Molecular Microbiology (2002) **45**(6), 1515–1526

Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, φRv1

Lori A. Bibb and Graham F. Hatfull*
Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA.

Summary

The genomes of *Mycobacterium tuberculosis* H37Rv and CDC1551 each contain two prophage-like elements, φRv1 and φRv2. The φRv1 element is not only absent from *Mycobacterium bovis* BCG but is in different locations within the two sequenced *M. tuberculosis* genomes; in both cases φRv1 is inserted into a REP13E12 repeated sequence, which presumably contains the bacterial attachment site, *attB,* for φRv1. Although φRv1 is probably too small to encode infectious phage particles, it may nevertheless have an active integration/excision system and be capable of moving from one chromosomal position to another. We show here that the *M. tuberculosis* H37Rv φRv1 element does indeed encode an active site-specific recombination system in which an integrase of the serine recombinase family (Rv1586c) catalyses integration and excision and a small, basic φRv1-encoded protein (Rv1584c) controls the directionality of recombination. Integration-proficient plasmid vectors derived from φRv1 efficiently transform BCG, can utilize four of the seven REP13E12 sites present in BCG as attachment sites, and can occupy more than one site simultaneously.

Introduction

Mycobacterium tuberculosis, the causative agent of human tuberculosis, is the leading cause of death from a single infectious agent (Dye *et al*., 1999). Our understanding of this organism has been advanced by the recent determination of the complete genome sequences of two *M. tuberculosis* strains, one of which is a well-studied laboratory strain (H37Rv) and the other a recent clinical isolate (CDC1551) (Cole *et al*., 1998) (GenBank accession number NC_002755). The two genomes share a common organization and are very similar at the sequence level, although there are a number of notable differences (Cole *et al*., 1998) (GenBank accession number NC_002755).

A comparison of the genomes of *M. tuberculosis* H37Rv, *Mycobacterium bovis* and the avirulent vaccine strain *M. bovis* BCG by subtractive genomic hybridization revealed three regions that are present in *M. tuberculosis* but absent from BCG (Mahairas *et al*., 1996). One of these, RD3, represents a 9.2 kbp segment that is also present in a virulent laboratory strain of *M. bovis* (Mahairas *et al*., 1996). The 9247 bp sequence of the H37Rv RD3 region encodes at least 14 complete open reading frames (ORFs) (see Fig. 1A) (Mahairas *et al*., 1996; Hendrix *et al*., 1999; Hatfull, 2000), five of which are predicted to encode proteins with sequence similarity to known phage proteins including a capsid subunit (Rv1576c), prohead protease (Rv1577c), terminase (Rv1578c), primase/helicase (Rv1582c) and integrase (Rv1586c; see Fig. 1A). Because the small number of ORFs and the limited subset of phage genes suggests that this DNA segment is unlikely to represent an intact prophage capable of independently generating infectious particles, it has been referred to as prophage-like element φRv1 (Cole *et al*., 1998; Hendrix *et al*., 1999). In H37Rv the RD3 segment is flanked by two identical copies of a 12 bp element (5′-GGTTGGCCGTGG) in direct orientation; BCG contains just a single copy of this 12 bp sequence (Mahairas *et al*., 1996). A simple interpretation of the relationship between the H37Rv and BCG genomes in this region is that BCG contains an unoccupied attachment site (*attB*) for the φRv1 element; the 12 bp segment thus corresponds to a common core shared by the putative *attB* site and *attP* of φRv1 (or a φRv1 precursor) as well as the attachment junctions *attL* and *attR* in H37Rv (Fig. 1A).

In *M. tuberculosis* H37Rv, φRv1 lies within a copy of the degenerate repetitive element REP13E12 that is present in seven copies in the genome (Cole *et al*., 1998). Although these elements represent a rather diverse group of sequences – the most distantly related pair having only 63% nucleotide identity – each of these could contain a potential chromosomal attachment site (*attB*) for the φRv1 element. The idea of alternative attachment sites

Accepted 25 June, 2002. *For correspondence. E-mail gfh@pitt.edu; Tel. (+1) 412 624 6975; Fax (+1) 412 624 4870.

is supported by the finding that *M. tuberculosis* CDC1551 also carries a copy of ϕRv1 that is in a different location (GenBank accession number NC_002755). In this strain, ϕRv1 is integrated into a different copy of the REP13E12 element with a concomitant duplication of the same 12 bp sequence that is duplicated in H37Rv. However, both H37Rv and CD1551 carry just a single copy of ϕRv1 and all of the other REP13E12 elements are unoccupied.

Interestingly, *M. tuberculosis* H37Rv contains a second prophage-like element, ϕRv2. This second element encodes several proteins with sequence similarity to those of ϕRv1, has a similar organization, and is a similar length (10 982 bp) (Hendrix *et al*., 1999). Most of the size difference between the two elements is accounted for by an IS6110 element in ϕRv2. CDC1551 also contains a copy of ϕRv2, although the IS6110 insertion is absent, and in both strains the element is integrated into the same $tRNA^{Val}$ gene (Cole *et al*., 1998) (GenBank accession number NC_002755).

The presence of nearly identical copies of ϕRv1 at two different chromosomal locations in H37Rv and CDC1551 supports the hypothesis that the element is mobile and is not just a derelict copy of a larger parental prophage. Comparison of the two ϕRv1 elements shows that, although they are nearly identical, there are eight differences (mostly single-base insertions or deletions) between them (Table 1 and Fig. 1A); these alter the coding potential of three genes. Assuming that both genome sequences are devoid of sequencing errors in this region, the simple interpretation is that the CDC1551 ϕRv1 element has non-functional copies of genes corresponding to *Rv1575*, *Rv1576c* and *Rv1585c*. It is noteworthy that the *Rv1575* product is related to the upstream protein coded by *Rv1574 (*and could perform redundant functions), and ϕRv2, which is also present in CDC1551, contains a close relative (73% amino acid identity) of Rv1576c.

The avirulent strains of *M. bovis* BCG have neither ϕRv1 nor ϕRv2 (Mahairas *et al*., 1996). Although not all clinical isolates of *M. tuberculosis* carry a copy of ϕRv2, all appear to have at least one copy of either ϕRv1 or ϕRv2, suggesting that these related elements could play a role in determining the physiology of *M. tuberculosis* (Brosch *et al*., 2000). These observations are consistent with the hypothesis that the presence of ϕRv1 and ϕRv2 is accompanied by phenotypic consequences, and several possible roles can be contemplated. For example, one or more of the encoded genes could influence the growth rate, metabolic requirements or pathogenicity of the host. Alternatively, ϕRv1 and/or ϕRv2 could confer the ability to form virus-like particles capable of generalized transduction similar to the genetic transfer agents in *Rhodococcus* and *Methanococcus* (Bertani, 1999; Lang and Beatty, 2000). Regardless of whether these elements can form virus-like particles, they may nevertheless be able to move from one chromosomal location to another provided that they encode an active site-specific recombination system and that there is more than a single attachment site in the *M. tuberculosis* genome.

Analysis of the ϕRv1 sequence shows that it includes a gene (*Rv1586c*) coding for 469 amino acids (or 463 amino acids depending on the start site; see Fig. 2B) putative integrase of the serine recombinase type, with an N-terminal ~140 residue segment that has sequence similarity to transposon resolvases and DNA invertases. This suggests that Rv1586c is a member of the relatively newly recognized family of large serine integrases, of which there are now more than 20 members. While these all contain an N-terminal segment (~140 amino acids) related to the N-terminal catalytic domain of transposon resolvases, the length of the proteins varies greatly (400–700 amino acids) and there is little sequence similarity among their C-terminal segments. The closest relative of Rv1586c is ORF469 of bacteriophage R4, although these share only 33% identity overall and just 29% in the region downstream of position 140 (Fig. 1B). The only mycobacteriophage known to encode a serine integrase is Bxb1 (Mediavilla *et al*., 2000), although the Bxb1 Int (gp35) and Rv1586c share only 26% identity over the C-terminal segment. Rather little is known about this class of recombinases, although active systems have been demonstrated for phages ϕC31 and TP901-1 (Thorpe and Smith, 1998; Breuner *et al*., 2001). The mechanisms for the control of

Table 1. Sequence differences in ϕRv1 between *M. tuberculosis* strains H37Rv and CDC1551.

Change. no	H37Rv ORF	H37Rv protein size (aa)	CDC1551 change (position in basepairs)	Consequence
1	Rv1573	136	G → C (57)	Conserved
2	Rv1575	117	G insertion (243) and	Frameshift, 102 aa protein
3			G → C (281)	
4, 5	Rv1576c	473	GC deletion (185, 186)	First frameshift, 68 aa protein
6			G deletion (250)	Second frameshift, 68 aa protein
7			G deletion (700)	Third frameshift, 68 aa protein
8	*Rv1585c*	170	C deletion (393)	Frameshift; fusion with *Rv1584c*

Changes are numbered as shown in Fig. 1A. Base changes listed are those in the coding strand and positions listed are with the ORF.

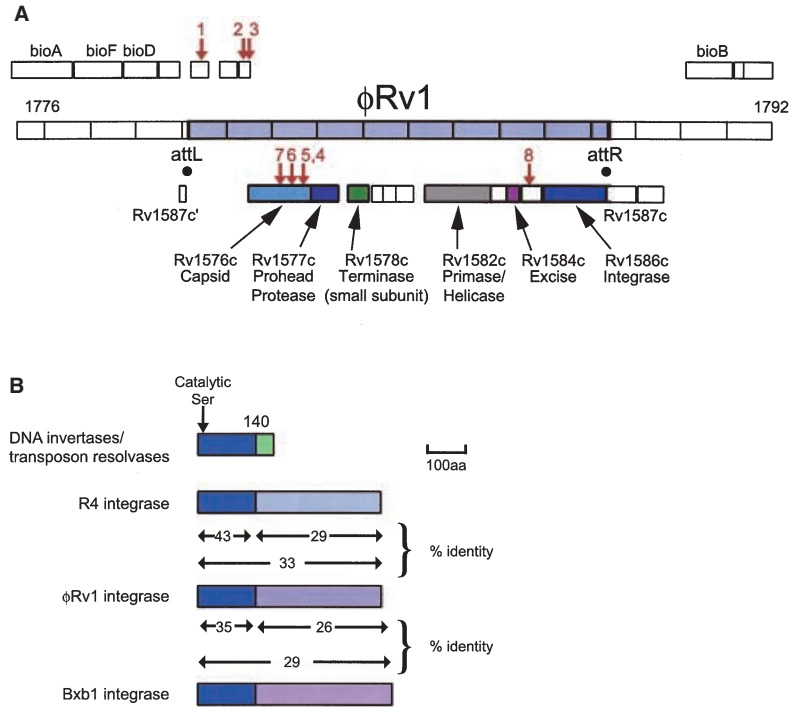

Fig. 1. Organization of the *M. tuberculosis* H37Rv prophage-like element ϕRv1.
A. The *M. tuberculosis* genome is represented by the bar with 1 kbp-spaced markers; the ϕRv1 element is shaded light blue. Open reading frames (ORFs) are shown as boxes above (transcribed rightwards) and below (transcribed leftwards) the genome and labelled with gene number and putative function where known or predicted. The ϕRv1 element is inserted into a copy of the REP13E12 repetitive element that is located within the biotin operon and interrupts gene *Rv1587c*. The vertical red arrows indicate sequence differences between the ϕRv1 elements in H37Rv and CDC1551 that are described in further detail in Table 1.
B. Structural organization of serine integrases. Phage-encoded serine integrases contain an N-terminal domain approximately 140 residues long (shown in navy blue) that has sequence similarity with the N-terminal catalytic domain of transposon resolvases and DNA invertases. The C-terminal segments are highly diverse in both sequence and length and are of unknown function although they presumably include the DNA-binding determinants. The closest relative to the ϕRv1 integrase is that encoded by phage R4. Phage Bxb1 is the only mycobacteriophage with a known serine integrase.

directionality in these recombination schemes are not understood, although TP901-1 ORF7 promotes excision *in vivo* (Breuner *et al*., 1999). In the ϕC31 system, integrase alone is unable to mediate excisive recombination, and a factor controlling directionality has yet to be identified (Thorpe *et al*., 2000).

We report here that the ϕRv1 element encodes a fully functional integration system which can utilize multiple attachment sites in slow-growing mycobacterial strains. Non-replicating plasmids carrying the ϕRv1 *integrase* and the presumptive *attP* site efficiently transform BCG through integration into the chromosome. Analysis of

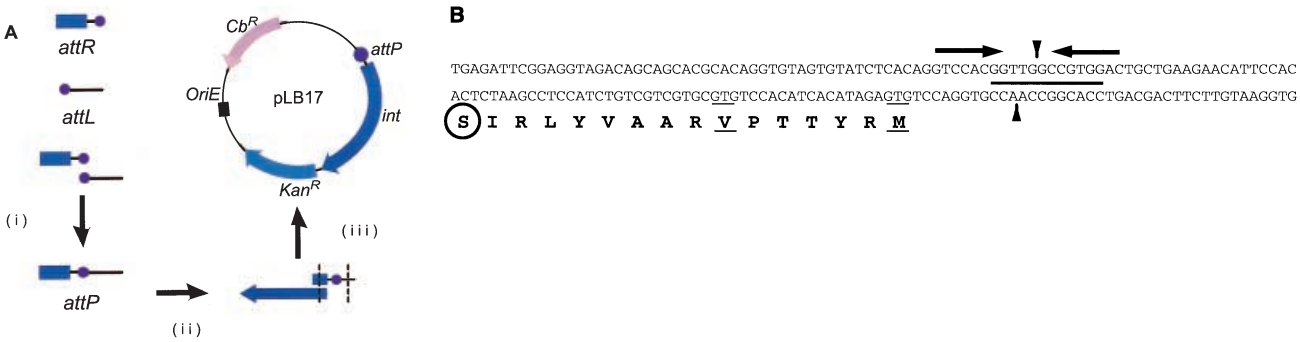

Fig. 2. Construction of ϕRv1 integration-proficient vector pLB17
A. In the first step (i) ϕRv1 *attP* was generated from H37Rv DNA using two PCR-amplified fragments containing *attL* and *attR* followed by annealing via the common 12 bp core segment (shown as a violet circle) and amplification by the outside primers. This *attP* fragment also contains the 5′ end of the integrase gene (blue box). Next (ii) this *attP* fragment was fused to a second DNA segment containing the remainder of the putative integrase gene, *Rv1586c*, and finally (iii) inserted into plasmid vector carrying an *E. coli* origin of replication (OriE) and resistance genes for ampicillin (Cb^R) and kanamycin (Kan^R).
B. Partial sequence of the ϕRv1 *attP* site. The 12 bp common core shared by *attP* and *attB* (underlined) overlaps 8 bp inverted repeats that are separated by a 4 bp spacer (horizontal arrows). Vertical arrowheads indicate the predicted positions of strand cleavage during recombination (see also Fig. 4B). The 5′ end of the leftwards-transcribed integrase gene lies immediately to the left of these inverted repeats; the putative catalytic serine is circled. There are two possible translational start sites for the integrase gene (underlined residues and codons). The start codon closest to the core is the assigned start in H37Rv, while the more distal GUG codon is annotated as the start in CDC1551. There is no experimental evidence to indicate which of these is used.

these transformants shows that four of the seven REP13E12 repeats are utilized as *attB* sites, and that multiple integration events can occur. These ϕRv1 integration-proficient plasmids do not efficiently transform the fast-growing *M. smegmatis* unless a functional *attB* site from BCG has been introduced by an alternative integration system. We also show that *Rv1584c* codes for a directionality factor that can be used to confer efficient loss of integrated DNAs in recombinant BCG strains.

Results

The M. tuberculosis *ϕRv1 prophage-like element encodes an active integration system*

To determine whether ϕRv1 encodes an active integration system, we constructed non-replicating integration-proficient plasmids carrying the putative integrase gene (*Rv1586c*) and the presumptive *attP* site, and tested whether they are capable of transforming BCG via site-specific integration. As a viral or free form of the ϕRv1 element has yet to be described, it was first necessary to reconstruct the presumptive *attP* site from the *attL* and *attR* junctions present in the H37Rv genome.

The putative *attP* site was generated from *attL* and *attR* using a polymerase chain reaction (PCR) scheme (Fig. 2A). In choosing how large a DNA fragment might be needed to encompass the functional *attP* site we took into account the following considerations. As the putative *int* gene is close to the 12 bp core of *attP*, we decided to use a contiguous *int–attP* DNA segment, thus eliminating any concern about the functional requirement of *attP* on this side. On the *int*-distal side of *attP,* we chose arbitrarily to include 177 bp to the right side of the core. Following amplification of a 371 bp *attP* fragment using *attL* and *attR* templates (Fig. 2A) and a second fragment containing the *Rv1586c* ORF, the products were inserted into a kanamycin-resistant plasmid vector. The resulting plasmid, pLB17, thus contains a 1621 bp fragment that includes the putative *attP* site and *int* gene of ϕRv1 (Fig. 2A), confers kanamycin resistance and can replicate in *Escherichia coli* but not in mycobacteria. It is noteworthy that the conceptual reconstruction of ϕRv1 *attP* reveals that there are 8 bp perfect inverted repeat sequences overlapping the boundaries of the 12 bp core with the centre of the site symmetry lying between the fourth and fifth positions of the 12 bp common core. If the pattern of strand cleavages were similar to that of other serine recombinases then it would be expected to generate two-base 3′ extensions during recombination (Fig. 2B).

The ability of pLB17 to integrate into the genomes of slow-growing mycobacteria was determined by transformation of the Connaught and Pasteur strains of *Mycobacterium bovis* BCG; both of these strains lack ϕRv1 and instead have an unoccupied *attB* site. As the plasmid is unable to replicate in mycobacteria, antibiotic-resistant transformants should arise only through stable plasmid integration into the BCG chromosome, similarly to those described previously for L5-, D29- and Ms6-derived integration-proficient vectors (Lee *et al*., 1991; Freitas-Vieira *et al*., 1998; Peña *et al*., 1998). We observed that pLB17 does efficiently transform BCG, yielding approximately 5×10^4 transformants per μg of DNA (Table 2); this is similar to transformation with an L5-derived integration vector or with an oriM-containing plasmid (Table 2). We conclude that the 1621 bp segment of *M. tuberculosis* DNA inserted into plasmid pLB17 contains a functional *attP* site and integrase gene and that ϕRv1 encodes a recombinationally active integration system.

Plasmid pLB17 utilizes multiple attB sites for integration

The 12 bp sequence that is common between ϕRv1 *attP* and the *attB* site that is occupied in H37Rv is present six times in the *M. tuberculosis* genome, five of which are within a REP13E12 sequence. The REP13E12 element itself is present seven times in *M. tuberculosis*, but two of these contain sequence variants of the 12 bp core (Cole *et al*., 1998). To determine which of the seven REP13E12 repeats in *M. tuberculosis* H37Rv are also present in BCG Connaught, seven pairs of primers were designed that uniquely amplify core-containing DNA fragments in each of the seven REP13E12 repeats. Amplification of BCG Connaught DNA with these primers indicates that all seven repeats are present (Fig. 3A), although there are several sequence differences between the sites in H37Rv

Table 2. Transformation efficiency of BCG Connaught and BCG Pasteur by extrachromosomal and integration proficient plasmids.

		Number of colonies	
Plasmid	Features	Connaught	Pasteur
pLB17(Kan^R)/pLB25(Hyg^R)	ϕRv1 *attP* and *integrase*	31	410
pMH94(Kan^R)/pGS67(Hyg^R)	L5 *attP* and *integrase*	39	610
pJL37(Kan^R)/pYUB415(Hyg^R)	Extrachromosomal (pAL5000 *oriM*)	14	500
No DNA	–	4	5

Electroporations were carried out using 100 ng of each DNA and the number of colonies obtained when plating ~10 ng is reported.

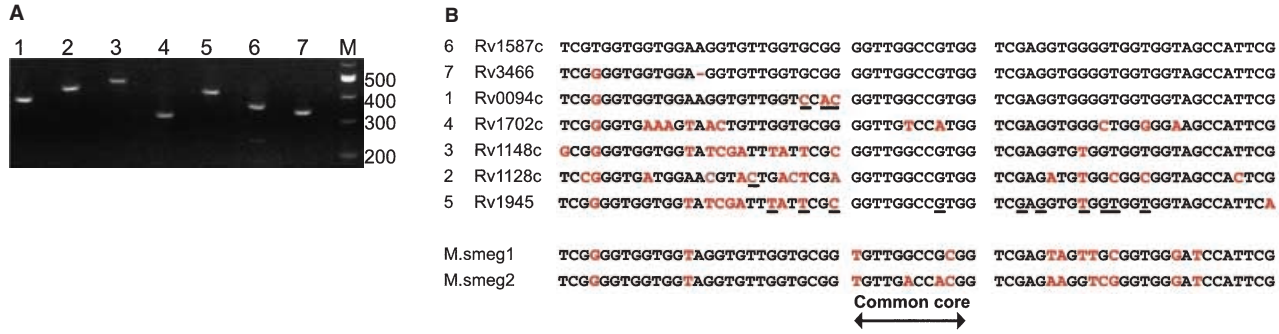

Fig. 3. *Mycobacterium bovis* BCG Connaught contains seven REP13E12 elements.
A. Seven pairs of primers designed from the sequences of the seven REP13E12 elements in *M. tuberculosis* H37Rv were used to amplify BCG Connaught DNA. The PCR products – each of which contains a putative *attB* site for the ϕRv1 element – are shown in lanes 1–7.
B. The DNA sequence around the 12 bp common core within each of the seven PCR products derived from BCG Connaught DNA was determined as shown. The number of each site shown on the extreme left corresponds to the lane number in part A. The gene number of each of the corresponding REP13E12 ORFs in *M. tuberculosis* H37Rv is also indicated. Site #6 is that occupied by ϕRv1 in *M. tuberculosis* H37Rv and sequence departures from this sequence are shown in red. Sequence differences between BCG-Connaught and H37Rv are shown as underlined bases. The sequences of two REP13E12 elements in *M. smegmatis* (M.smeg1 and M.smeg2) are also shown with departures from site #6 (*Rv1587c*) shown in red. These sequences differences between the *M. smegmatis* sites and *M. tuberculosis* site #6 probably contribute to the inefficient use for integration by plasmid pLB17. *M. smegmatis* sequences were obtained by BLAST analysis from the preliminary genome sequence available at http://www.tigr.org.

and BCG (Fig. 3B). Site 5 (which corresponds to the *M. tuberculosis Rv 1945* site, which has a single deviation from the consensus 12 bp sequence) has a change that generates a 12 bp core identical to the *attP* core (Fig. 3B).

To determine which of these sites are used for integration by pLB17, individual BCG transformants were analysed using seven sets of REP13E12 primers; these generate DNA fragments corresponding to either *attB* or *attL*, depending on whether integration has occurred at that particular site (Fig. 4A). A total of 41 transformants were tested, and in each case, at least one of the *attB* products was absent and replaced by the corresponding

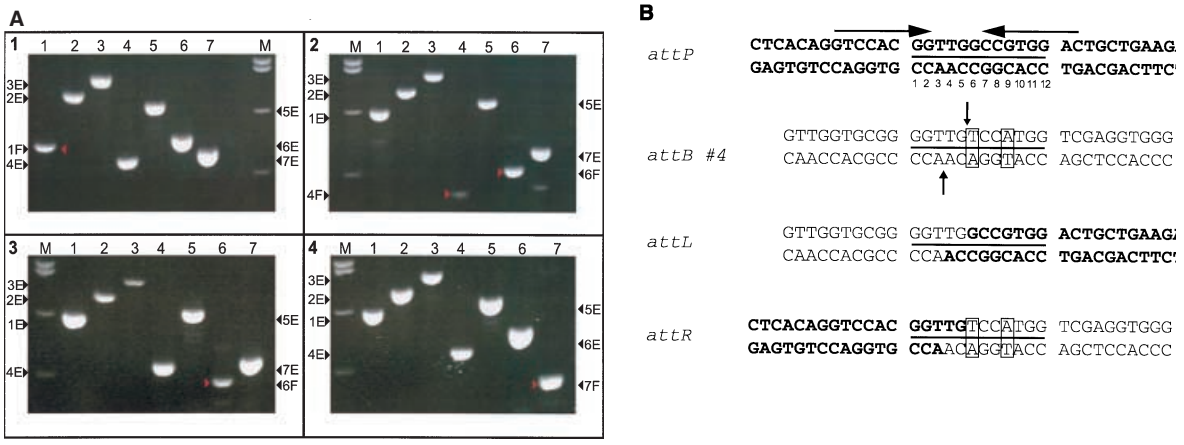

Fig. 4. Characterization of ϕRv1-mediated integration events in BCG Connaught.
A. Individual transformants recovered from electroporation of BCG Connaught with pLB17 DNA were used as substrates in seven separate PCR reactions; four representative samples are shown. Each panel represents the PCR analysis of a single transformant, and lanes 1–7 represent the amplification of the corresponding sites shown in Fig. 3. The reactions differ from those shown in Fig. 3 in that three primers were used to yield either *attB* or *attL* products depending on whether the site is vacant or occupied by pLB17. Amplification of *attL* in each panel is indicated by a red arrowhead. The transformants shown have pLB17 integrated into site 1 (panel 1), sites 4 and 6 (panel 2), site 6 (panel 3) and site 7 (panel 4).
B. Mapping the crossover site for ϕRv1 integration. The sequences of sites *attB* 4 and *attP* are shown in plain type and bold face respectively. The horizontal arrows represent the 8 bp perfect inverted repeats in *attP*. The 12 bp common core of the BCG Connaught REP13E12 site 4 differs from *attP* at positions 6 and 9. Sequence determination of the *attL* and *attR* junctions sites following pLB17 transformation shows that strand exchange has taken place to the left of position 6 of the 12 bp core. As the centre of symmetry within *attP* lies between positions 4 and 5, integrase presumably cuts to leave a two-base extension. It seems probable that the cleavage is made to generate 3′ extensions, as shown for transposon resolvases and DNA invertases.

Table 3. Distribution of ϕRv1 *attP–integrase* plasmid single-integration events in *M. bovis* BCG Connaught as determined by PCR analysis.

Site no.	Corresponding H37Rv ORF	Core sequence	Number of sites occupied/ total colonies assayed
1	*Rv0094c*	GGTTGGCCGTGG	11/39
2	*Rv1128c*	GGTTGGCCGTGG	0
3	*Rv1148c*	GGTTGGCCGTGG	0
4	*Rv1702*	GGTTGTCCATGG	0
5	*Rv1945c*	GGTTGGCCGTGG	0
6	*Rv1587c*	GGTTGGCCGTGG	15/39
7	*Rv3466c*	GGTTGGCCGTGG	13/39

Two transformants with double-integration events in sites 4 and 6 and sites 1 and 6 are not included.

attL product; this confirms that transformation was accompanied by site-specific integration of the plasmid. The majority of transformants (39/41) were missing just one of the *attB* fragments, although two had lost two of the products, suggesting that multiple integration events had occurred. Of the 39 single integration events, 11 occurred in site 1, 15 in site 6 and 13 in site 7 (Table 3). One of the double integration events utilized sites 4 and 6 and the other used sites 1 and 6. No integration events were seen in site 2, 3 or 5. As all of the transformants appeared to have a copy of pLB17 integrated into at least one of the REP13E12 repeats, it seems unlikely that there are any additional sites in the BCG genome that are efficient targets for the ϕRv1 integration system, including the one non-REP13E12 copy of the 12 bp core. The use of sites 6 and 7 is not unexpected as these are the sites occupied in *M. tuberculosis* strains H37Rv and CDC1551 respectively (Cole *et al*., 1998) (GenBank accession number NC_002755)**.** Moreover, this analysis shows that they do participate efficiently in recombination.

In BCG, alignment of the sequences surrounding the 12 bp core of the seven REP13E12 elements suggests strongly that the 12 bp core sequence alone is insufficient to confer identity as an *attB* site (Fig. 3B). Six of the seven repeats (1, 2, 3, 5, 6 and 7) possess the 12 bp sequence identical to that in the ϕRv1 *attP* site, but only three of these (1, 6 and 7) are used efficiently for integration. These three sites share extensive similarity to the right of the 12 bp core sequence, whereas to the left there are several differences between otherwise very similar sequences. Sites 2, 3 and 5 both contain the 12 bp core but have a number of sequence departures from sites 1, 6 and 7, particularly to the left of the core as represented in Fig. 3B. Presumably, the sequences flanking the 12 bp core are also required for *attB* identity. Moreover, as all of these sites reside within a repetitive element, it seems unlikely that the failure to recover integrants in these sites results from interruption of some essential function.

One integration event was seen in site 4, which contains two sequence departures from the 12 bp core sequence but shares the same 10 bp flanking the core to the left and to the right. Although only one event was observed in site 4, this suggests that the presence of all 12 bp within the core is not absolutely required for recombination. When the attachment junctions generated by integration into site 4 were sequenced, we found that recombination occurred between the left end of the core and the sequence difference at position 6 of the core (Fig. 4B). This is consistent with the interpretation of the likely recombination cut sites predicted from the site-symmetry seen in *attP* (Fig. 2B).

Integration of pLB17 in M. smegmatis

It is not known if the REP13E12 element found in *M. tuberculosis* and BCG is also present in *M. smegmatis*, and this raises the question as to whether pLB17 (and ϕRv1) are capable of efficient integration into *M. smegmatis*. Also, we cannot rule out the possibility that there may be host factors required for integration of pLB17 in the slow-growing mycobacteria that are not present in *M. smegmatis*. To address these issues, we determined the efficiency of transformation of pLB17 in *M. smegmatis* mc^2155 (Table 4). We observed that under conditions in which other integration-proficient or *oriM*-containing plasmids efficiently transformed *M. smegmatis* (~10^5 trans-formants per μg DNA) only a few hundred pLB17 transformants could be recovered. Although the number of transformants was small, it was consistently greater than the background number of antibiotic-resistant colonies obtained with non-replicating, non-integrating plasmid DNA (Table 4).

To determine the site of pLB17 integration in *M. smegmatis*, genomic DNA was prepared from two independent transformants, cleaved with *Nco*I (which does not cut within pLB17), ligated and recovered in *E. coli*. Sequencing of the attachment junctions showed that integration had indeed occurred via *attP* and that the same chromosomal site was used in both transformants. Comparison of this sequence with that of the unfinished genome of *M. smegmatis* (http://www.tigr.org) revealed that there are at least two segments that are similar to each other (>90% identity) and share sequence similarity (approximately

Table 4. Transformation efficiency of *Mycobacterium smegmatis* strains by extrachromosomal and integration-proficient plasmids.

		Number of colonies		
Plasmid	Features	mc^{2}155	LAB7	GS67
pLB17	ϕRv1 attP and *integrase*	6	~2300	8
pMD169	L5 *attP* and *integrase*	~1600	ND	ND
pJL37	Extrachromosomal (pAL5000 *oriM*)	~1400	~7600	~3100
pMD02	no *OriM*	0	ND	ND

Electroporations were carried out using 100 ng of each DNA, and the number of colonies obtained when plating ~10 ng is reported. GS67 is mc^{2}155 containing plasmid pGS67.
ND, not determined.

50% identity) to the H37Rv REP13E12 elements. However, these *M. smegmatis* REP13E12-like repeats differ from the *M. tuberculosis* repeats within both the 12 bp core and the flanking sequence (Fig. 3B). PCR analysis of an additional 25 pLB17 transformants showed that the majority had utilized one of these sites for integration (Fig. 5).

Efficient transformation of M. smegmatis *by pLB17 requires an attB site from BCG*

To determine if the reduced efficiency of pLB17 transformation of *M. smegmatis* arises from the inefficient use of the REP13E12-like repeats as *attB* sites, we constructed a recombinant strain (LAB7) in which a copy of BCG REP13E12 site 6 was introduced on a hygromycin-resistant L5-derived integration-proficient plasmid vector (pLB18). In contrast to *M. smegmatis* mc^{2}155, pLB17 transforms LAB7 cells with a similar efficiency to that of L5-derived integration-proficient vectors, approximately 300-fold greater than *M. smegmatis* or a recombinant lacking the BCG *attB* site (Table 4). PCR analysis of these transformants demonstrated that the BCG *attB* site was occupied in every transformant. These experiments suggest that the inefficient transformation of *M. smegmatis* by pLB17 reflects the lack of a suitable *attB* site for efficient integrative recombination rather than the absence of a host factor that is present only in slow-growing mycobacteria.

Plasmid pLB17 is stably maintained in M. smegmatis

To determine if the ϕRv1-derived integration-proficient plasmid pLB17 is stably maintained in the absence of selection, two pLB17 transformants of *M. smegmatis* strain LAB7 and two transformants containing an extrachromosomal plasmid (pMD30) were cultured in media without kanamycin. After ~40 generations of unselected growth, we detected little or no loss (<0.5%) of plasmid pLB17, whereas 18% of cells lost the extrachromosomal plasmid. The simple interpretation of this observation is that an additional ϕRv1-encoded protein is required for excision, although we cannot rule out the possibility that the integrase protein alone is capable of mediating excision but is not expressed in the integrated strains.

Rv1584c encodes a recombination directionality factor

As ϕRv1 contains only a small number of predicted ORFs (Fig. 1A), and several can be assigned putative functions based on sequence similarity, the number of ORFs that could encode a recombination directionality factor (RDF or Excise) is small. In the related ϕRv2

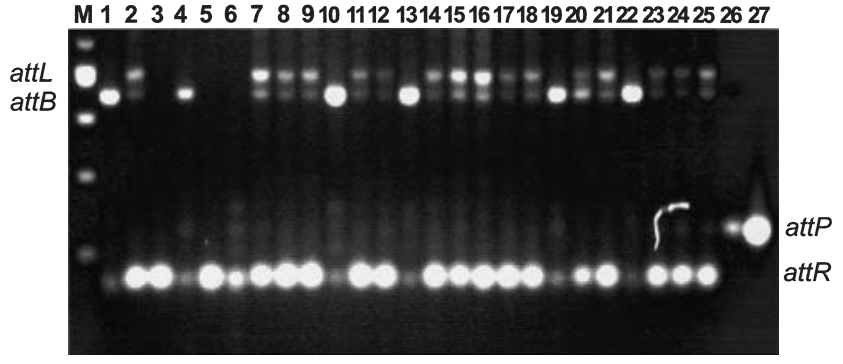

Fig. 5. Identification of integration sites in *M. smegmatis*. Twenty-five individual transformants derived by electroporation of *M. smegmatis* mc^{2}155 with pLB17 DNA were analysed by PCR using primers that amplify either *attL* or *attB* DNA (lanes 2–25). Control lanes show the products derived from *attB* (lane1) and *attP* (lane 26 and 27).

element – which uses a tyrosine recombinase – the second ORF (*Rv2657c*) downstream of the integrase encodes a protein with low but significant amino acid sequence similarity to the RDF encoded by phage L5 (gp36; Lewis and Hatfull, 2001). Although little is known about the control of directionality in the serine integrase reactions, we reasoned that Rv1584c – which is in a collinear position in ϕRv1 relative to Rv2657c in ϕRv2 – is a good candidate for the ϕRv1 RDF, and it also shows weak but perhaps meaningful sequence similarity with these proteins (Fig. 6A). To test this, we constructed a plasmid in which Rv1584c is fused to a low-level promoter (a mutant derivative of the L5 P_{left} promoter) in a tetracycline resistance extrachromosomally replicating plasmid (pLB36); similar plasmids fusing *Rv1584c* to the stronger hsp60 promoter do not efficiently transform mycobacteria, presumably because of toxicity of the gene product. Electroporation of *M. smegmatis* LAB7 with plasmid pLB36 efficiently yielded tetracycline-resistant transformants, but only in the absence of kanamycin (the kanamycin resistance gene is on pLB17); in the presence of kanamycin, no transformants were obtained. When individual tetracycline-resistant transformants were screened for kanamycin resistance, all of the 100 tested were kanamycin sensitive. All of the transformants maintained resistance to hygromycin, indicating that the loss of kanamycin resistance was not due to loss of the L5 integration plasmid vector. PCR analysis of the transformants demonstrated that loss of the kanamycin marker was indeed a consequence of precise excision of pLB17 (Fig. 6B). We conclude that *Rv1584c* encodes a directionality factor for controlling ϕRv1 site-specific recombination.

Discussion

We show here that the prophage-like element ϕRv1 encoded by *M. tuberculosis* H37Rv is fully active in both integrative and excisive recombination. Given that both ϕRv1 and ϕRv2 are considerably smaller than the 30–40 kbp that is typical of small but fully functional prophage genomes (i.e. those capable of generating infectious viral particles) they probably do not represent active prophages. None of the genes encoded by ϕRv1 in *M. tuberculosis* H37Rv appears to be a pseudogene, and the element is fully competent to integrate and excise. The putative capsid subunit encoded by *Rv1576c* is a close relative of the putative capsid subunit of a newly characterized mycobacteriophage, Che9c (our unpublished observations), and it seems plausible that the ϕRv1 element in *M. tuberculosis* H37Rv is capable of making DNA-containing particles. In contrast, the ϕRv1 element in *M. tuberculosis* CDC1551 has several mutations that interrupt three of the genes, including *Rv1576c* , which encodes the putative capsid subunit. It is possible, however, that the corresponding genes in the ϕRv2 element could complement these defects.

The activity of the ϕRv1 integrase is of interest as it is a member of the rather poorly understood class of phage

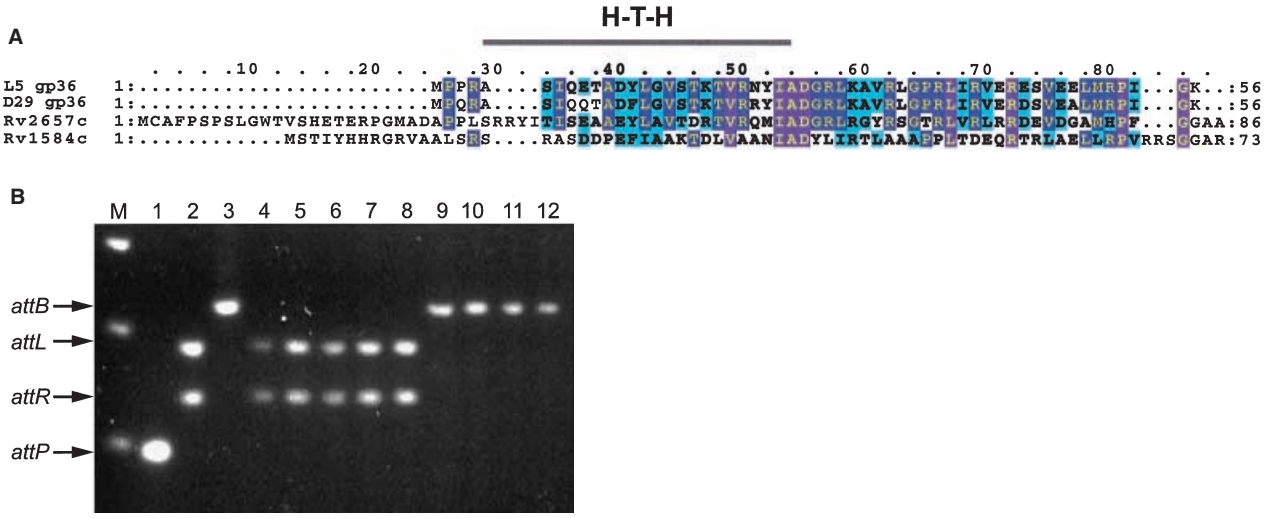

Fig. 6. Identification of ϕRv1 recombination directionality factor.
A. ClustalX alignment of the putative ϕRv1 Excise (Rv1584c) with the excise proteins of mycobacteriophages L5 and D29 (gp36) and the putative excise protein of ϕRv2 (Rv2657c). The grey bar represents the predicted helix–turn–helix DNA-binding motif in L5 gp36.
B. Rv1584c-mediated excision in *M. smegmatis*. Transformants of *M. smegmatis* LAB7 + pLB17 were analysed by PCR to determine the occupancy of the ϕRv1 *attB* site. Transformants containing plasmid pLB36 (which carries *Rv1584c*; lanes 9–12) all contain an unoccupied *attB* site, showing that excision of the resident plasmid has occurred. Transformants containing a vector control plasmid (pJL32; lanes 5–8) retain the *attL* and *attR* junction sites. Lanes 1–4 are control PCR reactions using plasmid templates for *attP*, *attL* and *attR* and *attB* (lanes 1–3 respectively) as well as untransformed *M. smegmatis* LAB7 + pLB17 (lane 5).

integrases that use a catalytic domain (and an active site serine) that is more typical of transposon resolvases and DNA invertases. Currently, 11 other serine integrases associated with phage or phage-like elements have been described, six of which (encoded by ϕC31, TP901-1, R4, A118, SPβc2 and ϕFC1) have been shown to be active either *in vitro* and/or *in vivo* (Kuhstoss and Rao, 1991; Christiansen *et al*., 1996; Matsuura *et al*., 1996; Thorpe and Smith, 1998; Lazarevic *et al*., 1999; Loessner *et al*., 2000; Breuner, 2001; Yang *et al*., 2002). All of the members of this class of integrases contain related ~140-residue N-terminal segments common to resolvases and DNA invertases, but the C-terminal domains are highly diverse in their lengths and sequences. The function of the C-terminal region is not known but presumably includes the DNA-binding determinants. However, as the presumptive recognition sites differ between *attP* and *attB* (Figs 2B and 3B), there may be more than a single DNA-binding domain within this C-terminal domain. Moreover, the protein–DNA interactions may be complex, as the attachment junctions at *attL* and *attR* contain combinations of these specificities.

The minimal DNA segment required for the *attP* function in integration systems that utilize tyrosine integrases is relatively large, and for both coliphage lambda and mycobacteriophage L5 it is approximately 250 bp (Hsu *et al*., 1980; Mizuuchi *et al*., 1981; Peña *et al*., 1997). In contrast, the minimum size required for a functional *attP* site in the serine integrase systems appears to be much smaller, and in ϕC31, TP901-1 and R4 *attP* function lies within a 39 bp, 56 bp and 64 bp segment respectively (Groth *et al*., 2000; Breuner, 2001; Olivares *et al*., 2001). As the crossover point within the *attP* site of ϕRv1 is very close to the 5′ end of the integrase gene (Fig. 2B), it seems probable that the minimum functional *attP* site of ϕRv1 is also relatively small. The symmetry of the *attP* site is indicated by the 8 bp inverted repeats overlapping the 12 bp core such that the centre of symmetry lies between the fourth and fifth bases of the 12 bp core. Because recombination occurs to the left of the base pair at position 6 (Fig. 4B) then the ϕRv1 integrase may cleave the DNA to generate two-base 3′ extensions as demonstrated for other serine recombinases (Hatfull and Grindley, 1988).

Control of the directionality of site-specific recombination catalysed by tyrosine integrases is typically mediated by a small DNA-binding protein (gpXis) that influences the architecture of complex protein–DNA structures (intasomes). An *in vivo* role for an excise protein has also been shown for directional control of the TP901-1-encoded serine integrase system, but its mechanism of action is not known (Breuner *et al*., 1999). The ϕRv1 excise protein encoded by *Rv1584c* does not share any obvious sequence similarity to the TP901-1 excise, although both proteins are extremely small (64 and 73 amino acids for TP901-1 and ϕRv1 respectively). The *Rv1584c* product does share weak sequence similarity to the known excise proteins encoded by L5, D29 and Rv2657c (ϕRv2) (Fig. 6A), one of which (L5 gp36) is known to bind specifically to L5 *attP* DNA. If Rv1584c is likewise a DNA-binding protein, it is not obvious where it could bind given the presumptively small size of the *attP* site of ϕRv1.

Because the ϕRv1 element integrates into a repetitive element, a comparison of the putative *attB* regions provides some insights into the possible sequence requirements for *attB* function. As seen in Fig. 3B, all of the seven repeats contain a segment corresponding to the common core, with only site 4 containing sequence departures from the 12 bp segment in ϕRv1 *attP*. The most preferred sites for integration are 6, 7 and 1 (in *Rv1587c*, *Rv3466* and *Rv0094c* respectively), which share strong sequence similarity. However, site 1 has several base differences immediately to the left of the 12 bp core (Fig. 3B) suggesting that this 4 bp region may be less important to *attB* function. The infrequently used sites (3, 2 and 5) all have a number of sequence departures in the region 5–12 bp to both the left and right of the core, implicating these regions as ones important for *attB* identity. None of the seven sites has any obvious direct or inverted repeats around the common core that might suggest possible protein-binding sites. We note however, that site 1 contains a copy of the 8 bp inverted repeat identified in *attP* immediately to the left of the core, giving it the appearance of an *attR* attachment junction.

Integration-proficient vectors are useful genetic tools for mycobacterial genetics as they enable the simple construction of recombinant strains in which a single copy of a plasmid can be introduced at a defined chromosomal location. Moreover, transformation with these vectors is efficient and the transformants have good genetic stability if the excise gene is absent; the stability can be further enhanced by removal of the integrase gene once it has performed its task in integration (Lee *et al*., 1991; Lewis and Hatfull, 2000). Vectors based on the ϕRv1 recombination system add a further dimension to these tools, as there are multiple *attB* sites available in BCG and *M. tuberculosis*. Thus, ϕRv1-based vectors can be used to insert multiple copies of recombinant genes in different chromosomal locations, or to transform a strain that already has one of the attachment sites occupied.

The finding that the ϕRv1 element encodes a fully functional site-specific recombination system – taken together with the observation that the other open reading frames in ϕRv1 are not obviously pseudogenes – argues strongly against the notion that this is simply a defective prophage. However, this then raises a number of questions as to what these prophage-like elements are doing in the *M. tuberculosis* genome. For example, do they make infectious phage-like particles and, if so, do these have tails?

Both ϕRv1 and ϕRv2 are notably devoid of recognizable tail genes, but it is possible that these are encoded elsewhere in the genome (there is sufficient diversity among phage tail genes that these may not be identifiable through searches for sequence similarity). If particles are generated, do these contain just ϕRv1 DNA or are they involved in generalized transduction like the genetic transfer elements of *Rhodococcus*, *Methanococcus* and *Serpulina* (Humphrey *et al*., 1997; Bertani, 1999; Lang and Beatty, 2000)? Finally, does either ϕRv1 or ϕRv2 encode genes that confer selective advantages to their hosts? While this is an attractive idea, there are rather few genes (14) present within the ϕRv1 element, and six of these can be assigned roles in phage virion assembly or recombination. All of the eight genes of unknown function are small (fewer than 170 residues), and two are pseudogenes in the CDC1551 strain of *M. tuberculosis.* It remains to be determined as to whether any of these contribute to the phenotypical characteristics of *M. tuberculosis* including its pathogenicity.

Experimental procedures

Bacterial strains and growth conditions

Mycobacterium smegmatis high-efficiency transformation strain mc^2155 has been described previously (Snapper *et al*., 1990); *Mycobacterium bovis* BCG Connaught and BCG Pasteur were laboratory stocks originally obtained from Dr William R. Jacobs Jr. and MedImmune Inc. *M. smegmatis* and *M. bovis* BCG were grown in Difco 7H9 liquid medium supplemented with 0.5% glycerol, 0.5% Tween 80 and 10% albumin–dextrose complex (ADC) for *M. smegmatis* and either 10% ADC or OADC for *M. bovis* BCG. Carbenicillin and cycloheximide were added to all cultures at concentrations of 50 μg ml^{-1} and 10 μg ml^{-1} respectively. When necessary, the following antibiotics were also added; kanamycin (8 μg ml^{-1}), hygromycin (50 μg ml^{-1}) and tetracycline (0.5 μg ml^{-1}). *M. smegmatis* was grown on Difco 7H10 agar and *Mycobacterium bovis* BCG was cultivated on Difco 7H11 agar; both were supplemented with 0.5% glycerol and 10% ADC. *E. coli* DH5α (Sambrook *et al*., 1989) was grown in Difco Luria broth (LB) or LB agar, supplemented with kanamycin (8 μg ml^{-1}), hygromycin (50 μg ml^{-1}) and tetracycline (0.5 μg ml^{-1}) when needed.

Plasmids and DNA

Plasmids pMH94 (Lee *et al*., 1991), pMD30 (Donnelly Wu *et al*., 1993), pMD169 (Donnelly Wu *et al*., 1993), pYUB415 (Peña *et al*., 1997) and pUC119 (Sambrook *et al*., 1989) have been described previously. Other plasmid vectors used were as follows: pGS67, which is a derivative of pUC119 containing a hygromycin resistance cassette and a DNA fragment carrying L5 *attP* and *integrase*; pJL37 is a kanamycin-resistant mycobacterial–*E. coli* shuttle plasmid derived from pMV261 (Stover *et al*., 1991) that contains an *Nde*I site at the hsp60 translation start site downstream of the *M. bovis* BCG hsp60 promoter; pJL32 which is a mycobacterial–*E. coli* shuttle vector was derived from pYUB53 (a kind gift from Dr William R. Jacobs Jr.) by removal of a *Pst*I fragment carrying kanamycin resistance; pMD02 is a pUC119 derivative with the kanamycin resistance cassette from Tn5 inserted.

The ϕRv1-based integration-proficient vector pLB17 was constructed as follows. First, the ϕRv1 integrase gene was PCR amplified from *M. tuberculosis* H37Rv genomic DNA using primers Rv1586c-F and Rv1586c-R (Table 5) and Pfu polymerase (Stratagene). The amplified product was digested with *Nde*I and *Bam*HI following the manufacturer's directions (New England Biolabs), and ligated with T4 ligase (New England Biolabs) into pET21a (Novagen) digested with the same enzymes. The resulting plasmid, pLB14, was sequenced to verify that the construction was correct. Next, DNA fragments containing the *attL* and *attR* sites of ϕRv1 were synthesized by PCR from *M. tuberculosis* H37Rv template DNA using attL-F2, attL-R, and attR-F, attR-P primer pairs respectively (Table 5). These were then used as templates in a second PCR reaction using attL-R and attR-F primers to create an *attP*-containing DNA fragment. This *attP* fragment was cut with *Acc*I and *Sac*I and joined in a three-way ligation with an *Acc*I–*Sal*I 1396 bp fragment from pLB14 containing the integrase gene and a *Sal*I–SacI vector fragment derived from pMH94. The resulting plasmid is pLB17.

A hygromycin-resistant derivative of pLB17 (i.e. pLB25) was constructed as follows. First the kanamycin resistance gene in pLB17 was removed by digestion with *Hin*dIII and religation to give plasmid pLB24. This was then digested with *Pst*I, dephosphorylated with alkaline phosphatase and ligated with a 1657 bp *Pst*I fragment containing a hygromycin

Table 5. Primers used in this study.

Name	5′ Sequence 3′
Rv1586c-F	CACGATGTTGTGGATCCGGCTC
Rv1586c-R	CCAACCGTGGACATATGAGATACAC
attL-R	AGCCGAACGAGCTCTTCCC
attL-F2	GGTTGGCCGTGGACTGCTG
attR-F	GCGCTGAATTCGTTGTCGAGG
attR-P	CCACGGCCAACCGTGGACCTG
attL-F	GCTCGCGAATTCGGGCGGG
attR-R	AGCGGGAGGTACCCCCAGG
Rv1584c-F	GTGCCGGATCCTCCGTGG
Rv1584c-R	AAGGAGTCGCATATGTCGACC
1	ACGTCGATGTGGGGATGTCC
2	CTCTTCCCTCACCTCCAAGG
4	CCAGCGGATCATGCTGTTCG
11	GCGTGGTGGTTAAAGCTCC
12	AGCGGTCGGACTACTCAGC
13	TAGCGAAGAAATCAAGTCCG
14	CAGAGGTTGCGCCTCTCG
15C	AGGCAGCTCGGCAACGTGC
17	CGTGTGGCTTTGACTGCTGG
19	GCAGCGAATCATGTTGTACG
20	ACAGCGAATCGTCCTCTACG
24	CCTGCTTCGCATGGGTCTCG
25	CCGTCTACGACGAACACACC
3′	CGGTTGATGCGCTGACGTCG
4′	ATGGTTGATCTCCTGGCGTGG
1C	GCGGCGATTCTCACGGATCG
4C	GCGGATCATGCTGTTCGCC
2C	TAGACAGCAGCACGCACAGG
3C	CGCCCGGATCGTCTCGGCC

resistance cassette. Plasmid pLB18 was constructed by amplification of the *M. bovis* BCG *attB* site 6 using primers attL-F and attR-R, *M. bovis* BCG Pasteur DNA and Pfu polymerase, followed by digestion of the product with Asp-718 and *Eco*RI and the resulting 476 bp fragment ligated with the 6912 bp Asp-718 and *Eco*RI fragment of pGS67. Finally, plasmid pLB36 was constructed by inserting a fragment containing *Rv1584c*, obtained by PCR amplification of H37Rv DNA with primers Rv1584c-F and Rv1584c-R (Table 5), under the control of an L5 P_{left} mutant promoter (11, unpublished) inserted into pJL32, which contains a tetracycline resistance cassette and *OriM*.

Transformation and electroporation

Plasmids were introduced into *E. coli* DH5α by standard heat shock transformation methods of $CaCl_2$-competent cells (Sambrook *et al*., 1989). Electrocompetent *M. smegmatis* and *M. bovis* BCG cells were prepared by growing a bacterial culture to mid-log phase (OD_{600} 0.8–1.0), harvesting and washing them three times in 10% glycerol and resuspending them in the same buffer. Plasmids were introduced by electroporation of 200 μl of electrocompetent cells with ~100 ng of DNA into using a Bio-Rad Gene Pulser II set at 1000 Ω, 2.5 kV, 25 μF. Cells were recovered in 1 ml of 7H9 broth with ADC and 0.05% Tween 80 at 37°C for 2 h (for *M. smegmatis*) or 4–12 h (for BCG).

PCR assays

Site integration assays were performed on genomic DNA obtained from transformant colonies as follows. Genomic DNA was prepared by picking a colony into 100 μl of TE (10 mM Tris, pH8; 1 mM EDTA) followed by vortexing and heating to 95°C for 10 min before using 1 μl for PCR using the Perkin Elmer 9600 or 2700 thermal cycler. For *M. bovis* BCG REP13E12 site usage assays, each of seven sites was analysed for integration using two methods. The first used two sets of two primers, one pair for amplification of *attB*, the other for *attL*, i.e. a total of 14 reactions for each transformant. For amplification of *attB*, primers 11 and 4 were used for site 1, yielding a 405 bp product, primers 12 and 19 for site 2 (447 bp), primers 13 and 19 for site 3 (485 bp), primers 14 and 20 for site 4 (331 bp), primers 15c and 19 for site 5 (424 bp), primers 1 and 4 for site 6 (362 bp), and primers 17 and 4 for site 7 (338 bp). To amplify *attL*, the following primers were used along with primer 2, which lies within the integrated plasmid, yielding products of the size given following the primer number; site 1 primer 11 (353 bp), site 2 primer 12 (395 bp), site 3 primer 13 (433 bp), site 4 primer 14 (276 bp), site 5 primer 15c (372 bp), site 6 primer 1 (310 bp), site 7 primer 11 (289 bp). The second method used a set of three primer pairs to amplify either *attB* or *attL* for each of the seven sites. For PCR assays with *M. smegmatis* strains carrying BCG site 6 *attB (*strain LAB7), four primers were used in each PCR reaction: 1C, 2C, 3C and 4C. These four primers could amplify *attB (*primers 1C and 4C – 325 bp product), *attP* (primers2C and 3C – 200 bp product), *attL* (primers 2C and 4C – 329 bp product) and *attR* (primers 1C and 3C – 240 bp product). In *M. smegmatis*, four primers were used in each PCR reaction: 24, 25, 3′ and 4′. Primers 3′ and 4′ amplify *attB,* giving a 228 bp product. Primers 24 and 25 amplify a 452 bp *attP.* Amplification of *attL* with primers 24 and 4′ gives a 131 bp product, and primers 3′ and 25 yield a 501 bp *attR*. These *M. smegmatis* primers were designed using preliminary sequence data obtained from The Institute for Genomic Research website at http://www.tigr.org.

Sequencing

Sequencing of clones and attachment sites was done using PE Applied Biosystems dRhodamine terminator chemistry and analysed on an ABI 310 capillary sequencer.

Acknowledgements

We would like to thank Brenda Wheeler for excellent technical assistance. We also thank John Lewis for assistance in bioinformatic analyses and critical reading of the manuscript. We thank Marisa Pedulla for critical reading of the manuscript and The Institute for Genomic Research for use of unpublished *M. smegmatis* and *M. tuberculosis* genome sequences available at http://www.tigr.org. This work was supported by grant AI045683 from the National Institute of Health.

References

Bertani, G. (1999) Transduction-like gene transfer in the methanogen *Methanococcus voltae*. *J Bacteriol* **181:** 2992–3002.

Breuner, A., Brondsted, L., and Hammer, K. (1999) Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. *J Bacteriol* **181:** 7291–7297.

Breuner, A., Bronsted, L., and Hammer, K. (2001) Resolvase-like recombination performed by the TP901-1 integrase. *Microbiology* **147:** 2051–2063.

Brosch, R., Gordon, S.V., Eiglmeier, K., Garnier, T., Tekaia, F., Yeramian, E., and Cole, S.T. (2000) Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex. In: *Molecular Genetics of the Mycobacteria.* Hatfull, G.F., and Jacobs Jr, W.R. (eds). Washington, DC: American Society for Microbiology Press, pp. 19–36.

Christiansen, B., Brondsted, L., Vogensen, F.K., and Hammer, K. (1996) A resolvase-like protein is required for the site-specific integration of the temperate lactococcal bacteriophage TP901-1. *J Bacteriol* **178:** 5164–5173.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.* (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393:** 537–544.

Donnelly-Wu, M.K., Jacobs, W.R. Jr and Hatfull, G.F. (1993) Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Mol Microbiol* **7:** 407–417.

Dye, C., Scheele, S., Dolin, P., Pathania, V., and Raviglione, R.C. (1999) Global burden of tuberculosis – estimated incidence, prevalence, and mortality by country. *J Am Med Assoc* **282:** 677–686.

Freitas-Vieira, A., Anes, E., and Moniz-Pereira, J. (1998) The

site-specific recombination locus of mycobacteriophage Ms6 determines DNA integration at the tRNA (Ala) gene of *Mycobacterium* spp. *Microbiology* **144:** 3397–3406.

Groth, A.C., Olivares, E.C., Thyagarajan, B., and Calos, M.P. (2000) A phage integrase directs efficient site-specific integration in human cells. *Proc Natl Acad Sci USA* **97:** 5995–6000.

Hatfull, G.F. (2000) Molecular genetics of mycobacteriophages. In: *Molecular Genetics of the Mycobacteria. .* Hatfull, G.F., and Jacobs Jr, W.R. (eds). Washington, DC: American Society for Microbiology Press, pp. 37–54.

Hatfull, G.F., and Grindley, N.D.F. (1988) Resolvases and DNA-invertases: a family of enzymes active in site-specific recombination. In: *Genetic Recombination.* Kucherlapati, R., and Smith, G.R. (eds). Washington DC: American Society for Microbiology Press, pp. 358–396.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., and Hatfull, G.F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci USA* **96:** 2192–2197.

Hsu, P.L., Ross, W., and Landy, A. (1980) The lambda phage att site: functional limits and interaction with Int protein. *Nature* **285:** 85–91.

Humphrey, S.B., Stanton, T.B., Jensen, N.S., and Zuerner, R.L. (1997) Purification and characterization of VSH-1, a generalized transducing bacteriophage of *Serpulina hyodysenteriae*. *J Bacteriol* **179:** 323–329.

Kuhstoss, S., and Rao, R.N. (1991) Analysis of the integration function of the streptomycete bacteriophage phi C31. *J Mol Biol* **222:** 897–908.

Lang, A.S., and Beatty, J.T. (2000) Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *Rhodobacter capsulatus*. *Proc Natl Acad Sci USA* **97:** 859–864.

Lazarevic, V., Dusterhoft, A., Soldo, B., Hilbert, H., Mauel, C., and Karamata, D. (1999) Nucleotide sequence of the *Bacillus subtilis* temperate bacteriophage SPbetac2. *Microbiology* **145:** 1055–1067.

Lee, M.H., Pascopella, L., Jacobs, W.R. Jr and Hatfull, G.F. (1991) Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette–Guerin. *Proc Natl Acad Sci USA* **88:** 3111–3115.

Lewis, J.A., and Hatfull, G.F. (2000) Identification and characterization of mycobacteriophage L5 excisionase. *Mol Microbiol* **35:** 350–360.

Lewis, J.A., and Hatfull, G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* **29:** 2205–2216.

Loessner, M.J., Inman, R.B., Lauer, P., and Calendar, R. (2000) Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. *Mol Microbiol* **35:** 324–340.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., and Stover, C.K. (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* **178:** 1274–1282.

Matsuura, M., Noguchi, T., Yamaguchi, D., Aida, T., Asayama, M., Takahashi, H., and Shirai, M. (1996) The sre gene (ORF469) encodes a site-specific recombinase responsible for integration of the R4 phage genome. *J Bacteriol* **178:** 3374–3376.

Mediavilla, J., Jain, S., Kriakov, J., Ford, M.E., Jacobs, W.R.J., Hendrix, R.W., and Hatfull, G.F. (2000) Genome organization and characterization of mycobacteriophage Bxb1. *Mol Microbiol* **38:** 955–970.

Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., Foeller, C., *et al.* (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.

Olivares, E.C., Hollis, R.P., and Calos, M.P. (2001) Phage R4 integrase mediates site-specific integration in human cells. *Gene* **278:** 167–176.

Peña, C.E., Lee, M.H., Pedulla, M.L., and Hatfull, G.F. (1997) Characterization of the mycobacteriophage L5 attachment site, attP. *J Mol Biol* **266:** 76–92.

Peña, C.E., Stoner, J., and Hatfull, G.F. (1998) Mycobacteriophage D29 integrase-mediated recombination: specificity of mycobacteriophage integration. *Gene* **225:** 143–151.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R. Jr (1990) Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4:** 1911–1919.

Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., *et al.* (1991) New use of BCG for recombinant vaccines. *Nature* **351:** 456–460.

Thorpe, H.M., and Smith, M.C. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci USA* **95:** 5505–5510.

Thorpe, H.M., Wilson, S.E., and Smith, M.C. (2000) Control of directionality in the site-specific recombination system of the streptomyces phage phiC31. *Mol Microbiol* **38:** 232–241.

Yang, H.Y., Kim, Y.W., and Chang, H.I. (2002) Construction of an integration-proficient vector based on the site-specific recombination mechanism of enterococcal temperate phage phiFC1. *J Bacteriol* **184:** 1859–1864.