Annu. Rev. Microbiol. 1993. 47:945–63 Copyright © 1993 by Annual Reviews Inc. All rights reserved

# THE TN5 TRANSPOSON

W. S. Reznikoff

Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53706

KEY WORDS: transposase, end sequences, gene regulation, cis-activity

# CONTENTS

Introduction	946
The Regulation of Transposase and Inhibitor Protein Synthesis	949
The Transposase N-Terminal Sequence Is Critical for Sequence-Specific DNA	
Binding and Other Transposase Activities	952
Recognition of Terminal 19-Base Pair Sequences	954
pl Is a Cis-Active Transposase (and a Trans-Active Transposition Inhibitor)	957
Transposition May Be Linked to Chromosome Replication and/or Cell Division	959
Conclusion	960

#### Abstract

The bacterial transposon Tn5 encodes two proteins, the transposase and a related protein, the transposition inhibitor, whose relative abundance determines, in part, the frequency of Tn5 transposition. The synthesis of these proteins is programmed by a complex set of genetic regulatory elements. The host DNA methylation function, *dam*, inhibits transposase promoter recognition and indirectly enhances the transposition inhibitor promoter. The inhibitor lacks the N-terminal 55 amino acids of the transposase, suggesting that this sequence plays a key role in the transposition process. An intact N-terminal sequence is required for the transposase's recognition of the 19-bp end DNA sequences. This is the first critical step in the transposition process. Transposase-end DNA interaction is itself regulated by an intricate series of reactions involving several host proteins: DnaA, Dam, and Fis. The transposase is a unique protein in that it acts primarily *in cis* and inhibits its own activity in trans. Models to explain these properties are described. Finally circumstantial evidence suggests that transposition occurs preferentially from newly replicated DNA that has yet to be partitioned to progeny cells. This

timing of transposition is likely to have a selective advantage for the host and the transposable element.

# Introduction

Transposition is a recombination process in which DNA sequences termed transposable elements move from an original site on a DNA molecule to a new site on the same or on a different DNA molecule. In addition, transposable elements can cause, and are associated with, other types of genetic rearrangements such as deletions, inversions, and chromosome fusions. The genomes of prokaryotic and eukaryotic organisms contain these elements. One could consider them as an ancient genetic machinery for causing genomic rearrangements and, therefore, for facilitating genome evolution. In addition, their associated biochemical reactions are likely to be similar to other interesting events involving the interaction of proteins and DNA. For these reasons transposable elements are of considerable interest.

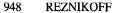
The transposable elements found in *Escherichia coli* fall into three general classes determined by their mechanisms of transposition. Transposons such as Tn3 and  $\gamma\delta$  transpose through a two-step replicative mechanism in which a cointegrate (fused replicon) structure is an intermediate. Transposons Tn10 and Tn903 transpose through a conservative cut-and-paste mechanism. Bacteriophage Mu and related viruses represent the third class of transposable element. In these cases the transposition can occur through either of the above two mechanisms, depending upon the proteins involved and the precise nature of the DNA strand cutting after an intermediate is formed between the transposable element and the target DNA sequence. Tn5 is generally assumed to transpose via a conservative mechanism (2); however, this presumption has not been critically tested. The reader is encouraged to examine the models and evidence for these transposition mechanisms in the recent monograph by Berg & Howe (3).

The conservative and replicative mechanisms of transposition share many basic characteristics. The transposable element encodes two critical functions required for the process—the end sequences and a protein termed the transposase. The element is defined by the specific sequences at its end. Transposition and related events remove the transposable element from its original sequence context precisely at the ends of these sequences. Changes in any base pair of these sequences typically reduces the frequency of or abolishes transposition. The transposable element also encodes a protein called the transposase. The transposase is a critical participant in many transposition functions including: specifically binding to the end sequences, bringing the two ends together through a protein oligomerization process, cutting or nicking the DNA adjacent to the end sequences, and inserting the transposable element DNA into a DNA target site. Host proteins also play critical roles in the transposition process such as facilitating the end-sequence binding of the transposase, nucleating the higher-order structure in which the ends are brought together, performing the necessary repair or replication functions, and regulating several steps in the transposition process.

Transposition is, in general, a quite rare, highly regulated process. Such tight regulation might enable the host cell to strike a balance between insuring proliferation of the transposable element and insuring the cell's own genetic survival—the very process of transposition causes chromosome breakage and rearrangements. The mechanisms of the regulation vary remarkably among the many cases studied, although some aspects are common to all, and the strategies, perhaps not surprisingly, are similar to ones found for other bacterial genetic systems. In addition to the role of host functions, transposable element encoded—functions often play a critical role in this regulation.

Implicit in the above general description is the fact that several very interesting molecular events are involved in and regulate the transposition process. Studying these events will help us understand other genetic processes. For instance, elucidating how transposition is regulated in a given system will tell us how complex DNA metabolizing processes might be controlled and give yet more examples of how gene expression can be modulated. The transposase is a protein that performs multiple complex functions. Protein structure/function studies would seek answers to the questions about the organization of the peptide domains that perform these functions and how they interact. The terminal DNA sequences are at first glance simply an example of a target for protein binding. However, their functionality is considerably more complex because they are often the target for more than one protein, and the protein-end sequence interaction is associated with at least two events: protein binding and DNA strand scission. Work done on other systems suggests that short DNA sequences can dictate several extremely intricate reactions, and transposable element end sequences may be a perfect example of such compact complexity. Transposable elements use host functions like biochemical parasites. Understanding how they use these functions will tell us much about the transposition process and about the functionality of the host functions themselves. Because many of these host functions are ones involved in host DNA metabolism, understanding their role in the transposition process may give us insights as to how host DNA metabolism is organized and regulated.

Finally, there are always surprising connections in biology. By studying transposition we embark on a largely unknown path into the cell's metabolism. For instance, if the host functions that Tn5 uses are organized in the cell in a unique spatial fashion, studying Tn5 transposition may reveal that arrangement.



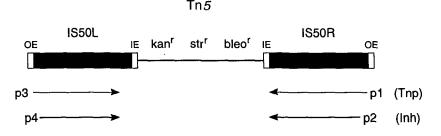


Figure 1 Transposon Tn5. Tn5 is a composite transposon in which genes encoding three antibiotic resistance proteins are bracketed by two IS50 elements, IS50L and IS50R. Both IS50 elements are delineated by 19-bp sequences, the inside end (IE) and the outside end (OE). IS50R encodes the transposase (Tnp or p1) and a second protein that inhibits transposition (Inh or p2). Tnp and Inh are translated in the same reading frame, but the Inh AUG is 55 codons downstream from the Tnp AUG. IS50L contains an ochre codon that results in the synthesis of p3 and p4, nonfunctional analogues of Tnp and Inh.

My laboratory and several other investigators have been studying the bacterial transposable element Tn5 as a model system. Tn5 is an example of a composite transposon in which antibiotic resistance genes are flanked by two nearly identical insertion sequences, IS50R and IS50L (see Figure 1 for a schematic). IS50R is a fully functional transposable element, while IS50L contains an ochre codon that results in the synthesis of inactive proteins (33).

IS50R encodes the transposase (p1 or Tnp) (15, 18, 19, 33), and the Tn5 Tnp has two opposing activities. *In cis* it acts as a transposase, catalyzing the transposition of the Tn5 or IS50 sequences from which it was encoded (15, 19). However, *in trans* it primarily acts as an inhibitor of transposition (42). The paradox of Tnp inhibiting the activity of other Tnp molecules is discussed in greater detail below. The Tnp inhibitory activity is one means by which Tn5 transposition is down-regulated.

A second protein [the inhibitor (Inh or p2)] is also encoded by IS50R (15, 18, 19, 45). Inh is translated in the same reading frame as Tnp but lacks the N-terminal 55 amino acids. Inh's only known function is to inhibit Tn5 transposition, which is thought to be the major means of down-regulating this process. The properties of Inh relative to the Tnp suggest that the N-terminal 55 amino acids of Tnp play an important role in its activity. The possible function of this sequence is an important topic of investigation, and this review discusses our current understanding of this function.

The relative abundance of Tnp and Inh plays a major role in determining the Tn5 transposition frequency. Interestingly, separate, apparently competing, promoters program Tnp and Inh syntheses (21, 43). Thus the regulation of these promoter activities becomes a crucial question. As discussed below, it is the host that regulates the relative activity of these two promoters, and the regulatory mechanism that has evolved appears to link the occurrence of transposition to the DNA replication process.

Tn5 has also evolved a mechanism for preventing the spurious synthesis of Tnp by virtue of accidental placement within an active transcription unit (21). The translation initiation signals were designed such that they will only function as part of a correctly initiated Tnp mRNA (21, 37). The strategy to accomplish this translation control may be widespread and is discussed below.

Each IS50 is bounded by two unique 19-bp end sequences [termed outside (OE) and inside (IE) ends] that are critical for transposition (17, 34). The distinguishing feature between IS50 and Tn5 transposition is the choice of the transposable element ends. Tn5 transposition utilizes two OEs, whereas IS50 transposition uses an OE and an IE sequence. Tnp binds to both of these sequences during the transposition process, and they likely perform other functions vital for Tnp activity. In addition, they are the sequences recognized by host functions. As we shall see, a host function binds to the OE to enhance transposition while the host functions affecting the IE down-regulate transposition. Two of the relevant host functions (DnaA and Dam) also link the transposition process to DNA replication.

Finally, we are just now beginning to obtain clues as to the possible relationship of Tn5 transposition with overall cell processes. Some of these clues, which point to DNA replication, have been suggested above. Others will become obvious during the discussion of the lethal effect of overproducing Tnp.

Douglas Berg (whose laboratory has contributed much of what we know about Tn5) recently published an excellent review of Tn5 (2). Therefore, this review is not a comprehensive treatment of the subject. Rather, this chapter concentrates on areas of current and future research interest raised by the questions implied above. The specific topics covered are:

- 1. How are the syntheses of Tnp and Inh regulated?
- 2. What are the possible functions for the Tnp N-terminal sequence?
- 3. What sort of protein recognition reactions occur at the 19-bp terminal sequences?
- 4. What is the molecular basis for the *cis*-active nature of the transposase?
- 5. How might transposition be linked to chromosome replication and/or cell division?

# The Regulation of Transposase and Inhibitor Synthesis

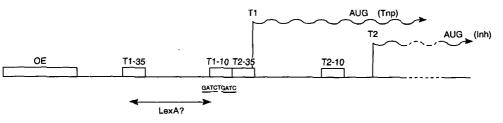
Studies by Biek & Roth (4) first indicated that the frequency of Tn5 transposition was regulated by Tn5-encoded functions. Tn5 sequences that were newly introduced into a cell transposed at dramatically lower frequencies in a cell already containing Tn5 as opposed to a cell lacking Tn5. In addition,

we (18) found that the Tn5 transposition frequency was constant regardless of the copy number of Tn5 (hence the transposition frequency of an individual Tn5 decreased in the presence of additional Tn5s). We now know that this down-regulation is primarily (but not entirely; see below) a consequence of the Tn5-encoded Inh protein. Inh functions *in trans* to block transposition. Moreover, the frequency of transposition is in part set by the abundance of Tnp and the ratio of Tnp to Inh. Thus the incoming Tn5 in the Biek & Roth experiment (4) encountered a preexisting pool of inhibiting Inh; as the copy number of Tn5 increases, the concentration of *trans*-acting Inh increases, but the amount of *cis*-acting Tnp per Tn5 remains constant.

What then determines the abundance of Tnp and Inh? The answers to this question are not only interesting for Tn5, they also reveal genetic regulatory motifs found in other systems.

Tnp and Inh are expressed from overlapping, probably competitive, promoters. This conclusion was deduced by a deletion analysis of in vivo promoter activities (21). As shown in Figure 2, the T2 transcript can only encode the Inh while the T1 transcript encodes both proteins [but the Inh is translated inefficiently from the T1 message (37)]. Such overlapping promoters are found in other systems (e.g. see 30) and can program contradictory functions. Thus by studying this Tn5 arrangement we will elucidate a more general regulation strategy.

DNA (*dam*) methylation down-regulates the synthesis of the T1 (Tnp) transcript and appears to up-regulate the synthesis of the T2 transcript (43). The frequency of Tn5 transposition is 10-fold higher in *dam* hosts, and this change in transposition seems to mirror (and presumably is the consequence of) a four- to fivefold increase in T1 mRNA and a twofold decrease in T2 mRNA. The opposite effect on promoter activity suggests that the promoters compete for RNA polymerase. An inspection of the DNA sequence indicates, and site-specific mutation studies prove, that the relevant GATC *dam* 



*Figure 2* Controlling elements for Tn5/IS50 Tnp and Inh synthesis. The 5' end of IS50 is defined by the 19-bp OE sequence (see Figure 4, below). The promoter for transposase synthesis (T1) initiates transcription 66 bp from the end of IS50 (21). Between the -35 and -10 regions of T1 is a weak LexA binding site (23). Overlapping the T1 -10 region are two Dam methylation sites that when methylated down regulate Tnp mRNA synthesis (43). The T2 (Inh) promoter overlaps T1 (21). The AUGs for Tnp and Inh are indicated (21).

methylation sites overlap with the -10 region of the T1 promoter. The specific relevance of *dam* regulation of Tnp (and Inh) synthesis is that it should couple transposition to DNA replication (about which more is discussed below) because newly replicated DNA is hemimethylated. Also, this regulation should stimulate transposition off of DNA that is newly introduced into the cell, which is exactly the observed result (27, 32).

Other transposon systems [in particular Tn10 (31)] also display Dam methylation control of Tnp synthesis and of transposition. Moreover, these studies demonstrate in a quite precise manner a type of gene regulation displayed in many systems—chemical modification of specific DNA sequences to modulate protein binding. This mechanism reappears in the discussion of protein recognition of the IS50 IE sequence.

Tn5 could transpose into highly expressed genes, thereby giving rise to read-through transcripts into the transposase gene. Production of these transcripts might result in spurious transposase synthesis. However, a set of experiments (21) has demonstrated that read-through expression of transposase does not occur because such messages do not program the translation of Tnp.

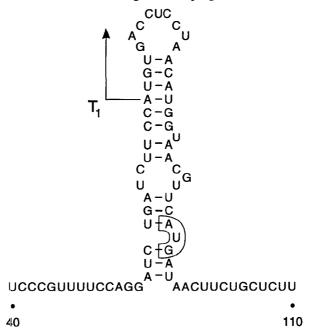


Figure 3 Secondary structure in read-through transcripts blocks Tnp translation. Transcripts that read through the end of IS50 do not express Tnp because of a secondary structure that occludes the Shine-Dalgarno sequence end and the AUG (21, 37). The start of the T1 transcript (which will not form a secondary structure) and the Tnp AUG are indicated. This figure is similar to one shown in Ref. 21.

Annual Reviews www.annualreviews.org/aronline

#### 952 REZNIKOFF

An RNA secondary structure in these read-through RNAs occludes the Tnp translation initiation signals (37). As shown in Figure 3, RNA sequences upstream of the T1 transcript start site are necessary for the formation of this secondary structure. This mechanism prevents read-through expression of Tn5 Tnp. The same mechanism appears to be shared by several other transposable elements [IS2 (13), IS3 (39), IS4 (20), IS5 (22), IS150 (38), and Tn10 (5)] and may be a general motif for *Escherichia coli* gene systems [for instance, the same pattern is seen for *lac* (36, 46) and *gal* (26)].

Do other mechanisms exist that regulate transposase synthesis? Recent reports present conflicting evidence regarding the possible role of LexA in regulating transposase synthesis. Our laboratory has found that LexA has no significant control over transposase levels (40), while Kuan & Tessman postulate that LexA represses transposase synthesis approximately fivefold, presumably by binding to a weak LexA binding site upstream of the T1 promoter (23) (see Figure 2). These studies differ in two respects: (a) Weinreich et al performed direct tests of a LexA effect (40), but the Kuan & Tessman studies were indirect (23), and (b) the sequence contexts of the two Tn5 systems were different. The former difference is a matter of technique, while the latter may suggest an interesting molecular phenomenon. Many genetic regulatory proteins are enhanced in their DNA-binding activities through cooperative binding to a secondary site (1). For instance, binding of the *lac* repressor to its operator is thought to be stabilized through a bond to a secondary binding site, thereby forming a looped DNA structure (10, 11). Future studies might examine whether such a nearby LexA binding site is present in the Kuan-Tessman system, but not in the one which we studied, and if so, whether such a fortuitous nearby LexA binding site could facilitate LexA binding to its weak Tn5 site.

# The Transposase N-Terminal Sequence Is Critical for Sequence-Specific DNA Binding and Other Transposase Activities

The transposase and its inhibitor are identical in sequence except that the Inh is missing 55 N-terminal amino acids (20). Because Inh does not promote transposition, it is missing one or more critical Tnp activities (15–17, 33, 45), suggesting that the N-terminal 55 amino acids encode a domain of importance. In vitro analyses of purified Tnp and purified Inh have shown that at least one property missing in the Inh preparations is a DNA sequence–specific binding activity. The Tnp, on the other hand, binds to OE and IE end sequences specifically and shows the expected sensitivity to OE sequence mutations and Dam methylation of the IE (N. de la Cruz, M. Weinreich, T. Wiegand, M. Krebs & W. Reznikoff, submitted; R. A. Jilk & W. Reznikoff, unpublished results). Recent studies have demonstrated that deletion of 11 N-terminal amino acids or introduction of 4 different N-terminal missense mutations destroys the DNA-binding activity of Tnp (M. D. Weinreich & W. Reznikoff, unpublished results). These experiments show that the N-terminal 55 amino acids are necessary for the sequence-specific DNA-binding activity and imply that the DNA-binding domain is in this region. Determining the precise DNA-binding domain will be instructive because the Tn5 transposase does not contain a previously classified DNA-binding motif, and yet the activity is clearly there.

Another surprising property requires the N-terminal three amino acids of Tnp. Overproduction of the transposase (in the absence of transposition) kills host cells. Deletion of as little as three amino acids prevents this killing (M. D. Weinreich & W. Reznikoff, unpublished results). A possible biological

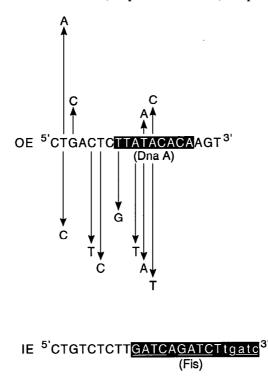


Figure 4 (Top) Outside and inside end sequences. The DnaA box is indicated for the outside end sequence (12, 17, 34, 44). Downward-pointing arrows denote mutations that act like a deletion in the adjacent deletion experiment. Upward-pointing arrows denote mutations that constitute the second class in the adjacent deletion experiment (16). The 11A change gave rise to one deletion in the second class. (Bottom) For the inside end sequence, the Dam sites and the overlapping Fis binding site are indicated (34, 41). Dam site methylation regulates both Tnp binding (43) and Fis binding (41). The lower-case letters indicate base pairs outside of the required 19-bp sequence. The Tnp/Inh termination codon is the complement to positions 12–14.

significance for the cell killing will be described in the section on chromosome replication below.

# **Recognition of Terminal 19–Base Pair Sequences**

The Tn5 and IS50 transposable elements are defined by the terminal 19-bp sequences. Tn5 is delineated by two inverted OE sequences and IS50 by an OE and an IE sequence (see Figure 4) (17, 34). All transposase-mediated genetic events (save one mentioned below) occur at the precise boundary of the relevant 19-bp sequences, and almost all changes in these sequences greatly reduce or abolish the frequency of transposition (24, 29). Presumably the importance of these sequences lies in the fact that they are recognized by proteins during the transposition process.

Although we are not certain of all of the salient facts, the protein-DNA interactions occurring at the OE and IE sequences clearly display a remarkable degree of complexity; at least two proteins recognize the OE and three proteins recognize the IE. In addition, different portions of the OE sequence may play roles in different steps in the transposition process.

The OE end is recognized by both the transposase and the host protein DnaA (17, 29, 44). The recognition of OE is fundamental to the transposition process. The binds specifically to OE sequences in vitro as judged by gel retardation assays (42; N. de la Cruz, M. Weinreich, T. Wiegand, M. Krebs & W. Reznikoff, submitted). However, transposase recognition of the OE may be much more complex than a simple binding reaction. For instance, some base pairs could be recognized during the strand-cleavage reaction.

The first evidence that various OE sequences dictate different steps in the transposition process came from a complicated set of in vivo observations (16). The can catalyze deletions adjacent to its site of insertion starting at the end of the OE sequence. These are most easily detected in situations in which the OE far removed (distal) from the location of the adjacent deletion is defective, greatly reducing or blocking transposition. Two classes of adjacent deletions have been found, and the nature of the class depends upon the precise change in the distal OE. If the distal OE is deleted, then the adjacent deletions start immediately adjacent to the functional OE. Some OE point mutations (which totally block transposition) also fall into this class. Mutant OE sequences of this class that have been tested for transposase binding in vitro bind transposase with a lower affinity than the wild-type OE (R. A. Jilk & W. Reznikoff, unpublished results). The second class of adjacent deletion is unusual in that this type starts one bp removed from the OE. These adjacent deletions are found in association with another subset of distal OE mutations, some of which are only partially defective in transposition and do not seem to impair transposase binding in vitro (16; R. A. Jilk & W. Reznikoff, unpublished results). Figure 4 displays the distribution of bp defects. Because

different subsets of OE point mutations give rise to different classes of adjacent deletions, this implies a functional fine structure to the OE. This fine structure is also seen in the binding of Tnp to the OE. Some OE mutations disrupt Tnp binding while others do not. A possible explanation for some of the mutations that do not impair transposase binding is that they are recognized by a different protein, DnaA. However, this explanation does not fit some cases (2A, 3C) because the mutations are located outside of the DnaA box as defined by sequence homology (12, 44). On the other hand, some mutations that presumably alter transposase binding (8G, 10T, 11A) are within the DnaA box, suggesting that these base pairs may be recognized by both proteins.

A closer examination of the results summarized in Figure 4 suggests an additional level of complexity. At positions 2 and 12 different mutations are associated with different classes of adjacent deletion formation—perhaps in these two cases the same base pair is recognized in different ways at different steps in the transposition process. Possibly both DnaA and transposase recognize position 12, and different base pair changes may discriminate differentially between these two protein-recognition processes. For position 2, Tnp might make different molecular contacts with the same base pair at different steps in the overall reaction.

DnaA also recognizes the OE sequence. This observation was first made through footprinting studies of the OE (12). Later experiments tested the relevance of this binding, showing that Tn5 and IS50 transposition occurs at a reduced efficiency in *dna* hosts, and that a Tn5 derivative with a mutation in the OE DnaA box did not appear to manifest sensitivity to the host *dna* genotype (29, 44). Current experiments are examining the effect of DnaA on transposase binding and are attempting to determine the in vitro properties of DnaA binding to various OE mutant sequences.

Thus, some end-sequence base pairs are recognized in very complex ways, at different steps in the transposition reaction and/or by different proteins. A similar complexity for end sequences has been suggested for other transposable elements (7, 14).

The IE sequence is even more complicated. It contains recognition sites for two proteins (transposase and Dam DNA methylase); it overlaps a binding site for a third protein, Fis, and also encodes the termination codon for the transposase (see Figure 4). The IE sequence and the OE sequence differ at 7 out of 19 positions, yet the transposase binds in vitro to an unmethylated IE with an affinity resembling its binding to the OE (R. A. Jilk & W. Reznikoff, unpublished results). This Tnp-IE binding result is consistent with the observed transposition preferences (9, 25). This result is quite surprising because it suggests considerable flexibility in the DNA recognition domain of Tnp. Alternatively, we could hypothesize that the positions that differ are not recognized by Tnp either because they do not play a critical protein-rec-

ognition role [position 4 may be an example of this (24)] or because they are recognized by some other protein (e.g. DnaA). However, the in vivo adjacent deletion studies suggest that mutations at some of these dissimilar positions do reduce Tnp binding directly (16). Thus we conclude that Tnp specifically recognizes two quite diverse sequences—another reflection of the complex substructures of IE and OE.

Dam DNA methylation inhibits transposase binding to the IE. Transposition of IS50 is sensitive to Dam DNA methylation because of the inhibition of two separate sequence-specific protein-DNA interactions. As described above, transcription initiation programmed by the T1 (transposase) promoter is directly inhibited by Dam DNA methylation of sites overlapping the -10region (43). However, Dam DNA methylation also inhibits IS50 (but not Tn5) transposition even when the transposase synthesis is programmed by a Dam-insensitive promoter (tac), indicating a direct effect of Dam methylation on IE recognition during transposition (24, 29, 43). This effect is presumably mediated through the Dam sites within the IE (see Figure 4). We recently showed that Dam methylation of IE DNA fragments reduces transposase binding in vitro (R. A. Jilk & W. Reznikoff, unpublished experiments). This observation is consistent with direct Dam DNA methylation control over transposition reactions using the IE. Very similar effects of DNA methylation that regulate Tn10/IS10 transposition have also been observed (31). As noted below, DNA methylation regulation of transposition may suggest a coupling between transposition and DNA replication.

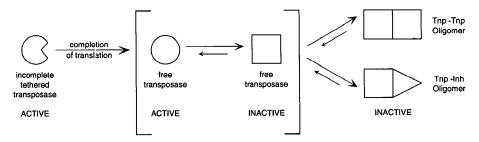
The role of Dam DNA methylation regulation of IS50 transposition (utilizing the IE) is complicated by the fact that a second protein (Fis) binds to a sequence overlapping the IE, and this binding (which is also blocked by Dam DNA methylation) is thought to compete with transposase-IE binding. Some in vivo experiments designed to study the frequency of IS50 transposition actually used various recombinant DNA constructs that have the same general structure as IS50 (OE-transposase gene-reporter gene-IE) but differ in the size of the IE sequence. These studies gave quite different results depending upon whether a 19- or a 24-bp sequence was used for IE. The 24-bp IE construct demonstrates much lower transposition frequencies in a dam host (9, 25). We now know that the 24-bp sequence (but not the 19-bp sequence) contains the overlapping Fis-binding site, and that Fis acts to inhibit transposition in a Dam-dependent manner (41). In vitro gel retardation and footprinting studies confirm that Fis binds to this sequence and that it binds efficiently only to the unmethylated site (41). The physiological significance of this Fis effect is unclear, but because Fis abundance varies with the stage of bacterial growth, it may act to dampen IS50 transposition from newly replicated DNA during the exponential phase, when it is most abundant.

The translation termination codon for the transposase (and its inhibitor) is located within the IE at positions 14–12. We have no information as to whether this influences transposition utilizing an IE or whether Tnp bound to the IE can influence the synthesis of the transposase.

# pl Is a Cis-Active Transposase (and a Trans-Active Transposition Inhibitor)

The Tn5 transposase functions primarily *in cis*, acting preferentially on OE-OE or OE-IE sequences located close to the site of transposase synthesis on the same relicon (15, 19). This observation has also been made for the transposase proteins encoded by other transposons such as Tn10 and Tn903 (8, 28). This preferential *cis* activity could result from either a chemical or functional instability in the transposase, and/or from a sequestration of the transposase during or shortly after synthesis. The Tn903 transposase is known to be chemically unstable (8), but various studies have indicated that Tn5 Tnp is, by and large, chemically stable in vivo (see 33).

My current hypothesis (pictured in Figure 5) is that the Tn5 Tnp is *cis*-active owing to protein conformation changes that are regulated by two factors; by the release of the Tnp protein from the translation apparatus and by the oligomerization of Tnp with monomers of Inh or Tnp. We propose that the nearly complete, tethered Tnp has a high affinity end-sequence binding activity and can initiate the transposition process prior to completion of its translation. Such a property would obviously lead to *cis* but not *trans* activity. The released completed Tnp is proposed to have two conformations in equilibrium, a low-abundance active conformation. Furthermore, the formation of Tnp-Tnp or Tnp-Inh oligomers probably occurs with the inactive conformation (or oligomerization inactivates the active conformation). The



*Figure 5* Transposase acts *in cis*. The Tn5 transposase functions primarily *in cis* (14, 18). Presumably, Tnp tethered to the translation apparatus can initiate the transposition reaction. Once Tnp synthesis is complete, the molecule favors an inactive conformation, and Tnp-Tnp or Tnp-Inh oligomerization prior to end-sequence binding will also inactivate Tnp.

longer a Tnp monomer exists without end-sequence binding the higher is the probability of oligomerization. Thus this property would also lead to *cis* activity.

The proposed model includes aspects of sequestration (the incomplete tethered Tnp is held in close proximity to its gene) and functional instability (a conformational change facilitated by oligomerization). The evidence for this model is quite circumstantial, but several of its predictions are clear.

The primary evidence supporting the notion that the tethered nascent protein may bind end sequences with an enhanced affinity came from gel retardation experiments examining the binding of Tnp subfragments to OE DNA. An in vitro synthesized Tnp subfragment lacking 100 carboxyl terminal amino acids binds specifically to OE DNA and does so with an affinity  $\sim$ 10-fold higher than that of full-length Tnp (L. Mahnke & W. Reznikoff, unpublished results). This observation is consistent with the OE-binding domain residing at the N terminus (see section 2 of this review) and suggests that this OE binding domain is occluded in most of the intact Tnp.

However, the completed Tnp is not completely inactive as Tnp does have some *trans* activity. We have recently described Tnp mutants that increase the frequency of transposition both *in cis* and *in trans* (42). These mutant proteins may have altered the equilibrium between active and inactive conformations of Tnp, and the mutations are believed to have a conformational effect because the mutant Tnp proteins are more sensitive to proteolytic degradation (T. Wiegand & W. Reznikoff, unpublished results).

In vivo studies suggest that protein oligomerization regulates Tnp activity. The Inh protein is known to inhibit transposase activity *in trans* (15, 18, 19, 45). Two simple models explain this inhibitory activity. Inh might bind to the end sequences forming inactive Inh-end sequence complexes, thereby blocking Tnp access to these sequences. However, Inh and other N-terminal deletions of Tnp cannot bind end-sequence DNA efficiently (N. de la Cruz, M. Weinreich, T. Wiegand, M. Krebs & W. Reznikoff, submitted); thus this model is not correct. Alternatively, Inh could form mixed oligomerize with Tnp and alter Tnp activity. Evidence suggests that Inh does oligomerize with Tnp and that the oligomers do have altered DNA-binding activity (N. de la Cruz, M. Weinreich, T. Wiegand, M. Krebs & W. Reznikoff, submitted).

Also suggesting that Tnp activity is regulated by oligomerization is the observation that Tnp itself inhibits transposition *in trans*. This property of Tnp was discovered during an in vivo analysis of a Tnp mutant that fails to make the Inh protein (42). The MA56 mutant destroys the Inh start codon and introduces an alanine in place of methionine at codon 56 of Tnp. This mutant is fully functional for transposition *in cis* but inhibits transposition *in trans*.

Although the model described above and in Figure 5 is consistent with all

the observations, direct experiments are needed to test it. For instance, do the hypertransposing mutants alter the conformation of Tnp? Will tethered, incomplete Tnp bind to OE DNA with a high affinity? Does oligomerization favor the inactive Tnp conformation?

# Transposition May Be Linked to Chromosome Replication

Some circumstantial evidence might link Tn5 transposition to the process of chromosome replication and partition. In particular the following host functions are known to influence the frequency of Tn5 transposition:

Dam: a protein catalyzing postreplicative DNA methylation; *dam* cells have elevated transposition frequencies (43).

DnaA: an *oriC*-binding protein required for the initiation of DNA replication (29, 44); *dnaA* cells have reduced transposition frequencies.

SulA: a protein that inhibits cell division; *sulA* cells have elevated transposition frequencies (35).

In addition, DeLong & Syvanen (6) reported that a small fraction of transposase is associated with the inner membrane fraction upon cell lysis. These correlations are intriguing but do not show a mechanistic link.

Observations that suggest a functional link between transposition and chromosome partition came from experiments analyzing the cellular consequences of transposase over expression. Cell death occurs in the absence of transposition as well as any other Tn5-encoded function (such as Inh or the end sequences) (M. D. Weinreich, unpublished results). The dying cells form long multinuclear filaments, indicating a failure in partition-dependent septation. *E. coli* mutants resistant to transposase overproduction killing continue to divide when transposase is overproduced. The mutations mapped using Hfr crosses reside at more than one locus. The one that has been most closely examined maps near the *fts* locus at 76 min (M. D. Weinreich & W. Reznikoff, unpublished results). *fts* genes encode functions required for septation.

These results point to a possible link between transposition and chromosome-partition function. This link is in addition to the preferential transposition of Tn5 or IS50 sequences off of newly replicated DNA (which results from *dam* regulatory effects) and to the sharing of DnaA in both transposition and replication processes. However, at least the link to chromosome partition can be bypassed; transposition does occur in the mutants resistant to transposase overproduction (M. D. Weinreich & W. Reznikoff, unpublished results).

How is this coupling accomplished and what is the advantage of this arrangement? We hope an analysis of the mutants mentioned above will suggest an answer to the first question. As for the second, many of the host functions participating in Tn5 transposition are assumed to be components of a sequestered organelle involved in host-chromosome replication. One possi-

ble advantage to Tn5 accruing from such an association is that a relationship between Tn5 and the same organelle might have evolved in order to insure that the organelle's transposition apparatus would have efficient access to these functions.

Alternatively, the proposed coupling of Tn5 transposition to chromosome replication and partition might decrease the risk of genomic "suicide" occurring as a result of the conservative cut-and-paste transposition process. Transposition is most likely to happen soon after DNA replication (when there are two copies of the donor DNA sequence) due to the influence of *dam* DNA methylation on transposase synthesis and inside-end usage. If, however, transposase inhibits chromosome partitioning, then this might delay cell division in the very cells in which transposition is most likely to have occurred, e.g. cells that contain transposase, assuring that the ongoing chromosome replication would be completed prior to cell division. Thus transposition would be biased towards cells in which a copy of the parental genome is maintained and can be inherited.

# Conclusion

The Tn5/IS50 system has been an amazingly productive object of experimental inquiry. The work to date has elucidated (and future work will continue to elucidate) the mechanisms by which transposition occurs and is regulated and will also continue to provide insights into important aspects of molecular biology.

One of the interesting aspects of Tn5 molecular biology is the density of information encoded within a short sequence. The IS50 sequence is 1533 bp in length. This element encodes all of the Tn5 functions required for transposition and all of the Tn5 regulatory mechanisms. Thus in the first 259 bp, starting at the OE, we find a complex Tnp-specific sequence, a DnaA-binding site, an imperfect symmetry element that blocks the expression of read-through mRNA, a promoter for Tnp mRNA synthesis, a possible LexA binding site, Dam methylation sites regulating the Tnp promoter activity, a promoter for Inh mRNA synthesis, translation initiation signals for Tnp and Inh, and sequences encoding the important N terminus of Tnp. Most of the rest of the sequence up to the Fis site, which overlaps the IE, is involved with encoding Tnp and Inh, although *cis*-active sites may also be in this region [e.g. a Fis binding site of unknown function (41)].

I have not discussed the bulk of the protein-coding sequences in this review because the required Tnp (and Inh) structure-function studies have not been performed. For instance, the Tnp presumably executes the following operations: binding of OE and IE, bringing the ends together (or inactivating the unbound Tnp) through oligomerization, cutting the DNA next to the end sequences, cleaving target DNA to give 9-bp 5' overhangs, and inserting the transposable element into target DNA. Each of the functions requires particular Tnp domains of critical importance. Where are they? How are they organized in a three-dimensional structure? Understanding the molecular basis of Tnp *cis* activity should also reveal fascinating insights into this protein's structure-function relationships. These and other questions will be the object of future experiments, which will require detailed genetic, biochemical, and structural analyses.

Even among the transposition operations for which we have substantial information, much work is left to be done. For instance, how does DnaA binding to the OE influence the Tnp-OE interaction? How do the different base pairs of the OE interact with Tnp and at what steps of the transposition reaction does this occur? How do the different base pairs of the IE function with regard to Tnp activity? How does Fis perturb the Tnp-IE interaction? What is the active form of the Tnp, and how many molecules are involved in synaptic complex formation?

Finally, research efforts should be directed at determining if transposition is coupled to the cell-division cycle and how that linkage is accomplished.

#### **ACKNOWLEDGMENTS**

The research in the author's laboratory has been supported by grants from the NIH (GM19670), the NSF (DMB-9020517), and the American Cancer Society (MV-554). The author is indebted to the many members of his laboratory who have contributed their imagination and work, and the editor of this volume for very helpful suggestions.

## Literature Cited

- Bellomy, G. R., Record, M. T. Jr. 1990. DNA loop formation: role in gene regulation and implications for DNA structure. In *Progress in Nucleic Acid Research and Molecular Biology*, ed. W. Cohn, K. Moldave, 39:81–128. New York: Academic
- Berg, D. E. 1989. Transposon Tn5. See Ref. 3, pp. 185-210
- Berg, D. E., Howe, M. M. 1989. Mobile DNA. Washington, DC: Am. Soc. Microbiol.
- Biek, D., Roth, J. R. 1990. Regulation of Tn5 transposition in Salmonella typhimurium. Proc. Natl. Acad. Sci. USA 77:6047-51
- Davis, M. A., Simons, R. W., Kleckner, N. 1985. Tn10 protects itself at two levels from fortuitous activation by external promoters. *Cell* 43:379-97
- 6. DeLong, A., Syvanen, M. 1990. Membrane association of the Tnp and Inh

proteins of IS50R. J. Bacteriol. 172: 5516-19

- Derbyshire, K. M., Hwang, L., Grindley, N. D. F. 1987. Genetic analysis of the interaction of the insertion sequence IS903 transposase with its terminal inverted repeats. *Proc. Natl. Acad. Sci. USA* 84:8049–53
- Derbyshire, K. M., Kramer, M., Grindley, N. D. F. 1990. Role of instability in the *cis* action of the insertion sequence IS903 transposase. *Proc. Natl. Acad. Sci. USA* 87:4048-52
- Dodson, K. W., Berg, D. E. 1989. Factors affecting transposition activity of IS50 and Tn5 ends. Gene 76:207-13
- Eismann, E., von Wilcken-Bergmann, B., Muller-Hill, B. 1987. Specific destruction of the second *lac* operator decreases repression of the *lac* operon in *Escherichia coli* fivefold. J. Mol. Biol. 195:949–52

- Flashner, Y., Gralla, J. D. 1988. Dual mechanism of repression at a distance in the *lac* operon. *Proc. Natl. Acad. Sci. USA* 85:8968–72
- Fuller, R. S., Funnell, B. E., Kornberg, A. 1984. The dnaA protein complex with the E. coli chromosomal replicative origin (oric) and other DNA sites. *Cell* 38:889–900
- Ghosal, D., Sommer, H., Saedler, H. 1979. Nucleotide sequence of the transposable DNA-element IS2. Nucleic Acids Res. 6:1111-22
- Huisman, O., Errada, P. R., Signon, L., Kleckner, N. 1989. Mutational analysis of IS10's outside end. *EMBO* J. 8:2101-9
- Isberg, R. R., Lazaar, A. L., Syvanen, M. 1982. Regulation of Tn5 by the right repeat proteins: control at the level of the transposition reactions? *Cell* 30:883–92
- Jilk, R. A., Makris, J. C., Borchardt, L., Reznikoff, W. S. 1993. Implications of Tn5 associated adjacent deletions. J. Bacteriol. 175:1264-71
- Johnson, R. C., Reznikoff, W. S. 1983. DNA sequences at the ends of transposon Tn5 required for transposition. *Nature* 304:280-82
- Johnson, R. C., Reznikoff, W. S. 1984. The role of the IS50R proteins in the promotion and control of Tn5 transposition. J. Mol. Biol. 177:645-61
- Johnson, R. C., Yin, J. C.-P., Reznikoff, W. S. 1982. The control of Tn5 transposition in Escherichia coli is mediated by protein from the right repeat. *Cell* 30:873–82
- Klaer, R., Kuhn, S., Tillmann, E., Fritz, H.-J., Starlinger, P. 1981. The sequence of IS4. Mol. Gen. Genet. 181:169-75
- Krebs, M. P., Reznikoff, W. S. 1986. Transcriptional and translational sites of IS50. Control of transposase and inhibitor expression. J. Mol. Biol. 192: 781-91
- Kroger, M., Hobom, G. 1982. Structural analysis of insertion sequence IS5. *Nature* 297:159-62
- Kuan, C.-T., Tessman, I. 1991. LexA protein of *Escherichia coli* represses expression of the Tn5 transposase gene. J. Bacteriol. 173:6406-10
- Makris, J. C., Nordmann, P. L., Reznikoff, W. S. 1988. Mutational analysis of IS50 and Tn5 ends. Proc. Natl. Acad. Sci. USA 85:2224-28
- Makris, J. C., Reznikoff, W. S. 1989. Orientation of IS50 transposase gene and transposition. J. Bacteriol. 171: 5212-14

- Merril, C. R., Gottesman, M. E., Adhya, S. L. 1981. Escherichia coli gal operon proteins made after prophage lambda induction. J. Bacteriol. 147: 875-87
- McCommas, S. A., Syvanen, M. 198. Temporal control of Tn5 transposition. J. Bacteriol. 170:889-94
- Morisato, D., Way, J. C., Kim, H.-J., Kleckner, N. 1983. Tn10 transposase acts preferentially on nearby transposon ends in vivo. *Cell* 32:799–807
- Phadnis, S. H., Berg, D. E. 1987. Identification of base pairs in the outside end of insertion sequence IS50 that are needed for IS50 and Tn5 transposition. Proc. Natl. Acad. Sci. USA 84:9118-22
- Reznikoff, W. S., Bertrand, K., Donnelly, C., Krebs, M., Maquat, L. E., et al. 1987. Complex promoters. In RNA Polymerase and the Regulation of Transcription, ed. W. S. Reznikoff, R. R. Burgess, J. E. Dahlberg, C. A. Gross, M. T. Record, Jr., M. P. Wickens, pp. 105–13. New York: Elsevier
- Roberts, D. E., Hoopes, B. C., Mc-Clure, W. R., Kleckner, N. 1985. IS10 transposition is regulated by DNA adenine methylation. *Cell* 43:117-30
- Rosetti, O. L., Altman, R., Young, R. 1984. Kinetics of Tn5 transposition. *Gene* 32:91-8
- Rothstein, S. J., Reznikoff, W. S. 1981. The functional differences in the inverted repeats of the Tn5 are caused by a single base pair nonhomology. *Cell* 23:191–99
- Sasakawa, C., Carle, G. F., Berg, D. E. 1983. Sequences essential for transposition at the termini of IS50. Proc. Natl. Acad. Sci. USA 80:7293-97
- Natl. Acad. Sci. USA 80:7293-97
  35. Sasakawa, C., Uno, Y., Yoshikawa, M. 1987. lon-sulA regulatory function affects the efficiency of transposition of Tn5 from \lambda b221 c1857 Pam Oam to the chromosome. Biochem. Biophys. Res. Commun. 142:879-84
- Schulz, V. P., Reznikoff, W. S. 1990. In vitro secondary structure analysis of mRNA from *lacZ* translation initiation mutants. J. Mol. Biol. 211:427-45
- Schulz, V. P., Reznikoff, W. S. 1991. Translation initiation of IS50R readthrough transcripts. J. Mol. Biol. 221: 65-80
- Schwartz, E., Kroger, M., Rak, B. 1988. IS150: distribution, nucleotide sequence and phylogenetic relationships of a new E. coli insertion element. Nucleic Acids Res. 16:6789-99
- 39. Timmerman, K. P., Tu, C.-P. D.

TN5 TRANSPOSITION 963

1985. Complete sequence of IS3. Nucleic Acids Res. 13:2127–39

- Weinreich, M. D., Makris, J. C., Reznikoff, W. S. 1991. Induction of the SOS response in *Escherichia coli* inhibits Tn5 and IS50 transposition. J. *Bacteriol.* 173:6910-18
- Weinreich, M. D., Reznikoff, W. S. 1992. Fis plays a role in Tn5 and IS50 transposition. J. Bacteriol. 174:4530– 37
- Wiegand, T. W., Reznikoff, W. S. 1992. Characterization of two hypertransposing Tn5 mutants. J. Bacteriol. 174:1229–39
- Yin, J. C.-P., Krebs, M. P., Reznikoff, W. S. 1988. The effect of *dam* methylation on Tn5 transposition. J. Mol. Biol. 199:35-46
- Yin, J. C.-P., Reznikoff, W. S. 1987. dnaA, an essential host gene, and Tn5 mutants. J. Bacteriol. 169:4637-45
- Yin, J. C.-P., Reznikoff, W. S. 1988, p2 and the inhibition of Tn5 transposition. J. Bacteriol. 170:3008-15
- Yu, X. M., Munson, L. M., Reznikoff, W. S. 1984. Molecular cloning and sequence analysis of *trp-lac* fusion deletions. J. Mol. Biol. 172:355-62



Annual Review of Microbiology Volume 47, 1993

# CONTENTS

Reflections of a Microbiologist, or How to Learn From the	
MICROBES, Norbert Pfennig	1
Regulation of Sulfur and Nitrogen Metabolism in	
FILAMENTOUS FUNGI, George A. Marzluf	31
AGROACTIVE COMPOUNDS OF MICROBIAL ORIGIN, Yoshitake Tanaka and Satoshi $\overline{O}$ mura	57
MOLECULAR ANALYSIS OF THE PATHOGENICITY OF STREPTOCOCCUS PNEUMONIAE: The Role of Pneumococcal Proteins, James C. Paton, Peter W. Andrew, Graham J. Boulnois, and Timothy J. Mitchell	89
MOLECULAR DETERMINANTS OF THE VIRULENCE AND INFECTIVITY OF	09
CALIFORNIA SEROGROUP BUNYAVIRUSES, Christian Griot,	
Francisco Gonzalez-Scarano, and Neal Nathanson	117
SUICIDAL GENETIC ELEMENTS AND THEIR USE IN BIOLOGICAL CONTAINMENT OF BACTERIA, S. Molin, L. Boe, L. B. Jensen, C.	
S. Kristensen, M. Givskov, J. L. Ramos, and A. K. Bej	139
TRANSPORT OF NUCLEIC ACIDS THROUGH MEMBRANE CHANNELS: Snaking Through Small Holes, Vitaly Citovsky and Patricia	
Zambryski	167
THE CELL CYCLE OF ESCHERICHIA COLI, William D. Donachie	199
CHROMOSOME SEGREGATION IN YEAST, Barbara D. Page and	
Michael Snyder	231
DISSIMILATORY METAL REDUCTION, Derek R. Lovley	263
ATP-DEPENDENT TRANSPORT SYSTEMS IN BACTERIA AND HUMANS: Relevance to Cystic Fibrosis and Multidrug Resistance, Carl A.	
Doige and Giovanna Ferro-Luzzi Ames	291
REGULATION OF THE HEAT-SHOCK RESPONSE IN BACTERIA, T. Yura, H. Nagai, and H. Mori	321
MOLECULAR BIOLOGY OF HYDROGEN UTILIZATION IN AEROBIC	
CHEMOLITHOTROPHS, B. Friedrich and E. Schwartz	351
THE SURFACE TRANS-SIALIDASE FAMILY OF TRYPANOSOMA CRUZI,	
George A. M. Cross and Garry B. Takle	385

vi (continued)

TRANS-SPLICING OF NEMATODE PREMESSENGER RNA, Timothy W.	
Nilsen	413
Regulation of the Phosphorelay and the Initiation of	
SPORULATION IN BACILLUS SUBTILIS, James A. Hoch	441
ADAPTIVE MUTATION: The Uses of Adversity, Patricia L. Foster	467
Genetics and Molecular Biology of Chitin Synthesis in	
FUNGI, Christine E. Bulawa	505
Antibiotics Synthesized by Posttranslational Modification, $J$ .	
Norman Hansen	535
STRUCTURE-FUNCTION AND BIOGENESIS OF THE TYPE IV PILI, Mark	
S. Strom and Stephen Lory	565
MOLECULAR BIOLOGY OF THE LYSR FAMILY OF TRANSCRIPTIONAL	
REGULATORS, Mark A. Schell	597
ENZYMES AND PROTEINS FROM ORGANISMS THAT GROW NEAR AND	()7
ABOVE 100°C, Michael W. W. Adams	627
GENETICS FOR ALL BACTERIA, B. W. Holloway	659
GENETICS OF DIFFERENTIATION IN STREPTOMYCES, Keith F. Chater	685
EVALUATING BIOREMEDIATION: Distinguishing Fact from Fiction, Michael J. R. Shannon and Ronald Unterman	715
Genetically Engineered Protection Against Viruses in	
TRANSGENIC PLANTS, John H. Fitchen and Roger N. Beachy	739
GENETIC MANIPULATION OF NEGATIVE-STRAND RNA VIRUS	
GENOMES, Adolpho García-Sastre and Peter Palese	765
THE CELLULOSOME: The Exocellular Organelle of Clostridium,	
Carlos R. Felix and Lars G. Ljungdahl	791
THE PROTEASES AND PATHOGENICITY OF PARASITIC PROTOZOA,	
James H. McKerrow, Eugene Sun, Philip J. Rosenthal, and Jacques Bouvier	821
THE STATIONARY PHASE OF THE BACTERIAL LIFE CYCLE, Roberto	021
Kolter, Deborah A. Siegele, and Antonio Tormo	855
POLYKETIDE SYNTHESIS: Prospects for Hybrid Antibiotics, Leonard	000
Katz and Stefano Donadio	875
RELEASE OF RECOMBINANT MICROORGANISMS, M. Wilson and S. E.	
Lindow	913
THE TN5 TRANSPOSON, W. S. Reznikoff	945
INDEXES	
Subject Index	965

CUMULATIVE INDEX OF CONTRIBUTING AUTHORS, VOLUMES 43-47987CUMULATIVE INDEX OF CHAPTER TITLES, VOLUMES 43-47989