

**SITE-SPECIFIC RECOMBINATION OF THE MYCOBACTERIUM TUBERCULOSIS
PROPHAGE-LIKE ELEMENT PHIRV1**

by

Lori Ann Bibb

B.S., Elizabethtown College, 1997

Submitted to the Graduate Faculty of
Arts and Sciences in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH
FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Lori Ann Bibb

It was defended on

March 1, 2004

and approved by

Karen Arndt

Joanne Flynn

Roger Hendrix

Craig Peebles

Graham F. Hatfull
Dissertation Director

SITE-SPECIFIC RECOMBINATION OF THE MYCOBACTERIUM TUBERCULOSIS PROPHAGE-LIKE ELEMENT PHIRV1

Lori Ann Bibb, Ph.D.

University of Pittsburgh, 2004

The site-specific recombination systems of bacteriophages and other mobile elements fall into two categories that are named for the integrase protein that catalyzes integration and excision of the phage genome. These integrases utilize either a tyrosine or a serine residue to carry out the nucleophilic attack of the DNA. In many cases the directionality of the reaction, that is whether the enzyme catalyzes integration or excision, is determined by an additional phage encoded protein referred to as RDF, or recombination directionality factor.

Two prophage-like elements, $\phi Rv1$ and $\phi Rv2$, were found through sequencing *Mycobacterium tuberculosis* strain H37Rv. These are absent from the vaccine bacillus *M. bovis* BCG, and are often found in virulent strains of *M. bovis* and *M. tuberculosis*. Through the work presented here, one of these elements, $\phi Rv1$, was found to encode an active recombination system, with a serine integrase and RDF. The $\phi Rv1$ element is found within a degenerate repeated element, REP13E12, which is present in seven non-identical copies in *M. tuberculosis* and *M. bovis* BCG. In vivo studies have revealed that four of the seven 13E12 elements can serve as integration sites (*attBs*) for a plasmid carrying a reconstructed *attP* and integrase, and that multiple integrations can occur. The fast growing saprophyte, *M. smegmatis*, also supports integration, although inefficiently. In *M. smegmatis*, the $\phi Rv1$ plasmid integrates into at least two 13E12 repeats that are quite different from those found in *M. bovis* BCG and *M. tuberculosis*. Inefficiency is overcome by providing *M. smegmatis* with an *attB* site from BCG.

These integrated plasmids are stable, and excision occurs in the presence of the ϕ Rv1 RDF encoded by *Rv1584c*. In vitro assays were developed for both integration and excision. All the substrate site requirements are relatively small. Integration occurs slowly but efficiently in the presence of excess *attB* or on an intramolecular *attP-attB* plasmid. In the presence of RDF integration is inhibited and excision is stimulated. The RDF binds to a specific sequence in both *attB* and *attL*, although the way in which it functions may be through protein-protein interactions with integrase.

TABLE OF CONTENTS

PREFACE	xii
I. INTRODUCTION	1
I. A. Recombination.....	1
I. B. Bacteriophages.....	2
I. C. Site-specific recombination.....	7
I. C. i. Tyrosine site-specific recombination systems.....	10
I. C. i. a. Simple tyrosine recombinases.....	16
I. C. i. b. Tyrosine integrases.....	20
I. C. i. c. Tyrosine recombinases and the control of directionality.....	24
I. C. ii. Serine recombinases.....	26
I. C. ii. a. Small serine recombinases.....	33
I. C. ii. b. Large serine recombinases.....	37
I. C. ii. c. Control of directionality in the serine recombinases.....	40
I. D. Mycobacteria.....	42
I. E. Prophages and prophage-like elements Rv1 and Rv2.....	42
I. F. Specific aims.....	48
I. F. i. To determine if Rv1 encodes a functional integrase.....	48
I. F. ii. To determine which sites serve as <i>attBs</i> in <i>M. bovis</i> BCG and <i>M. smegmatis</i>	48
I. F. iii. To establish an <i>in vitro</i> integration assay and determine the reaction requirements.....	48
I. F. iv. To determine the basis for directionality in Rv1.....	49
I. F. v. To determine how the RDF controls directionality of the recombination reactions.....	49
II. MATERIALS AND METHODS	50
II. A. Bacterial strains and growth conditions.....	50
II. B. Plasmids and DNA.....	51

II. C. Transformation and electroporation.....	58
II. D. PCR assays.....	58
II. E. Sequencing.....	60
II. F. Expression and purification of integrase.....	60
II. G. <i>In vitro</i> integration assays.....	61
II. H. Expression and purification of RDF.....	61
II. I. <i>In vitro</i> excision reactions.....	62
II. J. Radiolabeling DNA.....	62
II. K. DNA binding reactions.....	63
II. L. Genomic DNA preparation.....	63
II. M. DNaseI footprinting.....	64

III. INTEGRATION OF ϕ Rv1 IN MYCOBACTERIA..... 65

III. A. Introduction.....	65
III. B. <i>M. tuberculosis</i> ϕ Rv1 encodes an active integration system.....	67
III. C. ϕ Rv1 <i>attP-integrase</i> plasmids utilize multiple BCG <i>attB</i> sites.....	71
III. D. pLB17 integration in <i>M. smegmatis</i>	80
III. E. Efficient transformation of <i>M. smegmatis</i> by pLB17 requires a BCG <i>attB</i> site.....	82
III. F. ϕ Rv1 recombination crossover region.....	85
III. G. Stable maintenance of plasmid pLB17 in <i>M. smegmatis</i>	88
III. H. Discussion.....	88

IV. REQUIREMENTS FOR ϕ Rv1 INTEGRATION..... 94

IV. A. Introduction.....	94
IV. B. Expression and purification of ϕ Rv1 integrase (gpRv1586c).....	96
IV. C. <i>In vitro</i> integration assay establishment.....	96
IV. D. <i>attP</i> and <i>attB</i> site requirements.....	101
IV. E. <i>In vitro</i> efficiency of <i>M. bovis</i> BCG REP13E12 <i>attB</i> sites.....	111
IV. F. <i>In vitro</i> efficiency of <i>M. smegmatis</i> REP13E12 <i>attB</i> sites.....	122
IV. G. Activity of symmetrized <i>attB</i> sites.....	125
IV. H. Activity of <i>attB</i> mutants with changes on the left side.....	129
IV. I. Activity of <i>attB</i> core sequence mutants.....	130
IV. J. Role of supercoiling in ϕ Rv1 integrative recombination.....	130
IV. K. Additional factors in integration.....	137

IV. L. Discussion.....	140
------------------------	-----

V. CONTROL OF ϕRv1 RECOMBINATION DIRECTIONALITY.....	152
---	------------

V. A. Introduction.....	152
V. B. Rv1584c induces excision of an integrated <i>attP-integrase</i> plasmid.....	153
V. C. Rv1584c inhibits ϕ Rv1 integrative recombination <i>in vitro</i>	160
V. D. ϕ Rv1 RDF stimulates excisive recombination <i>in vitro</i>	163
V. E. Expression and purification of ϕ Rv1 RDF.....	163
V. F. Role of supercoiling in ϕ Rv1 excisive recombination.....	168
V. G. Intermolecular excision.....	168
V. H. Discussion.....	175

VI. MECHANISM OF ϕRv1 INTEGRATION AND EXCISION.....	182
--	------------

VI. A. Introduction.....	182
VI. B. ϕ Rv1 integrase binding to substrate sites.....	186
VI. C. RDF binding to substrate sites.....	189
VI. D. Integrase and RDF interactions at <i>attB</i>	192
VI. E. Integrase and RDF interactions at <i>attL</i>	198
VI. F. Integrase and RDF interactions at <i>attP</i>	205
VI. G. Integrase and RDF interactions at <i>attR</i>	211
VI. H. The role of the RDF binding site in <i>attB</i> and <i>attL</i> complex formation.....	217
VI. I. Role of the RDF binding region in integration and excision.....	224
VI. J. Integration complexes.....	231
VI. K. Excision complexes.....	240
VI. L. Discussion.....	241

VII. CONCLUSION.....	250
-----------------------------	------------

VII. A. ϕ Rv1 integration-putting the pieces together.....	250
VII. B. Control of ϕ Rv1 recombination directionality.....	252
VII. C. Role of ϕ Rv1 and ϕ Rv2?.....	254
VII. D. Future considerations.....	256

VII. D. i. Dissecting ϕ Rv1 recombination.....	256
VII. D. ii. Utility of ϕ Rv1 recombination.....	257
BIBLIOGRAPHY	261

LIST OF TABLES

Table 1. Transformation efficiency of <i>M. bovis</i> BCG Connaught and BCG Pasteur by extrachromosomal and integration proficient plasmids.....	72
Table 2. Transformation efficiency of <i>Mycobacterium smegmatis</i> mc ² 155 by extrachromosomal and integration proficient plasmids.....	79
Table 3. Transformation efficiency of <i>Mycobacterium smegmatis</i> strains by extrachromosomal and ϕ Rv1 integration proficient plasmids.....	81
Table 4. Sequence differences in ϕ Rv1 between <i>M. tuberculosis</i> strains H37Rv and CDC1551.....	92
Table 5. Sequence differences in ϕ Rv1 between <i>M. tuberculosis</i> H37Rv and <i>M. bovis</i>	92

LIST OF FIGURES

Figure 1. Life styles of a temperate bacteriophage.....	3
Figure 2. Integration and excision of a bacteriophage genome.....	5
Figure 3. Types of site-specific recombination reactions.....	8
Figure 4. Tyrosine recombinase domain organization.....	11
Figure 5. Mechanism and chemistry of tyrosine recombinases.....	13
Figure 6. Tyrosine recombinase substrate sites.....	18
Figure 7. Nucleoprotein complex models of tyrosine recombinases.....	21
Figure 8. Structural organization of serine-integrases.....	27
Figure 9. Mechanism and chemistry of serine recombinases.....	30
Figure 10. Serine recombinase substrate sites.....	34
Figure 11. Serine recombinase-DNA complexes.....	38
Figure 12. <i>M. tuberculosis</i> H37Rv prophage-like elements ϕ Rv1 and ϕ Rv2.....	44
Figure 13. Construction of ϕ Rv1 integration-proficient vector pLB17.....	68
Figure 14. <i>M. bovis</i> BCG-Connaught contains seven REP13E12 elements.....	73
Figure 15. Characterization of ϕ Rv1-mediated integration events in BCG-Connaught.....	76
Figure 16. Integration of ϕ Rv1 in <i>M. smegmatis</i>	83
Figure 17. Integration of ϕ Rv1 in <i>M. smegmatis</i> with BCG #6.....	86
Figure 18. Mapping the crossover site for ϕ Rv1 integration.....	89
Figure 19. Expression and purification of ϕ Rv1 integrase.....	97
Figure 20. <i>In vitro</i> integration assays.....	99
Figure 21. A molar excess of <i>attB</i> is required for efficient intermolecular integration.....	102
Figure 22. Time course <i>in vitro</i> integration reaction.....	104
Figure 23. Integrase titration integration assay.....	106
Figure 24. Minimum size of <i>attB</i> in integration.....	108
Figure 25. Minimal <i>attP</i> determination.....	112
Figure 26. Minimal ϕ Rv1 attachment and junction sites.....	115
Figure 27. <i>In vitro</i> activity of <i>attB</i> sites from <i>M. bovis</i> BCG.....	118
Figure 28. <i>In vitro</i> activity of <i>attB</i> sites from <i>M. smegmatis</i>	123
Figure 29. <i>In vitro</i> activity of synthetic <i>attB</i> sites.....	126
Figure 30. <i>In vitro</i> integration with left side mutants of <i>attB</i>	131
Figure 31. <i>In vitro</i> activity of <i>attB</i> core mutants.....	134
Figure 32. Role of supercoiling in integration <i>in vitro</i>	138

Figure 33. Additional factors in ϕ Rv1 integration.....	141
Figure 34. Intramolecular integration.....	145
Figure 35. <i>attB</i> sites.....	148
Figure 36. Alignment of Rv1584c with known and predicted recombination directionality factors.....	154
Figure 37. Rv1584c-mediated excision in mycobacteria.....	157
Figure 38. ϕ Rv1 RDF inhibits integrative recombination.....	161
Figure 39. Construction of excision substrate and excision reactions.....	164
Figure 40. Rv1584c stimulates excisive recombination in an intramolecular substrate.....	166
Figure 41. ϕ Rv1 RDF expression and purification.....	169
Figure 42. Purified RDF inhibits integration and promotes excision.....	171
Figure 43. Role of supercoiling in excisive recombination.....	173
Figure 44. Intermolecular excision.....	176
Figure 45. Attachment sites of ϕ Rv1.....	183
Figure 46. Integrase binding to attachment sites.....	187
Figure 47. RDF binding to attachment sites.....	190
Figure 48. Integrase and RDF binding to <i>attB</i>	193
Figure 49. DnaseI footprinting of integrase and RDF at <i>attB</i>	195
Figure 50. Integrase and RDF binding to <i>attL</i>	199
Figure 51. DnaseI footprinting of integrase and RDF at <i>attL</i>	202
Figure 52. Integrase and RDF binding to <i>attP</i>	206
Figure 53. DnaseI footprinting of integrase and RDF at <i>attP</i>	208
Figure 54. Integrase and RDF binding to <i>attR</i>	212
Figure 55. DNaseI footprinting of integrase and RDF at <i>attR</i>	214
Figure 56. Complex formation using substrate sites without the RDF binding site.....	218
Figure 57. Role of the RDF binding site in inhibition of integration.....	226
Figure 58. Excision with and without the RDF binding site.....	228
Figure 59. Integration complexes.....	232
Figure 60. Excision complexes.....	242
Figure 61. Models of integration and excision.....	248
Figure 62. REP13E12 sequences in other mycobacteria.....	259

PREFACE

Throughout my graduate career, many people have shaped to my development and enriched my experiences. Most notable of these is my advisor, Graham Hatfull, who has guided my research, provided a wealth of enthusiasm, and possesses infinite patience and a great sense of humor.

The members of the Hatfull laboratory have made an enjoyable work environment. Former graduate students Dr. John Lewis and Dr. Carol Pena, who were more senior than myself, shared their expertise, and acted as mentors. More recently, Laura Marinelli has been a friend and sounding board for research problems. She and other members of the department often supplied stress relief and distractions from my research (sometimes with other science).

My boyfriend, Benjamin Keating, facilitated the completion of this work by keeping me company when I had to stay late at the lab, and through his encouragement. Finally, I would like to thank my family for supporting all my endeavors.

I. INTRODUCTION

I. A. Recombination

Recombination plays an important role in countless biological processes. Various types of DNA rearrangements generate the antibody diversity that allows us to fight a myriad of microbes, repair mistakes in DNA, and resolve dimers created through transposition and circular chromosome replication. In other examples, recombination controls gene expression and developmental processes and allows bacteriophage and other mobile elements to move in and out of their host genome. Without recombination, the ecosphere would be lacking for diversity because genes could no longer be inserted, deleted, and exchanged.

There are several ways that DNA can be recombined, and most recombination mechanisms fall into one of three categories; homologous recombination, transposition, and site-specific recombination (Hallet and Sherratt, 1997). Homologous recombination occurs between two stretches of DNA with extensive similarity. Homologous or general recombination is part of DNA repair pathways and is accepted as being responsible for genomic rearrangements in many different organisms, and this type of recombination requires several protein complexes (Bishop and Schiestl, 2000; Eggleston and West, 1996; Sonoda *et al.*, 2001). Transposition is the movement of specific DNA segments into non-homologous loci. Transposable elements

themselves fall into several different categories, and transposons and insertion sequences are prevalent in many organisms; in fact ~40% of the human genome consists of transposons (Curcio and Derbyshire, 2003; Jurka *et al.*, 2004). The third type of recombination, site-specific recombination, recombines two specific target sites with relatively short stretches of homology. The recombination machinery consists of one or two enzymes and in some cases one or more additional factor (Hallet and Sherratt, 1997).

In addition to the importance of recombination in the living world, several of these systems have been adapted for use in the laboratory. By utilizing these recombination systems, mutagenesis, cloning and other techniques have become easier. Notably, many of these useful molecular systems are derived from bacteriophage.

I. B. Bacteriophages

Viruses that infect bacteria, known as bacteriophages, are estimated to be the most numerous life form on the planet at 10^{31} particles (Hendrix *et al.*, 1999). Thus, they can be found in nearly every environment (Frederickson *et al.*, 2003; Pedulla *et al.*, 2003; Walter and Baker, 2003) Phages are often described by their life style; that is whether they are temperate or lytic (Figure 1). Once it infects a cell by injecting its DNA, a temperate phage may adopt one of two fates; it can undergo lytic growth or lysogeny. In lytic growth, a phage utilizes the host machinery to produce progeny. A phage that only produces progeny after infecting a cell is called a lytic phage, while a temperate phage has a lytic phase as well as a lysogenic phase. During the lysogenic phase, the bacteriophage genome is stably maintained in the host cell and replicates as the host replicates. Although in lysogeny, some phages exist as extrachromosomal plasmids, there are many that reside within the host chromosome. This type of lysogen is formed

Figure 1. Life styles of a temperate bacteriophage

A temperate phage, after infecting a host bacterial cell by injecting its DNA, may undergo lytic or lysogenic growth. In lysogeny, a bacteriophage is often integrated into the host chromosome. Here, the phage genome, which is referred to as a prophage, is maintained stably until environmental cues cause it to excise (induction). At this point, the phage is said to have switched to lytic growth, where the phage uses the molecular machinery of the host to produce a number of progeny phage. The eventual lysis of the host cell will result in phage release.

Figure 1

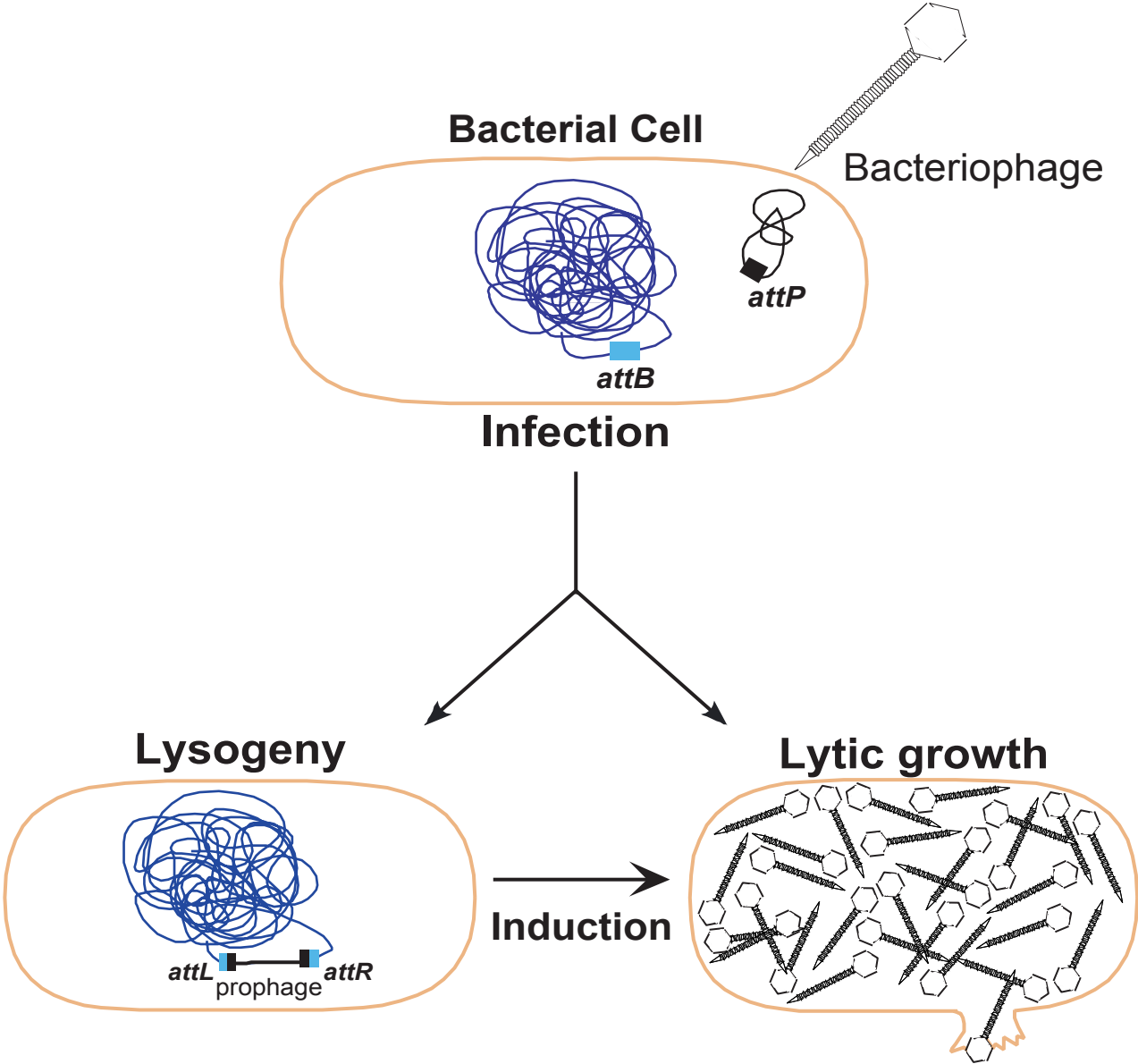
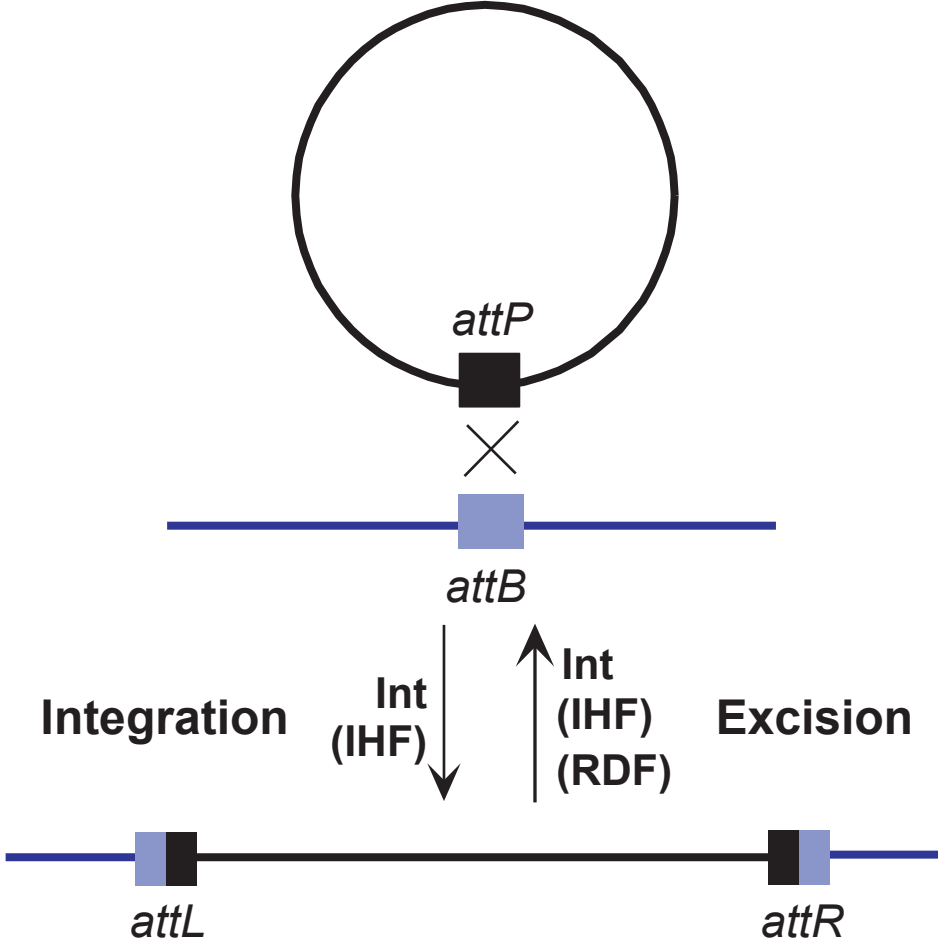


Figure 2. Integration and excision of a bacteriophage genome

This figure shows the general mechanism of bacteriophage integration and excision. In one type of lysogeny, the phage genome becomes integrated into the host chromosome by a site-specific recombination event, integration, between *attP* (attachment phage) on the phage genome and *attB* (attachment bacteria) on the bacterial chromosome. This reaction is catalyzed by a protein called integrase, but often a host derived accessory factor, such as IHF (integration host factor) is required for the reaction to occur. The bacteriophage genome in the integrated state is referred to as a prophage. It is flanked by the hybrid junctions *attL* and *attR*. The reverse reaction, excision, is essentially *attL* x *attR* recombination. Excision is also catalyzed by integrase but in all characterized systems, it requires an additional protein partner, a recombination directionality factor (RDF), which is encoded by the phage.

Figure 2



via a site-specific recombination event between a site in the phage chromosome called *attP* (*attachment phage*), and a site on the host chromosome, *attB* (*attachment bacteria*) (Figure 2). The product of this reaction, which is called integration, is the integrated phage genome, known as a prophage, flanked on the left and the right by the hybrid attachment junctions, *attL* and *attR*, which serve as the substrates for excision. Both the excisive and integrative reactions are catalyzed by an enzyme known as an integrase, and the direction of the recombination reactions must be controlled to ensure the success of the phage. Excision at an inappropriate time would fail to produce progeny phage. The integration and excision reactions are controlled by a recombination directionality factor or RDF (Lewis and Hatfull, 2001). Integrating phages utilize site-specific recombination, but this is not the only instance where site-specific recombination is used.

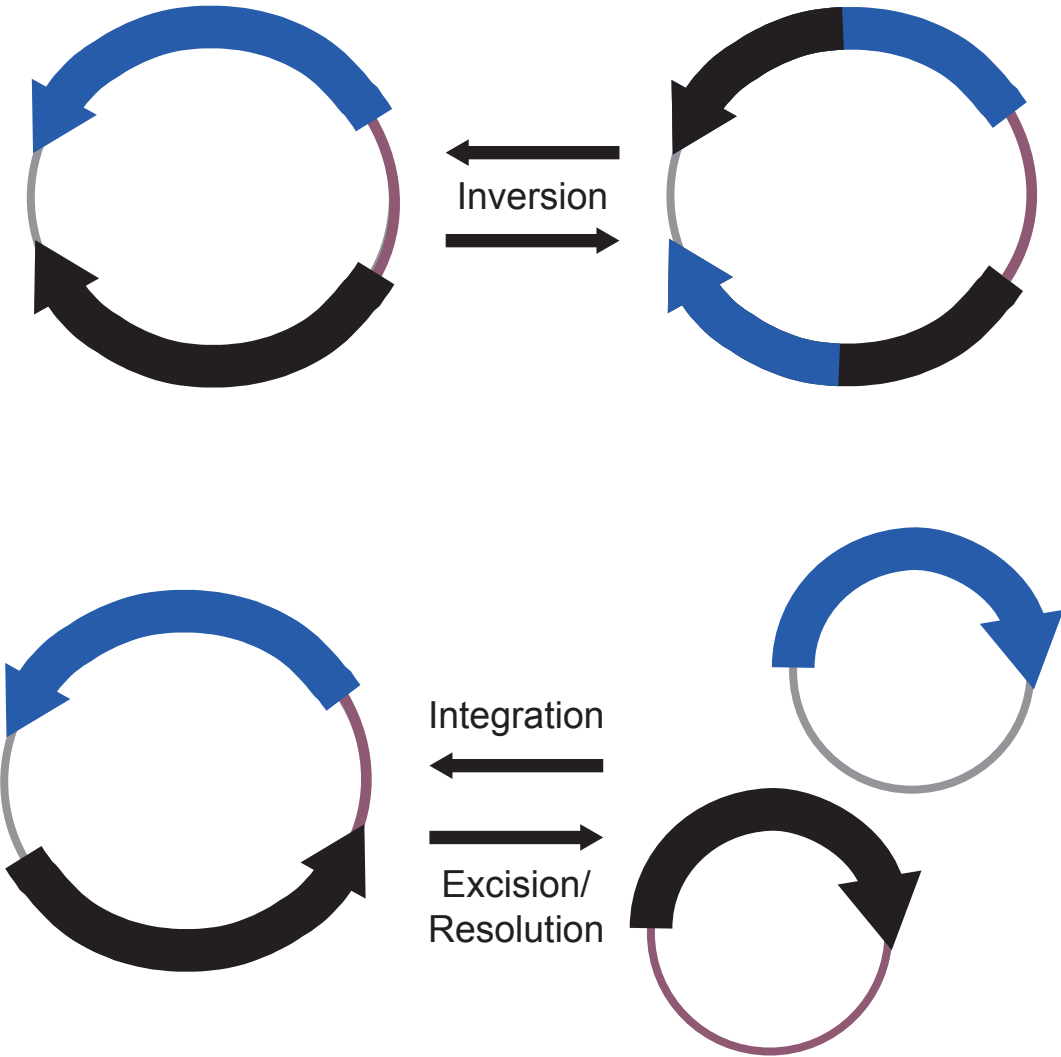
I. C. Site-Specific Recombination

The process of site-specific recombination occurs, as its name suggests, at sequence specific sites (Stark *et al.*, 1992). These reactions are conservative in that DNA is neither synthesized nor degraded, and no high energy cofactors are consumed (Stark *et al.*, 1992). One substrate DNA is cleaved at specific bonds and rejoined to the other substrate. Four cuts are made, one to each strand of the two substrates, and in all studied systems, four molecules of enzyme carry out the reaction, with one per cut (Kikuchi and Nash, 1979). The enzymes that carry out these reactions, called recombinases, catalyze one or more of the following reactions; resolution, inversion, and integration and excision (Figure 3) (Hallet and Sherratt, 1997). The type of reaction that occurs is determined by which sites are present, whether the sites are on two separate DNA molecules or on the same DNA, and the orientation of the sites

Figure 3. Types of site-specific recombination reactions

Schematics to describe the various types of site-specific recombination sites are shown. The sites are represented as black and blue arrows in the circles that represent double-stranded DNA molecules. A site-specific recombination reaction that occurs between two sites that are on separate DNA molecules is called integration. The reverse reaction, which occurs in an intramolecular substrate containing two sites in direct, or head-to-tail, configuration, removing the DNA between the crossover points within the two sites is either excision when the two sites are different or resolution when the sites are identical. Inversion reactions occur in an intramolecular substrate, where the DNA sequence between a pair of sites in indirect (head to head) configuration is flipped.

Figure 3



(Hallet and Sherratt, 1997). A recombinase is often more specifically named by the reaction or reactions that it catalyzes. Resolvases catalyze cointegrant resolution, which is an intramolecular reaction in which the DNA between a pair of directly oriented sites is removed, while invertases invert the sequence of DNA in between a pair of sites in indirect orientation (Hallet and Sherratt, 1997). Both integration and excision reactions are catalyzed by an integrase. In integration, two different DNA molecules are combined at their substrate sites. In excision, which is similar to resolution, DNA in between directly oriented sites is excised. These site-specific recombination reactions are catalyzed by enzymes that fall into two families: the tyrosine and serine recombinases, which are named for the catalytic amino acid (Hallet and Sherratt, 1997). Although the roles they fulfill are similar, and both families have members that catalyze each of the four types of reactions, the mechanism of catalysis differs, and the enzymes are unrelated at the sequence level, suggesting that they have evolved independently.

I. C. i. Tyrosine site-specific recombination systems

The tyrosine recombinases utilize a tyrosine residue to carry out the nucleophilic attack of the DNA. The enzymes of this family are diverse; there are resolvases, invertases, and integrases. Well-studied members of this family include Cre and Flp recombinases, XerCD, and bacteriophage integrases like λ , and L5. Despite their variability in action, these enzymes share a number of essential residues and have a similar domain structure (Argos *et al.*, 1986; Nunes-Duby *et al.*, 1998). The catalytic tyrosine is part of a conserved domain in the C-terminus (Figure 4) (Abremski and Hoess, 1992; Pargellis *et al.*, 1988; Tirumalai *et al.*, 1997). Other characterized domains are those responsible for DNA binding and protein-protein interactions (Figure 4) (Kazmierczak *et al.*, 2002; Moitoso de Vargas *et al.*, 1988).

Figure 4. Tyrosine recombinase domain organization

Schematic depictions of representative members of the tyrosine recombinases are shown. The tyrosine recombinases have three major domains that have been characterized. There is a C-terminal domain (shown in red) containing the tyrosine nucleophile and other residues important for catalysis. A central domain is involved in specific binding to core type sequences, or sites that flank the crossover site (gray). An N-terminal domain contains either arm-type binding determinants for the integrases, or non-specific binding for the recombinases (blue and purple, respectively). There is also a region in the extreme C-terminus that has been implicated in protein-protein interactions (yellow).

Figure 4

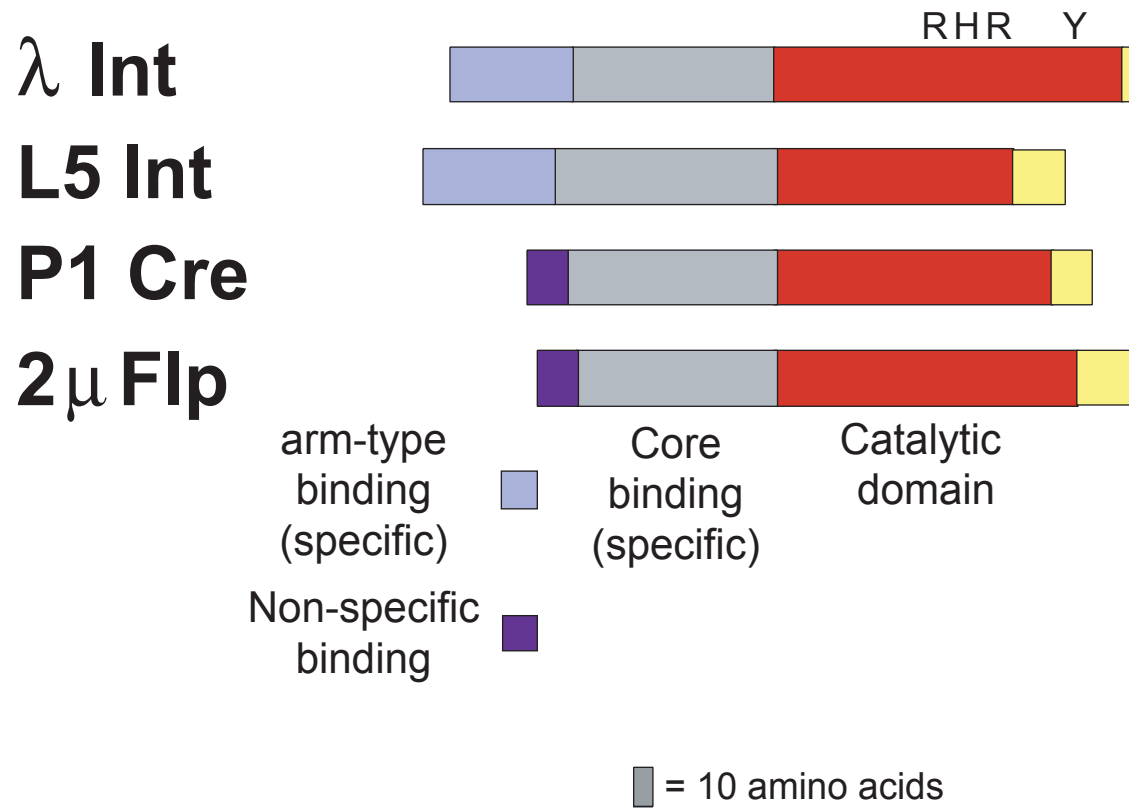


Figure 5. Mechanism and chemistry of tyrosine recombinases

A) In the mechanism of strand exchange for the tyrosine recombinases, there are four proteins (represented as ovals) that bind to the substrate sites to be recombined (shown as blue and black paired lines). The reaction consists of two sequential rounds of strand breakage and ligation. Only two molecules of recombinase are active at one time (active monomers are shown in red). These active recombinase molecules catalyze strand breakage and covalently bond to the 3' phosphate of the DNA and release a 5' OH leaving group (panels i&ii). The strand breaks on the top and bottom strands are separated by 6-8 bp. The free 5' OH then attacks the 3' phosphotyrosine of the other substrate (iii) and thus, a Holliday junction is formed (panel iv). An isomerization step follows and the other two recombinase units are activated (panels iv & v). A second round of strand breakage and exchange occurs (panels v & vi), resulting in the formation of the recombinant products (vii).

B) This figure shows the chemistry of the reaction catalyzed by the tyrosine recombinases at a single active site. Here, a single strand of DNA from one site is shown in black and a single strand from the second site is shown in blue. The OH of the catalytic tyrosine residue attacks the phosphate at the crossover site forming a 3' phosphotyrosine. The free 5' OH from the second site then attacks the phosphotyrosine, freeing the recombinase and forming one complete recombinant DNA strand.

Figure 5A

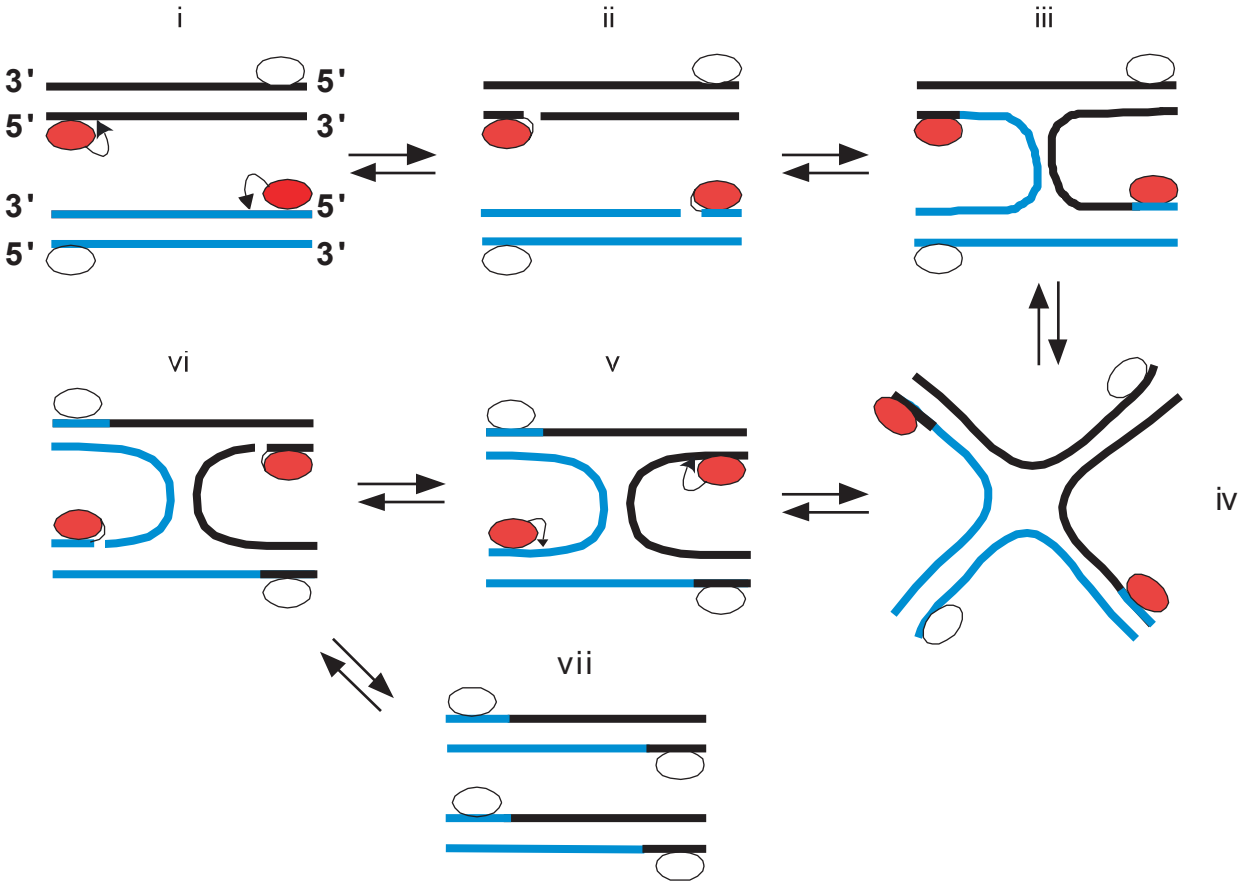
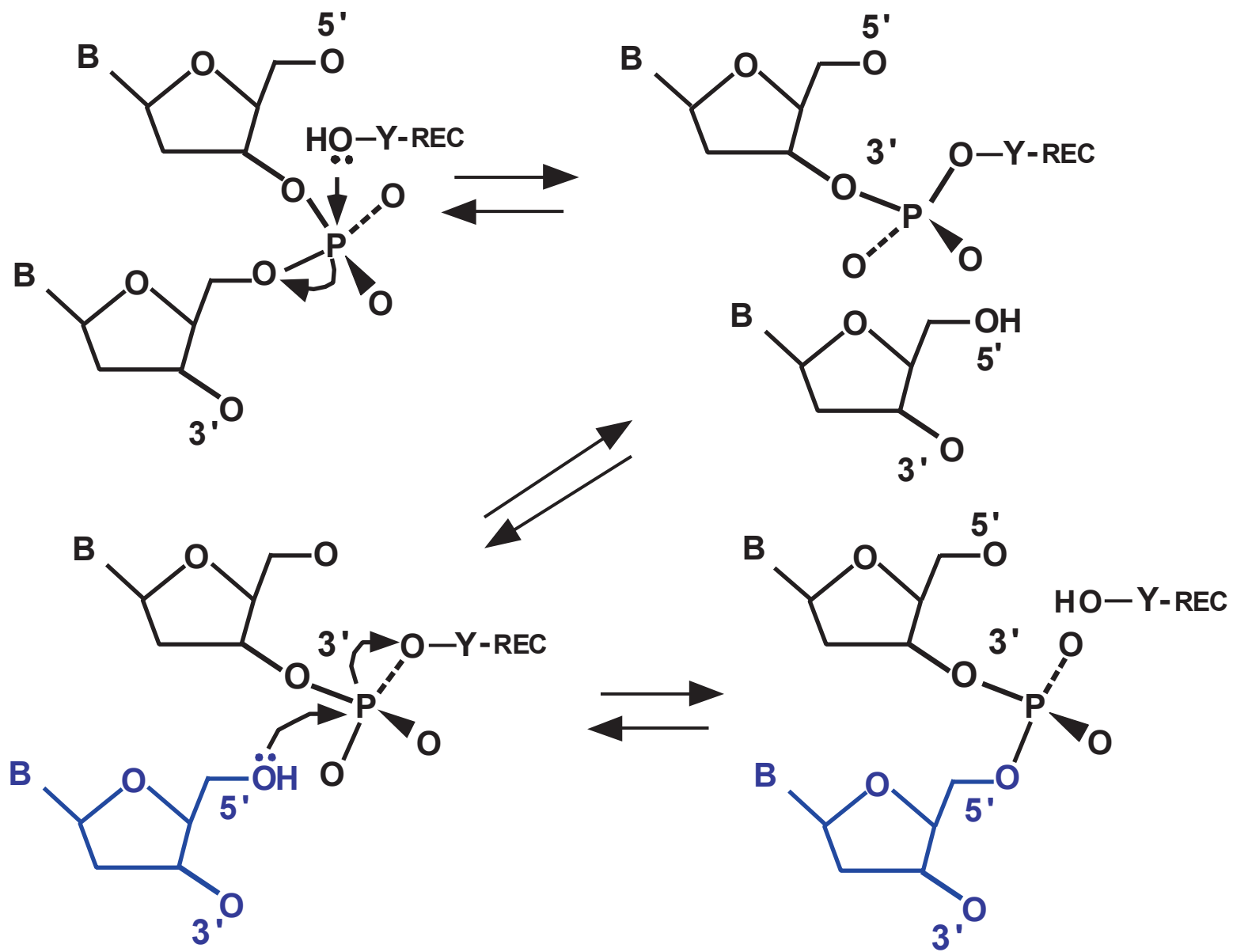


Figure 5B



The tyrosine recombinases function by carrying out two sequential rounds of strand cleavage and exchange, with a Holliday junction intermediate (Figure 5A) (Kitts and Nash, 1988). In the reactions catalyzed by these enzymes, a tetramer of recombinase protein is required to cut all four strands of the two substrate DNA molecules, and although four proteins are present throughout the reaction, only two monomers of recombinase are active at one time (Chen *et al.*, 1992; Nunes-Duby *et al.*, 1987; Van Duyne, 2002). The two activated integrase monomers cleave one strand on each of the substrates, and the hydroxyl group of the catalytic tyrosine then forms a covalent bond with the DNA, generating a 3' phosphotyrosine and a free 5' OH group (Figure 5B) (Pargellis *et al.*, 1988). This free hydroxyl group then executes the second attack on the phosphate of the other substrate, thus completing one round of strand exchange in the reaction. As a consequence of the first round of strand exchange, a Holliday junction is formed; and an isomerization step then activates the other two units of recombinase, thus initiating the second round of exchange (Cowart *et al.*, 1991; Kitts and Nash, 1988). The second exchange resolves the Holliday junction and results in the formation of the two recombinant products (Hsu and Landy, 1984). Although the chemistry of all tyrosine recombinase reactions is identical, and the mechanisms are similar, the substrate sites where this chemistry occurs are quite different.

I. C. i. a. Simple tyrosine recombinases

For several tyrosine recombinases, the sites upon which they act are relatively simple. Each contains a short asymmetric sequence (6-8bp) that is flanked by a set of short inverted repeats (11-13bp), which serve as recombinase binding elements (Grainge and Jayaram, 1999). It is within the asymmetric region that strand exchange occurs. In all cases where the sites are simple, the two substrate sites are identical. This simple substrate site organization is the case

for the *LoxP* (*locus of X-over, phage*) site acted upon by phage P1 Cre, the *frt* site (*Flp recombination target*) on the 2 μ m plasmid of *Saccharomyces cerevisiae*, which is utilized by the Flp recombinase, and the *dif* site of XerCD (Figure 6) (Abremski *et al.*, 1986; Hoess and Abremski, 1984; Jayaram, 1985). In vivo, Cre acts on two identical 34 bp intramolecular sites in direct orientation to achieve one of two ends; to circularize a linear P1 genome immediately following infection, and resolve genome dimers after replication (Abremski *et al.*, 1986; Sternberg *et al.*, 1981).

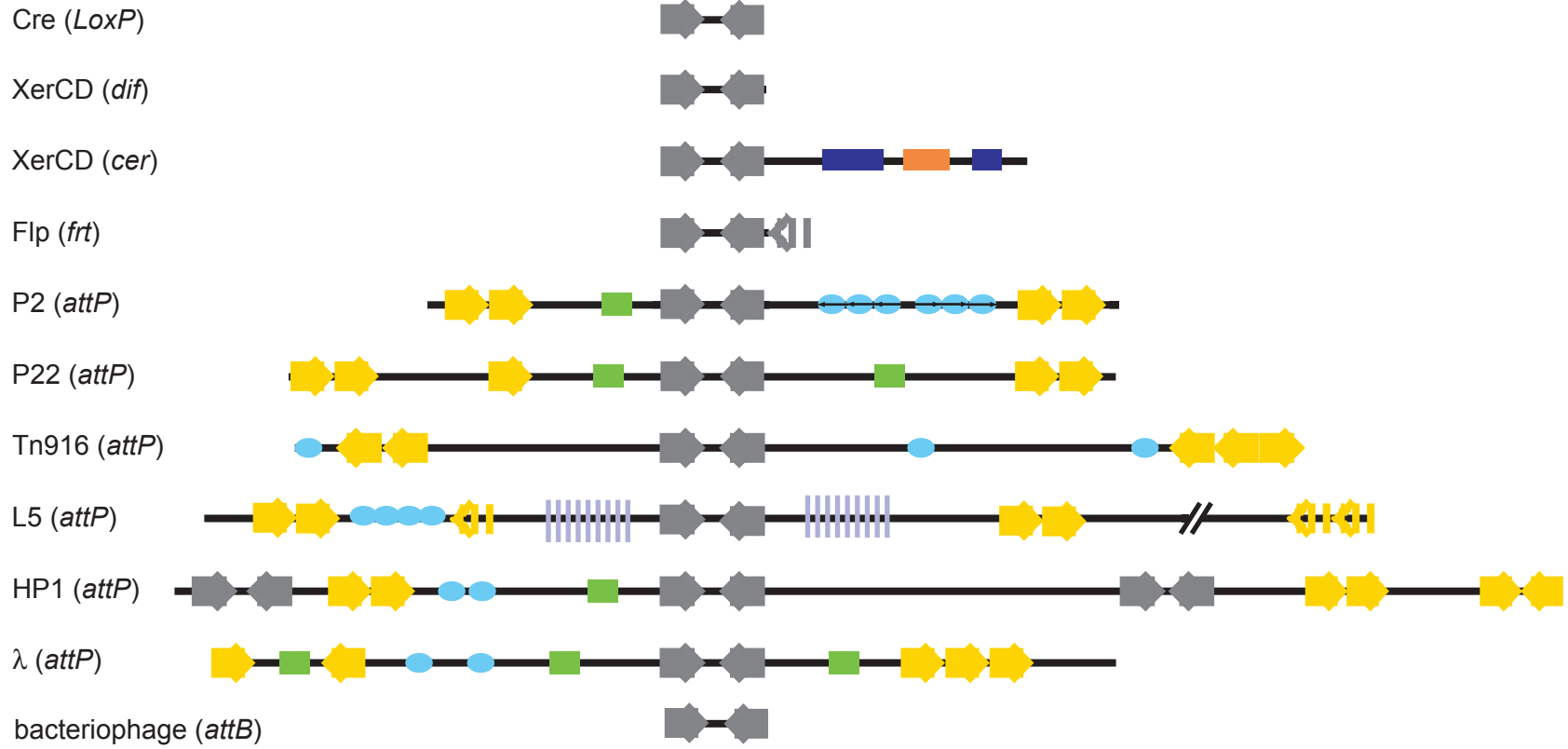
The Flp recombinase of *Saccharomyces cerevisiae* acts in a mechanism to maintain the high-copy number status of the 2 μ m plasmid (Jayaram *et al.*, 2002; Proteau *et al.*, 1986). Flp recombinase also works at two identical simple sites, although in *Saccharomyces cerevisiae frt* has an additional recombinase binding element located on the 3' end of the site, however, it has been shown to be dispensable both *in vitro* and *in vivo* (Jayaram, 1985).

The recombinase complex formed by XerC and XerD, collectively known as XerCD, has more than one substrate site. One of its sites, *dif*, is in the chromosome of *E. coli* and other bacteria near the replication terminus (Blakely *et al.*, 1993). The *dif* site is simple, and recombination events at these sites resolve chromosome dimers following replication (Blakely *et al.*, 1993). Another XerCD site, *cer*, which is located on ColE1 plasmids, consists of a simple site with an additional 190+ bp of regulatory sequences located at the 3' end (Summers and Sherratt, 1988). XerCD also acts at *cer* to resolve replication induced multimers, although in this case the site is part of the ColE1 plasmid (Summers and Sherratt, 1984, 1988). The regulatory region in *cer* is bound by two proteins PepA and ArgR, and recombination does not occur without these proteins (Colloms *et al.*, 1990; Stirling *et al.*, 1988; Stirling *et al.*, 1989). The recombinase proteins described here, Cre, Flp and XerCD, are some of the best studied of the

Figure 6. Tyrosine recombinase substrate sites

Schematic representatives of the tyrosine recombinase substrate sites are shown. Inverted repeat sequences that serve as recombinase recognition sites are shown as arrows. *LoxP*, *dif*, *prt* and the *attB* sites of bacteriophages are the simplest sites and have only a set of inverted repeats flanking a short asymmetric sequence in which strand exchange occurs. The *cer* site used by XerCD has a regulatory region that has sites for two factors, PepA and ArgR, the binding of which is necessary for recombination to occur. The *attP* sites used by several bacteriophages are much more complicated, containing the core type binding sequences (gray arrows) as well as several additional arm type binding sequences (yellow arrows), and sequences for the binding of other regulatory factors like IHF and RDF.

Figure 6



19

= core type rec. site
 = IHF site
 = PepA site
 or = not required
 = 20bp
 = arm type rec. site
 = RDF site (Xis/Cox)
 = ArgR site
 = mIHF binding region (non-specific)

tyrosine recombinases, and they act on two identical sites, however, not all tyrosine recombinases act alike.

I. C. i. b. Tyrosine integrases

In bacteriophages L5 and λ among others, an enzyme called integrase accomplishes two separate reactions involving two sets of non-identical sites, *attB* and *attP*, and *attL* and *attR*. The *attB* is small (≤ 50 bp) and simple, containing inverted repeats and an asymmetric sequence, much like the sites of the simple recombinases (Mizuuchi and Mizuuchi, 1985; Peña *et al.*, 1996). In the case of attachment (*att*) sites, these sequences are often referred to as the core. The *attP* also has a similar core type sequence, in addition, *attP* contains auxiliary recombinase binding sites, known as arm-type sites, as well as sites for the binding of other phage or host derived protein factors (Figure 6)(Hsu *et al.*, 1980; Peña *et al.*, 1997). These core type and arm type binding sites are recognized by different parts of the integrase protein (Figure 4) (Azaro, 2002; Moitoso de Vargas *et al.*, 1988). A single multivalent integrase is able to bind to both arm-type and core-type sites, and when a single monomer binds to both arm-type and core-type sites, it bridges these sites (Better *et al.*, 1982). Although there is no crystal structure available for any entire integrase protein with its paired substrates, it is believed that the *attP* DNA is wrapped around the proteins in a nucleoprotein complex (Figure 7A) (Pena *et al.*, 2000). Complex formation is facilitated by the binding of an additional host-derived protein, such as the *E. coli* integration host factor (IHF) in the case of λ , or mycobacterial Integration Host Factor (mIHF) for L5 (Nash, 1981; Pedulla *et al.*, 1996). These proteins bind to *attP* in between arm type sites and the core (Nash, 1981; Pedulla *et al.*, 1996). IHF has been shown to bend the *attP* DNA, and aid in the formation of nucleoprotein complexes (Robertson and Nash, 1988).

Figure 7. Nucleoprotein complex models of tyrosine recombinases

A) Models predicted for the bacteriophage L5 recombination system are shown. The components of the reaction are shown in (i). Integration intermediate complexes are shown in (ii). In the intasome, a pair of integrase molecules is believed to be bridging arm type sites P4 and P5 to the core within *attP*, with several molecules of mIHF in the region between these arm type sites and the core. The mIHF serves to aid in the formation of this DNA bend. In the synaptic complex, two more molecules of integrase interact with the pair already bridging core with P4 and P5. The second pair of integrase molecules binds to P2 and P1 and capture *attB*. In the presence of L5-Xis (iii), it is believed that Xis binds to and bends the DNA between P1 and P2 and core, preventing the formation of a ‘normal’ synaptic complex, and thus inhibiting integration. In the presence of mIHF, L5-Int is able to form a complex with *attL*, but it is not able to complex with *attR* (iv). However, in the presence of Xis, a stable complex is formed with *attR* and Int. These two complexes, *attL*-Int and *attR*-Int, can synapse and carry out excisive (*attL* x *attR*) recombination. Models are adapted from (Pena *et al.*, 2000), and (Lewis and Hatfull, 2003).

B) The predicted synaptic complex of XerCD with *cer* is shown. Here its typical substrate, a dimer of a ColE1 plasmid is shown as the black line with two *cer* sites. The accessory factors ArgA and PepA are shown in dark and light green, respectively. It is thought that these proteins act to form a complex trapping a specific DNA topology, bringing the two *cer* cross-over sites (red arrows) in close proximity with each other. The cross-over region is bound by XerC (in blue) and XerD (in violet), which carry out the chemistry of the reaction. This model is adapted from (Hallet and Sherratt, 1997).

Figure 7A

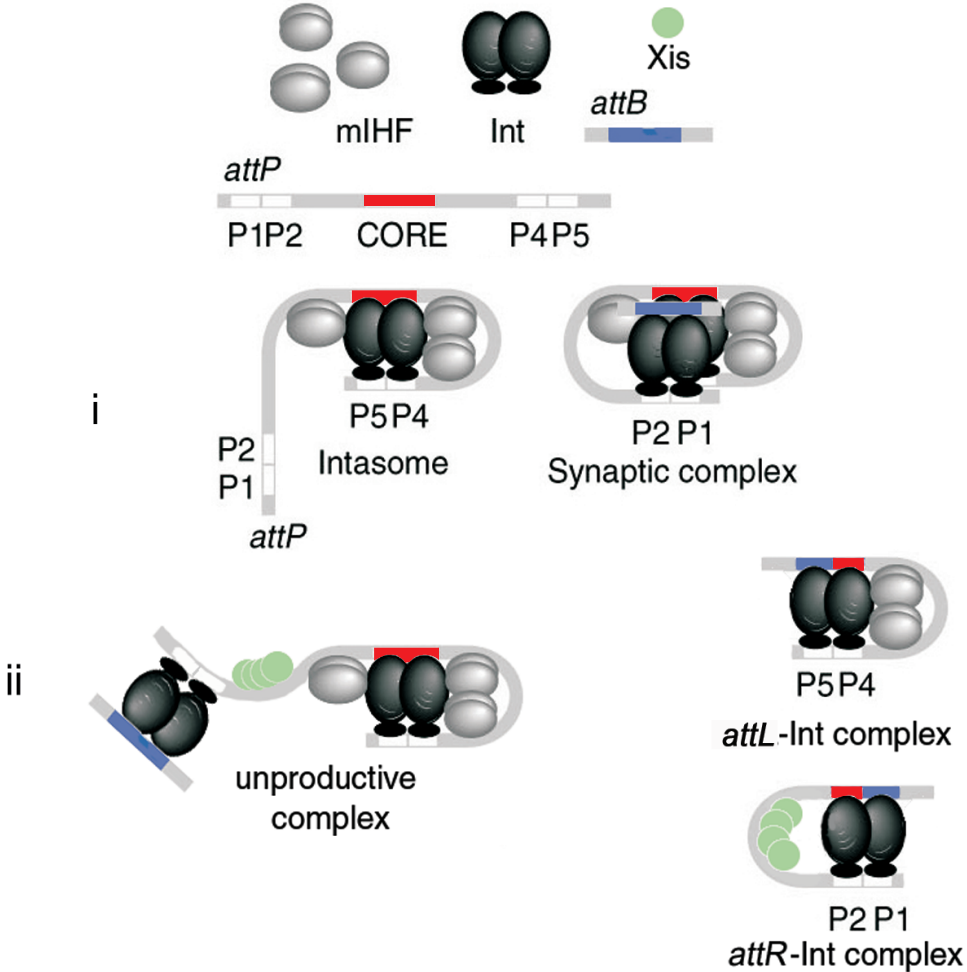
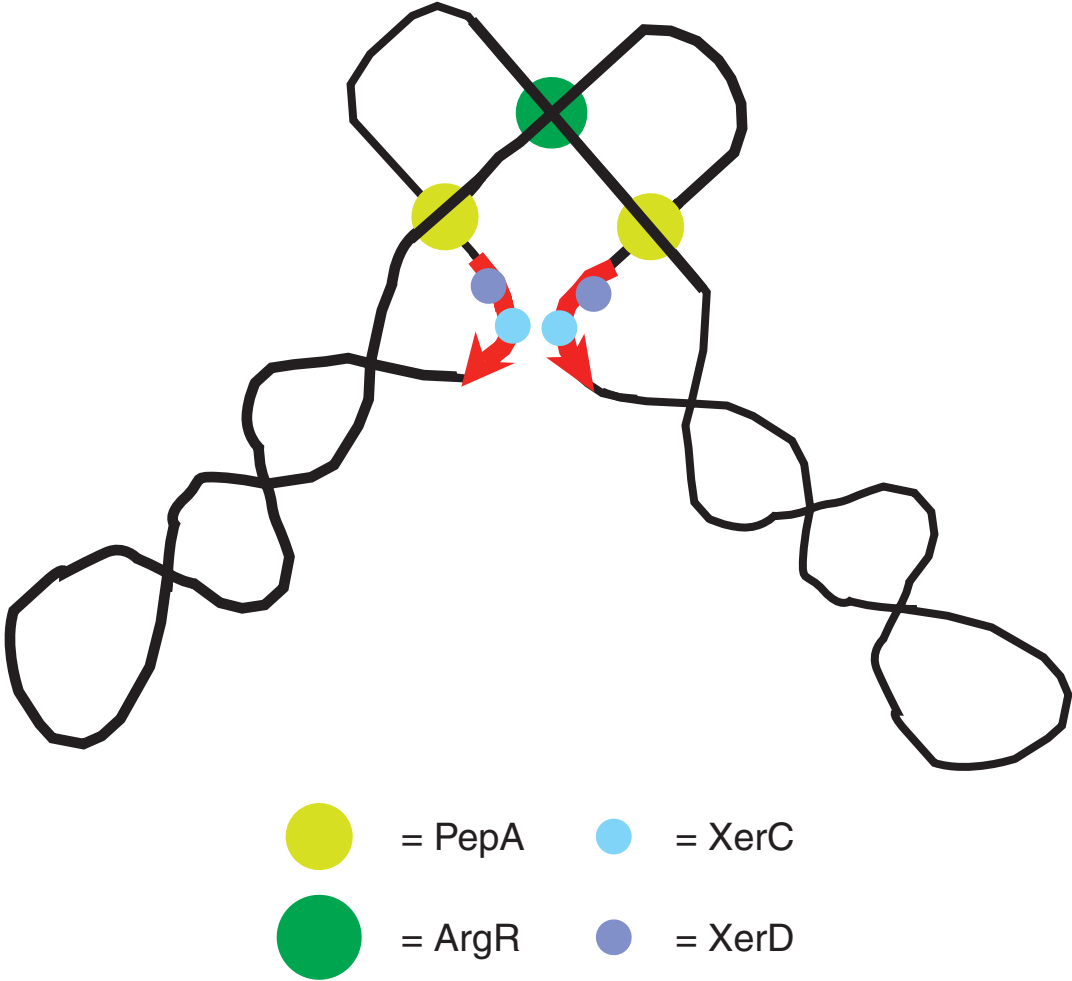


Figure 7B



Integrases also catalyze excision, which is essentially the reverse of integration. Here the *attL* and *attR* sites are recombined, and these two substrate sites are also non-identical, hybrids of *attP* and *attB*. Excisive recombination requires an additional phage-derived protein called a recombination directionality factor, or RDF (Abremski and Gottesman, 1982; Lewis and Hatfull, 2000).

I. C. i. c. Tyrosine recombinases and the control of directionality

The regulation of which site-specific reactions occur is important to ensure that the appropriate end is reached. For example, if XerCD were to carry out integration and create dimers, rather than catalyze dimer resolution, then plasmids would be lost and daughter cells would be without a complete chromosome. If a phage genome were to excise under the wrong conditions, no progeny would be produced. In the tyrosine recombinase systems, directionality is regulated in several different ways, and some of these systems are better understood than others. Regulation of XerCD at chromosomal *dif* is also complicated and poorly understood (Neilson *et al.*, 1999). There are several factors involved in the cell; the position of the *dif* site in the chromosome is important, cell division is required, a functional homologous recombination system is needed, and FtsK is known to play a role, but it is not yet known how all these elements are coordinated (Capioux *et al.*, 2002; Ip *et al.*, 2003; Steiner and Kuempel, 1998). At other XerCD substrate sites, such as the *cer* site in the ColE1 plasmid, the accessory proteins PepA and ArgR are necessary for recombination to occur (Stirling *et al.*, 1988; Stirling *et al.*, 1989). These factors and the recombinase bind to the site, and the DNA adopts a specific conformation, allowing recombination only at intramolecular sites in direct orientation; that is resolution (Figure 7B) (Alen *et al.*, 1997).

The directionality of the recombination reactions with bacteriophage-encoded tyrosine integrases is controlled by a phage-derived protein called a recombination directionality factor or RDF. In both λ and L5, the RDF is referred to as excisionase, excise or Xis. These proteins inhibit the integration reaction, and stimulate excision (Abremski and Gottesman, 1982; Lewis and Hatfull, 2000). The RDFs in the tyrosine systems are DNA binding proteins that bind specifically at sites within *attP*, and in the case of L5, Xis binds to four sites in between arm-type sites P2 and P3 (Figure 6) (Lewis and Hatfull, 2003). This binding does not appear to interfere with the binding of integrase, but Xis bends the *attP* DNA and it is thought that Xis does not allow for proper synapsis with *attB* (Figure 7A) (Lewis and Hatfull, 2003). To promote excision, L5-Xis directs the formation of an *attR*-Int complex, which can synapse with an L5-Xis independent *attL*-Int complex and promote excisive recombination (Figure 7A) (Lewis and Hatfull, 2003).

The control of directionality in λ is similar, and lambda Xis binds to the left arm of *attP* and *attR* at two 13bp sites in between P2 and the core (Figure 6) (Bushman *et al.*, 1984; Yin *et al.*, 1985). In the lambda system, there is experimental evidence that Xis interacts directly with integrase (Cho *et al.*, 2002). Like in the L5 system, the binding of lambda Xis bends the DNA and allows an *attR*-Int complex to form that participates in excisive recombination (Moitoso de Vargas and Landy, 1991; Thompson *et al.*, 1987).

In some phage systems that utilize tyrosine integrases, directionality control is slightly different. Bacteriophages P2, HP1, and 186 have related RDF proteins called Cox or Apl (Dodd *et al.*, 1993; Eriksson and Haggard-Ljungquist, 2000; Esposito and Scocca, 1994). These proteins are multifunctional, and in addition to their role in controlling recombination directionality, they function to regulate expression from certain promoters (Esposito and Scocca,

1997; Saha *et al.*, 1987). Cox proteins are similar to Xis in their role as RDFs, and act as a structural element of the substrate integrase complex (Eriksson and Haggard-Ljungquist, 2000).

I. C. ii. Serine recombinases

The second of the site-specific recombination enzyme families, the serine recombinases, utilize a serine as the catalytic residue (Hatfull and Grindley, 1988). Enzymes of this family, which include DNA invertases, transposon resolvases, and integrases, largely fall into two sub-families. The smaller of the two (180-210 amino acids) includes the majority of transposon resolvases and DNA invertases (Grindley, 2002). These enzymes have an amino-terminal catalytic domain of about 140 residues, which includes the serine residue (situated around position 10), and a small C-terminal domain that contains DNA binding determinants, including a Helix-Turn-Helix (H-T-H) motif (Figure 8A) (Hatfull and Grindley, 1988). The so-called large serine recombinases range in size from 441 to 772 residues (Smith and Thorpe, 2002). This group includes mostly phage integrases, and these large serine recombinases have a similar N-terminal catalytic domain, but in addition have a large C-terminal domain that presumably contains determinants for substrate binding. This domain has no recognizable H-T-H motif and is diverse among recombinases (Figure 8B) (Smith and Thorpe, 2002). Although most of the information regarding the serine recombinases has been gathered from work with invertases and resolvases, the data supports the idea that these large serine recombinase proteins act with a similar chemistry and strand exchange mechanism (Ghosh *et al.*, 2003). The serine recombinases are known to act through a concerted double strand cleavage of both DNA partners, followed by strand exchange and ligation (Figure 9A) (Stark *et al.*, 1989). As an intermediate in the reaction, these proteins covalently bind to the DNA creating a 5'

Figure 8. Structural organization of serine-integrases

A) Schematic representation of serine recombinases. Phage-encoded serine integrases contain an N-terminal domain that is approximately 140 residues long (shown in red), which has sequence similarity with the N-terminal catalytic domain of transposon resolvases and DNA invertases. The C-termini of the resolvases and invertases are small and have Helix-Turn-Helix motifs for substrate binding. The C-terminal segments of the large serine recombinases are diverse both in sequence and length and are of unknown function, although they presumably include the DNA-binding determinants. The closest relative to the ϕ Rv1 integrase is encoded by phage R4. Phages Bxz2, Bxb1, U2 and Bethlehem are the only mycobacteriophages that have predicted or known serine integrases.

B) Alignment of several serine recombinases. The gamma delta resolvase is abbreviated 'gd'. Rv1586c is the Integrase from the *Mycobacterium tuberculosis* prophage-like element ϕ Rv1. R4, Bxz1, and Bxb1 are the integrases from these bacteriophages. Residues that are absolutely conserved are shown in red, and include the catalytic serine, which in this alignment is shown at position 20. Residues that are conserved, with four of the five aligned proteins are shown in blue with yellow lettering, and similar residues are shown in violet with white lettering.

Figure 8A

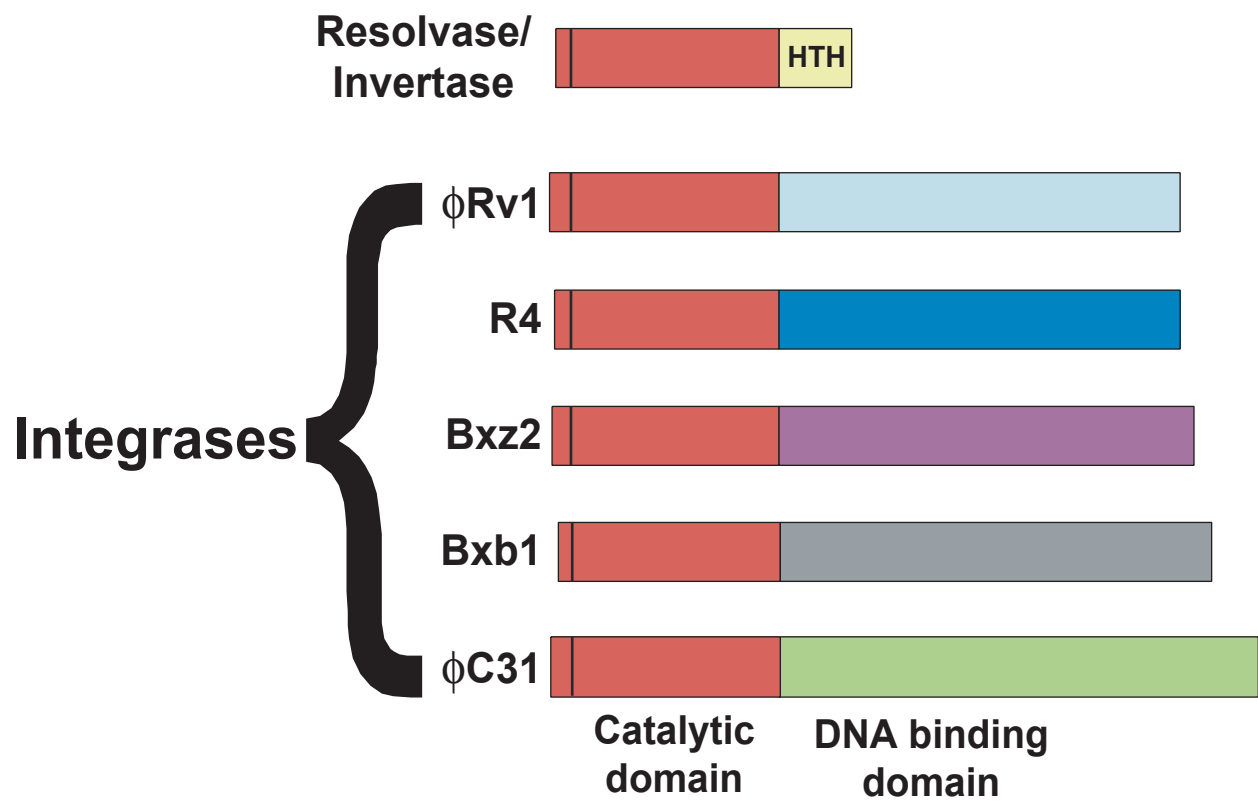


Figure 8B

```

Rv1586c 1:..MRYTTPVRAAVYLRLISEDRSGEQLGVARQREDCCLKCGQR.KWVPVEYLDNDVVSASTG.K...RRPAYE: 64
R4      1:MNRGGPTVRADIVVRISLDRTEGEEVGVERQESCRELCKSLGMEVGVQVWVDNDLSATKKNV...VRPDFE: 67
Bxz2   1:...MAQPLRALVGARVSVVQGPQKVSHIAQQETGAKWVAEQGHTVVGSFKDLVVSATVSPF...ERPDLG: 64
Bxb1   1:.....MRALVVIRLSR.VTDATTSPERQLESCQQCAQRGWVVGVAEDLDVSGAVDPFDRKRRLPNLA: 62
gd     1:.....MRLFGYARVST..SQQSLDIQVRALKDAGVKANR.....IFTDKASGSSSD.....RKGLD: 49

Rv1586c 65:QMLADITAGKIAAVVAVDLDRIHRPIELEAFMSLADEKR.LALATVAGDVDLATP...QGRVVARLKGS:130
R4     68:AMIASNPQ...AIVCWHTRDRIIRVTRDLERVIDLGVN...VHAVMAGHLDLSTE...AGRAVARTVTA:126
Bxz2   65:PWLSPELEGEWDILVFSKIDRMFRSTRDCVKFAEWAEAHG.KILVFAEDNMTLNYRDKDRSGSLESMMSE:133
Bxb1   63:RWLAFFEEQ.PFDVIVAYRVDRLTRSRHLQQLVHWAEHKKLVVSATEAHFDLTTTE...FAAVVIALMGT:128
gd     50:LLRMKVEEG..DVILVKLDRLGRDTADMIIQLIKEFDAQG.VSIRFIDDGISTDGE...MGKVVVTILSA:113

Rv1586c 131:VAAHETEHHKARQRRARQKAEERGHFNWSKAFGVLPGP.....NGPEFDERTAPLVKQAY:185
R4     127:WATYEGEQKAEKQLANIQNARAGKPYTPGIRPFGYGD.....DHMTIVTAEADAIRDGA:181
Bxz2   134:LFIYIGSFFAQLELNRFKSRARDSHRVLRGMDRWASGVPLGFRIVDHPSGKKGKLDLDFEGKALILEDMA:203
Bxb1   129:VAQMELEAIKERNRSAHFNRAG.KYRGSLLPPWGYLPTRVVG.....EWRLVDFVQERERILEVY:188
gd     114:VAQAERQRILERTNEGRQEAMAKG.....VVFGRKR.....KIDRDAVNLNMW:155

Rv1586c 186:ADILAGAS.LGDVCRQWNNDAG.....AFTITGRP.....WTTTTLTKFLRKPRNAGLRAYKGARYGPV:242
R4     182:KMILDGWS.LSAVARYWBEELKIQS..PRSMAAGKKG.....WSLRGVKVKVLTSPRYVGRSSYL:237
Bxz2   204:AKLLDGWS.FIRIAQDLNQRKVLNMDKAKIAK GKPPHPNPWTVNTVIESLTSPTQGIKMTKHGTRGGS:272
Bxb1   189:HRVVDNHEPLHLVAHDLNRRGVLSPKDYFAQLQGREPQGREWSATALKRSMISEAMLGYATLNG.....:252
gd     156:QQGLGASH.....ISKTMNIAR.....STVYKVLINESN.....:183

Rv1586c 243:DRDAIVGKAQWSFLVDEATFWA...AQAVLDAPGRAPGRKSVRR...HLLTGLAGCGKCGNHLAGSYRTD:306
R4     237:...EVVGDAQWPPIIDPDVYVG...VVAILNNPDRFSGGPRTGRTPGTLLAGIALCGECCKTVSG.RGYR:300
Bxz2   273:KIGTTVLDAEGNPIRLAPPTFPATWKOIQEAAARRQGNRRSKTYTANPMLGVGHCGACGASLAQQFTHR:342
Bxb1   252:...KTVRDDDGAFLVRAEPILTREQLEALRAELVKTSRAKPAVSTP.SLLLRVLFCAVCCPEPAYK.....:313
gd     :.....:

Rv1586c 307:GQVVVYCKACHGVAILADNIEFILYHIVAERLAMPDAVD..LLRREIHDA.....AEAETIRLELETL:367
R4     301:GVLVYGCCKDTH..TRTPRSIADGRASSSTLARLMFPPDFLPGLLASGQAEDEG.....QSAASKHSQAQTL:362
Bxz2   342:KLADGTEVTY..RTYRCGRTPFLNCSGISMRGDEADGLLEQLFLEQYGSQPVTEKVFVPEGEDHSELEQV:409
Bxb1   313:..FAGGRK..HPRYRCRSMGFPKHCNGTVA MAEWDAFCEEQVLDLLGDAERLEKVVVAGSDSAVELAEV:380
gd     :.....:

Rv1586c 368:YGELDRLAVERAEGLITAR...QVKISTDIVNAKITKLQARQDQERLRVFDGIPLGTPQVAGMIAELSPD:435
R4     363:RERLDGLATAYAEGLISLS...QMTAGSEALRKKLEVIEADLVGSAGIPPFDVPVAGVAGLISGWPTTPLPT:430
Bxz2   410:RATIDRLRRESDAGLIATAE.DERIYFERMKSLIDRRRLEAQPRRASGWTQETDKTNADEWTKASTPD:478
Bxb1   381:NAELVDLTSLIGSPAYRAGSPQREALDARI AALAAARQEELEGLEARPSGWEWRETGQRFQDWREQDTAA:450
gd     :.....:

Rv1586c 436:RFRAVLDVLAEVVVQPVG...KSGRTFNPERVQVNW...:469
R4     431:RRRAWDFCLVVTLNTQKGR..HASSMTVDDHVTIEWRDVAE.....:469
Bxz2   479:ERRRLLMKQGIKRFELVRCRKPDPFVRLFTPGEIPEGEPLPEPSR.....:522
Bxb1   451:KNTWLRSMNVRLTFDVRKGLTRTIDFGDLQEYEQHLRLGVSVERLHTGMS:500
gd     :.....:

```

Figure 9. Mechanism and chemistry of serine recombinases

A) A schematic representation of the strand exchange mechanism of serine recombinases is shown. Each substrate site is shown as paired black or blue lines, and recombinases are represented as ovals. Strand exchange is concerted with breaks occurring at all four strands at the same time (panels i and ii). These strand breaks on the top and bottom strands are separated by 2bp. After cleavage, strands are exchanged (iii) by an unknown mechanism involving an 180° rotation of the DNA strands. These breaks are then rejoined in the recombinant configuration (iv).

B) The breakage and rejoining mechanism at a single active site is illustrated here. The OH of the catalytic serine attacks the scissile phosphate at the crossover site, creating a 5' phosphoserine and a 3' OH leaving group. The strands are then exchanged via an unknown mechanism, and then the free hydroxyl group of the incoming strand attacks the phosphoserine resulting in a recombinant DNA strand.

Figure 9A

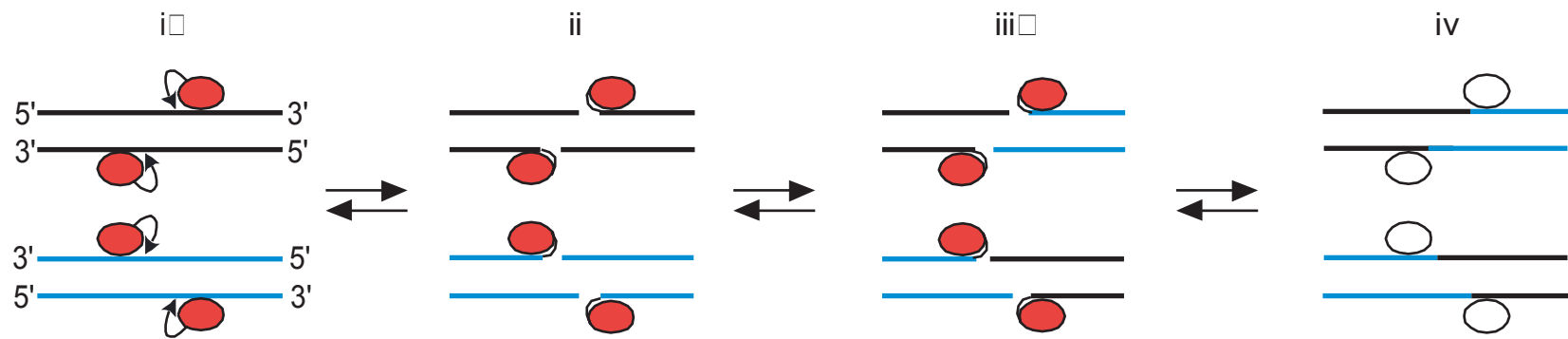
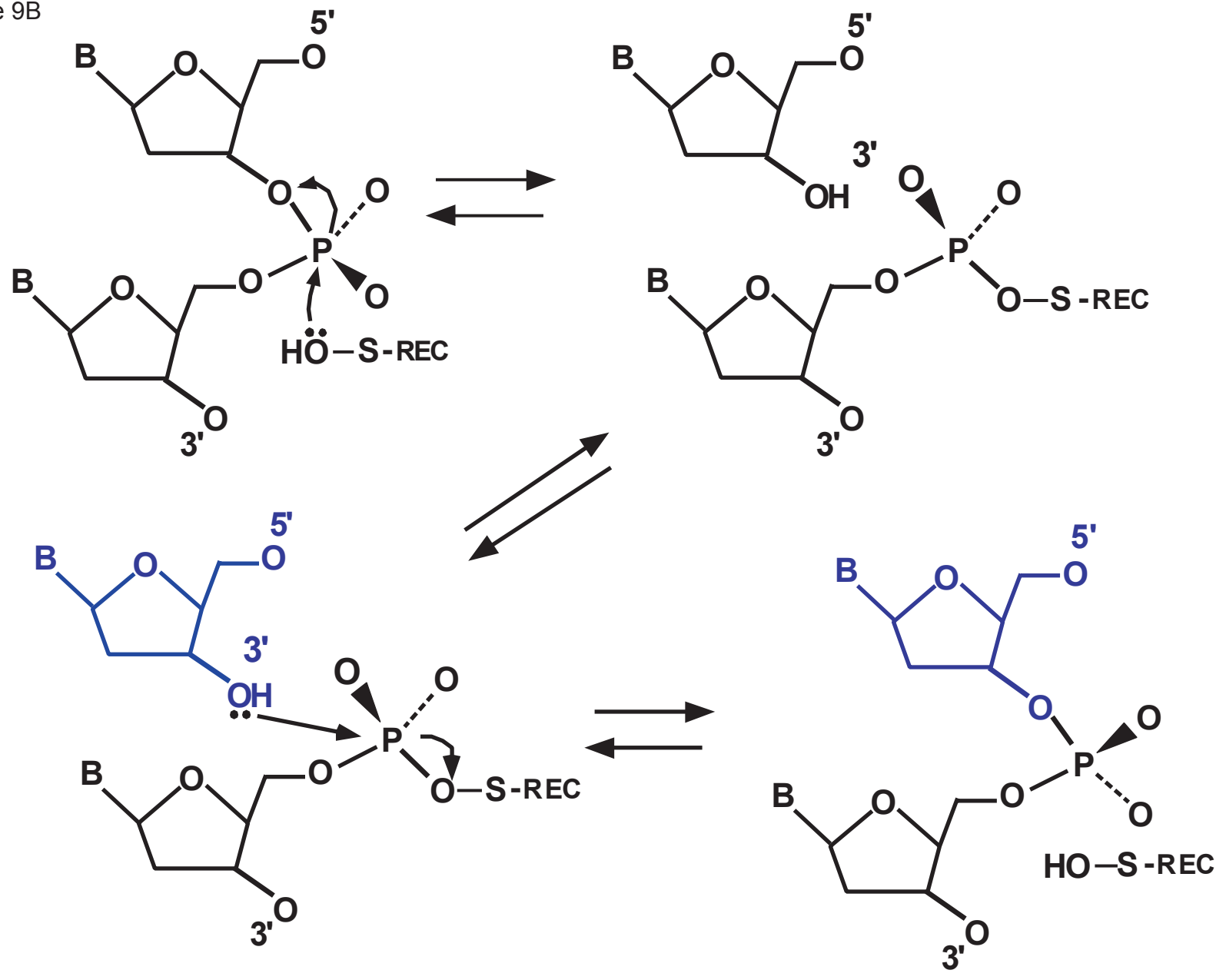


Figure 9B



phosphoserine and 3' OH leaving group (Figure 9B) (Reed and Grindley, 1981). The cuts made by the proteins are staggered by two base pairs, generating a 3' overhang (Reed and Grindley, 1981). Like the tyrosine recombinases, the chemistry of these diverse enzymes is identical, but variation exists in the sites upon which the proteins act, especially between the small and large recombinases (Figure 10A and B).

I. C. ii. a. Small serine recombinases

The majority of the small serine recombinases are invertases and resolvases, and these have been relatively well characterized. DNA invertases act at two identical sites that are simple and contain a pair of inverted repeats flanking an asymmetric spacer (Figure 10). These invertases include Gin of coliphage Mu, which inverts a DNA sequence that directs differential expression of tail fiber genes, and Hin of *Salmonella*, which generates flagellar variation (van de Putte, 1980; Zeig, 1977). Although the invertase sites are simple, there is a single enhancer element to which an accessory protein binds within the Hin invertible segment and provides 1,000-fold enhancement of inversion (Johnson *et al.*, 1986; Mertens *et al.*, 1986). The protein that binds to the enhancer is called Fis in *E. coli* and factor II in *Salmonella* (Johnson *et al.*, 1987; Johnson *et al.*, 1988).

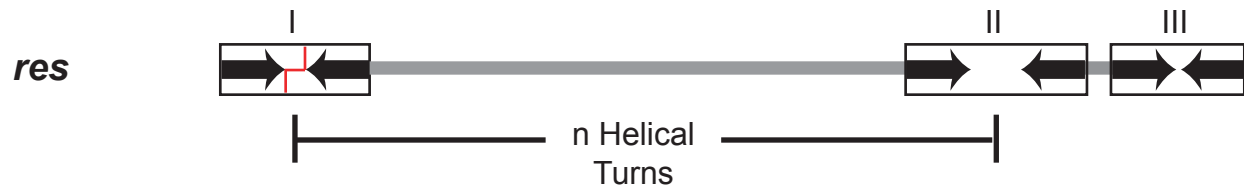
Resolvases, which are most often associated with replicative transposons such as $\gamma\delta$ and Tn3 are among the best understood of the small serine recombinases. A pair of identical sites are resolved that are 90-140bp long, with a series of two to three subsites each containing a pair of inverted repeats flanking a short asymmetrical segment (Figure 10) (Grindley, 2002). The canonical resolvase site (res) has three subsites, which are named I, II, and III. Subsites I and II are separated from the other by a number of helical turns, such that they lie on the same face of

Figure 10. Serine recombinase substrate sites

A) Schematic representations of the sites for two representative small serine recombinases are shown. These sites are for a resolvase (*res*) and an invertase (*inv/ix*). The *res* site has three subsites (I, II, and III), each consisting of a pair of inverted repeats that serve as binding sites for the resolvase enzyme, flanking a short asymmetric region. Although all three subsites are bound by resolvase, cross-over occurs within subsite I (red staggered line). The invertase substrate sites, known as *ix* (e.g. Hin sites are called *hix*) consist of a pair of inverted repeat sequences that are part of the recombinase binding site. The repeats flank a short (2bp) asymmetric sequence which is the central dinucleotide where cross-over occurs. Invertases also require a single enhancer element (~65bp) that is sometimes located within the invertible segment, but can be outside of it. This enhancer is bound by a protein that both stimulates the inversion reaction and imparts directionality control.

B) Features of *attP* and *attB* sites of serine integrases. The phage and bacterial host attachment sites (*attP* and *attB*, respectively) are shown for five serine integrases. For four of these systems the minimum size required for full activity is known, and is shown in parenthesis following *attB* and *attP*. Regions of identity between the two sites are shown as red bases, and the central dinucleotide is indicated by the boxed bases when it is known or has been predicted. Inverted repeats are shown as horizontal lines above each sequence.

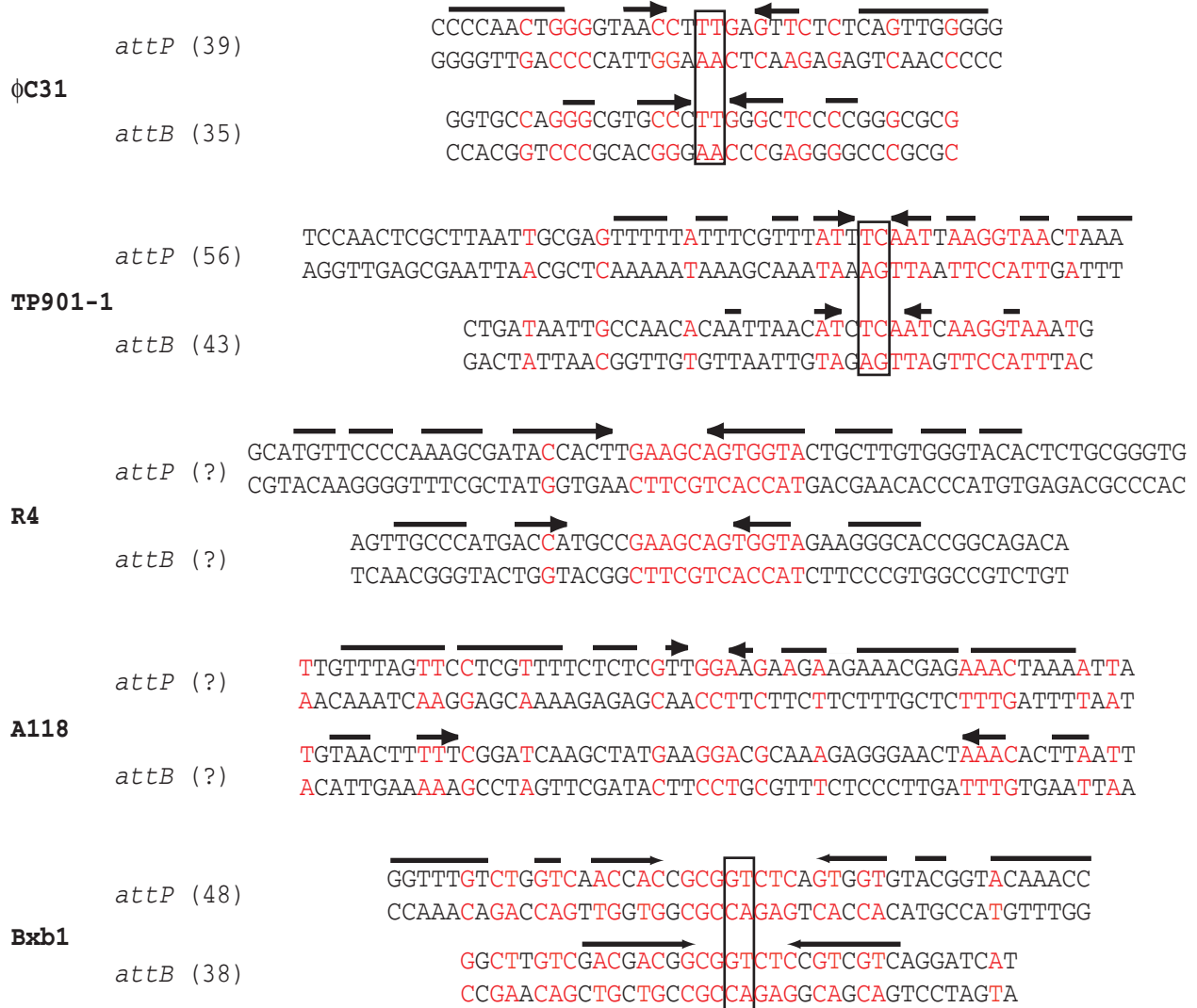
Figure 10A



inv (_ix)



Figure 10B



the DNA helix. A tetramer of resolvase binds at each pair of these synapsed subsites, for a total of 12 protomers (Rice and Steitz, 1994). Strand exchange occurs only at subsite I, but the other sites are also bound by the resolvase and are required for the correct topological complex to form and bring the subsite I's together (Figure 11) (Hallet and Sherratt, 1997).

I. C. ii. b. Large serine recombinases

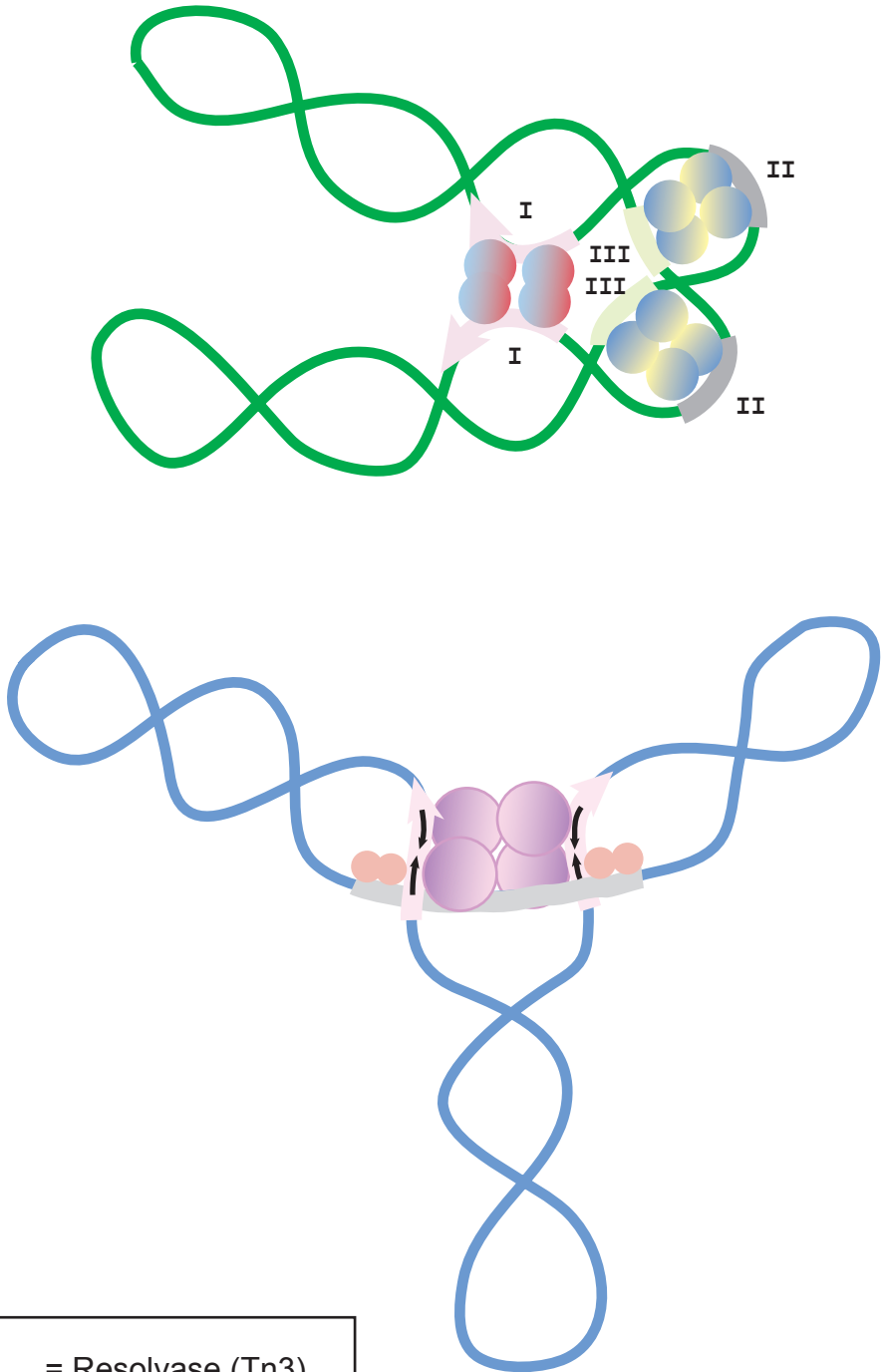
The family of serine recombinases largely consists of invertases and resolvases, however, more recently, a number of large serine recombinases have been found in the genomes of sequenced bacteria and bacteriophages. There are known examples that participate in developmental processes, such as spore formation in *B. subtilis*. During sporulation, SpoIVCA deletes a 48-kb DNA element, often referred to as the skin element (*sigK* intervening), which is within a sigma factor (*sigK*) required for mother-cell-specific gene expression (Kunkel *et al.*, 1990). Another large serine recombinase, XisF acts in heterocyst formation in *Anabaena* where it deletes an element within *fdnX* that is involved in nitrogen fixation (Carrasco *et al.*, 1994; Popham and Stragier, 1992). There are also enzymes involved in transposon mobility, such as TnpX and TndX (Bannam *et al.*, 1995; Wang *et al.*, 2000). Most of these large serine recombinase proteins, however, are associated with prophages and are integrases (Smith and Thorpe, 2002). The serine integrases carry out the same basic function as the tyrosine integrases, which is to integrate and excise their associated phage genome. Although the serine integrases are a relatively new group of recombinases, many systems are currently being investigated. Several systems have been studied, at minimum demonstrating that the integrase is active, these include bacteriophage R4, Bxb1, ϕ C31, TP901-1, ϕ FC1, A118, and SP β c2




Figure 11. Serine recombinase-DNA complexes

Resolvase utilizes twelve protomers bound to two *res* sites with three subsites each. The four molecules of resolvase bound at subsite I are the only catalytically active proteins, as crossover occurs within subsite I. The protomers bound to subsites II and III (eight total, four bound to each synapsed subsite pair) are thought to serve as structural elements to bring the two subsite I's into close proximity for catalysis. These resolvase molecules bound to subsites II and III also act as directionality control factors and allow only resolution to occur.

In contrast, the invertases have only four invertase subunits binding in complex, all of which participate in catalysis. These proteins bind to *inv* where catalysis occurs. In addition, an enhancer element is necessary for catalysis, and this element interacts with the invertase as well as *Fis*. These interactions cause the DNA to form a specific topology that aligns the crossover sites, again only allowing a certain reaction, in this case inversion, to occur.

Figure 11



	= Resolvase (Tn3)
	= Invertase (Hin)
	= Fis

(Christiansen *et al.*, 1994; Kim *et al.*, 2003; Lazarevic *et al.*, 1999; Loessner *et al.*, 2000; Matsuura *et al.*, 1996; Thorpe and Smith, 1998; Yang *et al.*, 2002), but few have been studied in depth. The substrate sites have been identified in at least six cases (all those cited above except ϕ FC1), and the minimal size required for activity is known in three phages (Bxb1, TP901-1 and ϕ C31) (Breuner *et al.*, 2001; Groth *et al.*, 2000; Kim *et al.*, 2003). In all these cases, both the *attP* and *attB* sites are small and appear to be simple. Like in the tyrosine integrase systems, *attB* consists of a pair of inverted repeats (IR) that flank a short asymmetrical sequence; however, unlike in the tyrosine systems, *attP* is also small and consists of a pair of IRs flanking an asymmetrical sequence (Figure 10) (Breuner *et al.*, 2001; Groth *et al.*, 2000; Kim *et al.*, 2003). Also unlike the tyrosine integrases, the IR sequences flanking the core in *attP* and *attB* are different. For the systems where an *in vitro* assay has been developed, ϕ C31 and Bxb1, it has been shown that supercoiling is not required for the reaction to occur (Kim *et al.*, 2003; Thorpe and Smith, 1998). For several systems, it is also known that there is no host factor requirement, and substrates are recombined efficiently *in vitro* and in heterologous hosts such as *E. coli*, *Schizosaccharomyces pombe*, and mouse and human cells (Groth and Calos, 2004). This is in contrast to tyrosine integrases, which require both a large *attP* sequence and a host factor and are often dependent on DNA supercoiling for efficient recombination.

I. C. ii. c. Control of directionality in the serine recombinases

As in the tyrosine recombination systems, it is important to maintain tight control over which reactions occur. For both the resolvase and invertase systems, directionality control is maintained by nucleoprotein complexes that cause the DNA to adopt a specific topology. Invertases require an accessory protein (e.g. Fis), which binds to an enhancer element (Johnson

and Simon, 1985; Johnson *et al.*, 1986). Experimental evidence suggests a complex with invertase, Fis and the substrate DNA that allows the invertase sites to align in only one orientation, and thus inversion occurs exclusively (Figure 11) (Bruist *et al.*, 1987; Johnson *et al.*, 1987). In several systems, mutants have been isolated that are Fis independent, and these are also defective in directionality control and can both invert and resolve substrates (Klippel *et al.*, 1988).

Resolvases also control directionality through the formation of protein-DNA complexes; however, resolution reactions require no additional factors. Instead, several molecules of resolvase are bound in a complex (Rice and Steitz, 1994). The *res* site has three subsites, and a dimer of resolvase binds to each of these subsites, thus, when the two *res* sites synapse, a tetramer of resolvase is bound to each set of paired subsites. Crossover occurs at subsite I, thus the molecules of resolvase bound here are catalytically active, while recombinase molecules that are bound at subsites II and III are architectural and act to capture the DNA in a specific topology and bring the subsite I's into synapse (Benjamin and Cozzarelli, 1990; Grindley, 1993; Stark *et al.*, 1989). Due to the specific topology of the complex, only resolution can occur (Stark and Boocock, 1995). Without subsites II and III, recombination does not normally occur, although mutants of resolvase have been isolated which can recombine two subsite I's in the absence of subsites II and III (Arnold *et al.*, 1999).

In the large serine recombinase systems, the basis for directionality is not well understood. *Anabaena* XisF has two accompanying ORFs, XisH and XisI, which are required for heterocyst-specific excision of the element in *fdxN* (Ramaswamy *et al.*, 1997). These RDFs appear to dictate the cell specificity of the recombination reaction, however, very little information is known about how these proteins function (Ramaswamy *et al.*, 1997).

Bacteriophage TP901-1 has an RDF that was identified as an open reading frame required for the excision of an integrated plasmid containing *attP* and *integrase* (Breuner *et al.*, 1999). Further characterization of this protein has not yet been reported. In ϕ C31, Integrase alone is unable to catalyze excisive recombination (*attL* x *attR*), but no factor involved in excision has been identified (Thorpe *et al.*, 2000).

I. D. Mycobacteria

Bacteria of the genus *Mycobacterium* have genomes with high GC content, and are characterized by their acid-fast staining. Although there are members that occur commonly in the environment like the fast-growing saprophyte *M. smegmatis*, several species are pathogenic. *M. avium* is an opportunistic pathogen causing disease in immuno-compromised individuals, *M. avium* subspecies *paratuberculosis* has been implicated in Crohn's Disease, which is a gastrointestinal disorder (Bull *et al.*, 2003). *M. leprae* is the causative agent of Leprosy, and *M. tuberculosis* causes tuberculosis. *Mycobacterium tuberculosis* remains a global health problem. It is estimated that in the year 2000, 8.2 million new cases of tuberculosis occurred worldwide, and approximately one-third of the world's population is believed to be infected with *M. tuberculosis* (Corbett *et al.*, 2003; Kochi, 1991). Recently several whole genome sequences mycobacteria strains have been made available; *M. tuberculosis* strains H37Rv, CDC1551, *M. bovis*, *M. avium* subspecies *paratuberculosis*, and *M. leprae* (Cole *et al.*, 1998).

I. E. Prophages and prophage-like elements ϕ Rv1 and ϕ Rv2

Many bacterial genomes harbor prophages or prophage-like elements, and these genetic elements often confer a phenotype to the host bacterium. In several cases the prophage carries a

toxin or virulence determinants. For example, the shiga toxin found in *Shigella* species and *E. coli* O157:H7 and the cholera toxin of *Vibrio cholera* are carried by prophages (Bhadra *et al.*, 1995; Bhargava *et al.*, 1990; Neely and Friedman, 1998). Salmonella also has several resident prophages that contribute to the pathogenicity of the bacterium (Figuroa-Bossi and Bossi, 1999). In other bacteria, prophage-like elements impart the host bacterium genetic transfer ability. These elements make bacteriophage head-like particles that package random bacterial DNA, which is then transferred to other cells, a process known as generalized transduction. Such elements have been described in *Methanococcus*, *Rhococcus*, and *Serpulina* (Bertani, 1999; Eiserling *et al.*, 1999; Humphrey *et al.*, 1997; Lang and Beatty, 2000).

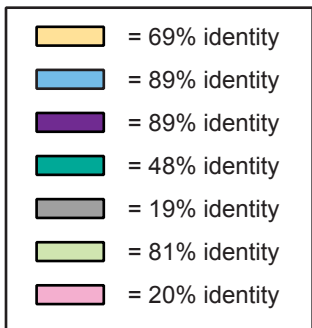
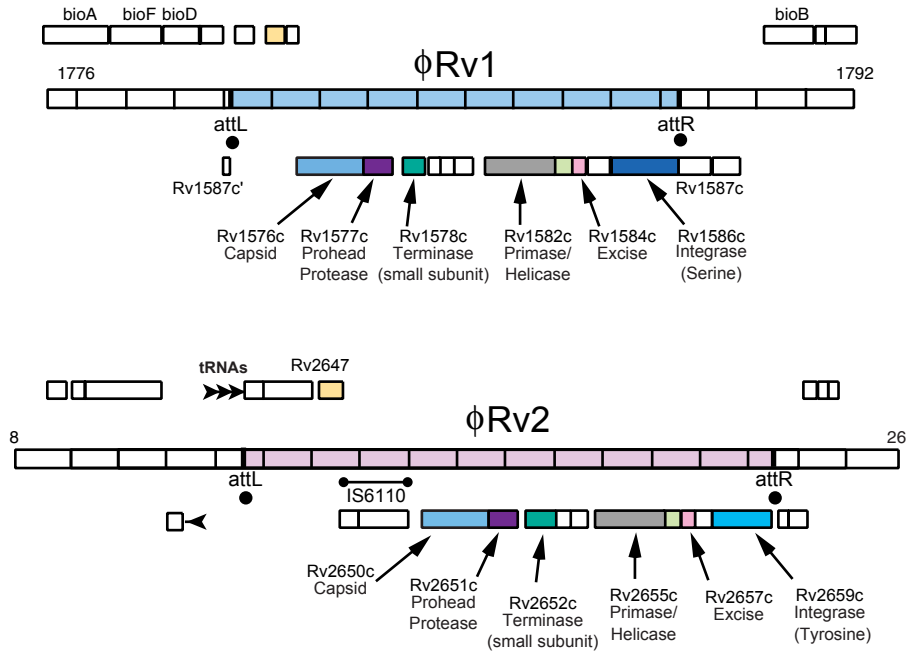
Examination of the genome sequence of *M. tuberculosis* H37Rv revealed several interesting features including two resident elements that appear to be related to bacteriophages called ϕ Rv1 and ϕ Rv2 (Figure 12) (Cole *et al.*, 1998; Hendrix *et al.*, 1999). Subtractive genomic hybridization comparing *M. tuberculosis* H37Rv, *M. bovis*, and the vaccine strain *M. bovis* BCG showed that these two elements, as well as a third region, are present in *M. tuberculosis* but absent from all BCG strains tested (Mahairas *et al.*, 1996). The deletion in BCG that is referred to as RD3 (Region of Difference), corresponds to ϕ Rv1. It is approximately 10 kilobases in length (9,247bp) and contains 14 predicted open reading frames (ORFs) (Figure 12) (Mahairas *et al.*, 1996). Although there are likely not a sufficient number of ORFs to make a complete dsDNA tailed bacteriophage, several of the proteins that are predicted by these ORFs have significant similarity to known phage proteins. There is a predicted capsid subunit (Rv1576c), a prohead protease (Rv1577c), a terminase (Rv1578c), a primase/helicase (Rv1582c), an integrase (Rv1586c) and a recombination directionality factor (RDF) (Rv1584c) (Figure 12). The small size of this element, along with the similarity to phage genes, has prompted the use of the term

Figure 12. *M. tuberculosis* H37Rv prophage-like elements ϕ Rv1 and ϕ Rv2

Prophage-like elements ϕ Rv1 and ϕ Rv2 are present in the genome of *M. tuberculosis* H37Rv (Hendrix et al., 1999). Although ϕ Rv1 and ϕ Rv2 are too small to make complete bacteriophage, both contain a predicted capsid protein, prohead protease, integrase and excise/RDF.

The *M. tuberculosis* genome is represented by the bar with 1kbp-spaced markers; the ϕ Rv2 is shown in lavender and the ϕ Rv1 element is shaded light blue. Open reading frames (ORFs) are shown as boxes above (transcribed rightwards) and below (transcribed leftwards) the genome and are labeled with the gene assignment and putative function where known or predicted. These elements have open reading frames that share colinearity as well as significant sequence similarity (shown in box at bottom left). The ϕ Rv1 element is inserted into a copy of the degenerate REP13E12 repetitive element (*Rv1587c and c'*), which is located within the biotin operon.

Figure 12



prophage-like element to describe Rv1 and Rv2 (Cole *et al.*, 1998; Hendrix *et al.*, 1999). The second prophage like element, Rv2 (BCG deletion RD11) has much in common with Rv1. The size of Rv2, at 10,982bp, is comparable to Rv1, with most of the difference being accounted for by an IS6110 element insertion present in Rv2 (Cole *et al.*, 1998). There are also several ORFs with similarity at the sequence level, as well as a similar organization of the ORFs themselves (Figure 12) (Hendrix *et al.*, 1999). The GC content of both Rv1 and Rv2 is approximately 66%, which is identical to that of the whole genome at 65.6% (Cole *et al.*, 1998). Thus, unlike typical pathogenicity islands which are characterized by the difference in %GC from that of the host, the prophage-like elements of *M. tuberculosis* H37Rv probably are derived from mycobacteria, or are not a recent transfer event.

These genetic elements are of interest for several reasons. We are interested in whether Rv1 and Rv2 are mobile and encode active recombination systems. The Rv1 element is of particular interest because it is predicted to encode a large serine integrase and a RDF and these recombination systems are relatively new and not well understood (Smith and Thorpe, 2002). Additionally, we would like to know if one of the ORFs in either element encodes a gene that plays some role in the physiology of *M. tuberculosis* or if these elements form viral-like particles and impart gene transfer ability.

Within H37Rv, Rv1 is found within a repeated element, REP13E12, which is located within the biotin operon. At either end of Rv1, there is a 12bp sequence (5'-GGTTGGCCGTGG-3'), while in *M. bovis* BCG, only one copy of this sequence is found (Cole *et al.*, 1998; Mahairas *et al.*, 1996). This 12bp sequence likely represents the common core of the attachment sites *attB*, *attP*, *attL* and *attR*. Thus, the duplicated sequence in *M. tuberculosis* is part of the recombination junctions *attL* and *attR*, while the single copy in *M.*

bovis BCG is the core of the bacterial attachment site, *attB*. The REP element in which Rv1 is located is present in seven non-identical copies in *M. tuberculosis* (Cole *et al.*, 1998); thus, it is possible that there are seven potential sites for Rv1 integration. A second strain of *M. tuberculosis*, that has been sequenced, clinical isolate CDC155, also has a copy of Rv1; however, it is in a different location (NC_002755).

The discreet absence of Rv1 from BCG and the fact that it is in two different locations in CDC1551 and H37Rv suggest that this element is mobile. The Rv1 element may be able to move from one chromosomal location to another if the integrase is functional and if there is more than one attachment site in the *M. tuberculosis* genome. Therefore, it should be possible to create integration proficient vectors, and these plasmids may be able to integrate into more than one site. The ability to insert and excise these elements will also allow for the determination of their role in *M. tuberculosis*. Their presence in *M. tuberculosis* and absence from *M. bovis* BCG also suggests that they may play a role in the physiology of *M. tuberculosis*. It is possible that one of the ORFs encodes a gene that influences the growth rate, metabolism or pathogenicity of the *M. tuberculosis*. It is also possible that either of these elements, Rv1 or Rv2, could form viral-like particles and impart gene transfer ability. In addition, these serine integrase systems, like the one predicted in Rv1, are poorly understood, especially in terms of directionality control. The study of Rv1 can be used to understand how these serine recombination systems function, since Rv1 is predicted to encode both an integrase and a RDF. Thus, a project was begun to investigate the site-specific recombination system of the Rv1 prophage-like element.

I. F. Specific aims

I. F. i. *To determine if ϕ Rv1 encodes a functional integrase*

The discreet absence of Rv1 from BCG and the fact that it is present in different locations in the two sequenced strains of *M. tuberculosis* leads us to believe that Rv1 does in fact encode a functional integrase. To determine if the integrase is functional, plasmids will be made that contain the reconstructed *attP* and *integrase* gene. We will then determine if these vectors are integration-proficient by transforming them into mycobacteria.

I. F. ii. *To determine which sites serve as attBs in M. bovis BCG and M. smegmatis*

After these plasmids are transformed into mycobacteria, colonies will be tested to determine where the plasmid has integrated. Since the sequence of the closely related organism *M. tuberculosis* is known, this can be done in BCG by PCR assay that amplifies the attachment sites or junctions. In *M. smegmatis*, this can be done by sequencing out from the integrated plasmid and blasting the results against the unfinished genome sequence

I. F. iii. *To establish an in vitro integration assay and determine the reaction requirements*

In order to dissect the requirements for integration, an *in vitro* integration assay must be established. The Rv1 integrase can be cloned, expressed in *E. coli* and purified. The substrate sites can also be cloned and used along with the purified integrase to reconstitute the recombination reaction. The conditions of the assay can then be altered to determine the requirements for integration. Substrate sites of different sizes can be utilized to determine the minimum size required for activity, and linear substrates can be used to determine the role of supercoiling in the reaction. Additionally, sequences of the substrate sites can be changed to determine which bases are important for integration.

I. F. iv. *To determine the basis for directionality in $\phi Rv1$*

The Rv1 element is predicted to encode a recombination directionality factor (RDF). This product of this ORF (*Rv1584c*) will be tested for activity based on what is known about other RDFs. In other systems, RDFs inhibit integration and promote excision. To assay for excision, first *Rv1584c* will be expressed in mycobacterial cells with an integrated *attP-integrase* plasmid. The putative Rv1 RDF will also be tested for activity *in vitro*. Integration assays will be done in the absence and presence of Rv1584c to look for inhibition. Excision activity will be assayed by adding Rv1584c to *attL* and *attR* substrates along with integrase.

I. F. v. *To determine how the RDF controls directionality of the recombination reactions*

The RDF encoded by *Rv1584c* will be used to reconstitute an *in vitro* excision reaction as described above. The assay conditions can be altered to determine the excision recombination requirements. To determine the minimum size required for activity, substrate sites of different sizes can be utilized. Linear substrates can also be used to determine the role of supercoiling in the reaction. Other RDFs act, at least in part, at the level of DNA binding, to specific sequences in *attP*. A combination of DNA band shift assays and DNaseI footprinting will be used to determine if the Rv1 RDF protein has DNA binding activity and how it affects integrase binding and recombination.

II. MATERIALS AND METHODS

II. A. Bacterial strains and growth conditions

The high efficiency transformation strain of *M. smegmatis*, mc²155, has been previously described . Laboratory stocks of *Mycobacterium bovis* BCG-Pasteur and BCG-Connaught were originally obtained from Dr. William R. Jacobs Jr. and MedImmune Inc, respectively. Both *M. smegmatis* and *M. bovis* BCG were grown in 7H9 liquid media (Difco) that was supplemented with 0.5% Glycerol, 0.5% Tween 80, and 10% Albumin Dextrose Complex (ADC) for *M. smegmatis* and either 10% ADC or OADC (Oleic acid, Albumin, Dextrose, Catalase) (Difco) for *M. bovis* BCG. *M. smegmatis* was cultivated on Difco 7H10 agar (Difco) and *Mycobacterium bovis* BCG was grown on 7H11 agar (Difco) both agars were supplemented with 0.5% glycerol and 10% ADC. The antibiotics carbenicillin and cyclohexamide were added to all mycobacterial culture media at concentrations of 50 µg/ml and 10µg/ml, respectively. When necessary the following antibiotics were added to liquid or solid media, kanamycin at 8µg/ml, hygromycin at 50µg/ml and tetracycline at 0.5µg/ml. *E. coli* strains DH5α (Sambrook *et al.*, 1989), GC5 (Gene choice Inc.), or the overexpression strain BL21(DE3)pLysS (Novagen) were grown in Luria broth (LB) or LB agar (Difco). When required, *E. coli* media was supplemented with kanamycin at 8µg/ml, hygromycin at 50µg/ml, chloramphenacol at 15µg/ml and tetracycline at 0.5µg/ml.

II. B. Plasmids and DNA

II. B. i. Plasmids constructed by others

II. B. i. a. *pGS67*

Plasmid pGS67 is similar to pMH94 (see below), both constructs have a fragment of mycobacteriophage L5 DNA that contains the *attP* sequence and the *integrase* gene. It is thus an integration-proficient vector in mycobacteria, and it transforms *M. smegmatis* and *M. bovis* BCG among others via a site-specific integration event. Like all plasmids described here, pGS67 contains an ColE1 derived *OriE* for extrachromosomal replication in *E. coli*. Plasmid pGS67 differs from pMH94 in that pGS67 has a gene encoding hygromycin resistance instead of the kanamycin resistance gene in pMH94.

II. B. i. b. *pJL plasmids*

pJL32 is a mycobacterial-*E. coli* shuttle vector that replicates extrachromosomally in both mycobacteria and *E. coli*. It also encodes tetracycline resistance. It is derived from pYUB53 (a gift from the laboratory of William R. Jacobs, Jr.), but it has been changed by removal of a PstI fragment carrying kanamycin resistance. pJL37 is a derivative of pMV261 (see below), it is a mycobacterial-*E. coli* shuttle vector that replicates extrachromosomally in both mycobacteria and *E. coli*. This plasmid is used for expression of genes in mycobacteria from the Hsp60 promoter from *M. bovis*. This pJL37 plasmid has been altered from the pMV261 parent by site directed mutagenesis to include an NdeI site at the transcriptional start.

II. B. i. c. *pMD plasmids*

The plasmid pMD02 is a pUC119-based vector into which a kanamycin cassette resistance from Tn5 has been inserted. It replicates extrachromosomally in *E. coli*, but it is not capable of replication in mycobacteria. Plasmid pMD30 mycobacterial-*E. coli* shuttle vector that

replicates extrachromosomally in both mycobacteria and *E. coli*. It is a pUC119 derivative with a mycobacterial origin of replication derived from pAL5000 and kanamycin resistance.

pMD169 is an integration-proficient vector derived from pMH94 (see below), which also contains a copy of gene 71L5c^{ts32} (Donnelly-Wu *et al.*, 1993).

II. B. i. d. pMH94

Plasmid pMH94 is a mycobacterial integration-proficient vector that has the *attP* and *integrase* sequences from mycobacteriophage L5 as a 2.1kb Sall fragment. This vector is a pUC119 derivative that has both ampicillin and kanamycin resistance (Lee *et al.*, 1991).

II. B. i. e. pMV261 and pMV261::LacZ

pMV261 is also a mycobacterial-*E. coli* shuttle vector that replicates extrachromosomally in both mycobacteria and *E. coli*. It also has the *M. bovis* BCG Hsp60 promoter for expression of genes in mycobacteria (Stover *et al.*, 1991). Plasmid pMV261::LacZ is identical to pMV261 except in this plasmid the *E. coli* β -galactosidase gene, LacZ, is expressed from the *M. bovis* BCG Hsp60 promoter.

II. B. i. f. pYUB415

Plasmid pYUB415 was a gift from the laboratory of W.R. Jacobs Jr. This plasmid is an *E. coli*-mycobacterial shuttle vector that replicates extrachromosomally in both *E. coli* and mycobacteria. This plasmid has hygromycin resistance (Peña *et al.*, 1997).

II. B. i. g. pYW1 and pYW1.11

The plasmid pYW1 is a derivative of pMV261::LacZ. It has been altered such that LacZ is being expressed from the P_{left} promoter from mycobacteriophage L5. This was accomplished by removing the hsp60 promoter with a BstEII digest and ligating the plasmid. This plasmid was

then digested with KpnI and XbaI and a fragment of L5 DNA cut with the same enzymes was inserted that contains the P_{left} promoter. The #11 P_{left} mutant derivative, pYW1.11, was created by random PCR mutagenesis and was inserted into the parent plasmid in the same fashion.

II. B. ii. pLB plasmid constructs

II. B. ii. a. pLB11

Plasmid pLB11 was constructed as a derivative of pBluescript SK-. This plasmid replicates extrachromosomally in *E. coli* and has ampicillin resistance. To construct pLB11, pBluescript was linearized by digestion with EcoRV and phosphatased. The ϕ Rv1 *attP* was obtained by digesting pLB25 (see below) with XmnI and PvuII and recovering the 774bp fragment. This fragment was then ligated into the linearized pBluescript vector to obtain pLB11.

II. B. ii. b. pLB12

Plasmid pLB12 is a pUC119 derivative that contains the ϕ Rv1 *attR*. The fragment with *attR* was amplified from *M. tuberculosis* H37Rv genomic DNA using primers attR-F (5'-GCGCTGAATTCGTTGTCGAGG-3'), and attR-R (5'-AGCGGGAGGTACCCCCAGG-3'). This fragment was cut with EcoRI and KpnI and inserted into pUC119 digested with the same enzymes.

II. B. ii. c. pLB13

Construct pLB13 is a pET21a (Novagen) derivative with the ϕ Rv1 RDF (Rv1584c) under the control of a T7 promoter and LacI. To clone the ϕ Rv1 RDF for overexpression, *Rv1584c* was amplified from H37Rv DNA with Pfu polymerase (Stratagene) using the following primers; Rv1584c-R (5'-AAGGAGTCGCATATGTCGACC-3') and Rv1584c-F (5'-

GTGCCGGATCCTCCGTGG-3'). The amplified product was then digested with Nde I and BamHI following directions supplied by the manufacturer (New England Biolabs), and ligated with T4 ligase (New England Biolabs) into pET21a (Novagen) that was digested with the same enzymes.

II. B. ii. d. pLB14

Plasmid pLB14 is a pET21a derivative plasmid used for expression of ϕ Rv1 integrase in *E. coli*, it has *Rv1586c* under the control of the T7 promoter and LacI. The *integrase* was amplified from *M. tuberculosis* H37Rv DNA with the following primers Rv1586c-R (5'-CCAACCGTGGACATATGAGATACAC-3') and Rv1586c-F (5'-CACGATGTTGTGGATCCGGCTC-3') this amplified fragment was then digested with NdeI and BamHI (sites underlined) and inserted into pET21a digested with the same enzymes.

II. B. ii. e. pLB15

The pLB15 construct is a pUC119 derivative that has the ϕ Rv1 *attL*. This plasmid was made by amplifying the *attL* sequence from H37Rv DNA with the following primers attL-R (5'-AGCCGAACGAGCTCTTCCC-3'), attL-F (5'-GCTCGCGAATTCGGGCGGG-3'). This PCR fragment was then digested with SacI and EcoRI and it was then ligated into pUC119 digested with the same enzymes.

II. B. ii. f. pLB17

Plasmid pLB17 contains the ϕ Rv1 integrase from pLB14 and the reconstructed *attP* sequence that was obtained by amplifying *attL* and *attR* from H37Rv with the following sets of primers; attL-F2 (5'-GGTTGGCCGTGGACTGCTG-3') and attL-R (5'-AGCCGAACGAGCTCTTCCC-3') for *attL*, and attR-R-P (5'-

CCACGGCCAACCGTGGACCTG-3') and attR-F (5'-GCGCTGAATTCGTTGTCGAGG-3') for *attR*. These two fragments were then used in a second PCR with the outside primers (attL-R and attR-F) to create *attP*. The *attP* fragment was digested with SacI (underlined) and AccI (internal to fragment) and the *integrase* was obtained from pLB14 by digestion with AccI and Sall, these two pieces of DNA were then ligated to pMH94 that was digested with Sall and SacI.

II. B. ii. g. pLB18

Construct pLB18 is a hygromycin resistance encoding plasmid that has the ϕ Rv1 integration site #6 from *M. bovis* BCG inserted into an L5-based integration proficient plasmid. To make this plasmid, the *attB* site #6 sequence was amplified from *M. bovis* BCG-Pasteur genomic DNA using primers attL-F and attR-R. This fragment was digested with Asp718 and EcoRI and ligated to the backbone fragment of pGS67 cut with the same enzymes.

II. B. ii. h. pLB24 and pLB25

Plasmid pLB25 is similar to pLB17 in that it is also a ϕ Rv1-based integration proficient vector with the *attP* and *integrase*, however, pLB25 has hygromycin resistance instead of kanamycin resistance. To create this plasmid, pLB17 was digested with HindIII, the 4759bp fragment was recovered and religated to remove the kanamycin resistance marker. The resulting plasmid, pLB24, was then digested with PstI, phosphatased, and ligated to a 1657bp hygromycin resistance containing fragment obtained by PstI digestion of pGS67.

II. B. ii. i. pLB32 and pLB19

pLB32 is a vector containing *M. tuberculosis* H37Rv ORF *Rv1584c* (ϕ Rv1 RDF) under the P_{left}^{Mut11} promoter and kanamycin resistance. This plasmid was created by inserting a PCR amplified fragment containing *Rv1584c* into pJL37 (described above). This fragment was obtained by amplification of H37Rv DNA with primers Rv1584c-F (5'-GTGCCGGATCCTCCGTGG-3')

and Rv1584c-R (5'-AAGGAGTCGCATATGTCGACC-3'). This fragment was digested with HindIII and NdeI and ligated to pJL37 digested with the same enzymes. This plasmid, pLB19 was digested with BstEII and religated to remove the Hsp60 promoter. The resulting plasmid, pLB30, was then digested with XbaI and KpnI and a 216 bp fragment from pYW1.11 digested with the same enzymes, which contains the P_{left}^{Mut11} promoter was inserted.

II. B. ii. j. pLB36

The plasmid pLB36 is a tetracycline resistance encoding construct that has the H37Rv ORF *Rv1584c* (the ϕ Rv1 RDF) under the control of an L5 P_{left} mutant promoter (Mut 11). This plasmid was created by inserting a fragment containing *Rv1584c* under the control of an L5 mutant promoter, P_{left}^{Mut11}, into pJL32, which contains a tetracycline resistance cassette and OriM. This was accomplished by digesting pLB32 with SacII, blunting the 3' overhangings, then purifying the 669bp fragment that contained P_{left}^{Mut11} driving *Rv1584c* expression. This fragment was then inserted into the MscI generated 5172bp piece of pJL32.

II. B. ii. k. pLB44

To create an intramolecular substrate for excision, a plasmid, pLB44, containing *attL* and *attR* in direct orientation was created. To construct such a plasmid, an *in vitro* integration reaction was performed using plasmid pLB17 which has the ϕ Rv1 *attP* (and *integrase*) and a 44bp *attB* (5'- GAAGGTGTTGGTGC GGGGTTGGCCGTGGTCGAGGTGGGGTGGTG-3'). This product of this reaction was then ligated to a 624bp HinCII DNA fragment from pMV261-lacZ.

II. B. ii. l. pLB45

Plasmid pLB45 is a pET21a derivative used for the expression of a C-terminal 6-histidine tagged version of ϕ Rv1 integrase. To create this plasmid, *Rv1586c* was amplified from H37Rv

DNA with Pfu polymerase using the following primers; Rv1586c-R2 (5'-
CCAACCGTGGACCATATGGAGATACAC-3') and LB-intH6 (5'-
TGCTCGAGTCGCCAATTCACCTGC-3'). The amplified product, as well as pET21a vector
were cut with NdeI and XhoI then ligated to create plasmid pLB45.

II. B. ii. m. pLB48-51

To obtain substrates for DNaseI footprinting, attachment sites were amplified with
primers containing restriction sites then inserted into pMOSBlue. The *attB* site was amplified
with attR-r-fp (5'-GCGAATTCACCGAGCTGACCCTGGCC-3')-E and attL-f-fp (5'-
GCGGATCCTCTCACGGATCGTCGTGG-3')-B from plasmid pLB18. The *attP* site was
amplified with attR-f-fp (5'-CGTAAAGCTTGCCACGCCGAGCTGTTCG-3')-H and attL-r-
fp (5'-GTCAGTCGACGTTACCGTCACCAAGTGG-3')-S from plasmid pLB17. Attachment
junction sites were amplified from pLB12 with the primers attR-f-fp and attR-r-fp for *attR* and
from pLB15 with attL-f-fp and attL-r-fp for *attL*. These amplified sites were kinased and cloned
blunt into a linearized vector cut with EcoRV. The pMOSBlue derivative plasmid with *attB* is
pLB48. The pMOSBlue derivative plasmid with *attP* is pLB49. The pMOSBlue derivative
plasmid with *attL* is pLB50 and pLB51 contains *attR*.

II. B. ii. n. pLB54

To create a substrate for intramolecular integration, a plasmid was constructed with *attP*
and *attB* in direct orientation. An intermolecular excision reaction was done using a plasmid
with *attR* (pLB12), and a linear 47bp *attL*. The product of this reaction is a linear DNA fragment
with *attP* and *attB*. This product fragment was then ligated to another DNA fragment containing
kanamycin resistance. The resulting plasmid, pLB54, has kanamycin resistance flanked by *attP*
and *attB*.

II. C. Transformation and electroporation

Plasmids were introduced into *E. coli* strains DH5 α or XL1-Blue by heat shock transformation methods using CaCl₂ competent cells (Sambrook *et al.*, 1989), and strains GC5 and BL21(DE3)pLysS were transformed according to the manufacturers (Gene Choice inc. and Stratagene, respectively). *M. smegmatis* and *M. bovis* BCG cells were made electrocompetent by growing a bacterial culture to mid-log phase (OD₆₀₀ 0.8-1.0), harvesting then washing them three times with 10% glycerol and resuspending them in 10% glycerol. Plasmids were introduced in electrocompetent mycobacterial cells by electroporation of 200 μ l electrocompetent cells with approximately 100ng of DNA into using a Bio-Rad Gene Pulser II set at 1000 Ω , 2.5 kV, 25 μ F at 0°C for *M. smegmatis* and at room temperature for *M. bovis* BCG. Pulsed cells were recovered in 1ml of 7H9 broth with ADC and 0.05% Tween80 at 37° C for ~2 hours for *M. smegmatis* or 4-12 hours for BCG.

II. D. PCR assays

To assay for site integrations, a PCR method was developed using genomic DNA obtained from transformant colonies. Genomic DNA was prepared by picking a colony into 100 μ l of TE (10mM Tris, pH 7.5, 1mM EDTA), which was then vortexed and heated to 95°C for 10 minutes. One microliter was used for PCR with Pfu polymerase in a Perkin Elmer 9600 or 2700 thermal cycler. For *M. bovis* BCG REP13E12 site usage assays, each of the seven sites was analyzed for integration using two methods. The first method used two sets of two primers, one pair for amplification of *attB*, the other for *attL*; for a total of 14 reactions for each transformant. For amplification of *attB*, primers 11 (5'-GCGTGGTGGTTAAAGCTCC-3') and 4 (5'-CCAGCGGATCATGCTGTTCG-3') were used for site #1, yielding a 405bp product,

primers 12 (5'-AGCGGTCGGACTACTCAGC-3') and 19 (5'-GCAGCGAATCATGTTGTACG-3') for site #2 (447bp), 13 (5'-TAGCGAAGAAATCAAGTCCG-3') and 19 for site #3 (485bp), 14 (5'-CAGAGGTTGCGCCTCTCG-3') and 20 (5'-ACAGCGAATCGTCCTCTACG-3') for site #4 (331bp), 15c (5'-AGGCAGCTCGGCAACGTGC-3') and 19 for site #5 (424bp), 1 (5'-ACGTCGATGTGGGGATGTCC-3'), and 4 for site #6 (362bp), and primers 17 (5'-CGTGTGGCTTTGACTGCTGG-3') and 4 for site #7 (338bp). To amplify *attL*, the following primers were used along with primer 2 (5'-CTCTTCCCTCACCTCCAAGG-3'), which lies within the integrated plasmid, to yield products of the size given following the primer number; site #1 primer 11 (353bp), site #2 primer 12 (395bp), site #3 primer 13 (433bp), site #4 primer 14 (276bp), site #5 primer 15c (372bp), site #6 primer 1 (310bp), site #7 primer 11 (289bp). The second method used a set of three primers to amplify either *attB* or *attL* for each of the seven sites, for a total of seven reactions for each transformant.

For PCR assays with *M. smegmatis* strains carrying BCG site #6 *attB* (strain LAB7), four primers were used in each PCR reaction; 1C (5'-GCGGCGATTCTCACGGATCG-3'), 2C (5'-TAGACAGCAGCACGCACAGG-3'), 3C (5'-CGCCCGGATCGTCTCGGCC-3'), and 4C (5'-GCGGATCATGCTGTTCGCC-3'). These four primers could amplify *attB* (primers 1C and 4C– 325bp product), *attP* (primers 2C and 3C - 200bp product), *attL* (primers 2C and 4C - 329bp product), and *attR* (primers 1C and 3C - 240bp product). In assays for integration in *M. smegmatis*, four primers were used in each reaction; 24 (5'-CCTGCTTCGCATGGGTCTCG-3'), 25, (5'-CCGTCTACGACGAACACACC-3') 3' (5'-CGGTTGATGCGCTGACGTCG-3'), and 4' (5'-ATGGTTGATCTCCTGGCGTGG-3'). Primers 3' and 4' amplify *attB* giving a 228bp product. Primers 24, and 25 amplify a 452bp *attP*. Amplification of *attL* with primers 24 and 4'

gives a 131bp product, and primers 3' and 25 yield a 501bp *attR*. Primers for *M. smegmatis* were designed using preliminary sequence data obtained from The Institute for Genomic Research website at <http://www.tigr.org>.

II. E. Sequencing

Sequencing of clones and attachment sites was done following protocols for PE Applied Biosystems dRhodamine terminator chemistry using a Perkin Elmer 9600 or 2700 thermal cycler. Sequencing reactions were then analyzed on an ABI 310 capillary sequencer (ABI).

II. F. Expression and purification of integrase

To obtain partially purified ϕ Rv1 Integrase for *in vitro* studies, a plasmid containing *Rv1586c* under the control of a T7 expression system was utilized. This plasmid, pLB14, was transformed into BL21(DE3)pLysS cells. The transformed cells were grown at 37°C to an OD600 of ~0.6, and expression was induced by the addition of IPTG at a final concentration of 1mM. Cells were then grown at 30°C for 4 hours, and harvested by centrifugation. The resulting cell pellets were resuspended in TED-150 buffer (20mM Tris HCl pH 8.0, 5mM EDTA, 1mM DTT, and 150mM NaCl), and lysed through freeze-thaw methods. DNaseI (Sigma) was added at 20 μ g/ml to remove DNA from the lysate. Ammonium Sulfate at 40% saturation was used to precipitate the protein. The pellet was resuspended in TED-150 buffer and then dialyzed overnight against the same buffer. The preparation was then run over a weak cation-exchange carboxymethyl column (CM) using the BioCad sprint system (Applied Biosystems).

To simplify the purification process, a C-terminal 6-His tagged version of the protein was also created. The expression conditions are identical, but after the cells are harvested, the pellets

were resuspended in Lysis buffer (50mM NaPO₄, 150mM NaCl, and 10mM Imidazole), lysed by a combination of freeze-thaw and sonication, and then purified through Ni-NTA chromatography (Qiagen).

II. G. *In vitro* integration assays

An *in vitro* recombination assay was developed for ϕ Rv1 integration. This assay most often utilized a supercoiled plasmid containing *attP*, pLB17 or pLB25, and a linear paired oligo *attB*. Plasmid DNA was added at 100ng, and the molarity for this amount of DNA is given in each experiment. The linear substrate was supplied in various amounts as described. Integrase was supplied either as a crude extract or as a purified protein. Proteins were diluted in a buffer containing 10mM Tris (pH 7.5), 1mM DTT, and 1mg/ml BSA. Reactions were most often done in a volume of 10 μ l, and 1 μ l of integrase, dilute or undiluted (neat) was added to each reaction. The reaction also contains 0.1 μ g BSA and 1x xis buffer (20mM Tris (pH 7.5), 10mM EDTA, 25mM NaCl, 10mM spermidine, and 1mM DTT). Reactions were incubated overnight at 30°C unless otherwise indicated, then heat killed at 75°C for 10 minutes, and electrophoresed on a 0.8% agarose gel.

II. H. Expression and purification of RDF

The ϕ Rv1 RDF encoded by *Rv1584c* was expressed from an inducible T7 promoter in pLB13, a plasmid derivative of pET21a, in BL21(DE3)pLysS cells. Freeze thaw was used to lyse the cells, then DNaseI (20 μ g/ml), MgCl₂ (10mM) and RNase A (10 μ g/ml) were added and incubated at 37°C for 30 minutes to remove nucleic acids. The lysate was then boiled for 15 minutes, an incubated on ice for 5 minutes. The suspension was then centrifuged to remove

contaminating proteins. The RDF was precipitated with 55% ammonium sulfate, and the pellet containing the RDF was resuspended in buffer containing 10mM Tris pH7.5, 5mM EDTA, and 100mM NaCl, the protein was then dialyzed overnight against 1L of the same buffer. This partially purified protein was then further purified by cation exchange chromatography using a carboxymethyl sepharose (Pharmacia) column eluting with a 100mM to 1M NaCl gradient.

II. I. *In vitro* excision reactions

In vitro excision reactions were done using integrase, RDF and a plasmid containing *attL* and *attR* in direct orientation (pLB40, pLB44, and pLB46). Intermolecular reactions utilize an *attR* plasmid (pLB15) and a linear paired oligo *attL* substrate or an *attL* plasmid (pLB12) and a linear paired oligo *attR* substrate. Integrase and RDF were diluted in a buffer containing 10mM Tris (pH 7.5), 1mM DTT, and 1mg/ml BSA. The reactions were carried out in 10 μ l volumes and 1 μ l of each protein was added to the reaction, unless otherwise indicated. Reactions also contain 0.1 μ g BSA and 1x xis buffer (20mM Tris (pH 7.5), 10mM EDTA, 25mM NaCl, 10mM spermidine, and 1mM DTT). All reactions were incubated overnight at 30 $^{\circ}$ C, unless otherwise indicated. Intermolecular reactions were heat killed at 75 $^{\circ}$ C for 10 minutes following incubation, then electrophoresed on a 0.8% agarose gel. Intramolecular reactions were heat killed, then diluted to 20 μ l for digestions, then treated with protease K prior to running on a 1% agarose gel.

II. J. Radiolabeling DNA

Substrate sites were labeled one of two ways. Plasmid fragments (\geq 150bp) were obtained by digestion with a restriction enzyme that cleaves to generate a 5' overhang that contains one or more threonines (e.g. BamHI G/AATTC). This overhang was filled using Klenow DNA

polymerase (Roche) with α ^{32}P dATP to label the fragment. In the case of footprinting, to obtain a substrate that was labeled at only one end, the other end of the fragment was cut with an enzyme that generates a blunt end or a 3' overhang. Smaller substrate sites ($\leq 80\text{bp}$) were labeled by treating synthesized oligonucleotides with polynucleotide kinase (New England Biolabs) in the presence of γ ^{32}P ATP. These oligos were then annealed to the complement and subjected to PAGE for purification.

II. K. DNA binding reactions

Binding reactions contained a radiolabeled substrate ($\sim 2000\text{-}3000$ cpm), integrase and/or RDF diluted in a buffer containing 10mM Tris (pH 7.5), 1mM DTT, and 1mg/ml BSA. The reaction also contains 0.1 μg BSA and 1x xis buffer (20mM Tris (pH 7.5), 10mM EDTA, 25mM NaCl, 10mM spermidine, and 1mM DTT). The 10 μl reactions were incubated for 30 minutes at 30 $^{\circ}\text{C}$, unless otherwise indicated. Binding reactions were then loaded onto a pre-running 5-10% acrylamide gel at 4 $^{\circ}\text{C}$.

II. L. Genomic DNA preparation

Genomic DNA was prepared from mycobacteria following the GTC protocol from Hatfull & Jacobs (Larsen, 2000). Briefly, late log cells were harvested and treated with freshly prepared chloroform:methanol (3:1) and vortexed. Phenol was then added and the sample was again vortexed. A solution containing 4M guanidine thiocyanate, 0.1M Tris, pH 8.0, 0.5% Sarcosyl, and 1% β -mercaptoethanol is added and mixed. Phenol and chloroform are added to extract the DNA, which is then precipitated with isopropanol, washed and resuspended in TE.

II. M. DNaseI footprinting

Binding reactions were performed as described above except that EDTA was removed from the xis buffer, and binding was done with singly labeled substrate at 10,000 to 15,000 cpm and incubated at room temperature (~25°C). DNaseI (Roche) was diluted to 0.125-0.0157 U/ μ l in a buffer containing 20mM Tris, pH 7.5, 100mM MgCl₂, and 20mM CaCl₂. One microliter of diluted DNaseI was added and incubated with the binding reactions for two minutes at room temperature. The reaction was promptly stopped by the addition of 100 μ l of phenol and vortexing, then 100 μ l of chloroform and 80 μ l of water were added. The mixture was centrifuged and the aqueous layer was transferred to a clean tube. Carrier tRNA, NaOAc and Ethanol were added to precipitate the reactions, which were then resuspended in loading buffer (95% formamide, 0.05% bromophenol blue, and 0.05% xylene cyanol ff) and run on a 6% sequencing gel. Sequencing reactions were run along side these reactions as a marker to determine where the proteins were binding. These reactions were done using the Sequenase version 2.0 kit (USB).

III. INTEGRATION OF ϕ Rv1 IN MYCOBACTERIA

Note: The data presented in this chapter were originally published in:

“Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, ϕ Rv1” Bibb, L.A. and Hatfull, G.F. (2002) *Mol Microbiol* 45: 1515-1526.

III. A. Introduction

The prophage-like element ϕ Rv1 contains an ORF (*Rv1586c*) that is predicted to encode a 469 amino acid integrase of the serine-recombinase type. This protein is part of a growing family of serine recombinases with an N-terminal segment of approximately 140 residues that has similarity to the catalytic domain of transposon resolvases and DNA invertases. The length of these large serine recombinase proteins is variable (444-720 residues), and the C-terminal segments have little similarity (Figure 8A). The integrase of bacteriophage R4 (ORF469) is the closest database match to *Rv1586c*, but it has only 33% identity overall and just 29% in the region downstream of position 140 (Figure 8B). Four mycobacteriophages are known or predicted to encode serine-integrases; Bxz2, Bxb1, U2 and Bethlehem (Mediavilla *et al.*, 2000; Pedulla *et al.*, 2003). These mycobacteriophage integrases are not highly related to *Rv1586c*. The ϕ Rv1 integrase shares only 23% identity overall with Bxz2 integrase and just 20% over the

C-terminal segment. Rv1586c shares 26% identity with Bxb1 Int over the C-terminal segment and 25% overall. The level of identity between Rv1586c and the integrases from Bethlehem or U2 is very similar to Bxb1 since these three phages are highly related and there is >94% identity between the three integrases.

A relatively straightforward first step in the characterization of integrases is the creation of integration-proficient vectors carrying the *integrase* gene and the putative *attP* sequence. Thus we set out to create integration proficient vectors derived from ϕ Rv1, and characterize the integration events. ϕ Rv1 is unique among those serine integrase systems that have been studied in two aspects. First, it is encoded by a prophage-like element, not a fully functional phage. Secondly, the site of integration in *M. tuberculosis* is a repetitive element REP13E12, of which there are seven non-identical copies in *M. tuberculosis*, and there are therefore seven potential sites for ϕ Rv1 integration (Cole *et al.*, 1998).

Through these studies, we show that ϕ Rv1 encodes a functional integration system, which can utilize multiple attachment sites in slow-growing mycobacteria. Plasmids carrying the ϕ Rv1 *integrase* and *attP* site efficiently transform *M. bovis* BCG through site-specific integration into the chromosome. Analysis of these transformants shows that four of the seven REP13E12 repeats in BCG can be utilized as *attB* sites, and that multiple integration events can occur. These ϕ Rv1 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.

III. B. *M. tuberculosis* ϕ Rv1 encodes an active integration system

To determine whether ϕ Rv1 encodes an active integration system, non-replicating plasmids were constructed that carry the putative integrase gene (*Rv1586c*) and the reconstructed *attP* site. These plasmids were tested for their ability to transform *M. bovis* BCG. The initial experiments were done in BCG for two reasons. First, *M. bovis* BCG is a slow-growing mycobacterial species closely related to *M. tuberculosis*; however, it is avirulent and used extensively as a vaccine. Secondly, despite its similarity to *M. tuberculosis*, BCG does not carry a copy of ϕ Rv1.

In order to create these ϕ Rv1 plasmids, the *attP* had to be reconstructed from the presumed *attL* and *attR* sequences present in *Mycobacterium tuberculosis* H37Rv. In H37Rv, we observe that ϕ Rv1 is flanked by two 12bp direct repeats (5'-GGTTGGCCGTGG-3'), while in BCG there is a single copy of this sequence. This 12bp sequence represents the core of the attachment junctions *attL* and *attR*, as well as *attP* and *attB*. The core is the sequence that is common to all recombination substrates and within this sequence strand exchange occurs. The putative *attP* site was generated from *attL* and *attR* using a two step PCR scheme (Figure 13A). A relatively large 371bp fragment of ϕ Rv1 DNA was amplified in order to be certain that all the activity of *attP* was retained. This fragment included 177bp on the right side of the core and 194bp on the left side. The core is quite close to the start of the integrase gene, thus much of this DNA is located within the coding region of integrase (Figure 13B). This fragment was combined with a second containing the *int* gene and was ligated into a vector with kanamycin resistance (Figure 13A). The resulting plasmid, pLB17, contains a 1,621bp *attP-Integrase* fragment, and is capable of replication in *E. coli*, but can not replicate in mycobacteria. Therefore, the only way that kanamycin resistant transformants should be obtained at high

Figure 13. Construction of ϕ Rv1 integration-proficient vector pLB17

A) To create a plasmid with the ϕ Rv1 *attP* and *integrase*, the *attP* had to be reconstructed from the *attL* and *attR* sequences in H37Rv. A two step PCR scheme was used to generate ϕ Rv1 *attP* from H37Rv DNA. In the first reaction, fragments containing *attL* and *attR* were amplified using a primer that annealed to the 12bp core (shown as a gray rectangle) and another primer that was within the ϕ Rv1 sequence (i-gray line). This was followed by a second reaction in which *attL* and *attR* were annealed via the 12 bp common core and amplified by the outside primers (ii). Because of the close proximity of the *Int* ORF (Figure 13B), this *attP* fragment also contains the 5' end of the *integrase* gene (green box). Next (iii) the *attP* fragment was ligated to a second segment of DNA that has the remainder of the putative integrase gene, *Rv1586c* and then (iv) inserted into plasmid vector carrying an *E. coli* origin of replication (OriE) and both ampicillin (Cb^R) and kanamycin (Kan^R) resistance.

B) Sequence of the ϕ Rv1 *attP* and the 5' end of *integrase* is shown here. The 12bp common core shared by all attachment sites, shown by the line in between the top and bottom strands, has some overlap with the inverted repeats in *attP* (horizontal arrows). The symmetry of these repeats extends over 26bp and 18 of these are a match. The 5' end of the leftwards-transcribed *integrase* gene lies within these inverted repeats. The putative catalytic serine is circled.

Figure 13A

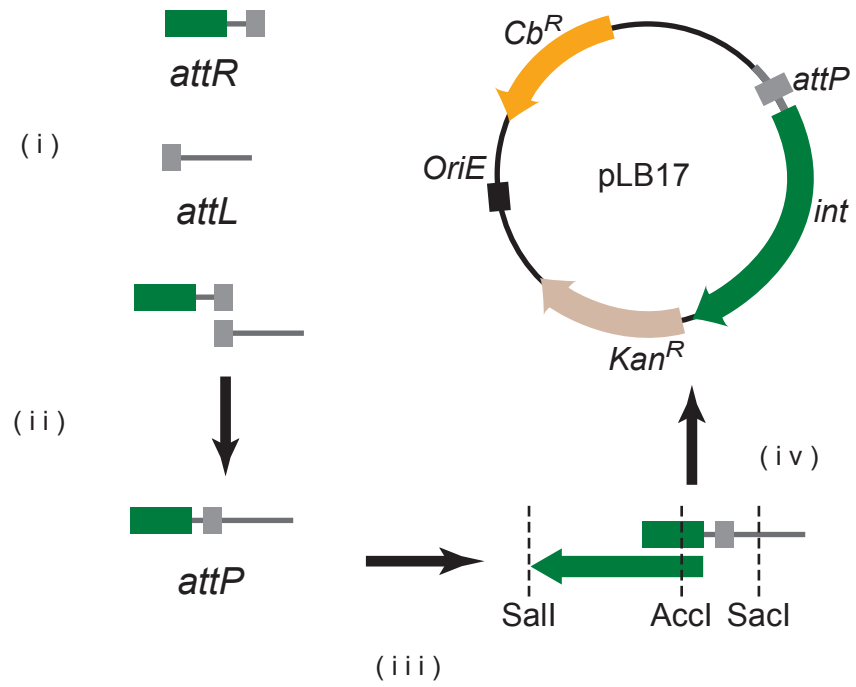


Figure 13B

TGAGATTTCGGAGGTAGACAGCAGCACGCACAGGTGTAGTGTATCTCACAGGTCCACGGTTGGCCGTGGACTGCTGAAGAACATTCCAC
ACTCTAAGCCTCCATCTGTCGTCGTGCGTGTCCACATCACATAGAGTGTCCAGGTGCCAACCGGCACCTGACGACTTCTTGTAAGGTG
(S) I R L Y V A A R V P T T Y R M

frequency in mycobacteria is through a site-specific integration event between *attP* on the plasmid and an *attB* site within the bacterial chromosome.

This plasmid was first transformed into *M. bovis* BCG Connaught and Pasteur strains and selected on kanamycin containing media. A second otherwise identical plasmid, pLB25, was constructed that has hygromycin resistance in the place of kanamycin resistance. We observed that both pLB17 and pLB25 efficiently transform BCG-Pasteur. Transformation with these ϕ Rv1 *attP-integrase* plasmids yielded 5×10^4 transformants/ μ g (Table 1), and this is comparable to the efficiency obtained by transformation with an L5-derived integration vector or with an oriM-containing plasmid (Table 1). BCG-Connaught is also transformed by these plasmids, but at a lower level ($\sim 10^3$ transformants/ μ g). We conclude that the segment of *M. tuberculosis* DNA in plasmids pLB17 and pLB25 contains a functional *attP* site and *integrase* gene, and thus ϕ Rv1 encodes an active integration system.

III. C. ϕ Rv1 *attP-integrase* plasmids utilize multiple BCG *attB* sites

The *M. tuberculosis* REP13E12 element in which ϕ Rv1 is found is present in seven non-identical copies. In order to determine the site into which the ϕ Rv1 *attP-int* plasmid had integrated upon transformation, seven unique primer pairs were designed for each of the sites to amplify either the REP/*attB* site or a junction (*attL*) created by an insertion event. An amplification experiment using each set of the *attB* primers with an untransformed colony of BCG-Connaught yielded a product. This shows that each of these repetitive elements is present in BCG and that the sequence is similar enough to *M. tuberculosis* for the primers to anneal (Figure 14A). However, subsequent sequencing of these products showed that there are some sequence differences (Figure 14B).

Table 1. Transformation efficiency of *M. bovis* BCG Connaught and BCG Pasteur by extrachromosomal and integration proficient plasmids.

Plasmid	Features	Efficiency (cfu μg^{-1})	
		Connaught	Pasteur
pLB17(Kan ^R)/pLB25(Hyg ^R)	ϕ Rv1 <i>attP</i> and <i>integrase</i>	3.1x10 ³	4.1x10 ⁴
pMH94(Kan ^R)/pGS67(Hyg ^R)	L5 <i>attP</i> and <i>integrase</i>	3.9x10 ³	6.1x10 ⁴
pJL37(Kan ^R)/pYUB415(Hyg ^R)	Extrachromosomal (pAL5000 <i>oriM</i>)	1.4x10 ³	5.0 x10 ⁴
No DNA	-	4.0x10 ²	5.0x10 ²

Electroporations were carried out using 100ng of each DNA.

Figure 14. *M. bovis* BCG-Connaught contains seven REP13E12 elements

A) Seven pairs of primers were designed from the sequences of the seven REP13E12 elements in *M. tuberculosis* H37Rv. These were used to amplify the corresponding regions from *M. bovis* BCG Connaught DNA. Lanes 1-7 show the PCR product for each of the REP elements, which are putative *attB* sites for the ϕ Rv1 element.

B) Partial DNA sequence of PCR products in A is shown. The DNA sequence surrounding the 12bp common core was determined. Each gene number of the corresponding REP13E12 ORF in *M. tuberculosis* H37Rv is also indicated. The site that is occupied by ϕ Rv1 in *M. tuberculosis* H37Rv is site #6, and differences from this sequence are shown in gray. Sequence differences between BCG-Connaught and H37Rv are shown as underlined bases. The sequences of two REP13E12 elements in *M. smegmatis* (Msmeg1 and Msmeg2) are also shown. *M. smegmatis* sequences were obtained by BLAST analysis from the preliminary genome sequence available at <http://www.tigr.org>.

Figure 14A

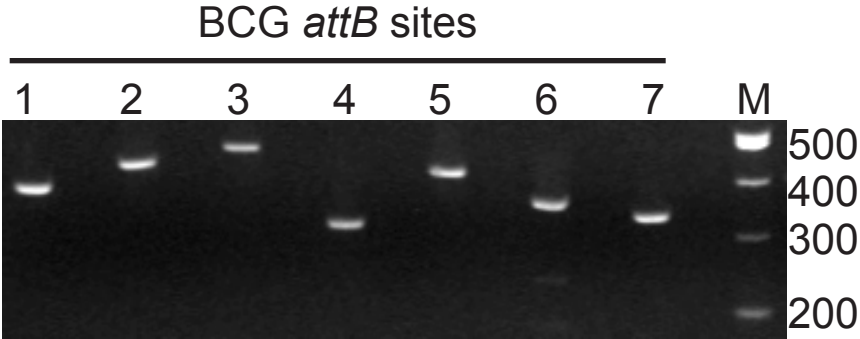


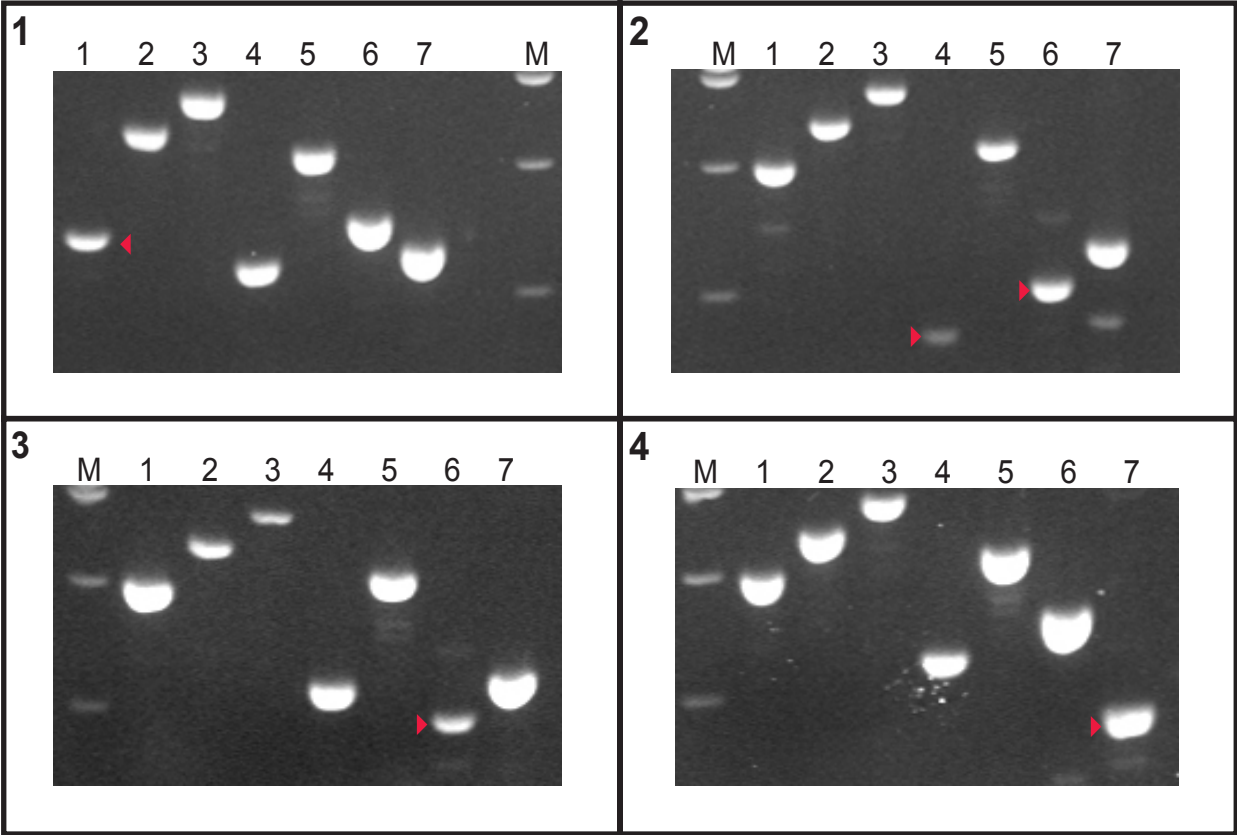
Figure 14B

6 (Rv1587c)	GTGGAAGGTGTTGGTGC <u>CGG</u>	GGTTGGCCGTGG	TCGAGGTGGGGTGGTGGTA
1 (Rv0094c)	GTGGAAGGTGTTGGT <u>CCAC</u>	GGTTGGCCGTGG	TCGAGGTGGGGTGGTGGTA
7 (Rv3466)	GGTGGAGGTGTTGGTGC <u>CGG</u>	GGTTGGCCGTGG	TCGAGGTGGGGTGGTGGTA
4 (Rv1702c)	AAAGTAA <u>CT</u> GTTGGTGC <u>CGG</u>	GGTTGTCCATGG	TCGAGGTGGGCTGGGGAA
2 (Rv1128c)	ATGGAACGTA <u>CTGACTCGA</u>	GGTTGGCCGTGG	TCGAGATGTGGCGGC <u>GGTA</u>
3 (Rv1148c)	GTGGTATCGATTTATTC <u>GC</u>	GGTTGGCCGTGG	TCGAGGTGTGGTGGTGGTA
5 (Rv1945)	GTGGTATCGATTT <u>ATTCG</u> <u>C</u>	GGTTGGCCGTGG	TCGAGGTGTGGTGGTGGTA
Msmeg1	GTGGTAGGTGTTGGTGC <u>CGG</u>	TGTTGGCCGCGG	TCGAGTAGTTGC <u>GGTGGGA</u>
Msmeg2	GTGGTAGGTGTTGGTGC <u>CGG</u>	TGTTGACCACGG	TCGAGAAGGTCGGGTGGGA
		←—————→	
		Common core	

Figure 15. Characterization of ϕ Rv1-mediated integration events in BCG-Connaught

Individual transformants recovered from electroporation of *M. bovis* BCG Connaught with pLB17 DNA were used as substrates in seven separate PCR reactions; one for each of the seven putative *attB* sites. Each panel represents the PCR analysis of a single transformant, and lanes 1-7 represent the amplification of the corresponding sites as shown in figure 14. However, the reactions here differ from those shown in figure 14 in that three primers were used to yield either *attB* or *attL* product depending on whether or not the site is occupied by pLB17. The presence of *attB* or *attL* is evidenced by the size of the product obtained by PCR. 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). The lane marked “M” is the size marker, and the bands are in descending size (from top to bottom); 500bp, 400bp, and 300bp.

Figure 15



Upon electroporation of mycobacteria with pLB17, kanamycin resistant colonies should only be obtained by site-specific integration between *attP* on the plasmid and a site in the bacterial genome. Plasmid pLB17 transformants of *M. bovis* BCG were analyzed in order to determine where the integration event had occurred. This was done for each of the seven sites by PCR, amplifying *attB* or *attL* (Figure 15). A total of 41 transformants were tested, and in each case, an *attB* product was absent and the corresponding *attL* product was present; confirming that transformation was accompanied by site-specific integration of the plasmid. In 39 out of 41 transformants just one integration event had occurred, while the remaining two appeared to have two integrations. One of the double integration events utilized sites #4 and #6, and the other used sites #1 and #6. Of the 39 single integration events; 11 occurred in site #1, 15 were in site #6 and 13 were in site #7. No integration events were seen in site #2, #3, or #5. The use of sites #6 and #7 is not unexpected since these are the sites occupied in *M. tuberculosis* strains H37Rv and CDC1551, respectively (Cole *et al.*, 1998; Fleischmann *et al.*, 2002).

Alignment of the seven REP13E12 elements points to bases in the sequence that are important for integration. These data suggest that an *attB* consists of more than just the 12bp core sequence (Figure 14B). Six of the seven BCG repeats (#1, #2, #3, #5, #6 and #7) have the same 12bp core sequence, but only three of these sites, #1, #6 and #7, are used frequently for integration, and the sequences of these three sites are quite similar (Figure 14B). Site #7 has a deletion which results in three base differences within the 50bp surrounding the core and these are located 14 bases to the left of the common core (Figure 14B). Site #1 also has three differences just to the left of the core (Figure 14B). Since both sites #1 and #7 are used frequently for integration, the changes present in both these sites must be well tolerated. A

Table 2. Transformation efficiency of *Mycobacterium smegmatis* mc²155 by extrachromosomal and integration proficient plasmids.

Plasmid	Features	Efficiency (cfu μg^{-1})
pLB17	ϕ Rv1 <i>attP</i> and <i>integrase</i>	6×10^2
pMD169	L5 <i>attP</i> and <i>integrase</i>	1.6×10^5
pJL37	Extrachromosomal (pAL5000 <i>oriM</i>)	1.4×10^5
pMD02	no <i>OriM</i>	0

Electroporations were carried out using 100ng of each DNA.

single integration event was found in site #4. This sequence has two differences in the core region and changes to the both left and right of the core but the 10bp flanking the core on both sides is identical. Since an integrant was found in site #4, bases six and nine of the core are not essential, but they may contribute to the infrequency of integration events at this site. Those sites in which no integrants were found, #2, #3, and #5, have several differences from those in which integrations occurred (Figure 14B). There are just a few changes to the right of the core, but there are many more on the left, suggesting that these bases may be important for catalysis.

III. D. pLB17 integration in *M. smegmatis*

Mycobacterium smegmatis is a species commonly used in the lab to understand the mycobacteria in general. If ϕ Rv1 *attP-integrase* plasmids transform *M. smegmatis*, these integrating vectors could be useful in research, and may also reveal something about the recombination system. To determine whether ϕ Rv1 based integration-proficient vectors are capable of transforming *M. smegmatis*, the high efficiency transformation strain mc²155 was transformed with the ϕ Rv1 *attP-integrase* plasmid pLB17 (Table 3). Integration-proficient vectors (L5-derived) or *oriM*-containing plasmids efficiently transform this strain of *M. smegmatis*, yielding $\sim 10^5$ transformants/ μ g DNA. However, transformation with plasmid pLB17 under these same conditions yielded only a few hundred colonies (Table 3).

At the time these studies were initiated, it was not known if the REP13E12 elements found in *M. tuberculosis* and *M. bovis* BCG are also present in *M. smegmatis*. To determine where the pLB17 plasmid had integrated, *M. smegmatis* genomic DNA was prepared from two independent transformants, digested with NcoI (which does not cut within pLB17), ligated and recovered in *E. coli*. Sequencing of the attachment junctions in these recombinant plasmids

Table 3. Transformation efficiency of *Mycobacterium smegmatis* strains by extrachromosomal and ϕ Rv1 integration proficient plasmids.

Plasmid	Features	Efficiency (cfu μ g ⁻¹)	
		LAB7	GS67
pLB17	ϕ Rv1 <i>attP</i> and <i>integrase</i>	2.3x10 ⁵	8x10 ⁵
pJL37	Extrachromosomal (pAL5000 <i>oriM</i>)	7.6x10 ⁵	3.1x10 ⁵

Electroporations were carried out using 100ng of each DNA. GS67 is mc²155 containing plasmid pGS67. LAB7 is mc²155 containing plasmid pLB18.

showed that integration had occurred via *attP* and that the same chromosomal site (Msmeg1) was used in both transformants. A comparison of the sequence of the recombinant plasmid with the unfinished genome of *M. smegmatis* (www.tigr.org) revealed that there are at least two segments with similarity to the H37Rv REP13E12 elements. However, these *M. smegmatis* REP13E12-like repeats differ from the *M. tuberculosis* repeats within both the 12bp core and the flanking sequence and have at most 50% identity to any H37Rv element (Figure 14B). PCR analysis of an additional 25 pLB17 transformants showed that the majority (15/24) had utilized one of these sites for integration (Figure 16). Because the genome sequence is still unannotated, we do not yet know if there are more of these 13E12 repetitive elements or if the primers used are able to discriminate between the sites.

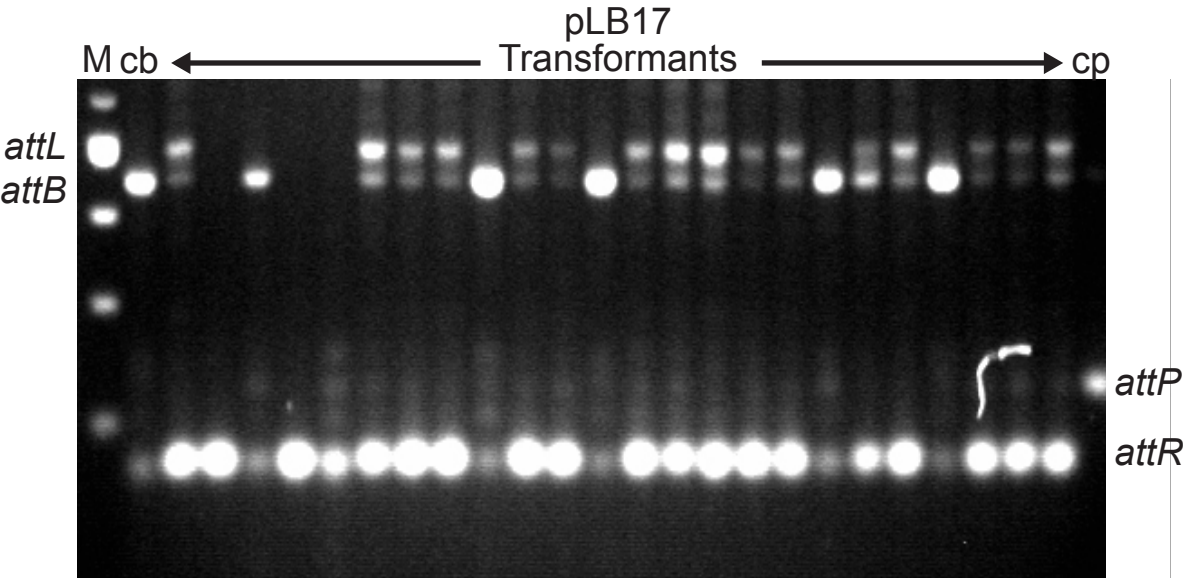
III. E. Efficient transformation of *M. smegmatis* by pLB17 requires a BCG *attB* site

Although the ϕ Rv1 integration-proficient plasmids transform *M. smegmatis*, the efficiency is 300-fold lower level than that of an extrachromosomal plasmid. There are several possible reasons for the reduced efficiency of transformation by pLB17; one of these is the lack of a proper *attB*. To determine if this reduced efficiency of ϕ Rv1 *attP-integrase* plasmid transformation of *M. smegmatis* is the result of inefficient use of the 13E12-like repeats as *attB* sites, we constructed a recombinant strain (LAB7) in which a copy of BCG *attB* site #6 was introduced. This was accomplished by cloning the high efficiency BCG site on a hygromycin-resistant L5-derived integration-proficient plasmid vector (pLB18) and transforming the construct into mc²155 cells. The pLB17 plasmid transforms LAB7 cells with an efficiency similar to that of extrachromosomally replicating vectors, and approximately 300-fold greater than pLB17 transforms *M. smegmatis* mc²155 or a recombinant lacking the BCG *attB* site (Table

Figure 16. Integration of ϕ Rv1 in *M. smegmatis*

A PCR assay was used to analyze pLB17 transformants of *M. smegmatis* mc²155. The assay can amplify *attL*, *attR*, *attB* and *attP*. The presence of *attB* indicates an empty site while amplification of *attL* and *attR* are indicative of integration event in Msmeg1. Twenty-four transformants were analyzed, and 16 out of 24 show amplification of *attL* and *attR*. Control lanes cb and cp show the products derived from *attB* and *attP*, respectively. Markers (M) are in descending size (top to bottom), 600, 500, 400, 300, and 200bp.

Figure 16



4). PCR analysis of several of these transformants demonstrated that the provided BCG site was occupied in every case (Figure 17). These experiments demonstrate that the inefficient transformation of *M. smegmatis* by pLB17 is due to the lack of a suitable *attB* site for integrative recombination.

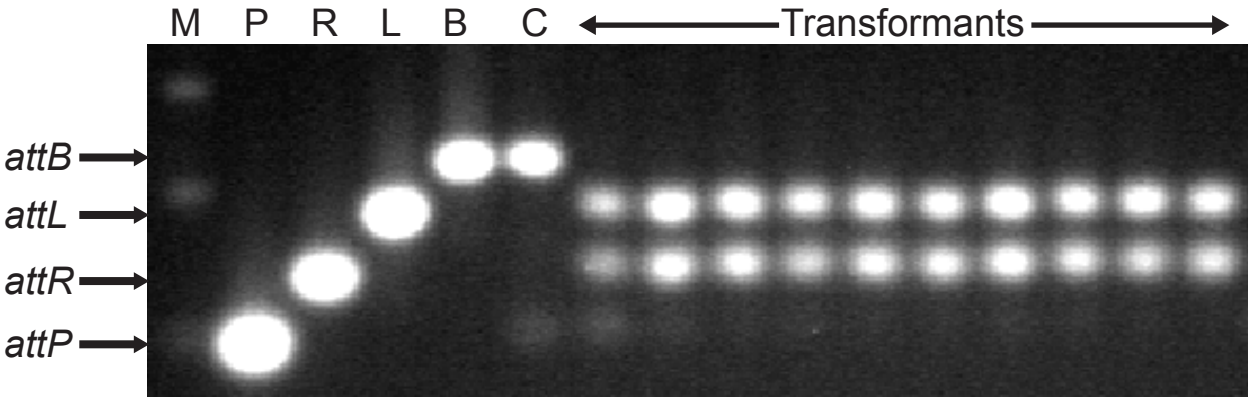
III. F. ϕ Rv1 recombination crossover region

Sequencing of the attachment junctions generated by integration into *M. bovis* BCG site #4 revealed that strand exchange occurred somewhere in the region between the left end of the core and the sequence difference in site #4 at position 6 (Figure 18). When the plasmids created by recovery of ϕ Rv1 *attP-integrase* transformants in *M. smegmatis* were sequenced, we also could determine that crossover had occurred between bases 1 and 8 of the core where two bases differ from the sequence of *attP* (Figure 18). Therefore crossover must occur between bases 1 and 6. Recombinases of the serine family catalyze double-stranded breaks in the substrates that are staggered by 2bp, and generate a 3' overhang. The two bases that make up this overhang are often referred to as the central dinucleotide. If the recombination chemistry of ϕ Rv1 is the same, there are three possible nucleotide pairs that could be the central dinucleotide; bases 2 and 3 (GT), bases 3 and 4 (TT), and bases 4 and 5 (TG). Examination of the sequence of *attP* showed that a pair of inverted repeats flank this region and that the repeats are centered about the 'TG' at bases 4 and 5 of the core. Thus, we believe that this TG dinucleotide represents the 2bp overhang generated by the double-stranded cleavage of the integrase (Figure 18).

Figure 17. Integration of ϕ Rv1 in *M. smegmatis* with BCG #6

Transformants were tested by PCR assay to determine if the ϕ Rv1 plasmid pLB17 had integrated into the *M. bovis* BCG *attB*. In this assay, the amplification of each of the four different attachment sites and junctions results in a product of different size. The first four lanes, B, R, L, and P, show amplification of *attB*, *attR*, *attL*, and *attP*, respectively. All transformants tested showed amplification of *attL* and *attR*, demonstrating that the pLB17 plasmid had integrated into the BCG site. Markers shown in lane M are in base pairs from top to bottom 400, 300, and 200.

Figure 17



III. G. Stable maintenance of plasmid pLB17 in *M. smegmatis*

To determine the stability of the ϕ Rv1-derived integration-proficient plasmid pLB17, mycobacteria containing this plasmid were cultured in the absence of selection. More specifically, two pLB17 transformants of *M. smegmatis* strain LAB7, and two LAB7 transformants containing an extrachromosomal plasmid (pMD30) were cultured in media without kanamycin. After approximately 40 generations of unselected growth in liquid media, no loss of plasmid pLB17 was detected (< 0.5%), whereas 18% of *M. smegmatis* cells lost the extrachromosomal plasmid. Thus, the ϕ Rv1 integrated plasmid is stable in the absence of selection.

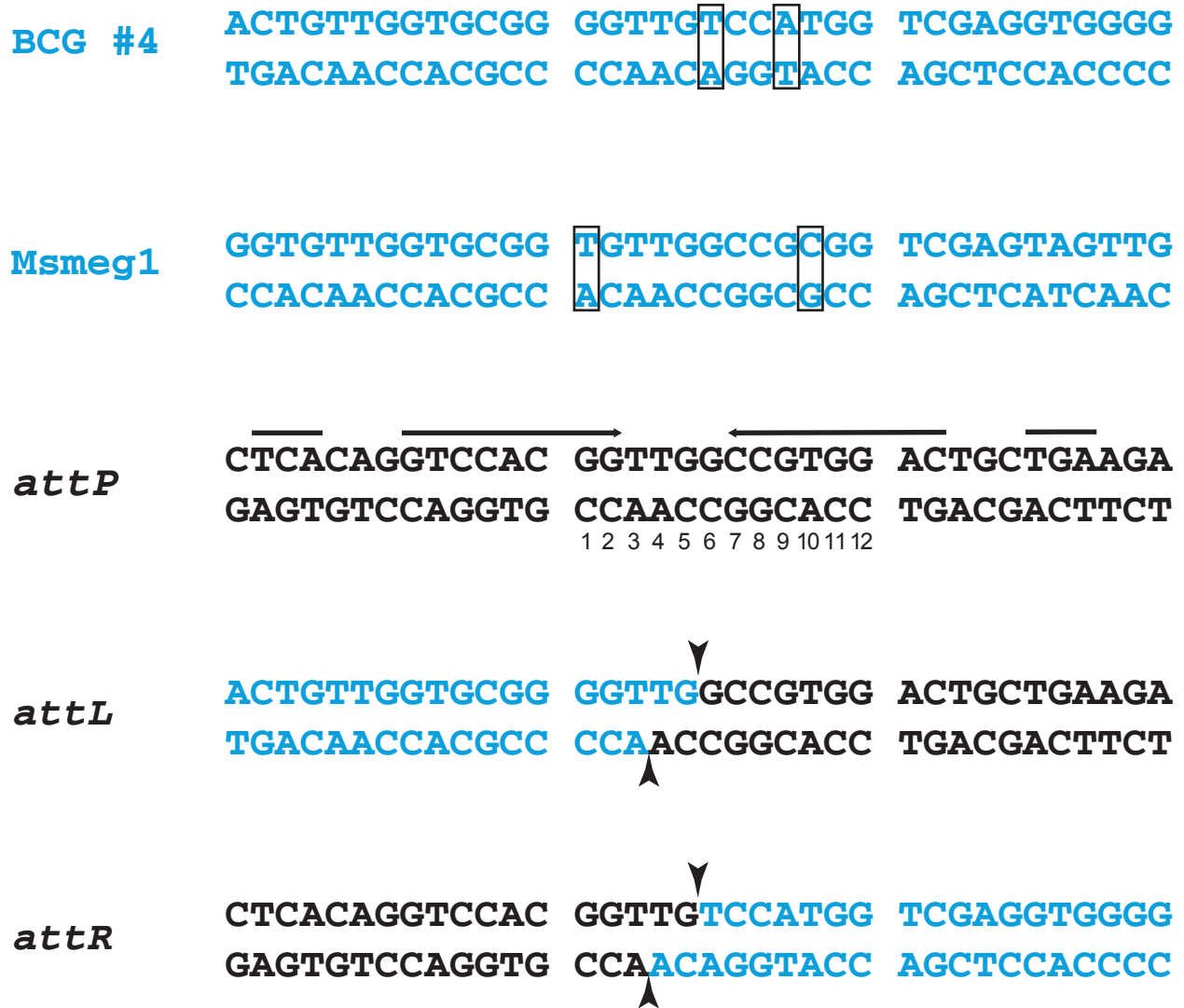
III. H. Discussion

These data show that the ϕ Rv1 prophage-like element of *M. tuberculosis* H37Rv has an active integrative recombination system. The serine integrase encoded by *Rv1586c* catalyzes recombination between a reconstructed *attP* provided on a plasmid and *attB* within the bacterial genome. In *M. bovis* BCG, this recombination reaction is efficient, and integration events were detected at four of the seven potential *attB* sites that are part of the REP13E12 repetitive elements. The sites where ϕ Rv1 is found in *M. tuberculosis* sequenced strains H37Rv and CDC1551 (sites #6 and #7, respectively) are frequently used along with a third site (#1) that is similar in sequence. Multiple integrations were observed in two cases (out of 41). Comparing the sequences of the *attB* sites and whether or not integration events were seen in them gives some information about bases that are important and those that are not. However, all these sites have multiple changes, and site usage may be influenced by other factors outside of sequence as *attB* (see discussion in chapter IV).

Figure 18. Mapping the crossover site for ϕ Rv1 integration

The region of recombination cross-over could be predicted from sequencing of integrant junctions. The sequences of *attB* BCG #4 and Msmeg1 are shown in blue and *attP* is shown in black. The 12bp common core of the BCG-Connaught REP13E12 #4 differs from *attP* at positions 6 and 9. The 12bp common core of the *M. smegmatis* site 1 (Msmeg1) differs from *attP* at positions 1 and 10. Sequence determination of the *attL* and *attR* junction sites following pLB17 transformation shows that strand exchange has taken place to the right of base 1 and to the left of position 6 of the 12bp core. The horizontal arrows represent the inverted repeat sequences found in *attP*. The center of symmetry within *attP* lies between positions 4 and 5 (thin vertical gray line) and the integrase presumably cuts to leave a 3' two-base extension, as shown for transposon resolvases and DNA-invertases. The recombinase most likely cuts 3' to the G in the top strand, and 3' to the A in the bottom strand (vertical arrows in *attL* and *attR*).

Figure 18



The 13E12 related sites in *M. smegmatis* are inefficient *attB* sites for ϕ Rv1 integration, and these two sites are divergent in sequence from those found in *M. bovis* BCG and *M. tuberculosis*. Efficient recombination in *M. smegmatis* can be achieved by providing an efficient site from BCG (site #6), suggesting that the reason for inefficient transformation is the lack of a true *attB* sequence and not that another component of the ϕ Rv1 integration reaction is absent from *M. smegmatis*. Transformation events in this BCG site #6 in *M. smegmatis* strain LAB7 are stable without selection, which suggests that they could be useful vectors for research.

Integration proficient plasmids have proven to be useful genetic tools for mycobacteria, and plasmids based on other site-specific recombination systems have been used to create genetically defined strains (Barsom and Hatfull, 1996; Lee *et al.*, 1991). Vectors based on ϕ Rv1 are stable in *M. smegmatis*, and may be in BCG as well. In addition, these vectors have the possibility to give multiple integrations that can easily be assayed by PCR analysis. A strain that already has one of the attachment sites occupied can also be transformed with a second ϕ Rv1 plasmid to create strains that are more complex.

The sequence of ϕ Rv1 in the second sequenced *M. tuberculosis* strain, CDC1551, is different than in H37Rv, although there are only eight base changes. There are five deletions, one insertion, and two base substitutions, and some of these changes result in alterations to the protein-coding regions (Table 4). The CDC1551 ϕ Rv1 element has non-functional copies of genes corresponding to *Rv1575*, *Rv1576c* and *Rv1585c*.

A strain of virulent *M. bovis* has also been sequenced that has a copy of ϕ Rv1, in this strain, ϕ Rv1 is located within the same 13E12 repeat as it is in H37Rv (Garnier *et al.*, 2003). Like CDC1551, there are some differences in the sequence of ϕ Rv1 in this *M. bovis* isolate (Table 5). There are nine changes, and two of these are conserved. The open reading frames in

Table 4. Sequence differences in ϕ Rv1 between *M. tuberculosis* strains H37Rv and CDC1551

H37Rv ORF	H37Rv protein size	CDC1551 change (position in base pairs)	Consequence	CDC1551 ORF
<i>Rv1573</i>	136	G → C (57)	conserved	MT3562
<i>Rv1575</i>	117	G insertion (243) and G → C (281)	frameshift, 102aa	MT3564
<i>Rv1576c</i>	473	GC deletion (185,186) G deletion (250) G deletion (700)	1 st frameshift, 68aa 2 nd frameshift, 68aa 3 rd frameshift, 68aa	MT3565c
<i>Rv1585c</i>	170	C deletion (393)	frameshift; <i>Rv1584c</i> fusion	MT3572c

Base changes listed are those in the coding strand, and positions listed are within the ORF.

Table 5. Sequence differences in ϕ Rv1 between *M. tuberculosis* H37Rv and *M. bovis*

H37Rv ORF	protein size	<i>M. bovis</i> change (position in base pairs)	Consequence	<i>M. bovis</i> ORF
<i>Rv1573</i>	136	G → C (57)	conserved	Mb1599
<i>Rv1575</i>	117	C deletion (118) and C insertion (241)	frameshift after 36 aa, 42 out of frame, then 36 in frame	Mb1601
<i>Rv1576c</i>	473	C → A (190) G → A (899) G → T (1032)	A → E A → T G → V	Mb1602c
<i>Rv1580c</i>	90	C → T (243)	conserved	Mb1607c
<i>Rv1582c</i>	471	G → C (378)	V → L	Mb1609c
<i>Rv1583c</i>	471	C → T (43)	P → S	Mb1610c

Base changes listed are those in the coding strand, and positions listed are within the ORF.

M. bovis corresponding to *Rv1582c* and *Rv1583c* each have a single amino acid change and the equivalent to *Rv1576c* has three amino acid changes. The *M. bovis* equivalent of *Rv1575* has two frame shifts that results in a protein with the first 36 amino acids in frame, the next 42 are out of frame, and then the final 36 are in frame. Because these ORFs have no database match, it is not known if there are consequences of the bacteria having these non-functional ORFs.

Sequencing of integrants within *attB* sites that have divergent core sequences revealed that strand exchange occurs within bases 2-5 of the 12bp common core. Inverted symmetry in *attP* is centered about the junction between bases four and five. We expect that the chemistry of strand exchange by these large serine integrases is identical to that of the resolvases and invertases. There is experimental evidence in both the ϕ C31 and Bxb1 systems that they cleave to produce a 2bp 3' overhang. If we assume that the mechanism of strand exchange in ϕ Rv1 is identical to that of other serine recombinases; with a concerted dsDNA break staggered by 2bp, we predict that the central dinucleotide is 'TG', bases four and five of the core.

IV. REQUIREMENTS FOR ϕ Rv1 INTEGRATION

IV. A. Introduction

The best understood site-specific integration system found in bacteriophage lambda (λ) has served as a paradigm for the study of other recombination systems. Much of the λ recombination mechanism has been ascertained through *in vitro* studies. The phage λ integrase (λ -Int) catalyzes the integration and excision of the λ genome in *E. coli*. The minimal substrate requirements for integration by λ -Int, are a small simple *attB* (21bp) and a relatively large complex *attP* (234bp) (Figure 6) (Mizuuchi and Mizuuchi, 1980, 1985). The *attB* has a pair of inverted repeat sequences, which serve as recognition sites for Int, that flank a short asymmetric region within which strand exchange occurs (Mizuuchi *et al.*, 1981). The *attP* site has a sequence that consists of the asymmetric region, referred to as the common core, and flanking inverted repeats called core type binding sites that are nearly identical to those found in *attB*. In addition, *attP* has another 200+ bp that includes sites where another domain of integrase binds as well as binding sites for other protein factors. Lambda integrase requires a supercoiled *attP* substrate, and a host-derived factor called IHF (Integration Host Factor), in order to carry out integrative recombination (Mizuuchi *et al.*, 1978; Nash and Robertson, 1981). Other tyrosine

integrases that have been studied act on similar substrates in terms of their relative size and complexity.

The relatively recently discovered serine integrases appear to accomplish integration differently. In all systems that have been studied in some detail, both *attB* and *attP* are small and simple (Figure 10B). These integration substrates consist of a pair of inverted repeats that flank a short asymmetric sequence that is common to both *attB* and *attP*, and it is thus referred to as the common core, and unlike the substrates of the tyrosine integrases, the two substrates used for integration by the serine integrases have distinct inverted repeats, and the repeats that flank *attB* differ from those that flank *attP* (Figure 10B) (Smith and Thorpe, 2002). In serine systems, supercoiling does not appear to play a role, as linear and supercoiled substrates work equally well, and integration occurs efficiently in the absence of additional factors (Kim *et al.*, 2003; Thorpe and Smith, 1998).

We have shown that the prophage-like element ϕ Rv1 encodes a serine integrase, which catalyzes integrative recombination of a plasmid containing *attP* and *integrase* in both *M. bovis* BCG and *M. smegmatis*. In this chapter, an *in vitro* integration reaction is established and the requirements for the assay are examined. Both the *attP* and *attB* minimal substrates are small and apparently simple. An excess of *attB* is required for efficient intermolecular recombination, however, efficient recombination occurs on an intramolecular substrate that has a single copy of *attP* and *attB* in direct orientation. Five out of seven REP13E12 sites from BCG act as *attB* substrates *in vitro* and yield recombinant product. *M. smegmatis* sites are inefficient *attBs* in integration reactions. Supercoiling is not required for integration, and no additional factors are required.

IV. B. Expression and purification of ϕ Rv1 integrase (gpRv1586c)

In vitro reactions have been established for many recombination systems, and have been invaluable in determining the mechanism of catalysis. In order to dissect the ϕ Rv1 integration reaction *in vitro*, a scheme was developed to express a recombinant form of ϕ Rv1 Integrase and purify it. A plasmid containing *integrase* under the control of T7 polymerase and LacI (a pET derivative) was grown in an *E. coli* over expression strain. A native version was used at first in order to establish the reaction. The protein was purified through salting out and ion exchange chromatography (Figure 19A). Subsequently a C-terminal 6-His tagged version was cloned in order to simplify the purification scheme (Figure 19B). This his-tagged version behaves equivalently to its native counterpart in all experiments (data not shown).

IV. C. *In vitro* integration assay establishment

An *in vitro* integration assay was established for ϕ Rv1 following what was previously done for mycobacteriophage L5; *attP* was provided as a supercoiled plasmid, and *attB* was supplied as a 472bp restriction fragment (Figure 20). For the *attB* substrate, site #6 from *M. bovis* BCG was utilized for two reasons. First, it is the site occupied in *M. tuberculosis* H37Rv, and secondly, the highest frequency of integration events were observed at this site in BCG (Table 3). Integration experiments with equal-molar *attP* and *attB* substrates showed a very low level of substrate conversion to recombinant product using either crude or partially purified integrase extract, even after long incubation (≥ 16 hours). The reason for this was unclear, although one possibility was that one of the components of the reaction is limiting. To test this

Figure 19. Expression and purification of ϕ Rv1 integrase

A) The ϕ Rv1 integrase was expressed in its native form in *E. coli* using a pET derivative plasmid (pLB14). The lane marked U shows the total cell lysate before the addition of the inducer IPTG. The induced sample (I) shows the appearance of a band just above the 45kDa marker that corresponds to the predicted size of integrase, 52kDa. A lysate was prepared from induced cells as described in the materials and methods. This lysate was the load (L) for a CM ion exchange column, and a subset of the collected fractions are shown. Lane M contains molecular weight markers with the sizes in kDa shown at the right.

B) In this panel, a C-terminal hexa-histidine tagged version of integrase has been cloned, and expressed in a similar fashion to the native protein in A. This crude lysate (L) was loaded on a Nickel-NTA agarose column. The wash (W) and flow-through (FT) are shown as well as a number of fractions. Several of these fractions contain a 52kDa band corresponding to integrase. Molecular weight markers are in lane M with the sizes in kDa shown at the right.

Figure 19 A&B

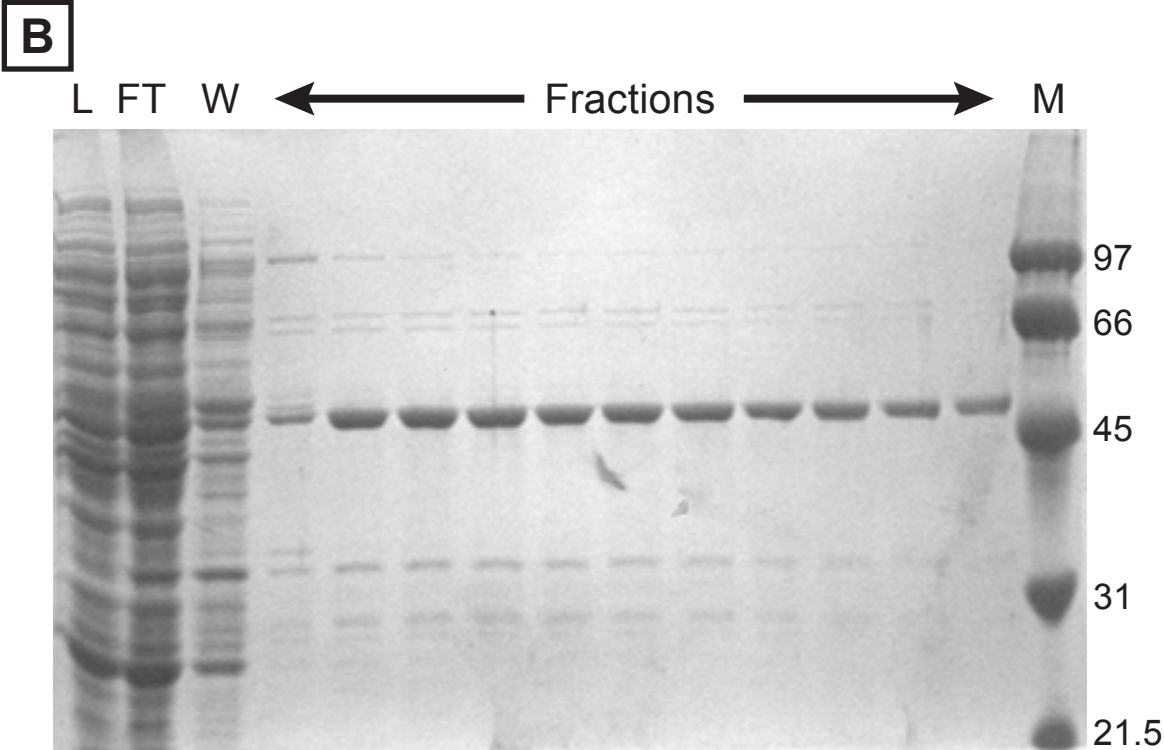
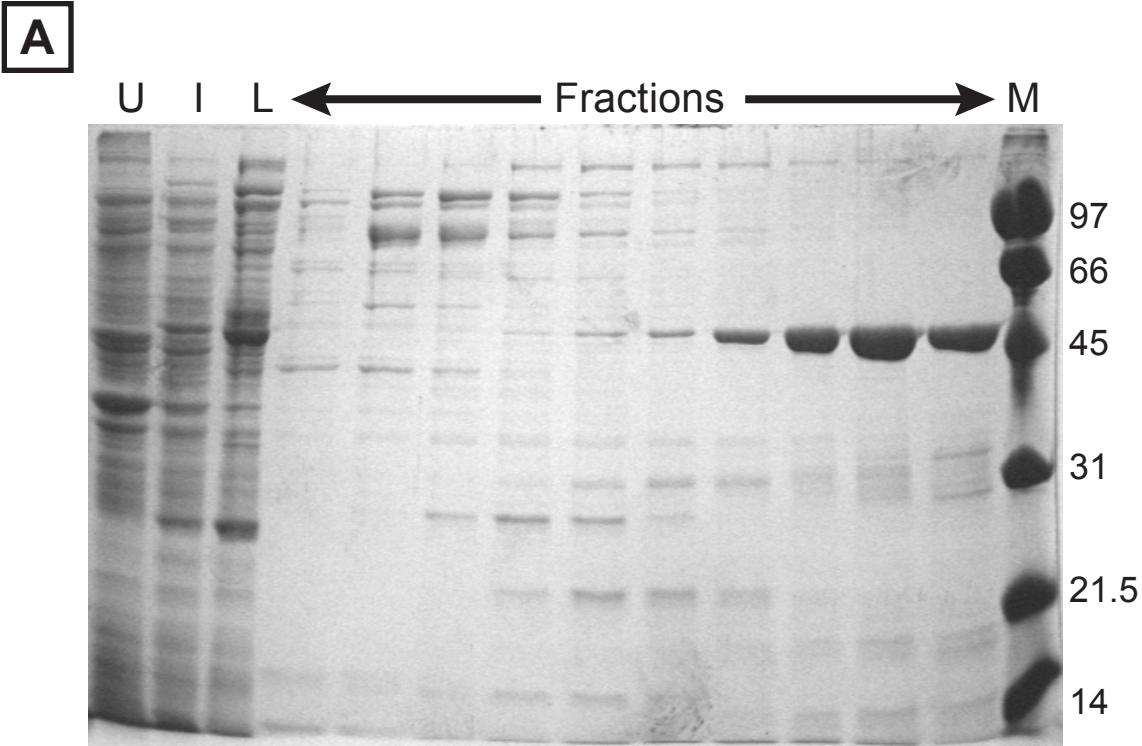
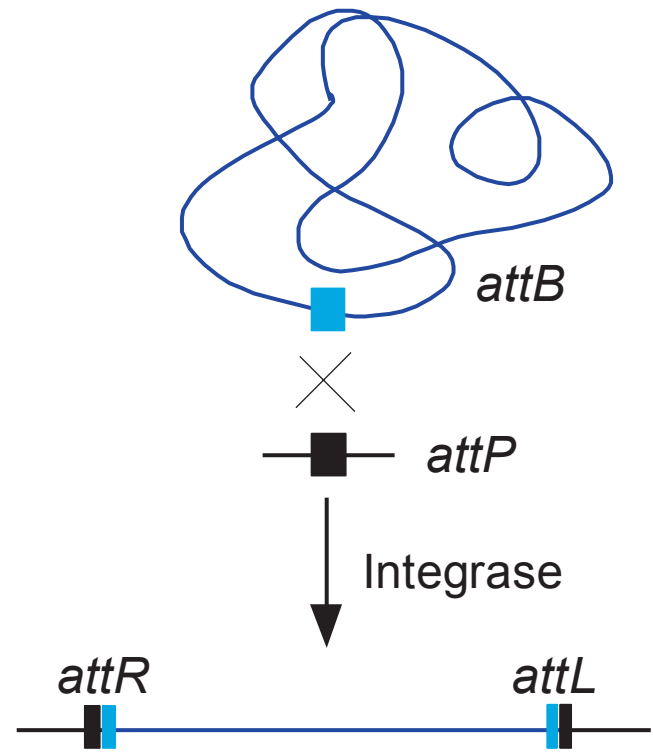
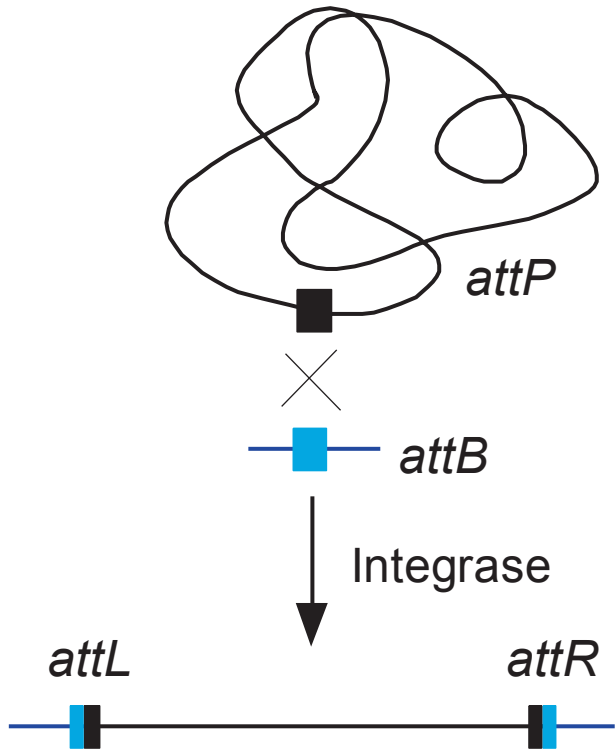


Figure 20. *In vitro* integration assays

Schematic representations of *in vitro* integration assays. Integration assays were done *in vitro* using one attachment site on a plasmid, and the other attachment site provided as a small linear fragment. These two DNA substrates are incubated with integrase. On the left is shown a reaction using a supercoiled *attP* plasmid (shown as the black coiled circle) and small linear *attB* (shown in blue). This is the reaction most commonly used. The substrates are incubated with integrase, which catalyzes recombination and gives rise to a single linear product containing *attL* and *attR* in direct orientation (head-to-tail). On the right, a reciprocal reaction is shown that uses a supercoiled *attB* plasmid (blue coiled circle), and a small linear *attP* (shown in black). This version of the reaction also gives rise to a linear product with *attL* and *attR*.

Figure 20

100



possibility, each component was titrated into the reaction. In order to add higher amounts of *attB*, 50 base complementary *attB* oligos were annealed and used as substrate in the reactions. The size of 50bp was chosen based on the small size requirements known for *attB* in other large serine integrase systems (Figure 10B). The addition of more *attB* to the reaction increased the amount of product formed. Experiments showed that efficient integrative recombination could be achieved by increasing the amount of *attB* in the reaction to at least a 30-fold molar excess over *attP* (Figure 21). Although the reaction is efficient under these conditions, it proceeds slowly, requiring five hours or more in order for the reaction to go to >90% completion (Figure 22). This slow reaction is not alleviated by increasing the size of *attB* to 60bp or 472 bp, or increasing the amount of Integrase (data not shown, Figure 23).

IV. D. *attP* and *attB* site requirements

In other serine integration systems, the minimal requirements for both substrates are rather short sequences (<60bp). To determine the minimum size of *attB* required for full integration activity, a series of paired oligo substrates were constructed ranging from 50 to 36 base pairs by deleting in 2 bp increments (one base pair from both the 5' and 3' end). These oligo pairs were annealed and used in integration assays where the amount of *attB* was varied. For these experiments, the minimum size required for activity was determined to be 40 base pairs, as reactions - at all tested concentrations - using an *attB* substrate of this size yielded product in amounts similar to larger *attB* sequences (Figure 24). Using a 38bp *attB* in an *in vitro* reaction yields a small amount of product, but only at the highest concentration tested, 900ng, which is a 1000-fold molar excess over *attP* (Figure 24).

Figure 21. A molar excess of *attB* is required for efficient intermolecular integration

In vitro reactions were performed with ϕ Rv1 *attP* as a supercoiled plasmid and *attB* provided as a linear 50bp substrate. This agarose gel shows supercoiled *attP* plasmid DNA (lower band in reaction lanes) being converted to linear product (indicated by the arrow). The linear 50bp *attB* substrate is not visible in this gel system due to its small size. Integrase is present at 3 pmoles. The *attP* is added at 100ng, which is 0.026 pmoles. The *attB* substrate was added at different amounts in each reaction. The amount of *attB* in the reaction shown in lane 1 is 7.2 pmoles, which is approximately a 300-fold (300x) molar excess over *attP*. There are 2.4 pmoles of *attB* substrate (100x) in lane 2, 0.72 pmoles (30x) in 3, 0.24 pmoles (10x) in 4, 0.072 pmoles (3x) in lane 5, 0.024 pmoles (1x) in 6, and 0.0072 pmoles (0.3x) in lane 7. The final lane (8) before the marker (M) has no *attB* added. These reactions were incubated for approximately 16 hours. Marker sizes shown to the right of the gel are given in kilobases (kb).

Figure 21

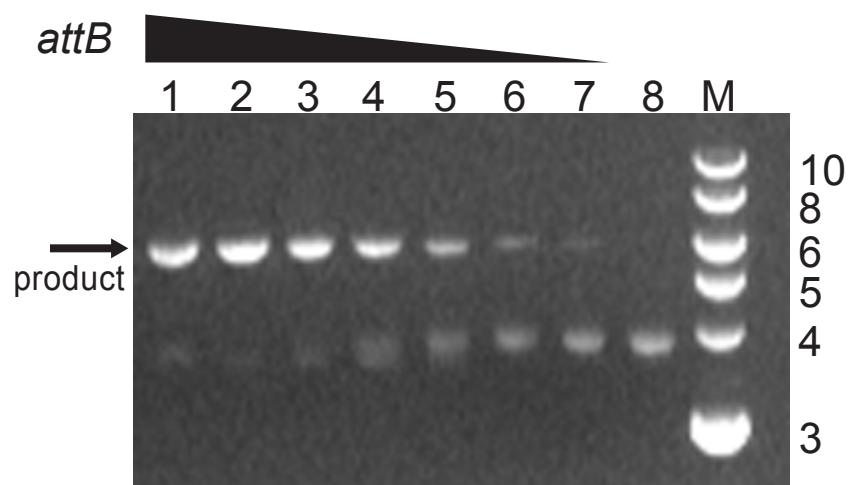


Figure 22. Time course *in vitro* integration reaction

In vitro integration reactions were done using supercoiled plasmid *attP* and linear 50bp *attB* at two different concentrations, with samples taken at different time points. This agarose gel shows two sets of reactions; on the left half of the gel the *attB* concentration is 0.072 μ M, which is three times the molar concentration of *attP* in the reaction (0.026 μ M). The right half of the gel has 0.72 μ M *attB*, which is a 30-fold molar excess over *attP*. The numbers given above each lane is the length of time each reaction was incubated in hours. Over time, the *attP* substrate, which is the lowest and highest bands on the gel representing the supercoiled and nicked circular forms, respectively, is converted to linear product, which is indicated by the arrow. The markers shown in lane M are in descending size (top to bottom) 10, 8, 6, 5, and 4 kb.

Figure 22

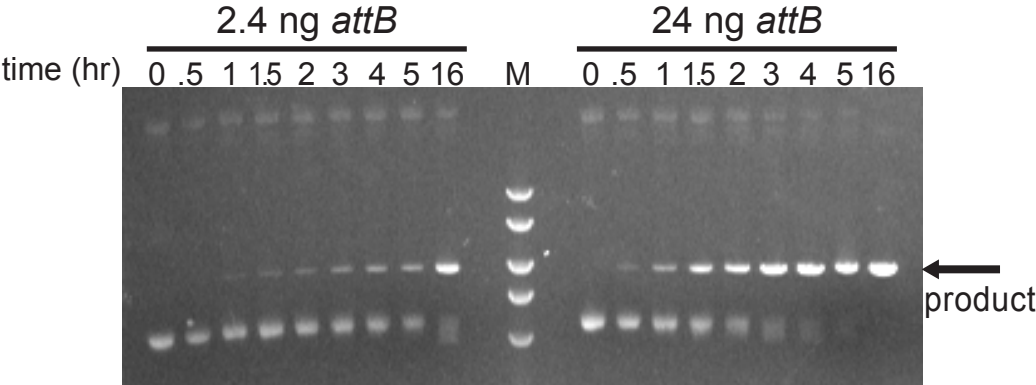


Figure 23. Integrase titration integration assay

Purified integrase was titrated in an *in vitro* integration assay using a supercoiled *attP* plasmid and linear 50bp *attB* at three different concentrations. There are three sets of reactions with five reactions in each, the amount of integrase is as follows; 1-90 fmoles, 2- 0.3 pmoles, 3- 0.9 pmoles, 4- 3 pmoles, 5- 9 pmoles. The set of reactions on the left part of the gel was done using 2.4 pmoles of *attB*, which is a 100-fold excess over *attP*, which is added at 0.026 μ M. The middle panel shows reactions that were completed with 0.72 μ M *attB* (30-fold excess over *attP*) and the set of reactions on the right side were done using 0.24 pmoles of *attB* (10-fold excess over *attP*). The first lane of the gel (-) has no integrase added. The substrate *attP* plasmid is converted to linear product as indicated by the arrow. These reactions were incubated for 5 hours. The sizes in the marker lane M are given on the right in kb.

Figure 23

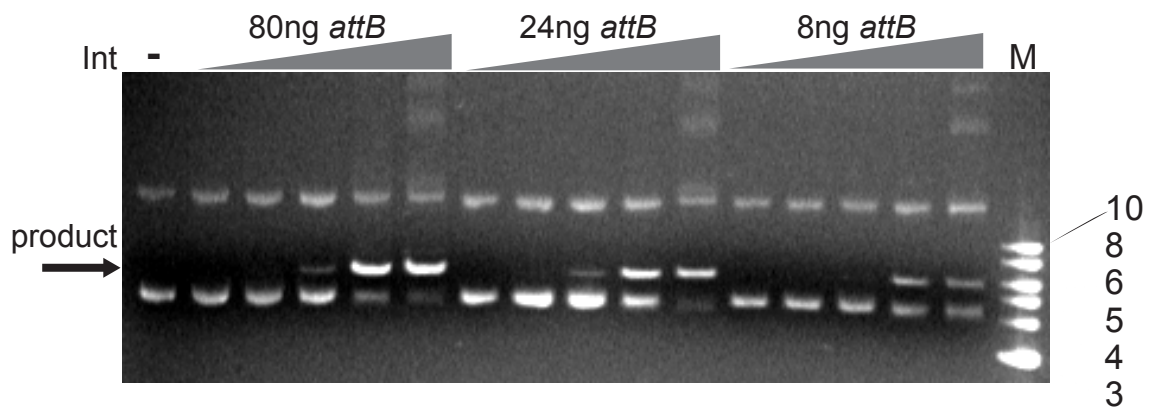


Figure 24. Minimum size of *attB* in integration

A) In vitro integration assays were performed with 3 pmoles of integrase, supercoiled 0.026 pmoles of *attP* plasmid and paired oligo *attB* substrates of various sizes as shown above each set of reactions. The amount of *attB* in each set of five is as follows: Lane 1 has 900ng of *attB*, lane 2 has 100ng, lane 3 has 10ng, lane 4 has 1ng, and lane 5 has 0.1ng. Because of the difference in size of the substrates, the concentration of *attB* is variable, and thus so is the molar ratio of *attB* to *attP*. In lane one, the ratio of *attB* to *attP* is between 1100:1 and 1500:1. In lane 2, it is 125-175:1. In lane 3, there is 13- to 18-fold more *attB* than *attP*. Lane four has *attB* at 1-2 times the amount of *attP*. In the fifth lane, the ratio of *attB* to *attP* is 0.1:1 to 0.2:1. The first lane of the gel, marked -, has no *attB* added. The position of the linear product is indicated by the arrow. Markers (M) are shown in kb.

B) Sequence of *attB* substrates used in A. The sequence of the common core is underlined and the predicted central dinucleotide is shown in red. The horizontal arrows show the regions of inverted symmetry in *attB*.

Figure 24A

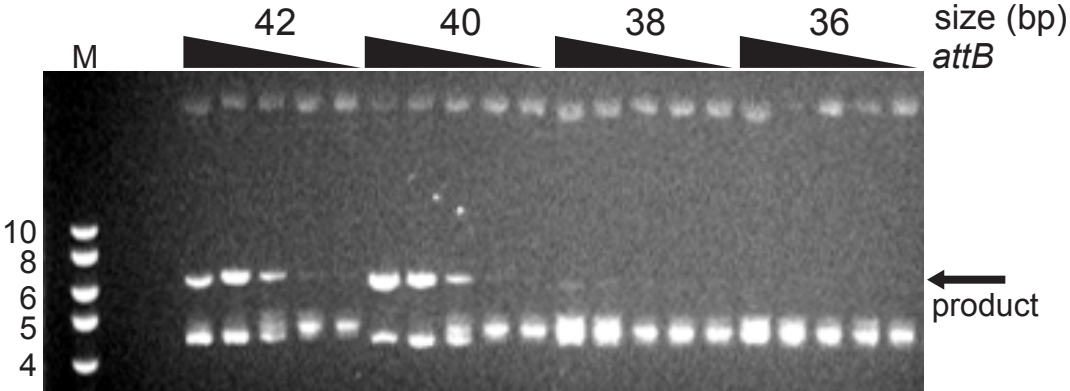
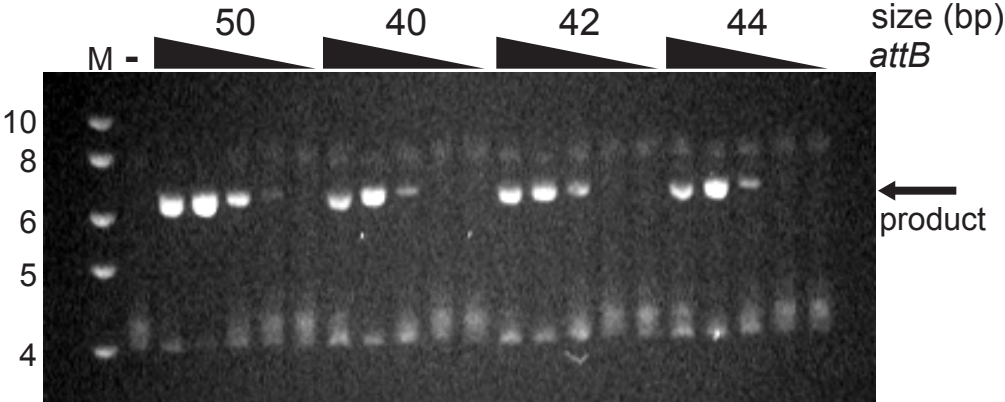


Figure 24B

50 TGGTGGAAAGGTGTTGGTGC GGGGT**TGG**CCGTGGTTCGAGGTGGGGTGGTGG
44 TGGAAGGTGTTGGTGC GGGGT**TGG**CCGTGGTTCGAGGTGGGGTGG
42 GGAAGGTGTTGGTGC GGGGT**TGG**CCGTGGTTCGAGGTGGGGTG
40 GAAGGTGTTGGTGC GGGGT**TGG**CCGTGGTTCGAGGTGGGGT
38 AAGGTGTTGGTGC GGGGT**TGG**CCGTGGTTCGAGGTGGGG
36 AGGTGTTGGTGC GGGGT**TGG**CCGTGGTTCGAGGTGGG

A similar set of experiments was done to determine the minimum size required for *attP*. In this case, *attB* was provided as a supercoiled plasmid and paired *attP* oligos were used. This version of the integration reaction does not go to completion at any tested concentration and is therefore not as efficient as reactions using supercoiled *attP* and linear *attB* (Figure 25A). This reaction is also more sensitive to excess *attP* and recombination is less efficient at higher concentrations of *attP*. A ten-fold excess of *attP* in the reaction yields the most product. This is different from the supercoiled *attP* and linear *attB* reaction, where a 100-fold excess of *attB* yields the most product, and a 300 to 1000-fold excess of *attB* only slightly reduces the amount of product formed. The minimum size of *attP* that is required to give full activity in the *in vitro* integration assay is 52bp (Figure 25B). In this experiment, substrates from 54 to 48bp were used. Larger substrates, even 68bp, did not yield more product than a 52bp *attP* substrate. Reactions with a 50bp *attP* yielded a small amount of product, and 48bp yielded no product in these experiments. Thus, in regards to substrate size requirements, ϕ Rv1 is similar to other serine integrases that have been studied, with the minimal sizes of *attP* and *attB* at 52 and 40 base pairs, respectively (Figure 26).

IV. E. *In vitro* efficiency of *M. bovis* BCG REP13E12 *attB* sites

In all initial *in vitro* assays, site #6 from *M. bovis* BCG was used as *attB*. Although it is the site most frequently used in BCG, other factors may effect site selection *in vivo*. It is therefore possible that this sequence does not represent the best *attB* sequence. In fact to obtain nearly complete conversion of substrate to product in recombination assays, an excess of *attB* is required and the reaction must be incubated for a long time. It is possible that another sequence may be a better *attB* substrate, and reactions with a different *attB* may go to completion at a

Figure 25. Minimal *attP* determination

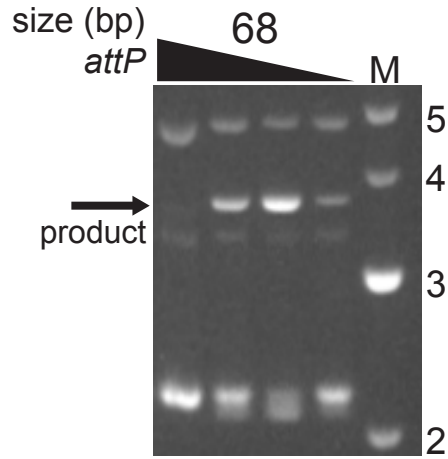
A) In vitro recombination reactions were done with supercoiled *attB* and linear 68bp *attP* substrate. The *attB* plasmid (pLB16) is held constant at 0.04 μ M, while the *attP* is added in different amounts in these reactions. The reaction shown in lane 1 has 900ng of *attP*, which is 20 μ M, and is a 500-fold molar excess over *attB*. Lane 2 has 2 μ M *attP*, or 50-fold molar excess. Lane 3 has 0.2 μ M *attP*, which is a 5-fold excess over *attB*, and lane 5 has *attP* at one-half the amount of *attB*, of 0.02 μ M. The product is indicated by the arrow at the left. DNA size marker (M) is given in kilobases on the right.

B) Integration reactions were done using linear substrate *attPs* of various sizes (shown above each set of reactions in bp) *in vitro* with an *attB* plasmid. Lane 1 of each set has 130 to 150ng of *attP* depending on the size, but in each case the amount was adjusted to a 100-fold molar excess over *attB*. Lane 2 has 10-fold excess of *attP*, lane 3 has an equal-molar amount of *attP* and *attB*, and lane 4 has *attP* at one-tenth the amount of *attB* in the reaction. The lane marked - has no *attP* added. The arrow shows the migration of the linear product. The markers in lane M are given in kilobases on the left.

C) The sequence of *attP* substrates used in A and B. The central dinucleotide is shown in red, and the horizontal arrows show the position of the inverted repeats.

Figure 25 A&B

A



B

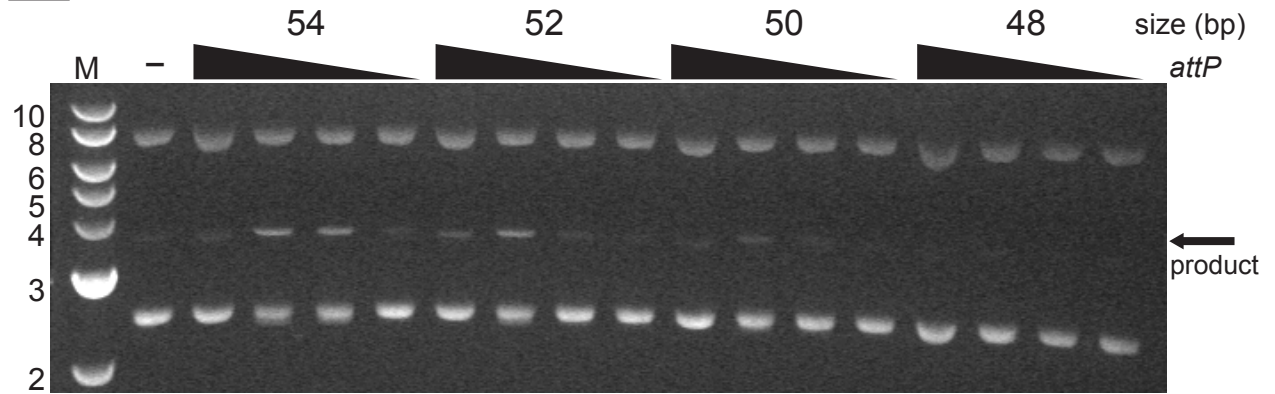


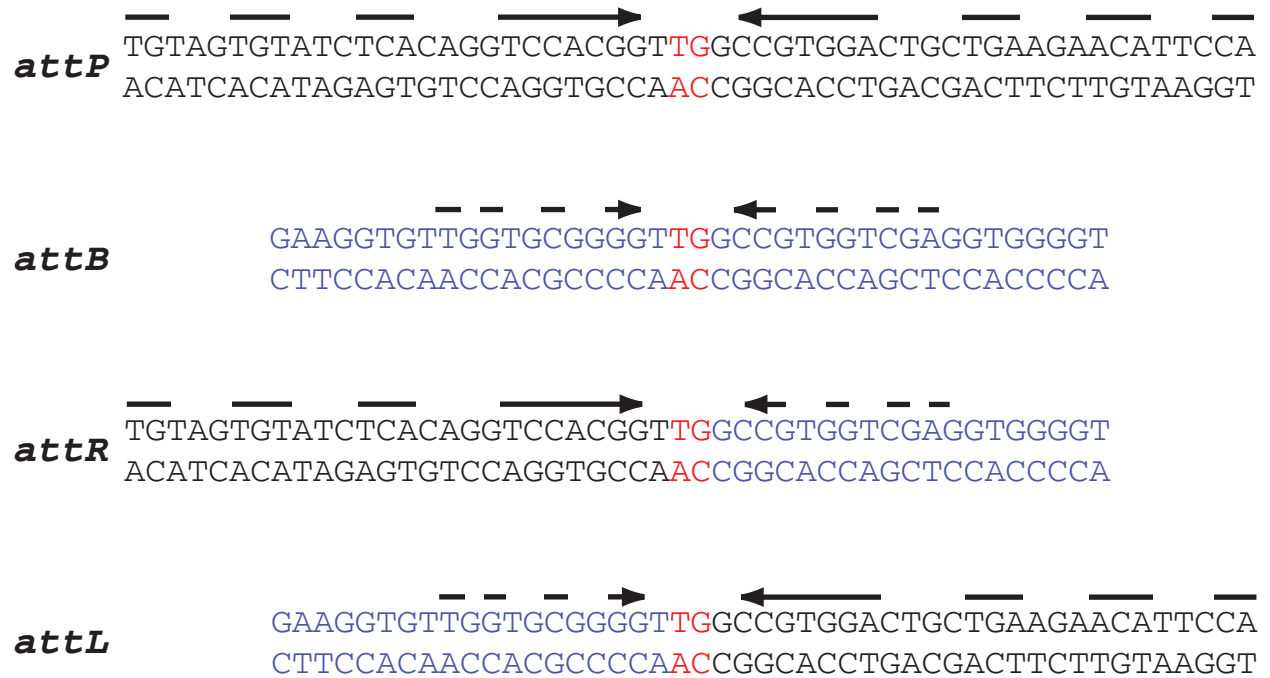
Figure 25C

68 ACAGGTGTAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGACTGCTGAAGAACATTCCACGCCAGGAG
54 TG TAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGACTGCTGAAGAACATTCCA
52 GTAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGACTGCTGAAGAACATTCC
50 TAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGACTGCTGAAGAACATTC
48 AGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGACTGCTGAAGAACATT

Figure 26. Minimal ϕ Rv1 attachment and junction sites

Minimal *attP* and *attB* sites that were determined by *in vitro* experimentation are shown here as well as the hybrid junctions *attL* and *attR* that are predicted as the result of integrative reactions. The *attP* substrate is shown in black, and *attB* is shown in blue and the sequences derived from each of these sites in *attL* and *attR* are colored accordingly. The predicted central dinucleotide is shown in red. Inverted repeat sequences are indicated by the horizontal arrows.

Figure 26



lower concentration of *attB* and possibly have a faster rate of reaction. Thus, a region of each of the REP13E12 elements surrounding the 12bp common core from *M. bovis* BCG and *M. smegmatis* were made as 50 base pair substrates (Figure 14B). These were titrated into *in vitro* assays to determine the efficiency at which they are used as *attBs* (Figure 27A). Using these *attB* substrates derived from BCG, product formation was seen in reactions with sites #1, #2, #4, #6, and #7. BCG site #1 was recombined with an efficiency similar to that of site #6 with >90% product conversion at both a 10-fold and 100-fold molar excess of *attB*. In fact, the amount of product in some site #1 reactions appeared to be somewhat higher, and all forms of the *attP* plasmid were converted to product. This is in contrast to what is seen in site #6 reactions, where some supercoiled and some nicked circular *attP* is still seen. When the density of the product bands is compared, more product is present in site #1 reactions than in site #6 reactions at all tested concentrations (Figure 27B). Reactions with sites #2, #4 and #7 yielded product at lower levels than #6 (Figure 27B). Sites #3 and #5 (3/5), which have the same sequence throughout 50 bp, showed no product formation when used as *attB* substrate in this assay (Figure 27A). These *in vitro* data for the most part agree with the *in vivo* integration results where sites #1, #6, and #7 are frequently used as integration sites. Site #4 was utilized in one instance, and sites #2, #3, and #5 had no integration events. However, there are two exceptions, site #2 did recombine *in vitro*, but no integration events were seen in this site. The amount of product formed in reactions with site #2 is less than the amount formed in reaction using site #1 or #6, thus recombination with site #2 is likely to be less efficient. It is possible that if more transformants were analyzed, integration events may be seen in site #2. A second possibility is that integrations into this site are not tolerated. Site #7 is also an exception; the intermediate level of activity *in vitro* does not support its frequent use *in vivo*. Less integration product is seen in reactions with site #7 as

Figure 27. *In vitro* activity of *attB* sites from *M. bovis* BCG

A) Sites from *M. bovis* BCG were used as *attB* substrates in integration assays with a supercoiled *attP* plasmid. The various BCG *attB* sites indicated above the gel were used as 50bp substrates and added in the following amounts. Lane 1 has 27 pmoles of *attB*, which is approximately a 1000-fold molar excess over *attP* [1000x], lane 2 has 2.7 pmoles (100x), lane 3 has 0.27 pmoles (10x), and lane 4 has 0.027 pmoles (1x). Lane - has no *attB* added. Sites #3 and #5 are identical over the 50 bases used in this assay, thus only one DNA was used for both these sites. The position of the linear product formed by recombination between the plasmid *attP* and linear *attB* is indicated by the arrow. Markers are shown in lane M, with the sizes in kb at the left.

B) Bar graph showing the product formed in reactions with each BCG site. This was calculated by density of product in the gel of the experiment shown in A. Here the amount of the product formed in the reaction with site #6 is set to 100 for each tested concentration, and the product formed in the other reactions is expressed as a percentage of the amount in site #6 reactions.

C) Substrate *attB* site sequences. *M. bovis* BCG sites #4 and #3 were further dissected and used in integration assays to determine which bases contribute to the reduction in conversion to product. Site #6 is 6/6/6, and each number in the series of three corresponds to the site from which the sequence is derived. The middle number is the sequence of the core; the first represents the sequence to the left of the core, while the one on the right is the sequence on the right.

D) Activity of hybrid *attB* sites. *In vitro* recombination assays were done with plasmid *attP* and Forty base-pair hybrid *attB* sites, whose sequences are as shown in C. Reactions were performed in sets of four. Lane 1 in each set has 23 pmoles of *attB* (1000 fold molar excess over *attP*), lane 2 has 2.3 pmoles, representing a 100 fold excess of *attB*. Lane 3 has 0.23 pmoles (10-fold), and

lane 4 has 0.023 pmoles, which is an equal-molar amount of *attB* and *attP*. The lane marked (-) has no *attB* added. Markers are shown in lane M with the sizes in kb at the right.

Figure 27 A&B

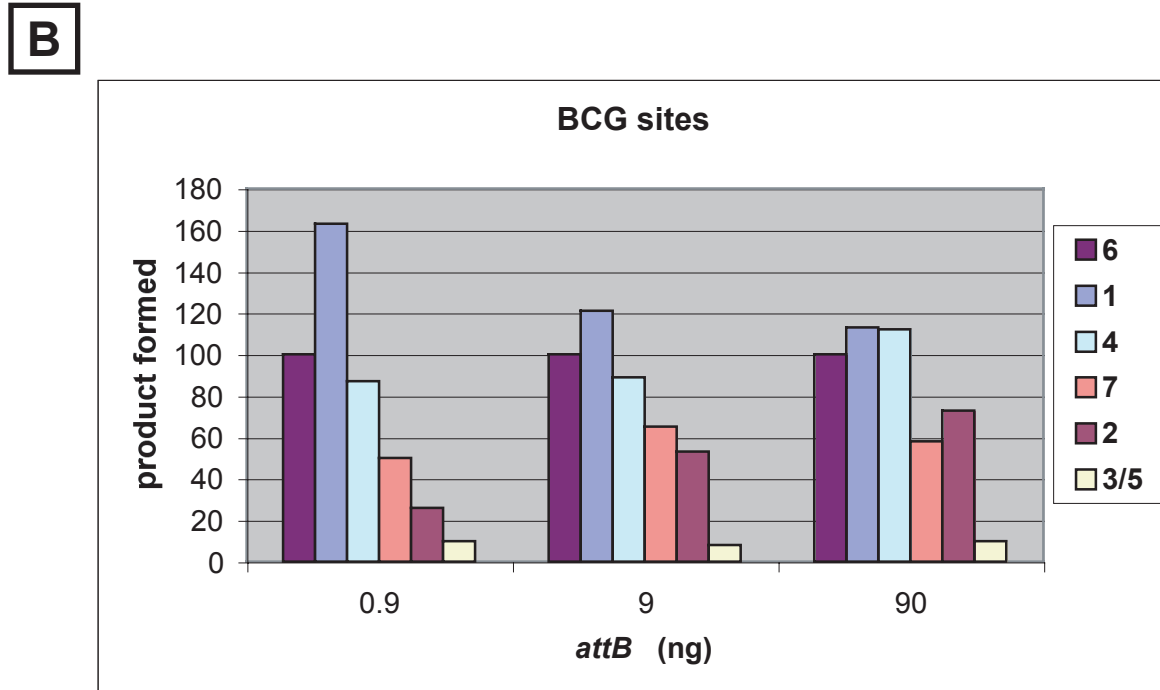
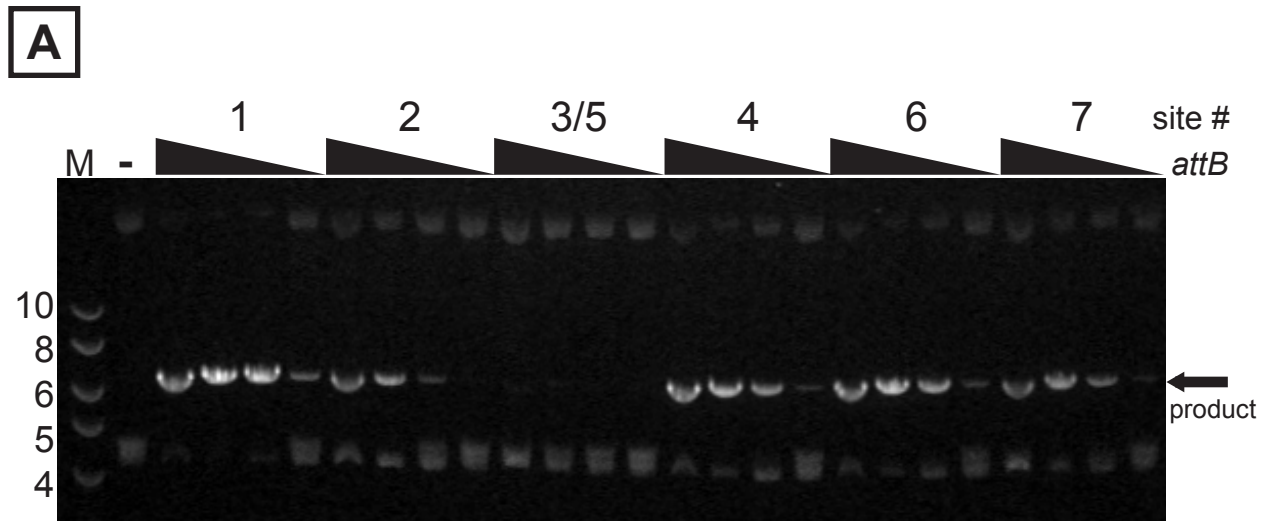
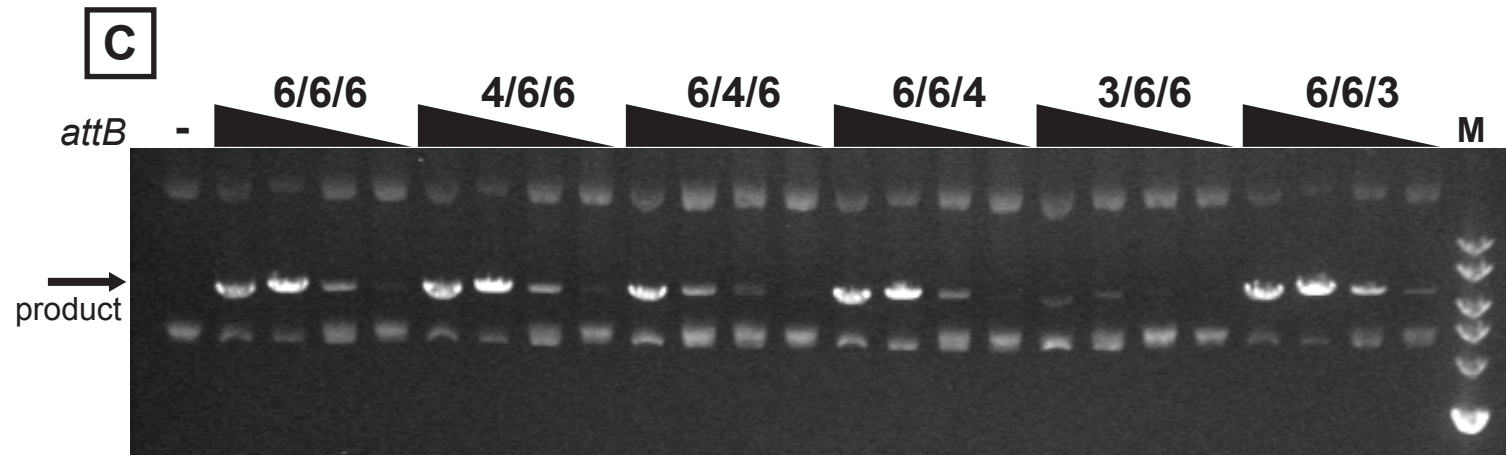


Figure 27 C & D



D

4 / 6 / 6	GTAAC	TGTTGGTGC	GG	GGTTGGCCGTGG	TCGAGGTGGGGT
6 / 4 / 6	GAAGGTGTTGGTGC	GGTTGTCC	ATGG	TCGAGGTGGGGT	
6 / 6 / 4	GAAGGTGTTGGTGC	GGTTGGCCGTGG	TCGAGGTGGG	CT	
3 / 3 / 6	GATCGA	TTTATTCGC	GGTTGGCCGTGG	TCGAGGTGGGGT	
6 / 3 / 3	GAAGGTGTTGGTGC	GGTTGGCCGTGG	TCGAGGTGT	G	

compared to reactions with site #1 or #6, and in fact, more product is seen in reactions with site #4. The reason for this is unclear but indicates that *in vitro* activity does not entirely correlate with site selection frequency.

Another set of reactions was done in which sites #3 and #4 were further dissected by separating changes within the core from those to the left or right of core. These were made as 50bp substrates. For site #4, three substrates were made in which the core or the region either to the left or right was made to match site #4, while the remainder of the sequence matched site #6 (Figure 27C). *In vitro* experiments using the substrate with the core sequence changed to #4 (6/4/6) showed a reduced level of product formation, while the substrates with changes to #4 sequence on the left or right showed levels of product formation similar to wild type (site #6 [6/6/6]) (Figure 27D). This indicates that these changes within the core sequence contribute to reduced activity as *attB*. Similar substrates were made for site #3, however, in this case the core is identical to the consensus, therefore, only two substrates were made. The substrate with the right half of the site matched to site #3 (6/6/3) has only a single base change, not surprisingly this substrate is converted to product in amounts similar to site #6. The substrate with the left portion changed to site #3 (3/6/6) sequences has several changes (9/16) and has a greatly reduced amount of product when used in reactions.

IV. F. *In vitro* efficiency of *M. smegmatis* REP13E12 *attB* sites

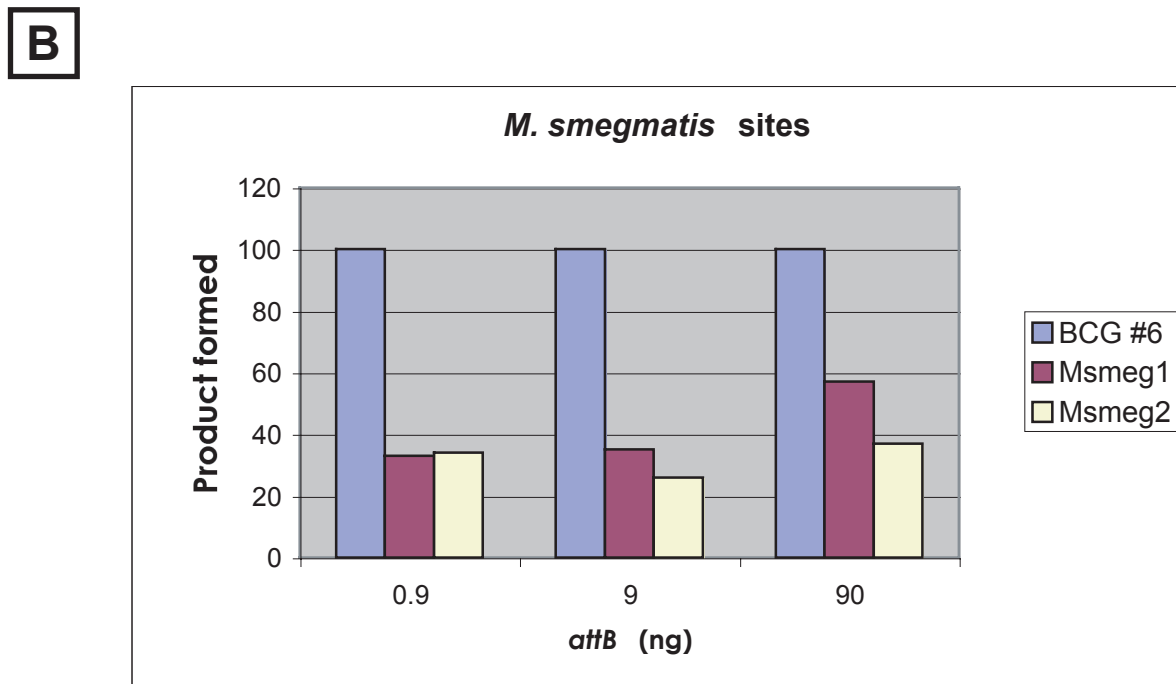
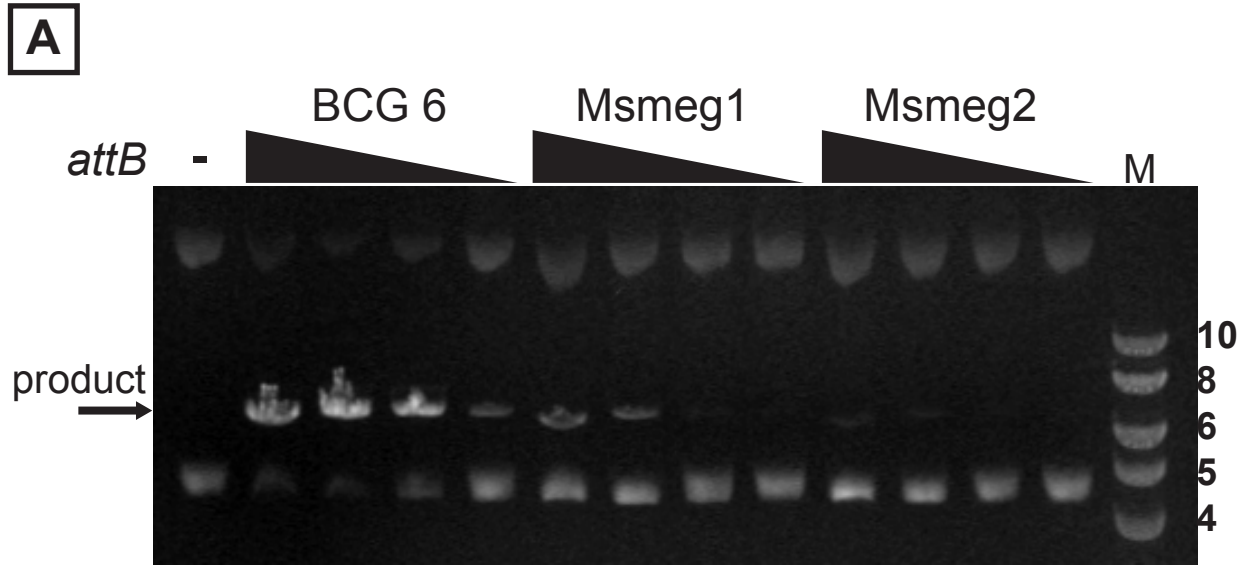
The two sites found by blast searching in *M. smegmatis* were also created as 50bp substrates and used with supercoiled *attP* in integration assays. These two sites, which are inefficient integration sites *in vivo*, also are utilized inefficiently as *attBs* *in vitro* (Figure 28A). In comparison to site #6 from BCG, only a small amount of product is formed at the two highest

Figure 28. *In vitro* activity of *attB* sites from *M. smegmatis*

A) Sites from *M. smegmatis* were used as *attB* substrates in integration assays with a supercoiled *attP* plasmid substrate. The Msmeg1 and Msmeg2 *attBs* were used as 50bp substrates along with BCG site #6 as a control. These sites were added in the following amounts. Lane 1 in each set has 23 pmoles of *attB* (1000-fold molar excess over *attP*), lane 2 has 2.3 pmoles, representing a 100-fold excess of *attB*. Lane 3 has 0.23 pmoles (10-fold), and lane 4 has 0.023 pmoles, which is an equal-molar amount of *attB* and *attP*. The position of the linear product formed by recombination between the plasmid *attP* (lowest and highest bands in the gel - supercoiled and nicked circular forms) and linear *attB* is indicated by the arrow. DNA size markers are shown in lane M with the sizes shown in kb at the right.

B) Graph of the experiment shown in A. The relative amount of product formed in reactions with the *M. smegmatis* sites and BCG site #6 is shown. This is calculated from the density of the product band regions. The amount of product produced by site #6 is set to 100%, and the product formed in the other reactions is expressed as a percentage of the amount in site #6 reactions.

Figure 28



tested concentrations for Msmeg1, and almost nothing for Msmeg2 (Figure 28B). Msmeg1 is the site that was recovered from the two ϕ Rv1 integration plasmid transformants, thus integration does occur in this site. We do not yet know if the Msmeg2 site is used as an integration site *in vivo*.

IV. G. Activity of symmetrized *attB* sites

Examination of *attB* reveals the presence of weak inverted repeat sequences flanking the central dinucleotide, where five out of ten bases match (Figure 26). The need for an excess of *attB* for all known ‘natural’ substrates in the *in vitro* reaction led us to believe that we were not using the best possible *attB* substrate. In an attempt to create a better *attB*, two DNAs were synthesized in which perfect inverted repeat sequences were made. One *attB* has the right inverted repeat altered to match the left (sym1), while the second had the left half changed to match the right (sym2), creating in both substrates 10bp perfect inverted repeats (Figure 29A). In reactions with the sym1 substrate, no product was produced at any concentration tested, while reactions with the sym2 substrate produced as much if not more than the wild type site #6 (Figure 29B). To further examine the possibility that sym2 maybe a better *attB*, wt (site #6) and sym2, were compared in an *in vitro* time course experiment done with either equal molar amounts of substrate or 10-fold excess of *attB* (Figure 29C). Comparison of the time course of “wt” to sym2 at 9ng of *attB* substrate (10-fold excess over *attP*), showed that sym2 reactions yielded more product at every time point tested (~1.5-3.0 times more) (Figure 29D). In contrast, the time course at equal molar ratio of *attB* to *attP* shows that reactions using sym2 or wt produced approximately equal amounts of product (Figure 29D). At higher concentrations,

Figure 29. *In vitro* activity of synthetic *attB* sites

A) Sequence of *attB* substrates used in these experiments. The set marked as “wt” uses site #6 from BCG as *attB* substrate. This *attB* is considered wild type (wt) because it is a native BCG sequence, and the changes made in other substrates are in the context of this sequence.

Differences from wild type sequence are shown in red, and regions of inverted symmetry are indicated by the horizontal arrows above each sequence.

B) *In vitro* integration assays were done using supercoiled *attP* plasmid and 50bp *attB* substrates. The *attB* substrate was used at different amounts in sets of four reactions. Lane 1 in each set has 2.3 pmoles of *attB* (100-fold molar excess over *attP*), lane 2 has 0.23 pmoles, representing a 10-fold excess of *attB*. Lane 3 has 0.023 pmoles, which is an equal-molar amount of *attB* and *attP*, and lane 4 has 0.0023 pmoles of *attB*, which is one-tenth the amount of *attP* in this reaction. The first lane of the gel (-) has no *attB* added. The position of the linear product formed by recombination between the plasmid *attP* and linear *attB* is indicated by the arrow. The size markers are shown in lane M, and are in descending size (top to bottom) 10, 8, 6, 5, and 4 kb.

C) Agarose gel of a time course *in vitro* integration assay using either BCG site #6 or sym2 as *attB* substrate. Reactions using a plasmid *attP* and either wt *attB* are shown on the left half of the gel and sym2 *attB* reaction are shown on the right half of the gel. The time course was done over 16 hours, and each reaction was incubated for the time indicated above each lane (in hours). Reactions were performed at two different amounts of *attB* substrate, 0.23 pmoles, which is 10-fold the amount of *attP* in the reaction (10X), and 0.023 pmoles, which is an equal-molar amount of *attB* and *attP* (1X). The position of the linear product is indicated by the arrow.

D) Graph of density data from the time course experiment shown in C.

Figure 29 A & B

A

6/wt GTGGAAGGTGTTGGTGC⁻GGGGT⁻ TG GCCGTGGT⁻CGAGGTGGGGTGGTGGTA⁻

TA GTGGAAGGTGTTGGTGC⁻GGGGT⁻ TA GCCGTGGT⁻CGAGGTGGGGTGGTGGTA⁻

sym1 GTGGAAGGTGTTGGTGC⁻GGGGT⁻ TG GCC⁻CCG⁻CACC⁻CAGGTGGGGTGGTGGTA⁻

sym2 GTGGAAGGTGTT⁻CGACC⁻ACGGT⁻ TG GCCGTGGT⁻CGAGGTGGGGTGGTGGTA⁻

B

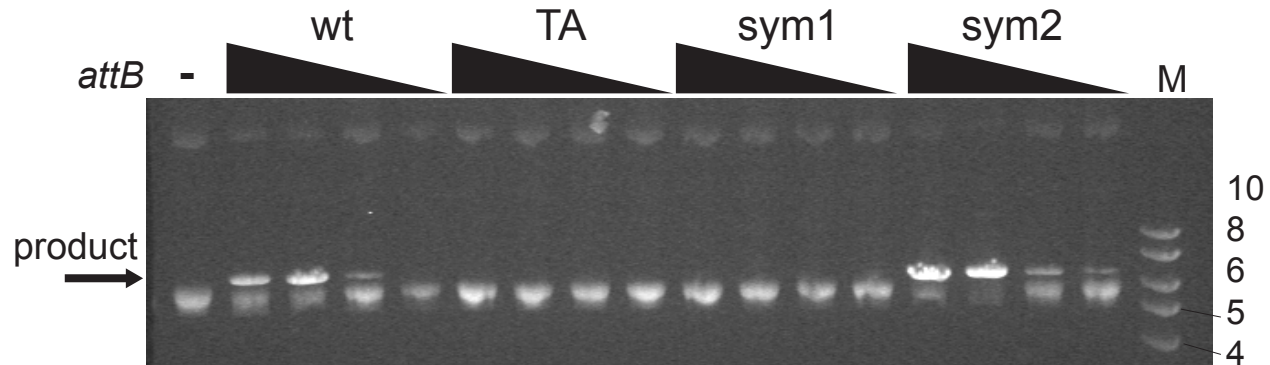
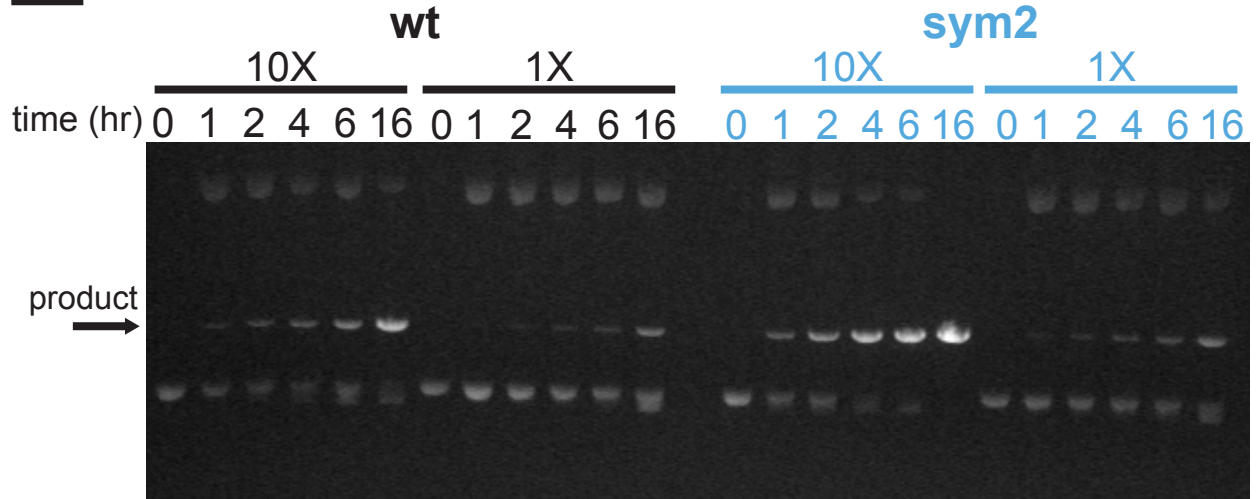
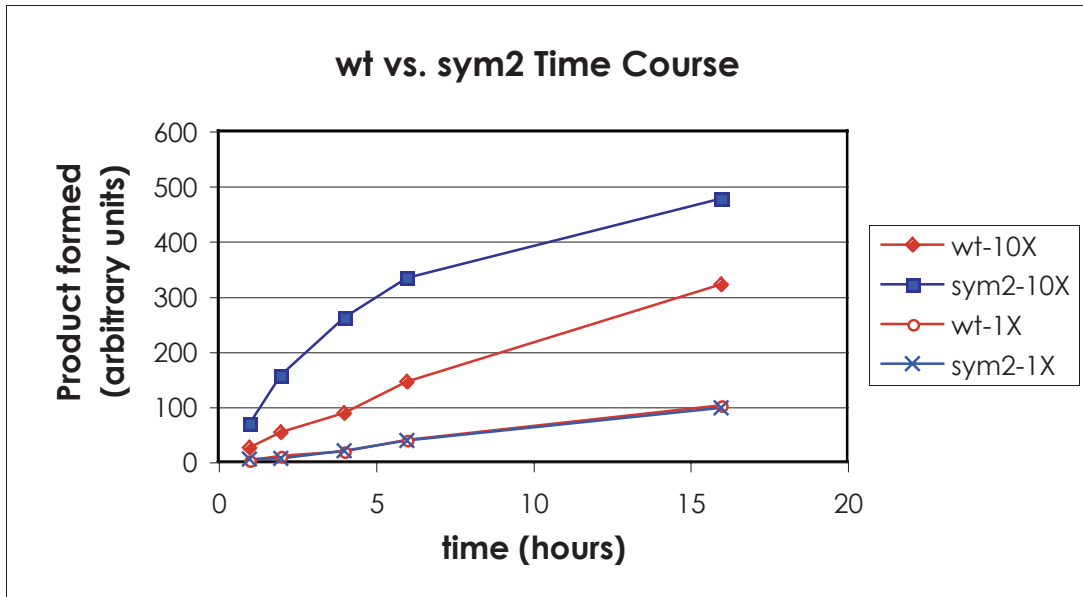


Figure 29 C & D

C



D



integration reactions using sym2 as the *attB* substrate yielded more product at each time point, suggesting that the speed of the reaction was increased.

An additional *attB* substrate was made in which the second base of the putative central dinucleotide was changed from G to A, thus symmetrizing the central dinucleotide (TA instead of TG). These 50bp *attBs* were added to *in vitro* recombination reactions with a supercoiled *attP* substrate. The central dinucleotide mutant (TA) was not converted to product at any tested concentration (Figure 29B). If the central dinucleotide is in fact the TG at bases 4 and 5 of the common core, this result is not surprising, as it has been observed in other systems. The TA mutant has a mismatched 3' overhang, which cannot be ligated into product.

IV. H. Activity of *attB* mutants with changes on the left side

A series of mutations were made in the left side of *attB* to test the importance of this region in integration. These changes were made within a 40bp minimal substrate (Figure 30A), and added to *in vitro* recombination assays as *attB* substrate. The 1-8 substrate made had the first eight bases of the sequence changed, which are just outside of the inverted repeats. In reactions using 1-8 as an *attB* substrate, product conversion is less than 50% of wild type, even at the highest concentration (Figure 30 B and C). When a substrate with the first four bases mutated (1-4) is used in an *in vitro* reaction, the amount of substrate converted to product is similar to wild type (wt), and at the highest concentration greater than 90% of substrate is converted to product. In fact, when the density of the product bands was compared, 1-4 reactions produced slightly more product than wt (Figure 30C). The 5-8 substrate, which has bases five through eight mutated, yielded an amount of product similar to the 1-8 mutant. This

suggests that bases 5-8 likely account for the low quantity of product formation in reactions with the 1-8 mutant, while bases 1-4 did not. A substrate with bases 9-12 mutated also had reduced *attB* activity (Figure 30C). This suggests that one or more of the bases from 5-8 and 9-12 are important in integrative recombination.

IV. I. Activity of *attB* core sequence mutants

Mutations were also made in each base pair of the 12bp common core (Figure 31A). These were made as single changes in a 40bp minimal *attB* substrate and tested by *in vitro* recombination assays. Changing the base at position 1, 2, 3, 6 or 12 had little effect on product formation, reactions using these substrates at 90ng yielded product at greater than 50% of wild type levels (Figure 31 B and C). Integration reactions that utilize a substrate with a change at positions 7, 8, or 9 yielded a reduced amount of product, less than 50% of wild type (Figure 31 B and C). This indicates that these bases contribute to *attB* recognition or a step in catalysis. Reactions using a substrate with a change at any of the four remaining positions (4, 5, 10, and 11) showed very little product formation, 20% of wild type or less (Figure 31 B and C). Thus these positions are important to integration *in vitro*. Most notably, reactions using an *attB* with a change at either of the central dinucleotide positions 4 or 5 gave almost no product, lending additional evidence that this is the central dinucleotide.

IV. J. Role of supercoiling in ϕ Rv1 integrative recombination

Supercoiling plays an important role in the integration reactions of tyrosine recombinases, using *attP* as a supercoiled substrate *in vitro* greatly stimulates the reaction in the λ system (Richet *et al.*, 1986). In L5, supercoiling of either *attP* or *attB* stimulates integrative

Figure 30. *In vitro* integration with left side mutants of *attB*

A) Sequence of *attB* substrates used in these assays. The mutant *attBs* were made as linear 40bp substrates, and they have changes in the region to the left of core. Bases that differ from the site #6 sequence are shown in red.

B) *In vitro* integration assays were carried out using a plasmid *attP* and linear *attB* substrates. Reactions were done in sets of four, and each reaction was done using a different amount of *attB*. Lane 1 has 600ng of *attB*, or 23 pmoles, which is a 1000-fold molar excess over the 0.024 pmoles of *attP* plasmid in the reaction. Lane 2 has 2.3 pmoles of *attB* (100-fold excess), lane 3 has 0.23 pmoles, which is a 10-fold excess over *attP* and lane 4 has 0.023 pmoles, which is equivalent to the amount of *attP* in the reaction. Size markers are shown in lane M, with the sizes in kb on the right.

C) Bar graph of product density from the experiment shown in A. Only the product formation at 100-fold and 1000-fold excess of *attB* has been quantitated. The intensity of product at lower amounts of *attB* is often too low to be quantitated. The amount of product in the reactions with wild type is set to 100, and the other reactions are expressed as a percentage of this amount.

Figure 30

A

wt #6	GAAGGTGTTGGTGCAGGGGT	TG	GCCGTGGTCGAGGTGGGGT
1-8	TCCTACTGTGGTGCAGGGGT	TG	GCCGTGGTCGAGGTGGGGT
1-4	TGCAGTGTGGTGCAGGGGT	TG	GCCGTGGTCGAGGTGGGGT
5-8	GAAGTCAGTGGTGCAGGGGT	TG	GCCGTGGTCGAGGTGGGGT
9-12	GAAGGTGTGCAAGCAGGGGT	TG	GCCGTGGTCGAGGTGGGGT

B

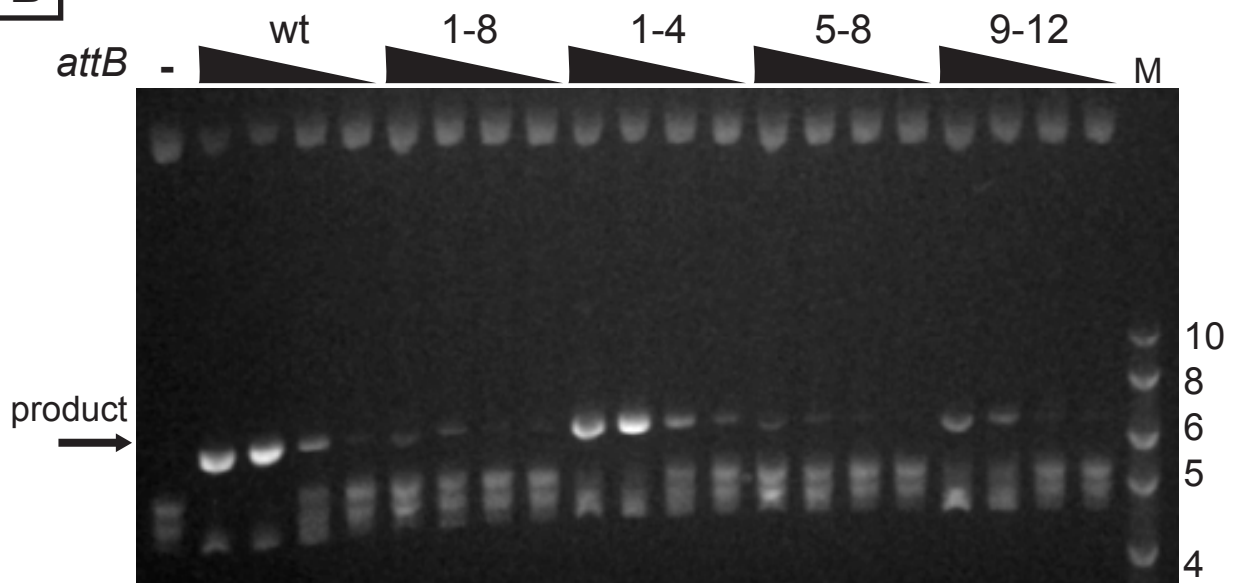


Figure 30C

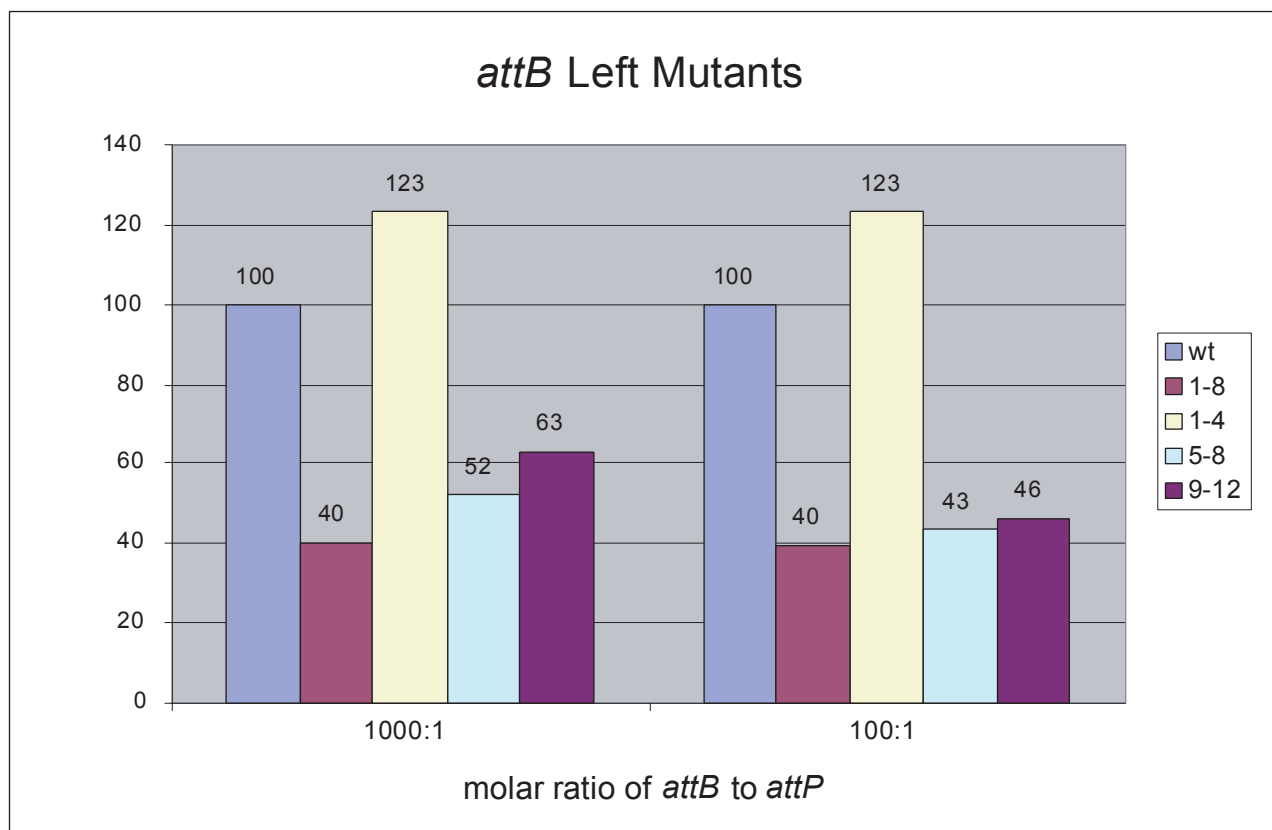


Figure 31. *In vitro* activity of *attB* core mutants

A) Sequence of *attB* core mutants used in the experiment, each has a single base change in one of the 12 bases of the common core. The base that has been changed in each substrate is shown in red. The central dinucleotide is boxed.

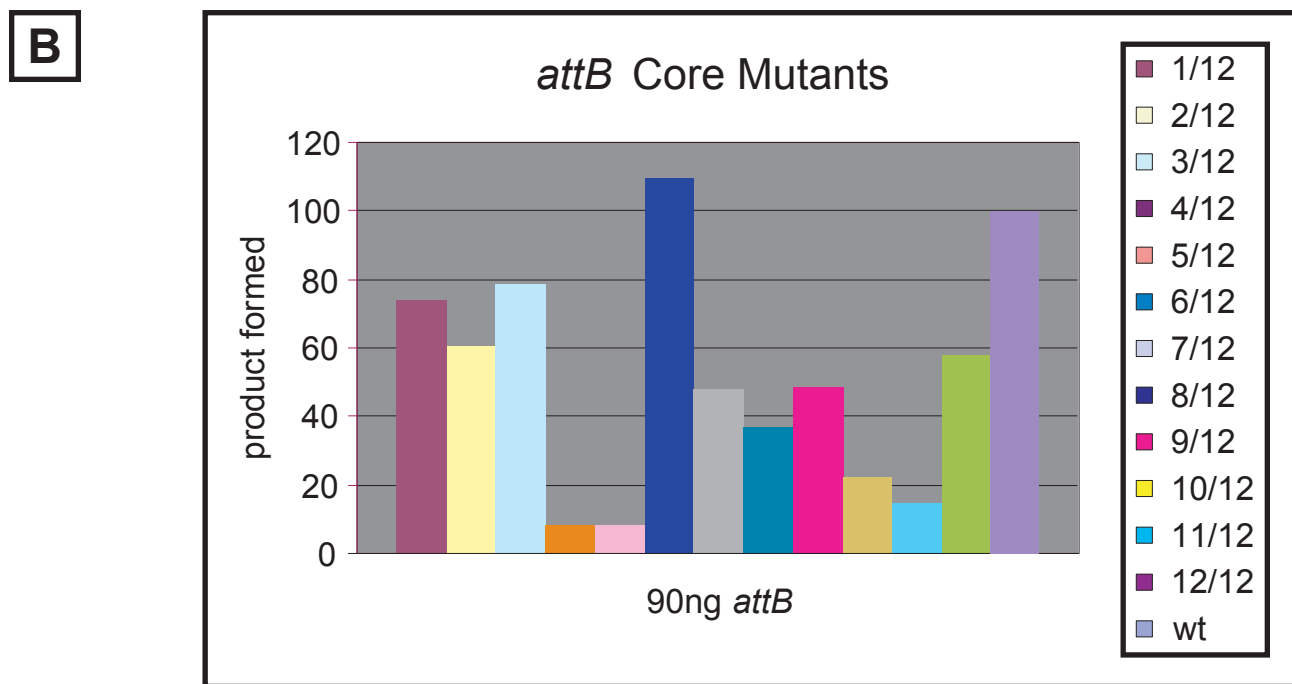
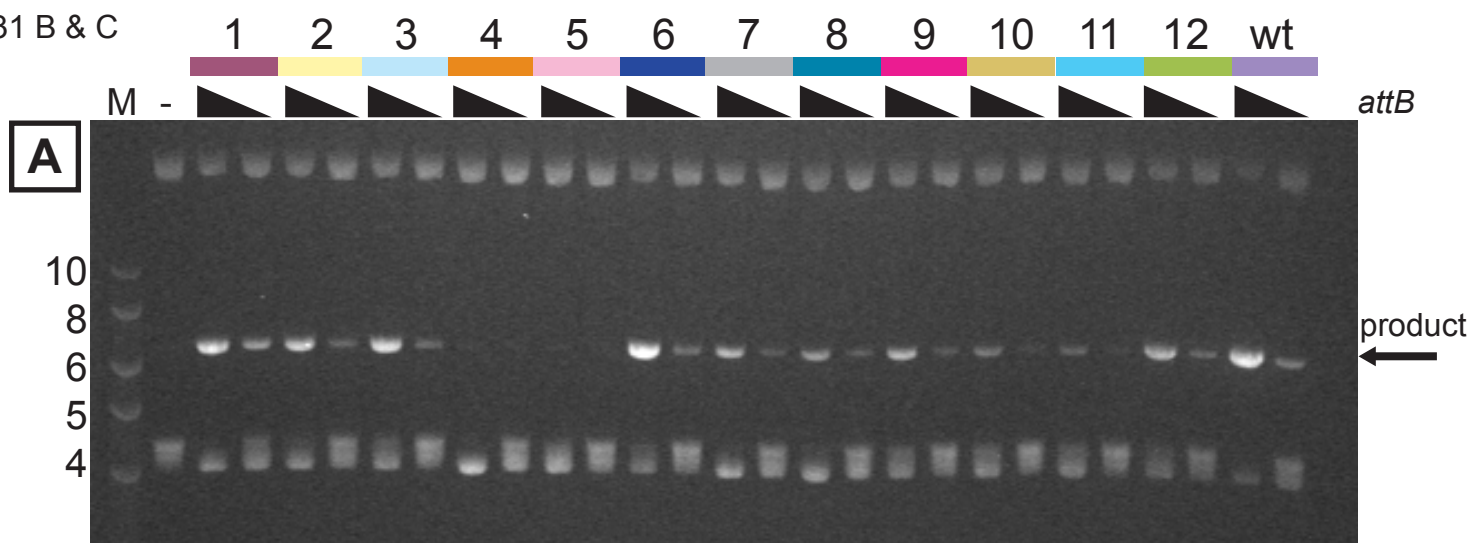
B) *In vitro* recombination assay with *attB* core mutant substrates. Reactions were done with an *attP* plasmid and linear 40bp *attB* at two different concentrations. The first lane in the pair of reactions for each substrate has 3.4 pmoles of *attB*, which is a 140-fold excess over the molar amount of *attP*. The second lane has 0.34 pmoles, a 14-fold molar excess of *attB*. The lane marked (-) has no *attB* added. The markers are in lane M, and the sizes are given in kilobases to the left of the gel.

C) Bar graph of product density data generated from the experiment shown in B. The density of product formed in the reactions with the wt sequence was set to 100%. Only the data for the reactions with 3.4 pmoles of *attB* is presented because the density of the product formed in reactions with 0.34 pmoles was too low to be accurately quantitated.

Figure 31A

wt	GAAGGTGTTGGTGCAGGGGT	TGGCCGTGGTCGAGGTGGGGT
1/12	GAAGGTGTTGGTGCAGG	CGTTGGCCGTGGTCGAGGTGGGGT
2/12	GAAGGTGTTGGTGCAGGG	CTTGGCCGTGGTCGAGGTGGGGT
3/12	GAAGGTGTTGGTGCAGGGG	ATTGGCCGTGGTCGAGGTGGGGT
4/12	GAAGGTGTTGGTGCAGGGGT	ATGGCCGTGGTCGAGGTGGGGT
5/12	GAAGGTGTTGGTGCAGGGGT	TTCGCCGTGGTCGAGGTGGGGT
6/12	GAAGGTGTTGGTGCAGGGGT	TGCCCGTGGTCGAGGTGGGGT
7/12	GAAGGTGTTGGTGCAGGGGT	TGGGCGTGGTCGAGGTGGGGT
8/12	GAAGGTGTTGGTGCAGGGGT	TGGCGGTGGTCGAGGTGGGGT
9/12	GAAGGTGTTGGTGCAGGGGT	TGGCCCTGGTCGAGGTGGGGT
10/12	GAAGGTGTTGGTGCAGGGGT	TGGCCGAGGTCGAGGTGGGGT
11/12	GAAGGTGTTGGTGCAGGGGT	TGGCCGTCTCGAGGTGGGGT
12/12	GAAGGTGTTGGTGCAGGGGT	TGGCCGTGCTCGAGGTGGGGT

Figure 31 B & C



recombination (Peña *et al.*, 1998). In the large serine integrase systems that have been studied, supercoiling is not required and provides no stimulating effect (Kim *et al.*, 2003; Thorpe and Smith, 1998). To test for the requirement of substrate supercoiling in the ϕ Rv1 integration reaction, an *in vitro* experiment was designed using linear *attB* and either a linear or supercoiled version of *attP* in a time course. The reaction set using the linear version of *attP* was generated by digesting the *attP* plasmid, pLB17, with HindIII prior to the recombination reaction. The set with supercoiled *attP* was digested with HindIII following the reaction. This was done so that amount of substrate (S) and product (P) could be compared in both sets of reactions (4.8 kb [S] vs. 3.3 & 1.5 kb [P]). In the reactions, a 30-fold excess of *attB* was utilized, and these reactions were incubated for different amounts of time. These reactions show that approximately equal amounts of substrate were converted to product at most time points in both linear and supercoiled sets of reactions (Figure 32). In some experiments, slightly more product was seen in reactions with supercoiled substrate, although at most, 25% more product was seen in supercoiled reactions in comparison to the linear counterpart.

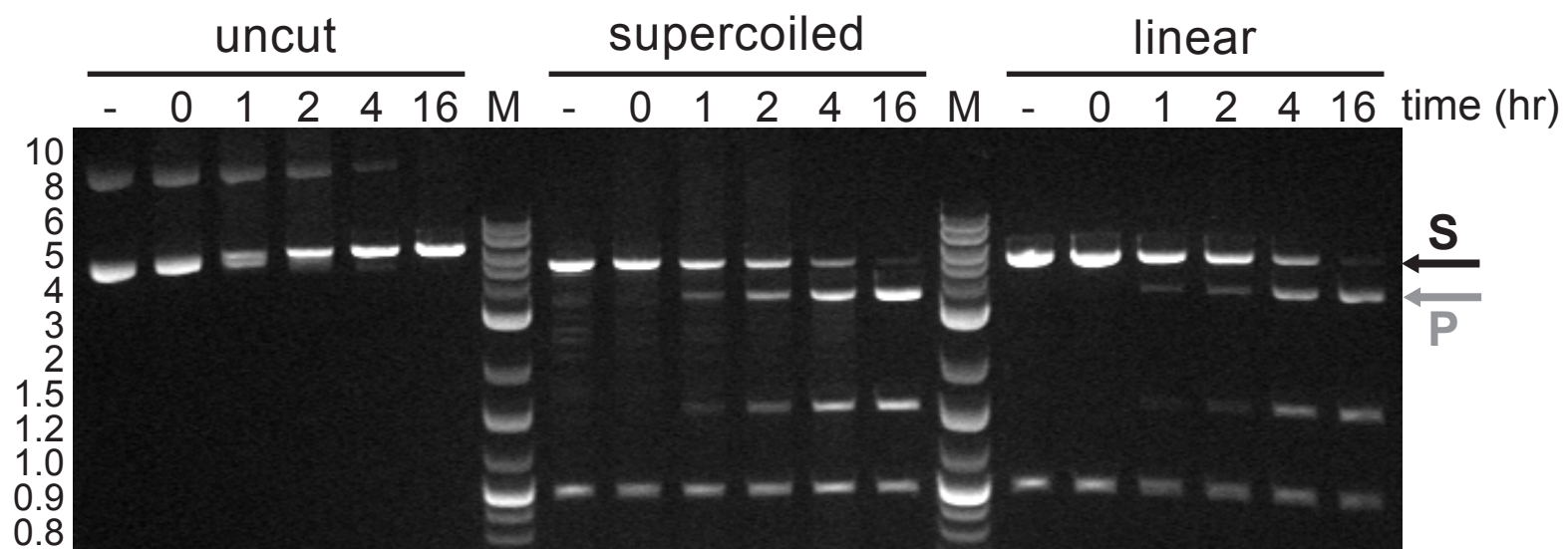
IV. K. Additional factors in integration

In vitro, an efficient ϕ Rv1 integration reaction requires an excess of *attB*, however, in the cell there is presumably a nearly 1:1 ratio of *attP* to *attB*. The data from the *in vivo* studies tells us that the recombination reaction must be efficient, as the number of transformants obtained with a ϕ Rv1 *attP-integrase* plasmid is similar to the number obtained with an extrachromosomal plasmid. A possible reason for the necessity of an excess of the linear substrate is that there is an additional factor present in mycobacteria that is absent from our reconstituted reactions. One candidate for the additional factor is mIHF (mycobacterial Integration Host Factor). The mIHF

Figure 32. Role of supercoiling in integration *in vitro*

A) In vitro integration assays were done with a linear 50bp *attB* substrate and either supercoiled or linearized plasmid *attP*. The *attB* substrate was used at 0.9 pmoles, which is 35-fold more than the amount of the *attP* plasmid used in the reactions. Linear *attP* substrate was created by digestion with Sall before carrying out the recombination reaction. Reactions using supercoiled (uncut) *attP* substrate were also digested with Sall, however this was done after recombination so that the amount of substrate (S-black arrow) and product (P-gray arrows), present in each reaction with linear or supercoiled *attP* could be compared. Samples were taken at different time points as indicated above each lane of the gel. One set of reactions using a supercoiled *attP* plasmid was not digested and is shown in the first third of the gel. Recombination reactions with supercoiled *attP* that were subsequently digested are in the middle of the gel, and integration reactions with linear *attP* are shown in the last third of the gel. Markers are shown in M lane with the sizes in kb on the left.

Figure 32



protein is a small 105 residue, 12kDa, non-specific DNA binding molecule that induces a bend in the DNA, and is required for efficient recombination of mycobacteriophage L5 (Pedulla *et al.*, 1996). This factor is present in *M. smegmatis* and *M. bovis* BCG among other mycobacteria; therefore, it could support ϕ Rv1 recombination in both species. To test if mIHF stimulates the ϕ Rv1 integration reaction, *in vitro* integration assays were carried out in the presence and absence of purified mIHF. Various amounts of mIHF were added, however, there was no detectable increase in product formed in reactions with mIHF over those without mIHF (Figure 33A). It is possible, however, that a different mycobacterial component will stimulate recombination. Because efficient transformation was achieved in both *M. bovis* BCG and *M. smegmatis* (Tables 1 and 3), both species must possess this factor. To look for a possible host factor, extracts were prepared from *M. smegmatis* and added to *in vitro* integration reactions. Again, no stimulation of the reaction was seen upon addition of these extracts (Figure 33B). Thus, we conclude that there is no host factor required for ϕ Rv1 integration.

IV. L. Discussion

The integration reaction of ϕ Rv1 - like other site-specific recombination systems can be reconstituted in a simple *in vitro* reaction using defined *attB* and *attP* substrates and either native or C-terminal 6-his-tagged integrase. Through these *in vitro* studies, we have demonstrated that the minimal substrate requirements for integration are 52bp for *attP* and 40 bp for *attB* (Figure 26). These small sizes for *attB* and *attP* are similar to what has been shown in other serine integrase systems (Figure 10B).

The *in vitro* integration reaction using a plasmid *attP* and a linear *attB* does not proceed efficiently when equal molar amounts of *attP* and *attB* are used. However, the addition of an

Figure 33. Additional factors in ϕ Rv1 integration

A) Role of mIHF in ϕ Rv1 integration. In vitro integration assays using *attP* plasmid and linear 50bp *attB* were incubated with various amounts of purified mIHF as indicated above each lane of the agarose gel. Reactions shown on the left half of the gel were done with an equal-molar amount of *attB* (0.9ng). The set of reaction on the right half of the gel was done with 90ng, which is a 100-fold molar excess of *attB*. Lanes marked with (-) have no integrase added, but mIHF is present at 32 pmoles. Markers are shown in M, and the sizes are in kilobases, 10, 8, 6, and 5 (from top to bottom).

B) Effect of mycobacterial extracts on *in vitro* integration. In vitro integrative recombination assays using plasmid *attP* and linear *attB* were done in the presence of increasing amounts of extract prepared from *Mycobacterium smegmatis* cells. These reactions were done with a plasmid *attP* and 50bp linear *attB*. The *attP* substrate was used at 100ng (pmoles and *attB* was at ng on the left half of the gel, and ng on the right. Increasing amounts of extracts were added as indicated by the triangles above the gel. The first lane has no integrase added. The reaction in the lane marked +I has integrase without extract added. All subsequent lanes have integrase added. The marker lane (M) has bands from top to bottom corresponding to 10, 8, 6, 5, and 4 kb.

Figure 33A

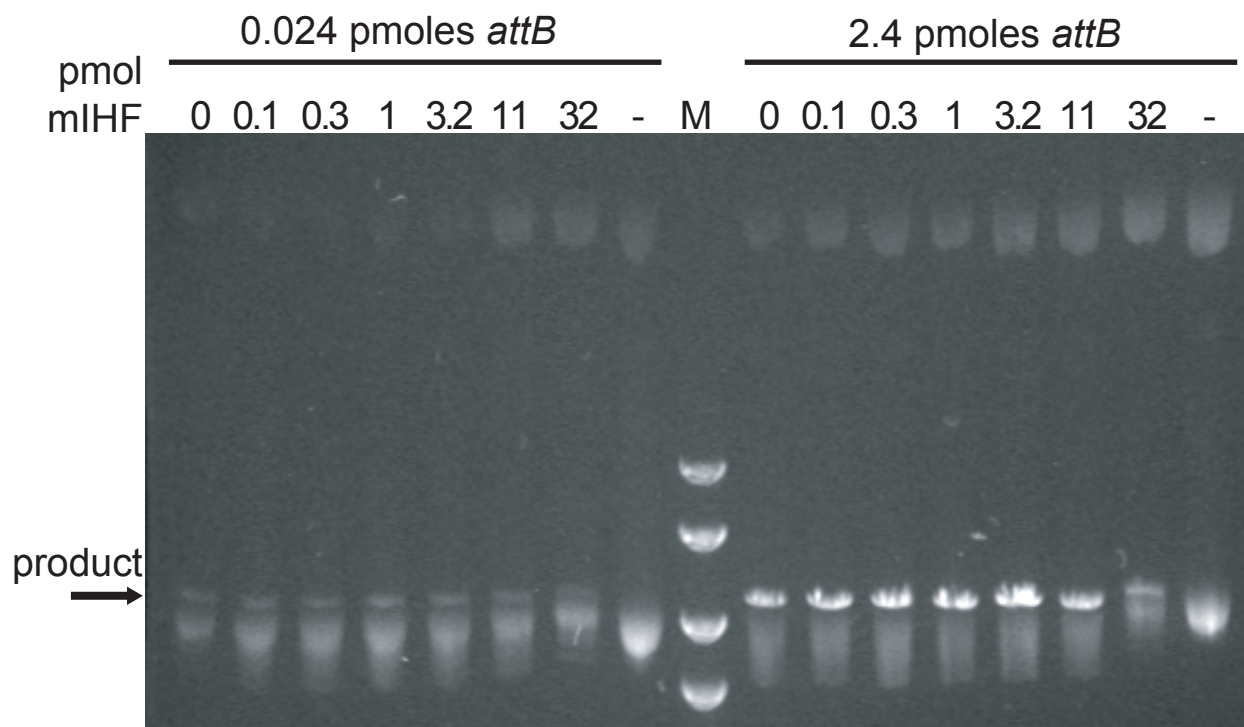
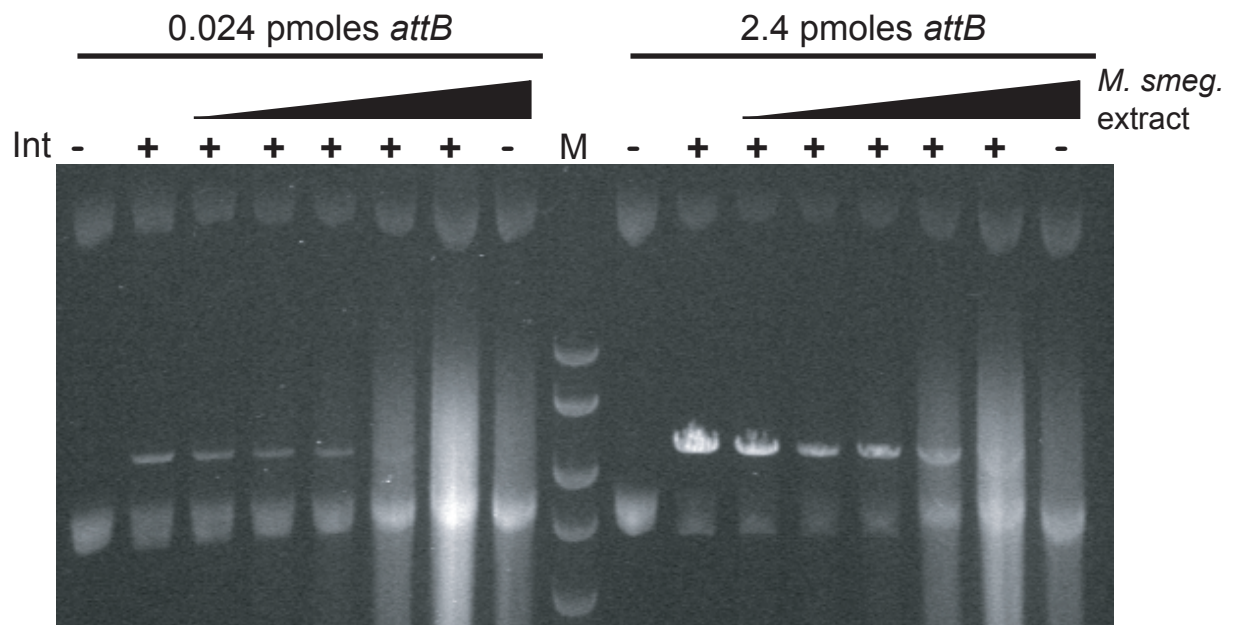


Figure 33B



excess of *attB* allows the reaction to go to completion. The reason behind the requirement for an excess of linear *attB* substrate is unclear, although it may reflect a deficiency of the interaction between integrase and *attB*. The deficiency may be at the level of integrase binding directly to *attB* or in bringing the *attB* in complex with *attP* (synapsis) or at some later stage in catalysis. Also, when integration reactions are done using a plasmid *attB* and a linear *attP* in equal molar amounts, there is also very little of the substrate is converted to product, while adding more linear *attP* substrate yields more product. A reaction using a 10-fold molar excess of *attP* yields the most product. The necessity of an excess of the linear substrate in this version of the reaction indicates a deficit in the ability of integrase to capture the linear substrate, whatever the substrate is, instead of a defect in *attB* interaction. In addition, these reactions utilizing a supercoiled *attB* and a linear *attP* are not as efficient as reactions with a plasmid *attP* and linear *attB*. Plasmid *attB* x linear *attP* reactions, in fact, do not go to completion at any concentration of *attP*.

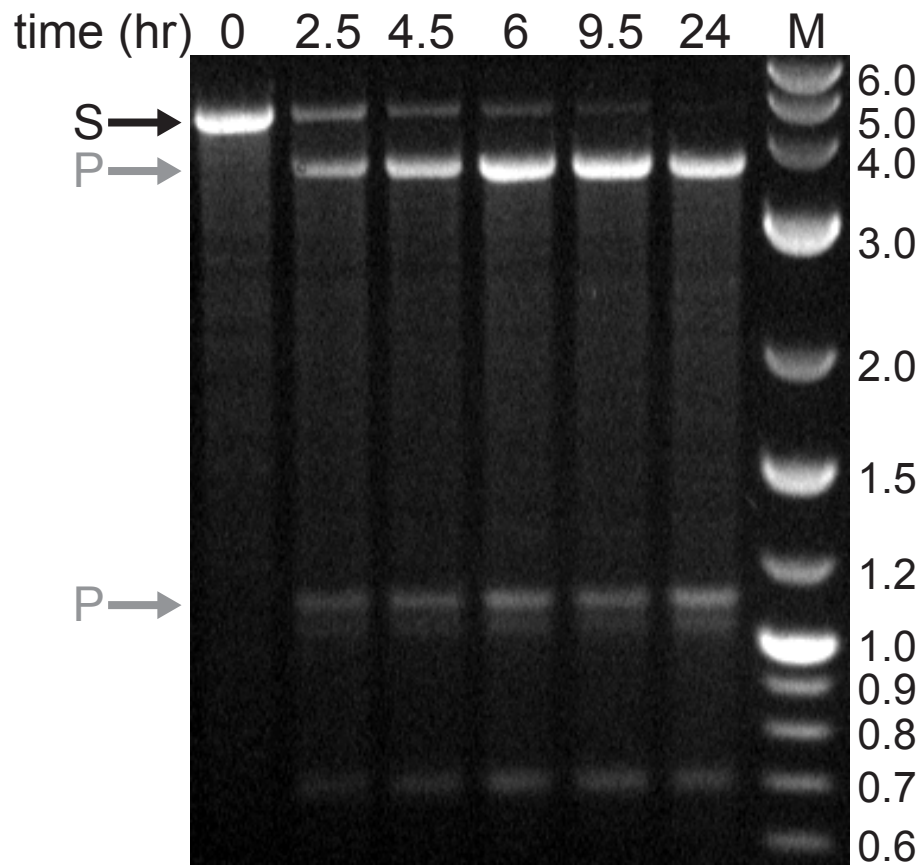
If a single plasmid with both *attP* and *attB* is used as substrate in an *in vitro* integration reaction, such that intramolecular recombination can occur, the reaction is efficient, with the reaction going to completion after extended incubation (Figure 34). This suggests that catalysis is efficient, but that some step before that is limiting the reaction, for example substrate binding or synapsis, as suggested above.

Although the ϕ Rv1 element is found within REP13E12 site #6 in H37Rv, this may not have been the sequence of the site when it first integrated there. The *in vitro* studies indicate that BCG site #1 recombines more efficiently than #6, thus may be more like the true *attB* sequence. Furthermore, a substrate in which the inverted repeats are a 10 out of 10 match (Sym2) is also converted to product more efficiently than #6 at higher concentrations. However, it is not converted to product more efficiently than site #6 at lower concentrations. This seemly

Figure 34. Intramolecular integration

Integrative recombination reactions were done using a single substrate with *attP* and *attB* in direct orientation. Samples were taken from the reaction at various time points, as shown above each lane (in hours). The reactions were then digested with HindIII after incubation to separate fragments indicative of product from those generated by cleavage of the substrate. The position of the substrate band (S) with a size of 4626bp is indicated by the black arrow. The product bands (P) at 3622bp and 1006bp are indicated by the gray arrows. The markers are shown in lane M, with the sizes in kb given on the right.

Figure 34



contradictory result may reflect a defect at the level of binding or synapsis. Once the sites synapse, catalysis occurs faster with sym2 than wild type. At low *attB* concentrations, product formation is low because the reaction is limited by the probability of supercoiled *attP* plus integrase colliding with *attB* plus integrase. At high concentrations, the probability of collision between *attP* plus integrase and *attB* plus integrase increases, and the difference in the rate of product formation can be detected.

Figure 35 shows the sequence of all naturally occurring or synthetic *attB* substrates used in this study. The sequences are organized according to how much of each substrate is converted to product in recombination reactions as compared to BCG site #6. Examination of the product formed in reactions using various *attB* sequences, both natural (derived from mycobacteria) and synthetic, gives some insight into what bases are important for catalysis and which are nonessential. However, several of the natural substrates have many changes, which make it difficult to dissect which specific bases contribute to the phenotype. For example, the *M. smegmatis* sites (Msmeg1 and Msmeg2) are inefficiently used as *attB* substrate as evidenced by the level of product formed in integration reactions, but these substrates have several changes. We can only say that one or more of these changes contribute to the reduction in recombinant product formed. The base changes in the left half of site #3 result in a reduction of *in vitro attB* activity, however there are nine changes. One or more of these changes contribute to the reduction in product formation. The changes in the core of site #4, which in this case are just two changes in positions 6 and 9 of the 12bp core, are responsible for reduced *attB* activity. From the core mutant data, the change at position 9 is most likely the greatest contributor to the low amount of product formed in reactions with 6/4/6 as *attB*. The changes in site #1 result in an

Figure 35. *attB* sites

The sequences of all *attB* for which *in vitro* integration recombination assays have been done are shown. Here, only the minimal 40bp of each site are shown. Bases shown in red are those that differ from *M. bovis* BCG site #6 which is considered the wild type (wt) or consensus sequence since all mutation made were based on this sequence. The sequences are grouped based on their relative activity as *attBs*. The first group, with three sequences, is followed by '+' show more product formation than site #6 (wt) under some tested conditions. Members of the second group, followed by a '=', have activity similar to wild type. The third group shows somewhat less product formation, thus the '-' indication. The fourth grouping has an even more reduced level of product formation *in vitro*, and is designated '- -'. The final grouping shows almost no activity *in vitro* and is designated by the 'X'.

Figure 35

site	sequence		activity
sym2	GAAGGTGTT CGAC ACGGT	TG GCCGTGGTCGAGGTGGGGT	+
1	GAAGGTGTTGGT CCAC GGT	TG GCCGTGGTCGAGGTGGGGT	+
4/6/6	GTAAC TGTTGGTGCGGGT	TG GCCGTGGTCGAGGTGGGGT	=
1-4	TCCT TGTTGGTGCGGGT	TG GCCGTGGTCGAGGTGGGGT	=
1/12	GAAGGTGTTGGTGCGG CG T	TG GCCGTGGTCGAGGTGGGGT	=
2/12	GAAGGTGTTGGTGCGGG CT	TG GCCGTGGTCGAGGTGGGGT	=
3/12	GAAGGTGTTGGTGCGGG GA	TG GCCGTGGTCGAGGTGGGGT	=
6/12	GAAGGTGTTGGTGCGGGT	TG CCC GTGGTCGAGGTGGGGT	=
12/12	GAAGGTGTTGGTGCGGGT	TG GCCGTG C TCGAGGTGGGGT	=
6/3/3	GAAGGTGTTGGTGCGGGT	TG GCCGTGGTCGAGGTG T GGT	=
6/6/4	GAAGGTGTTGGTGCGGGT	TG GCCGTGGTCGAGGTGGG CT	=
7	GG AGGTGTTGGTGCGGGT	TG GCCGTGGTCGAGGTGGGGT	-
9-12	GAAGGTGT GCA AGCGGGT	TG GCCGTGGTCGAGGTGGGGT	-
6/4/6	GAAGGTGTTGGTGCGGGT	TG TCC ATGGTCGAGGTGGGGT	-
7/12	GAAGGTGTTGGTGCGGGT	TG GG CGTGGTCGAGGTGGGGT	-
2	GA AC GT ACTGACT CGAGGT	TG GCCGTGGTCGAG A TG TGGC	--
1-8	TCCTACTG TGGTGCGGGT	TG GCCGTGGTCGAGGTGGGGT	--
5-8	GAAG TCAG TGGTGCGGGT	TG GCCGTGGTCGAGGTGGGGT	--
Msmeg1	G T AGGTGTTGGTGCGG TGT	TG GCCG CGG TCGAG TAGTTGC	--
8/12	GAAGGTGTTGGTGCGGGT	TG GC GG TGGTCGAGGTGGGGT	--
9/12	GAAGGTGTTGGTGCGGGT	TG GCC C TGGTCGAGGTGGGGT	--
10/12	GAAGGTGTTGGTGCGGGT	TG GCCG A GGTCGAGGTGGGGT	--
11/12	GAAGGTGTTGGTGCGGGT	TG GCCGT C GTCGAGGTGGGGT	--
3/3/6	GATCGA TT TATTC CGGGT	TG GCCGTGGTCGAGGTGGGGT	X
Msmeg2	G T AGGTGTTGGTGCGG TGT	TG ACC ACGGTCGAG AAGGTCG	X
4/12	GAAGGTGTTGGTGCGGGT	AG GCCGTGGTCGAGGTGGGGT	X
TA	GAAGGTGTTGGTGCGGGT	TA GCCGTGGTCGAGGTGGGGT	X
5/12	GAAGGTGTTGGTGCGGGT	TC GCCGTGGTCGAGGTGGGGT	X
sym1	GAAGGTGTTGGTGCGGGT	TG GCC CCGCA CCAGGTGGGGT	X

increase in product formation. In this substrate, there are three base changes just to the left of core, thus each of these differences is either neutral or has a positive effect on activity. Some changes made in the left part of *attB*, such as bases 5-8 and 9-12, are also converted to product in lower amounts than site #6, while changing the first four bases (1-4) has little effect on integrative recombination. Several positions of the core are important, as a change at bases 4, 5, 8, 9, 10, or 11 results in a reduction in the amount of *attB* converted to product. Two of these bases, at positions 4 and 5 are the putative central dinucleotide. Although, these substrates provide some useful insights, without a detailed mutagenesis study we can't say for sure how each base contributes to the minimal 40 bases required for activity.

From the *in vitro* studies, we also see that there are some differences between the efficiency of recombination with the various mycobacterial *attB* sites *in vitro* and how often an integration event was seen in the site. For example, site #7 from BCG is used frequently as an integration site *in vivo*, but *in vitro* it is only as efficient as site #2 or #4. These two latter sites, #2 and #4 are not used frequently *in vivo*. Only one event was seen in site #4, and no integration events were found in site #2. The *in vitro* assay is a simplified system and may reflect how efficiently an *attB* is converted to product while *in vivo* there may be other factors that effect site selection. Each *attB* site is in a REP13E12 element that is part of an ORF. Although there is similarity between the REP elements, because they are not identical we cannot be certain that there is functional redundancy. Sites #1, #6, and #7 are predicted to encode proteins that are greater than 90% identical; except that protein predicted by *Rv1587c* (#6) is 77 residues shorter. *Rv1945* (#5) and *Rv1148* (#3) are greater than 90% identical, and the two remaining predicted proteins are 40-60% identity, although much of the difference in the sites lies in size. Although the integration site is near the C-terminus of these ORFs, one or more of them may be essential

to *M. bovis*, and interruption of a particular ORF may be unfavorable. Another possibility is that the DNA at these sites may somehow be masked from the integrase. Bacterial chromosomes are condensed and organized in ways that we do not yet understand, and sites that are functional *attBs in vitro* (#2 and #4) may well be in regions of the chromosome that are not accessible.

Like other serine integrase systems, there is no requirement for a host-derived factor, and we have not detected a significant increase in product formation upon the addition of mIHF or mycobacterial extracts. Given the simple architecture of the sites is not clear how a host factor could play a role. There seems to be a single binding site for integrase unlike the core type and arm-type sites of the tyrosine *attP* substrates. The size of *attP* is insufficient for a host factor to function the way it does in the tyrosine systems, where it bends the DNA and aids in the formation of protein DNA complexes. Although a host factor in these serine systems could fulfill a different role, perhaps modulating substrate binding by integrase or synapsis. Further study is required to understand how the substrates are recognized, how they are discriminated, and how they are synapsed.

V. CONTROL OF ϕ Rv1 RECOMBINATION DIRECTIONALITY

V. A. Introduction

Integration and excision of a bacteriophage or other mobile element must be carefully regulated to ensure its success in propagation. In several tyrosine recombination systems, the control of directionality is achieved by the actions of a small protein, these directionality control proteins have been collectively referred to as recombination directionality factors, or RDFs (Lewis and Hatfull, 2001).

Many recombination directionality factors are small basic proteins, and their genes are also often located near their partner recombinase (Lewis and Hatfull, 2001). The ϕ Rv1 element has only 14 ORFs, and five of these can be assigned putative functions based on sequence similarity, leaving only nine for which we have no predicted function (Figure 12) (Cole *et al.*, 1998; Hendrix *et al.*, 1999). The ϕ Rv1 element has a similar organization to ϕ Rv2, the second prophage-like element of *M. tuberculosis*. This element, which encodes a tyrosine integrase, has a weak match to the mycobacteriophage L5 RDF, with a Blast e value of 10^{-05} (Lewis and Hatfull, 2001). The putative ϕ Rv2 RDF encoded by *Rv2657c* is two ORFs upstream of the integrase. In ϕ Rv1 there is a similar small ORF with 20% identity at the protein level to *Rv2657c*. This ORF, *Rv1584c*, is located in the same position as *Rv2657c*, two ORFs upstream

from the integrase gene (*Rv1586c*). This ϕ Rv1 *Rv1584c* gene product also has weak similarity to L5-Xis (19% identity) (Figure 36). Thus, Rv1584c is a candidate for the ϕ Rv1 RDF.

The mechanism for the control of directionality in the serine integrase schemes is not understood. It is known however, that TP901-1 ORF7 acts *in vivo* to promote excision, but how it functions is not known (Breuner *et al.*, 1999). In the ϕ C31 system, integrase alone is unable to mediate excisive recombination (*attL* x *attR*), but a factor controlling directionality has yet to be identified (Thorpe *et al.*, 2000). We have shown that a plasmid carrying the ϕ Rv1 *attP* and *integrase* is stably maintained in mycobacteria. Although it is possible that the integrase is not expressed when the plasmid is integrated since *attP* is very close to the start of the integrase gene, this stability suggests that an additional factor is required for excision.

In this chapter, the activity of Rv1584c as the ϕ Rv1 RDF is demonstrated. Expression of the ϕ Rv1 RDF in *M. smegmatis* LAB7 and *M. bovis* BCG strains with an integrated ϕ Rv1 plasmid results in excision of the plasmid. The addition of purified ϕ Rv1 RDF to *in vitro* integrative (*attP* x *attB*) recombination reactions results in inhibition of integration. Integrase, in the presence of RDF, catalyzes excisive (*attL* x *attR*) recombination. Excision occurs in an intramolecular supercoiled substrate as well as a linear substrate. Excision can also occur between a plasmid carrying *attL* and a small linear *attR*, or a plasmid *attR*, and a linear *attL*. The minimal sizes of *attL* and *attR* substrates are small at 47bp each. This is the first established excision reaction in a serine integrase system.

V. B. Rv1584c induces excision of an integrated *attP*-integrase plasmid

To test if *Rv1584c* indeed encodes a directionality factor, this ORF was cloned into an extrachromosomal vector under the control of a mycobacterial promoter. This promoter is a

Figure 36. Alignment of Rv1584c with known and predicted recombination directionality factors.

ClustalX alignment of the putative ϕ Rv1 RDF (Rv1584c) with the excise proteins of mycobacteriophages L5 and D29 (gp36) and the putative excise protein of ϕ Rv2 (Rv2657c). The residues that are conserved in all four sequences are shown in red. Those that are conserved in three out of four proteins are shown in yellow with blue outline. Similar residues that are conserved in at least three out of four sequences are shown in violet.

Figure 36

155

```

      . . . .10 . . . .20 . . . .30 . . . .40 . . . .50 . . . .60 . . . .70 . . . .80 . . . .
L5      1:.....MPPRA.....SIQETADYLGVSTKTVRNYIADGRLKAVRLGPRLIRVERESVEELMRPI...GK...:56
D29     1:.....MPPRA.....SIQQTADFLGVSTKTVRRYIADGRLKAVRLGPRLIRVERDSVEALMRPI...GK...:56
Rv2657c 1:MCAFPSPLGWTVSHETERPGMADAPLSRRYITISEAAEYLAVTDRTVQRMIADGRLRGYRSGLRLVRLRRDEV DGAMHPF...GGAA:86
Rv1584c 1:.....MSTIYHHRGRVAALSRS....RASDDPEFIAAKTDLVAANIADYLRITLAAAPPLTDEQRTRLAELLRPVRRSGGAR:73
  
```

mutant derivative of the L5 P_{left} promoter with diminished expression (Bibb and Hatfull, unpublished observations). This promoter was used because it was found that an extrachromosomally replicating plasmid with *Rv1584c* fused to the stronger hsp60 promoter of *M. bovis* does not efficiently transform mycobacteria. The tetracycline-resistance encoding plasmid expressing RDF (pLB36) was electroporated into *M. smegmatis* LAB7::pLB17 (this strain has BCG *attB* site #6 [Hyg^R] in which pLB17 [Kan^R] is integrated). Electroporation with plasmid pLB36 [Tet^R] efficiently yielded tetracycline resistant transformants. However, these could only be recovered in the absence of kanamycin (the kanamycin resistance gene is on pLB17); in the presence of both kanamycin and tetracycline, no transformants were obtained. When individual tetracycline resistant transformants were screened for kanamycin resistance by replica plating on media containing tetracycline and kanamycin, 100% of the tested transformants (100/100) 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-based integration-proficient plasmid carrying *attB* site #6. PCR analysis of the transformants, looking for either *attL* and *attR* or *attB*, demonstrated that loss of the kanamycin marker was indeed a consequence of precise excision of pLB17, as evidenced by the presence of *attB* in all pLB36 transformants (Figure 37A).

A similar experiment was conducted in *M. bovis* BCG-Connaught. A plasmid containing *Rv1584c* expressed from an L5- P_{left} derivative was transformed into cells with an integrated copy of a ϕ Rv1 *attP* and *integrase* plasmid. In this experiment, a kanamycin resistance-encoding vector, pLB32, was used in place of pLB36. This RDF plasmid was transformed into a strain of BCG carrying pLB25 (Hyg^R) integrated into site #6 (Figure 37B). Transformations were plated on both kanamycin and kanamycin-hygromycin containing media. Both control and pLB32

Figure 37. Rv1584c-mediated excision in mycobacteria

A) A PCR assay was used to test for excision of ϕ Rv1 in *M. smegmatis*. Transformants of *M. smegmatis* LAB7::pLB17 were analyzed by PCR to determine the occupancy of the ϕ Rv1 *attB* site. This assay is identical to the one shown in figure 17, where amplification of *attP*, *attB*, *attL* and *attR* yields different sized fragments. Lanes 1-4 show control reactions using plasmid or cosmid templates for *attP* (lane 1-pLB17), *attL* and *attR* (lane 2-MTCYC336-*M. tuberculosis* H37Rv cosmid containing ϕ Rv1), and *attB* (lane 3-pLB16), as well as untransformed control LAB7::pLB17 cells (lane 4). Four transformants with pLB36, a plasmid with *Rv1584c* (ϕ Rv1 putative RDF) are shown in lanes 9-12, all contain an unoccupied *attB* site showing that excision of plasmid (pLB17) had occurred. Transformants containing pJL32, a control plasmid, shown in lanes 5-8, all show amplification of *attL* and *attR* sites. Markers are shown in lane M, and are in descending size (top to bottom) 400, 300 and 200bp.

B) A PCR assay was also used to test for excision of ϕ Rv1 in *M. bovis* BCG. Transformants of *M. bovis* BCG-Connaught::pLB25 containing the integrated ϕ Rv1 plasmid in site #6 were analyzed to determine the occupancy of the ϕ *attB* site. This assay is identical to the one shown in figure 15A, where primers amplify *attB*, or *attL* for each of the sites, in this figure however, only sites #1, #4, #6, and #7 were analyzed, as indicated above each lane. An untransformed control of *M. bovis* BCG-Connaught::pLB25(6) is shown in the first panel. The second panel shows the PCR profile of a pJL32 (empty vector) transformant, this shows amplification of the *attL* junction for site #6. Four transformants with the plasmid (pLB32) that carries *Rv1584c* are shown in the bottom panel, all these contain an unoccupied *attB* #6 site showing that excision of the resident plasmid (pLB17) had occurred. Markers are shown in lane M, and the sizes of the markers are 500, 400, 300 and 200bp (from top to bottom).

Figure 37A

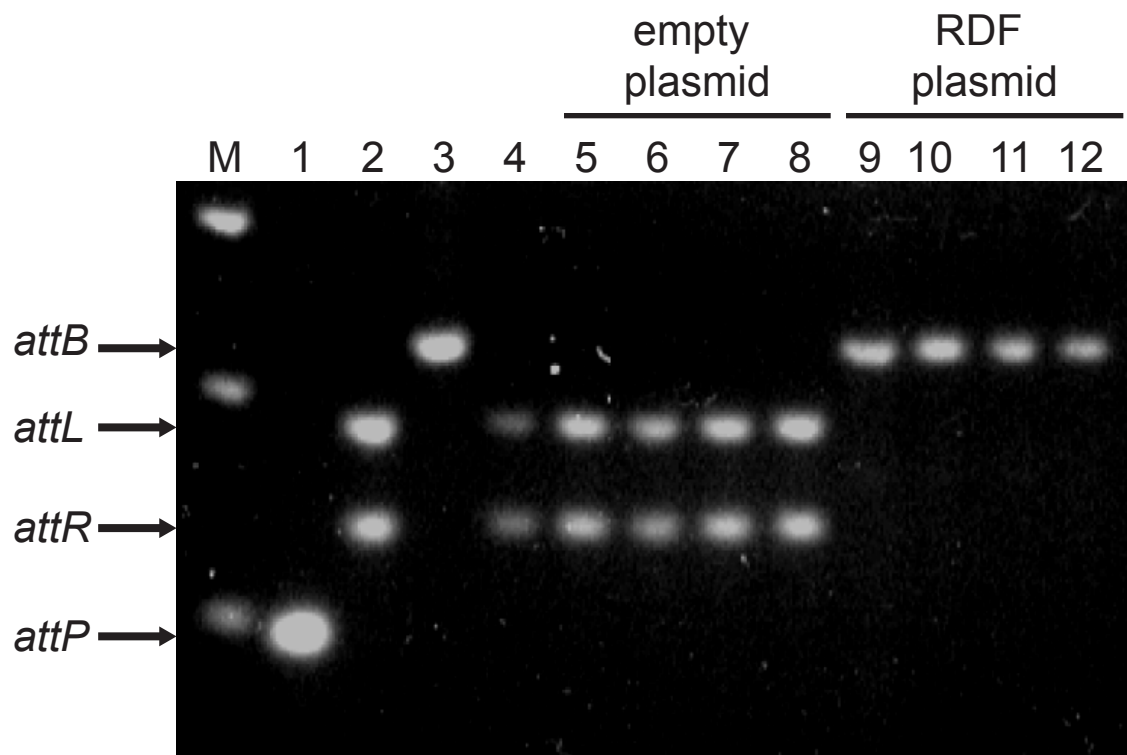
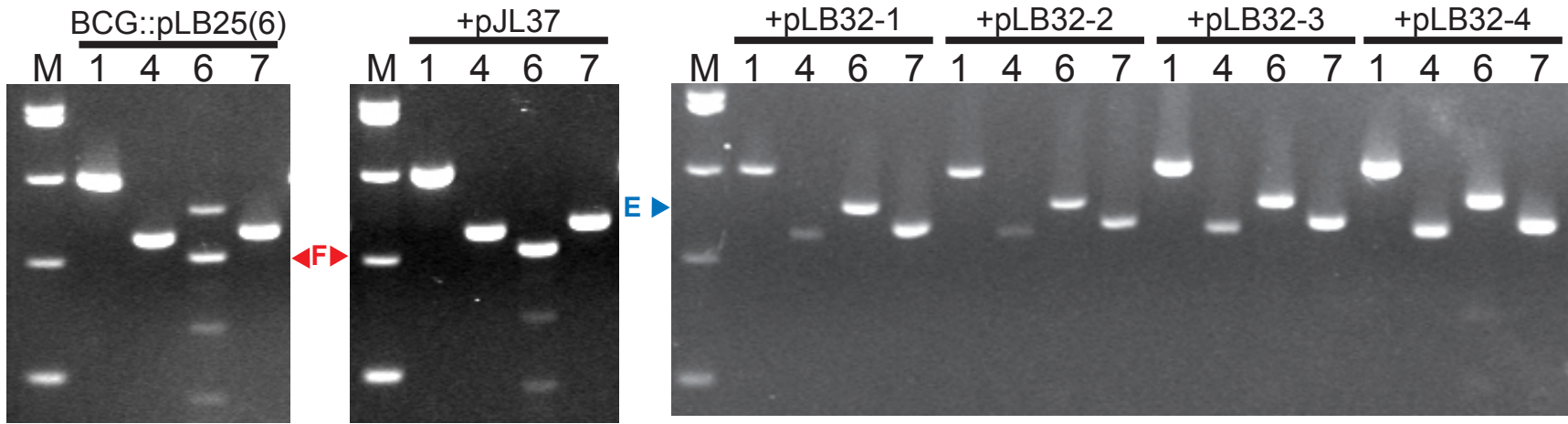


Figure 37B

159



transformants were recovered on kanamycin plates, and transformants with control plasmids were also recovered on plates with both kanamycin and hygromycin. However, no pLB32 transformants were recovered on media with kanamycin and hygromycin. Several pLB32 and control transformants were tested with the previously described BCG PCR assay, which surveys for the presence of *attB* or *attL*. However, in this case only the sites in which an integration event had been detected (#1, #4, #6 and #7) were assayed. It was determined that pLB32 transformants had an empty site #6, as evidenced by the amplification of the corresponding *attB* (Figure 37B). Thus, loss of hygromycin resistance correlates with loss of integrated pLB25, mediated by excision.

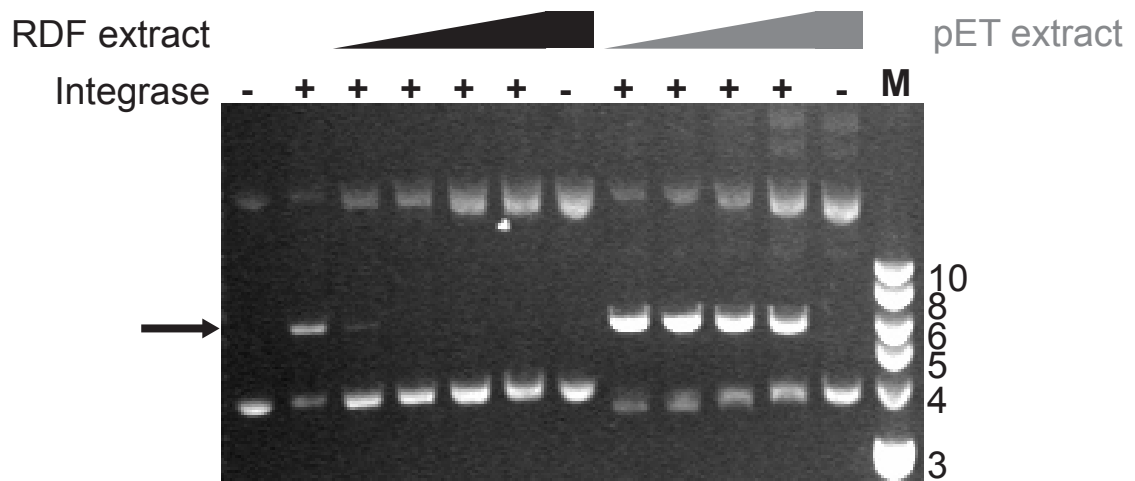
V. C. Rv1584c inhibits ϕ Rv1 integrative recombination *in vitro*

In order for the ϕ Rv1 RDF to control the directionality of site-specific recombination, it would likely inhibit one of the reactions, and promote the other reaction. In other recombination systems that utilize a tyrosine integrase such as lambda and L5, the RDF is required for excision, and also inhibits integration (Abremski and Gottesman, 1982; Nash, 1975). To test for inhibition of integration in the ϕ Rv1 system, a crude extract of ϕ Rv1 RDF was made from induced *E. coli* cells carrying an IPTG inducible expression vector with *Rv1584c*. A control extract was also prepared from cells with the empty parent vector (pET21a). These extracts were added at various dilutions to *in vitro* integration reactions. Reactions containing extracts from cells with RDF do not show integration product except a small amount in the reaction with the most dilute RDF extract. In parallel reactions with extract from cells containing an empty vector control, product was formed at all dilutions of extract tested. Thus, extract from cells with RDF inhibited integration while the control extract did not (Figure 38).

Figure 38. ϕ Rv1 RDF inhibits integrative recombination

Crude extract of ϕ Rv1 RDF expressing cells (black triangle) or control cells (pET vector-gray triangle) was added to *in vitro* integration assays with a supercoiled *attP* plasmid and a 50bp *attB*. The position of the linear recombinant product is indicated by the arrow. Addition of increasingly concentrated RDF crude extract inhibits the integration reaction, while addition of the control extract does not inhibit at any concentration. Markers in lane M are in kb with the sizes shown at left.

Figure 38



V. D. ϕ Rv1 RDF stimulates excisive recombination *in vitro*

In order to study the activity of the ϕ Rv1 RDF in further detail, we needed to establish an *in vitro* excision assay. First, an excision substrate was created which contained both *attL* and *attR* in direct orientation. This was made by carrying out a large scale integration reaction between a plasmid *attP* and a linear 50bp *attB*, and purifying the product, which is a linear DNA with *attL* and *attR* in direct orientation. This product was then ligated to another 627bp piece of DNA in order to put space between *attL* and *attR* (Figure 39). The resulting plasmid, pLB44, was incubated with integrase and crude extract from cells expressing Rv1584c (ϕ Rv1 RDF) or from cells with an empty vector. These reactions were incubated overnight (16 hours or more), then digested with a restriction enzyme to distinguish between the substrate *attL-attR* plasmid and the products, which are two plasmids, one carrying *attB* and one with *attP*. Extracts from cells carrying the ϕ Rv1 RDF (*Rv1584c*) but not those carrying the control plasmid stimulated conversion of substrate to product (Figure 40), as product associated bands are seen only lanes where RDF extract was added.

V. E. Expression and purification of ϕ Rv1 RDF

To confirm that the inhibition of integration and excision activities seen in crude extracts prepared from *E. coli* cells expressing RDF were indeed the result of the ϕ Rv1 RDF, this protein was purified. *E. coli* BL21(DE3)pLysS cells were transformed with pLB13, a plasmid with the *Rv1584c* ORF under the control of the T7 promoter and LacI, which is a derivative of pET21a. RDF expression was induced by the addition of IPTG, upon which, a band with the approximate expected molecular weight of gpRv1584c was seen on a coomassie-stained SDS-polyacrylamide gel. Although the expected molecular weight is 8kDa, the band

Figure 39. Construction of excision substrate and excision reactions

The top panel shows the integrative recombination reaction used to generate the excision substrate. This reaction used a plasmid containing *attP* (black coiled circle) and a linear *attB* (blue line) to yield a linear DNA with *attL* and *attR*. The middle panel outlines the construction of the plasmid pLB44. The linear product of the integration reaction was purified from an agarose gel and ligated with a DNA fragment (shown in gray) to generate pLB44, an excision substrate with *attL* and *attR* in direct orientation (as indicated by the arrows). The bottom panel shows a schematic of the excision reaction using pLB44. This plasmid is recombined by integrase in the presence of RDF to yield two smaller plasmids containing *attB* or *attP*. In the experiments shown subsequently, the reaction is digested to yield a series of linear fragments that indicate whether substrate or products are present.

Figure 39

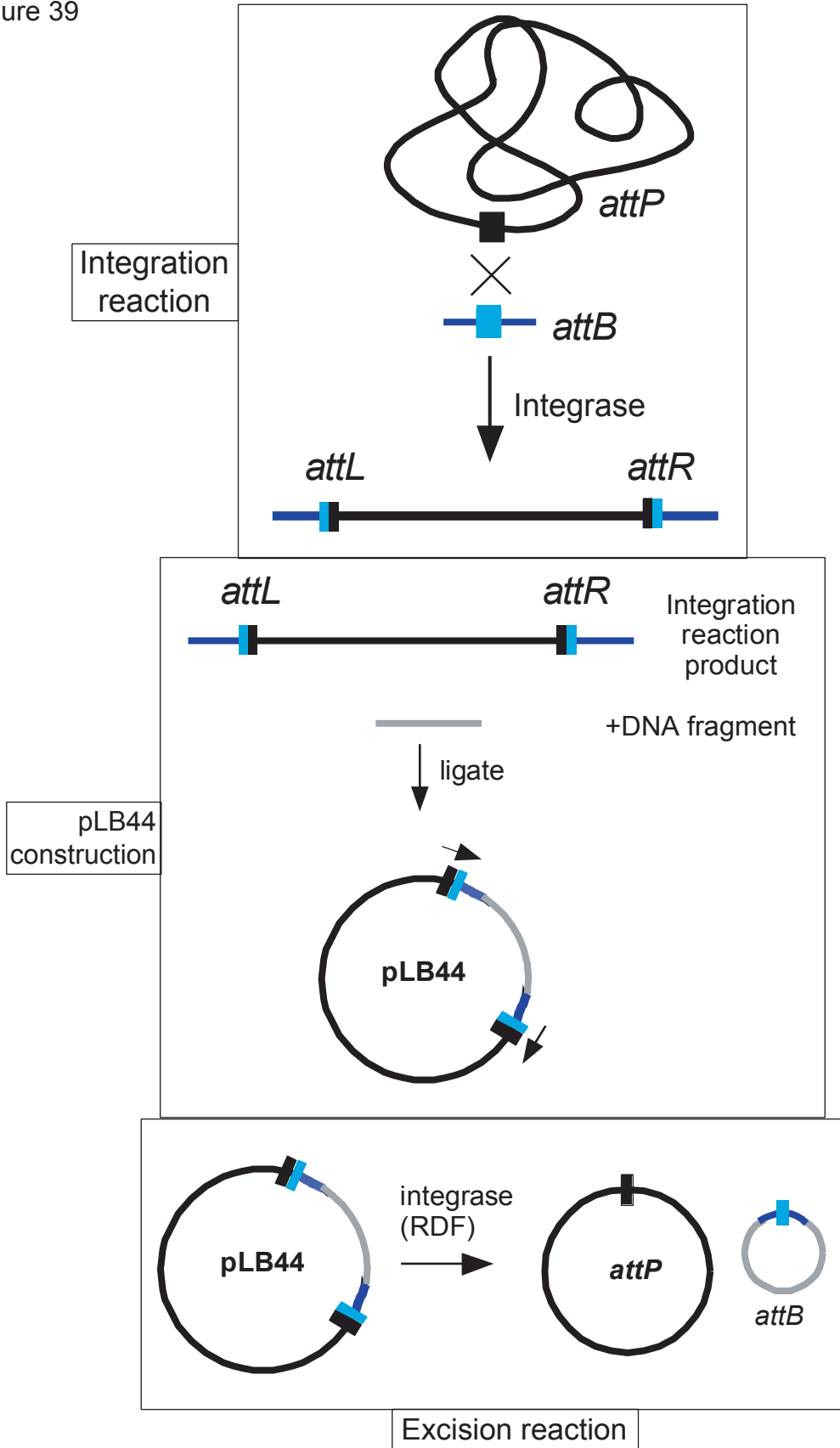
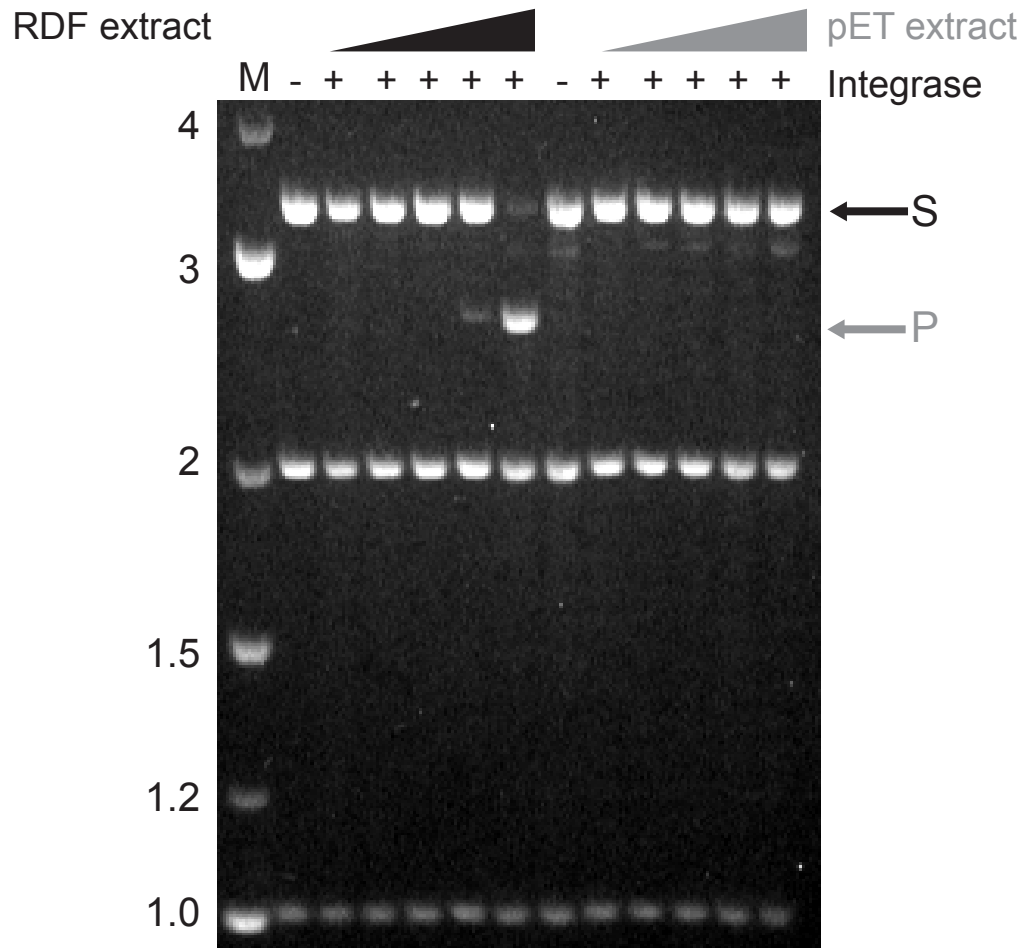


Figure 40. Rv1584c stimulates excisive recombination in an intramolecular substrate

An excision assay was devised using an intramolecular substrate with *attL* and *attR* in direct orientation. This substrate was incubated with extract from cells with ϕ Rv1 RDF or a control plasmid and integrase. The reactions were digested with BspHI following incubation to easily compare the substrate to the product. The migration of the digestion product that is indicative of substrate is 3269bp (shown by the (S) arrow). The fragment generated by digestion of the product is 2601bp (P-gray arrow). Addition of crude extract from RDF expressing cells (left half of gel) results in excisive recombination product, while the control extract does not.

Figure 40



actually migrates at less than 6.5kDa on the gel. The crude extract was purified by a combination of boiling, precipitation with 55% saturated ammonium sulfate, and CM-sepharose chromatography (Figure 41). This RDF preparation is greater than 90% pure by coomassie-blue staining. This purified RDF protein both inhibits integration and promotes excision (Figure 42), and excision is dependent upon the presence of both integrase and RDF (Figure 42).

V. F. Role of supercoiling in ϕ Rv1 excisive recombination

To test if substrate supercoiling plays a role in excisive recombination, excision reactions were performed with either supercoiled or linear substrate. To make linear substrate, pLB44 was digested before the excision reaction. Supercoiled reactions were done as previously described, and then these reactions were digested. These two sets of reactions were digested with the same enzyme either before (linear) or after (supercoiled) recombination so that the conversion of substrate to product could be compared between the two (Figure 43). The results of this experiment demonstrate that, much like integration, supercoiling is not required for excision and did not offer a significant stimulating effect. In fact, the set of reactions with the linear substrate had slightly more product at the end of the experiment (16 hours) as shown by comparison of the ratio of substrate (S) to product (P).

V. G. Intermolecular excision

An intramolecular substrate containing *attL* and *attR* is the natural excisive recombination substrate. Here, the integrated prophage is removed from the host genome via a site-specific recombination reaction between *attL* and *attR* in direct orientation with the products being two separate DNAs, one with *attB* and one with *attP*. We have shown that the ϕ Rv1 RDF

Figure 41. ϕ Rv1 RDF expression and purification

The ϕ Rv1 RDF was expressed in *E. coli* using a LacI inducible system with T7 polymerase.

Comparison of the induced sample (I) to the uninduced sample (U) shows a band near the expected size of RDF, 8kDa (near bottom of lane, indicated by the arrow). Partial purification of the lysate was completed then loaded (L) on a CM-Sepharose column and eluted with an increasing salt gradient. The flow through (FT) and fractions are shown, and late fractions have the 8kDa band corresponding to RDF. The second panel shows the load (L) of the column and the pooled dialyzed fractions (P). Molecular weight markers are shown in lane M. The sizes of the markers are in kDa, and are shown at the right.

Figure 41

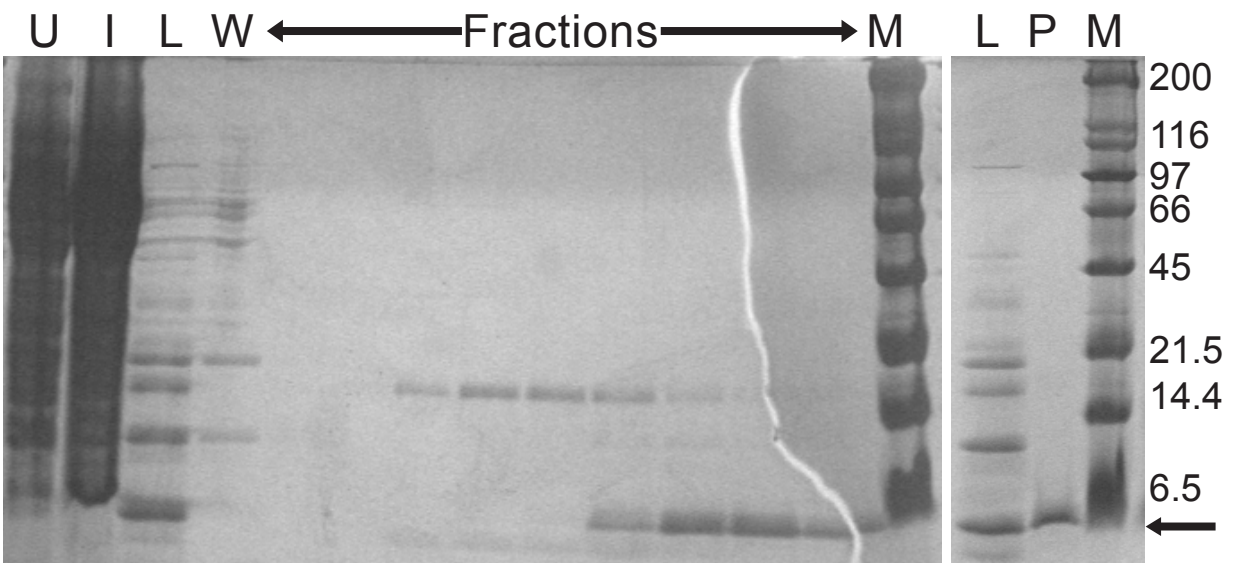


Figure 42. Purified RDF inhibits integration and promotes excision

The left half of the gel shows *in vitro* integration assays with increasing amounts of RDF. These reactions utilize an *attP* plasmid (0.024 pmoles) and a 50bp linear *attB* (~1.1 pmoles). A control reaction with integrase alone (4 pmoles) is shown in the first lane (-), and integration occurs in this reaction. The second lane (C), shows a reaction that has undiluted crude RDF extract, and integration is inhibited. The next six lanes show reactions with increasing amounts of the purified RDF (0.27, 0.4, 0.8, 1.6, and 2.7 pmoles) along with integrase. Lanes with the purified RDF show that it inhibits integration. The position of the integration product (IP) is indicated by the arrow on the left. The right half of the gel shows an excision assay, which is done using a supercoiled *attL-attR* plasmid (pLB44, at 0.047 pmoles). The product of excisive recombination (EP) is indicated on the gel by the arrow at right. The first lane (-) has integrase alone (4 pmoles), the second lane (c) has integrase and crude RDF extract. The next six lanes are reactions with integrase and increasing amounts of purified RDF (0.27, 0.4, 0.8, 1.6, and 2.7 pmoles). This purified RDF stimulates excision as shown by the presence of a band specific to excisive recombination. The DNA size markers are shown in lane M and the sizes are 10, 8, 6, 5, 4, 3, and 2 kb from top to bottom.

Figure 42

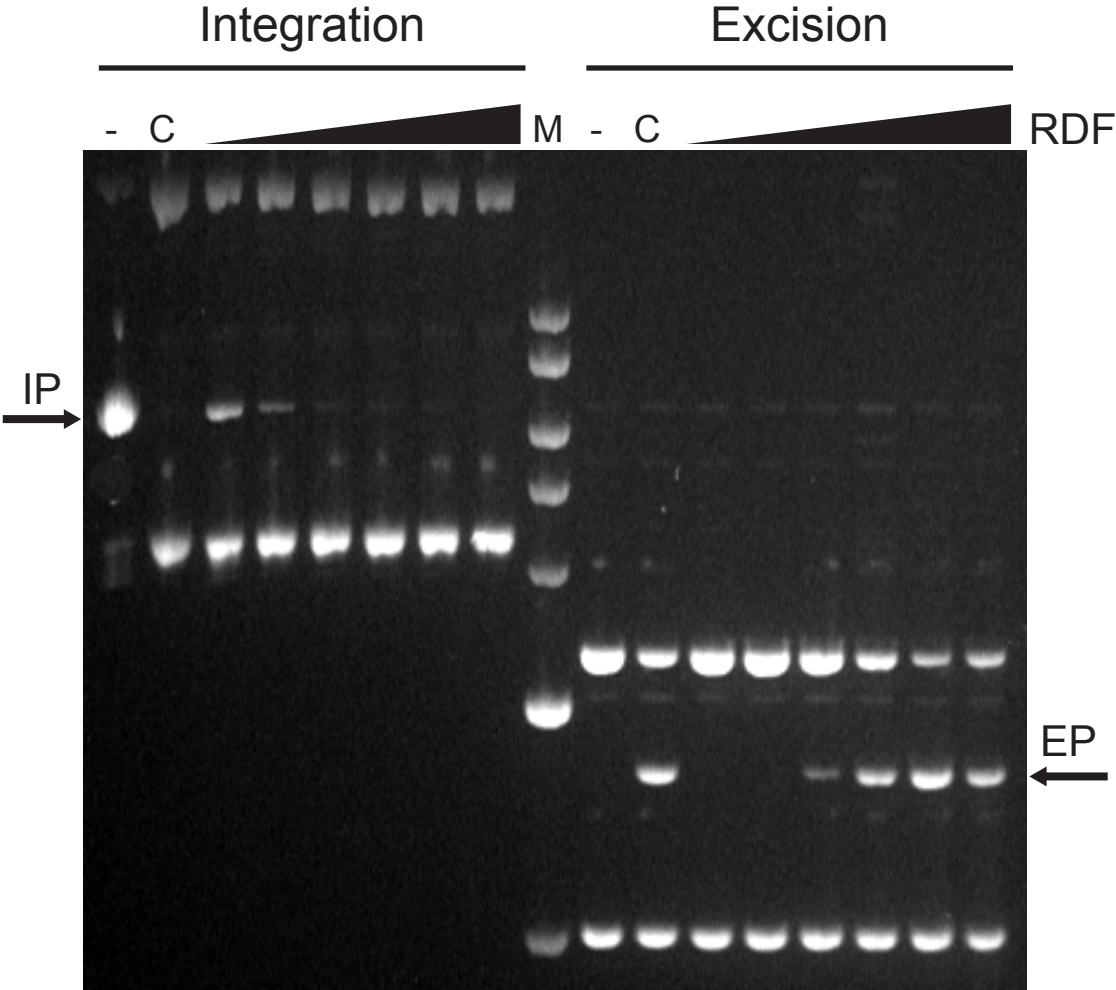
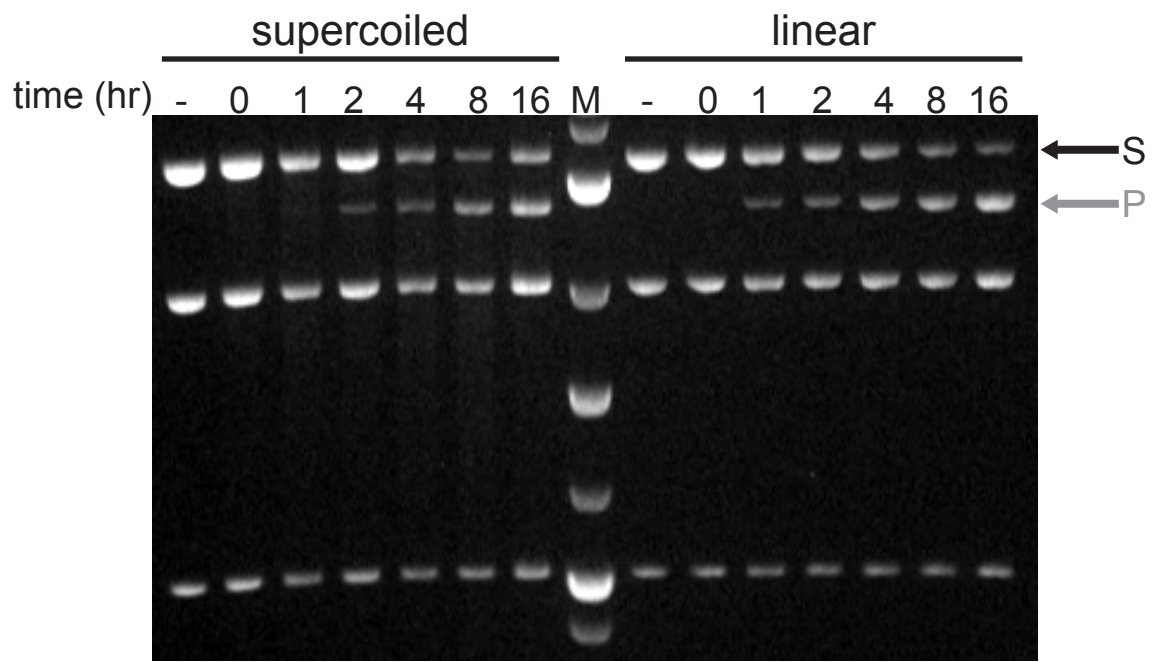


Figure 43. Role of supercoiling in excisive recombination

Excision reactions were done on a substrate that was either supercoiled, or linear. Linear substrate was made by digestion with restriction enzyme before the excision reaction.

Supercoiled reactions were also digested with the same restriction enzyme; however, this was done after recombination such that the two sets of reactions could be compared. Supercoiled reactions are shown on the left and linear reactions are shown on the right. Samples were taken at different time points and the time at which samples were incubated (in hours) is shown above each lane. The DNA size markers (M) are from top to bottom; 4, 3, 2, 1.5, 1.2, 1, and 0.9 kb.

Figure 43



indeed stimulates excisive recombination in an intramolecular substrate. To test if in the presence of RDF, the ϕ Rv1 integrase could also catalyze intermolecular *attL* x *attR* recombination, *in vitro* reactions were done two ways, either using a linear *attL* and a supercoiled *attR* or a linear *attR* and a supercoiled *attL*. In both versions of this intermolecular excision reaction, the product is a linear DNA containing *attP* and *attB*. When the *attL* and *attR* substrates are incubated with integrase and RDF, recombination does occur (Figure 44). These reactions are efficient, although the concentration requirements are similar to those seen for integrative recombination with a linear *attB*, where an excess of the linear substrate is required for efficient recombination. In both the supercoiled *attL* x linear *attR* and supercoiled *attR* x linear *attL* reactions, an approximately 100-fold excess of the linear substrate is required for efficient recombination. In these reactions with 90ng of linear substrate, approximately 90% of the plasmid substrate is converted to product (Figure 44). These intermolecular reactions are well suited to determining the minimal size required for each substrate. The minimum size of both *attL* and *attR* as linear substrate in these reactions is 47bp. Reactions using substrates of 45 or 43bp show very little product formation while a larger substrate of 51bp is recombined to product in similar amounts. These 47bp substrates are the same minimal sequences required for *attP* and *attB* (Figure 44B).

V. H. Discussion

The small basic protein encoded by the *M. tuberculosis* gene *Rv1584c* acts as a recombination directionality factor (RDF) for the prophage-like element ϕ Rv1. Expression of this RDF in mycobacteria containing an integrated ϕ Rv1 *attP* and *integrase* plasmid results in excision of the plasmid from the host genome. This is achieved by a site-specific recombination

Figure 44. Intermolecular excision

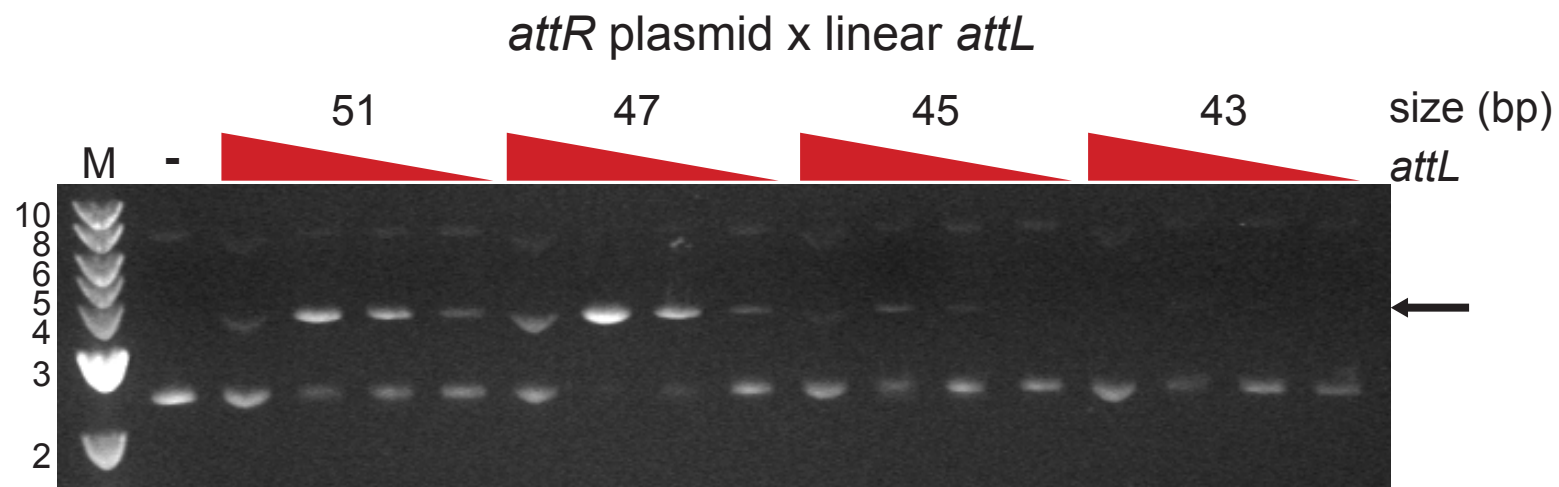
A) Intermolecular excision reactions were carried out with either a supercoiled *attR* and a linear *attL* (top panel) or a supercoiled *attL* and a linear *attR* (bottom panel). In the experiment shown in the top panel, a plasmid containing *attR* (0.04 pmoles) and linear *attL* substrates of various sizes were incubated with integrase (4 pmoles) and RDF and incubated overnight, then electrophoresed on a 0.8% agarose gel. The markers are shown in lane M, with the sizes of the bands in kb on the left. The size of the *attL* substrate used in the reactions (51, 47, 45, and 43) is shown above each set of four reactions. Each set of four has 900ng of *attL* in the first lane (27-32 pmoles), 90ng in the second lane (2.7-3.2 pmoles), 9ng in the third lane (0.27-0.32 pmoles), and 0.9ng (0.027-0.032 pmoles) in the fourth and final lane. The arrow to the right of the panel shows the position of the linear product fragment containing *attP* and *attB*. The lane marked C is a control reaction that has plasmid *attR*, 90ng of 51 bp *attL* and integrase alone. Comparison of the product formed in the reaction to that of the second lane in the 51bp *attL* set of reactions shows that the RDF is required for this reaction to occur.

In the experiment shown in the bottom panel, an *attL* plasmid and a linear *attR* substrate (0.04 pmoles) were incubated overnight with 4 pmoles of integrase and 2.7 pmoles of RDF. Each set of four has 900ng of *attL* in the first lane (27-32 pmoles), 90ng in the second lane (2.7-3.2 pmoles), 9ng in the third lane (0.27-0.32 pmoles), and 0.9ng (0.027-0.032 pmoles) in the fourth and final lane. The size of the *attR* substrate used in each set of four reactions (51, 47, 45, 43) is shown above the gel photo. The markers are shown in lane M, with the sizes of the bands in kb on the right. The arrow to the left of the panel shows the position of the linear product containing *attP* and *attB*. The lane marked C is a control reaction that has plasmid *attL*, 90ng of

51 bp *attR* and integrase alone. Comparison of the product formed in the reaction to that of the second lane in the 51bp set of reactions shows that the RDF is required for this reaction to occur.

B) The sequence of the linear *attL* and *attR* substrates used in the experiment in A are shown here.

Figure 44A



178

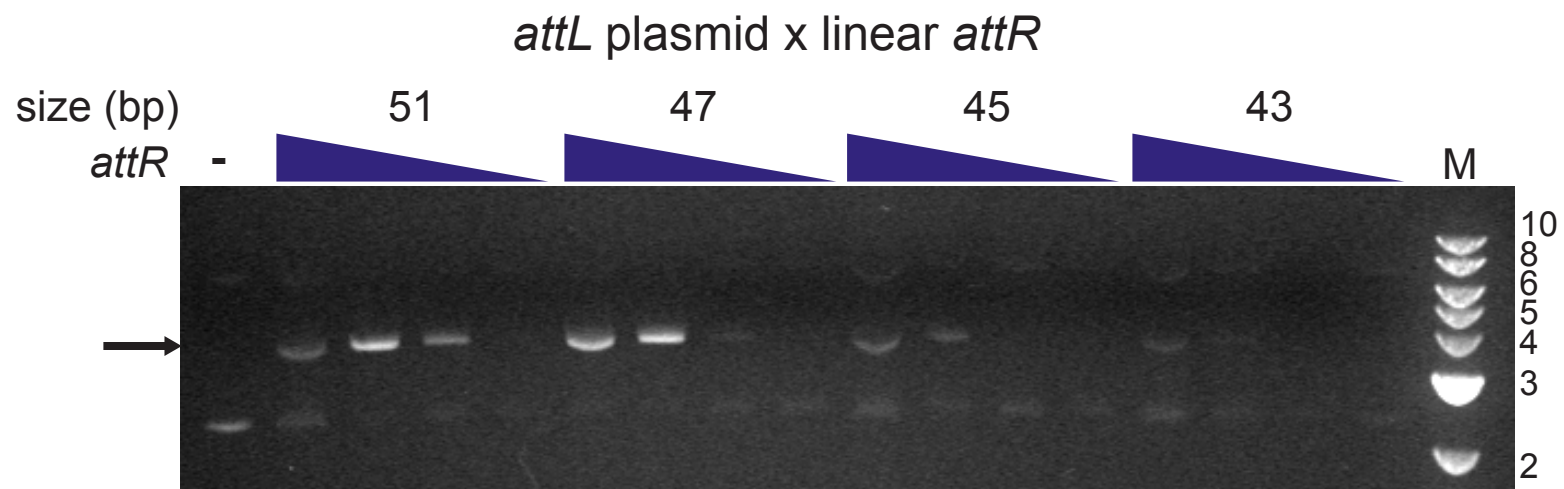


Figure 44B

attL

51 TGGAAGGTGTTGGTGC GGGGT**TGG**CCGTGGACTGCTGAAGAACATTCCACG
47 GAAGGTGTTGGTGC GGGGT**TGG**CCGTGGACTGCTGAAGAACATTCCA
45 AAGGTGTTGGTGC GGGGT**TGG**CCGTGGACTGCTGAAGAACATTCC
43 AGGTGTTGGTGC GGGGT**TGG**CCGTGGACTGCTGAAGAACATTC

attR

51 GGTGTAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGTCGAGGTGGGGTGG
47 TGTAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGTCGAGGTGGGGT
45 GTAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGTCGAGGTGGGG
43 TAGTGTATCTCACAGGTCCACGGT**TGG**CCGTGGTCGAGGTGGG

reaction between the attachment junctions *attL* and *attR*. Addition of the purified RDF protein to *in vitro* reactions with *attL* and *attR* results in excisive recombination yielding *attP* and *attB*. The presence of RDF in reactions with *attP* and *attB* causes inhibition of integrative recombination. The amount of protein required for these two activities of RDF is different. Complete inhibition of integration is seen at 80 fmoles of the purified protein while this amount of protein results in only a modest amount of excision, and the best results are obtained with 0.8 pmoles of RDF.

Excisive recombination can occur with both supercoiled and linear intramolecular substrates as well as between intermolecular substrates. All of these excision reactions, like integration, require extended incubation in order to go to completion (≥ 16 hours). The reaction is slow relative to other systems, such as Bxb1, which is complete after only two hours incubation (Kim *et al.*, 2003). Much like the intermolecular integration reactions with *attP* and *attB*, for efficient recombination, in intermolecular excision reactions an excess of the linear substrate is required, regardless of whether *attL* or *attR* is the linear form. In these reactions, the greatest product formation is seen when the linear substrate is present at approximately 100-fold molar excess over the plasmid substrate. This result gives further evidence that the requirement for excess linear substrate is a more global deficiency of integrase than a specific *attB* deficit. The intramolecular excision reaction, again like integration, does go to completion, and in this case, there are equivalent amounts of each substrate.

Although we have clearly demonstrated that Rv1584c functions as a recombination directionality factor for ϕ Rv1, whether it can function in *Mycobacterium tuberculosis* is not known. Since both the RDF and integrase are in fact functional, the mobility of the element is more likely. If a promoter for *Rv1584c* exists in ϕ Rv1, its expression must be regulated or the

gene product must be somehow suppressed to prevent the excision of ϕ Rv1. Whether or not it does in fact function in *M. tuberculosis*, in the laboratory we can use the plasmids created for the demonstration of excision in mycobacteria to remove ϕ Rv1 from *M. tuberculosis* to determine what role it may have in its physiology or virulence.

VI. MECHANISM OF ϕ Rv1 INTEGRATION AND EXCISION

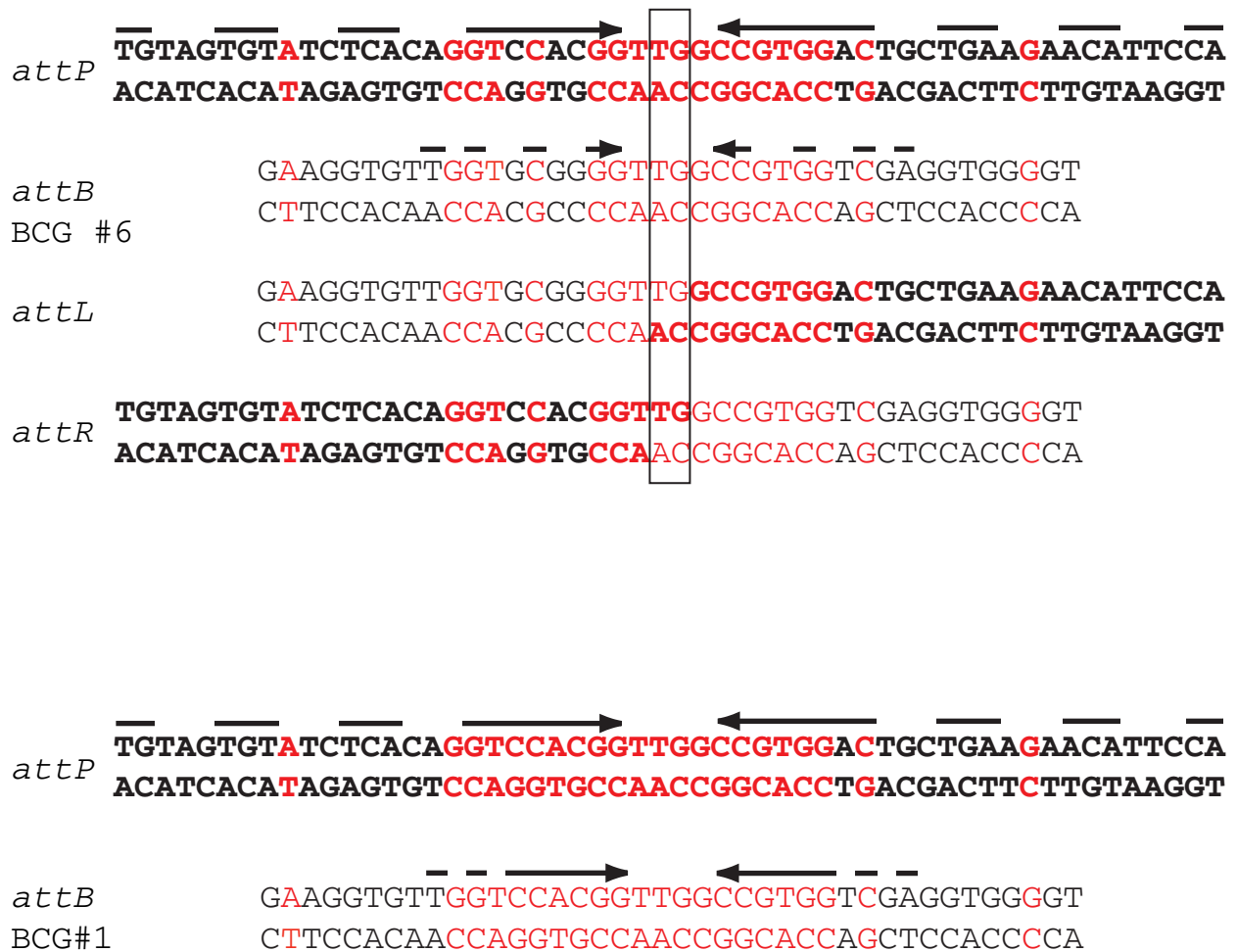
VI. A. Introduction

The ϕ Rv1 integrase catalyzes integration where *attP* and *attB* substrate sites are recombined to yield *attL* and *attR*. In the presence of the ϕ Rv1 recombination directionality factor (RDF), integrative recombination is inhibited, and excisive (*attL* x *attR*) recombination takes place. Integrase alone catalyzes only *attP* x *attB* recombination, and in the presence of RDF, integrase catalyzes only *attL* x *attR* recombination, and no other set of substrates (e.g. *attP* x *attP*, or *attL* x *attP*) are recombined *in vitro*. As shown in previous chapters, both the integrative and excisive recombination reactions require relatively small substrates. Integration reactions require just a 40bp *attB* and a 52bp *attP*, and excision requires a 47bp *attL* and a 47bp *attR*. Within the limited sequence of each site is all the information conferring its identity as *attP*, *attB*, *attL* or *attR*. Although the sites are different, the sequence flanking the central dinucleotide including the inverted repeats (IRs), is quite similar. This is shown in figure 45, where bases that are identical among the four sites are shown in red. BCG site #1 *attB* has even more identity with *attP* than site #6, with a 10 out of 10 match in the bases to the left of the central dinucleotide, and 8 out of 9 matches on the right. Yet, site #1 is a better *attB* substrate *in vitro* (Figure 27). Therefore, sequences outside of this region must contribute to identity as *attP*,

Figure 45. Attachment sites of ϕ Rv1

The sequence of each minimal attachment site is shown. The *attP* site is shown in bold, and *attB* (BCG site #6) is shown in plain text. Sequences in *attL* and *attR* are shown in plain or bold text depending on whether they are derived from *attB* or *attP*. Bases that are in common are highlighted in red. The predicted central dinucleotide (TG) is indicated by the box. Inverted repeats are shown by the horizontal arrows. At the bottom of the figure, the sequence of BCG site #1 *attB* is shown with *attP* to demonstrate the increased level of identity between these two sites, as shown by the bases in red.

Figure 45



attB, *attL* or *attR*. Each of these sites must be discriminated against the other three. We want to understand how the substrates are recognized by integrase and RDF, and how they are discriminated. The interactions of integrase with each site can not be equivalent because only specific sites are recombined. The RDF must also modulate its activity to inhibit integration, and to stimulate excision. In order to begin to understand how the integrase and RDF proteins function, DNA binding studies were done with all substrates.

These studies show that ϕ Rv1 integrase binds to all four substrate sites with different affinities. RDF also has DNA binding activity and binds to *attB* and *attL*. Addition of RDF to integrase and *attB*, *attL*, and *attP* results in a shift of integrase and the DNA to a slower migrating complex (Figures 48, 50, and 52). Addition of RDF to integrase and *attR* does not change the pattern of complex formation (Figure 54). DNaseI footprinting experiments show that integrase occupies a 47 to 60 base region in the substrate sites, while RDF binding results in a 34-36 base region of DNase I protection to the left of core in *attL* and *attB* (Figures 49, 51, 53, and 55). This RDF binding site is dispensable for both inhibition of integration and excision. This suggests that the interaction of RDF with integrase is essential for its activity. When integrase, *attP* and *attB* are present in a band shift assay, additional bands are seen (Figure 59). There is a complex with a slow migration that is likely a synaptic complex with integrase, *attP* and *attB*, as well as two additional bands that we believe are free and integrase bound product. These additional bands are not seen when the reactions are incubated on ice or in the presence of RDF, thus, RDF acts at a step before synapsis. A similar synaptic complex is seen when *attL*, *attR*, integrase and RDF are present, in the absence of RDF, no synapsis is seen, thus RDF is likely to act at this step in recombination (Figure 60).

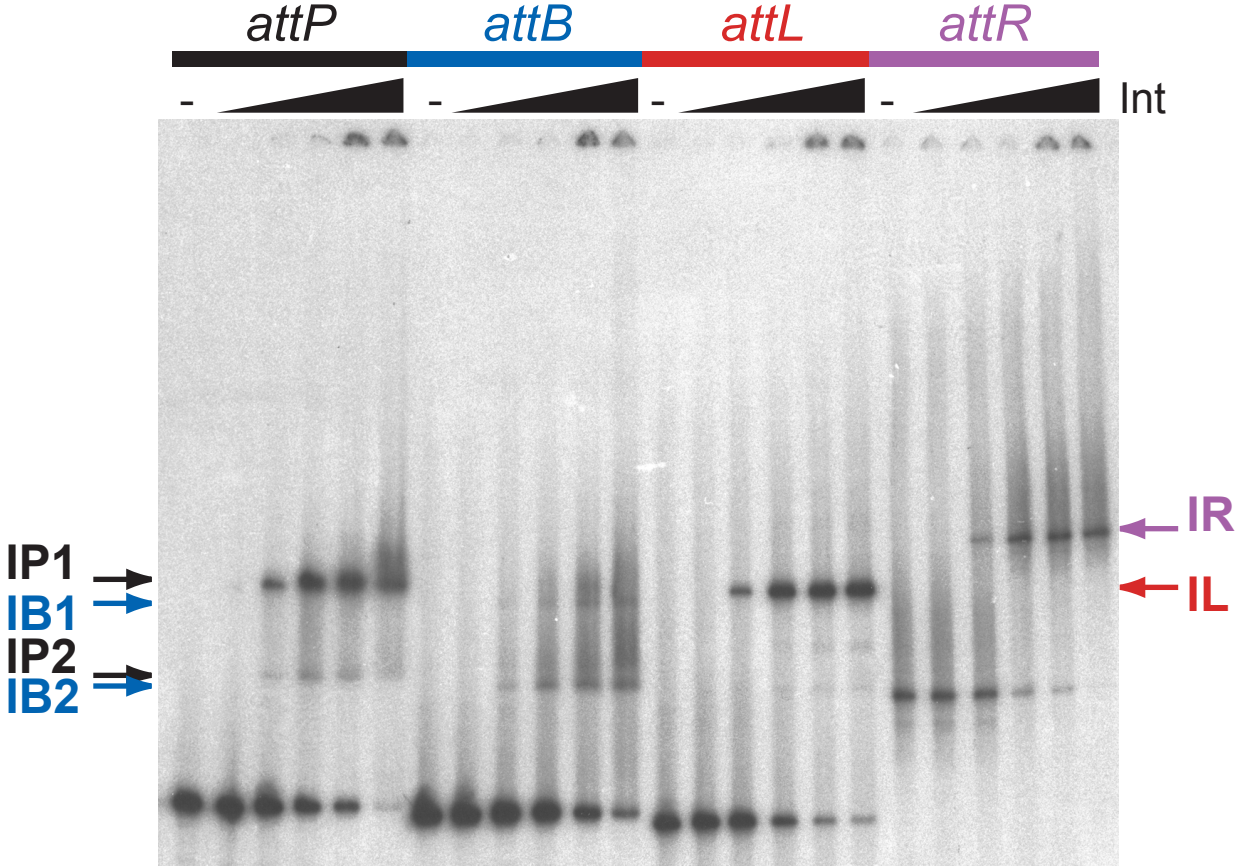
VI. B. ϕ Rv1 integrase binding to substrate sites

To catalyze both integrative and excisive recombination, the integrase must recognize and associate with each of its substrate sites. In order to detect the interaction between integrase and the substrate sites, DNA binding assays were performed with integrase and each attachment site. To do this, a radiolabeled DNA fragment containing one of the substrate sites was incubated with integrase until equilibrium was reached. This mixture was then loaded onto a polyacrylamide gel. These experiments show that integrase binds to all four sites, as evidenced by a shift of the free DNA to a slower migrating species (Figure 46). Although integrase binds to each site, the complexes formed with each site differ, as do the kinetics of complex formation. From these band shifts, an apparent affinity can be calculated as dissociation constant (K_d), which is the concentration of protein at which 50% of the DNA is bound, and 50% is unbound. These estimated affinities are as follows; *attB*-400nM, *attP*-130nM, *attL*-40nM, and *attR*-40nM. When integrase binds to *attB* or *attP*, two types of complexes are formed (IB1, IB2, or IP1, IP2, respectively) (Figure 46). These two complexes are formed with the same kinetics, as both 1 and 2 appear at the same concentrations, although they are not formed in the same amount, and the slower migrating complex (IP1, IB1) is the predominant form. When an *attP* and *attB* of the same size are used, these complexes formed by integrase binding have approximately the same migration. This result suggests that these complexes are similar in size and shape and that integrase of an equivalent quaternary structure (monomer, dimer, etc.) is likely to be binding to *attB* and *attP*. Integrase also binds to *attL* and *attR*, however, only a single complex is seen with each site (IL or IR). The *attL* site used in this experiment is approximately the same size as *attP*

Figure 46. Integrase binding to attachment sites

DNA band shift assays were done using ϕ Rv1 integrase and each of the attachment sites. A ^{32}P labeled DNA fragment was incubated with increasing amounts of integrase, and protein-DNA complexes were separated from free DNA on a non-denaturing polyacrylamide gel. A set of six reactions is shown for each site, as indicated by the colored bars and matching text above the gel. In each set, the first lane (-) has no integrase added and shows the migration of the free DNA molecule. The next lane in each set has $0.013\mu\text{M}$ integrase, the lane marked 2 has $0.04\mu\text{M}$ integrase. Lane 3 has integrase at $0.13\mu\text{M}$, the next lane has $0.4\mu\text{M}$ integrase, and the final lane has $1.3\mu\text{M}$ ϕ Rv1 integrase. The migration of the two complexes formed when integrase binds to *attP* are indicated by the black arrows on the left (IP1 and IP2). Complexes formed by the interaction of integrase with *attB* are indicated by blue arrows also on the left (IB1 and IB2). The complexes formed by the binding of integrase to *attR* (IR) or *attL* (IL) are shown on the right by the violet and red arrows, respectively.

Figure 46



and *attB*, and the migration of the complex formed by integrase binding to *attL* is similar to the predominant form in *attB* and *attP* binding (IP1, IB1) (Figure 46). This result suggests that integrase binds integration substrates and excision substrates differently; it binds in two ways to *attP* and *attB*, but in only one way to *attL* and *attR*.

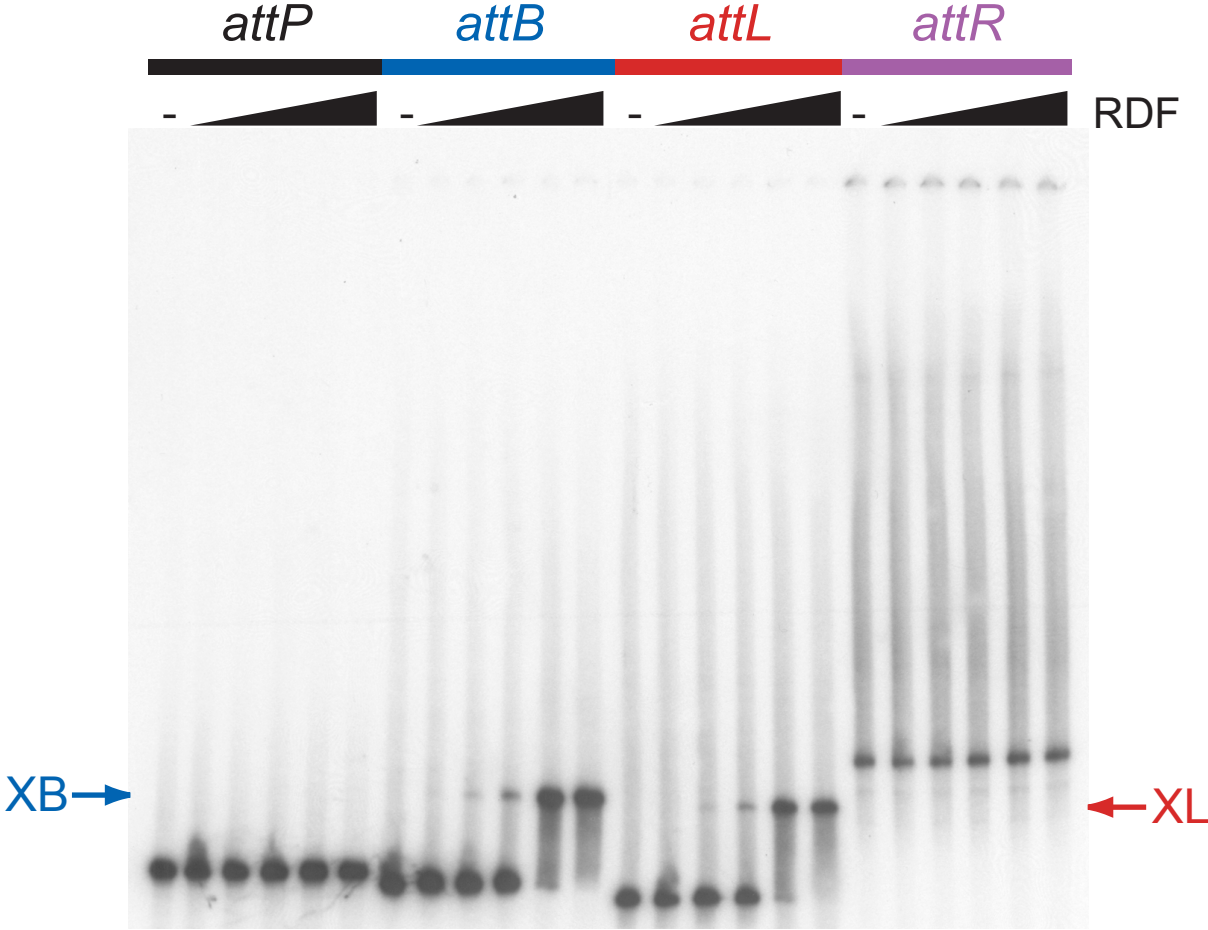
VI. C. RDF binding to substrate sites

Recombination directionality factors in other phage systems bind to specific DNA sequences. For example, the L5 RDF binds to *attP* and *attR* at four sites with the sequence 5'-CTTCNAC-3'. These are spaced at approximately 10 bp intervals (Lewis and Hatfull, 2003). To determine if the ϕ Rv1 RDF exhibits DNA binding activity, band shift assays were done with RDF and each of the integration and excision substrate sites. Addition of the recombination directionality factor to a labeled *attB* or *attL* results in a shift of the free DNA (Figure 47). In contrast, addition of RDF to *attR* or *attP* does not change the migration of the DNA. This demonstrates that RDF is a sequence-specific DNA-binding protein. In these experiments, a single complex is seen with both *attB* (XB) and *attL* (XL) and these complexes are formed with the same kinetics. RDF at 80nM shows very little complex formation, while at 267nM RDF, nearly all the DNA has been shifted into complex. When substrates of a similar size are used, these complexes have a similar migration, which suggests that RDF binds similarly to both sites and probably binds to the same sequence. Because RDF shifts *attL* and *attB*, but not *attP* or *attR*, RDF is likely binding to the left of the core in *attL* and *attB*, since both sites share sequence in this region (Figure 45).

Figure 47. RDF binding to attachment sites

DNA band shift assays were done using purified ϕ Rv1 Recombination Directionality Factor (RDF) and each of the attachment sites. A ^{32}P labeled DNA fragment containing an attachment site (*attP*, *attB*, *attL* or *attR*) was incubated with increasing amounts of RDF, and protein-DNA complexes were separated from the free DNA on a non-denaturing polyacrylamide gel. A set of six reactions is shown for each site as indicated above the gel. The first lane (-) in each set has no protein added and shows the migration of the free DNA molecule. The next lane in each set has 8nM RDF added. The lane marked 2 has 27nM RDF. Lane 3 has 80nM RDF, the next lane (4) has 270nM, and the final lane has 800nM ϕ Rv1 RDF. The migration of the complex formed by the interaction of RDF with *attB* (XB) or *attL* (XL) is indicated by blue and red arrows, respectively.

Figure 47



VI. D. Integrase and RDF interactions at *attB*

In the preceding chapter, the ability of the ϕ Rv1 RDF to inhibit integration and promote excision *in vitro* was demonstrated. We hope to gain insight into how RDF functions to carry out these two distinct activities by examining complex formation when both integrase and RDF are incubated with each attachment site. To this end, RDF was titrated into binding reactions with a constant concentration of integrase (1.3 μ M), and integrase was titrated into reactions with a constant amount of RDF (0.8 μ M). Addition of RDF to *attB* plus integrase shifts the DNA-integrase complexes (IB1, IB2) to a slower migrating complex (IXB1) (Figure 48). This shift occurs when RDF is at 0.27 μ M. This is the same concentration of RDF that shifts a significant amount of the free *attB*. At the highest tested RDF concentration (0.8 μ M), an even slower migrating complex (IXB2) is visible, as well as a small amount of the RDF-*attB* complex (XB). With the appearance of XB, there is also a concomitant reduction in the amount of free *attB* DNA. This indicates that the XB complex is likely to be formed by a shift of the free *attB* and not by displacement of integrase from the int-*attB* complex. Addition of Integrase to RDF with *attB* shifts the RDF *attB* complex (XB) to a slower migrating species, and the migration of this complex is identical to IXB2 (Figure 48). This complex is expected instead of IXB1 since in this set of reactions the RDF is at 0.8 μ M, the concentration at which the IXB2 complex is seen when RDF is titrated into integrase *attB* reactions. When additional integrase is added, more IXB2 complex is seen as well as a reduced amount of XB. This evidence suggests that both Integrase and RDF may be occupying the same DNA.

Solution DNaseI footprinting of integrase, RDF, and both integrase and RDF at *attB* was also done. Addition of integrase to *attB* showed protection over 47 bases including the central dinucleotide, and protection is seen at 0.24 μ M integrase (Figure 49A). This region of protection

Figure 48. Integrase and RDF binding to *attB*

DNA band shift assays were done using ϕ Rv1 Integrase, Recombination Directionality Factor (RDF) and *attB*. A ^{32}P labeled DNA fragment containing *attB* was incubated with increasingly concentrated integrase as indicated by the black triangle and box (13nM, 40nM, 130nM, 400nM, and 1.3 μM). Integrase was then held at the highest concentration of (1.3 μM) and RDF was titrated in (green triangle), and protein-DNA complexes were separated from the free DNA on a non-denaturing polyacrylamide gel. This is shown on the left half of the gel. The converse was also done where increasing amounts of RDF (8nM, 27nM, 80nM, 270nM, 800nM) were added to the *attB* DNA, then at the highest concentration of RDF (800nM), integrase was added in at different concentrations. This experiment is shown on the right. Complexes seen with integrase alone are indicated by the blue arrows (IB1, IB2). The complex that results from RDF interacting with *attB* (XB) is shown by the green arrow, and the two complexes formed when integrase and RDF are present are indicated by the violet arrows (IXB1, IXB2).

Figure 48

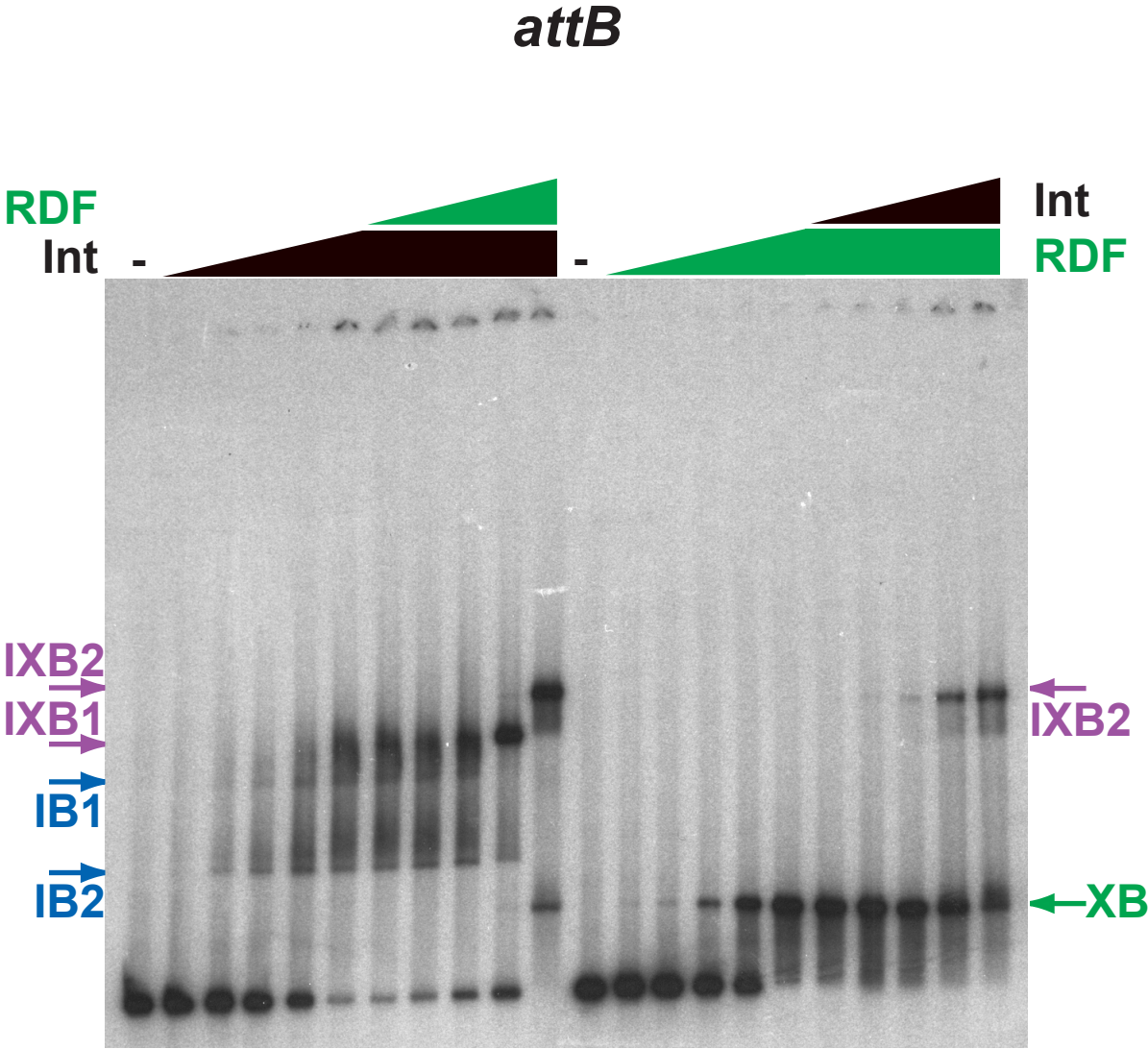
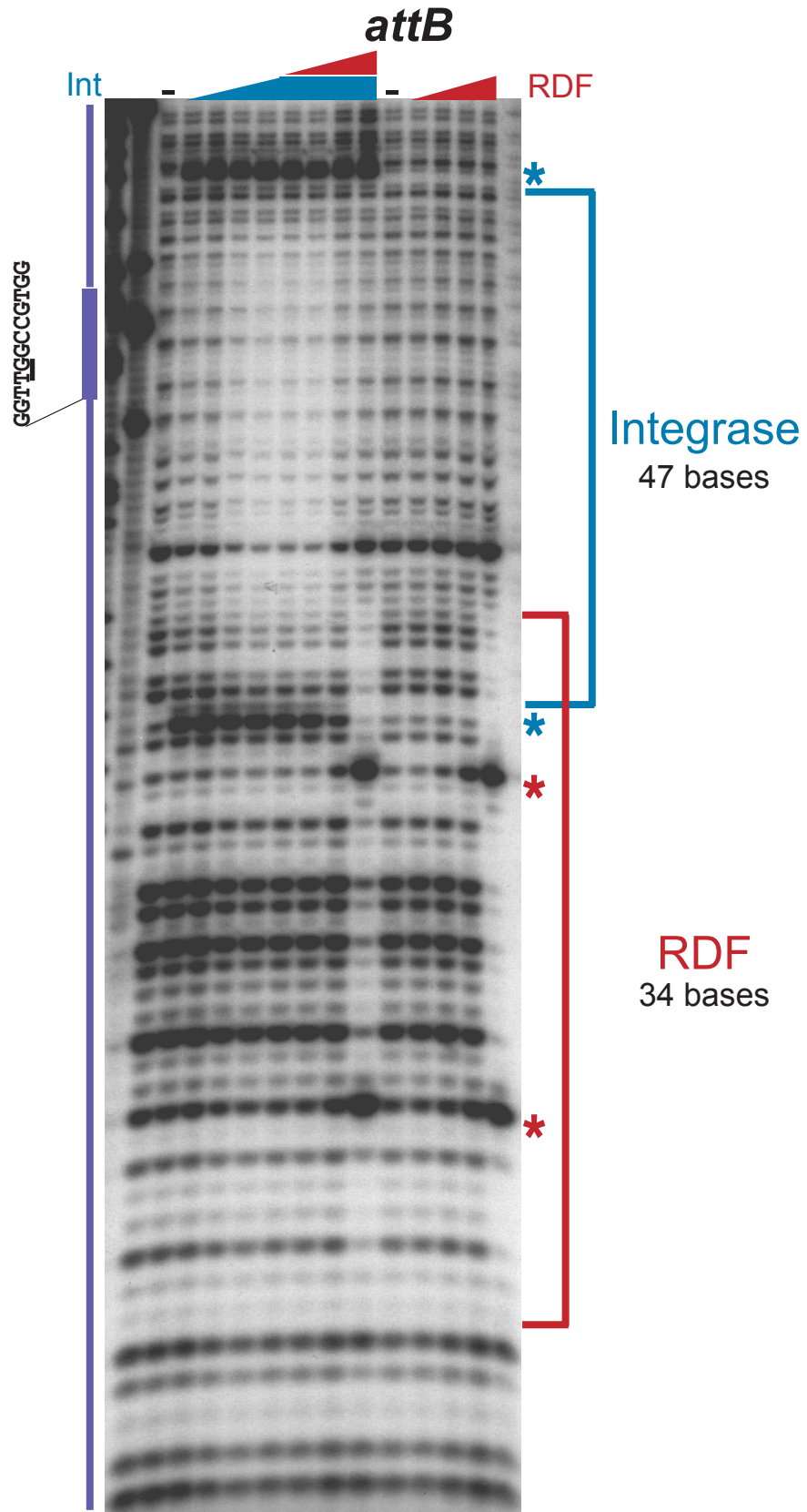


Figure 49. DnaseI footprinting of integrase and RDF at *attB*

A) DNaseI footprinting reactions were carried out on *attB* with Integrase and/or RDF. The first lane (-) shows the pattern of DNaseI digestion in the absence of protein. In the next four lanes, increasing amounts of Integrase (13nM (2), 40nM (3), 130nM (4), and 400nM (5)) were incubated with *attB*. A region of 47 bases overlapping the core is protected from digestion (indicated by the blue bracket). In the following four lanes, integrase is held at 400nM, and increasing amounts of RDF are present (8nM, 27nM, 80nM, 270nM). With increasing amounts of RDF, there is a reduction in the protection by integrase, and at the highest amount of RDF (270nM), an additional region of protection is seen to the left of the core. The next lane (-) has no protein added. The remaining four lanes have increasing amounts of RDF (8nM, 27nM, 80nM, 270nM). At the highest amount of RDF, an area of protection is seen to the left of the core (red bracket). This 34 base protected region includes two prominent enhancements (red stars).

B) Sequence of the *attB* region where protection is seen. The blue bar shown above the sequence indicates the region of integrase protection. The blue stars show where enhancements are seen. The red bar indicates the region of RDF protection, and red stars show the position of enhancements. The area of the bar that is both blue and red is where protection is seen for both integrase and RDF. The position of direct repeats (5'-TCGTNGTGG-3') that may be possible recognition sequences for RDF are shown by horizontal arrows in between the two strands. The arrows below the sequence show the indirect repeats flanking the central dinucleotide (boxed) in the core (uppercase). The black bases indicate the sequence of the minimal 40bp substrate.

Figure 49A



is flanked by enhancements (Figure 49A). Binding of RDF to *attB* shows protection over 34 bases to the left of the central dinucleotide, and this pattern is only seen at the highest concentration of RDF (0.27 μ M) (Figure 49A). The binding sites for these two proteins overlap slightly, and when both RDF and integrase are added together with *attB*, the pattern of protection shifts more to that of the RDF alone binding reactions; there is a reduction in the level of protection by integrase, and a footprint characteristic of RDF binding is seen. However, the integrase dependent enhancement at one end of the binding site (not overlapping the RDF binding site) is still present suggesting that the binding of integrase is not abolished, but instead is altered in some way. Reduction in the integrase protection is seen at 80nM RDF, but the RDF protection pattern is only seen at 0.27 μ M. This pattern may be indicative of RDF and integrase occupying *attB* at the same time. When the sequence of the region of *attB* protected by RDF is examined, two direct repeats are revealed with the sequence 5'-TCGTNGTGG-3' (Figure 49B). The first base of these repeats is spaced 20 bases apart. It is likely that these sequences represent recognition sequences for RDF. The enhancements seen in the footprint are at position 5 (N) of the repeat one (actually a 'T'), and at position 1 (T) of repeat two.

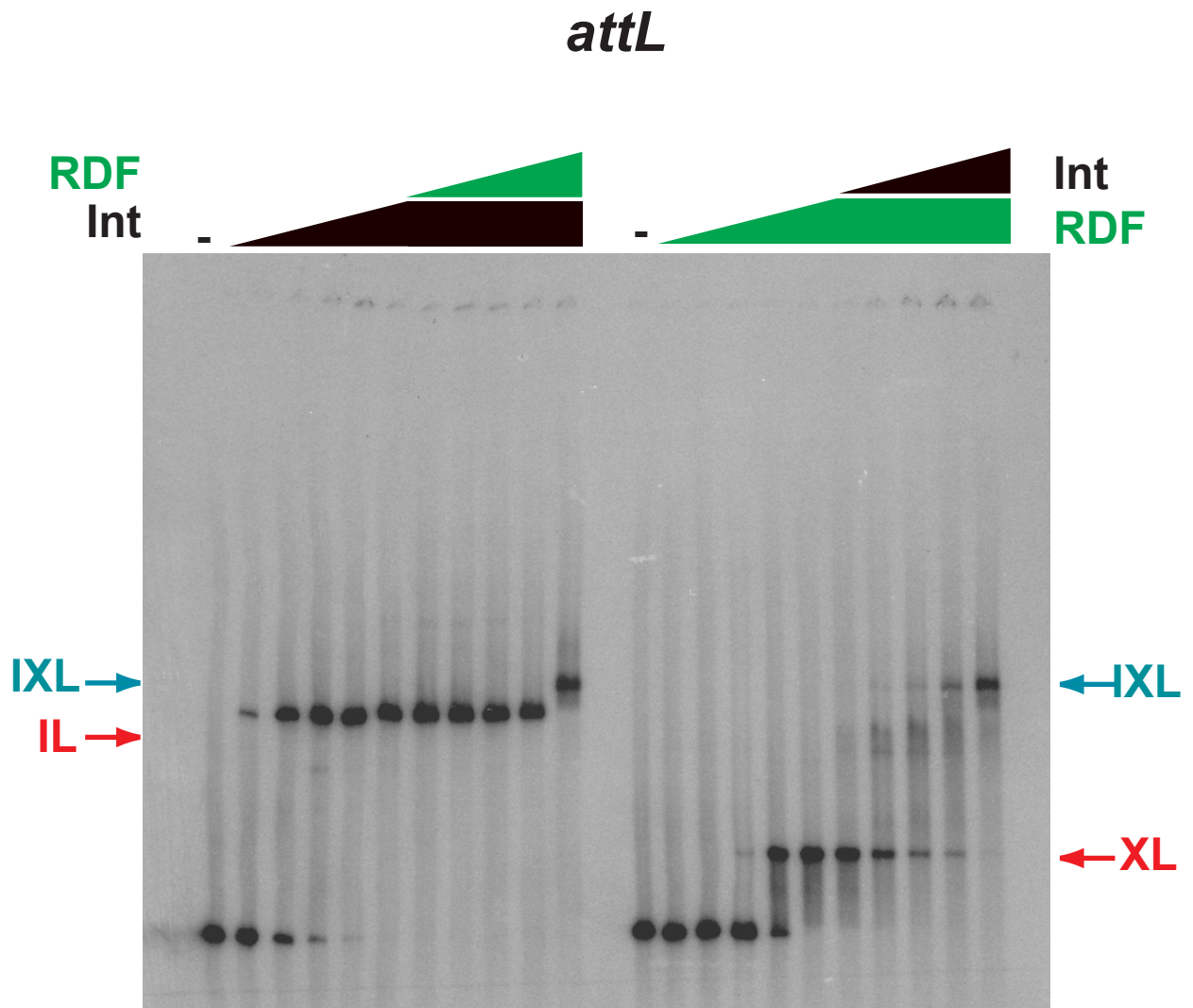
VI. E. Integrase and RDF interactions at *attL*

The pattern of complex formation with integrase and RDF at *attL* is similar to that seen at *attB*. Addition of RDF to *attL* binding reactions that contain a constant amount of integrase results in a shift of the integrase-*attL* complex (IL) to a species with a higher migration (IXL) (Figure 50). However, unlike when RDF and integrase are present with *attB*, only one complex is formed. This shift of IL to IXL is only seen at the highest concentration of RDF (0.8 μ M). At this concentration in RDF only binding reactions, all of the free *attL* is shifted into complex.

Figure 50. Integrase and RDF binding to *attL*

DNA band shift assays were done using ϕ Rv1 Integrase, Recombination Directionality Factor (RDF) and *attL*. A ^{32}P labeled *attL* fragment was incubated with increasing amounts of Integrase (13nM, 40nM, 130nM, 400nM, and 1.3 μM). Integrase was then held at the highest concentration (1.3 μM), RDF was titrated in, and protein-DNA complexes were separated from the free DNA on a non-denaturing polyacrylamide gel. This is shown on the left half of the gel. The converse was also done where increasing amounts of RDF (8nM, 27nM, 80nM, 270nM, 800nM) were added to the *attL* DNA, then at the highest amount of RDF (800nM), integrase was added in at different amounts. This experiment is shown on the right. The migration of the complex seen when integrase alone is added is indicated by the red arrow at the left (IL). The complex that results from RDF interacting with *attL* (XL) is shown by the black arrow at the right, and the complex formed when integrase and RDF are both present is indicated by the gray arrows (IXL).

Figure 50



Addition of integrase to reactions with *attL* and a constant amount of RDF, shifts the RDF-*attL* complex (XL) to a higher migrating band, with a migration identical to the complex seen when RDF is titrated into reactions with integrase (IXL) (Figure 50). Although with *attL* and integrase alone, only a single complex (IL) is seen, when integrase is titrated in the presence of RDF, an additional complex (IL[X]) is seen. This complex is smeary and has an intermediate migration, located in between XL and IXL. IL[X] is visible at lower integrase concentrations, below those at which the IXL complex is seen, and the intermediate complex disappears when IXL is present. Also, when integrase is titrated into reactions with RDF and *attL*, the concentration of integrase at which the RDF-*attL* complex (XL) disappears is different from the concentration which the RDF-*attB* complex (XB) disappears. At 0.8 μ M integrase, the RDF-*attL* complex is nearly gone while some *attB*-RDF (XB) complex is present at this integrase concentration. This difference may reflect the increased affinity of integrase for *attL*.

DNaseI solution footprinting was done to examine integrase and RDF interactions with *attL*. Integrase binding to *attL* results in a 55 base footprint overlapping the core region, and this protection is seen at 2.4 μ M Integrase (Figure 51). Addition of RDF to these integrase-*attL* binding reactions does not greatly affect the integrase protection pattern. However, there is a slight reduction in the level of protection by integrase, and an additional region of protection is seen including two enhancements. The additional pattern is equivalent to the pattern of protection seen with RDF alone, where a 36 base footprint is present with two prominent enhancements. The sequence of the *attL* region protected by RDF is identical to that of the protection region seen when RDF is bound to *attB*, except that here a slightly larger region is protected (36 instead of 34) (Figure 51B and 49B). The position of the enhancements is also

Figure 51. DnaseI footprinting of integrase and RDF at *attL*

A) DNaseI footprinting reactions were carried out on *attL* with Integrase and/or RDF. The first lane (-) shows the pattern of DNaseI digestion in the absence of protein. In the next four lanes, increasing amounts of Integrase were incubated with *attL*. Integrase was added at 40nM (2), 130nM (3), 400nM (4), and 1.3 μ M (5). In these lanes, a region of 55 bases overlapping the core is protected from DNaseI digestion (indicated by the blue bracket). In the following four lanes, integrase is held at 1.3 μ M, and increasing amounts of RDF are added (27nM, 80nM, 270nM, 800nM). At the highest amount of RDF, (800nM), there is little reduction in the protection by integrase and an additional protection region is seen to the left of the core. The next lane (-) has no protein added. The remaining four lanes have increasing amounts of RDF added (27nM, 80nM, 270nM, 800nM). At the highest amount of RDF, an area of protection is seen to the left of the core (indicated by the red bracket). This 35 base protected region includes two prominent enhancements (red stars).

B) Sequence of the *attL* region where protection by integrase is seen. The blue bar shown above the sequence indicates the region of protection by integrase. The blue stars show where enhancements are seen. The red bar shown above the sequence indicates the region of RDF protection. The red stars show the position of enhancements. The area of the bar where both blue and red are shown is where protection is seen for both integrase and RDF. The black horizontal arrows in between the two strands show the position of direct repeats (5'-TCGTNGTGG-3') that may be possible recognition sequences for RDF. The central dinucleotide is shown as the two boxed base pairs box in the 12bp core (uppercase sequence). The black bases indicate the sequence of the minimal 47bp substrate.

Figure 51A

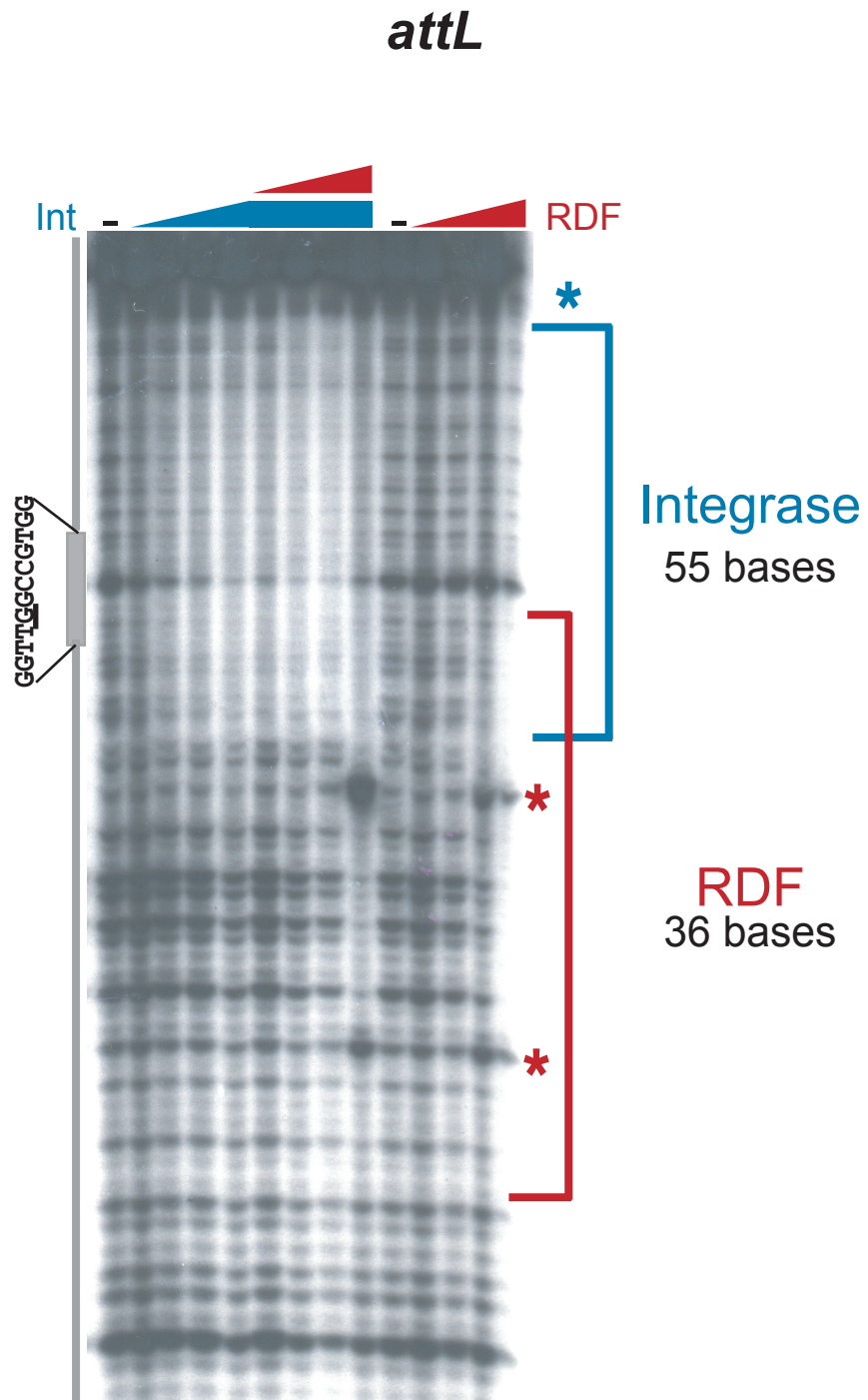


Figure 51B

attL

RDF

Integrase

204

ctcgtcgttgtggcgtagcagcttctcgtggtggtggaaggtggtggtgcggGGTTGGCCGTGGactgctgaagaacattccacgcca
gagcagcaacaccgcatcgtcgaagagcaccacaccttccacaaccacgcccCAACCGGCACctgacgacttcttgtaaggtgcggt



identical in both the *attL* and *attB* footprints, at position 5 (N) of the repeat one (actually a 'T'), and at position 1 (T) of repeat two.

VI. F. Integrase and RDF interactions at *attP*

Although RDF does not associate directly with *attP*, adding RDF to *attP*-integrase binding reactions results in a change in the pattern of complex formation. As was seen with both *attB* and *attL*, RDF addition to *attP*-Integrase shifts the DNA-integrase complexes IP1 and IP2 to a higher migrating complex (IXP) (Figure 52). In the absence of RDF, at the highest concentration of integrase, there is some material present at the position of the IXP complex, but it is smeary. Addition of RDF results in the formation of a clear complex. This complex is present at the two highest concentrations of RDF (0.24 μ M and 0.8 μ M), and both integrase-*attP* complexes are completely gone. When integrase is titrated into RDF-*attP* binding reactions, the free DNA is shifted into a complex (IXP) with a migration identical to that seen when RDF is titrated into *attP*-integrase binding reactions. In this set of reactions, the IXP complex is seen even at the lowest concentration of integrase provided, which is 24nM. A small amount of presumably integrase only complexes (IP1 and IP2) are also seen (Figure 52). Since RDF does not bind to *attP* directly, the shift of *attP* and integrase by RDF suggests that the RDF interacts directly with integrase, either as a member of this complex, or by inducing a conformational change in the *attP*-Integrase complex.

Similar DNA binding reactions of integrase and RDF with *attP* were subjected to solution DNaseI footprinting. Incubation of *attP* with RDF did not show any protection (Figure 53A). This result was not surprising since no shift of *attP* is seen when RDF alone is incubated with *attP* in a band shift assay. Also, the sequence of *attB* is quite different from *attP* in the

Figure 52. Integrase and RDF binding to *attP*

DNA band shift assays were done using ϕ Rv1 Integrase and the Recombination Directionality Factor (RDF) and *attP*. A ^{32}P labeled *attP* fragment was incubated with increasing amounts of Integrase (13nM, 40nM, 130nM, 400nM, and 1.3 μM). Then at the highest concentration of Integrase (1.3 μM), RDF was titrated in, and protein-DNA complexes were separated from the free DNA on a non-denaturing polyacrylamide gel. This is shown on the left half of the gel. The converse was also done where increasing amounts of RDF (8nM, 27nM, 80nM, 270nM, 800nM) were added to the *attP* DNA, then at the highest concentration of RDF, integrase was added in at different concentrations. This experiment is shown on the right. The migration of the complexes seen when integrase alone is added are indicated by the black arrows at the left (IP1, IP2), and the complex formed when integrase and RDF are both present is indicated by the teal arrow (IXP).

Figure 52

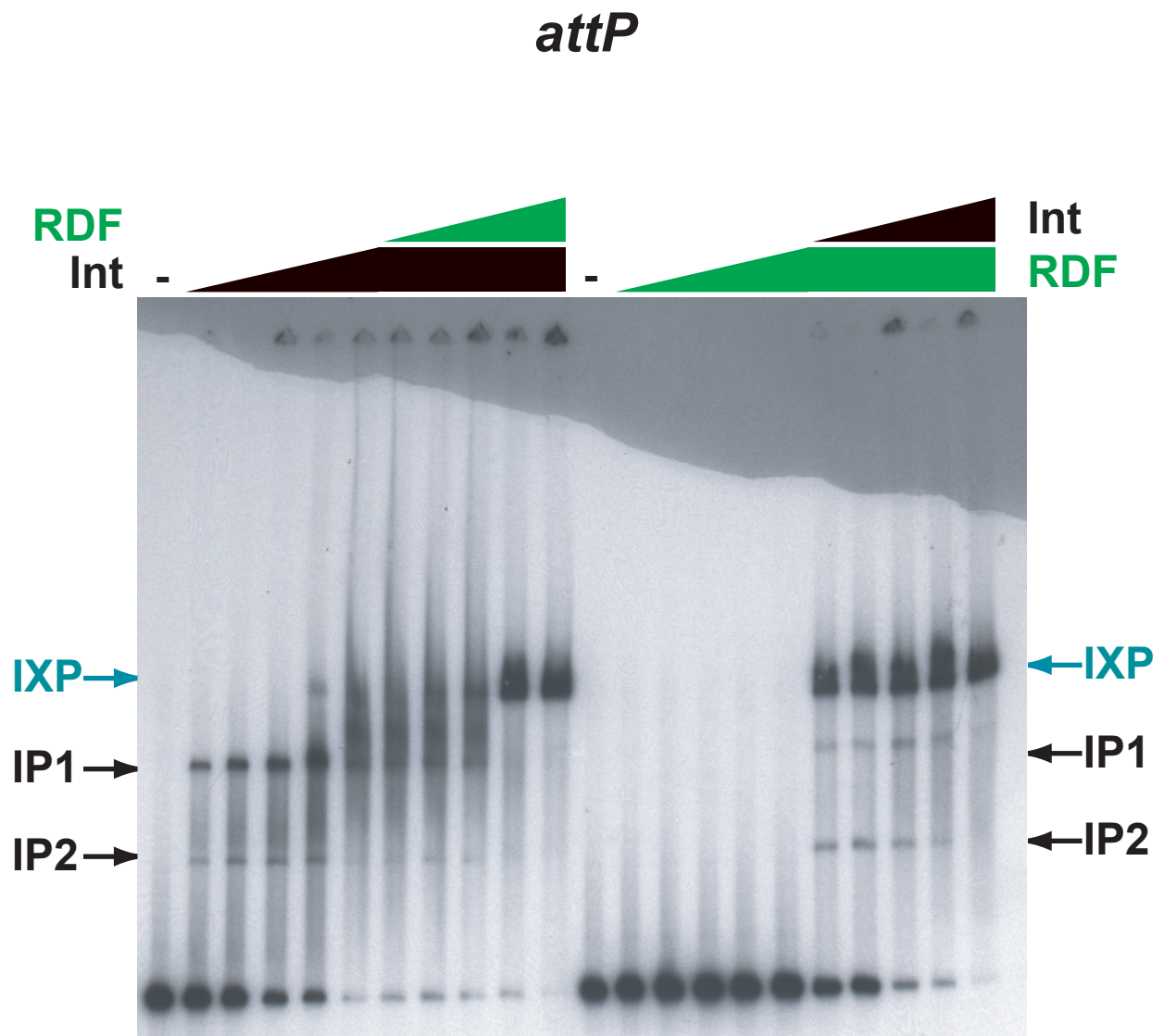


Figure 53. DnaseI footprinting of integrase and RDF at *attP*

A) DNaseI footprinting was done on *attP* with Integrase and/or RDF binding reactions. The first lane (-) shows the pattern of *attP* digestion by DNaseI in the absence of protein. In the next five lanes, increasing amounts of Integrase were incubated with *attP*. Integrase was added at 13nM (2), 40nM (3), 130nM (4), 400nM (5), and 1.3 μ M. In these lanes, a region of 60 bases overlapping the core is protected from DNaseI digestion (indicated by the blue bracket). In the following four lanes, integrase is held at 1.3 μ M, and increasing amounts of RDF are added (8nM, 27nM, 80nM, 270nM, 800nM). Even at the highest amount of RDF, (800nM), there is no change in the pattern of protection by integrase. The next lane (-) has no protein added. The remaining four lanes have increasing amounts of RDF. There is no protection of *attP* by RDF seen at any amount of RDF added.

B) Sequence of the *attP* region where protection by integrase is seen. The blue bar shown above the sequence indicates the region of protection by integrase. The blue stars show where enhancements are seen. The horizontal arrows below the sequence show the indirect repeats that flank the central dinucleotide (boxed) in the 12bp core (uppercase sequence). The black bases indicate the sequence of the minimal 52bp substrate. The bottom figure shows a comparison of *attP* and *attB* highlighting the region of *attB* where RDF binds (violet).

Figure 53A

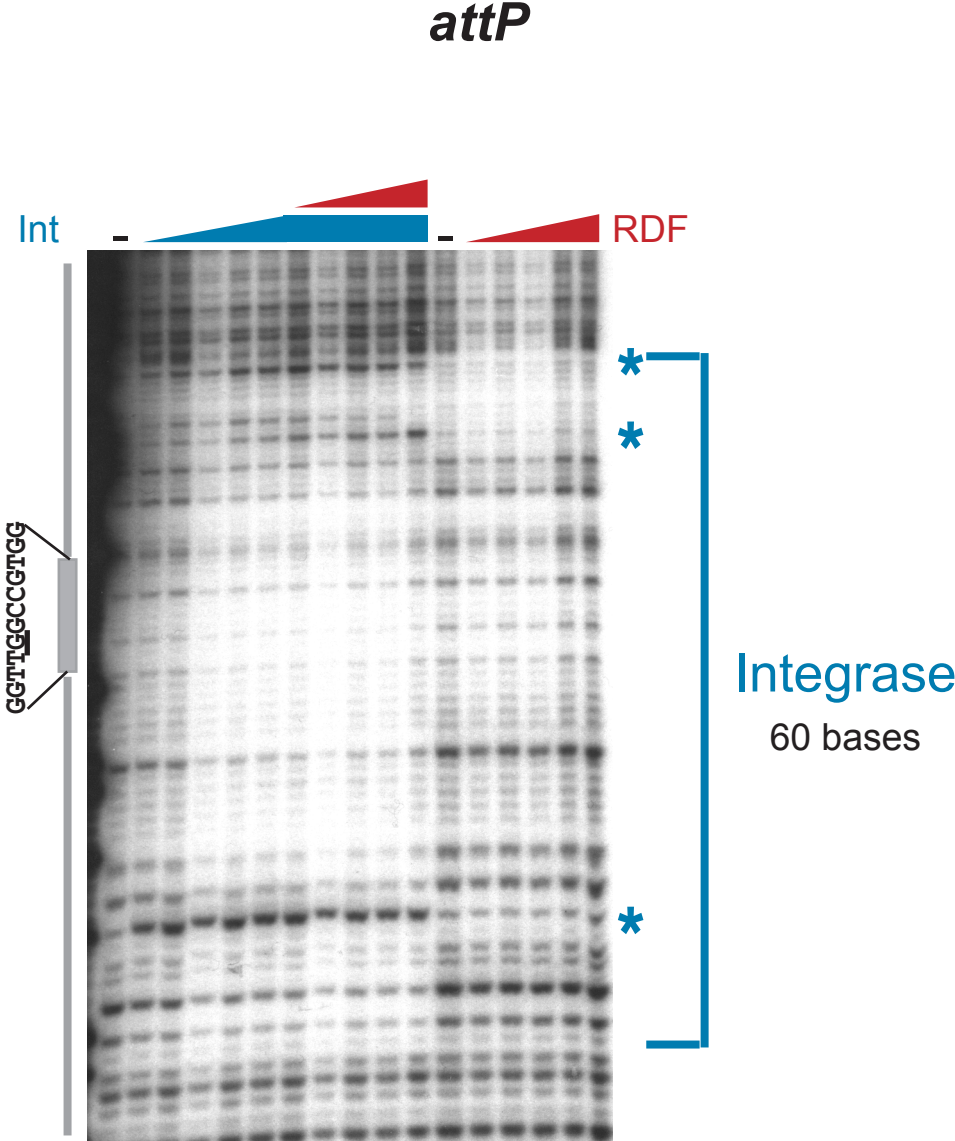


Figure 53B

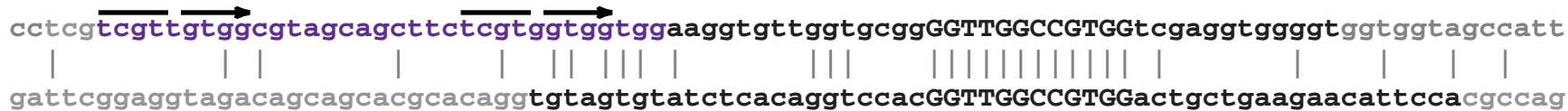
attP

Integrase



210

attB



attP

region where RDF binds (Figure 53B), thus we did not expect to see any protection of *attP* by RDF. Incubation of integrase with *attP* does result in a region of DNaseI protection that includes the core region (Figure 53A). This region spans 60 bases and there are three enhancements seen near the ends of this footprint. The addition of RDF to integrase and *attP* reactions does not change the pattern of footprinting by integrase, and no additional protection is seen (Figure 53A). This result, along with the complex seen when both integrase and RDF are added to *attP* in a band shift assay, suggest that the RDF is associating directly with the integrase that is bound to *attP*.

VI. G. Integrase and RDF interactions at *attR*

The interactions of RDF and integrase at *attR*, are still different from those seen at *attB*, *attL*, or *attP*. Unlike at the other three sites, addition of RDF to *attR* plus integrase does not change the pattern of complexes, and only the single integrase-*attR* complex (IR) is seen (Figure 54). Since RDF does not bind *attR* directly we would not expect a shift from RDF. However, when integrase is titrated into RDF plus *attR* binding reactions, two complexes are seen. The slower migrating one has the same mobility as the integrase-*attR* complex seen with integrase binding alone (IR). The second faster migrating complex (I(X)R) is only seen in binding reactions with RDF and integrase (Figure 54). This second complex is somewhat similar to what is seen when integrase is added to reactions with *attL* and RDF. The additional complex may be a different form of integrase binding to *attR*. For example, the upper complex may be a dimer and the lower complex may have a monomer of integrase bound to *attR*. The binding of integrase in this form seems to be mediated by RDF however, whether RDF is present in this complex is not known.

Figure 54. Integrase and RDF binding to *attR*

DNA band shift assays were done using ϕ Rv1 Integrase and the Recombination Directionality Factor (RDF) and *attR*. A ^{32}P labeled *attR* fragment was incubated with increasing amounts of Integrase (13nM, 40nM, 130nM, 400nM, and 1.3 μ M) then at the highest concentration of Integrase (1.3 μ M), RDF was titrated in, and protein-DNA complexes were separated from the free DNA on a non-denaturing polyacrylamide gel. This is shown on the left half of the gel. The converse was also done where increasing amounts of RDF (8nM, 27nM, 80nM, 270nM, 800nM) were added to the *attR* DNA, then at the highest concentration of RDF, integrase was added in at different concentrations. This experiment is shown on the right. The migration of the complexes seen when integrase alone is added are indicated by the black arrows at the left and right (IR), and the complex formed when integrase and RDF are both present is indicated by the green arrow at the right (I(X)R).

Figure 54

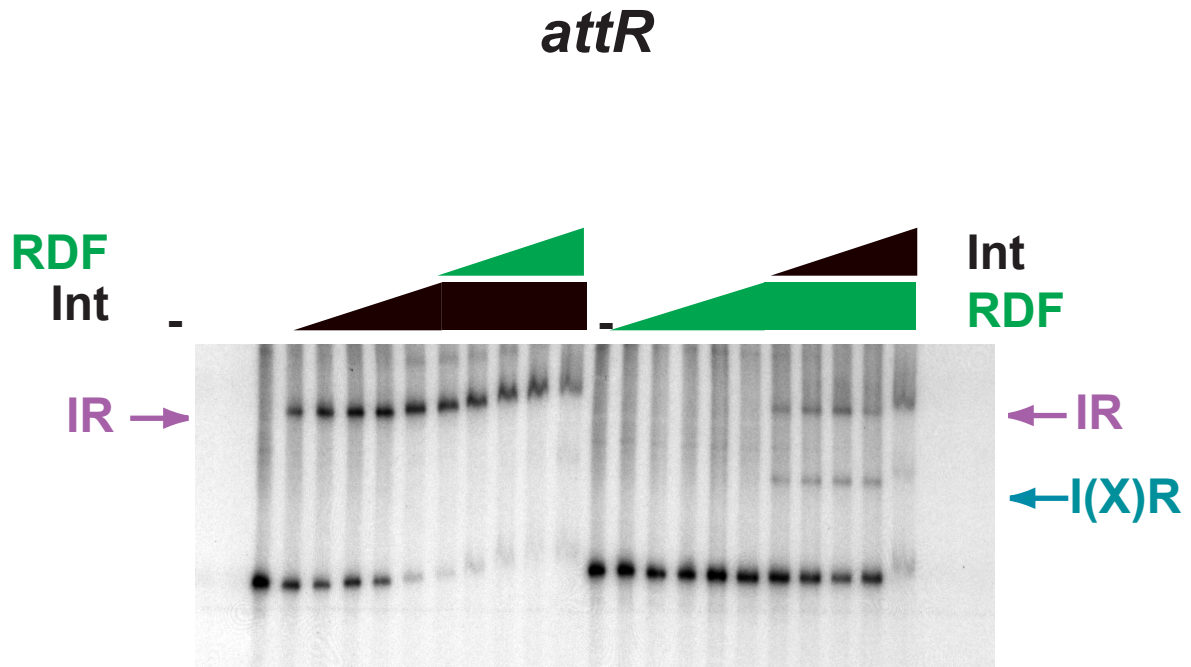


Figure 55. DNaseI footprinting of integrase and RDF at *attR*

A) DNaseI footprinting was done on *attR* with Integrase and/or RDF binding reactions. The first lane (-) shows the pattern of *attR* DNaseI digestion in the absence of protein. In the next five lanes, increasing amounts of Integrase were incubated with *attP*. Integrase was added at 13nM (1), 40nM (2), 130nM (3), 400nM (4), and 1.3 μ M (5). In these lanes, a region of 52 bases overlapping the core is protected from DNaseI digestion (indicated by the blue bracket). In the following four lanes, integrase is held at 1.3 μ M, and increasing amounts of RDF are present (8nM, 27nM, 80nM, 270nM, 800nM). Even at the highest amount of RDF (800nM), there is only a slight reduction in the pattern of protection by integrase. The next lane (-) has no protein added. The remaining four lanes have increasing amounts of RDF. There is no protection of *attR* by RDF seen at any amount of RDF added.

B) Sequence of the *attR* region where protection by integrase is seen. The blue bar shown above the sequence indicates the region of protection by integrase. The blue stars show where enhancements are seen. The central dinucleotide is shown as the boxed base pairs in the 12 bp core (uppercase sequence). The black bases indicate the sequence of the minimal 47bp substrate.

Figure 55A

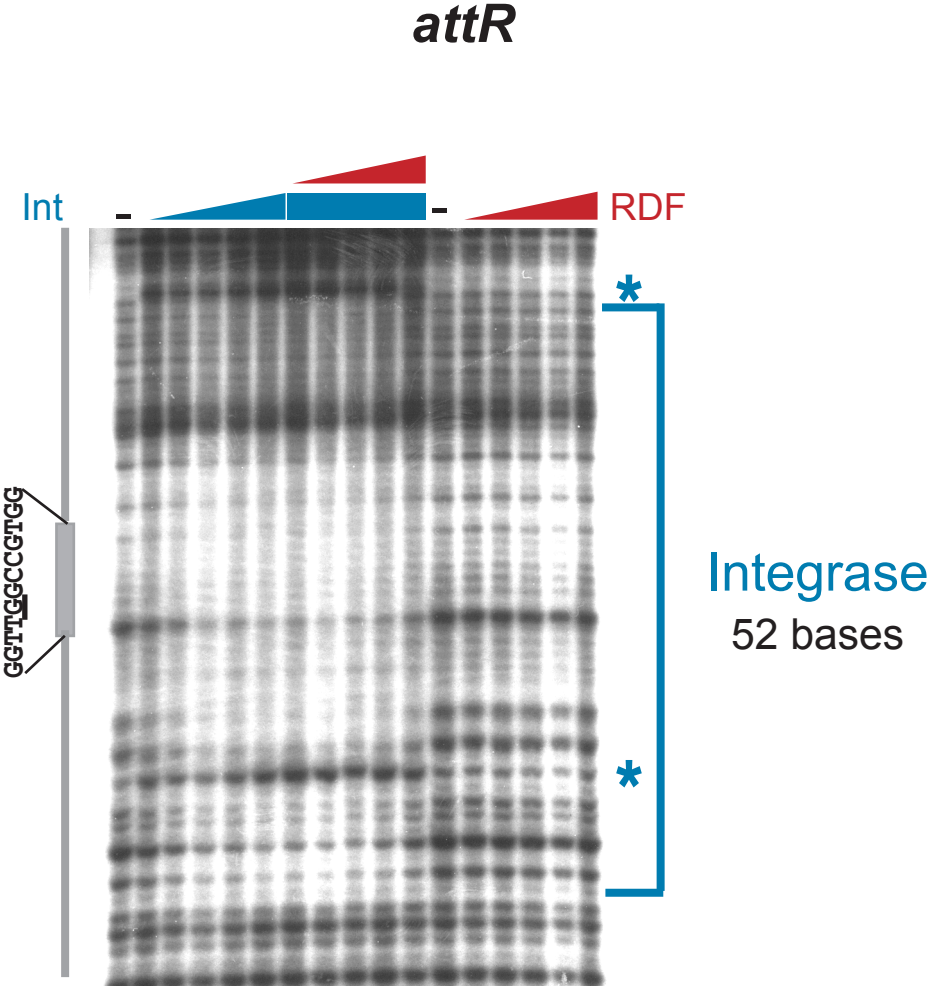


Figure 55B

attR

Integrase

aggtagacagcagcacgacaggtgtagtgtatctcacaggtccacGGTGGCCGTGGtcgaggtggggtggtggtagccatt
tccatctgtcgtcgtgctgtccacatcacatagagtgtccaggtgCCAACCGGCACCagtccaccccaccaccatcggtaa

When integrase is incubated with *attR* and subjected to limited DNaseI digestion, a region of protection spanning 52 bases is seen (Figure 55A). The footprint region extends over the core and includes two enhancements (Figure 55B). When RDF is incubated with *attR* and DNaseI footprinting reactions are done, there is no protection seen and the pattern of digestion with RDF is identical to what is seen with DNA alone. Addition of RDF to integrase and *attR* binding reactions does not change the pattern of digestion (Figure 55A). This result is not surprising since addition of RDF to integrase and *attR* does not shift the *attR*-integrase complex.

VI. H. The role of the RDF binding site in *attB* and *attL* complex formation

When integrase binds to *attB* two complexes are seen (IB1 and IB2). Incubation of *attB* with both integrase and RDF results in the formation of two additional complexes, one is seen at lower concentrations of RDF (IXB1), and the second is seen only at the highest concentration of RDF (IXB2) (Figure 48). In these binding experiments, the substrate is long enough to include the RDF binding site, thus, one or both of these complexes may be dependent on RDF binding directly to *attB*. To test the dependence of RDF binding to *attB* DNA in integrase-RDF-*attB* complex formation, two *attB* substrates were constructed with and without the RDF binding region. Complementary 80 base oligos were synthesized that included the sequences from the beginning of the RDF binding region across to the boundary of the minimal 40bp *attB*. Another set of 80 base oligos were synthesized that were identical except that the region of the RDF binding site was obliterated and random DNA sequence was put in its place (Figure 56A). These substrates were then labeled, annealed and used in DNA binding assays. Both substrates (*attB*+RDF_{bs} and *attB*-RDF_{bs}) support integrase binding as evidenced by a shift of the free DNA in the presence of integrase (Figure 56B). Much like what was seen with larger *attB* substrates,

Figure 56. Complex formation using substrate sites without the RDF binding site

A) Sequence of 80bp *attB* substrate sites with (*attB*+RDF_{bs}) and without (*attB*-RDF_{bs}) the RDF binding site. Identical bases between the two substrates are indicated by a vertical line in between the sequences. The region of *attB* where the RDF creates a footprint is shown as the boxed region. Horizontal arrows indicate the direct repeats that may be RDF recognition sequences. The central dinucleotide is shown in uppercase.

B) Integrase and RDF binding to *attB*+RDF_{bs} and *attB*-RDF_{bs}. DNA binding assays were done using an *attB* substrate site with or without the RDF binding site and integrase, RDF or both integrase and RDF. The reactions shown on the left utilized *attB* with the binding site, *attB*+RDF_{bs}, as indicated above the gel, while the reactions on the right used the *attB* without the site, *attB*-RDF_{bs}. The first lane in each set has DNA alone, and shows the migration of the free DNA. The next four lanes have increasing integrase as shown by the black triangle. The first of these has 13nM integrase, the second has 40nM integrase, the next lane has 130nM integrase, and the final lane has integrase at 400nM. In the following three lanes integrase is at 400nM and RDF is titrated as shown by the blue triangle. RDF is at 27nM, 80nM, and 270nM. The final three lanes of the set have RDF alone at the concentration mentioned above. The two complexes formed when integrase is added to *attB* are shown in black (IB1 and IB2). The complex seen upon RDF binding to *attB* is indicated by the blue arrow (XB). The complexes seen when both integrase and RDF are present are shown in green (IXB1 and IXB2).

C) Integrase and RDF binding to *attL*+RDF_{bs} and *attL*-RDF_{bs}. DNA binding assays were done using substrate *attL* sites with and without the RDF binding site and integrase, RDF or both integrase and RDF. The reactions on the left utilized the *attL* with the binding site, *attL*+RDF_{bs}, as indicated above the gel, while the reactions on the right used the *attL* without the site *attL*-

RDF_{bs}. The first lane in each set shows the migration of the free DNA. The next four lanes have increasing integrase as shown by the black triangle. The first of these has 13nM integrase, the second has 40nM integrase, the next lane has 130nM integrase, and the final lane has integrase at 400nM. In the following three lanes integrase is at 400nM and RDF is titrated as shown by the blue triangle. RDF is at 27nM, 80nM, and 270nM. The final three lanes of the set have RDF alone at the concentration mentioned above. The complex formed when integrase is added to *attL* is shown in black (IL). The complex seen upon RDF binding to *attL* is indicated by the blue arrow (XL). The complex seen when both integrase and RDF are present is shown in green (IXL).

Figure 56B

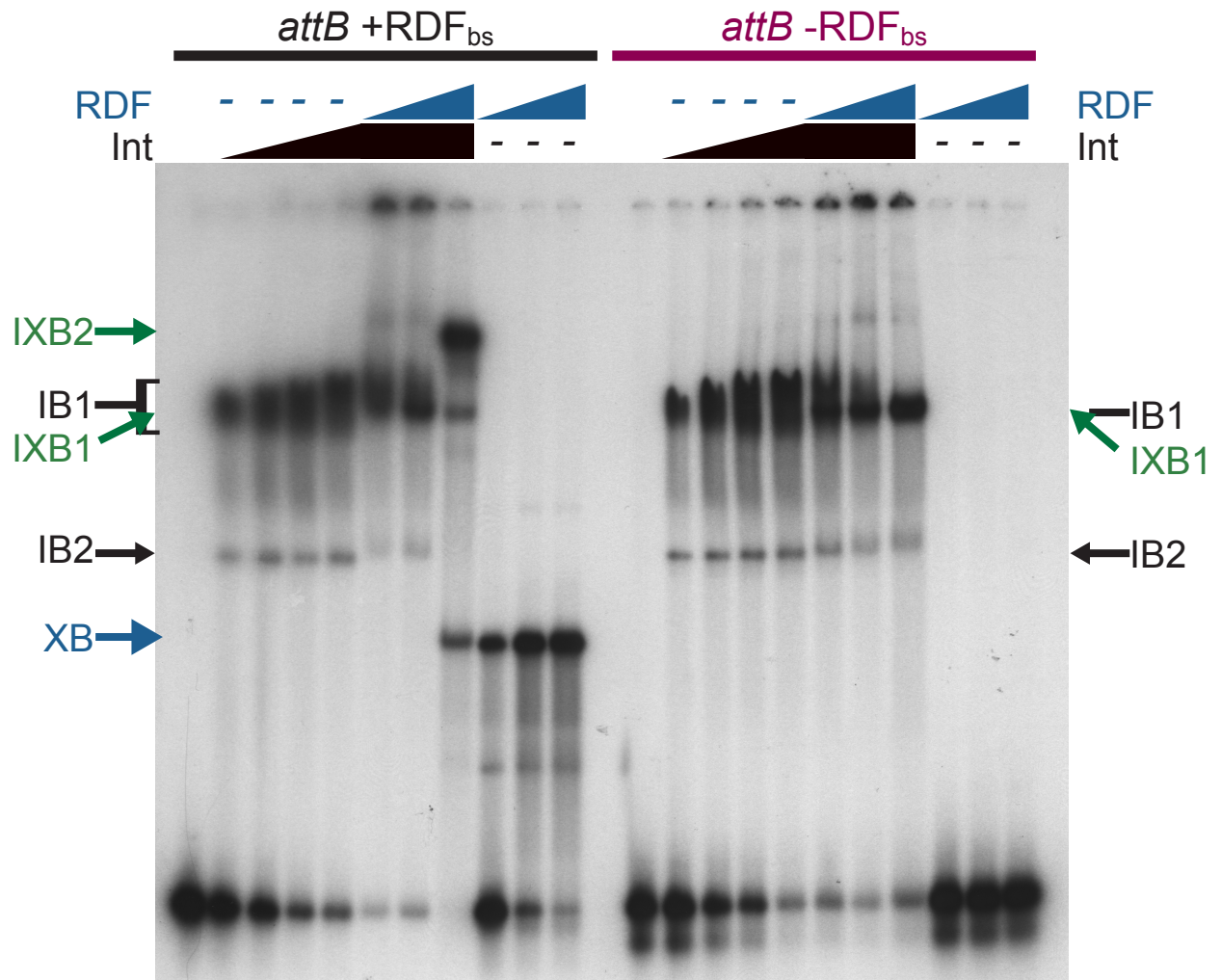
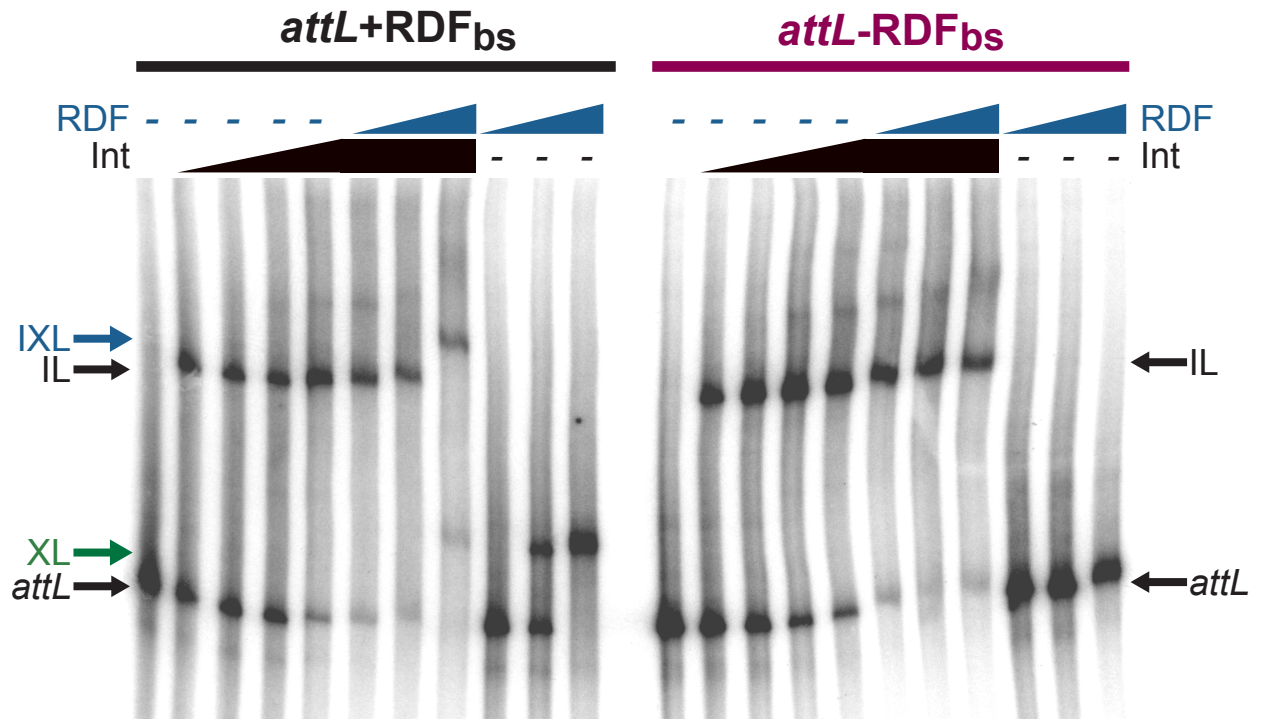


Figure 56C



two complexes are formed, IB1 and IB2, and the slower migrating complex (IB1) is the predominant one. Incubation of the wild type 80bp *attB* substrate (*attB*+RDF_{bs}) with RDF results in a shift of the free DNA. In contrast, RDF can not bind to the randomized sequence replacing the RDF binding site in the mutant *attB* (*attB*-RDF_{bs}), as evidenced by the absence of a shift the free DNA by RDF. When RDF is added to binding reactions with a constant amount of integrase and *attB*+RDF_{bs}, again two complexes are seen (Figure 56B). In this experiment however, the first complex seen at lower RDF concentrations (IXB1) appears to be at a migration similar to the slower migrating integrase-*attB* complex (IB1). Although this integrase-*attB* complex is expanded, when RDF is added a more compact band is produced. At higher RDF concentrations, a slower migrating complex is seen, which is likely to be equivalent to IXB2. In contrast, when RDF is added to integrase and *attB*-RDF_{bs}, the first integrase-RDF-*attB* complex (IXB1) is seen, but the slower migrating one (IXB2) is not seen. Therefore, IXB2 is RDF binding site dependent while IXB1 is not.

When integrase binds to *attL* a single complex (IL) is seen, while incubation of *attL* with both integrase and RDF results in a shift of this complex to a slower migrating species (IXL) (Figure 50). However, in these previous experiments the substrate is 341 bp, which includes the RDF binding site, thus this complex may be dependent on RDF binding directly to *attL*. To test the dependence of direct *attL* binding of RDF in the formation of the IXL complex, DNA band shift assays were done using two *attL* substrates with and without the RDF binding region. These substrates were obtained from *in vitro* integration assays using the 80bp *attB* substrates previously described. The product of an *in vitro* integration reaction using an *attP* plasmid with either *attB*+RDF_{bs} or *attB*-RDF_{bs} was gel purified and ligated to another piece of DNA, creating a plasmid with *attL* and *attR* in direct orientation. The *attL* obtained from the

reaction with *attB*+RDF_{bs} has a wild type sequence, while *attL* from the reaction with *attB*-RDF_{bs} has a mutant RDF binding site. Digestion of these plasmids with restriction enzyme FokI generated a 506bp fragment containing *attL*. These substrates were then labeled and used in DNA binding assays. Integrase binds to both *attL* substrates, shifting the free DNA into a single complex (IL) (Figure 56C). Incubation of the wild type *attL* substrate (*attL*+RDF_{bs}) with RDF also results in a shift of the free DNA. As shown in the *attB* experiment, the randomized sequence replacing the RDF binding site does not serve as a recognition sequence for RDF, as evidenced by the absence of a shift the free mutant *attL* (*attL*-RDF_{bs}) by RDF. When RDF is added to integrase at *attL*+RDF_{bs}, the IL complex is shifted to a higher migration (IXL). In contrast, when RDF is added to integrase and *attL*-RDF_{bs}, only the integrase-*attL* complex (IL) is seen, and the slower migrating IXL is not present (Figure 56C). Therefore, formation of the IXL complex is RDF binding site dependent.

VI. I. Role of the RDF binding region in integration and excision

Many of the *attB* and *attL* substrates used in the integrative and excisive *in vitro* recombination experiments have very little, if any, of the RDF binding site. The 40bp minimal *attB* substrate has just one base of the binding site, while the 50bp substrate that was most often used has just four additional bases. These *attB* substrates were used to generate the *attL* utilized in excision reactions, and thus, these excision substrates are equally devoid of the RDF binding site. Yet we have seen RDF dependent inhibition of integration and excision using these substrates lacking the RDF binding site. In order to carefully examine the role of the RDF binding site in integration inhibition, the *attB* substrates with and without the RDF binding site shown in figure 56A were used in integration assays and the ability of RDF to inhibit integration

at each of these sites was tested. The integration reaction used a plasmid-derived *attP* and integrase at 1.3 μ M. These two substrates, *attB*+RDF_{bs} and *attB*-RDF_{bs}, behaved identically in terms of the amount of product formed as a function of *attB* substrate concentration (Figure 57A). The ability of RDF to inhibit the integration reaction was tested as a function of concentration. Each of the integration reactions with and without the RDF binding site were inhibited to the same degree (Figure 57B). In the reactions with 0.0027 pmoles of RDF, some integration product is seen, however, at 0.008 pmoles of RDF reactions both with and without the binding site have been inhibited and thus have no detectable product. Thus, the presence of the RDF binding site in *attB* appears to have no effect on the ability of RDF to inhibit integration *in vitro*.

To test the dependence of direct *attL* binding of RDF in excisive recombination, the previously described intramolecular *attL-attR* plasmids were used in excision assays. These two *attL* substrates were obtained from *in vitro* integration reactions using the 80bp *attB* substrates with and without the RDF binding region. These two *attL-attR* intramolecular substrates were incubated separately with integrase at 0.8 μ M and increasing amounts of RDF. In both sets of reactions, excision was only observed at the highest tested concentration of RDF, 0.27 μ M (Figure 58A). Thus, the presence of the RDF binding site does not affect the concentration of RDF at which excision occurs. In another experiment, excision reactions were done using integrase at 0.8 μ M and RDF at 0.27 μ M, and substrate with or without the RDF binding site and samples were taken over time. These time courses look nearly identical in terms of the amount of product and substrate seen in each reaction. Thus, the presence or absence of the RDF binding site also does not affect the speed of the excision reaction.

Figure 57. Role of the RDF binding site in inhibition of integration

A) Integration reactions were done with supercoiled *attP* plasmid and an 80bp *attB* either with the RDF binding region (*attB*+RDF_{bs}- on the left) or without (*attB*-RDF_{bs}-on the right). These *attB* substrates were titrated into reactions with 0.040 pmoles *attP* plasmid and 4 pmoles integrase. These two *attB* substrates were added at ~ 0.038 pmoles, 0.11 pmoles (~3-fold excess), 0.38 pmoles (~10-fold excess), 1.1 pmoles (~30-fold excess), and 3.8 pmoles (~100-fold excess). The first lane (-) has no *attB* added. The two sites are recombined with similar efficiencies, as similar amounts of linear product are formed at the same amount of *attB* added. Lane M contains the DNA size markers with the sizes are in descending order; 10, 8, 6, 5, 4, 3, and 2 kb.

B) This panel shows a titration of RDF into integration reactions with 4 pmoles of integrase, *attP* plasmid and 1.1 pmoles (~30-fold excess) of *attB* with the RDF binding site (left) or without (right). The amounts of RDF are as follows: lane (-) no RDF, next lanes-0.0027 pmoles, 0.008 pmoles, 0.027 pmoles, 0.08 pmoles, and 0.27 pmoles. The amount of inhibition seen in reaction with *attB* without the RDF binding site to the product formation seen in reactions using the *attB* substrate with the binding site.

Figure 57 A&B

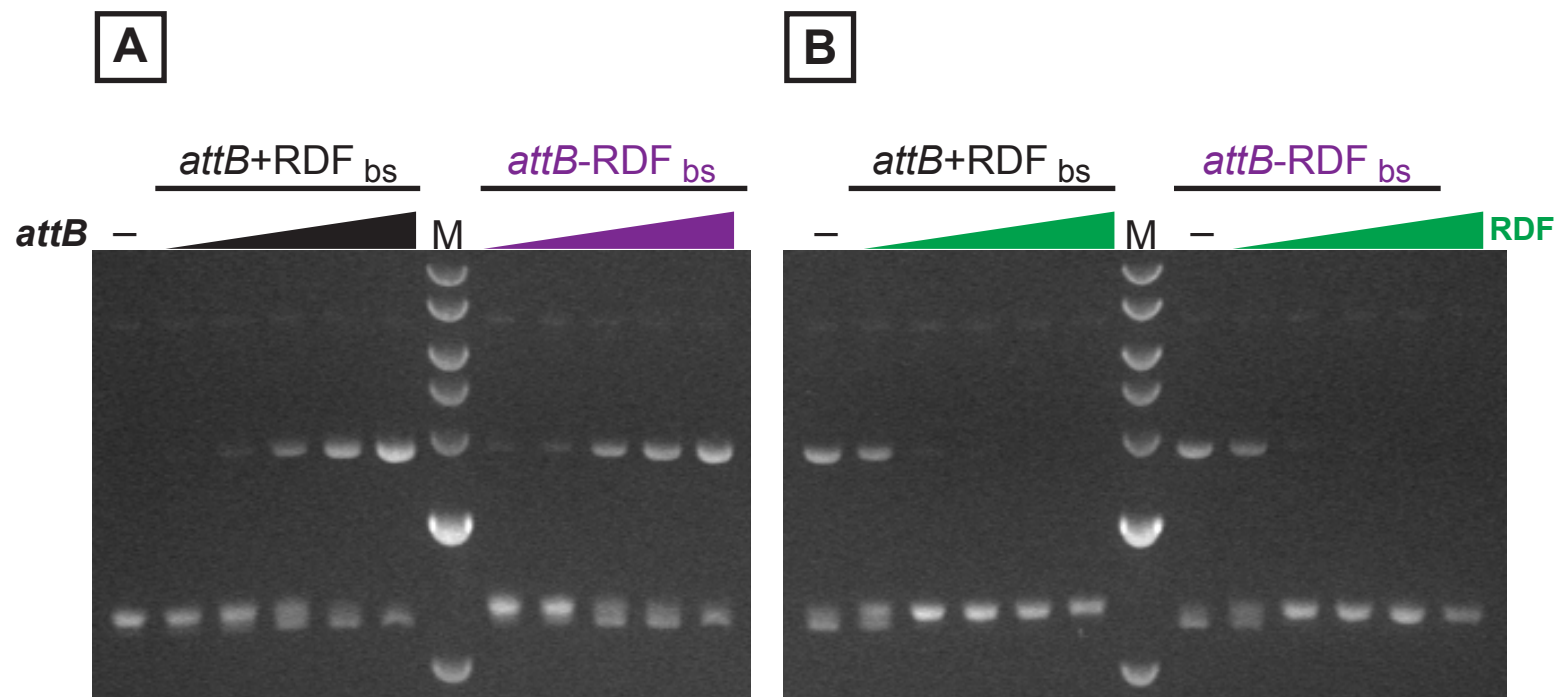


Figure 58. Excision with and without the RDF binding site

A) RDF was titrated into excision reactions with a single plasmid containing *attL* and *attR*, and 4 pmoles of integrase. The *attL-attR* plasmid used in the reactions on the left and the right are identical except for the RDF binding site. The plasmid used on the left has the wild type binding site and is thus designated +RDF bs, while the plasmid utilized on the right has a mutated sequence in place of the RDF binding site and is called –RDF bs. In each set, the lane marked (-) has no RDF, and the remaining lanes have RDF at 2.7nM, 8nM, 27nM, 80nM, and 270nM. Both sets of reactions were cleaved with *ClaI* following incubation. Digestion of the substrate with this enzyme yields a 4.0kb and a 740bp fragment, while product digestion produces 3.7kb and 1.0kb fragments. The markers are in lane M with the sizes given in kb at the right.

B) A time course of the excision reactions using substrate with and without the RDF binding site was also done. The substrates used are the same as in A, the plasmid used on the left has the wild type binding site (+RDFbs), while the plasmid on the right has a mutated RDF binding site (–RDFbs). In each set, the lane marked – has no RDF, and the remaining lanes have RDF at 270nM. The time at which each sample was taken is shown in hours above each lane. Both sets of reactions were then cleaved with *ClaI*. Digestion of the substrate with this enzyme yields a 4.0kb and a 740bp fragment, while product digestion produces 3.7kb and 1.0kb fragments. The markers are in lane M with the sizes given in kb at the right.

Figure 58A

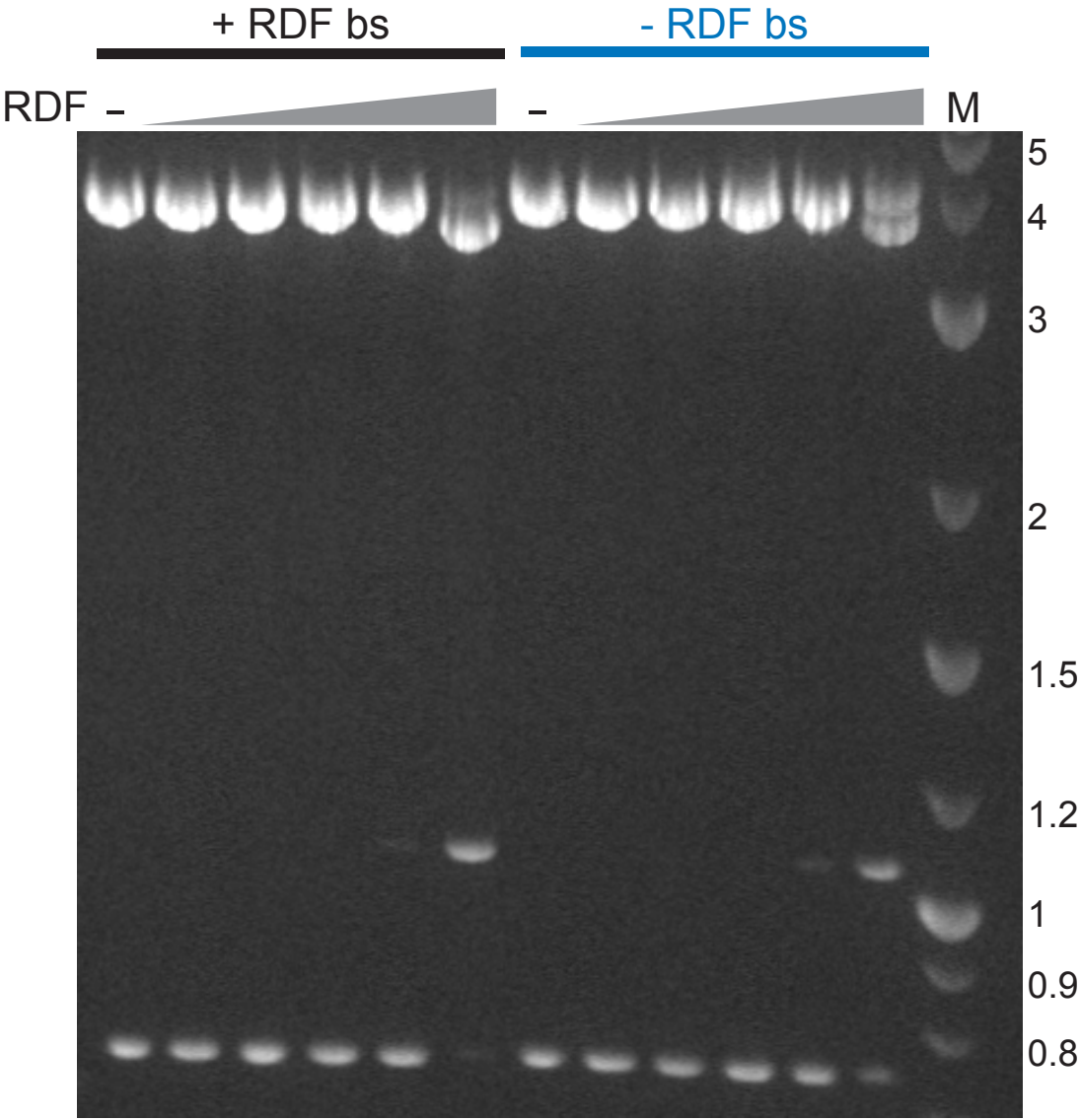
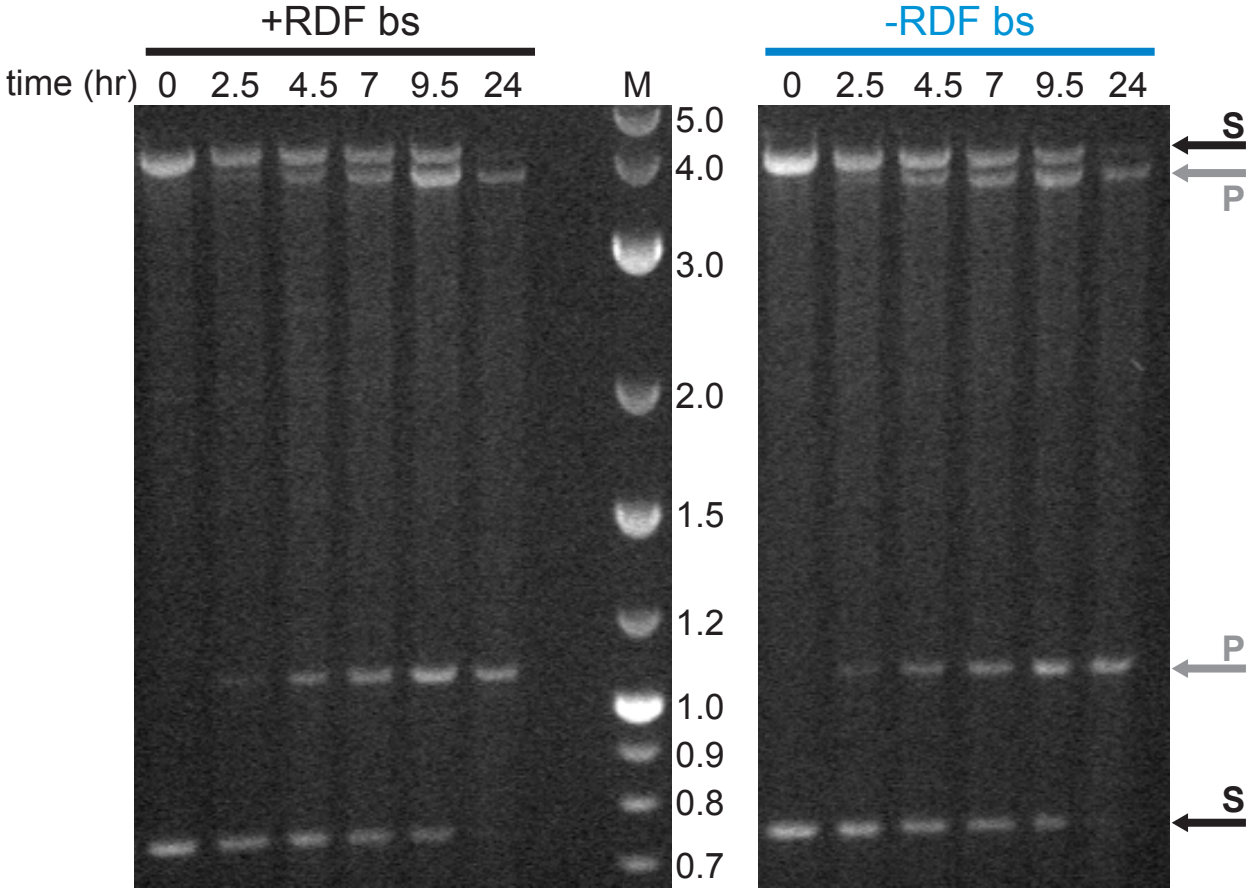


Figure 58B



VI. J. Integration complexes

In order for the integration reaction to occur, integrase and the substrate DNAs, *attP* and *attB*, must come together. A complex with *attP*, *attB* and all the integrase required for recombination is an intermediate in catalysis called a synaptic complex. To look for higher-order complexes that are intermediates in ϕ Rv1 integration, DNA binding assays were done with both *attP* and *attB*. When a cold 40bp *attB* is added to binding reactions containing radiolabeled *attP* and integrase at a temperature where recombination can occur (30°C), two additional bands are seen (Figure 59A). One of these (I) is present in small amounts and has a much slower migration than the other complexes and the second (II) migrates just below the lowest *attP*-integrase complex (IP2) (Figure 59A). When similar reactions with radiolabeled *attP*, integrase and *attB* are incubated for longer time periods at 30°C, more of these two additional bands are seen. A third band (III) is also seen, which migrates below the free 362bp *attP* DNA (Figure 59B). Because the reactions were incubated at a temperature at which recombination can occur, we expect one more of these bands to be product and the others to be recombination reaction intermediates. In support of this idea, when the reaction with hot *attP*, integrase and cold *attB* is incubated on ice, none of these additional bands are seen. All of these bands are also significantly reduced in the presence of RDF. When RDF is added to these reactions, a complex of slower migration is seen (IXP(B)). This band has a similar migration to an integrase-RDF-*attP* complex (IXP), but we do not know if *attB* is present in this complex.

The fastest migrating band seen near the bottom of the gel (III) is consistent with the size of free *attL* (215bp), one of the products of integration (Figure 59B). The other product of the recombination reaction, *attR*, would not be seen in these experiments because the *attP* DNA is singly labeled (at one end). The band that migrates just below the *attP*-integrase complex (II) is

Figure 59. Integration complexes

A) DNA band shift assays were done using ϕ Rv1 integrase, *attP* and *attB* to look for integration intermediates. A 32 P labeled *attP* fragment was incubated alone in lane one and with 3 pmoles of integrase, which is the amount used in integration assays, in lane 2. In the third and fourth lanes have *attP*, integrase and a cold 50bp *attB* was added at 9, and 90ng, respectively. The fifth through eighth lanes are identical to four except that 0.008 pmoles of RDF were added in lane 5, 0.027 pmoles in lane 6, 0.08 pmoles in lane 7, and 0.27 pmoles in lane 8. The complexes formed when integrase binds to *attP* (IP1, IP2) are indicated at the left by black arrows. Also on the left is an arrow indicating the position of a complex seen when integrase and RDF are present with *attP* and *attB*, IXP(B). On the left, the orange arrows show the position of two additional bands (I and II) seen when *attB*, *attP* and integrase are in these reactions.

B) This experiment is similar to A, except that samples were incubated for different times at 30°C or on ice (0°C) as indicated below the gel. The first lane has *attP* alone, and all other lanes (2-12) have both integrase at 3 pmoles and *attB* at 90ng. Lanes eight through 12 also have 2.7 pmoles of RDF. The complexes formed when integrase binds to *attP* are indicated by the black arrows at left. Shown on the right is an arrow indicating the position of a complex seen when integrase and RDF are present with labeled *attP* and cold *attB*, IXP(B). To the right, orange arrows indicate the three additional bands seen in reactions with *attB*, *attP* and integrase.

C) In this experiment the aberrant migration of the integrase-*attP* complex, IP2 is demonstrated. This gel shows three similar experiments, the first experiment shown on the left utilizes labeled *attP*, integrase and cold, wild type 50bp *attB*. The first lane shows the migration of free *attP*, and the second lane has 3 pmoles of integrase with *attP*. In the third and fourth lanes, 90ng of *attB* has been added to integrase and *attP*, and was incubated at 30°C (3) or on ice (4). Lane five

and six have 0.27 pmoles of RDF and were incubated at 30°C and 0°C, respectively. The set of reactions shown in the middle of the gel is similar to those shown on the right; with the difference being that in these middle reactions, a mutant *attB* was used that is defective in catalysis. The complexes formed when integrase binds to *attP* are indicated by the black arrows at the left. Also shown on the left is an arrow indicating the position of a complex present when integrase and RDF are present with labeled *attP* and cold *attB* [IXP(B)]. The orange arrows show the three additional bands seen when *attB* is added to *attP* and integrase reactions. The pink dots indicate the aberrant migration of IP2 in the presence of *attB*.

The final set of reactions shown on the right is similar to the other two; however, here a labeled *attB* and cold *attP* have been used. The complexes formed when integrase binds to *attB* are indicated at the right by blue arrows (IB1 and IB2). The complex formed when RDF binds to *attB* is also shown on the right in blue (XB). Shown on the right are arrows indicating the complexes seen when integrase, RDF, labeled *attP* and cold *attB* are present, IXP(P)1 and 2. Also, the orange arrows show the position of three additional bands seen when *attP* is added to *attB* and integrase reactions, (I_b, II_b, III_b).

D) This experiment shows the bands present when the proteins are degraded after incubation. The left and right halves of the gel are identical, except that after incubation the reactions on the right were treated with proteinase K. On both the left and right, lane one, like all other lanes has *attP*, however, in lane one only *attP* is present alone. Lane two, like all subsequent lanes has 3 pmoles of integrase. Lanes three and four and all subsequent lanes have 90ng of cold *attB*, and lane four has ethylene glycol at a final concentration of 40%. Lanes five and six have 0.27 pmoles of RDF added, but the two lanes differ in that ethylene glycol was added to 40% final

concentration in lane six. The following six lanes are the same as the first six, except that a mutant *attB* (AG) was used in the final set. Bands are labeled as previously described.

Figure 59A

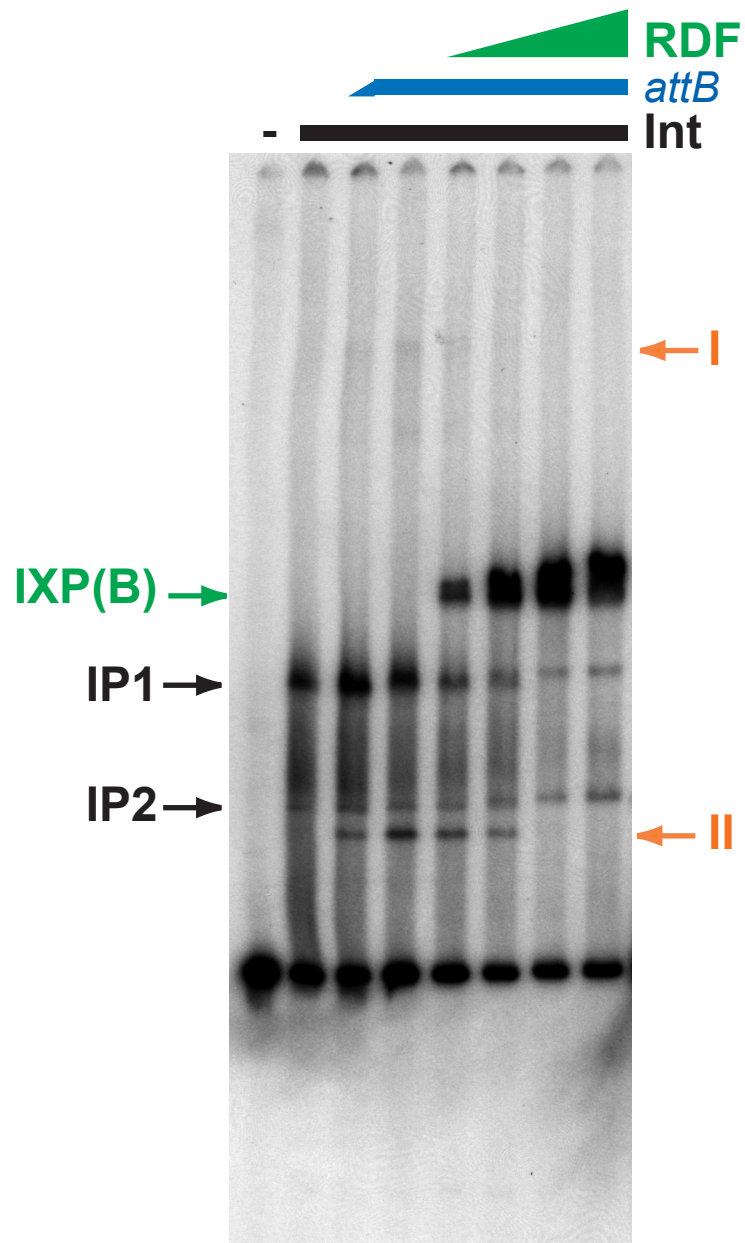


Figure 59B

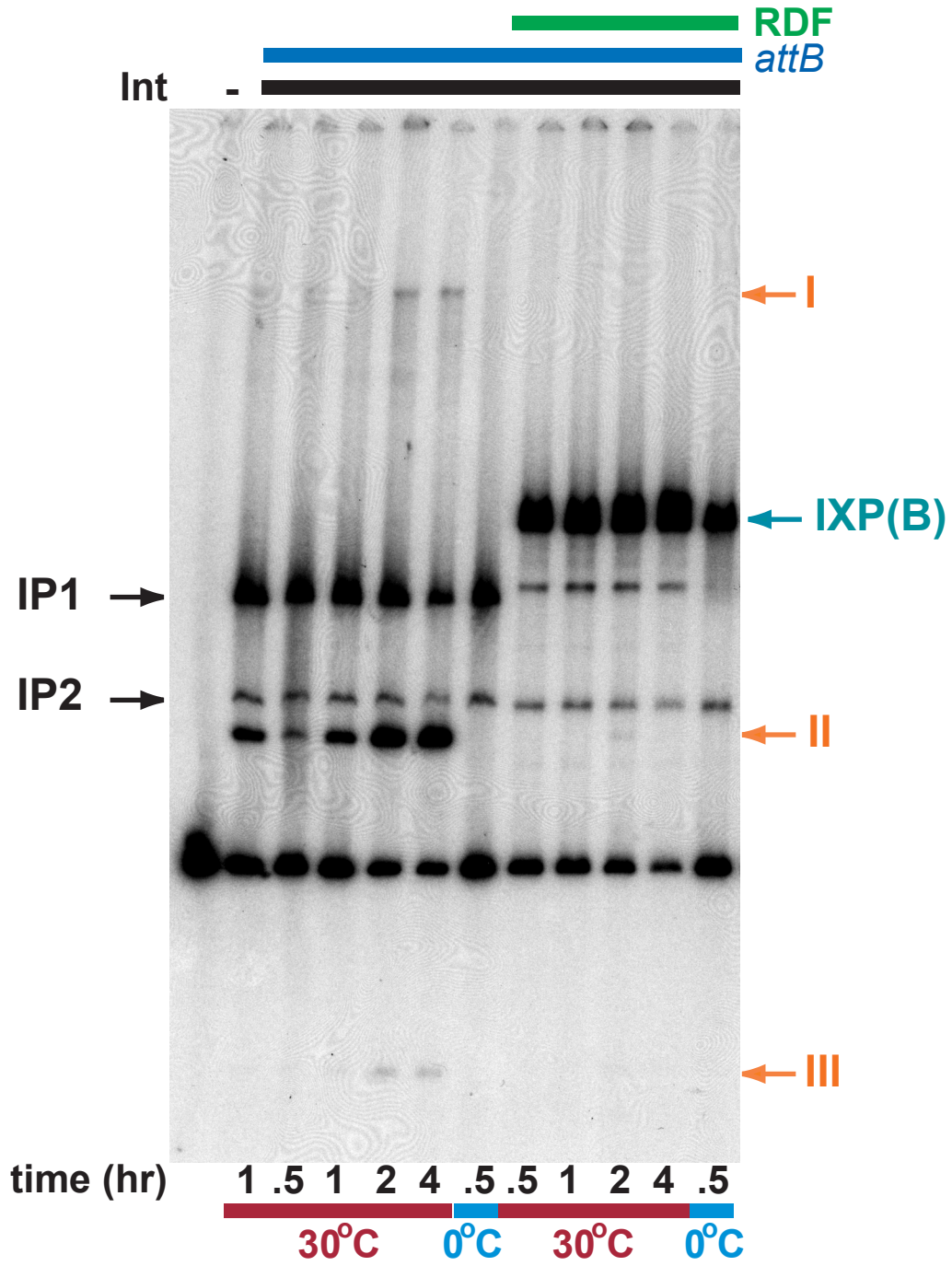


Figure 59C

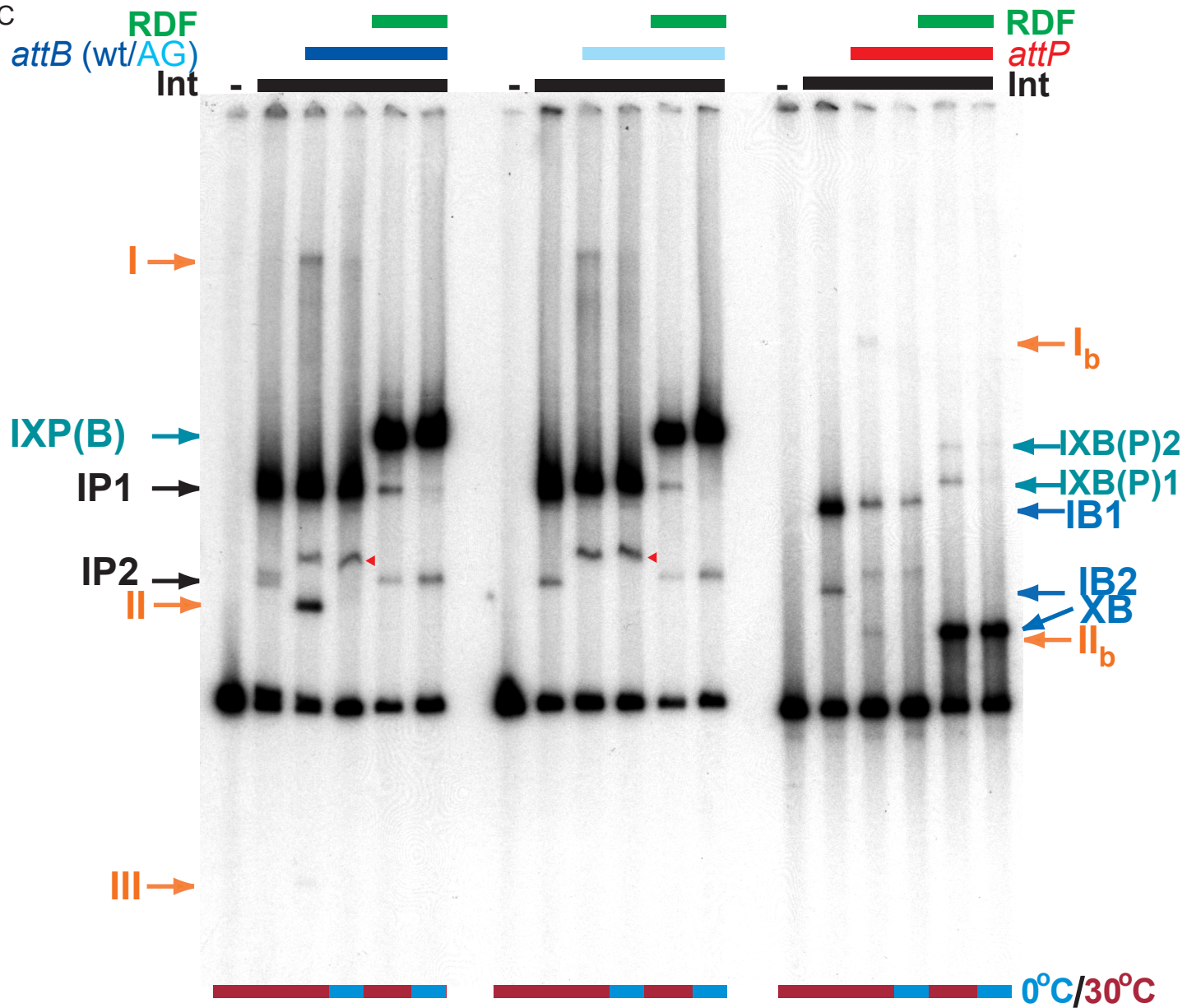
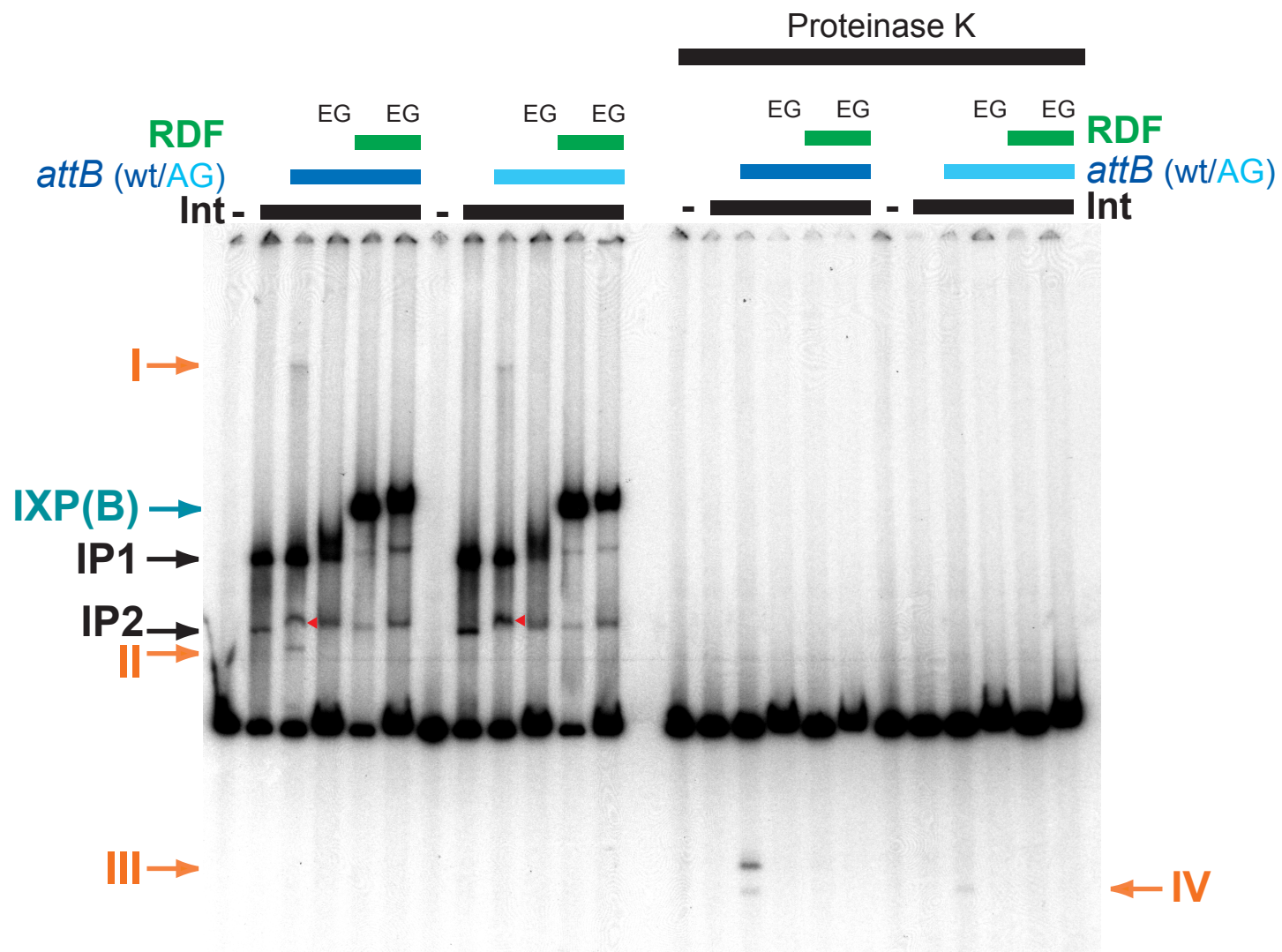


Figure 59D



consistent with *attL* product bound to integrase. Bound *attL* is also seen as a product in similar experiments with another serine integrase, Bxb1 (Ghosh *et al.*, 2003). When a mutant *attB* is used in the binding reactions, neither of these two bands (II and III) is seen (Figure 59C and D). This *attB* mutant has the fourth base of the common core changed from T to A, thus the putative central dinucleotide sequence is AG instead of TG. The AG mutant shows no *in vitro* recombination activity as *attB*, and thus it is not converted to product (4/12, Figure 31). The absence of bands II and III in reactions with this non-functional *attB* mutant is consistent with these bands being free and bound *attL* product, respectively. Unlike bands II and III, the slowest migrating band (I) is seen in reactions where either a wild type or a mutant *attB* (AG) is present along with *attP* and integrase. This band is likely to be a complex with *attP*, *attB*, and integrase. When a complementary experiment is done using radiolabeled *attB* and a cold linear *attP*, a similar complex (I_b) is seen (Figure 59D). There is one additional change that occurs when *attB* is added to binding reactions with *attP* and integrase. There is an apparent shift of the IP2 *attP*-integrase complex to a slower migrating species (red arrowheads in figure 59C and 59D). This shift occurs with both wild type and mutant *attB*, and at both 30°C and 0°C. Because this shift only occurs when *attP*, *attB* and integrase are present, presumably, this shifted complex contains *attB*, *attP* and all the integrase necessary for integration.

When reactions with *attP*, integrase and a wild type *attB* are treated with Proteinase K, all protein-DNA complexes disappear, and three bands are seen. One of these is the free *attB*. The larger and more predominant of the two other bands is probably free *attL* product (215bp). This band is absent from reactions using a recombination deficient *attB* mutant, while the smaller band is seen in reactions with either a wild type or mutant *attB*. It is likely that this smaller band

is a cleavage product that results from integrase cutting the *attP* substrate. This has been observed in reactions with other serine integrases (Ghosh *et al.*, 2003).

VI. K. Excision complexes

Much like in integration, in order for excision to occur, the substrate DNAs, *attL* and *attR*, and the integrase must synapse. The RDF must also act at some point before catalysis. It may do this either as a direct participant in a complex, or by altering the interactions between integrase and the substrate sites. Regardless of how RDF may function, all these participants must come together as an intermediate in catalysis. To detect intermediates in excisive recombination, band shift assays were done with a radiolabeled *attL* (341bp), integrase, RDF, and a linear cold *attR*. When a 47bp cold *attR* is added to binding reactions that contain radiolabeled *attL* and integrase, there is little change in the pattern of complexes seen, and no additional bands are visible. When RDF is added to reactions with *attR*, *attL* and integrase, four additional bands are seen (Figure 60). One of the bands seen at the highest concentrations of RDF (0.27 and 0.08 μ M) is the RDF *attL* complex (XL). Another band (III_{LR}) has a much faster migration than the other complexes. This band is this first band seen at lower RDF concentrations (0.027 μ M). The second (II_{LR}) migrates just below the free *attL*, and this band is only seen at 0.27 and 0.08 μ M RDF. The fastest migrating band (III_{LR}) is likely to be one of the products, *attB* (176bp). The other product, *attP*, is not seen in this experiment since the substrate *attL* is labeled only at one end (Figure 60). This *attB* product may be free, or bound to integrase. The band that migrates between III_{LR} and the free *attB*, II_{LR}, is RDF concentration dependent. Thus, II_{LR} is likely to be *attB* in complex with RDF. Whether integrase is present in this complex is not known. The third band seen when RDF is added has the slowest migration, (I_{LR})

(Figure 60). It is likely to be a synaptic complex with *attL*, *attR*, integrase, and perhaps RDF.

This complex is dependent on the presence of RDF. Thus, RDF is required for *attL* and *attR* to synapse.

VI. L. Discussion

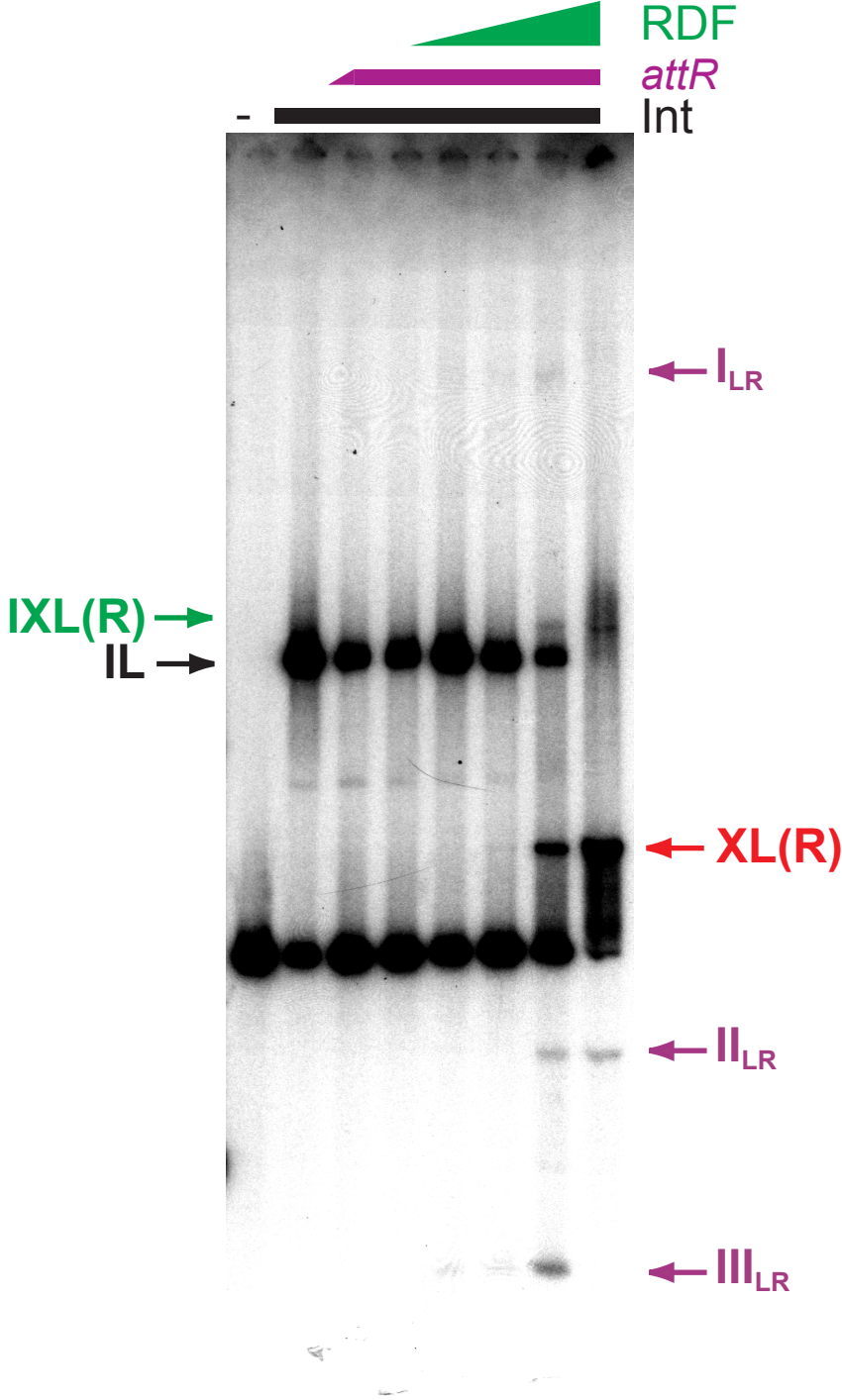
The investigation of the DNA binding behavior of integrase and RDF through band shift assay and solution DNaseI footprinting has revealed several things about the ϕ Rv1 recombination system. The ϕ Rv1 integrase binds to all four substrate sites, although the affinity for integrase at each site differs. Integrase has the highest affinity for *attL* and *attR* with a dissociation constant of approximately 40nM. Integrase binds to *attP* with a K_d of 130nM, and it has the lowest apparent affinity for *attB* ($K_d \sim 400$ nM). The ϕ Rv1 integrase, like other serine integrases, is able to bind to both *attP* and *attB*. This is in contrast to tyrosine systems like lambda. Lambda integrase does not form a stable complex with *attB* alone; instead integrase forms a complex with *attP* and host factors called an intasome, which then captures *attB* as naked DNA (Richet *et al.*, 1988).

The necessity for an excess of *attB* in the *in vitro* integration reaction may be the result of several factors, one of these may be the difference in the affinity of integration for the substrate sites. The apparent affinity of integrase for *attP* is at least 3-fold more than affinity of integrase for *attB*, thus an excess of *attB* may be required to overcome this difference. Also the affinity of integrase bound to *attB* for integrase bound to *attP* may also be low, thus having an effect on synapsis of *attB* and *attP*. Synapsis of these sites would be favored by an excess of *attB*. The deficiency in the reactions using a supercoiled *attB* with a linear *attP* is probably due to the difference in affinity for the sites. While in reactions with a supercoiled *attP*, the 2.6nM is

Figure 60. Excision complexes

A radiolabeled *attL* fragment was incubated with integrase, RDF and a small linear *attR*, to look for higher order protein-DNA complexes that are possible excisive recombination intermediates. The first lane (-) contains *attL* alone, and shows the migration of the free DNA. The second lane has 3 pmoles of integrase added and shows the position of the *attL*-integrase complex (IL-black arrow at left). All subsequent lanes have 3 pmoles of integrase as indicated by the black bar above the gel. The third lane has 9ng of a 47bp *attR*, the next lane and all subsequent lanes have 90ng of *attR* as shown by the purple bar. The fifth through eighth lanes have RDF added at; 0.008 pmoles (5), 0.027 pmoles (6), 0.08 pmoles (7), and 0.27 pmoles (8). Addition of RDF results in the appearance of four additional bands. The first of these is the RDF-*attL* complex (XL). There is also a slow migrating complex (I_{LR}) above all other bands, a fast migrating complex (III_{LR}) below all other bands, and a third (II_{LR}) that migrates just below free *attL* is seen at the highest amount of RDF (0.27 pmoles).

Figure 60



sufficient to bind enough integrase for recombination to occur, however, the concentration of *attB* in the supercoiled *attB* reactions (4nM) may be too low to support efficient recombination. When *attP* is present at equalmolar amounts or in slight excess, the reaction is most efficient, however, higher concentrations of *attP* compete away integrase from *attB*, and thus a reduction in recombination is seen.

The binding of integrase to *attP* results in a DNaseI protection footprint of 60 bases. When integrase binds to *attL* or *attR*, footprints are seen of 55 and 52 bases, respectively, and at *attB*, integrase binds to produce a 47 base footprint. These footprints are 5-7 bases longer than the size of the minimal substrates, however, the trend in the size requirement for the minimal substrates and footprints are the same, with *attB* being the shortest, and *attP* being the longest. In all these DNaseI footprinting experiments, the top strand was used, and the footprint extends across the common core including all the bases of the minimal substrate. When integrase binds to either *attB* or *attP*, two complexes are formed, although more of the slower moving complex is observed. When integrase binds to *attL* or *attR*, only one complex is seen. However, integrase binding in the presence of RDF, results in an additional complex that migrates in between the integrase alone complex and the free DNA. This is likely to be another form of integrase binding to the site.

As determined by DNaseI footprinting, RDF binds to a 34-36 base region that is to the left of the core in *attB* and *attL*. Upon RDF binding to *attB* or *attL*, just a single complex is formed, and RDF has the same apparent affinity for both sites, although we can only estimate that the dissociation constant is between 0.08 μ M and 0.27 μ M. At 0.08 μ M RDF, very little DNA is shifted while at 0.27 μ M nearly all the DNA is shifted. This is just a three-fold increase in concentration and may reflect cooperative binding of more than one molecule of RDF to the site.

Similarity in RDF binding to *attB* and *attL* is expected since RDF binds to an identical region in both sites. The region to which RDF binds has a pair of directly repeated sequences (5'-TCGTNGTGG-3') that are likely to serve as recognition sites. When the entire binding site is mutated, RDF can no longer bind directly to *attB* or *attL*.

Addition of RDF to binding reactions that contain integrase and *attL*, *attB*, or *attP*, results in a change in the pattern of complex formation. Addition of RDF to *attL* and integrase binding reactions shifts the integrase-*attL* complex (IL) to one with a slower migration (IXL). This shift is dependent upon the RDF binding site, and when the site is mutated, no shift is seen. When RDF is added to *attB* and integrase, both of the integrase-*attB* complexes (IB1 and IB2) are shifted to two slower migrating complexes (IXB1 and IXB2). The slowest migrating complex (IXB2) forms only at the highest concentration of RDF and is dependent upon the presence of the RDF binding site. The faster migrating complex (IXB1), which is seen at slightly lower RDF concentrations, is not RDF binding site dependent, and is seen in experiments using an *attB* with a mutant RDF binding site. This binding site independent complex, along with the complex seen when RDF is added to binding reactions with *attP* and integrase (IXP), suggest that RDF interacts directly with integrase.

When an *attB* substrate lacking the RDF binding site is used to examine inhibition of integration by RDF, a similar level of inhibition is seen in comparison to a reaction using an *attB* substrate with the RDF binding site. Also, when an excision reaction that uses a substrate without the RDF binding site is compared to a reaction with the RDF binding site, a similar level of excision is seen as a function of both RDF concentration and time. Therefore, the RDF binding site is dispensable for both its activities in inhibition of integration and excision. This

suggests that direct interactions between integrase and RDF are fundamental to directionality control by RDF.

Binding reactions done with the integration substrates *attP* and *attB*, show a slower migrating complex that could be a synaptic complex. This complex is only seen at 30°C, when recombination is actively occurring, and the complex can also be seen in reactions with a central dinucleotide mutant *attB*, although no recombinant product is detected. In the presence of RDF, this synaptic complex is not seen, which indicates that RDF acts at some step prior to synapsis. These studies also show that it is likely to be after substrate binding since RDF does not appear to disrupt binding of integrase to *attP* or *attB* as individual sites, although it does alter the complex pattern, shifting integrase-*attP* and integrase-*attB* to a slower migrating complex.

In similar binding reactions done with the excision substrates, *attL* and *attR*, no higher-order complex is observed in the presence of integrase alone. However, when RDF is added, a higher-order synaptic-like complex is observed as well as possible products. This result shows that in excision, RDF is also likely to be acting at step before synapsis.

The experiments presented here have uncovered some useful information about the ϕ Rv1 recombination system. Although at this point we do not know all the details of the mechanism, we can suggest a model for the recombination reactions that will provide a framework for future experimentation. The integration reaction begins with the binding of integrase to each of its substrates, *attB* and *attP*. It may either bind as a dimer to each, then these dimers could come together, or it may bind to one as a tetramer and then capture the other substrate. It is not yet known if ϕ Rv1 integrase binds to the sites as a dimer or a tetramer; however, we propose that it binds as a dimer. Separate dimers of integrase bind to *attP* and *attB*, and then specific

interactions between integrase bound to *attP* and integrase bound to *attB* bring the sites into synapsis (Figure 61).

The ϕ Rv1 Integrase bound to *attP* can only form a tetramer with integrase bound to *attB* because the interactions at each site are different. The conformation of integrase bound to *attP* is unlike the conformation of integrase bound to *attB* or any other site. Once the *attP* and *attB* sites synapse, the integrase becomes active and strand cleavage occurs (Figure 61). After 180° rotation of the DNA strands, they are then ligated in a recombinant configuration. The products of the integration reaction, *attL* and *attR*, often remain bound to integrase. However, the protein-protein interactions between these attachment site bound dimers are unstable thus, the sites dissociate.

When RDF is present, integrase bound at *attP* cannot synapse with integrase bound at *attB*. RDF interacts with both integrase bound to *attP* and integrase bound to *attB* as demonstrated by band shift assay. The interaction of RDF with integrase bound to these sites blocks synapsis, which thus prevents integration (Figure 61).

Integrase also binds as a dimer to *attL* and to *attR*. The interaction of integrase at these two sites is essentially a hybrid of the interactions at *attP* and *attB*, and thus the interactions at *attL* and *attR* are distinct from those at *attP* and *attB*. Excisive recombination, that is the recombination of *attL* and *attR*, can not occur in the absence of RDF. On its own, a dimer of integrase bound to *attL* cannot interact with a dimer of integrase bound to *attR*. However, in the presence of RDF and integrase, *attL* and *attR* can synapse (Figure 61). Transient interactions of RDF with integrase bound to *attL* and *attR* causes the complexes to adopt a conformation that allows them to interact and catalyze excisive recombination.

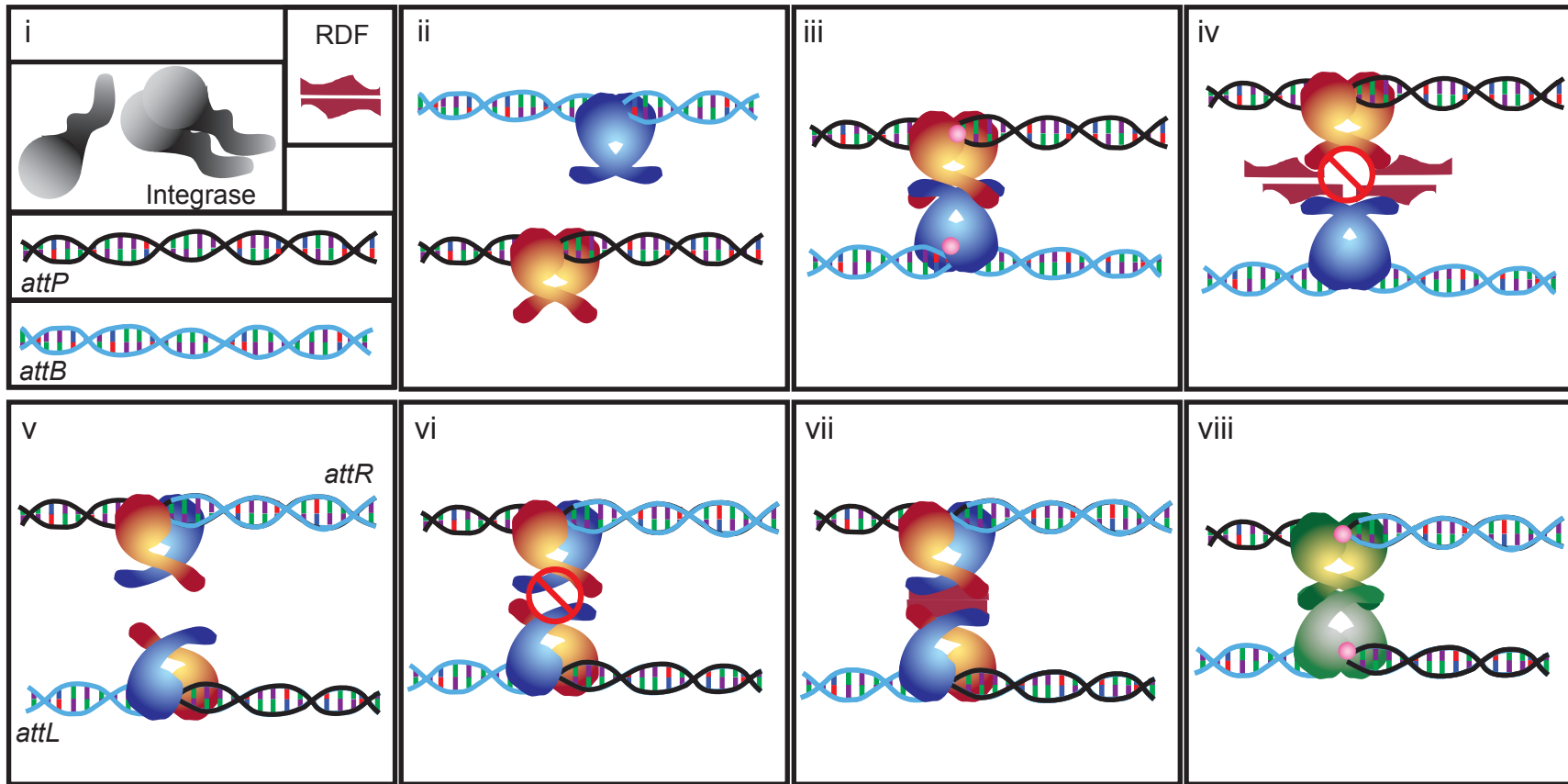
Figure 61. Models of integration and excision

Schematic representations of ϕ Rv1 recombination are shown here. Components of the reactions are shown in (i). Integrase is shown in gray although it is represented in slightly different shapes and colors when bound to *attP* (orange), *attB* (blue), *attL* and *attR* (orange and blue, or green when catalytically active). The recombination directionality factor (RDF) is in rust and *attP* DNA is black and *attB* DNA is blue.

Two protomers of integrase bind to *attB* and two protomers of integrase bind to *attP* (ii). Integrase binds differently to *attP* and *attB*, and a dimer of integrase bound to *attP* can synapse with a dimer of integrase bound to *attB* (iii). Synapsis causes each of the integrase monomers to become active (indicated by the pink dots) (iii), and integrative recombination takes place. In the presence of RDF, integrase bound to *attP* is blocked from interacting with integrase bound to *attB* (iv), thus blocking synapsis and integration.

Integrase also binds as a dimer to both *attL* and *attR* (v), although these dimers on their own cannot synapse (vi). When RDF is present, integrase bound to *attL* or *attR* changes conformation and the protein pairs bound to *attL* can now synapse with those bound to *attR* (vii). Excisive recombination can then occur (viii).

Figure 61



VII. CONCLUSION

VII.A. ϕ Rv1 integration-putting the pieces together

The *Mycobacterium tuberculosis* prophage-like element ϕ Rv1 encodes an active recombination system, with a serine integrase (*Rv1586c*) and a recombination directionality factor (RDF-*Rv1584c*). Although several serine integrases have been identified, only a few of them have been characterized in any considerable detail (Smith and Thorpe, 2002). The ϕ Rv1 integrase, like other serine phage integrases that have been studied, such as Bxb1, ϕ C31, R4, and TP901-1, utilize small substrates, and do not require substrate supercoiling, or an additional host factor (Breuner *et al.*, 2001; Kim *et al.*, 2003; Olivares *et al.*, 2001; Stoll *et al.*, 2002; Thorpe and Smith, 1998).

It is not known how the four different substrate sites (*attP*, *attB*, *attL*, and *attR*) are recognized and discriminated by integrase. The sequences that flank the crossover site in *attP* and *attB* are largely different (Smith and Thorpe, 2002). Presumably binding of integrase to *attB* is not equivalent to integrase binding to *attP*, and this difference in binding may be achieved either by different domains of integrase binding to each site or different contacts being made with each site and the same domain of integrase. The serine integrases have a large c-terminal region of

unknown function that could potentially contain two domains, one for *attB* recognition, and one for *attP*.

The details of synapsis are not known in any of these serine integrase systems, however, in the Bxb1 system, there is evidence that *attP* and *attB* substrates can synapse in either direct or indirect orientation (Ghosh *et al.*, 2003). Substrates that are synapsed in direct orientation will be cleaved and recombined to yield *attL* and *attR*, while substrates that synapse in the wrong orientation (indirect) will be cleaved but they will not be recombined. These misaligned sites cannot be recombined because the 2 base overlaps do not match. These cleavage products may be released or after a second 180° rotation, the DNA strands can be ligated to restore *attP* and *attB*. Evidence of this is seen in integration assays that utilize one substrate as a supercoiled plasmid, where misalignment of the substrates and subsequent re-ligation of *attP* and *attB* results in relaxation of the plasmid substrate (Ghosh *et al.*, 2003). Relaxation is seen as plasmid that runs slightly faster than the supercoiled substrate when electrophoresis of the reactions is done in the presence of ethidium bromide. These results show that the orientation of the substrates is given solely by the central dinucleotide (Ghosh *et al.*, 2003).

Some evidence of the role of the central dinucleotide in site orientation is seen in the ϕ Rv1 system, as relaxation of supercoiled substrates in recombination reactions (e.g. see figure 31, lanes [-] and [1-1]). Further evidence that in ϕ Rv1 the orientation of the sites is also determined by the central dinucleotide is seen when an *attB* substrate with a switched polarity central dinucleotide (CA instead of TG) is used in a recombination reaction. In reactions with this substrate, *attB* is recombined in the opposite orientation. If we refer to the left and right halves of *attP* and *attB* as P, P' and B, B', respectively, then the products of integration are *attL*

(B, P') and *attR* (B', P). The product of recombination between a wild type *attP* and *attB^{CA}* are (B', P') and (B, P).

In Bxb1, a higher-order nucleoprotein complex is seen in the presence of *attP*, *attB* and integrase (Ghosh *et al.*, 2003). This Bxb1 complex is not unlike that in the ϕ Rv1 system. Because the serine integrases require small substrates, and do not require substrate supercoiling, or an additional host factor, these substrate sites may synapse through protein-protein interactions between integrase bound (perhaps as a dimer) at each substrate site, in a manner similar to Cre recombinase. In the Cre recombinase system, no host factor is required, and the interactions between molecules of recombinase bound to the *LoxP* site are sufficient for synapsis (Mack *et al.*, 1992; Van Duynes, 2002). However, these serine integration systems differ from Cre because the integrase utilizes two different substrate sites. In addition, interactions of integrase with *attB*, *attP*, *attL* and *attR* are presumably not identical. If the interactions were identical, then any pair of sites could be recombined.

VII.B. Control of ϕ Rv1 recombination directionality

The protein encoded by *Rv1584c* controls the recombination directionality of ϕ Rv1. Expression of *Rv1584c* in mycobacteria cells that contain an integrated copy of the ϕ Rv1 *attP*-*integrase* plasmid results in excision of the plasmid. This recombination directionality factor (RDF) acts *in vitro* to inhibit integration and stimulate excision. The ϕ Rv1 RDF has DNA binding activity, and specifically binds to a region within *attB* and *attL*; however, this binding activity appears to be dispensable for the function of RDF. Both inhibition of integration and excision occur when this site is absent. These results indicate that interaction with integrase is likely to be central to the function of RDF. Band shift assays with either integration substrates or

excision substrates suggest that it probably acts at a step before synapsis. The proposed integrative synaptic complex does not form in the presence of RDF, and the excisive synaptic complex requires RDF.

Lactococcal bacteriophage TP901-1 is the only other serine integrase with a known RDF, and it has been demonstrated that an *attP-integrase* plasmid is excised in the presence of the ORF encoding this protein (Breuner *et al.*, 1999). As yet, no mechanistic data has been reported for this RDF, thus we do not know if it has binding activity or how it may stimulate excision. In the ϕ C31 system, where an RDF has not been identified, it has been shown that integrase alone cannot synapse any other pair of sites except *attB* and *attP*, which suggests that an RDF in this system may also act at or before synapsis (Thorpe *et al.*, 2000).

The ϕ Rv1 RDF shares some similarities and differences from the RDF proteins described in tyrosine recombination systems. RDFs that act in tyrosine systems bind to sites within *attP*, while the ϕ Rv1 RDF binds to *attB* (Connolly *et al.*, 2002; Esposito and Scocca, 1997; Lewis and Hatfull, 2003; Yagil *et al.*, 1989; Yin *et al.*, 1985; Yu and Haggard-Ljungquist, 1993). Much like these tyrosine systems the RDF in ϕ Rv1 acts to inhibit integration and stimulate excision. In lambda, it has been shown that the RDF (Xis) associates with the integrase, and when this interaction is interrupted by mutation, excision is also abolished (Swalla *et al.*, 2003; Warren *et al.*, 2003). Although we do not have direct evidence that the ϕ Rv1 RDF and integrase interact, the experimental evidence suggests that they do. Another study provided evidence that when the DNA binding activity of HK022 Xis (HK022 is closely related to lambda) is diminished either by mutation in the binding sites or in Xis itself, excision can still occur, although only at a fraction of wild type levels (Gottfried *et al.*, 2001; Gottfried *et al.*, 2003). The ϕ Rv1 also

functions in the absence of DNA binding. It is possible that in vivo this binding site plays a role; however, we have not explored this.

VII.C. Role of ϕ Rv1 and ϕ Rv2?

PhiRv1 is one of two prophage-like elements - the other being ϕ Rv2 - that are found in the widely used virulent laboratory strain of *M. tuberculosis*, H37Rv. Both ϕ Rv1 and ϕ Rv2 are absent from the vaccine bacillus *M. bovis* BCG. In fact, these two elements were originally identified as two regions of difference between *M. tuberculosis* H37Rv and *M. bovis* BCG, RD3 and RD11 (Brosch *et al.*, 2001; Mahairas *et al.*, 1996). The question arises as to why these elements are in *M. tuberculosis*. There are several possibilities; one is that these elements are decaying remnants of once active prophages. Another possibility is that they are active elements and may provide some function to the host, perhaps encoding virulence factors or imparting some other ability to the host. Yet another possibility is that ϕ Rv1 and ϕ Rv2 are parasitic elements located in *M. tuberculosis* simply as a means for propagation.

Although ϕ Rv1 and ϕ Rv2 are not obviously complete bacteriophages there are no apparent pseudogenes and they have several ORFs that are good matches to database entries (Hendrix *et al.*, 1999). As demonstrated by this study, the recombination system of ϕ Rv1 is functional, thus we do not favor the idea that the elements are just decaying remnants.

It is unlikely that both or either of these elements are absolutely required for virulence because these prophage-like elements are found with variability in virulent *M. bovis* strains and *M. tuberculosis* clinical isolates (Brosch *et al.*, 2001; Mahairas *et al.*, 1996). A recent study has shown that although most *M. tuberculosis* isolates have either ϕ Rv1 or ϕ Rv2 or both, some isolates (3 out of 46) lack both ϕ Rv1 and ϕ Rv2 (Brosch *et al.*, 2002). It is possible, however,

that there are compensating insertions in these strains, or that isolates that lack both ϕ Rv1 and ϕ Rv2 are phenotypically different than those that do have one or both of the prophage-like elements. However, H37Ra, an avirulent strain of *M. tuberculosis* has both ϕ Rv1 and ϕ Rv2. This strain has not been completely sequenced, and it is therefore possible that any virulence factors present in ϕ Rv1 or ϕ Rv2 could be mutated. Another study found that knocking the RD3 (which is ϕ Rv1) deletion back into *M. bovis* BCG, does not increase the virulence of BCG in a mouse model system (Pym *et al.*, 2002). If ϕ Rv1 does have virulence factors they are not sufficient to convert *M. bovis* BCG into a mouse pathogen.

Another possible role for ϕ Rv1 or ϕ Rv2 is gene transfer. The putative ϕ Rv1 capsid subunit encoded by *Rv1576c* is a close relative of the putative capsid subunit of mycobacteriophage Che9c (Pedulla *et al.*, 2003). This similarity lends support to the proposal that the ϕ Rv1 element in *M. tuberculosis* H37Rv is capable of making particles. These particles could package *M. tuberculosis* DNA in addition to or instead of ϕ Rv1 DNA, and thus carry out generalized transduction like the Genetic Transfer Elements described in *Rhodococcus*, *Methanococcus* and *Serpulina* (Bertani, 1999; Eiserling *et al.*, 1999; Humphrey *et al.*, 1997; Lang and Beatty, 2000).

Finally, it is possible that these elements provide no advantage to the host, and simply exist for their own benefit. With an understanding of ϕ Rv1 integration and excision it should be possible to excise this element from H37Rv and determine if it does in fact play a role in *M. tuberculosis* biology.

VII.D. Future considerations

VII.D.i. Dissecting $\phi Rv1$ recombination

One of our primary interests lies in understanding the recombination mechanism of $\phi Rv1$ recombination and serine recombinases in general. The proposed model needs to be tested and supported or rejected. The stoichiometry of integrase bound to each site is not known. We would like to know what the quaternary state of the integrase is both in solution and when bound to each of the substrate sites. The identity of each complex seen in band shifts needs to be verified, for example, the proposed synaptic complex in band shifts with both integration substrates, should contain both *attB* and *attP* and excision synaptic complex should contain *attL* and *attR*. We would also like to know if RDF is a component of the excision synaptic complex. In the band shift assays there are two additional RDF dependent complexes that are seen when RDF is added to integrase and *attB* or *attP*, and it is also not know if RDF is a physical component of the complexes.

We propose that the function of RDF is dependent on direct interactions with integrase. If RDF and integrase do interact directly, this interaction should be detected by genetic or biochemical methods. Furthermore, it should be possible to generate mutants of either RDF or integrase that do not interact with each other. A screen for mutants of RDF that cannot function in integration inhibition or excision may reveal an RDF that does not interact with integrase. It should also be possible to isolate mutants of integrase that are insensitive to integration inhibition by RDF or perhaps integrase mutants that recombine *attL* and *attR* in the absence of RDF.

VII.D.ii. Utility of ϕ Rv1 recombination

The serine integrases require no host factor and recombination has been demonstrated in heterologous systems such as other bacteria, eukaryotic cells and mammals (Kim *et al.*, 2003; Stoll *et al.*, 2002; Thomason *et al.*, 2001; Thorpe and Smith, 1998). Because the sequence requirements for the substrates are relatively short and no additional factors are required, work has been done to develop these systems for use in other organisms. These systems are currently being developed to generate transgenic mice, and engineer cell lines (Groth and Calos, 2004; Olivares *et al.*, 2002; Ortiz-Urda *et al.*, 2002; Ortiz-Urda *et al.*, 2003; Thyagarajan *et al.*, 2001). These phage integrase systems have an advantage over other recombinases because they utilize two different sequences (*attP* and *attB*) for recombination and the reactions are unidirectional (Thorpe *et al.*, 2000). Unlike what is seen in other recombinase systems, recombination cannot occur between the *attL* and *attR* hybrid junctions that flank the integrated cassette. It should be possible to develop the ϕ Rv1 system for the same applications. Because we have an understanding of directionality control in this system, an additional utility may be available to subsequently remove the integrated cassette.

The ϕ Rv1 integration-proficient *attP-integrase* vectors created as part of this work may also be useful vectors for the study of mycobacteria. We have shown that they transform *M. bovis* BCG and *M. smegmatis*. Although the efficiency of transformation in *M. smegmatis* is low relative to an extrachromosomal vector, these ϕ Rv1 plasmids are stably maintained in the absence of selection. The question arises as whether they could also be used in other mycobacteria. Blast searching of mycobacterial genome sequences showed that both *M. avium* subspecies paratuberculosis and *M. marinum* have regions of their genomes with identity to the *M. tuberculosis* 13E12 repeats (Genbank accession numbers NC_002944 and NC_004506,

respectively). Repeats are also found in *M. leprae* (Cole *et al.*, 2001). Although there are some differences within region where ϕ Rv1 integrates, the putative TG central dinucleotide is present, which suggests that ϕ Rv1 plasmids may transform these species (Figure 62). Integrating ϕ Rv1 vectors may be another tool to add to the mycobacterial (or perhaps even eukaryotic) tool box.

Although we have demonstrated here that the integrase and RDF are active, whether or not this element is mobile in *M. tuberculosis* or other members of the tubercule complex is unknown. If it is no longer actively of moving in and out of the *M. tuberculosis* genome, it certainly did at least once, perhaps in its former life as a bacteriophage.

Figure 62. REP13E12 sequences in other mycobacteria

The sequences of regions from other mycobacterial species with similarity to *M. bovis* BCG site #6 are shown. Blast searching revealed possible ϕ Rv1 *attB* sequences in *M. avium* subspecies paratuberculosis and *M. marinum*. These sites are aligned with the *M. bovis* BCG site which is shown in between the two sequences. Only one strand is shown, and the 12 base common core is shown in bold type, and the central dinucleotide is in red. Two repeats of the 13E12 family found in *M. leprae* are also shown.

Bibliography

- Abremski, K., and Gottesman, S. (1982) Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination. *J Biol Chem* **257**: 9658-9662.
- Abremski, K., Wierzbicki, A., Frommer, B., and Hoess, R.H. (1986) Bacteriophage P1 Cre-loxP site-specific recombination. Site-specific DNA topoisomerase activity of the Cre recombination protein. *J Biol Chem* **261**: 391-396.
- Abremski, K.E., and Hoess, R.H. (1992) Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng* **5**: 87-91.
- Alen, C., Sherratt, D.J., and Colloms, S.D. (1997) Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination. *Embo J* **16**: 5188-5197.
- Argos, P., Landy, A., Abremski, K., Egan, J.B., Haggard-Ljungquist, E., Hoess, R.H., Kahn, M.L., Kalionis, B., Narayana, S.V., Pierson, L.S.d., and et al. (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J* **5**: 433-440.
- Arnold, P.H., Blake, D.G., Grindley, N.D., Boocock, M.R., and Stark, W.M. (1999) Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *Embo J* **18**: 1407-1414.
- Azaro, M.A., A. Landy (2002) Lambda Integrase and the lambda Int family. In *Mobile DNA II*. Craig, N.L. (ed). Washington D.C.: ASM, pp. 118-148.
- Bannam, T., Crellin, P., and Rood, J. (1995) Molecular genetics of the chloramphenicol-resistance transposon Tn4451 from *Clostridium perfringens*: the TnpX site-specific recombinase excises a circular transposon molecule. *Molecular Microbiology* **16**: 535-551.
- Barsom, E.K., and Hatfull, G.F.-. (1996) Characterization of *Mycobacterium smegmatis* gene that confers resistance to phages L5 and D29 when overexpressed. *Mol Microbiol* **21**: 159-170.
- Benjamin, H.W., and Cozzarelli, N.R. (1990) Geometric arrangements of Tn3 resolvase sites. *J Biol Chem* **265**: 6441-6447.
- Bertani, G. (1999) Transduction-like gene transfer in the methanogen *Methanococcus voltae*. *J Bacteriol* **181**: 2992-3002.

- Better, M., Lu, C., Williams, R.C., and Echols, H. (1982) Site-specific DNA condensation and pairing mediated by the int protein of bacteriophage lambda. *Proc Natl Acad Sci U S A* **79**: 5837-5841.
- Bhadra, R.K., Roychoudhury, S., Banerjee, R.K., Kar, S., Majumdar, R., Sengupta, S., Chatterjee, S., Khetawat, G., and Das, J. (1995) Cholera toxin (CTX) genetic element in *Vibrio cholerae* O139. *Microbiology* **141** (Pt 8): 1977-1983.
- Bhargava, S., Tyagi, A.K., and Tyagi, J.S. (1990) tRNA genes in mycobacteria: organization and molecular cloning. *J Bacteriol* **172**: 2930-2934.
- Bibb, L.A., and Hatfull, G.F. (2002) Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1. *Mol Microbiol* **45**: 1515-1526.
- Bishop, A.J., and Schiestl, R.H. (2000) Homologous recombination as a mechanism for genome rearrangements: environmental and genetic effects. *Hum Mol Genet* **9**: 2427-2334.
- Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T., and Sherratt, D.J. (1993) Two related recombinases are required for site-specific recombination at dif and cer in *E. coli* K12. *Cell* **75**: 351-361.
- 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., Brondsted, L., and Hammer, K. (2001) Resolvase-like recombination performed by the TP901-1 integrase. *Microbiology* **147**: 2051-2063.
- Brosch, R., Pym, A.S., Gordon, S.V., and Cole, S.T. (2001) The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol* **9**: 452-458.
- Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., and Cole, S.T. (2002) A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* **99**: 3684-3689.
- Bruist, M.F., Glasgow, A.C., Johnson, R.C., and Simon, M.I. (1987) Fis binding to the recombinational enhancer of the Hin DNA inversion system. *Genes Dev* **1**: 762-772.
- Bull, T.J., McMinn, E.J., Sidi-Boumedine, K., Skull, A., Durkin, D., Neild, P., Rhodes, G., Pickup, R., and Hermon-Taylor, J. (2003) Detection and verification of *Mycobacterium avium* subsp. paratuberculosis in fresh ileocolonic mucosal biopsy specimens from individuals with and without Crohn's disease. *J Clin Microbiol* **41**: 2915-2923.
- Bushman, W., Yin, S., Thio, L.L., and Landy, A. (1984) Determinants of directionality in lambda site-specific recombination. *Cell* **39**: 699-706.

- Capiaux, H., Lesterlin, C., Perals, K., Louarn, J.M., and Cornet, F. (2002) A dual role for the FtsK protein in Escherichia coli chromosome segregation. *EMBO Rep* **3**: 532-536.
- Carrasco, C.D., Ramaswamy, K.S., Ramasubramanian, T.S., and Golden, J.W. (1994) Anabaena xisF gene encodes a developmentally regulated site-specific recombinase. *Genes Dev* **8**: 74-83.
- Chen, J.W., Lee, J., and Jayaram, M. (1992) DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell* **69**: 647-658.
- Cho, E.H., Gumport, R.I., and Gardner, J.F. (2002) Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. *J Bacteriol* **184**: 5200-5203.
- Christiansen, B., Johnsen, M.G., Stenby, E., Vogensen, F.K., and Hammer, K. (1994) Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. *J Bacteriol* **176**: 1069-1076.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Barrell, B.G., and al., e. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393**: 537-544.
- Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., and Barrell, B.G. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007-1011.
- Colloms, S.D., Sykora, P., Szatmari, G., and Sherratt, D.J. (1990) Recombination at ColE1 cer requires the Escherichia coli xerC gene product, a member of the lambda integrase family of site-specific recombinases. *J Bacteriol* **172**: 6973-6980.
- Connolly, K.M., Iwahara, M., and Clubb, R.T. (2002) Xis protein binding to the left arm stimulates excision of conjugative transposon Tn916. *J Bacteriol* **184**: 2088-2099.
- Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C., and Dye, C. (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* **163**: 1009-1021.

- Cowart, M., Benkovic, S.J., and Nash, H.A. (1991) Behavior of a cross-linked attachment site: testing the role of branch migration in site-specific recombination. *J Mol Biol* **220**: 621-629.
- Curcio, M.J., and Derbyshire, K.M. (2003) The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* **4**: 865-877.
- Dodd, I.B., Reed, M.R., and Egan, J.B. (1993) The Cro-like Apl repressor of coliphage 186 is required for prophage excision and binds near the phage attachment site. *Mol Microbiol* **10**: 1139-1150.
- 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.
- Eggleston, A.K., and West, S.C. (1996) Exchanging partners: recombination in *E. coli*. *Trends Genet* **12**: 20-26.
- Eiserling, F., Pushkin, A., Gingery, M., and Bertani, G. (1999) Bacteriophage-like particles associated with the gene transfer agent of methanococcus voltae PS. *J Gen Virol* **80**: 3305-3308.
- Eriksson, J.M., and Haggard-Ljungquist, E. (2000) The multifunctional bacteriophage P2 cox protein requires oligomerization for biological activity [In Process Citation]. *J Bacteriol* **182**: 6714-6723.
- Esposito, D., and Scocca, J.J. (1994) Identification of an HP1 phage protein required for site-specific excision. *Mol Microbiol* **13**: 685-695.
- Esposito, D., and Scocca, J.J. (1997) Purification and characterization of HP1 Cox and definition of its role in controlling the direction of site-specific recombination. *J Biol Chem* **272**: 8660-8670.
- Figuroa-Bossi, N., and Bossi, L. (1999) Inducible prophages contribute to Salmonella virulence in mice. *Mol Microbiol* **33**: 167-176.
- Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs Jr, W.R., Jr., Venter, J.C., and Fraser, C.M. (2002) Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. *J Bacteriol* **184**: 5479-5490.
- Frederickson, C.M., Short, S.M., and Suttle, C.A. (2003) The physical environment affects cyanophage communities in British Columbia inlets. *Microb Ecol* **46**: 348-357.
- Garnier, T., Eiglmeier, K., Camus, J.C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J.,

- Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V., and Hewinson, R.G. (2003) The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A* **100**: 7877-7882.
- Ghosh, P., Kim, A.I., and Hatfull, G.F. (2003) The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol Cell* **12**: 1101-1111.
- Gottfried, P., Kolot, M., and Yagil, E. (2001) The effect of mutations in the Xis-binding sites on site-specific recombination in coliphage HK022. *Mol Genet Genomics* **266**: 584-590.
- Gottfried, P., Silberstein, N., Yagil, E., and Kolot, M. (2003) Activity of coliphage HK022 excisionase (Xis) in the absence of DNA binding. *FEBS Lett* **545**: 133-138.
- Grainge, I., and Jayaram, M. (1999) The integrase family of recombinase: organization and function of the active site. *Mol Microbiol* **33**: 449-456.
- Grindley, N.D. (1993) Analysis of a nucleoprotein complex: the synaptosome of gamma delta resolvase. *Science* **262**: 738-740.
- Grindley, N.D.F. (2002) The Movement of Tn3-Like Elements: Transposition and Cointegrate Resolution. In *Mobile DNA II*. Craig, N.L., Craigie, R., Gellert, M., and A. M. Lambowitz (ed). Washington, D.C.: ASM Press.
- 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 U S A* **97**: 5995-6000.
- Groth, A.C., and Calos, M.P. (2004) Phage integrases: biology and applications. *J Mol Biol* **335**: 667-678.
- Hallet, B., and Sherratt, D.J. (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev* **21**: 157-178.
- 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: ASM 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 U S A* **96**: 2192-2197.
- Hoess, R.H., and Abremski, K. (1984) Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc Natl Acad Sci U S A* **81**: 1026-1029.
- 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.

- Hsu, P.L., and Landy, A. (1984) Resolution of synthetic att-site Holliday structures by the integrase protein of bacteriophage lambda. *Nature* **311**: 721-726.
- 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.
- Ip, S.C., Bregu, M., Barre, F.X., and Sherratt, D.J. (2003) Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination. *Embo J* **22**: 6399-6407.
- Jayaram, M. (1985) Two-micrometer circle site-specific recombination: the minimal substrate and the possible role of flanking sequences. *Proc Natl Acad Sci U S A* **82**: 5875-5879.
- Jayaram, M., Grainge, I., and Tribble, G. (2002) Site-Specific Recombination by the Flp Protein of *Saccharomyces cerevisiae*. In *Mobile DNA II*. al., N.L.C.e. (ed). Washington D.C.: ASM Press, pp. 192-218.
- Johnson, R.C., and Simon, M.I. (1985) Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. *Cell* **41**: 781-791.
- Johnson, R.C., Bruist, M.F., and Simon, M.I. (1986) Host protein requirements for in vitro site-specific DNA inversion. *Cell* **46**: 531-539.
- Johnson, R.C., Glasgow, A.C., and Simon, M.I. (1987) Spatial relationship of the Fis binding sites for Hin recombinational enhancer activity. *Nature* **329**: 462-465.
- Johnson, R.C., Ball, C.A., Pfeffer, D., and Simon, M.I. (1988) Isolation of the gene encoding the Hin recombinational enhancer binding protein. *Proc Natl Acad Sci U S A* **85**: 3484-3488.
- Jurka, J., Kohany, O., Pavlicek, A., Kapitonov, V.V., and Jurka, M.V. (2004) Duplication, coclustering, and selection of human Alu retrotransposons. *Proc Natl Acad Sci U S A* **101**: 1268-1272.
- Kazmierczak, R.A., Swalla, B.M., Burgin, A.B., Gumport, R.I., and Gardner, J.F. (2002) Regulation of site-specific recombination by the C-terminus of lambda integrase. *Nucleic Acids Res* **30**: 5193-5204.
- Kikuchi, Y., and Nash, H.A. (1979) Nicking-closing activity associated with bacteriophage lambda int gene product. *Proc Natl Acad Sci U S A* **76**: 3760-3764.
- Kim, A.I., Ghosh, P., Aaron, M.A., Bibb, L.A., Jain, S., and Hatfull, G.F. (2003) Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis* groEL1 gene. *Mol Microbiol* **50**: 463-473.
- Kitts, P.A., and Nash, H.A. (1988) Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *J Mol Biol* **204**: 95-107.

- Klippel, A., Cloppenborg, K., and Kahmann, R. (1988) Isolation and characterization of unusual *gin* mutants. *Embo J* **7**: 3983-3989.
- Kochi, A. (1991) The global tuberculosis situation and the new control strategy of the World Health Organization. *Tubercle* **72**: 1-6.
- Kunkel, B., Losick, R., and Stragier, P. (1990) The *Bacillus subtilis* gene for the development transcription factor sigma K is generated by excision of a dispensable DNA element containing a sporulation recombinase gene. *Genes Dev* **4**: 525-535.
- Lang, A.S., and Beatty, J.T. (2000) Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of *rhodobacter capsulatus* [In Process Citation]. *Proc Natl Acad Sci U S A* **97**: 859-864.
- Larsen, M. (2000) In *Molecular Genetics of the Mycobacteria*. Hatfull, G.F. and Jacobs Jr., W.R. (eds). Washington, DC: ASM Press. pp.313-320.
- Lazarevic, V., Dusterhoft, A., Soldo, B., Hilbert, H., Mael, 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 U S A* **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.
- Lewis, J.A., and Hatfull, G.F. (2003) Control of directionality in L5 integrase-mediated site-specific recombination. *J Mol Biol* **326**: 805-821.
- 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.
- Mack, A., Sauer, B., Abremski, K., and Hoess, R. (1992) Stoichiometry of the Cre recombinase bound to the *lox* recombining site. *Nucleic Acids Res* **20**: 4451-4455.
- 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**: 0000-0000.
- Mertens, G., Fuss, H., and Kahmann, R. (1986) Purification and properties of the DNA invertase gin encoded by bacteriophage Mu. *J Biol Chem* **261**: 15668-15672.
- Mizuuchi, K., Gellert, M., and Nash, H.A. (1978) Involvement of supertwisted DNA in integrative recombination of bacteriophage lambda. *J Mol Biol* **121**: 375-392.
- Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., Foeller, C., Hsu, P.L., Ross, W., and Landy, A. (1981) Structure and function of the phage lambda att site: size, int-binding sites, and location of the crossover point. *Cold Spring Harb Symp Quant Biol* **45**: 429-437.
- Mizuuchi, M., and Mizuuchi, K. (1980) Integrative recombination of bacteriophage lambda: extent of the DNA sequence involved in attachment site function. *Proc Natl Acad Sci U S A* **77**: 3220-3224.
- Mizuuchi, M., and Mizuuchi, K. (1985) The extent of DNA sequence required for a functional bacterial attachment site of phage lambda. *Nucleic Acids Res* **13**: 1193-1208.
- Moitoso de Vargas, L., Pargellis, C.A., Hasan, N.M., Bushman, E.W., and Landy, A. (1988) Autonomous DNA binding domains of lambda integrase recognize two different sequence families. *Cell* **54**: 923-929.
- Moitoso de Vargas, L., and Landy, A. (1991) A switch in the formation of alternative DNA loops modulates lambda site-specific recombination. *Proc Natl Acad Sci U S A* **88**: 588-592.
- Nash, H.A. (1975) Integrative recombination in bacteriophage lambda: analysis of recombinant DNA. *J Mol Biol* **91**: 501-514.
- Nash, H.A. (1981) Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. *Annu Rev Genet* **15**: 143-167.
- Nash, H.A., and Robertson, C.A. (1981) Purification and properties of the Escherichia coli protein factor required for lambda integrative recombination. *J Biol Chem* **256**: 9246-9253.
- Neely, M.N., and Friedman, D.I. (1998) Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. *Mol Microbiol* **28**: 1255-1267.

- Neilson, L., Blakely, G., and Sherratt, D.J. (1999) Site-specific recombination at dif by *Haemophilus influenzae* XerC. *Mol Microbiol* **31**: 915-926.
- Nunes-Duby, S.E., Matsumoto, L., and Landy, A. (1987) Site-specific recombination intermediates trapped with suicide substrates. *Cell* **50**: 779-788.
- Nunes-Duby, S.E., Kwon, H.J., Tirumalai, R.S., Ellenberger, T., and Landy, A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res* **26**: 391-406.
- 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.
- Olivares, E.C., Hollis, R.P., Chalberg, T.W., Meuse, L., Kay, M.A., and Calos, M.P. (2002) Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat Biotechnol* **20**: 1124-1128.
- Ortiz-Urda, S., Thyagarajan, B., Keene, D.R., Lin, Q., Fang, M., Calos, M.P., and Khavari, P.A. (2002) Stable nonviral genetic correction of inherited human skin disease. *Nat Med* **8**: 1166-1170.
- Ortiz-Urda, S., Thyagarajan, B., Keene, D.R., Lin, Q., Calos, M.P., and Khavari, P.A. (2003) PhiC31 integrase-mediated nonviral genetic correction of junctional epidermolysis bullosa. *Hum Gene Ther* **14**: 923-928.
- Pargellis, C.A., Nunes-Duby, S.E., de Vargas, L.M., and Landy, A. (1988) Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. *J Biol Chem* **263**: 7678-7685.
- Pedulla, M.L., Lee, M.H., Lever, D.C., and Hatfull, G.F. (1996) A novel host factor for integration of mycobacteriophage L5. *Proc Natl Acad Sci U S A* **93**: 15411-15416.
- Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs, W.R., Jr., Hendrix, R.W., and Hatfull, G.F. (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**: 171-182.
- Pena, C.E., Kahlenberg, J.M., and Hatfull, G.F. (2000) Assembly and activation of site-specific recombination complexes. *Proc Natl Acad Sci U S A* **97**: 7760-7765.
- Peña, C.E., Stoner, J.E., and Hatfull, G.F. (1996) Positions of strand exchange in mycobacteriophage L5 integration and characterization of the attB site. *J Bacteriol* **178**: 5533-5536.
- 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.A., Kahlenberg, J.M., and Hatfull, G.F. (1998) The role of supercoiling in mycobacteriophage L5 integrative recombination. *Nucleic Acids Res* **26**: 4012-4018.
- Popham, D.L., and Stragier, P. (1992) Binding of the Bacillus subtilis spoIVCA product to the recombination sites of the element interrupting the sigma K-encoding gene. *Proc Natl Acad Sci U S A* **89**: 5991-5995.
- Proteau, G., Sidenberg, D., and Sadowski, P. (1986) The minimal duplex DNA sequence required for site-specific recombination promoted by the FLP protein of yeast in vitro. *Nucleic Acids Res* **14**: 4787-4802.
- Pym, A.S., Brodin, P., Brosch, R., Huerre, M., and Cole, S.T. (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. *Mol Microbiol* **46**: 709-717.
- Ramaswamy, K.S., Carrasco, C.D., Fatma, T., and Golden, J.W. (1997) Cell-type specificity of the Anabaena fdxN-element rearrangement requires xisH and xisI. *Mol Microbiol* **23**: 1241-1249.
- Reed, R.R., and Grindley, N.D. (1981) Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. *Cell* **25**: 721-728.
- Rice, P.A., and Steitz, T.A. (1994) Model for a DNA-mediated synaptic complex suggested by crystal packing of gamma delta resolvase subunits. *Embo J* **13**: 1514-1524.
- Richet, E., Abcarian, P., and Nash, H.A. (1986) The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell* **46**: 1011-1021.
- Richet, E., Abcarian, P., and Nash, H.A. (1988) Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell* **52**: 9-17.
- Robertson, C.A., and Nash, H.A. (1988) Bending of the bacteriophage lambda attachment site by Escherichia coli integration host factor. *J Biol Chem* **263**: 3554-3557.
- Saha, S., Haggard-Ljungquist, E., and Nordstrom, K. (1987) The cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *Embo J* **6**: 3191-3199.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Smith, M.C., and Thorpe, H.M. (2002) Diversity in the serine recombinases. *Mol Microbiol* **44**: 299-307.
- Sonoda, E., Takata, M., Yamashita, Y.M., Morrison, C., and Takeda, S. (2001) Homologous DNA recombination in vertebrate cells. *Proc Natl Acad Sci U S A* **98**: 8388-8394.

- Stark, W.M., Sherratt, D.J., and Boocock, M.R. (1989) Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell* **58**: 779-790.
- Stark, W.M., Boocock, M.R., and Sherratt, D.J. (1992) Catalysis by site-specific recombinases. *Trends Genet* **8**: 432-439.
- Stark, W.M., and Boocock, M.R. (1995) Gatecrashers at the catalytic party. *Trends Genet* **11**: 121-123.
- Steiner, W.W., and Kuempel, P.L. (1998) Cell division is required for resolution of dimer chromosomes at the dif locus of Escherichia coli. *Mol Microbiol* **27**: 257-268.
- Sternberg, N., Hamilton, D., Austin, S., Yarmolinsky, M., and Hoess, R. (1981) Site-specific recombination and its role in the life cycle of bacteriophage P1. *Cold Spring Harb Symp Quant Biol* **45 Pt 1**: 297-309.
- Stirling, C.J., Szatmari, G., Stewart, G., Smith, M.C., and Sherratt, D.J. (1988) The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 cer locus. *Embo J* **7**: 4389-4395.
- Stirling, C.J., Colloms, S.D., Collins, J.F., Szatmari, G., and Sherratt, D.J. (1989) xerB, an Escherichia coli gene required for plasmid ColE1 site-specific recombination, is identical to pepA, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *Embo J* **8**: 1623-1627.
- Stoll, S.M., Ginsburg, D.S., and Calos, M.P. (2002) Phage TP901-1 site-specific integrase functions in human cells. *J Bacteriol* **184**: 3657-3663.
- Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., Bansal, G.P., Young, J.F., Lee, M.H., Hatfull, G.F., and et al., -. (1991) New use of BCG for recombinant vaccines. *Nature* **351**: 456-460.
- Summers, D.K., and Sherratt, D.J. (1984) Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**: 1097-1103.
- Summers, D.K., and Sherratt, D.J. (1988) Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the cer site. *Embo J* **7**: 851-858.
- Swalla, B.M., Cho, E.H., Gumport, R.I., and Gardner, J.F. (2003) The molecular basis of cooperative DNA binding between lambda integrase and excisionase. *Mol Microbiol* **50**: 89-99.
- Thomason, L.C., Calendar, R., and Ow, D.W. (2001) Gene insertion and replacement in Schizosaccharomyces pombe mediated by the Streptomyces bacteriophage phiC31 site-specific recombination system. *Mol Genet Genomics* **265**: 1031-1038.

- Thompson, J.F., de Vargas, L.M., Skinner, S.E., and Landy, A. (1987) Protein-protein interactions in a higher-order structure direct lambda site-specific recombination. *J Mol Biol* **195**: 481-493.
- 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 U S A* **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 [In Process Citation]. *Mol Microbiol* **38**: 232-241.
- Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., and Calos, M.P. (2001) Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol* **21**: 3926-3934.
- Tirumalai, R.S., Healey, E., and Landy, A. (1997) The catalytic domain of lambda site-specific recombinase. *Proc Natl Acad Sci U S A* **94**: 6104-6109.
- van de Putte, P., S. Cramer, and M. Giphart-Gissler (1980) Invertible DNA determines host specificity of bacteriophage Mu. *Nature* **286**: 218-222.
- Van Duyne, G.D. (2002) A Structural View of Tyrosine Recombinase Site-Specific Recombination. In *Mobile DNA II*. Craig, N.L., Craigie, R., Gellert, M., and A. M. Lambowitz (ed). Washington, D.C.: ASM Press, pp. 93-117.
- Walter, M.H., and Baker, D.D. (2003) Three Bacillus anthracis bacteriophages from topsoil. *Curr Microbiol* **47**: 55-58.
- Wang, H., Roberts, A., Lyras, D., Rood, J., Wilks, M., and Mullany, P. (2000) Characterization of the ends and target sites of the novel conjugative transposon Tn5397 from Clostridium difficile: excision and circularization is mediated by the large resolvase, TndX. *Journal of Bacteriology* **182**: 3775-3783.
- Warren, D., Sam, M.D., Manley, K., Sarkar, D., Lee, S.Y., Abbani, M., Wojciak, J.M., Clubb, R.T., and Landy, A. (2003) Identification of the lambda integrase surface that interacts with Xis reveals a residue that is also critical for Int dimer formation. *Proc Natl Acad Sci U S A* **100**: 8176-8181.
- Yagil, E., Dolev, S., Oberto, J., Kislev, N., Ramaiah, N., and Weisberg, R.A. (1989) Determinants of site-specific recombination in the lambdoid coliphage HK022. An evolutionary change in specificity. *J Mol Biol* **207**: 695-717.
- 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.

- Yin, S., Bushman, W., and Landy, A. (1985) Interaction of the lambda site-specific recombination protein Xis with attachment site DNA. *Proc Natl Acad Sci U S A* **82**: 1040-1044.
- Yu, A., and Haggard-Ljungquist, E.-. (1993) The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. *J Bacteriol* **175**: 7848-7855.
- Zeig, J., M. Silverman, M. Hilmen, and M. Simon (1977) Recombinational switch for gene expression. *Science* **196**: 170-172.