JOURNAL OF BACTERIOLOGY, Nov. 1993, p. 6836–6841
0021-9193/93/216836-06$02.00/0

Vol. 175, No. 21

Mycobacteriophage L5 Integrase-Mediated Site-Specific Integration In Vitro

MONG HONG LEE AND GRAHAM F. HATFULL*

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

Received 11 June 1993/Accepted 14 August 1993

Mycobacteriophage L5, a temperate phage of the mycobacteria, forms stable lysogens in *Mycobacterium smegmatis* via site-specific integration of the phage genome. Recombination occurs within specific phage and bacterial attachment sites and is catalyzed by the phage-encoded integrase protein in vivo. We describe here the overexpression and purification of L5 integrase and its ability to mediate integrative recombination in vitro. We find that L5 integrase-mediated recombination is greatly stimulated by extracts of *M. smegmatis* but not by *Escherichia coli* extracts, purified *E. coli* integration host factor, or purified HU, indicating the presence of a novel mycobacterial integration host factor.

The mycobacteria are an important group of organisms that includes *Mycobacterium tuberculosis* and *Mycobacterium leprae*, the causative agents of tuberculosis and leprosy, respectively. Mycobacteriophage L5 is a temperate phage that infects both fast- and slow-growing species of mycobacteria, including *Mycobacterium smegmatis* and *M. tuberculosis* (8). L5 forms stable lysogens in *M. smegmatis* that contain a single copy of an L5 prophage inserted into a chromosomal attachment site (*attB*) as a consequence of a conservative site-specific recombination event (16, 26). The phage genome contains an attachment site (*attP*) that is located close to its center and adjacent to the gene encoding the integrase protein (also called gpInt) (16). The *attB* and *attP* sites share a 43-bp core region of identical sequence that overlaps the 3′ end of a $tRNA_{Gly}$ gene of *M. smegmatis* (16) such that the integrity of the tRNA gene is maintained following prophage formation (16). The tRNA gene and *attB* site are also present in the avirulent antituberculosis vaccine strain bacille Calmette Guérin (BCG) (16). Plasmids containing the L5 *attP* site and integrase gene efficiently transform *M. smegmatis* and *M. tuberculosis* via site-specific recombination, indicating that no other phage genes are required for efficient integration (16). The integrated sequences are very stable, making these integration-proficient vectors important tools for the construction of recombinant BCG vaccines with novel protective repertoires (29). Inclusion of phage DNA to the right of the *attP* site confers a degree of instability to the integrated plasmids, suggesting the presence of a closely linked excisionase gene (16).

Mycobacteriophage L5 is one of only a few temperate phages for which the complete DNA sequence has been determined (8). The genomic organization has features that are reminiscent of phage lambda, although there is no obvious similarity at the sequence level (8). This is perhaps to be expected since the *Escherichia coli* and mycobacterial hosts are widely separated in evolutionary time. L5 also has some notable differences from phage lambda in that it encodes a DNA polymerase and three tRNA genes and appears to switch off host gene expression during lytic growth (8); these features may represent adaptations for infection of mycobacterial hosts, particularly those with very slow growth rates.

The mechanism and regulation of lambda integrase-mediated site-specific recombination have been extensively studied (15). The integration reaction requires the phage-encoded integrase (the recombinase) and the host-encoded integration host factor (IHF); prophage excision requires a second phage-encoded protein, excisionase, in addition to integrase and IHF (15). In both reactions, integrase is the recombinase that catalyzes strand cleavage and exchange and is related to a large class of related recombination proteins (1). The lambda integrase protein is composed of two domains, one of which (the C terminal) binds to the core region of *attP* and contains the catalytic residues, while the other (the N terminal) binds to arm-type sites located on either side of the core within *attP* (19). It has recently been demonstrated that recombination involves the formation of protein bridges in which one monomer of integrase can simultaneously occupy both a core-type and an arm-type site (12, 13, 18). The formation of these bridges is facilitated by IHF, which binds to specific sequences within *attP* and imparts a substantial DNA bend (24). The requirement for IHF in lambda integration can be attributed directly to its ability to bend DNA (5); moreover, these bends must be appropriately phased to facilitate formation of recombinagenic protein-DNA complexes (28). The nonspecific DNA-binding protein, HU, can at least partially substitute for IHF by binding and bending nonspecifically to *attP* DNA (7). *E. coli* IHF is also competent to stimulate recombination by the *Haemophilus influenzae* phage HP1 integrase (11).

Whereas IHF plays an architectural role, lambda integrase is the catalytic protein in recombination. Both proteins participate in formation of a higher-order structure at *attP* (the intasome) which then interacts with naked *attB* DNA to form a synaptic complex (23). Strand transfer is mediated through an ordered set of single-strand exchanges via a Holliday structure, each one accompanied by a covalent linkage between *attP* DNA and integrase (10, 14, 20, 22). This linkage involves a tyrosine residue at position 342, one of three absolutely conserved amino acid residues that must contribute to the catalytic site (22). Strand cleavage occurs seven base pairs apart within the core region of identical sequence shared by *attP* and *attB* (15).

While L5 integrase is clearly a member of the integrase family of recombinases and contains the absolutely conserved residues, it has no extensive amino acid similarity to lambda integrase (16); its closest relative appears to be the less well characterized integrase protein of the *Staphylococcus* phage L54a (unpublished observations). While the lambda and L5 integrases may use fundamentally similar mechanisms for

* Corresponding author.

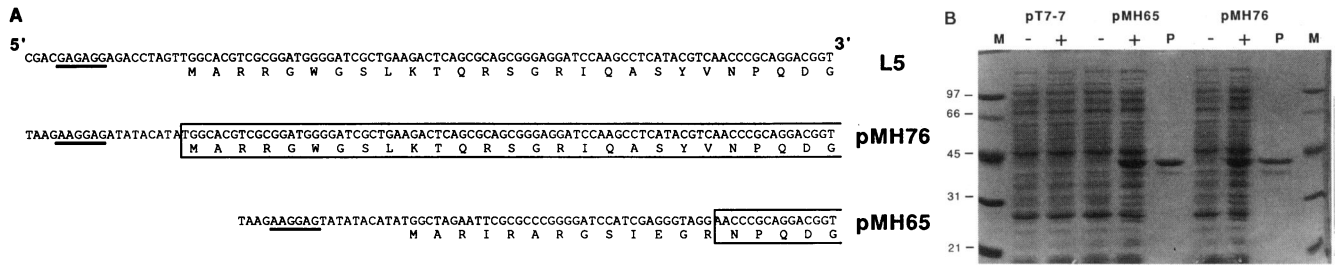

FIG. 1. Overexpression and purification of L5 integrase. (A) L5 integrase overexpression plasmids. The top part shows the DNA sequence of the 5′ end of the phage L5 integrase gene corresponding to coordinates 25387 (5′) to 25289 (3′) in the L5 genome (8); the deduced N-terminal amino acid sequence of gpInt is also shown. The L5 gpInt expression plasmid pMH76 contains the coding sequences of the 5′ end of the *int* gene fused to a vector ribosome-binding site and ATG translation initiation codon. Plasmid pMH65 does not contain *int* sequences corresponding to the first 22 amino acids of gpInt and encodes a fusion protein that includes 13 vector-derived amino acids. Both vectors contain a T7 promoter for expression of integrase. The DNA sequences in plasmids pMH76 and pMH65 originating from L5 are shown in boxes. Putative ribosome-binding sequences are underlined. (B) SDS-PAGE of L5 gpInt overexpression. *E. coli* cells containing either vector without L5 *int* (pT7-7), the truncated gene (pMH65), or the full-length gene (pMH76) were incubated either with (+) or without (−) the inducer IPTG, and total cell extracts were analyzed by SDS-PAGE. Partially purified forms of the overexpressed proteins were also analyzed (lanes P). Lanes M, molecular weight markers (sizes are indicated in kilodaltons). L5 gpInt has a predicted molecular size of 41.8 kDa.

recombination, the assembly of recombinagenic complexes, regulation of the integration and excision reactions, and the role of host factors may differ. We have sought to investigate these differences by establishing an in vitro reaction for L5 integrase-mediated recombination. We describe here the overexpression and purification of L5 integrase and its ability to mediate intermolecular integrative recombination. Efficient recombination requires addition of a heat-stable component of *M. smegmatis* extracts but is not assisted significantly by *E. coli* IHF, *E. coli* HU, or crude *E. coli* extracts.

MATERIALS AND METHODS

Construction of integrase expression plasmids. Plasmid pMH76 contains the entire L5 integrase gene linked to a vector translation initiation signal and the phage T7 promoter. pMH76 was constructed as follows. Two primers (5′-GGCG GCATATGGCACGTCGCGGATG and 5′-TCACTCTCGA ACGCGGGG) were used to amplify the integrase gene by the polymerase chain reaction, using plasmid pMH5 (16) as a template. The primers were designed such that the *int* gene would be flanked by *Nde*I (5′) and *Sal*I (3′) sites. Thirty cycles of amplification with *Taq* polymerase Stoffel fragment (Perkin Elmer) produced a major product of 1.5 kb which was purified by agarose gel electrophoresis and cloned into plasmid pT7-7 (a kind gift from Jude Samulski) that had been cleaved with *Nde*I and *Sal*I; the resulting plasmid was pMH74. As a consequence of the insertion, the presumed TTG translation initiation of L5 *int* was replaced by the vector ATG start codon (see Fig. 1 and Results). Since we were concerned about the introduction of mutations in the amplification step, we replaced the *Bam*HI-*Sal*I fragment of pMH74 with a *Bam*HI-*Sal*I fragment derived from plasmid pMH5 (16) to produce plasmid pMH76. In so doing, all but the first 15 codons of the *int* gene were replaced by a wild-type copy. The 5′ end of the *int* gene and the cloning junction were confirmed by DNA sequencing and shown to be as expected.

Plasmid pMH65 was constructed in the following way. A 1.4-kb *Hin*cII fragment derived from plasmid pMH5 (16) which contains the *int* gene was inserted into the *Stu*I site of plasmid pMal-c (New England Biolabs) such that *int* was in the same translational frame as the MalE moiety of the vector. The resulting plasmid was pMH37. Note that the *Hin*cII site at the 5′ end of the *int* gene is located at codons 22 and 23 (Fig. 1). pMH37 was then cleaved with *Bam*HI, and a 1.4-kb fragment containing the *int* gene and a small segment of the pMal-c vector (32 bp) was inserted into the *Bam*HI site of plasmid pT7-7 such that the *int* gene was in the same translational frame as the vector translation initiation signals (Fig. 1). The resulting plasmid was pMH65.

Plasmid pMH39 was constructed by inserting the 613-bp *Bam*HI-*Sal*I fragment from pMH5 into the *Bam*HI-*Sal*I sites of pUC119 (25). pMH57 is a pUC119 derivative in which a fragment of approximately 600 bp containing the *M. smegmatis attB* derived by sonication was inserted into the *Sma*I site of pUC119 (25). The *attB* site is located approximately 200 bp from one end of the inserted fragment.

Purification of L5 integrase. Plasmids pT7-7, pMH65, and pMH76 were introduced into *E. coli* BL21(DE3)pLysS (kindly provided by Jude Samulski) by transformation. Cultures were grown in L broth, and protein expression was induced by the addition of isopropyl-1-thio-β-D-galactoside (IPTG). For gel analysis, cells were collected by centrifugation, resuspended in sodium dodecyl sulfate (SDS) sample buffer (25), boiled for 3 min, and loaded onto an SDS–12% polyacrylamide gel (25).

L5 integrase was partially purified as follows. Two liters of cells containing the appropriate overexpression plasmid (either pMH65 or pMH76) was induced for 1 to 2 h by the addition of dithiothreitol, and cells were harvested by centrifugation at 14,000 rpm for 10 min at 4°C. The cell pellet was then resuspended in lysis buffer (10 mM phosphate [pH 7.0], 10 mM EDTA, 1 mM dithiothreitol, 200 mM NaCl), sonicated, and clarified by centrifugation at 14,000 rpm for 15 min at 4°C. L5 integrase was fractionated from other cellular proteins by differential ammonium sulfate precipitation by first adding ammonium sulfate to 25% and removing the precipitate by centrifugation. Additional ammonium sulfate was then added to the supernatant to a final concentration of 50%, and the protein precipitate was collected by centrifugation. A major portion of the integrase protein precipitated in this 25 to 50% window of ammonium sulfate concentration. The protein pellet was resuspended in lysis buffer and loaded onto a MonoQ FPLC column (Bio-Rad). The column was washed in buffer, and proteins were eluted with a 200 to 600 mM NaCl gradient. Fractions containing integrase were identified by SDS-polyacrylamide gel electrophoresis (PAGE). Integrase was found to elute at an NaCl concentration of approximately 300 mM. Peak fractions were typically 70 to 90% pure. Initially

gpInt was purified as a band identified by electrophoresis, since by itself it does not efficiently catalyze recombination. However, both crude extracts prepared from *E. coli* expressing gpInt and partially purified fractions are active in recombination when supplemented with an extract from *M. smegmatis* (see Results).

Preparation of bacterial cell extracts. Bacterial cell extracts were prepared from either *E. coli* DH5α or *M. smegmatis* mc^2155 (27). Typically, cells from a 1-liter culture of cells in late logarithmic growth were harvested by centrifugation at 14,000 rpm for 15 min and resuspended in lysis buffer. Cells were broken by sonication and clarified by centrifugation at 14,000 rpm for 15 min. In some cases, the extract was boiled for 10 min and then allowed to cool to room temperature for 15 min.

Assays for site-specific recombination. Assays for integrative intermolecular recombination contained approximately 1 μg of supercoiled pMH39 DNA and a 600-bp *attB* fragment derived by cleavage of pMH57 with *Eco*RI and *Hin*dIII. The *attB* DNA was radiolabeled at either the *Hin*dIII or *Eco*RI site by incorporation of radiolabeled deoxyribonucleotides with Klenow fragment (25) prior to cutting with the second enzyme. Other components of the recombination reaction were 10 mM Tris (pH 7.5), 25 mM NaCl, 1 mM dithiothreitol, 1 mg of bovine serum albumin per ml, and 10 mM spermidine. Typically, 2 μl of partially purified integrase, containing 0.1 to 1 μg of protein, was added to a total reaction volume of 20 μl. Reactions were carried out for 2 h at 25°C unless otherwise stated, terminated by the addition of SDS to 0.1%, and analyzed by electrophoresis through a 1% agarose gel (25). Recombination reactions were supplemented by addition of either 2 μl of cell extract, purified *E. coli* HU (a kind gift from Brigitte Lavoie and George Chaconas), or purified *E. coli* IHF (a kind gift from Howard Nash).

RESULTS

Overexpression and purification of L5 gpInt. Recombinant plasmids were constructed for L5 gpInt overexpression in *E. coli* on the basis of the T7 polymerase system (30, 31). Since it was not certain which codon is used for translation initiation of integrase in vivo, we constructed two separate expression plasmids containing different segments of the L5 gene that differ at the 5′ end. One contains a segment beginning at position 25369 in the L5 genome (8), corresponding to initiation at a UUG codon; the other begins at 25303 (note that integrase is transcribed leftward in the L5 genome [8]), close to an in-frame GUG codon (codon 28 relative to the UUG codon). The longer of the *int* segments was inserted such that an AUG initiation codon from vector sequences substituted for the presumed UUG initiation codon (Fig. 1A). The observation that the shorter protein is inactive in vitro (see below) is consistent with the conclusion that the UUG codon is probably used for translation initiation in vivo. Moreover, in the course of constructing these plasmids, we serendipitously isolated a mutant derivative in which the first position of codon 8 had a mutation resulting in substitution of proline for serine. This mutant protein is also defective in recombination (data not shown). L5 integrase appears to be one of at least five L5 genes that use UUG for translation initiation (8).

Induction of the T7 RNA polymerase resulted in the synthesis of a new protein of approximately 40 kDa that was not present prior to induction (Fig. 1B). Proteins of similar sizes were expressed from the two gpInt expression plasmids. The size of the expressed proteins corresponds well with the predicted molecular size of gpInt, which is 41.8 kDa. The N-terminally truncated form (gpInt 23–371) is predicted to have a very similar molecular size.

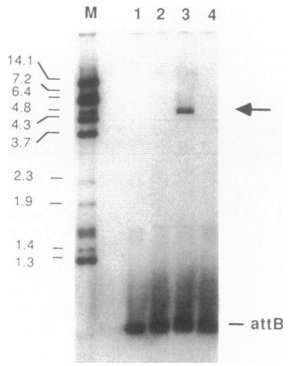

FIG. 2. L5 gpInt-mediated integrative recombination in vitro. Supercoiled plasmid DNA containing the L5 *attP* site (pMH39; 3.8 kb) and a 0.6-kb linear radiolabeled *attB* DNA were incubated with protein, and the reaction products were analyzed by agarose gel electrophoresis. Proteins were added to the reactions as follows: lane 1, no protein; lane 2, L5 gpInt; lane 3, L5 gpInt and a crude extract of *M. smegmatis*; lane 4, crude extract of *M. smegmatis*. Markers (lane M) are a radiolabeled *Bst*EII digest of lambda DNA with the sizes shown in kilobases. The arrow indicates the 4.4-kb product formed in presence of both L5 gpInt and *M. smegmatis* extract.

Both forms of L5 gpInt expressed from the *E. coli* plasmids were purified by a simple biochemical procedure. In brief, induced cells were harvested by centrifugation and broken open by sonication, and the lysate was clarified by centrifugation. L5 gpInt was then purified from other *E. coli* proteins by differential ammonium sulfate precipitation followed by chromatography through a MonoQ resin. Final preparations of gpInt were assessed to be 70 to 90% pure by SDS-PAGE (Fig. 1B).

L5 integrase catalyzes integrative recombination. To assay for recombination activity of the integrase preparations, purified L5 gpInt was added to reactions containing *attP* DNA (as supercoiled plasmid DNA) and a small radiolabeled linear *attB* DNA (Fig. 2); reaction products were then analyzed by agarose gel electrophoresis. We anticipated that recombination would result in the formation of a new radiolabeled DNA species with a size of approximately 4.4 kb (see Fig. 3A). Although in some reactions we did observe a small amount of such a product (in some experiments as much as 1% of the input *attB* substrate), this observation was not consistently reproducible.

We reasoned that efficient L5 gpInt-mediated recombination may require a factor encoded by the mycobacterial host. To test this, a crude extract was prepared from *M. smegmatis* and added to the reaction. This resulted in a large stimulation of the formation of a product migrating at the anticipated size of the recombinant molecule (Fig. 2). The stimulatory activity was heat resistant; activity was retained after boiling for 10 min, a property shared with *E. coli* IHF (21), HU (2), and Fis (3). Formation of the product required the addition of both L5 integrase and the *M. smegmatis* extract (Fig. 2).

To confirm that the reaction product was indeed that formed by recombination, the products were digested with restriction enzymes and the products were analyzed by agarose gel electrophoresis. The recombination scheme and the sizes of the expected products are shown in Fig. 3A. The results of restriction enzyme digestion (Fig. 3B) are consistent with the product being formed by site-specific recombination between the *attP* and *attB* sites.

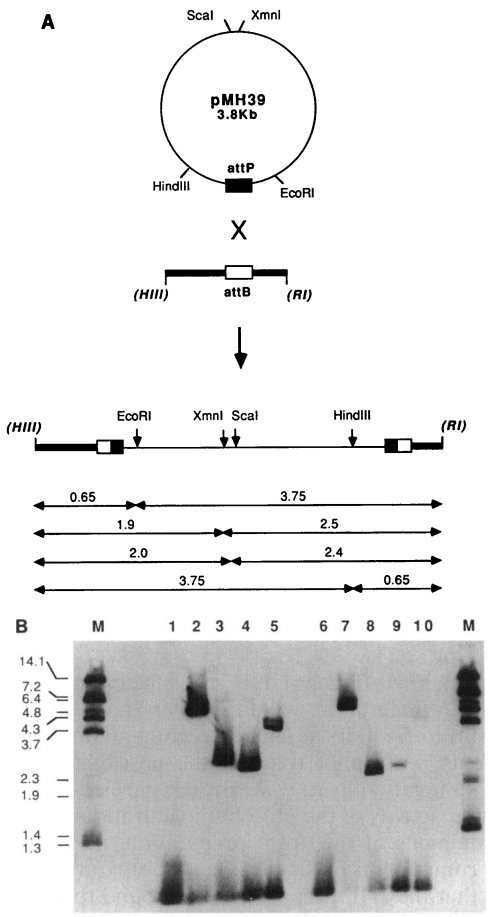

FIG. 3. Restriction enzyme analysis of recombination products. (A) Locations of relevant restriction enzyme sites. The recombination reaction contains a supercoiled 3.8-kb plasmid DNA, pMH39 (which contains the L5 *attP* site), and a small radiolabeled linear *attB* DNA (derived by *Hin*dIII and *Eco*RI digestion). Conservative site-specific recombination between the *attP* and *attB* sites is expected to give a linear 4.4-kb product. Digestion of this product with the enzymes indicated would yield fragments of the indicated sizes (shown in kilobases). If the *attB* DNA is labeled at one end only, then a single labeled fragment is observed. (B) Agarose gel electrophoresis of recombination products. The recombination reaction was performed as for Fig. 2 with L5 gpInt and an *M. smegmatis* extract. Lanes 1 to 5 contain *attB* DNA labeled at the *Eco*RI end; lanes 6 to 10 contain *attB* DNA labeled at the *Hin*dIII end. Reaction products were digested with no enzyme (lanes 2 and 7), *Xmn*I (lanes 3 and 8), *Sca*I (lanes 4 and 9), *Eco*RI (lane 5), and *Hin*dIII (lane 10). DNA size markers (lanes M) are shown in kilobases.

Other requirements for efficient recombination. The recombination reaction mediated by gpInt in combination with *M. smegmatis* extracts has been optimized with respect to time, temperature, and other requirements (Fig. 4). The reaction is reasonably slow and continues for at least 2.5 h (Fig. 4). Typically, between 50 and 80% of the *attB* substrate is converted to recombinant product in a 2.5-h period (Fig. 4). The reaction proceeds well at 25 or 37°C but is very inefficient at 42°C (Fig. 4B). Very little recombinant product is observed if spermidine is omitted from the reaction or if the *attP* DNA substrate is relaxed by topoisomerase I prior to the addition of integrase (data not shown). Efficient recombination requires neither Mg^{2+} ions nor ATP. In general, these parameters for L5 integrase-mediated recombination are very similar to those for lambda integrase recombination.

***E. coli* IHF and HU do not stimulate recombination.** The experiments described above show that a component of *M. smegmatis* is able to significantly stimulate L5 integrase-mediated recombination. The factor responsible shares the property of *E. coli* proteins HU and IHF in that it is heat stable and activity is retained even after boiling for 10 min. It seemed possible that in spite of the differences between mycobacterial and *E. coli* systems, *E. coli* IHF would be able to stimulate the reaction. Interestingly, we observed that *E. coli* IHF does not significantly stimulate recombination, although a small amount of recombination was observed at very high concentrations of IHF (Fig. 5). We also tested whether *E. coli* HU could stimulate recombination since it is less dependent on specific DNA sequences than is *E. coli* IHF. However, we observed no recombination in the presence of HU (Fig. 5). In addition, we saw no stimulation of recombination by addition of a crude *E. coli* extract (Fig. 5).

The N terminus of L5 gpInt is required for recombination. Plasmid pMH65 overexpresses a derivative of L5 integrase that contains only the segment from residues 23 to 371, fused to a short N-terminal 13-amino-acid leader (Fig. 1A). This N-terminally truncated protein has also been overexpressed and purified (Fig. 1B) but is unable to mediate recombination (Fig. 6). Crude extracts of the overexpressing *E. coli* cells were also tested and found to be inactive in recombination, indicating that we had not simply inactivated the protein during purification (data not shown). These observations are consistent with the hypothesis that the upstream UUG codon, rather than the downstream GUG codon previously suggested (16), is used in vivo. They also suggest that the extreme N-terminal segment of the protein is important for recombination function. While we cannot exclude the possibility that the 13-residue leader (vector-derived sequence) is positively interfering with activity, a mutant form of the protein with an amino acid substitution at position 8 was also unable to mediate recombination (data not shown). We note that the extreme N terminus of the putative full-length integrase protein is rich in amino acids with active side groups (Fig. 1).

DISCUSSION

We have described here the overexpression and purification of the integrase protein of mycobacteriophage L5 and established a simple in vitro recombination reaction. L5 integrase mediates efficient intermolecular integrative recombination in the presence of a host factor present in extracts of *M. smegmatis*. There are obvious similarities between L5 and lambda integration, even though their bacterial hosts are only distantly related. Moreover, while the two integrase proteins are members of the same integrase family of proteins, there is extremely little amino acid similarity.

Although the L5 and lambda integrase proteins may have diverged somewhat, they are of similar size and have functional as well as structural similarities. For example, integrative recombination is rather slow in vitro and somewhat temperature sensitive in both systems; neither reaction requires ATP, but both are stimulated by spermidine (21). The integrase from *H. influenzae* phage HP1 behaves similarly (6, 9). The lambda protein is composed of two domains, the N terminal of which is involved in specific interactions with the arm-type sites within *attP* (18). The extreme N-terminal part of L5 integrase also appears to be required for recombinational activity and contains many amino acid residues with the potential to interact with DNA. Moreover, we find that the N-terminally

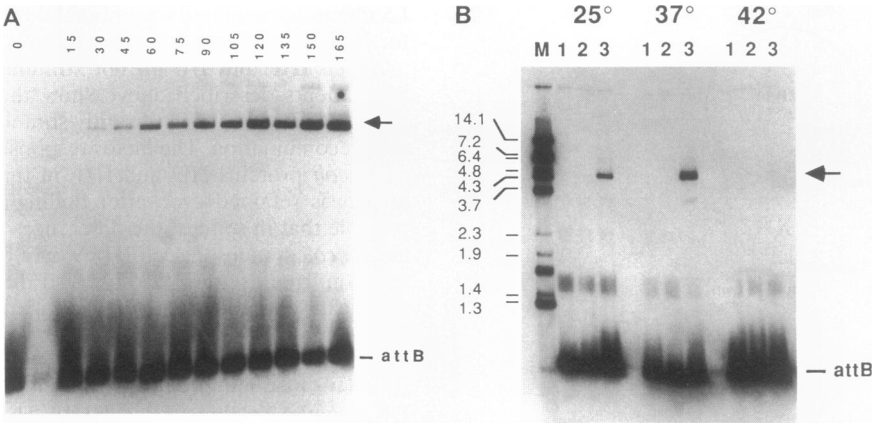

FIG. 4. Effects of time and temperature on recombination. (A) Kinetics of recombination. Recombination reactions were performed in the presence of L5 gpInt and *M. smegmatis* extract as described for Fig. 2 and 3. Reactions were terminated after incubation for the indicated times (minutes), and the products were analyzed by agarose gel electrophoresis. The *attB* substrate is indicated, and the recombination product is shown by an arrow. (B) Effect of temperature on recombination. Recombination reactions were performed as described for Fig. 2 and 3 but incubated at either 25, 37, or 42°C, as shown. For each set of reactions, lane 1 contained only L5 gpInt, lane 2 contained only *M. smegmatis* crude extract, and lane 3 contained both gpInt and *M. smegmatis* extract. DNA size markers (lane M) are shown in kilobases. The *attB* substrate is indicated, and the recombination product is shown by an arrow.

truncated integrase protein cannot bind to L5 *attP* arm-type sites even though it can bind to the *attP* core (our unpublished observations). We conclude that L5 integrase is likely to have an overall structure similar to that of lambda integrase.

Interestingly, L5 integration requires a novel mycobacterial host factor for efficient recombination. We propose that the factor responsible for stimulation should be referred to as mycobacterial IHF (mIHF) but emphasize that there is no evidence of any structural similarity with *E. coli* IHF. The factor shares the property of *E. coli* IHF and HU proteins in that it retains stimulatory activity even after boiling for 10 min. It seems possible that mIHF is a small heat stable protein of the general class to which IHF, HU, and Fis belong (2).

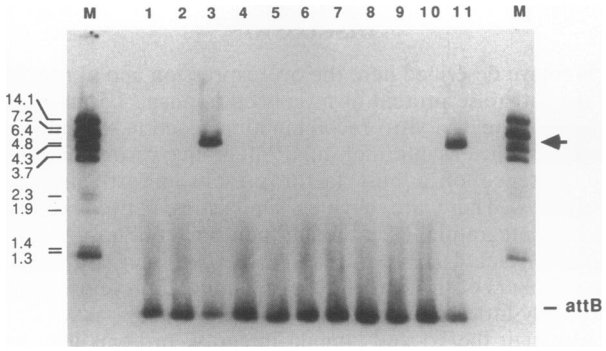

FIG. 5. Effects of host factors on L5 gpInt-mediated recombination. Recombination reactions were performed as for Fig. 2 and 3, and the products were analyzed by agarose gel electrophoresis. L5 gpInt was added to all reactions except those in lane 1 (no protein) and lane 4 (*M. smegmatis* extract only). Host proteins or extracts were added to the reaction as follows: lanes 3 and 11, crude boiled *M. smegmatis* extract; lane 5, crude extract of *E. coli*; lane 6, 67 ng of *E. coli* HU; lane 7, 6.7 ng of *E. coli* HU; lane 8, 1,100 ng of *E. coli* IHF; lane 9, 110 ng of *E. coli* IHF; lane 10, 11 ng of *E. coli* IHF. DNA size markers (lanes M) are shown in kilobases. The arrow indicates the position of the recombinant product. A small amount of recombinant product was observed at the highest concentration of IHF (lane 8) on the original autoradiograph.

However, neither IHF nor HU can efficiently substitute for mIHF in L5 integration; we do not yet know whether mIHF can substitute for IHF in lambda recombination. From these experiments, we cannot rule out the possibility that the absence of stimulation by *E. coli* extracts, purified HU, or IHF is due to poor activity of the proteins, inactivation of the proteins by contaminants in the gpInt preparation, or inactivation of gpInt by components of the host factor preparations. However, we note that the IHF preparation was found to be competent to give a DNA band shift and that HU was shown by Lavoie and Chaconas to be active in Mu recombination (15a). In addition, we observed that crude extracts of *E. coli* expressing gpInt were inactive in recombination even though they could be greatly stimulated by *M. smegmatis* extracts (data not shown).

It was found recently that expression of the *himA* gene of *Salmonella typhimurium* (which encodes an IHF subunit) is regulated in response to *Salmonella* infection of mice, using in vivo expression technology (17). Since IHF is known to be

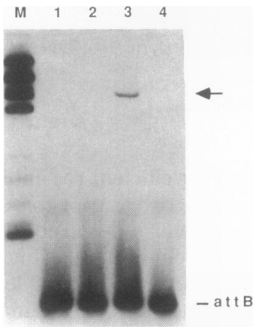

FIG. 6. The intact N terminus of L5 gpInt is required for recombination. Recombination reactions were performed as for Fig. 3, and the products were analyzed by agarose gel electrophoresis. Reactions contained the following proteins: lane 1, N-terminally truncated gpInt and *M. smegmatis* extract; lane 2, N-terminally truncated gpInt; lane 3, L5 gpInt and *M. smegmatis* extract; lane 4, L5 gpInt. Lane M, size markers.

involved in many aspects of gene regulation and DNA replication in addition to lambda recombination (4), it would not be surprising if it was also involved in some aspects of bacterial pathogenesis. The finding of a novel mycobacterial integration host factor raises the question as to whether it may be involved in mycobacterial virulence, including tuberculosis and leprosy.

ACKNOWLEDGMENTS

We thank Lisa Peterson and Lily Chang for excellent technical assistance, B. C. Wang for assistance in protein purification, and Gary Sarkis for comments on the manuscript.

This work was supported by a grant from the Vaccine Development Program of the World Health Organization.

REFERENCES

1. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L. Narayana, L. S. Peirson III, N. Sternberg, and J. M. Leong.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5**:433–440.
2. **Drlica, K., and J. Rouviere-Yaniv.** 1987. Histonelike proteins of bacteria. Microbiol. Rev. **51**:301–319.
3. **Finkel, S. E., and R. C. Johnson.** 1992. The Fis protein: it's not just for DNA inversion anymore. Mol. Microbiol. **6**:3257–3265.
4. **Friedman, D. I.** 1988. Integration host factor: a protein for all reasons. Cell **55**:545–554.
5. **Goodman, S., and H. Nash.** 1989. Functional replacement of a protein-induced bend in a DNA recombination site. Nature (London) **341**:251–254.
6. **Goodman, S., and J. Scocca.** 1989. Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of *Haemophilus influenzae*. J. Bacteriol. **171**: 4232–4240.
7. **Goodman, S. D., S. C. Nicholson, and H. A. Nash.** 1992. Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. Proc. Natl. Acad. Sci. USA **89**:11910–11914.
8. **Hatfull, G. F., and G. J. Sarkis.** 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. **7**:395–405.
9. **Hauser, M., and J. Scocca.** 1992. Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1. J. Biol. Chem. **267**: 6859–6864.
10. **Hsu, P. L., and A. Landy.** 1984. Resolution of synthetic att-site Holliday structures by the integrase protein of bacteriophage λ. Nature (London) **311**:721–726.
11. **Hwang, E., and J. Scocca.** 1990. Interaction of integration host factor from *Escherichia coli* with the integration region of the *Haemophilus influenzae* bacteriophage HP1. J. Bacteriol. **172**: 4852–4860.
12. **Kim, S., and A. Landy.** 1992. Lambda Int protein bridges between higher order complexes at two distant chromosomal loci *attL* and *attR*. Science **256**:198–203.
13. **Kim, S., L. Moisoto de Vargas, S. Nunes-Düby, and A. Landy.** 1990. Mapping of a higher order protein-DNA complex: two kinds of long-range interactions in λ attL. Cell **63**:773–781.
14. **Kitts, P., and H. Nash.** 1988. Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. J. Mol. Biol. **204**:95–107.
15. **Landy, A.** 1989. Dynamic, structural and regulatory aspects of λ site-specific recombination. Annu. Rev. Biochem. **58**:913–949.

15a. **Lavoie, B., and G. Chaconas.** Unpublished data.

16. **Lee, M. H., L. Pascopella, W. 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.
17. **Mahan, M. J., J. M. Slauch, and J. J. Mekalanos.** 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. Science **259**:686–688.
18. **Moisoto de Vargas, L., S. Kim, and A. Landy.** 1989. DNA looping generated by DNA binding protein IHF and the two domains of lambda integrase. Science **244**:1457–1461.
19. **Moisoto de Vargas, L., C. A. Pargellis, N. Hasan, E. Bushman, and A. Landy.** 1988. Autonomous DNA binding domains of λ integrase recognize two different sequence families. Cell **54**:923–929.
20. **Nash, H. A., C. E. Bauer, and J. F. Gardner.** 1987. Role of homology in site-specific recombination of bacteriophage λ: evidence against joining of cohesive ends. Proc. Natl. Acad. Sci. USA **84**:4049–4053.
21. **Nash, H. A., and C. A. Robertson.** 1981. Purification and properties of the *Escherichia coli* protein factor required for λ integrative recombination. J. Biol. Chem. **256**:9246–9253.
22. **Pargellis, C., S. Nunes-Düby, L. Moisoto de Vargas, and A. Landy.** 1988. Suicide recombination substrates yield covalent λ integrase-DNA complexes and lead to identification of the active site tyrosine. J. Biol. Chem. **263**:7678–7685.
23. **Richet, E., P. Abcarian, and H. A. Nash.** 1988. Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. Cell **52**:9–17.
24. **Robertson, C. A., and H. A. Nash.** 1988. Bending of the bacteriophage λ attachment site by Escherichia coli integration host factor. J. Biol. Chem. **263**:3554–3557.
25. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Snapper, S., L. Lugosi, A. Jekkel, R. Melton, T. Keiser, 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.
27. **Snapper, S., R. Melton, T. Keiser, and W. R. Jacobs, Jr.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol. Microbiol. **4**:1911–1990.
28. **Snyder, U., J. Thompson, and A. Landy.** 1989. Phasing of protein-induced DNA bends in a recombination complex. Nature (London) **341**:255–257.
29. **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.
30. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**:113–130.
31. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.