

Tn5: A Molecular Window on Transposition

William S. Reznikoff,^{*1} Archana Bhasin,^{*} Douglas R. Davies,^{*†} Igor Y. Goryshin,^{*}
 Lisa A. Mahnke,^{*} Todd Naumann,^{*} Ivan Rayment,^{*†}
 Mindy Steiniger-White,^{*} and Sally S. Twining^{*2}

^{*}Department of Biochemistry, University of Wisconsin, 433 Babcock Drive, Madison, Wisconsin 53706; and

[†]Institute for Enzyme Research, University of Wisconsin, 1710 University Avenue, Madison, Wisconsin 53705

Received September 28, 1999

DNA transposition is an underlying process involved in the remodeling of genomes in all types of organisms. We analyze the multiple steps in cut-and-paste transposition using the bacterial transposon Tn5 as a model. This system is particularly illuminating because of the existence of structural, genetic, and biochemical information regarding the two participating specific macromolecules: the transposase and the 19-bp sequences that define the ends of the transposon. However, most of the insights should be of general interest because of similarities to other transposition-like systems such as HIV-1 DNA integration into the host genome. © 1999 Academic Press

DNA transposition is an important mechanism for the remodeling of genomes and for the facilitation of horizontal genetic transfer. Transposition is likely to occur in all organisms. One specific example of this phenomenon is the root cause of a modern medical tragedy; the integration of HIV-1 DNA into target cell genomes. Another important medical problem (the rise and dissemination of antibiotic resistant determinants in bacteria) is due, in large part, to transposition. However, transposition is also the causative process behind various forms of beauty such as color variegation in maize kernels. In fact it was this phenomenon that was being studied by Barbara McClintock when she discovered transposition (1). Furthermore, developmental DNA rearrangement processes such as immunoglobulin gene formation are likely to occur by similar processes (2).

¹ To whom correspondence should be addressed. Fax: (608) 262-3452. E-mail: reznikoff@biochem.wisc.edu.

² Current address: Department of Biochemistry, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226-4801.

Transposition is a complex process giving rise to DNA insertions, inversions, deletions and chromosome fusions. The DNA structure behind this process is called a transposon or a transposable element. In its simplest manifestation, a transposable element is a DNA sequence that is defined by specific short inverted DNA sequences at its ends. The transposable element encodes a protein called a transposase (or in retroviruses an integrase) that catalyzes transposition. Although other proteins are required to repair DNA joints at the end of the transposition process and may modulate its frequency, the critical macromolecules in transposition are three: the transposon DNA defined by the end sequences, the transposase and target DNA.

Our laboratory has been studying the bacterial transposon Tn5 as a molecular window for understanding transposition. The detailed examination of any particular transposon yields valuable insights into all types of transposition since the underlying chemical mechanisms are very similar and indeed the transposase (and retroviral integrase) proteins have similar catalytic core architectures and active sites (3). The choice of Tn5 has been very fortuitous; however, because: the Tn5 transposase (and a related protein, the transposase inhibitor) is soluble and amenable to crystallographic analysis (3), there exist powerful genetic tools that have been used to isolate and characterize important mutant versions of the transposase (4–7) and the transposon end sequences (8, 9), and biochemical assays are available for studying many of the steps in the transposition process (7, 10, 11, and Bhasin, Goryshin, York and Reznikoff, unpublished).

The natural version of Tn5 is a very complex structure 5.8 kbp in length encoding 7 proteins (Fig. 1a) (12, 13). We can simplify our analysis by focusing on the transposase and the end sequences. The transposase is a 476 AA polypeptide; the structure of an N-terminal truncated version of the transposase (the inhibitor) is pictured in Fig. 1b (3). The missing N-terminus con-

tains an amino acid sequence critical for end DNA sequence binding (7, 14). Key components of this structure are the catalytic core, which is similar in architecture to those of avian sarcoma virus and HIV-1 integrases and virus Mu transposase (3, 15, 16, 17), and the catalytic active site containing critical D, D and E residues. There are two different natural 19 bp end sequences that the transposase can recognize leading to transposition (Fig. 1a). These are the outside end (OE) and inside end (IE) sequences normally found as components of IS50 (a Tn5 substructure) (8, 12, 13). We discovered that the OE and IE sequences were suboptimal for transposition when we created and analyzed a hyperactive related sequence called the mosaic end (Fig. 1a) (9).

The rest of this communication will briefly review our knowledge of each of the steps in Tn5 transposition. We will enrich our presentation by relating this story to information learned about other selected transposable elements. Transposition is a very disruptive process vis-à-vis the genome; therefore, it is not surprising that it is severely down-regulated. As we shall show, the analysis of Tn5 has indicated that there are multiple systems in place to accomplish this down-regulation. Finally, we shall discuss some issues related to host functions and Tn5 transposition, and briefly mention practical laboratory tools that are being developed using the Tn5 system.

THE PROCESS

Tn5 transposition is a multi-step cut and paste process schematically presented in Fig. 2. Other transposable elements appear to proceed through seemingly quite different transposition paths [for instance replicative transposition for Mu (18) or circle formation for IS911 (19)], but these likely result from minor variations in the basic chemistry as will be described.

END SEQUENCE BINDING

Tn5 transposase end sequence binding involves the interaction of an N-terminal amino acid sequence with the 19 bp end sequence. This is likely to involve a protein monomer-DNA monomer sequence interaction (see below). Our genetic data suggests that at least a helix turn helix motif is involved in the DNA sequence recognition (9). This motif was correlated with the end sequence binding function through the isolation of mutants that enhance the binding affinity or, in one case, alters the binding preference regarding OE versus IE. But co-crystallographic studies of the Tc3 transposase DNA binding domain bound to its cognate DNA sequence (20) and recent genetic data on Tn5 transposase (Naumann and Reznikoff, unpublished) suggest that the DNA binding motif is more complex.

A critical feature of the binding reaction is that it is highly disfavored probably due to two features. First, the DNA binding domain sequence is suboptimal (7). Secondly, the C-terminus of the protein interferes with the DNA binding (removal of the C-terminal sequences or occupation of the C-terminal sequences by protein-protein interaction is required to permit DNA binding detection) (3, 10, 14, 21). Thus, an unfavorable allosteric change in transposase structure removing the C-terminal inhibition is likely to occur prior to DNA binding. It is possible that the DNA binding reaction is the critical rate limiting step in Tn5 transposition.

The transposase C-terminal interference with the transposase DNA binding activity may be related to the intriguing observation that Tn5 transposase, like some other transposases, prefers to act *in cis* (12, 13, 47, 48). This explanation would posit that transposase still in the process of being translated might be better able to bind to nearby end DNA sequences since the inhibitory C-terminus would not yet exist.

As has been found for other sequence specific DNA binding proteins, the transposase has been found to bend end sequence containing DNA (21, 22). This bend (at least 35°) occurs at the transposon-donor backbone (DBB) boundary and might explain the observation that missing bases near the end sequence-DBB boundary enhance transposase binding (23). Perhaps the DNA bending locally destabilizes the helix and thereby facilitates the subsequent cleavage reactions.

SYNAPTIC COMPLEX FORMATION

The next step is highly favored. It is synaptic complex formation resulting from the dimerization of the monomeric transposase-end sequence complexes. It is the stoichiometry of the synaptic complexes formed at low transposase concentrations (two transposase molecules with two DNA sequences [Bhasin, Goryshin, York, and Reznikoff, unpublished]) which suggests that the preceding bound complexes involve a monomer of transposase with a monomer of DNA.

Most of the interactions leading to synapsis likely involve protein-protein interactions. Although the involved dimerization domains are uncertain (3, 24), recent genetic studies (Steiniger-White and Reznikoff, unpublished) suggest that dimerization mediated by the C-terminal alpha helix (3) is critical for this process. This is surprising since this domain is not shared by the related Tn10 transposase (24).

The observations that DNA cleavage involves trans catalysis (see below) and that transposase containing mutations of lysines 330 and 333 are defective in synapsis suggest trans protein-DNA interactions are also involved (Naumann and Reznikoff, unpublished). That is, nucleotide pairs at the very end of the specific binding sequence and/or at the adjoining DBB sequence may interact with opposite subunit lysines (residues

330 and 333, see Fig. 1b) near the catalytic active site. These are the contacts that possibly were detected by retroviral integrase–DNA cross-linking studies (25a).

Synaptic complex formation is typically studied experimentally by analyzing the formation of a comparable structure, the paired end complex. The paired end complex is formed when the two end sequences are located on separate DNA fragments.

Although our model postulates that transposase–end sequence binding precedes synapsis, it is possible that a more complex mixture of these two steps occurs. For instance, unstable dimerization of transposase monomers may expose the DNA binding domains allowing end sequence binding and then this is followed by protein–protein and protein–DNA rearrangements yielding the final synaptic complexes.

DNA CLEAVAGE

The final product of transposon–DBB cleavage is a blunt ended transposon DNA. How can one active site cleave the two DNA strands of opposite polarity? The generation of this blunt end cleavage product involves three phosphoryl transfer reactions in Tn5 transposition (Fig. 3) (25b) as it does for Tn10 transposition (26). First an OH group recruited from H₂O is used as a nucleophile to break the 3' strand phosphodiester bond bordering the end sequence. Second the released 3' OH group attacks the 5' strand phosphodiester bond causing release of the DBB and formation of a DNA hairpin. The third step in DNA cleavage is presumed to use an additional water donated OH group to break the hairpin phosphodiester bond. This three step mechanism allows both strands to be cleaved with no major reorientation of the active site.

The three-step cleavage mechanism distinguishes cut and paste transposition from replicative transposition. For instance, the active site for Mu transposase and Tn3 transposase must not be able to perform the cross strand hairpin attack thus the 5' strand remains attached to the DBB. Retroviral integrase is similarly limited to cutting the 3' strand, but in this case the attached 5' strand extension is limited to a few nucleotides (27). The details of the hairpin attack step also probably distinguishes Tn5 (and Tn10) linear type transposition from IS911 circle formation transposition. For Tn5 the cross strand attack occurs *in cis* within each bound DNA in the synaptic complex (Bhasin, Goryshin and Reznikoff, unpublished) while for IS911 the attack is *in trans* (between bound DNAs) (19).

Finally, all of the phosphoryl transfer reactions involved in DNA cleavage are catalyzed *in trans* within the synaptic complex (Naumann and Reznikoff, unpublished). That is, the transposase monomer bound to one end sequence catalyzes the strand cleavages at the other end. This was determined by studying the cleavage reactions for an OE–IE defined (IS50-like) trans-

poson using a mixture of IE specific, catalytically active transposase and OE specific, catalytically inactive transposase; cleavage occurred at the OE–DBB boundary. The reverse localization of the catalytically inactive monomer directs cleavage to the IE–DBB boundary. Trans cleavage has been found to occur during Mu transposition (28–30). It ensures that cleavages at both ends only occur within the context of a synaptic complex and not by transposase bound to only one end.

TARGET CAPTURE

Target capture is not well understood. Through a sequence analysis of inserts in an extensive collection of *in vivo* and *in vitro* transposition events, we know that there are some DNA target sequence biases and there is intriguing evidence that formation of multimeric transposase filaments on target sequences can facilitate this step (31). However, the transposase sequences involved in target capture are unknown.

STRAND TRANSFER INTO TARGET

Once bound to the target, the transposon 3'OH groups attack target strands with the individual attacks being spaced 9 bp apart thus leading to an insert with 9 base gaps bounding each side. This will lead, after repair, to the signature target site 9 bp duplication.

DISINTEGRATION

The reversal of the integration reaction is disintegration. This reaction has never been detected for Tn5 although it has been used as an assay for HIV-1 integrase core activity (32). On the basis of intuition, we believe that disintegration is highly disfavored for Tn5.

REMOVING TRANSPOSASE

At the end of this process transposase is bound tightly to the transposon containing products (11). How it is removed remains a mystery. For Mu, a host chaperone-protease is involved in removal (33). For Tn5 we have suggestive evidence that a proteolytic attack following a lysine residue at position 40 within the DNA binding domain may be involved (Twining, Goryshin and Reznikoff, unpublished).

It is assumed that host functions, DNA polymerase and ligase, repair the gaps.

DOWN-REGULATING TRANSPOSITION

Tn5 transposition is a very rare process *in vivo* as might be expected for such a destructive process. We believe that there are multiple mechanisms accomplishing this. These include (i) The wild-type protein is

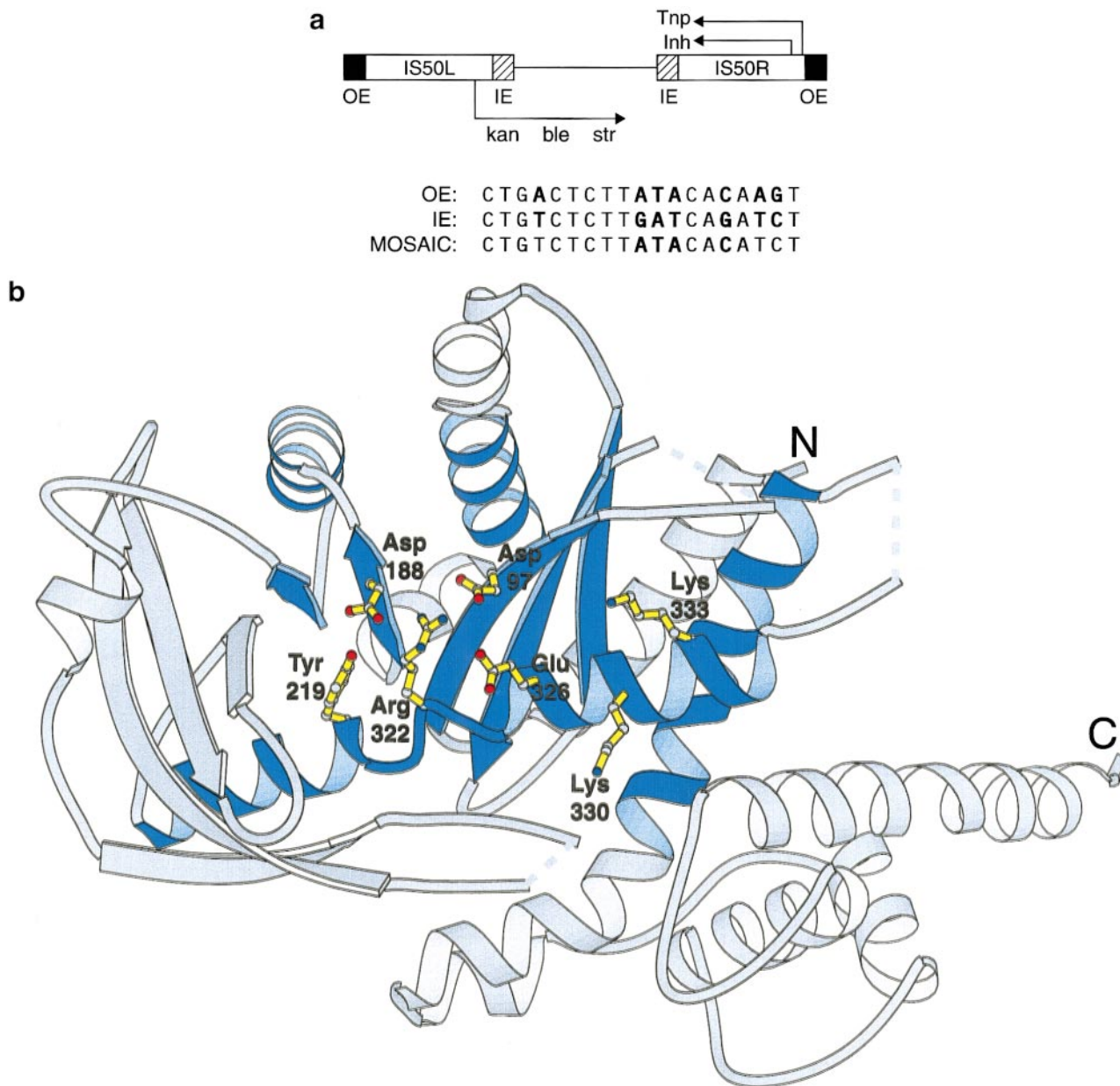


FIG. 1. (a) Tn5 structure and specific 19-bp end sequences. The structure of Tn5 has been described in detail (12, 13). Tn5 is a composite transposon in which two *IS50* elements bracket three antibiotic resistance genes. *IS50R* encodes the transposase (Tnp) and the trans-dominant transposition inhibitor (Inh). Inh lacks the N-terminal 55 amino acids of Tnp. *IS50L* encodes inactive truncated versions of Tnp and Inh (not shown). *IS50* is bordered by two 19-bp sequences (OE and IE) to which Tnp binds as the first step in transposition (see Fig. 2). The sequence of one strand of OE and IE are shown below. One strand of a 19-bp mosaic sequence found to be hyperactive for transposition (9) is also shown. (b) A ribbon diagram representation of Inh (3). The catalytic core whose architecture resembles that of HIV-1 and ASV integrases and Mu transposase is highlighted in dark blue. The DDE active site residues (Asp 97, Asp 188, and Glu 325) are shown as well as residues found in the conserved YRKK motif. Not visible are the 14 N-terminal amino acids of Inh (or the 69 N-terminal amino acids of Tnp).

almost a non-functional protein probably because of, in part, a suboptimal DNA binding domain and C-terminal interference with the DNA binding (3, 7, 10, 14, 21); (ii) The natural end DNA sequences OE and IE are suboptimal (9); (iii) Tn5 encodes an N-terminal deletion variant of the transposase called the inhibitor that acts as a trans-dominant negative regulator probably by forming inactive inhibitor-transposase complexes

(10, 34). The abundance of inhibitory peptides is likely augmented by proteolytic attack on the transposase (24); and (iv) Host functions modulate the transposition frequency. For instance dam methylase down-regulates transposase synthesis and use of the IE sequence (35). In addition, Topo-1 positively modulates transposition possibly by influencing the target capture step (36 and Yigit and Reznikoff, unpublished).

Other host functions may also modulate the frequency of Tn5 transposition.

NEW TOOLS FROM TRANSPOSITION

In vivo transposition systems are well established tools in genetics and genome research (see reference 37 for a review of microbial systems). One can easily imagine that *in vitro* transposition systems would be developed as tools as well. Several systems (those derived from Ty1(38), Tn5 (11), Tn7 (39), and Mu (40)) are currently available commercially as tools for intermolecular transposition; transferring transposon tags into target DNA molecules. One application of this technology is to use the transposon of choice as a mobile primer binding site for DNA sequencing.

Tn5 *in vitro* systems offer two additional technologies. Tn5 also transposes in an intra-molecular fashion thereby facilitating the one step formation of nested deletions and inversion/deletions of target genes (41a). Preformed Tn5 transposase-transposon complexes can be introduced into target cells by electroporation (41b) leading to transposition events. This latter technology bypasses limitations imposed by strictly *in vivo* techniques such as the need for transposase expression in the host cells.

GENERAL RELEVANCE

Since phosphoryl transfer reactions are universal in nucleic acid biochemistry, the molecular details of

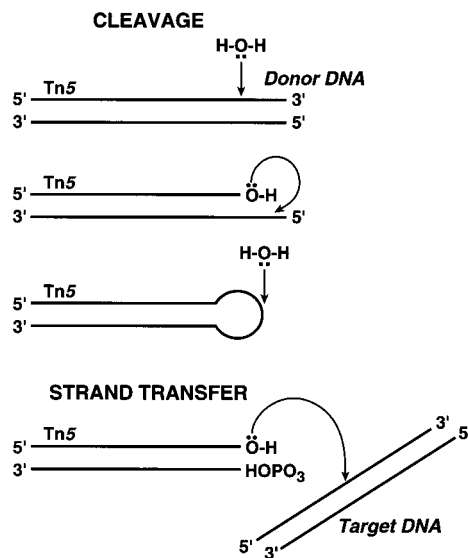


FIG. 3. Hairpin intermediate model for double strand cleavage and strand transfer during Tn5 transposition.

transposition catalysis are likely to be paradigms for understanding many important genetic reactions. In fact, the architecture of the Tn5 catalytic core not only resembles that of Mu transposase and the retroviral integrases, but also is similar to the recombination protein RuvC and RnaseH (42–46).

ACKNOWLEDGMENTS

The research described in this report was supported by National Institutes of Health Grants GM50692 (to W.S.R.), AR35186 (to I.R.), and EY12731 (to S.S.T.) and United States Department of Agriculture Hatch Grant 4184 (to W.S.R.). D.R.D. and T.N. are supported by National Institutes of Health Training Grant GM08349. L.A.M. was supported in part by National Institutes of Health Training Grant GM07215. W.S.R. is the Evelyn Mercer Professor of Biochemistry and Molecular Biology.

REFERENCES

1. McClintock, B. (1952) *Cold Spring Harbor Symp. Quant. Biol.* **16**, 13–47.
2. McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., and Oettinger, M. A. (1995) *Cell* **83**, 387–395.
3. Davies, D. R., Braam, L. M., Reznikoff, W. S., and Rayment, I. (1999) *J. Biol. Chem.* **274**, 11904–11913.
4. Krebs, M. P., and Reznikoff, W. S. (1988) *Gene* **63**, 277–285.
5. Wiegand, T. W., and Reznikoff, W. S. (1992) *J. Bacteriol.* **174**, 1229–1239.
6. Weinreich, M. D., Gasch, A., and Reznikoff, W. S. (1994) *Genes Dev.* **8**, 2363–2374.
7. Zhou, M., and Reznikoff, W. S. (1998) *J. Mol. Biol.* **271**, 362–373.
8. Makris, J. C., Nordmann, P. L., and Reznikoff, W. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2224–2228.
9. Zhou, M., Bhasin, A., and Reznikoff, W. S. (1998) *J. Mol. Biol.* **276**, 913–925.

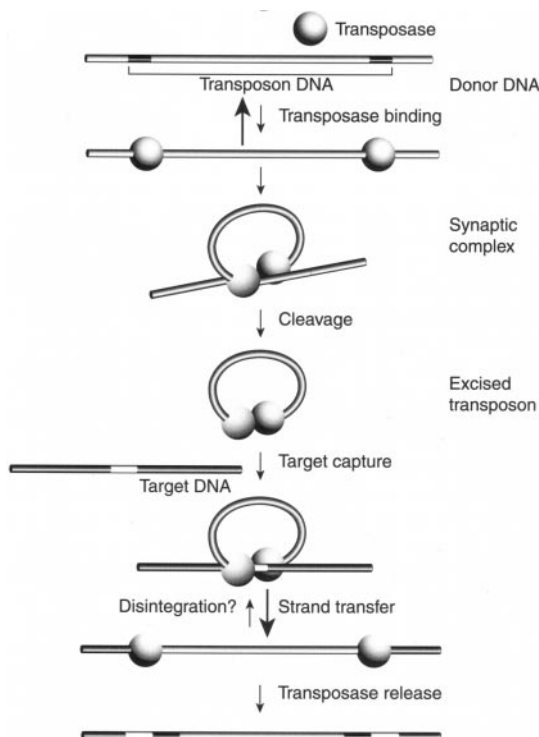


FIG. 2. Cut and paste transposition model.

10. De la Cruz, N. B., Weinreich, M. D., Wiegand, T. W., Krebs, M. P., and Reznikoff, W. S. (1993) *J. Bacteriol.* **175**, 6932–6938.
11. Goryshin, I. Y., and Reznikoff, W. S. (1998) *J. Biol. Chem.* **273**, 7367–7374.
12. Berg, D. E. (1989) in *Mobile DNA*, pp. 186–210, Am. Soc. Microbiol., Washington, DC.
13. Reznikoff, W. S. (1993) *Annu. Rev. Microbiol.* **47**, 945–963.
14. Weinreich, M. D., Mahnke-Braam, L., and Reznikoff, W. S. (1994) *J. Mol. Biol.* **241**, 166–177.
15. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science* **266**, 1981–1986.
16. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995) *J. Mol. Biol.* **253**, 333–346.
17. Rice, P., and Mizuuchi, K. (1995) *Cell* **82**, 209–220.
18. Pato, M. L. (1989) in *Mobile DNA*, pp. 24–52, Am. Soc. Microbiol., Washington, DC.
19. Polard, P., Prere, M. F., Fayet, O., and Chandler, M. (1992) *EMBO J.* **11**, 5079–5090.
20. Van Pouderooyen, G., Ketting, R. F., Perrakis, A., Plasterk, R. H. A., and Sixma, T. K. (1997) *EMBO J.* **16**, 6044–6054.
21. York, D., and Reznikoff, W. S. (1996) *Nucleic Acids Res.* **24**, 3790–3796.
22. York, D., and Reznikoff, W. S. (1997) *Nucleic Acids Res.* **25**, 2153–2160.
23. Jilk, R. A., York, D., and Reznikoff, W. S. (1996) *J. Bacteriol.* **178**, 1671–1679.
24. Braam, L. M., and Reznikoff, W. S. (1998) *J. Biol. Chem.* **273**, 10908–10913.
- 25a. Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J.* **16**, 6849–6859.
- 25b. Bhasin, A., Goryshin, I. Y., and Reznikoff, W. S. (1999) *J. Biol. Chem.*, in press.
26. Kennedy, A. K., Guhathakurta, A., Kleckner, N., and Haniford, D. (1998) *Cell* **95**, 125–134.
27. Mizuuchi, K. (1992) *J. Biol. Chem.* **267**, 21273–21276.
28. Aldaz, H., Schuster, E., and Baker, T. A. (1996) *Cell* **85**, 257–269.
29. Savilahti, H., and Mizuuchi, K. (1996) *Cell* **85**, 271–280.
30. Namgoong, S.-Y., and Harshey, R. M. (1998) *EMBO J.* **17**, 3775–3785.
31. Goryshin, I. Y., Miller, J. A., Kil, Y. V., Lanzov, V. A., and Reznikoff, W. S. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 10716–10721.
32. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3428–3432.
33. Levchenko, I., Luo, L., and Baker, T. (1995) *Genes Dev.* **9**, 2399–2408.
34. Yin, J. C.-P., and Reznikoff, W. S. (1988) *J. Bacteriol.* **170**, 3008–3015.
35. Yin, J. C.-P., Krebs, M. P., and Reznikoff, W. S. (1988) *J. Mol. Biol.* **199**, 35–46.
36. Yigit, H., and Reznikoff, W. S. (1998) *J. Bacteriol.* **180**, 5866–5874.
37. Berg, C. M., and Berg, D. E. (1996) in *Escherichia coli* and *Salmonella: Cellular and Molecular Biology*, pp. 2588–2612, Am. Soc. Microbiol., Washington, DC.
38. Devine, S. E., and Boeke, J. D. (1994) *Nucleic Acids Res.* **22**, 3765–3772.
39. Gwinn, M. L., Stellwagen, A. E., Craig, N. L., Tomb, J. F., and Smith, H. O. (1997) *J. Bacteriol.* **179**, 7315–7320.
40. Haapa, S., Taira, S., Heikkinen, E., and Savilahti, H. (1999) *Nucleic Acids Res.* **27**, 2777–2784.
- 41a. York, D., Welch, K., Goryshin, I. Y., and Reznikoff, W. S. (1998) *Nucleic Acids Res.* **26**, 1927–1933.
- 41b. Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R., and Reznikoff, W. S. (2000) *Nature Biotechnol.*, in press.
42. Arioshi, M., Vassilyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994) *Cell* **78**, 1063–1072.
43. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Marsuzaki, T., and Morikawa, K. (1990) *Nature* **347**, 306–309.
44. Yang, W., Hendrickson, W. A., Crouch, R. J., and Satow, Y. (1990) *Science* **249**, 1398–1405.
45. Yang, W., and Steitz, T. A. (1995) *Structure* **15**, 131–134.
46. Rice, P., Craigie, R., and Davies, D. R. (1996) *Curr. Opin. Struct. Biol.* **6**, 76–83.
47. Isberg, R. R., Lazaar, A. L., and Syvanen, M. (1982) *Cell* **30**, 883–892.
48. Johnson, R. C., Yin, J. C.-P., and Reznikoff, W. S. (1982) *Cell* **30**, 873–882.