Title

Interaction of the Tn7 encoded protein TnsB with the DNA sequences at the transposon termini that are required for transposition

Author

Arciszewska, Lidia Kamilla

Publication Date

1990

Peer reviewed|Thesis/dissertation

Interaction of the Tn7 encoded protein TnsB with the DNA sequences at the transposon termini that are required for transposition.

by

Lidia Kamilla Arciszewska

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology

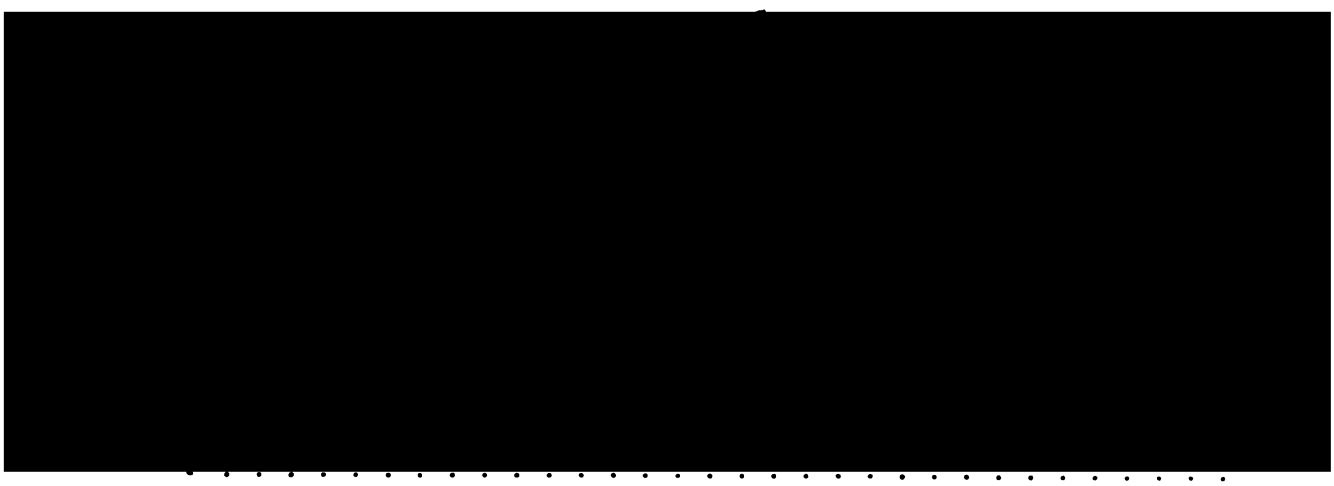
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



. . . . . .

Date University Librarian

Degree Conferred: 19

The unexamined life

is not worth living

Socrates

This thesis is dedicated to Andrzej Kasprzak

Acknowledgements

I am grateful to Bruce Alberts, Michael Bishop and Nancy Craig for their time and generous advice. I thank my colleagues for many helpful discussions and cooperation.

Interaction of the Tn7-encoded protein TnsB with the DNA sequences at the transposon termini that are required for transposition.

Lidia K. Arciszewska

Abstract

The bacterial transposon Tn7 encodes five transposition genes: *tnsA, tnsB, tnsC, tnsD,* and *tnsE*. This dissertation describes interactions of the TnsB protein with the termini of Tn7. The DNA sequences at the termini of the transposon that are required for transposition are identified in *in vivo* transposition studies. These sequences are extensive and complex but structurally different in both ends. The ends of Tn7 contain 30 base pair imperfect inverted terminal repeats, and, in addition, they are related by the presence of several copies of highly conserved 22 base pair repeats. My genetic studies suggest that these repeats are important for transposition. I have purified and characterized the TnsB protein. I show that TnsB is a sequence-specific DNA binding protein that recognizes multiple sites in the ends of Tn7. DNase I and hydroxyl radical protection experiments establish that TnsB binds to three sites in the Tn7 left end and to four sites in the Tn7 right end. The sites in the left end are directly oriented and separated by unrelated sequences; in

contrast, the sites in the right end are closely juxtaposed and overlap each other. The protection patterns of the binding sites show that the 22 base pair repeat constitutes a critical part of a TnsB binding site and indicate that the protein makes similar interactions with the seven sites in the Tn7 ends. I determine the apparent affinities of the various binding sites for TnsB. These studies reveal that TnsB has different affinity for the binding sites in each end and binds these sites in a sequential fashion. In each end, the binding begins at inner sites and progresses towards the termini of the transposon. I identify likely contacts that TnsB makes with the DNA of its binding site using a variety of high resolution protection and binding interference studies. I have also found that TnsB can bend DNA upon binding.

The presented in this thesis studies suggest that TnsB plays a critical role in Tn7 transposition by recognizing the ends of the transposon and predict that the protein is involved in promoting the formation of a higher order nucleoprotein transposition complex. The studies have also implicated TnsB in direct and indirect regulation of Tn7 transposition.

Table of Contents

Chapter I	Introduction	1
Chapter II	Transposon Tn7; *cis*-acting sequences in transposition and transposition immunity	68
Chapter III	Purification and characterization of TnsB, a transposition protein that binds to the ends of Tn7	87
Chapter IV	Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon	148
Chapter V	Conclusions	212

List of Tables

Plasmids carrying miniTn7 elements	73
Transposition of miniTn7s containing short Tn7 end segments	76
Identification of Tn7's *cis*-acting transposition sequences	77
Transposition of miniTn7 elements containing two Tn7L ends or two Tn7R ends	78
Tn7 transposition in the presence of *trans* multi-copy miniTn7 elements	79
Inhibition by individual multi-copy Tn7 ends	80
Transposition immunity in the *tnsABC+tnsE* pathway	80
Transposition immunity in the *tnsABC+tnsD* pathway	81
Inactivation of *attTn7* by adjacent Tn7 ends	82
TnsB purification	146

List of Figures

Structure of Tn7 and plasmids containing miniTn7 elements 71

Deletion variants of Tn7 ends 72

Plasmids containing Tn7 end segments 73

Structure of Tn7L and Tn7R 82

Column chromatography steps in the purification of TnsB protein 126

Protein composition of TnsB fractions 129

Determination of the native molecular weight of TnsB 131

TnsB binds to the ends of Tn7 133

Specific binding to the end of Tn7 copurifies with TnsB 135

DNase I footprinting analysis of the interaction of TnsB with the ends of Tn7 137

Nucleotide sequences of the Tn7 ends and of the regions protected by TnsB against DNase I attack 139

Sequence comparison of TnsB binding sites 142

Purified TnsB is active in Tn7 transposition *in vitro* 144

Interaction of TnsB with a single binding site as evaluated by protection and binding interference analysis 187

Summary of protection and interference analyses of a single TnsB binding site 191

Circular permutation analysis of TnsB-induced DNA bending at the γ binding site 194

Interaction of TnsB with intact Tn7 ends evaluated by hydroxyl radical footprinting 196

Summary of hydroxyl radical protection analysis of intact Tn7 ends 199

DNase I footprinting of Tn7L at various TnsB concentrations 203

Protein-DNA complexes formed between TnsB and DNA fragments containing Tn7L and Tn7R. 205

in situ phenanthroline-copper footprinting of TnsB-Tn7 complexes 207

The interaction of TnsB with a single binding site 209

CHAPTER I

Introduction

Transposable elements are specific DNA segments that can move from one genetic location to another. They were originally discovered in the late 1940's by Barbara McClintock during her pioneering genetic studies in *Zea mays* (McClintock, 1951). However, they did not receive much attention until some 20 years later when they were recognized again for their ability to be a cause of spontaneous mutation in *Escherichia coli* (Malamy, 1966; Jordon et al., 1967; Adhya and Shapiro, 1969), and also transfer of drug resistance between replicons (Hedges and Jacob, 1974; Kopecko and Cohen, 1975). Since then, it has become apparent that transposable elements are ubiquitous in nature; they have been found in the genomes of a wide variety of prokaryotic and eukaryotic species. In fact, they may comprise a substantial part of these genomes. For example, bacterial insertion sequences IS1, IS2, IS3, IS4, and IS5 constitute about 0.5% of the *Escherichia coli* genome (Shapiro, 1983), over 30 different transposable elements found in *Drosophila melanogaster* make up about 10% of the total genome (Finnegan and Fawcett, 1986), and a single LINE family of elements may comprise several percent of a mammalian genome (Fanning and Singer, 1987).

Transposable elements are responsible for a variety of genetic changes both in prokaryotes and eukaryotes. They have been shown to cause spontaneous mutations and genome rearrangements such as deletions, inversions, duplications, and replicon fusions in bacteria and chromosome translocations

in eukaryotes (Galas and Chandler, 1989; Roeder and Fink, 1983; Finnegan, 1989). Some of these genetic changes result directly from transposition events. Many others are caused by homologous recombination which can take place between multiple copies of the elements scattered within the genomes. It has become clear now that the presence and transposition of transposable elements in genomes must have important evolutionary consequences.

Another biological consequence of transposition is the transfer of genetic information within and between species. This has been well documented for bacterial transposable elements, many of which carry genes for antibiotic, heavy metal or toxin resistance. Some of these transposons have contributed significantly to the rapid spread of antibiotic resistance among bacteria. It has also been suggested recently that some eukaryotic transposons such as P elements in *Drosophila melanogaster* could be occasionally transmitted horizontally and perhaps carry along nontransposable sequences during such movement (Finnegan, 1989).

Studies on transposition have attracted many groups of researchers in recent years not only because of the role of transposons in genome rearrangements but also, and perhaps mostly, because transposons are excellent systems to study the mechanisms and enzymology of high-precision DNA transactions, protein-DNA interactions, assembly of highly organized nucleoprotein complexes, gene expression, and many other

molecular processes.

The subject of this thesis is the bacterial transposable element Tn7. The scope of my studies on Tn7 was to define the DNA sequences at the ends of the transposon that are necessary for transposition and to characterize protein-DNA interactions at these sequences. This introduction will provide a review of the interactions of proteins involved directly in transposition with DNA at the termini of the transposable elements. I shall begin with a definition of transpositional recombination and a brief description of the various mechanisms utilized by different transposable elements, and continue with the illustration of our current understanding of a transposition reaction. The major part of this review will comprise an account of the present knowledge of the protein-DNA interactions at the termini of the various transposable elements, beginning with the simplest and moving towards the more complex. Other transposition-related processes that involve binding of transposition proteins or their derivatives to DNA will also be discussed. Tn7 will be the last subject of this chapter, which will close by stating the objectives of the research work for this dissertation.

Definition of transpositional recombination

Transpositional recombination involves the joining of specific sites -- the termini of the transposon -- to a new,

generally nonspecific DNA target sequence. Several features distinguish transposition from other DNA rearrangements. Transposition is a non-reciprocal recombination event which does not require homology between the transposon and the target DNA sequences. Unlike conservative site-specific recombination -- which is another type of DNA rearrangement reaction that involves a single, specific, reciprocal crossover between two partners sharing short sequence homology -- transposition is not a conservative event and always involves at least some DNA replication. Usually, upon transposon insertion, a short sequence of the target DNA is duplicated with one copy present at each end of the newly inserted element. Target site duplication is a consequence of transposon insertion into the target DNA that has been cut with staggered breakes and subsequent filling of the single-stranded gaps. Duplication of the target site is considered to be a hallmark of transpositional recombination -only a few elements that have been classified as transposons do not obey this rule.

Although much of the current knowledge of transposition derives from studies on prokaryotic transposable elements, rapid advances have been made within the last few years in research on eukaryotic transposition.

Diverse mechanisms of transpositional recombination

All transposable elements studied can be divided into two major classes according to their mechanism of transposition. Class I - comprises those elements that transpose through a *DNA to DNA mechanism*. This class includes all identified bacterial transposable elements and among eukaryotic transposons, P elements and hobo in *Drosophila melanogaster*, Ac and En/Spm elements in *Zea mays*, and Tam elements in *Antirrhinum majus*. All these elements have short inverted repeats at their termini and encode at least one protein required for transposition -- a transposase. Class I transposable elements can be further subdivided according to how much DNA replication is involved in the transposition reaction. Three types of mechanisms have been identified: replicative, nonreplicative and conservative.

A: The *replicative transposition mechanism*. During replicative transposition, the entire transposable element is replicated by semi-conservative replication to form a cointegrate structure in which the donor and target molecules are joined together by two copies ot the element (Shapiro, 1979; Arthur and Sherratt, 1979; Pato, 1989; Sherratt, 1989). Cointegrate structures of some of the transposons can be further resolved by a site-specific recombination reaction between two internal

resolution sites *res* to produce a donor and target molecule, each carrying a copy of the element. Transposons belonging to the Tn3 family transpose using this mechanism. Other transposons that undergo replicative transposition but which do not encode cointegrate resolution systems are bacteriophage Mu and probably insertion sequences IS1 or IS903 (Galas and Chandler, 1989; Grindley and Reed, 1985). These elements can also transpose through a nonreplicative simple insertion mechanism; in fact, IS1 and IS903 predominantly transpose in this manner.

B: The *nonreplicative cut and paste mechanism*. During nonreplicative (cut and paste) transposition the transposon is cut from the donor molecule by double-stranded cleavage, and integrated into a target DNA molecule. This mechanism involves very little DNA synthesis, only enough to generate a target site duplication. Bacterial insertion sequence IS10 (Kleckner, 1989; Benjamin and Kleckner, 1989) likely IS50, and also probably eukaryotic Ac, Tam3, En/Spm and P elements transpose by this mechanism.

C: The *conservative transposition mechanism*. During this reaction the element is excised from the donor and inserted into a target molecule by conservative cleavage and ligation steps which do not involve any DNA synthesis (Caparon and Scott, 1989). Thus transposons Tn1554 and

Tn916 which transpose by this mechanism do not produce target site duplication.

Class II elements transpose through *DNA to RNA to DNA mechanism*. Transposition of transposons belonging to this class involves several steps. Transcription of the element is followed by reverse transcription of the RNA intermediate to produce a blunt-ended cDNA, which is subsequently integrated into a new site. The class II transposons include the retrotransposons that are structurally similar to integrated retroviruses, such as the copia-like elements in *Drosophila melanogaster*, Ty elements in *Saccharomyces cervisiae* and many others. Retrotransposons (Boeke, 1989; Bingham and Zachar, 1989) and retroviruses (Varmus and Brown, 1989) are similar in structural organization and in the proteins they encode. They are composed of an internal domain flanked by two several hundred base pair long, directly oriented, terminal repeats (LTR's). The proteins encoded by retrotransposons and retroviruses are functionally analogous and some of them also share sequence homology. Retrotransposons might be considered relatives of retroviruses which do not undergo an obligatory extracellular phase (Garfinkel et al., 1985, Boeke, 1989). Thus the retroviruses themselves can also be viewed as transposable elements.

Also belonging to class II are elements which are structurally distinct from retroviruses, like LINE elements in mammalian genomes, jockey and I, F, G elements in *Drosophila*

melanogaster. Their structure and coding capacity indicate that they also transpose through an RNA intermediate, however, their transposition is poorly understood and will not be discussed here.

As described above, various prokaryotic and eukaryotic transposable elements utilize a variety of routes to translocate within the genomes. It is therefore worth noting that these diverse transposition reactions are united by common features. Recent biochemical studies have revealed remarkable similarities in transposition of such distinct elements like Tn10 (Haniford et al., 1990), bacteriophage Mu (Surette et al., 1987; Craigie and Mizuuchi, 1987), Ty elements (Eichinger and Boeke, 1990), and integration of retroviruses (Craigie et al., 1990)(see below for details).

Transpositional recombination reactions

Transposition is unusual and thus particularly intriguing among recombination reactions in that it involves three DNA sites, the two transposon ends and a target site. These three sites are acted upon by a transposon-encoded protein(s) -- a transposase, which mediates the transposition reaction. Transposase is often aided by the host-encoded auxiliary proteins and almost always by DNA synthesis enzymes.

The ends of transposons define their borders and provide the specific DNA recognition sites for the proteins involved

directly in transposition. In many transposons, these recognition sites are within the short, terminal inverted repeats. In some transposons the terminal transposition sequences are much more extensive and contain repeated or different DNA sequence motifs. Comparison of the nucleotide sequence of the termini of various transposons does not reveal particular similarities among them, except among the closely related families of elements.

A target site is, in general, a more or less nonspecific DNA sequence which does not have homology to the ends of the transposon. Although most transposons insert into many different target sites, many elements display various degrees of target site preference.

Transposases are transposon-encoded proteins essential for transposition. The function of the transposase, (Grindley and Sherratt, 1978; Shapiro, 1979; Arthur and Sherratt, 1979) is the specific recognition of DNA sites at the ends of the element and the following enzymatic activities: DNA cleavage between the transposon ends and their flanking sequences, a staggered cleavage of a target site, and ligation of the transferred transposon strand with the overhanging strand of a target DNA. Many transposons encode only one protein necessary for transposition. This protein has been expected to have all the above mentioned, transposon-specific activities. One of these activities, the sequence-specific recognition of the elements ends, has been documented for several elements

(see below). However, to date, biochemical evidence regarding the enzymatic activities (cleavage and rejoining) is very limited. In one transposon, phage Mu, the enzymatic activities have been attributed to MuA transposase (still restricted evidence). The retroviral integrases have been also shown to carry all the functions required for retroviral integration (Craigie et al., 1990; Katz et al., 1990). Recent studies in some other transposable elements indicate that activities of the transposases might be shared between several transposon or even host-encoded proteins. Thus "transposase" is a functional term which refers to several specific activities that may be carried out by one or several proteins.

The proteins involved in conservative site-specific recombination such as resolvase/invertase or integrase families show striking regional sequence conservation (see in Berg and Howe, 1989). It is not apparent from the predicted amino acid sequences of the known transposases that these proteins share highly conserved structural features. Some structural similarities have been observed, however, among transposases of the related elements, for example, among groups of insertion sequences (Syvanen, 1988; Fayet et al., 1990) or within the Tn3 family (Sherratt, 1989). Structural similarities have been also found between the IS3 family and retroviral integrases (Fayet et al., 1990).

Several *host-encoded factors* have been implicated to be directly involved in transposition reactions. These are an

Escherichia coli histone-like proteins HU and integration host factor (IHF), as well as FIS and DnaA protein, and a *Drosophila melanogaster* protein that binds to the ends of the P element. The roles of these host proteins in transposition of different elements are discussed below.

How are the three DNA substrates of a transposition reaction brought together, broken and rejoined? The sequence of events described below is based mostly on biochemical studies on the best characterized prokaryotic transposon, bacteriophage Mu (Surette et al., 1987; Craigie and Mizzuchi, 1987) and on widely accepted models of transposition (Grindley and Sherratt, 1978; Shapiro, 1979, Arthur and Sherratt, 1979).

The two substrates of a transposition reaction, the termini of the transposon, are specifically recognized by a transposon-encoded protein, and brought together by protein-mediated interactions to form a synaptic complex. The formation of this complex in bacteriophage Mu *in vitro* requires a supercoiled transposon donor molecule and the presence of histone-like protein HU (Surette et al., 1987; Craigie and Mizuuchi, 1987). At this stage, the junctions between the transposon terminal nucleotides (on one strand at each end) and the flanking DNA are severed by a single-stranded cleavage to give a "cleaved donor" intermediate. The DNA nicking activity is carried out by MuA transposase (see below). The presence of cleaved donor complexes (Type I

transpososome) has been well documented by electron microscopy studies (Surette et al., 1987; Lavoie and Chaconas, 1990).

The cleaved donor intermediate is converted to a strand transfer intermediate in a transposase-mediated reaction. As originally proposed in the model of Shapiro (1979) and Arthur and Sherratt (1979), all three DNA substrates, the two transposon ends and a target site, are present in the strand transfer intermediate; The target site has been cleaved with a staggered break, and 5′ overhanging target ends have been ligated to the 3′ ends of the transposon. This intermediate complex has been also described by electron microscopy (Type II transpososome).

In bacteriophage Mu, the strand transfer intermediate is subsequently converted to a cointegrate product upon the replication of the Mu element by the host replication enzymes (Mizuuchi, 1983; Mizuuchi, 1984; Craigie and Mizuuchi, 1984). This replication reaction also generates a target site duplication. Alternatively, the strand-transfer intermediate can be resolved to a simple insertion product by nicking of a non-transferred strand and filling the gaps of a target staggered break to produce target sequence duplication. This stage of the reaction *in vitro* requires only host-encoded proteins (Mizuuchi, 1984; Craigie and Mizuuchi, 1985).

Recent biochemical studies of transposon Tn10 (Benjamin and Kleckner, 1989; Haniford et al., 1990), which undergoes nonreplicative cut-and-paste recombination, revealed that the

mechanism of its transposition is strikingly related to transposition of bacteriophage Mu. The polarity of DNA strand transfer is the same: the 3′ strand of transposon end is joined to 5′end of the overhanging target DNA. Also, the protein-DNA complexes analogous to bacteriophage Mu "cleaved donor" (excised transposon fragment, ETF) and "strand transfer intermediate" (strand transfer product) have been identified and found to have similar properties. However, unlike in phage Mu, the formation of "excised transposon fragment" intermediate proceeds through a double-stranded, not a single-stranded cleavage at the element's terminal nucleotides; so that at this stage, Tn10 is completely excised from its flanking donor DNA. A complex designated "strand transfer product" is related to phage Mu "strand transfer intermediate" and contains two single-stranded gaps which result from ligation of the staggered target DNA to flush ends of the transposon. These gaps are thought to be subsequently filled by host DNA repair synthesis.

Large nucleoprotein complex containing unintegrated viral DNA and protein activity required for its integration has been also identified in MLV retrovirus (Bowerman et al, 1989). The polarity of strand transfer reaction during retroviral integration is the same as in bacteriophage Mu and Tn10.

Rapidly accumulating genetic data and, in a few transposons, also biochemical data are beginning to unravel the protein-DNA interactions at the ends of the elements (see

below). Currently, the least understood step of the transposition reaction is the association of the target DNA with the nucleoprotein complex that contains the ends. In addition to MuA transposase, bacteriophage Mu encodes another transposition protein, MuB, whose role is to capture a target DNA (Harshey, 1983; Craigie and Mizuuchi, 1987). This protein is required for intermolecular Mu transposition and has been shown necessary for the formation of the strand transfer intermediate *in vitro* (Surette et al., 1987; Craigie and Mizuuchi, 1987; Maxwell et al., 1987). In other transposable elements, many of which encode only one protein required for transposition, this role could belong to transposase itself.

Protein-DNA interactions at the ends of the transposons

A central step in transposition is the interaction of the transposon ends with the proteins involved directly in recombination reaction and the formation of the recombinogenic complex. Although the basic goal of the protein interactions at the ends is believed to be the same in many systems -- sequence-specific recognition, subsequent DNA cleavage at the termini -- these interactions differ substantially in degree of complexity among the elements. In some transposons, specific protein interactions seem to be limited to a single protein-DNA recognition site at each transposon end and a single binding protein; in others, multiple sites and more

than one protein are required. Various transposon systems also have different requirements for host-encoded factors in transposition.

One transposase binding site.

A: Retrotransposons

Transposition of retrotransposons and retroviruses differs from transposition of other transposable elements in the early stages of this reaction. The steps leading to the usual excision of the transposon from its flanking DNA are, in retrotransposons and retroviruses, replaced with two other steps: the resident DNA segment is first transcribed into an RNA intermediate that is subsequently reverse-transcribed to generate the blunt-ended DNA intermediate, which, in turn, serves as a substrate for the integration reaction. The integration reaction itself, however, is very similar to transposition reaction: the ends of the DNA segment which usually contain imperfect, short inverted terminal repeats are recognized by the protein designated integrase (IN), the target DNA is cut with a staggered break, and the element is inserted between the overhanging strands. The inserted retroviral/retrotransposon genomes are nearly always flanked by two copies of duplicated genomic sequences.

Numerous genetic and biochemical studies and the development of cell-free systems for retroviral (Brown et al.,

1987; Fujiwara and Craigie et al., 1989) and Ty1 integration (Eichinger and Boeke, 1988) have facilitated the identification of the basic components of the integration reaction and its mechanism. Genetic studies have identified two classes of mutations which directly affect integration. In retroviruses, mutations within the region near the 3′ end of the *pol* gene abolish integration (Schwartzberg et al., 1984; Donehover and Varmus, 1984; Panganiban and Temin, 1984). This region encodes integrase (IN), which is generated by proteolytic processing of a larger precursor to yield 32kDa (ASLV and HIV) or 46 kDa (MLV) protein. Alignment of the predicted structures of all retroviral integrase proteins shows substantial similarities among them (Johnson et al., 1986). Like retroviruses, Ty elements encode a protein (p90-TYB in Ty1) which is essential for Ty integration (Boeke et al., 1988; Boeke, 1988) and bears regional sequence similarities to retroviral integrases. In yeasts retrotransposons, as well as in retrotransposons identified in other species, this protein is not encoded by the 3′ region of the *pol* gene (the coding domains are differently arranged in retrotransposons and retroviruses).

The second class of mutations that affect retroviral integration has been localized to the ends of the linear DNA segment that is a substrate in the integration reaction (Panganiban and Temin, 1984, Colicelli and Goff, 1985,1988; Cobrinik et al., 1987; Eichinger and Boeke, 1990). The ends

carry more or less perfect inverted repeats. These repeats vary in length in different viruses and retrotransposons and do not share significant sequence homology except for the presence of the invariant AC...GT, which will become terminal nucleotides of the integrated DNA segment. These dinucleotides are conserved in all retrotransposons and retroviruses. In retroviruses, they are usually situated 2 base pairs away from the termini. The 2 base pairs between the CA and the terminus are subsequently "lost" during integration (see below).

Mutational analyses of the best-characterized retroviral termini, those of MoMuLV (Colicelli and Goff, 1988; Roth et al., 1989), have shown that only 11-12 base pairs are needed for integration and that the sequence requirements for recognition of the termini are apparently quite flexible; DNA segments with only weak similarity to the original sequence can be utilized and even substitution of the highly conserved CA can be tolerated. These results indicate that substantial redundancy of contacts must be provided by the original sequence or that recognition of the ends is not very precise.

Retroviral terminal repeats are recognized by integrase (IN). DNase I protection experiments have shown that purified AMV pp32 integrase binds specifically to the DNA of the viral ends (Misra et al., 1982). Recent *in vitro* studies with purified fractions of MoMLV integrase (Craigie et al., 1990) and with purified ASLV integrase (Katzman et al., 1989; Katz et al., 1990) have established that, in addition to end

recognition, the integrase protein carries out several other functions to accomplish the integration reaction (Craigie et al., 1990). The blunt-ended viral DNA that is generated by reverse transcription requires processing -- removal of 2 nucleotides at the 3′ termini -- before it can undergo the integration reaction (Fujiwara and Mizuuchi, 1988; Brown et al., 1989). Craigie and coworkers (1990) have demonstrated that MoMLV integrase has a site-specific nuclease activity that produces the 3′ recessed ends and that it joins the 3′ recessed ends to 5′ overhanging ends of a target DNA that have been also cut by the IN protein. Similar evidence for ASLV integrase activities has been also reported by Skalka and coworkers (1990). *In vitro* studies in Ty1 have indicated that integration of this element is mechanistically related to retroviral integration, except that processing of the termini is not required; that is, the ends of unintegrated and integrated Ty DNA are the same (Eichinger and Boeke, 1990).

The above findings demonstrate that retroviral integrases are remarkably similar in function to the bacterial transposases. Both integrases and transposases (phage Mu) have a site-specific nuclease activity that cleaves at the 3′ ends of the element's DNA and an activity that joins the exposed 3′ends to 5′ ends of the target DNA that have been cleaved by the same protein.

The role of cellular proteins in retroviral integration is unclear. Fujiwara and Mizuuchi (1988) and Katz et al.

(1990) have observed that the efficiency of *in vitro* integration increases upon addition of cellular extracts. However, studies by Craigie et al. (1990) do not indicate stimulatory effects of the presence of cellular fractions on integration. The possible roles of either virion, nuclear or cytoplasmic cellular proteins in retroviral integration *in vivo* remain to be elucidated.

The recognition ot the termini of an unintegrated DNA segment by integrase does not seem to have stringent sequence requirements. This is quite unlike the end recognition in other transposons which, as described below, is rich in precision and/or complexity. Perhaps the simple recognition requirements in retroviruses/retrotransposons reflect the fact that the DNA element is already "excised" from its flanking donor molecule. Protein interactions with free termini might obey different rules; for example, even Tn3 transposase has been shown to bind to the terminus of any DNA fragment which carries protruding 5′ ends (Ichikawa et al., 1987).

B: Tn3 family of transposons

Transposons belonging to the Tn3 family transpose through a two-step replicative transposition mechanism (for review see Sherratt, 1989). They encode two recombination proteins: a transposase and a resolvase. The functions of resolvase are conservative site-specific resolution of cointegrate structures generated during the first step of transposition

reaction and control of the expression of transposase and itself. Transposases of the Tn3 family are very large, approximately 110,000 kDa, and display regions of similarity along their entire coding sequence (Mahillon and Lereclus, 1988). The inverted terminal repeats of these transposons contain all the sequences that are required at the ends for transposition. They are usually about 38 bp long and are also very similar among the elements. The inverted repeats contain several sequence motifs which are very highly conserved: the invariant 4 base pairs at the extreme termini and a few others within about a 25 base pair sequence that lies in the inner part of the repeat. Despite similarities between their proteins and terminal DNA sequences, transposases of many of the Tn3 family elements, even as closely related as Tn3 and Tn1000 ($\gamma\delta$), are not interchangable.

Biochemical studies in Tn3 and $\gamma\delta$ revealed that transposase binds specifically to the inverted repeats (Ichikawa et al., 1987; New et al., 1988; Wiater and Grindley, 1988; Ichikawa et al., 1990). DNase I protection experiments with Tn3 transposase performed by two groups gave different results. One group proposed that the protein binds the inverted repeat only within the internal 25 base pair region, positions 14 - 38, (Ichikawa et al., 1987). The other group suggested that transposase binds to the whole repeat (New et al., 1988). Also, DNase I and hydroxyl radical protection studies with $\gamma\delta$ transposase indicate that this protein makes

contacts along the entire inverted repeat (Wiater and Grindley, 1988; Wiater, 1990) and also with at least a 15 base pair region of the transposon's flanking sequence. This discrepancy in results may become a little less severe in the light of recent genetic and biochemical studies (Ichikawa et al., 1990; Wiater, 1990). Ichikawa et al. have shown that while mutations within the inner 25 base pair region of the Tn3 inverted repeat (positions 13 - 38) affect transposase binding, mutations within the outer 10 base pair region (position 1 - 10) do not, indicating that this region may not contribute to specific end recognition. However, the mutations within the outer region do affect transposition *in vivo*. Methylation interference studies with γδ transposase have shown that this protein makes base specific contacts only within the region between positions 10 - 37 (Wiater, 1990). These results suggest that the inverted repeat can be divided into two functional domains (Ichikawa et al., 1990). The inner domain is necessary for specific recognition by the transposase. The outer domain, which does not seem to be important for binding, may provide contacts with the transposase at a later step of transposition reaction or, as suggested by Ichikawa et al., it may contain sequences for specific recognition by some other factor involved in Tn3 transposition. The observed binding of γδ transposase at the outer regions of the inverted repeats and outside the transposon borders within the flanking DNA segments is likely

to be sequence nonspecific. Both, Tn3 and γδ transposases have been found to bind avidly in a sequence nonspecific fashion to certain DNA fragments (Ichikawa et al., 1987; Sherratt, 1989).

Wiater and Grindley (1988) have found that the ends of γδ are also recognized by another protein, *Escherichia coli* integration host factor (IHF), which binds immediately adjacent to the transposase binding site inside of the inverted repeat. The binding of the transposase and IHF is mutually cooperative. The role of IHF in γδ transposition is not known. Neither the frequency of transposition nor the distribution of insertions into a target replicon are altered in a host lacking IHF (Wiater and Grindley, 1990a). It should be remembered, however, that host proteins like IHF and HU can substitute for one another, and thus the *in vivo* effects of these factors on transposition are very difficult to evaluate. No IHF binding sites have been identified in the ends of Tn3.

C: Insertion sequences IS903, IS10, IS50, IS1 and the composite transposons

Insertion sequences are the shortest transposons and contain only determinants relevant to their own transposition. The insertion sequences considered in this section encode a single protein required for transposition, transposase. IS10 and likely IS50 transpose exclusively through a nonreplicative mechanism. IS903 and IS1 generally translocate by nonreplicative simple insertion but are also capable of

formation of cointegrate structures at low frequency (for reviews, see Grindley and Reed, 1985; Kleckner, 1989; Berg, 1989; Galas and Chandler, 1989). The composite transposable elements which evolved from these IS's, Tn903, Tn10, Tn5, and Tn1681, usually carry antibiotic resistance genes flanked by two inverted copies of an insertion sequence. All information necessary for transposition of composite transposons is encoded within the IS's. Tn903 contains two functional and identical copies of IS903. However, in both Tn10 and Tn5, only one IS copy is fully functional, the other copy is degenerate and carries a defective transposase gene.

The *cis*-acting DNA sequences which are essential for IS/Tn transposition are generally located within their terminal inverted repeats. These transposition sequences are believed to contain recognition sites for transposase and also, in many cases, for host-encoded protein(s). They may also carry sites for DNA methylation (see below). Since each composite transposon contains two insertion sequence modules, it therefore carries four "terminal" repeat sequences. Interestingly, in elements Tn10 and Tn5, the inside ends are somewhat different in structure from the outside ends (the inside and outside ends are named with respect to their position within the composite transposon), reflecting the possibility that transposition of these transposons and individual insertion sequences is controlled differently.

Transposases encoded by the insertion sequences IS10, IS50 and IS903 have several features in common. They are similar in size and their predicted amino acid sequences indicate significant regions of similarity in primary and secondary structure (Syvanen, 1988). Moreover, they are all *cis*-acting proteins, that is, they preferentially act at the site of their synthesis. One explanation for such activity could be that they are capable of very tight nonspecific DNA binding which could restrict their movement. Another possibility is that they are highly unstable. Recent studies (Derbyshire et al., 1990) suggest that the *cis*-action of IS903 transposase may be determined by its high instability.

The biochemical studies of interactions of the insertion sequence transposases with the DNA sequences at their ends have been complicated by the property of *cis*-action preference of their transposases, which made purification of these proteins very difficult. As a result, most information on protein interactions with the DNA at the ends comes from the indirect genetic studies. Recent detailed analyses of IS903, IS10 and IS50 indicate that inverted repeats of all three elements share a very similar functional organization.

IS903

The perfect 18 base pair inverted terminal repeats are the only sequences required for IS903 transposition (Weinert et al., 1984). The elegant mutational analysis of the repeat

revealed that almost every nucleotide position is important for efficient transposition (Derbyshire et al., 1987). These studies also suggest that the inverted repeat contains two functional domains -- a larger one for initial recognition by the transposase (positions 6 - 15) and the outer (positions 1 - 3) for another step, perhaps subsequent to transposase binding. Analysis of hierarchical effects of various nucleotide substitutions indicates that transposase makes major groove (or more complex) interactions with the outer portion of the repeat and minor groove contacts within a region between positions 13 - 16.

IS10

One of the most interesting features of IS10/Tn10 transposition is the different structural and functional requirements for inside- and outside-end transposition. Both the inside and the outside ends contain a 22 base pair long nearly perfect terminal inverted repeat. However, while the sequence of the repeat is sufficient for inside-end transposition (D. Ahmann, D. Morisato, L.Signon, P.R. Errada, and N. Kleckner, unpublished results, in Huisman et al., 1989), a much more extensive sequence (base pairs 1 - 42) is necessary for full outside-end activity (Way and Kleckner, 1984; Morisato and Kleckner, 1987; Huisman et al., 1989). *In vitro* studies have shown that outside-end transposition requires the presence of IHF or HU protein (Morisato and

Kleckner, 1987). The IHF consensus sequence has been found in the outside end sequence between base pairs 30 - 42. Moreover, DNase I footprinting analysis revealed that IHF binds specifically to the outside end and protects an extensive DNA segment between nucleotide positions 7 - 50 (D. Morisato and N. Kleckner, unpublished results, in Huisman et al., 1989).

Extensive mutational analysis of the transposition sequence of the outside end (Huisman et al., 1989) revealed that this sequence can be divided into two parts. The terminal region which comprises the inverted repeat, base pair 1 - 22; is presumably a primary site for transposase binding. The region between base pairs 23 - 42 provides the sequence information for IHF interaction.

Mutations within the inverted repeat fall into several phenotypic classes indicating the presence of several distinct functional domains (positions 1 - 3; 4 - 5; 6 - 19). The region between base pair 6 - 19 likely provides the most specific sequence determinants for transposase recognition. Interestingly, the terminal region, between base pair 1 - 3, constitutes a completely distinct domain; it appears to provide primarily nonspecific structural information and is not likely to be required for protein recognition. Finally, base pairs 4 and 5 seem to contain almost no sequence-specific information.

The effects of mutations within the region between base pair 23 - 42 on *in vivo* and *in vitro* outside end activity and

on IHF binding (Huisman et al., 1989; D. Morisato and N. Kleckner, unpublished results, in Huisman et al., 1989) indicate that binding of IHF is required for transposition involving this end and that IHF is a major host protein involved in Tn10 transposition. These results are consistent with earlier biochemical studies (Morisato and Kleckner, 1987) which showed that either IHF or HU stimulate outside end activity but that IHF is much more effective in this action.

Why do the outside ends, but not the inside IS10 ends, require host factors for transposition? There are two possible explanations. The sequences of the inverted repeats of the outside and inside ends vary at 6 positions; hence, the host factor may enhance transposase binding to the outside end. Alternatively, the host factor interaction may compensate for the presence of a possible inhibitory DNA sequence near the outside end.

It is worth noting that the inside IS10 ends, which do not require host factors for activity, contain GATC sequences and their transposition is regulated by adenine methylation which is controlled by *Escherichia coli dam* methylase (Roberts et al., 1985). The absence of methylation at this site activates the inside end as a substrate for transposition, most likely by affecting the interaction of the transposase with DNA of its recognition site. As a consequence, IS10 may only transpose efficiently when its DNA is hemimethylated, that is, just after the passage of the replication fork or

when the transposon is newly introduced into the cell by a mechanism involving a single strand DNA transfer. Therefore, transposition of the Tn10 element (outside end x outside end) and individual IS10 elements (outside end x inside end) is differentially regulated. The repression of IS10 transposition by adenine methylation may maintain the intactness of the Tn10 element.

IS50

The IS50, like IS10, also has two nonidentical ends, a 19 base pair outside end and a 19 base pair inside end (for review, see Berg, 1989). The outside and inside end sequences are identical at 8 positions within the first 9 base pairs (1 - 9). The next 10 base pairs are nonidentical. The nonidentical region of the outside end contains the site that matches the consensus sequence of the binding site for *Escherichia coli* DnaA protein. In contrast, the inside end nonidentical region contains two *dam* methylation sites. Mutational analyses of the inside and outside end sequence reveal that virtually every base is important to transposition (Phadnis and Berg, 1987; Makris et al., 1988; Dodson and Berg, 1989). Based on these studies, it has been proposed that the identical regions of the ends (positions 1 - 9) contain the primary sequence necessary for specific recognition by IS50 transposase.

Analysis of mutations within the proposed DnaA protein binding site suggests that this protein participates in IS50 transposition (Phadnis and Berg, 1987). Consistent with this hypothesis, it has been shown that DnaA protein is indeed essential for IS50 transposition (Yin and Reznikoff, 1987; Phadnis and Berg, 1987), and that the protein binds specifically to the outside end *in vitro* (Fuller et al., 1984). Several possible roles for DnaA protein in IS50 transposition have been proposed; among them, that the protein might assist in promoting or stabilizing a nucleoprotein transposition complex or that it might make the end sequence more accessible to transposition protein by DNA melting. The effects of inside end mutations support the idea that *dam* methylation directly regulates IS50 transposition (Makris et al., 1988). This regulation likely occurs by a mechanism similar to the one described above for IS10.

Recent studies have shown that integration host factor (IHF) modulates IS50 transposition under certain conditions (Makris et al., 1990). Two weak potential IHF binding sites which overlap the inside end sequence have been identified. However, several observations indicate that direct IHF interaction with DNA of the inside end does not mediate the observed modulation of transposition. Mutations that lie within the proposed IHF binding site and decrease the match to the IHF consensus sequence have no or little effect on transposition. In addition, mutations that increase the

consensus match actually decrease transposition (Dodson and Berg, 1989). Moreover, IHF fails to bind the inside end fragments *in vitro* (Makris et al., 1990).

IS1

IS1 is the only insertion sequence for which protein-DNA interactions have been biochemically studied. The 23 base pair long, nearly perfect inverted repeats, constitute all the sequences at the ends of the element required in *cis* for transposition (for review, see Galas and Chandler, 1989). Among several open reading frames identified within IS1, only two, *insA* and *insB*, are essential for transposition. The product of the *insA* gene has been identified and shown to bind specifically to the ends of IS1 (Zerbib et al., 1987). DNase I footprinting experiments (Zerbib et al., 1990) showed that *insA* protein protects almost the entire sequence of the 23 base pair inverted repeat. However, the InsA protein does not appear to be an IS1 transposase. Recent studies have shown that the active IS1 transposase is a fusion protein of InsA and InsB, generated by translational frameshifting (Sekine and Ohtsubo, 1989). The InsAB protein has been identified and shown to recognize specifically the ends of IS1 in a band shift assay (M. Chandler, personal communication). The proposed position of frameshifting between *insA* and *insB* indicates that InsAB and InsA may share a common DNA binding domain. Thus, the transposase and InsA protein may both

recognize the same DNA region within the IS1 end sequence (implications below).

The ends of IS1 are also bound by another protein, integration host factor, IHF (Gamas et al., 1987). IHF has been shown to bend DNA at the end upon binding (Prentki et al., 1987). Interestingly, the binding sites of InsA and IHF largely overlap. An intriguing possibility is that both proteins can bind the end sequence concomitantly, with InsA interacting in the major groove (InsA contains a helix-turn-helix binding motif) and IHF within the minor groove. The role of IHF binding to the ends of IS1 in transposition is not clear.

In summary, the protein-DNA interactions at the ends of the insertion sequences share several common features. First, and most striking, is their compactness. The end sequences are short and virtually every base pair appears to be essential for transposition. Second, the proposed sites of interactions with transposase can be divided into functional domains -- an inner domain for specific transposase recognition and an outer domain involved in another, perhaps later step of the transposition reaction. Third, the binding sites for transposase and the various host factors largely overlap. This last feature is particularly intriguing because of the possibility of the two proteins binding simultaneously to the same sequence. The insights from the genetic studies of

protein-DNA interactions at the ends of insertion sequences eagerly await biochemical confirmation.

One transposase - several binding sites (Bacteriophage Mu)

Bacteriophage Mu is a temperate phage which uses transposition throughout its life cycle (for reviews see, Pato, 1989; Mizuuchi and Craigie, 1986). Transposition of phage Mu occurs via both replicative and nonreplicative mechanisms. After infection, the initial integration of the phage into the host chromosome takes place by a nonreplicative, simple insertion mechanism. The replicative mechanism is used for phage multiplication during lytic growth. Mu transposition is extremely efficient; approximately 100 copies of the Mu genome are generated in less than one hour. To achieve such a high efficiency of transposition, bacteriophage Mu has developed a very complex organization of the reaction components: the *cis*-acting DNA sites and the proteins involved. The development of an *in vitro* transposition system (Mizuuchi, 1983) has provided extensive information on the nature, function and interactions between these components.

Bacteriophage Mu encodes two transposition proteins: a transposase, MuA, and MuB protein. MuA is the first transposase that has been shown to carry all the functions ascribed to these proteins. The MuA protein specifically

recognizes the ends of the transposon (Craigie et al., 1984) and, in the presence of *Escherichia coli* HU protein, can promote *in vitro* the formation of a higher-order nucleoprotein structure, a transpososome. In the transpososome, the ends of the transposon are held together in a synaptic complex (Surette et al., 1987; Craigie and Mizuuchi, 1987). *In vitro* studies have also implicated MuA for several enzymatic activities: a single-stranded cleavage at the termini of the element, a double stranded staggered cut at the target site, and ligation of the transferred transposon strand (3′) to an overlapping target strand (5′) (Surette et al, 1987; Craigie and Mizuuchi, 1987; Maxwell et al., 1987). Because MuA and HU alone are required for formation of the cleaved donor intermediate and are sufficient to generate a strand transfer intermediate, albeit extremely inefficiently (MuB is necessary for efficient reaction), and because HU is unlikely to provide enzymatic activity (Drlica and Rouviere-Yaniv, 1987), MuA must perform all these DNA breakage and rejoining steps.

In addition to MuA, bacteriophage Mu encodes another transposition protein, MuB, which is required for high-efficiency intermolecular transposition *in vivo* (Faelen et al., 1978; Harshey, 1983 and *in vitro* (Craigie et al., 1985). *In vitro* studies have determined that MuB is necessary for the formation of the strand transfer complex in which the target molecule is present (Surette et al., 1987; Craigie and Mizuuchi, 1987). However, the MuB protein does not seem to be

necessary for complex stability as it is only loosely associated with it (Lavoie and Chaconas, 1990). It has been proposed that a role of MuB, which binds DNA nonspecifically, is to facilitate interactions between the DNA target site and the pair of Mu DNA ends which are bound by MuA and HU proteins (Maxwell et al., 1987).

The role of HU in Mu transposition is not completely clear. HU is a histone-like, nonspecific DNA binding protein which is believed to have DNA wrapping properties (Drlica and Rouviere-Yaniv, 1987). As mentioned above, HU is required for the formation of the cleaved donor complex with which it remains loosely associated (Lavoie and Chaconas, 1990). Thus, the likely role of HU is organizing Mu DNA ends to allow MuA protein to promote the complex formation.

The complexity of phage Mu transposition is also reflected in the organization of the DNA sequences which directly participate in transposition. The DNA sequences at the ends of Mu are extensive and complex (not just inverted terminal repeats), and, in addition, a third *cis*-acting sequence element, an internal activation sequence or IAS, is required. Both the ends of Mu DNA and IAS are specifically recognized by MuA transposase, which contains two distinct, sequence-specific binding domains (Leung et al.,1989; Mizuuchi and Mizuuchi, 1989).

MuA binds to several sites within each end of Mu (Craigie et al., 1984). These sites are arranged differently within the

left and the right ends. The left end carries three binding sites, a terminal L1, and L2 and L3, that are approximately 90 base pairs away and are arranged in tandem. The right end contains three contiguous binding sites R1, R2 and R3; R3 is in opposite orientation with respect to the others. The proposed MuA binding sites share considerable sequence homology, mainly within several highly conserved clusters, and have no obvious internal symmetry. MuA binds noncooperatively to the sites within both ends; the terminal L1 and innermost L3 and R3 appear to be the strongest sites (Craigie et al., 1984; Groenen et al., 1987). Genetic studies have shown that mutations within the MuA binding sites affect transposition *in vivo* (Groenen and van de Putte, 1986) and binding of MuA *in vitro* (Groenen et al., 1987), providing the evidence that binding of MuA to the ends is required for Mu transposition. Genetic studies also indicate that no sequences other than the MuA binding sites in the Mu ends are necessary for transposition; however, the distance between L1 and L2/L3 is important (Groenen et al., 1985; Groenen and van de Putte, 1986).

A third *cis*-acting DNA segment, which is required *in vivo* and *in vitro* for efficient Mu transposition, an internal activation sequence or IAS, is located within the Mu operator-promoter region, approximately 900 base pairs away from the Mu left end (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989). This site is specifically recognized by the N-terminal domain

of MuA protein. *In vitro* studies have demonstrated that the interaction of the MuA protein with the IAS is necessary in an early step of the transposition reaction, the formation of a synaptic complex (Mizuuchi and Mizuuchi, 1989). The interaction of MuA with IAS is distance-independent and resembles interactions with enhancer sequences in other systems (for review see Ptashne, 1988; Wang and Giaever, 1988).

The IAS enhancer is also a site of action of another protein, IHF, which has been shown to bind specifically to the IAS (Krause and Higgins, 1986; Surette et al., 1989). The role of IHF in synaptic complex formation is not clearly understood. The standard *in vitro* transposition reaction does not require the presence of IHF. However, IHF becomes necessary in the *in vitro* reaction when a donor molecule which has *in vivo* levels of supercoiling is used (Surette and Chaconas, 1989). The supercoiling relief activity of IHF is mediated through the IAS enhancer sequence. *In vivo* studies have implicated the interaction of IHF with the IAS region also in regulating the expression of Mu repressor and transposition genes (Krause and Higgins, 1986; van Rijn et al., 1989).

Phage Mu lytic growth can be modulated by another *Escherichia coli* protein, FIS, which binds specifically to both ends of the Mu DNA (Beterminer et al., 1989). *In vivo* studies have shown that FIS affects phage Mu transposition

under certain conditions, for example when MuB is inactive or Mu repressor is overexpressed (van Drunen et al., 1990). The interaction of FIS with the Mu ends may influence the binding of transposase or other proteins to these sequences. Alternatively, as proposed by van Drunen et al. (1990), the action of FIS may have a role in the formation of a complex involving the Mu ends and the IAS.

As described above, multiple and complex interactions between the proteins and the Mu sequence elements take place during Mu transposition. One feature of these interactions invites comment. Substitution of a random DNA segment for two important for transposition binding sites, L2 and L3, in the Mu left end causes a decrease in transposition by only 10 fold (Groenen et al., 1985). This is in sharp contrast with observations made in insertion sequences such as IS50, IS903 or IS10 in which a substitution of virtually any single base pair within the essential sequences has more dramatic effects on transposition. Therefore, a reasonable conclusion might be that the complexity of the Mu transposition reaction ensures efficient phage transposition and propagation. This complexity of interactions also has a role in direct regulation of Mu transposition, as will be described below.

Several binding sites - more than one transposition protein

In many of the transposable elements that have been studied in detail, a single protein required for transposition has been found. This protein, called transposase, performs (or is believed to perform) all the specific functions necessary for transposon translocation: end recognition, formation of a transposition complex, and DNA cleavage and rejoining reactions. Recent studies indicate that in some transposons, the functions of transposase may be carried out by more than one protein. The transposons for which such a possibility has been raised are: the P element of *Drosophila melanogaster,* and plant transposons Ac/Ds and En/Spm (for reviews, see Engels, 1989; Fedoroff, 1989). These three eukaryotic transposons have several features in common. i) They are believed to transpose through a nonreplicative cut and paste mechanism. ii) The structure of the transposition sequences at their ends is complex and reflects elaborate protein-DNA interactions; the ends contain several different and/or repeated DNA sequence motifs. iii) The putative transposases of these elements recognize subterminal regions of the ends. I would like to note that these transposons are less thoroughly characterized than bacterial transposons described in previous sections, and therefore, notions regarding the nature and function of their transposition components are somewhat more speculative.

En/Spm

The ends of an En/Spm element contain 13 base pair terminal inverted repeats and have subterminal regions containing several direct and inverted copies of a 12 base pair sequence motif which are arranged differently in each end. The terminal regions consist of 180 base pairs at the 5′ end and 300 base pairs at the 3′ end (Gierl et al., 1988). Partial deletion of the 12 base pair motifs correlates with decreased excision of the element from its flanking DNA, and deletion of the terminal 2 base pairs dramatically reduces excision frequency, indicating that both structural motifs are important for transposition (Schwarz-Sommer et al., 1985).

En/Spm encodes at least two transposition proteins, TnpA and TnpB, which are derived from the same precursor transcript (Gierl et al., 1988; Masson et al., 1989). *tnpA* mRNA is a major product of this alternate maturation process (Pereira et al., 1986). TnpA is a sequence-specific DNA binding protein that recognizes the 12 base pair sequence motifs in the subterminal regions of transposon ends. However, *tnpA* alone is not sufficient to promote excision; mutant elements which still express *tnpA* but are defective in excision have been isolated; these mutants lack *tnpB* sequences. Hence, it has been speculated that a second protein necessary for transposition of En/Spm could be TnpB, which might interact

with the 13 base pair terminal inverted repeats and carry out the cleavage reaction at the element's ends (Gierl et al., 1989).

Ac

Only one Ac-encoded mRNA has been identified (Kunze et al., 1987; Finnegan et al., 1988). This transcript contains a long open reading frame whose product is believed to be the only transposon-encoded protein required for transposition (Coupland et al., 1988). The Ac-encoded protein, expressed in a baculovirus system, specifically recognizes a hexamer sequence motif (AAACGG), and binds to subterminal regions of the transposon ends (Kunze and Starlinger, 1989).

At both Ac ends, approximately 200 base pair sequences are necessary for efficient transposition (Coupland et al., 1989). These end sequences contain many copies of the hexamer sequence motif within the regions recognized by the Ac-encoded protein. Excision frequency clearly correlates with the presence of the motifs (Coupland et al., 1989), indicating that interactions of several Ac-protein molecules with DNA at each end are necessary for transposition.

The subterminal regions, which contain the hexamer motifs, are not the only sequences at the ends essential for transposition. A deletion of the portion of the 5' end located between the "motif box" and the terminus completely abolishes excision (Coupland et al., 1988). This region does not contain

the hexamer motifs and is not bound by Ac-encoded protein. The extreme termini of Ac element are 11 base pair inverted repeats. These repeats have been also shown to be essential to transposition (Hehl and Baker, 1989). Interestingly, they are not recognized by the Ac-encoded protein. Thus, sequences necessary for binding of the transposase protein are not sufficient for excision, suggesting that other protein(s) - cellular or Ac-encoded - may be involved directly in this element transposition.

P element

The ends of P element have a very complex structure. Genetic studies have shown that approximately 140 base pairs at the 5′ end and 163 base pairs at the 3′ end are necessary for transposition or excision (O'Hare and Rubin, 1983; Mullins et al., 1989). Mutational anlysis of the 3′ end has revealed that sequences throughout the end segment are important for transposition (Mullins et al., 1989). Both ends contain several sequence motifs: a 31 base pair terminal inverted repeat and an internal 11 base pair repeat. The 31 base pair repeat is essential for transposition (Karess and Rubin, 1984; Mullins et al., 1989). Mutations in the 11 base pair repeat also greatly reduce transposition. Moreover, base substitutions in several other regions throughout the essential sequences diminish the activity (Mullins etal., 1989). The structural complexity of P element ends differs

from other transposons that contain elaborate end sequences in that the ends of the P element do not contain repeated sequence motifs, and thus likely do not carry several binding sites for the same protein.

The P element encodes two polypeptides (Rio et al., 1986; Laski et al., 1986) an 87 kDa transposase which is the only transposon-encoded protein known to be necessary for transposition (Spradling and Rubin, 1982; Rubin and Spradling, 1982), and a 66 kDa repressor protein (see below). Both the transposase and the repressor share a hypothetical helix-turn-helix motif (Rio et al., 1986). Purified transposase protein binds specifically to an internal 10 base pair consensus sequence present at both 5′ and 3′ ends (Kaufman et al., 1989). These binding sites lie within the sequences known to be important for transposition (Mullins et al., 1989). The outer 16 base pair portion of the 31 base pair terminal inverted repeat is specifically recognized by a 66 kDa host protein whose function is unknown (Rio and Rubin, 1988). The fact that this DNA region is absolutely required in transposition suggests that the host factor might be involved in P element transposition. The above data strongly imply that the P element-encoded transposase may not carry all the functions necessary to accomplish transposition.

Summary

The above review of the current understanding of protein-DNA interactions at the ends of various transposable elements reveals remarkable diversity in these interactions in different systems. Such a diversity among various systems which perform basically the same reaction is not without a precedence; it has also been observed in conservative site-specific recombination. In both transposition and site-specific recombination the simplest systems -- one protein interacting with a single site per DNA partner -- provide sufficient means for the reaction to take place. The complexity of other systems, for example, Mu transposition, λ integration or Tn3 resolution, satisfies demands of either high efficiency, sophisticated direct regulation of the reaction, or its directionality (Pato, 1989; Thompson and Landy, 1989; Stark et al., 1989).

Regulation of transposition mediated by protein-DNA interactions at the ends

Transposable elements have developed intricate systems for regulating their transposition. The regulation of transposition may affect either transposase expression or the transposition reaction itself. Among the mechanisms that affect the transposition reaction directly are those that are mediated through interactions between proteins and DNA

sequences that are directly involved in transposition. Several such regulatory processes have been observed: i) direct inhibition of transposition mediated by proteins derived from or similar to transposases, ii) transposition immunity, and iii) control of transposition by host proteins interacting with the ends.

Direct inhibition of transposition

This type of inhibition has been observed in several transposons: the bacterial elements IS1, IS50 and phage Mu, and eukaryotic elements En/Spm and P element. The proteins involved in mediating direct inhibition of transposition derive from transposases. They are generated in different ways in various transposons. In IS50, the inhibitor protein is produced by translation of the transposase gene coding region from another start codon which is located downstream from the transposase start codon, i.e. the inhibitor lacks the N-terminal portion of the transposase (Isberg et al., 1982; Johnson and Reznikoff, 1984; Yin and Reznikoff, 1988; Krebs and Reznikoff, 1986). InsA, which is a negative regulator of IS1, is essentially transposase missing its C-terminal portion. InsA is produced from the same message as transposase but without the translational frameshifting which is required to generate fusion InsAB protein (Sekine and Ohtsubo, 1989; Machida and Machida, 1989; Zerbib et al., 1990). A 66 kDa repressor protein of P element is generated by tissue-specific

alternative splicing of the P element transposase message and lacks the last of the four ORF's present in the transposase gene (Misra and Rio, 1990; Roberton and Engels, 1989). The abberant polypeptide, TnpR of En/Spm, which reduces this element's transposition, shares homology with both TnpA and TnpB transposition proteins (Cuypers et al., 1988).

How can these proteins inhibit transposition? Since they likely share the specific DNA binding domains with their cognate transposases, they may recognize the same or similar DNA sequences and compete with the transposase in binding to the ends of the transposon. Alternatively, they may inhibit transposition by mixing with transposase to form heteromultimers which poison the reaction.

Recent data have revealed that InsA binds specifically to the inverted repeats at the ends of IS1 and inhibits transposition directly through competing with transposase for the binding sites (Zerbib et al., 1990). A mode of action of the two other inhibitors remains to be elucidated.

Bacteriophage Mu encodes a repressor protein that directly inhibits transposition (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989). However, unlike the negative regulators described above, Mu repressor is not a truncated version of MuA transposase but is highly related to it, particularly within the specific DNA binding domain. MuA repressor interacts with the enhancer-like IAS segment, preventing interaction of MuA protein with this site and thus directly

inhibiting the transposition reaction.

I would like to note that interactions of transposases or their relatives with the DNA sequences involved directly in transposition may also affect the transposase expression, since in some transposons, promoter regions are located within or close to these sequences. For example, InsA protein inhibits expression of IS1 transposase and itself (Zerbib et al., 1990). Binding of Mu repressor to the IAS region blocks expression of Mu transposition genes (Pato, 1989). Converesly, since MuA protein competes with repressor for the IAS binding sites, it is also likely to affect the expression of these genes (and itself; Craigie et al., 1984; Krause et al., 1986).

Transposition immunity

Several transposons have a property of inhibiting their own intermolecular transposition to a plasmid replicon they already reside in. This phenomenon, called transposition immunity, is *cis*-specific, since transposition to other DNA molecules in the cell is not affected. Transposition immunity is exhibited by the Tn3 family of transposons (Robinson et al., 1977; Sherratt, 1989), bacteriophage Mu (Adzuma and Mizuuchi, 1988), and Tn7 (Hauer and Shapiro, 1984). Numerous genetic studies have shown that the presence of the transposon ends (or even a single end) in a target molecule is necessary and sufficient to confer transposition immunity (Lee et al., 1983; Arthur et al., 1984; Huang et al., 1986; Goto et al.,

1987; Reyes et al., 1987; Darzins et al., 1988; this work Chapter II; Bainton et al., in preparation). Since the same sequences are required for transposition and transposition immunity, it has been proposed that the interactions of transposase with DNA at the ends mediate transposition immunity. Recent studies in Tn3 and the related transposon γδ have determined that, indeed, the specific binding of transposase to the ends in the target site is required as well as sufficient to exert immunity (Kans and Casabadan, 1989; Anemura et al., 1990; Wiater and Grindley, 1990b). These studies have shown that deletions or base substitutions in the end, within the sequence essential for specific recognition by the transposase, abolish transposition immunity, whereas those that do not affect transposase binding have no effect on immunity. Moreover, IHF, which stimulates binding of γδ transposase to the γδ ends, also stimulates immunity (Wiater and Grindley, 1990a). These studies also indicated that the end segment in a target molecule need not be transposition-proficient to provide immunity.

The mechanism of transposition immunity has been elucidated for bacteriophage Mu in *in vitro* studies (Adzuma and Mizuuchi, 1988,1989). Intermolecular transposition of Mu requires MuA transposase and MuB protein. MuB protein binds nonspecifically to DNA, and its role in Mu transposition is to capture target DNA and allow it to associate with the synaptic complex of transposon ends held by MuA protein. In the absence

of MuB, intermolecular transposition is extremely inefficient. The presence of MuA protein bound to a Mu end in the target molecule facilitates dissociation of MuB from this DNA, limiting its use as a target.

The mechanism of Tn3 immunity has not been elucidated. One model proposes that a requirement for a specific topology of the synaptic complex might be a basis for Tn3 immunity; the presence of the ends in the target may interfere with the productive formation of this complex (Sherratt, 1989). Alternatively, the mechanism of Tn3 immunity might be related to that of Mu. It has been suggested that Tn3 transposase might be a fusion of two functional components, equivalent to MuA and MuB proteins (Sherratt, 1989).

Control by host proteins interacting with the ends

Transposition of at least two bacterial elements, IS10 and IS50, is subject to regulation by adenine methylation (Roberts et al., 1985; Yin et al., 1988; Dodson and Berg, 1989). In both cases, this regulation affects transposase gene expression and the transposition reaction itself. Both elements contain GATC sites positioned within the transposase binding site at one of the ends; it has been proposed that methylation of these sites by the *Escherichia coli*-encoded *dam* methylase directly affects interaction of transposase with DNA of the end and therefore inhibits transposition. The likely consequence of *dam* regulation is that transposition can

occur only when DNA of the element is hemimethylated, that is during a certain time in cell cycle after the passage of the replication fork, and also when the element is introduced to a new host by a mechanism like conjugation.

Regulation of transposition by cytosine methylation has been also reported in plant transposable elements (Fedoroff, 1989). This regulation likely affects directly the transposition reaction. Binding of the Ac-encoded transposition protein and TnpA protein of En/Spm to their respective recognition sequence motifs at the ends depends on the state of methylation of a substrate DNA (Gierl et al., 1988; Kunze and Starlinger,1989).

As described above, several transposons require a direct participation in transposition reaction of host-encoded accessory proteins, like *Escherichia coli* IHF, HU, FIS and DnaA protein. Involvement of host factors in transposition raises the possibility that these proteins may play a regulatory role and link transposition to the physiological and environmental state of the host.

Transposon Tn7

Tn7 is a large (14 kilobase in length) bacterial transposon which encodes resistances to trimethoprim, and to streptomycin and spectinomycin (Barth et al., 1976; for review, see Craig, 1989). Although originally identified in

Escherichia coli (Hedges et al., 1972), Tn7 has been found in the chromosomes of many other bacterial species, including *Vibrio* species, *Caulobacter crescentis*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, and *Agrobacterium tumefaciens*.

Tn7 is distinguished among other transposable elements for its remarkable property of high-frequency site-specific transposition. In addition, Tn7 is capable of low frequency transposition to many other, apparently random target sites. To accomodate the unusual properties of its transposition, Tn7 encodes five transposition genes *tnsA*, *tnsB*, *tnsC*, *tnsD* and *tnsE* (Rogers et al., 1986; Waddell and Craig, 1988). *tnsA*, *tnsB* and *tnsC* are required in all transposition events; *tnsD* is also necessary for high frequency site-specific transposition, whereas *tnsE* mediates low frequency transposition to nonspecific target sites. Tn7 transposition proteins work efficiently *in trans*. The high frequency site-specific transposition and the complex array of transposition proteins make Tn7 an ideal system for studying transpositional recombination at the biochemical level.

Early genetic studies established several basic features of Tn7 transposition (Barth et al., 1976; Lichtenstein and Brenner, 1981, 1982; Smith and Jones, 1984, Hauer and Shapiro, 1984; Quartsi et al., 1985).

i) The specific site of Tn7 insertion into *Escherichia coli* chromosome (*attTn7*) lies at approximately

minute 84 between *phoS*, a gene involved in phosphate transport, and *glmS*, a gene involved in cell wall biosynthesis. Later studies (McKown et al., 1988; Gringauz et al., 1988) have shown that specific DNA sequences required for *attTn7* activity lie to one side and away from the insertion point. Tn7 inserts into *attTn7* in a unique orientation. In *attTn7*, the left end of Tn7, which is adjacent to its drug resistance determinants, is always adjacent to *phoS*, and the right end, which is adjacent to Tn7 transposition genes, is always proximal to *glmS*. Tn7 produces a 5 base pair target site duplication upon insertion.

ii) The sequences required in *cis* for Tn7 transposition lie at the ends of the transposon. Tn7 derivatives containing DNA sequences extending approximately 2000 base pairs from the terminus of the Tn7 left end and approximately 500 base pairs from the terminus of the right end are active in transposition when provided with Tn7-encoded transposition proteins *in trans*. It is unlikely that Tn7 encodes a cointegrate resolution system similar to that in the Tn3 family of transposons. The ends contain 30 base pair imperfect inverted terminal repeats and, in addition, carry several copies of a 22 base pair sequence motif. The

terminal inverted repeats are not sufficient for transposition. There is no apparent sequence homology between the nucleotide sequences at the ends of the transposon and DNA sequences at the target site.

iii) Tn7 displays transposition immunity. The presence of Tn7 or its derivative in a target plasmid reduces the frequency of subsequent Tn7 transposition to this molecule.

Objectives

The objective of this work was to define DNA sequences at the ends of transposon Tn7 that are necessary for transposition and to characterize protein interactions with these sequences.

I employed deletion analysis to identify the sequences at the ends required for transposition. I have also determined sequence requirements for transposition immunity. The results of these analyses are presented in Chapter II. I have purified TnsB and established that it is a sequence-specific DNA binding protein that recognizes multiple sites in both Tn7 ends. This work is the subject of Chapter III. The results of high resolution studies on the interactions of the TnsB protein with its binding sites at the ends are presented in Chapter IV.

Chapter II has been published in the *J. Mol. Biol.*: L. K. Arciszewska, D. Drake and N. L. Craig, "Transposon Tn7, *cis*-acting sequences in transposition and transposition immunity". D. Drake constructed the miniTn7 elements containing two left and two right ends and performed experiments showing the functional difference of the ends, Table 2.4, line 2 and 3, column 1 and 2. The text of Chapter III has been submitted to the *J.Biol.Chem.*: L. K. Arciszewska, R. L. McKown and N. L. Craig, "Purification and characterization of TnsB, a transposition protein that binds to the ends of Tn7". The purification procedure has been developed and all the experiments in this paper have been performed by myself. R. McKown's contribution to this work was: construction of a vector plasmid used as a source of *tnsB*, the initial TnsB purification by a different procedure, and providing the material for pilot footprinting experiments. The content of Chapter IV constitutes a manuscript which is to be submitted to *EMBO* Journal: L. K. Arciszewska and N. L. Craig "Interactions of the Tn7-encoded transposition protein TnsB with the ends of the transposon".

REFERENCES

Adhya, S. L., and Shapiro, J. A. (1969) *Genetics*, 62, 231-247

Adzuma, K., and Mizuuchi, K. (1988) *Cell* 53, 257-266

Adzuma, K., and Mizuuchi, K. (1989) *Cell* 57, 41-47

Amemura, J., Ichikawa, H., and Ohtsubo, E. (1990) *Gene* 88, 21-24

Amman, E., Brosius, J., and Ptashne, M. (1983) *Gene* 25, 167-178

Arciszewska, L. K., Drake, D., and Craig, N.L. (1989) *J. Mol. Biol.* 207, 35-52

Arthur, A., Nimmo, E. R., Hettle, S. J. H., and Sherratt, D. J. (1984) *EMBO J.* 3,1723-1729

Arthur, A., and Sherratt, D. (1979) *Mol. Gen. Genet.* 175, 267-274

Bainton, R., Gamas, P., and Craig, N. L., manuscript in preparation

Barth, P. T., Datta, N., Hedges, R. W., and Grinter, N. J. (1976) *J.Bacteriol.* 125, 800-810

Benjamin, H. W., and Kleckner, N. (1989) *Cell* 59, 373-383

Berg, D. E.,(1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds., American Society for Microbiology, Washington D. C., pp. 185-210

Betermier, M., Lefrere, V., Koch, C., Alazard, R., and Chandler, M. (1989) *Mol. Microbiol.* 3, 459-468

Bingham, P. M., and Zachar, Z. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds., American Society for Microbiology, Washington D. C., pp. 485-502

Boeke, J. D. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds., American Society for Microbiology, Washington D. C., pp. 335-374

Bowerman, B., Brown, P. O., Bishop, J. M., and Varmus, H. E. (1989) *Genes Develop.* 3, 469-478

Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1987) *Cell* 49, 347-356

Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 2525-2529

Caparon, M. G., and Scott, J. R. (1989) *Cell* 59, 1027-1034

Cobrinik, D., Katz, R., Terry, R., Skalka, A. M., and Leis, J. (1987) *J. Virol.* 61, 1999-2008

Colicelli, J., and Goff, S. P. (1985) *Cell* 42, 573-580

Colicelli, J., and Goff, S.P. (1988) *J. Mol. Biol.* 199, 47-59

Coupland, G., Baker, B., Schell, J., and Starlinger, P. (1988) *EMBO J.* 7, 3653-3659

Coupland, G., Plum, C., Chatterjee, S., Post, A., and Starlinger, P. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 9385-9388

Craig, N. L. (1989) in *Mobile DNA,* Berg, D. E., and Howe, M. M. eds.,American Society for Microbiology, Washington, D. C. pp. 211-226

Craigie, R., Arndt-Jovin, D. J., Mizuuchi, K. (1985) *Proc. Natl. Acad. Sci. U. S. A.* 82,7570-7574

Craigie, R., Fujiwara, T., and Bushman, F. (1990) *Cell* 62, 829-837

Craigie, R., and Mizuuchi, K. (1985) *Cell* 41, 867-876

Craigie, R., and Mizuuchi, K. (1987) *Cell* 51, 493-501

Craigie, R., Mizuuchi, M., and Mizuuchi, K. (1984) *Cell* 39, 387-394

Cuypers, H., Dash, S., Peterson, P. A., Saedler, H., and Gierl, A. (1988) *EMBO J.* 7, 2953-2960

Darzins, A., Kent, N. E., Buckwalter, M. S., and Casabadan, M. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* 85, 6826-6830

Derbyshire, K. M., Hwang, L., and Grindley, N. D. F. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 8049-8053

Derbyshire, K. M., Kramer, M., and Grindley, N. D. F. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 4048-4052

Dodson, K. W., and Berg, D. E. (1989) *Gene* 85, 75-81

Donehower, L. A., and Varmus, H. E. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 6461-6465

Drlica, K., and Rouviere-Yaniv, J. (1987) *J. Microbiol. Rev.* 51, 301-319

Eichinger, D. J., and Boeke, J. D. (1988) *Cell* 54, 955-966

Eichinger, D. J., and Boeke, J. D. (1990) *Genes Develop.* 4, 324-330

Engels, W. R. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds. American Society for Microbiology, Washington D. C., pp. 437-484

Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B., and Sved, J. (1990) *Cell* 62, 515-525

Faelen, M., Huisman, O., and Toussaint, A. (1978) *Nature (London)* 272, 580-582

Fanning, T. G., and Singer, M. F. (1987) *Biochim. Biophys. Acta* 910, 203-212

Fayet, O., Ramond. P., Polard, P., Prere, M. F., and Chandler, M. (1990) *Mol. Microbiol.* 4, 1771-1777

Fedoroff, N. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds. American Society for Microbiology, Washington D. C., pp 375-411

Finnegan, D. J. (1989) *Trends in Genet.* 5, 103-107

Finnegan, D. J., and Fawcett,D. H. (1986) *Oxford Surveys of Eukaryotic Genes* Maclean, N. ed., 3, pp. 1-62

Finnegan, E. J., Taylor, B. H., Dennis, E. S., and Peacock, W. J. (1988) *Mol. Gen. Genet.* 212, 505-509

Flores, C., Qadri, M. I., and Lichtenstein, C. (1990) *Nucl. Acids Res.* 18, 901-911

Fujiwara, T., and Craigie, R. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86,3065-3069

Fujiwara, T., and Mizuuchi, K. (1988) *Cell* 54, 497-504

Fuller, R. S., Funnell, B. E., and Kornberg, A. (1984) *Cell* 38, 889-900

Galas, D. J., and Chandler, M. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds., American Society for Microbiology, Washington D. C., pp. 109-162

Galas, D. J., and Schmitz, A. (1978) *Nucl. Acids Res.* 5, 3157-3170

Garfinkel, D. J., Boeke, J. D., and Fink, G. R. (1985) *Cell* 42, 507-517

Gay, N. J., Tybulewicz, V. L. J., and Walker, J. E. (1986) *Biochem. J.* 234, 111-117

Gamas, P., Bainton, R., and Craig, N. L., in preparation

Gamas, P., Chandler, M., Prentki, P., and Galas, D. J. (1987) *J. Mol. Biol.* 195, 261-272

Gierl, A., Lütticke, S., and Saedler, H. (1988) *EMBO J.* 7, 4045-4053

Gierl, A., Saedler, H., and Peterson, P. A. 1989 *Annu. Rev.Genet.* 23, 71-85

Goto, N., Mochizuki, A., Inagaki, Y., Horiuchi, S., Tanaka, T., and Nakaya, R. (1987) *J. Bacteriol.* 169, 4388-4390

Grindley, N. D., and Sherratt, D. J., (1978) *Cold Spring Harbor Symp. Quant. Biol.* 43, 1257-1261

Grindley, N., and Reed, R. (1985) *Annu. Rev. Biochem.* 54, 863-896

Gringauz, E., Orle, K. A., Waddell, C. S., and Craig, N. L. (1988) *J. Bacteriol.* 170, 2832-2840

Groenen, M. A. M., Kokke, M., and van de Putte, P. (1986) *EMBO J.* 5, 3687-3690

Groenen, M. A. M., Timmers, E., and van de Putte, P. (1985) *Proc. Natl. Acad. Sci. U. S. A.* 82, 2087-2091

Groenen, M. A. M., and van de Putte, P. (1986) *J. Mol. Biol.* 189, 597-602

Groenen, M. A. M., Vollering, M., Krijgsman, P., van Drunen, K., and van de Putte, P. (1987) *Nucl. Acids Res.* 15, 8831-8844

Haniford, D. B., Benjamin, H. W., and Kleckner, N. (1990) *Cell,* in press

Haniford, D. B., Chelouche, A. R., and Kleckner, N. (1989) *Cell* 59, 385-394

Hauer, B., and Shapiro, J. (1984) *Mol. Gen. Genet.* 194, 149-158

Hedges, R., and Jacob, A. (1974) *Mol. Gen. Genet.* 132, 31-40

Hehl, R., and Baker, B. (1989) *Mol. Gen. Genet.* 217, 53-59

Harshey, R. M. (1983) *Proc. Natl. Acad Sci. U. S. A.* 80, 2012-2016

Hedges, R., and Datta, N. (1972) *J. Gen. Microbiol.* 73, 573-575

Huang, C-J., Heffron, F., Twu, J-S., Schloemer, R. H., and Lee, C-H. (1986) *Gene* 41, 23-31

Huisman, O., Errada, P. R., Signon, L., and Kleckner, N. (1989) *EMBO J.* 8, 2102-2109

Ichikawa, H., Ikeda, K., Amemura, J., and Ohtsubo, E. (1990) *Gene* 86, 11-17

Ichikawa, H., Ikeda, K., Wishart, W. L., (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 8220-8224

Isberg, R., Lazaar, A., and Syvanen, M. (1982) *Cell* 30, 883-892

Johnson, M. S., McClure, M. A., Feng, D-F., Gray, J., and Doolittle, R. F. (1986) *Proc. Natl. Acad. Sci. U. S. A.* 83, 7648-7652

Johnson, R.C., and Reznikoff, W.S. (1984) *J. Mol. Biol.* 177, 645-664

Jordan, E., Saedler, H., Starlinger, P. (1967) *Mol. Gen. Genet.* 100, 296-306

Kans, J. A., and Casadaban, M. J. (1989) *J. Bacteriol.* 171, 1904-1914

Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) *Cell* 63, 87-95

Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) *J. Virol.* 63, 5319-5327

Kaufman, P. D., Doll, R. F., and Rio, D. C. (1989) 59, 359-371

Kleckner, N. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds., American Society for Microbiology, Washington D. C., pp. 227-268

Kopecko, D. J., Brevet, J., and Cohen, S. N. (1976) *J. Mol. Biol.* 108, 333-360

Krause, H. M., and Higgins, N.P. (1986) *J. Biol. Chem.* 261, 3744-3752

Krebs, P. M., and Reznikoff, W. S. (1986) *J. Mol. Biol.* 192, 781-791

Kubo, K., and Craig, N.L. (1990) *J.Bacteriol.* 5, 2774-2778

Kunze, R., and Starlinger, P. (1989) *EMBO J.* 8, 3177-3185

Lavoie, B. D., and Chaconas, G. (1990) *J. Biol. Chem.* 265, 1623-1627

Lee, C-H., Bhagwat, A., and Heffron, F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* 80,6765-6769

Leung, P. C., Teplow, D. B., Harshey, R. M. (1989) *Nature (London)* 338, 656-658

Lichtenstein, C., and Brenner, S. (1981) *Mol. Gen. Genet.* 183, 380-387

Lichtenstein, C., and Brenner, S. (1982) *Nature (London)* 297, 601-603

Machida, C., and Machida, Y. (1989) *J. Mol. Biol.* 208, 567-574

Mahillon, J., and Lereclus, D. (1988) *EMBO J.* 7, 1515-1526

Makris, J. C., Nordmann, P. L., and Reznikoff, W. S. (1988) *Proc. Natl. Acad Sci. U. S. A.* 85, 2224-2228

Makris, J. C., Nordmann, P. L., and Reznikoff, W. S. (1990) *J. Bacteriol.* 172, 1368-1373

Malamy, M. H. (1966) *Cold Spring Harbor Symp. Quant. Biol.* 31, 189-201

Masson, P., Rutherford, G., Banks, J. A., and Fedoroff, N. (1989) *Cell* 58 755-765

Maxwell, A., Craigie, R., and Mizuuchi, K. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 699-703

McClintock, B. (1951) *Cold Spring Harbor Symp. Quant. Biol.* 16, 223-257

Misra, T. G., Grandgenett, D. P., and Parson, J. T. (1982) *J. Virol.* 44, 330-343

Mizuuchi, K. (1983) *Cell* 35,785-794

Mizuuchi, K. (1984) *Cell* 39, 395-404

Mizuuchi, K., and Craigie, R. (1986) *Ann. Rev. Genet.* 20, 385-429

Mizuuchi, M., and Mizuuchi, K. (1989) *Cell* 58, 399-408

McKown, R. L., Orle, K. A., Chen, T., and Craig, N. L. (1988) *J. Bacteriol.* 170, 352-358

McKown, R. L., Waddell, C. S., Arciszewska, L. K., and Craig, N. L. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 7807-7811

Misra, S., and Rio, D. C. (1990) *Cell* 62, 269-284

Morisato, D., and Kleckner, N. (1987) *Cell* 51, 101-111

Mullins, M. C., Rio, D. C., and Rubin, G. M. (1989) *Genes Develop.* 3, 729-738

New, J. H., Eggleston, A. K., and Fennewald, M. (1988) *J. Mol. Biol.* 201, 589-599

O'Hare, K., and Rubin, G. M. (1983) *Cell* 34, 25-35

Orle, K. A., and Craig, N. L. (1990) *Gene* in press

Ouartsi, A., Borowski, D., and Brevet, J. (1985) *Mol. Gen. Genet.* 198, 221-227

Panganiban, A. T., and Temin, H. M. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 7885-7889

Pereira, A., Cuypers, H., Gierl, A., Schwarz-Sommer, Z., and Saedler, H. (1986) *EMBO J.* 5, 835-841

Pato, M. (1989) in *Mobile DNA* Berg, D. E. and Howe M. M. eds. American Society for Microbiology, Washington D. C. pp. 23-52

Phadnis, S. H., and Berg, D. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 9118-9122

Prentki, P., Chandler, M., and Galas, D. J. (1987) *EMBO J.* 6, 2479-2487

Ptashne, M. (1988) *Nature (London)* 335, 683-689

Reyes, O., Beyou, A., Mignotte-Vieux, C., and Richaud, F. (1987) *Plasmid* 18, 183-192

Rio, D. C., Laski, F. A., and Rubin, G. M. (1986) *Cell* 44, 21-32

Rio, D. C., and Rubin, G. M. (1988) *Proc. Natl. Acad. Sci. U. S. A.* 85, 8929-8933

Roberts, D., Hoopes, B. C., McClure, W. R., and Kleckner, N. (1985) *Cell* 43, 117-130

Robertson, H. M., and Engels, W. R. (1989) *Genetics* 123, 815-824

Robinson, M. K., Bennett, P. M., and Richmond, M. H. (1977) *J. Bacteriol.* 129, 407-414

Roeder, G. S., and Fink, G. R. (1983) in *Mobile Genetic Elements* Shapiro, J. ed., Academic Press, pp. 299-328

Rogers, M., Ekaterinaki, N., Nimmo, E., and Sherratt, D. J. (1986) *Mol. Gen. Genet.* 205, 550-556

Roth, M. J., Schwartzberg, P. L., and Goff, S. P. (1989) *Cell* 58, 47-54

Rubin, G. M., and Spradling, A. C. (1982) *Science* 218, 348-353

Schwartzberg, P., Colicelli, J., and Goff, S. P. (1984) *Cell* 37, 1043-1052

Schwarz-Sommer, Z., Gierl, A., Berntgen, R., Saedler, H. (1985) *EMBO J.* 4, 2439-2443

Sekine, Y., and Ohtsubo, E. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 4609-4613

Shapiro, J. A. (1979) *Proc. Natl. Acad. Sci. U. S. A.* 76, 1933-1937

Sherratt, D. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds. American Society for Microbiology, Washington D. C., pp. 163-184

Smith, G. M., and Jones, P. (1984) *J. Bacteriol.* 157, 962-964

Spradling, A. C., and Rubin, G. M. (1982) *Science* 218, 341-347

Stark, W. M., Boocock, M. R., and Sherratt, D. J. (1989) *Trends Genet.* 5, 304-309

Surette, M. G., Buch, S. J.,and Chaconas, G. (1987) *Cell* 49, 253-262

Surette, M. G., and Chaconas, G. (1989) *J. Biol. Chem.* 264, 3028-3034

Surette, M. G., Lavoie, B. D., and Chaconas, G. (1989) *EMBO J.* 8, 3483-3489

Syvanen, M. (1988) Bacterial Insertion Sequences. In *Genetic Recombination* Kucherlapti, R., and Smith, G. eds. American Society for Microbiology, Washington D. C., pp. 331-356

Thompson, J. F. and Landy, A. (1989) in *Mobile DNA* Berg, D. E. and Howe, M. M. eds. American Society for Microbiology, Washington D. C., pp. 1-22

van Drunen, C. M., van Zuylen, C., Mientjes, E., Goosen, N., and van de Putte, P. (1990) EMBO Workshop on Molecular Mechanisms and Control of Transposition, Roscoff, France

van Rijn, P. A., Goosen, N., Turk, S. C. H. J., and van de Putte, P. (1989) *Nucl. Acids Res.* 24, 10203-10212

Varmus, H., and Brown, P. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds. American Society for Microbiology, Washington D. C., pp. 53-108

Waddell, C. S., and Craig, N. L. (1988) *Genes Develop.* 2, 137-149

Wang, J. C., and Giaever, G. N. (1988) *Science* 240, 300-304

Way, J. C., and Kleckner, N. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 3452-3456

Weinert, T. A., Derbyshire, K. M., Hughson, F. M., and Grindley N. D. F. (1984) *Cold Spring Harbor Symp. Quant. Biol.* 49, 251-260

Wiater, L. C. (1990) Ph.D. Thesis, Yale University, New Heaven

Wiater, L. A., and Grindley, N. D. F. (1988) *EMBO J.* 7, 1907-1911

Wiater, L. C., and Grindley, N. D. F. (1990) *J. Bacteriol.* 4951-4958

Wiater, L. C., and Grindley, N. D. F. (1990) *J. Bacteriol.* 4959-4963

Yin, J. C. P., Krebs, M. P., and Reznikoff, W. S. (1988) *J. Mol. Biol.* 199, 35-45

Yin, J. C. P., and Reznikoff, W. S. (1987) *J. Bacteriol.* 169, 4637-4645

Zerbib, D., Jakowec, M., Prentki, P., Galas, D. J., and Chandler, M. (1987) *EMBO J.* 6, 3163-3169

Zerbib, D., Polard, P., Escoubas, J. M., Galas, D., and Chandler, M. (1990) *Mol. Microbiol.* 4, 471-477

Zerbib, D., Prentki, P., Gamas, P., Freund, E., Galas, D. J., and Chandler, M. (1990) *Mol. Microbiol.* 4, 1477-1486

CHAPTER II

Transposon Tn7; *cis*-acting sequences in transposition and transposition immunity

J. Mol. Biol. (1989) 207, 35–52

Transposon Tn7

cis-Acting Sequences in Transposition and Transposition Immunity

Lidia K. Arciszewska[1,3], Daniela Drake[1,3] and Nancy L. Craig[1,2,3]

[1]*Department of Microbiology and Immunology*
[2]*Department of Biochemistry and Biophysics*
[3]*The G. W. Hooper Foundation*
University of California, San Francisco
San Francisco, CA 94143, U.S.A.

(Received 10 June 1988, and in revised form 18 October 1988)

We have identified and characterized the *cis*-acting sequences at the termini of the bacterial transposon Tn*7* that are necessary for its transposition. Tn*7* participates in two kinds of transposition event: high-frequency transposition to a specific target site (*attTn7*) and low-frequency transposition to apparently random target sites. Our analyses suggest that the same sequences at the Tn*7* ends are required for both transposition events. These sequences differ in length and nucleotide structure: about 150 base-pairs at the left end (Tn*7*L) and about 70 base-pairs at the right end (Tn*7*R) are necessary for efficient transposition. We also show that the ends of Tn*7* are functionally distinct: a miniTn*7* element containing two Tn*7*R ends is active in transposition but an element containing two Tn*7*L ends is not. We also report that the presence of Tn*7*'s *cis*-acting transposition sequences anywhere in a target replicon inhibits subsequent insertion of another copy of Tn*7* into either an *attTn7* target site or into random target sites. The inhibition to an *attTn7* target site is most pronounced when the Tn*7* ends are immediately adjacent to *attTn7*. We also show that the presence of Tn*7*R's *cis*-acting transposition sequences in a target replicon is necessary and sufficient to inhibit subsequent Tn*7* insertion into the target replicon.

1. Introduction

Transposable elements are discrete DNA segments that can move from one genetic location to another. The bacterial transposon Tn*7* (Fig. 1(a); Barth *et al.*, 1976), which carries resistance to trimethoprim and to streptomycin and spectinomycin, can participate in two types of transposition event: high-frequency insertion into specific target sites in the chromosomes of many different bacteria and low-frequency insertion into many other random target sites (for a review, see Craig, 1989). The specific site of Tn*7* insertion into the *Escherichia coli* chromosome, *attTn7*, lies at about minute 84 between *phoS*, a gene involved in phosphate transport, and *glmS*, a gene involved in cell wall biosynthesis (Barth *et al.*, 1976; Lichtenstein & Brenner 1981, 1982; Walker *et al.*, 1984). Tn*7* inserts into *attTn7* in a unique orientation (Lichtenstein & Brenner, 1981). When Tn*7* occupies *attTn7* (*attTn7*::Tn*7*), the left end of Tn*7* (Tn*7*L), which is adjacent to its drug resistance genes, is always proximal to *phoS*, and the right end of Tn*7* (Tn*7*R), which is adjacent to its transposition genes, is always proximal to *glmS*. In addition to site and orientation-specific transposition to *attTn7*, Tn*7* also transposes at low frequency to many different plasmid and chromosomal sites (Barth & Grinter, 1977; Barth *et al.*, 1978; Moore & Krishnapillai, 1982; K. Kubo & N. Craig, unpublished results).

Transposable elements carry two kinds of information to promote their transposition: (1) genes encoding transposition proteins and (2) DNA sites at the element termini that participate directly in transposition (for reviews, see Grindley & Reed, 1985; Craig & Kleckner, 1987). Tn*7* contains five transposition genes: *tnsA*, *tnsB*, *tnsC*, *tnsD* and *tnsE* (Fig. 1(a); Rogers *et al.*, 1986; Waddell & Craig, 1988). These genes provide two distinct but overlapping transposition pathways (Rogers *et al.*, 1986; Waddell & Craig, 1988; K. Kubo & N. Craig, unpublished results). *tnsABC+tnsD* promote high-frequency transposition to *attTn7* and, when *attTn7* is unavailable, low-frequency transposition to other sites structurally related to *attTn7*. *tnsABC + tnsE*

0022-2836/89/090035-18 $03.00/0

promote low-frequency transposition to many different sites unrelated to *attTn7* and to each other. Thus, Tn*7* engages in transposition to both specific target sites and to random target sites.

Previous studies have shown that *cis*-acting transposition sequences lie at the ends of Tn*7* (Hauer & Shapiro, 1984; Smith & Jones, 1984; Ouartsi *et al.*, 1985; Barry, 1986; Rogers *et al.*, 1986). In this report, we define and characterize the *cis*-acting transposition sequences at the Tn*7* ends. We show that about 150 bp† at the terminus of Tn*7*L and at least 70 bp at the terminus of Tn*7*R participate directly in transposition. Thus, considerable sequence information at each end of Tn*7* beyond this element's terminal 30 bp inverted repeats (Lichtenstein & Brenner, 1982; Gosti-Testu & Brevet, 1982) is required for transposition. We also show that the ends of Tn*7* are functionally different.

The *cis*-acting transposition sequences of the transposons Tn*3* and bacteriophage Mu have been shown to provide another interesting activity, transposition immunity (Lee *et al.*, 1983; Reyes *et al.*, 1987; Adzuma & Mizuuchi, 1988). Transposition immunity is the ability of a transposable element present in a target DNA molecule to reduce substantially the frequency of subsequent insertion of another copy of the element into this target molecule. The same sequences that participate directly in transposition are necessary and sufficient to provide transposition immunity (Lee *et al.*, 1983; Huang *et al.*, 1986; Adzuma & Mizuuchi, 1988). Thus transposition immunity is fundamentally related to the process of transposition itself. The phenomenon of transposition immunity is an interesting example of "action at a distance", as short DNA segments containing the transposon ends can render large target molecules approaching 100 kb in size immune to further transposon insertion anywhere within the target molecule.

Hauer & Shapiro (1984) have shown that the presence of Tn*7*, or large DNA segments containing the Tn*7* ends, in a target plasmid reduces the frequency of subsequent Tn*7* insertion into these targets, i.e. that Tn*7* displays transposition immunity. In this report, we characterize Tn*7* transposition immunity and show that it is active in both the *tnsABC+tnsD* and the *tnsABC+tnsE* transposition pathways. We also present evidence that the sequences in Tn*7*R that are necessary for efficient Tn*7* transposition can provide transposition immunity to a target DNA molecule.

2. Materials and Methods

(a) *Media, chemicals and enzymes*

LB broth and agar were as described by Miller (1972) except that 1 mg glucosamine was added/ml agar. "Iso-sensitest Agar" (Oxoid) was used with trimethoprim. Supplements were: carbenicillin (100 μg/ml), chloramphenicol (30 μg/ml), gentamicin (10 μg/ml), kanamycin (50 or 100 μg/ml), rifampicin (100 μg/ml), tetracycline (5 or 20 μg/ml) and trimethoprim (100 μg/ml). DNA-modifying enzymes were obtained from commercial sources and used as recommended by the manufacturer.

(b) *Bacterial strains*

NLC51 is *E. coli* F⁻ *araD*139 Δ(*argF-lac*) U169 *rpsL*150 *relA*1 *flbB*5301 *deoC*1 *ptsF*25 *rbsR val*^{R} *recA*56. LA3 is NLC51 *attTn7*::Tn*7* (McKown *et al.*, 1987). The Tn*7* derivative used in this work, Tn*7S*, contains an IS*1* element inserted near its antibiotic resistance determinants (Fig. 1(a)); the transposition properties of Tn*7S* are indistinguishable from those of the canonical Tn*7* (Hauer & Shapiro, 1984). CW51 is *E. coli* F⁻ *ara*⁻ *arg*⁻ *lac pro*XIII *recA*56 Nal^R Rif^R (Waddell & Craig, 1988).

(c) *Manipulation and analysis of DNA*

Plasmid growth, isolation, transformation and restriction enzyme analysis were performed as described by Maniatis *et al.* (1982). Cloning procedures were as described by Maniatis *et al.* (1982) except that DNA fragments contained in slices excised from low-melting temperature agarose gels (Sea Plaque) were used directly in assembly of recombinant molecules as described by Struhl (1983). Phage lambda growth was performed as described by Maniatis *et al.* (1982).

(d) *Plasmids*

(i) attTn7 *plasmids*

In *attTn7*, the middle base of the 5 bp chromosomal sequence duplicated upon Tn*7* insertion is designated 0, sequences extending leftward toward *phoS* are given "−" numbers and sequences extending rightward toward *glmS* are given "+" numbers. pRM2 is a Cb^R pUC18 derivative containing *attTn7* −342 to +165. pKAO3 is a Cb^R pUC18 derivative containing *attTn7* −52 to +64 (McKown *et al.*, 1988).

(ii) tns *plasmids*

pCW4, pCW4::miniMu78 and pCW4::miniMu72 are Tc^R pACYC184 derivatives containing the *tns* genes (Waddell & Craig, 1988). pCW4 provides *tnsABCDE*, pCW4::miniMu78 provides functional *tnsABC+tnsD* and pCW4::miniMu72 provides functional *tnsABC+tnsE*.

(iii) *Plasmids containing miniTn*7L1900 × R537 *and miniTn*7L166 × R199

pRM2::miniTn*7L1900* × *R537* was constructed by *in vivo* transposition of the Cm^R miniTn*7* derivative Tn*7S*::Tn*9* Δ *Pst*I (Hauer & Shapiro, 1984) into *attTn7* of pRM2 (Fig. 1(b); McKown *et al.*, 1988).

pLA1 contains *attTn7*::miniTn*7L166* × *R199*; this miniTn*7* element contains a Km^R marker (Fig. 1(c); McKown *et al.*, 1987). It was obtained from pRM2::miniTn*7L1900* × *R537* by digestion with *Hpa*I and *Bal*I which cut in Tn*7*L and Tn*7*R, respectively (Lichtenstein & Brenner, 1982), ligation in the presence of *Sal*I linker (Pharmacia) and subsequent introduction of a fragment with terminal *Sal*I sites encoding Km^R from pUC-4K (Pharmacia).

pKAO3::miniTn*7L166* × *R199* contains *attTn7*::miniTn*7L166* × *R199* (Fig. 1(d); McKown *et al.*, 1988).

† Abbreviations used: bp, base-pairs; kb, 10^3 base-pairs; Cb, carbenicillin; Cm, chloramphenicol; Gen, gentamicin; Km, kanamycin; Rif, rifampicin; Tc, tetracycline; Tp, trimethoprim; nt, not tested; s.e.m., standard error of the mean.

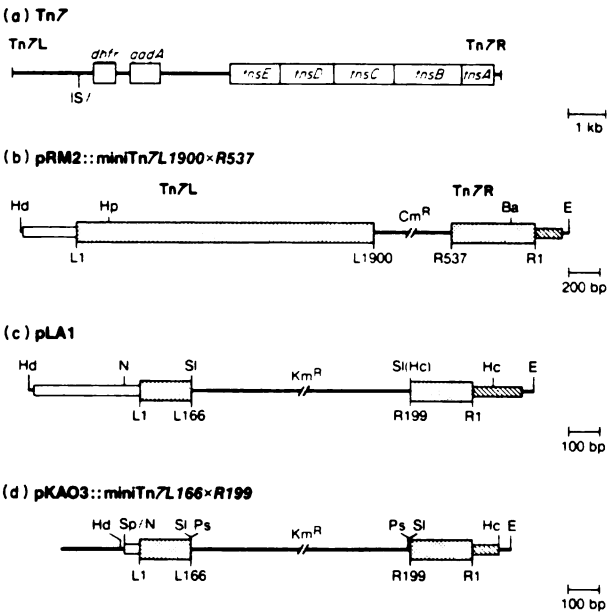


Figure 1. Structure of Tn*7* and plasmids containing miniTn*7* elements. (a) Tn*7*. *dhfr* encodes a dihydrofolate reductase (Fling & Richards, 1983; Simonsen *et al.*, 1983) that provides resistance to trimethoprim. *aadA* encodes a 3″(9)-*O*-nucleotidyltransferase (Fling *et al.*, 1985) that provides resistance to streptomycin and spectinomycin. and *tnsA*, *tnsB*, *tnsC*, *tnsD*, *tnsE* are Tn*7*'s transposition genes (Rogers *et al.* 1986; Waddell & Craig, 1988). The left end of Tn*7* (Tn*7*L) and the right end of Tn*7* (Tn*7*R) are indicated. The position of IS*1* in Tn*7*S is indicated (Hauer & Shapiro, 1984). In (b) to (d), the segments shown are contained within the polylinkers of the indicated plasmids. The Tn*7*L and Tn*7*R end segments are shown by stippled boxes and their boundaries are numbered. The Tn*7* end segments are flanked by the sequences that flank the Tn*7* ends in chromosomal *attTn7*. The chromosomal sequences which flank L1 and extend leftward towards *phoS* are indicated by an open box and the chromosomal sequences which flank R1 and extend rightward are indicated by a hatched box. The nucleotide sequence information required for *attTn7* target activity lies within the hatched boxes (Gringauz *et al.*, 1988). Cm^R and Km^R designate segments within the miniTn*7* elements that encode chloramphenicol resistance and kanamycin resistance. Restriction enzyme sites relevant to this work are shown. Ba, *Bal*I; E, *Eco*RI; Hc, *Hin*cII; Hd, *Hin*dIII; Hp, *Hpa*I; N, *Nla*III; Ps, *Pst*I; Sl, *Sal*I; and Sp, *Sph*I. (b) pRM2::miniTn*7L1900* × *R537*. (c) pLA1 (pRM2::miniTn*7L166* × *R199*). (d) pKAO3::miniTn*7L166* × *R199*.

pEM-1 contains miniTn*7L166* × *R199* flanked by *E. coli* chromosomal sequences, about 1600 bp flanking Tn*7*L and about 400 bp flanking Tn*7*R. It was obtained by *tnsABC*+*tnsE*-dependent transposition of miniTn*7L166* × *R199* to the *E. coli* chromosome at site EC-18 (*tnsE* chromosomal) and subsequent cloning of the *Eco*RV chromosomal fragment containing the miniTn*7* element into the *Hin*cII site of pUC19 (K. Kubo, R. Bainton & N. Craig, unpublished results).

(iv) *Plasmids containing deleted Tn7L end segments used in the assembly of miniTn7 elements*

Plasmids containing Tn*7*L end segments are shown in Fig. 2(a). pLA26 contains Tn*7*L166 flanked at its L1 terminus by *E. coli* chromosomal sequences that lie to the left (towards *phoS*) of the specific point of Tn*7* insertion in *attTn7*. It was obtained by treating the Tn*7*L-containing *Hin*dIII–*Pst*I fragment from pKAO3::miniTn*7L166* × *R199* with Klenow fragment and inserting it into the *Sma*I site of pUC18 with chromosomal sequences adjacent to vector p*lac*.

pLA30 (L149), pLA32 (L94), pLA57 (L75) and pLA34 (L23) contain shorter Tn*7*L segments obtained by restriction enzyme digestion. The Tn*7*L-containing *Pru*II fragment from pLA26 was isolated by electroelution, digested with *Mae*I, *Hga*I, *Mae*III or *Mnl*I and treated with Klenow fragment when necessary. The *Mae*I–*Pru*II, *Hga*I–*Pru*II and *Mnl*I–*Pru*II fragments containing Tn*7*L terminal sequences were introduced into the *Sma*I site and the *Mae*III–*Pru*II fragment was introduced into the *Hin*cII site of pUC18. In all cases, the chromosomal sequences flanking the Tn*7*L segment are adjacent to vector p*lac*.

pLA40 (L108) was constructed by isolation by electoelution of the Tn*7*L-containing *Hin*dIII–*Pst*I fragment from pKAO3::miniTn*7L166* × *R199* and digestion of this fragment with *Alu*I. The resulting *Alu*I–*Hin*dIII fragment containing Tn*7*L was inserted into *Hin*dIII–*Hin*cII site of pUC18.

(v) *Plasmids containing deleted Tn7R end segments used in the construction of miniTn7 elements*

Plasmids containing Tn*7*R segments are shown in Fig. 2(b). pLA28 contains Tn*7*R199 flanked at its R1

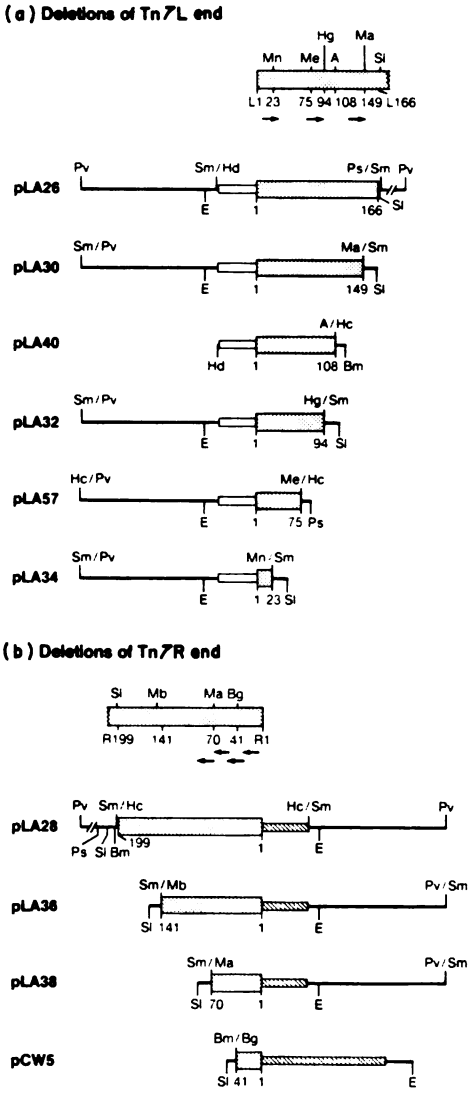


Figure 2. Deletion variants of Tn*7* ends. (a) Tn*7*L ends. (b) Tn*7*R ends. The Tn*7* end segments are shown at the top of each panel by stippled boxes and the end sequences are numbered. The positions of restriction enzyme sites used in the construction of deletion variants are indicated and the positions of the 22 bp sequence repeated within each Tn*7* end is shown (arrows; see the text). The bottom sections of each panel show portions of plasmids which contain Tn*7* end deletion variants; their construction is described in Materials and Methods. The Tn*7* end segments are indicated by stippled boxes, the chromosomal sequences that lie to the left of the specific insertion point in *attTn7* by open boxes, chromosomal sequences that lie to the right of the specific insertion point in *attTn7* by hatched boxes and vector sequences by lines. Restriction sites used in the construction of these plasmids are indicated on the top of the plasmid segments and restriction enzyme sites used in the assembly of miniTn*7* elements from these segments on the bottom of the plasmid segments as described in Materials and Methods. Restriction enzyme sites: A, *Alu*I; Bg, *Bgl*II; Bm, *Bam*HI; E, *Eco*RI; Hc, *Hin*cII; Hd, *Hin*dIII; Hg, *Hga*I; Ma, *Mae*I; Mb, *Mbo*II; Me, *Mae*III; Mn, *Mnl*I; Ps, *Pst*I; Pv, *Pru*II; Sl, *Sal*I; and Sm, *Sma*I.

terminus by *E. coli* chromosomal sequences that lie to the right (towards *glmS*) of the specific insertion point in *attTn7* (Tn7R/*attTn7*+64). It was obtained by insertion of the Tn*7*R-containing *Hin*cII fragment from pLA1 into the *Sma*I site of pUC18 with *attTn7* sequences adjacent to vector *plac*.

pLA36 (R141) and pLA38 (R70) contain shorter Tn*7*R segments obtained by restriction enzyme digestion. The Tn*7*R-containing *Pru*II fragment from pLA28 was isolated by electroelution, digested with *Mbo*II or *Mae*I, the Tn7R/*attTn7* fragments were treated with Klenow fragment and inserted into *Sma*I site of pUC18 with *attTn7* sequences adjacent to vector *plac*.

pCW5 contains Tn7R41/*attTn7*+165. It was constructed by isolation of the Tn*7*R-containing *Bgl*II *Eco*RI fragment from pRM2::Tn*7* and insertion of this fragment into the *Bam*HI-*Eco*RI site of pUC18 (C. Waddell & N. Craig, unpublished results).

(vi) *Construction of miniTn7 element plasmids*

All miniTn*7* elements contain a Km^R gene, except miniTn*7L1900* × *R537*, which contains a Cm^R gene. Plasmids containing miniTn*7* elements are listed in Table 1. pRM2::miniTn*7L1900* × *R537* and pLA1 are described above. All other plasmids containing miniTn*7* elements, except pDD3 and pLA5, were assembled by simultaneous ligation of fragments containing: (1) Tn*7* end segments (either a Tn*7*L and a Tn*7*R segment, 2 Tn*7*L segments or 2 Tn*7*R segments), (2) a Km^R segment (from pUC-4K) and (3) a vector backbone segment (either pUC18 or pACYC184). Tn*7*L166 fragments were obtained from either pLA26 or pLA1; fragments containing shorter Tn*7*L end segments were obtained from the plasmids shown in Fig. 2(a). Tn*7*R199 fragments were obtained from either pLA28 or pLA1; fragments containing shorter Tn*7*R segments were obtained from the plasmids shown in Fig. 2(b). The Tn*7* end restriction fragments used were dictated by the Km^R and vector fragments used. The Km^R fragment used had terminal *Sal*I sites except for pLA47, pLA66 and pLA5, where it had terminal *Bam*HI sites and pLA59, where it had terminal *Pst*I sites. The vector cloning sites used are shown in Table 1.

Several examples of miniTn*7* element plasmid constructions are given below. pLA48, which contains miniTn*7L94* × *R199*, was constructed by simultaneous ligation of the *Eco*RI-*Sal*I Tn*7*L94 fragment from pLA32, the *Sal*I-*Sal*I Km^R fragment from pUC-4K, the *Sal*I-*Eco*RI Tn*7*R199 fragment from pLA1 and *Eco*RI digested pUC18 and selection for Km^R. pLA47, which contains miniTn*7L108* × *R199*, was constructed by simultaneous ligation of the *Hin*dIII-*Bam*HI Tn*7*L108 fragment of pLA40, the *Bam*HI-*Bam*HI Km^R fragment from pUC-4K, the *Bam*HI-*Eco*RI fragment from pLA28 and *Hin*dIII-*Eco*RI-digested pUC18 and selection for Km^R. The orientations of the miniTn*7* elements and the

Table 1
Plasmids carrying miniTn7 elements

Plasmid†	miniTn7 elements and flanking sequences‡	Vector and cloning sites used§
pRM2 mini-Tn7*L1900* × *R537*	−342 *L1900* × *R537 att* + 165	pUC18
pLA1	−342 *L166* × *R199 att* + 165	pUC18
pLA43	−52 *L149* × *R199 att* + 165	pUC18-*Eco*RI
pLA47	−52 *L108* × *R199 att* + 64	pUC18-*Hin*dIII.*Eco*RI
pLA48	−52 *L94* × *R199 att* + 165	pUC18-*Eco*RI
pLA59	−52 *L75* × *R199 att* + 64	pUC18-*Eco*RI
pLA45	−52 *L23* × *R199 att* + 165	pUC18-*Eco*RI
pLA49	−342 *L166* × *R141 att* + 64	pUC18-*Hin*dIII.*Eco*RI
pLA50	−342 *L166* × *R70 att* + 64	pUC18-*Hin*dIII.*Eco*RI
pLA13	−342 *L166* × *R41 att* + 165	pUC18-*Hin*dIII.*Eco*RI
pLA66	*L166F* −342 × *R199 att* + 64	pUC18-*Eco*RI
pDD1	*att* + 165 *R199* × *R199 att* + 165	pACYC184-*Eco*RI
pDD2	−342 *L166* × *L166* −342	pACYC184-*Eco*RI
pDD3	−342 *R199* × *R199 att* + 165	pUC18
pLA55	*att* + 64 *R70* × *R70 att* + 64	pUC18-*Eco*RI
pLD1	*att* + 165 *R41* × *R199 att* + 165	pACYC184-*Eco*RI
pLA5	−342 *R41* × *R41 att* + 165	pUC18

† The plasmids were constructed as described in Materials and Methods.

‡ The miniTn7 elements contain Tn7L and Tn7R segments as indicated and a drug resistance segment. miniTn7*L1900* × *R537* provides Cm^R and all other miniTn7s provide Km^R. The miniTn7 elements are flanked by sequences that surround the specific point of Tn7 insertion in the *E. coli* chromosome. The numbering system for this region is described in Materials and Methods. −342 L or −52 L indicates that the terminus of the Tn7L segment at L1 is flanked by either 342 nucleotides or 52 nucleotides from the *E. coli* chromosome that lie between the point of Tn7 insertion and *phoS*. *R att* + 165 or *R att* + 64 indicates that the terminus of the Tn7R segment at R1 is flanked by sequences that extend from the point of insertion towards *glmS*. The designation *att* is used for these sequences because they contain the sequence information necessary for *attTn7* target activity (Gringauz *et al.*, 1988).

§ As described in Materials and Methods, the indicated restriction sites in the vector backbone were used in the assembly of miniTn7 element plasmids. When no site is indicated, plasmid is recovered in *in vivo* transposition into *attTn7*-containing plasmid.

Km^R fragment within these plasmids are not known for all plasmids.

pDD3 was constructed by *in vivo* transposition of miniTn7*R199* × *R199* into *attTn7* of pRM2. pLA5 was constructed by digestion of pDD3 with *Bgl*II and introduction of a Km^R segment with terminal *Bam*HI sites from pUC-4K.

(vii) *Plasmids containing single Tn7-end segments used in trans-inhibition experiments*

Plasmids containing single Tn7-end segments that were used in the *trans*-inhibition experiments are shown in Fig. 3.

pLA19, which contains Tn7L166, was constructed by isolating the *Hin*dIII-*Sal*I Tn7166 fragment from pLA1, treating it with Klenow fragment, and inserting it into the *Dra*II site of pUC18, which had been treated with Klenow fragment, with Tn7L sequences adjacent to vector *plac*.

pLA17, which contains Tn7R199/*attTn7* + 165, was constructed by isolating the *Eco*RI-*Sal*I Tn7R fragment from pLA1, treating it with Klenow fragment and inserting it into the *Dra*II site of pUC18, with Tn7 sequences adjacent to vector *plac*.

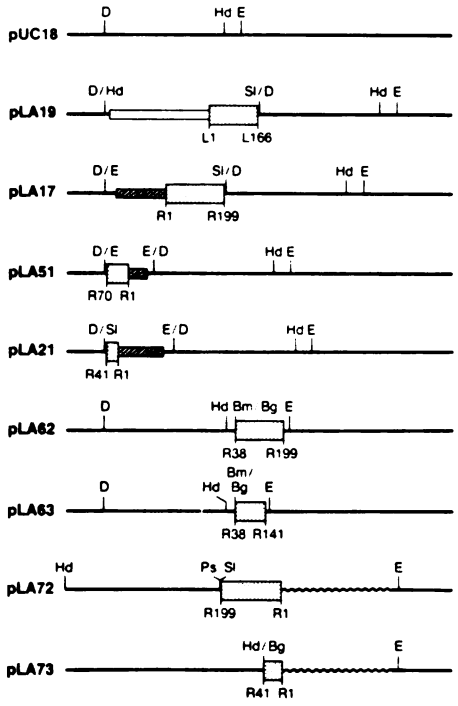


Figure 3. Plasmids containing Tn7 end segments. The construction of these plasmids, which were used in the *trans* inhibition experiments, is described in Materials and Methods. The stippled boxes represent the Tn7 end segments, the open boxes represent *E. coli* sequences that flank Tn7L in chromosomal *attTn7*, the hatched boxes represent chromosomal sequences that flank Tn7R in chromosomal *attTn7*, the wavy lines represent chromosomal sequences that flank Tn7R in site EC-18 and the straight lines represent vector sequences. The nucleotide positions at the boundaries of Tn7L and Tn7R segments are indicated. Restriction enzyme sites: Bg, *Bgl*II; Bm, *Bam*HI; D, *Dra*II; E, *Eco*RI;Hd, *Hin*dIII; Ps, *Pst*I; and Sl, *Sal*I.

pLA36, which contains Tn7R141/*attTn7* + 64, was constructed as described above in section (d)(v) and is shown in Fig. 2(b).

pLA51, which contains Tn7R70/*attTn7* + 64, was constructed by isolating the *Eco*RI Tn7R fragment from a plasmid pLA39, which is similar to pLA38, except that Tn7R70/*attTn7* + 64 fragment was inserted into the vector in opposite orientation, treating it with Klenow fragment and inserting it into the *Dra*II site of pUC18 with *attTn7* sequences adjacent to vector *plac*.

pLA21, which contains Tn7R41/*attTn7* + 165, was constructed by isolating the *Sal*I-*Eco*RI Tn7R fragment from pCW5, treating it with Klenow fragment and inserting it into the *Dra*II site of pUC18 with *attTn7* sequences adjacent to vector *plac*.

pLA62, which contains Tn7R sequences from positions R38 to position R199, is a derivative of pLA29. pLA29 is

similar to pLA28 except that the Tn7R199/*attTn7* + 64 fragment was inserted into the vector in opposite orientation. pLA62 was constructed by digestion of pLA29 with *Bam*HI and *Bgl*II and then religating the vector, thus removing the terminal Tn7R and *attTn7* sequences.

pLA63, which contains Tn7R sequences from position R38 to position R141, is a derivative of pLA37. pLA37 is similar to pLA36 except that the Tn7R/*attTn7* segment was inserted into the vector in opposite orientation. pLA63 was constructed by digestion of pLA37 with *Bam*HI and *Bgl*II and vector religation, thus removing the terminal Tn7R and *attTn7* sequences.

pLA72, which contains Tn7R199 flanked by *E. coli* chromosomal sequences unrelated to *attTn7* (Tn7R199/EC-18), was constructed by digestion of pEM-1 with *Hin*dIII and vector religation, thus removing the Tn7L sequences and part of the Km^R segment.

pLA73, which contains Tn7R41 flanked by *E. coli* chromosomal sequences unrelated to *attTn7* (Tn7R/EC-18) was constructed by digestion of pEM-1 with *Hin*dII and *Bgl*II, treatment with Klenow fragment and vector religation, thus removing the Tn7 sequences, the Km^R segment and part of the Tn7R sequences.

(viii) *Construction of pOX*-attTn7

NLC51 was lysogenized with λ 1046, which contains the Tn*10* transposase gene under control of a *ptac* promoter (Morisato *et al.*, 1983). The lysogen was then transformed with pRM4(McKown *et al.*, 1988), which contains an *attTn7* −342 to +165 segment within a Tc^R miniTn*10* element incapable of transposition in the absence of exogenous Tn*10* transposase. The F plasmid derivative lacking any known transposable element pOX38-Gen (Johnson & Reznikoff, 1984), referred here to as pOX, was then transferred into the lysogen by conjugation. pOX derivatives containing the *attTn7* miniTn*10* element resulting from Tn*10* transposition were obtained by conjugation of pOX and selection for Tc^R exconjugants. Southern analysis of transposition products into this plasmid confirmed the presence of a single *attTn7* sequence in this plasmid (Waddell & Craig, 1988).

(ix) *Construction of target plasmids containing miniTn7 elements used in transposition immunity and* attTn7 *inactivation experiments*

pOX derivatives containing various miniTn7 elements were obtained by transposition of the miniTn7 elements from donor plasmids in hosts containing chromosomal Tn7 and mating out of the episome.

pOX-*attTn7*::EP-18::miniTn7*L166* × *R199* (*tnsE* plasmid), pOX-*attTn7*::EP-19::miniTn7*L166* × *R199* and pOX-*attTn7*::EP-20::miniTn7*L166* × *R199* contain miniTn7*L166* × *R199* at positions other than *attTn7* (Fig. 3, lines 9, 12, 11 of Waddell & Craig, 1988). They were obtained by transposition of miniTn7*L166* × *R199* from chromosomal *attTn7* to pOX-*attTn7* in a host containing pCW4::miniMu72, which provides *tnsABC* + *tnsE*. Physical analysis of these plasmids showed that the miniTn7 insertions were in restriction fragments other than that containing *attTn7*.

pOX-*attTn7* derivatives containing miniTn7 elements in *attTn7* (pOX-*attTn7*::miniTn7) were obtained by transposition of the miniTn7 elements from donor plasmids in hosts containing chromosomal Tn7 and mating out of the episome. The high frequency of miniTn7 transposition to pOX-*attTn7* indicates that the elements inserted into *attTn7*.

(e) *Introduction of miniTn7 elements into chromosomal* attTn7

miniTn7*L1900* × *R537* and miniTn7*L166* × *R199* insertions into chromosomal *attTn7* were obtained as described (Waddell & Craig, 1988). Other miniTn7 insertions were obtained as follows: the miniTn7s were transposed from donor plasmids in LA3 (NLC51 *attTn7*::Tn7) to an F plasmid derivative temperature-sensitive for replication, F'ts*lac* (Beckwith *et al.*, 1966): F'ts *lac*::miniTn7 products were recovered by conjugation and appropriate selection. Subsequently, F'ts *lac*::mini-Tn7s were introduced by conjugation into an NLC51 derivative containing the *tns* gene plasmid pCW4; the chromosomal *attTn7* site of this strain was vacant. Temperature exclusion of the F'*ts* episome was then carried out by growth at high temperature in liquid media containing kanamycin (miniTn7 marker) and tetracycline (pCW4 marker). Cells lacking the episome and containing chromosomal miniTn7 insertions were identified by plating on Km + Tc + 5-bromo-4-chloro-3-indolyl-β-D-galactoside plates.

(f) *Mating out assay for transposition*

In this assay, a Tn7 element transposes in a donor cell to a conjugable target plasmid and transposition is detected by identification of a Tn7 element-containing transconjugant. All mating out transposition assays were carried out in derivatives of NLC51. The mating out assay was used to measure both *tnsABC* + *tnsD* and *tnsABC* + *tnsE*-dependent transposition; pOX-*attTn7* and pOX were used as target molecules, respectively (Waddell & Craig, 1988). Transposition was measured under 2 different conditions: (1) the transposable elements were located in plasmid donor molecules and the transposition proteins were supplied by chromosomal Tn7 in *attTn7*; or (2) transposable elements were located in chromosomal *attTn7* and transposition proteins were supplied by a *tns* plasmid such as pCW4. CW51 (Rif^R) was used as the recipient strain. Assays were carried out as described in Waddell & Craig (1988). Portions of the donor cultures were also plated on media containing appropriate antibiotics to determine that the donor plasmid was present in the cells used in the mating mixtures.

With both pOX and pOX-*attTn7* targets, the total number of exconjugants was determined by selection on Gen + Rif plates. The number of exconjugants containing Tn7 or miniTn7 elements was determined by selection on Tp + Rif, Cm + Rif or Km + Rif plates, as appropriate. Transposition efficiency was calculated as the ratio of the number of Tn7 element-containing exconjugants to the total number of exconjugants. In some experiments, the mating mixtures were also plated on media containing the appropriate antibiotics to identify exconjugants containing transferred donor plasmids.

(g) *Lambda hop assay for transposition*

In this assay, a miniTn7 element transposes from a replication and integration-defective phage lambda derivative upon infection into cells containing the *tns* genes; transposition is detected by isolation of cells containing the miniTn7 element by selection on plates containing appropriate antibiotics (McKown *et al.*, 1988). Lambda KK1 was used as the donor vehicle for miniTn7*L166* × *R199*, which carries Km^R. Lambda KK1 is lambda 780 *hisG* 9424::Tn*10 del*16 *del*17::*attTn7* −342 to +165::miniTn7*L166* − *KmR* − *R199* (McKown *et*

al., 1988). The recipient cells were derivatives of NLC51 containing various plasmids as described in Table 6. Assays were carried out as described by McKown *et al.* (1988).

(h) *Evaluation of site and orientation-specificity of insertion into* attTn7

To determine site and orientation-specificity of various miniTn*7* elements, we analyzed by restriction enzyme digestion the products of independent insertions of these elements into the *attTn7* plasmid pRM2. miniTn*7* insertions in pRM2 were recovered by 2 different methods. (1) Insertions of miniTn*7L166* × *R199*, miniTn*7L149* × *R199*, miniTn*7L166* × *R70* and miniTn*7R199* × *R199* into pRM2 were recovered by isolation of plasmid DNA from LA3 strains containing pOX::miniTn*7* derivatives and pRM2 and subsequent transformation to another host selecting for Km^R, the marker carried by the miniTn*7* elements. (2) Insertions of miniTn*7R199* × *R199* and miniTn*7R41* × *R199* into pRM2 were recovered by isolation of plasmid DNA from Km^RCb^R cells lacking the F episome, obtained after growth at 42 °C in liquid LB+Km+Cb of NLC51 derivatives containing Tn*7* in chromosomal *attTn7* and F'*ts*::miniTn*7* elements.

3. Results

(a) *Tn7 termini contain the* cis*-acting transposition sequences*

To identify Tn*7*'s *cis*-acting transposition sequences, we analyzed the transposition properties of miniTn*7* elements containing Tn*7*L and Tn*7*R end segments of different lengths. We use the nomenclature miniTn*7L1900* × *R537* to describe a miniTn*7* element in which a fragment of the left end of Tn*7* extending from the terminus L1 to position L1900 and a fragment of the right end of Tn*7* extending from the terminal position R1 to position R537 flank a drug resistance marker (Fig. 1(b)).

Transposition of miniTn*7*s, supported by Tn*7*-encoded transposition proteins supplied *in trans*, to conjugable target plasmids was measured by mating out assays. To evaluate low-frequency, random target site transposition mediated by *tnsABC'* + *tnsE*, we used as a target an F plasmid derivative, pOX38-Gen, hereinafter referred to as pOX (Johnson & Reznikoff, 1984; Waddell & Craig, 1988). To evaluate high-frequency, target site-specific transposition mediated by *tnsABC'* + *tnsD*, we used as a target a pOX derivative carrying Tn*7*'s specific insertion site, *attTn7*, hereafter referred to as pOX-*attTn7* (see Material and Methods; Waddell & Craig, 1988).

We examined transposition of miniTn*7* elements under two different conditions in which the copy number of the miniTn*7* elements, and the copy number of the Tn*7* transposition (*tns*) genes varied. In one condition, miniTn*7* elements in multi-copy, non-mobilizable donor plasmids were supplied with Tns proteins from intact Tn*7* in chromosomal *attTn7*. This situation allowed analysis of both simple insertion and cointegrate transposition products. In the other condition, single-copy miniTn*7*s in chromosomal *attTn7* were supplied with Tns proteins from a *tns* plasmid. This situation minimized differences in sequences flanking the miniTn*7*s and possible differences in both copy number of the miniTn*7* donor plasmids and *tns* gene expression mediated by regulatory sites within the transposon end segments (see below).

Intact Tn*7* transposes to pOX at a low frequency ($1{\cdot}7 \times 10^{-4}$ transposon-containing transconjugants/transconjugant) and at a much higher frequency ($9{\cdot}4 \times 10^{-3}$) to pOX-*attTn7*, reflecting the high efficiency of site-specific insertion in *attTn7*. Work by Hauer & Shapiro (1984) indicated that miniTn*7L1900* × *R537* contains Tn*7*'s essential *cis*-acting transposition sequences. Our analyses of multi-copy (Table 2A) and single-copy (Table 2B) miniTn*7L1900* × *R537* transposition support this view. Under either condition, miniTn*7L1900-R537* transposes at low frequency to pOX and much higher frequency to pOX-*attTn7*. It should be noted that the same total number of transposition events was observed with either multi-copy or single-copy miniTn*7L1900-R537*. Thus, the transposition frequency per miniTn*7* element was much lower with the multi-copy miniTn*7* elements. The failure to observe an increase in total transposition frequency when the number of transposition substrates was increased could mean that the transposition frequency was limited by the concentration of a transposition protein. However, the actual concentrations of transposition proteins are unknown under either the single-copy or multi-copy miniTn*7* conditions.

(b) cis*-acting transposition sequences are contained within the L166 and R199 segments*

We evaluated the transposition properties of miniTn*7L166* × *R199*, an element containing short Tn*7* end segments. We found that this element transposes as efficiently as miniTn*7L1900* × *R537* to both pOX and pOX-*attTn7* (Table 2A,B). These two elements have also been shown to transpose at equal frequency to *attTn7* using a different transposition assay (McKown *et al.*, 1988). Like intact Tn*7*, miniTn*7L166* × *R199* inserts into *attTn7* in site and orientation-specific fashion, as demonstrated by restriction enzyme analysis of a number of independent insertions into the *attTn7* plasmid pRM2 (obtained as described in Materials and Methods; data not shown) and by nucleotide sequences analysis of *attTn7*::miniTn*7L166* × *R199* insertions (McKown *et al.*, 1988). Thus the L166 and R199 segments contain the *cis*-acting sequences that are required for efficient Tn*7* transposition.

To determine whether miniTn*7L166* × *R199* transposition products are simple insertions or cointegrates, we examined the transfer of the donor plasmid backbone by screening miniTn*7*-containing transconjugants (as in Table 2A, line 3) for the presence of the donor plasmid backbone marker

Table 2

Transposition of mini-Tn7s containing short Tn7 end segments

Tn7 or miniTn7 element	Transposition proteins	Transposition frequency† A. *From multi-copy plasmid*‡ pOX	Transposition frequency† A. *From multi-copy plasmid*‡ pOX-*attTn7*	Transposition frequency† B. *From the chromosome*§ pOX	Transposition frequency† B. *From the chromosome*§ pOX-*attTn7*
Tn7	+	nt	nt	$4{\cdot}5(\pm 0{\cdot}9)\times 10^{-5}$ (9)	nt
L1900 × *R537*	+	$8{\cdot}9(\pm 2{\cdot}1)\times 10^{-5}$ (3)	$7{\cdot}3(\pm 1{\cdot}9)\times 10^{-2}$ (3)	$9{\cdot}1(\pm 1{\cdot}7)\times 10^{-5}$ (9)	$4{\cdot}0(\pm 0{\cdot}9)\times 10^{-2}$ (9)
L166 × *R199*	+	$9{\cdot}8(\pm 1{\cdot}6)\times 10^{-5}$ (12)	$6{\cdot}1(\pm 4{\cdot}1)\times 10^{-2}$ (6)	$4{\cdot}6(\pm 0{\cdot}7)\times 10^{-4}$ (5)	$5{\cdot}7(\pm 1{\cdot}7)\times 10^{-2}$ (5)
L166 × *R199*	−	$<1{\cdot}3(\pm 0{\cdot}3)\times 10^{-7}$ (3)	$2{\cdot}8(\pm 0{\cdot}6)\times 10^{-6}$ (10)‖	$<1{\cdot}6(\pm 0{\cdot}03)\times 10^{-7}$ (3)	$<7{\cdot}2(\pm 0{\cdot}9)\times 10^{-9}$ (4)
L166F × *R199*	+	$<9{\cdot}4(\pm 0{\cdot}8)\times 10^{-8}$ (3)	$8{\cdot}5(\pm 0{\cdot}3)\times 10^{-6}$ (3)¶	nt	nt

† Transposition to the indicated target plasmids was evaluated by mating out assays as described in Materials and Methods. The transposition values are the means ± s.e.m. (number of trials is given in parentheses) and were obtained by determining the ratio of the element containing transconjugants ($Tp^R Rif^R$, $Km^R Rif^R$ or $Cm^R Rif^R$) to the total number of transconjugants ($Gm^R Rif^R$).

‡ Tns proteins were provided in the indicated experiments by intact Tn7 in chromosomal *attTn7*. The donor plasmids were pRM2: miniTn7*L1900* × *R537*, pLA1 for miniTn7*L166* × *R199* and pLA66 for miniTn7*L166F* × *R199*.

§ Tns proteins were provided in the indicated experiments (including experiments with Tn7) by pOW4 except for miniTn7*L1900* × *R537* to pOX-*attTn7*, which was supported by pOW4 derivatives containing miniMu insertions in regions non-essential for Tn7 transposition (Waddell & Craig, 1988). The values for Tn7 transposition, miniTn7*L1900* × *R537* transposition to pOX-*attTn7* and miniTn7*L166* × *R199* transposition to pOX-*attTn7* in the absence of Tns proteins were kindly provided by C. Waddell (Waddell & Craig, 1988).

‖ All Km^R transconjugants in the indicated experiments contained the donor plasmid, i.e. they were $Km^R Cb^R Rif^R$. As described in the text, they likely contain hybrid plasmids generated by a homologous recombination-driven single crossover between *attTn7* sequences in the donor and target plasmids. $Km^R Cb^S Rif^R$ transconjugants that could result from a double crossover have never been observed (detectable frequency $<3{\cdot}5\times 10^{-7}$). The hybrid plasmids could be transferred to another host by conjugation and were stable in a Tn7-containing host.

¶ A very few $Km^R Cb^S$ transconjugants (frequency $6{\cdot}6\times 10^{-7}$) were recovered among the $Km^R Cb^R$ products. It is unclear whether they are the products of a very inefficient transposition reaction utilizing the inverted Tn7L end (or some other sequence in the donor plasmid) or result from transposition-independent processes.

Cb^R. No donor plasmid transfer, i.e. no cointegrate formation, was observed (detectable frequency $<1\%$ of the transposition products). Thus, formation of simple insertions is the predominant reaction in both the *tnsD* and *tnsE*-dependent transposition pathways.

Transposition of miniTn7*L166* × *R199* requires the Tns proteins (Table 2, line 4). We suggest that the very low level transfer of this element observed in the absence of the Tns proteins with the pOX-*attTn7* target (Table 2A, line 4) likely reflects a homologous recombination-driven single crossover between the donor and target plasmids; all such transconjugants contained the donor plasmid backbone marker Cb^R in addition to the miniTn7 marker Km^R. Such recombination could occur between *attTn7* sequences in the target pOX-*attTn7* plasmid and the *attTn7* sequences flanking the miniTn7 element in the donor plasmid. Low-level transfer of the donor plasmid backbone was also observed with a transposition-defective miniTn7 element (Table 2, line 5). Our experiments were performed in *recA*$^-$ strains; however, *recA*-independent recombination involving plasmids has been observed (O'Connor *et al.*, 1986; Cohen *et al.*, 1985). *tns*-independent transfer of miniTn7*L166* × *R199* without concommitant transfer of the donor plasmid backbone marker, which could result from a double crossover on both sides of the miniTn7 has not been observed (detectable frequency of $Km^R Cb^S$ products $<3{\cdot}5\times 10^{-7}$). We also examined the orientation requirement of the Tn7 ends by analyzing the transposition of miniTn7*L166F* × *R199*, an element in which L166 has been inverted with respect to its natural orientation in Tn7. No transposition of this element to pOX was detected (Table 2, line 5).

We conclude that sequences that participate directly in Tn7 transposition in both the *tnsABC* + *tnsD* and the *tnsABC* + *tnsE* pathways are contained within the L166 and R199 terminal segments and that a specific orientation of these segments is required for efficient transposition. The products of miniTn7*L166* × *R199* transposition are simple insertions.

(c) *Identification of transposition sequences within Tn7L166 and Tn7R199*

To localize further the *cis*-acting transposition sequences within the Tn7 end segments of miniTn7*L166* × *R199*, we constructed deletions of both end segments extending from their internal boundaries towards their termini and analyzed the transposition properties of miniTn7 elements containing various L ends and R199, and of elements containing L166 and various R ends. Similar results were obtained from analysis of both multi-copy and single-copy miniTn7s containing deleted end segments (Table 3). These experiments showed that extensive sequence information in both Tn7L and Tn7R is required for transposition. Furthermore, Tn7L and Tn7R segments of different lengths are required for transposition.

Table 3
Identification of Tn7's cis*-acting transposition sequences*

miniTn7 element	Transposition frequency (%)† A. *From multi-copy plasmids*‡ pOX		A. *From multi-copy plasmids*‡ pOX-*attTn7*		B. *From the chromosome* pOX		B. *From the chromosome* pOX-*attTn7*	
L166 × *R199*	100	(12)	100	(6)	100	(5)	100	(5)
L149 × *R199*	530	(4)	54	(5)	176	(5)	333	(3)
L108 × *R199*	≤0·3	(4)	0·1	(5)	1·1	(5)	1·0	(5)
L94 × *R199*	1·0	(4)	0·1	(4)	1·8	(4)	1·6	(5)
L75 × *R199*	≤0·5	(5)	0·1	(5)	0·7	(5)	nt	
L23 × *R199*	<0·1	(3)	<0·0002	(3)	nt		nt	
L166 × *R141*	184	(5)	102	(5)	141	(4)	nt	
L166 × *R70*	27	(5)	18	(5)	26	(5)	51	(5)
L166 × *R41*	<0·09	(3)	0·02	(5)	nt		nt	

† Transposition was evaluated by mating out assays as described in Materials and Methods and in the notes to Table 2. The transposition frequencies are expressed as a percentage of miniTn*7L166* × *R199* transposition from a multi-copy plasmid and from the chromosome as given in Table 2, line 3. The number of trials is given in parentheses.

‡ The donor plasmids used in these experiments were: line 1, pLA1; 2, pLA43; 3, pLA47; 4, pLA48; 5, pLA59; 6, pLA45; 7, pLA49; 8, pLA50; and 9, pLA13. Km^R transconjugants were tested for the presence of the donor plasmid backbone marker Cb^R. No Km^RCb^R transconjugants were recovered in experiments when pOX was a target, or when pOX-*attTn7* was the target with miniTn*7* elements that transposed efficiently (lines 1, 2, 7, 8). When pOX-*attTn7* was the target for elements that transposed poorly (lines 3 to 6 and 9), Km^RCb^R transconjugants were recovered at a frequency roughly the same as observed in the absence of Tns proteins (Table 2, line 4). Only Km^RCb^S transconjugants were considered in determining the transposition frequency of these miniTn*7* elements.

Analysis of miniTn*7*s containing various Tn*7*L segments revealed that miniTn*7L149* × *R199* transposed efficiently to both pOX and pOX-*attTn7* (line 2). However, transposition of miniTn*7*s containing L108 or smaller L end segments was markedly reduced (at least 100-fold; lines 3 to 5). Thus, although sequences within the region between positions L1 and L75 can support infrequent transposition, sequences within the region between positions L108 and L149 are required for efficient transposition. Analysis of elements containing deleted Tn*7*R ends revealed that whereas miniTn*7L166* × *R141* and miniTn*7L166* × *R70* transposed efficiently (lines 7, 8), transposition of miniTn*7L166* × *R41* was significantly decreased (more than 1000-fold, line 9). Thus, some sequences necessary for efficient transposition are located between positions R41 and R70. Although the end segment extending from the terminus R1 to position R41 could not support transposition in these experiments (as in miniTn*7L166* × *R41*), a result in good agreement with other observations (Smith & Jones, 1984; Ouartsi *et al.*, 1985), it contains sequences that do participate in transposition (see below). We note that although miniTn*7L166* × *R70* is highly active, its transposition frequency was slightly but consistently lower than that of miniTn*7*s containing longer Tn*7*R segments. This result, taken together with the results of other experiments described below, may indicate that, although sequences in Tn*7*R extending beyond position R70 are not essential for efficient transposition under our experimental conditions, they may participate in and influence transposition.

Restriction enzyme analysis of eight independent insertions of miniTn*7L149* × *R199* and miniTn*7L166* × *R70* into *attTn7* of pRM2 (data not shown) revealed that these elements insert site and orientation-specifically into *attTn7*. We also determined that the transposition products of these miniTn*7*s were predominantly simple insertions by examining transconjugants containing these elements for the presence of the donor plasmid backbone (data not shown). Therefore, transposition of these two miniTn*7*s has all the characteristics of wild-type Tn*7* transposition.

The above experiments demonstrate that the *cis*-acting transposition sequences in Tn*7*L and Tn*7*R required for efficient transposition are different in length, about 150 bp in Tn*7*L and about 70 bp in Tn*7*R. Because the end deletions similarly affect transposition to pOX and to pOX-*attTn7*, the same sequences in the Tn*7* ends are likely necessary for both *tnsD* and *tnsE*-dependent transposition.

(d) *Transposition of miniTn7s containing symmetrical ends*

We also analyzed transposition properties of miniTn*7*s containing symmetrical ends, i.e. two Tn*7*L segments or two Tn*7*R segments (Table 4). These experiments revealed that sequences in Tn*7*R can substitute for the *cis*-acting transposition sequences in Tn*7*L but that Tn*7*L sequences cannot substitute for Tn*7*R.

We were unable to detect transposition of miniTn*7L166* × *L166* to either pOX or to pOX-*attTn7* (Table 4, line 2). By contrast, we observed efficient transposition of miniTn*7R199* × *R199* and miniTn*7R70* × *R70* to both of these target plasmids (lines 3, 4). In fact, the transposition frequencies of single-copy miniTn*7R199* × *R199* and miniTn*7R70* × *R70* were close to that of miniTn*7L166* × *R199*. Transposition of miniTn*7R199* × *R199* required the same *tns* genes

Table 4

Transposition of miniTn7 elements containing two Tn7L ends or two Tn7R ends

	Transposition frequency (%)†							
	A. *From multi-copy plasmids*‡				B. *From the chromosome*			
miniTn7 element	pOX		pOX-*attTn7*		pOX		pOX-*attTn7*	
L166 × *R199*	100	(12)	100	(6)	100	(5)	100	(5)
L166 × *L166*	<0·2	(6)	<0·0005	(3)	nt		nt	
R199 × *R199*	57	(3)	64	(3)	59	(5)	77	(5)
R70 × *R70*	270	(7)	20	(4)	19	(5)	35	(5)
R41 × *R199*	3·4	(5)	0·2	(5)	1·2	(5)	1·5	(5)
R41 × *R41*	<0·09	(5)	0·14	(4)	nt		nt	

† Transposition was evaluated by mating out assays as described in Materials and Methods and in the notes to Table 2. The transposition frequencies are expressed as a percentage of miniTn7*L166* × *R199* transposition from a multi-copy plasmid or from the chromosome as given in Table 2, line 3. The number of trials is given in parenthesis.

‡ The donor plasmids used in these experiments were: line 1, pLA1; 2, pDD2; 3, pDD1; 4, pLA55; 5, pLD1; and 6, pLA5. It should be noted that pLA1, pLA55 and pLA5 are pUC18 derivatives, while pDD1, pDD2 and pLD1 are pACYC184 derivatives. It should also be noted that the flanking sequences of miniTn7*L166* × *L166* in pDD2, miniTn7*R199* × *R199* in pDD1 and miniTn7*R70* × *R70* in pLA55 are different from those of the other miniTn7 elements (see Table 1). Km^R transconjugants were tested for the presence of a donor plasmid backbone marker. In the experiments of lines 1 to 3 and 5, no transconjugants containing the donor plasmid backbone were observed (detectable frequency less than 1%). In experiments in line 4 when pOX was a target, transconjugants containing the donor plasmid marker (Cb^R) were recovered at low frequency, usually about 5%. The values in the Table represent both Km^RCb^R and Km^RCb^S transconjugants. The production of these Cb^R transconjugants is considered in the text. The transconjugant Cb^R plasmids could be transferred to another strain by conjugation and were stable in a Tn7-containing host.

as did transposition of miniTn7*L166* × *R199* (N. Craig, unpublished results). We also found that miniTn7*R41* × *R199* can transpose, albeit inefficiently (line 5). Thus the segment of Tn7R extending from positions R1 to R41 contains sequences that do participate in transposition. All miniTn7*R* × *Rs* transposed at higher frequency to pOX-*attTn7* than to pOX, suggesting that they have the capacity to recognize *attTn7*. Furthermore, restriction enzyme analysis of 22 independent miniTn7*R199* × *R199* insertions and 11 independent miniTn7*R41* × *R199* insertions into the *attTn7* plasmid pRM2 (obtained as described in Materials and Methods) demonstrated that these elements insert site-specifically in *attTn7* (data not shown)†.

These experiments revealed that the *cis*-acting transposition sequences in Tn7L and Tn7R are functionally different. Other analyses of the Tn7 ends described below support this conclusion.

As described above, the products of miniTn7*L166* × *R199* transposition to pOX were simple insertions, i.e. donor plasmid backbone sequences were not present in transconjugants containing the miniTn7 element. By contrast, we found that a small fraction (usually about 5%) of miniTn7*R70* × *R70* (Table 4A, line 4) and miniTn7*R199* × *R199* transposition products from pUC-derived plasmids contained the donor plasmid backbone marker (Cb^R) (data not shown). Such products could result from cointegrate formation during transposition of these elements. Alternatively, transconjugants containing the donor plasmid backbone could be generated by simple insertion transposition from a dimeric donor plasmid (Berg, 1983). In this case, the transposing segment would be the donor plasmid backbone flanked by two copies of the miniTn7 element. We favor the latter view as dimeric pUC-based donor plasmids were present when backbone-containing transconjugants were recovered (data not shown) and because we did not detect donor backbone-containing transconjugants when these miniTn7s transposed from other donor plasmids (as in Table 4, line 3; data not shown).

(e) trans *inibition by multi-copy Tn7 ends*

In many of the experiments presented above, we analyzed the transposition of miniTn7s from multi-copy plasmids in the presence of single-copy

† We also examined the orientation specificity of miniTn7*R41* × *R199* and miniTn7*R199* × *R199* insertion into *attTn7*. Physical analysis of 11 independent pRM2 :: miniTn7*R41* × *R199* insertions, obtained from a single F'ts :: miniTn7*R41* × *R199* donor as described in Materials and Methods, revealed that this miniTn7 element inserts into *attTn7* with its R199 end in the same orientation as does Tn7 and miniTn7*L166* × *R199*. We were surprised to find that the symmetrical element miniTn7*R199* × *R199* inserted into *attTn7* in a single orientation (as defined by this element's internal drug resistance marker) when 8 independent pRM2 :: miniTn7*R199* × *R199* insertions, obtained from a single F'ts :: miniTn7*R199* × *R199* donor, were examined. However, we recovered both orientations of miniTn7*R199* × *R199* in *attTn7* when 14 independent pRM2 insertions of this element from several different pOX :: miniTn7*R199* × *R199* donors were examined. It remains to be determined if the observed differences in miniTn7*R199* × *R199* orientation specificity reflect some influence of sequences flanking the Tn7 termini in the donor site or the different procedures used to recover the pRM2 insertions. Further experiments are required to define the determinants of Tn7 orientation specificity.

Table 5
Tn7 transposition in the presence of trans *multi-copy miniTn7 elements*

miniTn7 element	Location	Tn7 transposition frequency (%)† pOX		pOX-*attTn7*	
A. *Multi-copy*‡	Plasmid				
-	-	100	(9)	100	(4)
-	pUC18	28	(5)	nt	
-	pRM2	41	(6)	66	(5)
L166 × *R199*	pLA1	0·1	(8)	3	(3)
L166 × *L166*	pDD2	32	(3)	26	(3)
R41 × *R41*	pLA5	12	(5)	72	(4)
R199 × *R199*	pDD3	≤0·1§	(3)	nt	
R199 × *R199*	pDD1	1·8	(3)	3·9	(4)
B. *Single-copy*‖	Chromosome				
L166 × *R199*	Site no. 11	33	(4)	29	(5)
L166 × *R199*	Site no. 20	16	(5)	28	(3)

† Tn7 transposition was evaluated by mating out assays as described in Materials and Methods by determining the fraction of Tn7-containing transconjugants (Tp^RRif^R) among the total transconjugants (Gen^RRif^R). Transposition frequency is expressed as percentage of Tn7 transposition observed in the absence of an additional multi or single-copy miniTn7 element; to pOX, 100% is equal to $1·7 \times 10^{-4}$ and to pOX-*attTn7* 100% is equal to $9·4 \times 10^{-3}$. The number of trials is given in parentheses.

‡ Plasmids were derivatives of pUC18, except pDD1 and pDD2, which were pACYC184 derivatives. The transposition activity of miniTn7 elements was assayed simultaneously in these experiments. MiniTn7*L166* × *R199* and miniTn7*R199* × *R199* transposed efficiently (see Table 2A, line 3, and Table 4A, line 3, respectively). miniTn7*L166* × *L166* (Table 4A, line 2) and miniTn7*R41* × *R41* (Table 4A, line 6) were transposition-defective.

§ The more effective inhibition provided by pDD3 a pUC-derived plasmid, compared to pDD1, a pACYC-derived plasmid, probably reflects differences in copy number between these plasmids.

‖ In these experiments, Tn7 was located in chromosomal *attTn7* and miniTn7*L166* × *R199* was located at other chromosomal sites. The miniTn7*L166* × *R199* insertions were likely derived by *tnsE*-dependent transposition; their construction will be described elsewhere (K. Kubo and N.L.C.). The transposition frequency of the miniTn7*L166* × *R199* elements from the chromosome was also measured in these experiments and were: from no. 11, to pOX = $3·5 \times 10^{-5}$ and to pOX-*attTn7* = $1·6 \times 10^{-2}$, and from no. 20, to pOX = $1·0 \times 10^{-4}$ and to pOX-*attTn7* = $6·7 \times 10^{-2}$.

chromosomal Tn7. We have also examined Tn7 movement under these conditions (Table 5). The presence of multi-copy miniTn7*L166* × *R199 in trans* to both Tn7 and the target plasmid dramatically reduced the frequency of Tn7 transposition in both the *tnsABC* + *tnsE* and the *tnsABC* + *tnsD* pathways (Table 5, line 4). Because little decrease in Tn7 transposition was observed in the presence of multi-copy plasmids lacking the Tn7 ends (lines 2 and 3) or in the presence of single-copy chromosomal miniTn7*L166* × *R199* (lines 9, 10), we conclude that this inhibition is mediated by the multi-copy Tn7 ends. We also examined single-copy Tn7 transposition in the presence of multi-copy miniTn7s containing two Tn7L ends or two Tn7R ends. While little inhibition by multi-copy miniTn7*L166* × *L166* (line 5) or miniTn7*R41* × *R41* (line 6) was observed, multi-copy miniTn7*R199* × *R199* was a potent *trans* inhibitor of single-copy Tn7 transposition (lines 7, 8). Thus, the presence of Tn7R segments containing the sequences in this end required for effiicient transposition is required for *trans* inhibition. Other experiments described below suggest that the transposition sequences within the R199 segment mediate this multi-copy, *trans* inhibition of Tn7 transposition.

We emphasize that, although little transposition of single-copy Tn7 was observed in the presence of multi-copy miniTn7*L166* × *R199* or miniTn7*R199* × *R199*, the miniTn7s did transpose. Indeed, the transposition frequency of multi-copy miniTn7*L166* × *R199* (Table 2A, line 3) or miniTn7*R199* × *R199* (Table 4A, line 3) was comparable to the frequency of single copy Tn7 transposition in the absence of these multi-copy miniTn7s (see also note ‡ to Table 5). One interpretation of these results is that under these specific conditions the transposition proteins are limiting, so that the apparent inhibition of Tn7 transposition by the multi-copy miniTn7s actually reflects a competition for the available transposition protein(s) between the multi-copy miniTn7s and single-copy Tn7. It should also be noted that the level of transposition protein(s) could be different in the presence of the multi-copy miniTn7 (see Discussion).

In the above experiments, we analyzed *trans* inhibition by multi-copy Tn7 ends within miniTn7 elements, which could themselves transpose. We also examined the capacity of Tn7 ends to provide *trans* inhibition individually, i.e. when not contained within a miniTn7 element. In these experiments, we used a different transposition assay in which a miniTn7 transposed from a phage lambda derivative upon infection into cells containing a *tns* plasmid and vacant chromosomal *attTn7*. Transposition was measured in the presence and absence of an additional multi-copy plasmid containing either Tn7L or Tn7R end segments (Table 6). In these experiments, as in those described above, the multi-copy Tn7 end segments were *trans* to both the transposing miniTn7 element and to the target replicon. The presence of multi-copy R199 dramatically (about 200-fold) inhibited transposition of the miniTn7 from the infecting phage (Table 6, lines 2 and 4). However, *trans*-position was unaffected by the presence of multi-copy L166 or by multi-copy R41 (lines 3, 5, 6). Thus, the multi-copy R199 segment need not be part of a miniTn7 element to provide *trans* inhibition. We also found that the capacity of multi-copy R199 to provide *trans* inhibition was independent of the sequences flanking this segment; *trans* inhibition was observed with R199 segments flanked either by *attTn7* or by other sequences unrelated to *attTn7* (compare lines 2 and 4).

What sequences within the R199 segment mediate *trans* inhibition? As multi-copy R70 can provide *trans* inhibition (Table 6, line 8), the

Table 6
Inhibition by individual multi-copy Tn7 ends

Tn7-end segment flanking sequence	Plasmid†	Transposition frequency (%)‡	
	pUC18	100	(8)
R199-R1 *attTn7* + 165	pLA17	0·5	(7)
R41-R1 *attTn7* + 165	pLA21	120	(5)
R199-R1 EC-18	pLA72	1·0	(9)
R41-R1 EC-18	pLA73	186	(9)
−345 L1-L166	pLA19	110	(3)
R141-R1 *attTn7* + 64	pLA36	1·1	(4)
R70-R1 *attTn7* + 64	pLA51	1·5	(4)
R38-R199	pLA62	85	(4)
R38-R141	pLA63	83	(4)

† The plasmids containing the Tn7 end segments were pUC18 derivatives obtained as described in Materials and Methods. Tn7 end segments were flanked either by *attTn7* sequences or by sequences unrelated to *attTn7* as indicated. Similar results were also obtained with plasmids containing the end segments in different sites in the vector backbone and/or in orientations different from those used in the experiments shown in lines 2, 3, 7, 8 and 10 (data not shown).

‡ Transposition was evaluated by lambda hop assays performed as described in Materials and Methods. Transposition of miniTn7*L166* × *R199* from lambda KK1 was supported by transposition proteins from plasmid pCW4 (*tnsABCDE*) except lines 4 and 5 where pCW4::miniMu78 (*tnsABCD*) was used. The frequency of transposition is expressed as a percentage of miniTn7*L166* × *R199* transposition in the presence of pUC18. For hosts containing pCW4, 100% is equal to $9{\cdot}2 \times 10^{-5}$ transpositions/plaque forming unit and for hosts containing pCW4::miniMu78, 100% is equal $3{\cdot}7 \times 10^{-4}$ transpositions/plaque forming unit. The number of trials is given in parentheses.

Table 7
Transposition immunity in the tnsABC+tnsE *pathway*

Target plasmid	Tn7 transposition frequency (%)†	
pOX	100	(9)
pOX::miniTn7*L166* × *R199*	1·1	(6)‡
pOX::miniTn7*R199* × *R199*	0·2	(3)
pOX::miniTn7*R70* × *R70*	0·5	(9)
pOX::miniTn7*L94* × *R199*	1·5	(2)
pOX::miniTn7*R41* × *R199*	1·4	(9)
pOX::miniTn7*R41* × *R41*	43	(7)
pOX::miniTn7*L166* × *R70*	180	(4)
pOX::miniTn7*L166* × *R141*	1·9	(5)

† The transposition of Tn7 from chromosomal *attTn7* to pOX and the indicated pOX derivatives was evaluated by mating out assays as described in Materials and Methods. The transposition frequency is expressed as percentage of Tn7 transposition to pOX; 100% is equal to $1{\cdot}7 \times 10^{-4}$. The number of trials is given in parentheses. We emphasize that only *tnsE*-dependent transposition to pOX occurs (Waddell & Craig, 1988).

‡ The rare pOX products containing both Tn7 and miniTn7*L166* × *R199* were stable after many generations of growth under non-selective conditions.

sequences required must lie between positions R70 and R1. Moreover, the sequences between positions R38 and R1 are required but not sufficient to provide *trans* inhibition (lines 9, 10, 3). It is notable that sequences between positions R70 and R1 are also required for transposition itself. This correlation is consistent with the view that the phenomenon of *trans* inhibition of single-copy Tn7 transposition by multi-copy Tn7R ends reflects a competition between sequences in these segments that interact with a limiting transposition protein(s).

(f) *Tn7 transposition immunity*

We presented evidence above that multi-copy Tn7R segments can inhibit transposition when present *in trans* to a target molecule. The Tn7 ends can also inhibit transposition *in cis*, i.e. when present in a target molecule. Hauer & Shapiro (1984) found that the presence of miniTn7*L1900* × *R537* in the low-copy number plasmid RP4 substantially reduced the frequency of subsequent Tn7 insertion into this replicon. The ability of transposon ends in a target replicon to inhibit subsequent transposon insertion into that replicon is called transposition immunity. Transposition immunity reflects "action at a distance" as the presence of the transposon ends in the target plasmid can inhibit subsequent insertion of another copy of the transposon into any other site in the target plasmid. We have found that Tn7 transposition immunity is active in both the *tnsABC* + *tnsE* and the *tnsABC* + *tnsD* transposition pathways. We have also found that the presence of Tn7R end segments in a target molecule is sufficient to provide immunity and that Tn7R ends of sufficient length to contain the sequences necessary for efficient transposition are required to provide immunity to a target replicon.

(i) *The* tnsABC+tnsE *pathway*

We evaluated Tn7 transposition immunity in the *tnsE*-dependent transposition pathway by measuring in mating out assays the ability of miniTn7s in the low-copy number conjugative plasmid pOX to reduce the frequency of subsequent Tn7 insertion into these plasmids. The presence of miniTn7*L166* × *R199* in pOX substantially (about 100-fold) inhibited subsequent Tn7 insertion into this target molecule (Table 7, line 2). However, the presence of single-copy miniTn7 in a replicon other than the pOX target had little effect on Tn7 transposition to pOX (Table 5, lines 9, 10). We emphasize that this inhibition is different from the *trans* inhibition provided by multi-copy Tn7 ends described above in two ways: (1) the miniTn7 element resides within the target molecule, i.e. is *cis* rather than *trans* to the target; and (2) when resident in the target molecule, single-copy miniTn7 can inhibit Tn7 transposition, whereas multi-copy miniTn7 is required for *trans* inhibition. The ability of miniTn7*L166* × *R199* in pOX to inhibit subsequent Tn7 insertion does not reflect the prior occupancy of a highly preferred Tn7 insertion site as there are no preferred Tn7 insertion sites within pOX (Waddell & Craig, 1988). Moreover, miniTn7*L166* × *R199* at several different positions in other target plasmids related to pOX all

inhibited subsequent Tn*7* insertion into these plasmids (data not shown). It is also unlikely that the low-frequency recovery of plasmids containing both Tn*7* and miniTn*7L166* × *R199* reflects the activity of a transposon-encoded resolution system as we and others (data not shown: Moore & Krishnapillai, 1982; Hassan & Brevet, 1983) have observed that plasmids containing two copies of Tn*7* are stable. Thus, the presence of Tn*7* end segments containing this element's transposition sequences in a large (greater than 50 kb) target plasmid can provide this molecule with immunity to subsequent Tn*7* insertion.

To determine which sequences within the Tn*7* ends confer transposition immunity, we examined the capacity of various miniTn*7*s in pOX to inhibit subsequent Tn*7* insertion. We found that both miniTn*7R199* × *R199* and miniTn*7R70* × *R70* could inhibit subsequent Tn*7* insertion (Table 7, lines 3, 4). Thus the sequences within R70 are sufficient whereas Tn*7*L sequences are dispensable for immunity. We also found that miniTn*7* elements incapable of efficient transposition themselves, such as miniTn*7L94* × *R199* and miniTn*7R41* × *R199* can provide immunity (lines 5, 6). It should be noted that in these miniTn*7*s, at least one end of the element does contain the Tn*7*R sequences necessary for efficient transposition. We also found that miniTn*7R41* × *R41*, both ends of which lack sequences necessary for efficient transposition could not provide immunity to a target replicon (line 7). Thus, the ability of a miniTn*7* to provide transposition immunity correlates with the presence of Tn*7*R sequences necessary for efficient Tn*7* transposition. This correlation supports the hypothesis that the same sequences at the Tn*7* ends that participate directly in transposition mediate transposition immunity when present in the target molecule.

In view of the effective transposition immunity provided by miniTn*7R70* × *R70* (Table 7, line 4), it was intriguing to find that the presence of miniTn*7L166* × *R70* in a target replicon did not reduce the frequency of subsequent Tn*7* insertion into that target, i.e. this element did not provide immunity (line 8). We interpret this result to support our speculation that some sequences that lie beyond R70, i.e. between positions R70 and R141, can contribute to both transposition immunity and, as considered above, to transposition. Perhaps the differing abilities of miniTn*7R70* × *R70* and miniTn*7L166* × *R70* to provide immunity, i.e. miniTn*7R70* × *R70* does while miniTn*7166* × *R70* does not, reflect some interaction between the Tn*7* ends resident in the target replicon that occurs more efficiently in miniTn*7R70* × *R70* than in miniTn*7L166* × *R70*.

(ii) *The* tnsABC+tnsD *pathway*

To determine whether transposition immunity is also active in the *tnsD*-dependent transposition pathway, we examined *tnsABC* + *tnsD*-promoted transposition of miniTn*7L1900* × *R537* to pOX-*attTn7* target plasmids containing miniTn*7L166* × *R199* insertions at sites other than *attTn7*. The reactivity of *attTn7* in these plasmids was markedly decreased (about 100-fold, Table 8). Thus, transposition immunity is also active in the *tnsD*-dependent transposition pathway. We note that in both the *tnsD* and *tnsE*-mediated pathways, the presence of a miniTn*7* in the target plasmid decreased the frequency of subsequent Tn*7* transposition to the same degree (about 100-fold, compare Table 7, line 2, to Table 8, lines 2 to 4).

Table 8
Transposition immunity in the tnsABC+tnsD *pathway*

Target plasmid†		Transposition frequency (%)‡	
pOX-*attTn7*		100	(3)
pOX-*attTn7*::EP-18	miniTn*7L166* × *R199*	0·5	(3)
pOX-*attTn7*::EP-19	miniTn*7L166* × *R199*	1	(3)
pOX-*attTn7*::EP-20	miniTn*7L166* × *R199*	1·1	(3)

† The pOX-*attTn7* plasmids containing miniTn*7L166* × *R199* were obtained by *tnsABC* + *tnsE* transposition as described in Materials and Methods; the miniTn*7* elements lie in target sites other than *attTn7* as verified by Southern analysis (Waddell & Craig, 1988).

‡ Transposition of miniTn*7L1900* × *R537* from chromosomal *attTn7* to the indicated target plasmids was evaluated by mating out assays as described in Materials and Methods. Transposition was promoted by a *tnsABCD* plasmid, pUW4::miniMu78::miniTn*7L1900* × *R537*-containing transconjugants were identified by $Cm^R Rif^R$ selection. Transposition frequency is expressed as the percentage of miniTn*7L1900* × *R537* transposition to pOX-*attTn7*; 100% is equal to $2{\cdot}2 \times 10^{-2}$. The number of trials is given in parentheses.

(g) *Inactivation of* attTn7 *adjacent to Tn7*

In the above experiments, we demonstrated that the presence of the Tn*7* ends in a target replicon can inhibit subsequent Tn*7* insertion when they lie at a distance from either *attTn7* or a target site unrelated to *attTn7*. We have also investigated the effects of Tn*7* ends on *attTn7* target activity when they are located in *attTn7*, i.e. in the configuration that results from Tn*7* insertion into this site. This situation was of particular interest because the nucleotides required for *attTn7* target activity lie entirely to one side of the specific point of Tn*7* insertion and do not include the insertion point itself (Gringauz *et al.*, 1988). Thus, Tn*7* insertion into *attTn7* does not physically disrupt the sequences required for *attTn7* target activity. We analyzed *tnsABC* + *tnsD*-promoted transposition of miniTn*7L1900* × *R537* to pOX-*attTn7* plasmids containing various miniTn*7* elements already resident in *attTn7*. Transposition to *attTn7* was completely blocked by a number of miniTn*7* elements including miniTn*7L166* × *R199* and various miniTn*7R* × *R* elements (Table 9). We emphasize that this inactivation of an *attTn7* adjacent to a miniTn*7* was much greater than the effects of many of these miniTn*7*s in transposition

Table 9
Inactivation of attTn7 *by adjacent Tn7 ends*

Target plasmid†	Tn7 transposition frequency (%)‡	
pOX-*attTn7*	100	(3)
pOX-*attTn7* :: miniTn7*L166* × *R199*	<0·001	(4)
pOX-*attTn7* :: miniTn7*L166* × *R141*	<0·001	(2)
pOX-*attTn7* :: miniTn7*R199* × *R199*	<0·001	(4)
pOX-*attTn7* :: miniTn7*R70* × *R70*	<0·002	(3)
pOX-*attTn7* :: miniTn7*R41* × *R199*	<0·001	(3)
pOX-*attTn7* :: miniTn7*L166* × *R70*	<0·001	(3)
pOX-*attTn7* :: miniTn7*L166* × *R41*	0·17	(3)

† The pOX-*attTn7* plasmids containing miniTn7s in *attTn7* were obtained as described in Materials and Methods.

‡ Transposition of miniTn7*L1900* × *R537* from chromosomal *attTn7* to the indicated target plasmids was evaluated by mating out assays as described in Materials and Methods. Transposition was promoted by a *tnsABCD* plasmid, pUW4::miniMu78::miniTn7*L1900* × *R537*-containing transconjugants were identified by $Cm^R Rif^R$ selection. Transposition frequency is expressed as a percentage of miniTn7*L1900* × *R537* transposition to pOX-*attTn7*; 100% is equal to $2{\cdot}2 \times 10^{-2}$. The number of trials is given in parentheses.

immunity. The reactivity of *attTn7* already occupied by a miniTn7 element was reduced by at least 100,000-fold whereas the reactivity of *attTn7* at a distance from a miniTn7 element was reduced only about 100-fold (compare Table 9 to Table 8). It is also notable that miniTn7*L166* × *R70*, which failed to provide detectable transposition immunity (Table 7, line 8), completely inactivated an adjacent *attTn7* (Table 9, line 7). Furthermore, even the presence of the immunity-defective miniTn7*L166* × *R41* in *attTn7* substantially inhibited subsequent insertions into this site (about 600-fold, line 8). Thus, the capacity of Tn7 ends to inactivate a nearby *attTn7* is much more pronounced than their ability to inactivate a distant *attTn7* site.

4. Discussion

The termini of transposable elements provide specific *cis*-acting sequences, recognized and acted upon by the element's transposition proteins. The Tn7-encoded transposition genes can mediate two different but related transposition reactions: (1) *tnsABC* + *tnsD* promote transposition to *attTn7* and pseudo-*attTn7* sites and (2) *tnsABC* + *tnsE* promote transposition to sites unrelated to *attTn7* (Rogers *et al.*, 1986; Waddell & Craig, 1988; K. Kubo and N. Craig, unpublished results). We have not found any differences in the transposon-carried *cis*-acting sequences required for these two reactions. Thus, it is not unreasonable to suggest that similar protein–DNA interactions at the transposon ends mediate both *tnsD* and *tnsE*-dependent transposition and that these pathways differ only in their utilization of distinct target-specific proteins, likely TnsD and TnsE, bound to their cognate DNA sites.

Extensive sequence information at each end of Tn7 is required for efficient transposition. We have shown that sequences throughout the segments extending from the left end terminus at position L1 to position L149 and from the right end terminus at position R1 to at least position R70 are likely directly involved in transposition. Thus, the *cis*-acting transposition sequences in Tn7L and Tn7R are structurally different and asymmetric. The extreme (30 bp) termini of Tn7 are highly related inverted repeats (Fig. 4; Lichtenstein & Brenner, 1982; Gosti-Testu & Brevet, 1982). However, considerably more sequence information than these repeats is required for Tn7 transposition (this work; Smith & Jones, 1984; Ouartsi *et al.*, 1985).

Although different, the ends of Tn7 are related by the presence of several highly related copies of a 22 bp sequence (Fig. 4; Lichtenstein & Brenner, 1982). Tn7L contains three directly repeated 22 bp sequences separated by unrelated sequences; Tn7R contains four 22 bp sequences arranged contiguously in direct orientation. The 22 bp sequences in Tn7L lie in inverted orientation with respect to those in Tn7R. Tn7's 30 bp terminal inverted repeats contain one copy of the 22 bp sequence and an additional 8 bp at the extreme transposon termini. Two types of evidence suggest that these 22 bp sequences are important *cis*-acting transposition sites. The 22 bp repeats are prominent features of the short Tn7L and Tn7R end segments we have found to be required for efficient transposition. The smallest Tn7R segment capable of promoting efficient transposition is R70; the majority of this segment comprises three 22 bp repeats (Fig. 4). Removal of about 1·5 of these repeats (as in R41) reduces transposition at least 1000-fold. The smallest Tn7L segment we found to

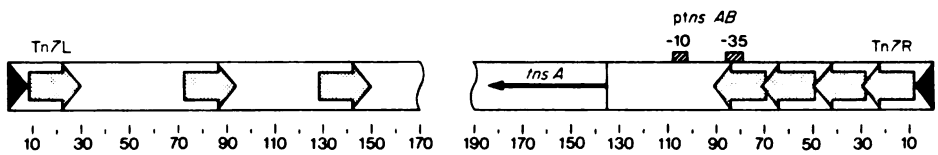


Figure 4. Structure of Tn7L and Tn7R. The Tn7L and Tn7R segments are shown and their sequences numbered. The stippled arrows represent the highly similar 22 bp repeats in each end of Tn7 (Lichtenstein & Brenner, 1982), the triangles at the extreme termini represent the 8 bp portion of Tn7's terminal inverted repeats, which are distinct from the 22 bp repeats. The proposed *tnsAB* promoter and amino-terminal coding sequence for *tnsA* are indicated (Gay *et al.*, 1984; Rogers *et al.*, 1986; Waddell & Craig, 1988).

be efficient in transposition is L149; this segment contains all three of Tn7L's 22 bp repeats. L108, which lacks one of these repeats as well as sequences between the repeats, is defective in transposition. Furthermore, other work from this laboratory has shown that TnsB can specifically recognize this sequence (McKown *et al.*, 1987; L. Arciszewska, R. McKown, N. Craig, unpublished results). It remains to be determined if other sequences in Tn7 ends, such as those which separate the 22 bp repeats in Tn7L, also have a specific role in transposition.

Although Tn7R contains four 22 bp repeats, Tn7R segments containing only three repeats are nearly as active in transposition as those with four. This finding need not contradict the hypothesis that the 22 bp repeats are important transposition signals. There are many examples of repeated protein-binding sites in which deletion of one site does not alter the apparent activity of the remaining sites (Fromm & Berg, 1983; Jayaram, 1985; Groenen *et al.*, 1985). The presence of four 22 bp repeats in Tn7R might be required under transposition conditions different than those we have used, for example when the concentrations of the proteins which act upon these repeats are lower. The 22 bp repeat sequence may also play a role in the regulation of *tns* gene expression. The proposed −35 sequence of the promoter, which directs expression of the *tnsABC* operon, is embedded within the fourth 22 bp repeat of Tn7R (Fig. 4; Gay *et al.*, 1986). Furthermore, expression from this promoter appears to be repressed in the presence of *tnsB* (Rogers *et al.*, 1986). There are known examples of other recombination sites that are also involved in the control of gene expression (Reed *et al.*, 1982; Wishart *et al.*, 1983; Wells & Grindley, 1984).

The ends of Tn7 are functionally as well as structurally asymmetric. We have demonstrated that the *cis*-acting transposition sequences in Tn7R can substitute for those in Tn7L in both *tnsD* and *tnsE*-dependent transposition, but that Tn7L cannot substitute for Tn7R. We have also found differences in the activities of the Tn7 ends in other processes related to transposition such as the ability to provide *trans* inhibition and in transposition immunity (see below). In all cases, the presence of Tn7R sequences is sufficient, while Tn7L sequences are dispensable or ineffective. These functional asymmetries of the Tn7 ends could reflect a more efficient interaction between Tn7R and a transposition protein(s) that recognizes sites common to both Tn7L and Tn7R, for example the proposed TnsB interaction with the 22 bp repeat sequences. Differential protein binding to Tn7L and Tn7R could result from different intrinsic affinities of the binding sites in each end or from different cooperative interactions between variously positioned sites within each end. Indeed, the *tnsB*-dependent end binding activity does have an apparently higher affinity for Tn7R than for Tn7L (McKown *et al.*, 1987). Alternatively, the observed differences in the activities of Tn7L and Tn7R may reflect a unique interaction of a transposition protein(s) with Tn7R.

We have found that Tn7R segments in a multicopy plasmid can inhibit the transposition of single-copy Tn7 to another target replicon. By contrast, we have not observed such *trans* inhibition by multi-copy Tn7L segments. This *trans* inhibition by multi-copy Tn7R requires the same sequences in Tn7R as are required for transposition itself. An attractive explanation is that this phenomenon actually reflects a competition for a rate-limiting transposition protein between transposition sequences in the many Tn7R segments in the multicopy plasmid and in the single copy Tn7. It should be noted that the presence of the multi-copy Tn7R segments could also affect the levels of the Tns proteins if the Tn7R segment contains a binding site(s) for a component of the *tns* regulatory circuitry. The smallest inhibitory segment (Tn7R70) lacks *tns* coding sequences or sequences complementary to the 5′ end of the proposed *tnsABC* mRNA (Gay *et al.*, 1986). Thus, it seems unlikely to us that *trans* inhibition could reflect the expression of a diffusible transposition inhibitor such as a truncated *tns* polypeptide or, as is the case with transposon Tn*10* (Simons & Kleckner, 1983), an anti-sense RNA.

The products of intact Tn7 transposition are simple insertions rather than cointegrates (Sherratt *et al.*, 1981). We have shown here that miniTn7 elements containing short Tn7 end segments also form simple insertions. Thus, Tn7 does not appear to encode an internal resolution site distinct from its terminal transposition sequences. Several other types of evidence are consistent with the view that Tn7 transposes through the direct formation of simple insertions rather than a two-step reaction involving cointegrate formation and resolution as occurs with transposons of the Tn*3* family (Gill *et al.*, 1978; Heffron *et al.*, 1978). DNA molecules containing two copies of Tn7 are stable and do not resolve (our unpublished observation; Hassan & Brevet, 1983; Hauer & Shapiro, 1984). Also, none of the *tns* genes encodes a protein that acts exclusively to resolve cointegrates (Waddell & Craig, 1988).

How can the organization of Tn7's *cis*-acting transposition sequences be compared to those of other transposable elements? Tn7 is unlike those transposons whose *cis*-acting transposition sequences are provided within terminal inverted repeats (Johnson & Reznikoff, 1983; Sasakawa *et al.*, 1983; Way & Kleckner, 1984; Weinert *et al.*, 1984; Gamas *et al.*, 1985; Huang *et al.*, 1986). The organization of Tn7's transposition sequences is perhaps most reminiscent of bacteriophage Mu, whose ends are also structurally asymmetric and contain multiple recognition sites for MuA transposase (Craigie *et al.*, 1984; Groenen *et al.*, 1985). However, we emphasize that there are no strong sequence similarities between the end sequences of Mu and Tn7 nor do the Tn7 transposition proteins promote the transposition of miniMu elements (Waddell & Craig, 1988).

(a) *Tn7 transposition to target molecules containing the Tn7 ends*

We have demonstrated here that the presence in a target molecule of miniTn*7* elements composed of short end segments containing the *cis*-acting transposition sequences prevents subsequent transposition of another copy of Tn*7* to this molecule. Such interference is most pronounced when the target site is immediately proximal to the Tn*7* element, as in *attTn7*::Tn*7*, a process we call inactivation of adjacent *attTn7*. The ends of Tn*7* also strongly, but to a lesser degree, inhibit transposition to target sites that lie at a distance from the transposon ends, i.e. provide transposition immunity. In the following sections, we discuss these transposition-related phenomena individually, although it should be noted that they could be mechanistically related.

(i) *Inactivation of* attTn*7* *by adjacent Tn7 ends*

The sequences required for *attTn7* target activity lie entirely to one side of the specific point of Tn*7* insertion so that Tn*7* insertion into *attTn7* does not physically disrupt these required sequences (Gringauz *et al.*, 1988). Thus, it might appear that *attTn7*::Tn*7* should be a target for subsequent insertion of another copy of Tn*7*. However, tandem insertions of Tn*7* into *attTn7* have not been observed (Lichtenstein & Brenner, 1981; Hauer & Shapiro, 1984) nor have insertions of miniTn*7*s into *attTn7*::Tn*7* been detected (K. Kubo & N. Craig, unpublished results). How is *attTn7*::Tn*7* protected against multiple Tn*7* insertions? We have shown here that the presence in *attTn7* of miniTn*7* elements containing the Tn*7*R sequences necessary for efficient transposition, such as miniTn*7L166* × *R70* or miniTn*7R70* × *R70*, is sufficient to provide a complete inactivation of *attTn7* (reduction of insertion frequency greater than 100,000-fold). A miniTn*7* element with a Tn*7*R segment containing only part of the *cis*-acting transposition sequences (miniTn*7L166* × *R41*) provides considerable, but not complete, inactivation of *attTn7* (more than 100-fold reduction in insertion frequency). These results suggest to us that the loss of target activity of *attTn7*::Tn*7* is related to the presence of the *cis*-acting transposition sequences within the proximal Tn*7*R end. Perhaps the binding of transposition proteins to the Tn*7*R end prevents an interaction between *attTn7* and a protein necessary for its target activity. The inactivation of *attTn7*::Tn*7* does not apparently result from changes in the primary DNA sequence at the specific point of insertion in *attTn7*::Tn*7*. We have found that a segment containing the sequences that result from Tn*7* insertion into *attTn7*, i.e. a junction between the extreme terminus of Tn*7*R and *attTn7*, can exhibit a target activity similar to that of *attTn7* itself (data not shown).

(ii) *Transposition immunity*

Transposition immunity is the phenomenon in which the presence of one copy of a transposable element in a target DNA molecule inhibits subsequent transposition of another copy of this element into any position in the target replicon. This *cis*-acting process has been observed with Tn*3*-like transposons (Robinson *et al.*, 1977) and bacteriophage Mu (Reyes *et al.*, 1987; Adzuma & Mizuuchi, 1988). It has been shown in these systems that the same *cis*-acting sequences are required for transposition and for transposition immunity, and that the presence of a single end of the transposon in the target is sufficient to provide immunity (Adzuma & Mizuuchi; Lee *et al.*, 1983; Arthur *et al.*, 1984; Huang *et al.*, 1986).

There are conflicting reports regarding Tn*7* transposition immunity. Hassan & Brevet (1983) observed that Tn*7* transposed at frequencies similar to those of plasmid RP4 and to its derivative carrying Tn*7* (RP4::Tn*7*). However, Hauer & Shapiro (1984) found that the presence of Tn*7* or miniTn*7L1900* × *R537* (Tn*7*S::Tn*9* Δ *Pst*I in their nomenclature) in plasmid RP4 substantially reduced the frequency of subsequent Tn*7* insertion in this replicon. As also noted by Hauer & Shapiro, these two studies were carried out under different conditions. For example, Hassan & Brevet analyzed Tn*7* transposition from a plasmid whereas Hauer & Shapiro analyzed transposition from the chromosome. Variations in transposition immunity have been observed under different experimental conditions with other transposable elements (Wallace *et al.*, 1981; Lee *et al.*, 1983; Heritage & Bennett, 1984).

We have observed that Tn*7* displays transposition immunity when it transposes from the chromosome to F-related target plasmids containing miniTn*7* elements. We found that miniTn*7L166* × *R199* provides transposition immunity in both the *tnsABC*′ + *tnsD* and the *tnsABC*′ + *tnsE* transposition pathways. Thus, the presence of short Tn*7* end segments containing this element's transposition sequences can make these large target plasmids (greater than 50 kb) immune to further Tn*7* insertion.

Which sequences in miniTn*7L166* × *R199* confer transposition immunity? Our experiments suggest that the *cis*-acting sequences of Tn*7*R that are necessary for efficient transposition are also required for transposition immunity. This view is indicated by our finding that miniTn*7R70* × *R70* confers immunity, whereas miniTn*7R41* × *R41* does not. We also found that miniTn*7* elements that are transposition-defective can provide transposition immunity if they contain an intact Tn*7*R end (as in miniTn*7R41* × *R199*). It remains to be determined if, with Tn*7* as with other transposons, the presence in the target molecule of a single transposon end containing the element's *cis*-acting transposition sequences is sufficient to provide immunity.

If the same sequences in Tn*7*R are involved in transposition and in transposition immunity, why does miniTn*7L166* × *R70*, an element highly active in transposition, fail to provide immunity? Our

interpretation is that R70 lacks sequences that contribute to transposition immunity and to transposition itself. These additional sequences must lie between positions R71 and R141 because miniTn*7L166* × *R141* does effectively provide immunity. We suggested above that the 22 bp repeats in Tn7L and Tn7R are essential transposition signals. R70 contains only three of the Tn7R's four 22 bp repeats whereas R141 contains all four of these repeats. Perhaps the presence of all four 22 bp repeats allows a much more effective association of transposition proteins with Tn7R. Transposition immunity may be a more sensitive reflection of the interaction between the transposon ends and transposition proteins under our experimental conditions. Moreover, immunity effects by Tn7L end segments might even be detectable under other conditions.

What is the mechanism of transposition immunity? Adzuma & Mizuuchi (1988) have shown that transposition immunity in bacteriophage Mu reflects the inability of MuB, a transposition protein required for efficient capture of the target DNA molecule, to bind effectively to a target DNA containing the transposase MuA bound to a transposon end. A somewhat different view suggests that transposition immunity of Tn*3*-like transposons involves an inspection of a target molecule by the incoming transposition complex and subsequent dissociation of the latter when a transposon end segment in the target molecule is encountered (Lee *et al.*, 1983; Sherratt *et al.*, 1983). An attractive hypothesis is that with Tn*7*, as with bacteriophage Mu, transposition immunity reflects the inability of Tn7's target proteins TnsD and TnsE to effectively interact with target DNA molecules containing Tn*7* end segments.

The relationship between transposition immunity, i.e. the ability of a transposon end to inhibit subsequent transposon insertion at a distant target site, and inactivation of adjacent *attTn7*, a target very close to a transposon end, is not known. We have observed that the inhibition of subsequent Tn*7* insertion is much stronger when the target site is immediately proximal to the Tn*7* end as in the inactivation of adjacent *attTn7*. Transposition immunity and adjacent *attTn7* inactivation might both reflect the same process whose potency is distance-dependent and much more pronounced at very short distances. Alternatively, adjacent *attTn7* inactivation may involve a mechanism distinct from that of transposition immunity. For example, the binding of a transposition protein to Tn7R might directly occlude the binding site of a protein required for *attTn7* target activity.

5. Summary

We have shown that the *cis*-acting sequences at the ends of Tn*7* required in transposition are extensive and that Tn7L and Tn7R are structurally as well as functionally different. In addition, we have shown that the Tn7R end sequences that participate directly in Tn*7* transposition also mediate other transposition-related processes such as *trans* inhibition of transposition by multi-copy Tn7R ends and the inactivation of target DNA molecules by transposition immunity and adjacent inactivation. In transposition as well as in these transposition-related phenomena, Tn7R is essential while Tn7L is dispensable.

We thank J. Shapiro for Tn*7*S, Tn*9*Δ*Pst*I and N. Kleckner for λ 1046. We thank Susan Michaelis and all members of the Craig laboratory for very helpful discussions during the course of this work. We also thank Brenda Andrews, Lorraine Marsh, Susan Michaelis and Candy Waddell for comments on this manuscript and Karyl Nakamura for its excellent preparation. L.K.A. was supported by a University of California Biotechnology and Education Program grant and by an L. P. Markey Charitable Trust Predoctoral grant. This work was supported by a grant from the NIH to N.L.C.

References

Adzuma, K., & Mizuuchi, K. (1988). *Cell*, **53**, 257–266.

Arthur, A., Nimmo, E., Hettle, S. & Sherratt, D. (1984). *EMBO J.* **3**, 1723–1729.

Barry, G. (1986). *Biotechnology*, **4**, 446–449.

Barth, P. T. & Grinter, N. J. (1977). *J. Mol. Biol.* **113**, 455–474.

Barth, P. T., Datta, N., Hedges, R. W. & Grinter, N. J. (1976). *J. Bacteriol.* **125**, 800–810.

Barth, P. T., Grinter, N. J. & Bradley, D. E. (1978). *J. Bacteriol.* **133**, 43–52.

Beckwith, J. R., Singer, E. R. & Epstein, W. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 393–401.

Berg, D. E. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 792–796.

Cohen, A., Zipora, S., Broido, S. & Laban, A. (1985). In *Plasmids in Bacteria* (Helinski, D. R., Cohen, S. N., Clewell, D. B., Jackson, D. A. & Hollaender, A., eds), pp. 505–519. Plenum, New York.

Craig, N. L. (1989). In *Mobile DNA* (Berg, D. E. & Howe, M. M., eds), pp. 211–225. American Society for Microbiology, Washington, D.C.

Craig, N. L. & Kleckner, N. (1987). In Escherichia coli *and* Salmonella typhimurium *Cellular and Molecular Biology*, pp. 1054–1070. American Society for Microbiology, Washington, D.C.

Craigie, R., Mizuuchi, M. & Mizuuchi, K. (1984). *Cell*, **39**, 387–394.

Fling, M. & Richards, C. (1983). *Nucl. Acids Res.* **11**, 5147–5158.

Fling, M., Kopf, J. & Richards, C. (1985). *Nucl. Acids Res.* **13**, 7095–7106.

Fromm, M. & Berg, P. (1983). *J. Mol. Appl. Genet.* **2**, 127–135.

Gamas, P., Galas, D. & Chandler, M. (1985). *Nature (London)*, **317**, 458–460.

Gay, N. J., Tybulewicz, V. L. J. & Walker, J. E. (1986). *Biochem. J.* **234**, 111–117.

Gill, R., Heffron, F., Dougan, G. & Falkow, S. (1978). *J. Bacteriol.* **136**, 742–756.

Gosti-Testu, F. & Brevet, J. (1982). *CR Acad. Sci. Paris*, **294**, 193–196.

Grindley, N. & Reed, R. (1985). *Annu. Rev. Biochem.* **54**, 863–896.

Gringauz, E., Orle, K. A., Waddell, C. S. & Craig, N. L. (1988). *J. Bacteriol.* **170**, 2832–2840.

Groenen, M. A. M., Timmers, E. & van de Putte, P. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 2087–2091.
Hassan, D. M. & Brevet, J. (1983). *Plasmid*, **10**, 31–44.
Hauer, B. & Shapiro, J. (1984). *Mol. Gen. Genet.* **194**, 149–158.
Heffron, F., McCarthy, B. J., Ohtsubo, M. & Ohtsubo, E. (1978). *Cell*, **18**, 1153–1163.
Heritage, J. & Bennett, P. M. (1984). *Plasmid*, **12**, 218–221.
Huang, C.-J., Heffron, F., Twu, J.-S., Schloemer, R. H. & Lee, C.-H. (1986). *Gene*, **41**, 23–31.
Jayaram, M. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 5875–5879.
Johnson, R. C. & Reznikoff, W. S. (1983). *Nature (London)*, **304**, 280–282.
Johnson, R. C. & Reznikoff, W. S. (1984). *Genetics*, **107**, 9–18.
Lee, C.-H., Bhaqwat, A. & Heffron, F. (1983) *Proc. Nat. Acad. Sci., U.S.A.* **80**, 6765–6769.
Lichtenstein, C. & Brenner, S. (1981). *Mol. Gen. Genet.* **183**, 380–387.
Lichtenstein, C. & Brenner, S. (1982). *Nature (London)*, **297**, 601–603.
Maniatis, T., Fritsch, E. & Sambrook, J. (1982). Editors of *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
McKown, R. L., Waddell, C. S., Arciszewska, L. K. & Craig, N. L. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 7801–7811.
McKown, R. L. Orle, K. A., Chen, T. & Craig, N. L. (1988). *J. Bacteriol.* **170**, 352–358.
Miller, J. (1972). Editor of *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Moore, R. J. & Krishnapillai, V. (1982). *J. Bacteriol.* **149**, 276–283.
Morisato, D., Way, J. C., Kim, H.-J. & Kleckner, N. (1983). *Cell*, **32**, 799–807.
O'Connor, M. B., Kilbane, J. J. & Malamy, M. H. (1986). *J. Mol. Biol.* **189**, 85–102.
Ouartsi, A., Borowski, D. & Brevet, J. (1985). *Mol. Gen. Genet.* **198**, 221–227.
Reed, R. R., Shibuya, G. J. & Steitz, J. A. (1982). *Nature (London)*, **300**, 381–383.
Reyes, O., Beyou, A., Mignotte-Vieux, C. & Richaud, F. (1987). *Plasmid*, **18**, 183–192.
Robinson, M. K., Bennett, P. M. & Richmond, M. H. (1977). *J. Bacteriol.* **129**, 407–414.
Rogers, M., Ekaterinaki, N., Nimmo, E. & Sherratt, D. (1986). *Mol. Gen. Genet.* **205**, 550–556.
Sasakawa, C., Carle, G. F. & Berg, D. E. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 7293–7297.
Sherratt, D. J., Arthur, A. & Burke, M. (1981). *Cold Spring Harbor Symp. Quant. Biol.* **45**, 275–281.
Sherratt, D. J., Arthur, A., Bishop, R., Dyson, P., Kitts, P. & Symington, L. (1983). In *Genetic Rearrangements* (Chater, K. F., Cullis, C. A., Hopwood, D. A., Johnston, A. W. B. & Woolhouse, H. W. eds), pp. 59–74. Sinauer Associates Inc., Sunderland, MA.
Simons, R. W. & Kleckner, N. (1983). *Cell*, **34**, 683–691.
Simonsen, C., Chen, E. & Levonson, O. (1983). *J. Bacteriol.* **155**, 1001–1008.
Smith, G. M. & Jones, P. (1984). *J. Bacteriol.* **157**, 962–964.
Struhl, K. (1983). *Gene*, **26**, 231–242.
Waddell, C. S. & Craig, N. L. (1988). *Genes Develop.* **2**, 137–149.
Walker, J. E., Gay, N. J., Saraste, M. & Eberle, A. N. (1984). *Biochem. J.* **224**, 799–815.
Wallace, L. J., Ward, J. M. & Richmond, M. H. (1981). *Mol. Gen. Genet.* **184**, 87–91.
Way, J. C. & Kleckner, N. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 3452–3456.
Weinert, T. A., Derbyshire, K. M., Hughson, F. M. & Grindley, N. D. F. (1984). *Cold Spring Harbor Symp. Quant. Biol.* **49**, 251–260.
Wells, R. G. & Grindley, N. D. F. (1984). *J. Mol. Biol.* **179**, 667–687.
Wishart, W. D., Machida, C., Ohtsubo, H. & Ohtsubo, E. (1983). *Gene*, **24**, 99–113.

Edited by M. Gottesman

CHAPTER III

Purification and characterization of TnsB, a transposition protein that binds to the ends of Tn7

SUMMARY

We have purified TnsB, a transposition protein encoded by the bacterial transposon Tn7. The purification procedure involves three chromatographic steps (DNA-cellulose, norleucine-sepharose and phosphocellulose) and yields milligram quantities of highly purified protein. The apparent molecular weight of denatured TnsB protein is approximately 85 kDa. Gel filtration chromatography and sucrose gradient sedimentation studies indicate that, in solution, native TnsB is a monomer of non-spherical shape. Using DNase I protection analysis, we have established that TnsB is a sequence-specific DNA binding protein that recognizes multiple sites in both ends of the transposon. The TnsB binding sites, three in the left end of Tn7 and four in the right end, are highly related in nucleotide sequence and are located in DNA segments that we have previously shown contain *cis*-acting sequences important for Tn7 transposition. Our results also show that one of the TnsB binding sites overlaps a proposed promoter for Tn7's transposition genes. These studies suggest that the specific binding of TnsB to the ends of Tn7 mediates recombination and may also regulate the expression of Tn7-encoded transposition genes.

INTRODUCTION

Transposable elements are DNA segments that can move from one genetic location to another. To promote this movement, transposons encode two types of information: transposition proteins, usually one but occasionally more, and *cis*-acting DNA sequences that are directly involved in transposition (for review, see Berg and Howe, 1989). In most transposons, these sequences are confined to the termini of the element. A key step in transposition is the recognition of the recombination sequences at the ends by transposition protein(s). The interactions between transposition proteins and the transposon ends have been analyzed at the biochemical level for only a few transposons (Craigie et al., 1984; Groenen et al., 1987; Ichikawa et al., 1987; Zerbib et al., 1987; Gierl et al., 1988; Wiater and Grindley, 1988; Kunze et al., 1989). We are interested in understanding the protein-DNA interactions that mediate the movement of the bacterial transposon Tn7 (Barth et al., 1976; for review, see Craig, 1989).

Tn7 is unique among transposable elements because it codes for five transposition genes, *tnsABCDE*[1], that mediate two transposition pathways (Rogers et al., 1986; Waddell and Craig, 1988; Kubo and Craig, 1990). *tnsABC + tnsD* mediate high-frequency transposition to a specific site in

the Escherichia coli chromosome, attTn7, while tnsABC + tnsE mediate low-frequency transposition to apparently random target sites. Because tnsABC are required in both pathways, the products of these genes are good candidates for recombination proteins that interact directly with the transposon ends.

We have previously established that the cis-acting DNA segments required for both Tn7 transposition pathways lie at the ends of the transposon (Arciszewska et al., 1989). These segments are extensive and differ in length and sequence: about 150 bp at the left end of Tn7 (Tn7L) and about 100 bp at the right end of Tn7 (Tn7R) are necessary for efficient transposition. The Tn7 ends are also functionally distinct: miniTn7 elements containing two Tn7R ends are active in transposition while those containing two Tn7L ends are not. As in many other transposons, the extreme termini of Tn7 are very similar: the terminal 30 bp of Tn7 are imperfect inverted repeats. In addition, the ends are related by the presence of multiple copies of highly homologous 22 bp repeats that are differently arranged in each end (Lichtenstein and Brenner, 1982). Tn7L contains three directly oriented repeats, separated by two segments of unrelated sequence. Tn7R contains four closely spaced repeats in inverted orientation with regard to those in Tn7L. Our in vivo analysis of the cis-acting transposition sequences at the Tn7 ends indicate that these

22 bp repeats are important for transposition (Arciszewska et al., 1989).

We have previously detected in cell lysates a tnsB-dependent DNA binding activity that specifically recognizes both ends of Tn7 (McKown et al., 1987). We demonstrate here that this activity reflects the specific interaction of TnsB protein with Tn7's cis-acting transposition sequences. We report the purification of TnsB protein and show that it binds to multiple sites in both ends of the transposon; these binding sites encompass the 22 bp repeats. The specific recognition of the Tn7 ends by TnsB is consistent with the requirement for tnsB in both Tn7 transposition pathways (Rogers et al., 1986; Waddell and Craig, 1988).

We have also found that one of the TnsB binding sites overlaps a proposed tns gene promoter (Gay et al., 1986). This observation supports the hypothesis (Rogers et al., 1986) that TnsB can also play a regulatory role in Tn7 transposition by modulating the expression of the Tn7 transposition genes (Waddell and Craig, 1988; Flores et al., 1990; Orle and Craig, 1990).

EXPERIMENTAL PROCEDURES

Bacteria and plasmids - tnsB and tnsABCDE cell lysates were prepared from CAG456 (E. coli K12 $lacZ_{am}$ trp_{am} pho_{am} $supC_{ts}$ rpsL $htpR_{165}$; Baker et al., 1984) carrying the indicated plasmids. TnsB protein was isolated from CAG456 carrying plasmids F'$lacI^{Q}$ lacZ:Tn5 (Sauer et al., 1988) and ptac-tnsB, a pBR322 derivative containing the tnsB gene downstream of the tac promoter. ptac-tnsB was constructed by inserting a 2.5 kb HindIII fragment containing tnsB from pCW12 (McKown et al., 1987) into the PvuII site of ptac12 (Amann et al., 1983) by blunt-end ligation. The tnsB-containing fragment extends from the BglII site approximately 900 bp from the right terminus of Tn7 (flanked by BamHI to HindIII of the pUC19 polylinker) to the HindIII site located approximately 3440 bp from the right terminus of Tn7 (Flores et al., 1990; Waddell and Craig, 1988). tnsABCDE lysates were prepared from CAG456 carrying pCW4, a pACYC184 plasmid containing tnsABCDE (McKown et al., 1987). Host lysates were prepared from CAG456. Cell lysates and protein fractions other than those containing TnsB that were used in the in vitro transposition assay were prepared from strains containing other tns plasmids as described in Bainton et al.,[2].

The Tn7L and Tn7R fragments used in the DNA binding assays were isolated from plasmids containing various Tn7

end segments. The fragment containing R1-199[3] was isolated from pKS$^+$Ra, which contains the EcoRI-BamHI Tn7R fragment of pLA28 (Arciszewska et al., 1989) inserted between the EcoRI and BamHI sites of Bluescript-KS (Stratagene). The fragment containing L1-166 was isolated from pKS$^+$La, which contains the EcoRI-BamHI Tn7L fragment of pLA26 (Arciszewska et al., 1989) inserted between the EcoRI and BamHI sites of Bluescript-KS. The fragment containing L109-166 was isolated from pLA77, which was constructed by inserting an AluI-HincII fragment, obtained by digestion of the EcoRI-HincII Tn7L fragment from pLA26, into EcoRV of Bluescript-KS; in pLA77, position L109 is adjacent to the vector EcoRI site. Plasmids used as substrates in the *in vitro* transposition assay are described elsewhere (Bainton et al.[2]).

Preparation and labeling of DNA fragments - Plasmid DNA was digested with appropriate restriction enzymes and electrophoresed through 6% polyacrylamide gels. Slices containing the DNA fragments of interest were cut out from the gels, the DNA electrophoretically transfered onto a DEAE membrane (Schleicher and Schuell) and recovered as suggested by the manufacturer. Purified DNA fragments were labeled at their 3' ends using DNA polymerase I Klenow fragment and appropriate [α-P^{32}]dNTPs (Tabor and Struhl, 1987). The EcoRI-BamHI fragment (294 bp) from pKS$^+$Ra

containing R1-199 was used for analysis of Tn7R; this fragment was labeled at EcoRI for analysis of the top strand[4] and at BamHI for analysis of the bottom strand. Fragments from pKS$^+$La were used for analysis of L1-166, the SmaI-HindIII (270 bp) fragment labeled at HindIII for the top strand and the EcoRI-BamHI (258 bp) fragment labeled at EcoRI for the bottom strand. For analysis of L109-166, the SmaI-HindIII (88 bp) fragment of pLA77 labeled at HindIII was used.

DNA binding assays - The band shift method (Fried and Crothers, 1981; Garner and Revzin, 1981) was used to evaluate the specific binding of TnsB fractions to Tn7 end fragments. Prior to assay, TnsB fractions were diluted in P-0.5 M NaCl buffer (see below) containing 1 mg/ml BSA (Fraction V - Boehringer Mannheim) (except in the experiments in Table 1 where TnsB fractions were diluted in host cell lysates prepared from strains lacking TnsB). For measurement of DNA binding activity, the TnsB fraction was sufficiently diluted that only a small fraction of the end fragment was bound by TnsB, conditions where the amount of fragment bound was proportional to TnsB concentration.
1 µl of diluted TnsB was included in 20 µl binding reactions. The binding reactions contained 13.8 mM Tris.HCl (pH 8.0), 1.05 mM EDTA, 2.05 mM DTT, 125 mM NaCl, 10.5% (v/v) glycerol, 350 µg/ml poly(dI-dC).poly(dI-dC),

1.05 mg/ml BSA, and approximately 0.15 pmol of end-labeled DNA fragment. After incubation for 7 min at room temperature, the reaction mixtures were electrophoresed through 6% polyacrylamide gels (29:1 acrylamide/N,N'methylenebisacrylamide) in Tris - Borate - EDTA buffer at 19 V/cm for 90 min. After drying, the gels were exposed to X-ray film.

Buffers - CL buffer was 150 mM Tris.HCl (pH 8.0), 0.1 M NaCl and 1 mM EDTA; P buffer was 25 mM Tris.HCl (pH 8.0), 1 mM DTT, 1 mM EDTA, 10% (v/v) glycerol and NaCl as indicated. S buffer was P-0.5 M NaCl but contained 25% glycerol.

Cell growth and recovery - Bacterial cultures used for protein purification were grown in LB (Miller, 1972) supplemented with carbenicillin at 100 μg/ml in a 60 L fermentor at 30°C. At OD_{600} 1.2, the cells were induced by the addition of isopropyl-β-D-thiogalactoside to 230 μg/ml and grown for an additional 2 h (final OD_{600} 2.0). The cells were harvested by centrifugation, washed in CL buffer, spun again, divided into approximately 40 g portions and frozen in liquid nitrogen. Cultures for small volume lysates were grown at 30°C without induction, harvested in exponential phase, washed and frozen.

tnsABCDE and host cell lysate preparation - CL buffer was added to frozen cell pellets (3 ml per 1 g cells) and lysis was achieved through a cycle of freezing and thawing in the presence of lysozyme at 200 μg/ml. Cell debris was removed by a 20 min centrifugation at 48000 x g. During the preparation of tnsABCDE lysate, DNA was removed by DNase I treatment using conditions as described for TnsB purification and the collected lysate was dialyzed against P-0.1 M NaCl buffer. Cell lysates and protein fractions used for the *in vitro* transposition assay were prepared as described elsewhere (Bainton et al.[2]).

TnsB purification - Unless otherwise indicated, all steps were carried out at 4°C.

- **Cleared cell lysate preparation** - 40 g of frozen cell pellet was thawed in the presence of 1 volume CL buffer (1 ml per 1 g cells). $MgCl_2$ (10 mM), $CaCl_2$ (2 mM) and pancreatic DNase I (10 μg/ml) were added to the cell suspension. The cells were disrupted by a single passage through a French press at 400 atm. After addition of another two volumes of CL buffer, the crude extract was centrifuged for 45 min at 48000 x g. The concentrations of $MgCl_2$ (10 mM) and $CaCl_2$ (2 mM) were readjusted in the collected supernatant, and another portion of DNase I (10 μg/ml) was added, followed by incubation at 16°C for 30 min. The cell lysate was dialyzed extensively against P-

0.1 M NaCl buffer at 0°C with three 6-h changes; the first change also contained 20 mM EDTA. After centrifugation at 90000 x g for 80 min, a clear cell lysate was collected (Fraction I).

- **DNA-Cellulose chromatography** - Fraction I was applied to a double-stranded DNA-cellulose column (40 ml packed volume equilibrated with P-0.1 M NaCl buffer) at a flow rate of 8 ml/h. After a 160 ml wash with P-0.1 M NaCl buffer (flow rate 20 ml/h), the column was eluted with a 10 column volume linear gradient of P-0.1 to 1.4 M NaCl buffer (flow rate 40 ml/h). tnsB-dependent binding activity eluted from the column between 0.38 - 0.51 M NaCl, coincided with the peak of 280 nm absorbance (Fig. 1A) and comigrated with the 85 kDa TnsB polypeptide on SDS-PAGE (data not shown). Fractions that contained the majority of the binding activity but the fewest contaminating polypeptides were pooled to give Fraction II.

- **Norleucine-Sepharose chromatography** - Solid ammonium sulfate was added to Fraction II to 20% saturation. After centrifugation at 90000 x g for 30 min, the supernatant was applied to a norleucine-sepharose column (8 ml packed volume, equilibrated with P-0.5 M NaCl + 20% $(NH_4)_2SO_4$ at a flow rate of 4 ml/h. The column was washed with 18 ml P-0.5 M NaCl + 20% $(NH_4)_2SO_4$ and then eluted with 48 ml of P-0.5 M NaCl + 10.3% $(NH_4)_2SO_4$ at a flow rate of 16 ml/h (Fig. 1B). Fractions that contained the majority of the

binding activity but the fewest contaminating polypeptides were pooled to form Fraction III.

- **Phosphocellulose chromatography** - Fraction III was diluted two-fold with P-0.5 M NaCl (to avoid protein precipitation upon dialysis into low salt buffer), dialyzed against P-0.25 M NaCl at 0°C (two 6 h changes), and applied to a phosphocellulose column (5 ml packed volume, equilibrated with P-0.25 M buffer) at a flow rate of 5 ml/h. The column was washed with 15 ml of P-0.25 M NaCl buffer and then eluted step-wise with 15 ml of P-0.39 M NaCl buffer and then with 15 ml of P-0.52 M NaCl buffer. TnsB eluted from the column at the second step (Fig. 1C). Fractions with the highest protein concentration were pooled to form Fraction IV.

- **Storage** - Purified protein was dialyzed against S buffer, divided into aliquots, frozen in liquid nitrogen and stored at -80°C.

- **Resins** - DNA-Cellulose was prepared from CF11 cellulose (Whatman) and calf thymus DNA (Pharmacia) as described by Alberts and Herrick, 1971. Norleucine-sepharose was prepared as described in Morris et al., 1979. Phosphocellulose (Whatman) was precycled as suggested by the manufacturer.

Nuclease test - Circular single-stranded M13mp18(+) phage DNA (5 μg/ml) and supercoiled Bluescript KS^+ plasmid DNA (3.5 μg/ml) were used as substrates in 50 μl reactions in 15 mM Tris.HCl (pH 8.0), 1.1 mM EDTA, 2.1 mM DTT, 150 mM NaCl, 15 mM $MgCl_2$, 11% (v/v) glycerol and 0.8 mg/ml BSA. TnsB (final concentration of Fraction IV was 5.5 μg/ml and of Fraction III was 2.0 μg/ml) were incubated with DNA for 15 min at 30°C. The reactions were stopped by addition of 2 ml of 0.2 M EDTA. DNA was subjected to electrophoresis on 0.7% agarose gels and visualized by ethidium bromide staining.

Gel filtration chromatography - A Sephadex G-150 column (98 cm x 1.6 cm) equilibrated with 25 mM Tris.HCl (pH 8.0), 1 mM DTT, 1 mM EDTA and 0.5 M NaCl was run at a flow rate of 4 ml/h. Protein standards (Pharmacia) at 2 mg/ml and TnsB (Fraction IV) at 40 μg/ml were run separately; the elution volumes for the standards and TnsB were determined by A_{280} and DNA binding activity, respectively.

Sucrose gradient sedimentation - 5 ml 5 - 20% (w/v) sucrose gradients were prepared in 25 mM Tris.HCl (pH 8.0), 1 mM DTT, 1 mM EDTA and 0.5 M NaCl. Protein standards (Pharmacia) at 2 mg/ml and TnsB (Fraction IV) at 140 μg/ml were layered onto the gradients and centrifuged in a 50.1 (Spinco) rotor at 40000 rpm for 15 h. Fractions were

collected from the bottom of the tubes. Positions of the protein standards were determined by protein content and the position of TnsB was determined by evaluation of DNA binding activity.

DNase I protection experiments - Protection assays were performed using the procedure of Galas and Schmitz, 1978. The reactions (100 µl) contained 15 mM Tris.HCl (pH 8.0), 0.1 mM EDTA, 1.1 mM DTT, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 11% (v/v) glycerol, 2 ng/ml poly(dI-dC).poly(dI-dC), 250 µg/ml BSA and approximately 0.15 pmol of 3' end-labeled DNA fragments. After incubation for 20 min at room temperature in the presence of TnsB, 10 µl of 5 µg/ml pancreatic DNase I (Worthington) was added and incubation continued for 2 min. Digestion was stopped by addition of 100 µl 0.65 M ammonium acetate, 0.1 M EDTA and 50 µg/ml of poly(dI-dC).poly(dI-dC). DNA was precipitated with 3 volumes of ethanol, washed in 0.3 M sodium acetate, reprecipitated and resuspended in formamide gel loading solution. DNA sequencing was performed using the chemical degradation method (Maxam and Gilbert, 1980).

in vitro transposition assay - in vitro assays were carried out as described by Bainton et al.[2] using Condition 2. Reactions (100 µl) were performed in 26 mM Hepes (pH 7.5), 2 mM Tris.HCl (pH 7.5), 2.5 mM KPO_4, 128 mM KCl, 15 mM

NaCl, 0.1 mM EDTA, 2.2 mM DTT, 0.8% (v/v) glycerol, 15 mM MgAc, 2 mM ATP and 5% polyvinyl alcohol in the presence of 100 ng of donor plasmid containing a miniTn7 element (pEM) and 2 μg of an attTn7 target plasmid. The reactions also contained approximately 40 μg tnsA cell lysate, 17 ng TnsB (Fraction IV), 10 μg tnsC cell lysate and 4 μg tnsD fraction (Stage III). After preincubation in the absence of MgAc for 7 min at 30°C, the reactions were incubated for 15 min at 30°C, stopped and the DNA purified from the other reaction components. Recovered DNA was linearized by restriction enzyme digestion, electrophoresed through 0.7% agarose, transferred to Nytran (Schleicher and Schuell) and analyzed by Southern hybridization using as a probe a DNA fragment specific for the miniTn7 element.

Other methods - Western analysis and preparation of anti-TnsB antibodies were done as described elsewhere (Orle and Craig, 1990). SDS-PAGE was carried out as described in Smith (1987). Protein content was assayed by the Bradford method (Bradford, 1976). The amino acid sequence of the N-terminus of TnsB was determined by the Protein Structure Laboratory, HSRL, Davis, CA on an AB1470 Gas Phase Sequencer.

RESULTS

We have previously reported (McKown et al., 1987) that crude *E. coli* lysates prepared from cells carrying Tn7 or a *tnsABCDE* plasmid contain a *tnsB*-dependent DNA binding activity that specifically recognizes the ends of Tn7. Here we report the fractionation of this binding activity and show that TnsB is a sequence-specific DNA binding protein that recognizes the ends of Tn7.

In our previous studies of the *tnsB*-dependent DNA binding activity, we used crude lysates from cells containing all the *tns* genes. We began our current characterization of the *tnsB*-dependent binding activity by establishing that *tnsB* is the only transposon-encoded gene required for specific end recognition as judged by the band shift assay method (Fig. 4; Fried and Crothers, 1981; Garner and Revzin, 1981). The *tnsB*-dependent binding activity produces several protein-DNA complexes when the intact Tn7 ends are used as substrates which presumably reflect the presence of several protein binding sites in each end (McKown et al., 1987). To simplify the assay for the purpose of monitoring the purification of TnsB, we selected a DNA fragment that contains only a single binding site: the Tn7L fragment containing L109-166 (see below). Our previous *in vivo* studies have shown that this region is critical for transposition (Arciszewska et al., 1989).

Purification of TnsB

TnsB protein was purified from E. coli containing a pBR322 derivative in which tnsB is located downstream of an inducible tac promoter. As a host, we used an htpR$^-$ strain that is known to be defective in proteolysis (Baker et al., 1984), because we observed substantial proteolytic degradation of TnsB in htpR$^+$ hosts (data not shown; Orle and Craig, 1990). Fractionation of the tnsB-dependent binding activity was monitored by band shift assays using the L109-166 DNA fragment as a substrate and protein composition was followed by SDS-PAGE. The purification procedure (described in detail in Experimental Procedures) involves three chromatography steps and yields about 2.5 mg of highly purified TnsB from 40 g of cells. The tnsB-dependent DNA binding activity and the 85 kDa TnsB polypeptide (Orle and Craig, 1990) copurified at all steps during fractionation. The results of a typical preparation are presented in Table I.

The cleared cell lysate (Fraction I) was prepared by cell disruption with a French press, followed by DNase I digestion to remove DNA which could interfere with the subsequent chromatographic steps. The amount of TnsB in cell lysates as estimated by Western analysis, was less than 0.5% of the total protein (data not shown). (Even in lysates from htpR$^-$ strains, several TnsB species were

evident in Fraction I, an intact 85 kDa polypeptide and several smaller species that likely result from proteolytic degradation.) Fraction I was applied to a double-stranded DNA-cellulose column and the bound proteins, which included TnsB, were eluted with a linear salt gradient. The fractions containing TnsB were pooled to give Fraction II. A substantial purification of TnsB occured at this step: there was a striking decrease in the number of contaminating polypeptides (Fig. 2A, compare lanes 1 and 2) that was accompanied by a greater than 90-fold increase in specific binding activity (Table I). After addition of ammonium sulfate to 20% saturation, Fraction II was subjected to hydrophobic chromatography on norleucine-sepharose. TnsB bound to the column and was eluted with 10% ammonium sulfate buffer. Fractions containing TnsB were pooled to yield Fraction III. This fraction contained only a few polypeptides: 85 kDa TnsB, several smaller TnsB-derived species ranging in size between 55 and 85 kDa and traces of other contaminants (Fig. 2A, lane 3). These contaminants were removed by a final chromatography step in which Fraction III was applied to a phosphocellulose column. TnsB bound to the column in low salt and was eluted with a high salt buffer. Fractions containing TnsB were pooled to form Fraction IV (Fig. 2A, lane 4). This purification procedure yields mg quantities of nearly homogenous TnsB protein with recovery of about 20% of the

original binding activity present in the cleared cell lysate. The specific DNA binding activity, as measured by band shift assays, increased about 150-fold (Table I).

The predominant species in Fraction IV is an 85 kDa polypeptide that is intact TnsB protein (Fig. 2A; Orle and Craig, 1990). However, several smaller polypeptides, ranging in size from 55 to 85 kDa, can also be seen. These smaller polypeptides are specifically recognized by anti-TnsB antibodies as demonstrated by Western analysis (Fig. 2B). We estimate that at least 90% of Fraction IV is intact TnsB polypeptide. Our observations suggest that the shorter TnsB-derived polypeptides result from proteolytic degradation (data not shown).

We also evaluated the purity of the TnsB fractions by testing for the presence of non-specific single- and double-stranded DNA endonucleases; the DNA substrates in these assays lacked Tn7 sequences. No endonucleolytic activity was detectable in Fraction IV or even Fraction III upon incubation of these TnsB fractions with DNA at protein concentrations several-fold higher than those used in other biochemical experiments (see Experimental Procedures).

Properties of TnsB

- **N-terminal amino acid sequence** - The amino acid sequence of the first 12 amino acids of Fraction IV TnsB protein was determined to be: X-X-Gln-Ile-Asn-Glu-Val-Val-Leu-Phe-Asp-

Asn. A nucleotide sequence encoding this amino acid sequence is present at the 5' end of the tnsB gene (Flores et al., 1990; Orle and Craig, 1990).

- Native Molecular Weight - The apparent molecular weight of denatured TnsB is approximately 85 kDa as determined by SDS-PAGE (Fig. 2A; Orle and Craig, 1990). This value is in good agreement with the value of 81 kDa calculated from the DNA sequence of the gene (Flores et al., 1990). Characterization of the native molecular weight of TnsB by gel filtration chromatography and sucrose gradient sedimentation analysis suggests that TnsB is a monomer in solution.

The observed Stokes radius of TnsB, as determined by gel filtration chromatography, is 5.4 nm (Fig. 3A). This value corresponds to a spherical protein of approximately 190 kDa. However, the sedimentation coefficient of TnsB as determined by sucrose gradient sedimentation is 4.3s (Fig. 3A). This value is similar to that observed for 67 kDa bovine serum albumin. Because different values of molecular weight were obtained with these methods, the native molecular weight of TnsB was calculated from the equation (Siegel and Monty, 1966):

$$M = 6 \pi \eta N a s / (1 - v \rho)$$

where a = the Stokes radius of the protein (nm), s = its sedimentation coefficient, v = the partial specific volume

(0.725 cm^3/g assumed), η = the viscosity of the medium, ρ = the density of the medium, and N = Avogadro's number. M computed from this equation is 95.7 kDa. The calculated frictional ratio, f/f_0, (Siegel and Monty, 1966) is equal to 1.8. These results suggest that TnsB is monomeric in solution and that it is a protein of markedly non-spherical shape.

TnsB binds specifically to the ends of Tn7

We used the band shift assay to examine the interaction of purified TnsB protein with the ends of Tn7 (Fig. 4). Several discrete complexes are observed with DNA fragments containing Tn7L (panel A) and Tn7R (panel B). The TnsB-DNA complexes (Fig. 4, A,B - lane 3) are identical in mobility to complexes formed with crude cell lysates from a tnsABCDE strain (Fig. 4, A,B - lane 1) and a tnsB strain (Fig. 4, A,B - lane 2). These results suggest that the specific interaction of TnsB with DNA underlies the tnsB-dependent complexes observed with the crude cell lysates. As considered in more detail below, both the Tn7L and Tn7R DNA segments contain several TnsB binding sites. A reasonable explanation for the several TnsB-DNA complexes observed with both Tn7L and Tn7R is that they reflect various distributions of TnsB among the multiple TnsB binding sites. Direct analysis of these complexes provides experimental support for this view (L.K.A. and N.L.C,

manuscript in preparation). Although our results indicate that TnsB is a monomer in solution, we have no information about the oligomeric state of TnsB at each TnsB binding site.

When a fragment of Tn7L containing a single TnsB binding site, L109-166, is used as a substrate in a band shift assay with crude tnsABCDE lysate, tnsB lysate and purified TnsB (Fig. 4C), a single major complex (marked "a") of retarded mobility is observed. The fact that complexes of identical mobility are observed with all the protein sources supports the view that TnsB mediates the tnsB-dependent DNA binding activity. Close inspection of the L109-166 complexes reveals the presence of several minor species (marked "b" and "c"). We suspect that these complexes contain smaller TnsB polypeptides that result from proteolytic degradation. Our TnsB preparation contains smaller TnsB species, and these "minor" TnsB-DNA complexes increase in amount when TnsB fractions containing more of the shorter TnsB polypeptides are used in such assays (data not shown). The major TnsB-derived polypeptides that we have observed (apparent molecular weight between 55 to 85 kDa, Fig. 2A) copurify with intact TnsB, making preparation of the homogenous, intact protein difficult.

Several lines of evidence suggest that TnsB protein binds directly to DNA. TnsB is, by far, the major protein

in our purified fraction (Fig. 2A). Moreover, we have observed that the TnsB polypeptide copurifies with the specific Tn7 end binding activity throughout the purification procedure. As shown in Figure 5, for example, the amount of shifted complex detected using the L109-166 fragment as a substrate parallels the amount of TnsB polypeptide when fractions resulting from the stepwise elution of TnsB from a phosphocellulose column are examined in a band shift assay. The binding activity of purified TnsB protein is not different when assayed in the presence or absence of host crude lysate (data not shown) and, as described above, complexes of apparently identical mobility are observed with tnsB-containing crude lysates and purified TnsB protein. Therefore, if a host protein is important for the end binding activity we have characterized, this host protein must also be present in our purified fractions. Although we have no obvious candidate for this host protein, it is difficult to exclude the possibility that it is present but not readily detected by the staining methods we have used. In addition, although this is unlikely, trace amounts of an enzymatic activity that modifies TnsB could be present in our purified preparations (Prives, 1990).

Identification of TnsB Binding Sites by DNase I protection analysis

To directly identify the positions of TnsB binding within the ends of Tn7, we carried out DNase I protection experiments with end-labeled DNA fragments containing Tn7L or Tn7R. The results of this analysis are shown in Fig. 6 and summarized in Fig. 7.

In Tn7L, three regions of protection, designated α, β and γ, were observed in the presence of TnsB. These protected regions are each about 30 bp in length and are separated by 20-30 bp of unprotected sequences. We note that the L109-166 segment used in our band shift assays encompasses a single region of TnsB protection. Also evident are positions of enhanced sensitivity to DNase I on both strands within the protected regions. Alignment of the sequences of the protected regions (Fig. 8) reveals that they are highly related in nucleotide sequence and that the positions of enhanced DNase I cleavage are similarly located within each protected region. The α, β and γ regions each encompass a copy of the highly conserved 22 bp sequence that is repeated several times in both ends of Tn7 (Lichtenstein and Brenner, 1982). These results are consistent with our previous observation (McKown et al., 1987) that a DNA fragment containing a single 22 bp repeat is specifically recognized by the tnsB-dependent binding activity in crude cell lysates.

In Tn7R, one continuous region of protection, approximately 90 bp in length, was observed. We propose that this region contains four TnsB binding sites designated ϕ, χ, ψ and ω. The presence of four TnsB binding sites in this region is consistent with the nucleotide sequence of the protected region, i.e. it contains four sequences similar to the TnsB binding sites in Tn7L, including the presence of four 22 bp repeats (Fig. 8). Furthermore, several positions of enhanced DNase I cleavage are evident within the Tn7R protected region whose pattern corresponds to the pattern of DNase I cleavage enhancements in the Tn7L binding sites. However, several slight deviations from the conserved positioning of the enhanced DNase I cleavage sites are observed: in ϕ, the enhancement is shifted by 1 bp on one strand and in ψ, partial protection, rather than enhancement, is observed on the same strand. These irregularities may reflect slightly different interactions of TnsB with the ϕ and ψ sites whose nucleotide sequences are very similar to each other and distinct from the other TnsB sites, or they may be caused by the close apposition of the TnsB binding sites in Tn7R.

Comparison of the seven proposed TnsB binding sites reveals that they share considerable sequence similarity (Fig. 8). The 22 bp repeats are prominent features of the TnsB binding sites, although some sequence similarities extend beyond the repeats. It is important to note that,

because of the apparent overlapping of the TnsB binding sites in Tn7R, the identity of some nucleotides may have multiple constraints. Thus, firm identification of the actual recognition determinants for TnsB binding must await mutational analysis.

An interesting feature of the TnsB binding sites is the presence in each site of 4 bp imperfect inverted repeats that flank a 3 bp region (Fig. 8, positions 21-23). These 3 bp regions are bracketed by the positions of increased DNase I cleavage. Interestingly, the most prominent sequence differences between the binding sites in Tn7L (α, β, γ) and in Tn7R (ϕ, χ, ψ, ω) occur within these 3 bp sequences (Fig. 8): the sites in the left end contain KGG while those in the right contain NTA, where K is T or G and N is any nucleotide. The functional significance of these structural differences remains to be determined.

Another feature of the DNase I protection patterns invites comment. The pattern of TnsB protection at the extreme termini of Tn7 appears to be slightly different than that at the internal binding sites. Two positions of considerably enhanced sensitivity to DNase I cleavage, adjacent to the 5' ends of the transposon, are evident at the edges of the terminal α and ω binding sites (Figs. 6 and 7). Such enhancements are not obvious at comparable positions in the β and γ sites. (In Tn7R, comparable positions in the internal TnsB sites cannot be clearly

evaluated because of the considerable overlap of these binding sites). The enhancements at the terminal sites may represent a distinctive interaction between TnsB and the transposon ends. The sequences of the α and ω sites are slightly different from the other TnsB sites; an obvious difference is that they are very GC rich between position 1 and position 8 (Fig. 8). These terminal DNase I cleavages could reflect a change in DNA structure promoted by the binding of TnsB to these sites (Gartenberg and Crothers, 1988). Unique interactions of TnsB with the terminal α and ω binding sites could play an important role in defining the junctions between each transposon end and flanking donor site DNA.

Activity of TnsB in Tn7 transposition in vitro

During purification, the activity of TnsB was measured by this protein's ability to bind specifically to the ends of Tn7. We have also tested the activity of purified TnsB in Tn7 transposition. For this purpose, we used a cell-free transposition system (Bainton et al.[2]) in which a miniTn7 element transposes from a donor plasmid to attTn7 in a target plasmid in the presence of lysates derived from cells carrying tnsABC + tnsD which are the genes required for in vivo Tn7 insertion into attTn7 (Rogers et al., 1986; Waddell and Craig, 1988). Experiments by Gamas et al. [5] have shown that partially purified TnsB (Fraction II) can

provide the tnsB function required in this system for in vitro Tn7 transposition.

We have determined that highly purified TnsB protein is active in transposition in vitro (Fig. 9). Comparison of the in vitro transposition activity of various dilutions of tnsB cell lysates and purified TnsB protein (Fraction IV) suggests that the transposition activity of the TnsB polypeptide is similar before and after purification. Both sources provided similar levels of transposition at equivalent concentrations of TnsB (data not shown). Thus, our purification procedure yields TnsB that is active both in sequence-specific DNA binding and in Tn7 transposition.

DISCUSSION

In this paper, we have reported the purification of the Tn7-encoded transposition protein TnsB. The purification is relatively simple and yields highly purified protein in quantities sufficient for most biochemical studies. Two pieces of evidence demonstrate that the polypeptide we have purified is the product of the tnsB gene: 1) the purified polypeptide is specifically recognized by anti-TnsB antibodies and 2) the sequence of the first twelve N-terminal amino acid residues of the purified polypeptide is in agreement with that predicted from the nucleotide sequence of the tnsB gene (Flores, et al., 1990; Orle and Craig, 1990). Our work has determined directly that the TnsB coding sequence begins at the first of three possible initiation codons that are present in the 5'- region of the gene. The TnsB coding sequence overlaps the end of the coding sequence of the preceding gene, tnsA (Flores et al., 1990; Orle and Craig, 1990), suggesting the possibility that expression of these genes may be translationally coupled.

We have established here that TnsB is a sequence-specific DNA binding protein that recognizes sequences at the ends of Tn7 that are required for transposition. The involvement of TnsB in transposon end recognition is

consistent with the requirement for tnsB in all Tn7 transposition reactions (Rogers et al., 1986; Waddell and Craig, 1988). The DNase I protection studies described here show that TnsB binds specifically to several sites in each end of Tn7; three binding sites were found in Tn7L and four are proposed in Tn7R. These binding sites lie within DNA segments we have established in genetic studies (Arciszewska et al., 1989) to be directly involved in and required for transposition. These studies also suggested that Tn7's transposition sequences would contain multiple protein binding sites. The binding of a transposition protein to multiple sites at the ends of the element has also been observed in bacteriophage Mu (Craigie et al., 1984) and several plant transposons (Gierl et al., 1988; Kunze and Starlinger, 1989). The presence of several binding sites at each transposon end suggests the possibility that TnsB may be involved in the formation of a higher-order nucleoprotein structure during the transposition reaction (Pato, 1989; Stark et al., 1989; Thompson and Landy, 1989).

The TnsB binding sites observed by protection against DNase I attack in Tn7L are about 30 bp in length; in Tn7R, the four proposed sites are closely juxtaposed and are not individually distinct. Comparison of the sequences of the sites in Tn7L and proposed sites in Tn7R reveals that they share considerable sequence similarity (Figure 8). Each of

the proposed TnsB binding sites includes the 22 bp repeat sequence originally noted by Lichtenstein and Brenner (1982) upon their sequencing of the ends of Tn7. The actual contribution of particular nucleotides within and outside these repeats to TnsB binding and function can only be determined through mutational analysis. A prominent feature of the interaction of TnsB with all the Tn7 end binding sites is the presence of positions of enhanced DNase I sensitivity at comparable positions within each binding site (Figure 7). This suggests that a common mode of TnsB-DNA interaction, perhaps a local distortion in DNA structure, occurs at all sites. Internal enhancements in protein binding sites have also been observed with Tn3 transposase (Ichikawa et al., 1987) and MuA protein (Craigie et al., 1984). Furthermore, there seems to be unique interaction between TnsB and the very termini of Tn7 as reflected by additional, distinctive DNase I cleavage patterns at these positions.

A key issue in unraveling the mechanism of a transposition reaction is understanding the DNA strand cleavages that disconnect the transposon from flanking DNA in a donor site and identification of the polypeptide that mediates this step. The edges of two of the TnsB binding sites, α in Tn7L and ω in Tn7R (as defined by protection against DNase I attack) are adjacent to the transposon termini. The junctions of the 3' transposon ends with

flanking donor DNA are very close to (or, in Tn7L, perhaps at) the edges of the TnsB-protected region. By contrast, positions within the transposon near the junctions of the 5' transposon ends with the flanking donor DNA are hypersensitive to DNase I attack. We note that the transposition protein MuA does not obviously protect the very terminal nucleotides of bacteriophage Mu on linear DNA fragments against DNase I cleavage (Craigie et al., 1984) and yet it does perform strand cleavage at these positions during transposition (Craigie and Mizuuchi, 1987; Surrette et al., 1987). The role of TnsB in cutting the transposon ends away from the donor site remains to be determined. We have been unable to detect a TnsB strand cleavage activity (data not shown). Perhaps certain interactions of TnsB with the termini would be evident only on an authentic transposition substrate, i.e. a supercoiled DNA molecule containing both Tn7 ends.

It is important to note, however, that TnsB may not be the only protein that acts at the ends of Tn7. We know that two other Tn7-encoded genes, <u>tnsA</u> and <u>tnsC</u>, are required in addition to <u>tnsB</u> in all transposition reactions (Rogers et al., 1986; Waddell and Craig, 1988) and we suspect that the products of these genes participate directly in transposition (Gamas et al.[5]). It is possible that TnsA or TnsC may mediate the strand breakages at the transposon termini. One role of TnsB binding adjacent to

the transposon termini may be to facilitate the interaction of TnsA, TnsC or even a host protein with the ends. The ends of many transposable elements are known to contain binding sites for several proteins (reviewed in Berg and Howe, 1989).

In addition to its direct role in transposition, TnsB also appears to play a regulatory role by controlling the expression of at least two transposition proteins, TnsA and itself. *tnsA* and *tnsB* form an operon whose transcription initiates near the right terminus of Tn7 (Waddell and Craig, 1988; Gay et al., 1986). The -35 sequence of the proposed *tnsAB* promoter is embedded within the ϕ TnsB binding site (Fig. 7). The binding of TnsB to this site could explain the observation that the expression from this promoter is negatively regulated in the presence of *tnsB* (Rogers et al., 1986).

We have observed considerable degradation of TnsB by host proteases. This proteolysis, which is markedly diminished in an *htpR*$^-$ strain, is manifested by the presence of several shorter TnsB-derived polypeptides. The relevance of such TnsB instability to Tn7 transposition is unclear; we observe TnsB degradation only when this protein is expressed at high level in the absence of the other *tns* genes (Orle and Craig, 1990). Some of the shorter species retain specific DNA binding activity but are inactive in Tn7 transposition *in vitro* (data not shown). Transposase-

related polypeptides are known to play inhibitory roles in regulating transposition in other systems, such as Tn5 (Johnson and Reznikoff, 1984), IS1 (Zerbib et al., 1990) and P elements (Robertson and Engels, 1989; Misra and Rio, 1990). Future studies will be required to reveal the role(s) that TnsB and its derivatives may play in both mediating Tn7 transposition and modulating its frequency.

REFERENCES

Alberts, B., and Herrick, G. (1971) Methods Enzymol. 21, 198-217.

Amman, E., Brosius, J., and Ptashne, M. (1983) Gene 25, 167-178.

Arciszewska, L.K., Drake, D., and Craig, N.L. (1989) J. Mol. Biol. 207, 35-52.

Baker, T.A., Grossman, A.D., and Gross, C.A. (1984) Proc. Natl. Acad. Sci. USA 81, 6779-6783.

Barth, P.T., Datta, N., Hedges, R.W., and Grinter, N.J. (1976) J. Bacteriol. 125, 800-810.

Berg, D.E., and Howe, M. (1989) (eds) Mobile DNA, American Society for Microbiology, Washington, D.C..

Bradford M. (1976) Anal. Biochem. 72, 248-254.

Craig, N.L. (1989) in Mobile DNA, Berg, D.E., and Howe, M.M., eds., American Society for Microbiology, Washington, D.C., pp.211-226.

Craigie, R., and Mizuuchi, K. (1987) Cell 51, 493-501.

Craigie, R., Mizuuchi, M., and Mizuuchi, K. (1984) Cell 39, 387-394.

Flores, C., Qadri, M.I., and Lichtenstein, C. (1990) Nucl. Acids Res. 18, 901-911.

Fried, M. and Crothers, D. (1981) Nucl. Acids Res. 9, 6505-6525.

Galas, D.J., and Schmitz, A. (1978) Nucl. Acids Res. 5, 3157-3170.

Gartenberg, M.R. and Crothers, D.M. (1988) Nature 333, 824-829.

Garner, M. and Revzin, A. (1981) Nucl. Acids Res. 9, 3047-3060.

Gay, N.J., Tybulewicz, V.L.J., and Walker, J.E. (1986) Biochem J. 234, 111-117.

Gierl, A., Lutticke, S., and Saedler, H. (1988) EMBO J. 7, 4045-4053.

Groenen, M.A.M., Vollering, M., Krijgsman, P., van Drunen, K., and van de Putte, P. (1987) Nucl. Acids Res. 15, 8831-8844.

Ichikawa, H., Ikeda, K., Wishart, W.L., (1987) Proc. Natl. Acad. Sci. USA 84, 8220-8224.

Johnson, R.C., and Reznikoff, W.S. (1984) J. Mol. Biol. 177, 645-664.

Kubo, K., and Craig, N.L. (1990) J. Bacteriol. 5, 2774-2778.

Kunze, R., and Starlinger, P (1989) EMBO J. 8, 3177-3185.

Lichtenstein, C., and Brenner, S. (1982) Nature (London) 297, 601-603.

Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499-559.

McKown, R.L., Orle, K.A., Chen, T. and Craig, N.L. (1988) J. Bacteriol. 170, 352-358.

McKown, R.L., Waddell, C.S., Arciszewska, L.K., and Craig, N.L. (1987) Proc. Natl. Acad. Sci. USA **84**, 7807-7811.

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

Misra, S., and Rio, D.C. (1990) Cell **62**, 269-284.

Morris, C.F., Hama-Inaba, H., Mace, D., Sinha, N.K., and Alberts, B. (1979) J. Biol. Chem. **254**, 6787-6796.

Orle, K.A., and Craig, N.L. Gene, (in press).

Pato, M. (1989) in Mobile DNA American Society for Microbiology, Washington, D.C., pp. 23-52.

Prives, C. (1990) Cell **61**, 735-738.

Robertson, H.M., and Engels, W.R. (1989) Genetics **123**, 815-824.

Rogers, M., Ekaterinaki, N., Nimmo, E., and Sherratt, D.J. (1986) Mol. Gen. Genet. **205**, 550-556.

Sauer, R.T., Smith, D.L. and Johnson, A.D. (1988) Genes and Develop. **2**, 807-816.

Siegel, L., and Monty, K. (1966) Biochim. Biophys. Acta **112**, 346-362.

Smith, J.A., (1987) in Current Protocols in Molecular Biology, Ausubel, F.M., et al., eds., Greene Publishing Associates and Wiley-Interscience, New York, New York, pp.10.2.1-10.2.9.

Stark, W.M., Boocock, M.R., and Sherratt, D.J. (1989) Trends Genet. **5**, 304-309.

Surette, M.G., Buch, S.J., and Chaconas, G. (1987) Cell 49, 253-262.

Tabor, S., and Struhl, K. (1987) in Current Protocols in Molecular Biology, Ausubel, F.M. et al eds., Greene Publishing Associates and Wiley-Interscience., New York, New York, pp.3.5.7-3.5.10.

Thompson, J.F., and Landy, A. (1989) in Mobile DNA, Berg, D.E. and Howe, M.M. eds., American Society for Microbiology, Washington, D.C., pp.1-22.

Waddell, C.S., and Craig, N.L. (1988) Genes & Develop. 2, 137-149.

Wiater, L.A., and Grindley, N.D.F. (1988) EMBO J. 7, 1907-1911.

Zerbib, D., Jakowec, M., Prentki, P, Galas, D.J. and Chandler, M. (1987) EMBO J. 6, 3163-3165.

Zerbib, D., Polard, P., Escoubas, J.M., Galas, D. and Chandler, M. (1990) Mol. Microbiol. 4, 471-477.

FOOTNOTES

1 Abbreviations: bp, base pair(s); kb, 1000 bp; tns, Tn7-encoded transposition gene.

2 Bainton, R., Gamas, P. and Craig, N.L. (in preparation).

3 Tn7 end segments are designated L (left end) or R (right end). The numbers denote the extent of Tn7 sequences; for example, R1-199 denotes a Tn7R segment extending from the terminal bp (R1) of the right end of Tn7 to position R199.

4 The top and bottom strands are as shown in Figure 7.

5 Gamas, P., Bainton, R. and Craig, N.L. (in preparation).

FIGURE LEGENDS

Figure 3.1 Column chromatography steps in the purification of TnsB protein

Chromatography was carried out as described in Experimental Procedures. Column fractions were assayed for protein content (solid line) and DNA binding activity (filled circles). DNA binding activity was analyzed by band shift assays containing 0.1 μl of the column fractions; at the resulting TnsB concentrations, the amount of substrate fragment bound is proportional to TnsB concentration. DNA binding activity is expressed as the fraction of the input L109-166 fragment that was present in a shifted complex as evaluated by densitometric scanning of gel autoradiograms.
A: DNA-cellulose chromatography. Fraction I was subjected to DNA-cellulose chromatography. The elution of bound TnsB from the column by an NaCl gradient is shown (NaCl concentration indicated by dotted line).
B: Norleucine-sepharose chromatography. Fraction II was subjected to hydrophobic chromatography on norleucine sepharose. The arrow indicates the addition of low-salt buffer which leads to the step-wise elution of TnsB.
C: Phosphocellulose chromatography. Fraction III was subjected to chromatography on phosphocellulose. The arrows indicate the addition of higher salt buffers, P-0.39

M NaCl at the first arrow and P-0.52 M NaCl at the second, which lead to the step-wise elution of TnsB.

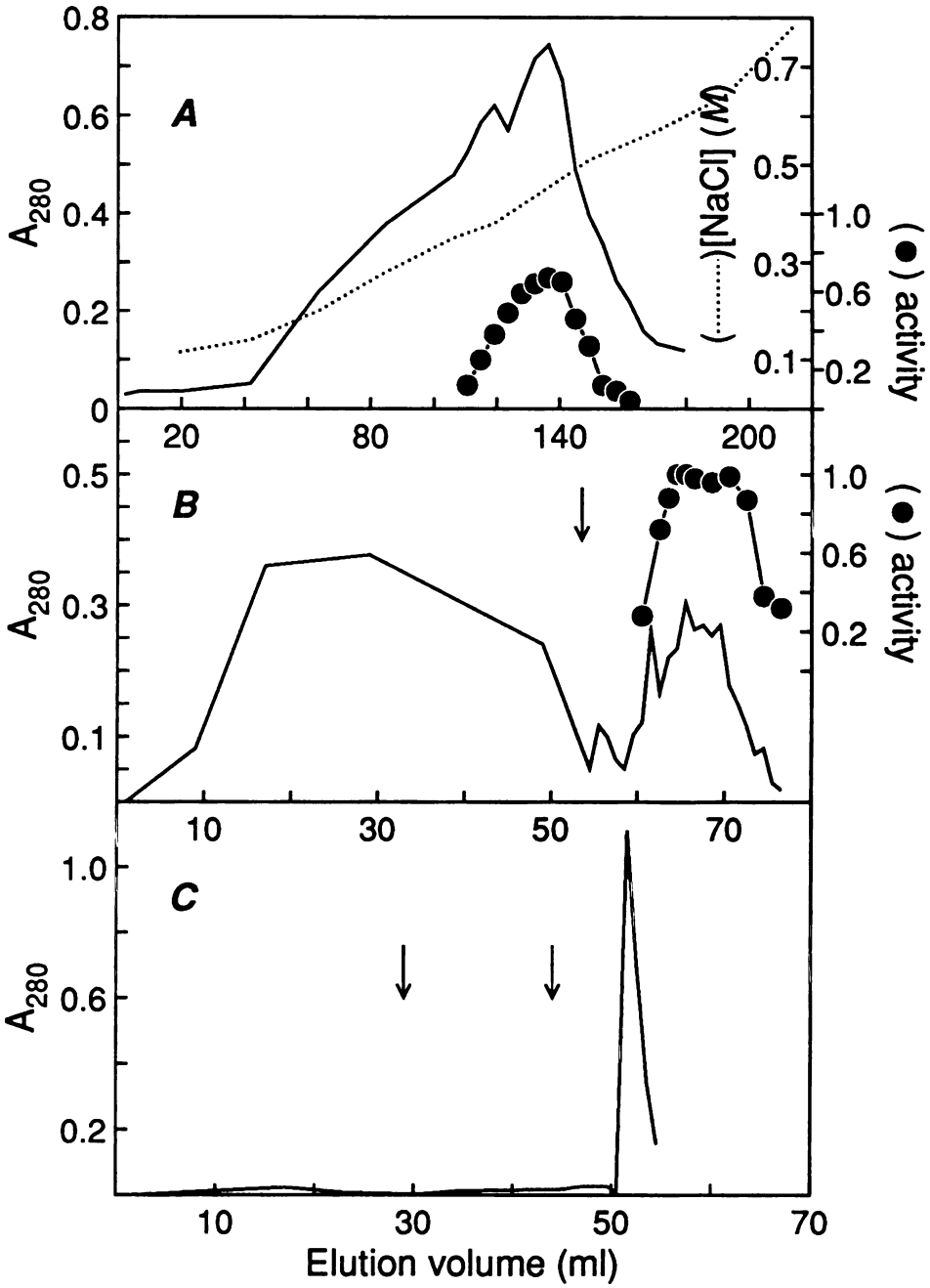

A
B
C
A_{280}
(●) activity
(……)[NaCl] (M)
Elution volume (ml)

Figure 3.2 Protein composition of TnsB fractions

A. SDS-PAGE analysis of TnsB fractions. Each fraction was separated by SDS-PAGE on a 5-15% gradient gel and stained with Coomassie Blue R-250. Lane 1, Fraction I, 75.2 μg; lane 2, Fraction II, 10.4 μg; lane 3, Fraction III, 3.2 μg; lane 4, Fraction IV, 8.8 μg. The numbers at the left indicate the positions of molecular weight markers.

B: Western analysis of purified TnsB protein. The indicated amounts of TnsB protein (Fraction IV) were separated by SDS-PAGE on a 7.5% gel and detected with affinity-purified anti-TnsB antibodies using conjugated alkaline phosphatase for detection. Lane 1, 320 ng; lane 2, 60 ng; lane 3, 13 ng. The numbers at the left indicate the positions of molecular weight markers.

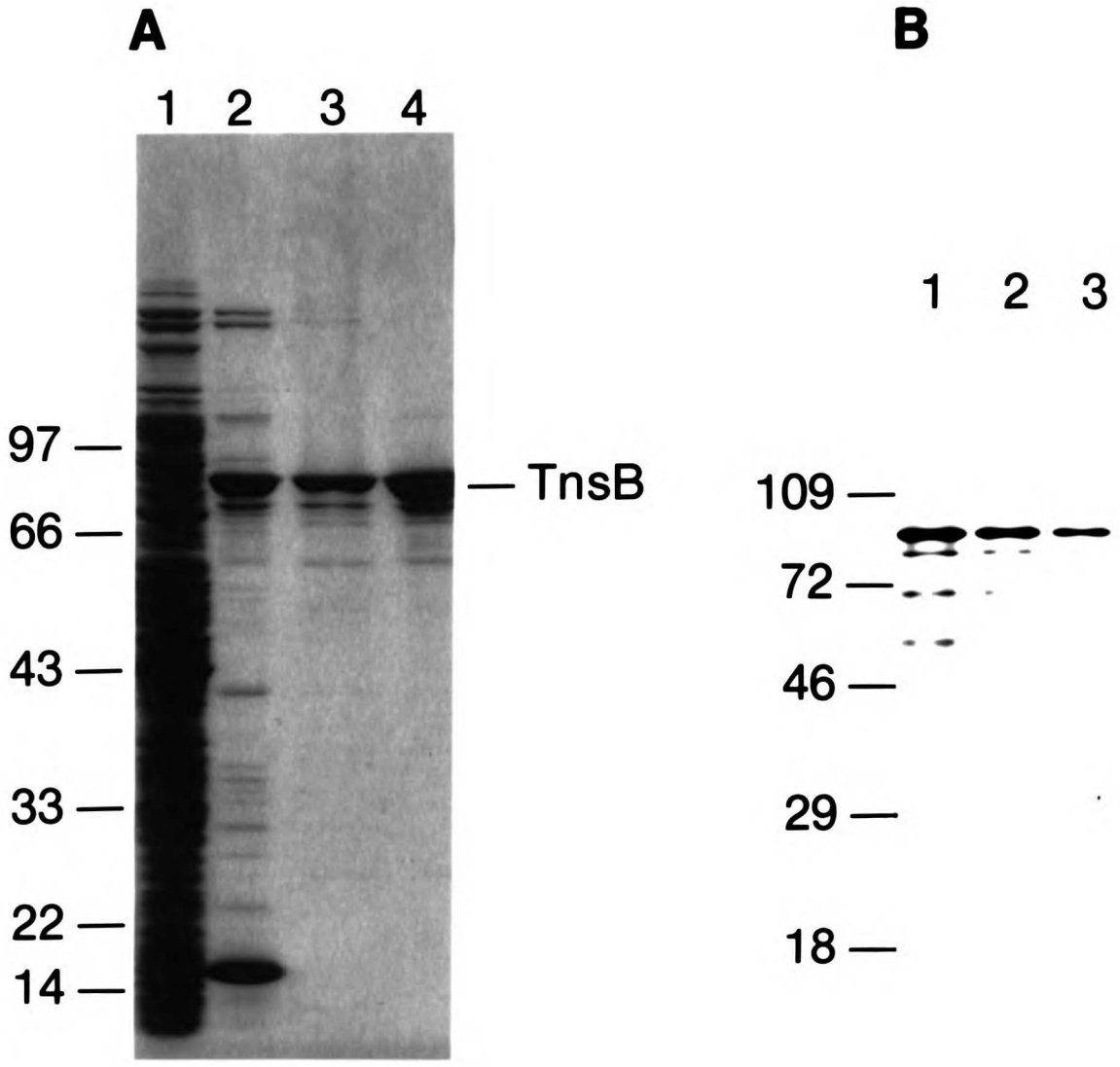

A
1
2
3
4
97
66
43
33
22
14
TnsB
B
1
2
3
109
72
46
29
18

Figure 3.3 Determination of the native molecular weight of TnsB

A: Gel filtration chromatography. TnsB (Fraction IV) and the indicated protein standards were subjected to gel filtration chromatography on Sephadex G-150 as described in Experimental Procedures. The Stokes radius of TnsB was determined to be 5.4 nm by comparison of its observed elution position (open square) relative to the observed elution positions and known Stokes radii (filled squares) of the protein standards: aldolase, 4.81 nm; BSA, 3.55 nm; ovalbumin, 3.05 nm; and chymotrypsinogen A, 2.09 nm.

B: Sucrose gradient sedimentation. TnsB (Fraction IV) and the indicated protein standards were subjected to sucrose gradient sedimentation analysis as described in Experimental Procedures. The sedimentation coefficient of TnsB was determined to be $s_{20,w}$ 4.3 by comparison to its observed sedimentation position (open circle) relative to the observed sedimentation positions and known $s_{20,w}$ values (filled circles) of the protein standards: catalase ($s_{20,w}$ 11.3), aldolase ($s_{20,w}$ 4.31), and ovalbumin ($s_{20,w}$ 3.66).

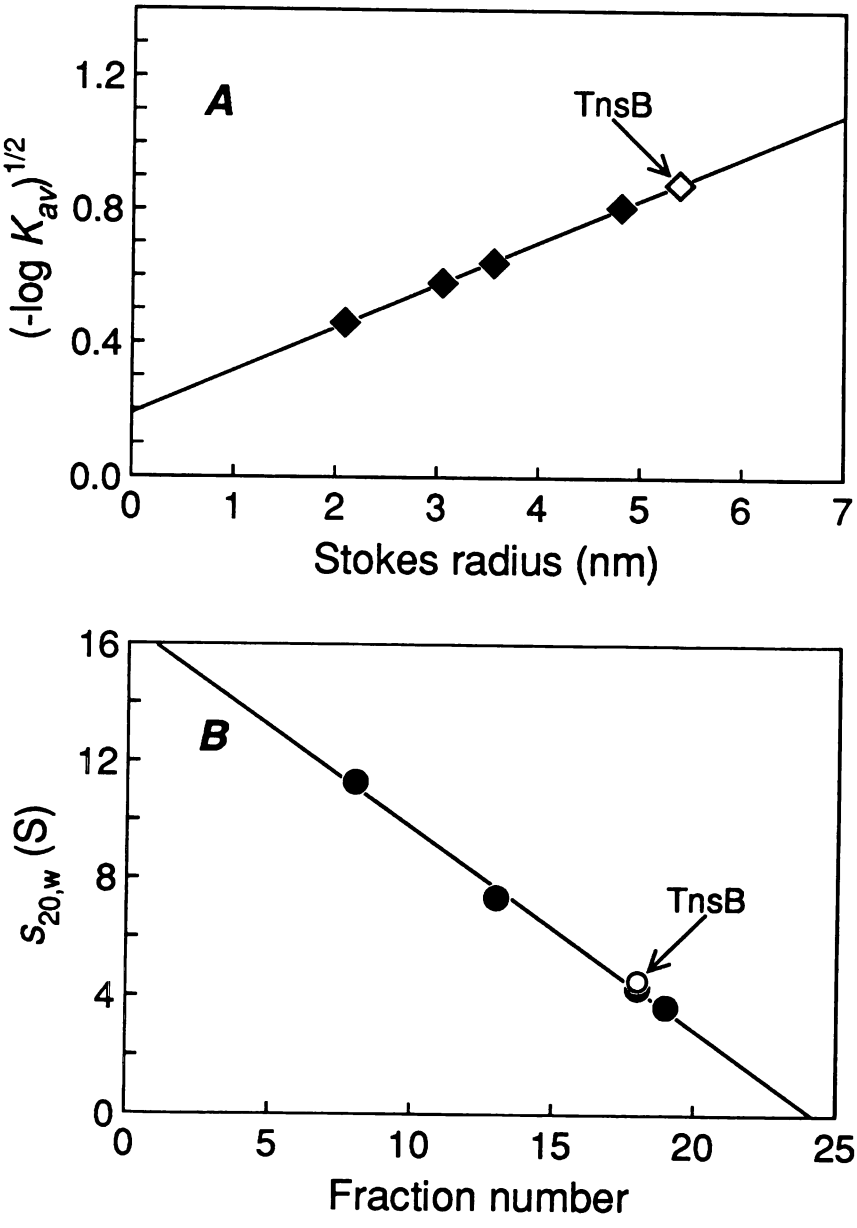

A
TnsB
$(-\log K_{av})^{1/2}$
1.2
0.8
0.4
0.0
0
1
2
3
4
5
6
7
Stokes radius (nm)
B
TnsB
$s_{20,w}$ (S)
16
12
8
4
0
0
5
10
15
20
25
Fraction number

Figure 3.4 TnsB binds to the ends of Tn7.

The DNA binding activity of various TnsB-containing fractions was evaluated by band shift assays using different Tn7 end fragments as substrates in assays that also contained 1 mg/ml BSA.

A: L1-166

B: R1-199

C: L109-166

Lane 1, tnsABCDE crude cell lysate, 280 μg/ml final concentration in reaction mixture; lane 2, tnsB crude cell lysate (Fraction I), 50 μg/ml in A and B and 120 μg/ml in C; lane 3, purified TnsB (Fraction IV), 1 μg/ml in A and B and 2.5 μg/ml in C; lane 4, no addition. In C, the complex labeled "a" is thought to contain intact TnsB whereas those labeled "b" and "c" are proposed to contain TnsB derivatives generated by proteolytic degradation.

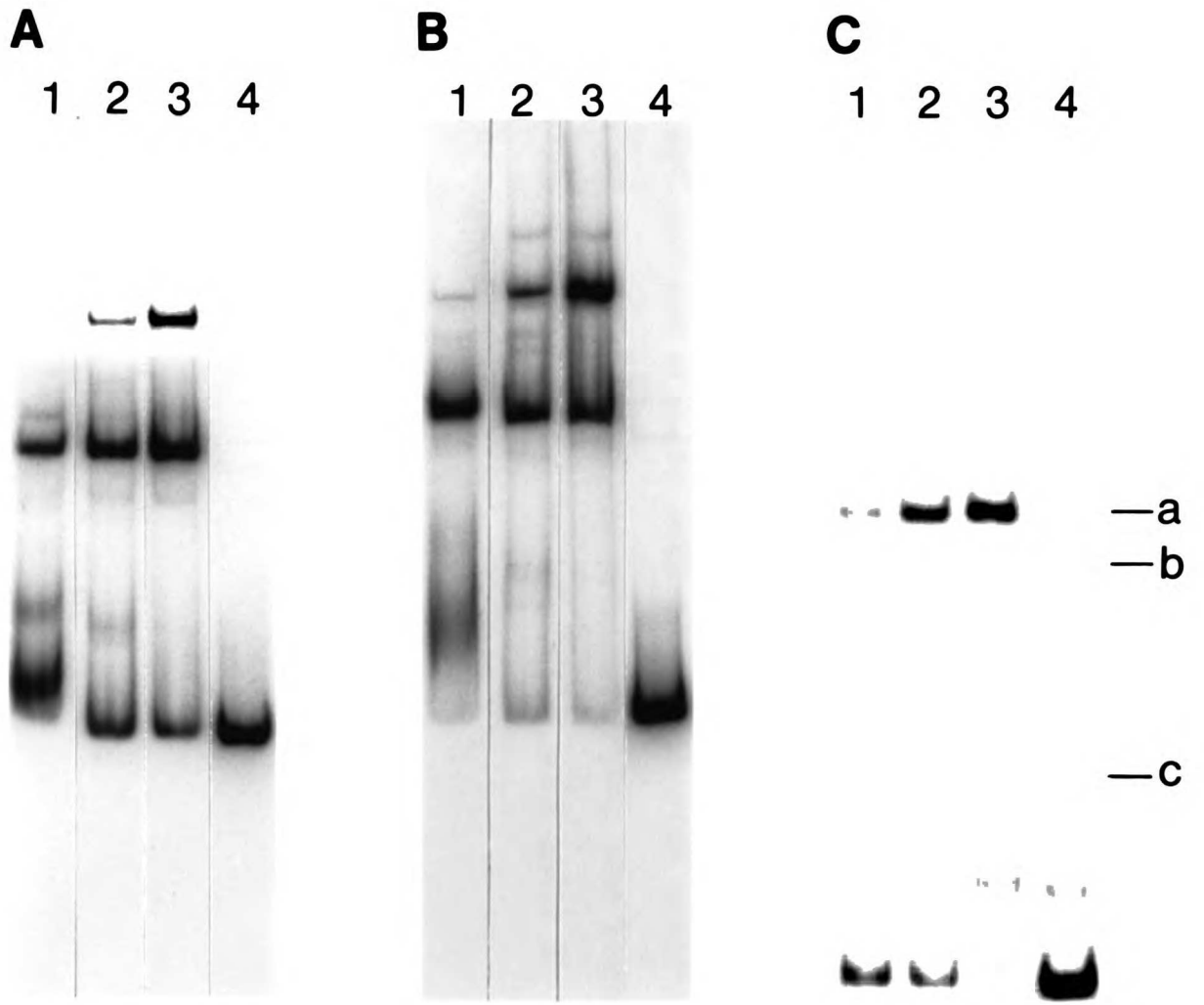

A
1 2 3 4
B
1 2 3 4
C
1 2 3 4
—a
—b
—c

Figure 3.5 Specific binding to the end of Tn7 copurifies with TnsB

1.1 mg of a Fraction III TnsB preparation was subjected to phosphocellulose chromatography as described in Experimental Procedures. Fractions containing TnsB protein that eluted from the column by the application of a P-0.52 M NaCl buffer step were evaluated for their protein content by SDS-PAGE (A) and their binding activity by a band shift assay (B).

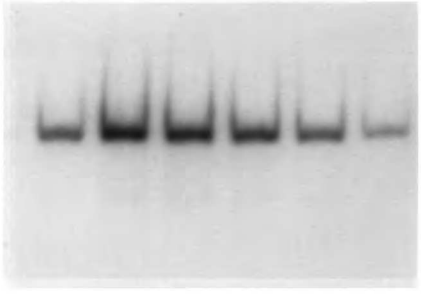
A: 16 μl of each fraction was separated by SDS-PAGE on a 5-15% gradient gel and stained with Coomassie Blue R-250.

B: 0.1 μl of each fraction was assayed using the L109-166 fragment as a substrate; no addition was made to the leftmost lane.

A

TnsB

B



TnsB + DNA

DNA

Figure 3.6 DNase I footprinting analysis of the interaction of TnsB with the ends of Tn7.

DNase I attack was performed in the absence (-) or presence (+) of 1.25 μ/ml (A) or 2.5 μ/ml (B-D) of Fraction IV TnsB to evaluate the interaction of TnsB with the indicated DNA strands. Products of A + G chemical sequencing reactions were used as markers. The nucleotide positions at the Tn7 ends are indicated, L1 being the terminal nucleotide in Tn7L and R1 being the terminal nucleotide in Tn7R. The vertical lines indicate the regions protected by TnsB. Top and bottom strands are as shown in Figure 7.

A. Bottom strand of Tn7L

B. Top strand of Tn7L

C. Bottom strand of Tn7R

D. Top strand of Tn7R

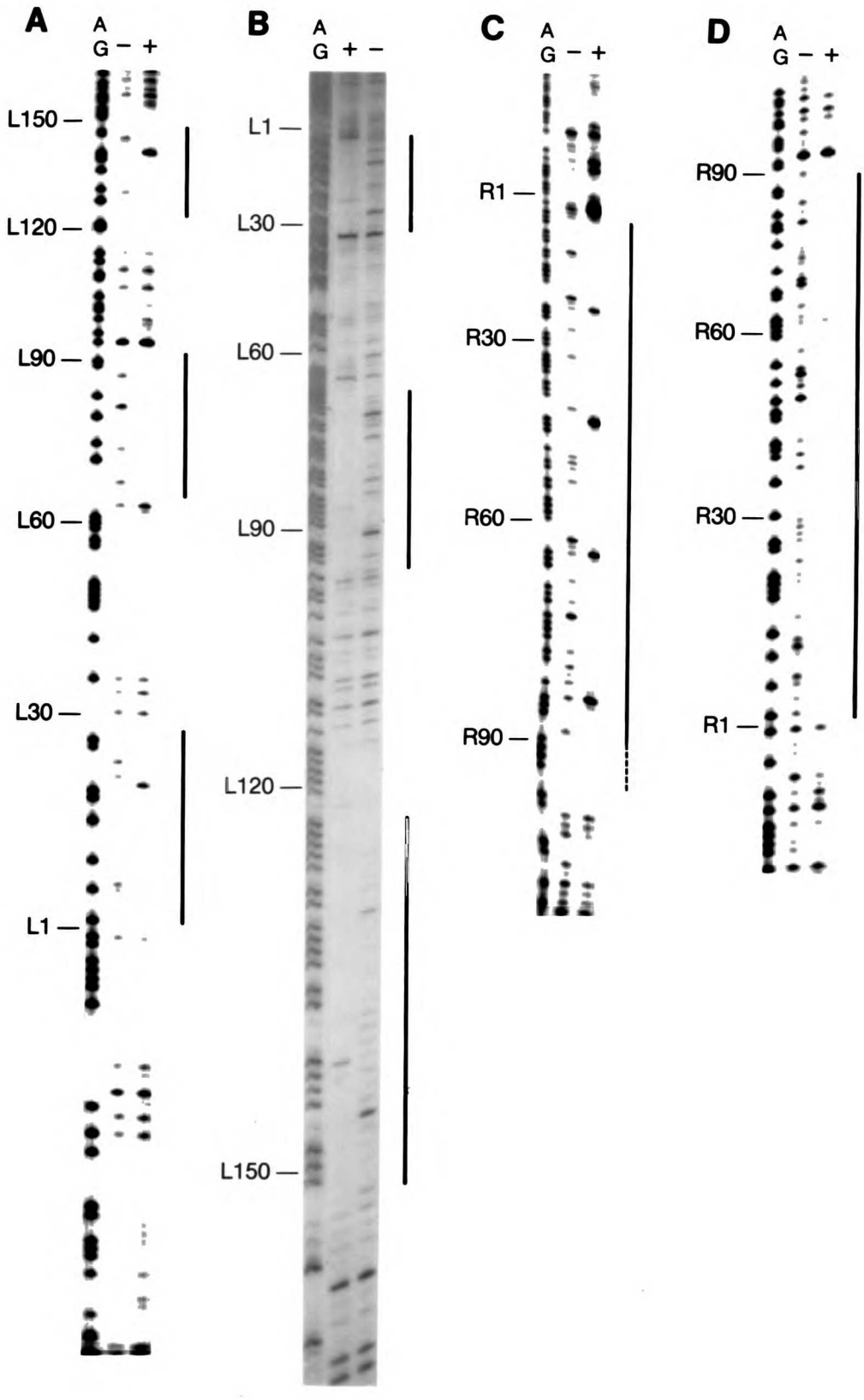

A
A G – +
L150
L120
L90
L60
L30
L1
B
A G + –
L1
L30
L60
L90
L120
L150
C
A G – +
R1
R30
R60
R90
D
A G – +
R90
R60
R30
R1

Figure 3.7 Nucleotide sequences of the Tn7 ends and of the regions protected by TnsB protein against DNase I attack.

The nucleotide sequences of Tn7L and Tn7R are shown. The terminal Tn7 nucleotides are L1 and R1 and numbers increase towards the inside of Tn7; the sequences flanking the ends are from attTn7 (Lichtenstein and Brenner, 1982; Gay et al., 1986; McKown et al., 1988). The arrows mark the 22 bp repeat sequences; the nucleotides of the terminal 8 bp of each end of Tn7, which together with the adjacent 22 bp repeats form Tn7's imperfect terminal inverted repeats, are denoted with bold italics and marked with dotted lines. The -35 sequence of the putative tnsAB promoter in Tn7R is indicated by a broken line. The regions of the Tn7 ends protected by TnsB and the proposed TnsB binding sites, a, β, and γ in Tn7L and ϕ, χ, ψ and ω in Tn7R, are marked. The dark grey boxes indicate bases whose 5' phosphodiester bonds are protected from DNase I attack, the light grey boxes indicate partially protected regions, the diamonds indicate positions of enhanced cleavage and circles mark positions that are not protected from attack by the presence of TnsB. In some cases, the status of regions

between the protected and unprotected regions cannot be evaluated because these areas are not markedly sensitive to DNAse I attack in the absence or presence of TnsB.

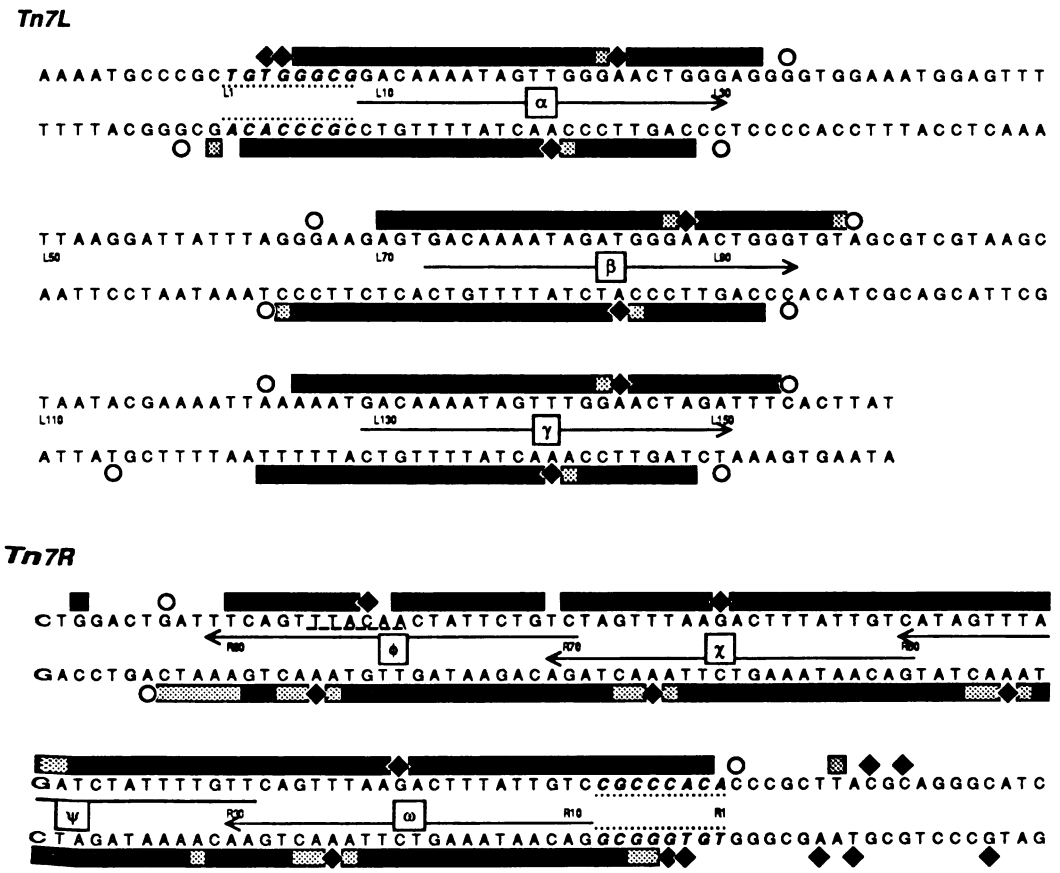

Tn7L
AAAATGCCCGCTGTGGGCGGACAAAATAGTTGGGAACTGGGAGGGGTGGAAATGGAGTTT
TTTTACGGGCGACACCCGCCTGTTTTATCAACCCTTGACCCTCCCCACCTTTACCTCAAA
TTAAGGATTATTTAGGGAAGAGTGACAAAATAGATGGGAACTGGGTGTAGCGTCGTAAGC
AATTCCTAATAAATCCCTTCTCACTGTTTTATCTACCCTTGACCCACATCGCAGCATTCG
TAATACGAAAATTAAAAATGACAAAATAGTTTGGAACTAGATTTCACTTAT
ATTATGCTTTTAATTTTTACTGTTTTATCAAACCTTGATCTAAAGTGAATA
Tn7R
CTGGACTGATTTCAGTTTACAACTATTCTGTCTAGTTTAAGACTTTATTGTCATAGTTTA
GACCTGACTAAAGTCAAATGTTGATAAGACAGATCAAATTCTGAAATAACAGTATCAAAT
GATCTATTTTGTTCAGTTTAAGACTTTATTGTCCGCCCACACCCGCTTACGCAGGGCATC
CTAGATAAAACAAGTCAAATTCTGAAATAACAGGCGGGTGTGGGCGAATGCGTCCCGTAG

Figure 3.8 Sequence comparison of TnsB binding sites

The sequence of one strand of each of the proposed TnsB binding sites is shown, the top strand of α, β, and γ in Tn7L and the bottom strand of ϕ, χ, ψ and ω in Tn7R (as shown in Fig. 7). The terminal α and ω sites are shown in bold type. The 3 bp regions enclosed by the boxes are bracketed by the positions of increased DNase I cleavage in each binding site, the 23-24 junction in the upper strand being hypersensitive and the 20-21 junction in the lower strand being hypersensitive. The underlined positions indicate regions of overlap between the Tn7R sites. The bottom two lines summarize the frequency of occurance of particular nucleotides at each position with the protected regions; K is T or G, R is A or G. The arrows indicate a 4 bp inverted repeat within the sites. The nucleotide positions enclosed by the grey box are the most highly conserved in sequence identity among the sites. The bottom arrow indicates the 22 bp repeats noted by Lichtenstein and Brenner (1982).

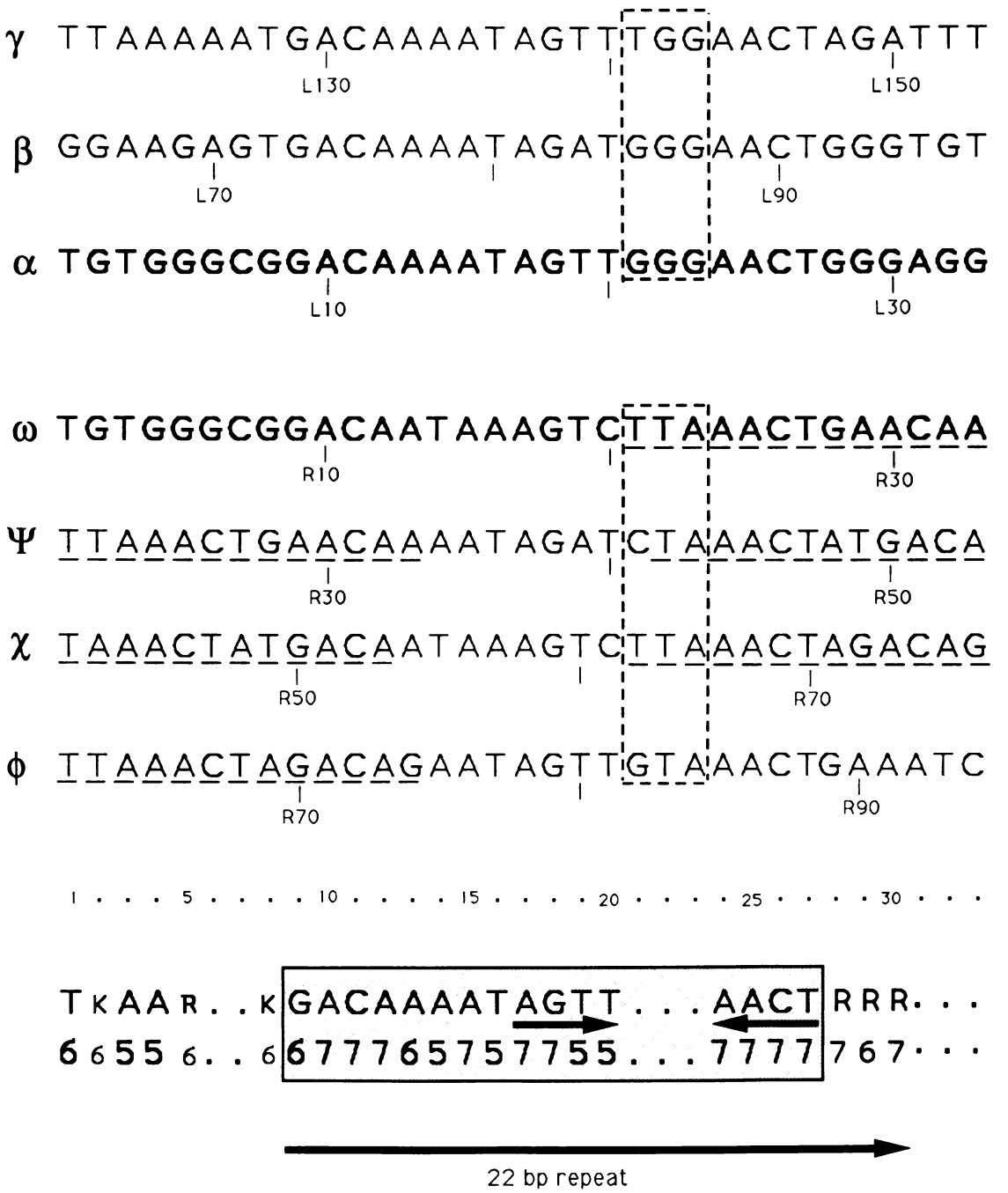

γ TTAAAAATGACAAAATAGTTTGGAACTAGATTT
L130
L150
β GGAAGAGTGACAAAATAGATGGGAACTGGGTGT
L70
L90
α TGTGGGCGGACAAAATAGTTGGGAACTGGGAGG
L10
L30
ω TGTGGGCGGACAATAAAGTCTTAAACTGAACAA
R10
R30
Ψ TTAAACTGAACAAAATAGATCTAAACTATGACA
R30
R50
χ TAAACTATGACAATAAAGTCTTAAACTAGACAG
R50
R70
ϕ TTAAACTAGACAGAATAGTTGTAAACTGAAATC
R70
R90
1 . . . 5 . . . . 10 . . . . 15 . . . . 20 . . . . 25 . . . . 30 . . .
TKAAR..KGACAAAATAGTT...AACTRRR...
6655 6..6777765757755...7777767...
22 bp repeat

Figure 3.9 Purified TnsB is active in Tn7 transposition *in vitro*

in vitro **Tn7 transposition reactions were carried out as described in Experimental Procedures in the presence (+) and absence (-) of TnsB protein (Fraction IV) at 0.17 μg/ml. After incubation, DNAs were recovered, linearized by digestion with a restriction enzyme, separated by gel electrophoresis and detected by Southern hybridization with a miniTn7-specific probe. Transposition was evaluated by following the translocation of miniTn7 element from a donor plasmid (position of migration "D") to an *attTn7* target plasmid (position of migration "P").**

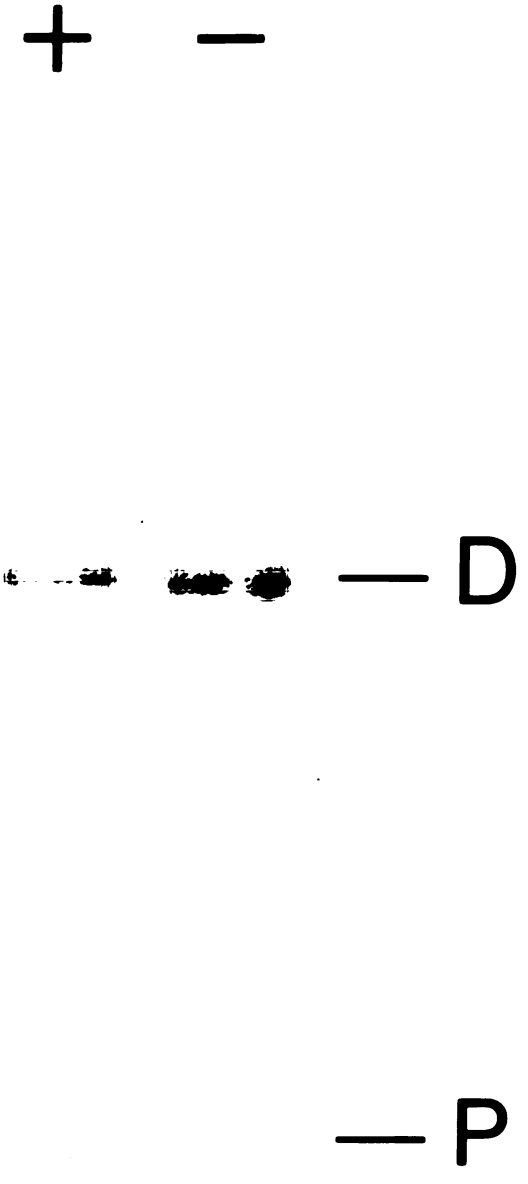

+
−
D
P

TABLE I

TnsB Purification

Fraction	Volume (ml)	Protein Concentration (mg/ml)	Total Protein (mg)	Relative activity[a]	Recovery of activity[b] (%)
I Cleared lysate	100	23.50	2350	(1)	(100)
II DNA cellulose	25	0.65	16	95	65
III Norleucine sepharose	18.5	0.20	3.7	265	42
IV Phosphocellulose	4.8	0.55	2.6	154	17

[a] - The specific binding activity (DNA binding activity/mg) of Fraction I was arbitrarily designated as 1. Specific activities of the pooled fractions were compared using the band shift assay to evaluate TnsB binding to the L109-166 fragment. Prior to assay, aliquots of each fraction were diluted in crude lysates from cells lacking TnsB and a constant volume of these dilutions was added to binding reaction mixtures (see Experimental Procedures). Binding activity was measured with dilutions such that only a small fraction of the input fragment was complexed with TnsB; under these conditions, the amount of complex shifted is proportional to the amount of TnsB. With purified TnsB (Fraction IV), about 50 % of the fragment is shifted in reactions containing 1.3 μg/ml TnsB. The amount of DNA fragment complexed with TnsB was determined by densitometric scanning of autoradiograms. Relative activities were determined by measuring the extent of dilution required of the various fractions to result in the shift of a small, arbitrarily designated amount of the input DNA fragment. The specific activity of the TnsB fractions was not different when the fractions were diluted and assayed in the presence of BSA, i.e. without the addition of E. coli crude lysate.

CHAPTER IV

Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon

ABSTRACT

We have used several high resolution methods to examine the interaction of TnsB, a transposition protein encoded by the bacterial transposon Tn7, with its binding sites at the ends of the transposon. These binding sites lie within the DNA segments that are directly involved in transposition. We show that the binding of TnsB to DNA can promote DNA bending, suggesting that the interaction of TnsB with the ends may result in formation of a highly organized protein-DNA complex. We also identify likely positions of close contact between TnsB and its binding sites. Analysis of the interaction of TnsB with intact Tn7 ends reveals that the various binding sites have differing apparent affinities for TnsB. Thus TnsB occupies its binding sites in a particular order, the sites immediately adjacent to transposon termini being occupied only after other inner sites are bound by TnsB.

INTRODUCTION

Transposable elements are discrete DNA segments that can move from one genomic location to another. A key step in transposition is the recognition by recombination proteins of DNA sequences at the ends of the element that are directly involved in transposition. Specific recognition of the ends is the initial step in a series of events that likely involves formation of higher-order protein-DNA complexes that contain both ends of the element, breakage of DNA strands at the transposon termini and joining of the transposon ends to target DNA. Both transposon- and host-encoded proteins may participate in end recognition (for review, see Berg and Howe, 1989).

We are interested in understanding the transposition mechanism of the bacterial transposon Tn7 (Barth et al. 1976, for review, see Craig, 1989). We have previously established that TnsB, a Tn7-encoded transposition protein, binds specifically to both ends of Tn7 (McKown et al, 1987; Arciszewska, L., R. McKown and N.L. Craig, in preparation). tnsB is one of an elaborate array of Tn7-encoded transposition genes, tnsABCDE, that mediate two distinct but overlapping transposition pathways differing in target site-selectivity (Rogers et al., 1986; Waddell and Craig, 1988; Kubo and Craig, 1990). tnsABC + tnsD promote high-frequency insertion into a specific site in the Escherichia

coli chromosome called attTn7 whereas tnsABC + tnsE promote low-frequency insertion in apparently random target sites. Characterization of Tn7 transposition to attTn7 in a cell-free system (Bainton, R., P. Gamas and N.L. Craig, in preparation; Gamas, P., R. Bainton and N.L. Craig, in preparation) suggests that the Tns proteins may all participate directly in recombination. The interaction of TnsB with the transposon ends accounts for the requirement for tnsB in all Tn7 transposition reactions.

Previous studies (Arciszewska et al, in preparation) revealed that TnsB binds to several sites that are highly related in nucleotide sequence at each end of Tn7. These sites lie within the Tn7 end segments that participate directly in recombination (Arciszewska et al, 1989). In the left end of Tn7 (Tn7L), TnsB binds to three distinct sites within the terminal 150 bp segment; in the right end of Tn7 (Tn7R), these studies suggested that TnsB binds to four closely juxtaposed sites within the terminal 100 bp. We report here the results of high resolution studies of the interaction of TnsB with the ends of Tn7 using various chemical footprinting and binding interference techniques. Analysis of the interaction of TnsB with a single binding site has identified likely positions of close contact between TnsB and DNA and has also shown that TnsB binding can introduce a bend into DNA.

We have also explored in more detail the interaction of TnsB with intact Tn7 ends; these studies support the view that Tn7R contains four TnsB binding sites. We have determined the relative apparent affinities of each site for TnsB and found that the various sites differ in affinity for TnsB. These results support the view that Tn7 recombination involves the ordered assembly of specialized protein-DNA complexes containing the ends of Tn7.

RESULTS

The interaction of TnsB with a single binding site.

We have previously characterized the specific interaction of TnsB with DNA by DNase I protection experiments (Arciszewska et al, in preparation). These studies demonstrated that TnsB binds to three regions in Tn7L, each approximately 30 bp in length, and to a single region in Tn7R approximately 90 bp in length. We proposed that the three protected regions in Tn7L represent three single TnsB binding sites and that the Tn7R region contains four closely juxtaposed sites. To obtain a more detailed picture of the interaction of TnsB with DNA, we have performed protection studies using hydroxyl radicals as an attack reagent. We have also carried out binding interference studies using DNA substrates modified by hydroxyl radical attack or by base methylation. For these studies we used a DNA segment containing a single TnsB binding site. This site, designated γ, is the innermost TnsB binding site in Tn7L and is well-separated from the other binding sites. We have previously established by <u>in vivo</u> transposition studies that the region containing the γ site is required for efficient recombination (Arciszewska et al., 1989).

Hydroxyl radical footprinting provides a high resolution picture of the interaction between a protein and

DNA because of the small size of the hydroxyl radical probing reagent which attacks deoxyribose in the DNA backbone (Tullius and Dombroski, 1986). Experiments showing TnsB protection of the γ binding site against hydroxyl radical attack are shown in Figure 1; the results of these experiments are summarized in Figure 2. In the γ site, TnsB protects several distinct clusters of nucleotides on both DNA strands against hydroxyl radical attack; these clusters, designated I, II, III, IV , and V, all lie within the region of DNA protected by TnsB against DNase I attack (Arciszewska et al, in preparation). There is also a cluster of nucleotides adjacent to protected cluster IV that is hypersensitive to hydroxyl radical attack in the presence of TnsB.

Three clusters of hydroxyl radical protection, III, IV and V, fall at positions that are highly conserved in sequence identity among the seven proposed TnsB binding sites (Arciszewska et al., in preparation). A notable feature of these conserved sequences is a four bp inverted repeat located asymmetrically, i.e. towards the right edge, within the TnsB binding site (Figure 2). (In the γ site, this inverted repeat is actually five bp in length.) Two clusters, IV and V, lie symmetrically on the inverted repeat. The "spacer" region between the five bp repeats is somewhat more susceptible to hydroxyl radical attack in the presence of TnsB; this region is also hypersensitive to

DNase I attack (Arciszewska et al., in preparation). Cluster III is located on one side (leftwards) of the inverted repeat, approximately in the middle of the binding site. Cluster II and the much weaker cluster I lie in a region of considerably less sequence similarity among the TnsB binding sites; this region is, however, also protected by TnsB from DNAse I attack (Arciszewska et al., in preparation).

The overall picture of a TnsB binding site that emerges from this study is of an extended (26 bp) region in which close contacts between TnsB and DNA occur at several positions that are highly conserved in sequence identity among all TnsB binding sites and also at several positions that are not as highly conserved. The short inverted repeats that appear to be specifically contacted by TnsB are asymmetrically located within the larger TnsB binding site.

A change in the susceptibility of DNA to hydroxyl radical attack can result either from direct protein contact or from protein-induced changes in the helical parameters of DNA. To further explore which nucleotide positions in a TnsB binding site might directly contact TnsB, we performed binding interference studies with γ site DNA that was previously modified by hydroxyl radical attack (Hayes and Tullius, 1989; Chalepakis and Beato, 1989). DNA fragments were first modified by hydroxyl

radical attack to introduce single, random nucleoside gaps. The gapped DNA was then incubated with TnsB under conditions where more than 95% of the input DNA bound to TnsB. Protein-bound and free DNA were then separated by native polyacrylamide gel electrophoresis, extracted and analyzed on a DNA sequencing gel (Figure 1). DNA fragments lacking a nucleoside whose presence is important for TnsB binding will be overrepresented in the DNA that did not bind to TnsB (unbound DNA) compared to the input (control) DNA; these fragments will also be underrepresented in bound DNA compared to the input control. DNA fragments lacking nucleosides that are not involved in TnsB binding will be present in the same proportion in the bound, unbound and input DNAs. The results of the missing nucleoside binding interference experiments are summarized in Figure 2.

We found that, in general, binding of γ site DNA to TnsB was most impaired by the absence of nucleosides at the same positions where the strongest protection from hydroxyl radical attack was observed, i.e. binding was impaired by missing nucleosides within protected clusters II, III, IV and V. This finding suggests that these clusters of protection reflect direct contacts between TnsB and DNA.

We also observed a modest effect on the binding of γ site DNA to TnsB when nucleosides were missing at several positions not obviously protected by hydroxyl radical

attack, i.e. at positions 9-14 on the top strand and 18-21 on the bottom strand. It is difficult to know whether these modest effects reflect changes in protein binding caused by the absence of specific nucleoside contacts or by local structural changes in the DNA backbone upon removal of the nucleosides at these positions (Koudelka et al., 1987). We also note that many of the positions where nucleoside removal has only a modest effect on TnsB binding are base-paired to nucleotides on the other strand that show strong protection and binding interference effects.

We have also probed the interaction of TnsB with DNA by evaluating the effect of DNA alkylation on TnsB binding to the γ site. Dimethylsulphate (DMS) alkylation of purines occurs in the major groove at N-7 of guanine and in the minor groove at N-3 of adenine. Such alkylation may interfere with TnsB binding through disruption of a specific base contact, by steric inhibition or by a perturbation of DNA structure upon introduction of an alkyl group (Siebenlist and Gilbert, 1980). DNA fragments containing the γ binding site were modified by treatment with DMS such that a single base per fragment was alkylated and incubated with TnsB protein under conditions where only a small fraction of the DNA was bound to TnsB. Protein-bound and free DNA were separated by electrophoresis on a native polyacrylamide gel, extracted and analyzed on a DNA sequencing gel (Figure 1). Nucleotide positions at which

alkylation interferes with the formation of a TnsB-DNA complex will be underrepresented in the bound DNA compared to the input control DNA. The results of this analysis are summarized in Figure 2.

We found that alkylation of the bottom strand guanines at position 11 and at position 26 strongly interferes with TnsB binding. In the top strand, alkylation of the position 23 guanine noticeably hindered TnsB binding and alkylation of the guanines at positions 9 and 18 had a weak effect on TnsB binding. Three of the guanines at which alkylation interfers with TnsB binding (positions 11, 18 and 26) are highly conserved among all the TnsB binding sites and are likely in close proximity to TnsB, as shown by the hydroxyl radical protection and missing nucleoside binding interference experiments. Inhibition of TnsB binding through guanosine alkylation suggests that TnsB makes base-specific contacts in the major groove at these positions. Whether TnsB makes base-specific contacts with guanosine at position 9 which is also a highly conserved position is less clear; only modest effects on TnsB binding are observed upon modification by alkylation or in the absence of this nucleoside. It would be reasonable to suspect that the interference with TnsB binding observed upon alkylation of guanine at position 23, which lies within the "spacer" of the inverted repeat, may reflect a secondary effect of a change in DNA structure upon

methylation, as this region is accessible to both hydroxyl radical and DNase I attack.

The results of the hydroxyl radical protection experiments and of binding interference experiments with DNA modified by alkylation or nucleoside removal consistently suggest that most of the specific interactions between TnsB and the γ site occur within the "core" region of the binding site, i.e. the region most highly conserved in base identity among all TnsB binding sites (Arciszewska et al., in preparation). These interactions include contacts with the inverted repeat motif within the core region and extensive contacts to one side (leftwards) of the inverted repeat. The methylation interference analysis also suggests that TnsB can make base-specific contacts in the DNA major groove. The moderate contacts observed outside the region of highly conserved sequence identity suggests that TnsB may interact non-specifically with the DNA backbone in these regions rather than with particular bases. Alternatively, base-specific contacts in this region may occur in only some TnsB binding sites and thereby contribute to determining the affinity of TnsB for a particular site. Future comparisons of the interaction of TnsB with different binding sites will be required to establish what DNA features determine the strength and specificity of the interaction of TnsB with DNA.

TnsB Can bend DNA

Several sequence-specific DNA binding proteins involved in recombination have been shown to change DNA structure upon binding to DNA (Prentki et al., 1987; Hatfull et al., 1987; Salvo and Grindley, 1988; Robertson and Nash, 1988; Thompson and Landy, 1988; Mertens et al, 1988). Such changes in DNA structure can result from DNA bending induced by the interaction of the protein with a single binding site or the looping of DNA when proteins bound to separate sites interact with each other. An attractive explanation for the existence of multiple TnsB binding sites at the ends of Tn7 is that TnsB promotes the organization of the ends into a higher-order nucleoprotein structure. Formation of such a complex would likely involve changes in DNA structure. We asked if TnsB can change the structure of DNA by inducing a bend in DNA at a single binding site. Wu and Crothers (1984) established that bent DNA migrates more slowly than does a linear fragment of the same size during gel electrophoresis. They also established that the relative mobility of bent DNA depends on the position of the bending center with respect to the fragment ends, fragments bent in the middle being the slowest migrating species. To examine the ability of TnsB to bend DNA, we introduced the L119-155 segment, which contains the γ binding site, into a tester plasmid (Kim et al, 1990). This plasmid is designed such that digestion

with different restriction enzymes generates a series of DNA fragments of the same length in which the position of the γ binding site is permuted with respect to the fragment ends. We examined the mobility of this set of free and protein-bound DNA fragments by polyacrylamide gel electrophoresis (Figure 3).

The free DNA fragments migrate with generally similar mobilities, indicating that there is no considerable intrinsic bend within the γ site DNA. By contrast, TnsB appears to induce a bend in the γ site as evidenced by the different mobilities of various TnsB - DNA complexes that contain the binding site in different relative positions. We estimate the TnsB-induced DNA bend angle to be near 50^{o}, based on calculations utilizing Rf values for the lowest and highest mobility complexes (Thompson and Landy, 1988). Similar analysis of the mobility of a number of different fragments suggests that the TnsB-induced bend lies between positions 1 and 10 of the γ binding site (data not shown). Thus, the bending center is located asymmetrically within the TnsB binding site, within the region of relatively weak TnsB-DNA interaction as defined by our footprinting and interference studies.

The interaction of TnsB with intact Tn7 ends

DNase I footprinting studies (Arciszewska et al., in preparation) revealed that the ends of Tn7 contain multiple

sites for TnsB binding. Three separate 30 bp regions of protection, designated α, β and γ, were evident in Tn7L; in Tn7R, a single 90 bp region of protection was observed which we proposed contains four overlapping TnsB binding sites, designated ϕ, χ, ψ, and ω. To further characterize the interaction of TnsB with the ends of Tn7, we performed hydroxyl radical footprinting experiments using DNA fragments containing L1-166 and R1-199. The results of the experiments are presented in Figure 4 and summarized in Figure 5.

In Tn7L, three groups of five clusters of protection were observed. These clusters are very similar in pattern and signal intensity to those observed with the single γ binding site. The positions of these groups of protected clusters correspond to the α, β and γ binding sites identified by DNaseI footprinting. We also observed several clusters of decreased sensitivity to hydroxyl radical attack between the α and β binding sites. We suspect that this weak protection represents non-specific binding of TnsB between TnsB bound specifically at the α and β sites; this region is also weakly protected from DNase I attack at high concentrations of TnsB (see below). We note, however, that there is no obvious sequence similarity of the region between α and β to the actual TnsB binding sites. It is interesting to note that the α and β sites occupy the same relative positions in Tn7L as ϕ and ω

do in Tn7R, so that the non-specific binding positions correspond to the spatial distribution of TnsB in the χ and ψ sites.

In Tn7R, four groups of protected clusters are evident, each group corresponding to clusters II, III, IV and V observed at the isolated Tn7L γ site. These results support the view that Tn7R contains four TnsB binding sites. Inspection of the Tn7R sequence reveals considerable overlap among the proposed TnsB binding sites ϕ, χ, ψ, and ω. Consistent with such overlap, protected cluster I is not observed except at the terminal ω site; at the other three sites, cluster V from the adjoining site overlaps these positions.

Comparison of the hydroxyl radical protection patterns observed at each of the seven TnsB binding sites reveals that these patterns are highly related, suggesting that TnsB makes similar contacts at each binding site.

Sequential occupancy of the TnsB binding sites

What roles do the multiple TnsB binding sites in the ends of Tn7 have in recombination? Does TnsB have a similar affinity for all sites or do these sites differ so that the sites are filled in a particular order? To determine the relative apparent affinities of the sites, we examined binding at different concentrations of TnsB to Tn7L and Tn7R by both DNase I footprinting and band shift

assays. In band shift experiments, TnsB forms several distinct protein-DNA complexes with fragments containing either Tn7L1-166 or Tn7R1-199 (McKown et al, 1987; Arciszewska et al, in preparation). An attractive explanation for these different complexes is that they represent different numbers of TnsB molecules distributed among the various binding sites. The relative ratios of the complexes was dependent on TnsB concentration, suggesting that TnsB might bind to each end in an ordered fashion as determined by binding sites with differing affinities. Through direct protection analysis of various TnsB-DNA complexes, we have now directly characterized the occupancy of the binding sites at different TnsB concentrations, thereby determining the relative apparent affinities of the TnsB binding sites.

Figure 6 shows the binding of TnsB at different concentrations to a Tn7L1-166 DNA fragment as evaluated by DNase I footprinting. A detailed analysis of TnsB protection against DNase I attack has been presented elsewhere (Arciszewska et al., in preparation) and is also summarized in Figure 5. Here, only the order of occupancy of the α, β and γ sites is considered. At the lowest TnsB concentration used, only the innermost, γ binding site is occupied. Thus, in Tn7L, the γ site has the highest apparent affinity for TnsB. With increasing TnsB, the β site is next filled and, at highest concentration, the

terminal α site is filled. These results demonstrate that TnsB has distinct apparent affinities for each of its Tn7L binding sites. We have no information about how the binding of TnsB to the γ site may influence the binding of TnsB to the other sites, but no remarkable cooperativity is evident from this study. We note that at high TnsB concentrations, weak protection of the region between the α and β sites is observed; we interpret this as non-specific binding as there is no sequence similarity between this region and the authentic TnsB binding sites. By contrast, no distinct order of site occupancy was observed in similar TnsB titration experiments with Tn7R (data not shown).

We have also examined the binding of TnsB to Tn7L and Tn7R at different TnsB concentrations using the band shift method (Figure 7). Three TnsB-DNA complexes, designated LI, LII and LIII, are observed with Tn7L and four complexes, designated RI, RII, RIII and RIV, are observed with Tn7R. Because there is a regular difference among the apparent molecular weights of the complexes, as judged by comparison to DNA markers (data not shown), we suspected that these various complexes contain DNA molecules bound by different numbers of TnsB molecules distributed among the multiple TnsB binding sites. The relative amounts of each complex is dependent on the TnsB concentration: at low concentrations, the faster migrating species (such as LI and RI) predominate, whereas at higher TnsB concentrations,

the slower migrating complexes (such as LIII and RIV) predominate. The gradual transition between these various forms is notable and suggests there is little cooperativity in the binding of TnsB to the ends of Tn7 under these conditions.

We analyzed these complexes by <u>in situ</u> footprinting, using copper-phenanthroline as an attack reagent on gels containing the complexes (Kuwabara and Sigmen, 1987), followed by extraction and analysis of the DNA (Figure 8). In the fastest migrating Tn7L species, complex LI, a single protected region corresponding to the γ binding site is observed. In complex LII, two regions of protection corresponding to the γ and β sites were observed. The slowest migrating species, complex LIII, displayed three regions of protection corresponding to the γ, β and α sites.

Thus, both direct analysis of TnsB-DNA complexes generated in band shift experiments and DNase I footprinting carried out in solution show that TnsB has a distinct affinity for each of its Tn7L binding sites and occupies them in sequential fashion, binding first to the innermost γ site, then to β and, finally, to the terminal α site. We have no information about the form of TnsB that occupies each binding site.

Similar analysis of the TnsB-DNA complexes formed with Tn7R reveal that the fastest migrating species, complex RI,

represents TnsB binding to the χ binding site; thus this site, which lies within Tn7R and not adjacent to the terminus, is the Tn7R site with the highest affinity for TnsB. In both complex RII and complex RIII, in addition to occupancy of the χ site, considerable protection of the ψ and terminal ω binding sites is observed. In complex RII, the ω site appears to be slightly more protected than does the ψ site whereas in complex RIII, comparable protection of both the ψ and ω sites is evident. We suspect that complex RII contains two molecules of TnsB and complex RIII contains three because of the regular increment in apparent molecular weight differences between the Tn7 end complexes (data not shown). Thus, we suggest that "complex RII" is heterogenous and is actually a mixture of two distinct protein-DNA complexes that have similar mobilities, one complex containing TnsB at the highest affinity site χ and also at the ψ site and the other complex containing TnsB at the highest affinity site χ and also at the terminal ω site. The slightly greater degree of protection of the terminal ω site compared to the ψ site in "complex RII" suggests that ω may have a slightly higher affinity for TnsB. In complex RIII, we imagine that three molecules of TnsB are present, occupying the χ, ψ and ω sites. In the slowest migrating complex RIV, all four TnsB binding sites are protected, suggesting that four TnsB molecules occupy the ϕ, χ, ψ and ω sites. Our studies suggest that the ϕ

site, which overlaps a likely tns promoter (Gay et al., 1986), is the site with the lowest apparent affinity for TnsB in Tn7R, filling only after the other Tn7R sites are occupied. Binding of TnsB to this site likely represses transcription of tns genes (Rogers et al., 1986; Waddell and Craig, 1988; Orle and Craig, 1990).

These studies have revealed that the multiple species observed in band shift assays which use the Tn7 ends as substrates are protein-DNA complexes that contain TnsB molecules distributed among its various binding sites. The binding sites in each end differ in their apparent affinity for TnsB; thus TnsB interaction with the ends is characterized by the progressive and sequential occupancy of its binding sites. The nucleotide sequence differences between the TnsB binding sites that are responsible for these differing affinities and the influences the sites have on each other remain to be determined.

We have previously shown that the ends of Tn7 are functionally distinct (Arciszewska et al., 1989). For example, miniTn7 elements containing two Tn7R segments transpose whereas elements containing two Tn7L segments do not. One attractive explanation for this observed functional difference is that the affinity of TnsB for Tn7L is very much lower than for Tn7R; however, in vitro comparisons of TnsB binding to each end provide no support for this view. We have observed no substantial difference

in the apparent affinity of TnsB for Tn7L and Tn7R when evaluated in several different ways including protein titration at low concentrations of substrate DNA, and challenge of TnsB binding by either increasing ionic strength or increasing non-specific DNA (data not shown). It should be noted that in previous experiments in which we examined the binding of TnsB in crude extracts to the Tn7 ends, the apparent affinity of TnsB for Tn7R appeared to be modestly higher than for Tn7L (McKown et al, 1987). It will be interesting to compare the binding of TnsB to the ends under authentic recombination conditions, i.e. ends located in supercoiled molecules containing both Tn7L and Tn7R in the presence of a full complement of recombination proteins (Bainton et al., in preparation).

DISCUSSION

In this paper, we have examined the interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon using several high-resolution methods. The results presented here, together with our previous work (Arciszewska et al., in preparation), show that TnsB binds to three sites in the left end of Tn7 and four sites in the right end. The Tn7L sites, α, β and γ, are directly oriented and are separated by segments of unrelated DNA sequence. The Tn7R sites, ϕ, χ, ψ and ω, are also directly oriented but are closely juxtaposed with some overlap between the edges of neighboring sites.

Our studies suggest that the binding of TnsB to the Tn7 ends may substantially change the overall structure of these DNA segments. We found that TnsB binding to the Tn7L γ site introduces a bend into this DNA. Although the degree of this bend is modest compared to some other protein-induced bends (Wu and Crothers, 1984; Kim et al, 1989; Thompson and Landy, 1988), the overall structure of the ends may be changed considerably if similar bends are introduced at all of the TnsB binding sites. Possible interactions between the various TnsB binding sites have yet to be evaluated. The Tn7L α and β sites are "phased", i.e. are separated by an even number (6) of helical turns, and thus are well-positioned for possible interactions

between TnsB molecules bound at these sites; the other Tn7L site, γ, is not, however, in phase with the adjacent β site (5.3 turns). In Tn7R, the four TnsB sites are phased, lying two helical turns apart. Because Tn7 transposition proceeds most efficiently when the Tn7 ends are located on a supercoiled DNA molecule (Bainton et al, in preparation), dissection of the interactions between the various TnsB sites on such a substrate will be required to identify and understand the TnsB interactions and resulting changes in DNA structure that may be critical to recombination. It has already been established that highly organized protein-DNA complexes mediate other recombination reactions (Thompson and Landy, 1989; Pato, 1989; Stark et al, 1989).

Sequential Occupancy of the TnsB Binding Sites

We have found that the TnsB binding sites within each end of Tn7 differ in their apparent affinity for TnsB. Because of these differences, the binding sites are sequentially occupied by TnsB in an ordered fashion as the concentration of TnsB is increased. A notable feature of this ordered binding is that sites inside the element are occupied first by TnsB and the sites that are adjacent to the transposon termini, are filled only after TnsB occupies the other inner sites.

In Tn7L, the innermost γ site has the highest apparent affinity for TnsB and thus is occupied first. With

increasing TnsB, the β site and then the terminal α site are also occupied. We have previously demonstrated that occupancy of the γ site is likely critical to recombination because the transposition of miniTn7 elements lacking the γ region is severely impaired (Arciszewska et al., 1989). The TnsB sites in Tn7R are also occupied in sequential fashion although the order of occupancy is not apparently as fixed as in Tn7L. The inner χ site has the highest apparent affinity for TnsB and thus is occupied first. Our studies suggest that TnsB subsequently occupies either the terminal ω site or the ψ site which thus appear to have similar apparent affinites for TnsB. After the χ, ψ and ω sites are filled, the innermost ϕ site is filled. Occupancy of the ϕ site is not essential to transpositio. We have established in <u>in vivo</u> studies that transposition of miniTn7 elements lacking the ϕ site can occur at frequencies only slightly lower than that of an element with intact ends; however, it is important to note that the behaviour of such deleted elements is distinctly different from that of intact elements (Arciszewska et al., 1989).

We imagine that the nucleoprotein structure active in recombination contains TnsB bound at multiple sites in each end including the terminal binding sites where strand breakage and joining must occur. Although it is not yet known whether TnsB actually executes breakage and rejoining or assists another Tn7-encoded or perhaps a host-encoded

protein in this function, it seems very likely that the binding of TnsB to the termini must play a critical role in end recognition and utilization. We have established here that the highest affinity sites for TnsB lie inside the transposon rather than at the extreme termini and that TnsB progressively occupies several sites in each end, culminating with binding to the termini. We imagine that recombination can occur only when the terminal TnsB sites are filled. An attractive hypothesis is that the sequentional binding of TnsB to the ends may play an important role in controlling transposition, no strand breakage occuring until a complete "transposome" is assembled at the ends of the transposon.

TnsB binding to Tn7R also likely has a role in regulating the expression of the Tn7-encoded tns genes. The promoter (Gay et al., 1986) of an operon that includes tnsA and tnsB and also perhaps the other tns genes (Flores et al, 1990; Orle and Craig, 1990) overlaps the ϕ TnsB binding site. Other studies have provided support for the hypothesis that the interaction of TnsB with Tn7R does indeed repress tns gene transcription (Rogers et al, 1986; Waddell and Craig, 1988; Orle and Craig, 1990). The observation that the ϕ site has the lowest relative affinity for TnsB suggests that this protein can act as negative regulator of gene expression only when its concentration actually exceeds the concentration required for recombination.

Recognition of DNA by TnsB

We have used several high-resolution techniques to examine the interaction of TnsB with a single binding site to begin to dissect how this protein may recognize and specifically contact DNA. Using protection from hydroxyl radical attack and several binding interference methods, we have detected contacts between TnsB and the Tn7L γ binding site that encompass 26 bp. These contacts extend widely from one side of the helix to another over two major and two minor grooves (Figure 9). The strongest contacts are observed within three clusters of nucleotides located within the portion of the binding site that is most highly conserved in sequence identity among all TnsB binding sites (Arciszewska et al, in preparation).

A prominent feature of the TnsB consensus sequence is a 4 bp inverted repeat located at one side of the binding site. Both the protection and binding interference studies of the Tn7L γ site indicate that symmetrical nucleotides within the inverted repeat are in close contact with TnsB, suggesting that these repeats play an important role in TnsB recognition. The inverted repeats flank a spacer segment that remains exposed to solvent in the presence of TnsB. We are intrigued by the fact that there is an apparently systematic difference between the sequences of these spacer regions in the Tn7L and Tn7R binding sites (Arciszewska et al, in preparation). However, it remains

to be determined what contribution this region may make to Tn7 transposition. The inverted repeat region is asymmetrically flanked by a highly conserved 7 bp sequence that also appears to make specific contacts with TnsB.

Our observations suggest that a TnsB binding site contains both symmetrical and asymmetrical features, i.e. the inverted repeat and an adjoining region. Thus it will be particularly interesting to determine the form of TnsB that interacts with each binding site and how these various highly conserved features are recognized.

In addition to the strong contacts that TnsB makes in the highly conserved region, markedly weaker interactions are observed outside this segment. These weaker contacts lie at one edge of the binding site in a region of little homology among the TnsB binding sites. Perhaps these contacts represent non-specific interactions between TnsB and DNA. Alternatively they may provide the critical features that differentiate the TnsB binding sites. We are intrigued by the possibility that this apparently non-conserved region may provide a special determinant for recognition of the extreme termini of the transposon (Arciszewska et al., in preparation).

What type of protein structural motif(s) in TnsB is used for DNA recognition? Our analysis of TnsB binding to DNA modified by alkylation suggests that specific recognition of DNA by TnsB involves contacts within the

major groove. We have observed that a TnsB derivative containing only the N-terminal 200 amino acids can bind specifically to the ends of Tn7 (data not shown). Flores et al. (1990) have noted that TnsB contains a region of weak similarity to the helix-turn-helix motif characterisitic of some sequence-specific DNA binding proteins which is located within this 200 amino acid segment. Understanding how TnsB does recognize DNA and what changes in DNA structure it may promote will help reveal the role(s) of this protein in recombination.

MATERIALS AND METHODS

Strains and Plasmids

Plasmids were propagated in JM109 = *E. coli recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* Δ(*lac-proAB*) [F'*traD36 proAB lacI*^{Q} Δ(*lacZ*)M15] (Stratagene).

Tn7 end fragments were prepared from plasmids containing various Tn7L and Tn7R segments. The fragments containing Tn7R (R1-199) and Tn7L (L1-166) were isolated from pKS^{+}Ra and pKS^{+}La (Arciszewska et al., in preparation), respectively. The fragments containing the Tn7L segments L109-166 and L89-166 were isolated from pLA77 (Arciszewska et al, in preparation) and pLA76. pLA76 was constructed by inserting by blunt-end ligation a *Hga*I-*Hinc*II fragment, obtained by digestion of the *Eco*RI-*Hinc*II Tn7L fragment from pLA26 (Arciszewska et al., 1989), into the *Eco*RV site of Bluescript-KS (Stratagene); in pLA76, position L166 is adjacent to the vector *Eco*RI site.

DNA fragments containing the Tn7L segment L119-155 used in the circular permutation experiments were isolated from pLA78. To construct pLA78, a fragment containing the L119-155 segment, flanked on both ends by *Xba*I staggered ends, was obtained by annealing two 42 bp synthetic oligonucleotides, 5'-CTAGAATTAAAAATGACAAAATAGTTTGGAACTAGATTTCAT and 5'-CTAGATGAAATCTAGTTCCAAACTATTTTGTCATTTTTAATT. This fragment was inserted into the *Xba*I site of pBend2, a

pBR322 derivative that contains two sets of directly oriented 17 restriction sites with unique XbaI and SalI sites between them (Kim et al., 1990). In pLA78, position L155 is adjacent to the vector EcoRI site.

Preparation and labeling of DNA fragments

Plasmid DNA was digested with appropriate restriction enzymes and electrophoresed through 6% polyacrylamide gels. Slices containing the fragments of interest were cut out from the gels, the DNA electrophoretically transferred onto a DEAE membrane (Schleicher and Schuell) and recovered as suggested by manufacturer. Purified DNA fragments were labeled at their 3' ends by incubation with DNA polymerase I Klenow fragment and appropriate α-^{32}P-NTPs (Tabor and Struhl, 1987). Fragments from pKS^{+}Ra were used for analysis of Tn7R (R1-199); the 306 bp SmaI-HindIII fragment labeled at HindIII was used for bottom strand analysis and the 294 bp EcoRI-BamHI fragment labeled at EcoRI was used for top strand analysis (top and bottom strands are as shown in Figure 5). Fragments from pKS^{+}La were used for analysis of Tn7L (L1-166); the 270 bp SmaI-HindIII fragment labeled at HindIII was used for top strand analysis and the 258 bp EcoRI-BamHI fragment labeled at EcoRI was used for bottom strand analysis. For analysis of Tn7 L109-166, fragments from pLA77 were used; the 85bp SmaI-HindIII

fragment labeled at HindIII was used for top strand analysis and the 91 bp EcoRI-HincII fragment labeled at EcoRI was used for bottom strand analysis. For analysis of Tn7 L89-166, the 104 bp HindIII-SmaI fragment labeled at HindIII from pLA76 was used for analysis of the bottom strand.

DNA Binding

Unless otherwise indicated, binding reactions (100 µl contained 13.1 mM Tris·HCl (pH 8.0), 1.03 mM DTT, 0.03 mM EDTA, 110 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 10.3% (v/v) glycerol, 2 ng/ml poly(dI-dC)·poly (dI-dC), 275 µg/ml BSA, approximately 0.15 pmol of 3' end-labeled DNA fragment and TnsB as indicated in the figure legends. After a 20 min incubation at room temperature, protein-DNA complexes were either subjected to DNaseI or hydroxyl radical attack or separated by electrophoresis and analyzed using the procedures described below.

DNaseI footprinting

DNaseI footprinting experiments were performed as described in Arciszewska et al, in prepration.

Hydroxyl radical footprinting

The reactions (150 µl) contained 13.3 mM Tris·HCl (pH 8.0), 0.003 mM DTT, 0.13 mM EDTA, 115 mM NaCl, 1 mM $CaCl_2$,

0.15 % (v/v) glycerol, 2 μg/ml poly(dI-dC)·poly(dI-dC), 250 μl BSA, approximately 0.23 pmol 3' end-labeled DNA fragment and TnsB as indicated in the Figure legends. After incubation with TnsB, hydroxyl radical cleavage of DNA was performed as described in Tullius and Dombroski (1986) except that the reaction was carried out for 6 min. Recovered DNA was washed in 0.3 M sodium acetate, reprecipitated, resuspended in formamide loading solution and analyzed on 8% sequencing gels. To decrease agents that interfere with hydroyxl radical cleavage, the TnsB protein used in these experiments was dialyzed against 25 mM Tris·HCl (pH8.0), 0.1 mM DTT, 1 mM EDTA, 500 mM NaCl and 5% (v/v) glycerol.

Missing nucleoside binding interference

The experiments were performed as described in Hayes and Tullius (1989). 3' end-labeled DNA fragments were modified by hydroxyl radical attack as described above for the footprinting experiments in 12.5 mM Tris.HCl (pH 8.0), 0.1 mM EDTA, 100 mM NaCl and 1 mM $CaCl_2$. The DNA was ethanol precipitated, washed in 0.3 M sodium acetate, reprecipitated and resuspended in 10 mM Tris·HCl (pH 8.0) and 0.1 mM EDTA. Protein-DNA complexes were formed in reactions as described. Free and protein-bound DNAs were separated by electrophoresis on 6% polyacrylamide gels (29:1 acrylamide/N, N'- methylenebisacrylamide) in TBE

buffer at 19 V/ cm for 90 min (Fried and Crothers, 1981; Garner and Revzin, 1981). The gels were autoradiographed for 1 hr and the bands containing free and bound DNA cut out. The DNA was electrophoretically transferred from the gel slices to DEAE membrane (Schleicher and Schuell), recovered as suggested by the manufacturer, ethanol precipitated, washed in 0.3 M sodium acetate, reprecipitated and resuspended in formamide loading solution.

Methylation binding interference

Alkylation of 3' end-labeled DNA fragments was performed as described by Maxam and Gilbert (1980). Protein-DNA complexes were formed in 100 µl reactions containing 13.7 mM Tris.HCl (pH 8.0), 1.1 mM DTT, 0.1 mM EDTA, 10.5% (v/v) glycerol, 130 mM NaCl, 300 µg/ml BSA, approximately 0.15 pmol modified DNA fragment and TnsB as described in the Figure legends. After incubation for 20 minutes at room temperature, free and protein-bound DNAs were resolved and recovered as described above. The recovered DNA was cleaved with piperidine and analyzed on a sequencing gel.

Band Shift Assays

Protein-DNA complexes were separated by electrophoresis as described above.

Copper-phenanthroline footprinting

DNA binding reactions were performed as above except that the reactions contained approximately 0.28 pmol 3' end-labeled DNA fragment. Protein-DNA complexes were resolved by electrohporesis in a 4% polyacrylamide gel. The gel was subjected to chemical attack with 1,10-phenanthroline copper ion following the procedure of Kuwabara and Sigman (1987). The gels were autoradiographed for 30 min, the slices containing the DNA of interest cut out and DNA electrophoretically transferred into a dialysis bag (Sambrook et al, 1989). DNA was recovered by ethanol precipitation, washed in 0.3 M sodium acetate, reprecipitated, resuspended in formamide loading buffer and analyzed on an 8% sequencing gel (Maxam and Gilbert, 1980).

Analysis of DNA bending by circular permutation

pLA78 DNA was digested with appropriate restriction enzymes and DNA fragments recovered as described above. Approximately 1 pmol of DNA was incubated in 20 μl reaction mixtures containing 13.7 mM Tris·HCl (pH 8.0), 1.05 mM EDTA, 2.05 mM DTT, 125 mM NaCl, 10.5% (v/v) glycerol, 2 ng/ml poly(dI-dC)·poly(dI-dC), 300 μg/ml BSA and 1.3 μg/ml TnsB for 7 min at room temperature and then electrophoresed through 6% polyacrylamide gels. DNA was visualized by ethidium staining.

Other methods

Preparation of purified TnsB protein was as described in Arciszewska et al (in preparation). DNA sequencing was performed using the chemical degradation method (Maxam and Gilbert, 1980).

REFERENCES

Arciszewska, L.K., Drake, D., and Craig, N.L. (1989) J. Mol. Biol. 207, 35-52.

Barth, P.T., Datta, N., Hedges, R.W., and Grinter, N.J. (1976) J. Bacteriol. 125, 800-810.

Berg, D.E., and Howe, M. (1989) (eds) Mobile DNA, American Society for Microbiology, Washington, D.C..

Chalepakis, G. and Beato, M. Nucl. Acids Res. 17, 1783.

Craig, N.L. (1989) in Mobile DNA, Berg, D.E., and Howe, M.M., eds., American Society for Microbiology, Washington, D.C., pp.211-226.

Craigie, R., Mizuuchi, M., and Mizuuchi, K. (1984) Cell 39, 387-394.

Flores, C., Qadri, M.I., and Lichtenstein, C. (1990) Nucl. Acids Res. 18, 901-911.

Fried, M. and Crothers, D. (1981) Nucl. Acids Res. 9, 6505-6525.

Garner, M. and Revzin, A. (1981) Nucl. Acids Res. 9, 3047-3060.

Gay, N.J., Tybulewicz, V.L.J., and Walker, J.E. (1986) Biochem J. 234, 111-117.

Hatfull, G.F., Noble, S.M. and Grindley, N.D.F. (1987) Cell 49, 103-110.

Hayes, J.J. and Tullius, T.D. (1989) Biochemistry 28, 9521-9527.

Kim, J., Zwieb, C., Wu, C. and Adhya, S. (1989) Gene 85, 15-23.

Koudelka, G.B., Harrison, S.C. and Ptashne, M. (1987) Nature (London) 326, 886-888.

Kubo, K., and Craig, N.L. (1990) J. Bacteriol. 5, 2774-2778.

Kuwabara, M.D. and Sigman, D.D. (1987) Biochemsitry 26, 7234-7238.

Mertens, G., Klippel, A., Fuss, H., Blocker, H., Frank, R. and Kahmann, R. (1988) EMBO J. 7, 1219-1227.

Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 499-559.

McKown, R.L., Waddell, C.S., Arciszewska, L.K., and Craig, N.L. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 7807-7811.

Orle, K.A., and Craig, N.L. (1990) Gene 95, 1-7.

Pato, M. (1989) in Mobile DNA American Society for Microbiology, Washington, D.C., pp. 23-52.

Prentki, P., Chandler, M. and Galas, D.J. (1987) EMBO J. 6, 2479-2487.

Robertson, C.A. and Nash, H.A. (1988) J. Biol. Chem. 263, 3554-3557.

Rogers, M., Ekaterinaki, N., Nimmo, E., and Sherratt, D.J. (1986) Mol. Gen. Genet. 205, 550-556.

Salvo, J.J. and Grindley, N.D.F. (1988) EMBO J. 7, 3609-3616.

Siebenlist, U. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 122-126.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, a Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. pp.6.26-6.29.

Stark, W.M., Boocock, M.R., and Sherratt, D.J. (1989) Trends Genet. 5, 304-309.

Tabor, S., and Struhl, K. (1987) in Current Protocols in Molecular Biology, Ausubel, F.M. et al eds., Greene Publishing Associates and Wiley-Interscience., New York, New York, pp.3.5.7-3.5.10.

Thompson, J.F. and Landy, A. (1988) Nucl. Acids Res. 16, 9687-9705.

Thompson, J.F., and Landy, A. (1989) in Mobile DNA, Berg, D.E. and Howe, M.M. eds., American Society for Microbiology, Washington, D.C., pp.1-22.

Tullius, T.D. and Dombroski, B.A. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 5469-5473.

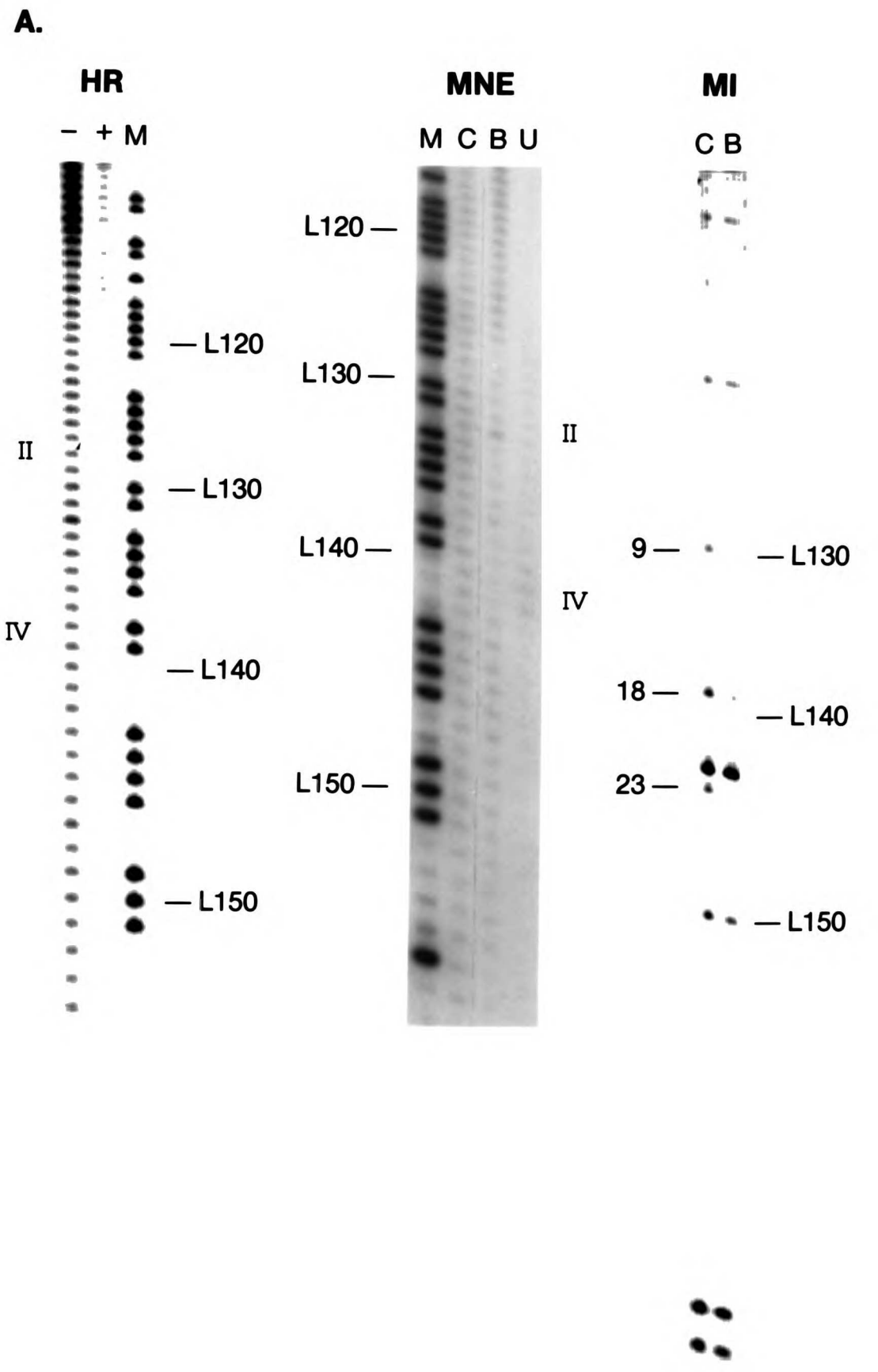
Waddell, C.S., and Craig, N.L. (1988) Genes & Develop. 2, 137-149.

Wu, H.-M. and Crothers, D.M. (1984) Nature (london) 308, 509-513.

Figure 4.1 Interaction of TnsB with a single binding site as evaluated by protection and binding interference analyses.

Interaction of TnsB with the Tn7L γ binding site was evaluated by protection from hydroxyl radical attack and by binding interference with DNA modified by either hydroxyl radical attack or alkylation. Panels A and B present analyses of the top and bottom strands, respectively. Hydroxyl radical attack (HR) was performed in the absence (-) or presence (+) of 2 μg/ml TnsB. Clusters of protection (I, II, III, IV and V) are indicated. In the binding interference experiments which used DNA modified by hydroxyl radical attack, i.e. missing nucleosides (MNE), the modified DNA was incubated with 2.5 μg/ml TnsB. Protein-bound DNA (B) was separated from unbound DNA (U) and compared to control DNA (C). Clusters of nucleosides whose absence interferes with TnsB binding (II, III, IV and V) are indicated. A + G chemical sequencing reactions were used as markers in the HR and MNE analyses. Nucleotides within the left end of Tn7 are indicated, L1 being the terminal nucleotide. Binding reactions for the methylation interference (MI) analysis, which used DNA modified by alkylation, were performed with 0.1 μg/ml TnsB. Protein-bound DNA (B) was separated from free DNA on native polyacrylamide gels and compared to control DNA (C). Bases

whose methylation affects TnsB binding are indicated by numbers which refer to their position within the γ binding site.


A.
HR
– + M
— L120
— L130
— L140
— L150
II
IV
MNE
M C B U
L120 —
L130 —
L140 —
L150 —
MI
C B
II
IV
9 —
18 —
23 —
— L130
— L140
— L150

B.

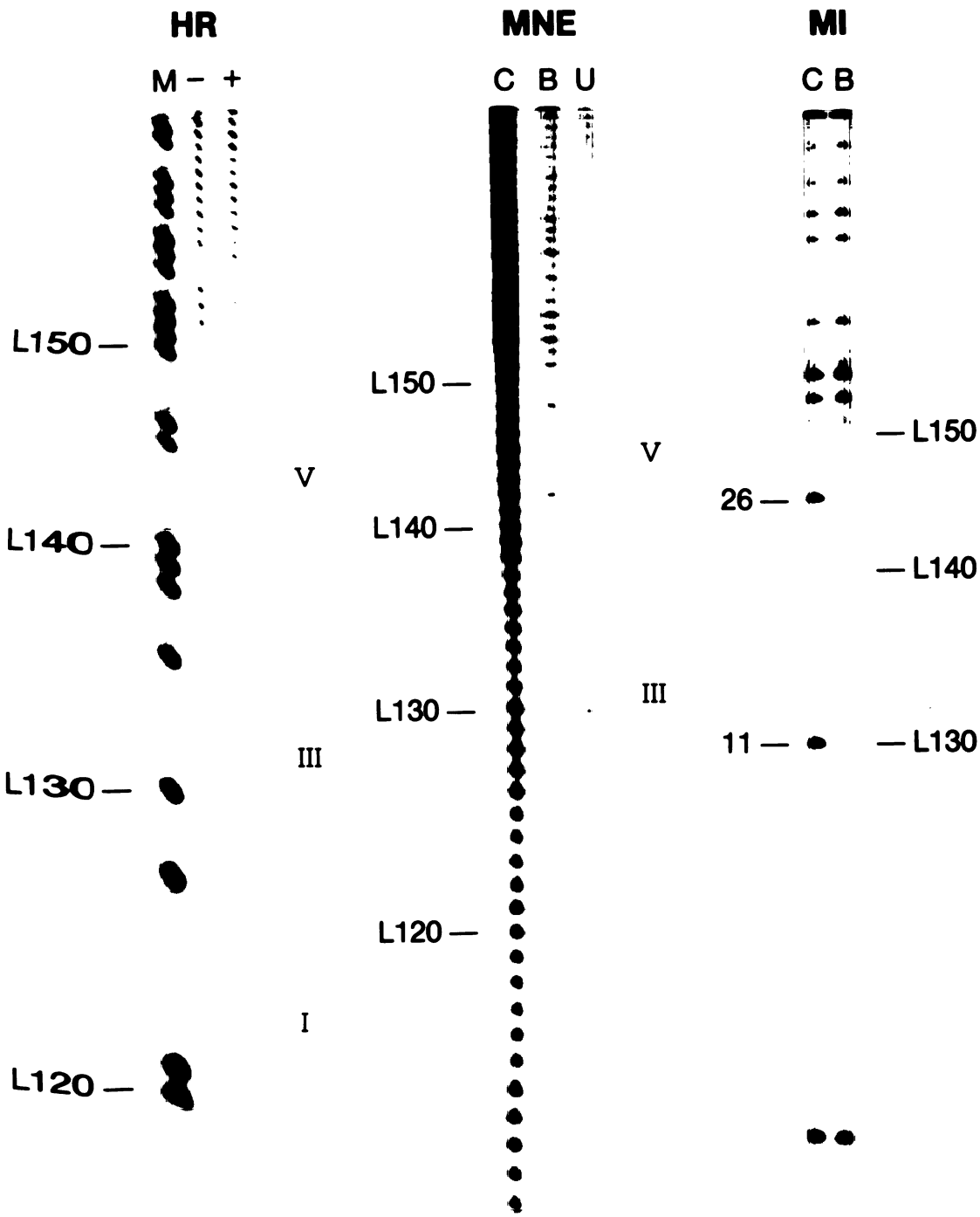

HR
M − +
L150 —
L140 —
L130 —
L120 —
V
III
I
MNE
C B U
L150 —
L140 —
L130 —
L120 —
V
III
MI
C B
— L150
26 —
— L140
11 —
— L130

Figure 4.2 Summary of protection and interference analyses of a single TnsB binding site.

The results of the experiments shown in Figure 1 are summarized diagrammatically. The top panel presents the results of the hydroxyl radical protection (HR) experiments and the bottom panel summarizes the results of binding interference studies that used DNA modified by hydroxyl radical attack, i.e. missing nucleosides (MNE) or by alkylation, i.e. methylation interference (MI). In both panels, the sequence of the γ TnsB binding site as defined by DNase I protection is shown (Arciszewska et al., in preparation). The numbers above the top strand in the top panel and below the bottom strand in the bottom panel refer to nucleotide positions within the Tn7L sequence, L1 being the terminal nucleotide; the numbers between the top and bottom strands refer to nucleotide positions within a TnsB binding site (Arciszewska et al, in preparation). The horizontal arrows indicate a 5 bp inverted repeat within the γ binding site; this inverted repeat is 4 bp in most TnsB binding sites. The nucleotides in the 3 bp spacer region between the inverted repeats are marked with bold letters. The grey box between the panels indicates the nucleotide positions that are highly conserved in sequence identity among all identified TnsB binding sites.

HR (top panel): Clusters of protection (I, II, III, IV and V) are marked. The degree of protection (strong,

moderate and weak) observed in the presence of TnsB is indicated by vertical bars (longest bar = strongest protection); filled dots indicate positions of very weak protection. The stars mark positions that are hypersensitive to hydroxyl radical attack in the presence of TnsB. The vertical arrows indicate positions that are hypersensitive to DNaseI attack.

MNE + MI (bottom panel): The effects of missing a particular nucleoside on TnsB binding as evaluated by examination of both bound and unbound DNA are indicated by vertical bars, the longest bars indicating the most severe effect. Nucleotides where base methylation affects TnsB binding are indicated by diamonds, filled diamonds indicating a strong inhibition of TnsB binding and open diamonds a modest inhibition of TnsB binding.

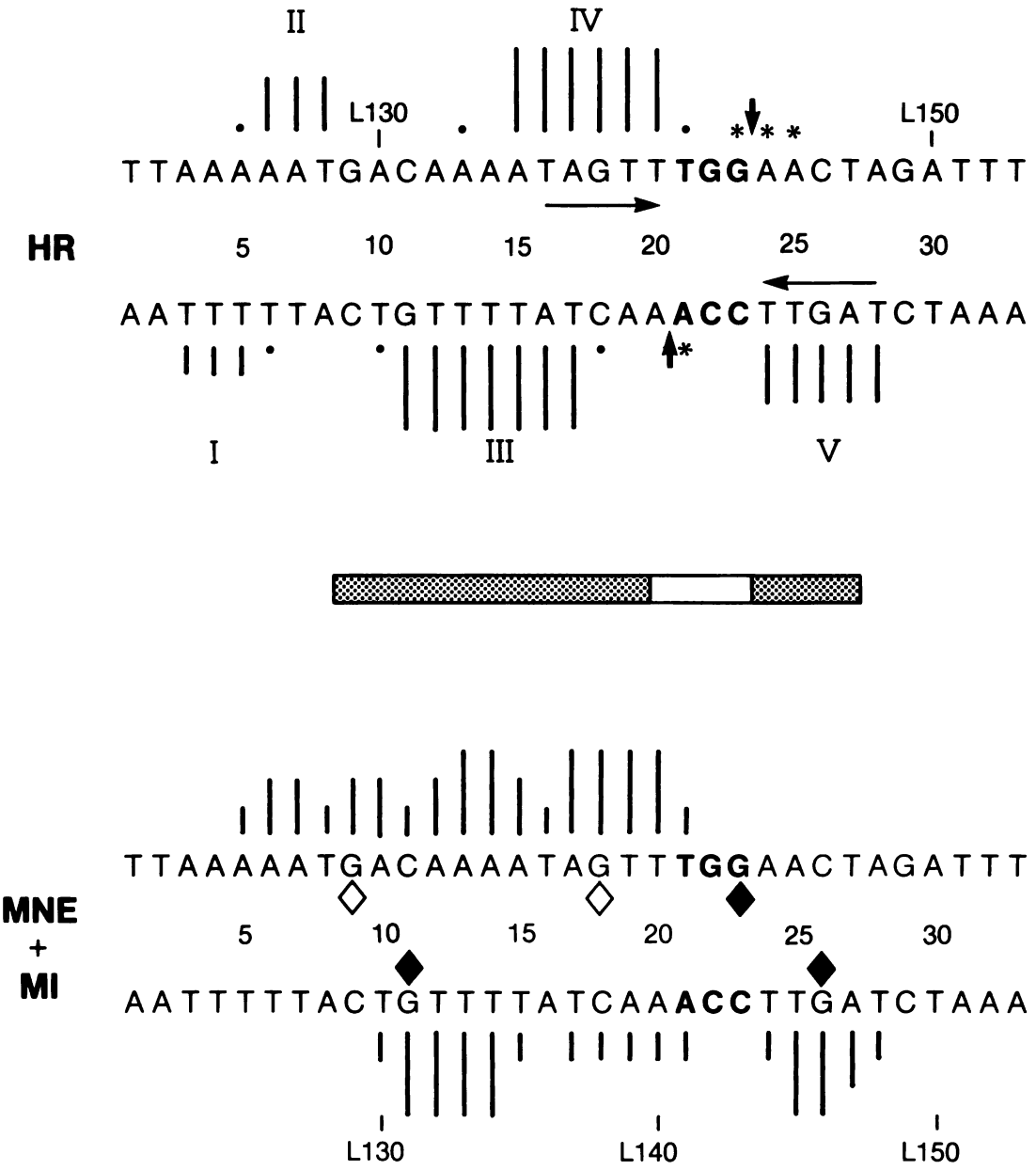

II
IV
L130
L150
TTAAAAATGACAAAATAGTTTGGAACTAGATTT
HR
5
10
15
20
25
30
AATTTTTTACTGTTTTATCAAACCTTGATCTAAA
I
III
V
MNE
+
MI
TTAAAAATGACAAAATAGTTTGGAACTAGATTT
5
10
15
20
25
30
AATTTTTTACTGTTTTATCAAACCTTGATCTAAA
L130
L140
L150

Figure 4.3 Circular permutation analysis of TnsB-induced DNA bending at the γ binding site.

The mobility of TnsB-DNA complexes in which the γ binding site was located at different positions in the DNA fragment was examined. Binding reactions were performed with 1.3 μg/ml TnsB. The mobility of free and protein-bound DNA was examined on a native polyacrylamide gel and DNA visualized with ethidium bromide. The DNA fragments used were obtained from pLA78 by digestion with the indicated restriction enzymes: lane 1, MluI; lane 2, NheI; lane 3, SpeI; lane 4, EcoRV; lane 5, PvuII; lane 6, StuI; lane 7, NruI; lane 8, KpnI; lane 9, BamHI.

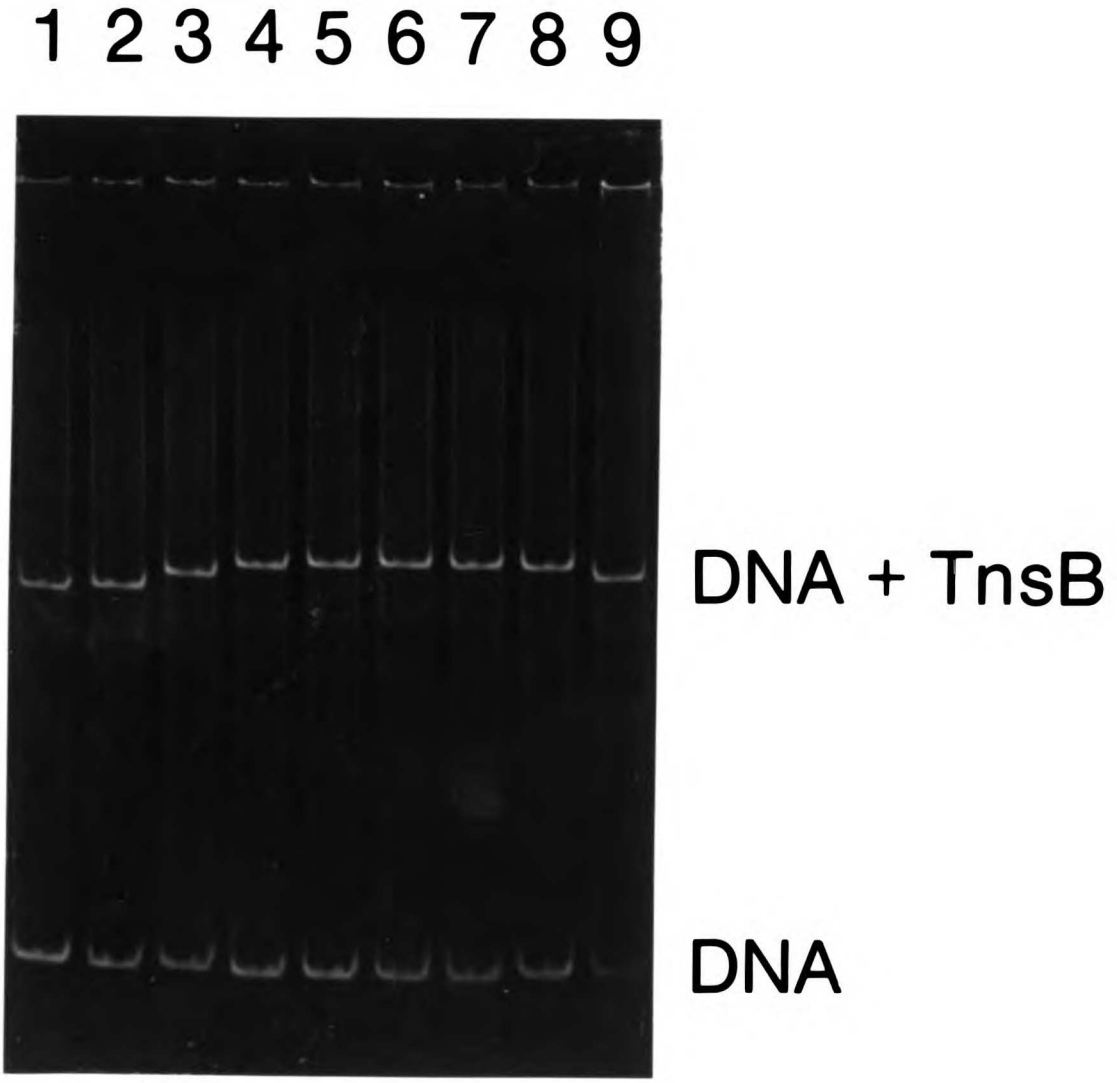
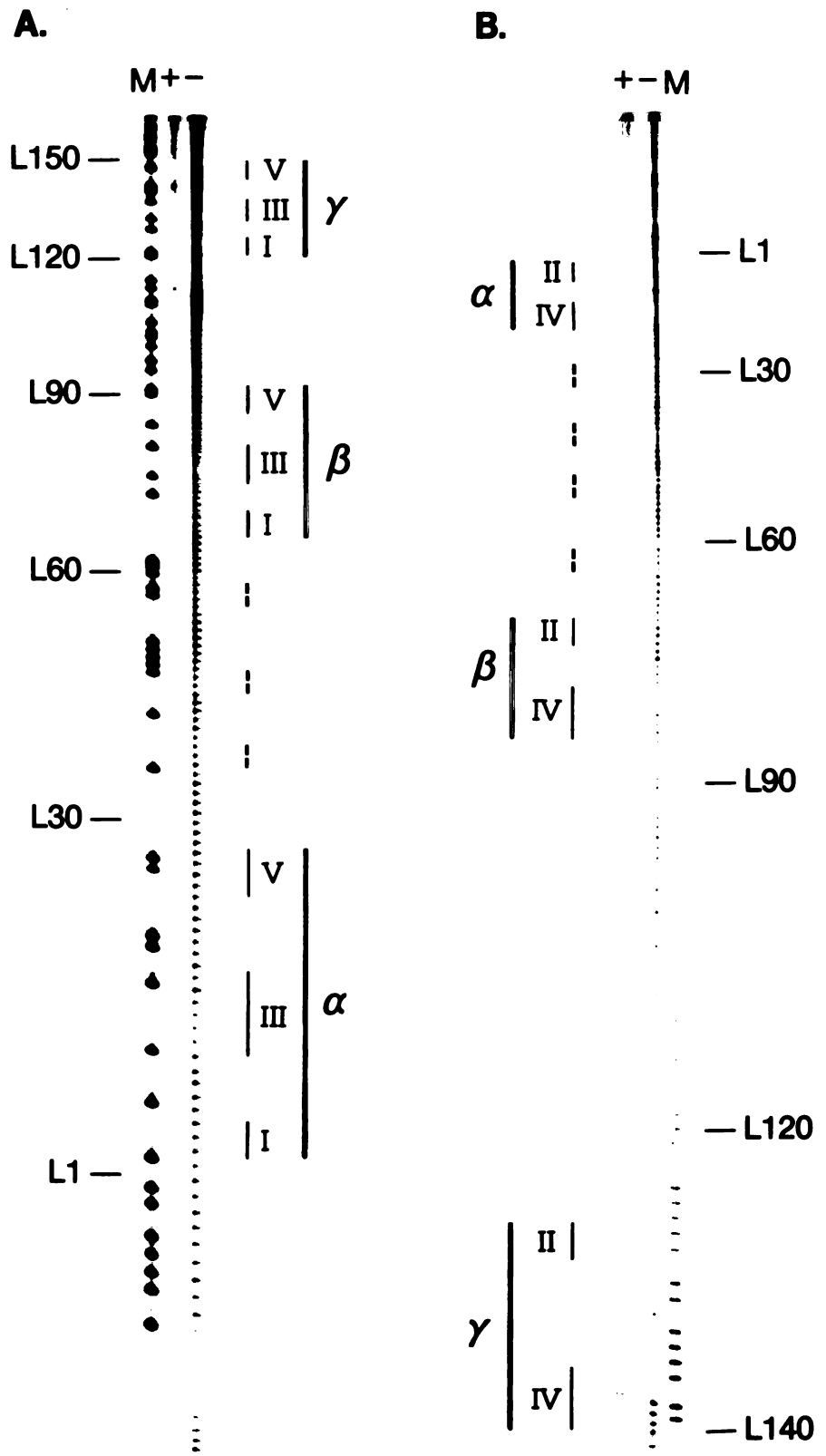

1 2 3 4 5 6 7 8 9
DNA + TnsB
DNA

Figure 4.4 Interaction of TnsB with intact Tn7 ends evaluated by hydroxyl radical footprinting

Hydroxyl radical attack of Tn7 end fragments, Tn7 L1-166 and Tn7 R1-199, was performed in the presence (+) or absence (-) of 1.5 μg/ml TnsB. Panel A, bottom strand of Tn7L; panel B, top strand of Tn7L; panel C, bottom strand of Tn7R and panel D, top strand of Tn7R (the top and bottom strands are as shown in Figure 5). A + G chemical sequencing reactions were used as markers. The nucleotide positions within the Tn7 ends are indicated, L1 and R1 being the terminal nucleotides of Tn7L and Tn7R, respectively. Protection clusters (I, II,III, IV and V) are indicated and marked with thin lines. The TnsB binding sites (α, β and γ in Tn7L and ϕ, χ, ψ and ω in Tn7R) are indicated with thick lines. The broken line indicates a region of weak protection between the α and β sites in Tn7L.

Tn7L


A.
M+−
L150
L120
L90
L60
L30
L1
V
III
I
γ
V
III
I
β
V
III
I
α
B.
+−M
L1
L30
L60
L90
L120
L140
II
IV
α
II
IV
β
II
IV
γ

Tn7R

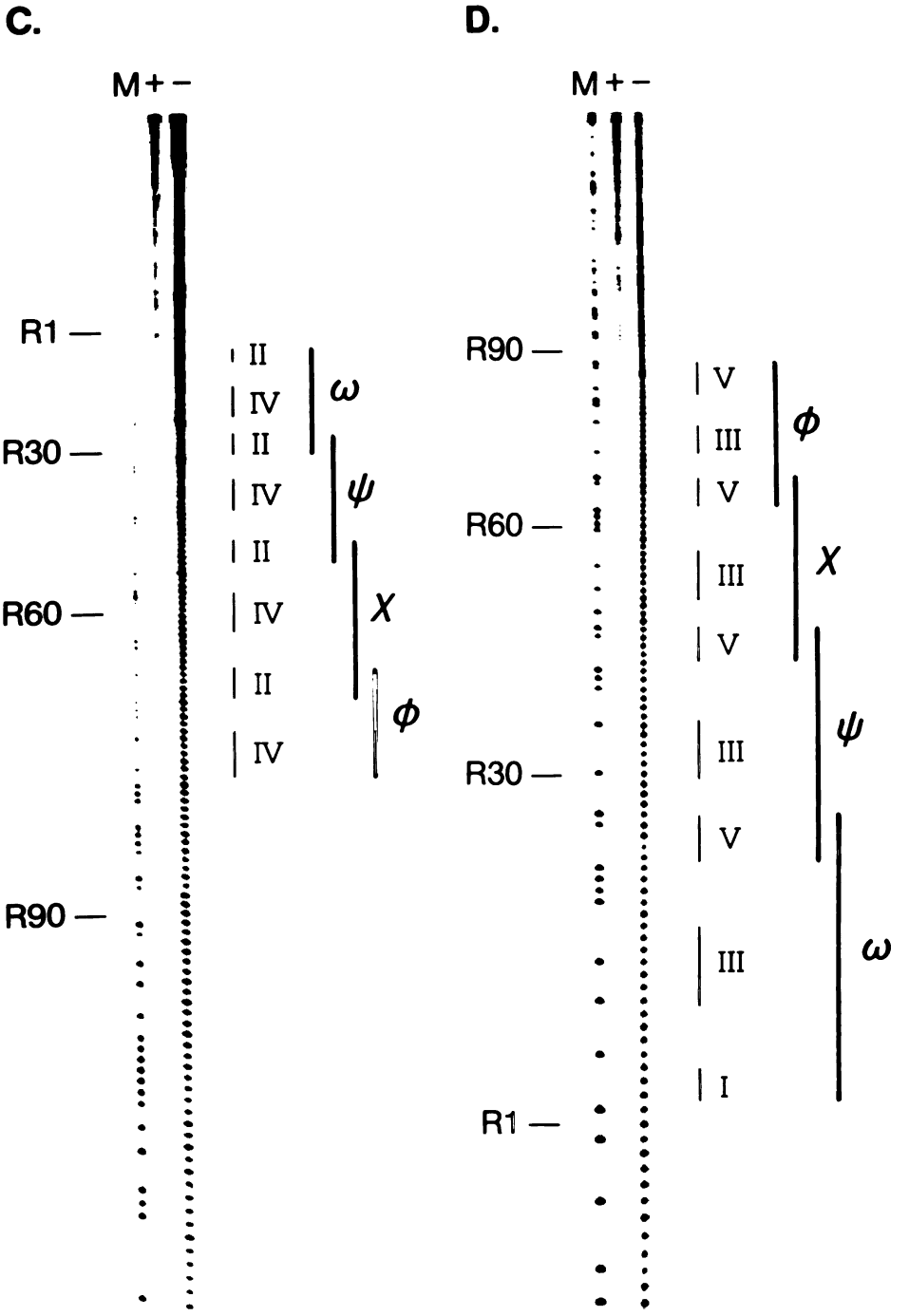
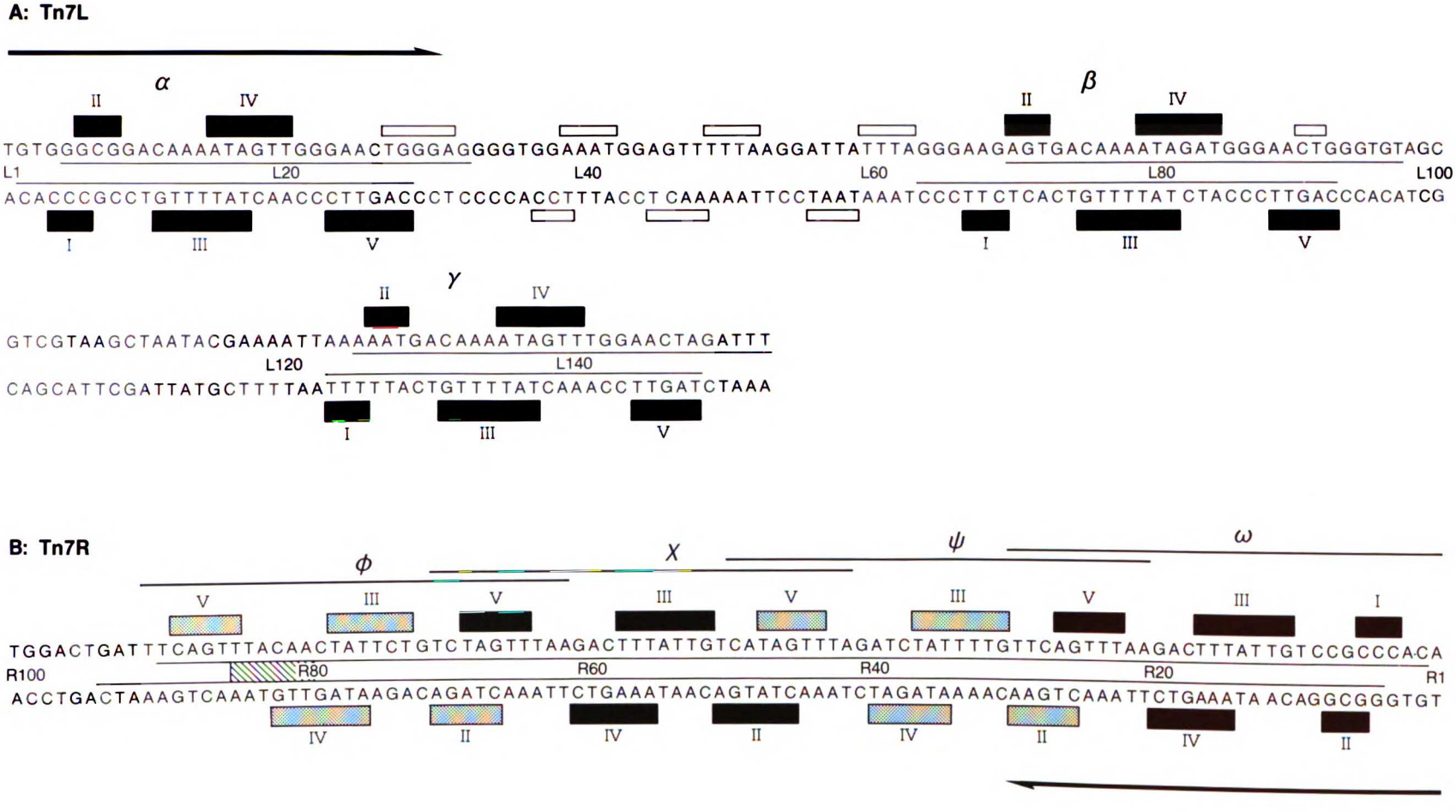

C.
M + −
R1 —
R30 —
R60 —
R90 —
II
IV
II
IV
II
IV
II
IV
ω
ψ
χ
φ
D.
M + −
R90 —
R60 —
R30 —
R1 —
V
III
V
III
V
III
V
III
I
φ
χ
ψ
ω

Figure 4.5 Summary of hydroxyl radical protection analysis of intact Tn7 ends.

The nucleotide sequence of Tn7 L1-166 (panel A) and Tn7 R1-100 (panel B) which contain the cis-acting sequences for Tn7 transposition (Arciszewska et al, 1989) are shown. L1 and R1 are the terminal nucleotides of each end; the bold arrows mark the 30 bp terminal inverted repeats. Regions protected against DNase I attack (Arciszewska et al, in preparation) are indicated by thin lines between the strands; the TnsB binding sites (α, β and γ in Tn7L and ϕ, χ, ψ and ω in Tn7R) are indicated. Clusters of nucleotides protected by TnsB against hydroxyl radical attack (I, II, III, IV and V) are marked with boxes (black in Tn7L and black and gray in Tn7R). The protection clusters within the intact ends correspond to those observed at a single TnsB binding site (Figures 1 and 2) except that in Tn7R, a single cluster of protection (V) is observed in sites ϕ, χ and ψ at the positions corresponding to clusters I and V of the isolated γ site because of the overlap between these sites. The open boxes indicate regions of weak protection. The -35 region of the putative tnsAB promoter in Tn7R (Gay et al, 1986; see text) is highlighted by a hatched box. Panel C shows the top strand of the consensus sequence for TnsB binding (with T or G indicated by K and A or G indicated by R (Arciszewska et al, in preparation). The

numbers below each nucleotide indicate the degree of conservation among the 7 characterized TnsB binding sites. The region of sequence identity that is most highly conserved is boxed. The horizontal arrows indicate a conserved 4 bp inverted repeat. The numbers above the sequence refer to nucleotide positions in a single TnsB binding site. Protection clusters (I, II, III, IV and V) are marked by open boxes.


A: Tn7L
α
β
γ
II
IV
I
III
V
TGTGGGCGGACAAAATAGTTGGGAACTGGGAGGGGTGGAAATGGAGTTTTTAAGGATTATTTAGGGAAGAGTGACAAAATAGATGGGAACTGGGTGTAGC
L1
L20
L40
L60
L80
L100
ACACCCGCCTGTTTTATCAACCCTTGACCCTCCCCACCTTTACCTCAAAAATTCCTAATAAATCCCTTCTCACTGTTTTATCTACCCTTGACCCACATCG
GTCGTAAGCTAATACGAAAATTAAAAATGACAAAATAGTTTGGAACTAGATTT
L120
L140
CAGCATTCGATTATGCTTTTAATTTTTACTGTTTTATCAAACCTTGATCTAAA
B: Tn7R
ϕ
χ
ψ
ω
TGGACTGATTTCAGTTTACAACTATTCTGTCTAGTTTAAGACTTTATTGTCATAGTTTAGATCTATTTTGTTCAGTTTAAGACTTTATTGTCCGCCCACA
R100
R80
R60
R40
R20
R1
ACCTGACTAAAGTCAAATGTTGATAAGACAGATCAAATTCTGAAATAACAGTATCAAATCTAGATAAAACAAGTCAAATTCTGAAATAACAGGCGGGTGT

C.

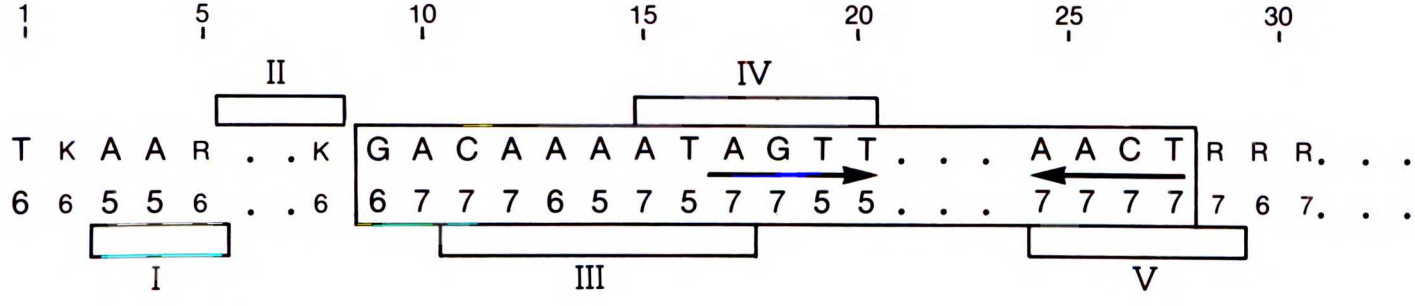

1
5
10
15
20
25
30
II
IV
T K A A R . . K G A C A A A A T A G T T . . . A A C T R R R. . .
6 6 5 5 6 . . 6 6 7 7 7 6 5 7 5 7 7 5 5 . . . 7 7 7 7 7 6 7. . .
I
III
V

Figure 4.6 DNase I footprinting of Tn7L at various TnsB concentrations

DNase I attack of a Tn7 L1-166 containing fragment was performed in the presence of TnsB at the indicated concentrations: lane 1, 0.16 μg/ml; lane 2, 0.31 μg/ml; lane 3, 0.62 mg/ml; lane 4, 1.25 μg/ml; lane 5, 2.5 μg/ml; lane D, no TnsB and lane M, A + G chemical sequencing reactions as markers. The Tn7 bottom strand (oriented as in Figure 5) is shown. The nucleotide positions are indicated with L1 being the terminal nucleotide of Tn7L. The regions of protection that define the α, β and γ binding sites are marked with a solid line. The broken line indicates that the region between α and β is somewhat protected by TnsB at the highest protein concentration used.

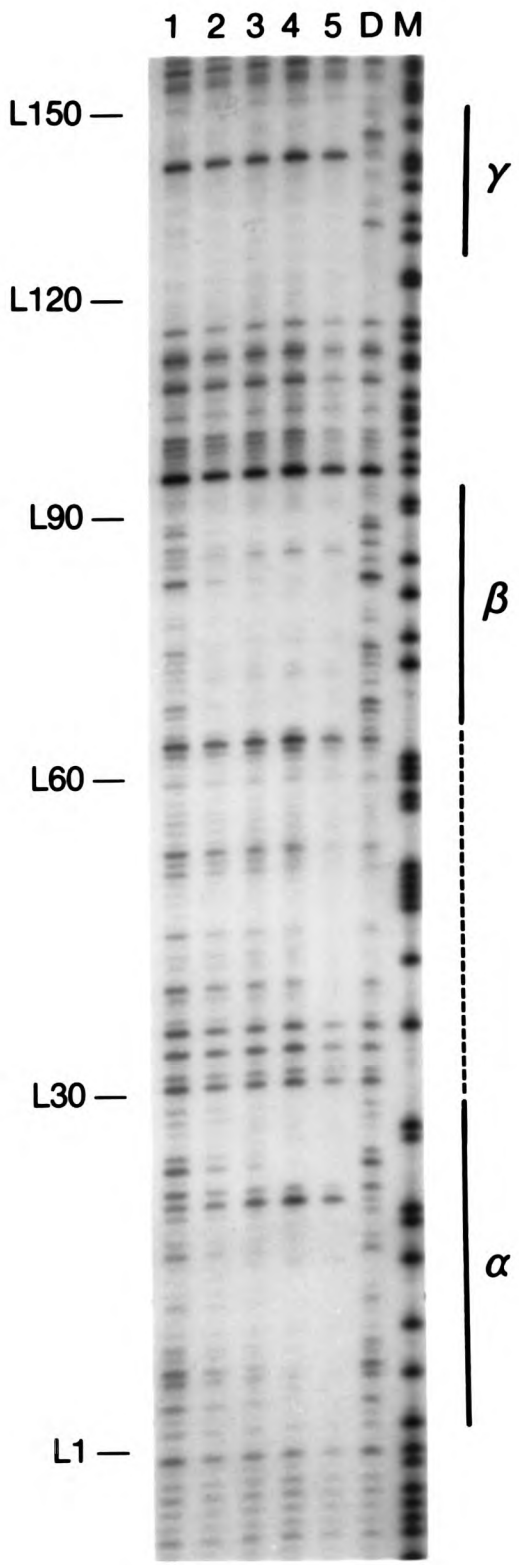

1 2 3 4 5 D M
L150
L120
L90
L60
L30
L1
γ
β
α

Figure 4.7 Protein-DNA complexes formed between TnsB and DNA fragments containing Tn7L and Tn7R

Protein-DNA complexes were formed in reactions containing the indicated amounts of TnsB and a Tn7L fragment (left panel) and a Tn7R fragment (right panel) and analyzed by native polyacrylamide gel electrophoresis. Three protein-DNA complexes (LI, LII and LIII) were observed with the Tn7L fragment and four complexes (RI, RII, RIII and RIV) were observed with the Tn7R fragment. Lane 1, no addition; lane 2, 0.31 μg/ml TnsB; lane 3, 0.62 μg/ml; lane 4, 1.25 μg/ml and lane 5, 2.5 μg/ml.

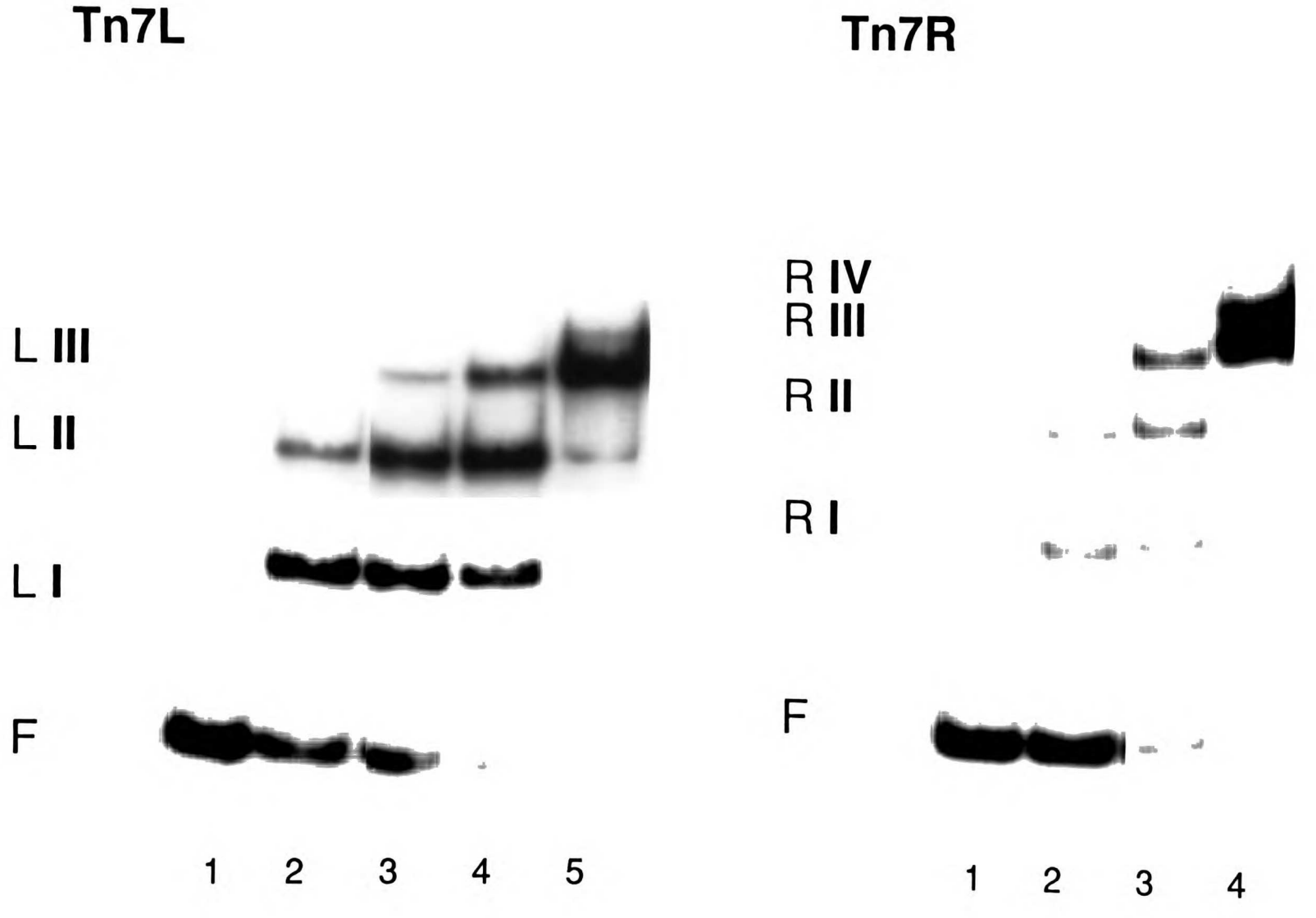

Tn7L
L III
L II
L I
F
1
2
3
4
5
Tn7R
R IV
R III
R II
R I
F
1
2
3
4

Figure 4.8 *in situ* phenanthroline-copper footprinting of TnsB-Tn7 end complexes

Gels containing the TnsB-DNA complexes decribed in Figure 7 were subjected to phenanthroline-copper attack and extracted DNA analyzed on DNA sequencing gels. The left panel (marked Tn7L) displays DNAs from Tn7L complexes. Lane M, A + G sequencing reaction as markers; lane I, complex LI; lane II, complex LII, lane III, complex LIII and lane C, no TnsB. The right panel (marked Tn7R) displays DNAs from Tn7R complexes: lane C, no TnsB; lane I, complex RI; lane II, complex RII; lane III, complex RIII, lane IV, complex RIV and lane M, A + G sequencing reaction as markers. In both panels, the nucleotide positions with Tn7L and Tn7R are indicated with L1 and R1 being the terminal nucleotides of each end. The positions of the TnsB binding sites (α, β and γ in Tn7L and ϕ, χ, ψ and ω in Tn7R) are shown.

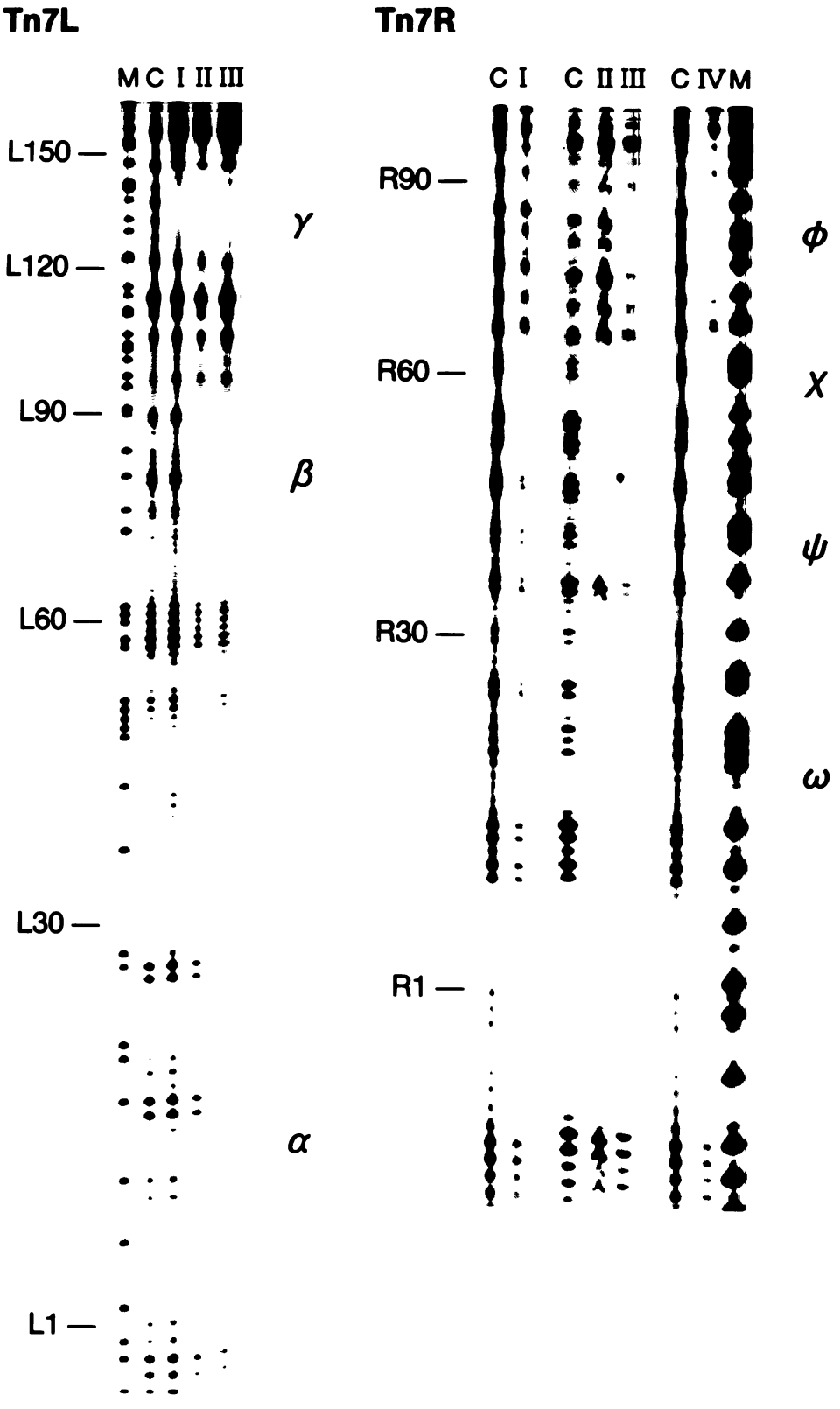
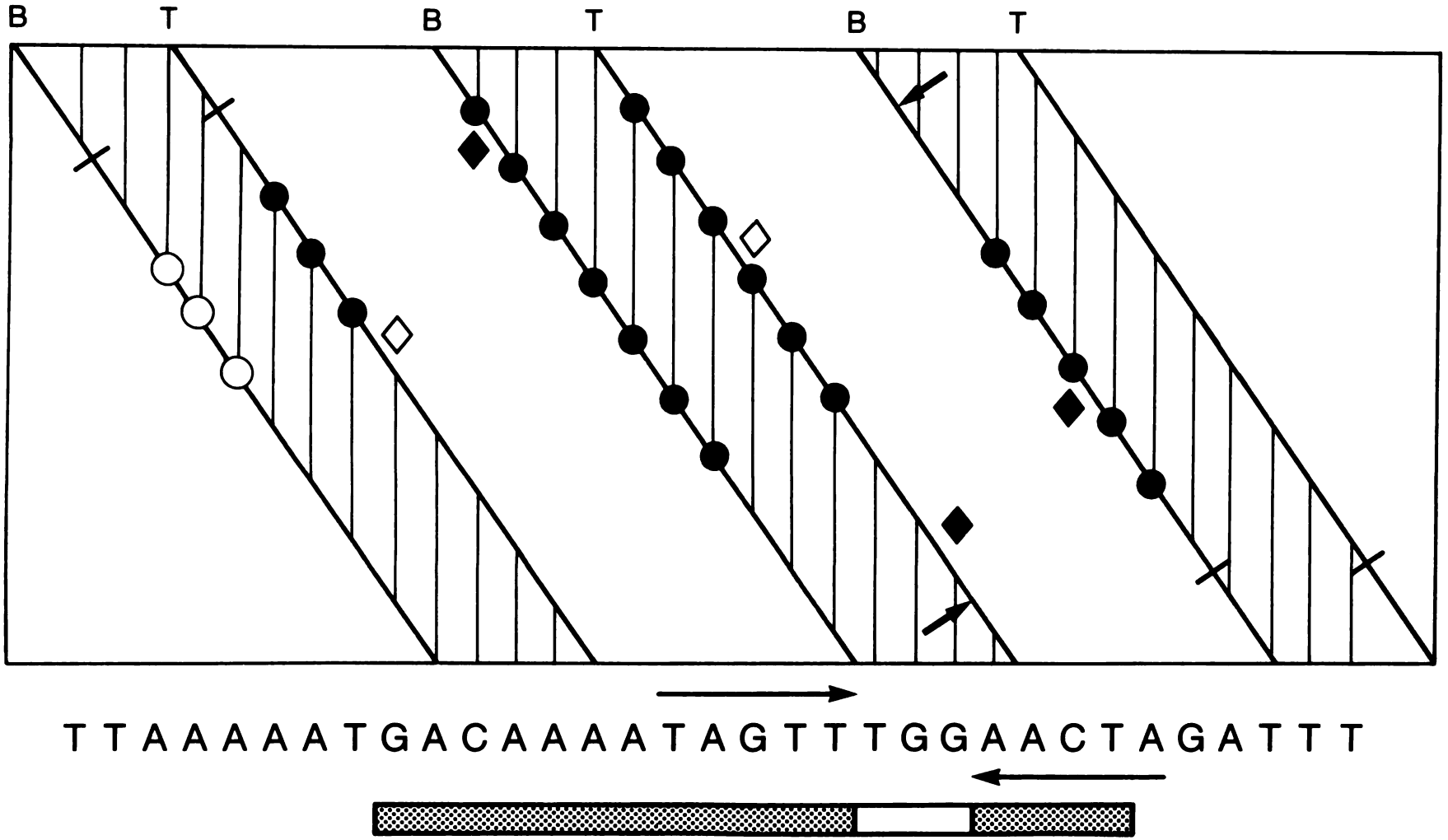

Tn7L
M C I II III
L150
L120
L90
L60
L30
L1
γ
β
α
Tn7R
C I
C II III
C IV M
R90
R60
R30
R1
φ
χ
ψ
ω

Figure 4.9 The interaction of TnsB with a single binding site

A DNA helix is diagrammed in planar representation (Siebenlist and Gilbert, 1980). The position of each base pair is indicated by a vertical line across the minor groove. The bottom (B) and top (T) strands are indicated. The nucleotide sequence of the top strand of the γ TnsB binding site is shown, horizontal arrows marking its 5 bp inverted repeats. The gray boxes below the sequence indicate the positions highly conserved in sequence identity among the TnsB binding sites; the open box indicates the 3 bp spacer between the conserved inverted repeats. Marked on the DNA helix are positions where alterations are observed in the presence of TnsB using a variety of probes. Filled circles indicate likely positions of close contact between TnsB and DNA as revealed by hydroxyl radical footprinting and by binding interference by DNA modified by hydroxyl radical attack, i.e. missing nucleosides. Open circles indicate positions protected from hydroxyl radical attack but whose absence does not obviously effect TnsB binding. Bases at which alkylation interferes with TnsB binding are marked with diamonds; filled diamonds indicating a strong block to TnsB binding and open diamonds a modest block to binding. The extent of the γ binding site as defined by protection

against DNase I attack is indicated by lines across the phosphodiester backbone; positions of hypersensitivity of DNase I attack are indicated by arrows (Arciszewska et al., in preparation).


B
T
B
T
B
T
TTAAAAATGACAAAATAGTTTGGAACTAGATTT

CHAPTER V

Conclusions

The major accomplishment of the research work for this dissertation has been understanding of one of the central steps in Tn7 transposition reaction - the recognition of the transposon ends by a transposition protein. To achieve this goal I have established genetically the DNA sequences at the Tn7 ends that are directly involved in the transposition reaction. Then I have purified the transposon-encoded TnsB protein and determined that it binds specifically to the ends. Finally, I have performed high resolution biochemical studies to characterize the interactions between TnsB and the Tn7 ends. Tn7 is the first transposable element whose protein-DNA interactions have been studied biochemically in such great detail. The major conclusions as well as detailed discussions of the results are presented in Chapters II, III, and IV. Here I would like to include only a few final comments that I have on this subject.

The termini of Tn7 are specifically recognized by the Tn7-encoded transposition protein TnsB. Several lines of evidence support the idea that TnsB binding to the ends is important for transposition. Firstly, the presence of the *tnsB* gene is required in all transposition events *in vivo* (Rogers et al., 1986; Waddell and Craig, 1988). Secondly, TnsB is absolutely necessary for transposition *in vitro* (R. Bainton, P. Gamas and N. Craig, in preperation). Finally, deletions of the ends that remove the TnsB binding site dramatically decrease the frequency of transposition (Chapter II). In the

light of abundant information regarding transposase binding in other systems (see Chapter I), these data strongly suggest, though do not unequivocally prove, that TnsB binding to Tn7 ends is essential for transposition.

The interactions of TnsB with DNA at the ends are very complex; each end of the transposon contains several binding sites for the protein (Chapter III, IV). Why the need to be so complex? Most bacterial transposons carry only a single transposase binding site located within the terminal inverted repeat (Ichikawa et al., 1987; Wiater and Grindley, 1988). Transposase interactions with the single terminal site at each end appear to be sufficient to ensure the formation of a synaptic complex and allow the precise cleavage at the transposon terminal nucleotides (Huisman et al., 1989; Haniford et al., 1990).

Several transposons contain multiple binding sites for transposition protein at the ends. These are bacteriophage Mu (Craigie et al., 1984) and eukaryotic transposons Ac and En/Spm (Kunze and Starlinger, 1989; Gierl et al, 1988). The complexity of the TnsB interactions with Tn7 ends is most reminiscent of MuA protein interactions with the ends of phage Mu. Interestingly, phage Mu and Tn7 have one property in common: they are both capable of very high efficiency transposition. During lytic growth, phage Mu can generate, using the transposition reaction, approximately 100 copies of its genome (Pato and Waggoner, 1981). Tn7, when introduced

into a naive cell, is able to transpose to its specific target site with up to 100% efficiency (Barth et al., 1976; Barth and Datta, 1977; L. Arciszewska and N. Craig, unpublished results). It is possible that the complex interactions of the TnsB protein with the ends might be necessary to secure high frequency site-specific Tn7 transposition. It is important to remember though, that both of Tn7's transposition pathways, the high frequency specific and low frequency nonspecific pathways, require the same transposition sequences at the ends. The latter statement, however, does not contradict the previous suggestion.

Alternatively, the complexity of the TnsB binding to the ends may reflect the requirement for some unique architecture of the Tn7 synaptic complex. A similar requirement has been observed in the Tn3 resolvase-mediated, site-specific recombination reaction (Sherratt, 1989).

Finally, the requirement for several protein binding sites may reflect low affinity of a transposition protein for a single binding site and thus a need for cooperative interactions of several protein molecules bound to multiple binding sites. This is likely to be the case in plant transposons Ac and En/Spm (Kunze and Starlinger, 1989; Gierl et al., 1988). In contrast, binding of the MuA protein to the ends of bacteriophage Mu appears to be noncooperative (Groenen et al., 1987). At the present time there is no available evidence regarding cooperativity of TnsB binding to the Tn7

ends. I have not found any indication of cooperativity in my binding studies but my experiments did not specifically address this issue.

What is the role of TnsB in the transposition reaction? Binding of TnsB to several sites in the Tn7 ends indicate that the protein may promote the formation of the specific, higher order nucleoprotein structure of the synaptic complex in which the Tn7 ends are held together. Formation of such a complex has been well documented in other recombination systems, like phage Mu (Pato, 1989), λ phage intasome (Thompson and Landy, 1989), and resolvase (Stark et al., 1989). This idea is additionally supported by the fact that TnsB induces DNA bending upon binding (Chapter IV). Although the degree of the bend is modest, TnsB-induced bending of the multiple sites within each Tn7 end may result in dramatic change in the DNA structure of the ends, and certainly could contribute to the complex formation. It would be interesting to know whether TnsB alone is sufficient to promote the assembly of the complex, or whether the presence of accessory proteins such as HU or IHF, or perhaps another Tns protein, is necessary for this step. The role of the host-encoded proteins in Tn7 transposition is amenable to analysis.

My studies determined that TnsB binds to the multiple sites in both Tn7 ends in a sequential fashion, beginning from the internal binding sites. This may indicate that the TnsB-mediated assembly of the synaptic complex is a very precise

process, and that protein interactions at the extreme termini, leading to the cleavage of transposon from flanking donor DNA, are very carefully controlled.

Does Tn7 carry other functions associated with transposases? This remains a mystery. I was unable to detect TnsB-dependent DNA nicking activity with supercoiled DNA substrates containing a transposition-proficient Tn7 end (unpublished). It is possible, that another Tns protein, for example TnsA or TnsC, both of which are required for both transposition pathways, is responsible for cleaving the Tn7 away from its flanking DNA. Alternatively, their presence may be required to stimulate TnsB-cleavage activity. Tn7 encodes several proteins whose functions are to capture target DNA for the transposition reaction. The TnsD protein recognizes the specific target site in *Escherichia coli* chromosome, *attTn7* (Waddell and Craig, 1989; K. Kubo and N. Craig, in preparation). By analogy, TnsE may interact nonspecifically with the random DNA target sites. TnsC is a nonspecific DNA binding protein (P. Gamas and N. Craig, unpublished results) and has been shown to interact with TnsD and *attTn7* DNA in the formation of the preinsertion complex (R. Bainton, K. Kubo and N. Craig, unpublished results). At present, there is no indication as to which of these proteins makes the staggered cut at the insertion site, and which catalyzes the strand transfer reaction.

The two ends of Tn7 are structurally different. The left end contains three TnsB binding sites separated by unrelated sequences; the right end contains four overlapping binding sites (Chapter IV). The difference in structure of Tn7 ends is not surprising; in all other transposons that contain repeated DNA sequences at each end, the several DNA sequence motifs/binding sites are arranged differently in each end (Craigie et al., 1984; Gierl et al., 1988; Kunze and Starlinger, 1989; Mullins et al., 1990). Surprisingly though, the ends of Tn7 are also functionally distinct (Chapter II). Other structurally different ends, for example in phage Mu, Ac and P element, are not interchangable; that is, one end cannot substitute for another (Mizuuchi and Craigie, 1986; Coupland et al., 1989; Mullins et al., 1990).

In Tn7, the right end can substitute for the left end but the left end cannot substitute for the right. In other words, the mini elements composed of two right Tn7 ends transpose while the elements composed of two left Tn7 ends do not. I have also found that the ends function differently in other transposition-related phenomena, like transposition immunity and "trans" inhibition of transposition (Chapter II), the right end always being "active" and the left being "inactive".

Why do the ends function differently? The finding that differences in function of the ends are also reflected in processes other than transposition itself suggest that the ends may differ quantitatively or qualitatively in protein

binding. One possibility is that the right end contains a site (s) for interaction with another Tns or host protein. Alternatively, the functional difference between the ends may simply reflect their different affinities for TnsB protein. The band shift binding experiments employing linear DNA fragments containing either the left or the right end did not reveal differences in TnsB binding (L. Arciszewska and N. Craig, unpublished results). However, the difference may only be evident with supercoiled DNA substrates. Does Tn7 utilize the difference between the ends in transposition? Tn7 inserts into *attTn7* in a unique orientation. The intriguing possibility is that the differential structure of the ends could determine the orientation specificity of Tn7 insertion.

Many transposable elements have developed complex systems for regulating their transposition. Amazingly, very little is known about the regulation of Tn7 transposition. My studies suggest that the TnsB protein may be involved in three regulatory processes: all of them are likely mediated by interactions of the protein with transposition sequences at Tn7 ends. These processes are: i) control of the expression of Tn7 transposition genes, ii) direct inhibition of the transposition reaction mediated by TnsB derived polypeptides (not well documented), and iii) transposition immunity. The first two processes were elaborated upon in the discussion section of Chapter III. Transposition immunity warrants additional comment. The mechanism of Tn7 immunity is not

known. The fact that Tn7 encodes proteins which interact with DNA of the target (TnsD/TnsE and TnsC) is very suggestive that the mechanism of immunity is similar to the one observed in Mu (Adzuma and Mizuuchi, 1988, 1989). The protein(s) interacting with the Tn7 end in a target would cause dissociation of a target-capturing protein(s) limiting the target availability for insertion. My studies indicate that TnsB is involved in mediating immunity. However, it is possible that TnsB acts in conjunction with another protein to confer immunity.

In summary, my studies indicate that TnsB plays a pivotal role in transposition by binding to the ends of the transposon. It is likely that TnsB is instrumental in forming a high order transposition complex and is also involved in the cleavage and rejoining steps. I have implicated TnsB in regulation of transposition, and have argued that TnsB may also be a central player in transposition immunity.

REFERENCES

Adzuma, K., and Mizuuchi, K. (1988) *Cell* 53, 257-266

Adzuma, K., and Mizuuchi, K. (1989) *Cell* 57, 41-47

Bainton, R., Gamas, P., and Craig, N. L., manuscript in preparation

Barth, P. T., and Datta, N. (1977) *J. Gen. Microbiol.* 102, 129-134

Barth, P. T., Datta, N., Hedges, R. W., and Grinter, N. J. (1976) *J.Bacteriol.* 125, 800-810

Coupland, G., Plum, C., Chatterjee, S., Post, A., and Starlinger, P. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 9385-9388

Craigie, R., Mizuuchi, M., and Mizuuchi, K. (1984) *Cell* 39, 387-394

Gierl, A., Lütticke, S., and Saedler, H. (1988) *EMBO J.* 7, 4045-4053

Groenen, M. A. M., Vollering, M., Krijgsman, P., van Drunen, K., and van de Putte, P. (1987) *Nucl. Acids Res.* 15, 8831-8844

Haniford, D. B., Benjamin, H. W., and Kleckner, N. (1990) *Cell*, in press

Huisman, O., Errada, P. R., Signon, L., and Kleckner, N. (1989) *EMBO J.* 8, 2102-2109

Ichikawa, H., Ikeda, K., Wishart, W. L., (1987) *Proc. Natl. Acad. Sci. U. S. A.* 84, 8220-8224

Kunze, R., and Starlinger, P. (1989) *EMBO J.* 8, 3177-3185

Lichtenstein, C., and Brenner, S. (1981) *Mol. Gen. Genet.* 183, 380-387

Mizuuchi, K., and Craigie, R. (1986) *Ann. Rev. Genet.* 20, 385-429

Mullins, M. C., Rio, D. C., and Rubin, G. M. (1989) *Genes Develop.* 3, 729-738

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

Pato, M. L., and Waggoner, B. T. (1981) *J. Virol.* 38, 249-255

Rogers, M., Ekaterinaki, N., Nimmo, E., and Sherratt, D. J. (1986) *Mol. Gen. Genet.* 205, 550-556

Sherratt, D. (1989) in *Mobile DNA* Berg, D. E., and Howe, M. M. eds. American Society for Microbiology, Washington D. C., pp. 163-184

Stark, W. M., Boocock, M. R., and Sherratt, D. J. (1989) *Trends Genet.* 5, 304-309

Thompson, J. F. and Landy, A. (1989) in *Mobile DNA* Berg, D. E. and Howe, M. M. eds. American Society for Microbiology, Washington D. C., pp. 1-22

Waddell, C. S., and Craig, N. L. (1988) *Genes Develop.* 2, 137-149

Waddell, C. S., and Craig, N. L. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 3958-3962

Wiater, L. A., and Grindley, N. D. F. (1988) *EMBO J.* 7, 1907-1911