JOURNAL OF BACTERIOLOGY, Sept. 1996, p. 5533–5536 Vol. 178, No. 18
0021-9193/96/$04.00+0

Positions of Strand Exchange in Mycobacteriophage L5 Integration and Characterization of the *attB* Site

CAROL E. A. PEÑA, JUDY E. STONER, AND GRAHAM F. HATFULL*

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received 9 May 1996/Accepted 11 July 1996

Mycobacteriophage L5 integrates into the genome of *Mycobacterium smegmatis* via site-specific recombination between the phage *attP* site and the bacterial *attB* site. These two sites have a 43-bp common core sequence within which strand exchange occurs and which overlaps a $tRNA^{Gly}$ gene at *attB*. We show here that a 29-bp segment of DNA is necessary and sufficient for *attB* function and identify the positions of strand exchange.

Mycobacteriophage L5 is a temperate phage of the mycobacteria and forms stable lysogens in *Mycobacterium smegmatis*; these lysogens contain a single copy of the L5 genome integrated into the *M. smegmatis* chromosome (4, 22). Comparison of the phage attachment site (*attP*), the site of insertion in the bacterial genome (*attB*), and the novel junctions formed by integration (*attL* and *attR*) reveals a common core sequence of 43 bp that is identical in all four DNAs (14). A similar sequence is also present in *Mycobacterium bovis* bacille Calmette-Guérin (BCG), although there is a single base difference between the BCG *attB* and L5 *attP* core sequences (14). This difference does not appear to prevent the use of this region as an attachment site for phage L5 DNA in BCG (14, 23). The 43-bp common core at the *M. smegmatis attB* site overlaps the 3′ end of a $tRNA^{Gly}$ gene such that the integrity of the tRNA gene is maintained following integration. Analysis of deletion derivatives of the L5 *attP* site suggests that *M. smegmatis* is not viable if this tRNA gene is interrupted (18).

We have previously described an in vitro reaction for L5-mediated integrative recombination (13). This reaction requires a supercoiled DNA substrate containing the *attP* site, a linear segment of DNA containing the *attB* site, the L5 integrase protein, and a protein (mycobacterial integration host factor [mIHF]) present in crude extracts of *M. smegmatis* (13, 17a). In this study, we determined the minimal segment of DNA required for *attB* function and identified the positions of strand exchange for integrase-mediated recombination.

The entire 43-bp common core is not required for *attB* function. Previous experiments showed that a DNA segment of approximately 600 bp containing the *attB* 43-bp common core is a good substrate for recombination in vitro (13). However, it is unlikely that this entire segment is needed for integration, and we note that only 20 to 25 bp of *attB* DNA is needed for lambda integrase-mediated recombination (16). We have recently observed that L5 integrase binds to the left side of the

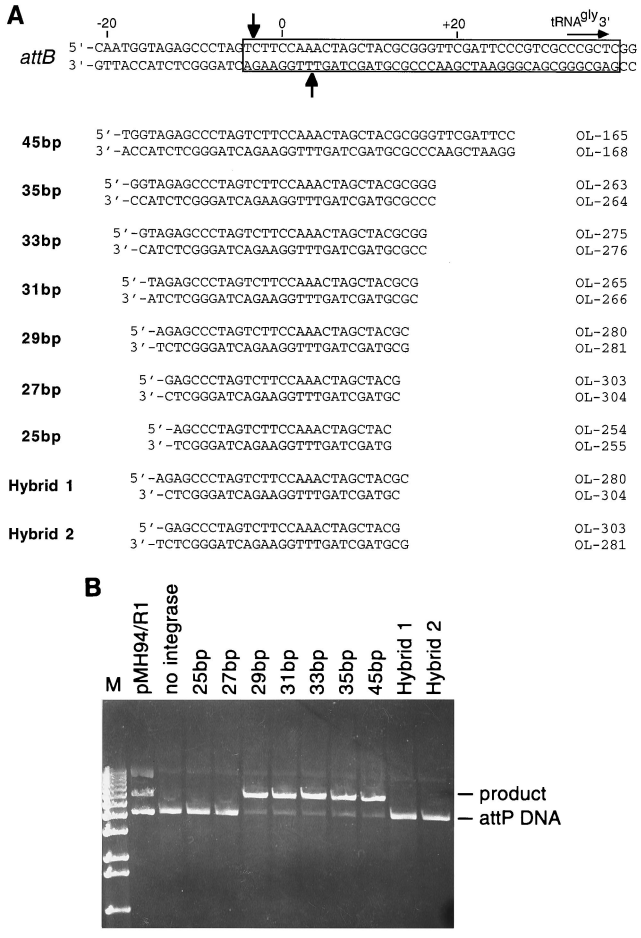

FIG. 1. Characterization of the *attB* site for L5 integration. (A) Oligonucleotide substrates used for in vitro recombination. The top part of the figure shows the sequence of *attB* DNA, including the 43-bp common core (boxed) and the 3′ end of the tRNA gene. Shown below this sequence are a series of DNA duplexes of different lengths derived by annealing synthetic oligonucleotides. The length of each duplex and the identity of each oligonucleotide are indicated. Hybrid duplexes 1 and 2 were derived by annealing oligonucleotides of different lengths. The positions of strand exchange as determined by the scheme illustrated in Fig. 2 are shown by vertical arrows. (B) In vitro recombination of *attB* oligonucleotide substrates. Each of the duplexed DNAs shown in panel A were tested as substrates for L5-mediated integration in vitro. Recombination reactions were performed similarly to those described previously (13) and contained 0.025 pmol of supercoiled pMH94 DNA (which contains *attP*), approximately 6 pmol of duplex *attB* DNA, 1.5 pmol of L5 integrase, and 0.2 μl of purified mycobacterial integration host factor (17a) and were incubated for 2 h at 37°C prior to electrophoresis through a 0.7% agarose gel. Recombination between the supercoiled *attP* DNA and *attB* DNA produces a linear product of 6.2 kb as indicated. Each lane contains recombination reactions performed with the oligonucleotide substrates indicated, with the exception of the DNA size markers (M), pMH94 DNA partially digested with *Eco*RI (pMH94/R1), and a negative control in which the 45-bp duplex was included but no integrase was added (no integrase). The 29- and 31-bp DNA fragments support similar levels of recombination as the 45-bp substrate when only 0.06 pmol of *attB* DNA is used (data not shown).

* Corresponding author. Phone: (412) 624-6975. Fax: (412) 624-4870. Electronic mail address: GFH@vms.cis.pitt.edu.

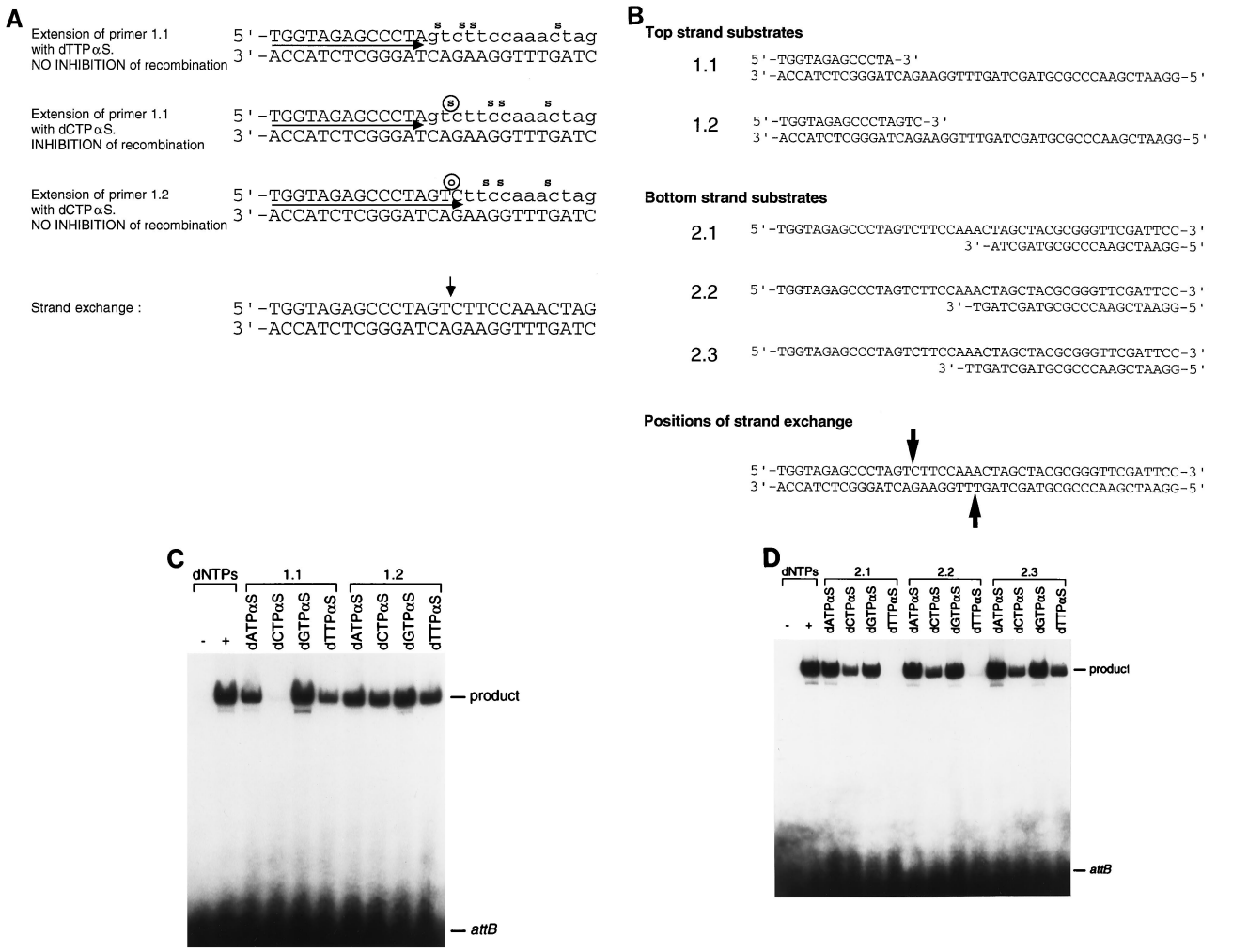

FIG. 2. Determination of positions of strand exchange. (A) Experimental scheme for determining positions of strand exchange. Recombination substrates are generated by extension of primer-template pairs (shown in uppercase letters) with the Klenow fragment of DNA polymerase I. The positions of the primers are indicated by horizontal arrows, with the arrowheads at the 3′ ends of the primers; nucleotides introduced by extension are shown in lowercase letters. Separate extension reactions are performed with each primer-template pair substituting one deoxynucleotide triphosphate (dNTP) with an α-thiosubstituted deoxynucleotide triphosphate (dNTPαS). If the position of strand exchange lies within the extended region (for example, by extension of primer 1.1), it is expected that incorporation of one dNTPαS will inhibit recombination (dCTPαS in the example shown) whereas incorporation of the other three dNTPαSs will not (for example, dTTPαS). However, if the position of strand exchange is within the primer (as in primer 1.2), then no inhibition will be observed when any of the dNTPαSs are incorporated. Thus, by careful choice of primers, the precise position of strand exchange can be deduced. It is important to note that the two recombination substrates generated by inclusion of dCTPαS in extension of either primer 1.1 or 1.2 differ only in the nature of one phosphodiester bond, which is an oxophosphate (o) when primer 1.2 is used and a thiophosphate (s) when primer 1.1 is used. The deduced position of strand exchange is indicated by the vertical arrow. (B) Primer-template oligonucleotide substrates used for mapping positions of strand exchange in L5. Two primer-template pairs (with primers 1.1 and 1.2) are shown for mapping the top strand exchange, and three (with primers 2.1, 2.2, and 2.3) are shown for the bottom strand exchange. The top and bottom template strands were provided by oligonucleotides OL-165 and OL-168, respectively, as shown in Fig. 1A. Substrates for recombination were generated by primer extension with the scheme illustrated in panel A. (C) Determination of position of top strand exchange. In vitro recombination reactions were performed with *attB* DNA substrates generated by extension of the primer-template pairs shown in panel B with thiosubstituted nucleotides as indicated. Reaction conditions were similar to those described in the legend to Fig. 1 except that 5′-radiolabeled *attB* substrates were used and crude extracts of *M. smegmatis* were used in place of purified mycobacterial integration host factor (see reference 13). Products were analyzed by electrophoresis through 0.8% agarose. Control reactions are shown in which the substrate was generated by extension of primer 1.1 with four dNTPs and incubated with all reaction components (+) or without the addition of integrase or *M. smegmatis* extract (−). Other lanes show recombination reactions performed with substrates generated with either dATPαS, dCTPαS, dGTPαS, or dTTPαS substituting for the cognate oxophosphate dNTP in extension of either primer 1.1 or primer 1.2. (D) Determination of position of bottom strand exchange. Reactions were performed as described for panel C but with substrates generated by extension of primers 2.1, 2.2, or 2.3 as indicated. Control reactions are shown in which the substrate was generated by extension of primer 2.2 with four dNTPs and incubated with all reaction components (+) or without the addition of integrase or *M. smegmatis* extract (−).

core of the L5 *attP* site (i.e., closest to the P1, P2, and P3 arm-type sites) and that the right part of the 43-bp core is not protected (18). Thus, while the extreme right part of the *attP* core may be required to restore the integrity of the chromosomal tRNA gene, it may not be needed for recombination (18). Likewise, not all of the 43-bp common core may be needed for *attB* function. We tested this by designing a 45-bp synthetic *attB* substrate containing the left side of the core and found it to be a good substrate for recombination in vitro (Fig. 1). Thus, the entire 43-bp common core at *attB* is not required for recombination.

A 29-bp segment of DNA is necessary and sufficient for *attB* function. To determine the minimum length of DNA necessary for *attB* function, we designed additional oligonucleotides of

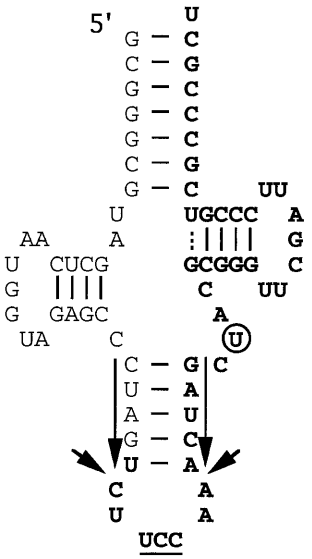

FIG. 3. Positions of strand exchange relative to tRNA structure. The predicted secondary structure of $tRNA^{Gly}$ of *M. smegmatis* is shown, with bases corresponding to the 43-bp *attB* core in bold type. The diagonal arrows indicate the corresponding positions at which strand exchange occurs at *attB*, and the region between these arrows also defines the anticodon loop. Vertical arrows indicate the sequences constituting the anticodon stem, and the anticodon 5′-UCC is underlined. The circled base in the variable loop is a G nucleotide in the corresponding $tRNA^{Gly}$ of BCG (14).

different lengths (Fig. 1A) and tested them as recombination substrates (Fig. 1B). These were designed such that the positions of strand exchange (described below) are at the center of each substrate. These experiments show that all of the substrates equal to or greater than 29 bp in length support efficient recombination (Fig. 1B). However, none of the substrates smaller than 29 bp recombined, including the 27-bp DNA segment (Fig. 1B). Thus, this 29-bp segment of *attB* DNA is both necessary and sufficient for recombination.

To test whether fully paired bases are required at the ends of the 29-bp substrate, we made two hybrid DNAs, each formed by annealing one of the 27-base oligonucleotides with one of the 29-bp oligonucleotides (hybrids 1 and 2) (Fig. 1A). No recombination was observed with either of these hybrid substrates, indicating a requirement for a fully duplexed 29-bp DNA segment for *attB* function (Fig. 1B).

Determination of the positions for strand exchange. The positions of strand exchange in the L5 *attB* site were determined with an experimental strategy described previously for phage HK022 integration (12). This approach utilizes the observation from the lambda integration system that a phosphothioate bond at the position of strand exchange strongly inhibits its recombination, whereas phosphothioates at other positions do not (9–11). A detailed description of the experimental scheme is illustrated in Fig. 2A.

To map the positions of both top and bottom strand exchanges, thiosubstituted *attB* substrates were generated by extension of primed oligonucleotide templates (Fig. 2B). As shown in Fig. 2C, inclusion of α-thiosubstituted dCTP (dCTPαS; but not dTTPαS, dGTPαS, or dATPαS) in the extension of primer 1.1 resulted in strong inhibition of recombination; however, no inhibition was observed when dCTPαS was used in extension of primer 1.2. Thus, the position of strand exchange must be the phosphodiester linkage between bases −3 and −4 on the top strand, the only position at which there is a chemical difference between these extension products (an oxophosphate bond in the primer 1.2 extension product and a thiophosphate bond in the primer 1.1 extension product). A similar analysis shows that the bottom strand exchange occurs between positions +3 and +4 (Fig. 2D). We

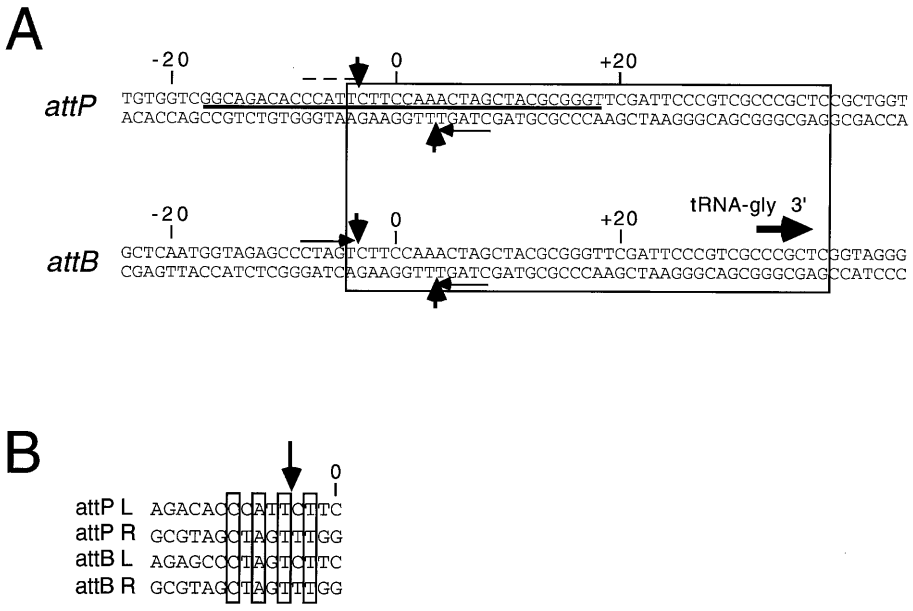

FIG. 4. Comparison of the core sequences of *attP* and *attB*. (A) Sequences of the L5 *attP* and *M. smegmatis attB* sites are aligned by the points of strand exchange. The horizontal arrows above and below the *attB* sequence indicate the sequences corresponding to the anticodon stem in the $tRNA^{Gly}$; similar sequences are indicated by the arrow and lines above and below the *attP* sequence. The 3′ end of the $tRNA^{Gly}$ gene is shown by the thick horizontal arrow above the *attB* sequence. The horizontal line between the top and bottom strands of *attP* represents the approximate region protected by L5 integrase from DNase I digestion (18). Sequences corresponding to the 43-bp common core are shown within the box. (B) Alignment of the half-site core sequences at *attB* and *attP*. The sequences from the left and right halves of *attP* and *attB* are aligned by the positions of strand exchange (shown by the arrow), and the bases in common for all half sites are shown in boxes.

deduce that L5 integrase cleaves DNA to produce 7-base 5′ extensions similar to those produced by the integrases of lambda (15), HK022 (12), HP1 (6), and P22 (21). This is, therefore, a highly conserved aspect of bacteriophage integration. The numbering system for the L5 attachment sites is based upon these observations, with the 0 position corresponding to the central base of the overlap region, and is similar to that used for phage HK022 (12).

Relationship between the tRNA anticodon loop and the 7-base overlap region. Several temperate bacteriophages use tRNA genes as attachment sites (1–3, 5–8, 12, 17, 19–21). Of these, the only two phages for which the points of strand exchange have been determined are HP1 (6) and P22 (21). In both HP1 and P22, the points of strand exchange lie 7 bp apart (as in phages lambda and HK022 [12, 15]), and these 7-base overlap regions correspond to the anticodon loops of the $tRNA^{Leu}$ and $tRNA^{Thr}$ genes, respectively (6, 21). Hauser and Scocca (6) proposed that one explanation for the use of tRNA genes as integration sites is that the dyad symmetry elements present within tRNA sequences provide appropriately spaced sites for integrase binding and cleavage. Similarly, the 7-base overlap region for L5 integration corresponds to the anticodon loop of the $tRNA^{Gly}$ gene at *attB* (Fig. 3). Although the reason for this association is not clear, it is possible that phages have acquired the use of tRNA genes as integration sites because of the shared requirements of both integrase and tRNA functions for two short inverted sequences separated by 7 bp. An alternative, more speculative explanation is that phage integrases are direct descendants of nucleases whose primary function was in tRNA processing.

The sequences of dyad symmetry that contribute to the tRNA anticodon stem may represent possible binding sites for integrase recognition, and related sequences are present in the left part of the L5 *attP* site (Fig. 4A). Alignment of these core-type sequences from the left and right parts of the *attP* and *attB* crossover sites suggests that the sequence 5′-CNA NTYT may be important for recognition by the core-type binding domain of L5 integrase (Fig. 4B).

We especially thank Ezra Yagil and Bob Weisberg, who discussed with us the experimental design for identifying the points of strand exchange prior to publication. We also thank Marisa Pedulla and Gary Sarkis for comments on the manuscript and Dawn Lever for excellent technical assistance.

This work was supported by grant GM49647 from the National Institutes of Health.

REFERENCES

1. **Campbell, A.** 1992. Chromosomal insertion sites for phages and plasmids. J. Bacteriol. **174:**7495–7499.
2. **Dupont, L., B. Boizet-Bonhoure, M. Coddeville, F. Auvray, and P. Ritzenthaler.** 1995. Characterization of genetic elements required for site-specific integration of *Lactobacillus delbrueckii* subsp. *bulgaricus* bacteriophage mv4 and construction of an integration-proficient vector for *Lactobacillus plantarum*. J. Bacteriol. **177:**586–595.
3. **Gabriel, K., H. Schmid, U. Schmidt, and H. Rausch.** 1995. The actinophage RP3 DNA integrates site-specifically into the putative tRNA(Arg)(AGG) gene of *Streptomyces rimosus*. Nucleic Acids Res. **23:**58–63.
4. **Hatfull, G. F.** 1994. Mycobacteriophage L5: a toolbox for tuberculosis. ASM News **60:**255–260.
5. **Hauser, M. A., and J. J. Scocca.** 1990. Location of the host attachment site for phage HP1 within a cluster of *Haemophilus influenzae* tRNA genes. Nucleic Acids Res. **18:**5305.
6. **Hauser, M. A., and J. J. Scocca.** 1992. Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1: location of the boundaries of the phage attachment site. J. Bacteriol. **174:**6674–6677.
7. **Hayashi, T., H. Matsumoto, M. Ohnishi, and Y. Terawaki.** 1993. Molecular analysis of a cytotoxin-converting phage, phi CTX, of *Pseudomonas aeruginosa*: structure of the attP-cos-ctx region and integration into the serine tRNA gene. Mol. Microbiol. **7:**657–667.
8. **Inouye, S., M. G. Sunshine, E. W. Six, and M. Inouye.** 1991. Retronphage phi R73: an *E. coli* phage that contains a retroelement and integrates into a tRNA gene. Science **252:**969–971.
9. **Kitts, P., E. Richet, and H. A. Nash.** 1984. λ integrative recombination: supercoiling, synapsis, and strand exchange. Cold Spring Harbor Symp. Quant. Biol. **45:**735–744.
10. **Kitts, P. A., and H. A. Nash.** 1988. An intermediate in the phage λ site-specific recombination reaction is revealed by phosphorothioate substitution in DNA. Nucleic Acids Res. **16:**6839–6856.
11. **Kitts, P. A., and H. A. Nash.** 1988. Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchange. J. Mol. Biol. **204:**95–107.
12. **Kolot, M., and E. Yagil.** 1994. Position and direction of strand exchange in bacteriophage HK022 integration. Mol. Gen. Genet. **245:**623–627.
13. **Lee, M. H., and G. F. Hatfull.** 1993. Mycobacteriophage L5 integrase-mediated site-specific recombination in vitro. J. Bacteriol. **175:**6838–6841.
14. **Lee, M. H., L. Pascopella, J. R. Jacobs, Jr., and G. F. Hatfull.** 1991. Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis* and bacille Calmette-Guérin. Proc. Natl. Acad. Sci. USA **88:**3111–3115.
15. **Mizuuchi, K., R. Weisberg, L. Enquist, M. Mizuuchi, M. Buraczynska, C. Foeller, P. L. Hsu, W. Ross, and A. Landy.** 1981. Structure and function of the phage lambda *att* site: size, *int*-binding sites, and location of the crossover point. Cold Spring Harbor Symp. Quant. Biol. **45:**429–437.
16. **Mizuuchi, M., and K. Mizuuchi.** 1985. The extent of DNA sequence required for a functional bacterial attachment site of phage lambda. Nucleic Acids Res. **13:**1193–1208.
17. **Papp, I., L. Dorgai, P. Papp, E. Jonas, F. Olasz, and L. Orosz.** 1993. The bacterial attachment site of the temperate rhizobium phage 16-3 overlaps the 3′ end of a putative proline tRNA gene. Mol. Gen. Genet. **240:**258–264.

17a. **Pedulla, M., M. H. Lee, D. Lever, and G. F. Hatfull.** A novel non-specific DNA-binding protein of mycobacteria required for phage L5 regulation. Submitted for publication.

18. **Peña, C., M. H. Lee, M. Pedulla, and G. F. Hatfull.** Characterization of the mycobacteriophage L5 attachment site, *attP*. Submitted for publication.
19. **Pierson, L. S., and M. L. Kahn.** 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in site-specific recombination. J. Mol. Biol. **196:**487–496.
20. **Reiter, W. D., P. Palm, and S. Yeats.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17:**1907–1914.
21. **Smith-Mungo, C., I. T. Chan, and A. Landy.** 1994. Structure of the P22 *att* site. J. Biol. Chem. **269:**20798–20805.
22. **Snapper, S. B., L. Lugosi, A. Jekkel, R. E. Melton, T. Kieser, B. R. Bloom, and W. R. Jacobs, Jr.** 1988. Lysogeny and transformation in mycobacteria: stable expression of foreign genes. Proc. Natl. Acad. Sci. USA **85:**6987–6991.
23. **Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, S. Snapper, R. G. Barletta, W. R. Jacobs, Jr., and B. R. Bloom.** 1991. New use of BCG for recombinant vaccines. Nature (London) **351:**456–460.