer; the fusion protein was still insoluble, as was the larger of the two other peptides, but the smaller appeared to show slight solubility. (This small peptide was too small to contain the whole of TnsA, and so purification of it was not investigated further.) It was thought that the fusion protein might be more soluble if cells were grown at a lower temperature; there is precedent for this with the *E. coli* protein XerC (G. May, pers. comm.); however, both the GST-TnsA fusion protein and the 30kDa protein were still completely insoluble in 1M salt. Neither was the problem alleviated by lysing the cells in the presence of 1.2% v/v Triton X-100. Because GST is readily soluble, it was considered likely that the 24kDa protein represented a substantial portion of GST and that the more insoluble 30kDa protein contained TnsA. As there seemed to be great problems in solubilising the fusion protein, before TnsA purification could even be considered, and because glutathione immobilised on agarose beads could not be used for purification of the cleaved product, it was decided to try an alternative approach to producing TnsA, as described below.

3.7 Expression from the T7 promoter

The pT7 series of expression vectors has been developed by S. Tabor (Tabor and Richardson, 1985; Tabor, unpublished). Each contains a polylinker downstream of the bacteriophage T7 ϕ10 gene promoter. T7 RNA polymerase differs from *E. coli* RNA polymerase in being a single large (99kD) polypeptide, It is specific for its own promoters; thus when T7 RNA polymerase is made available in cells, transcription of mRNAs under control of a T7 promoter is induced, with other operons unaffected. The induction can be made more effective by the addition of rifampicin to the cells; this disables the β subunits of the host RNA polymerase core so that

a



b

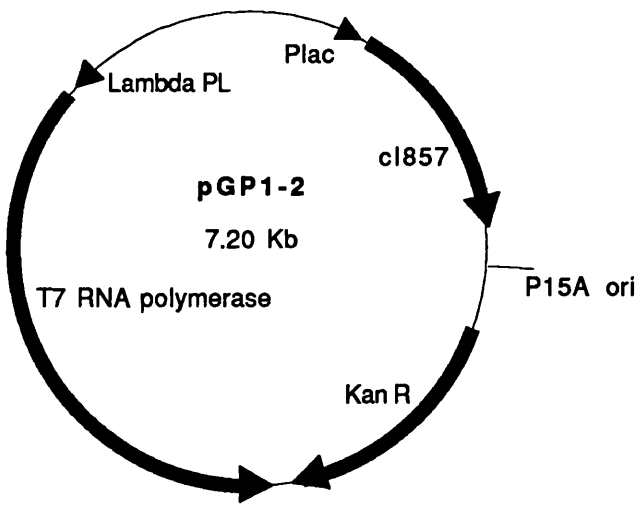


Figure 3.11 Structure of:

(a) the T7 promoter expression vector, pT7-7, and

(b) the T7 RNA polymerase encoding plasmid, pGP1-2

only genes under a T7 promoter are transcribed. In this system, the T7 RNA polymerase is encoded by a λ prophage, DE3, by an M13-derived phage, mGP1-2, or by the compatible plasmid pGP1-2 (see Figure 3.11b); in the first two cases, the RNA polymerase gene is controlled by p_{lac} and is thus IPTG inducible; in pGP1-2, the T7 RNA polymerase gene is under the control of a λ P_L promoter, with the temperature-sensitive allele cI_{857} of the λ repressor itself under a *lac* promoter as shown in Figure 3.11b. Induction of the T7 RNA polymerase gene in pGP1-2 can therefore be achieved by raising the growth temperature from 30° to 37° or 42°.

Among the pT7 series are several transcriptional fusions to the φ10 promoter; however the most promising for over-expression of *tnsA*, bearing in mind previous problems with protein levels, was considered to be pT7-7 (see Figure 3.11a), which also contains a strong ribosome binding site and the translation start site for the T7 gene 10; the start codon is included in a NdeI site (CATATG) at the beginning of the polylinker. Any gene also containing a NdeI site at the translational start can be cloned in and translated precisely as in the wildtype gene.

tnsA does not possess a NdeI site at the start codon; however it is possible to engineer one using site-directed mutagenesis, as described in the next section.

3.7.1 Site-directed mutagenesis of the *tnsA* translational start

Several methods are available for mutagenising a specific DNA sequence in a plasmid, for example see Smith (1985) for a review of various methods.

It was decided to use the Altered Sites™ *in vitro* mutagenesis system marketed by Promega (Lewis and Thompson, 1990), because of its simplicity. As shown in Figure 3.12, the basis of this system is a phagemid mutagenesis vector, pSELECT, which contains a *lacZα* gene with an inserted polylinker into which the sequence to be mutated is cloned (giving, in this case, the phagemid pAG401); and two genes encoding antibiotic resistance; the one encoding tetracycline resistance is functional, but the β-lactamase gene for ampicillin resistance has an engineered frameshift mutation inactivating it. Two mutagenising oligonucleotides are required, one for the desired site-specific mutation and the other to repair the β-lactamase gene; these are both 5' phosphorylated and annealed to the phagemid single strand template, as shown in Figure 3.13, with the former oligonucleotide in fivefold excess; and T4 DNA polymerase, which does not

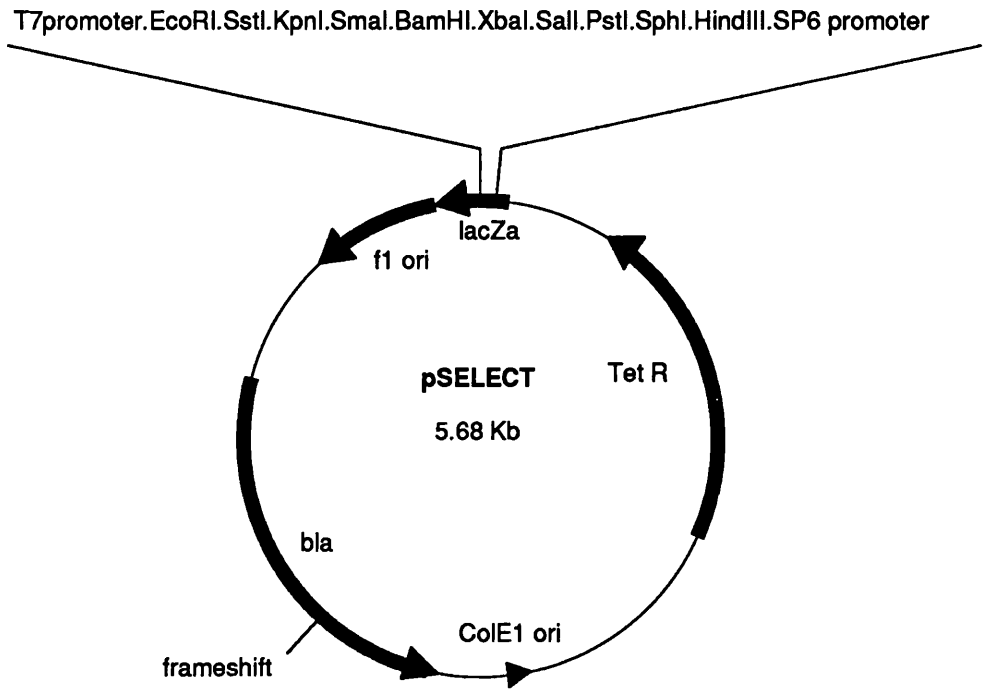


Figure 3.12 Structure of the phagemid pSELECT™ , supplied as part of the Promega Altered Sites mutagenesis kit and used for the site-directed mutagenesis described in Section 3.7.1.

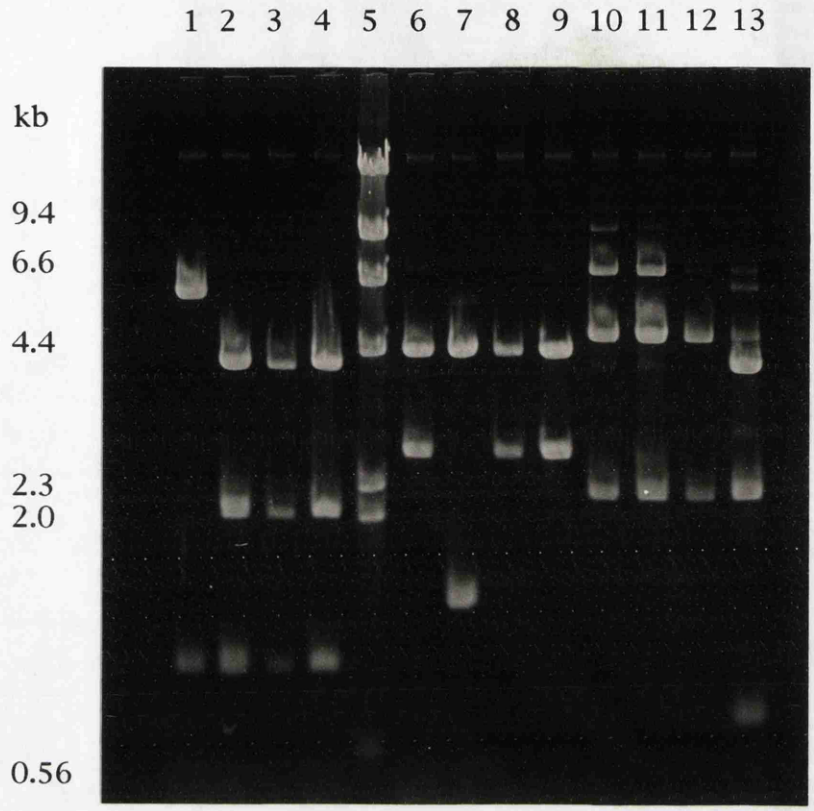


Figure 3.14 1% agarose gel showing diagnostic DNA restriction patterns of pAG401 (before mutagenesis), pAG402 (introduced Hin*c*II site), pAG403 (in which only the PstI site in the β-lactamase gene has been repaired) and pAG404 (introduced NdeI site)

1	pAG401	PstI
2	pAG402	PstI
3	pAG403	PstI
4	pAG404	PstI
5	λ	Hin*d*III
6	pAG401	Hin*c*II
7	pAG402	Hin*c*II
8	pAG403	Hin*c*II
9	pAG404	Hin*c*II
10	pAG401	NdeI
11	pAG402	NdeI
12	pAG403	NdeI
13	pAG404	NdeI

displace pre-bound DNA, is used to complete the second strand. The double stranded phagemid is transformed into the *E. coli* strain BMH71-18*mutS* which does not repair heteroduplex DNA; and ampicillin resistance is selected for. Because of the differential concentration of the two oligonucleotides, the majority of ampicillin resistant colonies are likely to contain the desired mutation, and plasmid DNA can be purified from them.

In this case, the Tn*7* sequence of *tnsA* and the 5' end of *tnsB* was cloned as a EcoRI-Hin*d*III fragment from pMR204 into the polylinker of pSELECT, and the oligonucleotide encoding the site-specific mutation had the sequence:

CTTAGGACTTTT*CA*TATGGCTAAAGC

where italics indicate mutated bases, the introduced NdeI site is shown underlined, and the translational start site is in bold type; there are 12 matching bases each side of the mutated ones, for optimal binding to the template.

Two ampicillin-resistant colonies resulting from the mutagenesis were screened: one phagemid (pAG403) had the original Tn*7* sequence, but the other (pAG404) had incorporated the desired site-directed mutation at the *tnsA* start, as confirmed by restriction mapping (Figure 3.14) and plasmid sequencing (Figure 3.6) and shown in Figure 3.5.

Since the Tn*7* sequence in pAG401 and pAG404 is downstream of a T7 promoter (see Figure 3.5) it was thought that high level *tnsA* expression might be possible without using the pT7-7 system. This was investigated by an experiment in which ^{35}S-methionine was used to label proteins produced following heat induction of cells containing pGP1-2, as a T7 RNA polymerase source, and pAG401; TnsA was not visible on the autoradiograph, even when rifampicin was used to block other transcription, although a band believed to be due to the cloned 5' end of *tnsB* could be seen (Figure 3.15). It appears that the strong T7 promoter is not sufficient to bring about a high cellular level of TnsA and that, as already hypothesised, post-transcriptional factors are also important.

It was therefore necessary to transfer the Tn*7* sequence, with the introduced NdeI site, to the translational fusion vector pT7-7; this was achieved by cutting both pT7-7 and pAG404 with NdeI and Hin*d*III and inserting the Tn*7* sequence into the vector to give the plasmid pAG411. The sequence was checked by restriction mapping and, around the 5' *tnsA* joint, by plasmid sequencing; it is shown in Figure 3.5.

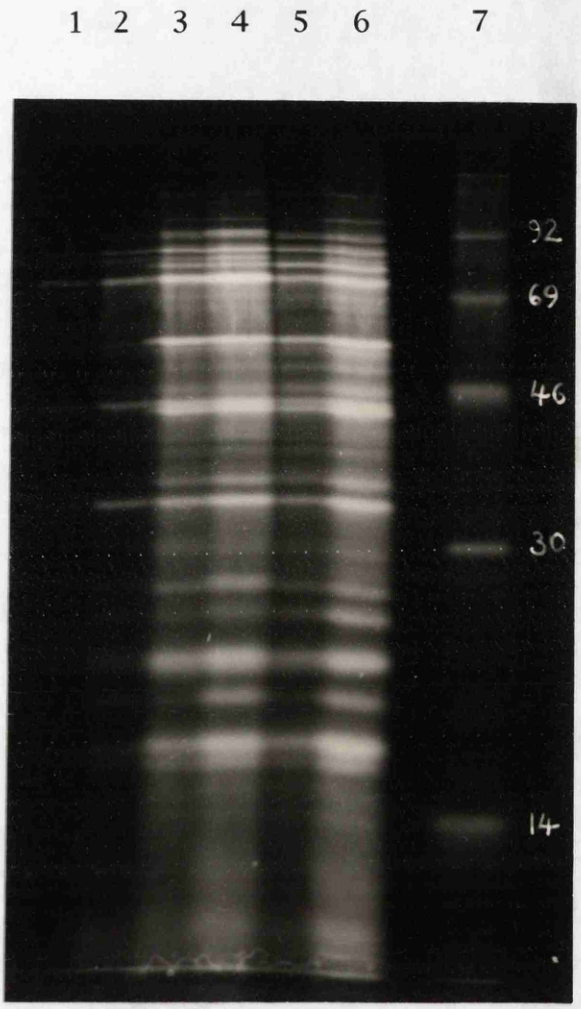


Figure 3.15 Autoradiograph of 12.5% polyacrylamide Laemmli gel showing 35S methionine labelled proteins in whole cells containing pGP1-2 as a source of T7 RNA polymerase and one of the plasmids pSELECT (depicted in Figure 3.12) and its derivative pAG401. The *tnsA* gene in pAG401 was induced as outlined below.

1	pSELECT	heat induction, rifampicin added, 2.5 hours expression
2	pAG401	heat induction, rifampicin added, 2.5 hours expression
3	pSELECT	heat induction, no rifampicin, 2.5 hours expression
4	pAG401	heat induction, no rifampicin, 2.5 hours expression
5	pSELECT	no heat induction, no rifampicin, 2.5 hours expression
6	pAG401	no heat induction, no rifampicin, 2.5 hours expression
7	molecular weight markers (sizes in kDa)	

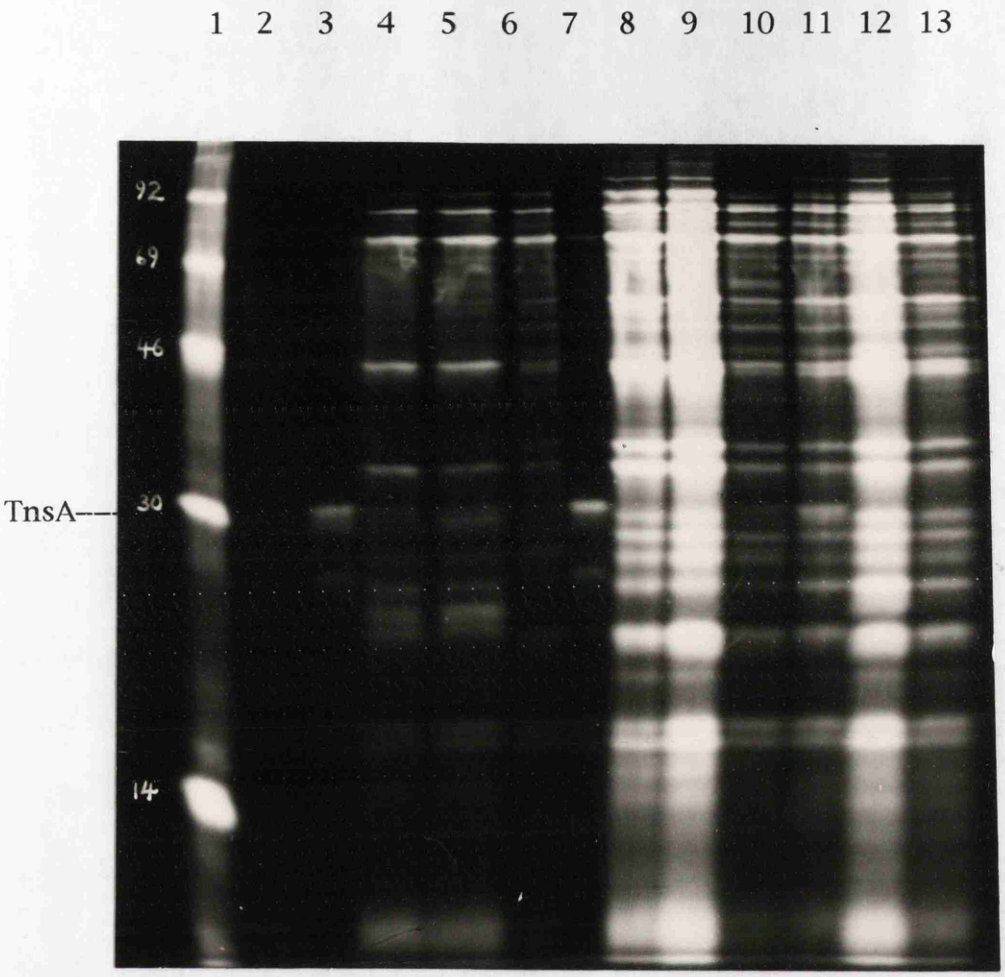


Figure 3.16 Autoradiograph of 12.5% polyacrylamide Laemmli gel showing ^{35}S methionine labelled proteins in whole cells containing pGP1-2 as a source of T7 RNA polymerase following induction of the T7 promoter by heat shock.

1	molecular weight markers (sizes in kDa)	
2	pSELECT	heat shock, rifampicin, 2.5 hours growth
3	pAG411	heat shock, rifampicin, 2.5 hours growth
4	pAG402	heat shock, rifampicin, 2.5 hours growth
5	pAG404	heat shock, rifampicin, 2.5 hours growth
6	pSELECT	heat shock, no rifampicin, 2.5 hours growth
7	pAG411	heat shock, no rifampicin, 2.5 hours growth
8	pAG402	heat shock, no rifampicin, 2.5 hours growth
9	pAG404	heat shock, no rifampicin, 2.5 hours growth
10	pSELECT	no heat shock, no rifampicin, 2.5 hours growth
11	pAG411	no heat shock, no rifampicin, 2.5 hours growth
12	pAG402	no heat shock, no rifampicin, 2.5 hours growth
13	pAG404	no heat shock, no rifampicin, 2.5 hours growth

As mentioned in Section 3.3, it was not certain that the presence of a translational start site in *tnsA* was necessary for transposition, and thus a second oligonucleotide for site-directed mutagenesis was synthesised with the sequence:

GGACTTTT*GTTAAC*GCTAAAGCAAAC

where italics indicate mutated bases, an introduced HpaI site is shown underlined, and the original site of the start codon is in bold type. Two ampicillin-resistant colonies resulting from the mutagenesis were screened, and both contained a phagemid (pAG402) in which the start site of *tnsA* had been lost and a HpaI/HincII site introduced, as confirmed by restriction mapping and plasmid sequencing (Figure 3.6). pAG402 was used to test for the requirement for a translational start codon for transposition; the mutated sequence was transferred to pT7-7, yielding the plasmid pAG416, as described fully in Section 3.3.

3.7.2 *tnsA* expression from pAG411

Once again, ^{35}S-methionine labelling was used to investigate expression of *tnsA* under the T7 ϕ10 promoter in pAG411, compared with the transcriptional fusions of pAG402 and pAG404. The method is described in Section 2.14. As can be seen in Figure 3.16, upon heat induction, protein production is almost exclusively of a species running on SDS-PAGE with a mobility expected for a protein the size of TnsA, even in the absence of rifampicin repression of the host RNA polymerase. It is interesting to observe that considerably more of TnsA is produced under these conditions than of the N-terminal TnsB fragment, contrary to all previous observations. The number of methionine residues (including the initial one) in TnsA is 3, the same number as in the TnsB, and so the strength of autoradiograph bands should be a reasonable indication of the relative concentrations of proteins present.

The ability of pAG411 to encode a TnsA protein active in *in vivo* transposition was tested using a standard plate mating assay with DS941::Tn*7*-1 as the donor strain and pAG101 as a source of *tnsB, C,* and *D*. mGP1-2 was used to supply T7 RNA polymerase because pGP1-2 and pAG101 both have kanamycin resistance as their only selectable marker. *In vivo* transposition was measured both in the presence and in the absence of mGP1-2 - encoded T7 RNA polymerase, the transcription of which can be

induced by IPTG, rather than heat shock. The following apparent transposition rates were obtained:

pT7-7 + mGP1-2	-IPTG	$<3.0 \times 10^{-8}$
pT7-7 + mGP1-2	+IPTG	$<2.0 \times 10^{-8}$
pAG411	-IPTG	4.5×10^{-2}
pAG411	+IPTG	5.0×10^{-2}
pAG411 + mGP1-2	-IPTG	3.8×10^{-2}
pAG411 + mGP1-2	+IPTG	2.4×10^{-2}

From the above results, the *tnsA* gene encoded by pAG411 is clearly active in transposition. It appears from the results with pAG411 in the *absence* of mGP1-2 that the production of T7 RNA polymerase is unnecessary for transposition to occur at wildtype frequencies; the particular T7 promoter sequence used here bears slight similarity to the consensus *E. coli* σ^{70} promoter sequence, and it is presumed that, in the absence of rifampicin, some transcription was thus initiating here.

Culture growth and production of TnsA under different regimes were investigated. 20ml cultures of DS941 containing pGP1-2 and either pT7-7 or pAG411 were grown at 30° in L broth with antibiotics and 1mM IPTG to repress transcription of the T7 RNA polymerase gene; after 2 hours, the cultures were split and one half of each was transferred to 42°. The OD_{600} was monitored for 4 hours after induction. The results are shown in Figure 3.17, and indicate that growth effectively ceases at 42°, upon production of T7 RNA polymerase, whether or not a gene is cloned downstream of the promoter.

A pulse-chase experiment, as described in Section 2.14, was carried out to determine the stability of TnsA produced under these conditions. As seen in Figure 3.4, the protein is stable for 15 minutes at the post-induction growth temperature of 30°, but after 60 minutes the TnsA level appears to have dropped slightly.

The optimum expression time for production of TnsA from pAG411 was investigated. 20ml cultures of DS941 containing pGP1-2 and either pT7-7 or pAG411 were grown, heat shocked at 42° to induce T7 RNA polymerase production and rifampicin treated as in Section 2.14; cultures were then transferred to 37°. Samples were collected after 1, 2, 3, 4, 5 and 6 hours and run on a Laemmli gel, as shown in Figure 3.18. The protocol on which this

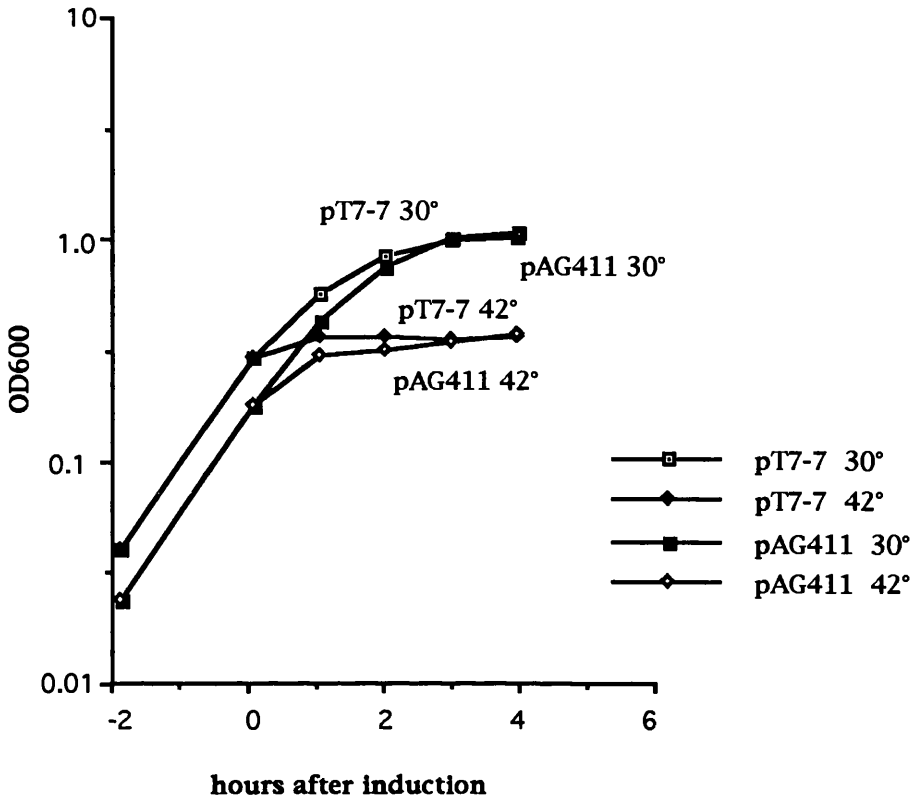


Figure 3.17 Growth of DS941(pGP1-2 pT7-7) and DS941(pGP1-2 pAG411) before and after induction by heat shock.

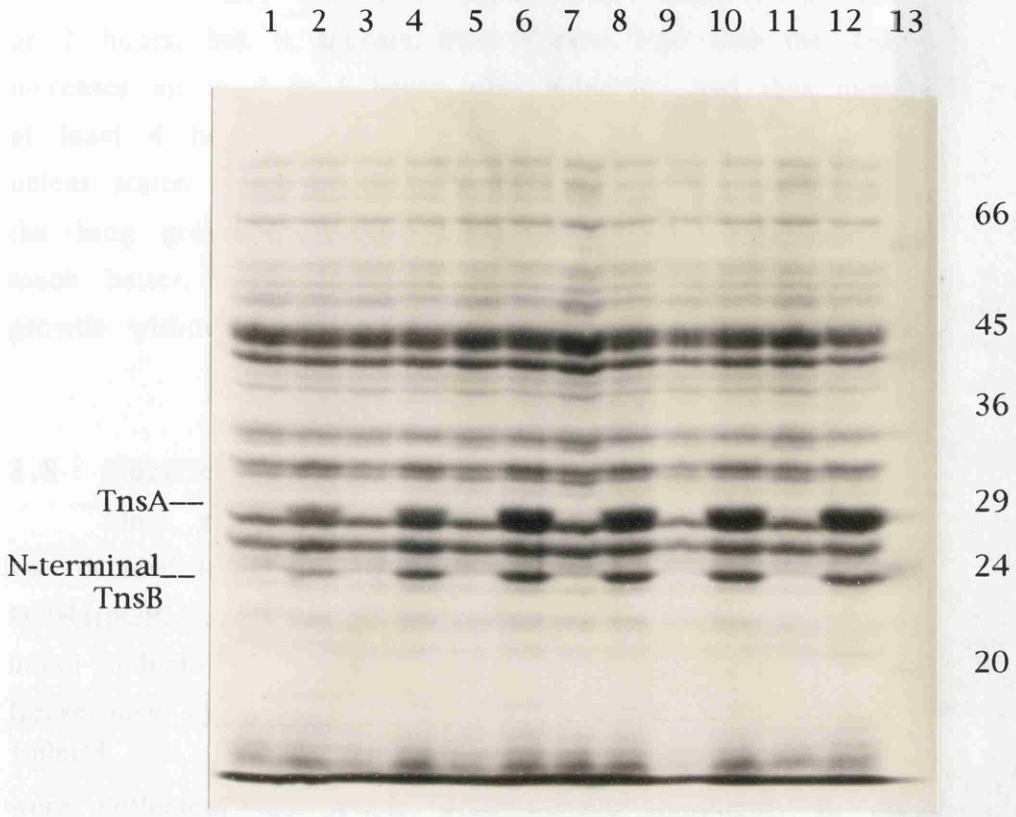


Figure 3.18 Coomassie stained 12.5% polyacrylamide Laemmli gel of whole cells illustrating production of TnsA from the plasmid pAG411 following induction of the T7 promoter in pGP1-2 by heat shock and addition of rifampicin to the induced cells. Expression of the *tnsA* gene was at 37° for the times indicated below. Alternate lanes show the control cloning vector, pT7-7.

1	pT7-7	1 hour induction
2	pAG411	1 hour induction
3	pT7-7	2 hour induction
4	pAG411	2 hour induction
5	pT7-7	3 hour induction
6	pAG411	3 hour induction
7	pT7-7	4 hour induction
8	pAG411	4 hour induction
9	pT7-7	5 hour induction
10	pAG411	5 hour induction
11	pT7-7	6 hour induction
12	pAG411	6 hour induction
13	molecular weight markers (sizes in kDa)	

method was based (S. Tabor, unpublished) recommended an expression time of 2 hours, but it appears from Figure 3.18 that the TnsA concentration increases up to 4 or 5 hours after induction, and thus induction periods of at least 4 hours were used in all further experiments with this system unless stated otherwise. Trials with and without rifampicin showed that, for the long growth times desirable for protein production, cell growth was much better, and TnsA yield was as good when rifampicin was omitted; growth without rifampicin was therefore the norm.

3.8 Purification strategies for TnsA

Once growth conditions had been optimised for TnsA production, it was decided to try to purify the protein. Initially, the small scale cultures of DS941(pGP1-2, pAG411) shown in Figure 3.18 were used; cells after 4, 5 and 6 hours induction were spun down. They were lysed with lysozyme and 2 freeze-thaw cycles in 150mM TrisCl pH8.0, 1mM EDTA, 1mM PMSF and 100mM KCl, then spun at 45000g at 2°C for 45 minutes. The supernatants were collected; the pellets were washed thoroughly in the same buffer, spun again, and the second supernatant removed. The pellets were then mixed with 150mM TrisCl pH8.0, 1mM EDTA, 1mM PMSF and 2M KCl, incubated on ice then spun down again. Samples run on a Laemmli gel (not shown) indicated that, as far as could be seen, all the high concentration protein believed to be TnsA was still in the pellet in the presence of 2M KCl, though it is possible that it was soluble at a concentration too low to be detected without the aid of Western blotting. The charge on a TnsA molecule at pH8.0 is predicted, as on Figure 3.10, to be ~-8, and so it had been hoped that the protein would be soluble in high salt concentration buffers. It is possible that, under these growth or lysis conditions, the TnsA molecules aggregate and form inclusion bodies within the cells; similar problems occur with other proteins expressed at a high level in *E. coli*, for example the native *E. coli* protein XerC. TnsA contains four cysteine residues which might be able to form intermolecular bonds in an *in vivo* or *in vitro* oxidation reaction; however, when whole cell samples *without* the reducing agent β-mercaptoethanol were loaded on Laemmli gels the band corresponding to TnsA appeared in the normal position with the normal intensity, suggesting that individual molecules, rather than large aggregates, were migrating through the gel. Thus, if aggregation occurs, it probably does not involve covalent intermolecular bonding. Since TnsA is

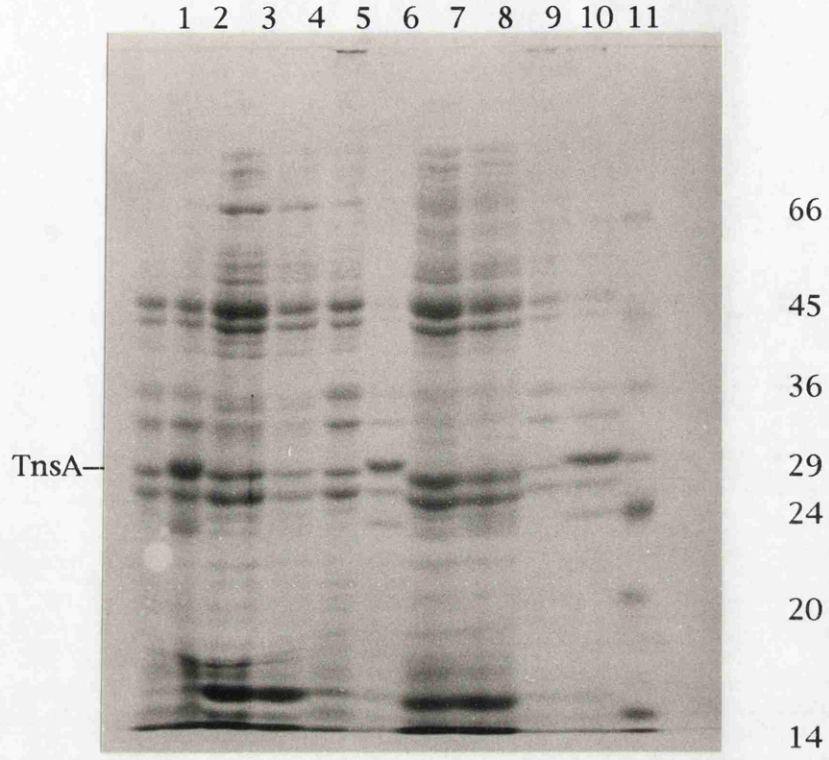


Figure 3.19 Coomassie stained 12.5% polyacrylamide Laemmli gel illustrating the insolubility under various conditions of overproduced TnsA, following induction for 6 hours in the presence of rifampicin. Alternate lanes show extracts of cells containing pGP1-2 and either the cloning vector pT7-7 or the *tnsA* expressing plasmid pAG411. Cell lysis was with lysozyme and freeze-thaw cycles, as described in Chapter 2. Each solubilisation step was followed by centrifugation to separate the supernatant and pellet.

1	pT7-7	whole cells
2	pAG411	whole cells
3	pT7-7	supernatant following solubilisation in 1M KCl pH8.0
4	pAG411	supernatant following solubilisation in 1M KCl pH8.0
5	pT7-7	pellet following solubilisation in 1M KCl pH8.0
6	pAG411	pellet following solubilisation in 1M KCl pH8.0
7	pT7-7	1M KCl pellet resuspended in 6M urea: new supernatant
8	pAG411	1M KCl pellet resuspended in 6M urea: new supernatant
9	pT7-7	1M KCl pellet resuspended in 6M urea: new pellet
10	pAG411	1M KCl pellet resuspended in 6M urea: new pellet
11	molecular weight markers (sizes in kDa)	

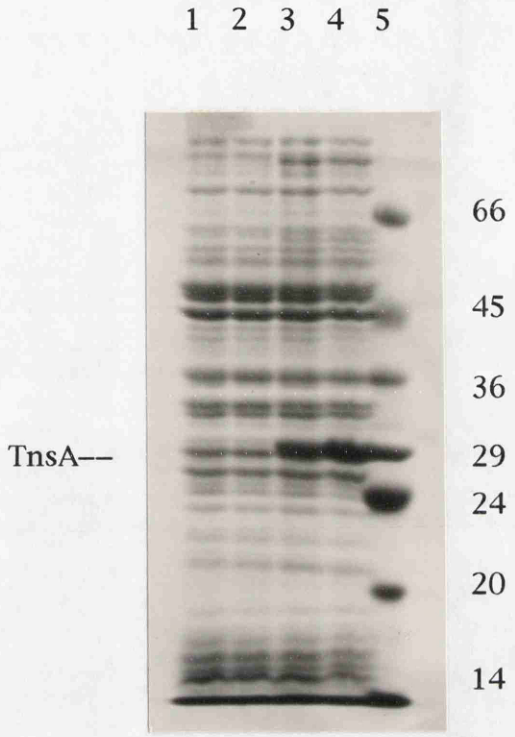


Figure 3.20 Coomassie stained 12.5% polyacrylamide Laemmli gel, showing expression in whole cells containing pGP1-2 and pSB58 of TnsA in the presence of TnsB, TnsC and TnsD. Cells containing the cloning vector pT7-7 are shown for comparison.

1	pT7-7	1 hour induction	whole cells
2	pT7-7	4 hour induction	whole cells
3	pSB58	1 hour induction	whole cells
4	pSB58	4 hour induction	whole cells
5	molecular weight markers (sizes in kDa)		

likely to interact in a transpososome with one or more of the other Tns proteins, it was thought possible that the 23kDa amino-terminal fragment of TnsB encoded by pAG411 was in some way inhibiting solubility of TnsA through TnsA-TnsB interaction; when pAG411 was substituted by pAG415, which encodes only the 22 amino-terminal residues of TnsB, exactly the same solubility characteristics were obtained, thus suggesting that TnsA-TnsB interactions were not significantly involved in the insolubility of TnsA.

Although maximum TnsA yield occurred after at least 4 hours growth, it was possible that soluble protein might be obtained after a shorter growth period, when the cellular TnsA level was lower. Cultures were therefore grown for only 10, 20 or 30 minutes after rifampicin addition. Samples were run on a Laemmli gel (Figure 3.19, lanes 1-6) and the remainder was lysed in a pH 8.0 buffer containing 1M KCl and spun; the TnsA was visible only in the pellet fraction under these conditions.

Another possibility was that TnsA required for solubility one of TnsB, TnsC or TnsD to be present. This was tested by using, instead of pAG411, pSB58, in which all of *tnsA, B, C* and *D* are cloned downstream of the same T7 promoter in pT7-7 with a translational fusion of the T7 sequence to *tnsA*. Cells were grown up, heat induced, and samples were collected 1 hour and 4 hours after rifampicin addition and run on a Laemmli gel. As can be seen in Figure 3.20, a protein believed to be TnsA can clearly be seen after both 1 hour and 4 hours; however, upon lysis in a 1M KCl buffer, all the TnsA was apparently once more in the pellet. It appears that co-translation of *tnsA* with *tnsB*, *tnsC* and *tnsD*, at least in the relative concentrations obtained in this system, does not facilitate solubility of TnsA.

It was possible that solubility was dependent on the salt used, and so further cultures were lysed in Tris glutamate buffers with potassium glutamate as the salt; potassium glutamate has been used successfully in protocols for the purification of other proteins, for example the Moloney murine leukaemia virus IN protein (Craigie *et al.*, 1990). Cultures containing pT7-7, pAG411 or pSB58 were grown for 3 hours after transfer from 42° to 37°, then split and lysed in either the normal Tris Cl/KCl buffer or in a 150mM Tris glutamate / 1M potassium glutamate buffer. Pellets and supernatants were run on a Laemmli gel (Figure 3.21), from which it appears that the TnsA was once more in the pellet. The pH of the buffer was also varied; at both pH6.5 (at which the protein would be expected to have a

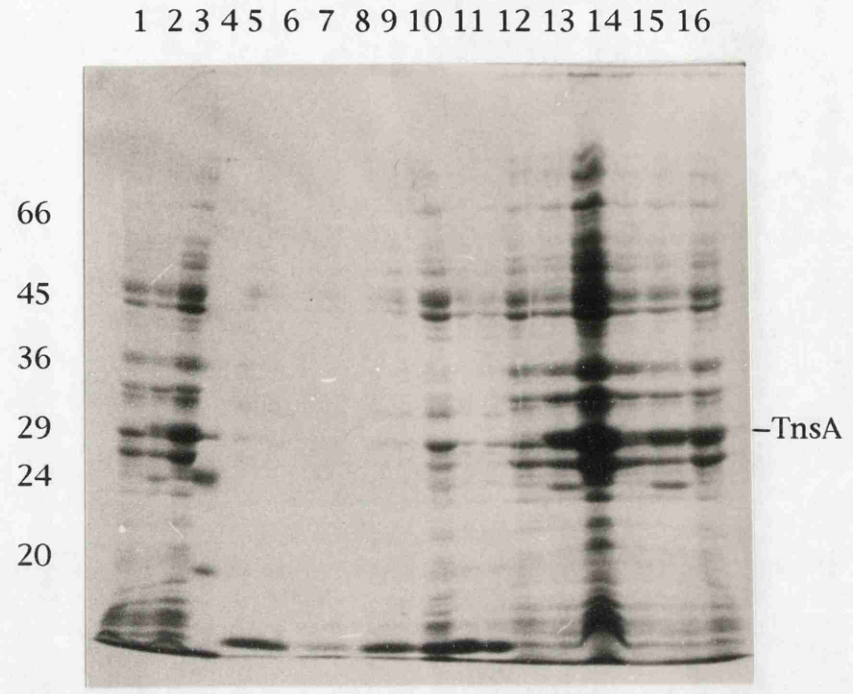


Figure 3.21 Coomassie stained 12.5% polyacrylamide Laemmli gel showing the effect of cell lysis at pH8.5 in the presence of glutamate anions on the solubility of TnsA. Cells contained pGP1-2 and one of pT7-7, pAG411 and pSB58, as indicated below, and were grown for 3 hours after induction by heat shock. They were lysed using both lysozyme and sonication in the presence of 1M salt of the anion indicated and the lysates were then centrifuged. (The samples containing 1M glutamate ions produced a distortion on the gel.)

1	pT7-7	whole cells	
2	pAG411	whole cells	
3	pSB58	whole cells	
4	molecular weight markers (size in kDa)		
5	pT7-7	glutamate buffer	supernatant
6	pAG411	glutamate buffer	supernatant
7	pSB58	glutamate buffer	supernatant
8	pT7-7	chloride buffer	supernatant
9	pAG411	chloride buffer	supernatant
10	pSB58	chloride buffer	supernatant
11	pT7-7	glutamate buffer	pellet
12	pAG411	glutamate buffer	pellet
13	pSB58	glutamate buffer	pellet
14	pT7-7	chloride buffer	pellet
15	pAG411	chloride buffer	pellet
16	pSB58	chloride buffer	pellet

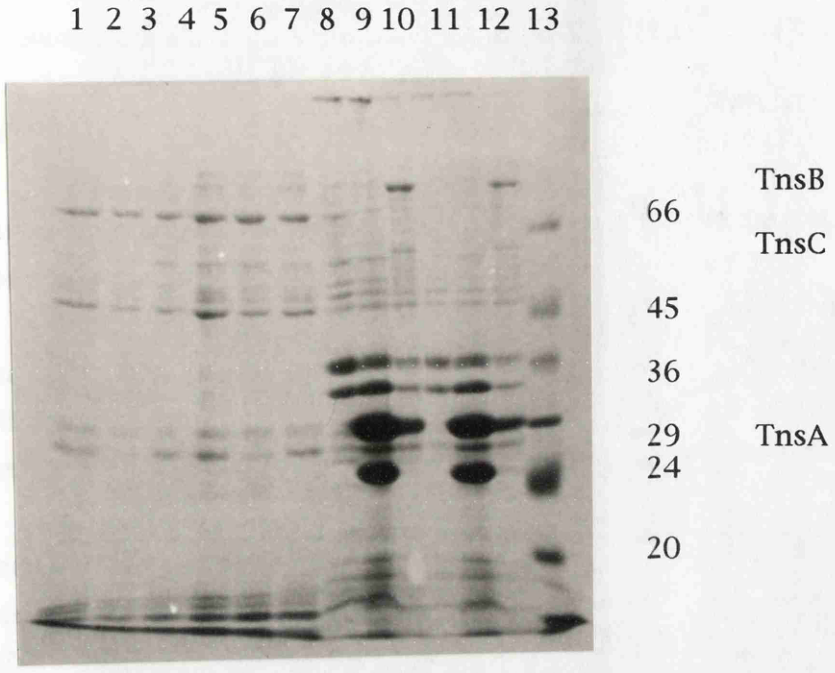


Figure 3.22 Coomassie stained 12.5% polyacrylamide Laemmli gel showing the effect of altering the anion of the lysis buffer and the pH of the resolubilisation buffer on the solubility of TnsA. These cells, containing pGP1-2 and one of pAG411, pSB58 or the vector pT7-7 were lysed in Tris glutamate pH8.5 with potassium glutamate using lysozyme treatment and sonication, and were subsequently centrifuged. The pellets were then resolubilised in 1.8M salt with Tris glutamate at the pH shown below and centrifuged once more.

1	pT7-7	pH6.5	supernatant following solubilisation of lysate pellet
2	pAG411	pH6.5	supernatant following solubilisation of lysate pellet
3	pSB58	pH6.5	supernatant following solubilisation of lysate pellet
4	pT7-7	pH9.6	supernatant following solubilisation of lysate pellet
5	pAG411	pH9.6	supernatant following solubilisation of lysate pellet
6	pSB58	pH9.6	supernatant following solubilisation of lysate pellet
7	pT7-7	pH6.5	new pellet following solubilisation of lysate pellet
8	pAG411	pH6.5	new pellet following solubilisation of lysate pellet
9	pSB58	pH6.5	new pellet following solubilisation of lysate pellet
10	pT7-7	pH9.6	new pellet following solubilisation of lysate pellet
11	pAG411	pH9.6	new pellet following solubilisation of lysate pellet
12	pSB58	pH9.6	new pellet following solubilisation of lysate pellet
13	molecular weight markers (size in kDa)		

charge of -2) and pH9.6 (at which TnsA has a predicted charge of -13) the protein was insoluble in the glutamate system (Figure 3.22). NaCl instead of KCl; the nucleases 30μg ml^{-1} DNase or 30μg ml^{-1} RNase; the detergents 1% Triton X-100 or 0.05% SDS; or 0.5 - 10mM DTT were also added to some lyses, without any apparent effect on TnsA solubility; neither did the temperature make any difference: TnsA was equally insoluble, whether lysed cells were incubated on ice or at room temperature, though it was encouraging to note that TnsA did not show obvious signs of degradation of the primary structure at room temperature. Incubation times of between 5 minutes and 36 hours were tested. Lysozyme treatment with freezing and thawing, sonication, and a combination of the two were all used; all gave a variable degree of lysis, and in all cases TnsA appeared to be insoluble.

Urea treatment, which denatures the protein, usually reversibly, has been used successfully as a step in the purification of some proteins insoluble in high salt alone, for example the Moloney murine leukaemia virus IN protein (Craigie *et al.*, 1990), in which case the protein was soluble in 4M urea. Solubility of TnsA was therefore tested in 6M urea. Cells were lysed in 150mM Tris Cl pH8.0, 1mM EDTA, 1M KCl and 6M urea and, as shown in Figure 3.19, lanes 7-11, the protein remains insoluble.

Another agent which solubilises proteins by denaturing them is guanidinium chloride. This was employed in the purification of the *Caenorhabditis elegans* Tc1 putative transposase (Schukkink and Plasterk, 1990), in which case the solubilised protein in 8M guanidinium chloride, 1M NaCl was dialysed against progressively lower concentrations of urea while maintaining the NaCl concentration, finally removing all urea from the sample to leave the renatured protein soluble in 1M NaCl. It was hoped that a similar method could be used in the purification of TnsA. The TnsA-containing pellets remaining after lysis in 1M KCl were treated with 6M guanidinium chloride, 10mM DTT and 5mM EDTA overnight at room temperature, then spun at 18000 rpm at 2° for 45 minutes. The supernatants were collected; they and the remaining pellets were run on a Laemmli gel, from which it was clear that TnsA had at last been solubilised. The next step was to try to transfer TnsA from guanidinium chloride to the slightly less severe urea without losing solubility. The supernatants were therefore dialysed against 8M urea, 1M KCl at room temperature in order to keep the urea in solution, after which, as can be seen on Figure 3.23, the TnsA remained soluble. Further dialysis against 4M urea, 1M KCl was carried out, with the TnsA remaining soluble, although direct solubilisation in 6M urea,

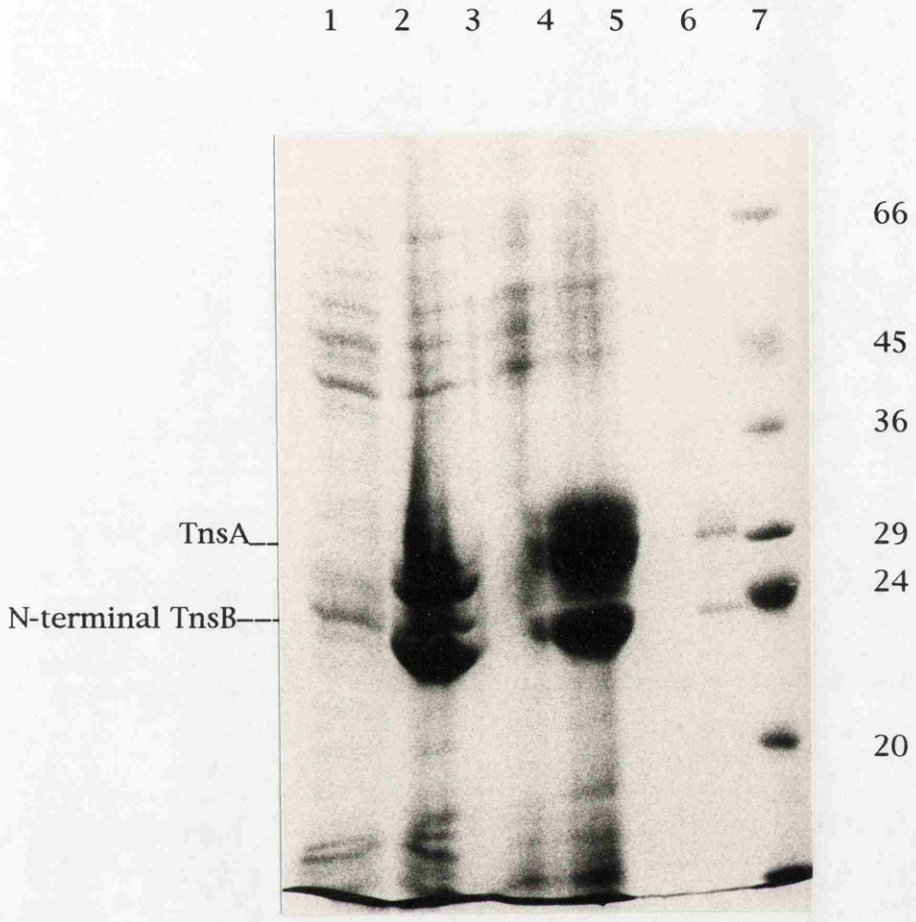


Figure 3.23 Coomassie stained 12.5% polyacrylamide Laemmli gel showing the solubilisation of TnsA in guanidinium chloride. Cells contained pGP1-2 and either pT7-7 or pAG411 in which the *tnsA* gene had been overexpressed. They were broken by incubation with lysozyme and sonication in the presence of 1M KCl, and the lysates centrifuged. The pellets were resuspended in 6M guanidinium chloride and centrifuged once more, giving the supernatants shown in lanes 1 and 2. These supernatants were then dialysed against 8M urea 1M KCl at room temperature (for reasons of solubility of urea) and then centrifuged again: the new supernatants and pellets are shown in lanes 3-6.

1	pT7-7	6M GuCl supernatant
2	pAG411	6M GuCl supernatant
3	pT7-7	8M urea supernatant
4	pAG411	8M urea supernatant
5	pT7-7	8M urea pellet
6	pAG411	8M urea pellet
7	molecular weight markers (sizes in kDa)	

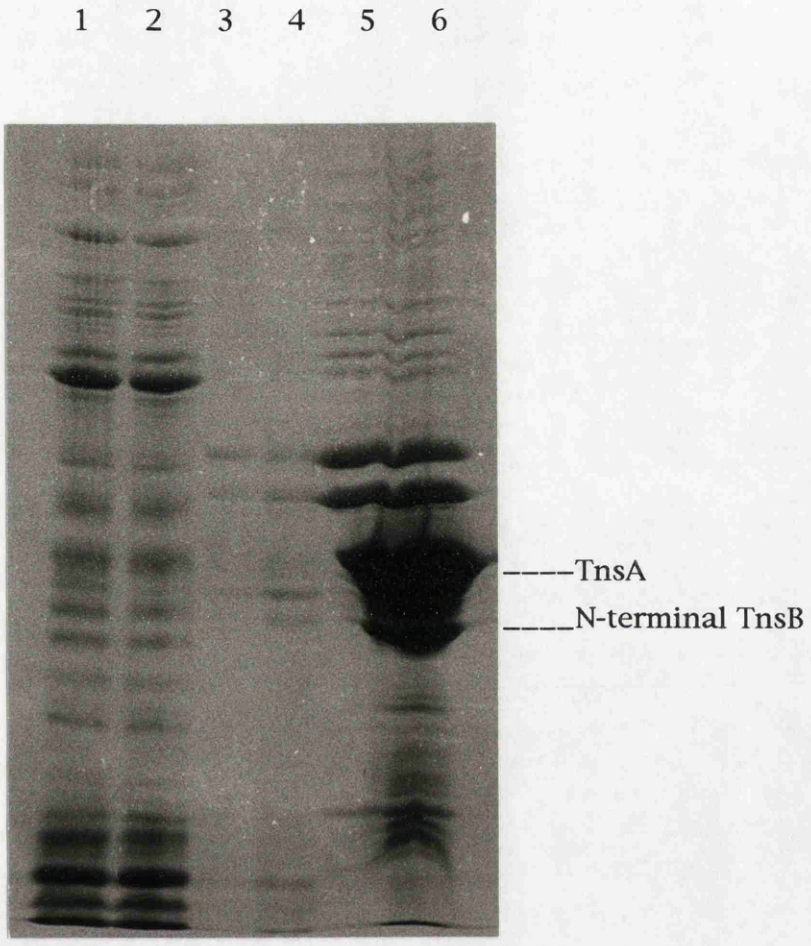


Figure 3.24 Coomassie stained 12.5% polyacrylamide Laemmli gel showing that precipitated TnsA cannot be resolubilised in the presence of 5M urea 1M KCl 25% ethandiol 1mM DTT 1mM PMSF 100mM Tris Cl pH 8.0. Cells overexpressing *tnsA* were broken with lysozyme and sonication in a pH8.5 buffer with 1M KCl. After centrifugation, the pellet fraction was resolubilised in the above buffer.

1	pT7-7	1M KCl supernatant
2	pAG411	1M KCl supernatant
3	pT7-7	pellet resuspended in 5M urea etc., new supernatant
4	pAG411	pellet resuspended in 5M urea etc., new supernatant
5	pT7-7	pellet resuspended in 5M urea etc., new pellet
6	pAG411	pellet resuspended in 5M urea etc., new pellet

1M KCl had failed. It was later determined that the KCl was not necessary in these dialyses. However, when subsequent dialysis against 2M urea, 1M KCl was performed, the TnsA precipitated.

The Tn*10* transposase protein is also intransigent, and solubility in the absence of DNA has only been maintained in 2M urea, 0.5M NaCl, 50% ethandiol (Benjamin and Kleckner, 1992). It was possible that TnsA could also be maintained in solution in the presence of ethandiol, and so the TnsA solution in 4M urea, 1M KCl was dialysed against 2M urea, 1M KCl, 25% ethandiol, 1mM DTT; the TnsA precipitated. It was considered that TnsA might be soluble if the cells were lysed in the presence of 5M urea, 25% ethandiol; cells were lysed in 75mM TrisCl pH8.5, 1mM EDTA, 1mM PMSF, 1mM DTT and spun; the resultant pellet was mixed into 5M urea, 1M KCl, 25% ethandiol, 1mM DTT, 1mM PMSF, 100mM TrisCl pH8.0, left on ice for 6 hours, then at room temperature overnight; after centrifugation, samples of the new pellets and supernatants were run on a Laemmli gel; as can be seen from Figure 3.24, TnsA remains in the pellet; it comprises, however, as judged by Coomassie staining, probably greater than 90% of the total protein in the pellet fraction.

This seemed a promising way to enrich extracts for the protein, which could then be solubilised in guanidinium chloride. The pellets shown on Figure 3.24 were therefore resuspended in 6M guanidinium chloride and, after centrifuging at 40000g, 2°C for 40 minutes, the supernatants obtained were dialysed against 5M urea, 1M KCl, 1mM DTT, 10mM Tris pH8.5. The extracts were again centrifuged, and the supernatants and pellets run on a gel; the TnsA was in the pellet fraction, which appeared very gelatinous, although previously it had remained soluble in 5M urea. It proved difficult to extract TnsA from this pellet.

(A more recent Tn*10* transposase purification protocol (R. Chalmers and N. Kleckner, unpublished) requires treatment of the crude cell lysate with 3M LiCl and 25% ethanol before solubilisation of the transposase in detergent 3-12 or Triton. As this latter method was only announced as this thesis was in press, its relevance to TnsA purification has not been evaluated, but it is certainly worth future investigation.)

As the purification of TnsA in detectable quantities was at the time proving very problematic, it was decided to wait until antibodies to the protein had been raised, after which the fate of very low concentrations of TnsA could be monitored by Western blotting.

3.9 *In vitro* transposition

It was necessary to devise a means of assaying the activity of TnsA protein when partially or completely purified. The rôle of TnsA in transposition was, and remains, unknown. It was therefore decided to attempt to develop an *in vitro* transposition system in which TnsA-containing extracts could be tested. *In vitro* transposition systems have not been obtained with the same success that has marked the study of site-specific recombination reactions. Indeed, the only known working *in vitro* transposition systems at the commencement of this work were those of bacteriophage Mu (Mizuuchi, 1983) and Tn*10* (Morisato and Kleckner, 1987). The transposition mechanisms of the *Drosophila* P element (Kaufman and Rio, 1992) and the *Caenorhabditis* Tc1 transposon (R. Plasterk, pers. comm.) are also currently being studied *in vitro*, as is the apparently mechanistically similar insertion mediated by the virally encoded IN protein of the DNA encoded by retroviruses such as HIV (Engelman *et al.*, 1991, Vink *et al.*, 1991) and Moloney murine leukaemia virus (Craigie *et al.*, 1990); see Section 1.5 for experimental details of these.

The protocol selected for Tn*7* reactions was essentially that of Mizuuchi (1983) used for the *in vitro* transposition of phage Mu and subsequently for Tn*10* (Morisato and Kleckner, 1987). Minor adjustments were made: the addition of components of the ATP regeneration system was eliminated because it was considered that firstly a high concentration of ATP could be provided and secondly dNTPs, creatine phosphate, creatine kinase and NAD were very likely to be in the crude cell extracts used initially; Morisato and Kleckner (1987) found that, in the presence of ATP, there were problems with DNA degradation, but it was considered that ATP was probably required for the Tn*7* reaction because of the homology to an ATP binding site found in the TnsC sequence. Also, PEG8000 was used in place of the PVA24000 which Mizuuchi used. Initially the plasmid pMR9 (Rogers, 1986) was used to express all the *tns* genes; it was selected because it was known that the encoded genes could complement a deletion of the *tns* region for transposition of Tn*7*-1 *in vivo*, and it was considered that where the *tns* genes were driven from their own promoter, the proteins were more likely to be produced in the correct stoicheiometry for formation of transpososomes; sonicated whole cell extracts were made and a range of dilutions of these were tested. The supercoiled plasmid pMR11 (Rogers, 1986) was the donor of the mini Tn*7*, Tn*7*-1; the supercoiled

recipient plasmid was chosen to contain the chromosomal Tn*7* *att* site. Incubations were carried out either at 4°, room temperature, 30° or 37° for a range of times from 30 minutes to 18 hours.

Three different methods of detection were tried. Firstly, the phenol/chloroform-extracted reaction mixes were run on an agarose gel which was subsequently ethidium stained; pEAL1 (Lichtenstein and Brenner, 1981) was used as the recipient because it and any transposon insertion derivative could easily be distinguished by size from both pMR9 and pMR11. Secondly, DS941 cells were transformed with the phenol/chloroform-extracted reaction mix, with selection for both the recipient plasmid, pEAL1, (Tc) and the mini Tn*7* (Cm). Thirdly, the reaction mixes were run on an agarose gel which was subsequently Southern blotted, using an internal DraI fragment of the chloramphenicol acetyl transferase gene as a probe; in this case, pMR86 was used as potential recipient instead of pEAL1 because the latter contains the remnants of a *cat* gene already. All three methods were tried several times, but no conclusive evidence for *in vitro* transposition having taken place was ever obtained.

It was later discovered (Bainton *et al.*, 1991) that *in vitro* transposition required absolutely, for reasons unknown, pre-incubation of the reaction mix in the absence of Mg^{2+} ions, whereas in the protocol of Mizuuchi, these ions were present at the setting up of the reactions; this could be a factor in the lack of success in observing *in vitro* transposition. However, the published protocol of Bainton *et al.* has been followed independently in two other laboratories (S. Bell, pers. comm.; C. Lichtenstein, pers. comm.) and in neither have the results reported by Bainton *et al.* been reproduced. Another important factor in determining the efficiency of the reaction may be the recently investigated initial incubation of TnsC, ATP, TnsD and target DNA before the other reaction components are added (Gamas and Craig, 1992), although it is implied (Bainton *et al.*, 1991) that this is not necessary for transposition to take place.

3.10 Concluding thoughts on TnsA

TnsA, when preceded by wildtype upstream sequence, appears to be very poorly translated. This phenomenon is not unique among transposons to TnsA: the transposases of, for example, IS*1*, IS*10* and Tn*3* are also subject to inefficient translation, although under the regulation of moderately

strong promoters (Kleckner, 1990b); this is viewed as a mechanism to control the level of transposition, and thus maintain the existence of the transposon without killing the host cells. The A and B genes of bacteriophage Mu are expressed from a single promoter upstream of A and, as with Tn*7*, the B protein is in excess over the A protein *in vivo*. This is believed to occur through autoregulation of translation in the case of Mu. In Tn*7*, however, there is no evidence of control by these means, and the poor ribosome binding site is probably the major factor in determining the low TnsA protein level. It is possible that translation of *tnsA* and *tnsB* is coupled, thus enhancing translation of *tnsB*, since the two reading frames overlap slightly; there is no experimental evidence for this.

The function of TnsA in transposition remains elusive. The proposed homology with the Tn*3* and IS*1* transposases (Gay *et al.*, 1986) is not strong, and the corresponding regions of these other transposases have not been characterised; that of IS*1* could be involved in strand exchange or simply play a structural rôle in protein-protein interaction; it is distinct from the identified DNA-binding domain, which is in the N-terminal half of the protein, in the InsA region. At the start of this work it was hoped that TnsA might prove to have a catalytic rôle in transposition; the low cellular level of the protein would have been compatible with catalytic turnover. There is, however no evidence for a catalytic function; further, recent work, both theoretical (Fayet *et al.*, 1990) and experimental (Engelman and Craigie, 1992) suggests that a conserved motif found in TnsB in common with many prokaryotic transposases and retroviral integrases may be implicated in DNA strand cleavage reactions catalysed by TnsB, and thus that a different function might have to be sought for TnsA.

The observed high insolubility of TnsA indicates that the molecules may have substantial hydrophobic surfaces, in spite of the predicted overall acidic nature of the protein, and computer-generated predictions of secondary structure (Figure 3.25) also show regions of hydrophobic α-helix; such surfaces could be involved in the protein-protein interactions necessary for assembly of a transpososome, and it is possible that TnsA acts as an anchor, coming into direct contact with each of TnsB, TnsC and TnsD, all of which are known to be capable of making DNA contacts. This would be supported by the observed co-precipitation of other Tns proteins with overexpressed TnsA.

It has been shown here that only a very low concentration of TnsA is required for transposition, and that this level of the protein is normally

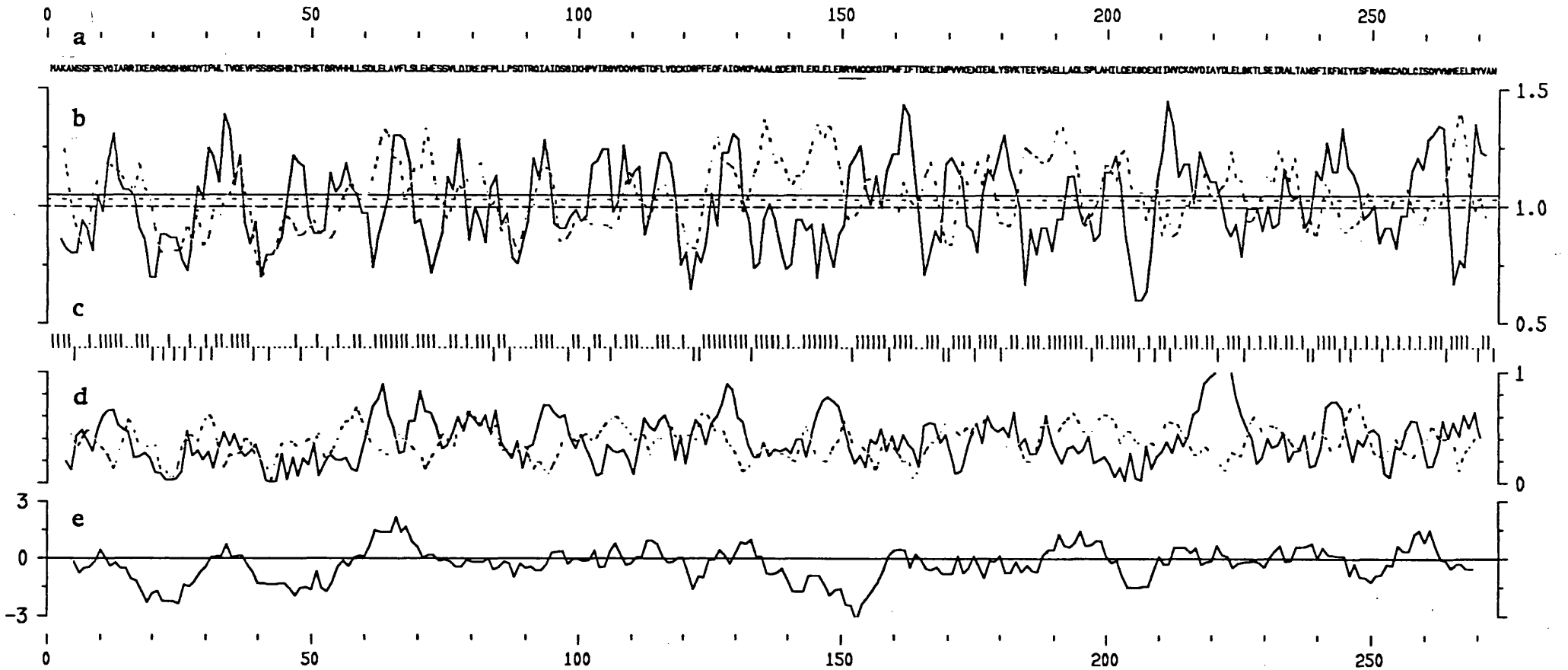


Figure 3.25 (a) Amino acid sequence of TnsA;
(b) Chou & Fasman prediction of alpha-helix (dotted) and beta-sheet (solid) structure;
(c) Chou & Fasman alpha helix forming (above) and breaking (below) residues;
(d) Eisenberg's helical hydrophobic moment, showing the probability of amphiphilicity in alpha helix (dotted) and beta sheet (solid);
(e) Kyte & Doolittle hydropathy, averaged over a window of 9 residues.

present in cells regardless of the status of the promoter, although the reaction does not occur in the complete absence of TnsA. It is therefore unlikely that TnsA concentration is rate-limiting for the transposition reaction. This might suggest that TnsA is not required at an early stage of the reaction; however, it has been demonstrated (Bainton *et al.*, 1991) that all of TnsA, TnsB, TnsC, and possibly also TnsD are required for the initial cleavage of the ends of the transposon.

In spite of the very low intracellular level of TnsA normally observed, overexpression of *tnsA* has no obvious deleterious effect on the host cells; however it may be that such effects are relieved by isolating the protein in inclusion bodies. Other proteins produced from T7 expression systems have exhibited a high degree of insolubility and lack of activity, and it has been suggested (M. Chandler, unpublished) that, where transcription occurs at such a high rate, the nascent proteins may be unable to fold correctly, thus altering their solubility and accounting for loss of activity. *In vivo* transposition experiments were conducted by S. Bell (in press) with a mini-Tn*7* and two different *tns*-encoding plasmids: pSB58, which is a translational fusion of *tnsABCDE* to the T7 gene 10 upstream region of pT7-7, and pSB84, which is a transcriptional fusion of the same Tn*7* sequence to the T7 gene 10 upstream region of pT7-5. Induction of the T7 promoter of pSB58 leads to a vast excess of TnsA over TnsB, as observed also with pAG411; induction of the promoter in pSB84 results in approximately the same level of TnsB but much less TnsA, as observed with wildtype Tn*7*. Transposition frequency was higher with pSB84 than with pSB58, suggesting that a very high level of TnsA may inhibit transposition by some mechanism, perhaps by sequestering all the other Tns proteins in a structure inaccessible to DNA.

In summary, TnsA may have one of several functions: it may be itself an enzyme, or it may have a structural rôle in transpososome formation; it may be a chaperonin, allowing other Tns proteins to adopt the correct conformation for transposition to take place; it may be involved in protein-protein interactions or DNA-protein interactions, although there is no evidence for the latter, and the former has not been tested. Whatever its function, it is clearly essential for transposition.

There is scope for much work still to be carried out on TnsA in the future.

CHAPTER 4

TnsB: DISSECTION OF FUNCTION

4.1 Introduction

The work described in this chapter relates to *tnsB*. As elaborated below, the protein product, TnsB, has been studied in several laboratories (Rogers *et al.*, 1986; McKown *et al.*, 1987; Arciszewska and Craig, 1991; Arciszewska *et al.*, 1991; Tang *et al.*, 1991), but it is not yet clear precisely how TnsB interacts with DNA or functions in conjunction with other proteins in transposition. Further studies of the TnsB protein were therefore undertaken. The majority of the investigations reported here address the hypothesis that Tn*7* transposition is regulated by some inhibitory action of TnsB peptides which are products of proteolysis; three different techniques involving truncated TnsB proteins were employed: gel binding assays, studies of promoter repression, and measurement of *in vivo* transposition frequencies. Site-directed mutagenesis of TnsB was used to clarify the position of the DNA binding site on the protein. Finally, novel methods of TnsB purification were investigated.

Analysis of the sequence (Flores *et al.*, 1990) of the *tnsB* gene shows an open reading frame encoding a protein of a maximum of 702 amino acids, with a corresponding size of 81kDa and pI of 8.7; this open reading frame is preceded by a relatively strong ribosome-binding site with the sequence:

GAGGAGUUGCGC

where underlining indicates bases which can pair with the 3' end of the 16S rRNA molecule. There is no indication from the sequence of a promoter directly preceding *tnsB*, and Waddell and Craig (1988) observed insertions in *tnsA* to be polar on *tnsB* activity, suggesting that *tnsA* and *tnsB* are co-transcribed. Ekaterinaki (1987), however, detected slight transcriptional initiation in this region, and so it is possible that there is a very weak *tnsB* promoter, and that the assays of Waddell and Craig were not sensitive enough to detect its activity. Gel retardation (reviewed by Lane *et al.*, 1992) has been used to show that TnsB can bind to 22bp motifs found at the left end (LE) and right end (RE) of the transposon, as described in Section 1.4.2.

The innermost motif, R4, at the Tn*7* RE coincides with the -35 box of the presumed promoter, and so TnsB most probably regulates its own transcription. The substantially higher concentration of TnsB than TnsA observed in cells encoding *tnsAB* under the control of either the wildtype promoter or a cloned *lac*, *tac* or T7 promoter must be post-transcriptional in origin, as discussed in Chapter 3.

Previous work with *tnsB* had established that a protein product of the predicted size could be seen on Laemmli gels of radioactively labelled proteins in minicells derived from cells containing a *tnsB*-encoding plasmid (Rogers, 1986), and of Coomassie-stained whole cells (Morrell, 1990; Arciszewska *et al*, 1991). Gel binding assays using crude cell extracts were used to demonstrate that a constituent of extracts from *tnsB*-encoding cells bound specifically to the Tn*7* right end (Ekaterinaki, 1987) and left end (Arciszewska and Craig, 1991), and to isolated 22bp motifs (Morrell, 1990). A protein of approximately 80kDa with this binding activity was purified partially by Morrell (1990) and to a greater degree by Arciszewska (Arciszewska *et al.*, 1991); in the former case nine retarded bands were observed on a binding gel using a DNA fragment containing only a single 22bp motif; in the latter case four retarded bands were reported in the presence of the complete right end sequence (containing four motifs). One objective of this work was to investigate the discrepancy between these sets of observations.

A large number of DNA binding proteins bind to their cognate site as dimers; in this case, the DNA binding site generally has dyad symmetry, and one monomer of the protein is associated with each half of the site. DNaseI and hydroxyl radical footprinting of the binding of partially purified TnsB to Tn*7* RE and LE fragments by Morrell (1990) and Arciszewska and Craig (1991) could not determine whether TnsB was binding to each individual 22bp motif as a monomer (as is believed to happen in the case of bacteriophage Mu, whose ends also have a series of 22bp motifs) or as a dimer, and there was disagreement over whether the TnsB form found in solution was monomer (Morrell, 1990) or dimer (Arciszewska *et al.*, 1991). The motifs show limited dyad symmetry, as shown below:

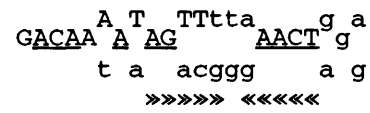


Figure 4.1 The consensus 22bp motif. Completely conserved bases are underlined. Upper case indicates that this base occurs in more than 50% of LE and RE motifs at this position. Chevrons mark the region of partial dyad symmetry.

The four retarded bands seen by Arciszewska and Craig (1991) in a binding assay with DNA containing four motifs can most simply be interpreted by hypothesising that one monomer of TnsB binds to each motif. It is also possible that a dimer binds to each site, but that individual molecules in a dimer either do not bind alone, or that binding of a dimer exhibits a high degree of cooperativity, such that monomer binding is not easily detected. If cooperativity were present, the multitude of additional retarded complexes seen on binding gels when TnsB is produced in a protease-competent strain such as DS941 might be due to a breakdown in cooperativity of truncated TnsB molecules, and not solely to a large number of different proteolytic fragments. Cooperativity in binding is addressed further in Section 4.6.4.

4.2 TnsB fragments: possible origins

The complete *tnsB* coding sequence has been cloned into several plasmid vectors, among which pMR207 has been much used. pMR207 is a *tnsB* expression plasmid based on the vector pMR78 (see Figure 4.3), and was used by M. Rogers (unpublished). When whole cells of DS941 containing pMR207, and thus expressing *tnsB*, were run on a Laemmli gel, several bands were seen which were apparently specific to the *tnsB*-encoding plasmid (Figure 4.2). It would have been instructive at this point to use anti-TnsB antibodies to identify precisely which observed bands were in fact due to fragments of TnsB, but unfortunately no such antibodies were available. The appearance and position of these bands was dependent

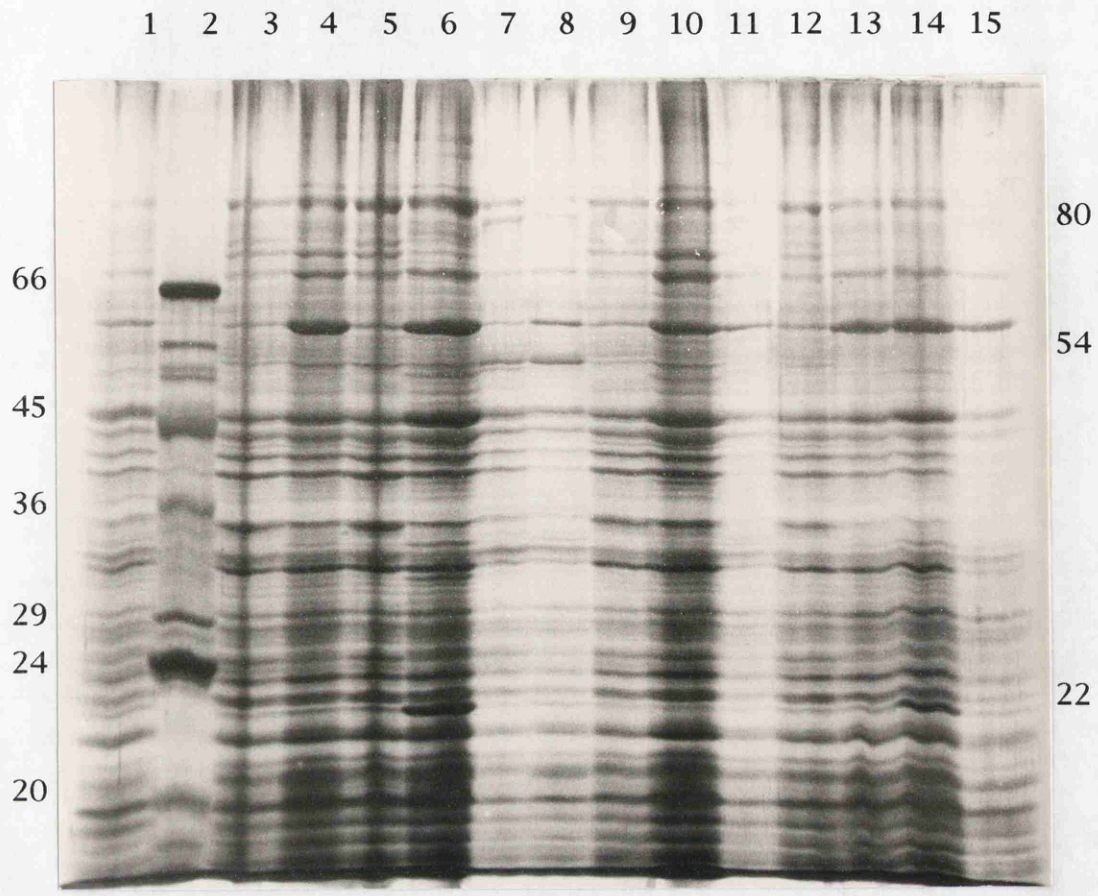


Figure 4.2 Silver stained 12.5% polyacrylamide Laemmli gel showing the production in whole DS941 cells of peptides with different electrophoretic mobilities following IPTG induction for $2^3/_4$ hours of *tnsB* from pMR207 and pMR205 (complete *tnsB*) and pAG205 (amino terminal 23kDa) under different growth conditions described below. Cells containing pMR78 (the relevant expression vector) were also grown as a control. Sizes above are given in kDa.

1	pMR78	100ml isosensitest broth in 250 ml flask.	IPTG
2	molecular weight markers		
3	pMR207	100ml isosensitest broth in 250 ml flask.	No IPTG
4	pMR207	100ml isosensitest broth in 250 ml flask.	IPTG
5	pMR207	100ml isosensitest broth in 2 l flask.	No IPTG
6	pMR207	100ml isosensitest broth in 2 l flask.	IPTG
7	pMR207	100ml M9 minimal broth in 2 l flask.	No IPTG
8	pMR207	100ml M9 minimal broth in 2 l flask.	IPTG
9	pAG205	100ml isosensitest broth in 250 ml flask.	No IPTG
10	pAG205	100ml isosensitest broth in 250 ml flask.	IPTG
11	pAG205	10ml M9 minimal broth in 250 ml flask.	IPTG
12	pMR205	100ml isosensitest broth in 250 ml flask.	No IPTG
13	pMR205	100ml isosensitest broth in 250 ml flask.	IPTG
14	pMR205	100ml isosensitest broth in 2 l flask.	IPTG
15	pMR205	100ml M9 minimal broth in 2 l flask.	IPTG

on the conditions under which the cells were grown; when M9 minimal liquid medium rather than isosensitest broth was used, proteins running with a mobility corresponding to a size of approximately 22kD and 52kD were seen (Figure 4.2 and M. Burke, pers. comm.), and the smaller of these proteins was also seen when cells were grown in isosensitest broth with increased aeration. It was considered possible that these proteins were in some way involved in a response of Tn*7* to changes in environmental conditions. It had earlier been suggested (Hauer and Shapiro, 1984) that Tn*7* transposition was stimulated by certain stress conditions, and it was possible that the observations noted here might be another manifestation of the phenomenon.

Three possible explanations for the appearance of additional proteins on Laemmli gels were that:

- translation of *tnsB* was initiating at several points;
- proteolysis of the full length protein was occurring;
- the observed proteins were not fragments of TnsB, but were other plasmid or host-encoded proteins, the production or stability of which was influenced by *tnsB*.

There are, according to the sequence, possible internal sites for translational initiation, as discussed by Flores *et al.* (1990), though none of those (with initial codon AUG, GUG or UUG) identified either by Flores *et al.* or in this work is preceded by any identifiable ribosome binding site; and thus (by analogy with *tnsA*; see Chapter 3) it is unlikely that peptides initiating at any of these internal points would be present at a high enough concentration to be clearly visible on a Laemmli gel of whole cells.

It is certainly likely that some of the apparently additional proteins observed on Laemmli gels were host-encoded, but this was not investigated further.

Several prokaryotic transposases are highly susceptible to proteolytic action: for example the IS*903* transposase is subject to *lon*-dependent instability (Derbyshire *et al.*, 1990). Proteolysis of TnsB appeared to be a plausible explanation of the origin of at least some of the extra bands on Laemmli gels, since extra bands indicating specific retardation of the Tn*7* RE were also seen on autoradiographs of binding gels using crude extracts or only partially purified TnsB (Morrell, 1990), compared with the purer protein used by Arciszewska (Arciszewska *et al.*, 1991). If proteolysis were indeed occurring, the question would arise as to whether this was

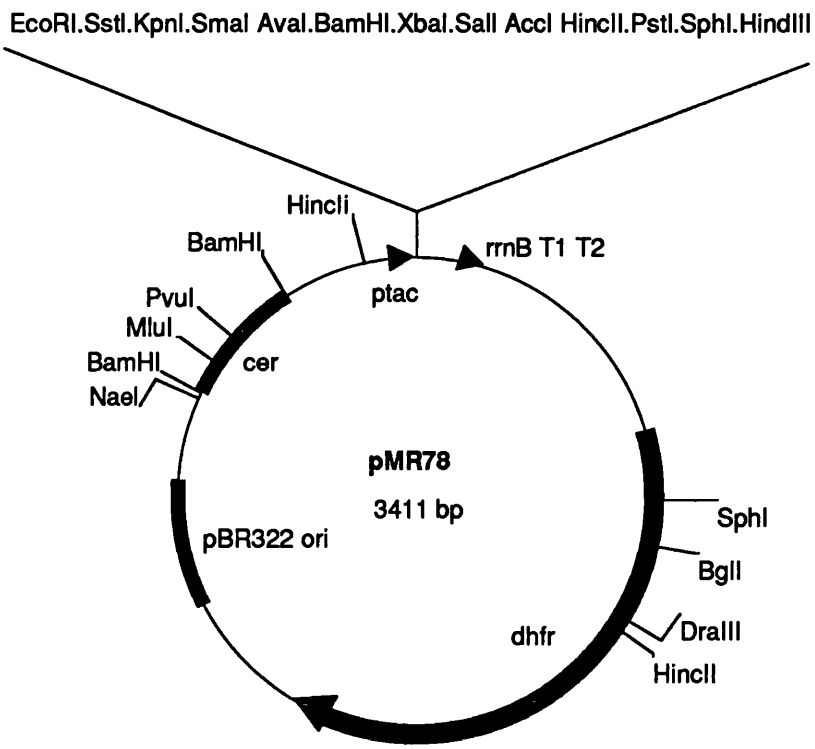


Figure 4.3 Structure of the expression vector pMR78, constructed by M. Rogers (unpublished). *cer* indicates the region involved in site-specific recombination in the maintenance of the monomer state in the naturally occurring plasmid ColE1. *dhfr* is the trimethoprim resistance gene of the naturally occurring plasmid R388. *rrnB* T_1T_2 is a region of strong ribosomal RNA transcription terminators found in *E. coli*.

The following plasmids were constructed by insertions of Tn*7* sequence into pMR78:

pMR204	pMR207	
pAG201	pAG202	pAG203
pAG204	pAG205	pAG206
pAG208	pAG209	pAG210
pAG211		

happening *in vivo* or, after harvesting the cells, *in vitro*.

If proteolysis were occurring *in vivo*, it might be possible, depending on the proteolysis mechanism involved, to reduce or completely avoid it by using an *E. coli* strain which was protease-deficient. Unfortunately such a strain was not available until this work was nearing completion; experiments using TnsB obtained in a *htpR* background are described in Section 4.6.4.

TnsB proteolysis *in vitro* could be reduced by the use of protease inhibitors such as PMSF and benzamidine, both of which are inhibitors of serine proteases, and EDTA, which blocks the action of proteases requiring divalent cations. All these were in fact used in the crude cell extracts which were used in binding experiments (the TnsB-specific binding activity does not require the presence of any divalent cation, and so EDTA can safely be added to a high concentration).

4.3 Preliminary gel-binding assays of TnsB function

The first studies on the binding of TnsB to the Tn*7* ends were by Ekaterinaki (1987) and Rogers (unpublished), both employing the method described in Chapter 2, using crude extracts of cells in which *tnsB* had been overexpressed. Gel binding was chosen instead of filter binding because the latter cannot distinguish between complexes of different stoicheiometry or conformation which may be formed; it has also been shown (Fickert and Müller-Hill, 1992) that not all protein-DNA complexes bind with equal affinity to nitrocellulose filters, and so quantification of binding is not possible by these means. Crude cell extracts had been successfully employed by Strauss and Varshavsky (1984) in gel binding experiments; a high concentration of unlabelled carrier DNA (such as the sheared salmon sperm which was used here) can be used as a sink for nonspecific binding to DNA by other proteins in the extracts, so that only binding by molecules which specifically recognise the DNA sequence under investigation is detected.

The extracts used by Rogers were from cells harbouring the plasmid pMR207, in which *tnsB* is under the control of the *tac* promoter of the vector plasmid pMR78, and much of the cloning described in this chapter is based on pMR78. The structure of pMR78 is shown in Figure 4.3.

Rogers also constructed the plasmid pMR204, which encodes *tnsA* and 542bp of the 5' end of the coding region of *tnsB*. It was observed by him

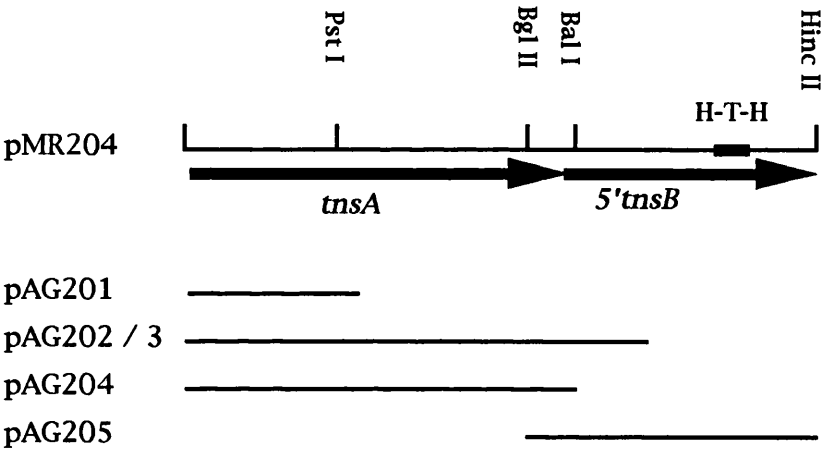


Figure 4.4 Deletions of the Tn *7* sequence in pMR204, shown at the top. The 5' end of the Tn *7* coding sequence in pMR204 was created by Exonuclease III deletion of the Tn*7* RE; the 3' end of the Tn *7* sequence in pAG201, pAG202 and pAG203 was also created by Exonuclease III deletion. The sequence in pAG204 terminates at the Bal I site marked above. The Tn *7* sequence in pAG205 is between the Bgl II and Hin *c* II sites shown.

H-T-H = putative helix-turn-helix; see Section 4.10 for further investigation of this.

1 2 3 4 5 6 7 8 9 10 11

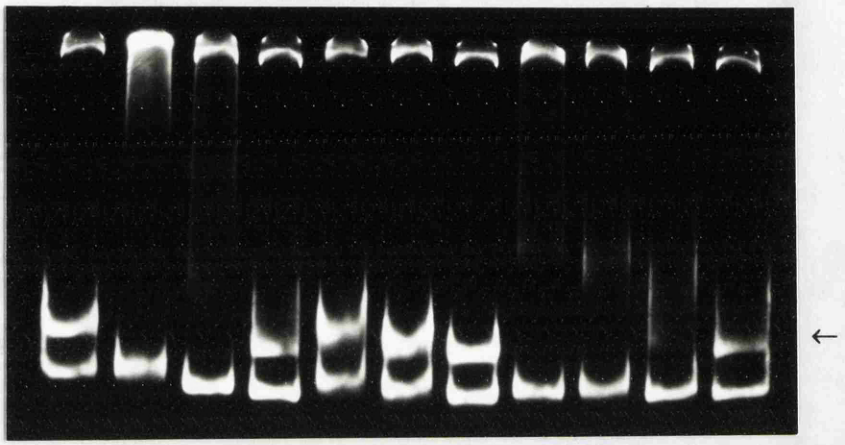


Figure 4.5 Gel retardation assay to show the specific binding activity associated with crude cell extracts expressing *tns* genes encoded by the plasmids pMR204, pAG204, pAG205 and pMR78 (vector with no insert); these inserts are shown diagrammatically on Figure 4.4. The ^{32}P labelled DNA is pNE200 (see Figure 4.11) cut with Hin*d*III, EcoRI and NdeI; the DNA fragment containing the Tn*7* sequence is the second one up, as indicated with an arrow. The reactions were carried out, and the gel poured and run in TE, as described in Chapter 2. The protein extracts used were as listed below.

1	pMR78 extract	2^{-3} dilution of extract stock
2	pMR204 extract	2^{-3}
3	pMR204 extract	2^{-4}
4	pMR204 extract	2^{-5}
5	pAG204 extract	2^{-3}
6	pAG204 extract	2^{-4}
7	pAG204 extract	2^{-5}
8	pAG205 extract	2^{-3}
9	pAG205 extract	2^{-4}
10	pAG205 extract	2^{-5}
11	pAG205 extract	2^{-6}

1 2 3 4 5 6 7 8 9

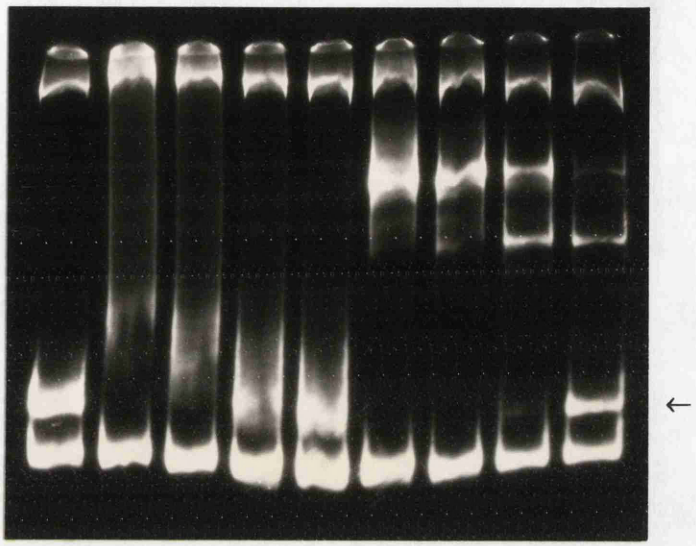


Figure 4.6 Gel retardation assay to show the specific binding activity associated with crude cell extracts expressing *tns* genes encoded by the plasmids pMR207, pAG205 and pMR78 (vector with no insert); these inserts are shown diagram-matically on Figure 4.9. The ^{32}P labelled DNA is pNE200 (see Figure 4.11) cut with Hin*d*III, EcoRI and NdeI; the DNA fragment containing the Tn*7* sequence is the second one up, as indicated with an arrow. The reactions were carried out, and the gel poured and run in TE, as described in Chapter 2. The protein extracts used were as listed below.

1	pMR78	2^{-3} dilution of extract
2	pAG205	2^{-3}
3	pAG205	2^{-4}
4	pAG205	2^{-5}
5	pAG205	2^{-6}
6	pMR207	2^{-3}
7	pMR207	2^{-4}
8	pMR207	2^{-5}
9	pMR207	2^{-6}

(unpublished) that extracts from cells containing this plasmid also exhibited binding activity to the right end of Tn*7*; at that time it was thought likely that the binding was due to TnsA, since at most only a quarter of the length of the TnsB protein was present. The initial objective of the work described here was to determine which of the two possible proteins was responsible directly or indirectly for the observed binding.

4.4 Construction of clones producing amino-terminal fragments of TnsB

Since the vector pMR78 had been used successfully for the initial binding studies, with adequate protein produced, and because experimental controls could thereby be minimised, it was decided to continue using it for cloning. As mentioned above, the first task was to subdivide the Tn*7* sequence encoded by pMR204 in order to find out where the amino acid sequence implicated in DNA binding lay. To this end a series of deletions of the Tn*7* sequence of pMR204 was made: pAG201, pAG202 and pAG203 (the latter two very similar) were exonucleaseIII deletions of the 3' end of the Tn*7* sequence; pAG202 and 203 removed approximately 300bp of the *tnsB* sequence, whereas pAG201 removed all the *tnsB* sequence and approximately 380bp of the 3' end of *tnsA*. Construction of pAG204 and pAG205 involved the endonucleases BalI and BglII respectively in cloning separately *tnsA* (pAG204) and the 5' end of *tnsB* (pAG205). The Tn*7* sequences cloned in these plasmids are shown schematically in Figure 4.4.

Gel binding assays using the same reaction conditions as those used by Rogers showed (Figure 4.5, Figure 4.6) that the only one of the above five plasmids to be associated with any binding activity to the Tn*7* right end (the Hin*d*III-EcoRI fragment of pNE200; see Figure 4.11) was pAG205. This implied that binding activity was specific to the N-terminal end of TnsB, rather than TnsA. Further attempts, using different reaction and gel conditions, were made at detecting TnsA-specific binding but none was successful. Observation of the N terminal-TnsB-specific binding was improved by running gels in Tris-glycine pH9.4 (Bednarz *et al.*, 1990), rather than the TE used by Rogers; compare Figures 4.6 and 4.7. In TE gels, binding was detected by reduction in intensity of the band corresponding to the unbound DNA, but no distinct band representing a protein-DNA complex could be seen; in Tris-glycine gels, retardation of bound DNA could be seen directly. This dependence of visualisation of retarded protein-DNA

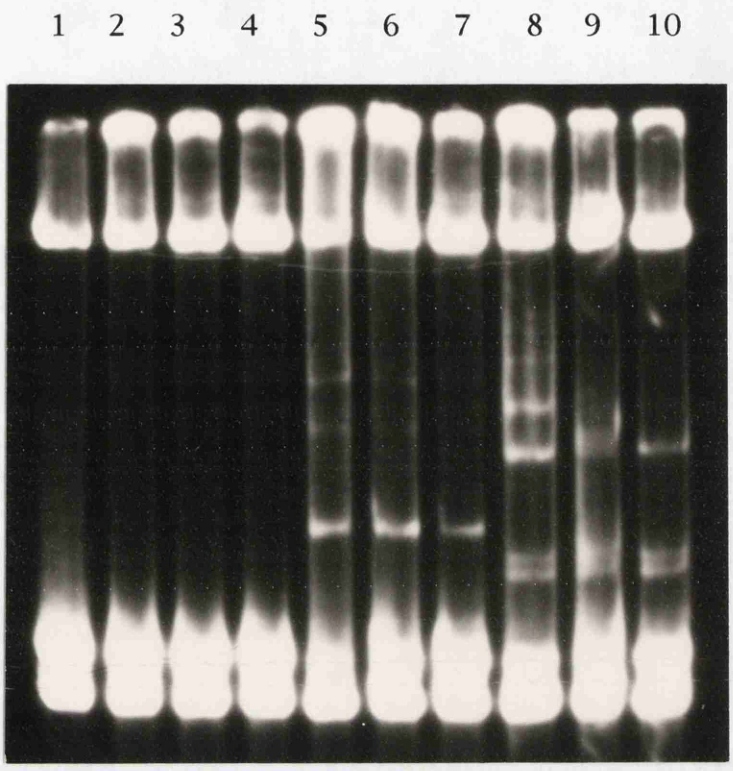
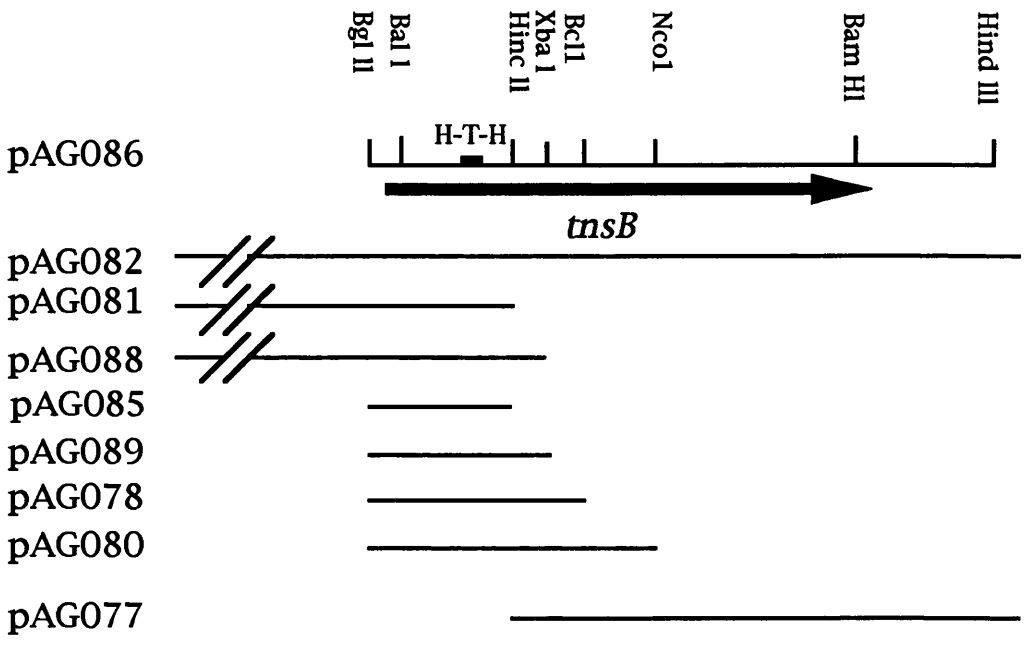


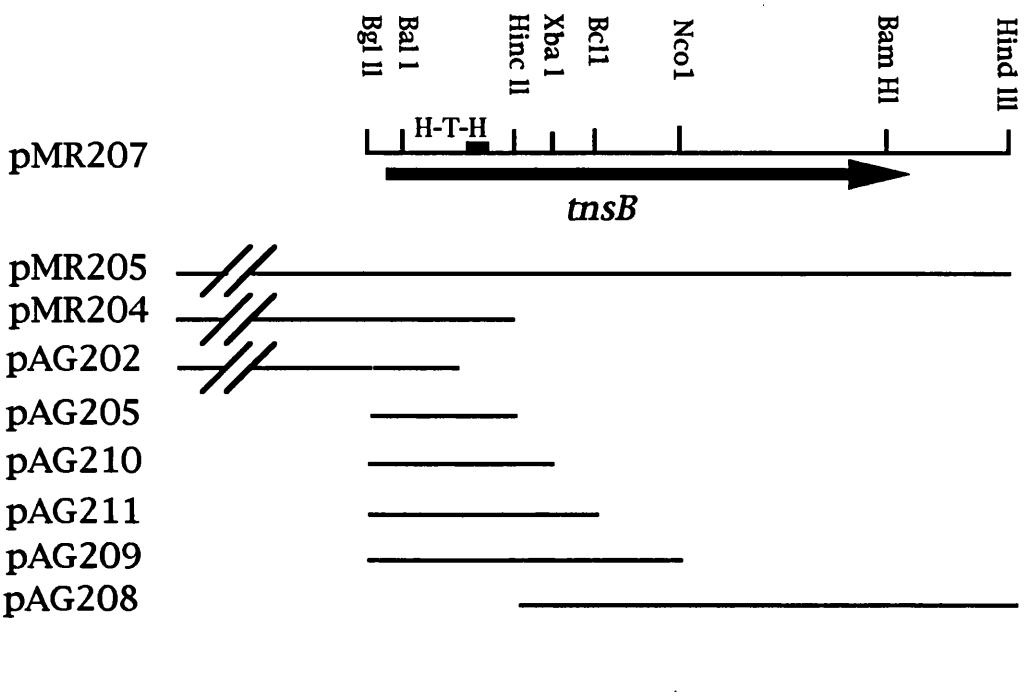
Figure 4.7 Gel retardation assay. This gel was made and run in Tris glycine pH9.4. The labelled DNA was as in Figure 4.6 and the reactions were carried out as described in Chapter 2. The protein extracts used were as listed below.

1	No protein extract; dilution buffer only added	
2	pMR78	2^{-2} dilution of extract
3	pMR78	2^{-4}
4	pMR78	2^{-6}
5	pAG205	2^{-2}
6	pAG205	2^{-4}
7	pAG205	2^{-6}
8	pMR207	2^{-2}
9	pMR207	2^{-4}
10	pMR207	2^{-6}



H-T-H = putative helix-turn-helix motif

Figure 4.8: TnsB-producing plasmids based on the vector pUC18. Construction of pAG077 is described further in Section 4.11.



H-T-H = putative helix-turn-helix motif

Figure 4.9: TnsB-producing plasmids based on the vector pMR78. Construction of pAG208 is described in Section 4.11.

complexes on the electrophoresis buffer has also been noted with the reaction between bacteriophage Mu repressor protein and its operator binding sites (Alazard *et al.*, 1992), where again Tris-glycine was the buffer of choice. It was clear from the lack of activity with pAG202 and pAG203 that sequences between 1170 and 1485bp from the Tn*7* right end were necessary for this binding activity; this region included the putative helix-turn-helix identified by Flores *et al.* (1990).

Further fragments of *tnsB* were cloned downstream of the *lac* and *tac* promoters in the vectors pUC18 and pMR78 respectively, as shown in Figures 4.8 and 4.9, thereby giving two series of plasmids encoding the same Tn*7* sequences but with different antibiotic resistance markers. Those based on pMR78 were used in binding gels as described in Section 4.6, whereas those based on pUC were used for studies of promoter repression (Section 4.7) and transposition inhibition *in vivo* (Section 4.9). The sizes of the predicted TnsB peptides produced from pMR78-based plasmids are listed in Figure 4.20; as will become clear later, the actual peptides observed were in some instances smaller than this.

Each of the cloned TnsB fragments was tested for its ability to complement deletion of *tnsB* in an *in vivo* assay of transposition of chromosomal Tn*7*-1 to pEN300, as described in Chapter 2. No transposition was detected in any of the assays, suggesting that sequence downstream of the NcoI site is essential for a transposition function other than binding to the 22bp motifs, perhaps interaction with another protein species or DNA cleavage or strand exchange. It had previously been shown (Rogers, 1986) that TnsB encoded by a gene truncated at the BamHI site has the same transposition activity as the wildtype protein; thus only the carboxy-terminal seven amino acids are known not to be necessary for transposition.

4.5 The possible rôle of TnsB amino-terminal fragments *in vivo* in transpositional regulation

Regulation of transposition of prokaryotic transposable elements is brought about in several ways, as discussed in Chapter 1. One method used by a number of elements apparently of diverse origin is the production of a truncated form of the transposase protein, which is able to bind to *cis*-acting sequences at the ends of the element in place of the transposase, but does not have the capability to catalyse the cleavage or strand transfer

reactions. In the case of IS*1* and several other insertion elements, as discussed in Chapter 1, this truncated protein is the N-terminal InsA, in which the translational frameshift necessary for transposase production does not occur. The IS*50* transposase coding region contains internal transcriptional and translational start sites, and the resultant amino-terminal deleted protein, p2, acts as a transpositional repressor by binding to DNA and blocking the action of the wildtype transposase (Yin and Reznikoff, 1988).

It is conceivable that a similar mechanism operates *in vivo* in regulation of Tn*7* transposition: one or more of the truncated TnsB protein species apparently produced *in vivo* could bind to the ends of Tn*7* and inhibit transposition through an inability to interact with other parts of the transposition complex to capture the target molecule or catalyse cleavage and strand transfer; transcription from the *tns* promoter might simultaneously be blocked by binding of such a peptide to the R4 motif, which overlaps the presumed -35 box of the promoter, thus further reducing the probability of transposition.

This hypothesis was investigated in three ways: by determining the ability of different fragments of TnsB to bind, singly or in combination, to the different motifs of the left and right ends (Section 4.6); to study the effect of these peptides, when supplied in excess, on Tn*7* transposition *in vivo* (Section 4.9); and to find out whether transcription could be inhibited by the binding of these peptides to the R4 motif (Section 4.7).

4.6 Binding of amino-terminal fragments of TnsB to the Tn*7* left and right ends *in vitro*

Each of the N-terminal fragments of TnsB constructed was tested for its ability to bind to 22bp motifs of the left and right end of Tn*7*, using gel retardation assays. Each *tnsB* deletion was cloned into the same vector, pMR78 (M. Rogers, unpublished), shown in Figure 4.3, to standardise results; this plasmid, derived from pKK223-3, contains the strong *rrnB* ribosomal RNA transcription terminators immediately downstream of the polylinker sequence; it can therefore be certain that truncated proteins translated from expressed cloned genes do not have significant carboxy-terminal extensions. For the purposes of gel binding assays, crude extracts of cells (which were the strain DS941 unless otherwise stated) containing the overproduced protein were always made in the same buffer

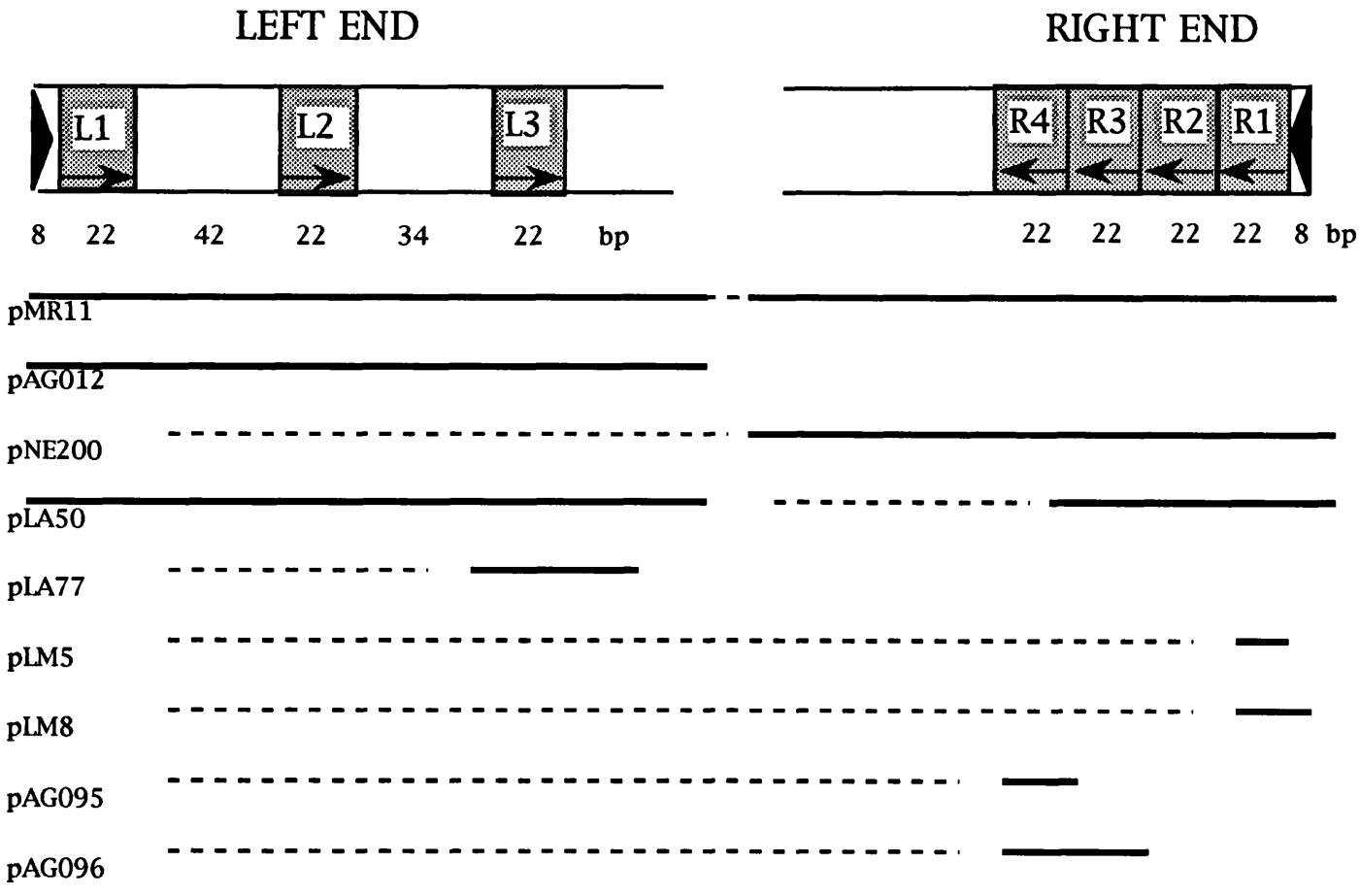


Figure 4.10 Tn*7* end motif fragments used in binding reactions

a The first 240 bases of the right end of *Tn7*, shown on one strand only.

```
                R1                       R2                          R3                       R4
TGTGGGCGGA CAATAAAGTC TTAAACTGAA CAAAATAGAT CTAAACTATG ACAATAAAGT CTTAAACTAG ACAGAATAGT

TGTAAACTGA AATCAGTCCA GTTATGCTGT GAAAAAGCAT ACTGGACTTT TGTTATGGCT AAAGCAAACT CTTCATTTTC

TGAAGTGCAA ATTGCCCGTC GTATTAAAGA GGGGCGTGGC CAAGGGCATG GTAAAGACTA TATTCCATGG CTAACAGTAC
```

b DNA fragments used ^{32}P labelled in binding gels.

```
pNE200                                          1
AGCTTGGCTG CAGGTCGATG AGTTCCCGGA CTTCTTGTGT GGGCGGACAA TAAAGTCTTA AACTGAACAA AATAGATCTA AACTATGACA

ATAAAGTCTT AAACTAGACA GAATAGTTGT AAACTGAAAT CAGTCCAGTT ATGCTGTGAA AAAGCATACT GGACTTTTGT TATGGCTAAA

GCAAACTCTT CATTTTCTGA AGTGCAAATT GCCCGTCGTA TTAAAGAGGG GCGTGGGGGA A
```

```
pLA50                                                                             1
AATTCAGCCG CGTAACCTGG CAAAATCGGT TACGGTTGAG TAATAAATGG ATGCCCTGCG TAAGCGGGTG TGGGCGGACA ATAAAGTCTT

AAACTGAACA AAATAGATCT AAACTATGAC AATAAAGTCT TAAACTAG
```

pLM5 9
AATTCGAGCT CGGTACCCGG GGATCCTCTA GAGACAATAA AGTCTTAAAC TGAA*GTCGAC CTGCAGGCAT GCAA*

pLM8 1
*AATTCGAGCT CGGTACCCGG GGATCCTCTA CTAGA*TGTGG GCGGACAATA AAGTCTTAAA CTGAA*GTCGA CCTGCAGGCA TGCAA*

pAG095 67
*AATTCGAGCT CGGTACCCGG GGATCCT*CTA GACAGAATAG TTGTAAACTG AA*GTCGACCT GCAGGCATGC AA*

pAG096 50
AATTCGAGCT CGGTACCCGG GGATCCTCTA GACAATAAAG TCTTAAACTA GACAGAATAG TTGTAAACTG AA*GTCGACCT GCAGGCATGC*

AA

Figure 4.11. Sequences of the Tn*7* right end fragments used in gel binding experiments. The Tn*7* sequence reads from the right end inwards and sequence extraneous to Tn*7* is italicised. The position of these sequences is indicated schematically on Figure 4.10, and in **b** above with the number of the first base encoded. 22bp motifs are indicated throughout by alternate underlining and overlining and are also numbered in **a**. The plasmid from which each fragment is obtained is indicated above the sequence. Only the top strand reading from the right end is shown.

(containing 1M KCl) in order to standardise results, although this was probably not necessary for solubility of all the peptides tested. While expression in a protease deficient strain would have been preferable, such a strain did not become available until the conclusion of this work (see Section 4.6.4). The high salt concentration in the binding reactions, while a necessary consequence of the TnsB solubilisation conditions, also helped to clarify the protein-DNA binding by disrupting non-specific interactions between molecules. The lack of salt in the electrophoresis buffer reduced the probability of preformed complexes dissociating while the gel was running. Glutamate buffers and potassium glutamate were also tested for binding reactions, as these were considered to be closer to physiological conditions than the Tris Cl and KCl routinely used, but no improvement in the clarity or stability of binding was observed (data not shown).

The different DNA fragments used ^{32}P end-labelled in gel binding assays are shown schematically in Figure 4.10, and the sequences of the corresponding RE fragments are listed in Figure 4.11. The Tn*7* sequences in pLM5, pLM8, pAG095 and pAG096 were obtained by synthesis of double-stranded oligonucleotides with added terminal restriction sites; other sequences were subcloned into plasmids after digestion with restriction enzymes.

4.6.1 Can high concentration TnsA interact with TnsB in binding to the RE?

This was tested by mixing extracts of DS941(pMR207) and DS941(pGP1-2, pAG411); this last plasmid, as described in Chapter 3, encodes *tnsA* under the control of a T7 promoter and, upon induction, a substantial proportion of the cellular protein is TnsA. In cells containing wildtype Tn*7*, or where plasmid-encoded *tnsA* and *tnsB* are coexpressed from the *tns* promoter, very little TnsA is produced, and it cannot be detected on a Laemmli gel, and so it was conceivable that TnsA-TnsB interactions did occur, but at too low a frequency to be observed on a binding gel; it was considered that some interaction might be detectable by an alteration in retardation pattern were more TnsA to be available. The ratio of molar concentrations of TnsA to TnsB in the DS941(pGP1-2, pAG411) extracts used was estimated, as judged by the relative staining intensity of the proteins on Coomassie-stained Laemmli gels, to be very approximately 30:1 but, while all the TnsB was believed to be in solution, the majority of the TnsA was sequestered in an insoluble form.

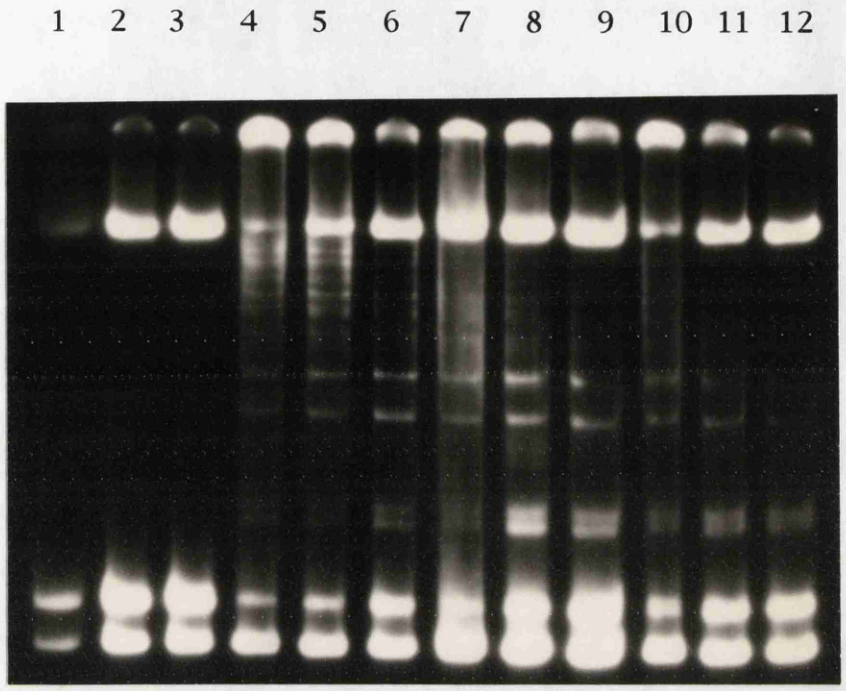


Figure 4.12 Gel retardation assay, showing the lack of effect of overproduced TnsA (encoded by pAG411) on the TnsB-specific binding pattern. The end labelled DNA was pNE200 cut with EcoRI, Hin*d*III and NdeI; the reactions were as described in Chapter 2. The gel was made and run in Tris glycine pH9.4.

1	pT7-7	undiluted extract			
2	pAG411	undiluted extract			
3	pAG411	2^{-2} dilution of extract			
4	pMR207	2^{-2}			
5	pMR207	2^{-4}			
6	pMR207	2^{-6}			
7	pMR207	2^{-2}	+	pT7-7	undiluted extract
8	pMR207	2^{-4}	+	pT7-7	undiluted extract
9	pMR207	2^{-6}	+	pT7-7	undiluted extract
10	pMR207	2^{-2}	+	pAG411	2^{-2} dilution
11	pMR207	2^{-4}	+	pAG411	2^{-2}
12	pMR207	2^{-6}	+	pAG411	2^{-2}

1 2 3 4 5 6 7 8 9

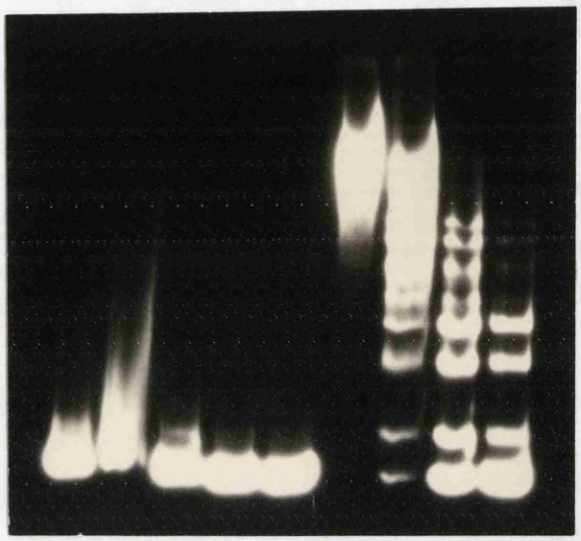


Figure 4.13 Tris glycine binding gel showing the lack of binding specific to the TnsB peptide encoded by pAG205 to the Tn*7* LE; the labelled DNA is pMR11 cut with EcoRI and HincII. Reactions were carried out as described in Chapter 2. Cell extracts contained the plasmids listed below.

1	pMR78	undiluted extract
2	pAG205	undiluted extract
3	pAG205	2^{-2} dilution
4	pAG205	2^{-4}
5	pAG205	2^{-6}
6	pMR207	undiluted extract
7	pMR207	2^{-2} dilution
8	pMR207	2^{-4}
9	pMR207	2^{-6}

1 2 3 4 5 6 7 8 9 10 11 12

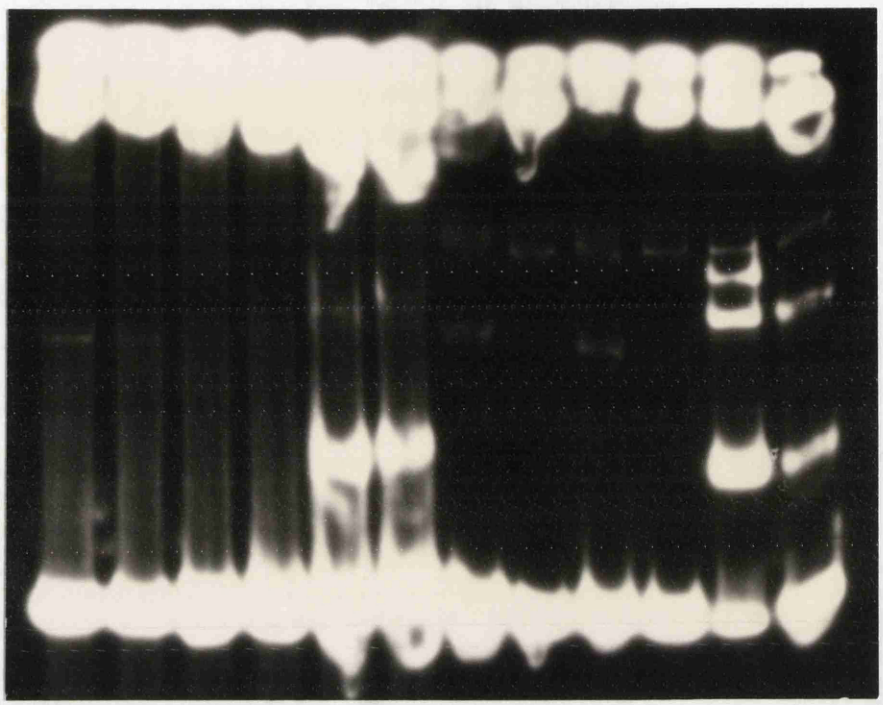


Figure 4.14 Gel retardation assay showing the lack of binding specific to the TnsB peptide encoded by pAG205 to the R1 and L3 end motifs of Tn*7*. The labelled R1 DNA is pLM5 cut with EcoRI and Hin*d*III; the labelled L3 DNA was pLA77 cut with EcoRI and Hin*d*III. The DS941 cell extracts used contained the plasmids indicated below. Reactions were carried out as described in Chapter 2; the gel was poured and run in Tris glycine.

1	R1 motif	with	pMR78	undiluted extract
2	R1 motif		pMR78	2^{-2} dilution of extract
3	R1 motif		pAG205	undiluted extract
4	R1 motif		pAG205	2^{-2} dilution
5	R1 motif		pMR207	2^{-2}
6	R1 motif		pMR207	2^{-4}
7	L3 motif	with	pMR78	undiluted extract
8	L3 motif		pMR78	2^{-2} dilution
9	L3 motif		pAG205	undiluted extract
10	L3 motif		pAG205	2^{-2} dilution
11	L3 motif		pMR207	2^{-2}
12	L3 motif		pMR207	2^{-4}

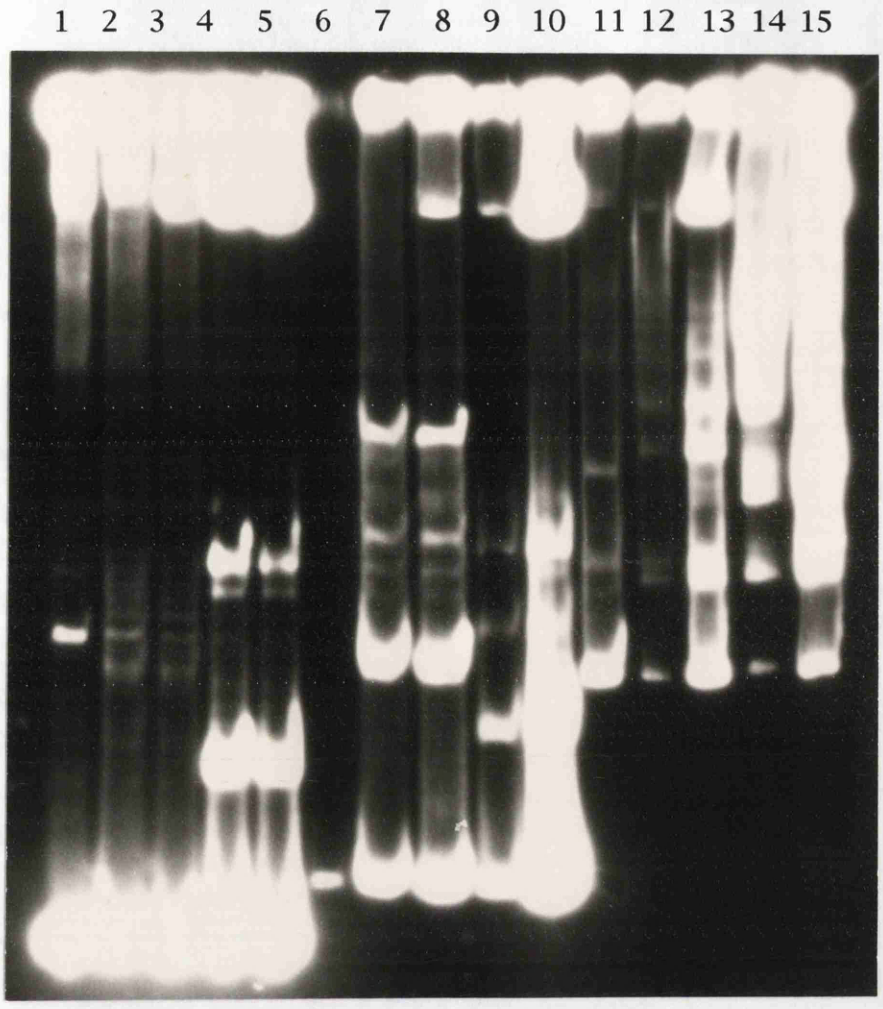


Figure 4.15 Gel retardation assay showing binding specific to various TnsB peptides to DNA fragments containing differing numbers of RE motifs. The peptides, produced in DS941, are encoded on plasmids by the sequences shown in Figure 4.9, and the labelled DNA fragments (from pAG095, pAG096 and pNE200) in Figure 4.10. The combinations tested are listed below. The gel was made and run in Tris glycine.

1	R4 motif	with	pMR78	undiluted extract
2	R4 motif		pAG205	undiluted extract
3	R4 motif		pAG205	2^{-2} dilution
4	R4 motif		pMR207	2^{-2} dilution
5	R4 motif		pMR207	2^{-4} dilution
6	R3 & R4 motifs	with	pMR78	undiluted extract
7	R3 & R4 motifs		pAG205	undiluted extract
8	R3 & R4 motifs		pAG205	2^{-2} dilution
9	R3 & R4 motifs		pMR207	2^{-2} dilution
10	R3 & R4 motifs		pMR207	2^{-4} dilution
11	R1-R4 motifs	with	pMR78	undiluted extract
12	R1-R4 motifs		pMR207	2^{-2} dilution
13	R1-R4 motifs		pMR207	2^{-4} dilution
14	R1-R4 motifs		pAG209	undiluted extract
15	R1-R4 motifs		pAG209	2^{-2} dilution

As can be seen on Figure 4.12, no difference in binding was seen in the presence of high concentration TnsA. There are two possible explanations for this: either the presence of TnsA does not affect the TnsB-specific binding to the RE, or not enough of the TnsA is soluble for interactions to be detectable. In the absence of antibodies to TnsA, it was not possible to quantify the concentration of soluble TnsA present in the cell extracts used.

4.6.2 Band patterns specific to TnsB and amino-terminal fragments of it

The experiments described in this section were carried out in the strain DS941, in which proteolysis was not inhibited. The results described below may therefore refer to smaller TnsB peptides than those encoded by the relevant plasmid. They are however relevant because both the studies of transcriptional repression reported in Section 4.7 and the *in vivo* transposition assays described in Section 4.9 were also carried out in DS941, there being no protease-deficient *galK* strain available.

The TnsB fragments were tested for their ability to retard the complete Tn*7* RE and LE, and various combinations of 22bp motifs from the RE and LE, as shown in Figure 4.10. As described in Section 4.4, the smallest TnsB fragment known to be associated with any binding activity was that encoded by pAG205, with sequence from the 5' end of the gene to the Hin*c*II site. While the peptide was able to bind to the complete RE, as encoded by pNE200 with two major retarded complexes (Figure 4.7), it was not competent to bind to the complete LE of pLA50 (Figure 4.13); further investigation demonstrated that binding did not occur at a detectable rate to a single motif from either the right or left end (Figures 4.14, 4.15), although retardation was observed in the presence of two contiguous RE motifs (Figures 4.15, 4.16), where again two retarded complexes could be clearly seen. Neither was pAG205-dependent binding seen to a hybrid site with two RE motifs, but spaced as L1 and L2 (Figure 4.17).

There are three possible explanations for these results. Firstly, the different 22bp motifs do not have identical sequence, and binding might occur to some motifs but not others. Secondly, the context of the motif might be important; bases outside the official motif might be required to stabilise binding. Thirdly. there may be TnsB-TnsB interactions which stabilise binding. The three possibilities are discussed below.

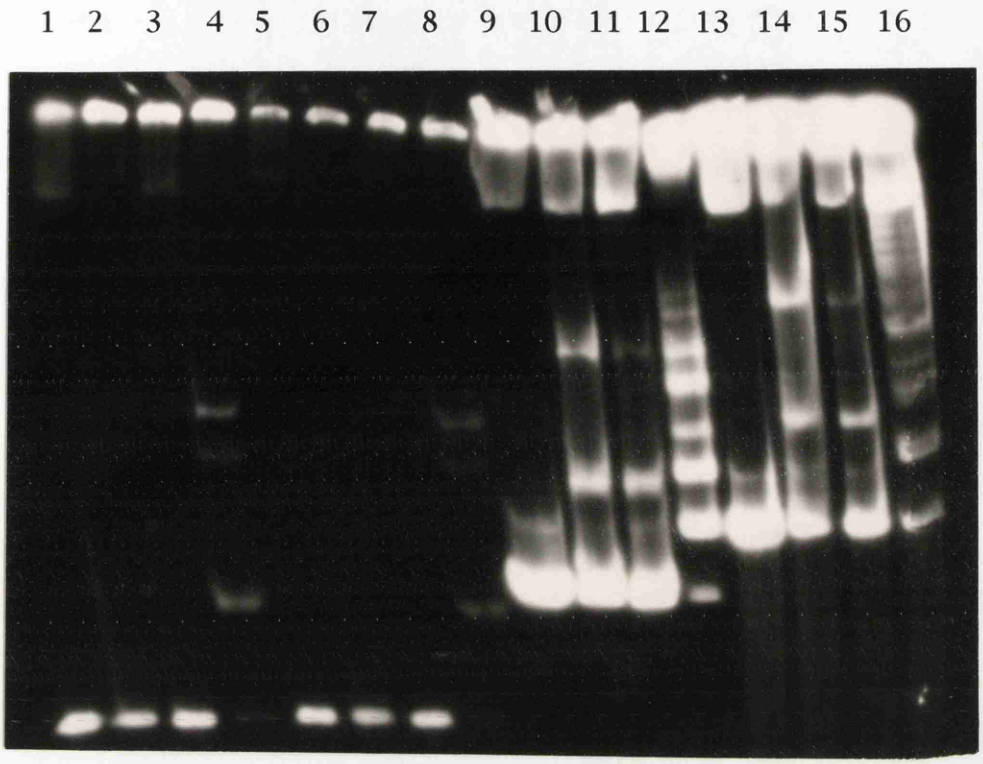


Figure 4.16 Gel retardation assay demonstrating the binding specific to the TnsB peptide encoded by pAG205 to a RE fragment containing $2^1/_2$ motifs (pNE200 cut with EcoRI and BglII), and the lack of binding to a single motif. The gel was poured and run in Tris glycine.

1	R1 motif	pMR78 extract	undiluted
2	R1 motif	pAG205	undiluted
3	R1 motif	pAG205	2^{-2} dilution
4	R1 motif	pMR207	2^{-2}
5	L3 motif	pMR78	undiluted
6	L3 motif	pAG205	undiluted
7	L3 motif	pAG205	2^{-2}
8	L3 motif	pMR207	2^{-2}
9	$^1/_2$R2, R3 and R4	pMR78	undiluted
10	$^1/_2$R2, R3 and R4	pAG205	undiluted
11	$^1/_2$R2, R3 and R4	pAG205	2^{-2}
12	$^1/_2$R2, R3 and R4	pMR207	2^{-2}
13	R1-R4 motifs	pMR78	undiluted
14	R1-R4 motifs	pAG205	undiluted
15	R1-R4 motifs	pAG205	2^{-2}
16	R1-R4 motifs	pMR207	2^{-2}

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

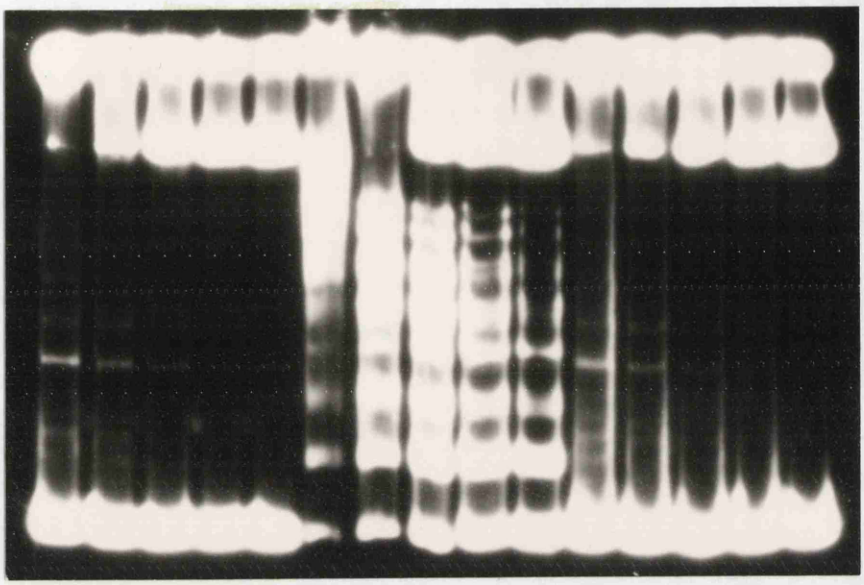


Figure 4.17 Gel retardation assay, demonstrating the lack of specific binding of the peptide encoded by pAG205 to a hybrid end, on the plasmid pAG091, formed of copies of the R1 motif with spacing as found between L1 and L2. Extracts were of DS941 containing one of the plasmids listed below. The gel was poured and run in Tris glycine.

1	pMR78	undiluted extract
2	pMR78	2^{-1} dilution
3	pMR78	2^{-2} dilution
4	pMR78	2^{-3} dilution
5	pMR78	2^{-4} dilution
6	pMR207	undiluted extract
7	pMR207	2^{-1} dilution
8	pMR207	2^{-2} dilution
9	pMR207	2^{-3} dilution
10	pMR207	2^{-4} dilution
11	pAG205	undiluted extract
12	pAG205	2^{-1} dilution
13	pAG205	2^{-2} dilution
14	pAG205	2^{-3} dilution
15	pAG205	2^{-4} dilution

The first option is that the peptide encoded by pAG205 has a very specific motif sequence requirement, which is satisfied by R2 and R3 (the binding to which as individual isolated motifs has not been tested), but not by R1, R4, L1, L2 or L3. While the LE motifs do vary substantially in sequence from those at the RE (see Figure 1.6), the difference between the individual RE motifs is very small, and the only base at which R1 and R4 are different from either R2 or R3 is the penultimate one as shown on Figure 4.1, where R1 and R4 have 'A', R3, L1, L2 and L3 have 'G' and R2 has 'T'. If this hypothesis were correct, binding of the peptide to a two-motif DNA RE fragment containing R3 and R4 would be expected to result in a single retarded band representing the binding of the peptide to R3; this is clearly not what is seen in Figure 4.15 or 4.16, where two bands are evident. This first hypothesis is therefore rejected.

The second consideration is the requirement for an extended DNA recognition sequence beyond the 22bp. The hydroxyl radical footprinting data of Arciszewska and Craig (1991), using what was believed to be full length TnsB, indicate some protection upstream of motifs (i.e. nearer the terminal nucleotides) but no downstream protection. As can be seen from Figure 4.11, the R1 motif of pLM5 was constructed without any flanking Tn*7* sequence, although the fragment used in binding gels included polylinker sequences; however, the R4 motif of pAG095 is accompanied by 3bp of upstream sequence, and the LE motifs of pLA77 and pLA50 are flanked by both upstream and downstream Tn*7* sequence (not shown on Figure 4.11). No binding is detectable to any of these three motifs in isolation. It is unlikely therefore that binding of this TnsB peptide could be achieved simply by the addition of DNA beyond the 22bp.

Finally, stabilisation by a second, contiguous truncated TnsB peptide molecule is considered. It appears, from the different results with the RE and LE motifs where spacing between the motifs differs, that binding of the pAG205-encoded peptide is only possible where motifs are contiguous. This is consistent with a situation in which a peptide molecule can be bound to each motif only as long as the peptide molecules can contact each other for stability. The dimensions and shape of the TnsB peptide encoded by pAG205 are not known, but it is plausible that it is too small to contact other similar molecules where there is a spacer region of 34bp or more (see Figure 1.6) between motifs. The fact that two, rather than one, retarded bands are seen with the R3+R4 DNA fragment probably indicates that binding is still subject to some instability and that transient binding to a single one of the

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

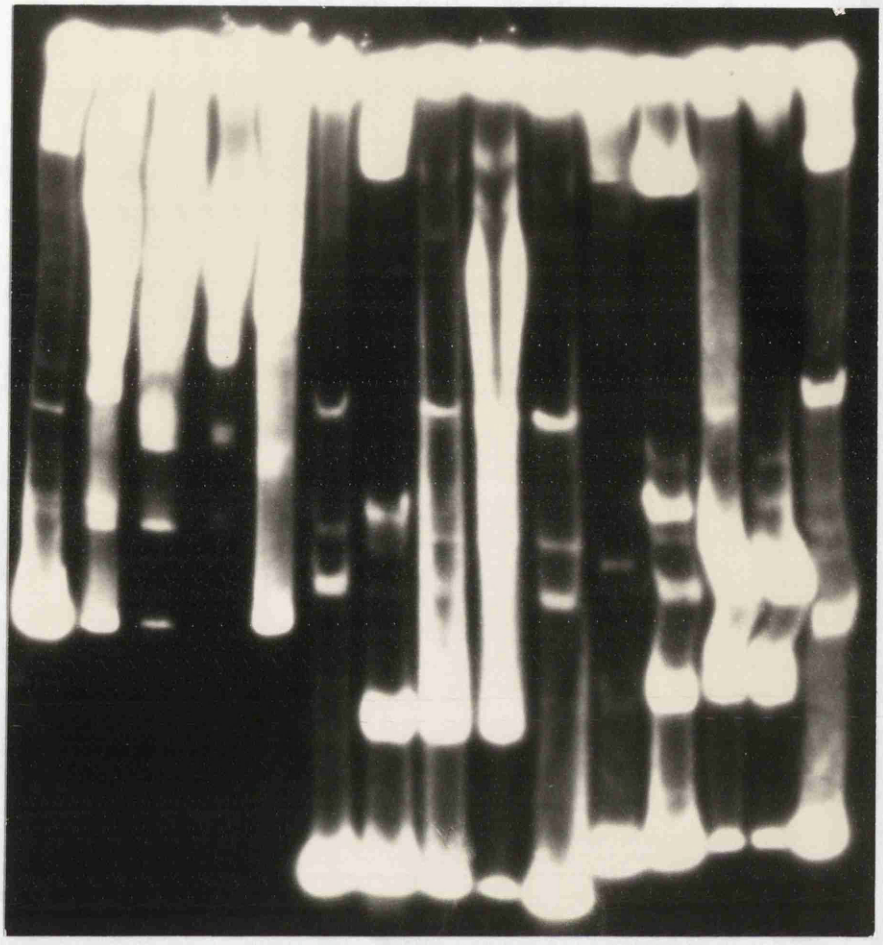


Figure 4.18 Tris glycine gel retardation assay showing the binding patterns specific to the peptides produced by DS941(pAG205), DS941(pAG209) and DS941(pAG210). The three labelled DNA fragments used are from pNE200 (R1-R4 motifs), pAG095 (R4) and pAG096 (R3 & R4).

1	R1-R4 motifs with	pMR78	undiluted extract
2	R1-R4 motifs	pMR207	2^{-3} dilution
3	R1-R4 motifs	pAG209	undiluted extract
4	R1-R4 motifs	pAG210	undiluted extract
5	R1-R4 motifs	pAG205	undiluted extract
6	R4 motif with	pMR78	undiluted extract
7	R4 motif	pMR207	2^{-3} dilution
8	R4 motif	pAG209	undiluted extract
9	R4 motif	pAG210	undiluted extract
10	R4 motif	pAG205	undiluted extract
11	R3 & R4 motifs with	pMR78	undiluted extract
12	R3 & R4 motifs	pMR207	2^{-3} dilution
13	R3 & R4 motifs	pAG209	undiluted extract
14	R3 & R4 motifs	pAG210	undiluted extract
15	R3 & R4 motifs	pAG205	undiluted extract

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

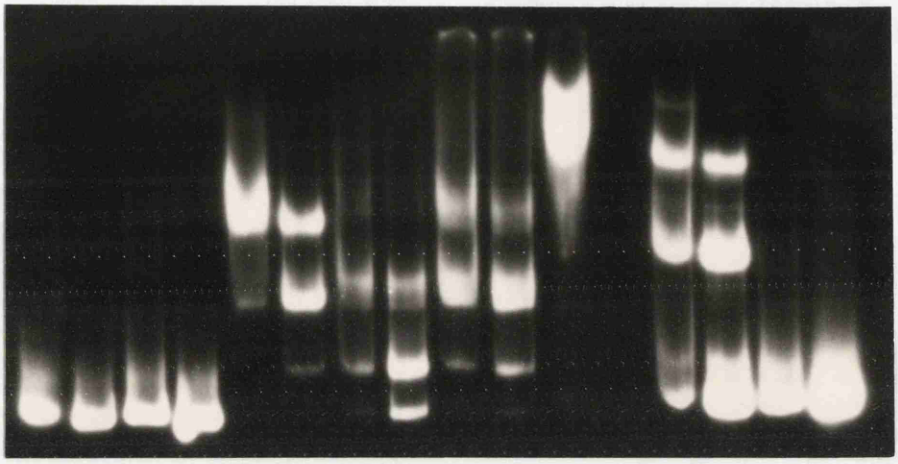


Figure 4.19 Gel retardation assay showing the binding specific to the TnsB peptides encoded by pAG205, pAG210, pAG211, pAG209 and pAG208 (the latter encoding a C-terminal TnsB peptide) in DS941 and also LA547(pMR78) and LA547(pMR207) extracts; the labelled pMR11 DNA contained the L1-L3 motifs.

1	pMR78	2^{-2} dilution of extract
2	pMR78	2^{-4} dilution
3	pAG205	2^{-2} dilution
4	pAG205	2^{-4} dilution
5	pAG210	2^{-2} dilution
6	pAG210	2^{-4} dilution
7	pAG211	2^{-2} dilution
8	pAG211	2^{-4} dilution
9	pAG209	2^{-2} dilution
10	pAG209	2^{-4} dilution
11	pMR207	undiluted extract
12	pMR207	2^{-2} dilution
13	pMR207	2^{-4} dilution
14	pMR207	2^{-6} dilution
15	pAG208	2^{-2} dilution
16	pAG208	2^{-4} dilution

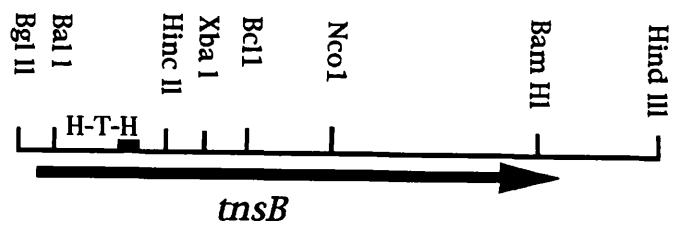
motifs occurs; it is possible that, once binding to both motifs has been established, one of the bound peptide molecules may dissociate from the DNA, leaving behind a peptide molecule which is now relatively stably bound to one of the two motifs.

The above discussion and Figures 4.7 and 4.13-4.16 suggest that the N-terminal TnsB peptide encoded by pAG205 requires two contiguous motifs to stabilise binding and that, where this condition is satisfied, two different protein-DNA complexes can be formed. It is unlikely that these complexes have the same stoicheiometry, with a degradation product of the peptide being responsible for one of them, since the observed spacing between the gel bands is not compatible with a small difference in protein size. Only two major retarded complexes could be seen with the complete right end, but it is possible that there was not a sufficiently high protein concentration for higher complexes to be stably formed.

The peptide encoded by the longer *tnsB* sequence of pAG210 is able to bind to single motifs (Figure 4.18), giving a single retarded complex on binding gels, and a second complex is detected when a second motif is encoded on the DNA fragment; compare the retardation of pAG095 and pAG096 DNA on Figure 4.18. There is also binding activity to the Tn*7* Left End DNA which is specific to this peptide and gives three retarded bands, as shown on Figure 4.19.

The peptides encoded by pAG211 and pAG209, with *tnsB* sequence to the 5' BclI and NcoI sites respectively, show similar binding properties to the pAG210 peptide (Figures 4.18, 4.19). It is interesting that the retardation specific to pAG209 is identical to that specific to pAG210, although the peptide encoded by the former is expected to be nearly twice the size of the pAG210-encoded peptide. It is possible that the pAG209-encoded peptide is cleaved before it is able to bind DNA; alternative explanations for this phenomenon are that the retardation is primarily due to bending of the DNA, that charge differences on the different TnsB peptides result in differing mobility; or that the TnsB peptides are having an indirect effect on other molecular species which are binding the DNA. This apparent discrepancy is discussed below and also in Section 4.6.4.

The predicted isoelectric points (the pH at which there is no net charge) of the different peptides are shown on Figure 4.20; Figure 4.21 shows a superimposition of the charge/pH plots for the different TnsB



	Predicted no. amino acids	Predicted peptide size/ kDa	Isoelectric point
pMR207	702	81	8.73
pAG205	198	23	7.77
pAG210	252	30	10.23
pAG211	316	37	10.26
pAG209	438	50	9.35

H-T-H = putative helix-turn-helix motif

Figure 4.20 Predicted sizes and isoelectric points of the TnsB peptides encoded by the plasmids listed above. The figures were calculated using the GCG ISOELECTRIC program.

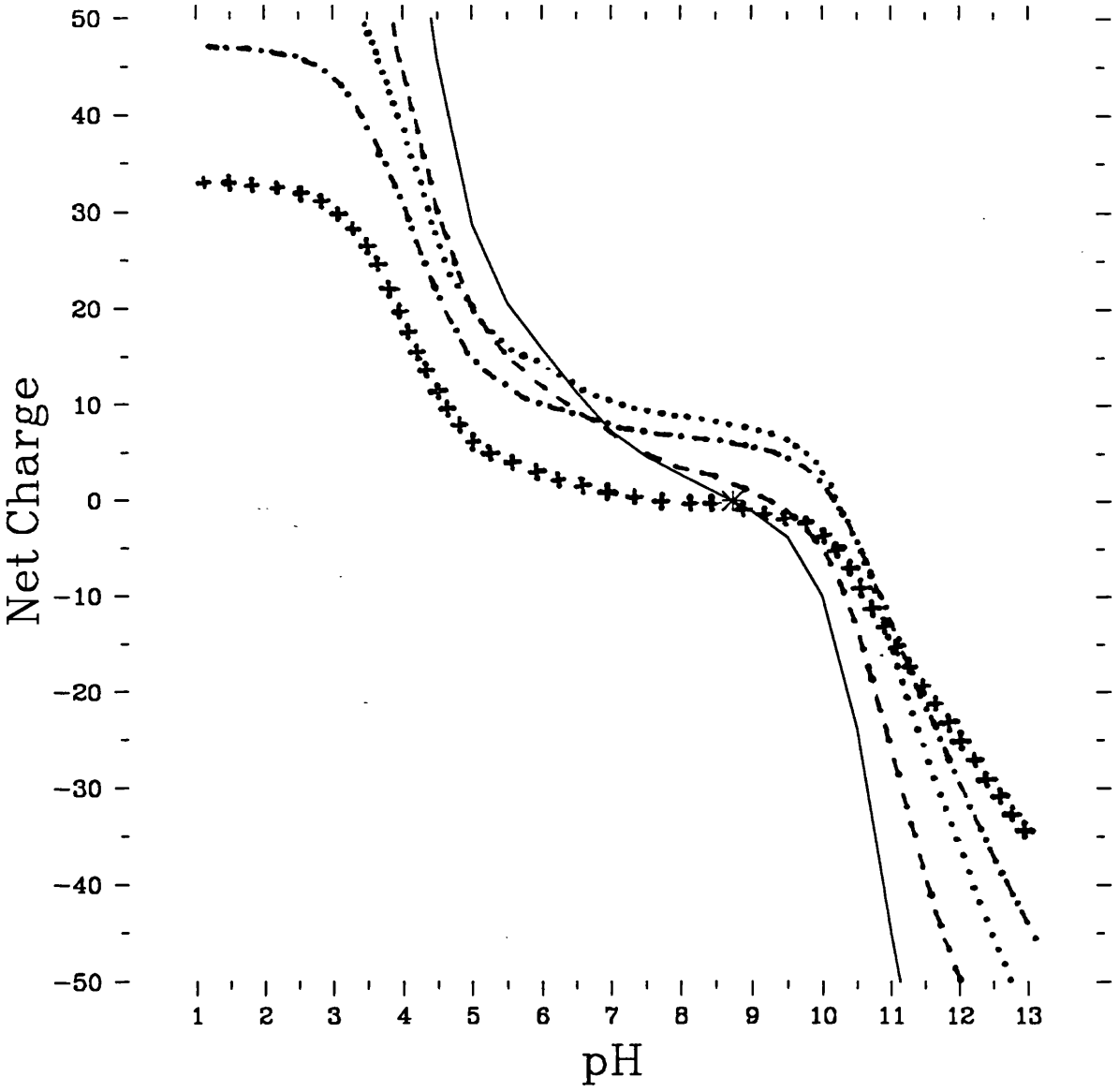


Figure 4.21. The variation of charge with pH in TnsB, encoded by pMR207, and the TnsB peptides encoded by the plasmids pAG205, pAG210, pAG211 and pAG209.

————	pMR207
+ + + + + + +	pAG205
_ . _ . _ . _ .	pAG210
. . . . .	pAG211
_ _ _ _ _ _ _	pAG209

peptides, as drawn by the ISOELECTRIC program. The computer predictions of pI values of the four peptides encoded by pAG205, pAG210, pAG211 and pAG209 are 7.77, 10.23, 10.26 and 9.35 respectively, compared with 8.73 for the complete TnsB protein, thus indicating the presence of a region rich in basic amino acids encoded by the DNA sequence between the HincII and XbaI sites used for subcloning. It can be seen from these pI values that, at the pH9.4 of the electrophoresis running buffer, the whole protein and the small peptide encoded by pAG205 might be expected to be slightly negatively charged, the peptides encoded by pAG210 and pAG211 positively charged, and the peptide encoded by pAG209 probably uncharged; this, together with their presumed greater molecular weight, would lead to the prediction that they would migrate through a gel considerably more slowly than the small peptide encoded by pAG205. Positively charged proteins might be expected to migrate slightly more slowly than negatively charged ones of similar molecular weight (Carey, 1988), whereas those encoded by pAG210 and pAG211 actually appear to migrate faster than the other, more negatively charged ones. Therefore difference in charge is unlikely to be the explanation for the difference in migration patterns observed on binding gels with the TnsB peptides encoded by the different plasmids.

It is well documented (Wu and Crothers, 1984) that a DNA fragment bent (through contact with a protein molecule or otherwise) is retarded relative to the unbent form during polyacrylamide gel electrophoresis, and that this retardation varies with the position of the bend in the DNA: DNA bent near one end is able to migrate faster through the gel matrix than DNA with a central bend. Bending of the DNA as a result of protein binding has not been investigated here, but any differential effect is unlikely to be substantial enough to account for the difference in mobility between the retarded bands specific to the pAG205 and pAG210 encoded peptides. A DNA-protein-DNA sandwich which is formed only in the presence of the pAG205 encoded protein is another formal possibility; according to this hypothesis, such a protein-DNA complex would only be stable in the, presumably dimeric, sandwich form since no lower band is seen; if it is indeed the TnsB peptide which is binding to the DNA, it is unlikely that binding would be enhanced in this peptide compared with the larger peptides where more domains are available for intermolecular interactions. Any sandwich complex is also likely to be even more retarded than the band seen on the autoradiographs; the formation of sandwich complexes in this case is therefore unlikely.

Support for the first hypothesis, of proteolytic cleavage, comes from the retardation pattern seen with pMR207 extracts; again, the mobility of the least retarded band appears the same, and this can be accounted for by the observation that several proteolytic products of TnsB are found. It seems that wildtype TnsB and the peptides encoded by pAG209 and pAG210 are all unstable and cleave to release a DNA-binding peptide which is probably smaller than the peptide encoded by pAG205 (assuming that the stoicheiometry of all the complexes under discussion is the same); this cannot be detected on Laemmli gels, but could be confirmed with the aid of antibodies.

Any explanation of the apparently aberrant mobility of the complexes retarded in the presence of the different peptides must also take into account the observation that binding of the small pAG205-encoded peptide was seen only to two or more contiguous 22bp motifs; it might be proposed that the other peptides can bind as a monomer, but that binding of this smallest peptide to 22bp motifs can only be stabilised in a dimeric form, with one protein monomer associated with each motif, where the motifs are in close enough proximity to enable this to occur. Were the pAG209- and pAG210-encoded peptides to be truncated by cleavage of amino acid residues from the amino-terminal end, additional carboxy-terminal sequence to that in the pAG205-encoded peptide could still be present in these former peptides. It is therefore a possibility that the extreme amino-terminal residues of TnsB are not required for binding, but that sequence encoded in *tnsB* between the HincII and XbaI sites is required for stability of peptide binding to a single 22bp motif.

The protease-deficient strain LA547 became available at the close of this work, and preliminary results of binding reactions using both wildtype TnsB and cloned peptides of it expressed in LA547 are reported in Section 4.6.4.

4.6.3 The stoicheiometry of binding

It is uncertain whether TnsB binds to each 22bp motif as a monomer or a dimer. The core DNA site of site-specific recombination systems is generally highly symmetric and thought to be bound by a dimer of the recombinase enzyme. Bacteriophage Mu, on the other hand, has repeated end motifs which do not display dyad symmetry, and it is believed that only one MuA protein molecule is bound to each terminal motif. The Tn*7* end

1 2 3 4 5 6 7 8 9 10 11 12 13

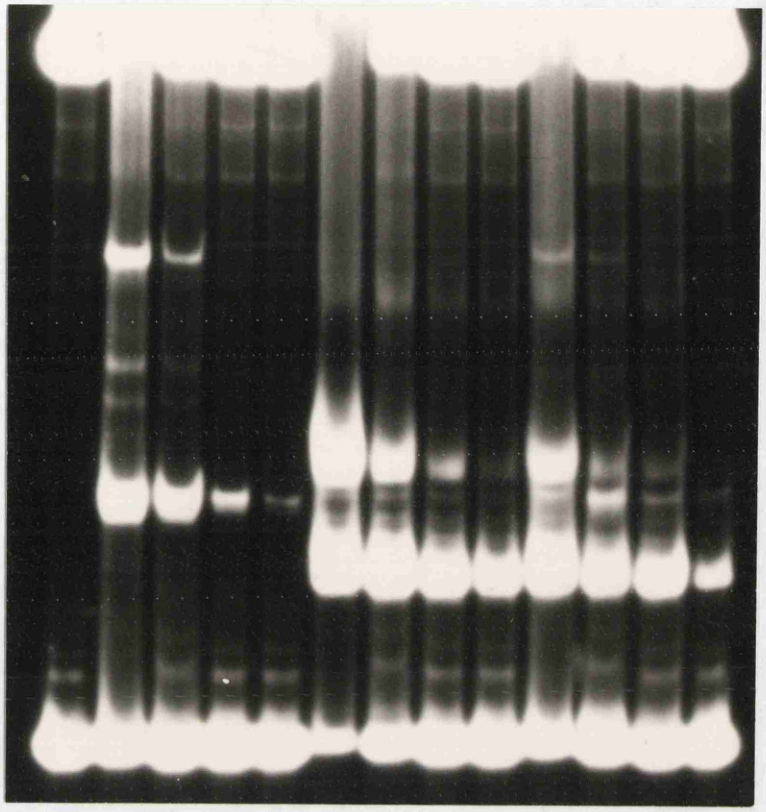


Figure 4.22 Tris glycine gel retardation assay showing the effect of mixing extracts containing two different TnsB peptides on the peptide-specific binding pattern. The labelled DNA is pAG096 (R3 and R4 motifs); the peptides are encoded by pAG205 and pAG209.

1	no extract; dilution buffer only				
2	pAG205	undiluted extract			
3	pAG205	2^{-2} dilution			
4	pAG205	2^{-4} dilution			
5	pAG205	2^{-6} dilution			
6	pAG209	undiluted extract			
7	pAG209	2^{-2} dilution			
8	pAG209	2^{-4} dilution			
9	pAG209	2^{-6} dilution			
10	pAG205	undiluted extract	+	pAG209	undiluted extract
11	pAG205	2^{-2} dilution	+	pAG209	2^{-2} dilution
12	pAG205	2^{-4} dilution	+	pAG209	2^{-4} dilution
13	pAG205	2^{-6} dilution	+	pAG209	2^{-6} dilution

motifs, as shown in Figure 4.1, have very limited dyad symmetry, which does not coincide with the centre of each motif, as defined by footprinting (Arciszewska and Craig, 1991), and so monomer binding to each motif is plausible.

A simple approach to investigating the stoicheiometry of TnsB binding, based on the method of Hope and Struhl (1987) was adopted here. Two different sized fragments of TnsB were used; these were encoded by pAG205 and pAG209, and the retardation patterns specific to each peptide were easily distinguishable. If the two cell extracts were mixed, monomer binding would result in a retardation pattern which was a superimposition of the patterns observed with the two individual extracts; if, however, a protein dimer could bind to each motif, it would be likely that heterodimers could be formed between the two different sized peptides, resulting in bands with novel mobility. With the assumptions that dimerisation could readily occur and that formation of a homodimer and a heterodimer were equally probable, the proportion of heterodimers expected from equal concentrations of the two peptides would be 50% of the total number of dimers.

A gel binding assay was therefore carried out with the two extracts above. The relative concentrations of the two cell extracts used were adjusted so that approximately the same proportion of the pAG096 DNA fragment was retarded by each. As can be seen from Figure 4.22, no novel bands were detected. This is evidence in support of monomer binding to each 22bp motif. By analogy with the bacteriophage Mu situation, monomer binding to each motif would be *a priori* more likely than that of dimers, and so the result obtained is reasonable.

It is however possible that the two peptides used were unable to form heterodimers. It is also conceivable that wildtype TnsB forms dimers, but that the domain required for dimer stability is at the carboxy end of the protein and thus has been deleted from all the truncated proteins used in these studies. A categorical pronouncement on the stoicheiometry of binding cannot therefore be made on the basis of this experiment alone.

A more rigorous investigation of the stoicheiometry of binding would require the use of pure protein of accurately determined concentration, labelled with tritiated amino acids; clearly the purification techniques for TnsB are not yet adequate for such a technique to be considered.

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

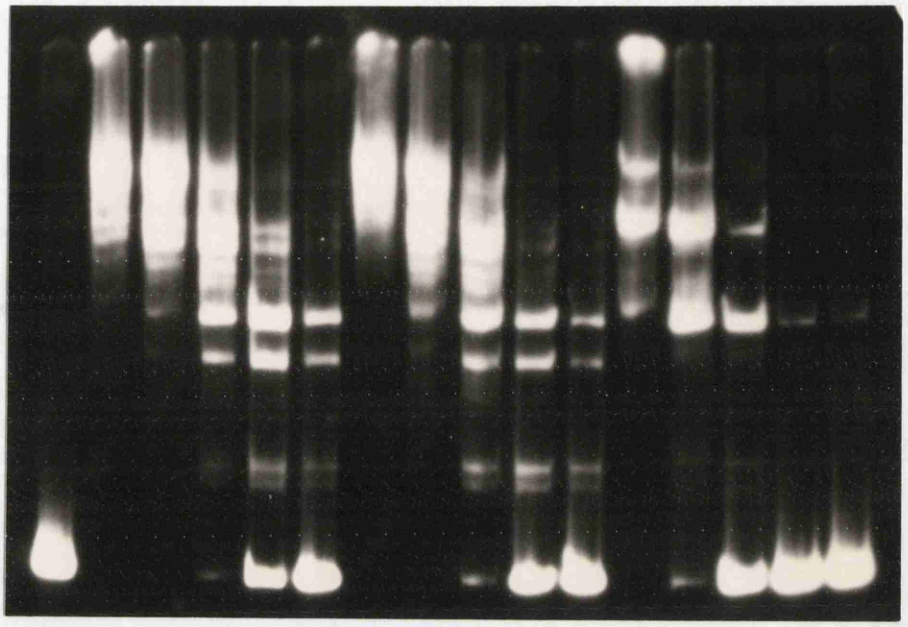


Figure 4.23 Tris glycine gel retardation assay showing the binding patterns specific to TnsB expressed from pMR207 in DS941 (grown at either 30° or 37°) and LA547 (grown at 30°). The labelled DNA is pNE200 cut with EcoRI and Hin*d*III to give the R1-R4 motifs.

1	no extract; dilution buffer only		
2	DS941(pMR207)	37°	undiluted extract
3	"	"	2^{-2} dilution
4	"	"	2^{-4} dilution
5	"	"	2^{-6} dilution
6	"	"	2^{-8} dilution
7	DS941(pMR207)	30°	undiluted extract
8	"	"	2^{-2} dilution
8	"	"	2^{-4} dilution
10	"	"	2^{-6} dilution
11	"	"	2^{-8} dilution
12	LA547(pMR207)	30°	undiluted extract
13	"	"	2^{-2} dilution
14	"	"	2^{-4} dilution
15	"	"	2^{-6} dilution
16	"	"	2^{-8} dilution

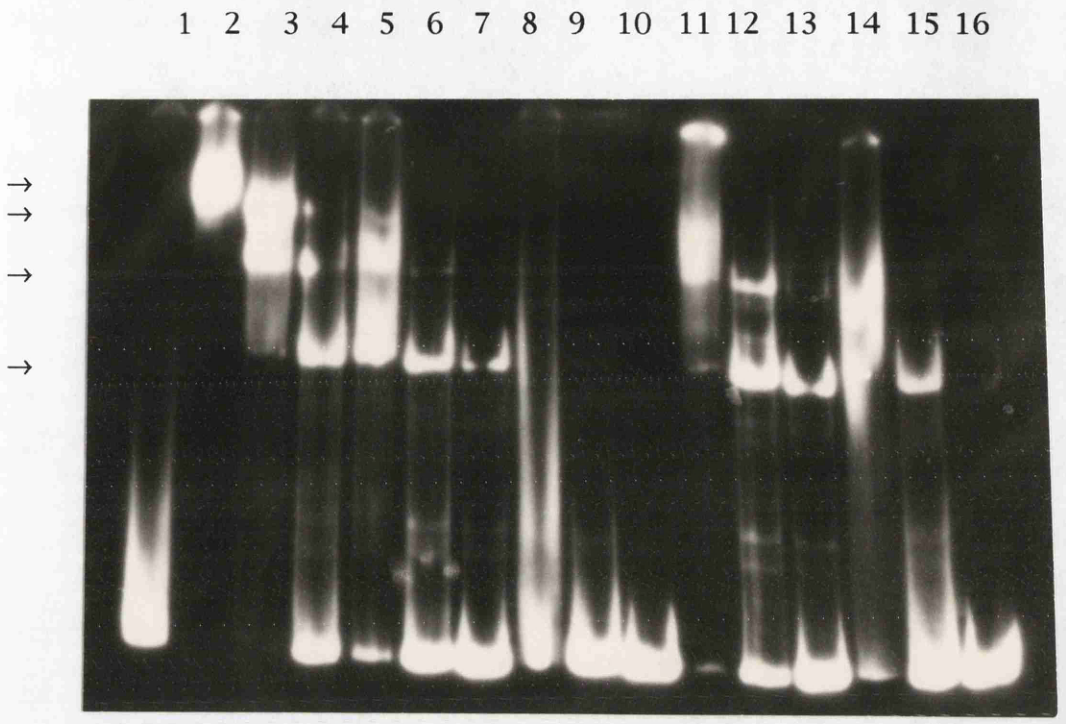


Figure 4.24 Tris glycine gel retardation assay with pNE200 labelled DNA, showing

(1) that there are 4 TnsB-specific retarded bands when using LA547(pAG082) extract and the labelled R1-R4 motifs of pNE200;

(2) that a similar binding pattern is obtained with extracts of LA 547 containing one of pAG431, pAG433 or pAG434, but not pAG432.

1	pMR78	undiluted extract
2	pAG082	undiluted extract
3	pAG082	2^{-2} dilution
4	pAG082	2^{-4} dilution
5	pAG431	undiluted extract
6	pAG431	2^{-2} dilution
7	pAG431	2^{-4} dilution
8	pAG432	undiluted extract
9	pAG432	2^{-2} dilution
10	pAG432	2^{-4} dilution
11	pAG433	undiluted extract
12	pAG433	2^{-2} dilution
13	pAG433	2^{-4} dilution
14	pAG434	undiluted extract
15	pAG434	2^{-2} dilution
16	pAG434	2^{-4} dilution

4.6.4 Gel binding assays using TnsB produced in a protease-deficient strain

It was not until towards the close of this research that the protease-deficient strain LA547 became available. LA547 has the *htpR* (otherwise known as *rpoH*) genotype, and thus does not produce the sigma factor σ^{32} which is required for expression of the *lon* protease-encoding gene. Because LA547 does not grow at 37°, different conditions were used for the expression of *tnsB* from pMR207: 400ml isosensitest broth with trimethoprim was inoculated with an overnight culture, grown at 30° until approximately $OD_{600}=0.6$, then diluted 1:1 in fresh isosensitest broth and trimethoprim, and IPTG was added to 1mM; growth was continued for between 6 and 7 hours before harvesting the cells, which were lysed in the normal way.

Extracts of pMR207-containing cells of DS941 and LA547 were compared on a binding gel, using labelled DNA containing the R1-R4 motifs; an extract of DS941(pMR207) grown at 30° was also made, to ensure that any differences observed could be assigned to either the strain or the growth conditions. It can be seen from Figure 4.23 that growth of DS941(pMR207) at 30° has no observable effect on the protein species produced which bind the Tn*7* right end; by contrast, the number of bands on the autoradiograph is reduced considerably when *tnsB* is expressed in LA547. Three or four retarded complexes are prominent when the LA547 extract is used, and it is thought likely that each of these represents binding to a different number of DNA motifs; as the protein concentration is increased, the DNA appears to be chased into higher complexes. It has been reported elsewhere (Arciszewska and Craig, 1991) that four distinct retarded complexes are seen when TnsB is produced in LA547; only three are seen here with extracts from cells containing the plasmid pAG205, perhaps because the TnsB concentration is lower than was used previously. A fourth, higher complex has been observed when the higher copy-number pUC-based plasmid pAG082 was used instead of pMR207 (Figure 4.24). Below the band corresponding to the least retarded complex seen with LA547, there are at least three other retarded complexes detected with DS941 extracts; it is likely that these are due to three different protein species, since DNA is not chased from one complex to another as the protein concentration is varied, and the difference in relative mobility between them is not compatible with one of them being a simple doubling of the number of protein molecules in a complex relative to that in any of the less

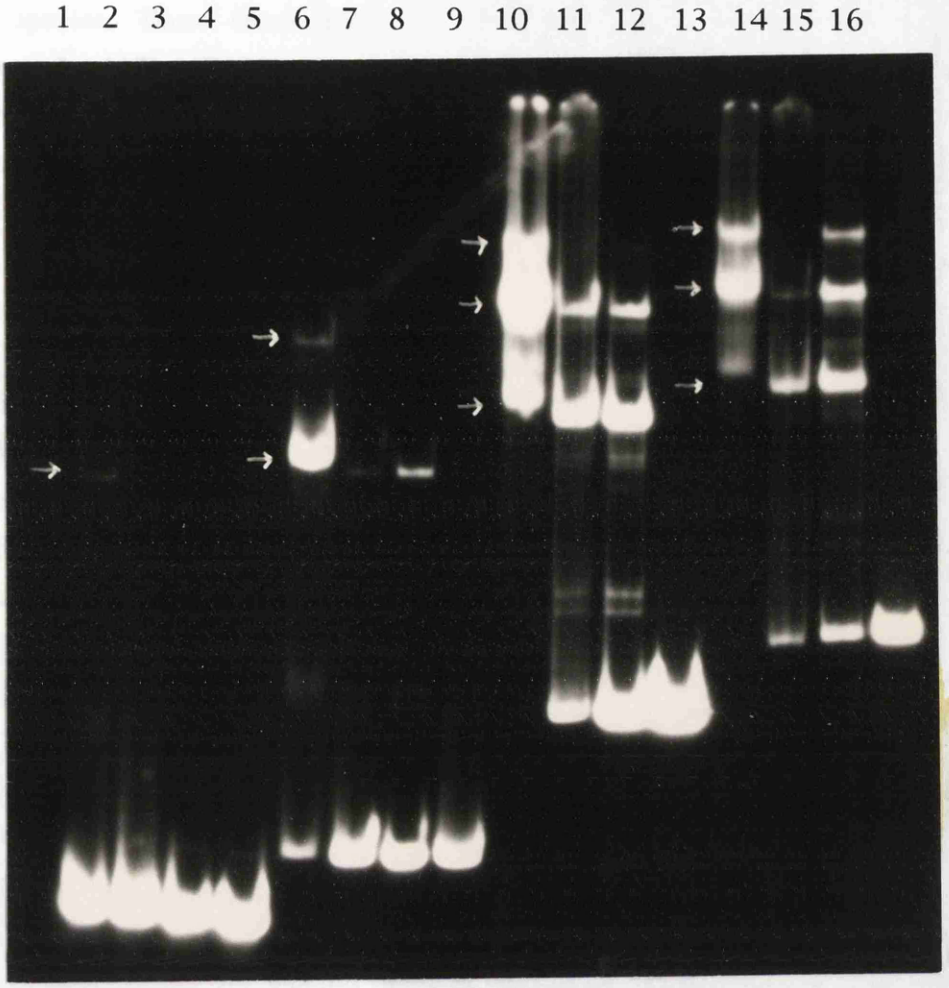


Figure 4.25 Tris glycine gel retardation assay to show the relative mobility of TnsB-specific retarded bands (marked with arrow) with labelled DNA of pAG095, pAG096, pLA50 or pNE200, containing 1, 2, 3 or 4 motifs respectively. The protein extract is LA547 (pMR207) throughout.

1	R4 motif	with	undiluted extract
2	R4 motif	with	2^{-2} dilution of extract
3	R4 motif	with	2^{-4} dilution
4	R4 motif	with	no extract; dilution buffer only
5	R3 & R4 motifs	with	undiluted extract
6	R3 & R4 motifs	with	2^{-2} dilution
7	R3 & R4 motifs	with	2^{-4} dilution
8	R3 & R4 motifs	with	dilution buffer only
9	L1-L3 motifs	with	undiluted extract
10	L1-L3 motifs	with	2^{-2} dilution
11	L1-L3 motifs	with	2^{-4} dilution
12	L1-L3 motifs	with	dilution buffer only
13	R1-R4 motifs	with	undiluted extract
14	R1-R4 motifs	with	2^{-2} dilution
15	R1-R4 motifs	with	2^{-4} dilution
16	R1-R4 motifs	with	dilution buffer only

retarded species. These three bands can also be seen faintly in the samples containing LA547(pMR207) extracts, where it is presumed that limited proteolysis is still occurring, either *in vivo* or *in vitro*. All three of the presumed TnsB proteolysis products must contain the DNA-binding domain, and so cleavage of TnsB must occur at a minimum of three different defined sites.

It was hypothesised that wildtype TnsB produced one retarded complex for each 22bp motif, with either a monomer or an inseparable dimer binding to each motif. To test this, a binding gel was run in which the LA547(pMR207) extract had been incubated with DNA containing either one, two, three or four motifs, as shown in Figure 4.25. While low intensity bands due to degradation products of TnsB were evident, one predominant band was seen with one motif, two bands with two motifs, and three bands with either three or four motifs. As discussed above, the lack of a fourth band with four motifs is probably due to low TnsB concentration in the cell extract; a fourth, more retarded band was seen with extracts from cells containing the pUC-based plasmid pAG082 (Figure 4.24). The relative mobilities of the complexes with the four different DNA fragments were compared: the fastest complex had a measured R_L value of 0.45, 0.48, 0.52 and 0.52 reading from left to right on Figure 4.25; the second complex had R_L of 0.30, 0.33 and 0.35; and the least mobile complex had values of 0.23 and 0.24. The consistency within each set of values suggests that complexes within a set have the same stoicheiometry, and thus lends weight to the hypothesis that one additional complex is added each time an additional motif is encoded on the DNA fragment.

It is noticeable in Figure 4.25 that TnsB-dependent binding to a single motif is weak, compared with the binding to a DNA fragment containing two motifs. This could be because ^{32}P-labelling of the former DNA had been inefficient, leading to the presence of a high concentration of competitor unlabelled DNA in the reaction. If this were the case, protein concentration would appear to be the limiting factor in determining binding to the motif, since a large proportion would be bound to unlabelled DNA. The TnsB concentration in the extract used was not sufficiently high to enable retardation to be seen with non-radioactive, ethidium-stained DNA (results not shown), and so a different approach was taken to resolving the problem. A standard radioactive gel binding assay was carried out in which the concentrations of both the specific DNA fragment

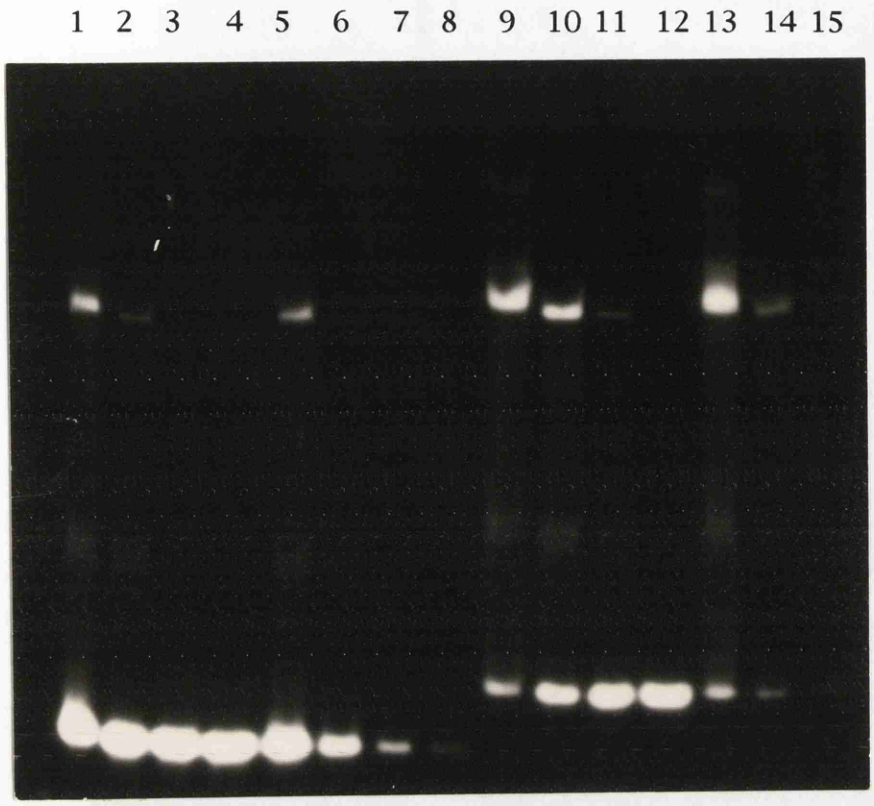


Figure 4.26 Gel retardation assay to investigate whether protein or DNA concentration is limiting in TnsB-specific binding to a single R4 motif. Dilutions of LA547(pMR207) extract were as indicated below; the labelled DNA, of pAG095 or pAG096, was also diluted from its standard concentration as shown.

1	pAG095	undiluted DNA	with	pMR207	undiluted extract
2	pAG095	undiluted DNA	with	pMR207	2^{-2} dilution
3	pAG095	undiluted DNA	with	pMR207	2^{-4} dilution
4	pAG095	undiluted DNA	with	pMR207	2^{-6} dilution
5	pAG095	undiluted DNA	with	pMR207	undiluted extract
6	pAG095	2^{-2} dilution	with	pMR207	undiluted extract
7	pAG095	2^{-4} dilution	with	pMR207	undiluted extract
8	pAG095	2^{-6} dilution	with	pMR207	undiluted extract
9	pAG096	undiluted DNA	with	pMR207	undiluted extract
10	pAG096	undiluted DNA	with	pMR207	2^{-2} dilution
11	pAG096	undiluted DNA	with	pMR207	2^{-4} dilution
12	pAG096	undiluted DNA	with	pMR207	2^{-6} dilution
13	pAG096	undiluted DNA	with	pMR207	undiluted extract
14	pAG096	2^{-2} dilution	with	pMR207	undiluted extract
15	pAG096	2^{-4} dilution	with	pMR207	undiluted extract

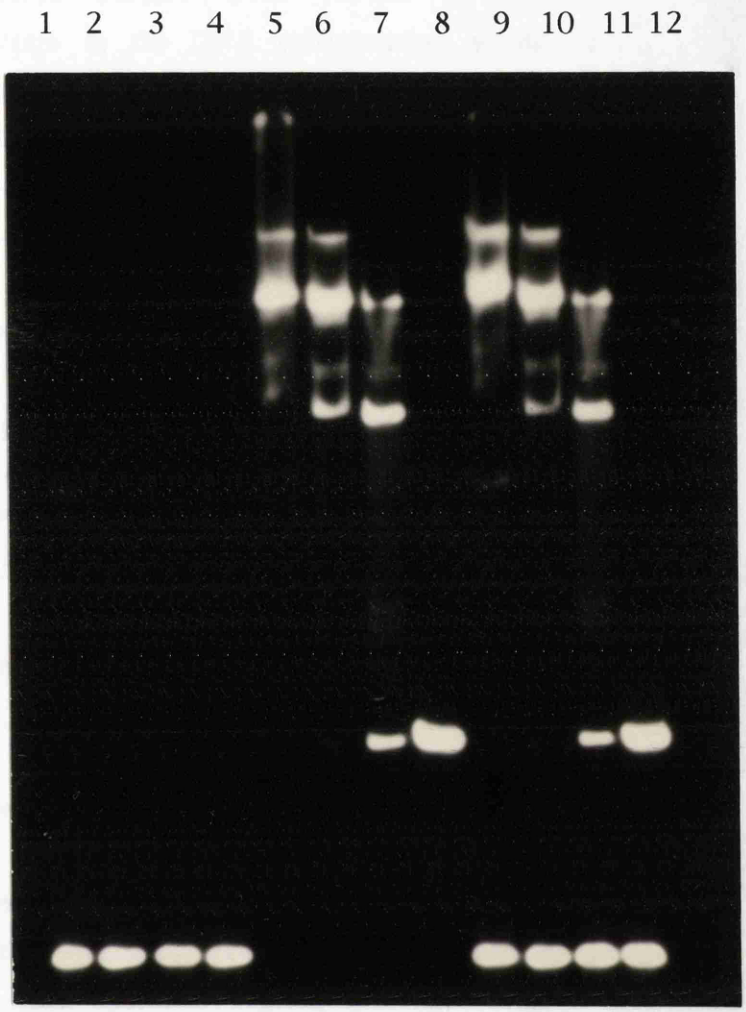


Figure 4.27 Gel retardation assay with competition between pAG095 (R4 motif) and pLA50 (L1-L3 motifs) labelled DNA for the TnsB produced in LA547(pMR207).

1	R4 motif	with	undiluted protein extract
2	R4 motif	with	2^{-2} dilution of extract
3	R4 motif	with	2^{-4} dilution
4	R4 motif	with	no extract; dilution buffer only
5	L1-L3 motifs	with	undiluted extract
6	L1-L3 motifs	with	2^{-2} dilution
7	L1-L3 motifs	with	2^{-4} dilution
8	L1-L3 motifs	with	dilution buffer only
9	R4 motif + L1-L3 motifs	with	undiluted extract
10	R4 motif + L1-L3 motifs	with	2^{-2} dilution
11	R4 motif + L1-L3 motifs	with	2^{-4} dilution
12	R4 motif + L1-L3 motifs	with	dilution buffer only

and the protein extract were varied. If protein concentration were limiting, reduction in the DNA concentration would be expected to result in a greater proportion of DNA bound by protein, and hence a higher ratio of retarded to unretarded DNA. The results, shown on Figure 4.26, suggest that this is not the case, although the proportion of labelled DNA retarded was so low that accurate quantification of the ratios was not possible. In a second investigation, DNA containing one 22bp motif (from pAG095) and a second fragment containing three RE motifs (from pLA50) were mixed together in reactions; it was anticipated that, as binding to a single motif appeared weak, TnsB-specific binding might occur preferentially to the pLA50 motif, rather than the pAG095 one if TnsB concentration were limiting. It does not appear from Figure 4.27 that binding to the pAG095 DNA fragment is visibly reduced in the presence of the competitor pLA50 DNA, and thus TnsB concentration is unlikely to be the limiting factor in the poor binding observed to a single 22bp motif. An alternative explanation must therefore be sought. It appears that additional DNA sequence must be necessary to stabilise the binding. It is not clear whether it is DNA *per se*, or the presence of protein bound to a second motif which is the stabilising factor; according to Arciszewska and Craig (1991), footprinting data indicate that TnsB protection of DNA extends upstream (as defined on Figure 1.6) of the 22bp motifs for up to 6bp, and it is possible that there is also some sequence preference in this upstream region, although none is obvious.

The maximum number of retarded bands seen in this system with one protein species is equal to the number of 22bp motifs in the labelled DNA fragment. If protein binding, either as a monomer or as an inseparable dimer, to the different motifs were a random phenomenon, it might be expected that a far larger number of retarded bands might be seen (up to $2^n - 1$, where n is the number of motifs), since binding at different positions on the DNA fragment would result in differential bending of the DNA and thus variation in mobility. Clearly, this is not seen. Two possible explanations for this observation are, firstly, that each retarded band represents a dynamic equilibrium between a number of different protein-DNA complexes each of the same stoicheiometry; or, secondly, that the complexes are formed in a specific order, and each band observed represents only one complex. There is some evidence (Arciszewska and Craig, 1991) that there is a degree of specificity in the order of formation of the complexes, but this specificity is not stringent

enough to account for the small number of bands seen. It is not possible to tell from the results reported here whether, in the presence of several motifs on the DNA, one particular motif is occupied preferentially. It is more likely that each band represents an average view of occupancy of the different sites; if this were indeed the case, mobility differences due solely to alternative bend centres of the DNA would be presumed to be indistinguishable.

A preliminary investigation was also carried out of the binding properties of cloned N-terminal peptides of TnsB expressed in LA547. It was attempted to transform LA547 with pAG205, pAG210, pAG211 and pAG209, all of which had been used previously for binding gels when the truncated *tnsB* genes were expressed in DS941. Unfortunately, despite several attempts, it proved impossible to transform LA547 with pAG211 and pAG209. This may be because truncation of TnsB at the BclI and NcoI sites of the gene results in a structure at the carboxy terminal end of the TnsB peptides which is in some way lethal to the host cells; in DS941, this structure could have been removed by *lon* encoded protease before the peptide had any deleterious effect on the host cells. An alternative explanation is that the *lacI^q* genotype of DS941 ensured a high degree of repression of the truncated *tnsB* genes during the establishment of colonies following transformation of the strain with the plasmids pAG211 and pAG209; whereas in LA547, which has the wildtype *lacI* genotype, limited gene expression following transformation could have been lethal, regardless of whether or not the peptide was post-translationally truncated.

Gel binding assays were carried out with the crude extracts of LA547 (pAG205) and LA547 (pAG210), using as DNA substrate the R4 motif of pAG095, and the R3+R4 motifs of pAG096. As can be seen on Figure 4.28, the mobility pattern of the pAG096 DNA fragment in the presence of the two pAG205-containing cell extracts is identical, indicating that *lon*-dependent proteolytic cleavage of the TnsB peptide encoded by pAG205 does not occur in DS941. When the retardation patterns of the two pAG210-containing cell extracts are compared, differences are evident: the DS941 extract gives two retarded complexes with the R4 motif of pAG095; only the upper of these two bands is seen with the LA547 extract, suggesting that some *lon*-dependent proteolytic cleavage of the peptide encoded by pAG210 occurs *in vivo*. If lanes 7 and 9 of Figure 4.28 are compared, it can be seen that the TnsB peptide encoded by pAG205 (Lane 7) apparently retards DNA more

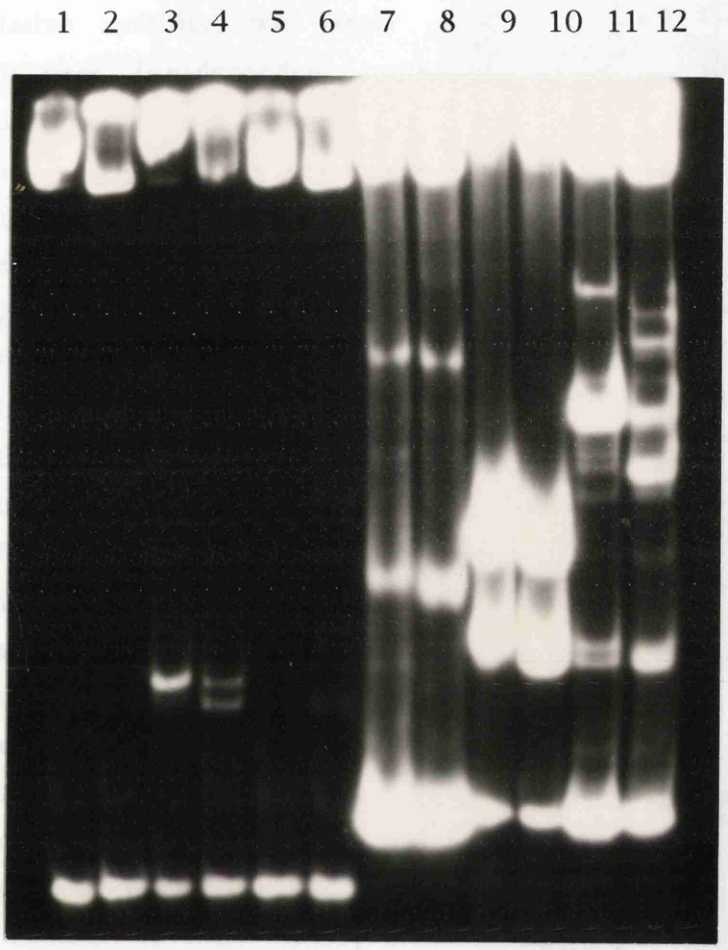


Figure 4.28 Gel retardation assay showing the comparative band patterns obtained using TnsB peptides encoded by pAG205 and pAG210 produced in LA547 and DS941; all extracts were a 4fold dilution of the stock. The two labelled DNA fragments were R4 (from pAG095) and R3+R4 (from pAG096).

1	R4 motif	with	LA547(pAG205) extract
2	R4 motif	with	DS941(pAG205)
3	R4 motif	with	LA547(pAG210)
4	R4 motif	with	DS941(pAG210)
5	R4 motif	with	LA547(pMR207)
6	R4 motif	with	DS941(pMR207)
7	R3 & R4 motifs	with	LA547(pAG205)
8	R3 & R4 motifs	with	DS941(pAG205)
9	R3 & R4 motifs	with	LA547(pAG210)
10	R3 & R4 motifs	with	DS941(pAG210)
11	R3 & R4 motifs	with	LA547(pMR207)
12	R3 & R4 motifs	with	DS941(pMR207)

than the considerably larger TnsB peptide encoded by pAG210 (Lane 9). There is no specifically defined relationship between protein molecular weight and relative mobility, but larger proteins generally retard DNA more than small ones. *lon*-dependent proteolysis of the TnsB peptides, suggested in Section 4.6.2, can now be ruled out, although proteolysis by other means may still occur. Possible alternative explanations for these retardation patterns are firstly that there is a substantial charge difference between the peptides encoded by pAG205 and pAG209, and that this is affecting mobility; or, secondly, that the stoicheiometry of binding to each motif differs from peptide to peptide.

Proteolysis of the peptides encoded by pAG205 and 210, pAG085, 089, 078, and 080 (the latter of which encode the same peptides as pAG205, 210, 211 and 209 respectively) was looked for. LA547 was transformed with these plasmids; transformants containing pAG205 and pAG210 grew very slowly and, as mentioned above, it proved impossible to transform LA547 with the other pMR78-based plasmids. Mid-log phase cells were induced with 1mM IPTG and grown for a further 4-6 hours; cells were pelleted and whole cells were run immediately on a 12.5% or 15% polyacrylamide Laemmli gel. On the three occasions on which this was carried out, there was no visible evidence of proteolysis of the expressed TnsB peptides (data not shown), although Western blotting of the gels, had suitable antibodies been available, might have revealed the presence of low concentrations of proteolytic products which were responsible for the extra autoradiograph bands seen.

4.6.5 Investigation of cooperativity in binding

In many systems involving binding of more than one DNA molecule to a segment of DNA, there is a cooperative effect: complex stability is increased as the number of protein molecules bound to the DNA is increased. This may occur with separated sites where DNA looping is possible, for example the three operator sites bound by the *deoR* repressor protein (Amouyal *et al.*, 1989) or where there are adjacent sites and no looping is necessary to bring protein molecules into direct contact, as seen with the Tn*3* *res* site and resolvase enzyme (Bednarz, 1989). The effect of increasing the protein concentration on the percentage occupancy of the protein binding sites on DNA is illustrated in Figure 4.29; in the absence of any cooperativity, a curve plotted is hyperbolic in shape whereas, when cooperativity occurs, the curve is sigmoid: there is an initial lag while the

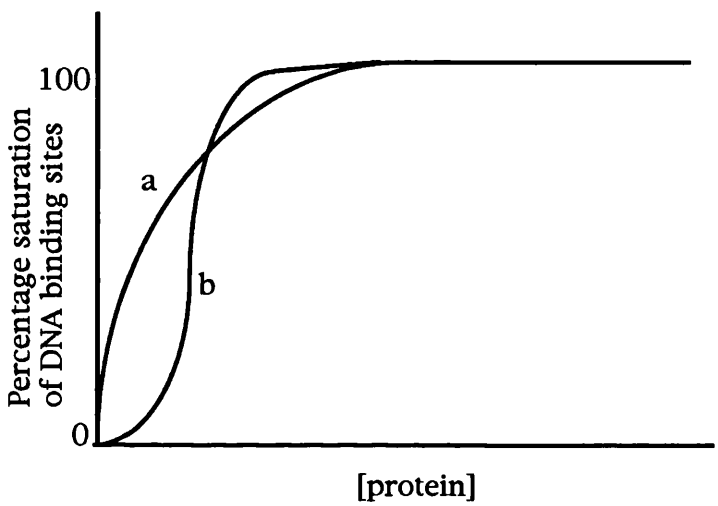


Figure 4.29 The effect of protein concentration on occupancy of the different binding sites on a DNA molecule
a in the absence of any cooperativity between different binding protein molecules;
b in the presence of cooperativity.

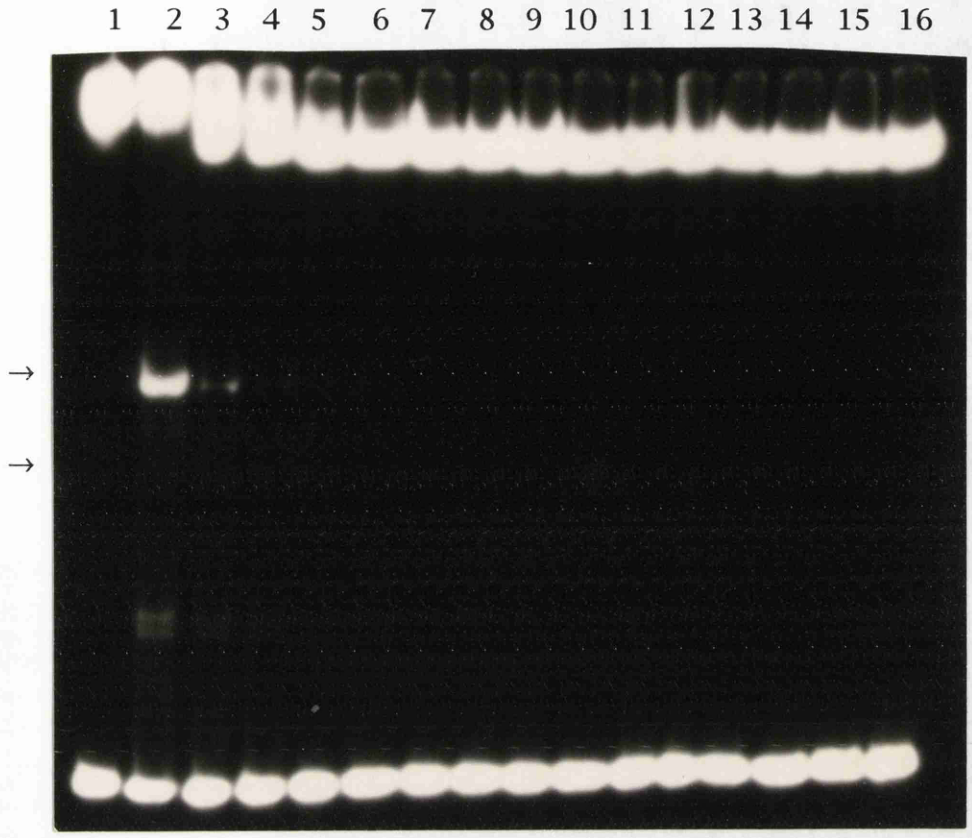


Figure 4.30 Gel retardation assay showing the effect of decrease in protein concentration on TnsB-specific binding to the labelled R3 and R4 motifs of pAG096 DNA. The protein extracts used here were made from LA547(pMR78) and LA547(pMR207). The retarded bands specific to full-size TnsB are indicated by arrows.

1	pMR78	undiluted extract
2	pMR207	undiluted extract
3	pMR207	2^{-1} dilution of extract
4	pMR207	2^{-2} dilution
5	pMR207	2^{-3} dilution
6	pMR207	2^{-4} dilution
7	pMR207	2^{-5} dilution
8	pMR207	2^{-6} dilution
9	pMR207	2^{-7} dilution
10	pMR207	2^{-8} dilution
11	pMR207	2^{-9} dilution
12	pMR207	2^{-10} dilution
13	pMR207	2^{-11} dilution
14	pMR207	2^{-12} dilution
15	pMR207	2^{-13} dilution
16	pMR207	2^{-14} dilution

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

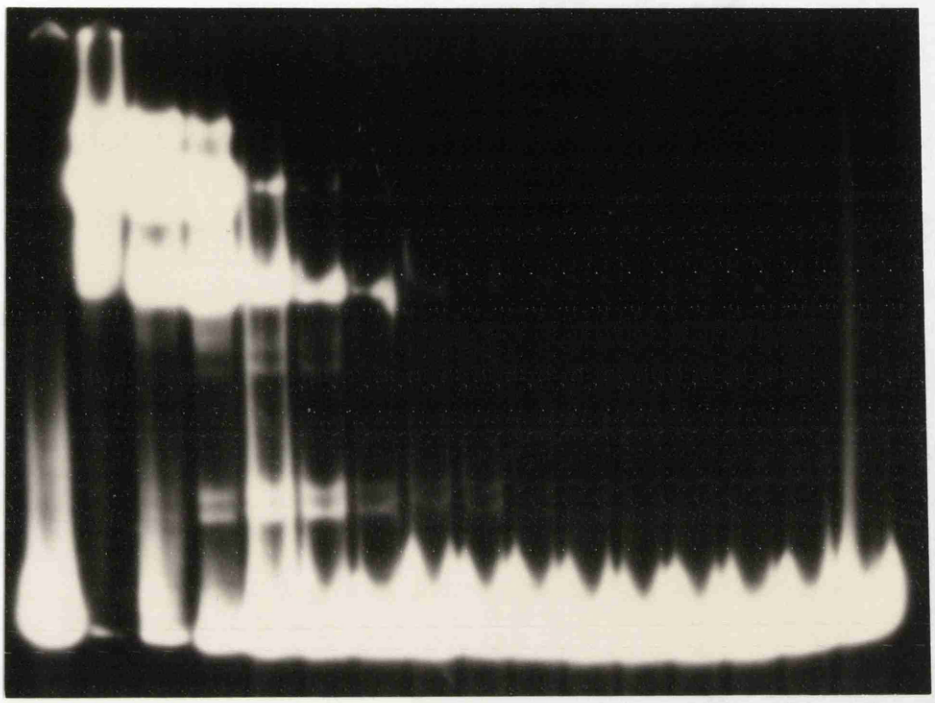


Figure 4.31 Gel retardation assay showing the effect of decrease in protein concentration on TnsB-specific binding to the labelled R1-R3 motifs of pLA50 DNA. The protein extracts used here were made from LA547(pMR78) and LA547(pMR207).

1	pMR78	undiluted extract
2	pMR207	undiluted extract
3	pMR207	2^{-1} dilution of extract
4	pMR207	2^{-2} dilution
5	pMR207	2^{-3} dilution
6	pMR207	2^{-4} dilution
7	pMR207	2^{-5} dilution
8	pMR207	2^{-6} dilution
9	pMR207	2^{-7} dilution
10	pMR207	2^{-8} dilution
11	pMR207	2^{-9} dilution
12	pMR207	2^{-10} dilution
13	pMR207	2^{-11} dilution
14	pMR207	2^{-12} dilution
15	pMR207	2^{-13} dilution
16	pMR207	2^{-14} dilution

1 2 3 4 5 6 7 8 9 10 11 12 13 14

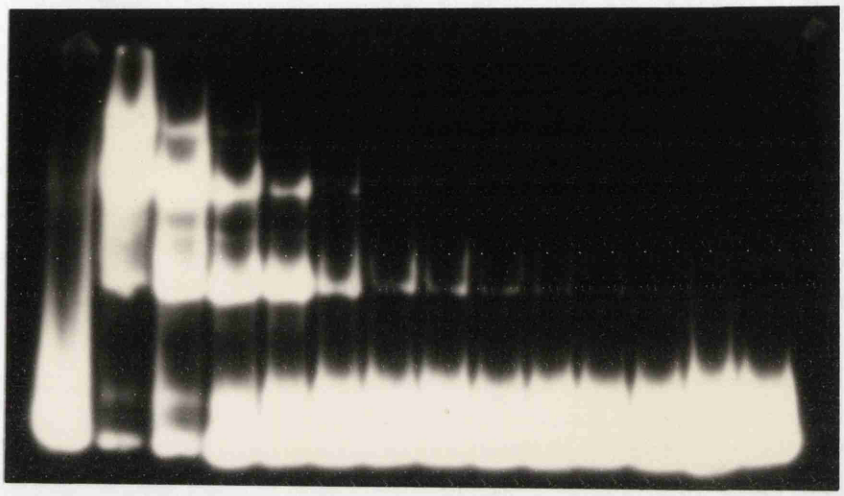


Figure 4.32 Gel retardation assay showing the effect of decrease in protein concentration on TnsB-specific binding to the L1-L3 motifs of pMR11 DNA cut with EcoRI and HincII and labelled. The protein extracts used here were made from LA547(pMR78) and LA547(pMR207).

1	pMR78	undiluted extract
2	pMR207	undiluted extract
3	pMR207	2^{-1} dilution of extract
4	pMR207	2^{-2} dilution
5	pMR207	2^{-3} dilution
6	pMR207	2^{-4} dilution
7	pMR207	2^{-5} dilution
8	pMR207	2^{-6} dilution
9	pMR207	2^{-7} dilution
10	pMR207	2^{-8} dilution
11	pMR207	2^{-9} dilution
12	pMR207	2^{-10} dilution
13	pMR207	2^{-11} dilution
14	pMR207	2^{-12} dilution

first site on each DNA molecule becomes loaded with protein, then the curve rises steeply as occupancy of other sites is facilitated by the presence of the first protein molecule. In terms of gel binding assays, lack of cooperativity is detected qualitatively as gradual transition from one retarded band to the next, more retarded one on an autoradiograph as the protein concentration is increased, whereas cooperativity is suggested by a rapid transition to the most retarded complex. Quantitative analysis of cooperativity where more than two protein molecules are involved is, however, probably not possible because the number of variables involved is apparently greater than the number of equations linking them (Senear and Brenowitz, 1991). Such analysis has therefore not been attempted here.

It was decided to investigate whether cooperativity could be detected in the binding of TnsB to the ends of Tn*7*. A TnsB-containing protein extract of LA547 was used for these assays since extraneous bands due to binding by proteolytic peptides would thus be minimised.

Three different ^{32}P end-labelled DNA fragments were used: the Tn*7* sequences encoded by pAG096 EcoRI-Hin*d*III (2 motifs of the RE), pLA50 EcoRI-SalI (3 motifs of the RE) and pMR11 EcoRI-Hinc*II* (all 3 motifs of the LE; see Figure 5.4), As depicted in Figure 4.10. In the first two of these, the DNA motifs are contiguous and thus any cooperativity in binding would be expected to be mediated by direct protein-protein contacts, whereas in the LE fragment the spacing between the motifs is greater and DNA looping might be required to bring about protein-protein contact.

Three binding assays were therefore carried out, each using one of the DNA segments listed above and twofold serial dilutions of LA547 (pMR207) protein extract. Results are shown in Figure 4.30 (2 RE motifs), Figure 4.31 (3 RE motifs) and Figure 4.32 (the 3 LE motifs). It can be seen that, in all three cases, where the protein concentration is lowest, binding is detected as occupancy of the least retarded band and, as the protein concentration is increased, higher bands representing less mobile complexes are gradually occupied. There is no sudden jump in occupancy from one band to another with increased protein concentration as might be expected in the presence of a cooperative effect, and in no case is a lower band of less intensity than an upper band in the same reaction; these observations suggest that there is probably no cooperativity in the sequential binding of TnsB molecules to either the right end or left end of Tn*7*. It is, however, possible that some cooperativity in binding is exhibited when the two Tn*7* ends are together, in other words, that synapsis of the

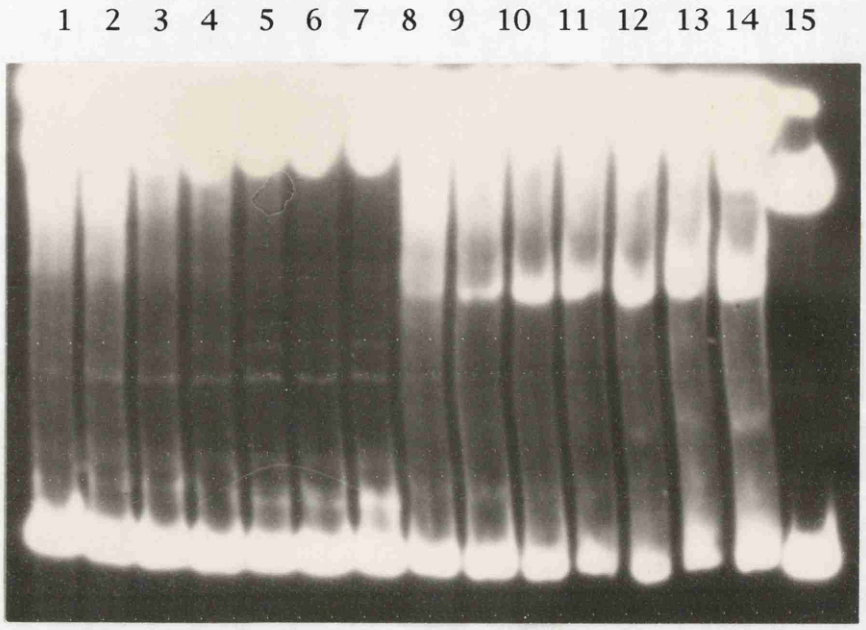


Figure 4.33 Gel retardation assay to investigate the stability of binding of the TnsB peptide encoded by pAG205 to the complete right end of Tn*7* cloned in pNE200 and radioactively labelled. Undiluted protein extracts of DS941(pMR78) and DS941(pAG205) were used; reactions were carried out at 37° for the times shown below before quenching on ice and loading on the gel. The results are shown graphically on Figure 4.34.

1	pMR78	90 minutes incubation
2	pMR78	60 minutes
3	pMR78	30 minutes
4	pMR78	20 minutes
5	pMR78	10 minutes
6	pMR78	5 minutes
7	pMR78	2 minutes
8	pAG205	90 minutes
9	pAG205	60 minutes
10	pAG205	30 minutes
11	pAG205	20 minutes
12	pAG205	10 minutes
13	pAG205	5 minutes
14	pAG205	2 minutes
15	no protein	30 minutes

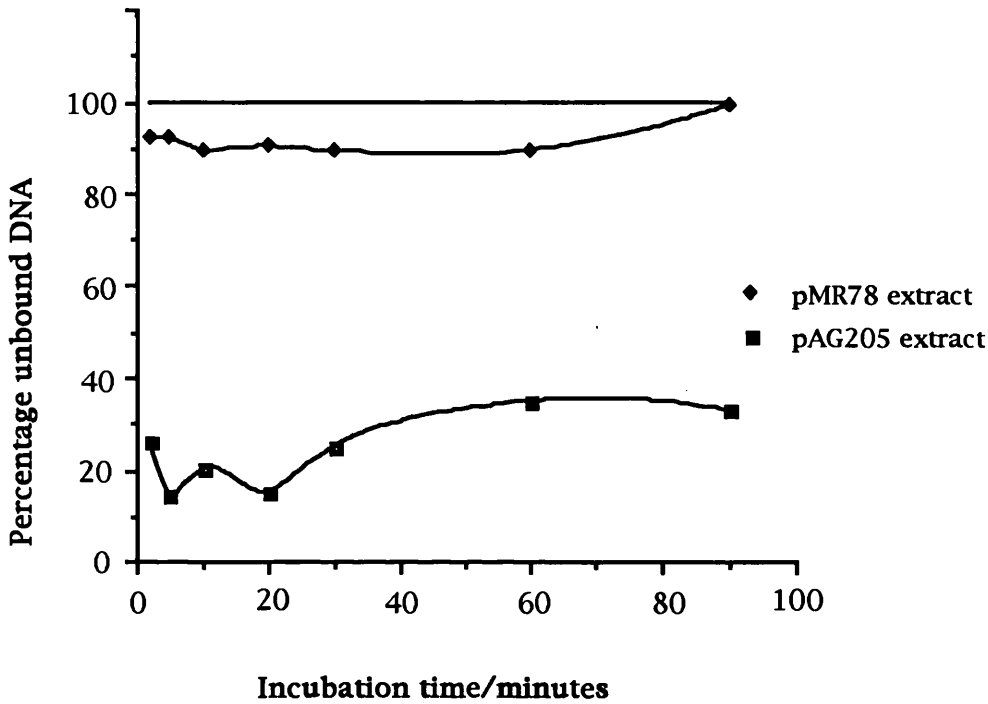


Figure 4.34 Stability of binding of the amino-terminal peptide of TnsB encoded by pAG205, compared with the control extract from cells containing the cloning vector pMR78.

Tn*7* ends occurs before the 22bp motifs are all occupied, and that occupation of the remaining motifs occurs rapidly upon synapsis.

4.6.6 Stability of binding

While the association of a protein-DNA complex in solution generally occurs by random collision and is dependent on physical factors such as the concentrations of the protein and DNA and the temperature of the solution, the dissociation rate depends on the strength of the specific interaction between the protein and DNA and is therefore the relevant factor in determining the relative stability of any specific complex when factors such as concentrations are fixed.

Stability of binding to the Tn*7* RE, as encoded by pNE200, was investigated for the minimal binding peptide encoded by pAG205, and also for the wildtype TnsB (the latter produced in the protease-deficient strain LA547).

A crude extract, made under standard conditions, of DS941(pAG205) was used to test the stability of the N-terminal TnsB peptide encoded by sequence truncated at the Hin*c*II site. This peptide produced in LA547 appears to have the same binding properties (as discussed in Section 4.6.4), and so no separate stability assay was carried out with LA547(pAG205) protein extract. It had previously been shown that specific binding of the peptide to DNA had reached equilibrium within five minutes, and so the procedure followed here would not be complicated by the concurrent establishment of an equilibrium state. As shown in Figure 4.33, incubation of the extract with ^{32}P-labelled pNE200 DNA at 37° was carried out for between 2 and 90 minutes, followed by quenching on ice before loading on the gel. After autoradiography of the dried gel, the unbound DNA was excised from the gel and the ^{32}P activity of each sample assayed in a Beckman scintillation counter. The unbound, rather than bound, DNA was used because loss of activity from this band is more clear-cut than appearance of activity elsewhere on the gel, and protein-DNA complexes which dissociate during electrophoresis can also be accounted for by assaying the unretarded DNA. As can be seen on Figure 4.34, incubation at 37° for between 5 and 20 minutes yields optimum binding. After 20 minutes incubation, less of the DNA is complexed with protein, possibly because the protein has been denatured.

A second assay of DNA-protein complex stability was carried out using wildtype TnsB made in the protease-deficient strain LA547, as

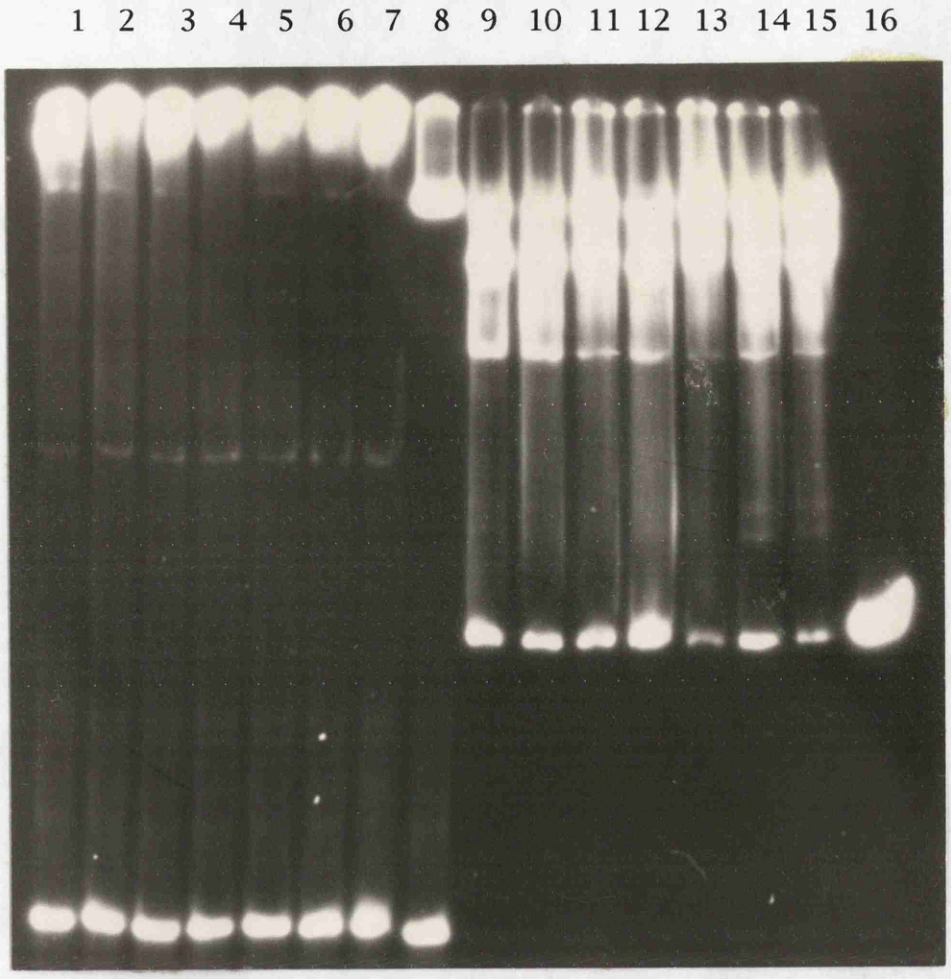


Figure 4.35 Gel retardation assay to investigate the stability of binding of TnsB to a single labelled R4 site of pAG095 and to the complete labelled Tn*7* right end on pNE200. The protein extract used throughout was undiluted LA547(pMR207). Reactions were at 37° for the times shown below, and were quenched on ice before loading on the gel.

1	R4 motif	60 minutes incubation	
2	R4 motif	45 minutes	
3	R4 motif	30 minutes	
4	R4 motif	20 minutes	
5	R4 motif	10 minutes	
6	R4 motif	5 minutes	
7	R4 motif	1 minute	
8	R4 motif	30 minutes	no protein
9	R1-R4 motifs	60 minutes	
10	R1-R4 motifs	45 minutes	
11	R1-R4 motifs	30 minutes	
12	R1-R4 motifs	20 minutes	
13	R1-R4 motifs	10 minutes	
14	R1-R4 motifs	5 minutes	
15	R1-R4 motifs	1 minute	
16	R1-R4 motifs	30 minutes	no protein

described in Section 4.6.4. As can be seen on Figure 4.35, contrary to what was observed above with the N-terminal peptide, binding of TnsB both to an isolated R4 motif and to a complete Tn7 right end is stable for at least 60 minutes.

It is not known how stability of binding of the other constructed TnsB peptides compares with that of the wildtype TnsB or the peptide tested above, but the results above suggest that the rôle of TnsB proteolysis products in *in vivo* transposition may become less significant, the longer binding to the end motifs takes place for. The *in vitro* transposition protocol of Bainton *et al.* (1991) requires incubation at 30° for 37 minutes, but the *in vivo* conditions are likely to be very different from those developed by Bainton *et al.*, and it is possible that transposition takes place much more quickly *in vivo* than *in vitro*.

4.7 Regulation of *tns* promoter activity by amino-terminal fragments of TnsB: galactokinase assays

The presumed -35 box of the *tns* promoter overlaps the R4 22bp motif, thus making possible the regulation of both transposition and transcription of the *tns* genes. It was shown by Ekaterinaki (1987) that TnsB could repress transcription from this promoter. It was not known however whether TnsB peptides, competent for binding but not for transposition, could also regulate transcription in a manner similar to that found in, for example, IS*1*; if this were indeed the case, repression of transcription and thus regulation of transposition might more efficiently take place.

Activity of the *tns* promoter in the presence of fragments of TnsB was studied by means of transcriptional fusions to a plasmid copy of the *galK* gene using extracts from the host strain DS941, which has the *galK* genotype. Galactokinase catalyses the reaction

galactose + ATP → galactose-1-phosphate.

Galactokinase activity can easily be assayed *in vitro* using ^{14}C labelled galactose as substrate, and measuring selective retention of the phosphorylated product on Whatman DE81 paper, using the method of McKenney *et al* (1981) as described in Chapter 2. It is assumed that any difference recorded in galactokinase activity is a reflection of difference in transcription of the *galK* gene, rather than any subsequent step in gene expression, since post-transcriptional steps, other than possibly stability of

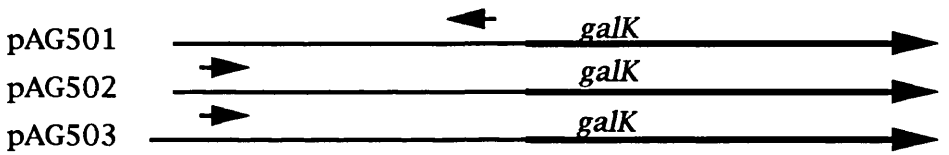


Figure 4.36 Arrangement of the Tn *7* sequence and *galK* gene in pAG501, pAG502 and pAG503. The thin line represents Tn *7* sequence, and the arrow above it indicates the position and orientation of the *tns* promoter.

a

	galK encoding plasmid			
	p253	pAG501	pAG502	pAG503
pUC18	2.7 ± 2.1	40.9 ± 23.6	512.8 ± 199.0	470.1 ± 198.2

b

		galK encoding plasmid		
		pAG501	pAG502	pAG503
plasmid encoding TnsB peptide	pAG085	1.18 ±0.39	1.04 ± 0.07	1.04 ± 0.13
	pAG089	0.56 ± 0.10	0.58 ± 0.05	0.80 ± 0.07
	pAG078	0.62 ± 0.06	0.47 ± 0.03	0.56 ± 0.01
	pAG080	0.76 ± 0.39	0.98 ± 0.10	0.76 ± 0.41
	pAG086	0.62 ± 0.33	0.26 ± 0.23	0.29 ± 0.32
	pAG077	1.03 ± 0.21	0.98 ± 0.06	0.96 ± 0.07

Table 4.1 **a** *tns* promoter activity measured as McKenney units of galactokinase activity, using the *galK* plasmid indicated at the top and in the presence of pUC18. The galactokinase units are expressed as nanomoles of galactokinase phosphorylated $minute^{-1}$ ml^{-1} of cells at OD_{650} of 1.0.

b Ratio of galactokinase activity in the presence of the TnsB peptide producer indicated, compared with activity in the presence of pUC18.

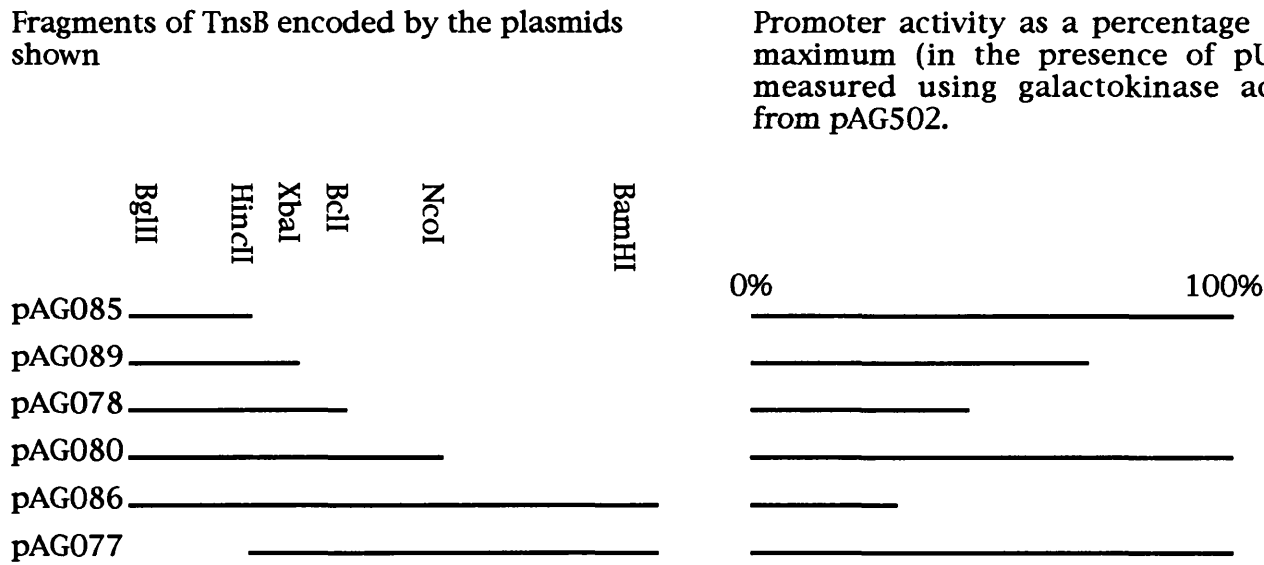


Figure 4.37 Left: the different TnsB peptides used in galactokinase assays;
Right: the promoter activity measured in the presence of the peptides indicated.

the 5' end of the transcript, are unlikely to be affected by the cloning procedure.

For reasons of plasmid compatibility, the standard *galK* transcriptional fusion vector, pKO500 (McKenney) was not able to be used. In its place the p15A origin vector p253, constructed by M. Burke (Slatter, 1987), was utilised; it contains a rudimentary polylinker sequence upstream of a promoterless *galK* gene; host cells were tested and showed no inherent galactokinase activity. Three different *tns* promoter insertions were constructed; these are shown in Figure 4.36. pAG503 contains the whole of the Tn*7* right end up to the AccI site in the *tnsA* coding sequence, taken from pMR88 as a Hin*d*III-AccI fragment and cloned into the SmaI and Hin*d*III sites of the p253 polylinker, such that the *tns* promoter is correctly oriented for controlling transcription of *galK*. pAG502 is similar to pAG503, but has the outermost 37bp of the Tn*7* right end, including R1 and half of R2, deleted by cleavage at the BglII site in R2, cloned into the BamHI and SmaI sites of p253; the 'traditional' promoter region with -35 and -10 boxes is intact, but it was considered possible that transcription was to some degree dependent on interactions between this region and sequences further upstream, and so this latter plasmid was also tested. Neither pAG502 nor pAG503 appears from sequence data to have a transcriptional terminator between the promoter and the *galK* gene. In pAG501, the insert is as in pAG502 but is in the reverse orientation so that transcription would be expected to be in the opposite direction to the *galK* reading frame; use of pAG501 is discussed further in Section 4.8.

TnsB polypeptides with various deletions were encoded by pAG085, pAG089, pAG078, pAG080, pAG077 and pAG086 (complete TnsB) as described in Section 4.4, and the vector pUC18 was also tested. The *tnsB* coding sequence in these plasmids is shown schematically in Figure 4.8. Construction of and use of pAG077, which encodes an N-terminal deletion of TnsB, is discussed further in Sections 4.11 and 4.12.

For each assay, DS941 was transformed with one of the *galK* plasmids and one of the TnsB-producing plasmids, and TnsB production was induced by addition of 1mM IPTG. Each plasmid combination was tested on separate occasions between 2 and 8 times, and the results are as shown in Table 4.1, and schematically in Figure 4.37. It was assumed that intracellular concentration of the TnsB peptides was the same; this could not be confirmed in the absence of antibodies, since the protein concentrations were too low for bands to be detectable on a silver-stained Laemmli gel,

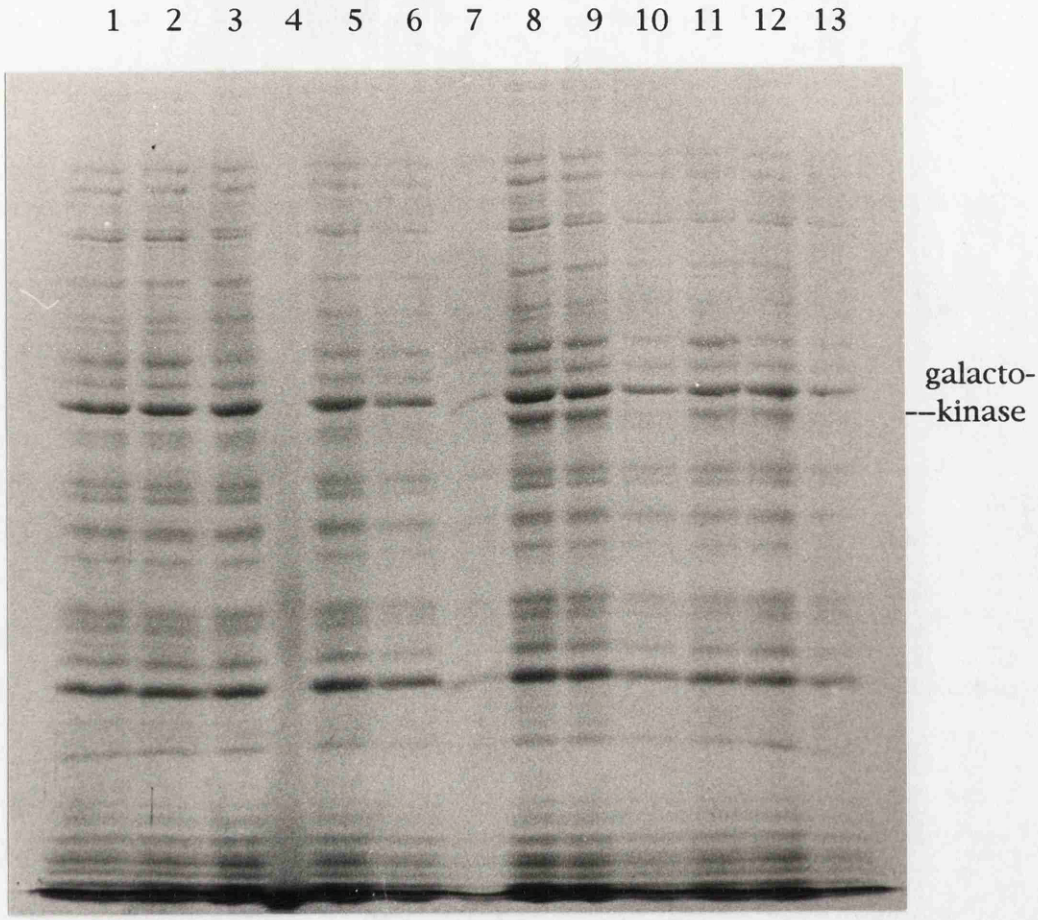


Figure 4.38 Coomassie stained 12.5% polyacrylamide Laemmli gel of whole cells used for galactokinase assays, showing the presence in cells of galactokinase where induced. DS941 harboured one of the four galactokinase-encoding plasmids p253, pAG501, pAG502 and pAG503, depicted on Figure 4.36, and one of the TnsB-producing plasmids, shown in Figure 4.8: pAG085 (N' peptide), pAG086 (complete TnsB) and pUC18 (cloning vector - no TnsB).

1	p253	+	pUC18
2	p253	+	pAG085
3	p253	+	pAG086
4	molecular weight markers		
5	pAG501	+	pUC18
6	pAG501	+	pAG085
7	pAG501	+	pAG086
8	pAG502	+	pUC18
9	pAG502	+	pAG085
10	pAG502	+	pAG086
11	pAG503	+	pUC18
12	pAG503	+	pAG085
13	pAG503	+	pAG086

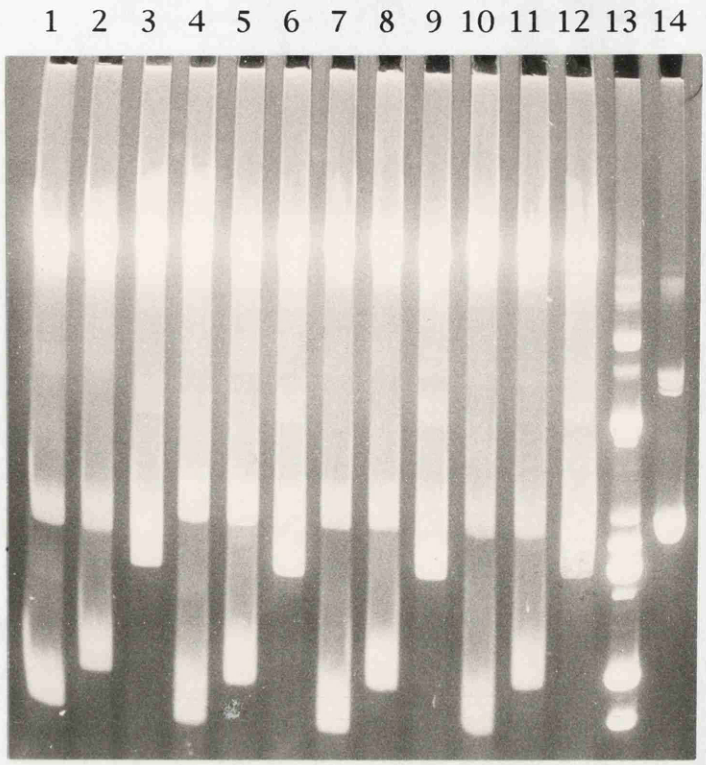


Figure 4.39 Ethidium stained single colony agarose gel of cultures used for galactokinase assays, showing the presence within cells of the plasmids encoding galactokinase and, at much higher copy number, the TnsB-encoding plasmids.

1	p253	+	pUC18
2	p253	+	pAG085
3	p253	+	pAG086
4	pAG501	+	pUC18
5	pAG501	+	pAG085
6	pAG501	+	pAG086
7	pAG502	+	pUC18
8	pAG502	+	pAG085
9	pAG502	+	pAG086
10	pAG503	+	pUC18
11	pAG503	+	pAG085
12	pAG503	+	pAG086
13	pUC18, pAG085, pAG086 DNA		
14	p253, pAG501, pAG502, pAG503 DNA		

although galactokinase could be seen (Figure 4.38). Copy number of the participating plasmids could be judged crudely by running cell lysates on an agarose gel and ethidium staining; see Figure 4.39 for an example of this.

It can be seen from the promoter strengths in the presence of pUC18 (Table 4.1a) that there is little difference between the pAG502 and pAG503 inserts; it is possible that in the presence of all four 22bp motifs (pAG503) transcription might be slightly reduced, perhaps due to intrinsic bending of the DNA (Morrell, 1990) inhibiting access of the transcription complex.

There was considerable variation between results obtained on different occasions, as can be seen from Table 4.1b; it is not known whether this was due to variable concentrations of the TnsB peptides, or to aspects of the *in vitro* assay. Also, in general, results with pAG503 exhibited greater variation than with pAG502; this observation may not be significant; alternatively, it may give an indication of the complete Tn*7* right end adopting different conformations on different occasions when larger portions of TnsB were used.

The results of the assays, as shown in Figure 4.37, show that no repression of the *tns* promoter occurs in the presence of TnsB N-terminal peptides encoded by pAG085 and pAG080. Since the peptide encoded by pAG085 apparently does not bind to the isolated 22bp motif encoding the presumed -35 box of the *tns* promoter, it is not altogether surprising that this peptide does not control transcription. It is more remarkable that the much larger peptide encoded by pAG080 also fails to repress the promoter, since clear repression is exhibited by smaller peptides terminating at the DNA XbaI and BclI sites. It is possible that the amino acid sequence encoded by the region between the BclI and NcoI sites adopts an aberrant conformation in the pAG080-encoded peptide, and that this inhibits binding by upstream sequence to the promoter; however, this peptide appears on binding gels (Section 4.6) to behave in a similar fashion to that of the smaller peptides encoded by pAG089 and pAG078, and does not obviously exhibit the lack of stability which might manifest itself as displacement from the DNA R4 motif when in competition with RNA polymerase for the binding site.

Repression of the *tns* promoter clearly occurs however in the presence of peptides encoded by pAG089, pAG078 and pAG086 (wildtype TnsB), with the latter exhibiting the greatest effect. All three of these species have also been shown to bind to the isolated R4 22bp motif, as

described in Section 4.6. It is therefore plausible that TnsB cleavage products, naturally produced by an unknown mechanism under certain growth conditions, may *in vivo* regulate transcription from the *tns* promoter, and thus control the frequency of transposition; if these were in excess over wildtype TnsB, then transposition would be inhibited both indirectly, through transcriptional repression, and directly, through binding to the terminal 22bp motifs of a protein species which was not competent for transposition. However, the results of investigations into the relative stability of binding, described in Section 4.6.6, suggest that the wildtype TnsB binds more stably over time at least than the amino-terminal peptide which was tested and thus that, unless the truncated peptide(s) were in vast molar excess over the wildtype protein, they might not have any significant effect on repression of transcription.

4.8 An antisense promoter near the Tn*7* right end

It is also notable from the galactokinase assays carried out that there is low, but consistent, promoter activity from pAG501, where the *tns* promoter is cloned in opposition to the *galK* gene. The DNA sequence of this region was examined for indications of a σ^{70}, σ^{54} or σ^{28} promoter, and a putative σ^{70} promoter was identified in the coding region of *tnsA*. This has the sequence:

TTGCAA..19bp..TAATAA

with the first of these bases 417bp from the Tn*7* right end and the corresponding transcriptional start probably 378bp from the right end. While the proposed -35 and -10 boxes bear reasonable similarity to the consensus sequence (see Section 3.2.1), the spacing between them is greater than usual, and thus local distortion such as unwinding of the double helix would be necessary for RNA polymerase to contact both boxes simultaneously. Promoters with this spacing are however known, for example that of the *mer* operon of Tn*501* (Lund *et al.*, 1986).

If transcription does indeed initiate here under normal *in vivo* conditions, the question arises as to what rôle, if any, this antisense transcript plays in the regulation of transposition. Involvement of antisense RNA is known in a wide variety of biological systems, such as replication of plasmid ColE1 (Tomizawa, 1984) and retroviruses (Borisenko *et al.*, 1992), plant photosynthesis (Hudson *et al.*, 1992) and fruit ripening (MacLachlan and Brady, 1992). In the case of transposon Tn*10*, a stable

transcript initiating at an antisense promoter, pOUT (Simons and Kleckner, 1983), base-pairs with the less stable mRNA encoding the transposase, blocks ribosome binding to this latter RNA (Ma and Simons, 1990) and also renders it susceptible to ribonucleaseIII cleavage (Case *et al.*, 1990).

Unlike the situation with Tn*10*, the Tn*7* antisense promoter is apparently much weaker than the *tns* promoter (compare the galactokinase activities in Table 4.1a). The length of the antisense transcript is uncertain; a search for transcriptional terminators downstream of this promoter failed to detect any in the Tn*7* sequence (the possible terminator sequence mentioned in Section 3.2.2 is upstream of this promoter), so it is possible that the RNA continues to be transcribed from target DNA until a host terminator is reached. There is no information on the relative stability of the two transcripts which might otherwise have indicated how the balance between them maintained transposition at the observed *in vivo* frequency. It is not known whether transcription from the antisense promoter is constitutive, or whether it is regulated by a Tn*7*- or host-encoded protein. There is no 22bp motif adjacent to the putative promoter sequence and so it might appear that TnsB is unlikely to control transposition; study of Table 4.1b, however, suggests that transcription decreases in the presence of a high concentration of TnsB and several of the larger C-terminal-truncated peptides of it, and so under normal circumstances TnsB concentration may in fact be a determinant of antisense transcription. It cannot be saturation of the cellular transcription system through induction of the plasmid-encoded *lac* promoter which is responsible for this effect, since no reduction in antisense transcription is observed in the presence of pAG085 or pAG077. The effect of TnsD concentration on antisense transcription has not been tested; like TnsB, TnsD requires a specific sequence for binding, and no such sequence is evident adjacent to the promoter. TnsC binds nonspecifically to DNA (Gamas and Craig, 1992), and so it is conceivable that it binds to an operator sequence associated with the antisense promoter; this has not been investigated. A host protein may also regulate transcription.

Since the antisense promoter strength is low relative to that of the *tns* promoter, unless its stability is much greater than that of the *tns* promoter, it is unlikely to play a significant rôle in transposition.

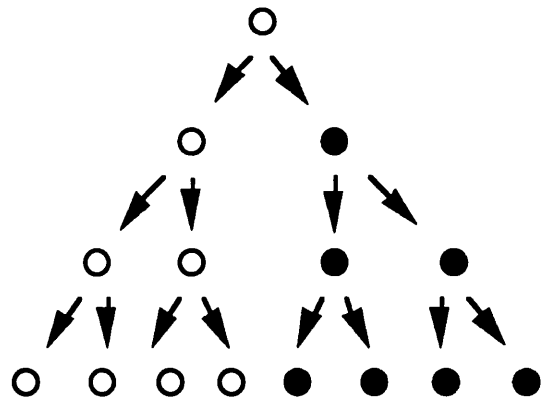
4.9 Regulation of transposition frequency: *in vivo* transposition assays

The various sized fragments of TnsB had been tested for their ability to bind DNA *in vitro* and to regulate transcription from the *tns* promoter. To ascertain whether they were likely to have any effect on transposition *in vivo*, it was also necessary to measure transposition rates in the presence and absence of an excess of the TnsB peptides over wildtype TnsB; an effect on transposition could be mediated either through regulation of *tns* promoter activity or through blocking of 22bp motifs to access by full-size TnsB.

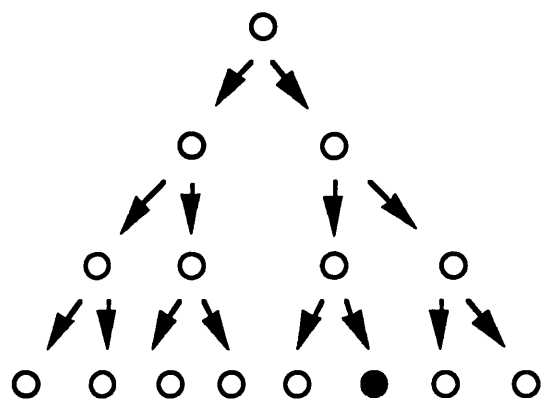
The system widely used for the measuring of transposition rates *in vivo* involves a mate-out assay. The transposon is initially at a plasmid or chromosomal *attTn7* site in a suitable donor strain; if a conjugative plasmid is also present, transposition into the plasmid may occur. This plasmid is then mated out into a second strain with an appropriate selectable marker. The number of recipient colonies harbouring the conjugative plasmid, and the proportion of those also containing the transposon are measured, and the transposition frequency is defined as

$$\frac{\text{total number of recipients containing plasmid with transposon}}{\text{total number of recipients of plasmid.}}$$

There is a slight difference in conjugation frequency between plasmids with and without a resident copy of Tn*7* (Rogers, 1986). More serious is the problem that, unless any external control mechanism (such as induction of a promoter normally completely repressed) is applied, transposition to the plasmid could occur at any time during which the two are both present in the donor cell. Since the mate-out assay is carried out the day following transformation of the cells with *tns* functions or acquisition of the target plasmid by the cells by conjugation, there are approximately 25 generations of growth overnight during which transposition could occur. It is not known whether transposition is equiprobable during this growth period; it is possible that there may be a burst of transposition immediately after all the necessary reaction components appear together in the cells, followed by quiescence.



One transposition event early on would result in all the daughter cells from this one reaction also containing the transposon in the target plasmid pEN300, as indicated by the presence of black circles in the diagram above; one transposition event immediately before conjugation would result in only this one cell containing the transposon in the plasmid, as shown below.



This method apparently therefore ought to give a frequency which is highly dependent on the time at which transposition takes place, with higher measured frequencies after longer growth periods. In practice, transposition frequencies measured in this way with the same system are remarkably constant, even when they are low and regardless of how long the transposon and target plasmid have coexisted in the cell (even for 3 or 4 days). One explanation for this is that cells in which transposition to the plasmid has occurred have lower viability than the others; this is not impossible, since the double strand chromosomal breaks believed to be formed as a result of Tn*7* excision may not always be repaired. Alternatively, the measured transposition rate may be a reflection of an

early burst of transposition, followed by a long transposition-free growth period during which the mate-out takes place.

Because of the constancy of frequencies measured in this way, and because no alternative system was available, the mate out assay was used here as an indication of the transposition competence of Tn*7* and its derivatives in different strain backgrounds. The conjugative plasmid used throughout this work was pEN300 (Rogers *et al.*, 1986), which is the naturally occurring plasmid R388 with a 1kb cloned chromosomal *attTn7* fragment. Samples of a population of colonies were used as starting cultures for the transposition assays; in this way, it was hoped to obtain a more representative indication of transposition frequency than by using a single colony.

Plate mating assays were therefore set up in DS902::Tn*7*; use of the *recA* strain was necessary to avoid possible complications in interpretation caused by recombination between the chromosomal Tn*7* and the plasmid-cloned *tnsB* sequence. The TnsB fragment was encoded on a pUC-based plasmid (pAG085, pAG089, pAG078, pAG080 or pAG086) as these had both a high copy number and good expression of the *tnsB* gene when induced with 1mM IPTG. As usual, the recipient plasmid was pEN300, which was selected for in this case by sulphonamide resistance; it was mated into MR1, with Tn*7* being now selected for by streptomycin resistance.

The measured mean transposition rates of wildtype Tn*7* in the presence of the plasmids listed were as follows:

pUC18	1.5 x 10-2
pAG085	1.3 x 10-2
pAG089	3.0 x 10-2
pAG078	8.1 x 10-3
pAG080	1.3 x 10-2
pAG086	1.1 x 10-4

It therefore appears that none of the truncated TnsB proteins tested could inhibit *in vivo* transposition to any significant extent when supplied *in trans* at high concentration. This is a little surprising since all the truncated peptides bind the Tn*7* right end at least; it may be that the relative stability of binding, as discussed in Section 4.6.5, is a factor in determining the influence of the TnsB fragments on transposition. It is possible however that, since the assays were not carried out in a protease-

deficient strain, truncated forms of TnsB were being produced naturally and that the transposition rate measured in the presence of pUC18 reflected an underlying level of TnsB fragments. It is unlikely that *in vivo* transposition *requires* TnsB fragments, since the *in vitro* assay developed by Bainton *et al.* (1991) functions with what is claimed to be pure full-size TnsB.

It was attempted to repeat these transposition assays in the *htpR* strain AG3, which was made from LA547 by P1 transduction of a chromosomal Tn*7*. It was found that the transposition rate of Tn*7* in this background was approximately 100-fold lower than in DS941 or DS902, at 10^{-4} although the conjugation frequency was unaltered, and that no transposition of Tn*7*-1 from the strain AG2 (8 isolates of which were made from LA547 by P1 transduction of chromosomal Tn*7*-1) could be detected when the *tns* genes were supplied *in trans*.

The effect of a high concentration of TnsB or peptides derived from it on transposition in AG3 was also studied. This system has the advantage over DS902::Tn*7* that the TnsB peptides in the cells are unlikely to be a truncated form of those coded for. It has the disadvantage that LA547 is rec^+, and so the results are likely to be complicated by homologous recombination between the chromosomally encoded *tnsB* and plasmid-encoded copies of it.

The measured transposition frequencies of Tn*7* from LA547 to pEN300 were:

pUC18	3.0×10^{-6}
pAG085	5.7×10^{-6}
pAG089	3.5×10^{-6}
pAG078	2.2×10^{-6}
pAG080	9.4×10^{-6}
pAG086	2.4×10^{-6}

This time, there was no significant change in the measured transposition frequency in the presence of any of the TnsB peptides encoded by the plasmids listed above. However, it can be seen that all the transposition rates determined in this assay are considerably lower than normally seen; the combination of transposition in the *htpR* background, and the presence of an extra plasmid appears to be highly detrimental to transposition, possibly because homologous recombination is occurring

between the transposon and the plasmid-borne transposon sequences, and the deleterious effects of this recombination on the transposon structure mask any variation in transposition frequency which might be due to the inhibiting effect of TnsB peptides. It is therefore difficult to interpret these results; it would be advisable for a future worker to construct a *recA htpR* strain for use in assays of this type.

4.10 Mutagenesis of the putative helix-turn-helix

An amino acid sequence with strong homology to that of the helix-turn-helix DNA binding motif has been identified in TnsB (Flores *et al.*, 1990). This is located in the region of TnsB shown to be involved in binding to the 22bp Tn*7* end motifs, and so the requirement for the precise amino acid sequence of this putative helix-turn-helix was investigated.

4.10.1 Construction of proteins with site-directed mutations

The standard helix-turn-helix motif of a DNA-binding protein consists of two α-helices, separated by helix-breaking amino acids which allow a bend angle of approximately 120°, and stabilised by interactions either within the individual molecule or with the partner in a dimer, as for example with FIS. The second (i.e. nearer the C-terminal) helix is the recognition helix, which lies in the major groove of the DNA and makes specific contacts with base pairs as well as hydrogen bonds to phosphate groups of the DNA backbone.

Four amino acids of the putative recognition helix identified by Flores *et al.* were subjected to site-directed mutagenesis using the method described in Chapter 2; the starting plasmid was pAG401, which encodes *tnsA* and the 5' end of *tnsB*, up to the HincII site, including the putative helix-turn-helix motif. The mutagenising oligonucleotides were selected so that each encoded a novel restriction site by which mutagenesis products could easily be identified; the oligonucleotides are shown in Figure 4.40. As can be seen from Figure 4.41, if this amino acid sequence is indeed in α-helical conformation, the four selected amino acids are dispersed around the helical axis and are not all expected to make contacts with the major groove of the DNA; it is most likely *a priori* that the side of the helix containing more charged and polar residues would be involved in interactions with the DNA, and hence the valine residue targeted for mutagenesis would probably not be in close contact with DNA. In the CAP

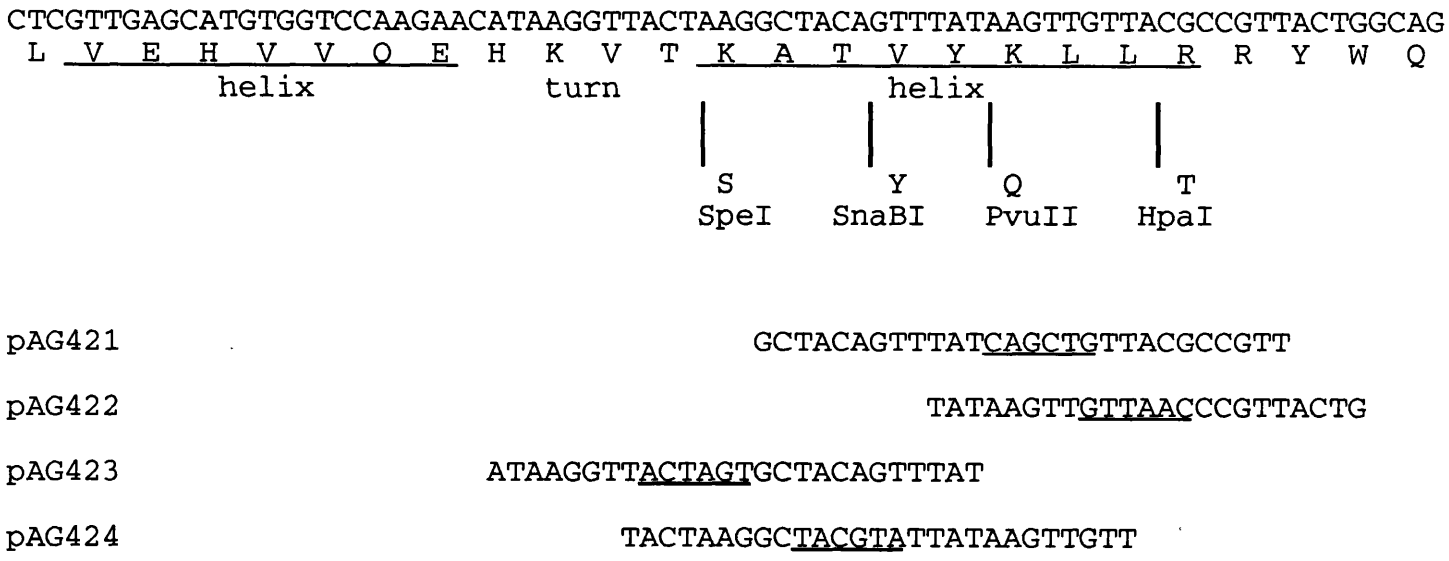


Figure 4.40 Construction of oligonucleotides used in the site-directed mutagenesis of the putative helix-turn-helix of TnsB.

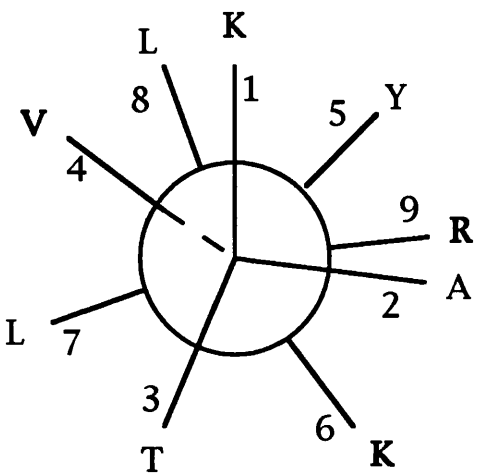


Figure 4.41 The putative recognition helix of TnsB, viewed along the helical axis, with numbering of residues from the N-terminal end. Amino acids selected for mutation are shown in bold type.

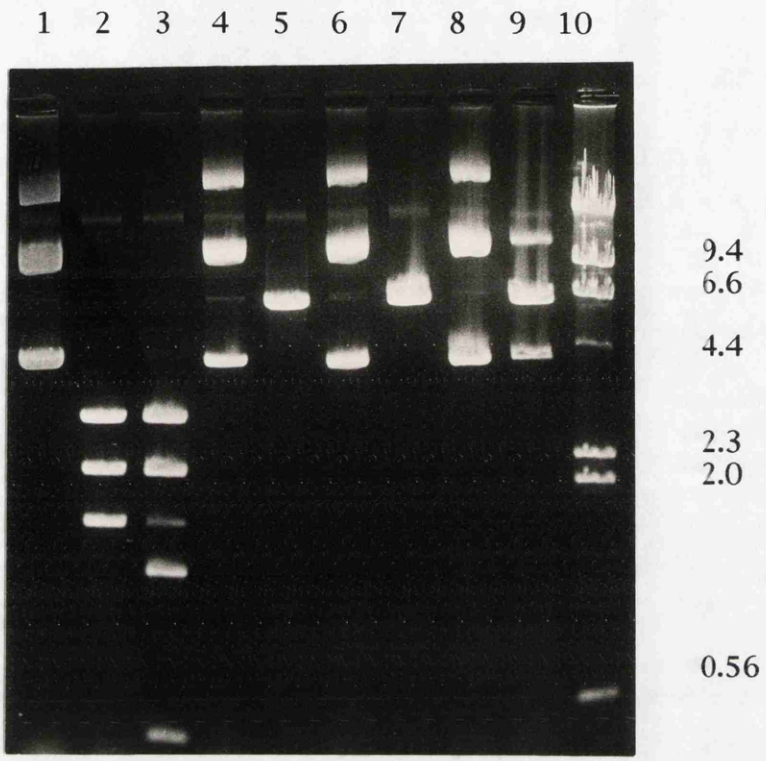


Figure 4.42 Ethidium stained 1% agarose gel showing digestion of pAG421, pAG422, pAG423 and pAG424 DNA by the diagnostic restriction enzyme. Sizes of the Lambda markers are in kb.

1	pAG401	uncut
2	pAG401	PvuII
3	pAG421	PvuII
4	pAG401	HpaI
5	pAG422	HpaI
6	pAG401	SpeI
7	pAG423	SpeI
8	pAG401	SnaBI
9	pAG424	SnaBI
10	λ	Hin*d*III

protein, it is known (S. Busby, pers. comm.) that the first two amino acids of the recognition helix are involved in specific binding to the DNA. The selected arginine residue is also part of the amino acid sequence which is common to TnsA and TnsB (see Section 3.1) and which, in TnsB, lies immediately downstream of the putative helix-turn-helix. Of the four mutations:

$K_{116}\rightarrow S$	(pAG423)
$V_{119}\rightarrow Y$	(pAG424)
$K_{121}\rightarrow Q$	(pAG421)
$R_{124}\rightarrow T$	(pAG422)

three substitute a charged residue with a polar one which would be predicted to maintain the integrity of any α-helix structure; the valine is substituted with tyrosine, which in some circumstances acts as a weak helix-breaker. The four resultant plasmids, pAG421, 422, 423 and 424 were checked by digestion with the appropriate diagnostic restriction enzyme, as shown in Figure 4.40 and Figure 4.42.

4.10.2 Binding assays with the mutagenised proteins

Crude cell extracts were made from DS941 harbouring pGP1-2, which encodes the bacteriophage T7 RNA polymerase gene, and one of the plasmids pAG421, pAG422, pAG423 or pAG424, in which the *tnsB* gene was under the control of the T7 gene 10 promoter; TnsB production had been induced by translation of the T7 RNA polymerase. pAG403 (which has a repaired *bla* gene but no other difference from pAG401) was also used as a control plasmid. No retardation of the R1-R4 DNA fragment of pNE200 was detected in the presence of crude extracts of cells containing any of the five plasmids. The reason for this is not clear, as the T7 promoter upstream of *tnsB* in these plasmids normally gives high expression.

The pUC-based plasmid pAG082 was used as the basis for a second set of plasmids harbouring the site-directed mutations in *tnsB*. These were made by substituting the BglII-Hin*c*II DNA fragment of pAG421, 422, 423 or 424 coding for the 5' end of *tnsB* for the corresponding fragment of pAG082. The resultant plasmids, named pAG431, 432, 433 and 434, retained the BglII and Hin*c*II sites at the junctions of the inserted sequence and that of pAG082, as confirmed by restriction analysis (not shown), and each also had the introduced site of its parent plasmid as shown in Figure 4.43. These

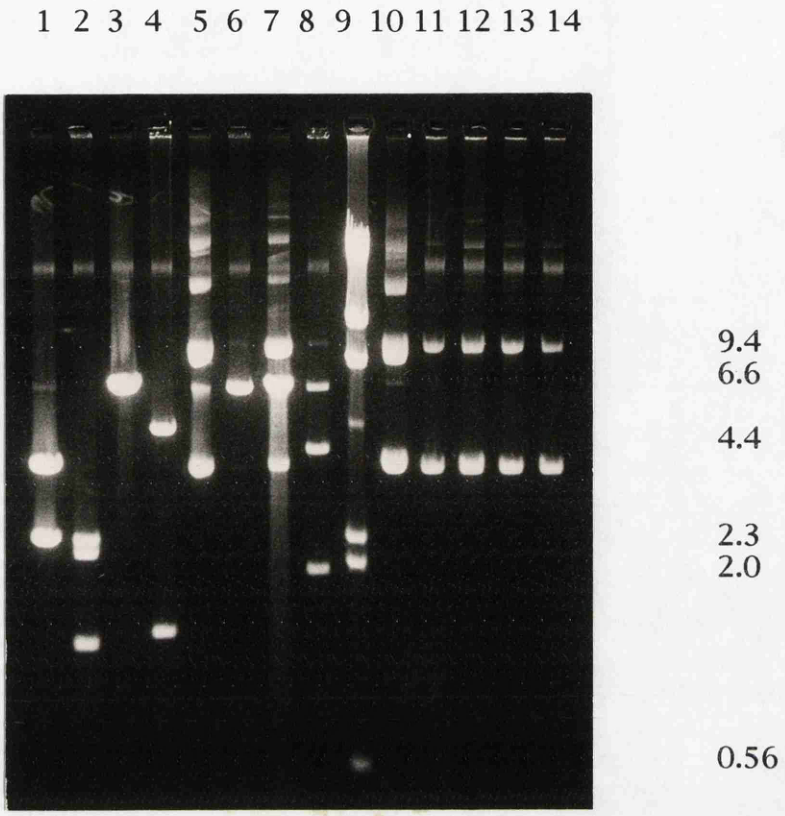


Figure 4.43 Ethidium stained 1% agarose gel showing digestion of pAG431, pAG432, pAG433 and pAG434 DNA by the diagnostic restriction enzyme.

1	pAG082	PvuII
2	pAG431	PvuII
3	pAG082	HpaI
4	pAG432	HpaI
5	pAG082	SpeI
6	pAG433	SpeI
7	pAG082	SnaBI
8	pAG434	SnaBI
9	λ	Hin*d*III
10	pAG082	uncut
11	pAG431	uncut
12	pAG432	uncut
13	pAG433	uncut
14	pAG434	uncut

four plasmids and the original pAG082 were transformed into LA547, and crude cell extracts of IPTG-induced cells were made in lysis buffer containing 1M KCl and 1mM PMSF.

A gel binding assay was carried out, using the five crude extracts described above, and labelled DNA from pNE200 containing all four motifs R1-R4. As is apparent from Figure 4.24, crude extracts of cells containing pAG431, pAG433 or pAG434 exhibit normal binding activity to this DNA fragment. The extract of cells containing pAG432, which encoded the TnsB with a mutation at the extreme carboxy end of the putative recognition helix, showed no binding activity. If the recognition helix has been correctly identified, it is surprising that none of the other three mutations affected binding to DNA. The $K_{116} \rightarrow S$ mutation of pAG423 and pAG433 in particular would have been expected to have a deleterious effect on DNA binding, since the presumptive amino terminal residue of the recognition helix was mutated, and this residue has been recognised as crucial for specific contacts with DNA in other proteins.

4.10.3 Transposition assays with the mutagenised proteins

A system for testing the competence of the mutagenised TnsB proteins was set up: DS941::Tn*7*-1(pEN300) was transformed with pMR106, which encoded *tnsC* and *tnsD*, and one of pAG431, 432, 433 and 434. pAG082, which differed from these four plasmids only in the lack of a site-directed mutation in *tnsB* was used as a positive control; it had previously been determined (Rogers, 1986) that no transposition took place in the absence of *tnsB*. Transposition was measured by the standard procedure of mating out the conjugative pEN300 into the strain MR1.

Mean results from these assays are as follows:

tnsA & *tnsB* encoding plasmid	apparent transposition frequency
pAG082	2.6×10^{-3}
pAG431	3.6×10^{-4}
pAG432	$< 2.3 \times 10^{-8}$
pAG433	1.0×10^{-3}
pAG434	3.4×10^{-2}

Transposition of Tn*7*-1 does not from the above data appear to be affected by any of the amino acid substitutions encoded by pAG431, 433 or

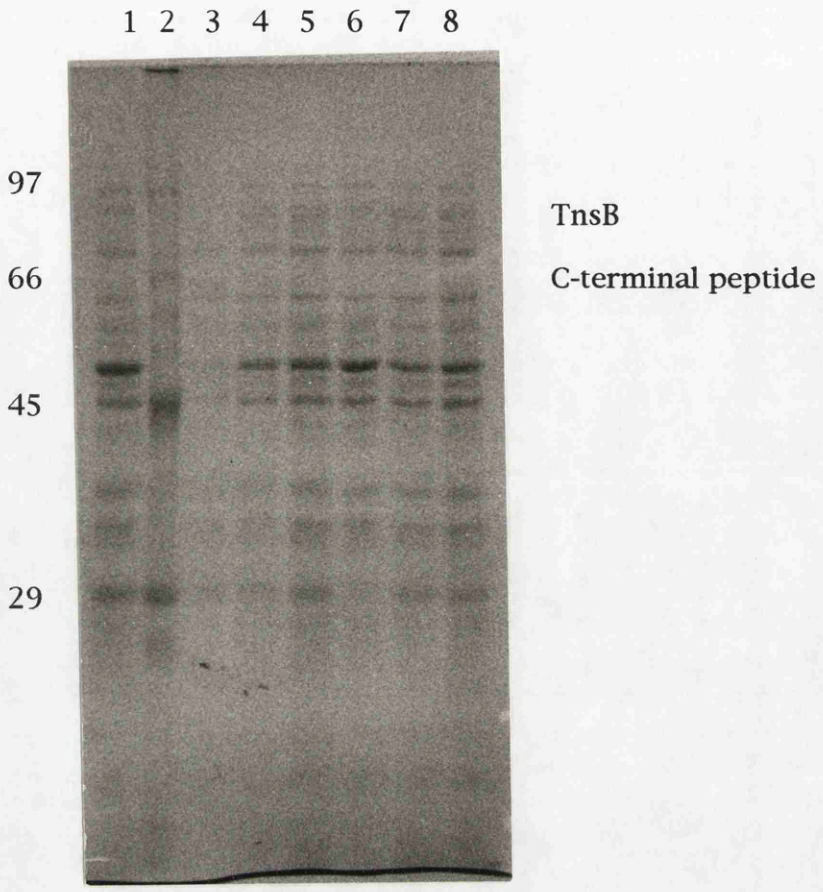


Figure 4.44 Coomassie stained 10% polyacrylamide Laemmli gel of IPTG-induced whole cells, containing the *tnsB*-encoding plasmids listed below, showing

(a) the production in DS941 of a carboxy-terminal TnsB peptide encoded by pAG077;

(b) the production in DS941 of full-size TnsB from pAG431, pAG433 and pAG434, but not from pAG432.

1 DS941(pUC18)
2 molecular weight markers
3 DS941(pAG077)
4 DS941(pAG082)
5 DS941(pAG431)
6 DS941(pAG432)
7 DS941(pAG433)
8 DS941(pAG434)

434. Transposition is totally abolished by the substitution in pAG432. These results are consistent with the observations from the gel binding experiments described above.

Further investigation of the proteins encoded by the plasmids pAG431, 432, 433 and 434 revealed that a protein running with the same mobility as TnsB could be detected in cells expressing the mutated *tnsB* gene in pAG431, 433 and 434. In the extract of cells containing pAG432, there was no visible protein of the expected mobility (Figure 4.44). This suggests that either a frameshift had been introduced somewhere in the *tnsB* gene during the mutagenesis procedure, or that the introduced mutation resulted in a protein which was very unstable. Future sequencing of the gene would resolve this question.

It was hoped that the experiments described here would have confirmed the importance of the region of TnsB tentatively identified by Flores *et al.* (1990) in DNA binding and in transposition. This has not happened. It is still possible that there is a helix-turn-helix structure as located by Flores *et al.*, but that the amino acids selected for mutagenesis were not the critical ones; saturation mutagenesis of the corresponding DNA sequence would resolve this question. Alternatively, there may be another, genuine, helix-turn-helix motif elsewhere in the amino terminal region of the protein which Flores *et al.* failed to identify. A third possibility is that DNA binding of TnsB is not through a helix-turn-helix at all, but through some other motif such as a β-sheet.

4.11 A carboxy-terminal fragment of TnsB: construction of a protein-producing clone

While it was known from gel binding assays (Section 4.4) that sequence 5' of the HincII site in *tnsB* (1485bp from the RE) encoded a protein which was associated, directly or indirectly, with binding to the right end of Tn*7*, it was not clear whether or not the remainder of the protein might also be involved in such an activity; for example, were TnsB to be the protein catalysing cleavage of the transposon ends, the N-terminal half might bind to the 22bp motifs, leaving the C-terminal region to bind at the very ends of the transposon. It was decided to construct a plasmid which would express that portion of *tnsB* downstream of the HincII site, to investigate whether the protein produced exhibited any binding

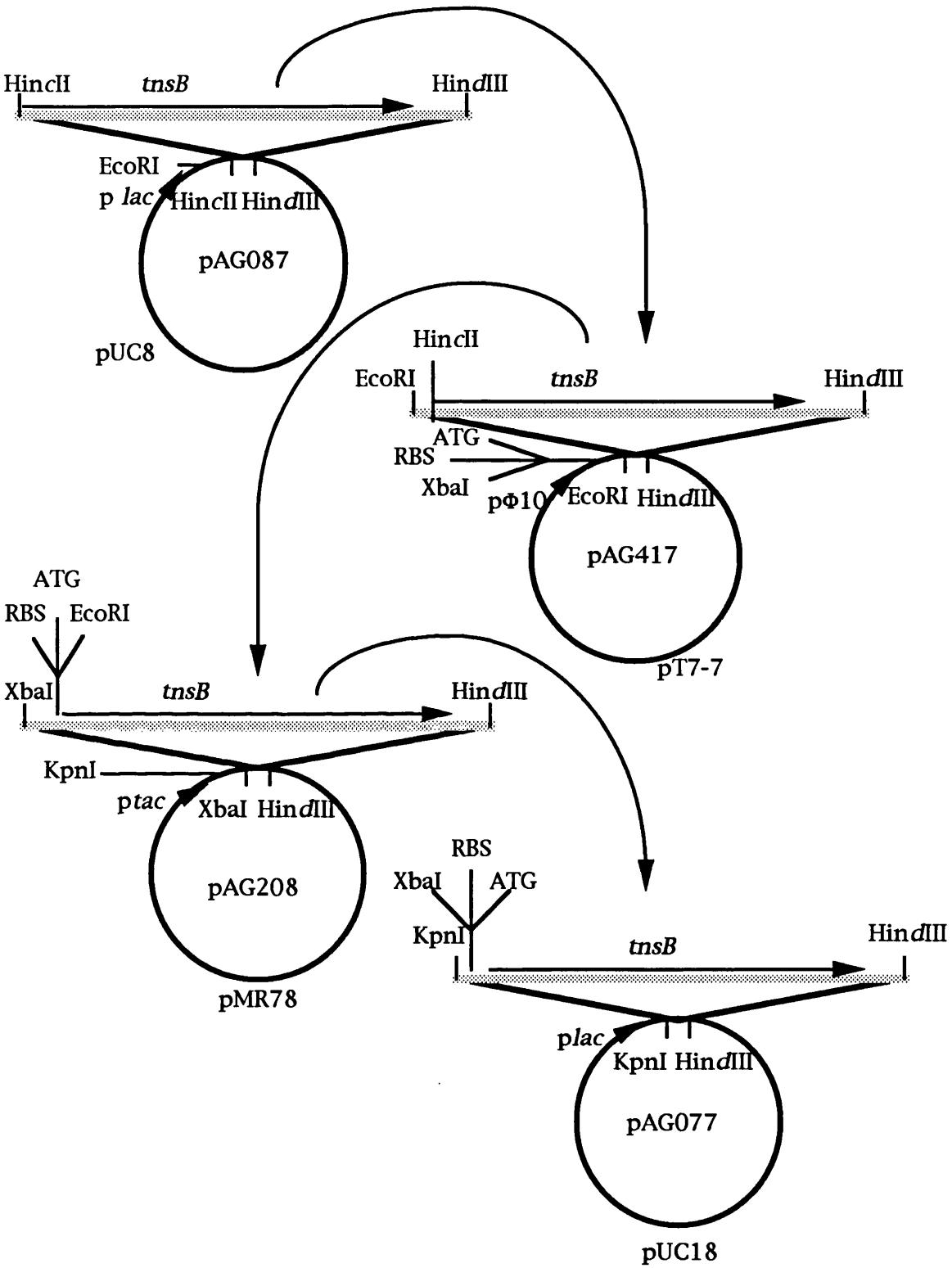


Figure 4.45 Construction of plasmids encoding the C-terminal peptide of TnsB.

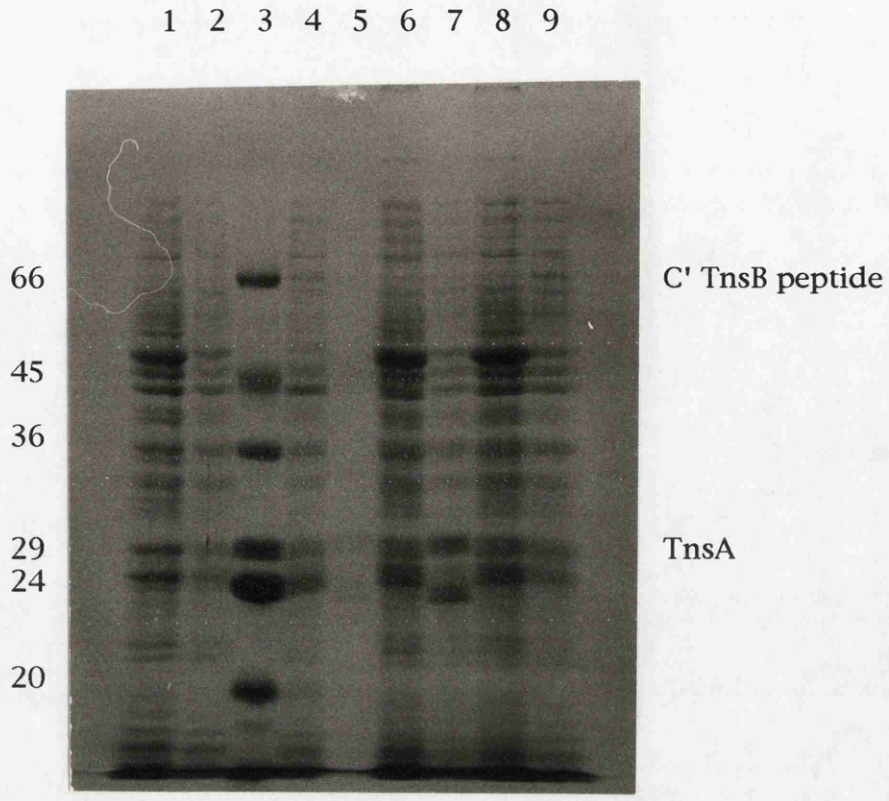


Figure 4.46 Coomassie stained 12.5% polyacrylamide Laemmli gel of whole cells; the T7 polymerase gene in pGP1-2 was heat-shock induced for three hours and rifampicin added where shown. Cells also contained one of the pT7-7 derived plasmids as indicated.

1	pT7-7	- induction
2	pT7-7	+ heat induction, with rifampicin added
3	molecular weight markers	
4	pAG411	- induction
5	pAG411	+ induction, rifampicin
6	pAG416	- induction
7	pAG416	+ induction, rifampicin
8	pAG417	- induction
9	pAG417	+ induction, rifampicin

activity, either by itself or when added to the N-terminal fragment. The strategy chosen was as illustrated in Figure 4.45.

The desired sequence was cloned into the polylinker of pUC8, to give pAG087; this plasmid lacks a suitable in-frame start codon and preceding ribosome binding site but, by cloning into pT7-7 as depicted in Figure 4.45, to give pAG417, both the ribosome binding site and in-frame start codon are obtained, and truncated TnsB protein can be seen on a Laemmli gel, as shown in Figure 4.46, lane 9. This protein-encoding sequence was then transferred to pMR78, to give pAG208 which could be used for expression of the protein for binding gels; and to pUC18, to give pAG077 which was suitable for use in galactokinase assays of repression of the *tns* promoter, similar to those described in Section 4.7. A protein specific to these plasmids and presumed to be the carboxy portion of TnsB could be seen migrating with the expected molecular weight of 66kD on Laemmli gels, as shown in Figure 4.44.

4.12 Activity of the carboxy-terminal peptide of TnsB

The possible activity of the constructed carboxy terminal peptide was studied by the three methods used for the investigation of the properties of the N-terminal peptides: gel binding assays of *in vitro* binding to the end motifs of Tn*7*, galactokinase assays of repression of promoter activity, and *in vivo* transposition assays.

4.12.1 Gel binding assays

Crude cell extracts were made of DS941(pAG208), using the method previously described since the C-terminal portion of the enzyme has a predicted pI of 9.06, similar to that of the complete enzyme, and these extracts were used in gel-binding assays, reproducing the conditions in studies of the binding of the N-terminal end of the protein (Section 4.6).

It was observed on binding gels, such as that with the complete right end sequence encoded by pNE200 shown in Figure 4.47, that there was sometimes a retarded band which could not be accounted for by host factor binding to the DNA fragment, as would be exhibited with addition of crude extract from cells containing the vector plasmid pMR78 in place of pAG208. It is possible therefore that the C-terminal portion of TnsB can bind DNA, either with low affinity specifically to the right end or, with high affinity, nonspecifically. Since only one band additional to those seen with pMR78-

1 2 3 4 5 6 7 8 9 10 11 12 13 14

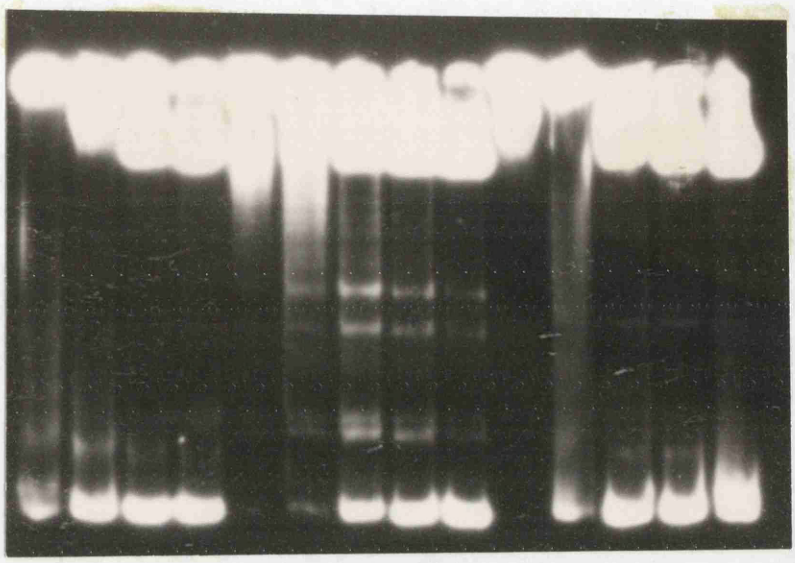


Figure 4.47 Gel retardation assay showing the retardation of labelled pNE200 DNA (encoding the complete Tn*7* RE) specific to extracts of pAG208-containing cells. The extracts of DS941 containing one of the plasmids listed were diluted for the reactions as shown below.

1	pMR78	undiluted extract
2	pMR78	2^{-1} dilution of extract
3	pMR78	2^{-2} dilution
4	pMR78	2^{-3} dilution
5	pMR207	2^{-2} dilution
6	pMR207	2^{-3} dilution
7	pMR207	2^{-4} dilution
8	pMR207	2^{-5} dilution
9	pMR207	2^{-6} dilution
10	pAG208	undiluted extract
11	pAG208	2^{-1} dilution
12	pAG208	2^{-2} dilution
13	pAG208	2^{-3} dilution
14	pAG208	2^{-4} dilution

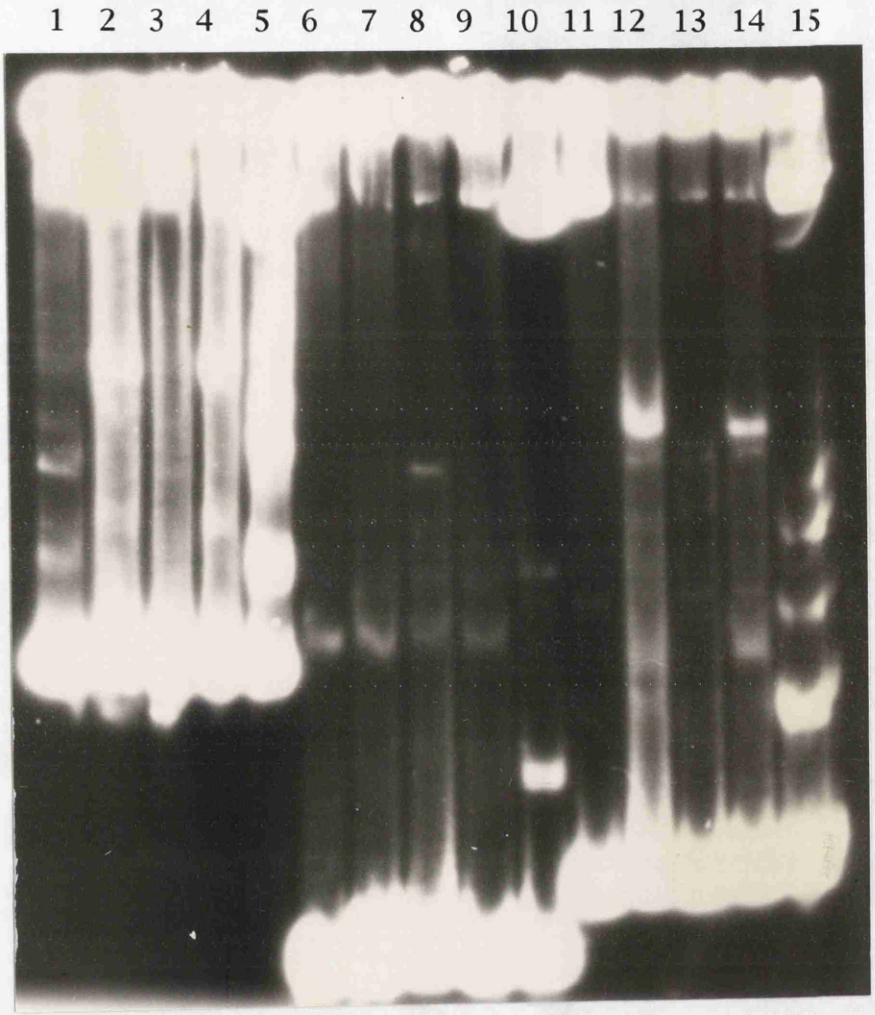


Figure 4.48 Gel retardation assay showing the effect of combining protein extracts containing the amino-terminal (pAG205) and carboxy-terminal (pAG208) TnsB peptides. All extracts were of DS941 containing one plasmid, as listed below; three different labelled DNA fragments were used: pNE200 (complete RE), pAG095 (R4 motif) and pAG096 (R3 and R4 motifs).

1	R1-R4 motifs	with	pMR78	undiluted extract
2	R1-R4 motifs	with	pAG205	undiluted extract
3	R1-R4 motifs	with	pAG208	undiluted extract
4	R1-R4 motifs	with	pAG205 & pAG208	2^{-1} dilution of each
5	R1-R4 motifs	with	pMR207	2^{-3} dilution of extract
6	R4 motif	with	pMR78	undiluted extract
7	R4 motif	with	pAG205	undiluted extract
8	R4 motif	with	pAG208	undiluted extract
9	R4 motif	with	pAG205 & pAG208	2^{-1} dilution of each
10	R4 motif	with	pMR207	2^{-3} dilution
11	R3 and R4 motifs	with	pMR78	undiluted extract
12	R3 and R4 motifs	with	pAG205	undiluted extract
13	R3 and R4 motifs	with	pAG208	undiluted extract
14	R3 and R4 motifs	with	pAG205 & pAG208	2^{-1} dilution of each
15	R3 and R4 motifs	with	pMR207	2^{-3} dilution

containing extract is observed, it is likely, though not certain, that any specific binding is to one site only on the pNE200 DNA fragment: perhaps to one of the 22bp motifs or to the terminal 8bp repeat; no binding to smaller DNA sequences has been observed.

The possibility of interaction between the N-terminal and C-terminal TnsB peptides was also considered. It was hoped that, even though they were on separate molecules, the two complementary (structurally if not functionally) parts of TnsB might associate to exhibit binding typical of the whole protein. To this end a binding assay was carried out in which extracts of cells containing pAG205 and pAG208 were mixed. The result, as shown in Figure 4.48, suggests that no such interaction occurs. It may still be possible to subdivide the protein and retain activity if a different cleavage site is chosen; alternatively, some overlap of the amino acid sequence between the two peptides may be necessary.

4.12.2 Repression of promoter activity

If the observed retardation specific to the carboxy-terminal moiety of TnsB were real, then it might be possible to detect some inhibition of transcription. The plasmid pAG077 was used for investigation of the ability of the carboxy-terminal portion of TnsB to repress transcription from the *tns* promoter, using galactokinase assays as described in Section 4.7. The results are summarised on Figure 4.37 and Table 4.1. It can be seen from these that there is no repression of transcription from the *tns* promoter in the presence of the carboxy-terminal peptide of TnsB encoded by pAG077.

4.12.3 *in vivo* transposition assays

The DNA- binding domain of TnsB appears, from the work described in Section 4.4 to reside in the N-terminal third of the protein; it is presumed that the remainder of the protein has some other function required for transposition (possibly catalysis of strand cleavage and / or transfer) since TnsB proteins truncated at the carboxy end cannot substitute for the whole protein in transposition *in vivo*. It was however considered possible that the N-terminal and C-terminal portions of TnsB, encoded by pAG205/pAG085 and pAG208/pAG077 respectively might be able to act *in trans* in place of a complete TnsB protein. To test this possibility, *in vivo* transposition assays were set up in which pMR25 (the Tn*7* sequence of which terminates at the same Hin*c*II site as pAG205/pAG085) was used as a source of TnsA and the N-terminal part of TnsB; pAG077 was provided as a

source of the remainder of TnsB; and TnsC and TnsD were encoded by pMR106. pAG086, which encodes the complete *tnsB* gene was used instead of pAG077 as a positive control. 1mM IPTG was added to induce the *tns* genes under *lac* or *tac* promoter control. The apparent transposition frequencies measured in this assay were:

pMR25/pAG086/pMR106	6.5×10^{-5}
pMR25/pAG077/pMR106	$<7.6 \times 10^{-9}$

The low transposition frequency (compared with others obtained in this work using pMR106, as described in Section 4.9) measured in the presence of pAG086 is of the same order of magnitude as transposition frequencies obtained by Rogers (1986) when testing cells containing the plasmids pMR25 and pMR106 together; the reason for this apparent reduction in transposition is not known.

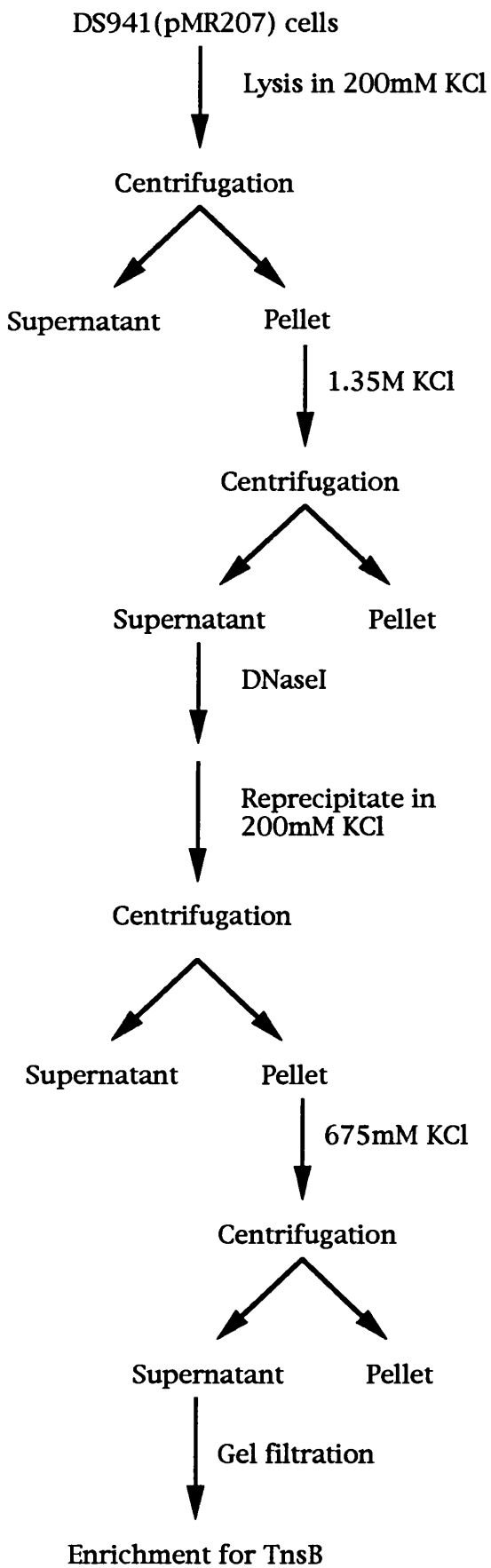
No transposition was detected when pAG077 was substituted for pAG086. Thus it seems that transposition either cannot take place at all, or takes place at an extremely low frequency, undetectable by this assay, when the TnsB protein is broken into two separate peptides at the point chosen here. It is possible that the chosen cleavage site is an integral part of a secondary or tertiary protein structure and that disruption of this structure results in large scale conformational changes in the protein, such that the two peptides can no longer fit together.

4.13 Purification strategies for TnsB

A previous attempt at TnsB purification had been carried out by Morrell (1990); this was not satisfactory because the enzyme obtained was extremely dilute, the sample contained many other proteins with TnsB at a low percentage of the total protein concentration (not greater than approximately 40% as judged by silver staining of gels), and TnsB proteolysis occurred during the procedure. Improvement of the purification technique was therefore attempted.

The method adopted by Morrell is depicted in Figure 4.49a, and an outline of the method used here is shown in Figure 4.49b. In the revised protocol, the cells were lysed as described in Chapter 2 in TrisCl pH7.0 or pH8.0 (as indicated below) and 100mM KCl in the presence of the protease inhibitors PMSF (1mM), benzamidine (1mM) and EDTA (30mM), and were

(a)


DS941(pMR207) cells
Lysis in 200mM KCl
Centrifugation
Supernatant
Pellet
1.35M KCl
Centrifugation
Supernatant
Pellet
DNaseI
Reprecipitate in 200mM KCl
Centrifugation
Supernatant
Pellet
675mM KCl
Centrifugation
Supernatant
Pellet
Gel filtration
Enrichment for TnsB

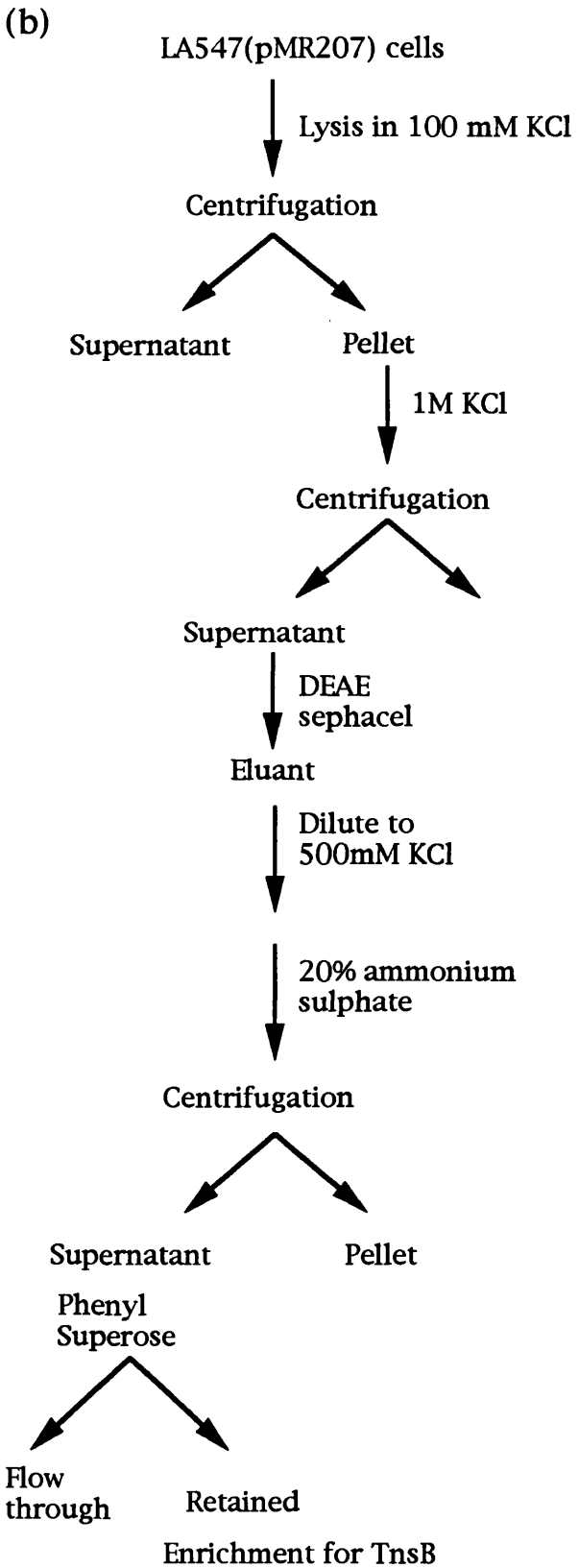


Figure 4.49 TnsB purification procedures
(a) that of Morrell (1990)
(b) as used here; see Section 4.13

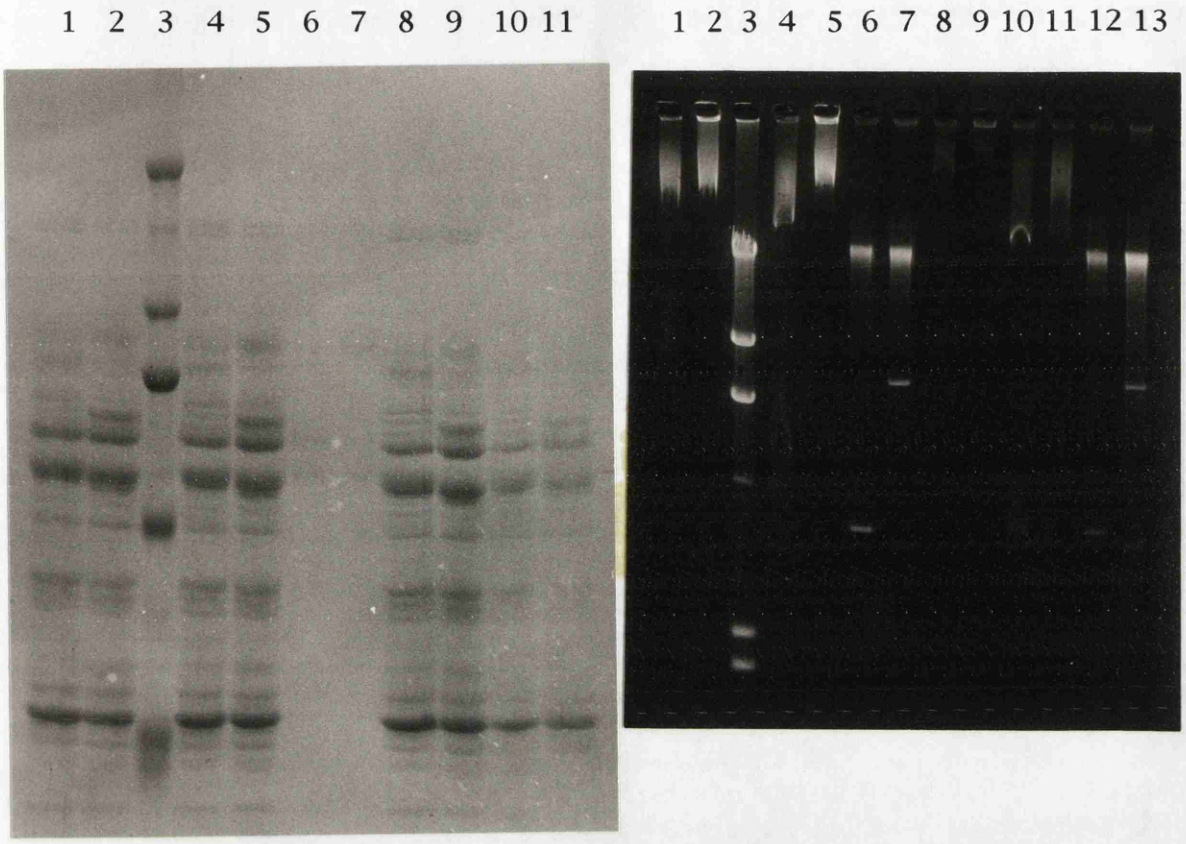


Figure 4.50 Separation of proteins and DNA in DS941(pMR78) and DS941(pMR207) by passage through DEAE sephacel:
(a) Coomassie stained 10% polyacrylamide Laemmli gel showing proteins at various steps in the purification;
(b) Ethidium stained 1% agarose gel showing the DNA content of the same samples.

1	pMR78	lysis in 1M KCl: supernatant after centrifugation	5µl
2	pMR207	lysis in 1M KCl: supernatant after centrifugation	5µl
3	size markers (a) protein molecular weight markers; (b) λ Hin*d*III		
4	pMR78	Lane 1 diluted to 0.5M KCl and centrifuged; s/n	10µl
5	pMR207	Lane 2 diluted to 0.5M KCl and centrifuged; s/n	10µl
6	pMR78	1st fraction from DEAE sephacel	15µl
7	pMR207	1st fraction from DEAE sephacel	15µl
8	pMR78	2nd fraction from DEAE sephacel	15µl
9	pMR207	2nd fraction from DEAE sephacel	15µl
10	pMR78	3rd fraction from DEAE sephacel	15µl
11	pMR207	3rd fraction from DEAE sephacel	15µl
12	pMR78	4th fraction from DEAE sephacel	15µl
13	pMR207	4th fraction from DEAE sephacel	15µl

centrifuged at 44000g at 4°C for 30 minutes. The supernatant was removed and the pellet was thoroughly washed and respun in the same buffer. The resultant pellet was resuspended overnight at 4°C in buffer of the same pH containing 1M KCl. Next day, the cell extracts were spun at 40000g at 4°C for 30 minutes and the supernatant was collected. Much of the TnsB remained in the pellet at this stage, but a high percentage was solubilised.

The supernatant at this stage contained DNA, which had to be removed before passing the sample through any HPLC column. As can be seen from Figure 4.49, the previous protocol had employed added DNAseI for removal of DNA; however, the action of DNAse is simply to nick the DNA, and other nucleases are required to degrade it further; these latter may have been separated from TnsB in the previous purification procedures. A different method was therefore adopted here: passing the sample by gravity through an equilibrated column of the weak anion exchanger DEAE sephacel, which retains the DNA while allowing proteins to pass through, as shown on Figure 4.50. Binding activity was retained on passage through DEAE sephacel (Figure 4.52).

It was hoped that ion exchange chromatography could be used to purify TnsB; TnsB has a calculated charge of +9 at pH7.0 (Figure 4.21) and so cation exchange was feasible, in principle at least. The HPLC column selected was MonoS HR5/5, which is a strong cation exchanger with an immobilised charged group $-CH_2-SO_3^-$. In theory, the positively charged TnsB molecules would bind to the negatively charged group on the column, until eluted in a sufficiently high salt concentration, this concentration being continuously increased. In practice, TnsB lysed in pH7.0 buffer could only be maintained soluble in a minimum of 500mM KCl and, at this concentration, TnsB flowed straight through the column along with virtually all the other proteins, so that there was no resultant enrichment for TnsB.

Because TnsB failed to adhere to a hydrophilic column, use of the hydrophobic Phenyl Superose column was investigated; the function of this is based on hydrophobic interactions between proteins and the column matrix and the selective elution of molecules from the matrix in a decreasing gradient of ammonium sulphate or other salt with a high salting out effect. The solubility of TnsB in ammonium sulphate was therefore investigated. It was found that in 1M KCl and at pH8.0, TnsB was soluble at 20% ammonium sulphate saturation and retained its binding activity, but that it began to precipitate at 25% (Figure 4.51). While 20% was

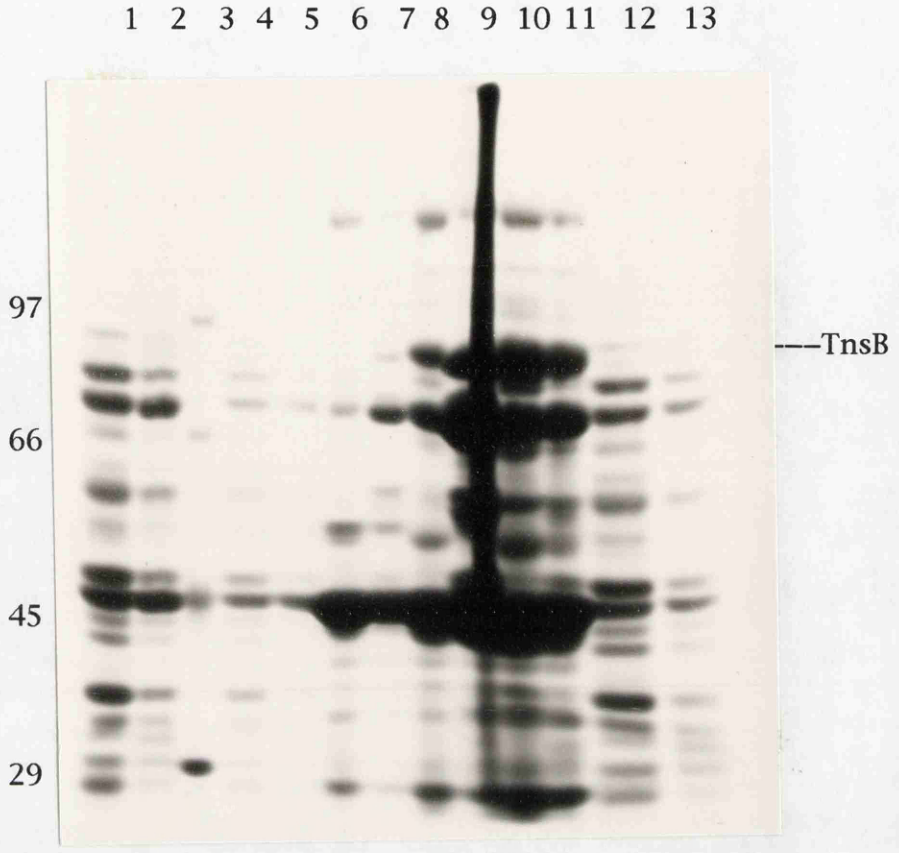


Figure 4.51 Coomassie stained 10% polyacrylamide Laemmli gel showing that TnsB precipitates in the 20-30% ammonium sulphate fraction. DS941 containing either pMR78 or pMR207 was grown, induced, centrifuged, lysed in 100mM KCl buffer and the pellets resuspended in 1M KCl buffer, then diluted to 500mM KCl and centrifuged once more. 20% ammonium sulphate was added to the supernatant on ice and dissolved, and then the extract was centrifuged and the supernatant and pellet collected. 25%, 30% and 35% ammonium sulphate fractions were obtained using the supernatant of the previous step.

1	pMR78	supernatant in 500mM KCl
2	pMR207	supernatant in 500mM KCl
3	molecular weight markers	
4	pMR78	20% ammonium sulphate pellet
5	pMR207	20% ammonium sulphate pellet
6	pMR78	25% ammonium sulphate pellet
7	pMR207	25% ammonium sulphate pellet
8	pMR78	30% ammonium sulphate pellet
9	pMR207	30% ammonium sulphate pellet
10	pMR78	35% ammonium sulphate pellet
11	pMR207	35% ammonium sulphate pellet
12	pMR78	35% ammonium sulphate supernatant
13	pMR207	35% ammonium sulphate supernatant

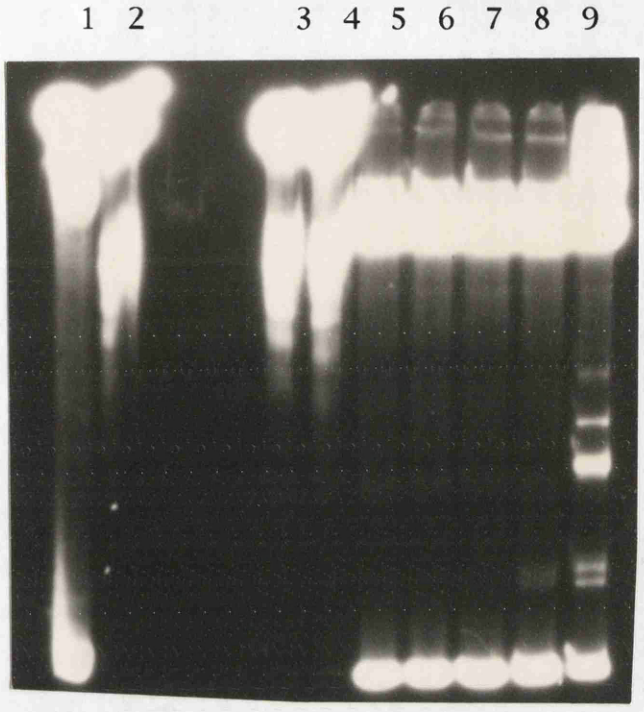


Figure 4.52 Gel retardation assay showing binding to the Tn*7* RE of pNE200 of proteins eluting from the phenyl superose column. DS941(pMR207) was grown, lysed and passed through DEAE sephacel as previously, and the 20% ammonium sulphate supernatant fraction was loaded on to the column; no proteins were detected eluting during operation of the ammonium sulphate gradient; the samples used here are from the initial flow through and from fractions collected after all ammonium sulphate had been removed.

1	DS941(pMR78)	1M KCl supernatant
2	DS941(pMR207)	1M KCl supernatant
3	DS941(pMR207)	DEAE sephacel
4	DS941(pMR207)	20% ammonium sulphate supernatant
5	DS941(pMR207)	phenyl superose flow through
6	DS941(pMR207)	1st fraction after end of gradient
7	DS941(pMR207)	2nd fraction after end of gradient
8	DS941(pMR207)	3rd fraction after end of gradient
9	DS941(pMR207)	4th fraction after end of gradient

lower than would be preferred for the starting concentration of ammonium sulphate (the manufacturers of the column recommend 40%), it was decided to proceed with this.

The first use of the Phenyl Superose column was of a 250µl sample with:

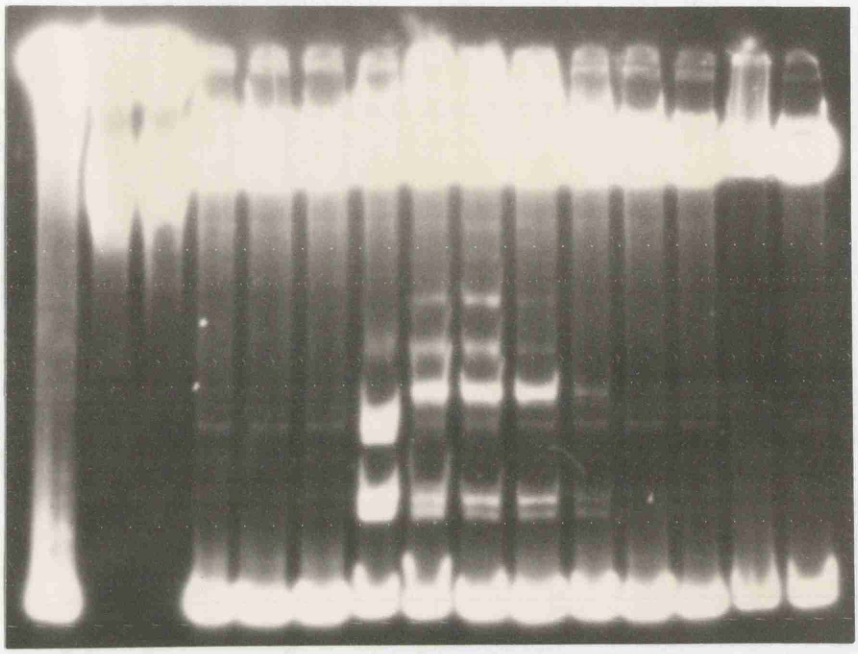
initial buffer: 50mM TrisCl pH8.0, 1mM EDTA, 1mM DTT, 1M KCl, 20%$(NH_4)_2SO_4$

final buffer: 50mM TrisCl pH8.0, 1mM EDTA, 1mM DTT, 1M KCl.

The ammonium sulphate gradient was in operation over 20 minutes, during which the flow rate was 0.5ml min^{-1}, the OD_{260} and OD_{280} were monitored, and samples were collected of the initial flow through the column, of proteins eluting during operation of the ammonium sulphate gradient (as judged by increase in absorption), and of proteins which adhered to the column throughout the gradient and were only eluted after all ammonium sulphate had been removed from the system; all samples were stored on ice. It was found that the only binding activity to the Tn*7* RE detected was associated with the final fractions after termination of the ammonium sulphate gradient, that TnsB concentration was very low, and that full size TnsB and presumed degradation products of it eluted together.

This situation was unsatisfactory because a large number of other proteins was eluting with TnsB (see Figure 4.52), and so different conditions, under which TnsB might elute earlier, were investigated. Hydrophobic interactions might be reduced by reduction of the temperature under which chromatography was carried out, but unfortunately the equipment available had no cooling system and the second buffer, even if initially chilled, could not be maintained at a low temperature. Alternatively, hydrophobicity might be reduced by reduction of the pH, thus making TnsB more highly charged, or by reduction of the KCl concentration. TnsB proved to be still soluble in 20% ammonium sulphate in the presence of 500mM KCl and so the KCl concentration of the buffer containing TnsB was reduced to 500mM for subsequent runs of the Phenyl Superose column; also, the cells were lysed in buffer of pH7.0, and this pH was maintained throughout subsequent steps of the purification. As before, the fractions eluting from the column were monitored on both silver stained Laemmli gels and on binding gels. While a much higher proportion of the total protein now flowed straight through the column, all

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



←

Figure 4.53 Gel retardation assay showing binding activity in the fractions eluting from the phenyl superose column in pH7.0 buffer, following lysis at pH7.0. Samples were diluted to 50% in glycerol, and were then used, without further dilution, in binding reactions as described previously. The labelled DNA is pNE200 cut with EcoRI and Hin*d*III, giving the complete RE. All protein samples were derived from DS941 (pMR207), except that in Lane 1 which was from DS941(pMR78). Retardation believed to be due to full-size TnsB is indicated with an arrow.

1	pMR78	DEAE sephacel supernatant
2	pMR207	DEAE sephacel supernatant
3	pMR207	20% ammonium sulphate supernatant
4	pMR207	phenyl superose column flow through
5	pMR207	1st fraction after end of gradient
6	pMR207	2nd fraction after end of gradient
7	pMR207	3rd fraction after end of gradient
8	pMR207	4th fraction after end of gradient
9	pMR207	5th fraction after end of gradient
10	pMR207	6th fraction after end of gradient
11	pMR207	7th fraction after end of gradient
12	pMR207	8th fraction after end of gradient
13	pMR207	9th fraction after end of gradient
14	pMR207	10th fraction after end of gradient
15	pMR207	11th fraction after end of gradient

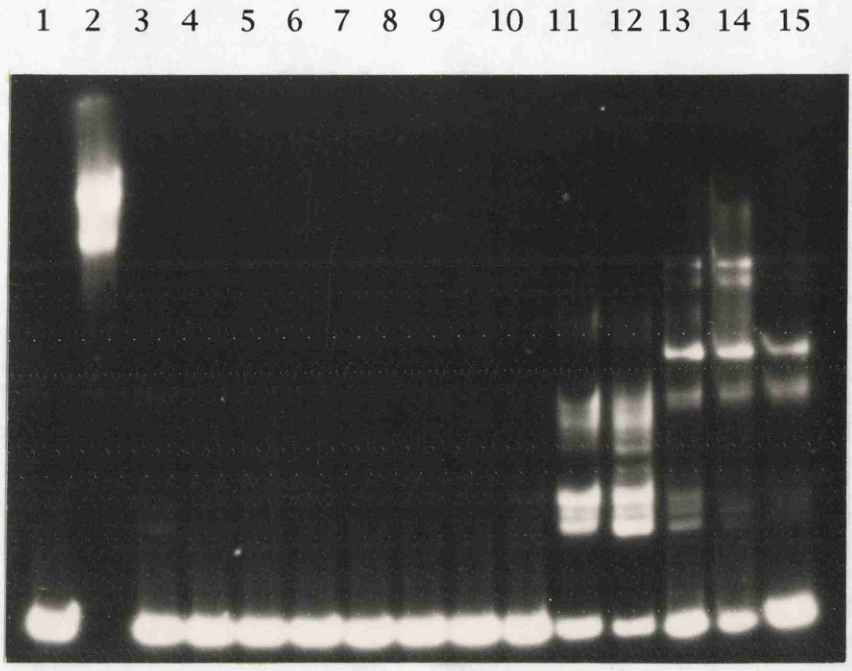


Figure 4.54 Gel retardation assay showing binding of fractions eluting from phenyl superose. These protein samples were derived from LA547(pMR207); samples, except in Lane 2, were prepared as described on Figure 4.49b. Reactions were carried out as in Figure 4.53.

1 no protein
2 crude cell extract made in 1M KCl
3 phenyl superose column main flow through
4 phenyl superose column minor flow through peak
5 1st fraction after end of gradient
6 2nd fraction after end of gradient
7 3rd fraction after end of gradient
8 4th fraction after end of gradient
9 5th fraction after end of gradient
10 6th fraction after end of gradient
11 7th fraction after end of gradient
12 8th fraction after end of gradient
13 9th fraction after end of gradient
14 10th fraction after end of gradient
15 11th fraction after end of gradient

binding activity was still associated with the fractions obtained after completion of the gradient, and TnsB was still eluting with other proteins with binding activity to the Tn*7* RE (Figure 4.53). The fraction collector was connected to a spectrophotometer with graph plotter; the proportion of the total protein eluting in the fraction with TnsB was estimated from graphical plots of absorbance against time to be between 0.04 and 0.12, so purification on the column was by a factor of ten to twenty. When the protease deficient strain LA547 became available, this was used for production of TnsB for Phenyl Superose chromatography; as shown in Figure 4.54, the binding pattern was still indicative of the presence of breakdown products of TnsB, and a comparison of the retardation patterns on binding gels (Figure 4.54 and Figure 4.25) revealed that considerable degradation of TnsB was occurring during the purification procedure, possibly during use of the Phenyl Superose column, where the TnsB remained at room temperature on the column for approximately 60 minutes. Clearly the Phenyl Superose matrix was far from ideal for purification of TnsB at room temperature; the problems encountered by Morrell of protein dilution and degradation were not overcome.

Refinement of this purification procedure was therefore halted. Any future purification would probably be most successful were it carried out completely at a maximum of 4°C, as in the purification strategy adopted by Arciszewska *et al* (1991), and using a system where such a large increase in volume were not necessary for elution from columns.

4.14 Discussion

The studies of TnsB function described in this chapter were based on the observations of M. Rogers (unpublished), Ekaterinaki (1987) and Arciszewska and Craig (1991) that there was TnsB-specific binding to the ends of Tn*7* which was detectable in gel binding assays, together with the fact that the binding region in the right end overlapped with the *tns* promoter (Gay *et al.*, 1986). Proteolytic products of TnsB were seen under various conditions; and it appeared that much of the detectable proteolysis occurred *in vivo* rather than *in vitro*, raising the question of whether this proteolysis was a natural part of the regulatory system of Tn*7* transposition, and whether perhaps these peptides could play a rôle analogous to that seen with truncated transposase proteins in several other systems, as described in Chapter 1.

Three different approaches were taken to addressing this question, all using artificially constructed peptides of TnsB which were intended to simulate the proteolysis occurring *in vivo*: gel binding assays using protein extracts containing TnsB peptides encoded by truncated *tnsB* sequences; studies of activity of the *tns* promoter in the presence of a high concentration of TnsB peptides; and *in vivo* transposition assays, again in the presence of TnsB peptides. The experiments were hampered until their later stages by the lack of a suitable protease-deficient strain in which to express the different truncated *tnsB* genes; thus it is believed that the peptides actually present were in some cases smaller than those encoded; however it is considered that the results presented here are still of value.

The combined results of these experiments indicate that some amino-terminal peptides of TnsB have the capacity to bind the 22bp motifs of the right and left ends of Tn*7*, although with reduced stability, and to have a slight repressive effect on *tns* transcription; their effect on *in vivo* transposition is, however negligible according to the assay used here. It therefore seems unlikely that these peptides play a significant part in the control of transposition in Tn*7*.

It had been suggested by Flores *et al.* (1990) that there was strong sequence homology to the helix-turn-helix motif of DNA-binding proteins in the amino-terminal region of TnsB. All the peptides studied here which exhibited clear binding to the 22bp motifs contained that part of the amino acid sequence which included the putative helix-turn-helix motif. Site-directed mutagenesis of the proposed recognition helix was carried out, with the intention of confirming the importance of this region for DNA binding. However, with the exception of one mutation which apparently resulted in total loss of the protein, there was no substantial difference in either *in vitro* binding activity or *in vivo* transposition rates when wildtype TnsB was substituted by the mutant protein. Thus it seems that either the recognition helix has not been correctly identified, or that the amino acids selected for mutation were not critical for function.

The carboxy terminal region of TnsB was also examined for evidence of DNA binding, but the results obtained were equivocal; it is possible that this carboxy-terminal domain makes weak contacts with the terminal nucleotides of Tn*7*, where strand cleavage would be expected to take place in a transposition reaction, but such conclusions cannot be drawn from the

available results. No protection of the terminal nucleotides of Tn*7* was detected by Arciszewska and Craig (1991) or Tang *et al.* (1991) during their hydroxyl radical footprinting experiments, but it is possible that this occurs, but only under certain conditions.

Section 1.6 discussed a region of possible homology with retroviral integrases and other prokaryotic transposases found in the centre of TnsB, between the XbaI and NcoI sites demarcating pAG205/pAG085 and pAG209/pAG086 respectively. Some of the conserved amino acids in this region have been shown (Engelman and Craigie, 1992) to be essential for catalytic function in the integrase protein of the retrovirus HIV-1. It is known (Section 4.3) that a TnsB peptide truncated at the NcoI site cannot complement a lack of wildtype TnsB in transposition *in vivo*, thus more than the 17 residues downstream of this conserved region to the NcoI site are also essential for TnsB function. As can be seen from Figure 1.8, both TnsB and the retroviral integrases have substantial amino acid sequence downstream of this conserved region; this may have a structural rôle, maintaining other parts of the protein in the correct conformation for catalytic activity.

An essential experiment for the future is to mutagenise the amino acid residues which are conserved across the transposase and retroviral integrase sequences to determine whether this similarity is purely coincidental, or whether the amino acids are also critical for transposition function in Tn*7* and other prokaryotic transposable elements. If the strand cleavage activity did indeed reside in TnsB, experiments could be carried out to determine whether this cleavage was *in cis*, i.e. cleavage of the Tn*7* end to which the protein was bound, or *in trans*: cleavage of the opposite end of the transposon. Cleavage *in trans* is proposed in the site-specific recombination reaction catalysed by the *Saccharomyces* Flp protein (Chen *et. al.*, 1992). Cleavage *in cis* is perhaps more likely with the Tn*7* ends, since single excised Tn*7* ends have been observed *in vitro* (Bainton *et al.*, 1991).

It can be seen from Figure 1.8 that TnsB is substantially larger than both retroviral integrase proteins and many other prokaryotic transposases, although that of Tn*3* is even larger at approximately 110,000 kDa; it remains to be seen whether this unusually large size is indicative of additional functions, yet to be determined, or whether the extra amino acid sequence plays a structural rôle or is in part redundant. Although much is

now known about TnsB, there is clearly scope for considerable future study.

CHAPTER 5

INVESTIGATION OF POSSIBLE SYNAPSE FORMATION MEDIATED BY TnsB

5.1 Introduction

Before any strand exchange can take place in a transposition or site-specific recombination reaction, the appropriate DNA sites must be brought together in the correct configuration with the aid of proteins. In the cases of site-specific recombination and replicative transposition, the entire molecule(s) encoding the reacting sites for strand exchange (transposon donor and target DNA in the case of transposition) must be incorporated in a strand exchange complex, although cleavage of 3' ends and strand transfer are not concerted in the case of transposition. Where non-replicative transposable elements such as Tn*7* and Tn*10* are concerned, however, a cut-and-paste reaction appears to take place and so it is theoretically possible that complete excision of the transposon from the donor molecule normally occurs before synapsis of the transposon ends with the target DNA molecule; excised linear transposons have been observed *in vitro* during transposition of both Tn*7* (Bainton *et al.*, 1991) and Tn*10* (Benjamin and Kleckner, 1992), and have been shown in the case of Tn*7* to proceed to target insertion, but it is not clear whether these are normal transposition intermediates; Tn*7* excised linear transposons have been observed at a low level in the absence of target DNA and TnsD (Bainton, 1992). Haniford *et al.* (1991) reported that excised Tn*10* was held together at the ends in a complex with the transposase protein; protein-DNA synapses have not been reported in any of the studies of Tn*7* transposition.

Synapse formation *in vitro* has however been studied in a variety of transposition and site-specific recombination systems, two of the most successful of which have proved to be those of the Mu transpososome, described in Section 5.2, and the Tn*3* *res*/resolvase complex, outlined in Section 5.3; in both cases, a low number of proteins is required and these have been available purified for use in reactions.

Whereas polyacrylamide gels are more commonly used for the study of protein-DNA interactions, agarose gels have also been used for several years where large DNA molecules are involved (for example, Berman *et al.*, 1987). Retardation of protein-DNA complexes occurs in agarose gels due to the additional molecular weight of the bound proteins; however, unlike the

situation with polyacrylamide gels, conformational changes in linear DNA do not by themselves result in lowered mobility on agarose gels (Lane *et al.*, 1992).

The experiments described in Sections 5.6 and 5.7 used both isolated Tn*7* ends on polyacrylamide gels, and a plasmid containing the mini Tn*7*, called Tn*7*-1, on agarose gels.

5.2 The Mu transpososome

Intermediates in the replicative transposition reaction of bacteriophage Mu have been studied *in vitro* by the groups of Chaconas and Mizuuchi; the intermediate protein-DNA structures have been given the name 'transpososomes'. As discussed in Section 1.3.5, three distinct progressively more stable transposition intermediates (Haniford and Chaconas, 1992) have been identified through the use of *in vitro* techniques; these are the Stable Synaptic Complex (SSC, Type 0 transpososome), Cleaved Donor Complex (CDC, Type 1 transpososome) and Strand Transfer Complex (STC, Type 2 transpososome), where in each case the name given by Mizuuchi *et al.* is given first. The existence of another, initial complex has been hypothesised (Mizuuchi *et al.*, 1992; Surette and Chaconas, 1992); the presence of the internal activating sequence in this complex is believed to be necessary in order to bring about the correct synapsis of the Mu ends. The Cleaved Donor Complex and Strand Transfer Complex are illustrated in Figure 5.1. The intermediates in Mu transposition have been obtained by altering the *in vitro* reaction conditions so that transposition cannot go to completion; for example, the use of Ca^{2+} ions instead of Mg^{2+} ions results in accumulation of the SSC intermediate, and in the absence of target DNA the CDC is obtained.

DNA-protein complexes involving both Mu ends have been visualised both by electron microscopy and by agarose and composite agarose/polyacrylamide gel electrophoresis, sometimes after crosslinking of the protein molecules (Lavoie *et al.*, 1991; Baker and Mizuuchi, 1992). The different transposition intermediate complexes have been identified on gels by their mobility both before and after treatment with SDS, which disrupts non-covalent protein binding, Topoisomerase I, which relaxes DNA supercoils, and restriction endonucleases which were selected to cleave specifically in one of the Mu, donor or target DNA domains (Surette *et al.*, 1987).

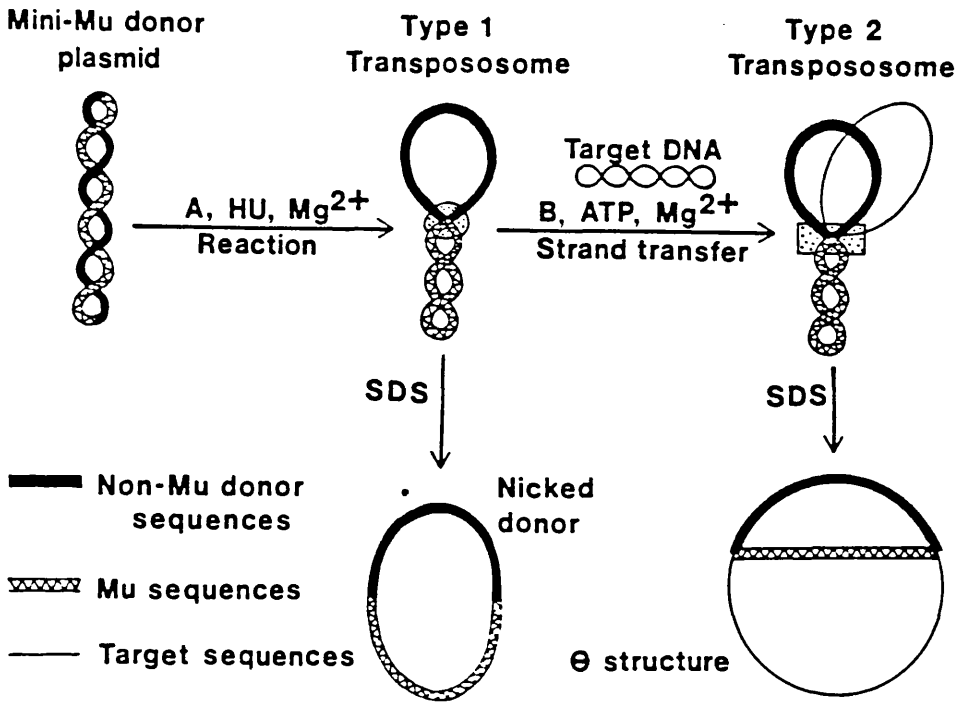


Figure 5.1 Intermediates in the bacteriophage Mu *in vitro* transposition reaction, and their characteristic forms on treatment with SDS. The SSC (Type 0 transpososome) and the hypothesised IAS-containing precursor complex are not illustrated. This figure is from Surette and Chaconas (1989)

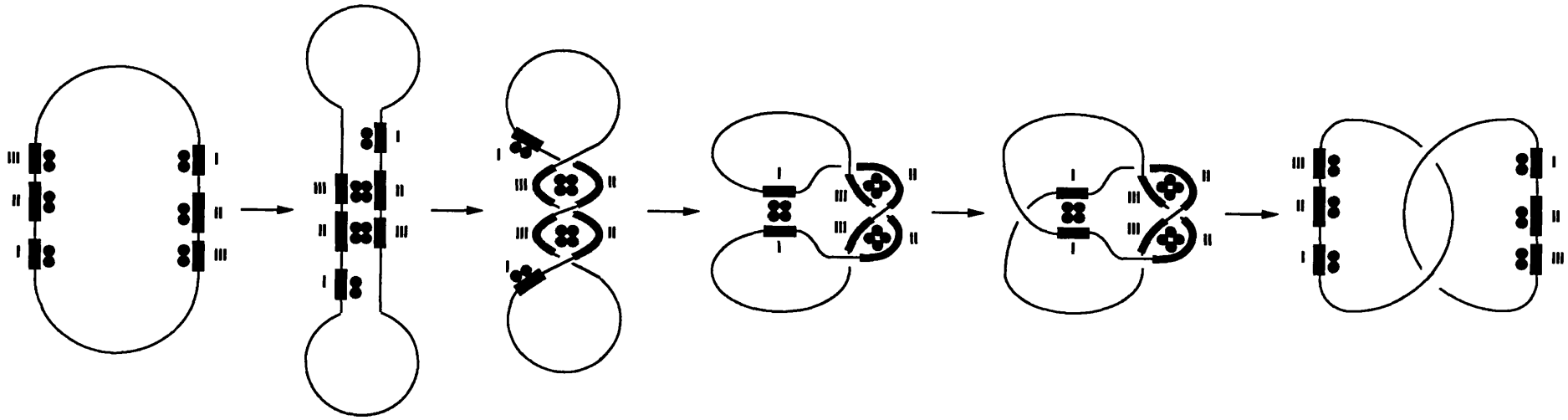


Figure 5.2 Synapsis and resolution of a cointegrate molecule containing two *res* sites, according to the two step synapsis model of M. Boocock. The *res* site is the only part of Tn*3* required for this reaction. The subsites I, II and III of each *res* site are marked; it is believed that each is bound by a resolvase dimer (D. Blake, unpublished). This Figure is reproduced by kind permission of its creator, W. M. Stark.

5.3 The Tn*3* *res*/resolvase complex

The Tn*3* resolvase synaptic intermediate has been isolated in high percentage yield and characterised on agarose gels and by electron microscopy by Benjamin and Cozzarelli (1988), and work is continuing in this field (M. Boocock, M. Stark, M. Watson, S. Akhter, unpublished results). Steps in the formation of a possible synaptic structure, proposed by M. Boocock, and its resolution are shown in Figure 5.2. The Tn*3* resolution reaction with a supercoiled substrate is exclusively intramolecular, in contrast to the transposition reactions of Mu and Tn*7*; thus only one DNA molecular species, containing two *res* sites, is required for the reaction, in addition to the resolvase protein and a simple buffer (in which Mg^{2+} was omitted by Benjamin and Cozzarelli in order to minimise completed recombination reactions). When the recombination intermediates of Benjamin and Cozzarelli were electrophoresed on an agarose gel, synapsis of the two *res* sites by resolvase was detected as a slight retardation of the 2.4kb supercoiled DNA molecule containing the *res* sites. That this retardation was genuinely due to synapsis and not simply to independent binding of resolvase to the *res* sites was demonstrated by cleaving the resolvase-bound DNA molecule either with EcoRI, which had one recognition site, or Hin*d*III, which had two sites in different domains; cleavage with EcoRI yielded an α-shaped product with two free arms, as seen by electron microscopy and inferred from gel mobility; similarly cutting with Hin*d*III gave a χ-shaped product with four free arms. It was found that stability of these synaptic complexes was greatly improved by crosslinking the protein molecules in each synapse using glutaraldehyde, a bifunctional compound which reacts with amine groups, particularly those in lysine residues. Topoisomerase relaxation of resolvase-bound DNA substrate was employed in an attempt to demonstrate a linking number change in the DNA during synapsis, and hence to derive information about the synaptic structure.

5.4 The probable Tn*7* transposition mechanism

Tn*7* transposition *in vivo* is most likely to proceed by a two-stage mechanism: in the first stage, the Tn*7* ends are cleaved, apparently with a 3-base staggered cut, so that donor DNA remains attached to the 5' ends of the transposon (Bainton *et al.*, 1991); in the second stage, the target DNA is subjected to a 5bp staggered cut, and the Tn*7* DNA is inserted.

Chromosomally-encoded DNA repair mechanisms are probably responsible for ensuring that the resultant single-stranded target DNA is filled in and that the remaining donor DNA nucleotides are excised, as has been postulated in the case of the *Drosophila* P element (Engels *et al.*, 1990). Bainton *et al.* (1991) reported double strand breakage of the Tn*7* ends *in vitro* only when TnsA, B, C and D, a host cell extract and *attTn7*-containing target DNA were all supplied in the reaction, although Bainton (1992) observed a low level of double strand cleavage at one or both transposon ends in the absence of TnsD and target DNA; it is possible that *in vitro* cleavage may occur without associated target DNA under reaction conditions very dissimilar from those encountered *in vivo*. Artificially pre-cleaved Tn*7* can, however, proceed to insert in target DNA *in vitro*, provided the Tns proteins, ATP and Mg^{2+} are added (Bainton *et al.*, 1991); this implies that the two stages *can* be uncoupled spatially, temporally and energetically, although this may not normally occur. The observations also suggest that, in this particular case, TnsA, B, C and D are all required for the second stage of the reaction, although one or more of them may simply be involved in bringing together the two Tn*7* ends which had been presented in an artificially separated form. Since a host cell extract was always added to the reactions of Bainton *et al.*, it is likely that a host encoded protein such as IHF or HU may be required for correct synapse formation. (HU binds DNA non-sequence-specifically, although it has been suggested that it shows a preference for kinked DNA (Pontiggia *et al.*, 1993), and at high concentration is believed to wrap the DNA around it into a form similar to that of a eukaryotic nucleosome (Rouvière-Yaniv *et al.*, 1979). IHF has limited DNA sequence specificity and binding of it to DNA results in a bend in the DNA at this point.) FIS has been shown to bind, apparently non-specifically, to the Tn*7* RE (Morrell, 1990), and may also have a part in synapsis. No Tn*7* protein-DNA transposition intermediate complexes were observed by Bainton *et al.*; their samples were treated with 55% urea before loading on a gel, thus disrupting any non-covalent protein-DNA association which might have occurred.

The precise rôle of the individual proteins in the transposition reaction is not known, although clearly TnsB and TnsD are involved in binding to the Tn*7* ends and the *attTn7* site respectively; and TnsC shows binding activity both non-specifically to DNA and to nucleotides (Gamas and Craig, 1992). It was considered likely that, while TnsC and TnsD were

involved in target selection, TnsB, possibly in conjunction with TnsA, might bring the two ends of the transposon together in a pre-transpososome complex, with the end regions containing the 22bp motifs wrapped around each other, although with the differing number and arrangement of motifs at the two ends, the structure of such a complex is not obvious.

Two Tn*7* right ends can be substituted for a left end and a right end (Arciszewska *et al.*, 1989) in a functional transposon; thus the precise configuration of the motifs and associated bound proteins of the two ends appears to be unimportant for transposition to take place. A copy of Tn*7* deleted for R4 transposes with nearly the wildtype frequency *in vivo* and still shows orientation specificity when transposing to *attTn7* (Arciszewska *et al.*, 1989), so the asymmetry in *number* of motifs cannot be involved in orienting the transposon in the target DNA, although the difference in *spacing* between L1, L2 and L3, and R1, R2 and R3 is likely to play a part in this; it is also possible that orientation is determined by the small sequence differences in the motifs at the two ends. In the case of the transposition of bacteriophage Mu, the terminal motifs are all occupied initially by Mu A protein, but binding to only the L1, R1 and R2 sites is necessary for steps subsequent to formation of the CDC (Type 1 transpososome) (Mizuuchi *et al.*, 1991). It is not known whether all the Tn*7* terminal motifs remain occupied by TnsB throughout the transposition process or whether, as with Mu, binding to some motifs is dispensable after a certain step.

5.5 The search for Tn*7* synaptic complexes *in vitro*

The search for such TnsB-mediated complexes was made in two different ways, both employing gel retardation: polyacrylamide gels were used to look for complexes between isolated linear DNA fragments containing Tn*7* end sequences; and agarose and composite agarose-polyacrylamide gels were used to study the intramolecular interactions of the left and right ends of mini-Tn*7* Tn*7*-1 located on the plasmid pMR11, which is illustrated in Figure 5.4.

5.6 Binding of TnsB to linear fragments

This was investigated using gel binding techniques similar to those described in Chapter 4. "Sandwich" DNA-protein-DNA complexes have been identified on polyacrylamide binding gels using Lac repressor protein and a DNA fragment containing the *lac* operator sites (Krämer *et al.*, 1987); the tetrameric Lac repressor protein acts as a bridge, holding together two linear DNA molecules and, when these DNA molecules are of different sizes, the heterodimeric structure can easily be distinguished by its gel mobility. It was hoped that an analogous structure containing TnsB and the Tn*7* ends might be formed and be detectable on binding gels. The interaction between the different motif-bound TnsB molecules during the (presumed) synapsis of the Tn*7* ends is completely unknown, but it is likely that TnsB facilitates this synapsis. It was therefore decided to look for Tn*7* end-TnsB sandwich-type complexes.

Because Tn*7* transposition is known to take place *in vivo* at the same measured frequency when two right ends are substituted for the normal situation of a left end and a right end (Arciszewska *et al.*, 1989), both the Tn*7* fragments used here included the cloned Tn*7* right end sequence of pNE200, but different fragments representing the two ends were cut at the EcoRI site and either the Hin*d*III site or the NdeI site, and could thus be distinguished by the length of flanking pUC vector sequence. In the experiment described below, the smaller 239bp EcoRI-Hin*d*III fragment was ^{32}P end labelled, and the larger 450bp EcoRI-NdeI fragment was used unlabelled in an excess of approximately 100fold over the labelled fragment.

Crude cell extracts from DS941(pMR207) and DS941(pMR78) were used as the TnsB source and its negative control; these were diluted as shown in Figure 5.3. The labelled, unlabelled or both fragments were added and incubated under standard conditions with the carrier DNA and cell extract as described in the caption to Figure 5.3. The reaction mix was then run on a 6% Tris-glycine EDTA pH9.4 polyacrylamide gel at 200V and 4°. As can be seen from Figure 5.3, a novel band is clearly visible on the autoradiograph in lane 13, where the TnsB-containing extract is incubated first with the labelled DNA and then with the unlabelled fragment in addition.

It is thought for several reasons that this band may represent an intermolecular complex containing both DNA fragments, possibly bringing the two end sequences together in a defined synaptic complex. Firstly, the additional band is only seen in the presence of both the labelled and

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

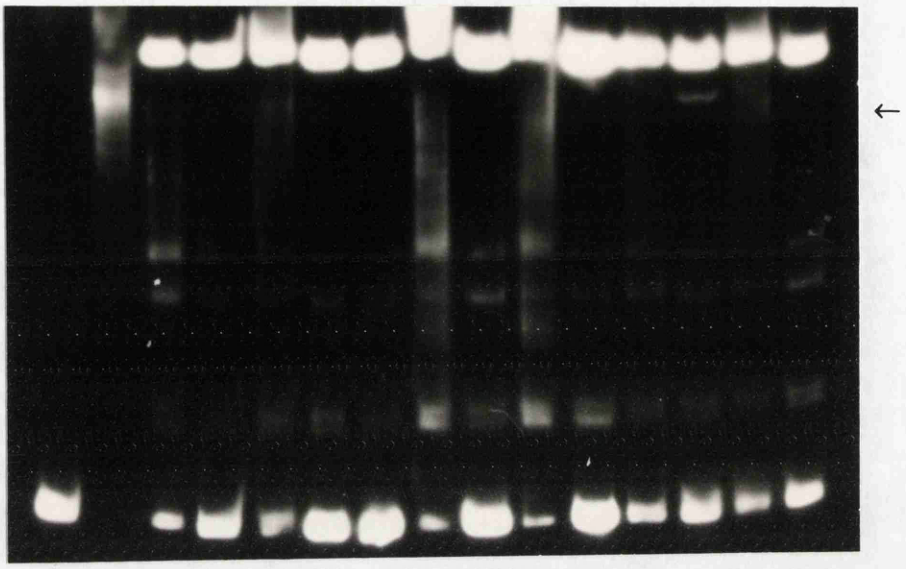


Figure 5.3 Gel retardation assay, showing possible formation of DNA-protein-DNA complexes. Two different DNA fragments both encoding the complete Tn*7* RE were used: pNE200 cut with EcoRI and Hin*d*III was ^{32}P end-labelled at 4ng per reaction; pNE200 cut with EcoRI and NdeI was used unlabelled at 8ng per reaction (but at the same molarity of Tn*7* ends as the labelled DNA). The protein extracts were of DS941(pMR78) and DS941(pMR207). Reactions were carried out as described in Section 2.17, with incubation at 37° for the times below and loaded on a 6% polyacrylamide gel.

1	pMR78	undiluted extract	)
2	pMR207	2^{-2} dilution of extract	) labelled pNE200 only;
3	pMR207	2-4 dilution	) 10 minutes incubation
4	pMR207	2^{-6} dilution	)
5	pMR207	2^{-2} dilution	) labelled and unlabelled pNE200
6	pMR207	2^{-4} dilution	) added together;
7	pMR207	2^{-6} dilution	) 10 minutes incubation.
8	pMR207	2^{-2} dilution	) unlabelled pNE200 for 5 mins; then
9	pMR207	2^{-4} dilution	) labelled & unlabelled for 10 mins.
10	pMR207	2^{-2} dilution	) unlabelled pNE200 for 15 mins;
11	pMR207	2^{-4} dilution	) labelled & unlabelled for 10 mins.
12	pMR207	2^{-2} dilution	) labelled pNE200 for 10 mins; then
13	pMR207	2^{-4} dilution	) labelled & unlabelled for 5 mins.
14	pMR207	2^{-2} dilution	) labelled pNE200 for 10 mins; then
15	pMR207	2^{-4} dilution	) labelled & unlabelled for 15 mins.

unlabelled fragments; there may also be bands indicative of homodimer complexes of the labelled fragment, but these cannot so easily be distinguished from simple monomolecular DNA-protein retarded complexes. Secondly, the labelled DNA has been retarded more than had previously been observed, as can be seen by comparison of Figures 5.3 and 4.23; retardation can in theory be due to either increased molecular weight of a complex or to bending of the DNA, or to a combination of the two, but a shift of the magnitude seen here is unlikely to be due to additional bending alone, and so an intermolecular DNA complex is plausible. Thirdly, the larger unlabelled fragment would be expected to cause more retardation in a DNA dimeric complex than would a second, smaller, labelled fragment. It can be postulated that pre-incubation with the labelled fragment is necessary for radioactively identifiable intermolecular protein-DNA complexes to form; initially homodimers of the labelled fragment are created, but the second, unlabelled fragment, being in excess, then displaces one of the labelled DNA fragments to form the observed complex; when the pre-incubation is with the unlabelled fragment instead, protein-DNA complexes form as before, but the concentration of the labelled fragment is too low to bring about appreciable displacement of unlabelled DNA. This hypothesis is in agreement with the observation that the additional band is much weaker after a longer incubation with the unlabelled DNA: the unlabelled DNA could now also be displacing the radioactive fragment of heterodimers. It is possible that, because the extracts were made in DS941 rather than the protease-deficient LA547, a truncated TnsB protein is involved in the formation of the complex; most of the other retarded bands are clearly specific to a TnsB peptide rather than the full-length TnsB protein; compare Figure 5.3 with Figure 4.23.

It is not clear why this retardation occurred more strongly with the lower concentration of TnsB, than with the higher.

5.7 Binding of TnsB to circular DNA molecules

A second line of study was to try to observe synapsis of the two Tn*7* ends in the plasmid pMR11, which contains Tn*7*-1, (shown in Figure 5.4) mediated by TnsB, possibly in conjunction with other Tns proteins. This was tested both in the absence and in the presence of a target plasmid encoding *attTn7*; chemical crosslinking of proteins was also used to stabilise any synapse which might be formed.

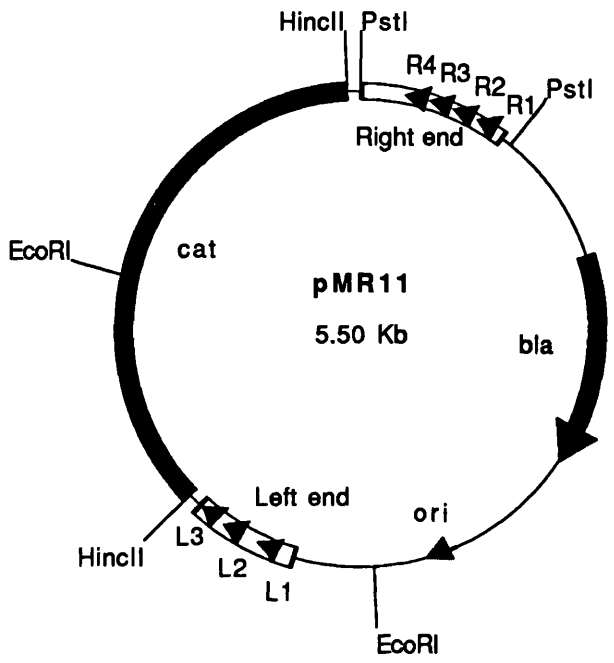


Figure 5.4 Structure of the mini-Tn*7* containing plasmid pMR11. The Tn*7* terminal sequences, including all the 22bp motifs, flank and are able to mobilise a chloramphenicol resistance gene. The Tn*7*-1 is cloned into pUC8.

Initial studies used non-radioactive supercoiled pMR11 DNA; binding reactions with crude cell extracts were carried out using the standard conditions for binding of TnsB to linear DNA fragments, as described in the previous chapter. The sheared salmon sperm DNA, used with radioactively labelled Tn*7* DNA as a carrier DNA sink for non-specific DNA binding by proteins, could not be used here as the gels were to be ethidium stained, and the blur of chromosomal DNA would have been likely to obscure any novel bands; instead, supercoiled pUC18 was generally used (the stock also contained a high proportion of nicked plasmid). Reaction mixes were loaded on a 0.8% agarose gel, run in TAE buffer at 2.3-2.5 Vcm^{-1} at 4° for 14-

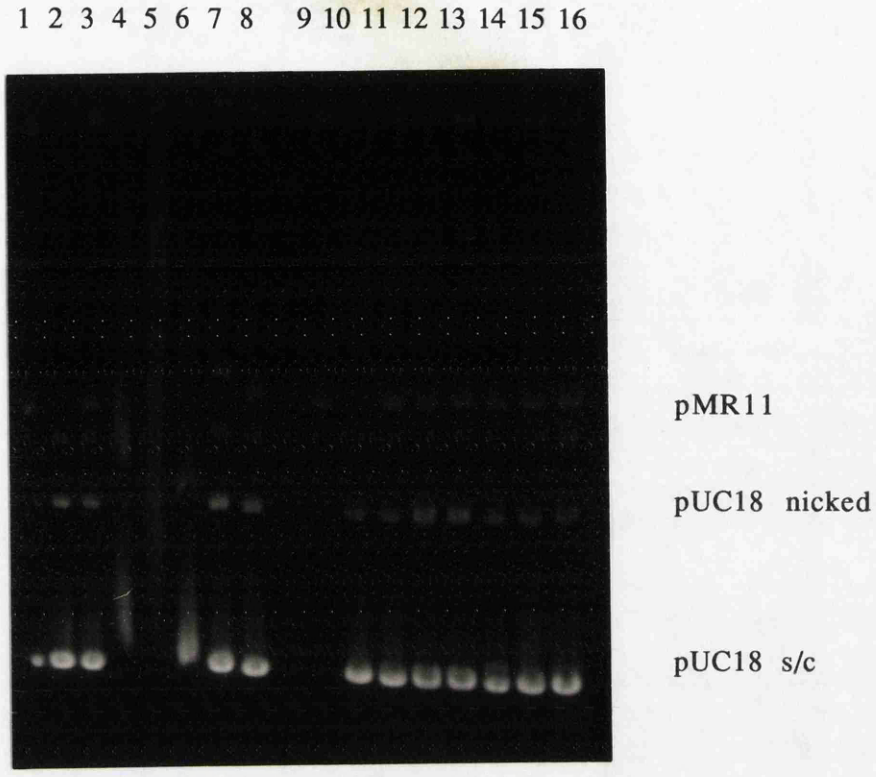


Figure 5.5 Gel retardation assay, with samples run on a 1% agarose gel. The reactions were as described in Section 2.17, using supercoiled plasmids; pMR11 contained Tn7-1, as shown in Figure 5.4, and pUC18 was the carrier DNA. Protein extracts were of DS941(pMR78) or DS941(pMR207). 0.5% SDS was added after incubation where shown below.

1	pMR11	with	no protein extract; dilution buffer only		
2	pUC18	with	no protein extract; dilution buffer only		
3	pMR11 + pUC18	with	no protein extract; dilution buffer only		
4	pMR11 + pUC18	with	pMR78	undiluted extract	
5	pMR11 + pUC18	with	pMR207	undiluted extract	
6	pMR11 + pUC18	with	pMR207	2^{-2} dilution of extract	
7	pMR11 + pUC18	with	pMR207	2^{-4} dilution	
8	pMR11 + pUC18	with	pMR207	2^{-6} dilution	
9	pMR11	with	dilution buffer only		+ SDS
10	pUC18	with	dilution buffer only		+ SDS
11	pMR11 + pUC18	with	dilution buffer only		+ SDS
12	pMR11 + pUC18	with	pMR78	undiluted extract	+ SDS
13	pMR11 + pUC18	with	pMR207	undiluted extract	+ SDS
14	pMR11 + pUC18	with	pMR207	2^{-2} dilution	+ SDS
15	pMR11 + pUC18	with	pMR207	2^{-4} dilution	+ SDS
16	pMR11 + pUC18	with	pMR207	2^{-6} dilution	+ SDS

18 hours. (TAE buffer was found to give better resolution under these conditions than TBE, which was used as an alternative buffer in some experiments.) It was observed that, at low total protein concentrations, TnsB-specific retardation of pMR11 was detected; at high protein concentrations, retardation was no longer specific either to TnsB protein or to pMR11 DNA, due to the other DNA-binding proteins present in the cell extracts (see Figure 5.5). The retardation disappeared when SDS was added to reaction samples before loading on the gel, suggesting that protein binding was responsible for the retardation; there was no indication that any nicking or double-strand cleavage of the Tn*7* ends had taken place during contact with TnsB. When pre-nicked pMR11 was used in place of the supercoiled molecule in the reactions, very slight retardation was observed; however the unbound nicked molecule runs with such a low mobility that substantial retardation would not be expected. It should be noted that SDS-treated and SDS-free samples can safely be run side by side on a binding gel without loss of binding activity in the latter.

Similar reactions were also run on a vertical composite 0.8% agarose-1.7% polyacrylamide TBE gel as described by Higgins *et al.* (1989). TnsB-specific retardation of the DNA was not however detected by this method (data not shown), and so further experiments used agarose gels.

Because DNA conformational changes do not affect mobility through agarose (Lane *et al.*, 1992), it was not clear from the initial experiments whether the specific retardation observed was indicative of synapsis of the two Tn*7* ends by TnsB, or whether it was merely the result of a large number (possibly up to 14) of TnsB molecules binding to the 22bp motifs in each pMR11 plasmid molecule. Various strategies could be employed to distinguish between these possibilities:

- Cut TnsB-bound pMR11 with an appropriate restriction enzyme. *If* synapsis of the ends were occurring this could, depending on the choice of enzyme, yield either a χ-shaped complex (with one cut in each domain), or an α-shaped molecule (with a cut in one domain only); and these complexes might be distinguished electrophoretically from the uncut species; the α-form would be expected to run slightly faster than the full-length linear molecule, and the χ-form more slowly than the linear plasmid.

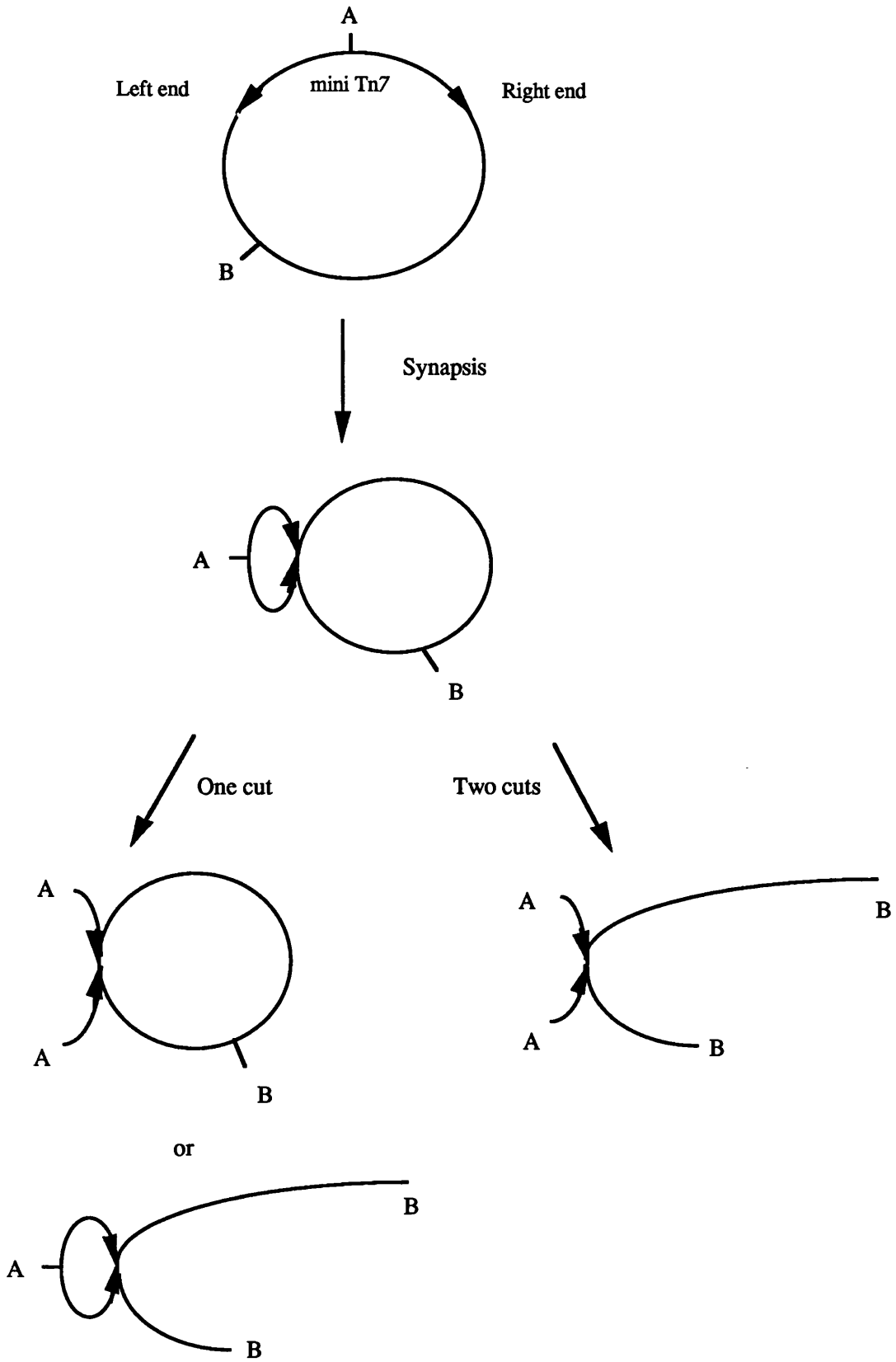


Figure 5.6 Synapsis of the two Tn*7* ends in a plasmid containing a mini-Tn*7* (not showing the supercoiling of the domains or DNA-binding proteins), followed by cleavage at one or both of the restriction sites A and B. Cleavage of the complex at one site (shown on the left with a break at B) yields an α-shaped structure, whereas cleavage at both sites results in a χ form, shown on the right.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

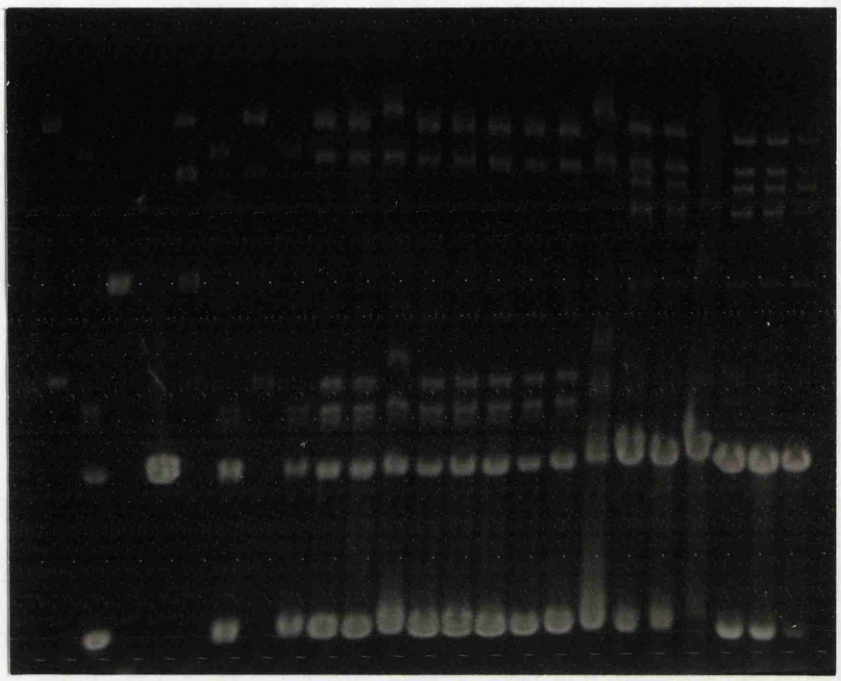


Figure 5.7 Gel retardation assay as in Figure 5.5. Where indicated, 9mM $MgCl_2$ was added and the DNA was cut with 100U EcoRI for 15 mins.

1	pMR11		uncut
2		pUC18	uncut
3	pMR11		EcoRI in standard restriction buffer
4		pUC18	EcoRI in standard restriction buffer
5	pMR11		EcoRI in binding buffer+protein dilution buffer+Mg^{2+}
6		pUC18	EcoRI in binding buffer+protein dilution buffer+Mg^{2+}
7	Lane 5	with SDS added	
8	Lane 6	with SDS added	
9	pMR11+	pUC18	binding buffer + protein dilution buffer
10	pMR11+	pUC18	binding buffer + pMR78 2^{-4} dilution of extract
11	pMR11+	pUC18	binding buffer + pMR207 2^{-4} dilution of extract
12	Lane 9	with SDS added	
13	Lane 10	with SDS added	
14	Lane 11	with SDS added	
15	Lane 9	with $MgCl_2$ added during binding	
16	Lane 10	with $MgCl_2$ added	
17	Lane 11	with $MgCl_2$ added	
18	Lane 15	with EcoRI added after binding	
19	Lane 16	with EcoRI added after binding	
20	Lane 17	with EcoRI added after binding	
21	Lane 18	with SDS added after digestion with EcoRI	
22	Lane 19	with SDS added	
23	Lane 20	with SDS added	

- Reverse the orientation of one of the Tn*7* ends on pMR11. Formation of a synapse with the resultant supercoiled plasmid would be energetically different and, if still favourable, the resultant complex would be topologically different from synapse formation with pMR11; and thus the two might possibly be distinguishable electrophoretically. If, however, synapsis were not occurring, the complex would run on a gel with the same mobility as that with pMR11.
- Use electron microscopy to examine the structure of protein-DNA complexes, both before and after digestion.

5.7.1 The search for α and χ complexes

As can be seen from Figure 5.4, pMR11 contains two EcoRI sites, one within the mini-Tn*7* and the other in the vector sequence; thus, were synapsis to take place, complete cutting with EcoRI would result in a χ-form, and partial cutting would yield an α-shaped complex, as shown in Figure 5.6. There are unfortunately no common restriction enzymes recognising only a single site in pMR11 with which an α-form might be created. It had previously been determined that, when using cell extracts of DS941 as a source of protein, the TnsB-Right End DNA complexes had a half life in reaction tubes at 37° of only about 15 minutes (see Section 4.6.5) and so it was believed that incubations of complexes with EcoRI could not be carried out for the standard 1 hour; it was hoped to compensate for this by adding a high concentration of the restriction enzyme. Although EcoRI was active in the TnsB binding buffer when 5mM or 10mM Mg^{2+} was added, and TnsB still bound the DNA in the presence of Mg^{2+}, significant digestion of pMR11 with EcoRI in the presence of crude cell extracts was not possible in 15 minutes, even when 2000U EcoRI (supplied at 2000 U μl^{-1}) was used (see Figure 5.7). There was no evidence of bands on the gel which could be attributed to α or χ structures; this could be either because no or very little synapsis took place, or because the complexes were formed but were not stable enough to survive the subsequent incubation with the restriction enzyme, as has been observed in other systems (for example Benjamin and Cozzarelli, 1988).

The second of these possibilities was further investigated by crosslinking the DNA and bound protein using glutaraldehyde. Binding reactions were set up in the absence of Tris (which would be attacked by the glutaraldehyde), treating the incubated reaction mixes with 0.1% glutaraldehyde for 5 minutes at 37° to crosslink the TnsB molecules in

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

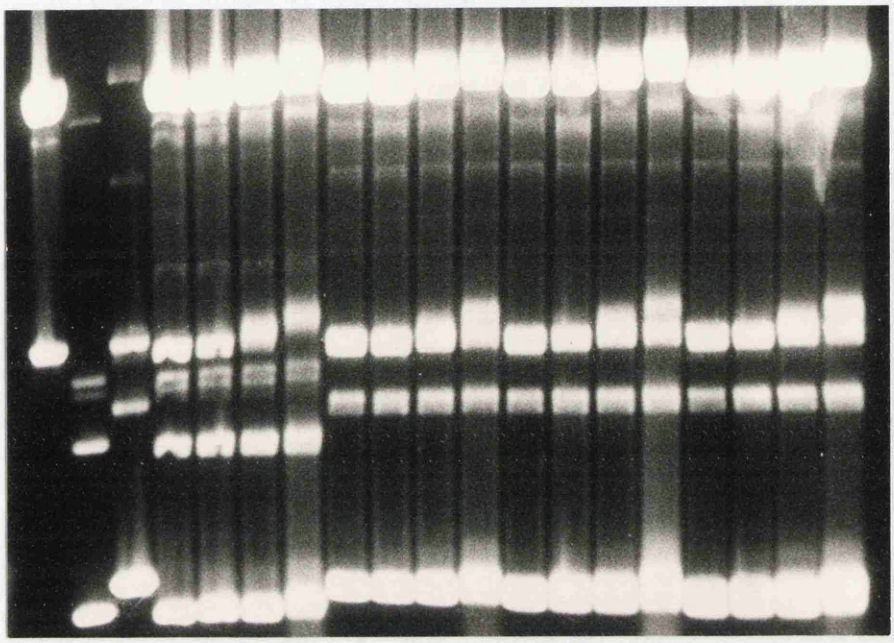


Figure 5.8 Gel retardation assay as in Figure 5.7. A protein extract of DS941(pSB58), with TnsA, TnsB, TnsC and TnsD was also tested. pMR80 (pUC with a cloned *attTn7* site) was used instead of pUC18 where shown. 10mM $MgCl_2$ was added in certain reactions either initially or after binding, as indicated.

1 pMR11
2 pUC18
3 pMR80
4 pMR11+ pUC18, binding buffer + protein extract dilution buffer
5 pMR11+ pUC18, binding buffer + pMR78 2^{-4} dilution of extract
6 pMR11+ pUC18, binding buffer + pMR207 2^{-4} dilution of extract
7 pMR11+ pUC18, binding buffer + pSB84 undiluted extract
8 pMR11+ pMR80, binding buffer + protein dilution buffer
9 pMR11+ pMR80, binding buffer + pMR78 2^{-4} dilution
10 pMR11+ pMR80, binding buffer + pMR207 2^{-4} dilution
11 pMR11+ pMR80, binding buffer + pSB84 undiluted extract
12 pMR11+ pMR80, binding buffer + protein dilution buffer + Mg^{2+} initially
13 pMR11+ pMR80, binding buffer + pMR78 2^{-4} dilution + Mg^{2+} initially
14 pMR11+ pMR80, binding buffer + pMR207 2^{-4} dilution + Mg^{2+} initially
15 pMR11+ pMR80, binding buffer + pSB84 undiluted + Mg^{2+} initially
16 pMR11+ pMR80, binding buffer + protein dilution buffer + Mg^{2+} after 8 mins
17 pMR11+ pMR80, binding buffer + pMR78 2^{-4} dilution + Mg^{2+} after 8 mins
18 pMR11+ pMR80, binding buffer + pMR207 2^{-4} dilution + Mg^{2+} after 8 mins
19 pMR11+ pMR80, binding buffer + pSB84 undiluted + Mg^{2+} after 8 mins

contact with each other, then quenching the reaction with 50mM Tris.Cl pH8.0 before adding EcoRI. Because concentrated crude cell extracts containing probably less than 1% protein as TnsB, rather than pure TnsB, were added, the glutaraldehyde would have had a wide variety of potential substrates, and so a multitude of other, uncontrollable crosslinking reactions would also have occurred, possibly affecting specific binding of TnsB. It is likely that, were pure TnsB protein to be used instead of the crude cell extracts, as well as improved specificity of crosslinking, digestion with EcoRI would be complete, since neither of the restriction sites in pMR11 should be occluded by bound TnsB. This line of investigation was suspended until such time as substantially purified TnsB became available (see Chapter 4).

At the close of this work, the protease deficient strain LA547 was made available. It then became clear (see Section 4.6.4) that the unstable TnsB-specific binding observed previously had in fact been due to proteolytic fragments of TnsB, rather than to the whole enzyme; when *tnsB* was expressed in LA547, the TnsB-specific binding to the Tn*7* Right End was stable for at least sixty minutes. The crosslinking experiment was repeated using cell extracts of LA547(pMR207) with incubation times with EcoRI for up to 60 minutes. Although cleavage by EcoRI now appeared to be complete after 50 minutes in the presence of cell extracts, there was large scale degradation of the DNA which was associated with some cellular component and brought about by addition of Mg^{2+} to the restriction digests, and it was not possible to detect any new bands in the DNA smears on the gel which resulted from this degradation (data not shown).

It is possible that, in addition to TnsB, one or more of TnsA, TnsC and TnsD is required to bring about synapsis of the Tn*7* ends *in vitro*. A crude cell extract containing TnsA, B, C and D was made from cells harbouring the plasmid pSB84, in which the four genes were under T7 transcriptional control, and was used in binding reactions using the same conditions as with extracts containing TnsB alone, and adjusting the extract concentration so that the TnsB concentration (judged by the pattern of binding to a ^{32}P-labelled pNE200 Right End fragment) was the same as with the pMR207 extract. Addition of 10mM Mg^{2+}, either initially or after 8 minutes pre-incubation, was also investigated with the pSB84 extract. (Figure 5.8) It was found that the gel retardation of pMR11 was slightly greater after incubation with pSB84 extract than with pMR207 extract, but it was not possible to characterise the different complexes. The retardation

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

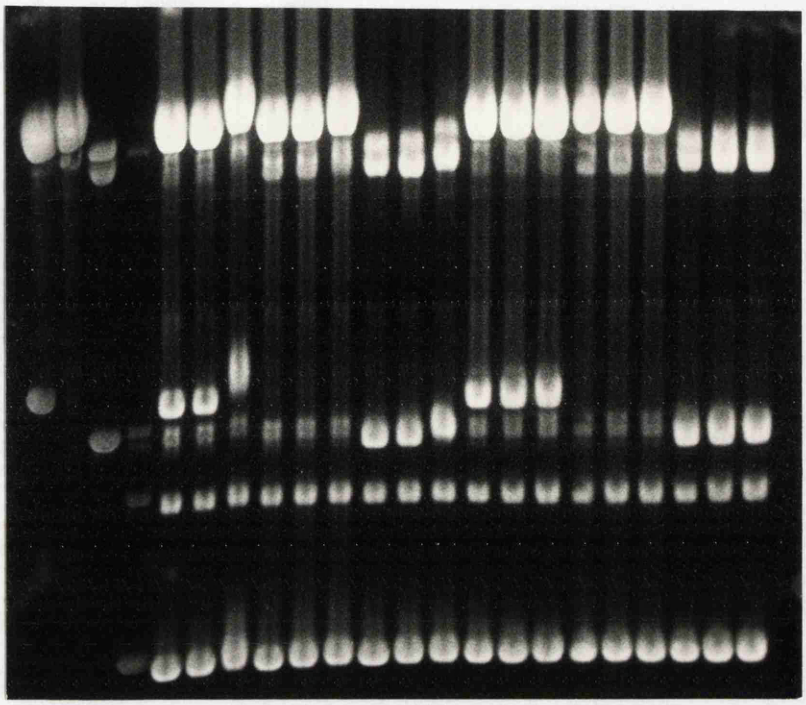


Figure 5.9 Gel retardation assay as in Figure 5.6. pMR11 was either supercoiled (s/c) or nicked (n; using DNaseI in the presence of ethidium bromide). pAG012, containing only one Tn7 end was also tested. 0.5% SDS was added after binding, where shown.

1	pMR11	supercoiled
2	pMR11	nicked
3	pAG012	supercoiled
4	pUC18	supercoiled
5	pMR11 s/c + pUC18	no protein; dilution buffer only
6	pMR11 s/c + pUC18	pMR78 2-4 dilution of extract
7	pMR11 s/c + pUC18	pMR207 2-4 dilution
8	pMR11 n + pUC18	dilution buffer only
9	pMR11 n + pUC18	pMR78 2-4 dilution
10	pMR11 n + pUC18	pMR207 2-4 dilution
11	pAG012 s/c + pUC18	dilution buffer only
12	pAG012 s/c + pUC18	pMR78 2-4 dilution
13	pAG012 s/c + pUC18	pMR207 2-4 dilution
14	Lane 5 with SDS added	
15	Lane 6 with SDS added	
16	Lane 7 with SDS added	
17	Lane 8 with SDS added	
18	Lane 9 with SDS added	
19	Lane 10 with SDS added	
20	Lane 11 with SDS added	
21	Lane 12 with SDS added	
22	Lane 13 with SDS added	

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

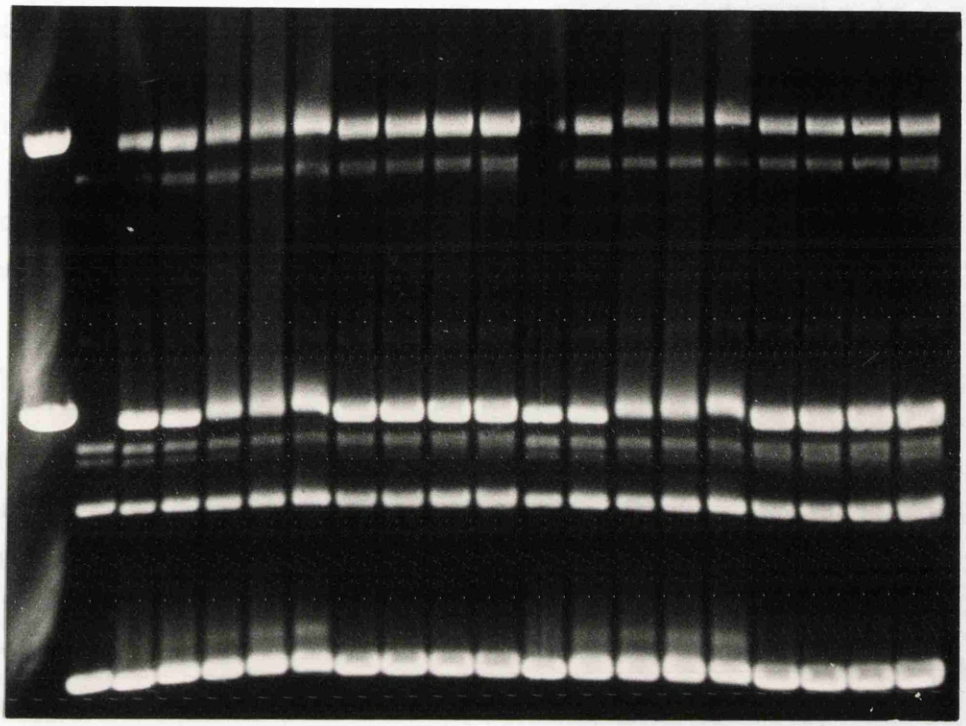


Figure 5.10 Gel retardation assay, as in Figure 5.6, except that the pMR11 concentration was halved here. pAG013 (two Tn7 ends in direct repeat) was also tested here. Protein extracts were made from DS941(pMR78), DS941(pMR207) (two different extracts) and DS941(pAG210), the latter encoding a truncated TnsB protein.

1	pMR11			
2		pUC18		
3	pMR11 +	pUC18	no protein; extract dilution buffer only	
4	pMR11 +	pUC18	pMR78	2^{-3} dilution of extract
5	pMR11 +	pUC18	pMR207	2^{-3} dilution of extract 1
6	pMR11 +	pUC18	pMR207	2^{-3} dilution of extract 2
7	pMR11 +	pUC18	pAG210	2^{-3} dilution of extract
8	Lane 4	with SDS added		
9	Lane 5	with SDS added		
10	Lane 6	with SDS added		
11	Lane 7	with SDS added		
12	pAG013 +	pUC18	dilution buffer only	
13	pAG013 +	pUC18	pMR78	2^{-3} dilution
14	pAG013 +	pUC18	pMR207	2^{-3} dilution of extract 1
15	pAG013 +	pUC18	pMR207	2^{-3} dilution of extract 2
16	pAG013 +	pUC18	pAG210	2^{-3} dilution
17	Lane 13	with SDS added		
18	Lane 14	with SDS added		
19	Lane 15	with SDS added		
20	Lane 16	with SDS added		

observed with the pSB84 extract was diminished when pMR80 (pUC with *att Tn7* cloned into the polylinker) was used as the carrier DNA for non-specific binding instead of pUC18. This may be because TnsD stabilises the Tn7 end binding / synapsis by TnsB when it is not sequestered by *attTn7*. These results were not altered by the presence of Mg^{2+} ions. There was no evidence of synapsis of the donor (pMR11) and target (pMR80) DNA molecules, which would have resulted in a species which migrated with very low mobility in gels.

5.7.2 Tn*7* ends in direct repeat: *in vitro* complex formation

As suggested above, inversion of one of the Tn7 ends in pMR11 should lead to complications in synapsis of the two ends in a supercoiled plasmid. Two derivatives of pMR11 were constructed: pAG012 has the Tn7 right end sequence completely deleted, and in pAG013 this sequence is reversed in orientation. These two plasmids were used as substrates for binding reactions alongside pMR11. (Figure 5.9, Figure 5.10). It can be seen that retardation of both pAG012 and pAG013 occurs in the presence of TnsB. pAG012 exhibits less retardation than pMR11; it cannot form synaptic complexes as it encodes only one Tn7 end, and so this retardation must be due solely to binding to the three left end motifs, L1, L2 and L3. (Intermolecular synapsis, were it to occur, would result in a DNA-protein complex much higher up the gel, and there is no evidence for this.) This observation with pAG012 demonstrates that binding not involving gross structural changes can indeed be detected with plasmids this large on an agarose gel. The retardation of pAG013 is apparently identical to that of pMR11, which is exactly the same size, strongly suggesting that the observed retardation is not in fact due to synapsis of the Tn7 ends, but simply to binding to the two ends independently.

Thus, although it may be possible to bring about synapsis of the Tn7 ends *in vitro*, it appears that the attempts described above have not been successful.

5.7.3 Tn*7* ends in direct repeat: *in vivo* transposition

As far as is known, no previous attempt had been made to investigate whether anything akin to transposition occurred when two Tn7 ends in direct repeat were provided with Tns proteins and target DNA. By analogy with other transposition and site-specific recombination systems (reviewed for example by Grindley and Reed, 1985), it was believed that correct

orientation of the transposon ends was crucial for transposition. Lack of observed transposition in the presence of directly repeated transposon ends could well be because of a failure of synapsis for topological reasons, although there are other possibilities.

In the experiments conducted here, the plasmids pMR11, pAG012 and pAG013 were used as sources of Tn*7* ends, as for the *in vitro* investigations above. The strain used was DS903, which is *recF*; although some homologous recombination may occur between mini-Tn*7* elements on the donor plasmids, any extra ends in a pAG013 dimer would remain in the same orientation, and so should not invalidate the results; dimerisation of pAG012 would similarly lead to a species with ends in direct repeat.

The standard mate-out assay was used to measure transposition; pMR121 encoded the *tns* genes. No chloramphenicol-resistant recipient colonies were recovered where either pAG012 or pAG013 was tested. Because of the large size of pEN300, it was not possible to use standard agarose gel electrophoresis to detect any insertions due to abortive transposition reactions in the donor cells. Although this may be an indication of a complete lack of strand transfer activity to the *attTn7* site of pEN300, it is also possible that some reaction occurred but that the resultant plasmids were degraded or were incapable of conjugation.

5.8 Other studies of potential synapse formation by TnsB

A different approach to the investigation of potential synapse formation was taken by Morrell (1990): ligation experiments were carried out, using as substrate a linear DNA molecule containing a 1kb EcoRI *att Tn7* fragment into which had been transposed a 0.7kb mini-Tn*7*, derived from Tn*7*-1 by deletion of the *cat* gene; control substrates with the Tn*7* ends in alternative configurations were also used. Upon incubation of mini-Tn*7* DNA and TnsB-containing fractions in the presence of T4 DNA ligase, a TnsB-specific difference in the linking number of the intramolecular product DNA was observed. It was believed that most of the trapped supercoiling was due to interactions between the transposon ends mediated by TnsB. In these experiments there was, however, no observable difference in the rate or final concentration of ligated product in the presence of TnsB, and so the part played by TnsB remains uncertain.

5.9 Discussion

DNA-protein intermediates in the cointegrate resolution reaction of Tn*3* and the transposition reaction of bacteriophage Mu have been detected by gel electrophoresis technology and electron microscopy. Excised transposon fragments have also been observed with Tn*10* (Benjamin and Kleckner, 1992) and Tn*7* (Bainton *et al.*, 1991), but in neither of these latter cases has a reaction intermediate protein-DNA complex been identified. It was hoped that the work described here would lead to evidence for one or more Tn*7*-Tns protein synapses, thus clarifying the role that the Tns proteins play in the transposition reaction. Unfortunately, perhaps due to the crude nature of the protein extracts used, there is no incontrovertible evidence for such synapses, although the results with DNA fragments on polyacrylamide gels look promising. Electron microscopy might be a feasible way of confirming the structure of protein-DNA complexes if the pure Tns proteins were available; however, it is impractical with crude cell extracts, as clear images would be unobtainable.

It is not certain that any synapsis of the Tn*7* ends occurs prior to their cleavage, but it is likely that, given the right conditions, complexes involving the transposon ends (as seen with Tn*10*) and both the transposon and the target DNA will be conclusively identified in the near future. Achieving transposition *in vitro* at all has proved, and is still proving, problematical and so perhaps it is not altogether surprising that the more difficult task of isolating protein-DNA intermediates in the *in vitro* reaction still awaits success. Four years elapsed between the setting up of a Mu *in vitro* transposition system (Mizuuchi, 1983) and the identification of protein-DNA complexed intermediates in this reaction (Surette *et al.*, 1987). It is hoped that this period will be shorter in the case of Tn*7*.

CHAPTER 6

EPILOGUE

The work reported here has been concerned with elucidating the mechanism of transposition of the prokaryotic element Tn*7*. This transposition process is clearly more complicated in Tn*7* than in the vast majority of other prokaryotic transposable elements because of the large number of genes shown to be necessary for the reaction, and much has yet to be found out about the rôles of the different genes and their protein products. The three topics studied here and described in chapters 3, 4 and 5 were principally related to two of the five *tns* genes, *tnsA* and *tnsB*.

Chapter 3 deals with *tnsA*, the gene immediately downstream of the *tns* promoter. Firstly, a wide variety of possible reasons for the poor wildtype expression of the gene was explored, and it was concluded that the poor ribosome binding site upstream of the gene played a major part in this expression. The necessity for a protein product was demonstrated. Several different strategies for overexpressing the gene in order to obtain substantial quantities of the protein product were described; the most satisfactory of these was that using a phage T7 expression system with a translational fusion which removed the dependence on the wildtype ribosome binding site.

Finally, partial purification of the protein was achieved, although it was not certain that this protein retained any activity at the end of the process. The TnsA protein proved disappointingly intransigent to solubilisation, at least in detectable concentrations, and denaturation with guanidinium chloride and urea was necessary to obtain and maintain solubility. It might have been possible to achieve greater purification by passing the protein, in a high molar concentration of urea, down a gel filtration or affinity chromatography column, but no such facility was available at the time this work was carried out. Had purification been simpler and to a greater degree, the TnsA thus obtained could have been used in functional assays.

Attempts at setting up an *in vitro* transposition system for Tn*7* were made; these were unsuccessful for reasons unknown. Such a system, if

yielding a detectable proportion of transposition product, would be invaluable in confirming the continued activity of the TnsA at each step in purification; it should also enable the isolation of reaction intermediates under various conditions, as has been achieved with bacteriophage Mu, and thus help elucidate the different steps in the transposition reaction and the requirements for each step.

When this study was started, the function of *tnsA* was totally unknown; despite the work carried out here and elsewhere, no rôle for the TnsA protein has yet been assigned. No DNA-binding activity has been detected; co-precipitation of TnsB with TnsA has been observed here, but this may simply be a reflection of the sequestration of very high concentrations of both proteins in inclusion bodies, rather than a natural tendency for the two proteins to bind each other in some multi-protein complex. It is possible that TnsA has catalytic activity in DNA cleavage or strand exchange, although the sequence similarity of TnsB with other transposases and retroviral integrases makes this latter a more likely candidate for the rôle. One possible solution is that TnsA makes essential protein-protein contacts which hold together TnsB, TnsC and TnsD (all of which have been demonstrated to bind DNA) in the correct configuration for exposure of the transposon ends and recognition of the target DNA. Experimental crosslinking of the protein molecules in the transposition complex, together with identification by antibodies, might resolve this question.

Chapter 4, the longest chapter in this thesis, dealt with various aspects of the part TnsB, or regions of it, played in transposition. The majority of the chapter was involved with testing the hypothesis that naturally produced proteolytic fragments of TnsB regulated transposition. This was examined in three distinct ways, all using artificially generated TnsB peptides to mimic the observed *in vivo* proteolysis: by gel retardation assays, studies of transcriptional repression, using galactokinase assays, and *in vivo* transposition assays with both wildtype TnsB and an excess of TnsB peptides present. Most of this work was carried out in the strain DS941, which is not protease-deficient, and thus interpretation of the results was made more difficult by the *in vivo* cleavage of the TnsB peptides which were produced; the ideal background for these studies would be a *galK htpR recA* strain. Nevertheless, some conclusions can be drawn. A specific DNA-binding activity appears to reside in the amino-terminal part of the TnsB

protein with the 22bp motifs at both the RE and LE of the transposon being bound; and *tnsB* DNA sequence between 1170bp and 1485bp from the Tn*7* RE is necessary for this activity. Where a single 22bp motif or motifs separated by other sequence are involved, further downstream TnsB residues are necessary for detectable binding. There is no evidence for either binding of a protein dimer to each copy of the motif, or cooperativity of TnsB binding. There is no indication that binding is influenced by the presence of TnsA. Under the conditions used, wildtype TnsB appears to bind more stably than small proteolytic fragments of it.

The results of the galactokinase assays suggest that some repression of the *tns* promoter is brought about by binding of the TnsB amino-terminal peptides, but that repression by wildtype TnsB is much tighter and thus, bearing in mind also the differential stability in binding, these peptides are unlikely to play a significant rôle in regulation of transposition unless present in vast excess over the wildtype protein.

Similarly, there was no detectable reduction in the *in vivo* Tn*7* transposition rate when amino-terminal peptides of TnsB were present in the cells in excess over the wildtype protein. It therefore seems unlikely that naturally produced cleavage products of TnsB play a significant part in the regulation of transposition in the way originally envisaged.

The region of TnsB shown to be essential for DNA binding activity contains the putative helix-turn-helix motif identified by Flores *et al.* (1990). Site-directed mutagenesis of the proposed recognition helix was carried out, with the intention of obtaining support for the existence of this structure; however, mutagenesis of residues thought likely to be important in DNA recognition had no effect on the binding properties of TnsB.

The carboxy-terminal region of TnsB was also examined for indications of DNA binding and associated properties; there was no clear evidence for binding and no indication of any other effects.

A purification strategy for TnsB different to that of Morrell (1990) was investigated, but there was no improvement on the previous yield; it was considered likely that only by carrying out all steps at low temperature could proteolysis be reduced to an acceptable level.

Chapter 5 described attempts to observe synapsis of the Tn*7* ends brought about by TnsB, either alone or in conjunction with other Tns proteins. These experiments were hampered by the crude nature of the

TnsB extracts used, particularly where ethidium-stained agarose gels were involved. TnsB-dependent binding could be detected, but this could not be identified with synapsis.

In view of the reported similarity between the amino acid sequences of TnsB, other prokaryotic transposases and retroviral integrases, and the absolute requirement in the case of HIV-1 for two of the residues conserved among these enzymes, it would be important for a future worker to investigate the requirement for these residues for function of TnsB; site-directed mutagenesis could be used to obtain mutant proteins which could be tested, both alone for DNA binding activity and in conjunction with other Tns proteins for evidence of impaired DNA cleavage and strand transfer. This might lead to the identification of TnsB as the enzyme in the Tn*7* system responsible for catalysing the cleavage and ligation reactions, and new insight into the rôles of the other Tns proteins.

Transposition systems generally exhibit regulation, though the particular method employed varies. Regulation of transposition can be considered as a mechanism for self-preservation on the part of the transposon, since the host cell's viability is decreased in the presence of chromosomal deletions, rearrangements and gene inactivation resulting from transposition of an element. It appears from the work described in Chapter 4 that negative regulation of transposition by TnsB peptides is unlikely to be significant. The mobility of many transposable elements is believed to be affected by the methylation state of an adenine nucleotide in a GATC sequence in one of the terminal inverted repeat sequences and/or promoter sequences; Tn10 and IS10, for example exhibit enhanced transposition immediately after passage of a replication fork, when the GATC sequence located in the pIN promoter is only hemimethylated (Roberts *et al*, 1985). Similar regulation of Tn*7* transposition might involve the methylation state of the R2 motif, although in this case it would be unlikely to be transcriptional initiation which was affected. Ekaterinaki (1987) however believed that this dependence on *dam* methylation did not occur in the case of Tn*7*. It is still not known how many transcriptional units are present in the *tns* system; if, as is believed by Waddell and Craig (1988), each of *tnsC*, *tnsD* and *tnsE* has its own promoter, there are clearly additional opportunities for regulation at the transcriptional level. Transcript mapping of the entire *tns* region would therefore be a useful activity for a future worker.

The work reported in this thesis has involved both *in vitro* and *in vivo* studies. There are two schools of thought as to whether *in vitro* studies should be an end in themselves, or whether they should be used as a means to better understanding of the *in vivo* processes involved. It is the belief of the present author that, as a life scientist, the ultimate aim of this work should be understanding how Tn*7* transposition occurs and is regulated within bacterial cells, and how the transposition process can be clinically manipulated to prevent the spread of antibiotic resistance within bacterial populations in host animals, including *Homo sapiens*. *In vitro* experimentation is beneficial insofar as it helps elucidate *in vivo* mechanisms, but it should be recognised that *in vitro* reaction conditions often bear little resemblance to those found within cells, and thus reactions *in vitro* may not occur at all *in vivo*, or they may occur by a different mechanism, and the reaction kinetics are likely to be different. The Cre-*lox* site-specific recombination reaction, for example, has been shown to display *in vivo* an orientational specificity which does not occur under the conditions used for the *in vitro* reaction (Adams *et al.*, 1992). Extrapolation of *in vitro* data to *in vivo* conditions should therefore be carried out with great caution.

Bacterial trimethoprim resistance was discovered in 1972, only three years after the introduction of trimethoprim to medical practice, in *E. coli* and the related bacterium *Klebsiella aerogenes* (Fleming *et al.*, 1972), and it was initially identified as being on a R plasmid. The recognition of the resistance as being on a transposon, originally designated Transposon C, came later (Barth *et al.*, 1976). Trimethoprim resistance due to Tn*7* subsequently spread widely through *E. coli* infections of both *Homo sapiens* and domesticated mammals such as pigs.

Tn*7*, in common with some other transposable elements such as Tn*21*, contains two distinct systems by which antibiotic resistance genes may have been transmitted: in addition to insertion of the whole transposon, it is possible that in the past these resistance genes were mobilised as part of an integron-like structure, although no functional copy of the Tn*7* integrase gene has been identified. Clinical control of the spread of antibiotic resistance must therefore take account of both possible modes of transmission of the genes. It is to be hoped that the results of the work reported here will play a part in the future management of bacterial infection in animals.

REFERENCES

Adams,D.E., Bliska,J.B. and Cozzarelli,N.R. (1992) Cre-*lox* recombination in *Escherichia coli* cells - mechanistic differences from the *in vitro* reaction. J. Mol. Biol. **226**:661-673

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

Adzuma,K. and Mizuuchi,K. (1991) Steady-state kinetic-analysis of ATP hydrolysis by the B-protein of bacteriophage-Mu - involvement of protein oligomerization in the ATPase cycle. J. Biol. Chem. **266**:6159-6167

Alazard,R., Bétermier,M. and Chandler,M. (1992) Escherichia coli integration host factor stabilises bacteriophage Mu repressor interactions with operator DNA in vitro. Mol. Microbiol. **6**: 1707-1714

Alberti,S., Oehler,S., Wilcken-Bergmann,B., Krämer,H. and Müller-Hill,B. (1991) Dimer-to-tetramer assembly of Lac repressor involves a leucine heptad repeat. New Biologist **3**:57-62

Amouyal,M., Mortensen,L., Buc,H. and Hammer,K. (1989) Single and double loop formation when DeoR repressor binds to its natural operator sites. Cell **58** :545-551

Ananthan,J., Goldberg,A.L. and Voellmy,R. (1986) Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes. Science **232**:522-524

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

Arciszewska,L.K., Drake,D. and Craig,N.L. (1989) Transposon Tn7: *cis*-acting sequences in transposition and transposition immunity. J. Mol. Biol. **207**:35-52

Arciszewska,L.K., McKown,R.L. and Craig,N.L. (1991) Purification of TnsB, a transposition protein that binds to the ends of Tn7 J. Biol. Chem. **266**:21736-21744

Bainton,R. (1992) PhD thesis, University of California at San Francisco, USA.

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

Baker,T.A. and Mizuuchi,K. (1992) DNA-promoted assembly of the active tetramer of the Mu transposase. Genes Dev. **6**:2221-2232

Baker,T.A., Mizuuchi,M and Mizuuchi,K. (1991) MuB protein allosterically activates strand transfer by the transposase of phage Mu. Cell **65**:1003-1013

Barth,P.T., Datta,N., Hedges,R.W. and Grinter,N.J. (1976) Transposition of a deoxyribonucleic acid sequence encoding trimethoprim and streptomycin resistances from R483 to other replicons. J. Bacteriol. **125**:800-810

Barth,P.T. and Grinter,N.J. (1977) Map of RP4 derived by insertion of Transposon C. J. Mol. Biol. **113**:455-474

Bastos,M.C.F. and Murphy,E. (1988) Transposon Tn*554* encodes three products required for transposition. EMBO J. **7**:2935-2941

Bednarz,A.L. (1989) Dissection of the Tn3 resolution site. PhD thesis, University of Glasgow.

Bednarz,A.L., Boocock,M.R. and Sherratt,D.J. (1990) Determinants of correct *res* site alignment in site-specific recombination by Tn*3* resolvase. Genes Dev. **4**:2366-2375

Bender,J. and Kleckner,N. (1986) Genetic evidence that Tn*10* transposes by a nonreplicative mechanism. Cell **45**:801-815

Bender,J., Kuo,J. and Kleckner,N. (1991) Genetic evidence against intramolecular rejoining of the donor DNA molecule following IS*10* transposition. Genetics **128**:687-694

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

Benjamin,H.W. and Kleckner,N. (1989) Intramolecular transposition by Tn10. Cell **59**:373-383

Benjamin,H.W. and Kleckner,K. (1992) Excision of Tn*10* from the donor site during transposition occurs by flush double-strand cleavages at the transposon termini. Proc. Natl. Acad. Sci. USA **89**:4648-4652

Berg,D.E. (1989) Transposon Tn*5*. *In* D.E.Berg and M.M.Howe, eds., "Mobile DNA". American Society for Microbiology

Berman,J., Eisenberg,S. and Tye,B.K. (1987) An agarose gel electrophoresis assay for the detection of DNA binding activities in yeast cell extracts. Methods Enzymol. **155**:528-537

Bethesda Research Laboratories (n.d.) M13 cloning/dideoxy sequencing instruction manual. BRL, Gaithersburg, MD, USA.

Biel,S.W. and Berg,D.E. (1984) Mechanism of IS*1* transposition in *E. coli*: choice between simple insertion and cointegration. Genetics **108**:319-330

Birnboim,H.C. and Doly,J. (1979) A rapid alkaline extraction procedure for screening recombinant DNA. Nucleic Acids Res. **7**:1513-1523

Bissonnette,L. and Roy,P.H. (1992) Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multi-resistance plasmids and transposons of Gram-negative bacteria. J. Bacteriol. **174**:1248-1257

Borisenko,A., Miroschnickhenko,O. and Tikchonenko,T.I. (1992) Inhibition of bovine leukemia-virus replication by the antisense RNA in cell-line CC81. Virus Res. **23**:89-97

Bowerman,B., Brown,P.O., Bishop,J.M. and Varmus,H.E. (1989) A nucleo-protein complex mediates the integration of retroviral DNA. Genes Dev. **3**:469-478

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

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

Carey,J. (1988) Gel retardation at low pH resolves *trp* repressor-DNA complexes for quantitative study. Proc. Natl. Acad. Sci. USA **85**:975-979

Caruso,M. and Shapiro,J.A. (1982) Interactions of Tn7 and temperate phage F116L of *Pseudomonas aeruginosa*. Mol. Gen. Genet. **188**:292-298

Case,C.C., Simons,E.L. and Simons,R.W. (1990) The IS*10* transposase mRNA is destabilized during antisense RNA control. EMBO J. **9**: 1259-1266

Cech,T.R. (1990) Self splicing group I introns. Annu. Rev. Biochem. **59**: 543-568

Chaconas,G., Gloor,G. and Miller,J.L. (1985) Amplification and purification of the bacteriophage Mu encoded B transposition protein. J. Biol. Chem. **260**:2662-2669

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

Chen.E.Y. and Seeburg,P.H. (1985) Supercoil sequencing - a fast and simple method for sequencing plasmid DNA. DNA **4**:165-170

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

Chou,J., Casadaban,M.J., Lemaux,P.G. and Cohen,S.N. (1979) Identification and characterization of a self-regulated repressor of translocation of the Tn*3* element. Proc. Natl. Acad. Sci. USA **76**:4020-4024

Cohen,S.N. (1976) Transposable genetic elements and plasmid evolution. Nature **263**:731-738

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

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

Craig,N.L. (1989) Transposon Tn7. *In* D.Berg and M.Howes, eds., "Mobile DNA", American Society for Microbiology.

Craig,N.L. and Kleckner,N. (1987) Transposition and site-specific recombination. *In* F.C.Neidhardt. ed., "*Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology", American Society for Microbiology.

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

Craigie,R. and Mizuuchi,K. (1985) Cloning of the A gene of bacteriophage Mu and purification of its product, the Mu transposase. J. Biol. Chem. **260**:1832-1835

Craigie,R. and 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

Dahlberg,A.E., Dingman,C.W. and Peacock,A.C. (1969) Electrophoretic characterization of bacterial polyribosomes in agarose-acrylamide composite gels. J. Mol. Biol. **41**:139-147

Danilevich,V.N. and Kostyuchenko,D.A. (1985) Immunity to repeated transposition of insertion sequence IS21. Molekulyarnaya Biologiya (Moskva) **19**:1242-1250

Datta,N. and Hedges,R.W. (1972) Trimethoprimresistance conferred by W plasmids in Enterobacteriacae. J. Gen. Microbiol. **72**:349-355

de Boer,H.A. and Kastelein,R.A. (1986) Biased codon usage: an exploration of its role in optimization of translation. *In* W.Reznikoff and L.Gold, eds. "Maximizing gene expression", Butterworths.

Derbyshire,K.M., Kramer,M. and Grindley,N.D.F. (1990) Role of instability in the cis action of the insertion sequence IS*903* transposase. Proc. Natl. Acad. Sci. USA **87**:4048-4052

Diekmann,S. (1987) DNA methylation can enhance or induce DNA curvature. EMBO J. **6**:4213-4217

Dodd,I.B. and Egan,J.B. (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18**: 5019-5026

Dodson,K.W. and Berg,D.E. (1989) Factors affecting transposition activity of IS*50* and Tn*5* ends. Gene **76**:207-213

Dumans,A.T., Carvalho,M.F. and Coelho,A. (1989) Tn*4527*, a Tp Sp/Sm transposon related to Tn*7* and flanked by IS*1*. Plasmid **22**:256-259

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

Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α helices: crystal structure of the protein-DNA structure. Cell **71**:1243-1237

Ely,B. (1982) Transposition of Tn*7* occurs at a single site on the *Caulobacter crescentus* chromosome. J. Bacteriol. **151**:1056-1058

Engelman,A. and Craigie,R. (1992) Identification of conserved amino acid residues critical for Human Immunodeficiency Virus Type I integrase function in vitro. J. Virol. **66**:6361-6369

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

Engels,W.R., Johnson-Schlitz,D.M., Eggleston,W.B. and Sved,J. (1990) High-frequency P element loss in Drosophila is homolog dependent. Cell **62**:515-525

Farnet,C.M. and Haseltine,W.A. (1990) Integration of human immunodeficiency virus type 1 DNA *in vitro*. Proc. Natl. Acad. Sci. USA **87**:4164-4168

Fayet,O., Ramond,P., Polard,P., Prere,M.F. and Chandler,M. (1990) Functional similarities between retroviruses and the IS*3* family of bacterial insertion sequences? Mol. Microbiol. **4**:1771-1777

Fickert,R. and Müller-Hill,B. (1992) How Lac repressor finds *lac* operator *in vitro*. J. Mol. Biol. **226**:59-68

Fleming,M.P., Datta,N. and Grüneberg,R.N. (1972) Trimethoprim resistance determined by R factors. Brit. Med. J. **1**:726-728

Fling,M.E., Kopf,J. and Richards,C. (1985) Nucleotide sequence of the transposon Tn*7* gene encoding an aminoglycoside-modifying enzyme, 3"(9)-*O*-nucleotidyl transferase. Nucleic Acids Res. **13**: 7095-7106

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

Flores,C., Qadri,M.I. and Lichtenstein,C. (1990) DNA sequence analysis of five genes; *tnsA, B, C, D* and *E*, required for Tn7 transposition. Nucleic Acids Res. **18**:901-911

Fuller,R.S., Kaguni,J.M. and Kornberg,A. (1981) Enzymatic replication of the origin of the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **78**:7370-7374

Galas,D.J. and Chandler,M. (1989) Bacterial insertion sequences. *In* D.E.Berg and M.M.Howe, *eds.*, "Mobile DNA", American Society for Microbiology.

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

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

Grindley,N.D.F. and Reed,R.R. (1985) Transpositional recombination in prokaryotes. Annu. Rev. Biochem. **54**:863-896

Hagemann,A.T. and Craig,N.L. (1993) Tn7 transposition creates a hotspot for homologous recombination at the transposon donor site. Genetics **133**:9-16

Hager,D.A. and Burgess,R.R. (1980) Elution of proteins from sodium dodecyl sulfate polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity - results with sigma subunit of Escherichia coli RNA polymerase, wheat-germ DNA topoisomerase, and other enzymes. Anal. Biochem. **109**:76-86

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. Mol. Microbiol. **5**:1941-1959

Halling,S.M. and Kleckner,N. (1982) A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. Cell **28**:155-163

Haniford,D.B., Benjamin,H.W. and 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. and Chaconas,G. (1992) Mechanistic aspects of DNA transposition. Current Opinion Genet. Dev. **2**:698-704

Hauer,B. and Shapiro,J. (1984) Control of Tn7 transposition. Mol. Gen. Genet. **194**:149-158

He,Y.-Y., McNally,T., Manfield,I., Navratil,O., Old,I.G., Phillips,S.E.V., Saint-Girons,I. and Stockley,P.G. (1992) Probing *met* repressor-operator recognition in solution. Nature **359**:431-433

Heim,U., Tietze,E., Weschke,W., Tschäpe,H. and Wobus,U. (1989) Nucleotide sequence of a plasmid born [sic] streptothricin-acetyl-transferase gene (sat-1) Nucleic Acids Res. **17**:7103

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

Higgins,N.P., Collier,D.A., Kilpatrick,M.W. and Krause,H.M. (1989) Supercoiling and integration host factor change the DNA conformation and alter the flow of convergent transcription in phage Mu. J. Biol. Chem. **264**:3035-3042

Hirel,P.H., Schmitter,J.M., Dessen,P., Fayat,G. and Blanquet,S. (1989) Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. USA **86**:8247-8251

Hope,I.A. and Struhl,K. (1987) GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J. **6**:2781-2784

Hudson,G.S., Evans,J.R., von Caemmerer,S., Arvidsson,Y.B.C. and Andrews,T.J. (1992) Reduction of ribulose-1,5-bisphosphate carboxylase oxygenase content by antisense RNA reduces photosynthesis in transgenic tobacco plants. Plant Physiol. **98**:294-302

Huisman,O., Errada,P.R., Signon,L. and Kleckner,N. (1989) Mutational analysis of IS10's outside end. EMBO J. **8**:2101-2109

Joachimiak,A., Kelley,R.L., Gunsalus,R.P., Yanofsky,C. and Sigler,P.B. (1983) Purification and characterization of the *trp* aporepressor. Proc. Natl. Acad. Sci. USA **80**:668-672

Kaufman,P.D. and Rio,D.C. (1992) P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. Cell **69**:27-39

Kay,A., Mandart,E., Trepo,C. and Galibert,F. (1985) The HBV hbx gene expressed in *Escherichia coli* is recognised by sera from hepatitis patients. EMBO J. **4**:1287-1292

Kelley,R.L. and Yanofsky,C. (1985) Mutational studies with the *trp* repressor of *Escherichia coli* support the helix-turn-helix model of repressor recognition of operator DNA. Proc. Natl. Acad. Sci. USA **82**:483-487

Kleckner,N. (1990a) Regulating Tn*10* and IS*10* transposition. Genetics **124**: 449-454

Kleckner,N. (1990b) Regulation of transposition in bacteria. Annu. Rev. Cell Biol. **6**:297-327

Krämer,H., Niedmöller,M., Amouyal,M., Revet,B., Wilcken-Bergmann,B. and Müller-Hill,B. (1987) Lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. EMBO J. **6**: 1481-1491

Krebs,M.P. and Reznikoff,W.S. (1986) Transcriptional and translational initiation sites of IS*50*. Control of transposase and inhibitor expression. J. Mol. Biol. **192**:781-792

Kuo,C.-F., Zou,A., Jayaram,M., Getzoff,E. and Harshey,R. (1991) DNA-protein complexes during attachment-site synapsis in Mu DNA transposition. EMBO J. **10**:1585-1591

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

Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759-1764

Lane,D., Prentki,P. and Chandler,M. (1992) Use of gel retardation to analyze protein-nucleic acid interactions. Microbiol. Rev. **56**:509-528

Lavoie,B.D., Chan,B.S., Allison,R.G. and Chaconas,G. (1991) Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu Type 1 transpososome. EMBO J. **10**:3051-3059

Lee,Y. and Schmidt,F.J. (1985) Characterization of the in vivo RNA product of the pOUT promoter of IS*10*$_R$. J. Bacteriol. **164**:556-562

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

Lichtenstein,C. and Brenner,S. (1981) Site-specific properties of Tn*7* transposition into the *E. coli* chromosome. Mol. Gen. Genet. **183**:380-387

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

Linder,P., Lasko,P.F., Ashburner,M., Leroy,P., Nielsen,P.J., Nishi,K., Schnier, J. and Slonimski,P.P. (1989) Birth of the D-E-A-D box. Nature **337**:121-122

Lund,P.A., Ford,S.J. and Brown,N.L. (1986) Transcriptional regulation of the mercury-resistance genes of transposon Tn*501*. J. Gen. Microbiol. **132**: 465-480

Lüthi,K., Moser,M., Ryser,J. and Weber,H. (1990) Evidence for a role of translational frameshifting in the expression of transposition activity of the bacterial insertion element IS*1*. Gene **88**:15-20

Ma,C. and Simons,R.W. (1990) The IS*10* antisense RNA blocks ribosome binding at the transposase translation initiation site. EMBO J. **9**:1267-1274

McClintock,B. (1948) Mutable loci in maize. Carnegie Inst. Wash. Year Book **47**:155-169

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

McCommas,S.A. and Syvanen,M. (1988) Temporal control of transposition in Tn5. J. Bacteriol. **170**:889-894

Machida,C. and Machida,Y. (1989) Regulation of IS*1* transposition by the *insA* gene product. J. Mol. Biol. **208**:567-574

McKenney,K., Shimatake,H., Court,D., Schmeissner,U., Brady,C. and Rosen berg,M. (1981) A system to study promoter and terminator sig nals recognised by the *E. coli* RNA polymerase. *In* Gene amplification and analysis, Vol.II, pp383-415, *eds*. Chirikjian, J.C. and Papas,T.S. Elsevier/North Holland, Amsterdam.

McKown,R.L., Waddell,C.S., Arciszewska,L.K. and Craig,N.L. (1987) *In vitro* identification of a transposon Tn7 - dependent DNA binding activity that recognises the ends of Tn7. Proc. Natl. Acad. Sci. USA **84**:7807-7811

MacLachlan,G. and Brady,C. (1992) Multiple forms of 1,4-beta-glucanase in ripening tomato fruits include a xyloglucanase activatable by xyloglucan oligosaccharide. Australian J. Plant Physiol. **19**: 137-146

Maekawa,T., Amemura-Mekawa,J. and Ohtsubo,E. (1993) DNA binding domains in Tn*3* transposase. Mol. Gen. Genet. **236**:267-274

Mahillon,J. and Lereclus,D. (1988) Structural and functional analysis of Tn*4430*: identification of an integrase-like protein involved in the cointegrate-resolution process. EMBO J. **7**:1515-1526

Maniatis,T., Fritsch,E. and Sambrook,J. (1982) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Martin,C., Timm,J., Rauzier,J., Gomez-Lus,R., Davies,J. and Gicquel,B. (1990) Transposition of an antibiotic resistance element in mycobacteria. Nature **345**:739-743

Martinez,E. and de la Cruz,F. (1990) Genetic elements involved in Tn*21* site-specific integration,a novel mechanism for the dissemination of antibiotic resistance genes. EMBO J. **9**:1275-1281

Maxam,A.M. and Gilbert,W. (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560-564

Mendiola,M.V. and de la Cruz,F. (1992) IS91 transposase is related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. Nucleic Acids Res. **20**:3251

Miller,J.H. (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory, NY, USA

Miller,J.L. and Chaconas,G. (1986) Electron microscopic analysis of in vitro transposition intermediates of bacteriophage Mu DNA. Gene **48**:101-108

Mizuuchi,K. (1983) In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. Cell **35**: 785-794

Mizuuchi,K. (1992) Transpositional recombination: mechanistic insights from studies of Mu and other elements. Annu. Rev. Biochem. **61**:1011-1051

Mizuuchi,K. and 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,K. and Craigie,R. (1986) Mechanism of bacteriophage Mu transposition. Annu. Rev. Genet. **20**:385-429

Mizuuchi,M., Baker,T.A. and Mizuuchi,K. (1991) DNase protection analysis of the stable synaptic complexes involved in Mu transposition. Proc. Natl. Acad. Sci. USA **88**:9031-9035

Mizuuchi,M., Baker,T.A. and Mizuuchi,K. (1992) Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. Cell **70**:303-311

Morisato,D. and Kleckner,N. (1987) Tn10 transposition and circle formation in vitro. Cell **51**:101-111

Morrell,E. (1990) Transposition of Tn7. PhD thesis, University of Glasgow.

Murphy,E. (1983) Inhibition of Tn*554* transposition: deletion analysis. Plasmid **10**:260-269

Murphy,E. and Löfdahl,S. (1984) Transposition of Tn*554* does not generate a target duplication. Nature **307**:292-294

Parsell D.A., Silber,K.R. and Sauer,R.T. (1990) Carboxy-terminal determinants of intracellular degradation. Genes Dev. **4**: 277-286

Parsons,R.L. and Harshey,R. (1988) Autoregulation of phage μ transposase at the level of translation. Nucleic Acids Res. **16**:11285-11301

Pharmacia LKB Technology AB. (n.d.) Glutathione S-transferase gene fusion system. Pharmacia data file GSDF 70-01-098

Polard,P., Prère,M.F., Chandler,M. and Fayet,O. (1991) Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS*911*. J. Mol. Biol. **222**:465-477

Polard,P., Prère,M.F., Fayet,O. and Chandler,M. (1992) Transposase-induced excision and circularization of the bacterial insertion sequence IS*911*. EMBO J. **11**:5079-5090

Pontiggia,A., Negri,A., Beltrame,M. and Bianchi,M.E. (1993) Protein HU binds specifically to kinked DNA. Mol. Microbiol. 7:343-350

Poyart-Salmeron,C., Trieu-Cuot,P., Carlier,C. and Courvalin,P. (1990) The integration-excision system of the conjugative transposon Tn*1545* is structurally and functionally related to those of lambdoid phages. Mol Microbiol. **4**:1513-1521

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

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

Ringquist,S., Shinedling,S., Barrick,D., Green,L., Binkley,J., Stormo,G.D. and Gold,L. (1992) Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. Mol. Microbiol. **6**:1219-1229

Roberts,D.E., Hoopes,B.C., McClure,W.R. and Kleckner,K. (1985) IS10 transposition is regulated by adenine methylation. Cell **43**:117-130

Roberts,D. and Kleckner,N. (1988) Tn*10* transposition promotes RecA-dependent induction of lambda prophage. Proc. Natl. Acad. Sci. USA **85**:6037-6041

Robinson,M.K., Bennett,P.M., Grinsted,J. and Richmond,M.H. (1977) Inhibition of TnA translocation by TnA. J. Bacteriol. **129**:407-414

Rogers,M.S. (1986) Transpositional properties and organisation of Tn7. PhD thesis, University of Glasgow.

Rogers,M., Ekaterinaki,N., Nimmo,E. and Sherratt,D. (1986) Analysis of Tn7 transposition. Mol. Gen. Genet. **205**:550-556

Rouvière-Yaniv,J., Yaniv,M. and Germond,J.-E. (1979) E. coli DNA binding protein HU forms nucleosome-like structures with circular double-stranded DNA. Cell **17**:265-274

Rowland,S.J. and Dyke,K.G. (1990) Tn*552*, a novel transposable element from *Staphylococcus aureus*. Mol. Microbiol. **4**:961-975

Schauder,B. and McCarthy,J.E.G. (1989) The role of bases upstream of the Shine-Dalgarno region and in the coding sequence in the control of gene expression in *Escherichia coli*: translation and stability of mRNAs in vivo. Gene **78**:59-72

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

Schurter,W. and Holloway,B.W. (1987) Interactions between the transposable element IS21 on R68.45 and Tn7 in *Pseudomonas aeruginosa* PAO. Plasmid **17**:61-64

Scott,J.R., Kirchman,P.A. and Caparon,M.G. (1988) An intermediate in transposition of the conjugative transposon Tn*916*. Proc. Natl. Acad. Sci. USA **85**:4809-4813

Sekine,Y. and Ohtsubo,E. (1989) Frameshifting is required for production of the transposase encoded by insertion sequence *1*. Proc. Natl. Acad. Sci. USA **86**:4609-4613

Senear,D.F. and Brenowitz,M. (1991) Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. J. Biol. Chem. **266**:13661-13671

Senghas,E., Jones,J.M., Yamamoto,M., Gawron-Burke,C. and Clewell,D.B. (1988) Genetic organization of the bacterial conjugative transposon Tn*916*. J. Bacteriol. **170**:245-249

Shapiro,J.A. (1979) Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. USA **76**:1933-1937

Simons,R.W. and Kleckner,N. (1983) Translational control of IS10 transposition. Cell **34**:683-691

Singer,B.S., Gold,L., Shinedling,S.T., Colkitt,M., Hunter,L.R., Pribnow,D. and Nelson,M.A. (1981) Analysis *in vivo* of the rIIB cistron of bacteriophage T4. J. Mol. Biol. **149**:405-432

Slatter,R. (1987) The regulation and properties of Tn*1/3* transposase. PhD thesis, University of Glasgow.

Smith,D.B. and Johnson,K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. Gene **67**:31-40

Smith,M. (1985) In vitro mutagenesis. Annu. Rev. Gen. **19**:423-462

Somers,W.S. and Phillips,S.E.V. (1992) Crystal structure of the *met* repressor-operator complex at 2.8Å resolution reveals DNA recognition by β-strands. Nature **359**:387-393

Sørensen,M.A., Kurland,C.G. and Pedersen,S. (1989) Codon usage determines translation rate in *Escherichia coli*. J. Mol. Biol. **207**:365-377

Stark,W.M., Sherratt,D.J. and Boocock,M.R. (1989) Site-specific recombination by Tn3 resolvase: topological changes in the forward and reversed reactions. Cell **58**:779-790

Stokes,H.W. and Hall,R.M. (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions - integrons. Mol. Microbiol. **3**:1669-1683

Strauss,F. and Varshavsky,A. (1984) A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell **37**:889-901

Sundström,L., Roy,P.H. and Sköld,O. (1991) Site-specific insertion of three structural gene cassettes in transposon Tn*7*. J. Bacteriol. **173**: 3025-3028

Surette,M.G. and Chaconas,G. (1989) A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction is *Escherichia coli* integration host factor. J. Biol. Chem. **264**:3028-3034

Surette,M.G., Buch,S.J. and Chaconas,G. (1987) Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. Cell **49**:253-262

Surette,M.G. and Chaconas,G. (1992) The Mu transpositional enhancer can function in trans: requirement of the enhancer for synapsis but not for strand exchange. Cell **68**:1101-1108

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

Tang,Y., Lichtenstein,C. and Cotterill,S. (1991) Purification and characterisation of the TnsB protein of Tn7: a transposition protein that binds to the ends of Tn7. Nucleic Acids Res. **19**: 3395-3402

Taylor,A.L. (1963) Bacteriophage-induced mutation in *E.coli*. Proc. Natl. Acad. Sci. USA **50**:1043-1051

Thomasian,S.K. and Voll,M.J. (1989) Specificity of transposition of Tn*7* in *Vibrio parahaemolyticus*. Plasmid **22**:82-85

Thomson,J.A., Hendson,M. and Magnes,R.M. (1981) Mutagenesis by insertion of drug resistance transposon Tn*7* into a Vibrio species. J. Bacteriol. **148**:374-378

Tietze,E. and Brevet,J. (1990) Nucleotide sequence of the streptothricin-acetyl-transferase gene sat-2. Nucleic Acids Res. **18**:1283

Tietze,E., Brevet,J. and Tschäpe,H. (1987) Relationships among the streptothricin resistance transposons Tn1825 and Tn1826 and the trimethoprim resistance transposon Tn7. Plasmid **18**:246-249

Tobias,J.W., Shrader,T.E., Rocap,G. and Varshavsky,A. (1991) The N-end rule in bacteria. Science **254**:1374-1377

Tomizawa,J.-I. (1984) Control of ColE1 plasmid replication: the process of binding of RNA1 to the primer transcript. Cell **38**:861-870

Tschäpe,H., Tietze,E., Heier,H., Hummel,R., Prager,R., Seltmann,G., Steinrück, H., Trolldenier,H., Witte,W.and Rische,H. (1986) Zur Resistenzentwicklung von *Escherichia coli* gegen Streptothricine (Nourseothricin, Grisin). Arch. Exper. Vet. Med., Leipzig **40**: 658-669

Turner,P., Barber,C. and Daniels,M. (1984) Behaviour of the transposons Tn*5* and Tn*7* in *Xanthomonas campestris* pv *campestris*. Mol. Gen. Genet. **195**:101-107

van Gent,D.C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1992) Mutational analysis of the integrase protein of human immunodeficiency virus type 2. Proc. Natl. Acad. Sci. USA **89**:9598-9602

Varshavsky,A. (1992) The N-end rule. Cell **69**:725-735

Vink,C., Yeheskiely,E., van der Marel,G.A., van Boom,J.H. and Plasterk,R.H.A. (1991) Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. Nucleic Acids Res. **19**:6691-6698

Vogele,K., Schwartz,E., Welz,C., Schiltz,E. and Rak,B. (1991) High-level ribosomal frameshifting directs the synthesis of IS*150* gene-products. Nucleic Acids Res. **19**:4377-4385

Waddell,C. and Craig,N. (1988) Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. Genes Dev.**2**: 137-149

Woerner,A.M., Klutch,M., Levin,J.G. and Marcus-Sekura,C.J. (1992) Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein. AIDS Res. Human Retroviruses **8**:297-304

Wu,H.-M. and Crothers,D.M. (1984) The locus of sequence-directed and protein-induced DNA bending. Nature **308**:509-513

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

Yin,J.C.P. and Reznikoff,W.S. (1988) p2 and inhibition of Tn*5* transposition. J. Bacteriol. **170**:3008-3015

Zerbib,D., Polard,P., Escoubas, J.M., Galas,D. and Chandler,M. (1990) The regulatory role of the IS*1*-encoded InsA protein in transposition. Mol. Microbiol. **4**:471-477

Zwieb,C., Kim,J. and Adhya,S. (1989) DNA bending by negative regulatory proteins: Gal and Lac repressors. Genes Dev. **3**:606-611