Exploring the Requirements of the Mycobacterium Phage Brujita Integrase

by

Bryce Leslie Lunt

B.S. Molecular Biology, Brigham Young University, 2011

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

THE DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Bryce Leslie Lunt

It was defended on

January 10, 2017

and approved by

Craig L. Peebles, Professor, Department of Biological Sciences

Saleem A. Khan, Professor, Department of Microbiology and Molecular Genetics

Anthony Schwacha, Associate Professor, Department of Biological Sciences

Jon P. Boyle, Associate Professor, Department of Biological Sciences

Dissertation Advisor: Graham F. Hatfull, Eberly Family Professor of Biotechnology, HHMI

Professor, Department of Biological Sciences

EXPLORING THE REQUIREMENTS OF

THE MYCOBACTERIUM PHAGE BRUIJTA INTEGRASE

Bryce Leslie Lunt, PhD

University of Pittsburgh, 2017

Temperate phages establish lysogeny and immunity to superinfection upon phage integration. Brujita Integrase (Int) is a model for a recently discovered subclass of tyrosine integrases, which lack traditional N-terminal arm-binding domains and contain additional C-terminal degradation tags. Typically, the arm-binding domain recognizes binding sites in the phage attachment DNA (*attP*) and serves to control the directionality in integration and excision. The absence of this domain raises the question of how Brujita Int binds attachment site substrates to control site selectivity and directionality. In this work, we describe the DNA and protein requirements for Brujita Int function, how it implements site selection, directionality control, and higher order complex formation. We show that Brujita Int is a simple recombinase, which resembles Cre, FLP, and XerC/D more closely than it does the tyrosine phage integrases. In contrast to Lambda Int, Brujita Int uses small DNA substrates, does not discriminate between *attP* and *attB* sites, and lacks directional control. It also shows unusual binding to its attachment sites. Binding to *attB* occurs at two half sites, B and B' in a cooperative manner, such that binding at B' is required for binding at B. Binding to *attP* is also unusual. The *attP* site is comprised of three half sites, P, P', and an additional site to the left of P, P1. However, binding is not detected at P, although this site is

required for recombination. Additionally, P1 is required for recombination, presumably to stabilize binding at P. Finally, we have solved the crystal structure of the Brujita Int and have found residues responsible for higher order complex formation. In total, our results indicate that Brujita Int may represent an evolutionary precursor to the canonical tyrosine integrases. Moreover, given its relative simplicity, Brujita Int presents a useful tool for the creation of synthetic genetic circuits as well as genome manipulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This work would not be possible without the support of many different people. First, I would like to thank my mentor, Dr. Graham Hatfull, for his guidance and direction, as well as for training me to become an independent scientist. I would like to thank all of the members of the Hatfull lab, especially those involved in the genetic switch paper, Gregory Broussard, Lauren Oldfield, Valerie Villanueva, and Emily Shine for their collaboration, encouragement, and advice.

Additionally, several people were very supportive and helpful in my development as a scientist. I thank Ching-Chung Ko for his intellectual support, including scientific discussion covering nearly every aspect of my project. Shweta Singh and Gregory Broussard were great mentors in the early stages of my work in helping gain a perspective and grasp of my project. Bob Duda was also an exceptional resource to me both for technical and intellectual support.

I especially thank the VanDemark Lab for their help with all of my protein work, both in technical support and training, specifically, Swarna Mohan, Adam Wier, Chris Amrich, Aubrey Lowen, Joel Rosenbaun, and Jenna Zalewski. I am also grateful to Dr. Andy VanDemark for his training in structural biology, and for his collaboration and solving of the Brujita Int crystal structure.

Finally, I would like to thank my parents, Jim and Irene Lunt for their encouragement as well as my wonderful wife, Anna, and my sweet boys, Hyrum and Alvin, for making long hard days more bearable.

Some of the work from this dissertation is from the Broussard *et al., Molecular Cell* 2013 paper. Most of the in vivo work is published in Lunt and Hatfull *J. Mol. Bio*. 2016, and the structural work and protein mutants will be published in Lunt *et al*., in process. This work was

supported by a National Science Foundation pre-doctoral fellowship 1247842 (for fellow 2013141947).

1.0 INTRODUCTION

Human tuberculosis (TB) is caused by *Mycobacterium tuberculosis*. This disease has posed a global health risk in the past and currently exhibits one of the largest single risks of the present and future (WHO, 2016). Nearly one-third of the world's population is infected with tuberculosis currently and 5% are estimated to have drug resistant infections (WHO, 2016). During 2015 alone, 6.1 million new cases were reported and an estimated 1.4 million deaths occurred as a result of tuberculosis infections, making tuberculosis the leading cause of death by an infectious agent (WHO, 2016). At heightened risk are those with HIV infections, which escalates the probability of rapid TB progression soon after infection (Corbett *et al.*, 2003). Given its prevalence, new tools are needed to understand, treat, and mitigate tuberculosis infections (Parisien *et al.*, 2008).

Bacteriophages are a powerful tool for studying mycobacteria. Mycobacteriophages, or phages that infect mycobacteria, are abundant and diverse, with nearly 1300 unique phages having been isolated and sequenced to date. These phages can be categorized by sequence similarity into at least 32 unique groups (http://www.phagesdb.org). Although most of these phages were isolated on *M. smegmatis*, a portion can also infect *M. tuberculosis* (Jacobs-Sera *et al.*, 2012). Interestingly, the majority of the annotated genes in these phages have no known function (Hatfull *et al.*, 2010, Casjens, 1992). Exploiting such novel genes may allow us to develop tools to manipulate, detect, and treat tuberculosis (Hatfull *et al.*, 2010, Pope *et al.*, 2015, Pope *et al.*, 2011).

A novel integration-dependent switch has been found in mycobacterium phage BPs, Charlie, and Brujita (Broussard *et al.*, 2013, Broussard & Hatfull, 2013). This system is much simpler than its counterparts in the *E. coli* phage Lambda and mycobacteriophage L5 systems, exhibiting characteristics of a streamlined and less regulated system and may represent an evolutionary precursor to the more complex genetic switches. Integration-dependent immunity consists of three components: an integrase (Int), which catalyzes site-specific recombination, repressor (Rep), a pro-lysogeny protein, and Cro, a pro-lysis protein. In particular, mycobacterium phage Brujita Int has been used as a model for studying these novel integration proteins. In Brujita, Int operates in novel and more simple ways when compared to traditional tyrosine integration systems.

This study aims to give a comprehensive picture of mycobacterium phage Brujita integration system by examining the requirements for integration, binding, and complex formation.

1.1 BACTERIOPHAGES

Bacteriophages represent the largest single group of biomass on the planet – with 10^{31} in existence – and are predicted to be the most abundant source of novel genetic information (Hendrix, 2003, Wommack & Colwell, 2000, Suttle, 2005). With 10^{23} infections occurring every second, total population turnover happens every few days (Suttle, 2007, Pedulla *et al.*, 2003). This large and rapidly evolving population of phages provides excellent models for studying evolution and for finding novel genetic systems. Indeed, much of the early work and techniques in molecular biology were completed and pioneered in model systems such as *E. coli* phage Lambda (Casjens & Hendrix, 2015, Abbani *et al.*, 2007, Aihara *et al.*, 2003, Biswas *et al.*, 2005, Herskowitz & Hagen,

1980, Hsu *et al.*, 1980, Hochschild *et al.*, 1986). Phage Lambda was the subject of the first direct sequencing of biological DNA, as well as the first sequenced dsDNA virus (Wu & Kaiser, 1968, Onaga, 2014, Sanger *et al.*, 1982).

Mycobacteriophages, have proven to be exceptional tools for studying the mycobacteria. Their study has provided understanding of host-phage interactions, of evolution of both mycobacteria and the phage, and of insights into new genetic systems. These phages are used for manipulating genomes (Marinelli *et al.*, 2008, van Kessel & Hatfull, 2007, Bardarov *et al.*, 1997, Donnelly-Wu *et al.*, 1993, Hatfull *et al.*, 2010, Huff *et al.*, 2010, Lee *et al.*, 1991, van Kessel & Hatfull, 2008a, van Kessel *et al.*, 2008, van Kessel & Hatfull, 2008b, Bardarov *et al.*, 2002), transposon and reporter gene delivery systems (Bardarov *et al.*, 2002, Bardarov *et al.*, 2003, Bardarov *et al.*, 1997), selectable markers (Donnelly-Wu *et al.*, 1993, Petrova *et al.*, 2015), integrating vectors (Huff *et al.*, 2010, Morris *et al.*, 2008, Lee *et al.*, 1991, Pham *et al.*, 2007), and vaccine development (Hatfull *et al.*, 1994). Additionally, mycobacterium phages have the potential to be effective tools for the treatment of disease, especially tuberculosis. Work in the Hatfull laboratory has shown that these phages have toxin-antitoxin systems, endolysins (Payne *et al.*, 2009, Payne & Hatfull, 2012), lysterases, and additional toxic proteins that may prove useful for treatment of this disease. With the enormous amount of new genetic information still untapped among the nearly 1200-sequenced mycobacterium phages (http://www.phagesdb.org) there is a great likelihood that additional novel genes and technologies remain to be discovered and developed (Pope *et al.*, 2015, Pope *et al.*, 2011).

The recent discovery of the integration-dependent bacteriophage immunity systems in mycobacteriophages represents a novel genetic switch that is much smaller and simpler than the canonical Lambda system and may constitute an evolutionary precursor to such systems

(Broussard & Hatfull, 2013).

1.2 THE GENETIC SWITCH

One major contribution gained from the study of phages is the discovery and characterization of the bistable genetic switches that determine a fate based on an external signal. The Lambda genetic switch controls the life cycle of the phage, determining whether productive infection proceeds through lysis, or whether the phage enters a dormant mode known as lysogeny (Ptashne, 2011). In Lambda, the switch is governed by two repressor proteins, which are negative regulators of each other (Ptashne *et al.*, 1982, Dubnau & Losick, 2006, Ptashne, 2004). The Lambda genetic switch is quite complicated involving a variety of factors, DNA sequences, and proteins.

When a temperate bacteriophage infects its host, the phage will proceed toward one of two fates, either productive infection through lysis, or establishment of a latent infection through lysogeny (Ptashne, 2011, Ptashne *et al.*, 1982, Ptashne, 2004, Oppenheim *et al.*, 2005). This decision between the two fates constitutes the genetic switch. If the switch is flipped toward lysis, the phage genome is replicated, viral proteins are made and progeny phage are packaged and released through lysis of the host cell. Alternatively, if the switch is flipped toward lysogeny, a repressor protein is expressed, which shuts down transcription of lytic genes. Additionally, a site-specific DNA recombinase, known as an integrase (Int) is expressed. This protein catalyzes the recombination reaction between attachment sites in the host chromosome and the phage genome, resulting in an integrated phage particle, known as the prophage. The resulting cell is now a lysogen, in which the phage DNA is replicated with the host genome and as the bacterial cell divides, the phage DNA is passed to daughter cells. The repressor protein allows the prophage to

block expression of lytic genes, as well as prevent superinfection by other members of the same immunity system, thus preserving the lysogenic state (Little, 2010, Ptashne, 2011, Ptashne, 2004, Oppenheim *et al.*, 2005). The lysogen is induced when a signal, such as DNA damage or cell stress, stimulates production of integrase and accessory proteins to exit from lysogeny. These accessory proteins – such as Xis – are needed to promote the excisive recombination reaction and remove the phage DNA from the host genome. Following excision, the phage enters lytic growth (Astromoff & Ptashne, 1995, Ptashne, 2011, Ptashne *et al.*, 1982, Bloch *et al.*, 2013).

1.2.1 Bacteriophage Lambda

The temperate bacteriophage Lambda has been extensively studied and its genetic switch is perhaps the best understood. During Lambda infection, the phage adsorbs to the exterior of the host cell and injects its genome, which then circularizes. RNA synthesis is initiated at two promoters, P_R and P_L, which transcribe two genes, *cro* and *N*, before transcriptional termination at terminators *tL1* and *tR1* by a protein called Rho (Casjens & Hendrix, 2015, Roberts, 1969). During lysis, protein N can prevent termination by binding to *tR1* and allow transcription through early lytic genes by desensitizing RNA polymerase to Rho-dependent and Rho-independent terminators (Echols, 1971, Franklin, 1974, Friedman & Gottesman, 1983, Friedman & Court, 1995, Reed *et al.*, 1997). One of the products made after N-mediated anti-termination is protein Q, which is also an anti-terminator, allowing expression of the late genes (Roberts *et al.*, 1998). Q binds to the DNA at the transcription complex and migrates with this complex as it transcribes the late genes, which leads to lysis of the host cell (Deighan & Hochschild, 2007, Shankar *et al.*, 2007, Marr *et al.*, 2001).

The Lambda lysogeny switch consists of two major players, the pro-lysis protein Cro, and repressor (CI), which promotes lysogeny (Hochschild *et al.*, 1986). *Cro* and *cI* are separated by 3 operator DNA sequences, OR_1, OR_2, and OR_3, with each protein having different affinities for each site (Cro prefers OR_3, CI prefers OR_1 and OR_2) (Campbell, 1994). The repressor binds weakly – but cooperatively – in dimers to OR_1 and OR_2, blocking the ability of RNA polymerase to bind and transcribe *cro*. At the same time it recruits polymerase to transcribe its own gene (Ptashne, 2011). At high concentrations of CI, binding will occur at OR_3, which loops the DNA using protein-protein interactions between the other occupied OR sites, preventing RNA polymerase from transcribing repressor (Lewis *et al.*, 2011, Cui *et al.*, 2013a, Cui *et al.*, 2013b, Priest *et al.*, 2014). Thus, Lambda repressor is able to both positively and negatively regulate its own expression in response to protein levels (Campbell, 1994, Reichardt & Kaiser, 1971). In a similar manner, Cro binds to OR_3 and blocks expression of CI. At high concentrations, Cro also binds to OR_2 and OR_1, preventing its own transcription (Campbell, 1994).

Among the genes transcribed with the help of anti-terminator N, are *cII* and *cIII,* which encode two accessory factors that play important roles in the switch. CII is a transcription factor that is required for initial expression of CI at the promoter P_{RE}, yet CII is rapidly degraded in *E. coli*; Because of this, CIII is required to prevent degradation and stabilize CII (Ptashne, 2004, Campbell, 1994, Halder *et al.*, 2007, Kobiler *et al.*, 2002, Shotland *et al.*, 2000). Though CI and Cro are the central players, CII levels determine whether the phage will enter lytic growth or lysogeny (Kobiler *et al.*, 2002). When host protease levels are low, CII will accumulate, lead to expression of *cI*, and establishment of lysogeny. Following UV treatment or when proteases are abundant, CII will be rapidly degraded, *cI* will not be expressed, and lytic growth will follow (Lederberg & Lederberg, 1953, Court *et al.*, 2007).

Although there are occasional differences, the Lambda genetic switch is quite typical of many temperate phage switches (Campbell, 1994, Knight *et al.*, 1989). In such lambdoid genetic switches, production of integrase occurs after the lysogeny decision has been made and CI has repressed lytic gene expression. In this case, integration contributes to stability of the lysogen, but has no role in the decision to enter lysogeny.

1.2.2 Integration-Dependent Phage Immunity

In contrast to the lambdoid phages, we have recently discovered a much simpler genetic switch in the mycobacteriophages, where integrase is the key player in this decision. Integration-dependent phage immunity is a novel type of genetic switch that consists of only three genes: *int* (Int), *rep* (Rep), and a predicted *cro* (Figures 1-2) (Broussard *et al.*, 2013). There are no analogs for the Lambda *cII, cIII,* or *xis* in this system (Broussard & Hatfull, 2013, Broussard *et al.*, 2013).

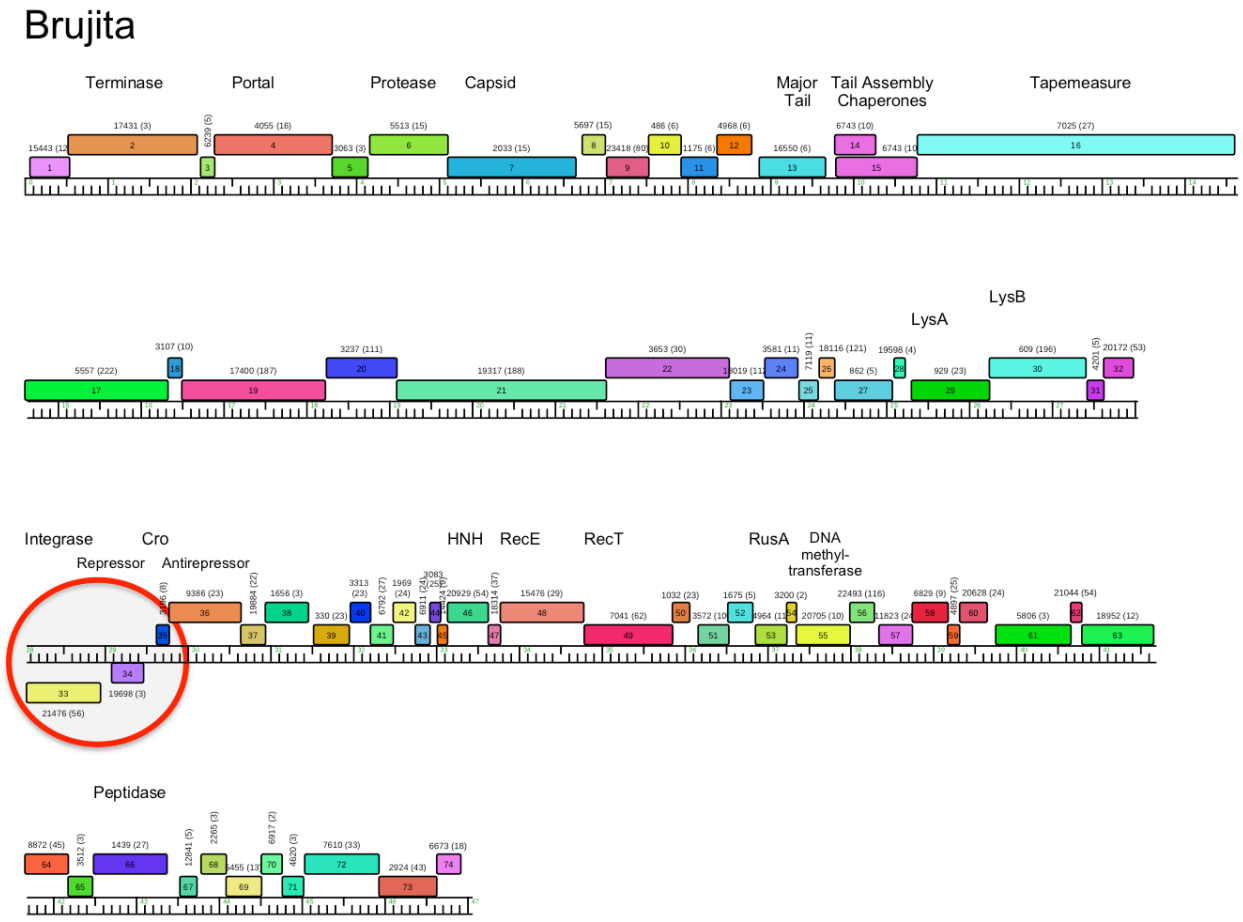


Figure 1: Genome map of mycobacterium phage Brujita

Mycobacteriophage Brujita genome map is shown with a ruler denoting genome length. Boxes above the ruler represent genes transcribed in a rightward direction and those below are transcribed leftward. Predicted functions are listed above the genes. The circled region contains the lysogenic genetic switch.

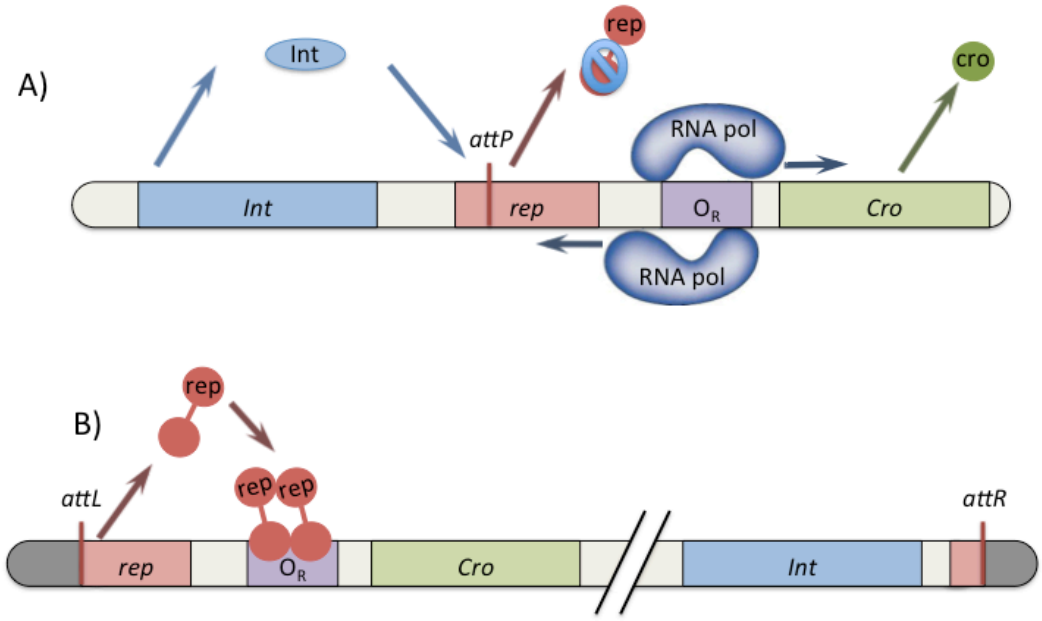


Figure 2: Integration-dependent phage immunity

A. The lysogenic genetic switch of Mycobacterium phage Brujita consists of three genes, *int, rep,* and a putative *cro.* The *attP* core is located within the repressor gene, such that, upon integration, **(B)** the 3' end of the gene is truncated and a shorter functional form of the repressor is made.

Integration-dependent immunity has been studied using mycobacteriophages Brujita, Charlie, and BPs. Genes *int* and *rep* are transcribed in the leftward direction and a putative *cro* is transcribed rightward (Figure 2). *Rep* and *cro* are separated by 155 bp, a region that contains the operator sites and two promoters, P_{Rep} and P_R (Broussard & Hatfull, 2013, Villanueva *et al.*, 2015). P_{Rep} drives expression of *rep* and *int*, while P_R drives expression of *cro*. The operator, O_R is a 12 base pair sequence, which can be bound by Rep to block expression of *cro* (Villanueva *et al.*, 2015).

In contrast to phage Lambda, integration-dependent immunity is largely regulated through protein stability as both Int and Rep contain DNA encoded proteolytic degradation tags with limited similarity to the ssrA tag found in *M. smegmatis* (Broussard *et al.*, 2013). These tags cause

degradation when attached to GFP *in vivo* and are likely used to maintain a rapid turnover of both Int and Rep (Broussard *et al.*, 2013). Since Int is the determining factor in this system, the degradation tag allows Int to be controlled at the host-protease level, where low protease levels would favor Int accumulation, followed by integration (Broussard & Hatfull, 2013). This provides an elegant feedback system, comparable to CII-degradation in *E. coli* (see section 1.2.1).

As in the Lambda system, Brujita integration occurs between *attB* and *attP*. *attB* is found in the bacterial chromosome located in a $tRNA^{Thr}$. Though they integrate at different *attB* sites, other tyrosine integrase systems also use tRNA-associated *attB* sequences (Campbell, 1992, Dupont *et al.*, 1995, Gabriel *et al.*, 1995, Hauser & Scocca, 1992b, Hayashi *et al.*, 1993, Inouye *et al.*, 1991, Kolot & Yagil, 1994, Papp *et al.*, 1993, Pierson & Kahn, 1987, Reiter *et al.*, 1989, Smith-Mungo *et al.*, 1994). Because the host cannot readily alter its tRNA without losing functionality, this presumably is advantageous; thus, the insertion site is maintained over long evolutionary history. Additionally, many of these integrases recombine near the anticodon loop of the tRNA, likely because the built-in dyad symmetry of the anticodon loops provides appropriate spacing for both binding and recombination (Hauser & Scocca, 1992a).

The second insertion site is *attP*, which in Brujita-like systems, is found near the 3' end of *rep* (Figure 2). Upon integration between *attP* and *attB*, *rep* is disrupted and a premature stop codon is introduced, which removes the C-terminal degradation tag. This allows for the expression of a stable and active truncated version of the repressor (Broussard & Hatfull, 2013, Broussard *et al.*, 2013). This tag has been shown to not only cause repressor turnover, but also prevent the normal function of repressor (Broussard *et al.*, 2013). The full-length form of repressor cannot accumulate to bind O_R *in vivo* (though it can bind these sequences *in vitro* (Villanueva *et al.*, 2015))

and does not accumulate to high enough levels to provide superinfection immunity against phages in the same repressor group.

The putative Brujita Cro currently lacks a known function (Broussard *et al.*, 2013). Cro has not been successfully deleted and is assumed to be required for the lytic cycle (Broussard & Hatfull, 2013). Although Cro does reduce lysogenization frequencies when expressed from a plasmid, it does not repress the P_{Rep} promoter (Broussard & Hatfull, 2013, Broussard *et al.*, 2013). It is possible in this system that the putative Cro function is mis-annotated, as it does not appear to have any similarity to repressor or other Cro proteins, but does have limited similarity to Xis and other DNA binding proteins.

1.3 RECOMBINATION

The family of tyrosine recombinases is broad and includes the integrases and simple recombinases, such as Cre, FLP, and XerC/D. Brujita Int is a member of the tyrosine integrase family, though it exhibits many characteristics of the tyrosine recombinases. In order to understand how Brujita Int fits within the context of the family of tyrosine recombinases, this section will detail the characteristics of the well-known recombinases as well as those of canonical integrases.

The lysogenic genetic switch discussed above depends on the function of an integrase protein for lysogen stability. Integrases are members of the larger family of site-specific DNA recombinases, which come in two flavors – the serine and tyrosine recombinases – categorized based on the catalytic residue that attacks the DNA backbone. Serine integrases are characterized by relatively large and diverse two-domain proteins (500-900 residues), small *attB* and *attP* sites (<50bp) (Ghosh *et al.*, 2003, Bibb *et al.*, 2005), and recombination by staggered double-stranded

DNA breaks creating two-base 3' overhangs (Ghosh *et al.*, 2003). Serine integrases also do not require any accessory proteins, but do require recombination directionality factors (RDFs). For example, Bxb1-Int and φC31-Int require an RDF for excision, but instead of performing this role by DNA binding, the RDFs function by binding to Int, using protein-protein interactions to modulate Int activity (Ghosh *et al.*, 2006, Khaleel *et al.*, 2011). In the phage integrases described to date, both serine and tyrosine, all require an Xis for excision (Khaleel *et al.*, 2011).

More than 1300 tyrosine recombinases have been identified (Van Houdt *et al.*, 2012). They form two groups: the phage integrases (Lambda Int, L5 Int, P22 Int, HK022 Int), and the simple recombinases (XerC/D, Cre, FLP) (Groth & Calos, 2004b). Both groups share several features. They require a region of identity located near the positions of cleavage and strand exchange, known as the core. Both classes also share a basic two-domain structure consisting of a core-DNA binding domain and a catalytic domain. Though there is little DNA or amino-acid similarity between any two integrases or recombinases, these proteins perform their respective functions using the same reaction mechanism (Chen *et al.*, 1992, Hoess & Abremski, 1985, Gronostajski & Sadowski, 1985, Evans *et al.*, 1990, Pargellis *et al.*, 1988). This mechanism consists of a pairwise exchange of single DNA strands on opposing duplexes followed by ligation and formation of a Holliday junction intermediate. A second set of cleavages and strand exchange occurs, and the intermediate is resolved into the recombination product. Although sequences differ greatly between the members of the integrase and recombinase families, several conserved residues are found, which make up the active site tetrad (RHRY) (Chen *et al.*, 1992, Groth & Calos, 2004b).

There are many differences between the recombinases and integrases. The recombinases are bidirectional and exhibit no directionality control, a major difference from the integrases (Van

Duyne, 2015, Midonet & Barre, 2014, Jayaram *et al.*, 2015). They also do not require cofactors such as IHF or Fis.

In contrast, the tyrosine integrases use two non-identical DNA sites for recombination (Figure 3). In addition, they contain an N-terminal arm-binding domain. This combination – required for arm-type binding of DNA and control of directionality – is a characteristic shared by nearly all of the previously discovered tyrosine integrase proteins (Groth & Calos, 2004b). Additionally, all discovered integrases, both tyrosine and serine, require an RDF (such as Xis) for excision (Khaleel *et al.*, 2011). Tyrosine recombinases can use two identical sites for recombination and as such, have no recombination directionality control (Groth & Calos, 2004b, Radman-Livaja *et al.*, 2006).

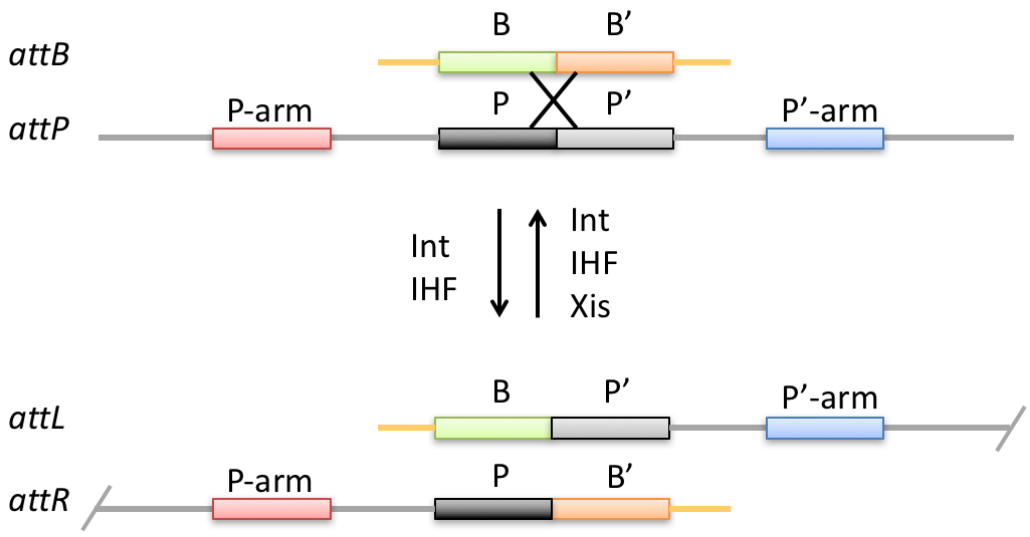


Figure 3: Diagram of integrase mediated site-specific recombination

Tyrosine integrases, such as Lambda Int perform site specific recombination between two sites, *attP* and *attB* (top). The reaction requires four subunits of Int to bind the core DNA sites (B, B', P, and P') as well as the arm-type sites (P-arm and P'-arm) to form synapsis. A cofactor, IHF is also required for the reaction. Integration produces *attL* and *attR* products (bottom), consisting of a different arrangement of the core and arm-sites. The excision reaction is a reverse of integration, requiring an additional protein, Xis to reform *attP* and *attB*.

1.3.1 Mechanism of Tyrosine Mediated Site-Specific Recombination

Though they have largely different sequence and cofactor requirements, all of the tyrosine site-specific recombinases use the same reaction mechanism (Grindley *et al.*, 2006, Groth & Calos, 2004b). Recombination depends upon a conserved active-site tetrad, RHRY. As the tyrosine recombinases do not use metal cofactors, the RHRY tetrad instead provides the positively charged residues to activate the scissile phosphodiester bond (Grindley *et al.*, 2006). Mutation of any one of these residues prevents cleavage of DNA (Chen *et al.*, 1992).

In tyrosine-mediated recombination, the hydroxyl group of the catalytic tyrosine (from the RHRY tetrad) acts as the nucleophile for strand cleavage (Chen *et al.*, 1992). This results in a transesterification and a 3' phosphotyrosyl bond between the recombinase and DNA, and a free 5' hydroxyl group on the cleaved DNA strand (Cao *et al.*, 1997). Subsequently, this 5' hydroxyl group now acts as a nucleophile in a second cleavage step and attacks the phosphotyrosyl bond of the opposite DNA duplex. This results in cleavage of the protein-DNA bond and formation of recombinant DNA with four interlinked DNA strands in a Holliday junction intermediate (Kitts & Nash, 1987, Arciszewska *et al.*, 1995, West, 1997, West, 1996). From this point, the reaction can proceed in one of two directions. The second cleavage step can occur using the same two subunits, resulting in regeneration of original substrate DNAs. Alternatively, the complex can proceed through branch migration where the DNA-protein complexes shift to activate the second pair of proteins, causing a separating and ligation of the spacer DNA. In the tyrosine recombinases, the spacer is 6-8 base pairs, which must be identical between the two DNA duplexes for migration to occur (Furth *et al.*, 1983, Groth & Calos, 2004b, Rajeev *et al.*, 2009). Once branch migration has finished, the second pair of proteins now attacks the phosphate backbone to form the

phosphotyrosine bond. This bond is once again broken by attack of the free hydroxyl group, resulting in the recombinant DNA product (Grindley *et al.*, 2006).

1.3.2 Tyrosine Recombinases

Several tyrosine recombinases have been studied in detail. Of particular importance are FLP, Cre, and XerC/D. This class of proteins perform varied functions in a variety of hosts including chromosome maintenance of replicons and plasmid copy number control. In all cases, these recombinases and integrases require a spacer region of 6-8 bp, which is identical in both DNA sites. This spacer is placed between the cleavage points, and the requirement for spacer identity between attachment sites is due to branch migration, which is inhibited by any mismatches in the spacer region (Furth *et al.*, 1983, Nunes-Duby *et al.*, 1989, Nunes-Duby *et al.*, 1987, Kitts & Nash, 1987, Richet *et al.*, 1988, Nash & Robertson, 1989, Rajeev *et al.*, 2009).

1.3.2.1 FLP Recombinase

FLP recombinase is encoded by the 2 μm plasmid of *Saccharomyces cerevisiae.* It facilitates an increase in copy number of the plasmid during replication to allow daughter cells to maintain the plasmid during cell division (Chen *et al.*, 2000, Sadowski, 1995, Senecoff *et al.*, 1985). FLP inverts sequences between two 599 bp repeats, of which, fewer than 65-bp are required for inversion (Broach *et al.*, 1982). The FLP recombination site contains three 13-bp repeats, the second and third of which are inverted relative to each other and are separated by an 8-bp spacer (Senecoff *et al.*, 1985, Qian & Cox, 1995). The first repeat is positioned directly next to the second, in the same orientation, and is not required for recombination activity (Andrews *et al.*, 1985). The minimal sequence required for *in vitro* recombination is 28-bp, including two 10-bp inverted

repeats and the 8-bp spacer (Senecoff *et al.*, 1985). There is also some flexibility with the length of the spacer region, as FLP can recombine using 7 or 9-bp spacer (Senecoff *et al.*, 1985).

The FLP synaptic complex contains four monomers. Recombination in *trans*, such that the catalytic tyrosine of one protein is shared with its partner and cleaves the DNA bound by the partner protein (Chen *et al.*, 1992, Lee *et al.*, 1996). FLP mediates recombination without additional proteins or accessory factors (Cox, 1983).

1.3.2.2 Cre/*loxP* Recombination

The Cre/*loxP* system is found in *E. coli* bacteriophage P1 and functions to circularize P1 DNA during host infection (Sternberg *et al.*, 1981, Segev & Cohen, 1981). It also resolves concatenates that form during genome replication in the lysogen for efficient segregation (Austin *et al.*, 1981, Sternberg *et al.*, 1981).

Cre recombines between two identical 34 bp *loxP* sites that consist of two 13 bp inverted repeats separated by 8 bp (Hoess *et al.*, 1982). Cre cleavage creates 6 bp 5' DNA overhangs (Shaikh & Sadowski, 1997). Cre binds cooperatively to *loxP* sites (Hoess & Abremski, 1984), cleaves DNA in a *trans*-mechanism (Shaikh & Sadowski, 1997), and requires no additional proteins for recombination (Hamilton & Abremski, 1984, Abremski *et al.*, 1986, Abremski & Hoess, 1985).

1.3.2.3 XerC/D Recombination

The E. coli XerC/D system converts multimers of ColE1 based plasmids, arising by homologous recombination, to monomers. This leads to stable segregation and inheritance of the plasmids (Summers & Sherratt, 1984, Blakely *et al.*, 1991). It also aids in chromosome segregation during cell division by a similar mechanism (Blakely *et al.*, 1991). XerC/D like systems are widely

found in bacteria (Arciszewska & Sherratt, 1995), with BLAST results showing hundreds of potential homologs. This unusual recombination system requires two closely related recombinases, XerC and XerD, which function in tandem (Blakely *et al.*, 1993, Colloms *et al.*, 1990).

Recombination products differ depending on whether the substrate is a plasmid or the bacterial chromosome (Blakely *et al.*, 1991, Blakely *et al.*, 1993, Kuempel *et al.*, 1991). Recombination at *cer*, the ColE1 plasmid site, is exclusively intramolecular and requires accessory proteins, ArgR and PepA, as well as 190-bp of accessory DNA adjacent to the 30-bp core (Blakely *et al.*, 1993, Stirling *et al.*, 1988, Stirling *et al.*, 1989, Summers & Sherratt, 1988, Colloms *et al.*, 1990). These accessory factors and sequences are required for maintaining the intramolecular recombination preference (Blakely *et al.*, 1993). In contrast, recombination at the *dif* recognition sequence in the bacterial chromosome occurs both inter-molecularly and intra-molecularly, requiring only XerC, XerD, and 28-bp of DNA (Blakely *et al.*, 1991, Blakely *et al.*, 1993). The core site consists of 11-bp XerC and XerD binding sites separated by a 6-8 bp spacer (6 bp for *dif* and 8bp for *cer*) (Sherratt *et al.*, 1995). These binding sites have partial dyad symmetry (Hayes & Sherratt, 1997).

XerC/D bind cooperatively to core sites (Colloms *et al.*, 1990, Blakely *et al.*, 1993). Recombination proceeds with top strand exchange, mediated by XerC and occurs in *cis* (Blakely *et al.*, 1993, Arciszewska & Sherratt, 1995). XerD then mediates cleavage and strand exchange of the bottom strand. *Dif* core sequences have a 6-bp spacer region, though recombination can occur at reduced rates with an 8-bp spacer, as is found in the *cer* site (Blake *et al.*, 1997).

1.3.3 Tyrosine Integrases

1.3.3.1 Lambda Integration

Tyrosine integrases differ from tyrosine recombinases in that they contain an additional N-terminal arm-type binding domain, which is used to control directionality of recombination. Their *attB* sequences are similar to recombinase binding sites (~24 bp). However, integrases require large *attP* sites (~240 bp in Lambda) (Mizuuchi & Mizuuchi, 1980).

Lambda Int is perhaps the best understood tyrosine integrase and is prototypical of the other tyrosine integrase systems. This 3-domain protein recombines an attachment sites in the bacterial chromosome with a site in the phage genome (*attB* and *attP,* respectively). During Lambda lysogeny, CII activates the P_I promoter, which transcribes *int*, but not *xis* (Casjens & Hendrix, 2015). The *int* mRNA from the P_L promoter contains a *sib* site, which leads to its degradation, causing Int to accumulate only in the presence of high levels of CII-activated P_I transcription, ensuring that the phage is committed to lysogeny (Schindler & Echols, 1981, Gottesman *et al.*, 1982). If Int is present in large enough concentration, it will facilitate recombination of the phage genome into the host chromosome, leading to a stable prophage existence.

Upon DNA stress caused by UV light, mitomycin C, or other factors, the prophage induces expression of *int* and *xis* and undergoes excision and induction of lytic expression (Casjens & Hendrix, 2015, Oppenheim *et al.*, 2005). The mRNA control ceases after integration, and moderate levels of Int and Xis are sufficient to cause excision (Gottesman *et al.*, 1982, Schindler & Echols, 1981).

Although the Lambda *attB* sequence is 24 bp, the *attP* site requires 240 base pairs of DNA (Mizuuchi & Mizuuchi, 1980). *attP* contains binding sites for IHF (H1, H2, and H3), Xis (X1 and

X2) (Craig & Nash, 1984, Gardner & Nash, 1986), and arm-type interactions (P1, P2, P'1, P'2, P'3) (Hsu *et al.*, 1980, Ross & Landy, 1983). Both *attP* and *attB* share a region of identity, known as the core (C and C').

Lambda Int contains three domains: The N-terminal arm-type binding domain (NTD), the core DNA binding domain, and the C-terminal catalytic domain (CTD). The NTD binds with high affinity to the *attP* arm-sites, which flank the core by 40-120 bp. Subsequently, the core DNA sites C and C' are recognized by the core-binding domain (Seah *et al.*, 2014, Gottfried *et al.*, 2000). Lambda Int simultaneously binds to both arm-sites and the core-sites, which requires severe DNA bending facilitated by IHF for integrative recombination (Seah *et al.*, 2014, Kim & Landy, 1992, Miller *et al.*, 1979). The Int-*attP* complexes then capture *attB* DNA, forming a synaptic complex and allowing recombination to occur. Lambda can cleave DNA using a *cis*-mechanism, meaning that the catalytic tyrosine attacks DNA that is bound by the same subunit (Nunes-Duby *et al.*, 1994). However, some data suggests a *trans*-mechanism – or nucleophilic attack of DNA bound by a neighboring subunit – can occur (Han *et al.*, 1993). It is possible that Lambda is capable of both methods of cleavage and may prefer one or the other depending on the environment.

The Lambda Int C-terminal tail has a novel function in that it stabilizes the nucleophilic tyrosine of the neighboring Int subunit in *trans* (Landy, 2015). This region (approximately 30 residues) appears to reposition between the monomeric form and the tetrameric Int bound to Holliday junction (HJ) DNA (Kwon *et al.*, 1997, Aihara *et al.*, 2003). This region includes the catalytic tyrosine (Y342), which repositions 26 Å between the two structures. Additionally, the extreme C-terminus (including strand β-7) extends from one promoter of Int to the active site of another, forming a *trans*-interaction that may coordinate activity of the subunits during synapsis and recombination (Landy, 2015).

The Recombination Directionality Factor (RDF), also known as Xis, is a DNA binding and bending protein (Yin *et al.*, 1985, Sam *et al.*, 2004, Abremski & Gottesman, 1982, Kim & Landy, 1992). Xis alters the macromolecular architecture of the complex, partnering with previously used IHF, Int, and *attL* and *attR* DNA to allow the formation of the excisive synaptic complex. Xis is not only required for excisive intasome formation, but also heavily inhibits the formation of integrative complexes by binding to X1 and X2 on *attP* DNA (Moitoso de Vargas *et al.*, 1989, Ross & Landy, 1983, Richet *et al.*, 1988). Therefore, the presence or absence of Xis controls the directionality of the recombination reaction. Fis is a host-encoded factor that stimulates inversions (Seah *et al.*, 2014). It can also aid excisive recombination under low Xis concentrations by binding to its site, which corresponds to the Xis binding site, X2 (Numrych *et al.*, 1991, Ball & Johnson, 1991, Thompson *et al.*, 1987b)

1.3.3.2 L5 Integration

Another well-studied, canonical tyrosine integrase is L5 Int. Mycobacterium phage L5 is a temperate phage that infects *M. smegmatis,* as does phage Brujita. However, the L5 genetic switch is similar to that of Lambda, but lacks CII and CIII. It is not known if L5 has anti-termination, but it does have some terminators (Brown *et al.*, 1997, Donnelly-Wu *et al.*, 1993). There are approximately 30 asymmetric repressor binding sites throughout the L5 genome, known as stoperators, to which the repressor binds in order to cumulatively shut down gene expression (Pope *et al.*, 2011, Brown *et al.*, 1997). Even with these differences, the general scheme for integration and excision are similar to those of Lambda (Lewis & Hatfull, 2003, Pena *et al.*, 2000, Pena *et al.*, 1997b, Azaro & Landy, 2002).

The L5 integration system has a 43-bp core sequence that is found in all four DNA sites (*attB, attP, attL,* and *attR*), and overlaps a host $tRNA^{Gly}$ gene (Lee *et al.*, 1991, Pena *et al.*, 1996).

Upon integration, the *attP* core sequence replaces part of the tRNA gene, preserving the integrity of the tRNA, which is required for *M. smegmatis* growth (Pena *et al.*, 1997b). Only 29-bp of the *attB* core sequence are required for recombination (Pena *et al.*, 1996). L5 *attP* is a 240 bp sequence, which contains binding sites for Int-arm interactions (P1-P7), Int-core interactions, mIHF binding, and Xis binding. Although L5 *attP* has seven arm-sites, only four of them are needed for integration (P1, P2, P4, and P5) (Pena *et al.*, 1997b). mIHF is required for recombination, but does not bind DNA independent of Int. mIHF is also thought to make protein-protein contacts with Int (Lewis & Hatfull, 2003, Pena *et al.*, 1999, Pena *et al.*, 2000)

Like Lambda Int, L5 Int is a three-domain protein containing N-terminal arm binding (NTD), core binding, and C-terminal catalytic domains (CTD). L5 uses a different host factor, mIHF to facilitate DNA bending and the site locations and orientations differ from that of Lambda (Pena *et al.*, 1997b, Pedulla *et al.*, 1996). Also like Lambda, L5 forms higher order intasome complexes with IHF binding on both sides of the core (Pedulla *et al.*, 1996, Pena *et al.*, 1999, Richet *et al.*, 1986, Thompson *et al.*, 1987a). However, L5 only uses the right arm sites, P4 and P5, to form this complex. The left sites, P1 and P2 are bound by Int, which can then capture *attB* DNA (Pena *et al.*, 2000, Pena *et al.*, 1999, Lewis & Hatfull, 2003).

L5 Int cleavage generates 7 bp 5' extensions, similarly to Lambda (Pena *et al.*, 1996). Like the other tyrosine integrases, it requires an RDF, or Xis protein, for excision and directionality control (Lee & Hatfull, 1993, Pedulla *et al.*, 1996). Unlike the Lambda system, L5 Xis does not interact with Int, but functions by DNA bending alone (Lewis & Hatfull, 2003). Additionally, supercoiling of either *attB* or *attP in vitro* increases the frequency of the reaction, but is not required (Pena *et al.*, 1998).

1.4 INVESTIGATING BRUJITA INTEGRASE

Brujita Int is a model for a new subclass of tyrosine integrases that deviate from the canonical systems (such as Lambda Int and L5 Int). Members of this new subclass do not contain N-terminal arm binding domains and do not appear to use these interactions to control recombination directionality. Additionally, Brujita Int appears to be more similar to the tyrosine recombinases. How then does Brujita Int control directionality? How does it function and what are its DNA and protein requirements?

The following chapters will address these questions using *in vivo* and *in vitro* recombination assays to investigate the DNA and protein requirements for recombination. DNA binding assays will be used to identify binding sites, important nucleotides, and cleavage points. Finally, X-ray crystallography combined with analysis of specific protein mutants will help identify requirements for synapsis and complex formation.

2.0 DISSECTION OF SEQUENCE REQUIREMENTS FOR BRUJITA RECOMBINATION

Work discussed in this chapter was published in the following articles:

Broussard, G.W., Oldfield, L.M., Villanueva, V.M., **Lunt, B.L.**, Shine, E.E., Hatfull, G.F., January 2013. Integration-Dependent Bacteriophage Immunity Provides Insights into the Evolution of Genetic Switches. *Molecular Cell,* Volume 49, Issue 2, Pages 237-248.

Lunt, B.L., and Hatfull, G.F., January 2016. Brujita Integrase: A Simple, Arm-Less, Directionless, and Promiscuous Tyrosine Integrase System. *J Mol Biol*, Volume 428, Issue 11, Pages 2289-2306.

Figures 4, 6, and 7 are adapted from Broussard *et al.,* 2013. Figures 8, 9, and 10 are adapted from Lunt and Hatfull, 2016. The experiments discussed below were performed by the author unless otherwise stated. Emily Shine performed the experiments testing P_{Int} for activity. Many of the figures and tables were made in collaboration with Gregory Broussard and Graham Hatfull.

2.1 INTRODUCTION

Broussard *et al.* showed that Brujita Int contains no arm-type binding domain and most likely does not contain the cognate arm-binding DNA sites (Broussard *et al.*, 2013). Without the arm-type binding domain, the questions arise of whether Brujita Int has an alternative mechanism for arm-type interactions, whether it controls directionality of recombination, and whether it forms a similar synaptic complex to the prototypical Lambda integration system. Interactions between the

arm-binding domain and DNA arm sites are used by canonical tyrosine integrases to control recombination directionality (Van Duyne, 2005). Additionally, Lambda Int requires arm-type interactions for integration as Lambda Int without this domain can nick DNA, but lacks recombination activity (Landy, 2015, Tirumalai *et al.*, 1997).

Toward this end, I will first characterize the Brujita DNA site requirements. This chapter discusses the DNA and cofactor requirements of Brujita Int.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial Growth

M. smegmatis cultures were grown in 7H9 broth or 7H10 agar (Difco) with carbenicillin (50 μg/ml), cyclohexamide (10 μg/ml), and appropriate antibiotics. *E. coli* cultures were grown in LB broth or agar (Difco) with appropriate antibiotics.

2.2.2 Plasmid Construction

Plasmid pBL03 was created by site-directed mutagenesis, which deleted the *attP* core (Brujita 29,173-29,211) from pGWB87. Plasmid BL04 was created by site-directed mutagenesis, changing TGG to TAG by making the mutation in pGWB87 synonymous with Brujita C28935T. Plasmid BL05 was created by site-directed mutagenesis, changing TTG to TAG by making the mutation synonymous with Brujita A28980T in pGWB87. Plasmid BL06 was created by cloning the PCR amplified fragment Brujita 28,970-29,411 into the Hind III and Nhe I sites of pYUB854 (Bardarov

et al., 2002). Plasmid BL08 was created by cloning the PCR amplified fragment Brujita 28,970-29,311 into the Hind III and Nhe I sites of pYUB854. Plasmid BL10 was created by cloning the PCR amplified fragment Brujita 28,970-29,231 into the Hind III and Nhe I sites of pYUB854. Plasmid BL14 was created by cloning the PCR amplified fragment Brujita 28,970- 29,211 into the Hind III and Nhe I sites of pYUB854. Plasmid BL26 was created by cloning the PCR amplified fragment Brujita 28,970-29,206 into the Hind III and Nhe I sites of pYUB854.

Plasmid BL15 was created by cloning the PCR amplified fragment Brujita 29,073-29,411 into the Hind III and Nhe I sites of pYUB854. Plasmid BL23 was created by cloning the PCR amplified fragment Brujita 29,138-29,411 into the Hind III and Nhe I sites of pYUB854. Plasmid BL24 was created by cloning the PCR amplified fragment Brujita 29,143-29,411 into the Hind III and Nhe I sites of pYUB854. Plasmid BL39 was created by cloning the PCR amplified fragment Brujita 29,138-29,211 into the Hind III and Nhe I sites of pYUB854. Plasmid BL40 was created, using site-directed mutagenesis and changing ATG to ATC by making the mutation synonymous with Brujita C28946G in pGWB87. Plasmid BL41 was created, using site-directed mutagenesis and changing the TTG to CTC by making the mutations synonymous with Brujita C28979G and A28981G in pGWB87. Plasmid BL42 was created, using site-directed mutagenesis and changing the ATG to ATC by making the mutation synonymous with Brujita C29009G in pGWB87.

Plasmid pBL61 expressing Int^{Long} was constructed as follows: plasmid pET21a was digested with Nde I and Hind III and a 957 bp Int^{Long} fragment corresponding to Brujita coordinates 29011 - 28057 was PCR amplified and joined to pET21 using Gibson Assembly (Invitrogen). Plasmid pBL60 expressing Int^{Short} was constructed similarly using a 894 bp PCR fragment containing Brujita coordinates 28948 – 28057. Plasmid pBL51 expressing an MBP-Int fusion was constructed by linearizing plasmid pLC3 (Owens *et al.*, 2006) with Nde I and Hind III and ligating

it with the 957 bp Int^{Long} PCR product. Plasmid pBL83 was made by digestion of pBAD28, which contains an arabinose inducible promoter, (Guzman *et al.*, 1995) with Kpn I and Xba I and ligating it with a PCR amplified fragment consisting of Brujita Int^{Long} cleaved with Kpn I and Xba I. All Int expressing plasmids contain a stabilizing mutation (A296E).

Plasmids containing *attP* and *attB* were created from a pJL37 backbone (Bibb & Hatfull, 2002). The hsp60 promoter was removed by digestion with Xba I and Nde I and replaced with a 999 bp hygromycin resistance marker, PCR amplified from pYUB854 (Bardarov *et al.*, 2002). *attP* constructs of differing lengths (shown in Figure 9), were PCR amplified and inserted into the Bam HI and Nde I sites, and *att*B constructs were similarly inserted into the Not I and Kpn I sites using either Infusion (Clontech) or Gibson Assembly (Invitrogen). All plasmid constructs were verified by Sanger sequencing. Counter-selection plasmids (pBL90 derivatives) were modified from pBL50-derived plasmids by digestion with Dra I and cloning of a PCR amplified *ccdB* gene and upstream sequence (including the promoter region, 669 bp total) from pDONR/Zeo (Thermo Fisher Scientific). All *ccdB* plasmids were grown in One Shot *ccdB* survival competent cells (Thermo Fisher Scientific).

Attachment site mutants were generated using Quick Change Site-Directed Mutagenesis (Stratagene) to generate mutants shown in Figure 10. Mutants were PCR amplified and cloned into the pBL90 backbone as described above. Mutants were recombined using integrase expressing pBL83 cells as before. Following recombination, plasmid was extracted from transformed colonies and sequenced to determine the presence of the mutation. Start site mutants (Figure 4) were generated using Site-Directed Mutagenesis with pGWB87 as a template.

2.2.3 Integration and Stability in Mycobacteria

The Brujita *attP* requirements for integration were determined by co-transformation of *M. smegmatis* mc^2155 with *attP*-containing plasmids and *int*-expressing plasmid pBL03. Plasmid stability was determined by growing *M. smegmatis* containing pBL06 (containing 442bp *attP*), pGWB81 (an integrative plasmid containing *rep* and WT, unstable *int*), and pGWB87 (an integrative plasmid containing *rep* and stabilized *int*) (Broussard *et al.*, 2013) to saturation with selection, diluting 1:10,000 into selection-free media, and growing to saturation twice, for a total of approximately 26 generations. Cultures were plates for individual colonies and were tested for drug resistance.

2.2.4 Expression and Purification of Brujita Int

Plasmid pBL61, which places Int^{Long} under the *E. coli* lac promoter, was transformed into BL21-CodonPlus (DE3)-RIPL cells (Agilent). An isolated colony was used to inoculate 50 ml of LB and antibiotics and incubated overnight at 37°C. After incubation, the cells were pelleted and used to inoculate 2 liters of autoinduction media (Studier, 2005). The cells were incubated with shaking for 18 hours at room temperature, pelleted, and frozen at -80° C. Thawed pellets were resuspended in 10 ml/gram lysis buffer (10 mM phosphate pH 7.0, 400 mM NaCl, 5 mM BME, 5 mM imidazole, 5% glycerol), protease inhibitors added (PMSF, leupeptin, aprotinin, and pepstatin) and lysed by three passes through a homogenizer. The lysed cells were spun at 30,000 x g in a Sorvall Lynx 6000 centrifuge for 40 minutes at 4° C. Untagged Int was isolated using nickel affinity chromatography as native integrase has been found to bind well to nickel, followed by heparin affinity chromatography. Integrase containing fractions were further purified by size exclusion

chromatography, dialyzed into freezer storage buffer, (10 mM Phosphate pH 7.0, 400 nM NaCl, 50% glycerol, 5 mM BME) and stored at -20° C.

2.2.5 *In Vivo* Recombination Assays

To assay for integrase-mediated recombination in *M. smegmatis*, 50 ng of an *int*-expressing plasmid pBL03 (Broussard *et al.*, 2013) and 50 ng plasmid pLB50 were co-electroporated into *M. smegmatis* mc^2155 and transformants were recovered on solid media containing kanamycin. Colonies were then picked onto plates containing either kanamycin or kanamycin and hygromycin. Recombination frequency was recorded as the percentage of transformants isolated on kanamycin related to those selected with both antibiotics.

For *E. coli* recombination assays, plasmid pBL83 was transformed into Stellar competent cells (Clontech), an isolated colony selected, and grown in LB broth with antibiotics overnight to saturation. Three ml of overnight culture were inoculated into 100 ml fresh broth with antibiotics and grown at 37° C with vigorous shaking to $O.D._{600} = 0.5$. Forty-five minutes after inoculation, arabinose was added to final concentration of 40 mM to induce Int expression. Cells were pelleted and washed 4 times with 10% ice-cold glycerol, frozen on dry ice, and stored at -80° C until use. Un-induced control cells were made similarly without the addition of arabinose. Recombination was measured by electroporating 50 ng of substrate plasmid DNA containing directly oriented attachment sites into Stellar *E. coli* cells carrying pBL83 and selecting kanamycin resistant transformants. Recombination frequencies were determined as cfu/μg DNA. Error values represent the average efficiency ± one standard error of the mean value for three separate transformations (six for the experiment shown in Fig. 9).

2.2.6 *In Vitro* Reaction Conditions

In vitro reactions were performed in XerD reaction buffer (Colloms *et al.*, 1996), containing 50 mM Tris, pH 8.0, 50 nM NaCl, 1.25 mM EDTA, 5 mM spermidine, 1 mM L-arginine, 25 µg/ml BSA, and 10% glycerol. Approximately 100 ng DNA and dilutions of Int ranging from 0 to 625 nM (final concentration) were added and the reactions were incubated at 37°C for 2 hours. Following incubation, half of the reaction was linearized by adding 1 ul NEB Buffer 2.1 and 0.5 µl NotI to each sample, incubating for 2 hours at 37°C. Each sample was then added to 1 µl 10mg/ml Proteinase K and 1 µl 10% SDS and incubated at 65°C for 20 minutes. Linearized and non-linearized samples were run on 0.6% agarose TBE gels overnight at 20V and stained with ethidium bromide.

Table 1. Plasmids used in Chapter 2.

Plasmid Name	Description	Reference
pBL03	pGWB87 with SDM attP core (39bp) deletion	Broussard *et al*., 2013
pBL04	pGWB87 with SDM TGG-TAG (downstream of third start site)	Broussard *et al*., 2013
pBL05	pGWB87 with SDM TTG-TAG (second start site)	Broussard *et al*., 2013
pBL06	pYUB854 with 442 bp attP (203L, 239R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL08	pYUB854 with 342 bp attP (203L, 139R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL10	pYUB854 with 262 bp attP (203L, 59R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL14	pYUB854 with 242 bp attP (203L, 39R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL15	pYUB854 with 339 bp attP (100L, 239R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL23	pYUB854 with 274 bp attP (35L, 239R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL24	pYUB854 with 269 bp attP (30L, 239R) cloned into Hind III and NheI sites	Broussard *et al*., 2013

pBL26	pYUB854 with 237 bp attP (203L, 34R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL39	pYUB854 with 74 bp attP (35L, 39R) cloned into Hind III and NheI sites	Broussard *et al*., 2013
pBL40	pGWB87 with SDM ATG-ATC (third start site)	Broussard *et al*., 2013
pBL41	pGWB87 with SDM TTG-CTC (second start site)	Broussard *et al*., 2013
pBL42	pGWB87 with SDM ATG-ATC (first start site)	Broussard *et al*., 2013
pBL45	pGWB87 with SDM ATG-ATC (third start site) and TTG-CTC (second start site)	This study
pBL46	pGWB87 with SDM ATG-ATC (third start site) and ATG-ATC (first start site)	This study
pBL47	pGWB87 with SDM TTG-CTC (second start site) and ATG-ATC (first start site)	This study
pBL50	pJL37 with Hsp60 replaced with HygR (from pYUB854) cloned into XbaI and NdeI sites, attB436 cloned into KpnI and NotI sites, and attP442 cloned into BamHI and NdeI sites	Lunt and Hatfull, 2016
pBL51	pLC3 with Brujita IntLong cloned into NdeI and HindIII sites	Lunt and Hatfull, 2016
pBL61	pet21a with Brujita IntLong cloned into NdeI and HindIII sites. Stop codons were introduced before His tag	Lunt and Hatfull, 2016
pBL83	pBAD28 with Brujita Int cloned into KpnI and XbaI sites	Lunt and Hatfull, 2016
pBL90	pBL50 with ccdB (from pDONRZeo) cloned into DraI site	Lunt and Hatfull, 2016
pBL103	pBL90 with 242 bp attP (203L, 39R)	Lunt and Hatfull, 2016
pBL105	pBL90 with 274 bp attP (35L, 239R)	Lunt and Hatfull, 2016
pBL106	pBL90 with 269 bp attP (30L, 239R)	Lunt and Hatfull, 2016
pBL111	pBL90 with 270 bp attB (35L, 235R)	Lunt and Hatfull, 2016
pBL116	pBL90 with 240 bp attB (201L, 39R)	Lunt and Hatfull, 2016
pBL118	pBL90 with 284 bp attP (45L, 239R)	Lunt and Hatfull, 2016
pBL120	pBL90 with 232 bp attP (203L, 29R)	Lunt and Hatfull, 2016
pBL144	pBL90 with attL440 and attR438 replacing attB and attP	Lunt and Hatfull, 2016
pBL145	pBL90 with 217 bp attP (203L, 14R)	Lunt and Hatfull, 2016
pBL146	pBL90 with 212 bp attP (203L, 9R)	Lunt and Hatfull, 2016
pBL147	pBL90 with 215 bp attB (201L, 14R)	Lunt and Hatfull, 2016
pBL148	pBL90 with 210 bp attB (201L, 9R)	Lunt and Hatfull, 2016
pBL149	pBL90 with 250 bp attB (15L, 235R)	Lunt and Hatfull, 2016
pBL150	pBL90 with 245 bp attB (10L, 235R)	Lunt and Hatfull, 2016
pBL151	pBL90 with 205 bp attB (201L, 4R)	Lunt and Hatfull, 2016
pBL152	pBL90 with SDM changing position 1 of the attB core T-A	Lunt and Hatfull, 2016

pBL153	pBL90 with SDM changing position 2 of the attB core A-T	Lunt and Hatfull, 2016
pBL154	pBL90 with SDM changing position 3 of the attB core A-T	Lunt and Hatfull, 2016
pBL155	pBL90 with SDM changing position 4 of the attB core T-A	Lunt and Hatfull, 2016
pBL156	pBL90 with SDM changing position 5 of the attB core G-C	Lunt and Hatfull, 2016
pBL157	pBL90 with SDM changing position 6 of the attB core A-T	Lunt and Hatfull, 2016
pBL158	pBL90 with SDM changing position 7 of the attB core A-T	Lunt and Hatfull, 2016
pBL159	pBL90 with SDM changing position 8 of the attB core T-A	Lunt and Hatfull, 2016
pBL160	pBL90 with SDM changing position 9 of the attB core A-T	Lunt and Hatfull, 2016
pBL161	pBL90 with SDM changing position 1 of the attP core T-A	Lunt and Hatfull, 2016
pBL162	pBL90 with SDM changing position 2 of the attP core A-T	Lunt and Hatfull, 2016
pBL163	pBL90 with SDM changing position 3 of the attP core A-T	Lunt and Hatfull, 2016
pBL164	pBL90 with SDM changing position 4 of the attP core T-A	Lunt and Hatfull, 2016
pBL165	pBL90 with SDM changing position 5 of the attP core G-C	Lunt and Hatfull, 2016
pBL166	pBL90 with SDM changing position 6 of the attP core A-T	Lunt and Hatfull, 2016
pBL167	pBL90 with SDM changing position 7 of the attP core A-T	Lunt and Hatfull, 2016
pBL168	pBL90 with SDM changing position 8 of the attP core T-A	Lunt and Hatfull, 2016
pBL169	pBL90 with SDM changing position 9 of the attP core A-T	Lunt and Hatfull, 2016
pBL171	pBL90 with 14 bp attB (5L, 9R)	Lunt and Hatfull, 2016
pBL173	pBL90 with 49 bp attP (35L, 14R)	Lunt and Hatfull, 2016
pBL177	pBL144 with 240 bp attL (201L, 39R)	Lunt and Hatfull, 2016
pBL178	pBL144 with 215 bp attL (201L, 14R)	Lunt and Hatfull, 2016
pBL179	pBL144 with 210 bp attL (201L, 9R)	Lunt and Hatfull, 2016
pBL180	pBL144 with 274 bp attL (35L, 239R)	Lunt and Hatfull, 2016
pBL181	pBL144 with 249 bp attL (10L, 239R)	Lunt and Hatfull, 2016
pBL182	pBL144 with 244 bp attL (5L, 239R)	Lunt and Hatfull, 2016
pBL184	pBL144 with 19 bp attL (5L, 14R)	Lunt and Hatfull, 2016
pBL186	pBL144 with 242 bp attR (203L, 39R)	Lunt and Hatfull, 2016
pBL187	pBL144 with 212 bp attR (203L, 9R)	Lunt and Hatfull, 2016

pBL188	pBL144 with 207 bp attR (203L, 4R)	Lunt and Hatfull, 2016
pBL189	pBL144 with 335 bp attR (100L, 235R)	Lunt and Hatfull, 2016
pBL190	pBL144 with 270 bp attR (35L, 235R)	Lunt and Hatfull, 2016
pBL191	pBL144 with 260 bp attR (30L, 235R)	Lunt and Hatfull, 2016
pBL192	pBL144 with 44 bp attR (35L, 9R)	Lunt and Hatfull, 2016
pBL200	pBL144 with 217 bp attR (203L, 14R)	Lunt and Hatfull, 2016
pBL202	pBL144 with 250 bp attR (25L, 235R)	Lunt and Hatfull, 2016
pBL204	pBL90 with 240 bp attB (5L, 235R)	Lunt and Hatfull, 2016
pBL205	pBL90 with 235 bp attB (0L, 235R)	Lunt and Hatfull, 2016
pBL206	pBL144 with 205 bp attL (201L, 4R)	Lunt and Hatfull, 2016
pBL207	pBL144 with 239 bp attL (0L, 239R)	Lunt and Hatfull, 2016
pBL208	pBL90 with 264 bp attP (25L, 239R)	Lunt and Hatfull, 2016
pBL213	pBL90 with SDM changing position 1 of the attP and attB cores T-A	Lunt and Hatfull, 2016
pBL214	pBL90 with SDM changing position 2 of the attP and attB cores A-T	Lunt and Hatfull, 2016
pBL215	pBL90 with SDM changing position 3 of the attP and attB cores A-T	Lunt and Hatfull, 2016
pBL216	pBL90 with SDM changing position 4 of the attP and attB cores T-A	Lunt and Hatfull, 2016
pBL217	pBL90 with SDM changing position 5 of the attP and attB cores G-C	Lunt and Hatfull, 2016
pBL218	pBL90 with SDM changing position 6 of the attP and attB cores A-T	Lunt and Hatfull, 2016
pBL219	pBL90 with SDM changing position 7 of the attP and attB cores A-T	Lunt and Hatfull, 2016
pBL220	pBL90 with SDM changing position 8 of the attP and attB cores T-A	Lunt and Hatfull, 2016
pBL221	pBL90 with SDM changing position 9 of the attP and attB cores A-T	Lunt and Hatfull, 2016
pBL222	pBL144 with 24 bp attL (5L, 19R)	Lunt and Hatfull, 2016
pBL223	pBL144 with 49 bp attR (35L, 14R)	Lunt and Hatfull, 2016
pBL224	pBL90 with attB436 replacing attP (attB in both locations)	Lunt and Hatfull, 2016
pBL225	pBL90 with attP442 replacing attB (attP in both locations)	Lunt and Hatfull, 2016
pBL226	pBL224 with attB74 (35L, 39R) in both sites	Lunt and Hatfull, 2016
pBL227	pBL224 with attB74 (35L, 39R) in the KpnI and NotI sites, and attB19 (5L, 14R) in the BamHI and NdeI sites	Lunt and Hatfull, 2016
pBL228	pBL224 with attB19 (5L, 14R) in both sites	Lunt and Hatfull, 2016

2.3 RESULTS

2.3.1 Analysis of Brujita Int Start Sites

The Brujita Int coding sequence contains three different translation start sites within the first 22 codons and it is not clear from bioinformatic data which of these is used *in vivo*. To determine which start site is correct, synonymous or conservative site-directed mutants were made in each translation start codon within an integrating plasmid, to test their impact upon Int activity (Figure 4). Integrating plasmids were created to express integrase protein and *attP*, so that when introduced to the host containing an *attB*, the plasmid will integrate into the *attB* site. The integrating plasmid contains a selection marker and does not contain a mycobacterial origin of replication, so that selection of this marker eliminates non-integrating events. Plasmids containing these mutants were transformed into *M. smegmatis* and selected for integration ability.

There is little difference in integration efficiency compared to wild type when mutating the second or third start sites (TTG and ATG, mutants 4 and 3, respectively), suggesting that these mutations do little to alter the activity of Int (Figure 4). Mutant 2 creates a stop codon downstream of all three start sites, and results in no integration, showing that the translational start occurs before this sequence. Additionally, altering the TTG to a TAG stop codon (mutant 1, Figure 4) reduces recombination 100-fold, suggesting that the main start codon occurs at or before this site. However, mutant 1 yields active protein *in vivo*, though perhaps not as efficient as the full-length version, implying that some translation initiates after this codon. Mutation of the first ATG start reduces integration efficiency 10-fold, which we interpret to mean that this is the primary site used *in vivo.* Interestingly, this is not responsible for the entirety of translation initiation, as there is still Int functionality without this site (Figure 4).

In combination, our results show that translation begins with the first ATG start, as it yields the most active protein. Moreover, mutating both of the other two start codons (pBL45, mutants 3,4) results in efficient recombination. Additionally, every combination of mutants that includes allele 5 results in poor integration efficiency. The fact that start positions 2 and 3 yield some activity is likely due to lower levels of translation initiation at these sites or that a less effective Int protein produced.

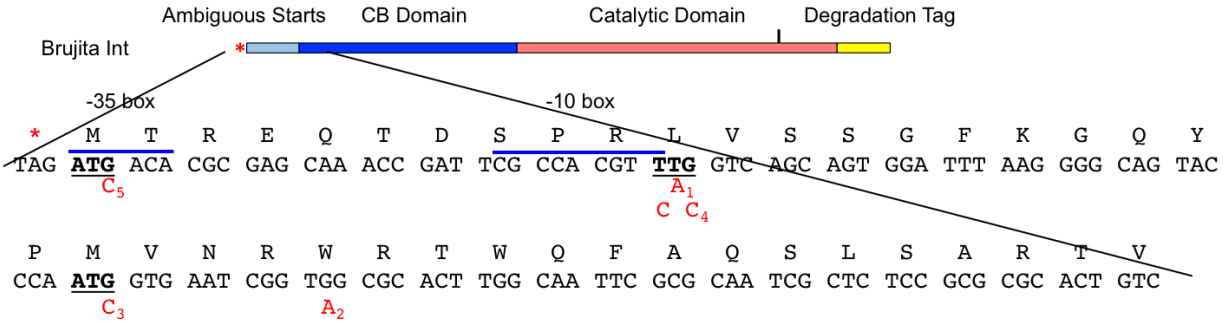
2.3.2 Integrant Stability

One unique feature of the integration-dependent phage immunity systems is the integrase C-terminal degradation tag (Broussard *et al.*, 2013). In Brujita, this tag can be eliminated by mutation of the penultimate alanine to glutamic acid A317E (A296E in early nomenclature due to a mis-annotation). Stabilizing Int by mutating this tag increases integration efficiency nearly 800-fold by stabilizing the protein (Broussard *et al.*, 2013).

Broussard *et al.,* hypothesized that this degradation tag may have some effect on the ability to form lysogens. However, phages containing this mutation decrease the numbers of lysogens formed and instead create clear plaques (Broussard *et al.*, 2013). This effect could result from either poor integration or increased excision. To further clarify, we compared the excision ability of both wild type and the stabilized mutant Int allele (Figure 4). Although the stabilizing Int allele increases integration activity, it also causes unstable integrants (compare excision frequencies of pGWB81 and pGWB87, Figure 4), suggesting that activity of the protein is controlled by Int accumulation, instead of DNA topology, as in Lambda (Landy, 2015).

Brujita integration is likely controlled by Int accumulation through constant turnover of the protein, using the C-terminal degradation tag (Broussard *et al.*, 2013). Upon integration, Int is

separated from its native promoter, P_{Rep}, and placed under the control of some unknown promoter, which would be required for expression of Int in a lysogen (Broussard *et al.*, 2013). One potential source of Int transcription in the lysogen is a predicted promoter in the 5' region of the Int coding sequence (P_{Int}), the resulting mRNA would contain translation start 3, the most downstream site, but not start site 1. To examine the effect of each of these starts on excision, the mutants described above were also tested for integrant stability (Figure 4).



Plasmid	Mutation	Nucleotide Change	Transformation Efficiency (cfu/µg)	Excision Frequency (% of control)
pGWB81	WT (unstable)		1×10^2	0
pGWB87	WT (stable)		6.03×10^3	100
pBL05	1	TTG-TAG	2.9×10^2	109
pBL04	2	TGG-TAG	0	N/A
pBL40	3	ATG-ATC	2.5×10^4	0
pBL41	4	TTG-CTC	8.0×10^3	0
pBL42	5	ATG-ATC	1.1×10^2	0
pBL45	3,4		6.0×10^3	0
pBL46	3,5		8.3×10^1	1
pBL47	4,5		3.7×10^2	0

Figure 4: Brujita Int translational start sites and stability

Brujita Int contains three potential translational start-sites. A variety of mutations were made to test the effect of each site on integration (transformation efficiency) and integrant stability (excision). Mutant alleles are shown in red below the wild type nucleotide, with the corresponding

mutant number. A predicted promoter, P_{Int}, is positioned with the -35 box overlapping the first start site and the -10 box overlapping a single base of the second start site.

Integrated test plasmids containing unstable wild type Int are preserved in every colony tested over 26 generations of unselected growth. In contrast, analogous experiments using the stabilizing mutation leads to only 27% retention of the integrative plasmid over the same time frame. It appears that the action of removing this degradation tag stimulates excision, suggesting that this process (like integration) is also controlled by Int concentration. This control may be due to promoter activation or by a decrease in degradation levels, allowing Int accumulation to lead to excision. However, as noted above, the full-length version of Int (Int^{Long}) cannot be expressed from the predicted promoter P_{Int} (Figure 5). Instead, the only translational product would be Int^{Short}, translated from the most 3' ATG start site (Figure 4). This version of Int is capable of integration, but at much lower levels than Int^{Long}.

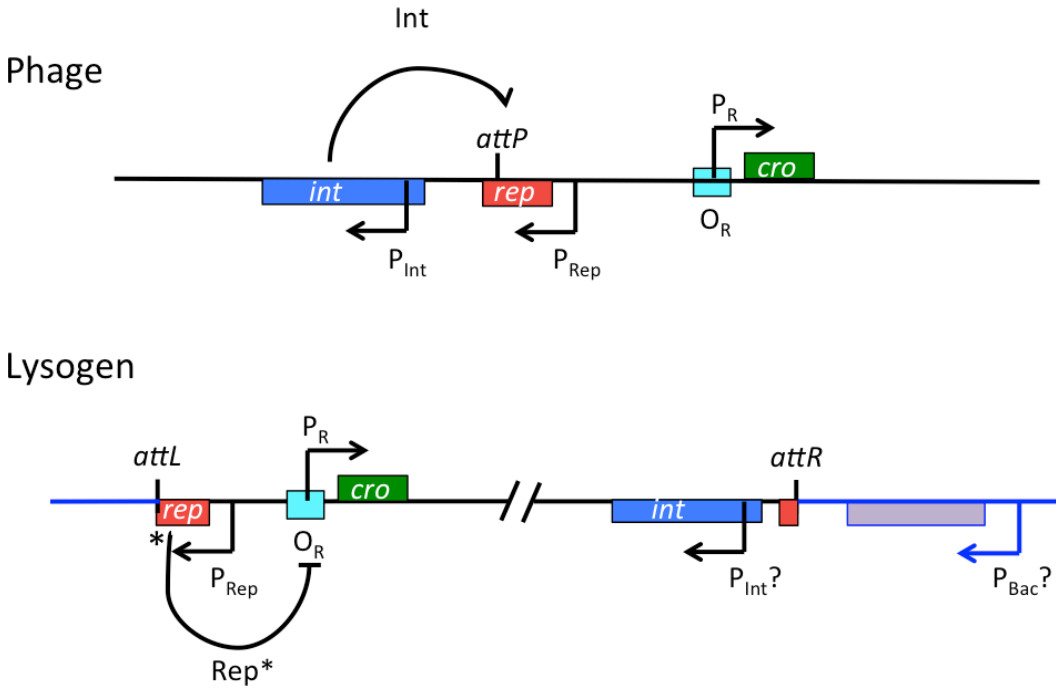


Figure 5: Brujita Int is separated from its native promoter upon integration

The Brujita genetic switch is organized with the *attP* site in the 3' region of *rep*. Upon integration by Int, *rep* is cleaved, producing a truncated, active form of the repressor. After integration, P_{Rep} is separated from Int, and further expression of Int must come from either predicted promoter P_{Int}, or a nearby P_{Bac}.

Mutations in Brujita Int translational start sites show intriguing differences in integration and excision frequencies. Int with mutations in the second or third start site (mutants 4 and 3, pBL41 and pBL40 respectively) only slightly alter recombination and result in stable integrants. Additionally, pBL45 – which contains a competent version of Int^{Long} – is incapable of causing excision, implying that something other than Int length is important for this process. Indeed, a portion of the -10 box of the predicted P_{Int} overlaps mutant 4, which may be sufficient to disable this promoter *in vivo* (Figure 4). Unfortunately, attempts to show activity of this promoter *in vivo* have not been successful (Emily Shine, personal communication), so the validity of this sequence as a promoter is unclear.

2.3.3 Dissection of Brujita Int Attachment Sites

Brujita Int exhibits an unusual integrase that lacks an N-terminal arm-type binding domain common to the other tyrosine integrases, and thus, Brujita may not utilize corresponding arm-type DNA sites that are normally present in *attP*. To determine the minimal *attP* sequences needed for *in vivo* recombination, deletion analyses were performed using a two-part integrative plasmid system. A plasmid expressing A317E stabilized Int was co-transformed into *M. smegmatis* cells along with plasmid containing various sized *attP* sequences. The mixture of these plasmids allows Int to be produced and integrate the second plasmid into the host *attB*. However, since Int containing plasmid cannot replicate, and the integrating (*attP* containing) plasmid lacks Int, it

cannot be excised in future generations. The transformants were selected using a kanamycin resistance marker on the integrating plasmid.

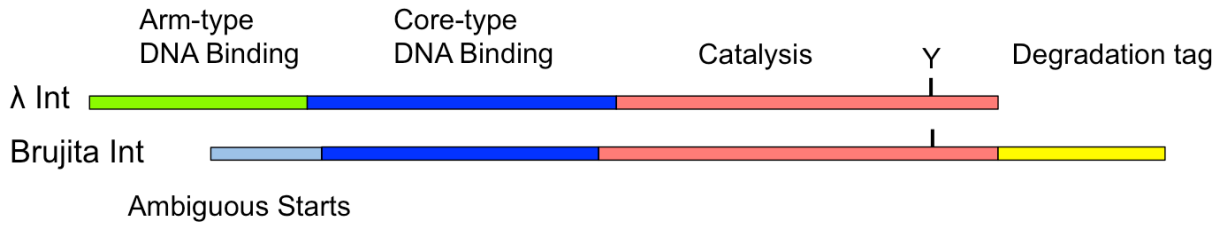


Figure 6: Brujita Int does not contain an arm-binding domain

Schematic of the domains of Lambda Int compared to Brujita Int. Brujita Int does not contain an arm-binding domain, but has an additional C-terminal degradation tag not found in Lambda Int. The Brujita Int start site is ambiguous, as there are three potential starts in the first 22 codons. See Figure 4 for examination of these sites.

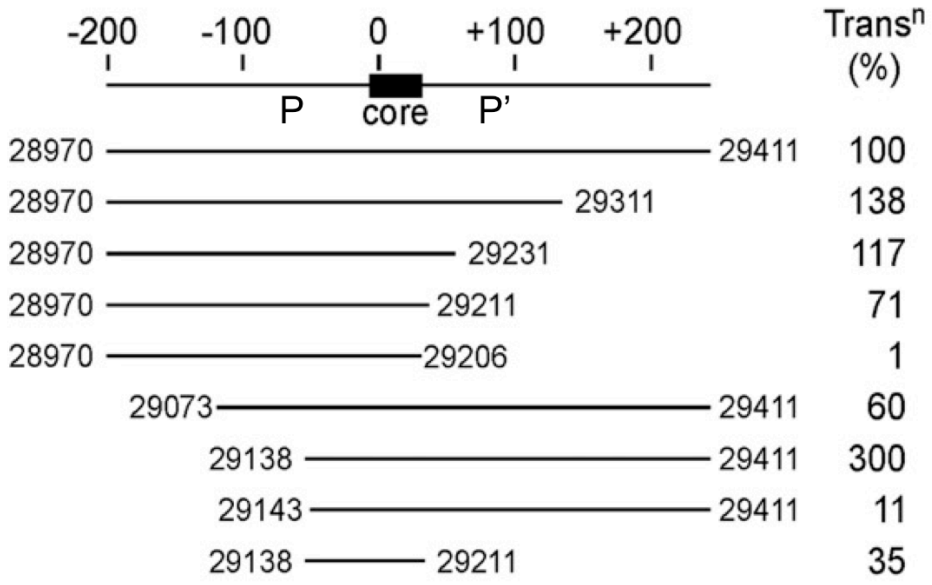


Figure 7: *attP* Mapping in *M. smegmatis*

Brujita *attP* was tested for the minimal DNA requirements by transforming integrating plasmids containing different sized *attP* segments into *M. smegmatis*. The number of transformants were

counted and expressed as a percent of control. Brujita genome coordinates are given as boundaries of tested fragments.

The core region is a segment of identity between *attP* and *attB*. The Brujita *attB* site is located in an essential $tRNA^{Thr}$ (Broussard *et al.*, 2013). Using a normal *attP,* this integration results in a complete $tRNA^{Thr}$ with the *attP* core becoming the 3' half of a functional $tRNA^{Thr}$, due to identity of the core sequences. This identity of the core region is not only required for the recombination mechanism, but additionally is potentially needed to preserve the function of the tRNA upon integration.

Lambda integration requires a 240 bp *attP* (Landy & Ross, 1977, Radman-Livaja *et al.*, 2006). We used the integration assay described above to determine the sequence required for Brujita integration. Truncation of the *attP* sequence from the left side (P) shows little difference in integration until approximately 35 bp left of the core. When this region is removed, transformation efficiencies are decreased 10-fold, suggesting that this sequence is important for integration (Figure 7). Similarly, truncation from the right-side results in little or no loss of transformation affinity until the core region is truncated, a finding that is not surprising, due to its presumed role in replacing a portion of the tRNA upon integration. However, in Brujita, the core is 39 bp, and it is likely that this entire sequence is not required for recombination – but instead for the maintenance of tRNA activity – as the core requirement in other tyrosine recombinases is limited to 6-8 bp (Landy, 2015). From this data, we conclude that the Brujita *att*P site is much smaller than in the classical tyrosine integrase systems, such as Lambda, (240 bp) (Landy & Ross, 1977, Radman-Livaja *et al.*, 2006), whereas this data shows that Brujita functions with 74 base pairs and thus lacks arm-type DNA binding domains (Broussard *et al.*, 2013).

2.3.4 Developing an *In Vivo* Recombination Assay

The Brujita *attP* and the *M. smegmatis attB* share 39 bases of identity, known as the core sequence. As *attB* occurs in this tRNA, a portion of the 74-base pair *attP* requirement is likely needed to preserve this tRNA following integration. Moreover, the actual biochemical requirements of this site may be much smaller, as is the case in the L5 Int system (Pena *et al.*, 1997a). To test this hypothesis, we developed a plasmid integration system, lacking this tRNA, where both *attP* and *attB* flank a hygromycin resistance marker (see Figure 8A for reference). The construct under discussion lacks *ccdB*, such that upon recombination, the marker is lost. Using this system, we can test *attP* constructs smaller than the 74 bases previously found, as well as *attB* fragments, both of which would be impossible in a *M. smegmatis*-based chromosomal integration system.

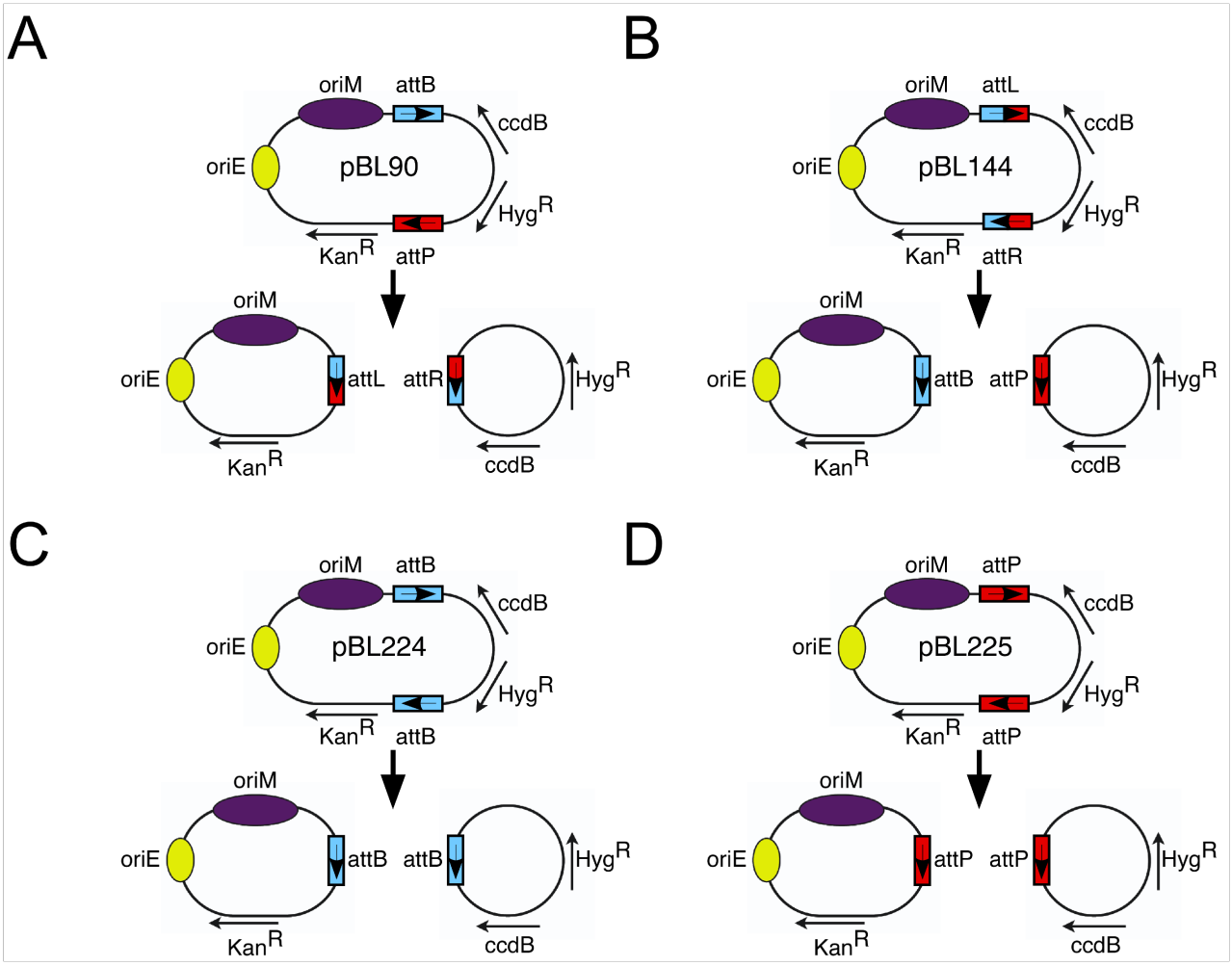


Figure 8: Brujita intra-molecular recombination assays

Plasmids used for recombination assays in vivo. Plasmids pLB90 (**A**), pBL144 (**B**), pBL224 (**C**), and pBL225 (**D**), each contain an E. coli replication origin (oriE), a mycobacterial origin of replication (oriM), kanamycin and hygromycin resistance genes, and the counter-selectable *ccdB* gene (Bernard *et al*., 1994, Hiraga *et al*., 1986). Plasmid pBL90 contains Brujita *attP* and *attB* sites in direct orientation such that recombination between them yields two products, one containing the origins of replication and kanamycin resistance and the other carrying hygromycin resistance and *ccdB*; plasmids pBL144 (*attL* × *attR*), pBL224 (*attB* × *attB*) and pBL225 (*attP* × *attP*) are constructed similarly. Plasmid pBL50 (not shown) is similar to pBL90 but lacks the *ccdB* gene.

To validate this assay, we first verified its function in *M. smegmatis* by co-transforming the plasmid pBL50 (containing *attP-HygR-attB*) along with a plasmid expressing a stabilized Brujita integrase mutant, pBL03. Transformants were screened for loss of hygromycin resistance and compared to control levels. We found that recombination occurs in approximately 41% of colonies tested, and that recombination frequencies of truncated *attP* showed similar trends to those found previously (Broussard *et al.*, 2013).

As we sought to understand whether any cofactors were required, we needed to examine these requirements outside of the context of the mycobacteria. To accomplish this, we first had to verify that the system functions in *E. coli.* We proceeded using a pBAD derived plasmid (pBL83) to express A317E stabilized integrase. These cells were additionally transformed with the pBL50 derived plasmids of interest, and sensitivities of the transformed colonies were determined. Both the *M. smegmatis* and *E. coli* data agree with previous results (data not shown), demonstrating that this plasmid system is a valid substitute for chromosomal integration assays.

While both of these approaches are helpful for validation of the intra-molecular recombination assay, it became clear that a simpler and more efficient system was needed, as the above screens were unable to easily distinguish between good and poor recombination efficiencies. To solve this problem, we developed a counter-selection plasmid by inserting a *ccdB* gene – which is toxic when expressed in a non-permissive strain (Hiraga *et al.*, 1986) – into the region between *attP* and *attB* sites (Figure 8A). This construction allows for selection of recombined plasmids, as a failure to recombine results in cell death. The plasmid of interest was then transformed into cells expressing Brujita Int and the numbers of transformed colonies were recorded and expressed as

recombination efficiency (cfu/μg transformed DNA).

2.3.5 Examination of *attP* and *attB* Requirements

We used our intra-molecular recombination system to probe the DNA sequence requirements of *attP*. We examined progressively smaller *attP* fragments and found that the right side of the *attP* core (bases 14 through 39 of the core) is not needed for integrative recombination (see Figure 9A). When we remove the 5 bases from positions +9 to +14 in the core, a distinct drop in efficiency is observed (~100 fold), suggesting that this region is required for efficient recombination. Similarly, when the region left of the core is truncated, the region from positions -35 to -30 shows a ~50-fold drop in efficiency, also indicating that this region is important for integrative recombination. A fragment consisting of the region from -35 to +14 yields recombination levels similar to wild type, demonstrating that this region is both necessary and sufficient for recombination *in vivo*.

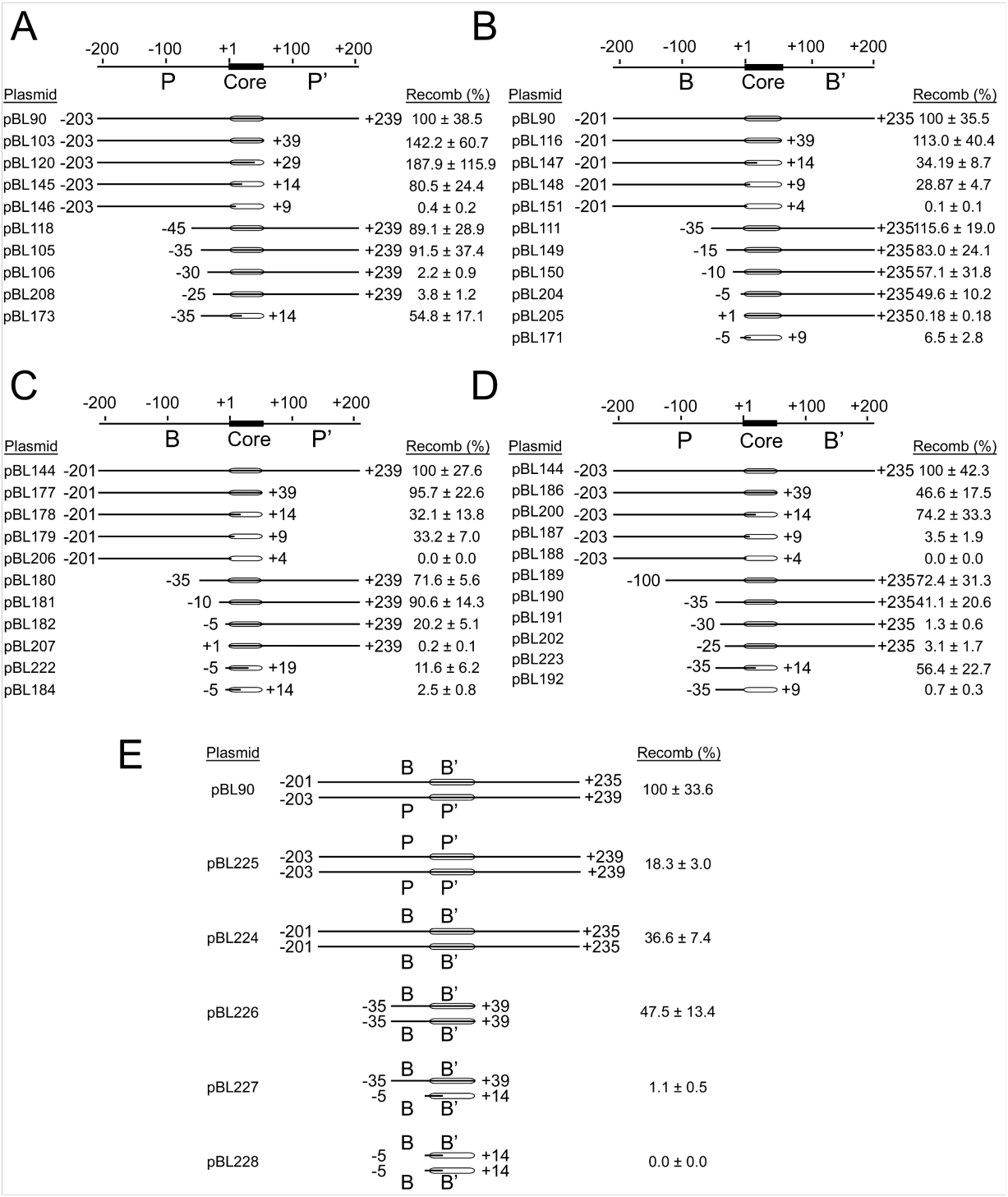


Figure 9: Length requirements of the Brujita attachment sites

Length requirements for *attP* (**A**), *attB* (**B**), *attL* (**C**), and *attR* (**D**). Plasmid pBL90 contains 442 bp *attP* DNA (phage coordinates 28,970 to 29,411) and a 436-bp *attB* site with the common cores centrally located. When transformed into E. coli cells expressing Brujita Int (pBL90 +) and *ccdB* is counter-selected, approximately 8×10^4 colonies per microgram (cfu/μg) DNA are recovered;

recombination is Int-dependent and few colonies are recovered in the absence of Int (pBL90 –). Plasmid derivatives of pBL90 were constructed in which segments of either *attP* or *attB* (A and B, respectively) were deleted from the left or right ends, as indicated by the deletion endpoint coordinates, and the numbers of recombinants recovered reported as a percent of control (shown at right). Derivatives of pBL144 (see Figure 8) were similarly tested for *attL* (C) and *attR* (D) length requirements. **E.** Int-mediated recombination of pBL224 (*attB* × *attB*) and pBL225 (*attP* × *attP*) and derivatives. Errors values represent the standard errors of three separate transformations.

Using the intra-molecular recombination assay, the sequence requirements of *attB* were similarly examined (Figure 8B). When the region from +4 to +9 is removed, recombination is eliminated (10,000-fold decrease, Figure 9B). When the region from -5 to 0 is removed, efficiency similarly drops ~100-fold. Surprisingly, when the fragment from -5 to +9 is tested, efficiency is lower than when either individual side is tested, implying that, though not required, the exterior plasmid sequences may be contributing to sequence recognition.

Since the *attP* site is comprised of P and P' half sites and the *attB* of B and B' half sites (Figure 9A-B) (Ross *et al.*, 1979), we expected that when testing the requirements for *attL* and *attR*, we would see a combination of the individual sequence requirements (*attL* being the composite of B and P' and *attR* being the composite of P and B'). However, this was not entirely the case. The *attR* half site P was similar to the *attP* requirement (a 10-fold drop from -35 to -30, Figure 9A, D), but the B' half site fragments contained in pBL187 (+9 segment) was 10-fold lower than the control, even though in *attB* this fragment was sufficient for recombination (Figure 9B, D).

The B half site of *att*L shows similar results to those found earlier (see Figure 9B-C), with ~100-fold decrease in efficiency after deleting the region -5 to 0, suggesting this region is necessary in both *att*B and *att*L. In the P' prime half-site of *att*L, deleting +9 to +14 does not change the efficiency of recombination, even though this deletion shows a 100-fold change in *att*P (Figures 9A, C). Also of interest is the minimal *att*L sequence (pBL184 fragment). Even though

the regions of -5 and +14 are both necessary, the fragment -5 to +14 is not sufficient for excisive recombination, suggesting that additional sequence may be needed for this reaction. Additionally, this assay is performed in a plasmid background, which may be contributing required sequences, making it difficult to identify precise boundaries. Regardless of these restrictions, it is clear that the Brujita attachment sites differ from those of the other tyrosine recombinases. It contains a small ~24 bp *attB* sequence, similar to the Lambda system, and a small, asymmetrical 49 bp *attP* sequence, much smaller than traditional systems, which require ~240 bp for proper arm-type interactions (Mizuuchi & Mizuuchi, 1985, Mizuuchi & Mizuuchi, 1980). The observation that sequence requirements for integration and excision are similar, if not the same (*attP* is the same as *attR* and *attB* is the same as *attL*), suggests that there is no means of directionality control and no recombination directionality factor (RDF) in the Brujita system, a feature that is characteristic of the tyrosine recombinases (Groth & Calos, 2004b).

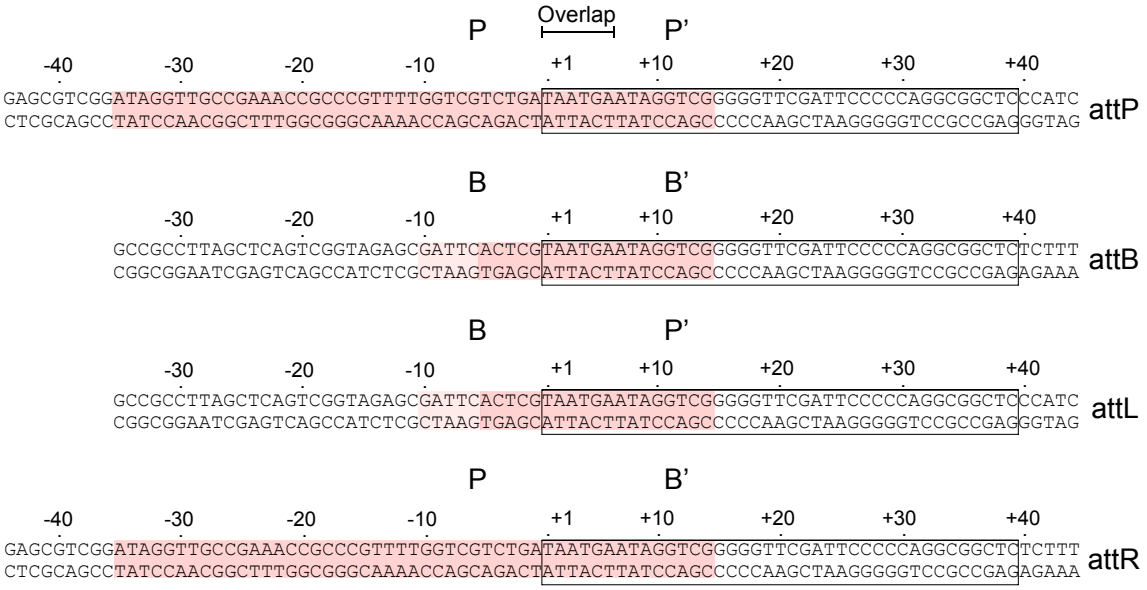


Figure 10: Summary of Brujita attachment sites

attP and *attB* share a 39-bp common core sequence (boxed) that overlaps a host $tRNA^{Thr}$ gene. Int-mediated strand exchange occurs within the common core to form *attL* and *attR* product sites. Because the recombination requirements do not extend beyond the common core to its right, the B′ and P′ sites are identical, and thus, *attP* and *attR* as well as *attB* and *attL* are functionally equivalent. The regions required for in vivo recombination are highlighted in red.

2.3.6 Brujita Site Selectivity

Because Brujita Int does not use arm-type interactions and there appear to be difference in the requirements between integrative and excisive recombination substrates, we sought to understand whether Int differentiates between *attP* and *attB* sequences. Plasmids containing either *attP* x *attP* or *attB* x *attB* (pBL225 and pBL224, respectively, Figure 8C-D, 9E) were constructed as before and tested for recombination ability. Surprisingly, both plasmids are competent for recombination, although at slightly lower levels than with *attP* x *attB* substrates. Plasmids containing smaller *attB* sequences also recombine, but this does not proceed with *attB* sequences are smaller than ~74 bp (pBL226, pBL227, and pBL228, Figure 9E). These results were verified to be Int-dependent, and not a result of homologous recombination. The ability of Int to recombine identical sites, as well as normal *attB* x *attP* recombination is unexpected. This suggests that Int is promiscuous and has little (if any) preference between DNA sites. This is especially surprising, considering that *attP* and *attB* have different boundaries and sequences.

2.3.7 Mapping the Brujita Crossover Positions

In Lambda, Int binds to *attP* and *attB* sites separated by a 7-base spacer between cleavage points (Grindley *et al.*, 2006, Groth & Calos, 2004a). Cre also exhibits similar binding using a 6-base

spacer (Van Duyne, 2001). In both cases, recombination results in cleavage at the 5' end of the spacer, creating overhangs that then participate in strand exchange (Van Duyne, 2001, Grindley *et al.*, 2006). Since the attachment sites are much smaller in the Brujita Int system than in Lambda, we questioned whether the positions of Int binding and strand exchange would also be different. To test this, we designed several mutations in the *attP* core of an integrating *M. smegmatis* plasmid. However, because this core replaces the 3' half of a functional $tRNA^{Thr}$ in *M. smegmatis*, only a few specific positions would give interpretable data, while producing a functional tRNA and allowing the integrant to survive. Mutations were made separately in the 3rd, 6th, and 9th bases of the core (A-T, A-G, and A-C respectively) and selected for integration of the plasmid. Successful integrants were recovered and sequenced in the *attL* and *attR* regions to identify the location of the mutant base. In all cases tested, the A-T mutation in the third position of the core segregated to *attR* (n=4) and the other two mutations segregated to *attL* (n=4, n=3), suggesting that strand exchange occurs between positions three and six of the *attP* core (data not shown).

Due to the limitations of performing this experiment in *M. smegmatis*, the plasmid-based counter selectable intra-molecular recombination system was adapted to use *attP* or *attB* with each of the first nine bases individually mutated (Figure 10). Integrative recombinants were selected, recorded, and the recovered colonies were sequenced as before to identify the final location of the mutated bases. Because this system results in the loss of *ccdB* and the *attR* containing DNA, mutants can only be recovered if they segregate to the *attL* site. Mutations in positions 1, 2, and 9 of the core reduce recombination approximately 100-fold and mutations in positions 3, 4, and 5 show ~10 fold reduction in recombination when mutated. Positions 6-8 show a slight change in recombination (2-fold). These observations are true whether *attP* or *attB* bases were mutated.

A

Plasmid	Mutation Position	*attB* +1 +9	*attP* +1 +9	Recomb^n (%)	Mutation in *attL*
pBL90	wt	TAATGAATA	TAATGAATA	100 ± 25%	N/A
pBL152	+1	aAATGAATA	TAATGAATA	0.5 ± 0.3%	1/5
pBL153	+2	TtATGAATA	TAATGAATA	0.4 ± 0.3%	2/5
pBL154	+3	TAtTGAATA	TAATGAATA	5.2 ± 2.5%	7/13
pBL155	+4	TAAaGAATA	TAATGAATA	17.7 ± 2.4%	2/13
pBL156	+5	TAATcAATA	TAATGAATA	11.0 ± 4.6%	2/10
pBL157	+6	TAATGtATA	TAATGAATA	43.9 ± 16.8%	2/17
pBL158	+7	TAATGAtTA	TAATGAATA	41.5 ± 11.6%	0/12
pBL159	+8	TAATGAAaA	TAATGAATA	66.0 ± 17.9%	0/11
pBL160	+9	TAATGAATt	TAATGAATA	1.3 ± 0.7%	0/9
pBL161	+1	TAATGAATA	aAATGAATA	0.1 ± 0.1%	0/4
pBL162	+2	TAATGAATA	TtATGAATA	1.2 ± 0.6%	3/7
pBL163	+3	TAATGAATA	TAtTGAATA	4.0 ± 1.0%	7/10
pBL164	+4	TAATGAATA	TAAaGAATA	43.5 ± 12.1%	9/10
pBL165	+5	TAATGAATA	TAATcAATA	26.7 ± 5.6%	9/10
pBL166	+6	TAATGAATA	TAATGtATA	122.2 ± 6.3%	4/10
pBL167	+7	TAATGAATA	TAATGAtTA	192.3 ± 31.9%	9/9
pBL168	+8	TAATGAATA	TAATGAAaA	96.7 ± 18.7%	10/10
pBL169	+9	TAATGAATA	TAATGAATt	3.4 ± 1.1%	8/8
pBL213	+1	aAATGAATA	aAATGAATA	0.9 ± 0.4%	
pBL214	+2	TtATGAATA	TtATGAATA	34.0 ± 12.7%	
pBL215	+3	TAtTGAATA	TAtTGAATA	50.1 ± 8.4%	
pBL216	+4	TAAaGAATA	TAAaGAATA	29.8 ± 10.3%	
pBL217	+5	TAATcAATA	TAATcAATA	0.6 ± 0.3%	
pBL218	+6	TAATGtATA	TAATGtATA	37.4 ± 28.8%	
pBL219	+7	TAATGAtTA	TAATGAtTA	27.3 ± 12.5%	
pBL220	+8	TAATGAAaA	TAATGAAaA	13.1 ± 6.9%	
pBL221	+9	TAATGAATt	TAATGAATt	0.2 ± 0.2%	

B

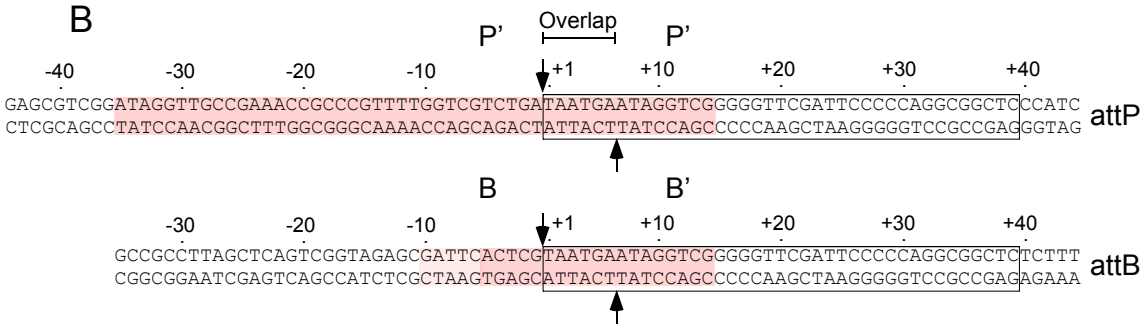


Figure 11: Mutational analysis of the putative overlap region in *attP* and *attB*

(**A**) Derivatives of plasmid pBL90 were constructed with single-base substitutions between coordinates + 1 and + 9 in *attP*, *attB*, or both *attP* and *attB* as indicated, and tested for recombination. Positions of the base substitutions are shown, together with the numbers of recombinant products in which the substitution is present in the *attL* product. (**B**) Diagram of *attP* and *attB* sequences showing positions of cleavage (arrows). Minimal sequences required for *in vivo* recombination are highlighted in red. The boxed region represents the 39 bp core.

In the Lambda system, mutating nucleotides near the positions of strand exchange can be very informative. If a base is located in a binding site, the mutation will interfere with recombination and lower its efficiency (Campbell, 1992, Rajeev *et al.*, 2009). A similar result will occur of the mutation is located in the crossover, or spacer region. However, when compensating mutations are made in the corresponding attachment site, it becomes clear whether the mutated base is located in a binding site or the spacer. If the position is in the spacer, the compensating mutation will restore the recombination defect, while no such change will occur for a position involved in binding (Furth *et al.*, 1983, Rajeev *et al.*, 2009).

With this in mind, we made compensating mutations in the *attP* core and tested recombination efficiencies (Figure 11, pBL213-221). Positions 1 and 9 are not restored by the compensating mutation, suggesting that these positions are involved in binding to Brujita Int. When the compensating mutation is made in positions 2 and 3, recombination is largely restored, implying that these are located in the spacer region. However, no other locations exhibit similar behaviors to this position, which is unexpected. Also unexpected is the behavior of position 5 (pBL156 and pBL217). When one attachment site contains the mutation, recombination is decreased by ~10 fold. However, the compensating mutation causes a 100-fold decrease in recombination, which implies that this position has some sort of binding role and that mutation here has an additive effect. Unfortunately, these data do not present a clear picture of what is occurring in these positions. Further work will need to be done in order to clarify these questions.

We recovered and sequenced transformants from each of the above mutants to track the final location of mutations in each position (Figure 11). We found that mutations in *attB* positions 7, 8, and 9 are never found in *attL* and are likely to segregate to *attR*, meaning that they are to the

right of the cleavage point. Mutants in positions 1-6 are found at intermediate frequencies in *attL*, suggesting that these are between cleavage positions. Sequence data from *attP* agrees, with the sole exception of position 1, in which no mutants are found in *attL*. However, very few transformants could be recovered from parents with one mutant in this position, so it is likely that this is due to small sample size. These data are consistent with a model in which the positions of cleavage are to the left of core position 1 and to the right of core position 6, which would create 6-base overhangs (Figure 11).

2.3.8 *In Vitro* Recombination

In the tyrosine recombinase family of proteins, much has been learned about recombination mechanisms, cleavage positions, and *cis*- versus *trans*-cleavage using *in vitro* recombination assays, as they allow a careful dissection of the requirements and features of the system (Arciszewska *et al.*, 1995, Arciszewska & Sherratt, 1995, Lee & Hatfull, 1993, Nash, 1975, Ghosh *et al.*, 2007, Gibb *et al.*, 2010). One goal in exploring the Brujita system was to develop a similar assay for Brujita Int.

To identify conditions likely to facilitate *in vitro* recombination, we began experimenting with various conditions used for Lambda Int, XerC/D, and L5 Int *in vitro* recombination (Arciszewska *et al.*, 1995, Arciszewska & Sherratt, 1995, Colloms *et al.*, 1996, Lee & Hatfull, 1993, Nash, 1975, Syvanen, 1974). Using these various buffer conditions, we tested whether Int, *attP,* and *attB* were sufficient to cause recombination *in vitro.* Since we have had considerable success with the intra-molecular plasmid based *in vivo* recombination system, discussed in Chapter 2, we also used this substrate *in vitro.* To increase the sensitivity of this assay, recombination was

tested by transforming our *in vitro* reactions into *E. coli* and selecting for the desired recombinant molecules.

Assayed by gel electrophoresis, the conditions described in section 2.2.6 yielded recombinant products from intra-molecular recombination of the plasmid with pBL90 and pBL224 (Figure 12). In this gel, as the concentration of Int increases, there is a decrease of supercoiled plasmid, accompanied with an appearance of two product bands. Samples of the product were transformed into Neb5α cells to test for the loss of the *ccdB* marker, which will allow transformed cells to grow and signifies successful recombination. The results of this experiment are shown in Table 2. Together, these results confirm that we are observing Int-mediated plasmid recombination.

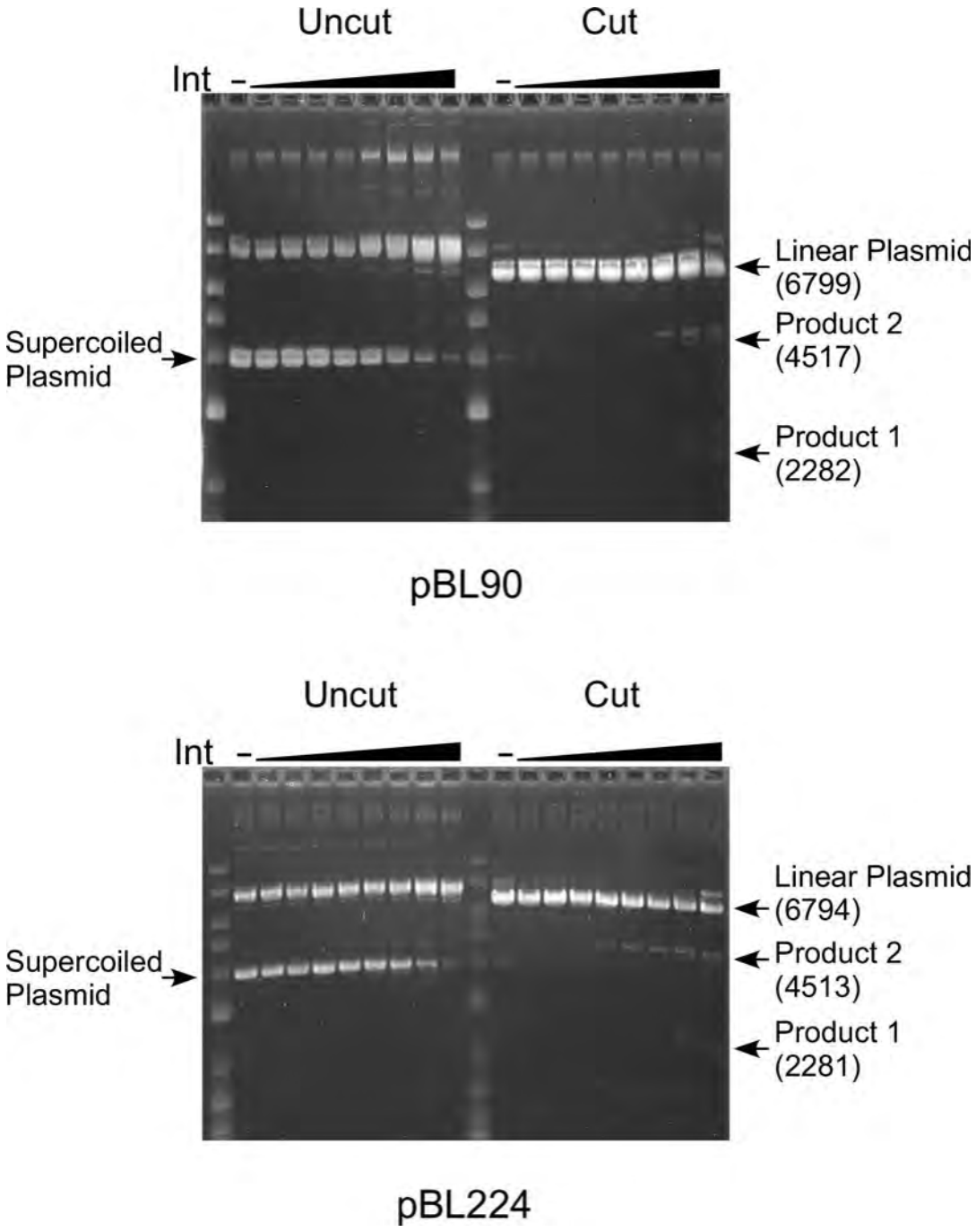


Figure 12: *In vitro* recombination using plasmid substrates

pBL90 (top) and pBL224 (bottom) were used as substrates for *in vitro* recombination under conditions described in section 2.2.6. Recombination products were digested with proteinase K and SDS and run on 0.6% agarose gel, either uncut, or linearized by restriction digestion. The sizes of parent plasmid and product are denoted.

Table 2. *In vitro* recombination frequencies using pBL90

	Int Concentration						
	0	0.61 nM	2.4 nM	9.8 nM	39.1 nM	156.3 nM	625 nM
Colonies Recovered	2	2	3	7	48	206	131

An interesting observation from these reactions is that though there are bands corresponding to the proper size for recombination products, additional bands of similar size are also observed, resulting in double bands for each product. The parental plasmid was linearized with NotI, and this restriction site is present in only one of the products, resulting in a linear 4517 bp product and a circular 2282 bp product that may be supercoiled. Alternatively, these DNAs may result from some aberrant recombination, yielding non-standard products. Thus, these additional bands and any implications they have will require further study in an optimized system.

Although successful *in vitro* recombination was achieved on a few occasions, reproducibility and optimization have proven difficult. Preliminary assays showed that Int is most active at 37°C and 42°C *in vitro*. Shortly after these preliminary experiments, *in vitro* recombination reactions ceased to function. Several batches of Int were expressed and purified as well as many preparations of plasmid DNA, linear oligos, and various combinations of *attB* and *attP* sequences were attempted, to no avail. The source of the DNA substrates, however, appear to be important. Many different preparations were made using either mini-prep kits (ThermoFisher) or phenol-chloroform extraction (Ausubel *et al.*, 1996). Phenol-extracted plasmid preparations in general worked poorly relative to mini-prepped plasmid DNA.

Although an optimized *in vitro* system could not be developed, several important pieces of information can be gleaned from the successful reactions. First, it appears that Int only converts supercoiled plasmid DNA into products, suggesting that there is a requirement for supercoiled DNA. However, it is possible that only supercoiled plasmid orients the two attachment sites in close proximity to form a synaptic complex, and integration in a phage-host scenario would have no such issues. Second, these reactions do not require a cofactor, such as mIHF, as these reactions only include Int and DNA. However, this does not preclude the possibility that a cofactor may stimulate the reaction *in vivo*. We note that *in vivo* recombination in Brujita occurs at significantly lower rates than in Lambda Int or L5 Int under normal conditions (Broussard *et al.*, 2013), suggesting that there is an inherently lower efficiency with Brujita Int than the other integrases.

We note that recombination occurs more efficiently *in vitro* using *attB* x *attB* (pBL224) than *attB* x *attP* (pBL90) substrates (Figure 12). This is unexpected, as the opposite trend was observed *in vivo* (Figure 9E).

2.4 DISCUSSION

We have shown that Brujita Int uses an *attB* site of approximately 24 bp, similar to the requirements for L5 Int and Lambda Int (Mizuuchi & Mizuuchi, 1980, Pena *et al.*, 1996). However, we demonstrated that the *attP* sequence is unusual, in that it only requires 49 bp of DNA, much less than the 240 bp requirement for Lambda Int and L5 Int (Pena *et al.*, 1997b, Mizuuchi & Mizuuchi, 1980). Brujita does not use arm-type interactions and we provide evidence that directionality is instead controlled by manipulating protein expression and stability (Broussard *et al.*, 2013). Additionally, *attB* has the same DNA sequence as *attL and attP* is the same as *attR*,

meaning that there is no method of directionality control and that integration and excision are identical events at the molecular level.

Brujita Int is promiscuous in its ability to recombine with any combination of attachment sites and has little (if any) preference between DNA sites. This is especially surprising, considering that *attP* and *attB* are different sequences and have different length requirements. This raises the question as to why this system uses two non-identical sites (*attP* and *attB*), when recombination can occur in a simpler manner using identical sites, presumably easier to maintain. It appears that maintaining the two different sites confers some selective advantage that we are not aware of, such as the ability to modulate activity.

We have also presented evidence that Brujita Int cleaves DNA to create 6 bp overhangs. This spacing is uncommon for tyrosine integrases, which generally create 7 bp overhangs (Groth & Calos, 2004b, Pena *et al.*, 1996), but typical for recombinases (Cre and XerC/D create 6 bp spacers) (Shaikh & Sadowski, 1997, Blake *et al.*, 1997). It is not clear whether the space between cleavage points has important implications for other aspects of recombination.

The *in vitro* reactions shown here verify that Brujita Int mediated site-specific recombination occurs, though at very low efficiency. This agrees with previous data that Brujita Int is 10,000-fold less effective *in vivo* than its L5 Int counterpart, and even 100-fold less efficient than L5 Int when stabilized *in vivo* (Broussard *et al.*, 2013). Thus, it appears that the degradation tag is not the only reason for lower levels of recombination in Brujita, but that the protein itself is biochemically less efficient than other systems described. With this in mind, our *in vitro* results may represent optimal activity for this protein. The ability to achieve *in vitro* recombination, even inefficiently, shows that there are no essential cofactor requirements such as mIHF. Additionally, recombination appears to work better *in vitro* using *attB* x *attB* (pBL224) substrates than *attB* x

attP (pBL90) substrates. It is not understood why this is the case, when *in vivo* recombination shows the opposite trend. Further experiments are needed to answer this question.

In conclusion, Brujita Int contains several features of the simple tyrosine recombinases, while the only features it seems to contain from the integrases is the ability to recombine between dissimilar phage and host DNA sites. The control of directionality also proceeds through a novel mechanism. This system, therefore, represents a simple tyrosine recombinase masquerading as an integrase by using a degradation tag and non-identical attachment sites.

3.0 REQUIREMENTS FOR BRUJITA INT-DNA BINDING

Work discussed in this chapter was published in the following article:

Lunt, B.L., and Hatfull, G.F., January 2016. Brujita Integrase: A Simple, Arm-Less, Directionless, and Promiscuous Tyrosine Integrase System. *J Mol Biol*, Volume 428, Issue 11, Pages 2289-2306.

The experiments discussed below were performed by the author. Figures 13-15, 17-21, and 23-25 were adapted from Lunt and Hatfull, 2016 and were made in collaboration with Graham Hatfull.

3.1 INTRODUCTION

The DNA requirements for Brujita recombination *in vivo* reveal that Brujita Int requires small DNA sites: *attB* and *attP* are ~24 and ~49 bp respectively. However, how Brujita Int binds DNA to form the synaptic complex is unknown. This chapter will discuss the experiments performed and the data that examines how Brujita Int binds DNA as well as the nucleotides required for these interactions.

3.2 MATERIALS AND METHODS

3.2.1 Annealing of Oligos

Oligos were prepared for binding assays by resuspension to 100 μM in TE. 10 μl of each single stranded oligo were combined and incubated at 95°C for 5 minutes, then cooled gradually to room temperature. Oligos were then radiolabeled as described below.

3.2.2 Electrophoretic Mobility-Shift Assays

EMSAs were performed by established protocols (Ausubel *et al.*, 1996, Villanueva *et al.*, 2015). DNA oligonucleotides were 5' radiolabeled with γ-^{32}P ATP and T4 polynucleotide kinase (Roche). Non-oligo DNAs (*attP*292 and *attB*345) were PCR amplified and gel purified prior to radiolabelling as above. Binding reactions were prepared by adding 5 ng radiolabeled probe, 1 μg calf-thymus DNA, binding buffer (20 mM Tris pH 8.0, 10 mM EDTA, 25 mM NaCl, 10 mM Spermidine, 1 mM DTT) and 1 μl of varying concentrations of Int in a final volume of 10 μl. After 30 minutes incubation at room temperature, the binding reactions were run on a 5% native TBE polyacrylamide gel at 250 V for 2-3 hours. Detection was performed using autoradiography and phosphorimaging.

3.2.3 Competition with Unradiolabeled DNA

Radiolabeling and binding reactions were performed as above with 125 nM Int and 5 ng radiolabeled DNA incubated for 30 minutes at room temperature. Following incubation,

unradiolabeled *attB* was added in increasing concentrations to binding reactions containing radiolabeled *attP* and incubated again for 30 minutes at room temperature. Samples were run on 5% native polyacrylamide gels for 2 hours at 250 V and imaged as before. The same procedure was used with radiolabeled *attB* and unradiolabeled *attP* oligos.

3.2.4 Generation of DNA Spacing Mutants

attP spacer mutants described in this chapter were generated using site-directed mutagenesis with pBL90 as a template. Mutant plasmids were Sanger-sequenced for verification. Plasmids were transformed into Int-expressing Stellar *E. coli* cells and transformed colonies were recovered as described in Chapter 2.

3.2.5 Binding Assays for Ferguson Plots

Native Tris acrylamide gels were prepared in 7.5, 10, 12.5, and 15% concentrations (29:1 acrylamide:bisacrylamide) for Trials 1 and 2. Trial 3 used gels of 10, 12.5, 15, and 20% low cross-linking acrylamide (100:1 acrylamide:bisacrylamide). Binding assays were prepared as described above and in (Lunt & Hatfull, 2016). Briefly, 5 ng *attB44* DNA (18 nM final concentration) was added to binding buffer with or without calf-thymus DNA and incubated with 0, 39, or 156 nM Int for 30 minutes. The mixtures were run on each gel until the loading dye band migrated to the end of the gel (4-5 hours at 250 V). The gels were fixed and silver stained using a Silver Stain Plus kit (BioRad).

3.2.6 Ferguson Plots

Gels were imaged and distances of migrating complexes and standards were measured using ImageJ (Schneider *et al.*, 2012, Abràmoff *et al.*, 2004, Rasband, 1997). Relative mobilities and standard curves were calculated and plotted according to established protocols (Ausubel *et al.*, 1996).

Table 3. Plasmids used in Chapter 3.

Plasmid Name	Description	Reference
pBL90	pBL50 with ccdB (from pDONRZeo) cloned into DraI site	Lunt and Hatfull, 2016
pBL258	pBL90 with SDM 5'AG-TC in B'	Lunt and Hatfull, 2016
pBL259	pBL90 with SDM 5'AG-TC in B (reverse complement)	Lunt and Hatfull, 2016
pBL260	pBL90 with SDM 5'AG-TC in P'	Lunt and Hatfull, 2016
pBL261	pBL90 with SDM 5'AG-TC in P (reverse complement)	Lunt and Hatfull, 2016
pBL262	pBL90 with SDM 5'AG-TC in P1	Lunt and Hatfull, 2016

3.3 RESULTS

3.3.1 Binding of Brujita Int to *attB* and *attP* DNA

To identify the DNA sequence specificity of Brujita integration, we first needed to develop DNA binding assays. We resorted to using Electrophoretic Mobility Shift Assays (EMSAs), where increasing concentrations of Int are incubated with radiolabeled *attP* or *attB* DNA. Binding events are observed as a shift to form a slower migrating band, representing a DNA-protein complex.

Untagged Brujita Int^{Long} and Int^{Short} were expressed and purified using a nickel affinity column, as Int is capable of binding to this resin in its native form. Binding reactions were prepared using 292 bp *attP*, or 345 bp *attB* PCR amplified and purified sequences. These sequences contain the core DNA positioned in the center. We found that binding occurs similarly in both proteins with both *attP* and *attB* (Figure 13). Binding affinity however, is reduced slightly in Int^{Short} (approximately 5-fold). Interestingly, *attP* forms two complexes with Int, while *attB* forms three complexes.

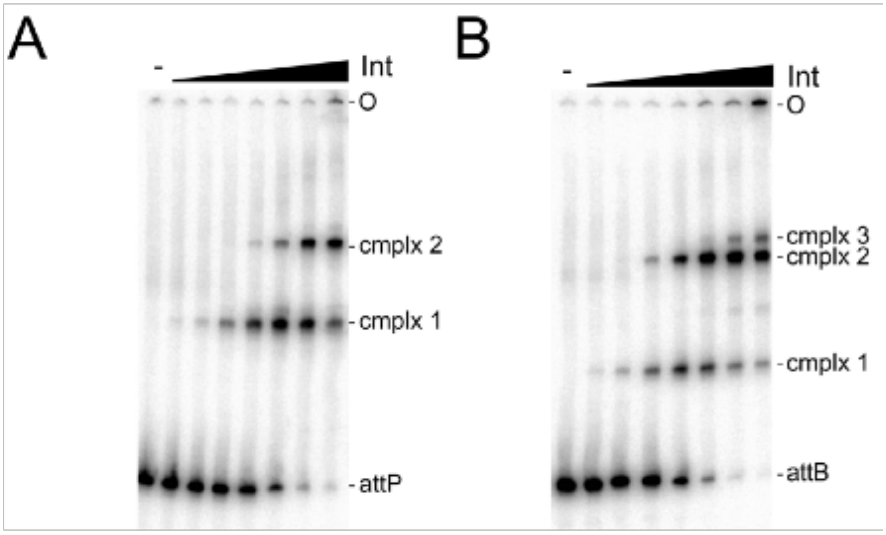


Figure 13: Binding of Brujita Int to *attP* and *attB* DNA substrates

A. Binding of Brujita Int to *attP* DNA. The positions of the origin (O), free DNA (*attP*), and complexes (cmplx1, cmplx2) are indicated. Int concentrations are, from left to right, 0, 0.15, 0.61, 2.4, 9.8, 39, 156, and 625 nM. **B.** Binding of Brujita Int to *attB* DNA. The positions of free DNA (*attB*) and complexes (cmplx1, cmplx2, and cmplx3) are indicated. Int concentrations are the same as in A.

3.3.2 Competition with Unradiolabeled DNA

We questioned whether the Int-DNA binding reaction is transient – in which Int binds and releases DNA rapidly – or a more permanent interaction. To answer this question, we performed competition experiments, where radiolabeled DNA is pre-bound to Int, non-radiolabeled DNA is added in increasing concentrations for 30 minutes, and the mixture is run on a gel. We observe that when radiolabeled *attB* is competed with cold *attP*, the concentration of *attP* required to displace half of the Int is 14 nM, similar to the K_d of the binding reaction (see Figure 14). Likewise, 25 nM cold *attB* is required to compete off half of the radiolabeled *attP* probe. This data suggests that the binding reaction is not permanent, as it appears Int is able to release and bind probe to equilibrium within the 30-minute timeframe of this experiment.

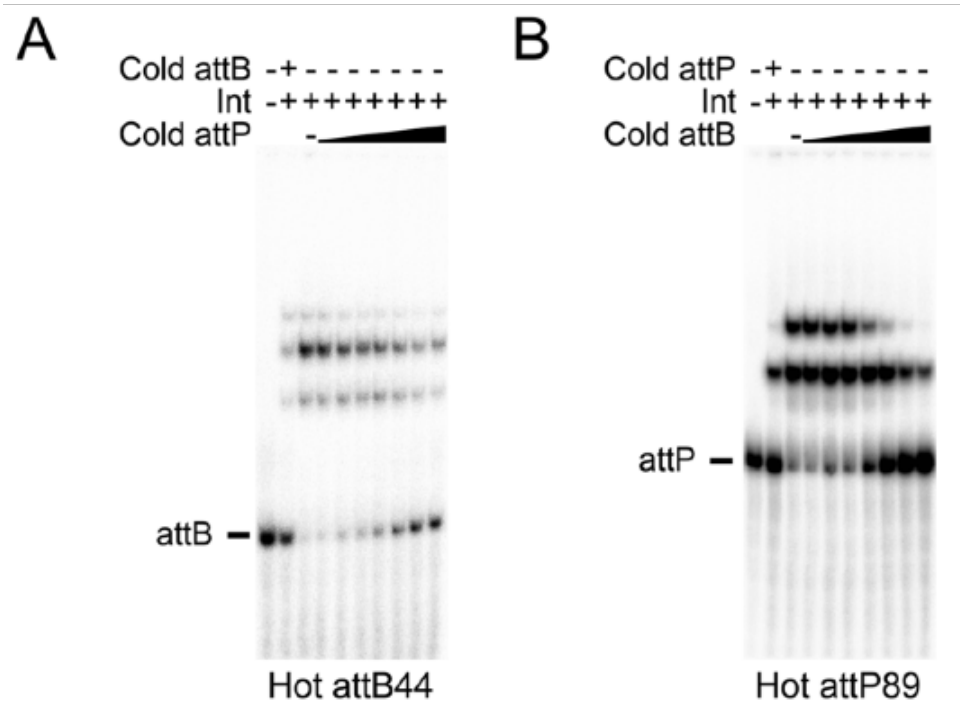


Figure 14: Competition experiments between hot and cold DNA

A. Competition experiments between ^{32}P radiolabeled *attB* and cold *attP* with detection of radiolabeled probe by phosphorimaging. Concentrations of cold *attP* are, from left to right: 0, 0, 0, 5, 9, 18, 36, 73, 145, and 290 nM. Int concentration is 53 nM and hot and cold *attB* are 18 nM. **B.** Competition experiments between radiolabeled *attP* and cold *attB*. Concentrations of cold *attB* are, from left to right: 0, 0, 0, 7, 14, 28, 55, 110, 220, and 440 nM. Int concentration is 53 nM and hot and cold *attP* are 9 nM.

3.3.3 Int-DNA Stoichiometries

When free in solution, size exclusion analysis shows that Brujita Int elutes at a size consistent with monomeric protein (data not shown). The complexes observed in binding assays could consist of DNA bound by either monomers or dimers of Int to multiple sites. To identify the molecular nature of complex-1 (fastest migrating) and complex-2 (next fastest migrating) in the binding assays, we bound *attP* and *attB* DNA with native Int, MBP-tagged Int, and a mixture of both proteins (Figure 15A-B). If there is more than one binding site on each DNA, we expect to see a hybrid band formed when the two proteins are mixed. Indeed, such a hybrid is seen with both *attP* and *attB* DNA (Figure 15A-B). A faint band of the same size is also found in the MBP-Int alone sample, due to the tendency of MBP-fusion proteins to self-cleave (Feher *et al.*, 2004, Arnau *et al.*, 2006). However, *attP* appears to only form a single hybrid, indicating that it contains only two binding sites, yet *attB* shows two or more hybrid bands. The nature of the slower migrating hybrid formed with *attB* is unknown, but it is clear that *attB* also contains at least two Int binding sites. Therefore, these data suggest that complex-1 in each DNA consists of a monomer of Int bound to the input DNA and complex-2 represents a second Int monomer occupying an additional site.

Additionally, because Int can recombine using two *attB* or two *attP* sequences, it is possible that recombination is occurring in these assays. However, these products would be identical to reactants and would therefore be unidentifiable. To address this, we performed binding

experiments using different sized DNAs. We observed that there are no hybrids formed when *attP* sequences of varying sizes are mixed, suggesting that only a single DNA fragment is present in each band. However, when *attB* DNA is used, a single hybrid is formed, larger than any other complex observed. The nature of this complex is not known, nor why this hybrid runs slower than any other complex. It is possible that *attB* complex-3 represents two molecules of DNA – each with two Int proteins bound – involved in interactions resembling a synaptic complex, but this does not explain the apparent size of the observed hybrid. We do not observe the formation of recombination products with either *attP* or *attB* under these conditions.

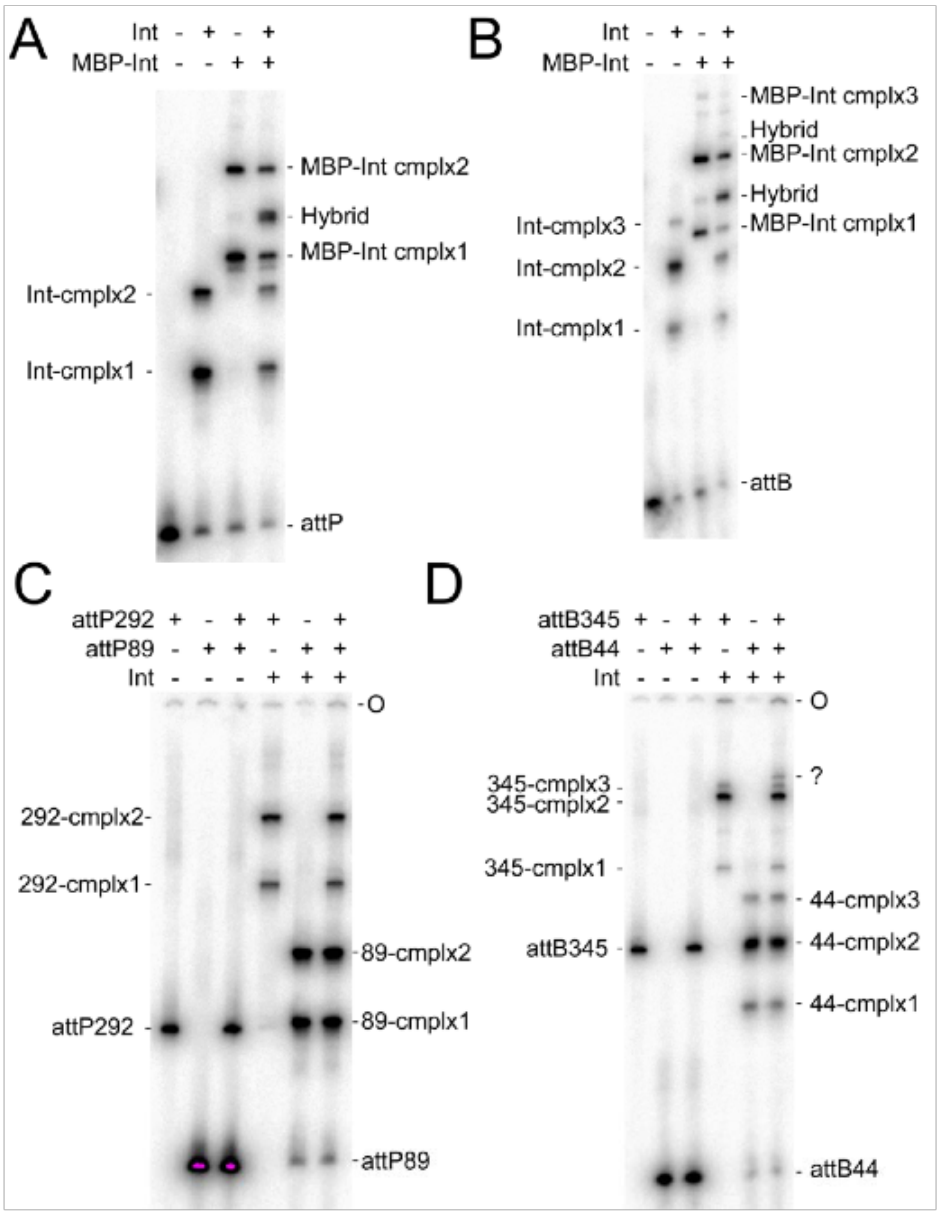


Figure 15: Hybridization assays

A. Binding of Int and MBP–Int (156 nM and 137 nM, respectively) separately and in combination (as indicated) to ^{32}P radiolabeled *attP* DNA. The positions of complexes and hybrid complexes containing both types of Int are indicated. **B.** Binding of Int and MBP–Int (156 nM and 137 nM, respectively) separately and in combination (as indicated) to radiolabeled *attB* DNA. The positions of complexes and hybrid complexes containing both types of Int are indicated. **C.** Binding of Int to two differently sized *attP* DNA substrates (89 bp and 292 bp, *attP89* and *attP292*, respectively; as indicated), separately, and in combination. **D.** Binding of Int to two differently sized *attB* DNA substrates (44 bp and 345 bp; *attB44* and *attB345*, respectively; as indicated). The nature of the lowest migrating complex with both DNAs (marked-?) is unclear.

3.3.4 Estimating Sizes of Int-DNA Complexes

The Int-DNA binding assays described herein have been helpful in elucidating the nature of Int-DNA interactions. However, the question remains of what the molecular components of each band are. The hybridization assays demonstrate that complex-1 is comprised of a DNA molecule bound by one Int protein and complex-2 contains a second Int protein bound at a second site. However, the nature of complex-3 remains unclear. Complex-3 may comprise two DNAs bound with Int proteins, or alternatively may contain up to four Int proteins in a structure resembling a synaptic complex. Although these two possibilities predict different molecular weights for the complex, the non-denaturing gels used for the binding assays are not directly useful for measuring molecular weights.

To distinguish between these possibilities, DNA binding assays were combined with Ferguson plots, where unknown DNA-protein complexes are run on varying concentrations of native poly-acrylamide and the relative mobilities of the complexes are compared to standard proteins (Ausubel *et al.*, 1996). Under these conditions, larger proteins are more strongly affected by gel concentration. By generating standard curves, the actual molecular weights of unknown complexes can be estimated.

The Ferguson plots are shown in Figure 16 and a summary of the calculated and actual sizes of the proteins is shown in Table 4. Unfortunately, some complexes were inconsistently visible in the four gels and mobilities could not be calculated for all complexes in each experiment. Calculated complex sizes are most accurate from ~20-150 kDa, which includes complexes-1 and 2, but not complex-3. The precision of these measurements was poor, resulting in two of the three experiments generally yielding reasonable values, with a third containing an outlier for each complex. Oddly, these outliers are scattered through all three experiments, such that there is no

one data set that is superior to the others. Trial three seems to provide the best fit for the native standards, yet the calculated Int-DNA complex sizes contain a large percentage of outliers from reasonable or known sizes.

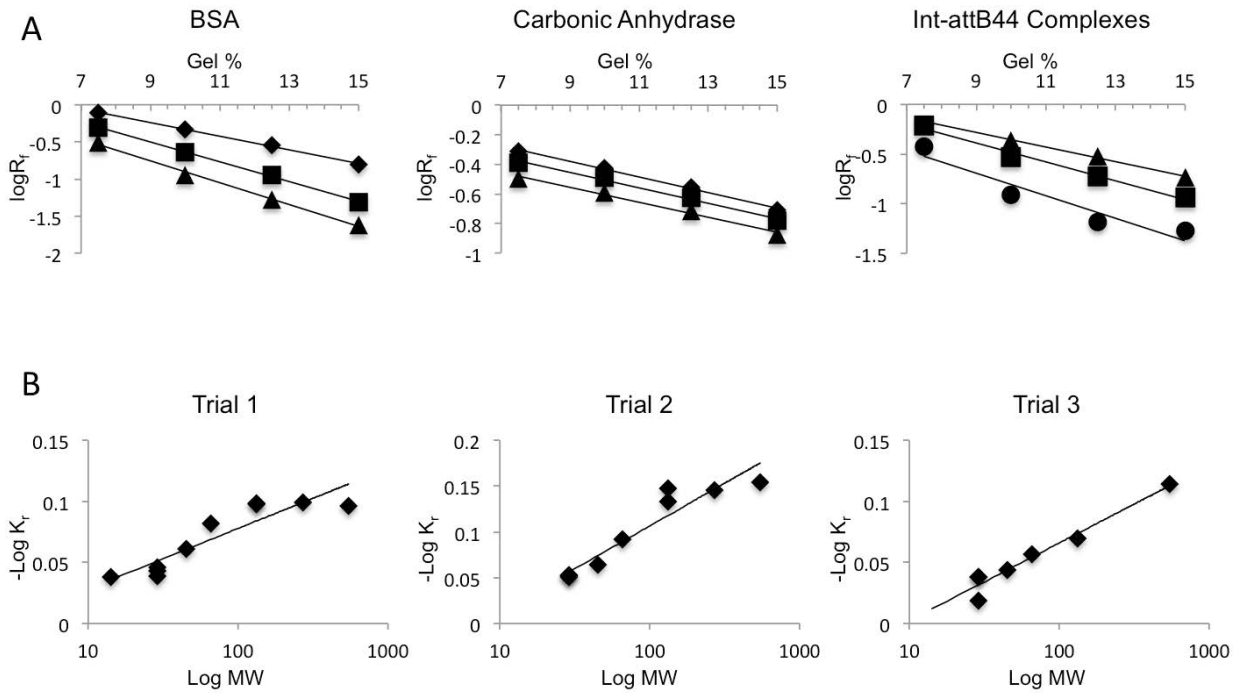


Figure 16: Mobilities of protein standards and Int complexes

A. Representative plots of the -logR_f values, from left to right: BSA, Carbonic Anhydrase, and Brujita Integrase-DNA complexes. The BSA plot shows relative mobilities of the monomer (diamonds) and two isoforms of a dimer (squares and triangles). The two complexes with parallel slopes are the same size, but have different mobilities due to charge isomerization, while those with different slopes are different sizes. The Carbonic Anhydrase plot shows three charge isomers with the same mass and slope. **B.** Standard curves are attained by plotting the -logK_r (slopes of the lines in A) against the log molecular weight of known standards. Unknowns can be calculated using the equation of the trend line. Shown here are standard curves from three trials.

Table 4. Calculated sizes of native protein complexes

Top. Three trials were run using the listed standards. Actual sizes and calculated sizes based on the standard curve are shown. **Bottom.** Known complex sizes are shown in black in the Actual Size column, with estimates, based on the proposed complex shown in blue. Red signifies calculated values that are outliers. Complexes-1, -2, and -3 for each experiment are indicated by "c1", "c2", and "c3".

	DNAs	Ints	Actual Size	Trial 1	Trial 2	Trial 3
α-Lactalbumin			14.2	18.7		
Ovalbumin			45	45.2	36.8	44.4
Carbonic Anhydrase Isoform 1			29	25.4	27.6	18.0
Carbonic Anhydrase Isoform 2			29	19.6	27.2	36.3
Carbonic Anhydrase Isoform 3			29	23.1	26.1	35.9
BSA Monomer			66	101.0	75.5	70.2
BSA Dimer			132	187.6	219.6	113.6
Urease Trimer			272	198.8	299.1	
Urease Hexamer			545	175.8	379.2	555.6
Int-attB37c1	1	1	58.8			65.7
Int-attB37c2	1	2	94.9			85.8
	2	2	117.6			
Int-attB37c3	2	3	153.7			81.0
	2	4	189.8			
Int-attB44c1	1	1	63.2	74.0	62.1	20.8
Int-attB44c2	1	2	99.3	169.8	105.6	111.2
	2	2	126.4	341.9	319.1	
Int-attB44c3	2	3	162.5			41.8
	2	4	198.6			
MBPInt-attB44c1	1	1	106.7			100.2
MBPInt-attB44c2	1	2	186.3			194.2
	2	2	213.4			
MBPInt-attB44c3	2	3	293			1478.8
	2	4	372.6			

It is clear from previous experiments that complex-1 consists of an Int monomer bound to a single DNA molecule, while complex-2 contains two monomers bound to a single DNA. Thus we can use these complexes to validate the molecular weight standard curves. Two of our trials (1

and 2) show a similar trend of complex-3 consisting of 2-3 times the mass of complex-2. However, the calculated mass for complex-3 is much larger than we would expect if this complex does indeed consist of four Int monomers and two DNA molecules. Unfortunately, the data does not appear to be consistent enough to make any reliable conclusions regarding this complex.

3.3.5 Requirements for Int-*attB* Binding

To identify the binding sites and DNA requirements for Int-*attB* binding, we first analyzed deletion and substitution derivatives of *attB*. Binding patterns are shown in Figure 17 and a summary of the effects of mutations is shown in Figures 18-19. First, we showed, by deletion analysis, that all of the requirements for binding are contained within approximately 37 bp of *attB* sequence (coordinates -14 to +23); when *attB37* is truncated by two nucleotides on either side (*attB33*), binding is reduced significantly. Next, we sought to identify the requirements for each binding half site. Truncation from the left side show that complex-2 is lost (*attB28*), but not complex-1 and complex-3. Further reduction shows that complex-1 is formed well until *attBR3*, which consists of positions +4 to +27, suggesting that the left boundary of the B' half site is located in this region. Truncation from the right side (*attBL1* through *attBL9*) results in no binding whatsoever, suggesting that binding at B is dependent on binding at B', this is supported by observation that the sequences of B and B' share no recognizable similarity, implying very different affinities for each site. Additionally, nearly full-length sequences are required to allow any binding from right-side truncations (*attBL10* through *attBL13*), confirming that binding is dependent upon B' occupancy, presumably using cooperative interaction between Int subunits. This data suggests that the boundary for B' binding is between coordinates +4 and +21. However, since we cannot isolate B half site binding from B', these boundaries are not clear. It does appear that sequence specificity

requirements extend into the first and last 4 bases of *attB37*, as replacing these nucleotides with corresponding positions from pBL90 based plasmid assays does impact binding (*attB37NS4* and *attB38NS5*).

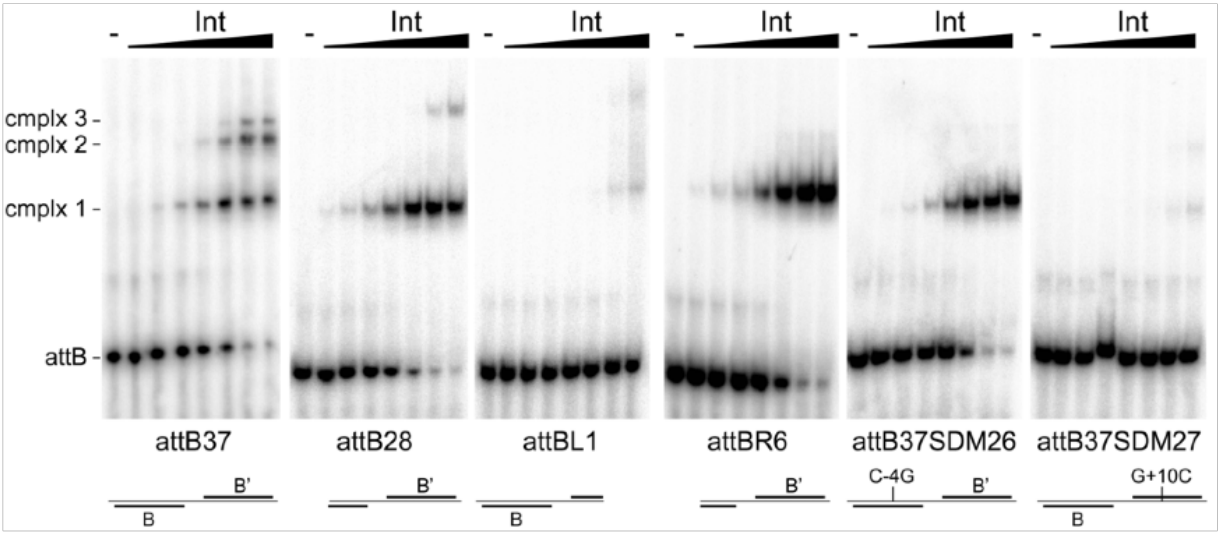


Figure 17: Int binding to mutant derivatives of *attB*

Binding of Int to representative *attB* DNA substrates. Int concentrations are, from left to right, 0, 0.15, 0.61, 2.4, 9.8, 39, 156, and 625 nM. See Figure 18 for substrate sequences.

We next made a series of transversion mutations in each position in *attB* to identify specific binding site sequences (Figure 19). We first tested binding to B' using variations of *attBR6*. Substitutions in most positions either do not affect binding, or have little effect. However, transversions at positions +9 and +10 eliminate binding entirely (*attBR6SDM8* and *attBR6SDM9*), indicating that these two nucleotides are critical for proper Int binding.

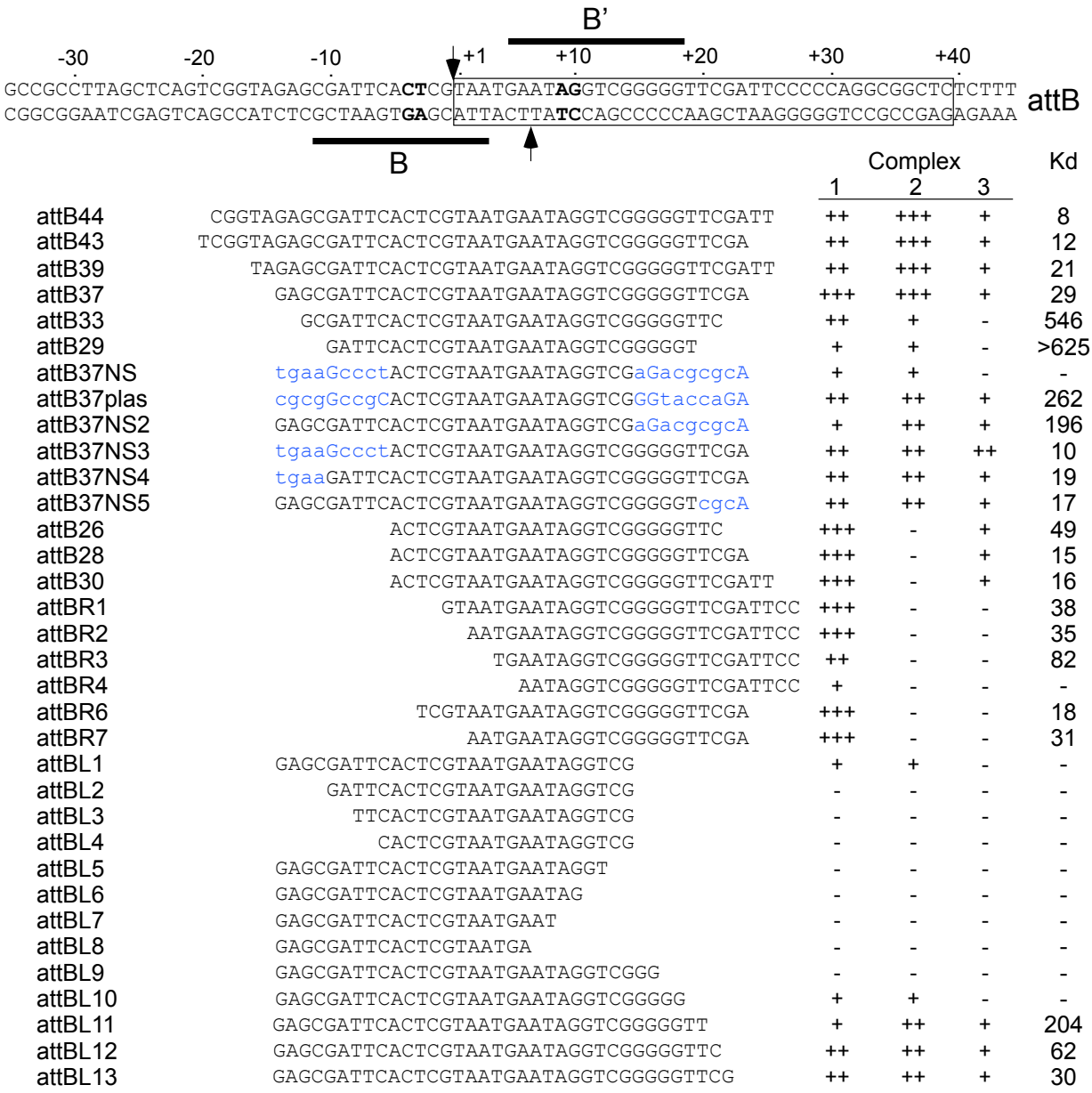


Figure 18: Summary of Brujita Int binding to *attB* substrates.

attB DNA substrates of varying lengths were tested for ability to bind Int. The intensities of complexes 1, 2, and 3 (as shown in Figures 15 and 17) are scored from weakest (–) to strongest (+++), with the +++ designation corresponding to a similar intensity to input DNA (see Figure 15 and 17). Overall binding affinities (K_ds) are shown, calculated as the Int concentration at which 50% of the input DNA is bound, regardless of the complexes formed. Base positions in upper case black type correspond to wild-type *attB* sequences, and those in blue lower-case type are non-native *attB* sequences corresponding to either plasmid-derived sequences (substrates *attB37plas*) or random sequence. Selected gels are shown in Figure 17. The box encompassing positions +1-+39 indicates the core region.

Next, we made mutations in the larger *attB37* DNA. Binding patterns were similar for those positions tested in *attBR6*. When B half site positions were examined, we found that these mutations have no effect on complex-1 formation, as expected, yet several nucleotides reduce complex-2 formation when mutated (*attB37SDM13* through *attB37SDM16*). Most prominent among these are substitutions at positions -3 and -4 (*attB37SDM26* and *attB37SDM6*), which eliminate all binding activity of complexes-2 and 3. This is especially interesting, because these two positions make up a 5'-AG dinucleotide, corresponding to the positions found in B', which also form a 5'-AG dinucleotide. The dinucleotides are also symmetrically positioned around the cleavage points identified in Chapter 2. While this sequence may be necessary for binding at B, it is not sufficient, as this binding does not occur in the absence of B'.

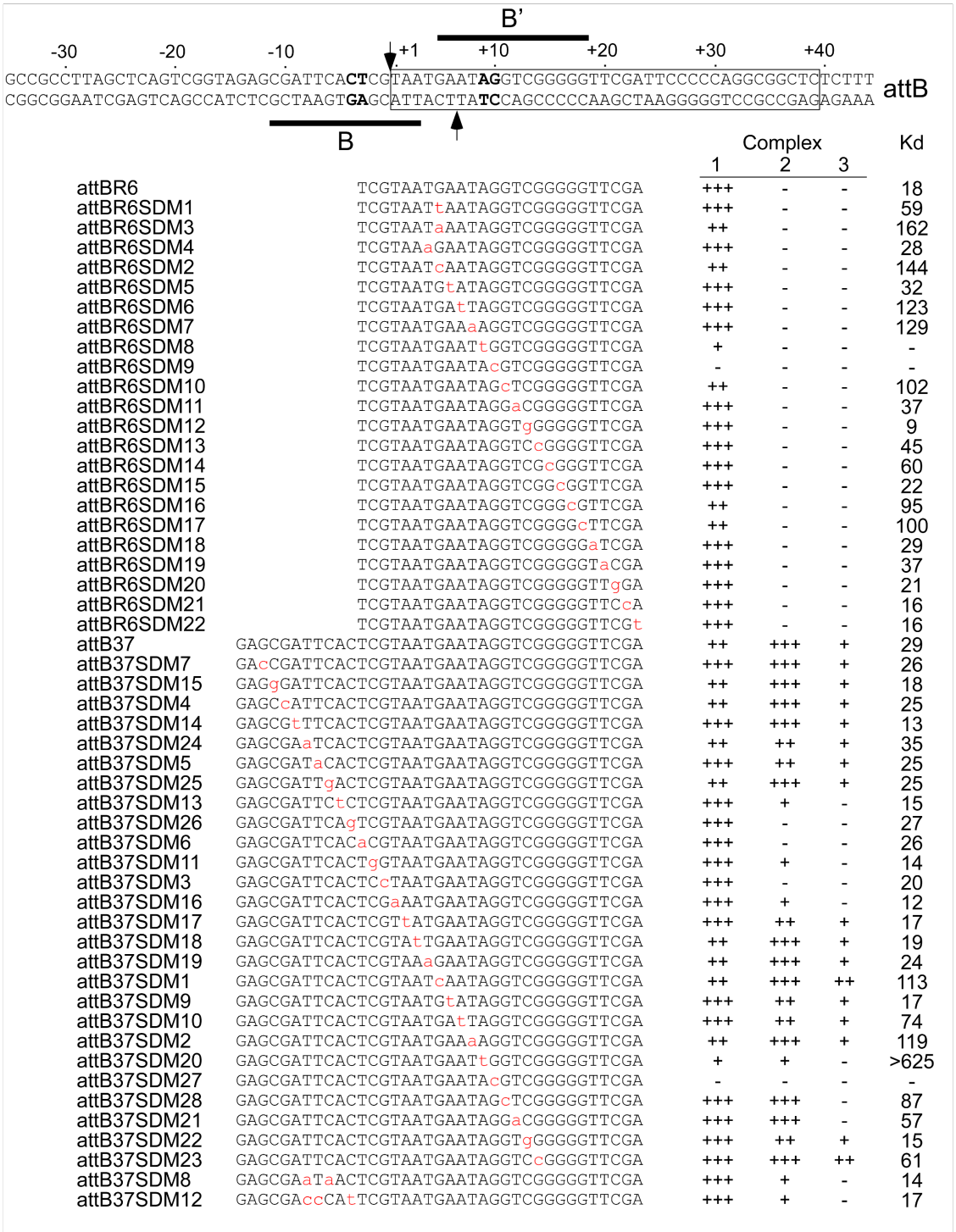


Figure 19: Summary of Brujita Int binding to *attB* substrates with single-base substitutions

attB substrates with single-base changes from the wild-type sequence were tested for Int binding. Complex intensities and affinities are as described for Figure 17. The box encompassing positions +1 to +39 indicates the core region.

3.3.6 Requirements for Int-*attP* Binding

The *in vivo* site requirements show that *attP* is longer than *attB* and that thus binding must differ between them. Additionally, the right half sites, B' and P' are both located within the core and are identical, so the binding requirements for B' already examined must also apply to P'. Thus any differences between the two sequences must be due to the left half of the attachment site. All requirements for Int-*attP* binding are contained within coordinates -42 to +21 (*attP63*) and some of these requirements are located in the extreme 9 bp at the 5' end of this segment (*attP67plasmid, attP67NS5,* and *attP67NS3*). Binding occurs differently between *attP* and *attB*, as *attP37* (which includes the same relative positions as *attB37*) only forms complex-1. This complex presumably contains a monomer of Int bound to DNA, as it migrates similarly to *attB* complex-1. The remaining 30 bp (*attP30*) also result in complex-1 formation. However, the DNA included by *attP30* does not encompass the P half site, so there must be an additional binding site present. Additionally, *attP28* – which lacks a full P half site – shows no difference in binding when compared to *attP37*, indicating that binding does not occur at the P half site. Thus *attP* complex-1 must be comprised of two different forms, one with Int bound at the P' half site, and one with it bound to the second site, which we name P1. Complex-2 – formed with *attP* – would then consist of two Int subunits bound to *attP*, one to each site, and each binding independently.

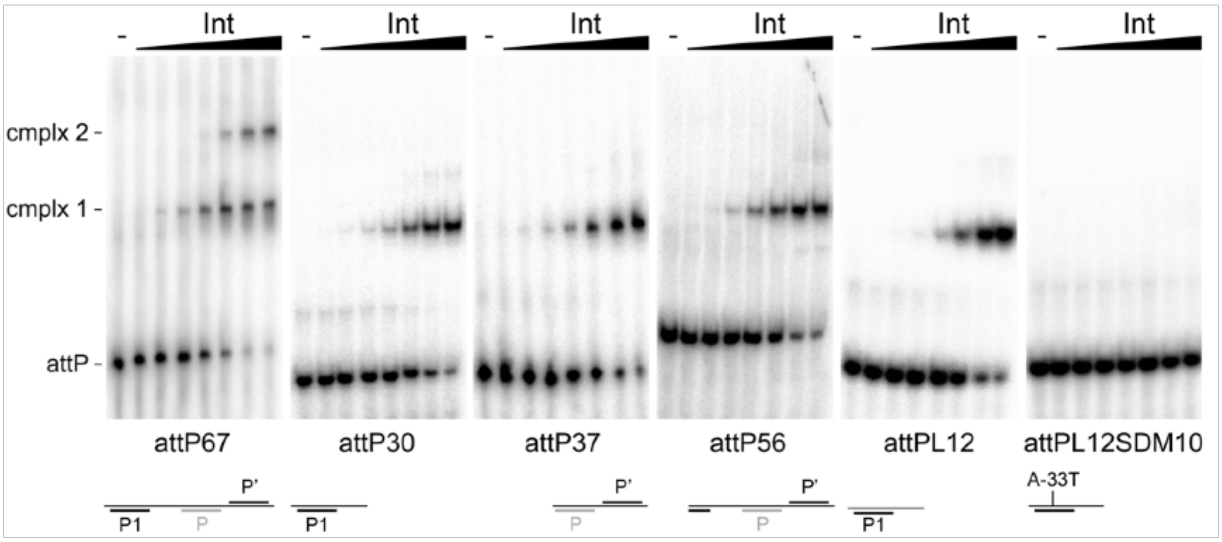


Figure 20: Int binding to mutant derivatives of *attP*

Binding of Int to representative *attP* DNA substrates. Int concentrations are, from left to right, 0, 0.15, 0.61, 2.4, 9.8, 39, 156, and 625 nM. See Figure 21 for substrate sequences.

Substitutions in the *attP* DNA reveal that several positions alter binding to P1 (Figures 20-21). Positions -32 and -33 are particularly important and substitution of these nucleotides eliminates binding entirely. These positions also comprise a 5'-AG dinucleotide, in direct orientation to that found in P' (and B'). When either of these dinucleotides is mutated in a larger substrate (*attP67*), it results in loss of complex-2 and formation of only complex-1 (Figure 23B). When the dinucleotide is mutated in both P1 and P', the resulting DNA has no binding activity, indicating that no binding occurs at P. Additionally, this mutation in P does not influence binding. Thus it appears that P is not used in this system, a particularly unexpected result, as binding at the P half site is likely to be required for proper nucleophilic attack of the DNA backbone during recombination.

Figure 21: Summary of Brujita Int binding to mutant *attP* substrates

attP DNA substrates of varying lengths or with mutations were tested for ability to bind Int. The intensities of complexes 1 and 2 (as shown in Figures 15 and 20) are scored from weakest (–) to strongest (+++), with the +++ designation corresponding to a similar intensity to input DNA. Overall binding affinities (K_ds) are shown, calculated as the Int concentration at which 50% of the input DNA is bound, regardless of the complexes formed. Base positions in upper case black type correspond to wild-type *attP* sequences, and those in blue lower case type are non-native *attP* sequences corresponding to either plasmid-derived sequences (substrates *attP67plas*) or random sequence. Selected gels are shown in Figure 20.

3.3.7 Positions of *attP* Half-Sites and DNA Spacing

The Brujita *attP* site consists of two half-sites that exhibit *in vitro* binding, P1 and P'. There is no site with *in vitro* binding at the traditional P position, though this position is likely to be required for recombination. There are two conceivable reasons for this. First, predictive modeling of the *attP* DNA shows that it may have a bend of 22° between the P1 and P half-sites, due to sequence alone (Model.it, data not shown) (Munteanu *et al.*, 1998). Additionally, half-sites P1 and P' are positioned approximately 30 bp apart, which is similar to the spacing needed for IHF or HU to cause large bends approaching 140° in DNA (Moitoso de Vargas *et al.*, 1989, Bonnefoy & Rouviere-Yaniv, 1992). Finally, the P1 half site is oriented in same orientation as P', opposite to the orientation that would be required for the P half site (see Figure 21). Thus it seems possible that the P1 site is compensating for the lack of a site at P by using a large DNA bend to position the P1 half site in the proper orientation and in proximity to the other binding sites. A second possibility is that P1 facilitates binding at P by protein-protein interactions; in which case, a new assay would be needed to test this hypothesis.

In order to test the DNA bending hypothesis, various insertions or substitutions were made in the region between P1 and P' and the effects on recombination were determined (Figure 22). There are two A/T rich regions in this portion of *attP*: a triad of adenines, and a tetrad of thymines. Regions rich in A/T nucleotides are often involved in intrinsic DNA bends (Crothers *et al.*, 1990) and these nucleotides may be contributing to a bend in this system. To determine whether these sequences play a role in this system, they were substituted with C/Gs, which results in a loss of recombination ability (Figure 22), indicating that these sequences are important for recombination. One, two, five, and eleven base C/G or random nucleotide insertions reduce recombination heavily, though it still occurs at low levels (1-4% of wild type). This suggests that the spacing between P1

and P' (or P) is important, as any deviation from this spacing causes a similar defect. If this DNA were undergoing a severe bend similar to that caused by HU and IHF, we would expect to see eleven base insertions rescue recombination, as this would introduce an additional full turn in the DNA and preserve the orientation of the sites on the double helix. Since this is not the case, it is unlikely that severe DNA bending is occurring in this region.

Plasmid	Sequence	Mutation	Integration Efficiency (%)
	P1 P P'		
pBL90	GAGCGTCGGATAGGTTGCCGAAACCGCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	WT	100
pBL238	GAGCGTCGGATAGGTTGCCGcgcCCGCCCGcgcgGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	3 and 4 bp substitutions	0
pBL239	GAGCGTCGGATAGGTTGCCGAAACCGggCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	2 bp insertion	4.0
pBL240	GAGCGTCGGATAGGTTGCCGAAACCGggccgCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	5 bp insertion	1.3
pBL241	GAGCGTCGGATAGGTTGCCGAAACCGggccgggccggCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	11 bp insertion	2.6
pBL248	GAGCGTCGGATAGGTTGCCGAAACCGgtacagtctCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	9 bp insertion	2.7
pBL247	GAGCGTCGGATAGGTTGCCGAAACCGgtacagtctgaCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	11 bp insertion	0
pBL242	GAGCGTCGGATAGGTTGCCGAAACCGCCCGTTTTGGTCgcGTCTGATAATGAATAGGTCGGGGGTTCGA	2 bp insertion	3.5
pBL243	GAGCGTCGGATAGGTTGCCGAAACCGCCCGTTTTGGTCgccgcGTCTGATAATGAATAGGTCGGGGGTTCGA	5 bp insertion	0
pBL244	GAGCGTCGGATAGGTTGCCGAAACCGgCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	1 bp insertion	4
pBL245	GAGCGTCGGATAGGTTGCCGAgACCGCCCGTTTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	1 bp substitution	115.2
pBL246	GAGCGTCGGATAGGTTGCCGAgACCGCCCGTgTTGGTCGTCTGATAATGAATAGGTCGGGGGTTCGA	2-1 bp substitutions	0

Figure 22: Mutations in the proposed bending region

Substitutions (blue) or insertions (red) were made in the proposed bending region between P1 and P. The mutations and the resulting effect on recombination (integration efficiency) are shown as a percentage of wild type.

Insertions of two and five bp into the middle of the P half site also reduce recombination (3.5 and 0%), which may be due to disruption of the P half site. A single base substitution in the adenine triplet (pBL245) does not alter recombination, but a second substitution in the thymine tetrad eliminates it, suggesting that the tetrad is important for either binding or recombination. The reason for this importance is not clear, as it lies outside of the P half site.

3.3.8 Half-Site Conserved Dinucleotides

Mutational analysis of *attB* and *attP* binding has revealed an important 5'-AG dinucleotide present in each of the five half sites (B, B', P1, P', and putative P). In each case, with the exception of P, this dinucleotide is critical for Int binding, as mutation of either position is sufficient to abolish this interaction. Although binding is not observed at the P half site, it is difficult to conceive a method where cleavage could occur without this site. The hypothesis that the DNA was bending between P1 and P has no supporting data. Alternatively, we hypothesize that although it is undetectable *in vitro*, binding occurs at P and that this half site is important for recombination. We mutated the 5'AG dinucleotide in each of the five half sites and tested their effects on recombination (Figure 23A). As expected, recombination is heavily reduced when the dinucleotide is mutated in B, B', P1, and P'. However, alteration of the P half-site dinucleotide results in a moderate reduction in recombination, suggesting that this sequence and the P half site are important in the recombination reaction. The effect of altering this dinucleotide is less severe than with the other half sites, which implies that, though important for recombination, binding at this site is not as critical as with the other half sites. Additionally, since binding is undetectable here, it is possible that Int is held in position at this site by protein-protein interactions with protein bound to the other half sites.

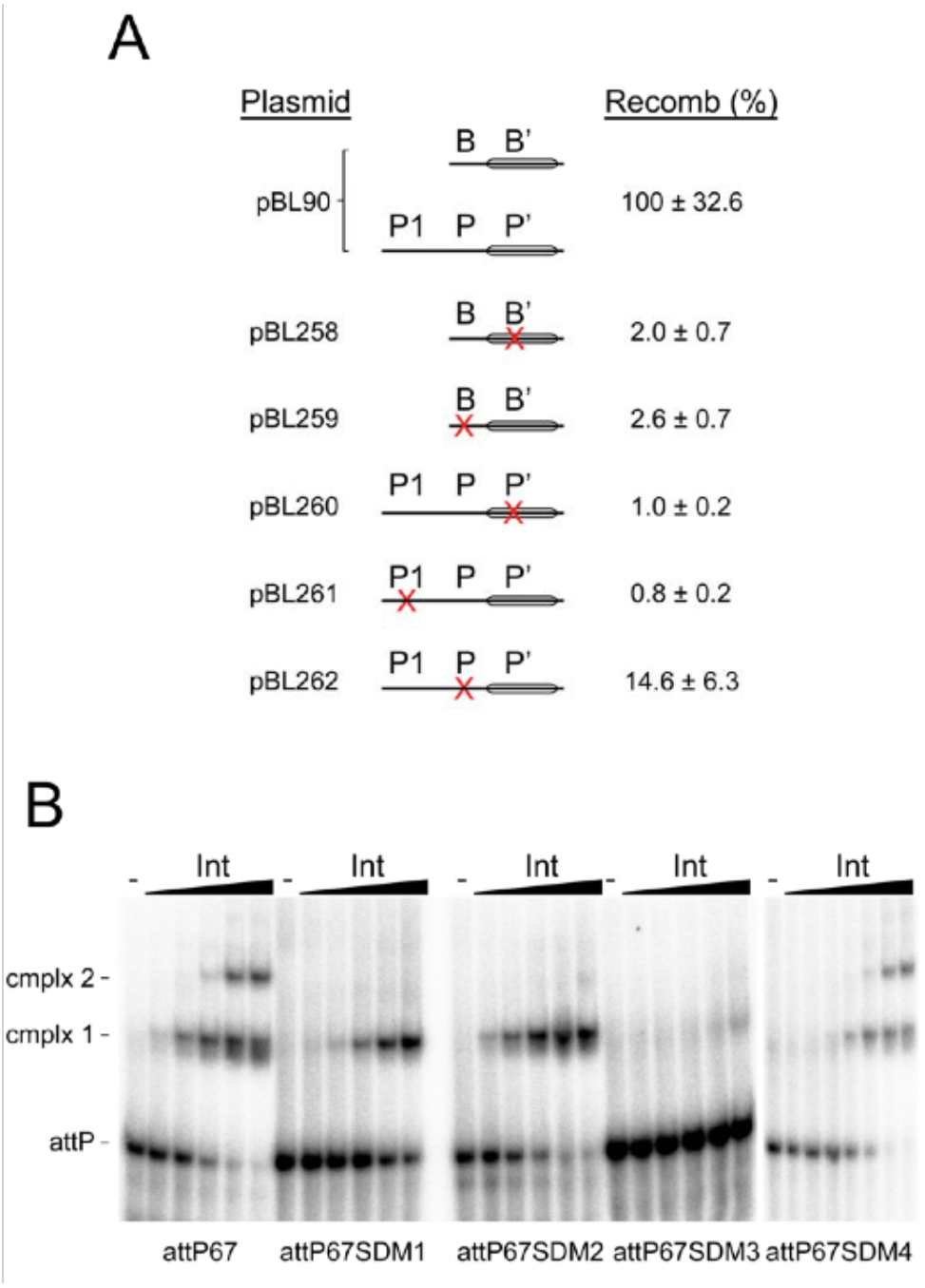


Figure 23: Brujita binding at the P half site is required for recombination

A. *attP* and *attB* substrates were mutated in B′, B, P′, P1, or P half sites as indicated (X), changing the 5′-AG bases in each half site important for Int binding to 5′-TC. The recombination frequencies are shown compared to wild-type (pBL90). **B.** Int binding to *attP* substrates with 5′-AG mutations. From left to right, wild- type *attP*, mutations in P′, mutations in P1, mutations in both P′ and P1, mutations in P. Concentrations from left to right for *attP67* through *attP67SDM4*; 0, 2.4, 9.8, 39, 156, and 625 nM.

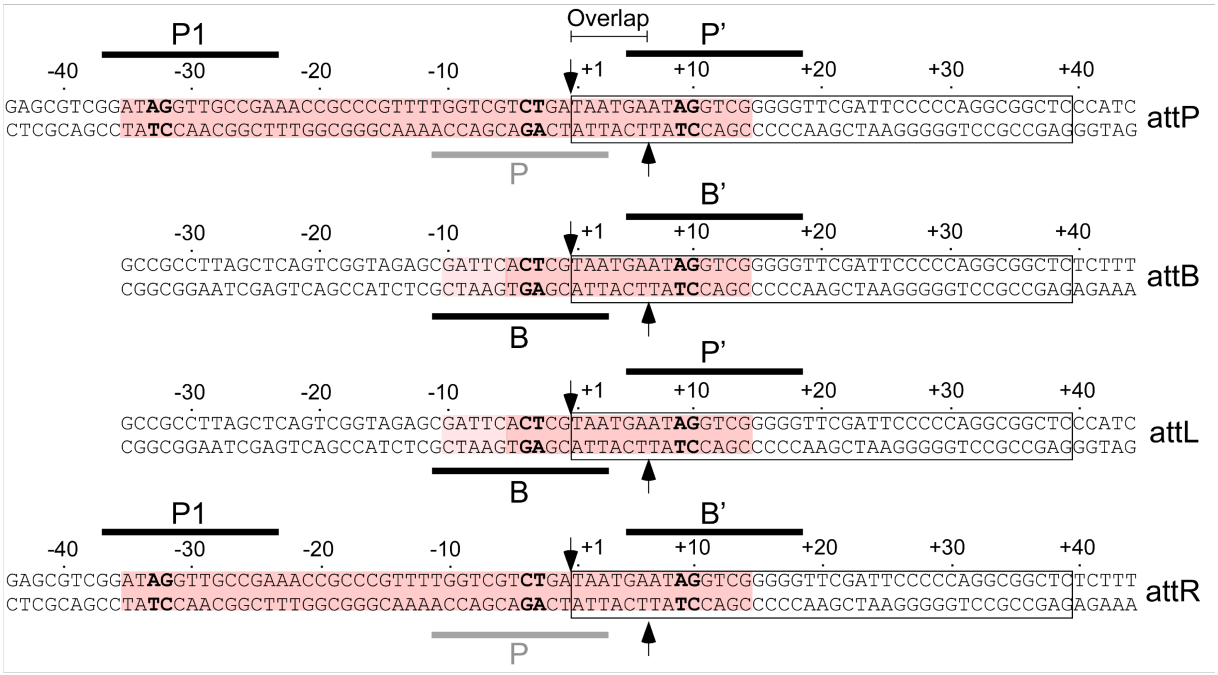


Figure 24: Organization of Brujita attachment sites

attP and *attB* share a 39-bp common core sequence (boxed) that overlaps a host $tRNA^{Thr}$ gene. Int-mediated strand exchange occurs within the common core to form *attL* and *attR* product sites, and mutations analysis suggests stand cleavages at the left end of the common core as indicated by the vertical arrows. DNA binding studies show that Int binds to two half sites, B, and B′, in *attB* and to P′ and a leftmost site designated P1 in *attP*. Int does not bind to the putative P half site *in vitro*, but the site is required for recombination *in vivo*. Because the recombination requirements do not extend beyond the common core to its right, the B′ and P′ sites are identical, and thus, *attP* and *attR* as well as *attB* and *attL* are functionally equivalent. The regions required for in vivo recombination are highlighted in red. 5′-AG dinucleotides important for Int binding are shown in bold.

3.4 DISCUSSION

In this chapter, we have shown that Brujita Int binds *attB* DNA to form three complexes and forms two with *attP* DNA. Additionally, Int^{Long} and Int^{Short} bind DNA to form similar complexes and both exhibit recombination activity, but Int^{Short} does not bind as strongly, nor recombine as efficiently as does Int^{Long}. Thus, it appears that the N-terminal 21 residues of Int^{Long} that are absent in Int^{Short} may improve DNA binding and recombination.

Competitive binding with unradiolabeled DNA reveals that recombinant products or synaptic complexes are not formed in these binding reactions. Presumably, this is due to the buffer system, as we achieved recombinant products with a different buffer (see section 2.2.6). Additionally, from these assays, we learn that these complexes are relatively unstable, as unradiolabeled probe is fully capable of competing with pre-bound radiolabeled probe. Thus, the DNA-Int binding interactions appear to be short-lived, though the extent of which is unknown. This differs from other integration systems, such as Lambda Int, whose synaptic complexes are stable for more than 60 minutes (Segall & Nash, 1993).

The experiments in this chapter also show that Int binds to both *attP* and *attB* independently as monomers at P1, P', and B'. We identify the stoichiometries of complex-1 and complex-2: with complex-1 containing a monomer of Int bound to a single DNA and complex-2 containing an additional monomer bound to the same DNA at a second site. In *attP*, these binding events are independent, but in *attB* they are cooperative, with binding at B completely relying on occupancy of the B' site.

The effective range of the Ferguson plots confirms the size and composition of complexes-1 and 2, but not complex-3. From the data, we confirm not only that the mass of complexes-1 and 2 conform to the previously demonstrated compositions, but also that Int binds *attB* DNA in

monomeric form, as we see the size of complex-1 consistent with a single monomer and complex-2 consistent with two monomers. The size of complex-3 may be too large to measure accurately, as its mobility is very low in all gels. It is also possible that the data are variable due to a non-spherical shape of the Int-DNA complexes, which may result in an unusual tumbling radius while migrating through the gel matrix. It is unclear, however, whether this would have a significant impact on migration. Another potential cause of the high propensity of outliers is poor resolution and visualization of the complexes. Three attempts were made to measure these complexes, but in each case, staining was poor and some bands could not be easily seen or differentiated. Repeating the experiment using longer gels or a better silver staining method may alleviate this problem.

In this chapter, we identify the Int binding requirements for *attB* and *attP*. These binding sites agree well with the previously discovered recombination requirements in Chapter 2. Approximately 37 bp is required for *attB* binding and *attP* requires approximately 60 bp. These sequences share a common right side, with identical B' and P' half sites, meaning that *attL* and *attR* are equivalent at the sequence level to *attB* and *attP*, respectively. We observe that individual nucleotides generally have little effect on binding, with the exception of a conserved 5'-AG dinucleotide, found in each half site (B, B', P1, P, and P'). This dinucleotide is required for binding to each half site and recombination is heavily reduced in the absence of a single dinucleotide.

The lack of binding at P is surprising, especially due to the strong cooperative interactions between B and B'. One may expect a similar cooperative interaction to recruit binding at P, but this was not observed in any arrangement, even with occupancy of the P' and P1 half sites. Oddly, even though binding is not detectable at P, this dinucleotide is required for *in vivo* recombination, demonstrating that this site is indeed important. It may be that binding at this site is so transient and weak that it is undetectable *in vitro*, or that some cofactor – which facilitates binding – is

missing. We prefer the former hypothesis, as *in vitro* recombination shows that no cofactor is required. Interestingly, the P1 half site, which extends to the left in the asymmetrical portion of *attP*, is also required for recombination. Why this site is required is unclear, though it may help to stabilize interactions at the P half site. The positioning of this sequence to the left of the core binding sites implies that its role in recombination is not catalytic.

The dissimilarity between the left boundaries of *attP* and *attB* may cause site selectivity such that recombination with identical sites would be less likely than *attP* x *attB* recombination. However, we do observe self-recombination (*attB* x *attB* and *attP* x *attP*, discussed in section 2.3.6), although at reduced rates compared to *attP* x *attB* recombination *in vivo*. The differences between *attP* x *attB* recombination and self-recombination may be magnified *in vivo* with a wild type version of Int that does not carry the stabilizing mutation.

Insertions between P1 and P show that there is some requirement for proper spacing between these sites, and that any deviation nearly eliminates recombination. It is possible that there are protein-protein interactions occurring between P1 and P and that this is allowing a monomer to bind at the P half site. The insertion experiments support this, as a single base pair insertion would rotate the two binding sites approximately 34° (360°/10.5 bp) with respect to each other, significantly altering their position and ability to interact. Additionally, substitution of a single position of the thymine tetrad eliminates recombination, even though this site is not predicted to have a role in binding. Instead, this may be important for allowing DNA to bend slightly, allowing Int bound at P1 to interact with a monomer at P.

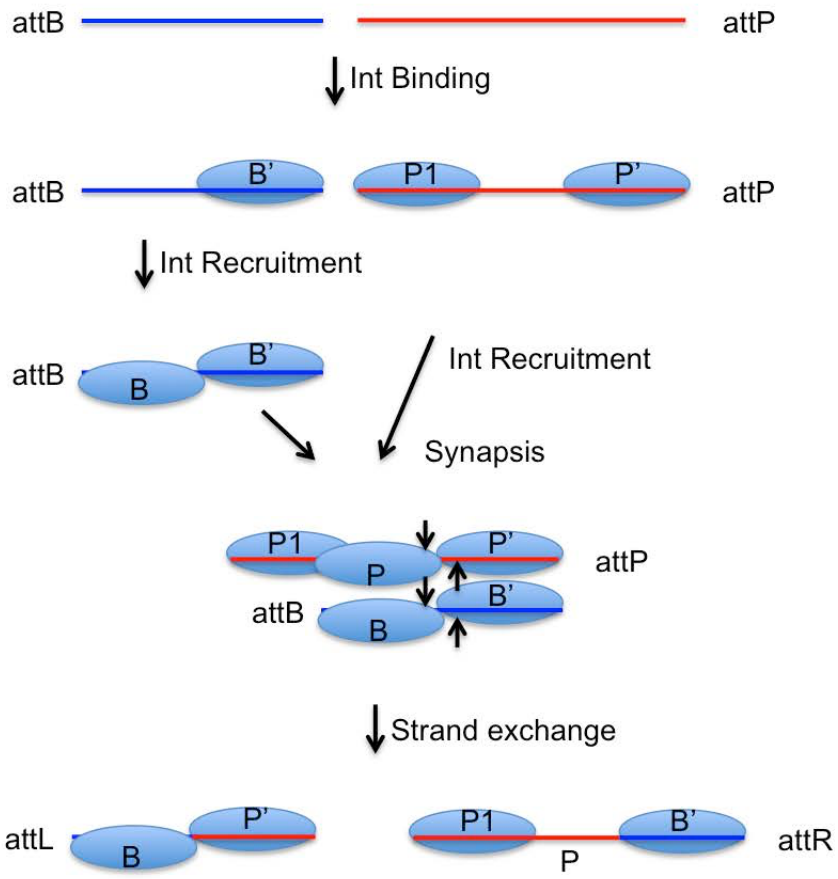


Figure 25: A model for Brujita Int-mediated recombination

Int protomers (blue ovals) bind independently to B′, P, and P′ half sites. Binding at B′ facilitates recruitment of an Int protomer to the B half site, presumably through subunit–subunit interactions, whereas Int binds independently to P1 and P′ half sites, but not to P. We propose that recruitment of Int to the P half site occurs at synapsis, followed by cleavage at the crossover sites (vertical arrows), strand exchange, and release of *attL* and *attR* products.

We propose a model for Int binding to attachment sites (Figure 25). First, independent binding events occur at P1, P', and B'. Binding at B' cooperatively recruits Int binding to B. Binding at P only occurs at synapsis, with potential interactions from P1, P', and possibly B to stabilize the very weak Int-DNA binding. Cleavage and strand exchange occurs, followed by release of the products *attL* and *attR*.

4.0 STRUCTURAL COMPONENTS OF THE BRUJITA INT SYSTEM

Work discussed in this chapter is being prepared in the following article:

Lunt. B.L., VanDemark, A.P., and Hatfull, G.F., Structure of the Mycobacteriophage Brujita Integrase.

The crystal structure data collection and structure determination were performed by Andy VanDemark. Many of the figures and tables were made in collaboration with Graham Hatfull and Andy VanDemark.

4.1 INTRODUCTION

The next step in understanding Brujita Int is to identify how this system forms the synaptic complex. This is particularly interesting because of the requirement for the fifth Int binding site at P1, as the role for this site is not yet known. We hypothesize that the fifth Int protein is required to help recruit a protomer of Int to the weakest interacting core site (P half site, see section 3.4) (Lunt & Hatfull, 2016). If true, this would require specific protein-protein interactions that may be unique to this system.

4.2 MATERIALS AND METHODS

4.2.1 Expression of Brujita Integrase

pBL61 was made as described in section 2.2. To generate a catalytically inactive Int (Y269A), pBL61 was mutated to create pBL231 using site-directed mutagenesis.

4.2.2 Protein Expression

Brujita Int was expressed as in Chapter 2 and (Lunt & Hatfull, 2016). Briefly, plasmids expressing Int were transformed into BL21-CodonPlus (DE3)-RIPL cells (Agilent). A single colony was used to inoculate a 50 ml overnight culture grown at 37°C in appropriate antibiotics. This culture was pelleted and resuspended in 2 L fresh autoinduction media with antibiotics (Studier, 2005). Cells were shaken vigorously (160-250 rpm) at 23°C for 18 hours. The cells were pelleted, frozen at -80°C and resuspended in lysis buffer (10 mM phosphate pH 7.0, 400 mM NaCl, 5 mM imidazole, 5% glycerol, 1 mM BME). PMSF, leupeptin, aprotinin, and pepstatin were added and cells were lysed by 3 passes through a homogenizer. The insoluble protein was pelleted in a Sorvall centrifuge at 13,000 rpm, 4°C for 30 minutes and the supernatant recovered. Brujita Int mutants were expressed using 1.25 ml overnight starter cultures to inoculate a 50 ml culture.

4.2.3 Purification of Brujita Int

The soluble protein was subjected to nickel affinity chromatography, followed by heparin affinity chromatography. We have found that like XerC/D, untagged Brujita Int binds avidly to nickel

(Subramanya *et al.*, 1997). Following the heparin column, the protein was further purified by size exclusion chromatography. The resulting protein was concentrated using a spin concentrator and stored at 4°C for future use in 10 mM Tris pH 7.0, 250 mM NaCl, 5% glycerol, and 1 mM BME.

4.2.4 Crystallization Conditions

Brujita Int Y269A crystal screens were prepared using the sitting drop vapor diffusion method (Stevens, 2000). Initial crystals were observed using a 1:1 mixture of 7 mg/ml protein in 100 mM Tris-HCl pH 8.5, 8% PEG-8000. Crystals form very quickly and could be observed within hours. Diffraction quality crystals were grown in 100 mM Tris pH 7.5, 5% PEG-3350.

4.2.5 Seleno-Methionine Crystals

Protein was prepared as before using established protocols for selenomethionine growth. Selenomethionine crystals were grown in 100 mM Tris pH 8-8.5, 4-10% PEG-3350 using 8 mg/ml Int.

4.2.6 Structure Solution

Brujita Int crystals described here are members of the space group C2, with unit cell dimensions a = 133.2 Å, b = 50.1 Å, c = 60.7 Å. Diffraction data from the native Brujita Int crystals were collected at the University of Pittsburgh BST3 facility, and selenomethionine crystal data were collected at the Advanced Light Source synchrotron. Experimental phasing was performed by the single wavelength anomalous dispersion technique using a selenomethionine crystal collected at

peak wavelength. The initial model was refined against native data at 1.85 Å using positional refinement and anisotropic *B*-factor refinement.

4.2.7 Generating Mutants

Targeted mutants were made by site directed mutagenesis, using pBL61 as a template for expression and purification, or pGWB87 as a template for integration assays. Mutants were Sanger-sequenced for verification.

4.2.8 Binding Assays

Binding assays were performed as described in Chapter 3 and (Lunt & Hatfull, 2016).

Table 5. Plasmids used in Chapter 4.

Plasmid Name	Description	Reference
pBL90	pBL50 with ccdB (from pDONRZeo) cloned into DraI site	Lunt and Hatfull, 2016
pBL258	pBL90 with SDM 5'AG-TC in B'	Lunt and Hatfull, 2016
pBL259	pBL90 with SDM 5'AG-TC in B (reverse complement)	Lunt and Hatfull, 2016
pBL260	pBL90 with SDM 5'AG-TC in P'	Lunt and Hatfull, 2016
pBL261	pBL90 with SDM 5'AG-TC in P (reverse complement)	Lunt and Hatfull, 2016
pBL262	pBL90 with SDM 5'AG-TC in P1	Lunt and Hatfull, 2016

4.3 RESULTS

4.3.1 The Crystal Structure of Brujita Int

In order to understand how Brujita Int complex formation and synapsis occur, we formed crystals and collected X-ray diffraction data from Brujita Int containing the A317E stabilizing mutation, as well as Y296A, inactivating the catalytic tyrosine, as this mutant yielded higher quality diffraction data. Initial diffraction data was collected at the University of Pittsburgh BST3 facility. However, attempts to solve the phage problem using molecular replacement were unsuccessful. We instead resorted to using the anomalous dispersion method with selenomethionine crystals. These crystals were formed as described in section 4.2.5 and selenomethionine crystal data were collected at the Advanced Light Source synchrotron. Details of the structural solution are found in section 4.2.6. The structure of Brujita Int has been solved to 1.9 Å resolution. 292 of the 318 amino acids are present in the structure, including residues S14 to E305.

Brujita Int shares a similar structure to the other members of the tyrosine recombinases, including XerD and Lambda Int (Aihara *et al.*, 2003, Biswas *et al.*, 2005, Subramanya *et al.*, 1997). This is unexpected considering these proteins are distantly related with 18% identity to Lambda Int and 26% identity to XerD. The greatest differences in structures occur at the extreme C-terminus, where there is an increased sequence and structural diversity than in the rest of the proteins (Figure 26A, C) (Subramanya *et al.*, 1997).

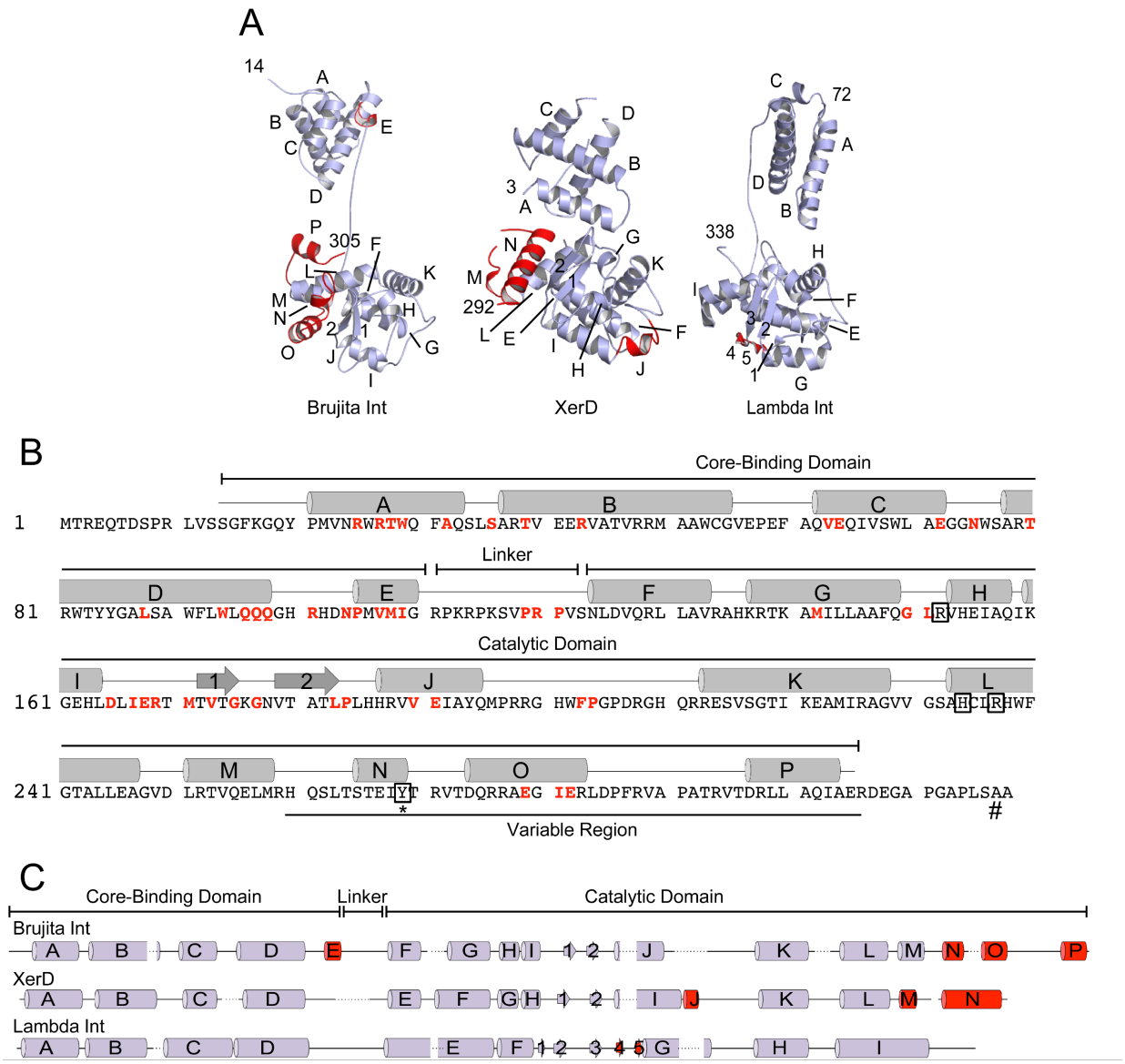


Figure 26: Structural organization of Brujita Int

A. Brujita Int consists of two domains, a core-DNA binding domain (top) and a catalytic domain (bottom), separated by a linker region of ~12 residues. The structure is mostly α-helical, with helices labeled A-P, and contains two β-strands arranged in an antiparallel sheet in the C-terminal, catalytic domain. Residue numbers indicate the N- and C-termini of the structure. The structures of Lambda Int (1Z19_A) and XerD (1A0P) are shown for comparison. Red colored residues denote differences from the conserved structure (see also 24D). **B.** Secondary structure map of Brujita Int. The Core-Binding Domain, Linker, and Catalytic Domain are marked by brackets and the variable region is underlined. The active site residues R152, H234, R237 and Y269 are boxed and an asterisk denotes the position of the catalytic tyrosine, which was mutated to an alanine for structural determination. The penultimate alanine residue (A317) was mutated to glutamic acid to promote stabilization of the protein. Point mutations were made to test residues important for

protein-protein interactions and are shown in red. The full details of these mutants are shown in Table 6. **C.** The Core-Binding Domain, Linker, and Catalytic Domain are marked as in C. Secondary structure alignment of Brujita Int, Lambda Int, and XerD showing conservation of secondary structure among these proteins. Red regions are deviations from this conserved structure. Compare to Figure 26A.

Brujita Int consists of two domains joined by a 12-residue linker. It lacks a characteristic N-terminal arm-binding domain, which is present in the other tyrosine integrases (Lunt & Hatfull, 2016, Aihara *et al.*, 2003, Biswas *et al.*, 2005, Pena *et al.*, 1999, Esposito *et al.*, 2001, Broussard *et al.*, 2013). The Brujita Int N-terminal, or core-binding (CB) domain, consists of residues 14-111 and the C-terminal, catalytic domain consists of residues 122-318 (Figure 26A-B). The CB domain contains four α-helices (A-D), organized in two perpenticular groups of antiparallel pairs, similar to the core DNA binding domain of Lambda Int, Cre, and XerD (Aihara *et al.*, 2003, Biswas *et al.*, 2005, Guo *et al.*, 1997, Gopaul *et al.*, 1998, Subramanya *et al.*, 1997). Additionally, this domain contains a small fifth α-helix (helix-E) just before the linker region, not found in Lambda Int or XerD (Aihara *et al.*, 2003, Biswas *et al.*, 2005, Subramanya *et al.*, 1997). The Brujita Int catalytic domain contains 11 α-helices (F-P) and two β-strands arranged in an anti-parallel sheet. Also present in this domain are the conserved active site residues (R152, H234, and R237), including the catalytic tyrosine (Y296) (Figure 26A-B).

The structure of the CB domain of Brujita Int is particularly similar among all the related structures and can be directly superimposed to them, with the exception of Brujita Int helix-E. The inter-domain linker is interesting in that it is largely ordered in Brujita Int crystals, despite having no organized secondary structure. There is limited electron density found toward the middle of the linker, suggesting some flexibility. This long linker is also found in Lambda Int, but not in the structures for XerD or Cre, which are also related to Brujita Int (Figure 26A, C) (Ghosh *et al.*,

2007, Subramanya *et al.*, 1997, Ghosh *et al.*, 2005). These linkers may be present in all four proteins, but it is possible that they are unobserved due to disorder in the crystal structures.

4.3.2 Domain Rotation

Though the Brujita CB domain is traditionally considered the DNA binding domain, *in vitro* studies show that protein fragments consisting of this domain (residues 1-114) are unable to bind DNA (data not shown). This is supported by the electrostatic analysis of the protein, which in Brujita Int and XerD show DNA binding surface in both domains (Subramanya *et al.*, 1997). Thus, DNA binding is likely a collaborative action involving both domains.

A difference between Brujita Int and the structures of Lambda Int and XerD is the orientation of the two domains (Subramanya *et al.*, 1997, Aihara *et al.*, 2003, Biswas *et al.*, 2005). In Brujita Int, the CB and catalytic domains are oriented approximately 180° relative to Lambda Int and XerD, such that a rearrangement would need to occur in order to align them to the latter structures. Electrostatic analysis of Brujita Int reveals patches of positively charged residues on the lower part of the CB domain, surrounding the linker and on the top of the catalytic domain (Figure 27). This is likely the portion of the protein that binds DNA as it is in XerD and Lambda Int (Subramanya *et al.*, 1997, Aihara *et al.*, 2003, Biswas *et al.*, 2005). The potential for DNA binding is difficult to see in the Brujita Int structure, due to the orientation of the domains (Figure 27A). Cleavage of Brujita Int in the linker region and subsequent alignment of each domain to Lambda Int-DNA (PDB: 1Z19_B) reveals a likely model for how Brujita Int rearranges upon contact with DNA (Figure 27B), with the DNA passing through the cavity that is created between the two domains and the linker after this rotation. We deduce that active version of Int occupies

the same orientation as Lambda Int. The linker region and its flexibility may permit the two domains to rotate and rearrange for DNA binding.

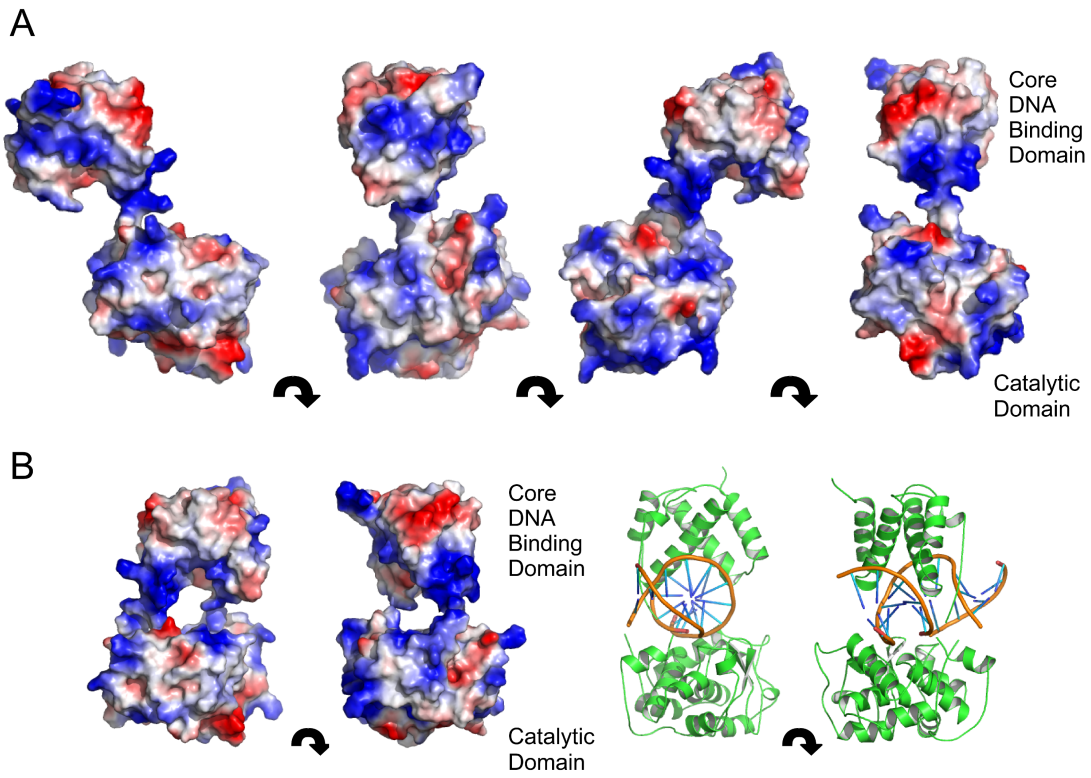


Figure 27: Brujita Int surface electrostatics

A. Brujita Int shown in four different angles (90° rotations) show positively charged surfaces (blue) around the lower face of the CB domain, the linker region and patches on the top of the catalytic domain. It is not obvious how Brujita Int could bind DNA in this arrangement. **B.** Brujita Int domains cleaved and aligned separately to Lambda Int-DNA (1Z19) showing a model for Brujita Int-DNA structure. The two views are 90° rotations. Surface electrostatics (left) and cartoon models (right) display Int in the same orientations.

A structural change is required for XerD to bind DNA. In this case, there must be a large conformational change, or separation or the two domains, in order to allow DNA to access the active site (Subramanya *et al.*, 1997). Because there is evidence for a conformational change in XerD, it is plausible that there is a similar change when Brujita Int binds DNA.

4.3.3 Organization of the Active Site

The catalytic domain of Brujita Int is folded similarly to HP1 Int, XerD, and Lambda Int, (Hickman *et al.*, 1997, Subramanya *et al.*, 1997, Biswas *et al.*, 2005), with the exception of Brujita Int helices N-P (residues 263-305), comprising the variable region (see Figure 26B-C). This region has increased variability in both sequence and structure when compared to related proteins. In the Brujita Int structure, these helices (N-P) are organized and folded differently from XerD, HP1 Int, and Lambda Int (residues 335-356, Figure 26A, C) (Hickman *et al.*, 1997, Biswas *et al.*, 2005, Subramanya *et al.*, 1997, Argos *et al.*, 1986). The variable region includes the catalytic tyrosine (A269 in Brujita Int), which must be positioned correctly for attack of the phosphate backbone in recombination.

The conserved RHR triad coordinates the phosphate backbone to allow attack by the catalytic tyrosine (Grindley *et al.*, 2006, Landy, 2015). These residues are oriented to act in catalysis, with the three residues found in similar locations in each structure solved (Figure 28A) (Aihara *et al.*, 2003, Biswas *et al.*, 2005, Subramanya *et al.*, 1997, Gopaul *et al.*, 1998, Guo *et al.*, 1997). In Lambda Int, the structure of the catalytic domain without DNA (PDB: 1AE9_A) reveals the catalytic tyrosine positioned far (26 Å) from its position when bound to DNA (Figure 28C) (PDB: 1Z1G, 1Z1B, 1P7D, and 1Z19). This suggests that the variable region (including the last 23 residues) in Lambda Int rearranges upon binding to DNA (Aihara *et al.*, 2003, Landy, 2015, Kwon *et al.*, 1997). In Brujita Int, the catalytic tyrosine (mutated to alanine in the structure) is similarly positioned to that of the Lambda catalytic structure, though the position of the backbone in the rest of the variable region is quite dissimilar (Figure 28B). This provides a potential model for how the rearrangement can occur in Brujita Int, where – upon DNA binding – the backbone

may undergo a simple rotation at the top of helix-M (equivalent to helix-I in Lambda) and rearrange the variable region to position the nucleophilic tyrosine for catalysis (Figure 28D).

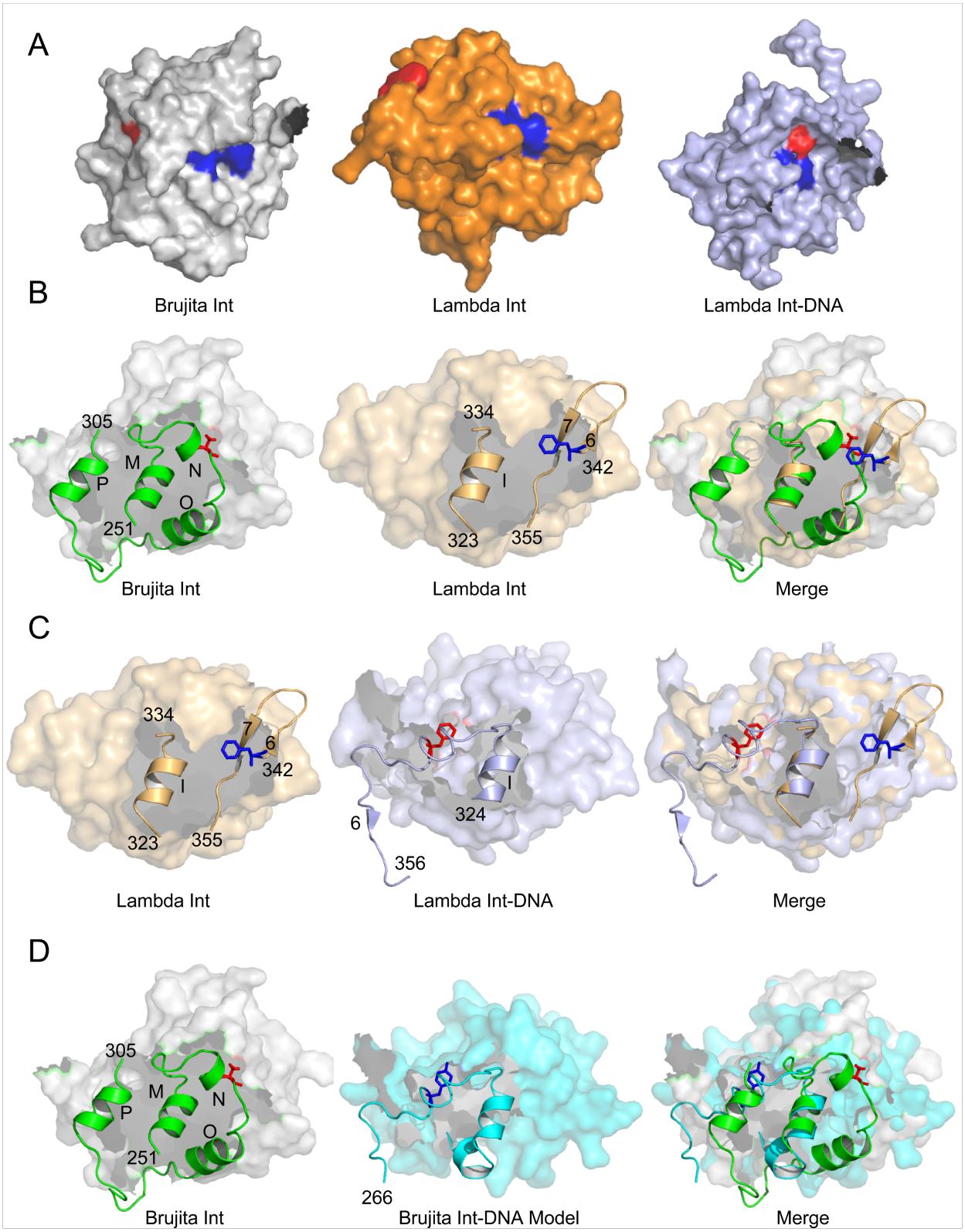


Figure 28: Rearrangement of catalytic tyrosine

A. From left to right, the catalytic domain of Brujita Int, the catalytic domain of Lambda Int (PDB: 1AE9_A), and the catalytic domain of Lambda Int-DNA (PDB: 1Z1G_A). The Brujita Int active site residues R152, H234, R237 are shown in blue and the catalytic tyrosine Y269 is shown in red. The Brujita Int catalytic tyrosine is positioned away from the rest of the catalytic residues, similarly to the structure of Lambda Int, not near the rest of the active site, as it is in Lambda Int-DNA. **B.** From left to right, the variable region of Brujita Int (green), Lambda Int catalytic domain (1AE9_A, orange), and merge of both structures. Immediately following Brujita Int helix-M (Lambda Int helix αI), the two structures diverge, with Brujita Int forming three helices, N, O, and P in a different fold than occurs in Lambda Int. The catalytic residues in both are not positioned to perform catalysis. The Brujita Int catalytic residue position is shown in red (Brujita Int A269), and Lambda Int in blue (F342 in this structure). **C.** From left to right, the variable region of Lambda Int without DNA (1AE9_A, orange), Lambda Int bound to DNA (1Z1G_A, light blue), and merge of both structures, showing the likely DNA-induced conformational shift of the variable region. The location of the catalytic residue position is shown in red or blue (Lambda Int F342). **D.** From left to right, the variable region of Brujita Int without DNA, Brujita Int modeled to Lambda Int-DNA (1Z1G_A), and merge of both structures, showing the potential conformational change that may occur upon DNA binding, allowing the catalytic position to be relocated into the active site.

4.3.4 Designing Mutations to Analyze Protein Structure and Function

As shown in Chapter 3 and explained in section 4.1, Brujita attachment sites contain 5 half sites, B, B', P1, P, and P', which are all required for recombination. Binding at *attB* exhibits cooperativity with binding at the B site depending on occupancy of B'. Presumably, this cooperativity is mediated by protein-protein interactions between subunits of Int (interaction-1 in Figure 29). We postulate that the P1 half site is required to mediate binding of Int at the P half site, likely through additional protein-protein interactions. Since binding to the P site is unobserved *in vitro*, this interaction may be solely due to cooperative interaction with the P1 bound subunit (interaction-3 in Figure 29). Additionally, the Int-DNA complexes must synapse for catalysis and strand exchange, which must also form subunit interactions between the members of the active tetramer (interaction-2 in Figure 29).

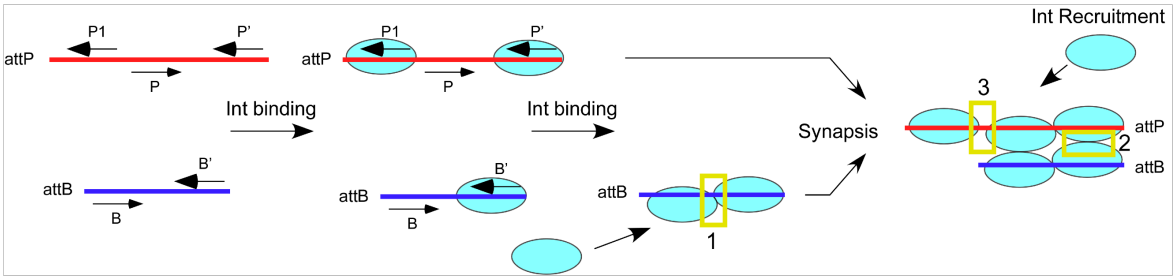


Figure 29: Model of Brujita Int binding

Schematic representation showing Brujita Int binding and formation of synapsis. Brujita Int binds independently to sites at B', P1 and P'. Binding at B' recruits Int binding at B and these four Int protomers recruit another protomer to bind at P, forming a synaptic complex as reported in (Lunt & Hatfull, 2016). Proposed subunit interactions are denoted by yellow boxes.

These predicted interactions can be examined by assessing DNA binding and recombination properties of strategically positioned mutations in Brujita Int. We note that efforts to obtain DNA-Int co-crystals have so far been unsuccessful. As a result, potential interactions were examined by modeling the Brujita Int structure to a Lambda Int tetramer bound to Holliday junction DNA (PDB: 1Z1G). The four subunits are labeled I, II, III, and IV (see Figure 31B). The solved crystal structure of Brujita Int contains a single protein in the asymmetric unit and the packing arrangement of the subunits reveals many types of different interactions between subunits. Three interfaces in particular are large (690, 681, and 1100 $Å^2$) (Figure 30), yet these do not seem to correlate with interactions in the modeled tetramer. We designed mutations in locations that were predicted to disrupt subunit interaction as well as residues that are conserved in the tyrosine recombinases. The DNA binding and recombination properties of these mutants are shown in Table 6.

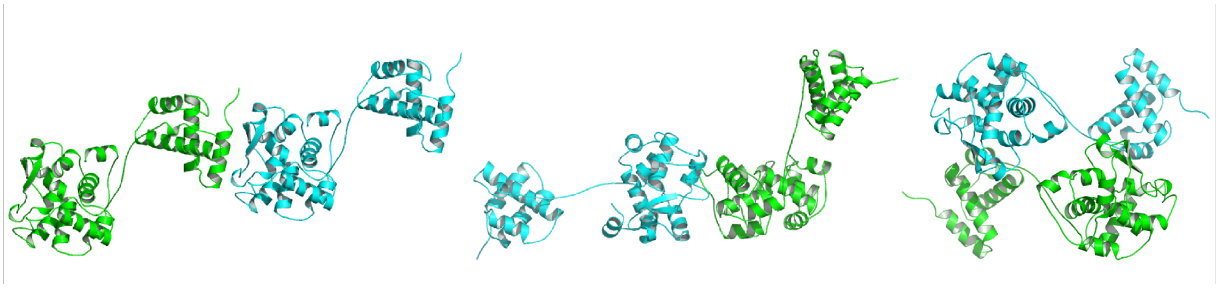


Figure 30: Brujita Int crystal subunit interactions

Three large interfaces were observed in the Brujita Int crystal packing arrangement of (from left to right) 690, 681, and 1100 Å^2. These do not appear to have any relation to interactions formed during synapsis, but may be related to interactions formed between Int bound to P1 and P (see section 5.1.4)

4.3.5 Effects of Mutations in the Core Binding Domain

We made a total of 20 Brujita Int mutants with amino-acid substitutions in the CB domain of Brujita Int. These mutants were tested for *in vivo* recombination and DNA binding ability. Thirteen of these mutants have strong defects in recombination (<1% of wild-type, see Table 6). Three of these are insoluble *in vitro*, likely due to misfolding (L88D, W94D, N104W/P105W). Three more have strong defects in DNA binding (T39W, R43E, and R101E). None of these residues is in a position to contact DNA – even with a 180° rotation of the subunits – and are probably also misfolded.

Table 6. Binding and recombination properties of Brujita Int mutants

Mutants were tested for their effect on binding and recombination. Each mutant was tested for its ability to bind a 44bp oligo representing *attB*. The relative binding efficiency is shown with the amount of each complex formed relative to control. Strong binding is represented by +++, weak binding by +, and no binding by -. The integration efficiency of each mutant protein is shown relative to control Int. All proteins contain the A317E stabilizing mutation with the sole exception of pBL263 WT (unstable).

Plasmid	Mutation	Location	Complex			attB44 K_d	Integration
			1	2	3	(nM)	(% of control)
pBL61	WT (stable)		++	+++	+	12	100
pBL263	WT (unstable)		++	+++	+	25	0.16
pBL268	R25A/T28A	α-A	++	+++	-	13	54.86
pBL269	R25E/R27E	α-A	++	+++	-	18	0.11
pBL324	R25A/R27A/ T28A/W29A	α-A	++	+++	-	23	24.3
pBL341	R27E	α-A	++	+++	-	43	48.82
pBL342	R27A	α-A	++	+++	-	32	40.75
pBL318	A32W	α-A	++	+++	-	44	0.0
pBL319	S36L	A-B Loop	++	++	++	89	0.15
pBL321	T39W	α-B	+	+	-	>775	0.1
pBL322	R43E	α-B	+	-	-	>657	0.3
pBL340	V63R/E64A	α-C	++	+++	-	33	33.68
pBL272	E72A/N75A	α-C	++	+++	+	15	140.26
pBL323	T80R	α-D	-	-	-	>442	0.0
pBL326	L88D	α-D				Insoluble	0.1
pBL327	W94D	α-D				Insoluble	0.0
pBL267	Q96A/Q97A/Q98A	α-D	++	+++	-	31	3.12
pBL328	R101E	D-E Loop	++	-	-	>472	0.0
pBL329	R101A	D-E Loop	++	++	-	86	0.2
pBL330	N104W/P105W	α-E				Insoluble	0.0
pBL316	V107A/M108A/I109A	α-E	+++	+	+	45	0.0
pBL331	I109R	α-E	++	+	-	304	0.0
pBL332	P119W/R120E/P121W	Linker	++	++	-	160	0.0
pBL333	M142H	α-G				Insoluble	0.0
pBL334	G150Y/L151D	G-H Loop				Insoluble	0.0
pBL343	D165A	I-1 Loop	++	+++	+	42	32.0
pBL271	I167A/E168A/R169A	I-1 Loop	+	+++	-	19	87.62
pBL325	I167E/E168R/R169E	I-1 Loop	++	+++	-	30	15.0
pBL335	M171H/V173R	β-1	+	++	-	97	0.0
pBL336	G175Y/G177Y	β-1	++	+++	+	65	0.0
pBL315	L183E/P184H	β-2	+	++	-	194	0.0
pBL337	P184W	β-2				Insoluble	0.6
pBL270	V190A/E191A	α-J	++	+++	++	16	63.56
pBL317	V190R/E191R	α-J	++	+++	+	54	0.34
pBL338	F203E/P204W	J-K Loop				Insoluble	0.0
pBL320	E279A/I281A/E282A	α-O	++	+++	-	24	0.0

Six of the 20 mutants (R25A/T28A, R25E/R27E, R25A/R27A/T28A/W29A, R27E, R27A, and A32W) bind *attB* similarly to wild type and form complex-1 and complex2, but fail to form complex-3 (Figure 31A). This indicates that these mutants can recognize DNA normally and are sufficient to mediate interactions between subunits occupying the B and B' half sites. Though the nature of complex-3 is unknown, it may form from interactions which are similar to those needed for synaptic complex formation. However, these mutants are not restricted to any single face of the protein, and only two are strongly recombination defective (R25E/R27E, A32W, <1% of wild type Int) (Table 6). Additionally, mutants such as S36L form complex-3, but are strongly recombination defective. Therefore, it appears that though complex-3 may have implications for synapsis, this complex is not required for recombination. Helix-A mutants exhibit normal recognition of DNA needed for cooperative binding of B and B' half sites. We propose that helix-A is important for subunit interactions that form the synaptic complex, specifically, interface-2 between protomers I and IV (Figure 31B). The mutants R25E/R27E and A32W appear to block recombination by either interfering with the interface or preventing domain dynamics, which may be needed for strand exchange. Other helix-A mutants may weaken these interactions and thus reduce recombination efficiency, but to a lower degree.

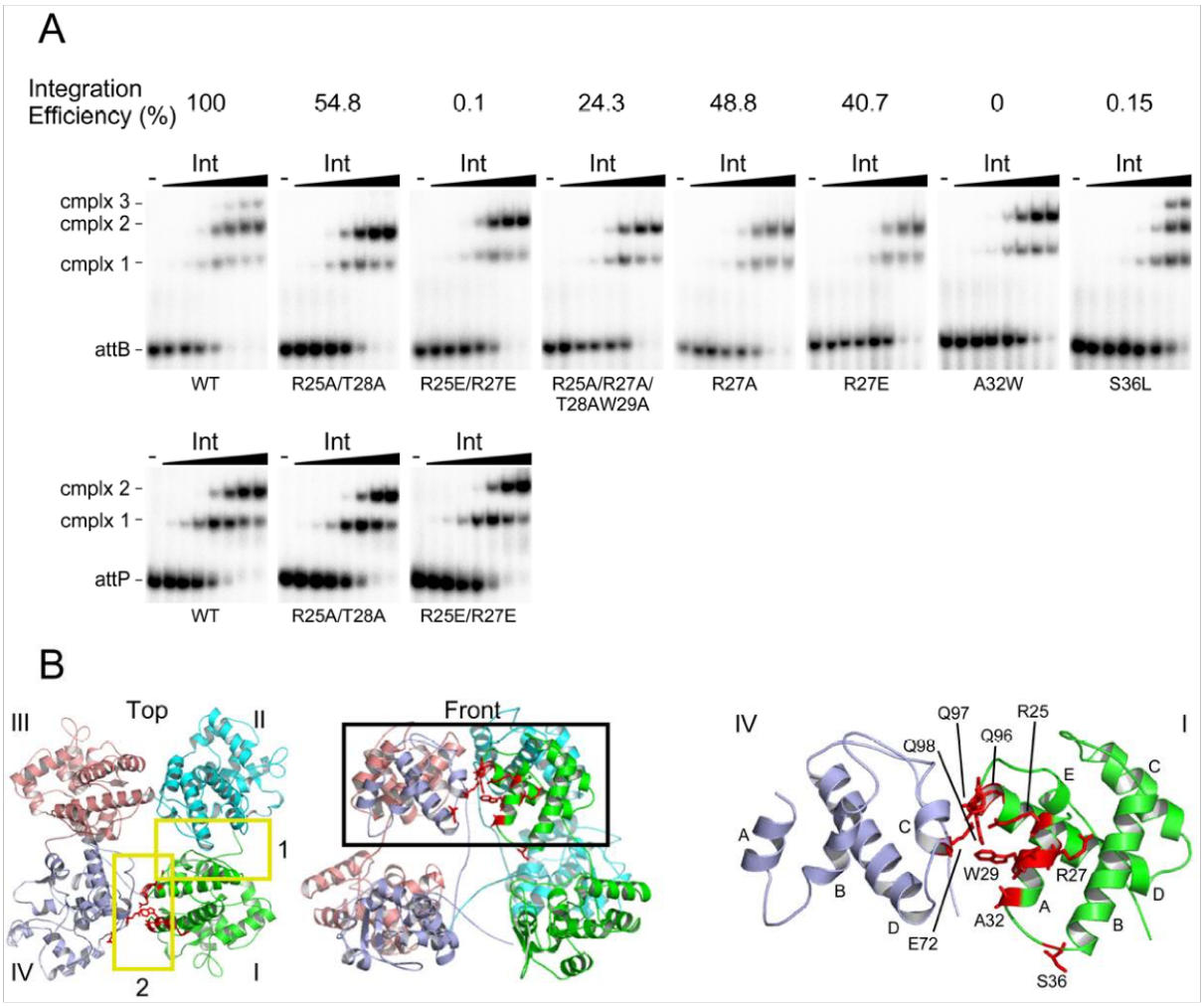


Figure 31: Brujita Int mutants affecting interface 2

A. Binding of Brujita Int mutants to a 44 base pair *attB* substrate (top) and an 89 base pair *attP* substrate (bottom). The positions of free DNA (*attB*) and complexes (cmplx 1, cmplx 2, cmplx 3) are indicated. Int concentrations are from left to right, approximately, 0, 0.15, 0.61, 2.4, 9.8, 39, 156, and 625 nM. **B.** Top and side views of a tetramer of Brujita Int modeled to Lambda Int-Holliday Junction DNA (PDB: 1Z1G). A close up of two core-binding domains (right), showing potential interactions involving mutated residues R25, W29, and E72. Other residues shown (R27, A32, S36) effect recombination but are not predicted to be involved in protein-protein interactions.

Residues Q96, Q97, and Q98 are located in the C-terminal end of helix-D and are oriented in interface-2 (Figure 31B). These substitutions result in protein highly defective at recombination, yet binding is nearly unaffected (Table 6 and Figure 32). When this region of Brujita Int is

examined in the modeled tetramer, we observe that these residues are positioned within interface-1 between subunits I and II (Figure 33), though they are also closely positioned to interface-2 between subunits I and IV (Figure 31B). We also observe that E72 and N75 of subunit I, are positioned in close proximity to Q96, Q97, and Q98 in interface-2 and these may play a role in the interface. However, the mutant E72A/N75A has no recombination or DNA binding defects.

Two of the mutations in helix-C (V63R/E64A, E72A/N75A) have only minor defects in recombination and successfully bind *attB* (Table 6). Furthermore, three of the mutants within helix-D (T80R, L88D, and W94D) are recombination defective. This is likely due to poor solubility (L88D, W94D) or defective binding (T80R), possibly due to defective folding. Alternatively, residue T80 may play a role in DNA-recognition, as it is positioned close to the region in which DNA is predicted to bind. This portion of helix-D is proposed to contribute to DNA binding in XerD (Subramanya *et al.*, 1997, Spiers & Sherratt, 1997).

We have observed that the Q96A/Q97A/Q98A mutation may interrupt both interface-1 and interface-2. However, this mutant appears to bind to *attB* cooperatively, implying that interface-1 between subunits I and II is not adversely affected. Thus, the effect on recombination is likely due to interference in interface-2. In contrast, mutations in the loop between helix-D and helix-E (R101A, R101E) or in helix-E (V107A/M108A/I109A, I109R) exhibit severe recombination defects and poor formation of *attB* complex-2 (Figure 32). These phenotypes suggest that the residues contribute to the formation of cooperative interactions between B and B'. The mutant V107A/M108A/I109A binds normally to the *attP* half sites, P1 and P', as these do not rely on cooperative interactions.

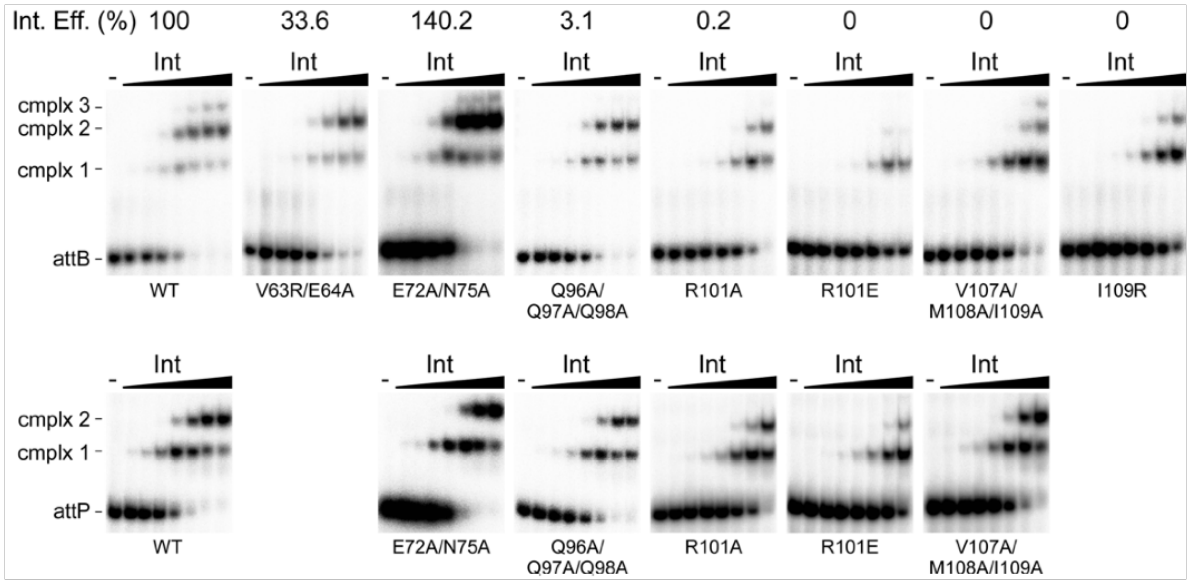


Figure 32: Core binding domain mutants affect *attB* cooperativity

Binding of Brujita Int mutants to a 44 bp *attB* substrate. The positions of unbound DNA (*attB*) and complexes (cmplx 1, cmplx 2, cmplx 3) are indicated. Int concentrations are from left to right, approximately, 0, 0.15, 0.61, 2.4, 9.8, 39, 156, and 625 nM.

Mutant R101E appears to interfere with the formation of subunit interactions between B and B', as it loses the ability to form *attB* complex-2. Additionally, it exhibits a reduced capacity to form complex-2 with *attP* DNA, likely due to decreased affinity, as *attP* binding does not rely on cooperative interactions. However, this residue is not positioned to interact with any other subunits in our model. Instead, R101 is positioned in the linker between helices-D and E, directed toward helix-C (Figure 33). It is possible that this residue stabilizes these helices and that the R101E mutant destabilizes them so they can no longer interact with neighboring subunits.

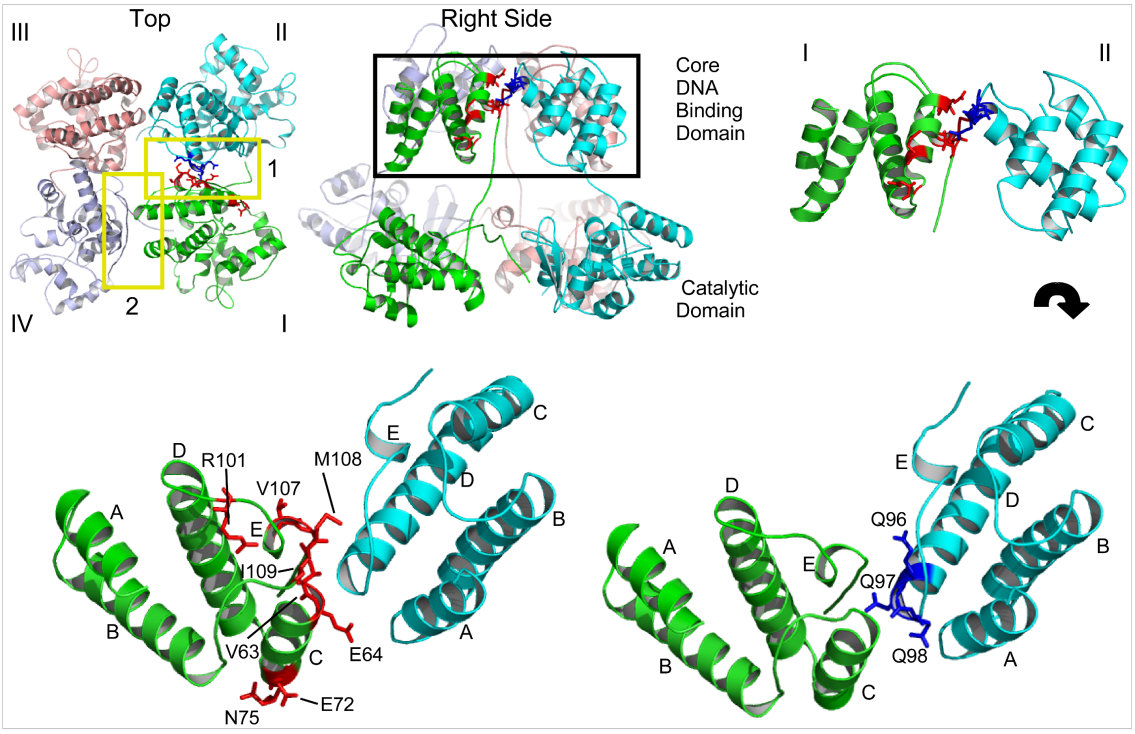


Figure 33: Mutants affect formation of interface-1

Top and side views of a tetramer of Brujita Int modeled to Lambda Int-Holliday Junction DNA (PDB: 1Z1G). A close up of two core-binding domains (right), showing potential interactions involving mutated residues M108, V63, E64, and Q96, Q97, Q98 on a neighboring molecule of Int. V107, E72, N75 also appear to effect protein-protein interactions.

We created an additional mutant in the inter-domain linker (P119W/R120E/P121W). This mutant shows a moderate reduction in binding to *attB* DNA, but a complete loss of recombination (Figure 34A). These residues are predicted to be in close proximity to bound DNA and may be interfering with DNA-protein contacts. However, it is not clear why this would alter recombination. It is possible that this mutant alters flexibility of the linker region needed for recombination dynamics.

4.3.6 Effects of Mutations in the Catalytic Domain

An additional thirteen mutations were designed to interrupt subunit interactions in the catalytic domain and were tested for integration and binding ability. Of the thirteen mutants, eight cause the protein to be strongly defective in recombination (Table 6), four of which are insoluble *in vitro* (M142H, G150Y/L151D, P184W, and F203E/P204W), likely due to protein misfolding. Three mutants located in the linker between helix-I and strand β-1 (D165A, I167A/E168A/R169A, and I167E/E168R/R169E) and two mutations located in helix-J (V190A/E191A and V190R/E191R) are situated at the bottom of the catalytic domain (Figure 34C). Two of these mutants (I167E/E168R/R169E and V190R/E191R) have slightly reduced binding affinity, but all bind normally to *attB*, suggesting that they form interaction-2. An intriguing hypothesis is that these residues are involved in *trans*-interactions with the extreme C-terminus of a neighboring subunit, much like the *trans*-interactions formed with the β-7 strand of the Lambda Int tetramer in synapsis (PDB: 1Z1G) (Figure 34C-E). One of these mutants (V190R/E191R) has a strong recombination defect and one (I167E/E168R/R169E) has a moderate defect, possibly due to interference with this *trans*-interaction. It is also plausible that the rearrangement we propose for positioning the catalytic tyrosine also facilitates normal binding, which may explain why these mutants exhibit moderate binding defects. The residues I167, E168, and R169 correspond to the region in Lambda that is involved in this *trans*-interaction.

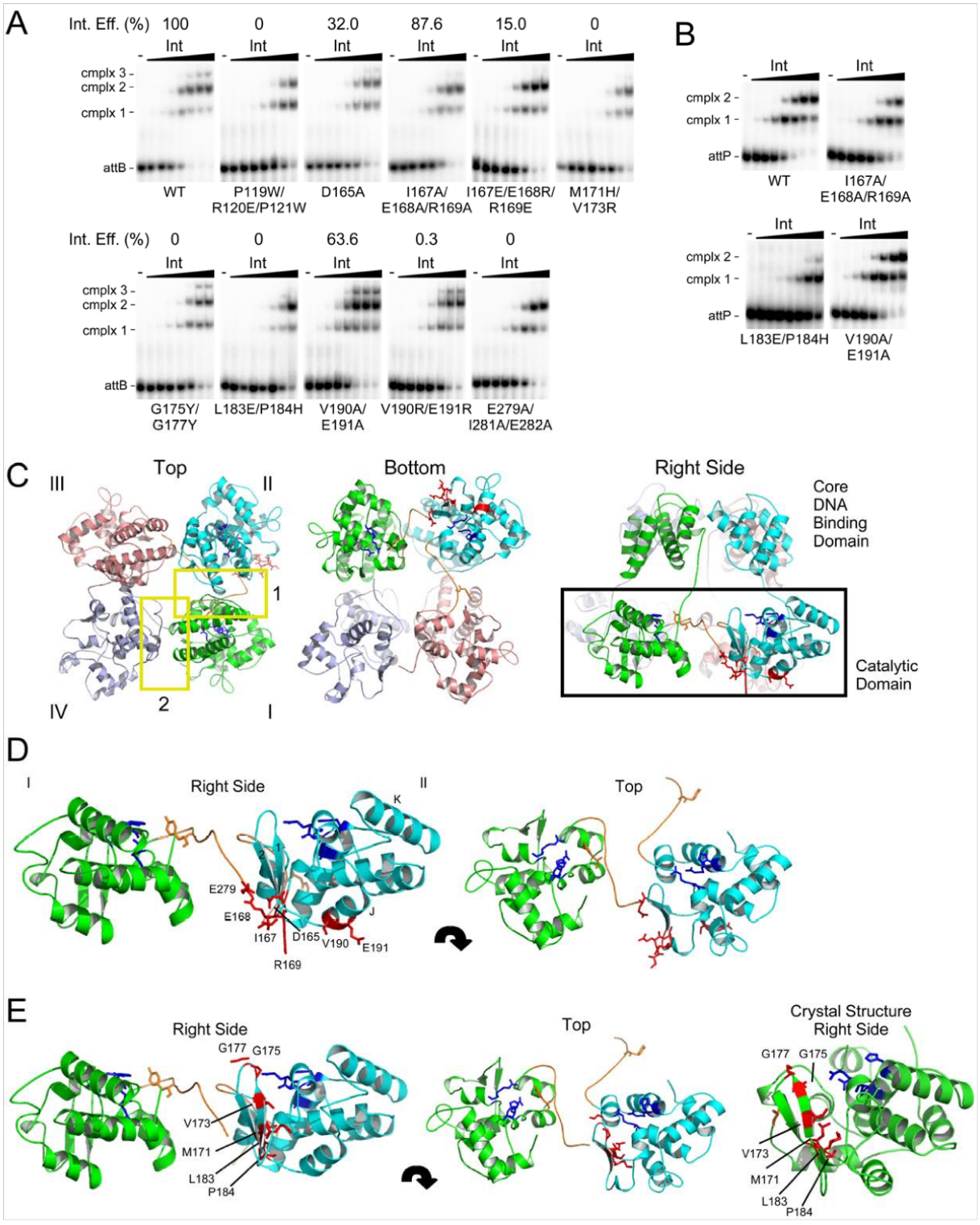


Figure 34: Brujita Int mutations affect catalysis

A. Binding of Brujita Int mutants to a 44 bp *attB* substrate. The positions of free DNA (*attB*) and complexes (cmplx 1, cmplx 2, cmplx 3) are indicated. Int concentrations are from left to right, 0, 0.15, 0.61, 2.4, 9.8, 39, 156, and 625 nM. **B.** Binding of Brujita Int mutants to an 89 bp *attP* substrate. The positions of free DNA (*attP*) and complexes (cmplx 1, cmplx 2) are indicated. Int concentrations are as in panel A. **C.** Top, bottom, and right side views of a tetramer of Brujita Int modeled to Lambda Int-Holliday junction DNA (PDB: 1Z1G). **D.** A close up of two catalytic domains, showing potential interactions involving mutated residues E279, E168, D165. Additional residues (280-305) were not included in the model output. Several residues exhibit interesting effects when mutated, but are not clearly involved in subunit interaction, recombination, or binding (V190, E191, I167, and R169). **E.** A close up of two catalytic domains modeled to Lambda Int-HJ DNA, (left and middle) showing residues with effects on recombination. The Brujita crystal structure (right) showing that these residues are not positioned near the active site. For C-E, active site tetrad (RHR) is shown in blue, the nucleophilic tyrosine (Y269) and the variable region are shown in orange, and the mutated residues in red.

Three mutations in strands β1 and β2 are of interest. The mutant G175Y/G177Y shows slightly reduced binding affinity, but forms all three complexes similarly to wild type *in vitro*, yet cannot facilitate recombination. These residues are not close to the active site – but are near the catalytic tyrosine – and their disruption may alter its flexibility or active site formation. Additionally, the M171H/V173R and L183E/P184H mutations cause a decrease in binding affinity to *attB* DNA, accompanied by poor *in vitro* complex-1 formation and a lack of complex-3. The L183E/P184H mutation forms complex-2 with *attB*, suggesting cooperative interactions can occur, yet fails to form *attP* complex-2. This may be a result of defective binding that does not inhibit interface formation. It is possible that these mutants are defective due to destabilizing necessary interactions within the catalytic domain, particularly those involving the active site. This is likely due to the importance of this β sheet region in DNA-binding and catalysis, as is the case in XerD (Subramanya *et al.*, 1997).

The final mutation, E279A/I281A/E282A is located in the variable region at the extreme C-terminus, which we propose rearranges upon DNA binding. This region is predicted by the

Lambda model to extend to a neighboring protomer, contributing to the catalytic domain of the neighbor protein, and may be interacting to form *trans*-subunit bridges to stabilize the active site.

4.4 DISCUSSION

The differences in organization of Brujita attachment sites present questions as to the structural differences between Brujita Int and the other characterized tyrosine integrases and recombinases. In this chapter, a structural comparison shows that Brujita Int is similar in its fold to XerD and Lambda Int, a feature that the larger family of tyrosine recombinases seems to share, even though they exhibit limited similarity at the sequence level. The greatest differences between the proteins are threefold: the presence of a small helix E in Brujita Int (not found in Lambda Int or XerD), an inter-domain linker in Brujita Int and Lambda Int (but not XerD), and a C-terminal variable region, which is differs among the three proteins. The variable region is particularly interesting in that it contains the nucleophilic tyrosine residue. Additionally, this region of Brujita folds much differently than either Lambda Int or XerD, and is not positioned to act in catalysis in the current structure.

Mutational analysis reveals several groups of residues that appear to be important for subunit interactions within a model of a Brujita Int tetramer. Additionally, there are many residues that effect binding and recombination for reasons not readily identifiable. Some of these mutants may destabilize critical interactions, which normally allow proper movement between active and inactive forms of Int, or that these mutants interfere with catalytic residues.

The role of the P1 half-site remains mysterious. As we have previously reported, it is required for recombination *in vivo* using *attB* x *attP* sites, though recombination can also occur

between *attB* x *attB* (Lunt & Hatfull, 2016). Thus it seems that P1 is only necessary in combination with P. This is presumably because Int does not readily bind the P half-site, as no binding is observed *in vitro*, and the P1 site may be required to facilitate binding at P (Lunt & Hatfull, 2016). The most obvious mechanism would be direct protein-protein interactions between Int bound to P1 and P. However, none of the mutants described here necessarily validate this hypothesis.

The structure of Brujita Int presented here contains two features that must be rearranged in order for the protein to act in binding and catalysis. This is an unusual characteristic when compared to the other tyrosine recombinases, though many of them have the benefit of DNA in their structures. The structure of XerD, which does not contain DNA, similarly requires one of these rearrangements (though in a different manner) requiring a separation of the two domains in order to bind DNA (Subramanya *et al.*, 1997). Brujita Int likewise necessitates a rotation of the two domains approximately 180° in order to create a DNA binding cavity. The C-terminal tail in Brujita undergoes a more striking rearrangement, we believe by changing its fold entirely, allowing it to position the catalytic tyrosine within the active site, and forming inter-subunit bridges within a synaptic tetramer, as has been shown in Lambda Int (Kwon *et al.*, 1997, Aihara *et al.*, 2003, Landy, 2015).

The question of the role of the variable region is intriguing. The variable region has been implicated in determining whether tyrosine mediated cleavage occurs in a *cis*- or *trans*-mechanism. In the case of FLP recombinase, an unusually long variable region may be partially responsible for FLP using *trans*-cleavage (Subramanya *et al.*, 1997, Landy, 1993). Brujita Int similarly has a longer than usual variable region, containing ~30 residues that are not present in Lambda Int. Considering the large evolutionary distance between the members of the tyrosine recombinase family, it is interesting that the most critical residue – the tyrosine nucleophile – is so

closely followed by this region that has unusually high variability (Subramanya *et al.*, 1997). This variability implies that the region is not under active selection to preserve a function, but mutation of these residues reveals that they do play roles in binding and catalysis. Perhaps the variable region is used as a method of controlling the activity of the catalytic residue, such as maintaining its inactivation until recognition of specific sequences, in order to prevent cleavage of non-specific DNA.

We observe a similar positioning of the C-terminal helices of Brujita Int as has been found in Lambda Int catalytic domain (Kwon *et al.*, 1997). In the structures of both these proteins, the extreme C-termini are not in position to catalyze recombination. In Lambda Int, this region appears to rearrange upon DNA-binding, placing the nucleophile into the proper position in two of the four Int proteins involved in the synaptic complex (Aihara *et al.*, 2003). The tyrosines in the remaining two proteins are not in position to act in catalysis, and are disordered in the crystal structure (Biswas *et al.*, 2005). After initial cleavage and strand exchange, the other two proteins become activated and their C-termini rearrange into position for catalysis (Landy, 2015, Grindley *et al.*, 2006). This C-terminus has been proposed as a mechanism for coordinating the activity of the proteins in recombination, ensuring that only the proper subunits are active (Landy, 2015). We propose that a similar system functions in Brujita Int and that the protein is held in an inactive form until DNA is introduced, whereupon a structural rearrangement occurs, allowing proper positioning of the tyrosine nucleophile within the active site. It is also conceivable that this C-terminus is involved in coordinating the activities of the Int subunits in synapsis. Some of the mutants we describe, such as M171H/V173R, G175Y/G177Y, and L183E/P184H may be interfering with the ability of the active tyrosine to relocate to its active position.

As discussed above, the extreme C-terminus of Brujita Int is likely to change conformations to reposition the nucleophile. However, Brujita Int not only contains ~30 additional residues in the variable region, but protein levels are controlled through degradation encoded by the tag at the C-terminus of the protein (Broussard *et al.*, 2013). Assuming this occurs, how does the relocation of the C-terminus into a *trans*-subunit interaction affect the activity of the degradation tag? Additionally, how does this tag affect the assembly of synapsis and recombination? These questions may provide further answers into the still unknown differences between Brujita Int and canonical tyrosine integrases.

5.0 CONCLUSIONS AND FUTURE PERSPECTIVES

5.1 SUMMARY OF THE BRUJITA INT SYSTEM

In this work, we have described the Brujita integration system – a model for a new subclass of tyrosine integrases – which function and are controlled differently than the canonical systems. Though recently discovered, this class has many family members, with 58 currently identified among the actinobacteriophages (http://www.phagesdb.org). We have shown here that Brujita Int has no N-terminal arm-binding domain and no arm-DNA sites. It also uses an unusually small *attP* site and a normal-sized *attB* (approximately 60 bp and 37 bp, respectively). Brujita Int does not require any cofactors, and is capable of integrative and excisive recombination without the use of a recombination directionality factor, such as Xis. This is due to the identity of the right side binding sites, P' and B', which results in no sequence differences between *attP* and *attR*, and *attB* and *attL*. Thus the integration and excision reactions are the same at the molecular level. This is a unique example among the bacteriophage tyrosine integrases, of an integrase not employing an RDF for directionality control (Khaleel *et al.*, 2011, Fogg *et al.*, 2014). However, a single mobile element recombinase has been reported which uses a similar method of controlling directionality by differential expression, though to our knowledge, this system does not use pair expression with protein degradation, as Brujita does (Miyazaki & van der Meer, 2013).

Integrases in the Brujita family contain C-terminal degradation tags, which have been shown to increase turnover of the protein, creating a convenient and simple method of controlling directionality – not by accessory proteins and DNA topology – but instead by coordinating Int expression and degradation (Broussard *et al.*, 2013). This system would respond similarly to that

of Lambda Int, where host protease levels regulate stability of the CII protein, which in turn regulates the lysogeny decision. In Brujita these factors are combined into Int, which responds to host protease levels, governs the lysogeny decision, and performs the recombination reaction independently of any other proteins.

We note that Brujita Int is much less efficient in recombination, both *in vivo* and *in vitro*, than its counterparts in L5 and Lambda Int (Lee & Hatfull, 1993, Broussard *et al.*, 2013, Nash, 1975), a feature that may represent an additional level of control in the absence of RDFs. The cause of this lower efficiency is unknown, but may be due to the low affinity for the B and P half sites.

5.1.1 Is Brujita Int an Integrase or Recombinase?

The lack of directionality control at the protein DNA level, and control of levels of Int accumulation present a new method of directionality control for a tyrosine integrase (Broussard & Hatfull, 2013, Broussard *et al.*, 2013, Fogg *et al.*, 2014). Without a traditional RDF and no arm-type interactions, Brujita is essentially a simple tyrosine recombinase, like XerC/D, FLP, and Cre. The combination of the C-terminal degradation tag and control of expression allow Brujita Int to function like an integrase with directionality control. One question still remains unanswered in this area; how is Brujita Int expressed from a lysogen? The best option we can identify is the predicted promoter P_{Int}, which occupies the 5' region of the *int* coding sequence. The activation of this promoter in a Brujita lysogen would produce Int^{Short}, which we have shown is capable of binding and recombination, though at lower efficiencies than Int^{Long}.

Additionally, we have shown that in an integrative plasmid containing Brujita Int is fully stable over at least 26 generations, but that the A317E stabilizing mutation is sufficient to destabilize it, allowing excision to occur at moderate levels. This supports the idea that Int

expression must be induced in the lysogen in order for excision to occur. This hypothesis must be further tested, as no activity has yet been shown by this promoter and it is not clear which promoter or which conditions stimulate this expression (Emily Shine, personal communication). However, a similar system has been employed by *Pseudomonas knackmussii* integrase IntB13 in the mobile element ICE*clc* (Miyazaki & van der Meer, 2013). In this system, *intB13* is expressed constitutively for integration, but upon integration, is separated from the promoter P_{circ}. Excision requires the use of a different promoter, P_{int}, from which expression occurs at low frequency (Miyazaki & van der Meer, 2013).

The implications of having two potential versions of Brujita Int – which both have binding and recombination activity – are interesting. This may be an artifact, but more likely implies some additional level of modulating the recombination reaction. The two versions differ by the N-terminal 21 residues, which are absent in Int^{Short}. These residues confer some effect on DNA binding, as Int^{Long} binds with higher affinity than Int^{Short}. Mutational analyses have shown that translation can occur from multiple start sites, so there is likely to be a mixed population of both forms in normal phage infection.

These features represent a fascinating system, one that is neither a true tyrosine integrase, nor a simple recombinase. Brujita Int exhibits features of both systems and likely represents an evolutionary precursor to more complex tyrosine integration systems, such as Lambda Int (Broussard & Hatfull, 2013, Broussard *et al.*, 2013).

5.1.2 Why is Int Promiscuous?

We have also shown that Int is capable of promiscuous recombination, or recombination using any combination of its sites (*attP* recombines with *attP*, and *attB* with *attB*, in addition to normal *attP*

x *attB* recombination). This is an odd feature and to our knowledge is the first tyrosine integrase that shows this promiscuity. The lack of site selectivity begs the question that if recombination can occur between *attB* and *attB*, why has there been selective pressure to develop a more complex system, in which two different DNA sites must be maintained? This conundrum is added to by the fact that *attP* is larger than *attB* and asymmetrical. For this system to have been maintained, it must give the phage some selective advantage over a system involving two identical sites. It is possible that there is some level of recognition or directionality control exhibited as a result of the difference between the two DNAs. However, because there are no arm-type interactions and no RDF, this would have to employ a previously uncharacterized method of control.

5.1.3 Identification of Interfaces Required for Synapsis

The Brujita system uses five attachment sites, B and B' in *attB*, and P1, P, and P' in *attP*. Binding occurs independently to P1, P' and B'. There are strong cooperative interactions between protomers at B and B', such that binding at B only occurs once B' is occupied. We have shown that this cooperativity is likely due to subunit-subunit interactions involving interface-1 and that several residues contribute to this interface including residues V107, M108, and I109.

Another interface must form between subunits on different DNAs, interface-2. We have shown that the Q96, Q97, Q98 triad appears to play a role in this interface, along with residues in helix-A of the neighboring Int subunit (R25, R27, and W29). Additional residues may be involved in interfaces in the catalytic domain, but these are less clear.

5.1.4 What is the Role of P1?

The Brujita *attP* is unusual in its small and asymmetrical nature. It contains an added binding site, P1, to the left of the core sites P and P'. We have demonstrated that P1 and P' are bound independently and that binding does not occur to P *in vitro*, though this site is required for recombination. Unlike other recombination systems, the fifth site, P1, is also required for recombination.

We propose that binding at the P1 site is required for recruitment of Int to the P half site. However, Int must also interact with DNA at the P half site to some degree, as the conserved 5'AG dinucleotide at P is critical for recombination. It seems probable that an Int subunit-subunit interaction allows the synaptic complex to form, followed by cleavage, strand exchange, and resolution.

The P1 and P sites are separated by approximately 12 bp, much more than the separation between P and P', or B and B', which is predicted to be 1-2 bp by mutational analysis (Figure 24). Thus the interaction between P1 and P must be dissimilar to interaction-1 (Figure 29). Predictive modeling of the *attP* DNA suggests that *attP* has a predisposition to bend between these two half sites. Furthermore, mutational analysis reveals that any deviation from the original spacing disrupts the interaction between P1 and P. Substitution of a thymine in the tetra-thymine region also appears to disrupt the interaction (Figure 22), possibly because this alters the natural bend of the DNA. We propose that this region of *attP* bends naturally and that this bend orients the CB domains of two Int subunits in close proximity, allowing the formation of a subunit-subunit interaction, possibly similar to interface-1. Alternatively, the proteins may form a new type of interface, such as one seen in the crystal contacts (Figure 30). Whichever interface is formed

appears to allow the stabilization or recruitment of Int to the P half site – where binding is much weaker – followed by assembly of synapsis and recombination (Figure 35).

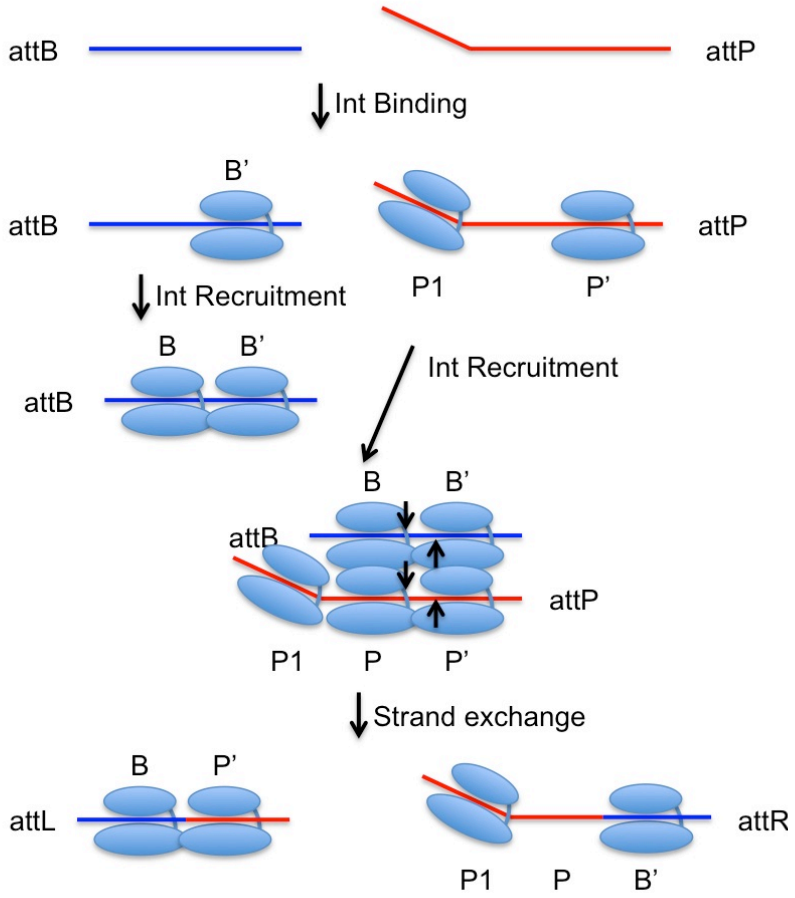


Figure 35: Model for Brujita Int recombination

Brujita Int binds to *attP* at P1 and P' and *attB* at B'. Binding at B' allows cooperative binding of Int to B. DNA sequence between P1 and P cause a bend in the DNA, allowing recruitment of Int to P, stabilized by subunit-subunit interactions between protomers at B, P1, and P' in synapsis. Strand exchange occurs and the complex is released, yielding recombinant products.

There are several experiments that may be used to test this model. First, a more thorough examination of the DNA in the bending region would clarify if the tetra-thymine region is important for bending or binding. If these four residues share the same effect on recombination, they are most likely important for bending, as they are positioned outside the P binding site and would not affect this interaction (see Figure 22). Additionally, we could confirm that this region has a DNA bend with and without Int by using the pBend system and measuring the bend *in gello* (Kim *et al.*, 1989, Zwieb & Adhya, 2009). If this bending is confirmed, FRET analysis could be used to verify that the two Int molecules form an interaction between the P1 and P half sites (Gabba *et al.*, 2014, Kempe *et al.*, 2016).

One of the most beneficial methods of clarifying this model would be to form a Brujita Int-DNA co-crystal. To date, attempts have been made using various sized *attB* sequences, suicide substrates, and engineered binding sites, with no success. Even though binding can be readily achieved to the B' half site, the low affinity of Int for the B site, even with cooperativity, makes obtaining a purified complex difficult. It appears that this interaction is too weak to allow Int dimerization on the *attB* site for long periods of time.

As co-crystallization has remained elusive so far, new strategies will be needed. In order to achieve this, modifications may need to be made to Int or the DNA. For instance, using a suicide DNA fragment containing a 3-base flap has been instrumental in crystallizing Lambda Int-DNA complexes (Aihara *et al.*, 2003). The presence of this flap allows cleavage to occur, creating a phosphotyrosine bond. The flap diffuses away, leaving DNA with no phosphate nucleophile to liberate the catalytic tyrosine. This method has been tested with Brujita, but additional changes to the oligos or binding conditions may be needed to allow cleavage of the DNA.

5.1.5 Int-DNA Contacts

Int-DNA binding appears to be dependent on the 5'AG dinucleotide present in each binding site. Though these are not the only positions that affect binding, mutation at any of these positions eliminates binding (in the case of B, B', P1, and P') and recombination (in the case of P). Thus it appears that these nucleotides are directly contacted by Int. Other nucleotides may also have protein contacts, but these do not exhibit the severe binding and recombination defects seen with this dinucleotide.

5.2 APPLICATIONS

The Brujita Int system represents a simple recombination system with directional control employed presumably by differential expression of two versions of Int, paired with protein degradation. Phage tyrosine integrases have been effective tools for biotechnology for decades (Fogg *et al.*, 2014, Groth & Calos, 2004b, Murphy, 2012). However, these are complicated and bulky systems that use large DNA sequences and two to three proteins for integration and excision. Much simpler options are FLP and Cre, which require no cofactors, but do not have directionality control. The Brujita system may be an effective alternative as Int can be expressed within the cell for integrating large DNA sequences into bacterial and eukaryotic chromosomes. Brujita Int functions well in *E. coli* and *M. smegmatis*, and we expect that it will also function in other systems.

Another use for the Brujita Int system is in synthetic biology, where biological computer circuits rely on bi-directional or directional recombinases such as Brujita Int (Bonnet *et al.*, 2012, Avlund *et al.*, 2009, Hasty *et al.*, 2002, Fogg *et al.*, 2014). Canonical tyrosine integrases have

shown little promise in this field due to their complexity, but the size and simplicity of Brujita Int may allow its effective use these circuits (Fogg *et al.*, 2014, Bonnet & Endy, 2013, Meinke *et al.*, 2016).

5.3 FUTURE DIRECTIONS

A complete understanding of the Brujita Int system lacks answers to several questions. First, a verification of the method used to express lysogenic form of Int used in excision is needed. Secondly, a more detailed understanding of the synaptic complex is required, particularly how Int bound to P1 and P interact and the specific protein-protein and DNA-protein contacts that are made. Finally, more work is needed to clarify the role of the C-terminal variable region and the catalytic rearrangement of Brujita Int. These questions would be greatly aided by additional protein and protein-DNA co-crystals.

This work has laid the foundation for understanding the function of a new subclass of tyrosine integrases, integrases of the Brujita Int family. These integrases represent a growing population with very different requirements than the canonical Lambda Int system. More work is needed to understand the Brujita Int family, yet already these proteins appear to have abundant uses as well as information regarding the evolution of the tyrosine recombinases.

APPENDIX A

PHAGE HUNTING WITH MYCOBACTERIUM MINNESOTENSE

Mycobacterium minnesotense is a recently discovered non-tubercule mycobacterium, found in sphagnum peat bogs in Minnesota (Hannigan *et al.*, 2013). It is a non-pathogenic member of the same group as the disease causing *M. tuberculosis*, *M. leprae*, and *M. ulcerans* and is a close relative of *M. arupense*, a potential human pathogen (Hannigan *et al.*, 2013). This bacterium is of interest to the Hatfull lab, because it is a member of the actinomycetes, which we seek to understand from an evolutionary perspective. In particular, we are interested in understanding how phages evolve in closely related bacteria, such as between *M. smegmatis* and *M. minnesotense.* To this end, we sought to develop *M. minnesotense* as a suitable host for phage hunting.

We acquired six *M. minnesotense* isolates as a kind gift from the Dahl lab. These isolates are classified as provisional BSL1 bacteria (DSM45633), and have potential for use in phage hunting projects as an alternative host. They grow slowly, requiring 5 days growth for liquid cultures and 7-10 days for plate cultures.

A.1 MATERIALS AND METHODS

A.1.1 Growth Conditions for *M. minnesotense*

M. minnesotense grows best at 30°C and isolates will not grow at 37°C (Hannigan *et al.*, 2013). *M. minnesotense* isolates were found to grow best on 7H11 media supplemented with OADC and cycloheximide. Plates were grown at 30°C. Individual colonies appear after 7 days, but are very small until 10 days. Colonies begin as a pale yellow, and slowly turn pink to salmon when grown in light

For liquid cultures, isolates grow best in 7H9 supplemented with OADC and cycloheximide. Growth in liquid media was performed by shaking vigorously (250 rpm) in baffled flasks at 30°C. Cultures reached saturation after 5-7 days.

A.1.2 Infection Tests Using *M. smegmatis* Phages

Phage high-titer lysates were diluted 10-fold to 10^{-5} in phage buffer. Spot tests were prepared by mixing 0.5 ml bacterial culture with 4.5 ml 1 x MBTA and plating on large 7H11 plates supplemented with OADC and cycloheximide. The plates were allowed to cool and 3 µl of each phage dilution was spotted on the top agar. The plates were incubated for 5-7 days before plaques began to appear.

A.1.3 Phage Plating

Soil or peat moss samples were added to 20 ml phage buffer with $CaCl_2$ and shaken vigorously for one hour. The samples were pelleted and the phage buffer was filter sterilized. Following sterilization, 3.5 ml of this phage buffer was added to 50 ml 7H9 with OADC, 10 μg/ml cycloheximide, 3 ml $CaCl_2$, and 500 μl *M. minnesotense* and shaken at 250 rpm, 30°C for 4-6 days. Enrichments were pelleted, filtered, and 50 μl of enrichment was added to 0.5 ml bacteria, mixed with 4.5 ml MBTA, and plated on 7H11 plates with OADC and cycloheximide.

A.2 RESULTS

A.2.1 Phage Hunting

M. minnesotense grows on 7H11 media and will not grow in the absence of albumin (Hannigan *et al.*, 2013). *M. minnesotense* bacteria are naturally non-clumpy and do not need to be grown in the presence of tween, although tween does not appear to inhibit growth. Plated colonies appear after 7 days, but are very small until ~10 days. Liquid cultures require 5 days to reach saturation and must be grown in baffled flasks with vigorous shaking. Little to no growth is observed without shaking or with shaking in non-baffled flasks.

Since these isolates were found in sphagnum peat bogs, soil and samples were used from commercially available sphagnum peat moss from two sources. Several soil samples and commercially available peat moss were used in enrichment cultures with each of the six isolates.

Many rounds of enrichment yielded a few plates that appeared to have very tiny pinprick plaques, however, these could not be isolated for reinfection.

A.2.2 Infection Using *M. smegmatis* Phages

Another question of interest to us is whether *M. smegmatis* phages are able to cross-infect *M. minnesotense* isolates, since these two bacteria are related. We selected, purified, and made high titer lysates of fifteen phages. These lysates were diluted and spotted on *M. smegmatis* and each of the six *M. minnesotense* isolates. Results are summarized in Table 7 and representative plates are shown in Figure 36.

Many of the phages tested show the ability to cause lysis of all of the *M. minnesotense* isolates at high concentrations (>10^8 pfu/ml). As shown in Table A1, phages D29, Bxz2, HelDan, Saintus, Alma, Hedgerow, BPs, and Halo are able to cause lysis of all or nearly all isolates. These phages are from different clusters, including A2, A3, A8, A9, B2, and G1. However, oddly, even closely related phages are incapable of causing this lysis, as is the case with KBG, an A1, Peaches, an A4, and Acadian, a B5. KBG results in some clearing, but only in three of the tested isolates and very poorly at maximum concentrations. This raises the question of what differences are present that allow some more distant or unrelated phages to lyse the same host, while phages that are fairly closely related cannot. Similarly, what differences have arisen to allow certain isolates to escape lysis, while others are susceptible, considering they are all highly related at the sequence-level? Some phages, such as T2 and T4, exhibit a resistance to lysis from without (Abedon, 2011, Visconti, 1953) and it is possible that some similar mechanism is at work here.

Interestingly, though lysis is seen in many of the spots, clearing is only observed at high phage concentrations and no individual plaques can be seen, suggesting lysis from without or other

abortive infection (Abedon, 2011). This is reasonable, especially when it is noted that many phages have distinct host preferences (Jacobs-Sera *et al.*, 2012) and considering the evolutionary distance between these two species, in which receptors may be sufficiently different to prevent successful infection. Other factors that may contribute to this failure to infect include CRISPR, DNA restriction, and immunity systems (Jacobs-Sera *et al.*, 2012), though due to the evolutionary distance between these bacteria, abortive infection is most likely a result of an incompatibility rather than immunity against infection. The clearing observed may be a result of enzymatic activity from the phage tails, such as the lysozyme activity of T4 gp5 (Arisaka *et al.*, 2003), or the endorhamnosidase activity of the P22 tails-pike (Goldenberg & King, 1982).

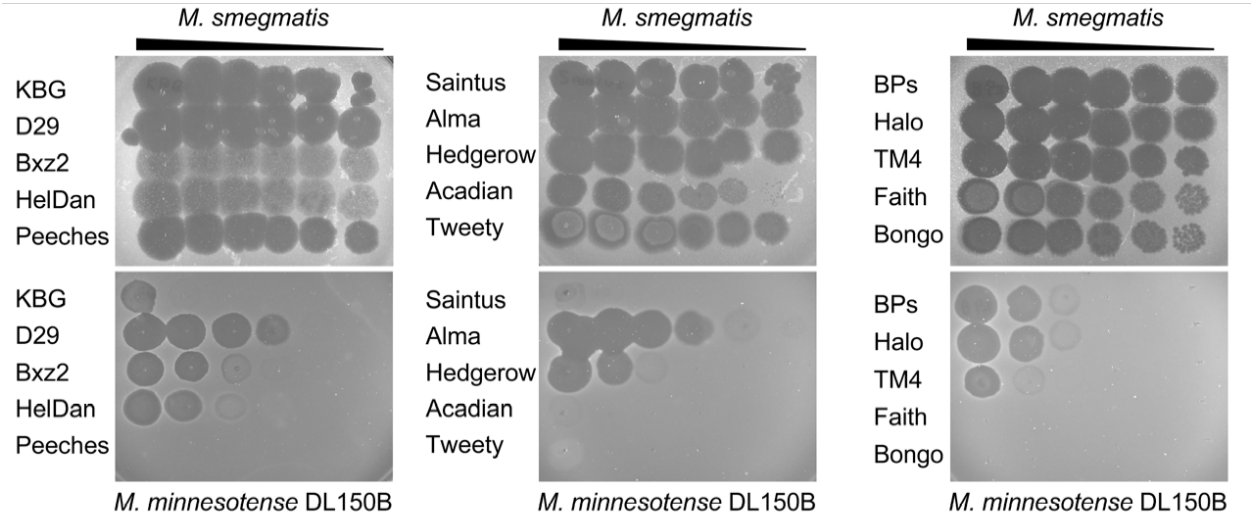


Figure 36: Lysis of *M. minnesotense* isolates by mycobacterium phages

Shown are dilutions of phages plated on *M. smegmatis* (top) or *M. minnesotense* (bottom) with phages dilutions from concentrated (left) to dilute (right). A summary of lytic ability is shown in Table 7.

Table 7. Infection patterns of various *M. smegmatis* phages.

Phages from several different clusters were tested for their ability to infect *M. minnesotense* isolates. Lytic ability is scored based on whether a zone of clearing was observed at high (+), medium (++), or low concentration (+++), or no clearing (-).

Phage	Cluster	Lytic Ability						
		M. smeg	DL47	DL49	DL150B	DL184	DL630	DL678A
KBG	A1	+++	-	+	+	-	-	+
D29	A2	+++	+	++	++	+	-	+
Bxz2	A3	+++	+	++	++	+	+	+
HelDan	A3	+++	+	++	++	+	++	++
Peaches	A4	+++	-	-	-	-	-	-
Saintus	A8	+++	+	+	+	+	+	+
Alma	A9	+++	++	++	+++	++	+	++
Hedgerow#	B2	+++	+	++	++	+	++	++
Acadian	B5	+++	-	-	-	-	-	-
Tweety	F1	+++	-	+	-	-	-	-
BPs	G1	+++	+	++	++	+	++	++
Halo	G1	+++	+	++	++	+	++	++
TM4	K2	+++	+	+	+	+	-	-
Faith	L2	+++	-	-	-	-	-	-
Bongo	M1	+++	-	-	-	-	-	-

A.3 CONCLUSIONS

M. minnesotense appears to be sufficiently different from *M. smegmatis* that phages cannot cross-infect, though many of the phages tested can cause clearing through abortive infection or lysis from without. As a result *M. minnesotense* would be a prime candidate for studying the evolution of related phages, assuming that infecting phages can be isolated. The conditions that were attempted were insufficient to find phages for *M. minnesotense*, but this could be due to lack of ideal soil samples (particularly those of the peat bogs from which the bacteria were isolated) or proper infection conditions. Particularly, varying levels of $CaCl_2$ or $MgCl_2$ may improve infection, or there may be some other metal ion or cofactor required for infection to take place.

APPENDIX B

BRUJITA INT-DNA CO-CRYSTALLIZATION

B.1 MATERIALS AND METHODS

B.1.1 Annealing of Oligos

Oligos were annealed as described previously in Chapter 3. Briefly, oligos were prepared for binding assays by resuspension in approximately 1 ml Annealing Buffer (10 mM Tris pH 7.5, 50 mM NaCl). Equal molar amounts of each single stranded oligo were combined and diluted 10-fold in Annealing Buffer. The oligos were then incubated in a boiling water bath for 5 minutes and allowed to cool gradually to room temperature. Oligos were vacuum concentrated and run over a size exclusion column, then concentrated in a spin concentrator prior to Int binding trials.

B.1.2 DNA Binding Reactions

The Int binding ability of each oligo was tested by mixing increasing concentrations of Int with 1 µl oligo at near equal molar ratios and buffer was added to a total volume of 10 µl. The mixture was incubated at room temperature for 20 minutes and run on a 3% agarose gel. Gels were visualized by ethidium bromide staining.

B.1.3 Buffers

Both Int and DNA oligos were dialyzed into buffer containing either 10 mM Tris pH 7.0, 250 mM NaCl, 5% glycerol, and 1 mM BME, or 10 mM Tris pH 7.0, 800 mM NaOAc, 5% glycerol, and 1 mM BME.

B.2 RESULTS

To better understand how synapsis and Int-DNA binding occur biochemically, we attempted to form Int-DNA co-crystals for X-ray diffraction. Many different preparations of stabilized Int (A317E), as well as stabilized, catalytically inactive Int (Y269A/A317E) were tested with a variety of oligos. To form Int-DNA crystals, we first needed oligos that form stable complexes with DNA. Binding reactions were examined for stability of the complex *in gello*, beginning with *attB35* derivatives. However, due to the weak binding nature of the B half site, all constructs similar to wild type *attB* failed to form stable complexes. We next examined constructs containing a single binding site (*attB1/2hairpin*, attB1*/2.6*, and *attBR7*). These constructs were more promising and formed relatively stable complexes. However, large scale preparation of these complexes results in rapid precipitation of Int, when mixed with DNA at high concentration. To avoid this, we increased the NaCl concentration in the buffer. A concentration of 400 mM-500 mM NaCl prevents precipitation, but effectively inhibits DNA-Int binding. We attempted to dialyze the complex into lower salt concentrations which would allow a more gradual change, hopefully stimulating binding once the salt concentration was low enough. However, Int is not stable below

200 mM NaCl, and this pattern held true for the DNA-Int mixture. Thus, stable binding could not be achieved with these oligos.

To achieve a more permanent interaction, we designed suicide oligos, containing a 3-base flap, similar to that previously described (Aihara *et al.*, 2003). This oligo is designed so that Int-mediated cleavage can occur, leaving a small, free tri-nucleotide to diffuse away. The 3-base flap then anneals in place of the diffused DNA, but since it lacks a free hydroxyl group, it cannot attack the phosphotyrosyl bond between Int and the DNA backbone. This creates a stable covalent linkage between Int and DNA, and may help to isolate the complex. Unfortunately, we were unable to get significant levels of cleavage *in vitro*. This may first require the development of an efficient *in vitro* recombination or cleavage reaction, which has been difficult to this point.

Since binding can be achieved with a single B' half site, we designed an oligo containing two half sites based on the sequence found at B' (*attB40*). Both sites should now bind DNA efficiently, and they presumably can also participate in cooperative binding, as occurs in wild type *attB*. Indeed, EMSAs using this oligo show increased formation of complex-2, while complex-1 is relatively reduced, indicating that this oligo participates in better complex-2 formation than our previously used DNAs. However, we still observe the rapid precipitation of Int when mixed with this oligo at high concentrations. Thermal shift assays revealed that Int is most stable in NaOAc, so we adjusted our buffers accordingly, though this did not alter the precipitation of Int when mixed with DNA. We resorted to mixing Int and DNA at low concentration in high salt (800 mM NaOAc) buffer, then using a spin concentrator (Millipore) to concentrate the mixture while buffer exchanging to lower salt (100 mM NaOAc). We observed a large loss in Int using this method, but it appeared to give us stable complex, so we proceeded to crystallization trials.

Crystallization trials were prepared using a twelve-standard condition screen prepared by the VanDemark Lab. Int-*attB40* complex at ~7 mg/ml was mixed with each condition in the sitting drop vapor diffusion method (Stevens, 2000). After several days, small, micro-, pseudo-crystals could be observed in a single well (100 mM Tris pH 8.5, 30% PEG 3350, 200 mM $MgCl_2$). Further analysis revealed that these were likely DNA crystals. Crystals were also observed in another condition (1 M Na Cacodylate pH 6.5, 50% PEG 8000, 2M NaOAc) with Int A317E. These may represent DNA crystals, as they have not been reproduced. Additional screens could not reveal any conditions forming promising initial Int-*attB40* crystals.

Table 8. Oligos for Int-DNA co-crystallization

Forward and reverse oligo sequences are shown with the predicted binding sites in blue.

Oligo Name	Sequence
attB35F	GAGCGATTCACTCGTAATGAATAGGTCGGGGGTTC
attB35R	GAACCCCCGACCTATTCATTACGAGTGAATCGCTC
attB35ATF	GAGCGATTCACTCGTAATGAATAGGTCGGGGGTTCA
attB35ATR	GAACCCCCGACCTATTCATTACGAGTGAATCGCTCT
attB35GCF	GAGCGATTCACTCGTAATGAATAGGTCGGGGGTTCG
attB35GCR	GAACCCCCGACCTATTCATTACGAGTGAATCGCTCC
attBR7F	AATGAATAGGTCGGGGGTTCGA
attBR7R	TCGAACCCCCGACCTATTCATT
attB1/2.6F	GTAATGAATAGGTCGGGGGTTCGA
attB1/2.6R	TCGAACCCCCGACCTATTCATTAC
attB1/2suicide1F	TCGAACCCCCGACCTATTCATTA
attB1/2suicide1R	TTACGAAGCGTAATGAATAGGTCGGGGGTTCGA
attBsuicide2F	TCGAACCCCCGACCTATTCA
attBsuicide2R	TCATTACGAAGCGTAATGAATAGGTCGGGGGTTCGA
attB1/2hairpin	TCGAACCCCCGACCTATTCATTACGAAGCGTAATGAATAGGTCGGGGGTTCGA
attB40F	TCGAACCCCCGACCTATTAATGAATAGGTCGGGGGAAGCT
attB40R	AGCTTCCCCCGACCTATTCATTAATAGGTCGGGGGTTCGA

B.3 CONCLUSIONS

Though we were unable to find any oligos that allow the formation of stable complexes, we have identified some features that would be important in oligos in future trials. First, the suicide substrates, though unsuccessful here, have the greatest potential for allowing stable complex formation. It is likely that the substrates or buffer conditions we used were not properly designed for *in vitro* cleavage. Future work in developing *in vitro* cleavage assays would give necessary information for the development of these oligos and conditions.

It appears that the B half site is far too weak to form stable DNA-Int complex, even in cooperation with the B' site. Thus, any oligos using multiple half sites would require an engineered B site to improve this binding. The oligo we developed, *attB40*, appears to be a modest improvement, but still does not form lasting stable complex with Int.

REFERENCES

Abbani, M.A., C.V. Papagiannis, M.D. Sam, D. Cascio, R.C. Johnson & R.T. Clubb, (2007) Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *P Natl Acad Sci USA* **104**: 2109-2114.

Abedon, S.T., (2011) Lysis from without. *Bacteriophage* **1**: 46-49.

Abràmoff, M.D., P.J. Magalhães & S.J. Ram, (2004) Image processing with ImageJ. *Biophotonics international* **11**: 36-42.

Abremski, K., B. Frommer & R.H. Hoess, (1986) Linking-number changes in the DNA substrate during Cre-mediated loxP site-specific recombination. *J Mol Biol* **192**: 17-26.

Abremski, K. & S. Gottesman, (1982) Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination. *J Biol Chem* **257**: 9658-9662.

Abremski, K. & R. Hoess, (1985) Phage P1 Cre-loxP site-specific recombination. Effects of DNA supercoiling on catenation and knotting of recombinant products. *J Mol Biol* **184**: 211-220.

Aihara, H., H.J. Kwon, S.E. Nunes-Duby, A. Landy & T. Ellenberger, (2003) A conformational switch controls the DNA cleavage activity of lambda integrase. *Mol Cell* **12**: 187-198.

Andrews, B.J., G.A. Proteau, L.G. Beatty & P.D. Sadowski, (1985) The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. *Cell* **40**: 795-803.

Arciszewska, L., I. Grainge & D. Sherratt, (1995) Effects of Holliday junction position on Xer-mediated recombination in vitro. *Embo J* **14**: 2651-2660.

Arciszewska, L.K. & D.J. Sherratt, (1995) Xer site-specific recombination in vitro. *Embo J* **14**: 2112-2120.

Argos, P., A. Landy, K. Abremski, J.B. Egan, E. Haggard-Ljungquist, R.H. Hoess, M.L. Kahn, B. Kalionis, S.V. Narayana, L.S. Pierson, 3rd & et al., (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. *Embo J* **5**: 433-440.

Arisaka, F., S. Kanamaru, P. Leiman & M.G. Rossmann, (2003) The tail lysozyme complex of bacteriophage T4. *Int J Biochem Cell Biol* **35**: 16-21.

Arnau, J., C. Lauritzen, G.E. Petersen & J. Pedersen, (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein expression and purification* **48**: 1-13.

Astromoff, A. & M. Ptashne, (1995) A variant of lambda repressor with an altered pattern of cooperative binding to DNA sites. *Proc Natl Acad Sci U S A* **92**: 8110-8114.

Austin, S., M. Ziese & N. Sternberg, (1981) A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell* **25**: 729-736.

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith & K. Struhl, (1996) *Current Protocols in Molecular Biology*. Wiley Intersciences, New York.

Avlund, M., I.B. Dodd, K. Sneppen & S. Krishna, (2009) Minimal gene regulatory circuits that can count like bacteriophage lambda. *J Mol Biol* **394**: 681-693.

Azaro, M.A. & A. Landy, (2002) λ integrase and the λ Int family. *Mobile DNA II* **118**.

Ball, C.A. & R.C. Johnson, (1991) Efficient excision of phage lambda from the Escherichia coli chromosome requires the Fis protein. *J Bacteriol* **173**: 4027-4031.

Bardarov, S., S. Bardarov Jr, Jr., M.S. Pavelka Jr, Jr., V. Sambandamurthy, M. Larsen, J. Tufariello, J. Chan, G. Hatfull & W.R. Jacobs Jr, Jr., (2002) Specialized transduction: an

efficient method for generating marked and unmarked targeted gene disruptions in Mycobacterium tuberculosis, M. bovis BCG and M. smegmatis. *Microbiology* **148**: 3007-3017.

Bardarov, S., Jr., H. Dou, K. Eisenach, N. Banaiee, S. Ya, J. Chan, W.R. Jacobs, Jr. & P.F. Riska, (2003) Detection and drug-susceptibility testing of M. tuberculosis from sputum samples using luciferase reporter phage: comparison with the Mycobacteria Growth Indicator Tube (MGIT) system. *Diagnostic microbiology and infectious disease* **45**: 53-61.

Bardarov, S., J. Kriakov, C. Carriere, S. Yu, C. Vaamonde, R.A. McAdam, B.R. Bloom, G.F. Hatfull & W.R. Jacobs, Jr., (1997) Conditionally replicating mycobacteriophages: a system for transposon delivery to Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **94**: 10961-10966.

Bernard, P., P. Gabant, E.M. Bahassi & M. Couturier, (1994) Positive-selection vectors using the F plasmid ccdB killer gene. *Gene* **148**: 71-74.

Bibb, L.A., M.I. Hancox & G.F. Hatfull, (2005) Integration and excision by the large serine recombinase phi Rv1 integrase. *Molecular Microbiology* **55**: 1896-1910.

Bibb, L.A. & G.F. Hatfull, (2002) Integration and excision of the Mycobacterium tuberculosis prophage-like element, phiRv1. *Mol Microbiol* **45**: 1515-1526.

Biswas, T., H. Aihara, M. Radman-Livaja, D. Filman, A. Landy & T. Ellenberger, (2005) A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* **435**: 1059-1066.

Blake, J.A., N. Ganguly & D.J. Sherratt, (1997) DNA sequence of recombinase-binding sites can determine Xer site-specific recombination outcome. *Mol Microbiol* **23**: 387-398.

Blakely, G., S. Colloms, G. May, M. Burke & D. Sherratt, (1991) Escherichia coli XerC recombinase is required for chromosomal segregation at cell division. *New Biol* **3**: 789-798.

Blakely, G., G. May, R. McCulloch, L.K. Arciszewska, M. Burke, S.T. Lovett & D.J. Sherratt, (1993) Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12. *Cell* **75**: 351-361.

Bloch, S., B. Nejman-Falenczyk, J.M. Los, S. Baranska, K. Lepek, A. Felczykowska, M. Los, G. Wegrzyn & A. Wegrzyn, (2013) Genes from the exo-xis region of lambda and Shiga toxin-converting bacteriophages influence lysogenization and prophage induction. *Arch Microbiol* **195**: 693-703.

Bonnefoy, E. & J. Rouviere-Yaniv, (1992) HU, the major histone-like protein of E. coli, modulates the binding of IHF to oriC. *Embo J* **11**: 4489-4496.

Bonnet, J. & D. Endy, (2013) Switches, switches, every where, in any drop we drink. *Mol Cell* **49**: 232-233.

Bonnet, J., P. Subsoontorn & D. Endy, (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc Natl Acad Sci U S A* **109**: 8884-8889.

Broach, J.R., V.R. Guarascio & M. Jayaram, (1982) Recombination within the yeast plasmid 2mu circle is site-specific. *Cell* **29**: 227-234.

Broussard, G.W. & G.F. Hatfull, (2013) Evolution of genetic switch complexity. *Bacteriophage* **3**: e24186.

Broussard, G.W., L.M. Oldfield, V.M. Villanueva, B.L. Lunt, E.E. Shine & G.F. Hatfull, (2013) Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches. *Mol Cell* **49**: 237-248.

Brown, K.L., G.J. Sarkis, C. Wadsworth & G.F. Hatfull, (1997) Transcriptional silencing by the mycobacteriophage L5 repressor. *Embo J* **16**: 5914-5921.

Campbell, A., (1994) Comparative molecular biology of lambdoid phages. *Annu Rev Microbiol* **48**: 193-222.

Campbell, A.M., (1992) Chromosomal insertion sites for phages and plasmids. *J Bacteriol* **174**: 7495-7499.

Cao, Y., B. Hallet, D.J. Sherratt & F. Hayes, (1997) Structure-function correlations in the XerD site-specific recombinase revealed by pentapeptide scanning mutagenesis. *J Mol Biol* **274**: 39-53.

Casjens, S., G. Hatfull, and R. Hendrix, (1992) *Evolution of dsDNA tailed-bacteriophage genomes*. HARCOURT BRACE JOVANOVICH.

Casjens, S.R. & R.W. Hendrix, (2015) Bacteriophage lambda: Early pioneer and still relevant. *Virology* **479-480**: 310-330.

Chen, J.W., J. Lee & M. Jayaram, (1992) DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell* **69**: 647-658.

Chen, Y., U. Narendra, L.E. Iype, M.M. Cox & P.A. Rice, (2000) Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell* **6**: 885-897.

Colloms, S.D., R. McCulloch, K. Grant, L. Neilson & D.J. Sherratt, (1996) Xer-mediated site-specific recombination in vitro. *Embo J* **15**: 1172-1181.

Colloms, S.D., P. Sykora, G. Szatmari & D.J. Sherratt, (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.

Corbett, E.L., C.J. Watt, N. Walker, D. Maher, B.G. Williams, M.C. Raviglione & C. Dye, (2003) The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* **163**: 1009-1021.

Court, D.L., A.B. Oppenheim & S.L. Adhya, (2007) A new look at bacteriophage lambda genetic networks. *J Bacteriol* **189**: 298-304.

Cox, M.M., (1983) The FLP protein of the yeast 2-microns plasmid: expression of a eukaryotic genetic recombination system in Escherichia coli. *Proc Natl Acad Sci U S A* **80**: 4223-4227.

Craig, N.L. & H.A. Nash, (1984) E. coli integration host factor binds to specific sites in DNA. *Cell* **39**: 707-716.

Crothers, D.M., T.E. Haran & J.G. Nadeau, (1990) Intrinsically bent DNA. *J Biol Chem* **265**: 7093-7096.

Cui, L., I. Murchland, I.B. Dodd & K.E. Shearwin, (2013a) Bacteriophage lambda repressor mediates the formation of a complex enhancer-like structure. *Transcription* **4**: 201-205.

Cui, L., I. Murchland, K.E. Shearwin & I.B. Dodd, (2013b) Enhancer-like long-range transcriptional activation by lambda CI-mediated DNA looping. *Proc Natl Acad Sci U S A* **110**: 2922-2927.

Deighan, P. & A. Hochschild, (2007) The bacteriophage lambdaQ anti-terminator protein regulates late gene expression as a stable component of the transcription elongation complex. *Mol Microbiol* **63**: 911-920.

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

Dubnau, D. & R. Losick, (2006) Bistability in bacteria. *Mol Microbiol* **61**: 564-572.

Dupont, L., B. Boizet-Bonhoure, M. Coddeville, F. Auvray & P. Ritzenthaler, (1995) Characterization of genetic elements required for site-specific integration of Lactobacillus delbrueckii subsp. bulgaricus bacteriophage mv4 and construction of an integration-proficient vector for Lactobacillus plantarum. *J Bacteriol* **177**: 586-595.

Echols, H., (1971) Regulation of lytic development. *Cold Spring Harbor Monograph Archive* **2**: 247-270.

Esposito, D., J.S. Thrower & J.J. Scocca, (2001) Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation. *Nucleic Acids Res* **29**: 3955-3964.

Evans, B.R., J.W. Chen, R.L. Parsons, T.K. Bauer, D.B. Teplow & M. Jayaram, (1990) Identification of the active site tyrosine of Flp recombinase. Possible relevance of its location to the mechanism of recombination. *J Biol Chem* **265**: 18504-18510.

Feher, A., P. Boross, T. Sperka, S. Oroszlan & J. Tozser, (2004) Expression of the murine leukemia virus protease in fusion with maltose-binding protein in Escherichia coli. *Protein expression and purification* **35**: 62-68.

Fogg, P.C., S. Colloms, S. Rosser, M. Stark & M.C. Smith, (2014) New applications for phage integrases. *J Mol Biol* **426**: 2703-2716.

Franklin, N.C., (1974) Altered reading of genetic signals fused to the N operon of bacteriophage lambda: genetic evidence for modification of polymerase by the protein product of the N gene. *J Mol Biol* **89**: 33-48.

Friedman, D.I. & D.L. Court, (1995) Transcription antitermination: the lambda paradigm updated. *Mol Microbiol* **18**: 191-200.

Friedman, D.I. & M. Gottesman, (1983) Lytic mode of lambda development. *Cold Spring Harbor Monograph Archive* **13**: 21-51.

Furth, M., S. Wickner, R. Hendrix, J. Roberts, F. Stahl & R. Weisberg, (1983) Lambda II.

Gabba, M., S. Poblete, T. Rosenkranz, A. Katranidis, D. Kempe, T. Zuchner, R.G. Winkler, G. Gompper & J. Fitter, (2014) Conformational state distributions and catalytically relevant dynamics of a hinge-bending enzyme studied by single-molecule FRET and a coarse-grained simulation. *Biophys J* **107**: 1913-1923.

Gabriel, K., H. Schmid, U. Schmidt & H. Rausch, (1995) The actinophage RP3 DNA integrates site-specifically into the putative tRNA(Arg)(AGG) gene of Streptomyces rimosus. *Nucleic Acids Res* **23**: 58-63.

Gardner, J.F. & H.A. Nash, (1986) Role of Escherichia coli IHF protein in lambda site-specific recombination. A mutational analysis of binding sites. *J Mol Biol* **191**: 181-189.

Ghosh, K., F. Guo & G.D. Van Duyne, (2007) Synapsis of loxP sites by Cre recombinase. *J Biol Chem* **282**: 24004-24016.

Ghosh, K., C.K. Lau, F. Guo, A.M. Segall & G.D. Van Duyne, (2005) Peptide trapping of the Holliday junction intermediate in Cre-loxP site-specific recombination. *J Biol Chem* **280**: 8290-8299.

Ghosh, P., A.I. Kim & G.F. Hatfull, (2003) The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol Cell* **12**: 1101-1111.

Ghosh, P., L.R. Wasil & G.F. Hatfull, (2006) Control of phage Bxb1 excision by a novel recombination directionality factor. *Plos Biol* **4**: 964-974.
Gibb, B., K. Gupta, K. Ghosh, R. Sharp, J. Chen & G.D. Van Duyne, (2010) Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Res* **38**: 5817-5832.
Goldenberg, D. & J. King, (1982) Trimeric intermediate in the in vivo folding and subunit assembly of the tail spike endorhamnosidase of bacteriophage P22. *Proc Natl Acad Sci U S A* **79**: 3403-3407.
Gopaul, D.N., F. Guo & G.D. Van Duyne, (1998) Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *Embo J* **17**: 4175-4187.
Gottesman, M., A. Oppenheim & D. Court, (1982) Retroregulation: control of gene expression from sites distal to the gene. *Cell* **29**: 727-728.
Gottfried, P., E. Yagil & M. Kolot, (2000) Core-binding specificity of bacteriophage integrases. *Mol Gen Genet* **263**: 619-624.
Grindley, N.D., K.L. Whiteson & P.A. Rice, (2006) Mechanisms of site-specific recombination. *Annu Rev Biochem* **75**: 567-605.
Gronostajski, R.M. & P.D. Sadowski, (1985) The FLP recombinase of the Saccharomyces cerevisiae 2 microns plasmid attaches covalently to DNA via a phosphotyrosyl linkage. *Mol Cell Biol* **5**: 3274-3279.
Groth, A.C. & M.P. Calos, (2004a) Phage integrases: biology and applications. *Journal of Molecular Biology* **335**: 667-678.
Groth, A.C. & M.P. Calos, (2004b) Phage Integrases: Biology and Applications. *Journal of Molecular Biology* **335**: 667-678.
Guo, F., D.N. Gopaul & G.D. van Duyne, (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**: 40-46.
Guzman, L.M., D. Belin, M.J. Carson & J. Beckwith, (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.
Halder, S., A.B. Datta & P. Parrack, (2007) Probing the antiprotease activity of lambdaCIII, an inhibitor of the Escherichia coli metalloprotease HflB (FtsH). *J Bacteriol* **189**: 8130-8138.
Hamilton, D.L. & K. Abremski, (1984) Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites. *J Mol Biol* **178**: 481-486.
Han, Y.W., R.I. Gumport & J.F. Gardner, (1993) Complementation of bacteriophage lambda integrase mutants: evidence for an intersubunit active site. *Embo J* **12**: 4577-4584.
Hannigan, G.D., B. Krivogorsky, D. Fordice, J.B. Welch & J.L. Dahl, (2013) Mycobacterium minnesotense sp. nov., a photochromogenic bacterium isolated from sphagnum peat bogs. *International journal of systematic and evolutionary microbiology* **63**: 124-128.
Hasty, J., D. McMillen & J.J. Collins, (2002) Engineered gene circuits. *Nature* **420**: 224-230.
Hatfull, G.F., L. Barsom, L. Chang, M. Donnelly-Wu, M.H. Lee, M. Levin, C. Nesbit & G.J. Sarkis, (1994) Bacteriophages as tools for vaccine development. *Dev Biol Stand* **82**: 43-47.
Hatfull, G.F., D. Jacobs-Sera, J.G. Lawrence, W.H. Pope, D.A. Russell, C.-C. Ko, R.J. Weber, M.C. Patel, K.L. Germane & R.H. Edgar, (2010) Comparative Genomic Analysis of 60 Mycobacteriophage Genomes: Genome Clustering, Gene Acquisition, and Gene Size. *Journal of Molecular Biology* **397**: 119-143.
Hauser, M.A. & J.J. Scocca, (1992a) Site-specific integration of the Haemophilus influenzae bacteriophage HP1: location of the boundaries of the phage attachment site. *J Bacteriol* **174**: 6674-6677.

Hauser, M.A. & J.J. Scocca, (1992b) Site-specific integration of the Haemophilus influenzae bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. *J Biol Chem* **267**: 6859-6864.

Hayashi, T., H. Matsumoto, M. Ohnishi & Y. Terawaki, (1993) Molecular analysis of a cytotoxin-converting phage, phi CTX, of Pseudomonas aeruginosa: structure of the attP-cos-ctx region and integration into the serine tRNA gene. *Mol Microbiol* **7**: 657-667.

Hayes, F. & D.J. Sherratt, (1997) Recombinase binding specificity at the chromosome dimer resolution site dif of Escherichia coli. *J Mol Biol* **266**: 525-537.

Hendrix, R.W., (2003) Bacteriophage genomics. *Curr Opin Microbiol* **6**: 506-511.

Herskowitz, I. & D. Hagen, (1980) The lysis-lysogeny decision of phage lambda: explicit programming and responsiveness. *Annu Rev Genet* **14**: 399-445.

Hickman, A.B., S. Waninger, J.J. Scocca & F. Dyda, (1997) Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. *Cell* **89**: 227-237.

Hiraga, S., A. Jaffe, T. Ogura, H. Mori & H. Takahashi, (1986) F plasmid ccd mechanism in Escherichia coli. *J Bacteriol* **166**: 100-104.

Hochschild, A., J. Douhan, 3rd & M. Ptashne, (1986) How lambda repressor and lambda Cro distinguish between OR1 and OR3. *Cell* **47**: 807-816.

Hoess, R.H. & K. Abremski, (1984) Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc Natl Acad Sci U S A* **81**: 1026-1029.

Hoess, R.H. & K. Abremski, (1985) Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J Mol Biol* **181**: 351-362.

Hoess, R.H., M. Ziese & N. Sternberg, (1982) P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc Natl Acad Sci U S A* **79**: 3398-3402.

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

Huff, J., A. Czyz, R. Landick & M. Niederweis, (2010) Taking phage integration to the next level as a genetic tool for mycobacteria. *Gene* **468**: 8-19.

Inouye, S., M.G. Sunshine, E.W. Six & M. Inouye, (1991) Retronphage phi R73: an E. coli phage that contains a retroelement and integrates into a tRNA gene. *Science* **252**: 969-971.

Jacobs-Sera, D., L.J. Marinelli, C. Bowman, G.W. Broussard, C. Guerrero Bustamante, M.M. Boyle, Z.O. Petrova, R.M. Dedrick, W.H. Pope, G. Science Education Alliance Phage Hunters Advancing, P. Evolutionary Science Sea-Phages, R.L. Modlin, R.W. Hendrix & G.F. Hatfull, (2012) On the nature of mycobacteriophage diversity and host preference. *Virology* **434**: 187-201.

Jayaram, M., C.H. Ma, A.H. Kachroo, P.A. Rowley, P. Guga, H.F. Fan & Y. Voziyanov, (2015) An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective. *Microbiol Spectr* **3**.

Kempe, D., M. Cerminara, S. Poblete, A. Schone, M. Gabba & J. Fitter, (2016) Single-Molecule FRET Measurements in Additive-Enriched Aqueous Solutions. *Anal Chem*.

Khaleel, T., E. Younger, A.R. McEwan, A.S. Varghese & M.C. Smith, (2011) A phage protein that binds phiC31 integrase to switch its directionality. *Mol Microbiol* **80**: 1450-1463.

Kim, J., C. Zwieb, C. Wu & S. Adhya, (1989) Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene* **85**: 15-23.

Kim, S. & A. Landy, (1992) Lambda Int protein bridges between higher order complexes at two distant chromosomal loci attL and attR. *Science* **256**: 198-203.

Kitts, P.A. & H.A. Nash, (1987) Homology-dependent interactions in phage lambda site-specific recombination. *Nature* **329**: 346-348.

Knight, K.L., J.U. Bowie, A.K. Vershon, R.D. Kelley & R.T. Sauer, (1989) The Arc and Mnt repressors. A new class of sequence-specific DNA-binding protein. *J Biol Chem* **264**: 3639-3642.

Kobiler, O., S. Koby, D. Teff, D. Court & A.B. Oppenheim, (2002) The phage lambda CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis. *P Natl Acad Sci USA* **99**: 14964-14969.

Kolot, M. & E. Yagil, (1994) Position and direction of strand exchange in bacteriophage HK022 integration. *Mol Gen Genet* **245**: 623-627.

Kuempel, P.L., J.M. Henson, L. Dircks, M. Tecklenburg & D.F. Lim, (1991) dif, a recA-independent recombination site in the terminus region of the chromosome of Escherichia coli. *New Biol* **3**: 799-811.

Kwon, H.J., R. Tirumalai, A. Landy & T. Ellenberger, (1997) Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276**: 126-131.

Landy, A., (1993) Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Curr Opin Genet Dev* **3**: 699-707.

Landy, A., (2015) The lambda Integrase Site-specific Recombination Pathway. *Microbiol Spectr* **3**: MDNA3-0051-2014.

Landy, A. & W. Ross, (1977) Viral integration and excision: structure of the lambda att sites. *Science* **197**: 1147-1160.

Lederberg, E.M. & J. Lederberg, (1953) Genetic Studies of Lysogenicity in Escherichia Coli. *Genetics* **38**: 51-64.

Lee, J., I. Whang & M. Jayaram, (1996) Assembly and orientation of Flp recombinase active sites on two-, three- and four-armed DNA substrates: implications for a recombination mechanism. *J Mol Biol* **257**: 532-549.

Lee, M.H. & G.F. Hatfull, (1993) Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. *J Bacteriol* **175**: 6836-6841.

Lee, M.H., L. Pascopella, W.R. Jacobs, Jr. & G.F. Hatfull, (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, D., P. Le, C. Zurla, L. Finzi & S. Adhya, (2011) Multilevel autoregulation of lambda repressor protein CI by DNA looping in vitro. *Proc Natl Acad Sci U S A* **108**: 14807-14812.

Lewis, J.A. & G.F. Hatfull, (2003) Control of directionality in L5 integrase-mediated site-specific recombination. *Journal of Molecular Biology* **326**: 805-821.

Little, J.W., (2010) Evolution of complex gene regulatory circuits by addition of refinements. *Curr Biol* **20**: R724-734.

Lunt, B.L. & G.F. Hatfull, (2016) Brujita Integrase: A Simple, Arm-Less, Directionless, and Promiscuous Tyrosine Integrase System. *J Mol Biol* **428**: 2289-2306.

Marinelli, L.J., M. Piuri, Z. Swigonova, A. Balachandran, L.M. Oldfield, J.C. van Kessel & G.F. Hatfull, (2008) BRED: A Simple and Powerful Tool for Constructing Mutant and Recombinant Bacteriophage Genomes. *Plos One* **3**: -.

Marr, M.T., S.A. Datwyler, C.F. Meares & J.W. Roberts, (2001) Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. *Proc Natl Acad Sci U S A* **98**: 8972-8978.

Meinke, G., A. Bohm, J. Hauber, M.T. Pisabarro & F. Buchholz, (2016) Cre Recombinase and Other Tyrosine Recombinases. *Chem Rev*.

Midonet, C. & F.X. Barre, (2014) Xer Site-Specific Recombination: Promoting Vertical and Horizontal Transmission of Genetic Information. *Microbiol Spectr* **2**.

Miller, H.I., A. Kikuchi, H.A. Nash, R.A. Weisberg & D.I. Friedman, (1979) Site-specific recombination of phage lambda: the role of host gene products *Cold Spring Harb. Symp. Quant. Biol.* **43**: 1121-1126.

Miyazaki, R. & J.R. van der Meer, (2013) A new large-DNA-fragment delivery system based on integrase activity from an integrative and conjugative element. *Appl Environ Microbiol* **79**: 4440-4447.

Mizuuchi, M. & K. Mizuuchi, (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. & K. Mizuuchi, (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., S. Kim & A. Landy, (1989) DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. *Science* **244**: 1457-1461.

Morris, P., L.J. Marinelli, D. Jacobs-Sera, R.W. Hendrix & G.F. Hatfull, (2008) Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. *J Bacteriol* **190**: 2172-2182.

Munteanu, M.G., K. Vlahovicek, S. Parthasarathy, I. Simon & S. Pongor, (1998) Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem Sci* **23**: 341-347.

Murphy, K.C., (2012) Phage recombinases and their applications. *Advances in virus research* **83**: 367-414.

Nash, H.A., (1975) Integrative recombination of bacteriophage lambda DNA in vitro. *Proc Natl Acad Sci U S A* **72**: 1072-1076.

Nash, H.A. & C.A. Robertson, (1989) Heteroduplex substrates for bacteriophage lambda site-specific recombination: cleavage and strand transfer products. *Embo J* **8**: 3523-3533.

Numrych, T.E., R.I. Gumport & J.F. Gardner, (1991) A genetic analysis of Xis and FIS interactions with their binding sites in bacteriophage lambda. *J Bacteriol* **173**: 5954-5963.

Nunes-Duby, S.E., L. Matsumoto & A. Landy, (1987) Site-specific recombination intermediates trapped with suicide substrates. *Cell* **50**: 779-788.

Nunes-Duby, S.E., L. Matsumoto & A. Landy, (1989) Half-att site substrates reveal the homology independence and minimal protein requirements for productive synapsis in lambda excisive recombination. *Cell* **59**: 197-206.

Nunes-Duby, S.E., R.S. Tirumalai, L. Dorgai, E. Yagil, R.A. Weisberg & A. Landy, (1994) Lambda integrase cleaves DNA in cis. *Embo J* **13**: 4421-4430.

Onaga, L.A., (2014) Ray Wu as Fifth Business: Deconstructing collective memory in the history of DNA sequencing. *Stud Hist Philos Biol Biomed Sci* **46**: 1-14.

Oppenheim, A.B., O. Kobiler, J. Stavans, D.L. Court & S. Adhya, (2005) Switches in bacteriophage lambda development. *Annu Rev Genet* **39**: 409-429.

Owens, R.M., F.F. Hsu, B.C. VanderVen, G.E. Purdy, E. Hesteande, P. Giannakas, J.C. Sacchettini, J.D. McKinney, P.J. Hill, J.T. Belisle, B.A. Butcher, K. Pethe & D.G. Russell, (2006) M. tuberculosis Rv2252 encodes a diacylglycerol kinase involved in the biosynthesis of phosphatidylinositol mannosides (PIMs). *Mol Microbiol* **60**: 1152-1163.

Papp, I., L. Dorgai, P. Papp, E. Jonas, F. Olasz & L. Orosz, (1993) The bacterial attachment site of the temperate Rhizobium phage 16-3 overlaps the 3' end of a putative proline tRNA gene. *Mol Gen Genet* **240**: 258-264.

Pargellis, C.A., S.E. Nunes-Duby, L.M. de Vargas & A. Landy, (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.

Parisien, A., B. Allain, J. Zhang, R. Mandeville & C.Q. Lan, (2008) Novel alternatives to antibiotics: bacteriophages, bacterial cell wall hydrolases, and antimicrobial peptides. *J Appl Microbiol* **104**: 1-13.

Payne, K., Q. Sun, J. Sacchettini & G.F. Hatfull, (2009) Mycobacteriophage Lysin B is a novel mycolylarabinogalactan esterase. *Mol Microbiol* **73**: 367-381.

Payne, K.M. & G.F. Hatfull, (2012) Mycobacteriophage endolysins: diverse and modular enzymes with multiple catalytic activities. *Plos One* **7**: e34052.

Pedulla, M.L., M.E. Ford, J.M. Houtz, T. Karthikeyan, C. Wadsworth, J.A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N.R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J.G. Lawrence, W.R. Jacobs, R.W. Hendrix & G.F. Hatfull, (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**: 171-182.

Pedulla, M.L., M.H. Lee, D.C. Lever & G.F. Hatfull, (1996) A novel host factor for integration of mycobacteriophage L5. *Proc Natl Acad Sci U S A* **93**: 15411-15416.

Pena, C.E., J.M. Kahlenberg & G.F. Hatfull, (1998) The role of supercoiling in mycobacteriophage L5 integrative recombination. *Nucleic Acids Res* **26**: 4012-4018.

Pena, C.E., J.M. Kahlenberg & G.F. Hatfull, (1999) Protein-DNA complexes in mycobacteriophage L5 integrative recombination. *J Bacteriol* **181**: 454-461.

Pena, C.E., J.M. Kahlenberg & G.F. Hatfull, (2000) Assembly and activation of site-specific recombination complexes. *Proc Natl Acad Sci U S A* **97**: 7760-7765.

Pena, C.E., M.H. Lee, M.L. Pedulla & G.F. Hatfull, (1997a) Characterization of the mycobacteriophage L5 attachment site, attP. *Journal of Molecular Biology* **266**: 76-92.

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

Pena, C.E., J.E. Stoner & G.F. Hatfull, (1996) Positions of strand exchange in mycobacteriophage L5 integration and characterization of the attB site. *J Bacteriol* **178**: 5533-5536.

Petrova, Z.O., G.W. Broussard & G.F. Hatfull, (2015) Mycobacteriophage-repressor-mediated immunity as a selectable genetic marker: Adephagia and BPs repressor selection. *Microbiology* **161**: 1539-1551.

Pham, T.T., D. Jacobs-Sera, M.L. Pedulla, R.W. Hendrix & G.F. Hatfull, (2007) Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. *Microbiology* **153**: 2711-2723.

Pierson, L.S., 3rd & M.L. Kahn, (1987) Integration of satellite bacteriophage P4 in Escherichia coli. DNA sequences of the phage and host regions involved in site-specific recombination. *J Mol Biol* **196**: 487-496.

Pope, W.H., C.A. Bowman, D.A. Russell, D. Jacobs-Sera, D.J. Asai, S.G. Cresawn, W.R. Jacobs, R.W. Hendrix, J.G. Lawrence, G.F. Hatfull, G. Science Education Alliance Phage Hunters Advancing, S. Evolutionary, R. Phage Hunters Integrating, Education & C. Mycobacterial Genetics, (2015) Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *eLife* **4**: e06416.

Pope, W.H., D. Jacobs-Sera, D.A. Russell, C.L. Peebles, Z. Al-Atrache, T.A. Alcoser, L.M. Alexander, M.B. Alfano, S.T. Alford, N.E. Amy, M.D. Anderson, A.G. Anderson, A.A. Ang, M. Ares, Jr., A.J. Barber, L.P. Barker, J.M. Barrett, W.D. Barshop, C.M. Bauerle, I.M. Bayles, K.L. Belfield, A.A. Best, A. Borjon, Jr., C.A. Bowman, C.A. Boyer, K.W. Bradley, V.A. Bradley, L.N. Broadway, K. Budwal, K.N. Busby, I.W. Campbell, A.M. Campbell, A. Carey, S.M. Caruso, R.D. Chew, C.L. Cockburn, L.B. Cohen, J.M. Corajod, S.G. Cresawn, K.R. Davis, L. Deng, D.R. Denver, B.R. Dixon, S. Ekram, S.C. Elgin, A.E. Engelsen, B.E. English, M.L. Erb, C. Estrada, L.Z. Filliger, A.M. Findley, L. Forbes, M.H. Forsyth, T.M. Fox, M.J. Fritz, R. Garcia, Z.D. George, A.E. Georges, C.R. Gissendanner, S. Goff, R. Goldstein, K.C. Gordon, R.D. Green, S.L. Guerra, K.R. Guiney-Olsen, B.G. Guiza, L. Haghighat, G.V. Hagopian, C.J. Harmon, J.S. Harmson, G.A. Hartzog, S.E. Harvey, S. He, K.J. He, K.E. Healy, E.R. Higinbotham, E.N. Hildebrandt, J.H. Ho, G.M. Hogan, V.G. Hohenstein, N.A. Holz, V.J. Huang, E.L. Hufford, P.M. Hynes, A.S. Jackson, E.C. Jansen, J. Jarvik, P.G. Jasinto, T.C. Jordan, T. Kasza, M.A. Katelyn, J.S. Kelsey, L.A. Kerrigan, D. Khaw, J. Kim, J.Z. Knutter, C.C. Ko, G.V. Larkin, J.R. Laroche, A. Latif*, et al.*, (2011) Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. *Plos One* **6**: e16329.

Priest, D.G., L. Cui, S. Kumar, D.D. Dunlap, I.B. Dodd & K.E. Shearwin, (2014) Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and lambda repressors. *Proc Natl Acad Sci U S A* **111**: 349-354.

Ptashne, M., (2004) A genetic switch: Phage lambda revisited, 3rd edit. *Cold Spring Laboratory Press, New York, NY*.

Ptashne, M., (2011) Principles of a switch. *Nature chemical biology* **7**: 484-487.

Ptashne, M., A.D. Johnson & C.O. Pabo, (1982) A genetic switch in a bacterial virus. *Scientific American* **247**: 128-130, 132, 134-140.

Qian, X.H. & M.M. Cox, (1995) Asymmetry in active complexes of FLP recombinase. *Genes Dev* **9**: 2053-2064.

Radman-Livaja, M., T. Biswas, T. Ellenberger, A. Landy & H. Aihara, (2006) DNA arms do the legwork to ensure the directionality of lambda site-specific recombination. *Curr Opin Struct Biol* **16**: 42-50.

Rajeev, L., K. Malanowska & J.F. Gardner, (2009) Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol Mol Biol Rev* **73**: 300-309.

Rasband, W., (1997) ImageJ. US National Institutes of Health, Bethesda, MD. In., pp.

Reed, M.R., K.E. Shearwin, L.M. Pell & J.B. Egan, (1997) The dual role of Apl in prophage induction of coliphage 186. *Mol Microbiol* **23**: 669-681.

Reichardt, L. & A.D. Kaiser, (1971) Control of lambda repressor synthesis. *Proc Natl Acad Sci U S A* **68**: 2185-2189.

Reiter, W.D., P. Palm & S. Yeats, (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* **17**: 1907-1914.

Richet, E., P. Abcarian & H.A. Nash, (1986) The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell* **46**: 1011-1021.

Richet, E., P. Abcarian & H.A. Nash, (1988) Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell* **52**: 9-17.

Roberts, J.W., (1969) Termination factor for RNA synthesis. *Nature* **224**: 1168-1174.

Roberts, J.W., W. Yarnell, E. Bartlett, J. Guo, M. Marr, D.C. Ko, H. Sun & C.W. Roberts, (1998) Antitermination by bacteriophage lambda Q protein. *Cold Spring Harb Symp Quant Biol* **63**: 319-325.

Ross, W. & A. Landy, (1983) Patterns of lambda Int recognition in the regions of strand exchange. *Cell* **33**: 261-272.

Ross, W., A. Landy, Y. Kikuchi & H. Nash, (1979) Interaction of int protein with specific sites on lambda att DNA. *Cell* **18**: 297-307.

Sadowski, P.D., (1995) The Flp recombinase of the 2-microns plasmid of Saccharomyces cerevisiae. *Prog Nucleic Acid Res Mol Biol* **51**: 53-91.

Sam, M.D., D. Cascio, R.C. Johnson & R.T. Clubb, (2004) Crystal structure of the excisionase-DNA complex from bacteriophage lambda. *J Mol Biol* **338**: 229-240.

Sanger, F., A.R. Coulson, G.F. Hong, D.F. Hill & G.B. Petersen, (1982) Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol* **162**: 729-773.

Schindler, D. & H. Echols, (1981) Retroregulation of the int gene of bacteriophage lambda: control of translation completion. *Proc Natl Acad Sci U S A* **78**: 4475-4479.

Schneider, C.A., W.S. Rasband & K.W. Eliceiri, (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671-675.

Seah, N.E., D. Warren, W. Tong, G. Laxmikanthan, G.D. Van Duyne & A. Landy, (2014) Nucleoprotein architectures regulating the directionality of viral integration and excision. *Proc Natl Acad Sci U S A* **111**: 12372-12377.

Segall, A.M. & H.A. Nash, (1993) Synaptic intermediates in bacteriophage lambda site-specific recombination: integrase can align pairs of attachment sites. *Embo J* **12**: 4567-4576.

Segev, N. & G. Cohen, (1981) Control of circularization of bacteriophage P1 DNA in Escherichia coli. *Virology* **114**: 333-342.

Senecoff, J.F., R.C. Bruckner & M.M. Cox, (1985) The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. *Proc Natl Acad Sci U S A* **82**: 7270-7274.

Shaikh, A.C. & P.D. Sadowski, (1997) The Cre recombinase cleaves the lox site in trans. *J Biol Chem* **272**: 5695-5702.

Shankar, S., A. Hatoum & J.W. Roberts, (2007) A transcription antiterminator constructs a NusA-dependent shield to the emerging transcript. *Mol Cell* **27**: 914-927.

Sherratt, D.J., L.K. Arciszewska, G. Blakely, S. Colloms, K. Grant, N. Leslie & R. McCulloch, (1995) Site-specific recombination and circular chromosome segregation. *Philos Trans R Soc Lond B Biol Sci* **347**: 37-42.

Shotland, Y., A. Shifrin, T. Ziv, D. Teff, S. Koby, O. Kobiler & A.B. Oppenheim, (2000) Proteolysis of bacteriophage lambda CII by Escherichia coli FtsH (HflB). *J Bacteriol* **182**: 3111-3116.

Smith-Mungo, L., I.T. Chan & A. Landy, (1994) Structure of the P22 att site. Conservation and divergence in the lambda motif of recombinogenic complexes. *J Biol Chem* **269**: 20798-20805.

Spiers, A.J. & D.J. Sherratt, (1997) Relating primary structure to function in the Escherichia coli XerD site-specific recombinase. *Mol Microbiol* **24**: 1071-1082.

Sternberg, N., D. Hamilton, S. Austin, M. Yarmolinsky & R. Hoess, (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.

Stevens, R.C., (2000) High-throughput protein crystallization. *Curr Opin Struct Biol* **10**: 558-563.

Stirling, C.J., S.D. Colloms, J.F. Collins, G. Szatmari & D.J. Sherratt, (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.

Stirling, C.J., G. Szatmari, G. Stewart, M.C. Smith & D.J. Sherratt, (1988) The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 cer locus. *Embo J* **7**: 4389-4395.

Studier, F.W., (2005) Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**: 207-234.

Subramanya, H.S., L.K. Arciszewska, R.A. Baker, L.E. Bird, D.J. Sherratt & D.B. Wigley, (1997) Crystal structure of the site-specific recombinase, XerD. *Embo J* **16**: 5178-5187.

Summers, D.K. & D.J. Sherratt, (1984) Multimerization of high copy number plasmids causes instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**: 1097-1103.

Summers, D.K. & D.J. Sherratt, (1988) Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the cer site. *Embo J* **7**: 851-858.

Suttle, C.A., (2005) Viruses in the sea. *Nature* **437**: 356-361.

Suttle, C.A., (2007) Marine viruses--major players in the global ecosystem. *Nature reviews. Microbiology* **5**: 801-812.

Syvanen, M., (1974) In vitro genetic recombination of bacteriophage lambda. *P Natl Acad Sci USA* **71**: 2496-2499.

Thompson, J.F., L.M. de Vargas, S.E. Skinner & A. Landy, (1987a) Protein-protein interactions in a higher-order structure direct lambda site-specific recombination. *J Mol Biol* **195**: 481-493.

Thompson, J.F., L. Moitoso de Vargas, C. Koch, R. Kahmann & A. Landy, (1987b) Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. *Cell* **50**: 901-908.

Tirumalai, R.S., E. Healey & A. Landy, (1997) The catalytic domain of lambda site-specific recombinase. *Proc Natl Acad Sci U S A* **94**: 6104-6109.

Van Duyne, G.D., (2001) A structural view of cre-loxp site-specific recombination. *Annu Rev Biophys Biomol Struct* **30**: 87-104.

Van Duyne, G.D., (2005) Lambda integrase: armed for recombination. *Curr Biol* **15**: R658-660.

Van Duyne, G.D., (2015) Cre Recombinase. *Microbiol Spectr* **3**: MDNA3-0014-2014.

Van Houdt, R., R. Leplae, G. Lima-Mendez, M. Mergeay & A. Toussaint, (2012) Towards a more accurate annotation of tyrosine-based site-specific recombinases in bacterial genomes. *Mob DNA* **3**: 6.

van Kessel, J.C. & G.F. Hatfull, (2007) Recombineering in Mycobacterium tuberculosis. *Nat Methods* **4**: 147-152.

van Kessel, J.C. & G.F. Hatfull, (2008a) Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol Microbiol* **67**: 1094-1107.

van Kessel, J.C. & G.F. Hatfull, (2008b) Mycobacterial recombineering. *Methods Mol Biol* **435**: 203-215.

van Kessel, J.C., L.J. Marinelli & G.F. Hatfull, (2008) Recombineering mycobacteria and their phages. *Nat Rev Microbiol* **6**: 851-857.

Villanueva, V.M., L.M. Oldfield & G.F. Hatfull, (2015) An Unusual Phage Repressor Encoded by Mycobacteriophage BPs. *Plos One* **10**: e0137187.
Visconti, N., (1953) Resistance to lysis from without in bacteria infected with T2 bacteriophage. *J Bacteriol* **66**: 247-253.
West, S.C., (1996) The RuvABC proteins and Holliday junction processing in Escherichia coli. *J Bacteriol* **178**: 1237-1241.
West, S.C., (1997) Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet* **31**: 213-244.
WHO, (2016) Global Tuberculosis Report 2016. In. Geneva, Switzerland: World Health Organization, pp.
Wommack, K.E. & R.R. Colwell, (2000) Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69-114.
Wu, R. & A.D. Kaiser, (1968) Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J Mol Biol* **35**: 523-537.
Yin, S., W. Bushman & A. Landy, (1985) Interaction of the lambda site-specific recombination protein Xis with attachment site DNA. *Proc Natl Acad Sci U S A* **82**: 1040-1044.
Zwieb, C. & S. Adhya, (2009) Plasmid vectors for the analysis of protein-induced DNA bending. *Methods Mol Biol* **543**: 547-562.