HETEROMERIC TRANSPOSASE ELEMENTS TARGET INSERTIONS INTO SPECIFIC GENOMIC LOCI AND METHODS TO STUDY TRANSPOSITION

A DISSERTATION

PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF CORNELL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

ASHWANA DEVI FRICKER

MAY 2015

HETEROMERIC TRANSPOSASE ELEMENTS TARGET INSERTIONS INTO SPECIFIC GENOMIC LOCI AND METHODS TO STUDY TRANSPOSITION

Ashwana Devi Fricker, PhD

Cornell University 2015

Transposons are mobile genetic elements found in all domains of life that are capable of moving between positions in a genome. The bacterial transposon Tn7 and related elements accumulate in specific insertion sites of the chromosome called genomic islands. The ability to recognize a specific insertion site and control the frequency of insertion involves five transposon encoded genes (*tnsABCDE*). The Tn7 core machinery is comprised of the atypical heteromeric transposase, TnsAB, and a regulator, TnsC. By using the core machinery and one of two targeting proteins, TnsD or TnsE, Tn7 is capable of inserting into either a chromosomal locus (*attTn7*) or horizontally transferred DNA. Tn7 is one of at least three families of transposons containing a heteromeric transposase that are found in diverse bacteria across a range of environments. In this work, I analyze ten elements from each of these three heteromeric transposase families and discuss how the conserved proteins within each family relate to TnsD and how those proteins may be responsible for targeting a specific conserved site on the chromosome. I then shift my focus to the other protein involved in target site selection, TnsE, which has been suggested to be species-specific in its interactions with the β-clamp. I find that cells expressing β-clamps from foreign hosts are viable in the absence of the *dnaN* allele and that TnsE-mediated transposition in strains containing foreign clamps only occurs in the presence of both proteins from the same host. In order to study how TnsE interacts with other host proteins, such as SeqA, a greater set of experimental tools is required. I develop an expression vector to express SeqA and a set of six mutants and suggest an *in vitro* assay to analyze multiple SeqA-

TnsE interactions within one experiment. I then work to develop a high-throughput method for mapping TnsE-mediated transposition events. I use a *dam*- strain and draw a few conclusions based on where insertions map in *seqA*- and wild type strains. More importantly, I suggest improvements for the method moving forward that will enable future generations of scientists to map millions of insertion events within a single experiment.

BIOGRAPHICAL SKETCH

As a child, Ashwana's parents stimulated her interest in the biological sciences- taking her into lab and letting her fill pipette boxes and soups of random paper and pencil shreds. She began doing neurobiological research sophomore year of high school, studying the physiological effect of neurons through serotonin receptor activation. Initially choosing to work in neurobiology because of personal interactions with friends, she realized that microbiology could affect more people in worse conditions. Because of this in college she chose a path in environmental microbiology.

Ashwana started in microbiology studying ferric iron mobilization by a bacterial consortium under the tutelage of the "grandfather" of geomicrobiology, Prof. Henry Ehrlich. In pursuit of financially-stable research, she then switched to studying dechlorination of chlorinated aromatics by a mixed bacterial culture under the watchful eye of Prof. Donna Bedard. She spent the rest of her time as an undergraduate working to isolate the bacterium responsible for the extensive dechlorination patterns.

Finishing college one year ahead of her peers, upon her acceptance to Cornell University in 2008, she decided that she would focus on learning as much as she could during her time here. She first settled in the laboratory of Dr. Ruth Ley, studying the effect of the innate immune system on the bacteria associated with the host. She then made the transition to the laboratory of Dr. Joseph Peters where she studied similarities between heteromeric transposase elements, and developed methods that will aid future generations in studying transposition.

THIS THESIS IS DEDICATED TO CAITI

ACKNOWLEDGEMENTS

I would like to acknowledge the people who have supported and advised me during my academic career. My two greatest mentors, Lakshmi Devi and Lloyd Fricker have encouraged the pursuit of knowledge, while reminding me to enjoy life. Together, they have inspired my curiosity in science, and led me into scientific research from an early age. There are too many friends to acknowledge independently, including many who have since left Ithaca and those who I have lived with, all who have taught me to work hard and party hard.

I am thankful for the initial committee members of my graduate study, Dr. Ruth Ley, Dr. Ian Hewson, Dr. Esther Angert, and Dr. Cynthia Leifer. Additionally, I am grateful to my current committee members, Dr. John Helmann and Dr. Jeffrey Roberts, whose advice has been insightful and invaluable. I particularly appreciate the guidance from my major adviser, Dr. Joseph Peters, who has continued to push me along the right track.

I am appreciative of my friends and colleagues in the Microbiology program and at Cornell who have provided me with insightful discussions and technical advice. Marco Straus, Mike Petassi, and Bennett Kapili have all been helpful during my time in the lab.

Finally, I am grateful for the three people that have supported me the most, for late-nights: my boyfriend Shreyas, for difficult times: my best friend Crystal, and for always being there: my cat Caiti.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION*

1.1 Summary

Mobile genetic elements have the ability to move between positions in a genome. Some of these elements are capable of targeting one of the template strands during DNA replication. Examples found in bacteria include (*a*) Red recombination mediated by bacteriophage lambda, (*b*) integration of group II mobile introns that reverse splice and reverse transcribe into DNA, (*c*) HUH endonuclease elements that move as single-stranded DNA, and (*d*) Tn*7*, a DNA cut-and-paste transposon that uses a target-site-selecting protein to target transposition into certain forms of DNA replication. In all of these examples, the lagging-strand template appears to be targeted using a variety of features specific to this strand. These features appear especially available in certain situations, such as when replication forks stall or collapse. In this review, we address the idea that features specific to the lagging-strand template represent vulnerabilities that are capitalized on by mobile genetic elements.

1.2 Introduction

After DNA replication, each daughter cell receives an original and a copied strand of parental DNA. Although this view might suggest that the two daughter chromosomes are equal in all ways, in reality, processing events related to the direction of DNA replication can impose differences in the mutation potential of the two strands. Depending on the direction of DNA replication, each stretch of nascent DNA an individual cell receives is derived from either the

*This chapter was prepared as a review in the journal Annual Reviews in Genetics Fricker, A.D., Peters, J.E., (2014). Vulnerabilities on the Lagging-Strand Template: Opportunities for Mobile Elements

leading-strand template or the lagging-strand template. It is known that repair biases lead to a general skew toward G > C in one strand of the chromosome in bacteria and archea (20), and that there are other processing differences between the strands (see below). In this review, we focus on the finding that the lagging-strand template is more vulnerable to mobile DNA elements than is the leading-strand template. We focus on bacteria because there are exceptional numbers of mobile elements accumulated by horizontal transfer in bacteria and because the regions of the genome derived from the leading- and lagging-strand templates are easier to unambiguously define. However, these findings likely apply to all three domains of life.

We suggest that some processing events found more frequently on the lagging-strand template make this strand more accessible or vulnerable, and therefore serve as an opportunity for mobile DNA elements seeking to insert into the host genome. Furthermore, we propose that the lagging-strand template is especially vulnerable in regions where DNA replication is perturbed. We also address the idea that hosts may have evolved systems to protect the vulnerable lagging-strand template.

1.2.1 Differences between the template strands

In all bacteria analyzed to date, DNA replication proceeds bi-directionally from one origin of replication per chromosome (*oriC*). Each replication fork complex (called a replisome) is responsible for replicating approximately half of the chromosome, and each of these regions is called a replichore. DNA polymerases in the replisome are held to the DNA by association with a protein ring called the sliding-clamp processivity factor (or sliding clamp). Another protein complex (collectively called tauin bacteria) couples the polymerases together across the strands (**Figure 1*a***). DNA polymerases replicate DNA in a 5′->3′ direction. Owing to the antiparallel nature of the complementary strands, the polymerase associated with the leading-strand template

moves continuously in the same direction as the larger replisome, whereas DNA replication on the lagging-strand template progresses away from the replisome. To provide the 3′ OH needed for DNA polymerase, new RNA primers are produced by primase every 1--2 kb in bacteria. The resulting DNA fragments on the lagging-strand template found prior to completion are called Okazaki fragments. The transient single-stranded DNA (ssDNA) in the lagging-strand template occurring between priming events is coated with single-stranded DNA binding protein (SSB).

In *Escherichia coli*, and likely all bacteria, two coordinated forces move the replisome: the action of the DNA polymerase on the leading-strand template and the action of the DnaB helicase (the enzyme responsible for separating the DNA strands) on the lagging-strand template (40). Primase (DnaG) is known to closely associate with DnaB. The process of initiating a new Okazaki fragment involves making a new RNA primer, loading a sliding clamp at this primer, and signaling DNA polymerase to switch from the last sliding clamp to the new one, leaving a sliding clamp free on the DNA for other interactions (reviewed in 13). Therefore, there are at least two regions of ssDNA on the lagging-strand template during replication: one between the Okazaki fragment that shrinks as the polymerases advance and one between the helicase and Okazaki fragment that grows in size until a new primer is started (**Figure 1**, arrows). There is compelling evidence suggesting that three DNA polymerases are active in a replisome, leaving the possibility that two Okazaki fragments may be replicated in a single replisome, adding an additional stretch of ssDNA (26, 48, 80). Multiple nonexclusive models have been suggested for how replication of a new Okazaki fragment is signaled. These basically differ on whether a second gap filling activity is needed to complete each fragment (44, 114). Multiple processing events need to occur in order to mature the lagging-strand template into a continuous DNA strand. In *E. coli*, DNA polymerase I (Pol I) fills in any remaining ssDNA gaps and, using its

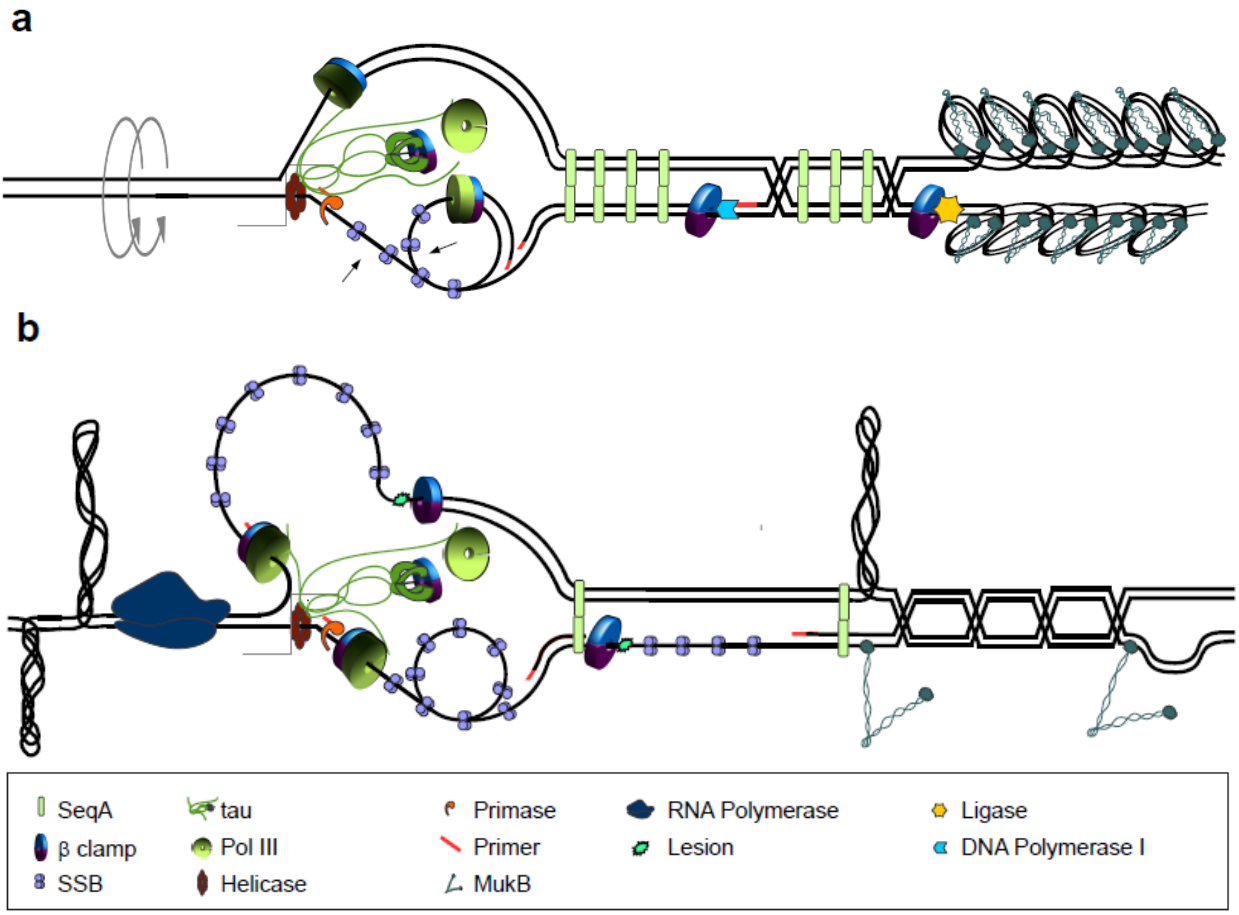

Figure 1.1

Normal and damaged DNA replication forks. The three-polymerase model, in which the polymerases on the leading- (top) and lagging- (bottom) strand template are coupled by tau (green tentacles), is shown. Pol III (DNA polymerase III) (green circles) is held to each strand by the sliding clamp (blue/purple homodimeric circles). The helicase (red hexamer) tracks along the lagging-strand template and opens the two parental strands. Each Okazaki fragment is started with an RNA primer (red line) generated by primase (orange comma). The ssDNA (single-stranded DNA) on the lagging-strand is loaded with SSB (single-stranded DNA binding protein) (light purple homotetramer).
A) Normal DNA replication fork. Okazaki fragments are processed by the progressive action of RNA primer removal by Pol I (DNA polymerase I)(blue chevron) and repair of the remaining nick by ligase (pink star). After replication, the hemimethylated duplex DNA has been suggested to be tethered together by SeqA (green rods) until MukB (teal balls and chains) forms a secondary structure. The two arrows indicate regions of ssDNA during normal DNA replication.
B) Damaged DNA replication fork. DNA replication forks can stall or collapse due to a number of processes. Many of these are shown, but any one could stall or collapse a DNA replication fork. Supercoiling, large protein complexes, such as RNA polymerase (dark blue crab claw), or lesions (yellow blob) can lead to replication fork stalling. The ssDNA on both strands is loaded with SSB. After replication, a failure in the coupling protein SeqA or in the secondary structure forming protein MukB may lead to altered supercoiling, delayed precatenane removal, and other issues with daughter-strand segregation.

flap endonuclease (54) and/or 5′--3′ exonuclease (52) activity, may displace and remove the RNA primer. After processing by Pol I, the 5′ PO_4 and 3′ OH remain until DNA ligase seals the nick (50). Ligase may associate with the sliding clamp previously occupied by DNA Pol I to help identify its substrates (50).

1.2.2 Exogenous and endogenous forces complicate DNA replication

Damaged DNA and physical or topological blockages in the template strands perturb the orderly coordination of events described above. Many types of lesions can block DNA polymerase on the leading-strand or lagging-strand template, resulting in significant gaps that require special repair processes, such as homologous recombination and lesion repair systems, and the recruitment of various different DNA polymerases (Figure 1B). The entire replisome can be stalled at cross-links that occur across the DNA strands at collisions with protein complexes (such as RNA polymerase) or in regions of high supercoiling (12, 77, 78). In many of these cases, a stalled replication fork needs to be restarted or completely re-assembled using a special machinery involving the Pri proteins and associated factors (30, 84). Multiple pathways for restarting DNA replication forks have been reconstituted in vitro, where, following reloading of the DnaB helicase, other replisome components can be coordinated. In some bacteria with circular chromosomes, replication forks are also actively stalled at a specific DNA sites in the chromosome (called *ter* sites), where a *trans*-acting protein binds to the site, inhibiting the progression of the replicative DNA helicase (reviewed in 14). In other bacteria, replication may terminate when the two forks collide or by some unknown mechanism. It remains unclear how replication fork stalling specifically affects processing events on the lagging-strand template, but this template appears to be more vulnerable to mobile elements in these situations (see below).

1.3 Mobile DNA Elements that take Advantage of Attributes of the Lagging Strand Template

The following examples present cases in which mobile elements appear to take advantage of aspects found predominantly on the lagging-strand template, including stretches of ssDNA, free 3′ OH ends that can prime replication, and interactions found with specific proteins enriched on the lagging-strand template. The lagging-strand template may be especially accessible in cases in which a replisome encounters DNA damage, stalls or collapses, and at sites where DNA replication actively terminates. We refer to this accessibility of the lagging-strand template as a vulnerability in the host-mediated process of DNA replication, given that the insertion of mobile elements is more likely to be detrimental than beneficial to the host organism. As addressed below, it can be argued that both the mobile element and the host may derive some benefit from greater accessibility during periods of stress.

1.3.1 Lambda Red recombination and other phage systems

Host-mediated recombination is extremely important to the cell for restarting stalled or collapsed replication forks. Homologous recombination systems that are unrelated to those found in the bacterial host can be found in some viruses of bacteria called bacteriophages (phages). It is still under debate as to why phages would have recombination systems, although multiple ideas have been proposed. Phage strains lacking lambda Red recombination function grow poorly, which may be attributed to the formation of concatemers of the lambda phage during replication (87). Alternatively, lambda Red recombination may play an important role in the highly mosaic nature of the bacteriophage (58). By being able to prime interphage recombination using very small patches of homology, there are more opportunities to stitch together new phage variants. However, phage-mediated recombination events are not needed for integration into the

chromosome (2). The best-studied system for phage-mediated recombination is the Red recombination system from bacteriophage lambda, although other phage systems seem to show the same properties (79, 93, 99) (see below).

The lambda Red recombination system involves two proteins, Exo (or Redα) and Bet (or Redβ). The lambda Red system has been reviewed in detail recently (30a, 64a) and is reviewed briefly here. The Bet recombinase is a single-strand annealing protein that binds to ssDNA, scans DNAs in the cell for a homologous sequence, and anneals the bound DNA to a homologous or mostly homologous sequence, allowing it to be present in progeny. Host-encoded genetic recombination generally utilizes a different type of protein, the well-studied RecA recombinase, in bacteria (reviewed in 53). Exo is an exonuclease that is responsible for processing duplex DNAs to ssDNA that can be bound by Bet without an accessory activity. A third protein that is found with the lambda Red system is Gam, which protects DNAs with blunt or nearly blunt ends from the action of RecBCD (and other host exonucleases). Without Gam, RecBCD aggressively digests lambda DNA because it lacks *chi* sites that normally act as a switch which converts the process from one in which the DNA is degraded to one in which the RecA recombinase is actively loaded (64). In fact, work with *gam* mutants identified the *chi* sequence by the ability of this sequence to protect the phage from RecBCD (65).

An interesting aspect of lambda Red recombination is that there is an apparent dependency on DNA replication and an enhanced recombination with the lagging-strand template, indicating that discontinuously replicated DNA is more active for Red-mediated recombination. The bias for the lagging-strand template is revealed from work with recombination with ssDNA oligonucleotides. To monitor recombination into the chromosome or plasmid, successful recombination of oligonucleotides containing mismatches can be screened or

directly selected. In experiments utilizing oligonucleotides, an early observation indicated that recombination was more efficient if it was complementary to the lagging-strand template compared to when it was complementary to the leading-strand template (15, 45). One model to explain a bias for recombination on the lagging-strand template holds that, unlike the RecA recombination system, which has the capacity to survey duplex DNA for homologous sequences, the phage recombinase must have access to DNA that is already single stranded.

Given that DNA replication is an essential process, it is difficult to test whether DNA replication is required for Red recombination. However, a great many experiments are highly consistent with the idea that ssDNA oligonucleotides recombine into transient ssDNA regions at the replication fork. For example, a decrease in recombination efficiency is found with oligonucleotides that lack a 3′ OH, suggesting that DNA replication must be primed from this end (32). Furthermore, a mutant allele of DnaG primase, which extends the length of Okazaki fragments and therefore expands the regions of ssDNA, increases the frequency of Red recombination (42). Important hints about the degradation of incorporated DNA during Red recombination also come from experiments which can determine the extent of oligonucleotide processing after recombination. This processing can be monitored using oligonucleotides containing mismatches to the host target sequence. After Red recombination has integrated the oligonucleotide, sequencing can determine which regions of the oligonucleotide were recombined into the host genome (46, 63). In addition to observing that oligonucleotides are frequently processed when they are recombined into the host, it was also found that the exonuclease activities of the normal DNA polymerases used in *E. coli* replication, DNA polymerase I and III, appeared to be responsible for these processing events (46). Neither exonucleases involved in DNA processing nor a DNA polymerase that is induced with DNA

damage, Pol II, appeared to be involved in processing the oligonucleotides, supporting the idea that Red recombination occurs during normal DNA replication.

The effect of host-encoded methyl-directed mismatch repair (MMR) on the Red recombination process is also consistent with a process that occurs during DNA replication. A requirement for monitoring Red recombination is that at least one mismatch is present on the oligonucleotide used in the experiment. Red recombination occurs at ~ 100-fold increased frequency in *E. coli* hosts that are deficient in MMR (7). The *E. coli* MMR system is responsible for identifying and correcting mismatches that occur following DNA replication (reviewed in 22, 36). Upon recognition of the mismatched bases, MutS recruits MutL and MutH to nick one strand, which is removed and replaced. One important feature of the MMR system involves an ability of the system to recognize the newly replicated strand (that presumably contains a polymerase mistake). In *E. coli* and related bacteria, this recognition process involves DNA methylation. A separate enzyme, DNA adenine methyltransferase (Dam) is responsible for methylation of the A site at GATC/CTAG sequences. The methylase tracks behind the replication fork, leaving the new GATC sequence transiently unmethylated. This so-called hemimethylated state at GATC/CTAG sequences where one A is methylated and the other is not is recognized by the MutH protein. MutH will then nick the new (unmethylated) DNA strand. A helicase removes the strand with the nick along with the incorrect base, and subsequent replication replaces the missing strand. In bacteria, archaea, and eukaryotes that lack the Dam methylation system, a different and incompletely understood process is involved in identifying the newly replicated strand. Although the type of mismatch can affect the overall effect of MMR deficiency on the final recombination process, the loss of MMR does not change the bias for the lagging-strand template (45). In practice, high efficiency recombination can occur in wild-type

cells by using oligonucleotides with consecutive mismatches that prohibit MutS binding, with mismatches that are not recognized by the MMR system, or using oligonucleotides containing chemically modified bases that subvert the MMR system (7, 86, 101, 111).

There is reason to believe that the phage-encoded recombination systems take advantage of a vulnerability that is already present during DNA replication. In support of this, oligonucleotides can recombine in bacteria that have no obvious homologs to the phage systems. At a very low frequency, RecA-independent oligonucleotide recombination can be monitored in various members of gamma proteobacteria with the same bias for the lagging-strand template. These systems show many of the same hallmarks of the lambda Red system, in that the concentration of the oligonucleotide and the ability of certain DNA sequences to anneal more stably are important for obtaining higher frequencies of recombination (86, 94). However, there is an interesting difference between phage-mediated oligonucleotide recombination and oligonuceotide recombination associated with the host, concerning the length of the oligonucleotide used for recombination. In phage-mediated oligonucleotide recombination, the Red recombination frequency of smaller oligonucleotides (minimally 23 bp) increases exponentially as the length increases to 40 bp but plateaus up to 70 bp (86). In one set of experiments, the host oligonucleotide recombination systems integrated shorter oligonucleotides (~20 bp in length) with approximately the same efficiency as oligonucleotides of up to 70 bp in length in *E. coli* and 120 bp in *Pseudomonas syringae* (94). The observation that, with Red-independent oligonucleotide recombination, smaller oligonucleotides (shorter than 20 bp) can recombine with a low frequency that does not increase with the length of the oligonucleotide has been used to suggest that the naked oligonucleotide simply binds to regions of ssDNA at the replication fork (94). Conversely, the observation that the efficiency of Red recombination

increases with progressively longer oligonucleotides has been used to support the importance of Bet-binding (86). Work with a phage system derived from *Mycobacteria* suggests that these same attributes found in the proteobacteria are also found in highly diverged bacteria; the mycobacterial phage Che9c encodes the exonuclease gp60 and recombinase gp61, which carry out the same functions as Exo and Bet, respectively, (99).

Work with lambda Red and other phage recombination systems suggests that regions of transiently exposed ssDNA are available for Bet on the lagging-strand template but not generally on the leading-strand template. A single oligonucleotide containing sequences flanking a region up to 45 kb can recombine with the DNA causing a deletion of the region (Figure 1.2A and (100)). This finding suggests that annealed oligonucleotides are stable and that sequences found in an oligonucleotide can bridge regions that are well separated during active replication (100). Generally, phage recombination systems only have high efficiency in the hosts from which they were originally isolated (93, 99). This is consistent with the idea that these systems do not passively seek out ssDNA but, instead, rely on interactions between the phage proteins and host proteins to facilitate the entry process into the replisome. It remains unclear which mechanisms afford the replisome protection from phage recombination, but work in this area could synergize with work in other systems that target the lagging-strand template.

1.3.2 Group II mobile introns

Another type of mobile element that appears able to capitalize on features on the lagging-strand template is group II mobile introns (reviewed in 43). Full integration of these elements can be associated with host DNA replication. Group II mobile introns move as RNA after self-splicing out of an mRNA. The element itself is a catalytic intron RNA (ribozyme) that carries out two transesterification reactions for excision from an mRNA as a lariat structure. Group II

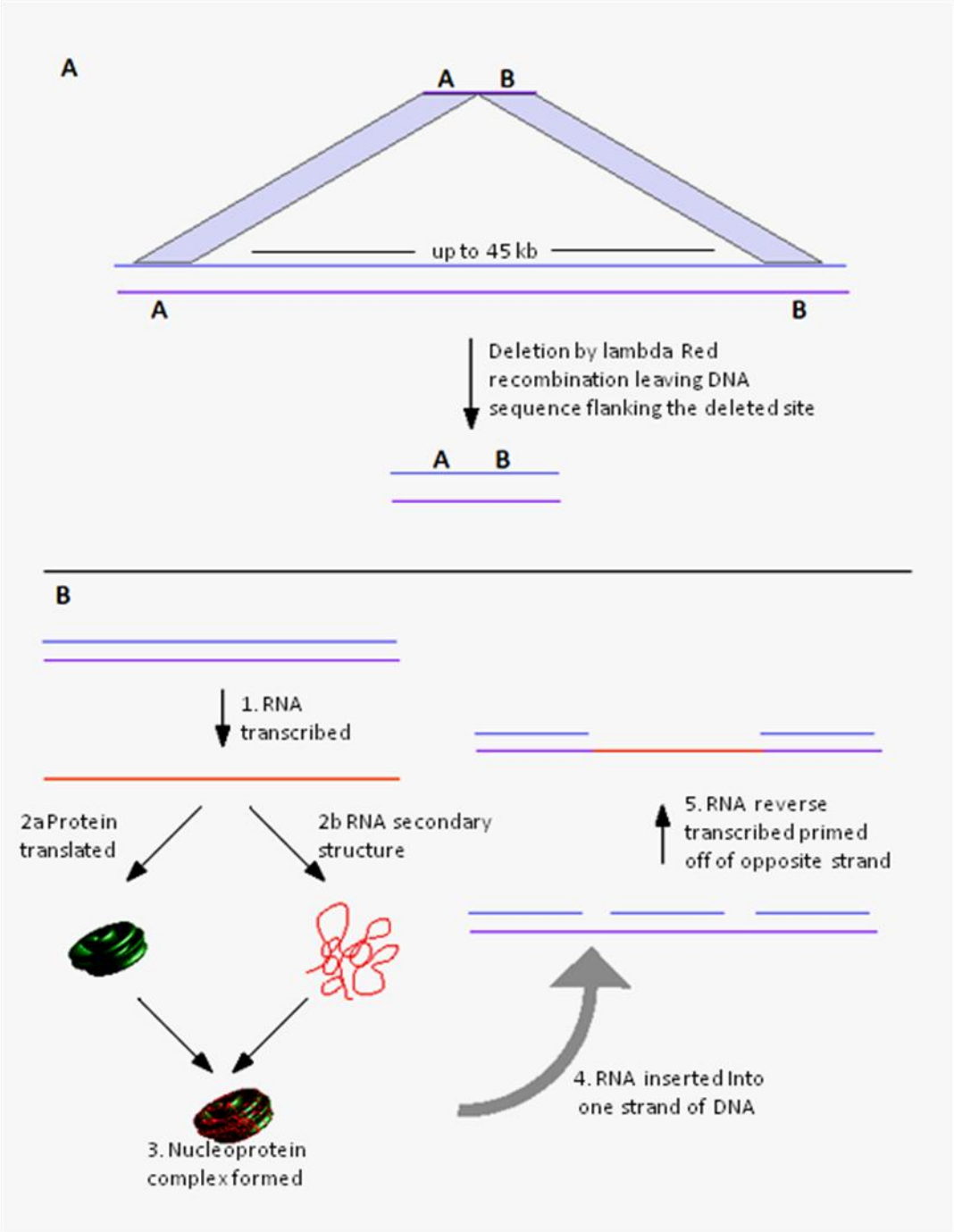

Figure 1.2

Targeting of the lagging strand template during replication by either (**A**) deletions using a single-stranded oligo with DNA flanking the target site and lambda Red recombination or (**B**) insertions by group II mobile introns that insert using a ribozyme complex made up of protein (green) and RNA (red).

mobile introns additionally use an intron-encoded protein (IEP) that acts as a maturase to help the mRNA fold into the correct structure and a reverse transcriptase activity to make a DNA copy of the RNA element after integration (Figure 1.2B (59)). The IEP can also have endonuclease activity, for reasons described below. As the word intron implies, group II introns use the same chemistry as introns that splice out of mRNA in eukaryotes (and are likely relatives of introns found in mRNA). After excision, the element integrates into DNA by using one of two pathways that differ by the targets that are recognized. In one pathway, called retrohoming, the element inserts into a single conserved site that base pairs with the RNA (and typically is recognized by the IEP). A second targeting pathway that occurs at a much lower frequency, called retrotransposition, involves a broader array of sites that have imperfect matches to the site used in retrohoming.

Integration events in DNA involve reverse splicing, where the mobile RNA is joined to the target DNA (81a). This process is facilitated by the ability of the IEP to unwind the DNA strands to allow the element to search for homology for target recognition (91). After the single-stranded RNA (ssRNA) is integrated into one DNA strand, the subsequent steps can be variable, depending on the element and the host (63a, 114b). However, in all cases, a 3′ OH must be available to initiate reverse transcription to make a DNA copy of the ssRNA element that has been integrated (114a, 114b). In some cases, the IEP also contains an endonuclease activity that is capable of breaking the other strand (85). This ability to cleave the second strand for priming reverse transcription is often used in the retrohoming reaction (8a). However, the process of retrotransposition that allows the movement of the element to other sites often requires another mechanism for initiating reverse transcription (34). Retrotransposition favors integration into

DNA replication forks, and depending on the element and host, can show a bias to the lagging-strand template found during DNA replication (33, 34).

One well-studied group II mobile intron is the Ll.LtrB intron found in the relaxase gene *ltrB*, involved in conjugal plasmid transfer in *Lactococcus lactis* (60). Ll.LtrB preferentially retrohomes into the *ltrB* gene using its IEP, LtrA. *In vivo*, intron movement can be monitored using a system in which L1.LtrB insertion events in the chromosome can be detected by direct antibiotic selection (8). Using this system, retrohoming occurs at a high frequency (approximately half the substrates that are available result in integration). The importance of the endonuclease activity of LtrA to retrohoming was assessed using a triple LtrA mutant, Y529A-R531A-T533A ($LtrA^{YRT}$), which has normal maturase and DNA binding ability but lacks endonuclease activity (85). The $LtrA^{YRT}$ protein showed an ~100-fold drop in retrohoming compared with the wild-type protein, indicating that retrohoming is highly dependent on the endonuclease activity of LtrA (8, 34). However, in the process of retrotransposition, where a broader variety of sites is recognized, movement does not appear to be affected as strongly. Although retrotransposition with wild-type LtrA only occurred at a frequency of 0.02%, this translated into an unexpectedly modest drop (~40%) in the frequency of retrotransposition with $LtrA^{YRT}$ (34). Interestingly, retrotransposition was also strongly biased into the lagging-strand template, as was particularly obvious in a unidirectionally replicating plasmid in the host, where 32 out of 33 events were in the lagging-strand template (33). Multiple factors could account for a bias for this strand: a requirement for ssDNA over dsDNA, a need for more frequent priming via 3′ OH ends, or interactions with lagging-strand-dependent host replication factors. Using an in vitro system, retrotransposition into ssDNA was found to be preferred over retrotransposition into dsDNA, a bias that was markedly more evident in some DNA sequences. Analysis of natural

retrotransposition events found in sequenced genomes also indicated a bias to the lagging-strand template (34). An explanation for why retrotransposition, but not retrohoming, is biased to lagging-strand replication comes from the knowledge that LtrA-mediated dsDNA unwinding and endonuclease activities are dependent upon binding of specific target sequences (62, 91). The contacts that are important for unwinding and second-strand nicking would therefore be unlikely to be found in sites used for retrotransposition. Thus, these integration events could be expected to be dependent on the ssDNA nature of the lagging-strand template and the availability of 3′ OH from Okazaki fragments to prime reverse transcription (34). Although other factors cannot be ruled out, these results suggest there is a strong bias for the lagging-strand template because it provides more availability of ssDNA targets. This would remove the need for unwinding a dsDNA template and may provide more opportunities for priming on the lagging-strand template that involves discontinuous replication.

Together, these data suggest that ssDNA and Okazaki fragments on the lagging-strand template provide an opportunity for group II elements to utilize a wider variety of insertions sites during retrotransposition, suggesting that these elements may specifically target these regions of the DNA because they are naturally available or accessible. It is hypothesized that more ancestral group II mobile elements, those without endonuclease domains, may have used ssDNA preferentially and biased retrotransposition into the lagging-strand. Elements that acquired IEPs with a C-terminal DNA-binding domain could increase DNA unwinding ability (27), allowing for insertions into dsDNA, thereby providing a gateway for evolving more stringent target-site selection. There is also evidence that insertion during DNA replication and into the lagging-strand template may reduce or eliminate, via priming from Okazaki fragments, the need for cleavage of the second strand. This is consistent with the finding that RmInt1, a group II mobile

intron from *Sinorhizobium meliloti* that naturally lacks the C-terminal endonuclease domain, has a distinct preference for the lagging-strand template (57). The acquisition of an endonuclease domain could be seen as another advance that would obviate the need for insertion into actively replicating DNA, which would provide a primer for reverse transcription. This pathway may have been of paramount importance for early group II mobile elements and may still be important to many simpler elements found today.

1.3.3. HUH endonuclease elements

Another class of mobile elements that takes advantage of aspects of the lagging-strand template are the HUH (where H is a histidine and U is a hydrophobic residue) endonuclease elements. They are distinct from other classes of mobile elements in that they move as ssDNA. As described below, the movement of these elements is highly linked to the availability of ssDNA, both for excision and for integration. Two well-studied IS*200*/*605* family members that represent this class are IS*608* and IS*Dra2*.

IS*608* was originally discovered in the pathogen *Helicobacter pylori* and is particularly well studied because it is able to mobilize in a heterologous *E. coli* system (39). Using this *E. coli* system, it has been shown that IS*608* integration occurs at many positions but is always immediately downstream of a tetranucleotide sequence (5′-TTAC) (39). IS*608* naturally carries two genes: *tnpA*, which encodes a transposase, and *tnpB*, which is not required for transposition and is of unknown function (39). The TnpA transposase is a member of a large family of proteins that includes a conserved HUH domain to coordinate metal ions involved in catalysis (3). Members of the HUH family are found in all domains of life and appear to have adapted to many cellular processes involving cleavage and ligation at specific sequences in ssDNA (3).

TnpA contains the catalytic HUH motif, and work with this protein provides insight into the function of other HUH family proteins that have additional domains. Reconstitution of the transposition pathway in vitro indicated that TnpA catalyzes excision from ssDNA substrates (but not from dsDNA substrates) flanked by the terminal left and right ends (28). This ssDNA requirement inherently biases transposition to the lagging-strand template during replication. Consistent with this idea, a bioinformatics analysis of the natural occurrence of this family of elements indicated that there is a strong bias for elements in the lagging-strand template (97). To address strand differences *in vivo*, excision of the transposon from either the leading- or lagging-strand template of a plasmid was monitored (97). After overnight growth, a high degree of excision was found to occur only when the element resided in the lagging-strand template and was hardly detectable when situated in the leading-strand template (97). Furthermore, the frequency of excision correlated to what is known about the length of Okazaki fragments: transposition frequencies were highest in a 0.3-kb element but decreased as the length was increased to 4 kb (97). This suggests the IS*608* element may need to reside entirely in one ssDNA region found on the lagging-strand template. Additional studies related IS*608* movement to DNA replication; transposition was dependent on replication in experiments that used strains containing temperature-sensitive mutations in essential DNA replication genes (97). Mutants that produce less DnaG primase, which increases the length of stretches of ssDNA on the lagging-strand template, had higher excision frequencies (97). This phenotype can be rescued by overexpression of DnaG, which consistently resulted in fewer excision events (97). Moreover, the excision frequency with less DnaG as compared with the wild type was greater in longer synthetic constructs than in shorter constructs (97).

Examining where transposition events occurred also provided interesting information about the presumed availability of ssDNA in cells. Insertion events that occurred into mobile plasmids and the chromosome were biased in an orientation expected for insertions into the lagging-strand template (97). Chromosomal insertions showed hot spots in highly transcribed rRNA operons, presumably because replication forks overtake RNA polymerase in these regions (97). In cases where there is a rear-end collision between the faster moving DNA polymerase and RNA polymerase, DNA replication is not believed to be stalled (76), but this event could impart other effects that can be capitalized on by these elements. The replication termination system of *E. coli* could be used as a tool to relate transposition and replication fork pausing (1). In *E. coli*, the progression of DNA replication forks is inhibited at specific sites in the DNA, called *ter* sites, through the action of a *trans*-acting protein, Tus (reviewed in 66). Interestingly, a unique insertion hot spot was observed proximal to an active *ter* site in a plasmid-based system and was dependent on the Tus protein (97). These results indicate that ssDNA in the lagging-strand template is the preferred transposition target and that access to this structure may increase when replication forks stall.

Similar results have been obtained for IS*Dra2*, another member of the IS*200*/*605* family from *Deinococcus radiodurans* (97). As was found with IS*608*, sequenced transposition events across the chromosome show a bias with the direction of DNA replication (97). Additionally, in a transposition assay based on the ability of excision events to activate a gene (where presumably orientation would not matter), events were biased to the lagging-strand template (97). The bias toward ssDNA found in other instances could also be shown with *D. radiodurans* as a function of its ability to withstand very high levels of radiation (70a). When exposed to such extreme conditions, the chromosome experiences high levels of fragmentation, and extended sections of

ssDNA are used to rebuild the chromosome. It was found that under these conditions, transposition is stimulated (70a), and as expected, any bias to the leading- or lagging-strand is lost (97). This is consistent with the idea that ssDNA is the most important feature of the lagging-strand template for transposition and not necessarily some of the other aspects specific to DNA replication. Many transposons seem to regulate transposition with the physiologic state or stressed state of the cell (see below). It is possible that sensitivity of the IS*Dra2* element to ssDNA could also be an adaptive mechanism to stimulate transposition in response to host stress.

These results suggest that the ssDNA found on the lagging-strand template during DNA replication provides an opportunity for the very existence of HUH endonuclease elements. Work with these elements suggests that ssDNA may also be a signal of cell stress from extreme DNA damage stimulating the movement of this class of element. These studies also support the notion that not all lagging-strand templates are the same. Places where DNA replication forks stall, overtake RNA polymerase, or are subject to frequent priming by DnaG may lead to differential opportunities for movement of HUH endonuclease elements involving the lagging-strand template.

1.3.4 TnsE-mediated Tn7 transposition

Transposons are discrete elements that can move within a genome. Transposons that move as ssDNA appear to require the lagging-strand template during replication in a donor or target DNA (see above). There are examples of DNA transposons that move as dsDNA that are also sensitive to DNA replication. Transposon Tn*7* provides an interesting example where the lagging-strand template is actively targeted for transposition (73). Tn*7* transposes via a dsDNA intermediate in a process known as cut-and-paste transposition, in which the element is excised from one site and inserted into another position in the cell (47, 69, 74). The process of Tn*7*

transposition requires multiple proteins that act on nonidentical *cis*-acting left and right ends that flank the element. The transposase that removes and rejoins the element is composed of two proteins, TnsA and TnsB (TnsAB), which work with a regulator protein, TnsC. Transposition targets are identified by one of two dedicated targeting proteins, TnsD or TnsE. The TnsD protein specifically targets a unique sequence found in bacterial chromosomes. The ability to target the lagging-strand template involves one of the Tn*7*-encoded proteins, TnsE, which is able to specifically recognize components found on the lagging-strand template during replication.

TnsE-mediated transposition was initially of interest because of its ability to specifically direct transposition into mobile plasmids called conjugal plasmids, which are capable of moving between bacteria (110). When a conjugal plasmid is present in the strain, transposition is stimulated >100-fold, and the vast majority of these insertions are targeted into the mobile plasmids, despite the fact that they make up ~1% of the DNA in the cell (73, 110). Analysis of these insertions indicated that there is no sequence specificity with the TnsE-mediated pathway (110). However, a striking orientation bias of insertions is found in these conjugal plasmids (110). During conjugation, the relaxase nicks the origin of transfer (*oriT*) of the conjugal plasmid, and one strand of the plasmid DNA is transferred into the recipient cell, initiating DNA processing events (108). Host-mediated DNA replication synthesizes the complementary strands in the donor and recipient cell during the process of transfer. In the donor cell, a continuous process akin to processing on a leading-strand template occurs that is initiated by the liberated 3′ OH end. However, in the recipient cell, DNA replication is continuously reprimed in a process more similar to events found on the lagging-strand template. In early work with TnsE (110), it was unclear which specific molecular target was recognized and how it could drive the orientation bias with transposition. Targeting was dependent on active conjugation; conjugal

plasmids were activated as transposition targets in recipient cells, but the plasmid-encoded or host-encoded proteins involved in the process were not identified.

The essential role of host-mediated processes on the lagging-strand template was indicated by experiments in strains without conjugal plasmids (73). In these strains, TnsE-mediated transposition events occur at a very low frequency into the chromosome, with a regional bias centered around DNA replication termination sites. Strikingly, the rare TnsE-mediated insertion events in the chromosome occurred only in a single orientation across each replichore (73). The strict orientation bias indicated that a replication process from the host was the preferred target for TnsE-mediated transposition. This finding, in addition to the previous observation that TnsABC+E transposition targets events in conjugal plasmid in recipient cells, indicates that the lagging-strand template provides a DNA structure and/or protein complex that is recognized by TnsE-mediated transposition.

Additional experiments revealed that two components found during DNA replication on the lagging-strand template are essential for recognition by TnsE. Although TnsE generally has a strong affinity for DNA, competition experiments revealed that its preferred binding substrates are structures that contain a 3′ recessed end. Supporting the role of this structure, it was found that TnsE gain-of-activity mutant proteins show an enhanced ability to interact with 3′ recessed-end structures (73). These structures are abundant during replication, an observation that reinforces the idea that TnsABC+E transposition targets the lagging-strand template. Further investigation revealed that there was also an essential protein component for TnsE-mediated transposition, the sliding-clamp processivity factor (68). As noted above, the sliding-clamp proteins are deployed with each new priming cycle on the lagging-strand template, and sliding clamps left behind appear to be important for recruiting proteins that are responsible for actions

that mature the lagging-strand template, including RNA primer removal, ligation of Okazaki fragments, and mismatch repair (50, 75). Proteins that interact with the sliding clamp have a conserved motif that facilitates part of this interaction (9a). Interestingly, a putative clamp-interacting motif was identified in TnsE and was also found to be conserved across homologs of the TnsE protein (68). Interaction between TnsE and the sliding clamp was confirmed biochemically and genetically (68). Mutations in TnsE that perturbed the TnsE--sliding clamp interaction also either abolished or significantly decreased TnsE-mediated transposition but did not affect transposition targeted in other pathways (68). The interaction between TnsE and the sliding clamp was found to be weak, a possible adaptation that might mitigate some of the consequences of interacting with an essential component in the cell (68).

The TnsABC+E transposition system was also reconstituted in an in vitro system and showed TnsABC+E-dependent transposition if the target plasmid had a 20-bp gap(68). Of note, plasmids were not used as targets for TnsABC+E-dependent transposition if they were only nicked. When the sliding-clamp protein was preloaded onto this substrate, insertion events were strongly biased to the same orientation as found with *in vivo* TnsABC+E transposition (68). In this assay, the sliding clamps are loaded onto the DNA substrate in a single orientation and are believed to preferentially reside at gaps (25). Consistent with this idea was the finding that a specific interaction with sliding clamps on the lagging-strand template directed insertions in a single orientation (68). Although the in vitro reaction seems to recapitulate the minimum requirements for TnsE-mediated transposition, there are likely to be other components in the system. The in vitro reaction required gain-of-activity mutants that have ~1,000-fold higher transposition frequencies than wild-type proteins (68). These mutants showed the same bias in transposition as the wild type (73). It is unclear whether these mutants simply amplify a low

signal or whether they are compensating for an unknown component that normally must be present for transposition *in vivo*.

TnsE-mediated transposition not only targets the lagging-strand template, but these events are also strongly biased to regions of natural terminators of DNA replication (72, 73). As explained above with IS*608*, transposition is stimulated at sites where DNA replication terminates. A similar increase in the frequency of transposition in response to interfering with DNA replication is also observed in other elements, as explained below. Perturbing DNA replication forks also appears to make the lagging-strand template more vulnerable to TnsE-mediated transposition. These together suggest that forks become more vulnerable under periods of replication stress. Enhanced targeting of TnsE-mediated transposition events to conjugal plasmids could also be due to a loss of some type of protection found with normal DNA replication forks. One possibility is that unknown features may protect replication forks where both the leading- and lagging-strand templates are processed in a coordinated fashion. The protection may be lost during conjugation in which only one strand is replicated in each cell. This result would also be consistent with the finding that the filamentous bacteriophage M13, which replicates in a process where the replication of both strands is not spatially coordinated, is also a target for TnsE-mediated Tn*7* transposition (19).

TnsABC+E transposition is also aggressively stimulated by replication events associated with DNA double-strand break (DSB) repair (72, 90). Repair of DSBs in bacteria usually involves initiating DNA replication with one of the broken ends using homologous recombination (reviewed in 41, 53). In *E. coli*, RecBCD exonuclease loads onto a DSB, degrading both strands of dsDNA until it encounters a specific DNA sequence called a *chi* site. *Chi* sites are recognized only in one orientation and are overrepresented in the chromosome in

such a way that would be expected to quickly facilitate the reestablishment of DNA replication forks progressing in the normal direction toward the terminus region. After RecBCD engages a *chi* site, it then degrades from the 5′ end while actively loading RecA onto the 3′ end of extended ssDNA. RecA-coated ssDNA can invade a sister copy of the chromosome, forming a structure called a D-loop, in which one of the strands of the duplex DNA is displaced by the incoming ssDNA. The D-loop structure is recognized by the Pri proteins and associated factors that assemble and initiate a DNA replication fork from the 3′ end. Subsequent work with the system ruled out specific interactions with proteins involved in replication restart needed for repair and instead showed that replication initiated at the break was very likely a target (90). Of further interest, the insertion events primarily occurred at hot spots that were dependent on a regional DSB at an origin-proximal position. Hot spots for TnsE-mediated insertion were not found during normal DNA replication. These results suggest that not all DNA replication is the same with regard to Tn*7* transposition because normal DNA replication events initiated at *oriC* did not result in highly focused hot spots for Tn*7*, but DNA replication initiated for DNA repair resulted in highly active hot spots that attracted most of the transposition events. This would suggest that Tn*7* might be sensitive to differences in the replisome that stem from how they were originally initiated. It is also possible that proteins expressed during DNA damage alter the replisome, something that has been suggested in other work (35, 49, 56). A third possibility is that other changes in the cell may alter the ability of the replisome to proceed.

TnsE-mediated transposition appears to be specifically adapted to target transposition into the lagging-strand template, especially in regions where replication forks tend to stall or during atypical DNA replication, such as that found during conjugation or replication-mediated DSB repair. Like Tn*7* and IS*608*, there are examples in which other very different types of DNA

transposons preferentially transpose into replication forks that are actively stopped. For example, IS*903* insertion in *E. coli* shows strong biases to places where DNA replication terminates and shows an orientation dependency during conjugation (31, 95). A bias for transposition to a region of replication termination can also be found in some Firmicutes with Tn*917*. This bias may also involve an interaction with the sliding clamp because the Tn*917* transposase contains a putative sliding-clamp-interacting motif (24, 89) (also see below). A major difference between Tn*7* and the other elements discussed above is that Tn*7* appears to choose to actively target the lagging-strand template without an obvious need for a feature (i.e., ssDNA and/or a 3′ OH end) found on this strand for cut-and-paste transposition.

1.4 Is the Lagging-Strand Template Vulnerable, and are Molecular Systems in place to Reduce this Vulnerability?

Features of the lagging-strand template provide an opportunity for the mobilization of a variety of genetic elements. Other processes could also be considered vulnerabilities specific to the lagging-strand template. For example, the frequency of mutation differs depending on the placement of *lacZ* alleles on the leading- verses lagging-strand template (18). Constitutive expression of the bacterial DNA damage response genes (the SOS response) can magnify this effect, possibly by an ability of the increased levels of activated RecA found during the SOS response overwhelming SSB or by the fact that the discontinuous nature of replication on the lagging-strand template allows more opportunities for SOS-induced DNA polymerases to pirate free 3′ OH ends. Hairpins formed from inverted repeats are normally not energetically favorable in dsDNA but may be favored when a stretch of ssDNA is available on the lagging-strand template, making it vulnerable to processing by enzymes that cleave hairpins like SbcCD (16). In yeast, triplet repeats, which are able to form hairpins, are more unstable on the lagging-strand

template than on the leading-strand template (21), something that may relate to the abundance of ssDNA on this strand and the ability of these sequences to be expanded or deleted during Okazaki-fragment processing (38). Although stretches of ssDNA, free 3′ OH ends, and free sliding-clamp proteins may occasionally be found on the leading-strand template (112), they are more common on the lagging-strand template by the nature of discontinuous replication. Therefore, it seems fair to consider whether these features represent potential vulnerabilities for the host. This characterization seems especially appropriate given that these features are preferentially targeted by mobile elements in certain atypical situations when replication is perturbed, actively terminated, or initiated as a result of DSB repair, or during replication initiated by other genetic elements. In the following sections, we address the idea that there may be specific benefits for mobile elements to target features on the lagging-strand template and that there may be molecular systems in hosts that help protect these features.

1.4.1 Is it advantageous for mobile elements to respond to DNA replication?

Many transposons have been shown to upregulate their movement during DNA replication. This is an important advantage for transposons that use cut-and-paste transposition in that the DSB created in donor DNA at the site left by the transposon can then be repaired by homologous recombination from the sister chromosome. Because the sister chromosome would still have a copy of the transposon, this process would also re-establish the element at the site that it vacated. A second benefit would be the ability to test a new insertion site in only one daughter cell: if insertions occur into an essential gene, only the daughter cell with the chromosome that received the insertion would be lost. Specific molecular systems have been identified for upregulating transposition after replication (81, 113) and an association of transposition with replication has been shown with some eukaryotic elements (82, 92). In a

variety of other cases, eukaryotic transposases have been found to interact with the sliding clamp or to contain putative sliding-clamp-interacting motifs (17, 68, 96, 105, 106). Association with the sliding clamp could act as a way to coordinate other interactions on DNA (96). Alternatively, association with sliding clamps could provide a mechanism to regulate transposition. This stems from the fact that sliding clamps are only found loaded on DNA during DNA replication and during certain DNA repair events. Therefore, transposons that require an interaction with sliding clamps on DNA would be active only during these times and at these places. In the case of bacteria, if DNA replication associated with conjugal plasmid transfer and bacteriophage replication is readily used by mobile genetic elements, these elements would also have the advantage of facilitated horizontal transfer, as is found with TnsE-mediated Tn*7* transposition.

1.4.2 Is replication fork stress an important indicator of host stress?

Genetic elements have been suggested to monitor the growth state and stress level of their host. Perhaps the most classic example is the activation of the lytic cycle of bacteriophage lambda following SOS induction. In this process, ssDNA-bound RecA acts as the SOS signaling molecule, inducing cleavage of the lambda repressor, thereby initiating the lysis program (9, 23). This feature of induced mobility in response to host stress is also observed by other mobile elements. Molecular systems exist that allow transposons to eavesdrop on the metabolic state of the host, permitting them to increase the frequency of transposition in cells that are subject to nutrient stress (4-6, 10, 88, 95, 98, 107). DNA replication is also acutely sensitive to many natural processes on the DNA (61), and the elongation phase of DNA replication is actively regulated in response to nutrient stress with a variety of molecular mechanisms (11, 55, 71, 83). Therefore, genetic elements that move more frequently when DNA replication is perturbed also end up mobilizing more frequently under the same stress conditions that affect replication.

If replication found during DNA DSB repair is differentially recognized by some genetic elements, there could be consequences on the frequency of transposition in natural populations. Unlike bacterial growth in the laboratory, the majority of natural environments are nutrient limited, and it is generally accepted that there are protracted periods of slow or no growth (40a, 82a). Under these conditions, DNA replication associated with the duplication of cells is limited; however, DNA replication associated with DNA repair may take on a more dominant role. Given these proposed environmental DNA replication patterns, transposition rates measured in the laboratory using actively growing cells may be misleading for understanding the mobility of genetic elements that are sensitive to the type of DNA replication. In addition, natural transformation also provides an opportunity for repair-initiated replication when ssDNA fragments are integrated into the genome (51, 102). Therefore, if replication associated with DNA repair is a more general target found with some genetic elements, it could also be an indicator of cell stress in multiple distinct ways.

1.4.3 Do cells have distinct mechanisms to allow protection?

If the lagging-strand template is vulnerable, it seems reasonable to ask whether systems are in place to protect the lagging-strand template. Technically, any system that helps to ensure the stability and orderly progress of replication forks would also limit access to the types of genetic elements discussed above (**Figure 1**). Ahead of the replication fork, gyrase is responsible for removing positive supercoils. Behind the replication fork, precatenanes can accumulate and must be unlinked by Topo IV for the sister chromosomes to eventually be segregated (104). Unlinking of chromosomes via Topo IV appears to be coordinated with the bridging activity of SeqA between (and/or within) sister chromosomes in a process that is also affected by the eventual condensation of DNA by the MukBEF system (37, 67). The eventual segregation of the

two sister chromosomes also appears to be a highly regulated process (103, 109). Organizing processes controlling supercoiling and strand separation, coupling DNA replication on both template strands, and orchestrating primer removal, gap filling, and ligation before the chromosomes are unlinked and segregated may limit many negative outcomes, including access to mobile genetic elements. The coordination of these features might also provide a steric barrier to proteins and nucleic acids from horizontally acquired genetic elements. Although it would be hard to implicate any part of replication and segregation systems as processes that evolved specifically to protect the chromosome from genetic elements, it will be interesting to know whether specific molecular systems in these elements have evolved to disrupt aspects of replication as a mechanism to get access to the lagging-strand template. In a practical sense, any mechanisms that we uncover in mobile genetic elements may find use as tools to help manipulate bacterial chromosomes.

1. 5 Acknowledgments

Work in the Peters lab is funded by the National Science Foundation (MCB-1244227).

REFERENCES

1. Bidnenko V, Ehrlich SD, Michel B. 2002. Replication fork collapse at replication terminator sequences. *EMBO J*. 21(14):3898--907

2. Bobay LM, Touchon M, Rocha EPC. 2013. Manipulating or superseding host recombination functions: a dilemma that shapes phage evolvability. *PLOS Genet*. 9(9):e1003825

3. Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat. Rev. Microbiol*. 11(8):525--38

4. Claverys JP, Prudhomme M, Martin B. 2006. Induction of competence regulons as a general response to stress in Gram-positive bacteria. *Annu. Rev. Microbiol*. 60(1):451--75

5. Coros AM, Twiss E, Tavakoli NP, Derbyshire KM. 2005. Genetic evidence that GTP is required for transposition of IS*903* and Tn*552* in *Escherichia coli*. *J. Bacteriol*. 187(13):4598--606

6. Coros CJ, Piazza CL, Chalamcharla VR, Smith D, Belfort M. 2009. Global regulators orchestrate group II intron retromobility. *Mol. Cell* 34(2):250--56

7. Costantino N, Court DL. 2003. Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. USA* 100(26):15748--53

8. Cousineau B, Lawrence S, Smith D, Belfort M. 2000. Retrotransposition of a bacterial group II intron. *Nature* 404(6781):1018--21

8a. Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, et al. 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* 94(4):451--62

9. Craig NL, Roberts JW. 1980. *E. coli* RecA protein-directed cleavage of phage λ repressor requires polynucleotide. *Nature* 283(5742):26--30

9a. Dalrymple BP, Kongsuwan K, Wijffels G, Dixon NE, Jennings PA. 2001. A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc. Natl. Acad. Sci. USA* 98(20):11627--32

10. DeBoy RT, Craig NL. 2000. Target site selection by Tn*7*:*att*Tn*7* transcription and target activity. *J Bacteriol.* 182(11):3310--13

11. DeNapoli J, Tehranchi AK, Wang JD. 2013. Dose-dependent reduction of replication elongation rate by (p)ppGpp in *Escherichia coli* and *Bacillus subtilis*. *Mol. Microbiol.* 88(1):93--104

12. De Septenville AL, Duigou S, Boubakri H, Michel B. 2012. Replication fork reversal after replication-transcription collision. *PLOS Genet.* 8(4):e1002622

13. Duderstadt KE, Reyes-Lamothe R, van Oijen AM, Sherratt DJ. 2014. Replication-fork dynamics. *Cold Spring Harb. Perspect. Biol.* 6(1):a010157

14. Duggin IG, Wake RG, Bell SD, Hill TM. 2008. The replication fork trap and termination of chromosome replication. *Mol. Microbiol.* 70(6):1323--33

15. Ellis HM, Yu D, DiTizio T, Court DL. 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA* 98(12):6742--46

16. Eykelenboom JK, Blackwood JK, Okely E, Leach DRF. 2008. SbcCD causes a double-strand break at a DNA palindrome in the *Escherichia coli* chromosome. *Mol. Cell* 29(5):644--51

17. Feschotte C, Mouchès C. 2000. Evidence that a family of miniature inverted-repeat transposable elements (MITES) from the *Arabidopsis thaliana* genome has arisen from a *pogo*-like DNA transposon. *Mol. Biol. Evol*. 17(5):730--37

18. Fijalkowska IJ, Jonczyk P, Tkaczyk MM, Bialoskorska M, Schaaper RM. 1998. Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc. Natl. Acad. Sci. USA* 95(17):10020--25

19. Finn JA, Parks AR, Peters JE. 2007. Transposon Tn*7* directs transposition into the genome of filamentous bacteriophage M13 using the element-encoded TnsE protein. *J. Bacteriol.* 189(24):9122--25

20. Frank AC, Lobry JR. 1999. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* 238(1):65--77

21. Freudenreich CH, Stavenhagen JB, Zakian VA. 1997. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol*. 17(4):2090--98

22. Fukui K. 2010. DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids* 27:260512

23. Galkin VE, Yu X, Bielnicki J, Ndjonka D, Bell CE, Egelman EH. 2009. Cleavage of bacteriophage λ CI repressor involves the RecA C-terminal domain. *J. Mol. Biol*. 385(3):779--87

24. Garsin DA, Urbach J, Huguet-Tapia JC, Peters JE, Ausubel FM. 2004. Construction of an *Enterococcus faecalis* Tn*917*-mediated-gene-disruption library offers insight into Tn*917* insertion patterns. *J. Bacteriol*. 186(21):7280--89

25. Georgescu RE, Kim SS, Yurieva O, Kuriyan J, Kong XP, O'Donnell M. 2008. Structure of a sliding clamp on DNA. *Cell* 132(1):43--54

26. Georgescu RE, Kurth I, O'Donnell ME. 2012. Single-molecule studies reveal the function of a third polymerase in the replisome. *Nat. Struct. Mol. Biol*. 19(1):113--16

27. Guo H, Zimmerly S, Perlman PS, Lambowitz AM. 1997. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J.* 16(22):6835--48

28. Guynet C, Hickman AB, Barabas O, Dyda F, Chandler M, Ton-Hoang B. 2008. In vitro reconstitution of a single-stranded transposition mechanism of IS*608*. *Mol. Cell* 29(3):302--12

30. Heller RC, Marians KJ. 2006. Replication fork reactivation downstream of a blocked nascent leading strand. *Nature* 439(7076):557--62

30a. Hillyar CRT. 2012. Genetic recombination in bacteriophage lambda. *Biosci. Horiz.* 5:hzs001

31. Hu WY, Derbyshire KM. 1998. Target choice and orientation preference of the insertion sequence IS*903*. *J. Bacteriol.* 180(12):3039--48

32. Huen MSY, Li X, Lu LY, Watt RM, Liu DP, Huang JD. 2006. The involvement of replication in single stranded oligonucleotide-mediated gene repair. *Nucleic Acids Res.* 34(21):6183--94

33. Ichiyanagi K, Beauregard A, Belfort M. 2003. A bacterial group II intron favors retrotransposition into plasmid targets. *Proc. Natl. Acad. Sci. USA* 100(26):15742--47

34. Ichiyanagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M. 2002. Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.* 46(5):1259--72

35. Indiani C, Patel M, Goodman MF, O'Donnell ME. 2013. RecA acts as a switch to regulate polymerase occupancy in a moving replication fork. *Proc. Natl. Acad. Sci. USA* 110(14):5410--15

36. Iyer RR, Pluciennik A, Burdett V, Modrich PL. 2006. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* 106(2):302--23

37. Joshi MC, Magnan D, Montminy TP, Lies M, Stepankiw N, Bates D. 2013. Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. *PLOS Genet*. 9(8):e1003673

38. Kantartzis A, Williams GM, Balakrishnan L, Roberts RL, Surtees JA, Bambara RA. 2012. Msh2-Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions. *Cell Rep*. 2(2):216--22

39. Kersulyte D, Velapatiño B, Dailide G, Mukhopadhyay AK, Ito Y, et al. 2002. Transposable element ISHp*608* of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity. *J. Bacteriol*. 184(4):992--1002

40. Kim S, Dallmann HG, McHenry CS, Marians KJ. 1996. Coupling of a replicative polymerase and helicase: a τ-DnaB interaction mediates rapid replication fork movement. *Cell* 84(4):643--50

40a. Kolter R, Siegele DA, Tormo A. 1993. The stationary phase of the bacterial life cycle. *Annu. Rev. Microbiol.* 47:855--74

41. Kowalczykowski SC. 2000. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci*. 25(4):156--65

42. Lajoie MJ, Gregg CJ, Mosberg JA, Washington GC, Church GM. 2012. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res.* 40(22):e170

43. Lambowitz AM, Zimmerly S. 2010. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol*. 3(8):a003616

44. Leu FP, Georgescu R, O'Donnell M. 2003. Mechanism of the *E. coli* tau processivity switch during lagging-strand synthesis. *Mol. Cell* 11(2):315--27

45. Li X, Costantino N, Lu L, Liu D, Watt RM, et al. 2003. Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*. *Nucleic Acids Res.* 31(22):6674--87

46. Li X, Thomason LC, Sawitzke JA, Costantino N, Court DL. 2013. Bacterial DNA polymerases participate in oligonucleotide recombination. *Mol. Microbiol*. 88(5):906--20

47. Li Z, Craig NL, Peters JE. 2013. Transposon Tn*7*. In *Bacterial Integrative Mobile Genetic Elements,* ed. AP Roberts, P Mullany, pp. 1--32. Austin, TX: Landes Biosci.

48. Lia G, Michel B, Allemand JF. 2012. Polymerase exchange during Okazaki fragment synthesis observed in living cells. *Science* 335(6066):328--31

49. Lia G, Rigato A, Long E, Chagneau C, Le Masson M, et al. 2013. RecA-promoted, RecFOR-independent progressive disassembly of replisomes stalled by helicase inactivation. *Mol. Cell* 49(3):547--57

50. Lopez de Saro FJ, O'Donnell M. 2001. Interaction of the β sliding clamp with MutS, ligase, and DNA polymerase I. *Proc. Natl. Acad. Sci. USA* 98(15):8376--80

51. Lorenz MG, Wackernagel W. 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev*. 58(3):563--602

52. Lundquist RC, Olivera BM. 1982. Transient generation of displaced single-stranded DNA during nick translation. *Cell* 31(1):53--60

53. Lusetti SL, Cox MM. 2002. The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu. Rev. Biochem*. 71:71--100

54. Lyamichev V, Brow MA, Dahlberg JE. 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 260(5109):778--83

55. Maciąg M, Kochanowska M, Łyżeń R, Węgrzyn G, Szalewska-Pałasz A. 2010. ppGpp inhibits the activity of *Escherichia coli* DnaG primase. *Plasmid* 63(1):61--67

56. Maliszewska-Tkaczyk M, Jonczyk P, Bialoskorska M, Schaaper RM, Fijalkowska IJ. 2000. SOS mutator activity: unequal mutagenesis on leading and lagging strands. *Proc. Natl. Acad. Sci. USA* 97(23):12678--83

57. Martínez-Abarca F, Barrientos-Durán A, Fernández-López M, Toro N. 2004. The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res*. 32(9):2880--88

58. Martinsohn JT, Radman M, Petit MA. 2008. The λ Red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLOS Genet*. 4(5):e1000065

59. Matsuura M, Saldanha R, Ma H, Wank H, Yang J, et al. 1997. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev*. 11(21):2910--24

60. Mills DA, McKay LL, Dunny GM. 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J. Bacteriol.* 178(12):3531--38

61. Mirkin EV, Mirkin SM. 2007. Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev*. 71(1):13--35

62. Mohr G, Smith D, Belfort M, Lambowitz AM. 2000. Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev*. 14(5):559--73

63. Mosberg JA, Lajoie MJ, Church GM. 2010. Lambda Red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* 186(3):791--99

63a. Muñoz-Adelantado E, San Filippo J, Martínez-Abarca F, García-Rodríguez FM, Lambowitz AM, Toro N. 2003. Mobility of the *Sinorhizobium meliloti* group II intron RmInt1 occurs by reverse splicing into DNA, but requires an unknown reverse transcriptase priming mechanism. *J. Mol. Biol*. 327(5):931--43

64. Murphy KC. 1991. Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme. *J. Bacteriol*. 173(18):5808--21

64a. Murphy KC. 2012. Phage recombinases and their applications. In *Advances in Virus Research*, ed. M Łobocka, W Szybalski, pp. 367--414. Waltham, MA: Acad. Press

65. Myers RS, Stahl FW. 1994. Chi and the RecBCD enzyme of *Escherichia coli*. *Annu. Rev. Genet*. 28:49--70

66. Neylon C, Kralicek AV, Hill TM, Dixon NE. 2005. Replication termination in *Escherichia coli*: structure and antihelicase activity of the Tus-*ter* complex. *Microbiol. Mol. Biol. Rev*. 69(3):501--26

67. Nicolas E, Upton AL, Uphoff S, Henry O, Badrinarayanan A, Sherratt D. 2014. The SMC complex MukBEF recruits topoisomerase IV to the origin of replication region in live *Escherichia coli*. *mBio* 5(1):e01001-13

68. Parks AR, Li Z, Shi Q, Owens RM, Jin MM, Peters JE. 2009. Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell* 138(4):685--95

69. Parks AR, Peters JE. 2009. Tn*7* elements: engendering diversity from chromosomes to episomes. *Plasmid* 61(1):1--14

70a. Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M, Sommer S. 2010. Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence. *PLOS Genet*. 6(1):e1000799

71. Persky NS, Ferullo DJ, Cooper DL, Moore HR, Lovett ST. 2009. The ObgE/CgtA GTPase influences the stringent response to amino acid starvation in *Escherichia coli*. *Mol. Microbiol.* 73(2):253--66

72. Peters JE, Craig NL. 2000. Tn*7* transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Mol. Cell* 6(3):573--82

73. Peters JE, Craig NL. 2001. Tn*7* recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev.* 15(6):737--47

74. Peters JE, Craig NL. 2001. Tn*7*: smarter than we thought. *Nat. Rev. Mol. Cell Biol.* 2(11):806--14

75. Pluciennik A, Burdett V, Lukianova O, O'Donnell M, Modrich P. 2009. Involvement of the β clamp in methyl-directed mismatch repair in vitro. *J. Biol. Chem.* 284(47):32782--91

76. Pomerantz RT, O'Donnell M. 2008. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature* 456(7223):762--66

77. Possoz C, Filipe SR, Grainge I, Sherratt DJ. 2006. Tracking of controlled *Escherichia coli* replication fork stalling and restart at repressor-bound DNA *in vivo*. *EMBO J.* 25(11):2596--604

78. Postow L, Ullsperger C, Keller RW, Bustamante C, Vologodskii AV, Cozzarelli NR. 2001. Positive torsional strain causes the formation of a four-way junction at replication forks. *J. Biol. Chem.* 276(4):2790--96

79. Poteete AR, Fenton AC. 1993. Efficient double-strand break-stimulated recombination promoted by the general recombination systems of phages λ and p22. *Genetics* 134(4):1013--21

80. Reyes-Lamothe R, Sherratt DJ, Leake MC. 2010. Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science* 328(5977):498--501

81. Roberts D, Hoopes BC, McClure WR, Kleckner N. 1985. IS*10* transposition is regulated by DNA adenine methylation. *Cell* 43(1):117--30

81a. Roitzsch M, Pyle AM. 2009. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA* 15(3):473--82

82. Ros F, Kunze R. 2001. Regulation of activator/dissociation transposition by replication and DNA methylation. *Genetics* 157(4):1723--33

82a. Roszak DB, Colwell RR. 1987. Survival strategies of bacteria in the natural environment. *Microbiol. Rev*. 51(3):365--79

83. Rymer RU, Solorio FA, Tehranchi AK, Chu C, Corn JE, et al. 2012. Binding mechanism of metal·NTP substrates and stringent-response alarmones to bacterial DnaG-type primases. *Structure* 20(9):1478--89

84. Sandler SJ. 2000. Multiple genetic pathways for restarting DNA replication forks in *Escherichia coli* K-12. *Genetics* 155(2):487--97

85. San Filippo J, Lambowitz AM. 2002. Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J. Mol. Biol.* 324(5):933--51

86. Sawitzke JA, Costantino N, Li X, Thomason LC, Bubunenko M, et al. 2011. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J. Mol. Biol.* 407(1):45--59

87. Segawa T, Tomizawa J. 1971. Formation of concatemers of lambda phage DNA in a recombination-deficient system. *Mol. Gen. Genet.* 111(3):197--201

88. Sharpe PL, Craig NL. 1998. Host proteins can stimulate Tn*7* transposition: a novel role for the ribosomal protein L29 and the acyl carrier protein. *EMBO J.* 17(19):5822--31

89. Shi Q, Huguet-Tapia JC, Peters JE. 2009. Tn*917* targets the region where DNA replication terminates in *Bacillus subtilis*, highlighting a difference in chromosome processing in the firmicutes. *J. Bacteriol*. 191(24):7623--27

90. Shi Q, Parks AR, Potter BD, Safir IJ, Luo Y, et al. 2008. DNA damage differentially activates regional chromosomal loci for Tn*7* transposition in *Escherichia coli*. *Genetics* 179(3):1237--50

91. Singh NN, Lambowitz AM. 2001. Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J. Mol. Biol*. 309(2):361--86

92. Spradling AC, Bellen HJ, Hoskins RA. 2011. *Drosophila* P elements preferentially transpose to replication origins. *Proc. Natl. Acad. Sci. USA* 108(38):15948--53

93. Swingle B, Bao Z, Markel E, Chambers A, Cartinhour S. 2010. Recombineering using RecTE from *Pseudomonas syringae*. *Appl. Environ. Microbiol.* 76(15):4960-68

94. Swingle B, Markel E, Costantino N, Bubunenko MG, Cartinhour S, Court DL. 2010. Oligonucleotide recombination in Gram-negative bacteria. *Mol. Microbiol*. 75(1):138--48

95. Swingle B, O'Carroll M, Haniford D, Derbyshire KM. 2004. The effect of host-encoded nucleoid proteins on transposition: H-NS influences targeting of both IS*903* and Tn*10*. *Mol. Microbiol.* 52(4):1055--67

96. Taylor MS, LaCava J, Mita P, Molloy KR, Huang CRL, et al. 2013. Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* 155(5):1034--48

97. Ton-Hoang B, Pasternak C, Siguier P, Guynet C, Hickman AB, et al. 2010. Single-stranded DNA transposition is coupled to host replication. *Cell* 142(3):398--408

98. Twiss E, Coros AM, Tavakoli NP, Derbyshire KM. 2005. Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress: host factors and transposition. *Mol. Microbiol.* 57(6):1593--607

99. Van Kessel JC, Hatfull GF. 2007. Recombineering in *Mycobacterium tuberculosis*. *Nat. Methods* 4(2):147--52

100. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, et al. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460(7257):894--98

101. Wang HH, Xu G, Vonner AJ, Church G. 2011. Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res.* 39(16):7336--47

102. Wang JD, Sanders GM, Grossman AD. 2007. Nutritional control of elongation of DNA replication by (p)ppGpp. *Cell* 128(5):865--75

103. Wang X, Liu X, Possoz C, Sherratt DJ. 2006. The two *Escherichia coli* chromosome arms locate to separate cell halves. *Genes Dev.* 20(13):1727--31

104. Wang X, Reyes-Lamothe R, Sherratt DJ. 2008. Modulation of *Escherichia coli* sister chromosome cohesion by topoisomerase IV. *Genes Dev.* 22(17):2426--33

105. Warbrick E. 2000. The puzzle of PCNA's many partners. *Bioessays* 22(11):997--1006

106. Warbrick E, Lane DP, Glover DM, Heatherington W. 1998. PCNA binding proteins in *Drosophila melanogaster*: the analysis of a conserved PCNA binding domain. *Nucleic Acids Res.* 26(17):3925--32

107. Wardle SJ, O'Carroll M, Derbyshire KM, Haniford DB. 2005. The global regulator H-NS acts directly on the transpososome to promote Tn*10* transposition. *Genes Dev.* 19(18):2224--35

108. Waters VL, Guiney DG. 1993. Processes at the nick region link conjugation, t-DNA transfer and rolling circle replication. *Mol. Microbiol.* 9(6):1123--30

109. White MA, Eykelenboom JK, Lopez-Vernaza MA, Wilson E, Leach DRF. 2008. Non-random segregation of sister chromosomes in *Escherichia coli*. *Nature* 455(7217):1248--50

110. Wolkow CA, DeBoy RT, Craig NL. 1996. Conjugating plasmids are preferred targets for Tn*7*. *Genes Dev*. 10(17):2145--57

111. Yang Y, Sharan SK. 2003. A simple two-step, "hit and fix" method to generate subtle mutations in BACs using short denatured PCR fragments. *Nucleic Acids Res*. 31(15):e80

112. Yeeles JTP, Marians KJ. 2013. Dynamics of leading-strand lesion skipping by the replisome. *Mol. Cell* 52(6):855--65

113. Yin JCP, Krebs MP, Reznikoff WS. 1988. Effect of DAM methylation on Tn*5* transposition. *J. Mol. Biol*. 199(1):35--45

114. Yuan Q, McHenry CS. 2013. Cycling of the *E. coli* lagging strand polymerase is triggered exclusively by the availability of a new primer at the replication fork. *Nucleic Acids Res.* 42(3)1747--56

114a. Zhong J, Lambowitz AM. 2003. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.* 22(17):4555--65

114b. Zimmerly S, Guo H, Perlman PS, Lambowitz AM. 1995. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 82(4):545--54

CHAPTER 2

HETEROMERIC TRANSPOSASES: GENERATORS OF GENOMIC ISLANDS ACROSS DIVERSE BACTERIA*

2.1 Summary

Horizontally acquired genetic information in bacterial chromosomes accumulates in blocks termed genomic islands. Tn7-like transposons form genomic islands at a single programmed insertion site in bacterial chromosomes, *attTn7*. Transposition involves five transposon-encoded genes (*tnsABCDE*) including an atypical heteromeric transposase. One transposase subunit, TnsB, is from the large family of bacterial transposases, the second, TnsA, is related to endonucleases. A regulator protein, TnsC, functions with different target site selecting proteins and the TnsAB transposase. TnsD directs transposition into *attTn7*, while TnsE encourages horizontal transmission by targeting mobile plasmids. Recent work suggests that distantly related elements with heteromeric transposases exist with alternate targeting pathways that also facilitate the formation of genomic islands. Tn6230 and related elements can be found at a single position in a gene of unknown function (*yhiN*) in various bacteria as well as in mobile plasmids. Another group termed Tn6022-like elements form pathogenicity islands in the *Acinetobacter baumannii comM* gene. While Tn7 has TnsD, the other two transposons appear to have a protein likely related to TnsD. Furthermore, all three transposons seem to direct transpostion to a specific site in the genome, suggesting they share a basic mechanism for inserting into their respective loci. Bioinformatics work supports the idea that the Tn6022-like and Tn6230-like elements can also mobilize to other sites, including plasmids. Exciting future

*Parts of this chapter appear in a paper published in the journal Molecular Microbiology. Peters, J.E., Fricker, A.D., Kapili, B.J., Petassi, M.T. (2014). Heteromeric transposase elements: generators of genomic islands across diverse bacteria. All of the work presented in this chapter was conducted by A.D.F. Figures 2.1 and 2.2 are published in Peters, J.E., et al. (2014).

work will involve determining the molecular mechanisms that allow or facilitate these other types of targeting pathways.

Horizontal transfer between bacteria plays an important role in chromosome evolution. Horizontally acquired genes are typically found grouped together in blocks of sequence called pathogenicity islands, fitness islands, or more generally, genomic islands (16, 32). Genomic islands are not found in all strains within a species and when found the contents of these islands often differ between strains. Genomic islands are frequently associated with tRNA genes, presumably formed through the action of bacteriophage integration (8, 40). An iterative process of integration and subsequent inactivation and partial degradation of bacteriophages at a given position provides a plausible model for their formation. Transposition with a specialized transposon family, called Tn7-like elements, also forms genomic islands (34, 35). Transposons are discrete DNA segments that can move between positions within a genome. Tn7-like elements can form genomic islands at their single preferred attachment site (*attTn7*) (reviewed in (25)), where accretion of genetic information is likely catalyzed by their ability to maximize horizontal transfer by targeting mobile plasmids (Figure 2.1A) (51). In this chapter I discuss two families of elements that are related to, but distinct from, Tn7-like elements that form genomic islands at positions other than the chromosomal *attTn7* site. I term these transposons Tn6230-like (31, 36) and Tn6022-like (17, 41) based on representatives of these types of elements. These relatives of the Tn7-like elements appear to have properties for increasing their dispersal and also their own heterogeneity. However, little is known about the movement of these elements which can carry with them many genes including genes relevant to pathogenesis like antibiotic resistance, metal resistance, secretion systems, effectors, and many others.

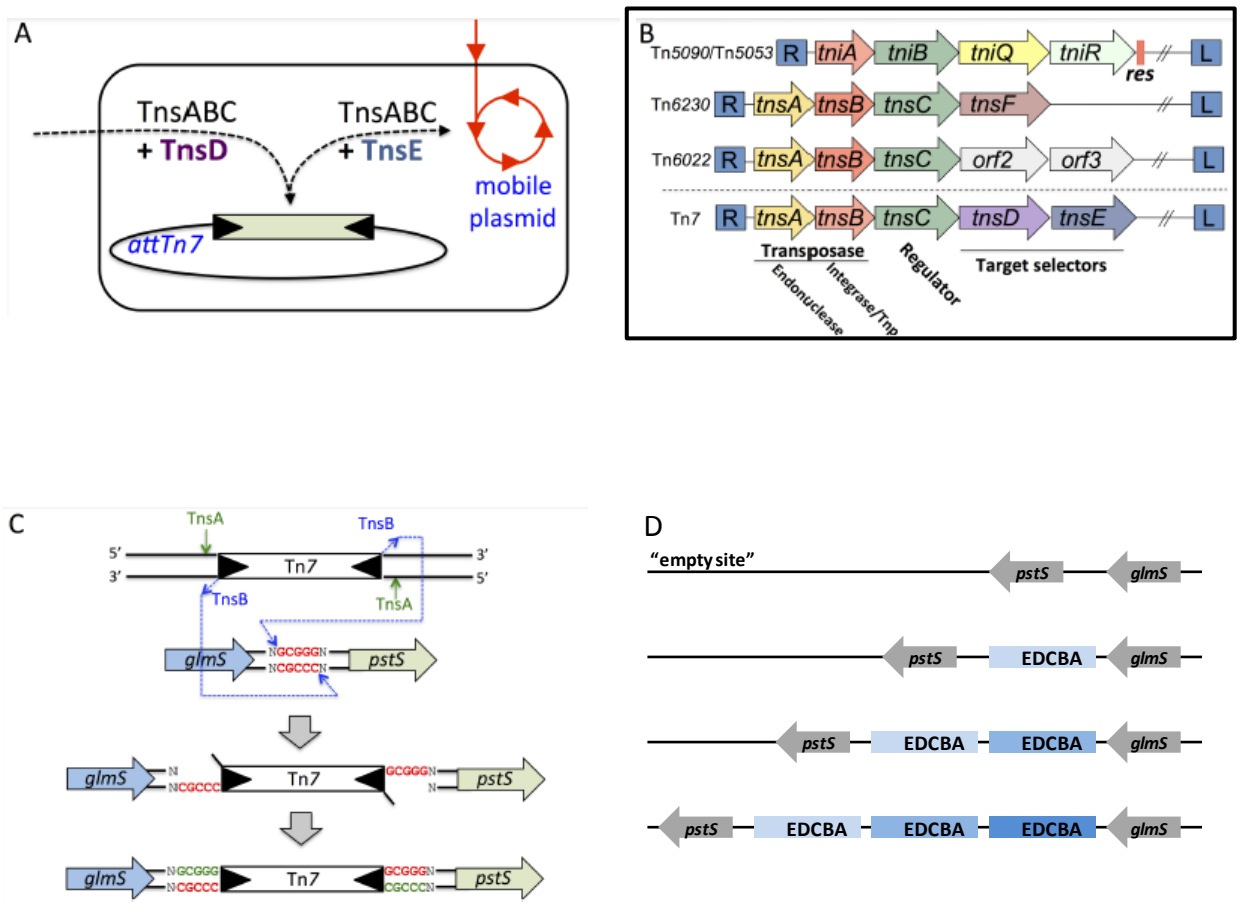

Figure 2.1

Features of Tn7 and other transposons.
A) Tn7 has two targeting pathways for transposition. TnsABC+TnsD transposition directs transposition into a single chromosomal site, *attTn7*, found in bacteria likely facilitating vertical transmission by targeting a neutral site. TnsABC+TnsE transposition preferentially directs transposition into mobile plasmids (shown in red) likely facilitating horizontal dissemination of the element. **B)** Tn5090/Tn5053 direct transposition using three proteins encoded by *tniABQ*, but require the action of a site-specific recombinse encoded by *tniR* to act at the element resolution site (*res*). Tn7, Tn6022, and Tn6320 encode a number of proteins as described in the text. Similar colored block arrows indicate proteins sharing conserved motifs. Note that *tnsA* of Tn6022 has previously been called *tniC* or *orf1*. **C)** Tn7 transposition with TnsABC+TnsD occurs into a single position located between *glmS* and *pstS* in *E. coli*. TnsB catalyzes breakage and joining events, while TnsA is responsible for making breaks in the flanking DNA at the 5' ends of the element allowing excision via a cut-and-paste mechanism. Gaps at the ends of the element result from the staggered joining events found with transposition that are repaired by DNA polymerase to form the target site duplication indicative of transposition ("old" DNA in red, "new" DNA in green). See text for details and references. **D)** Tn7 transposition can occur serially, that is each new insertion occurs proximal to the glmS gene.

2.2 Tn7-like Elements

Tn7 and Tn7-like elements are widespread across different types of bacteria found in a broad variety of environments and typically encode numerous genes that increase the fitness of the host and the element (35). Some functions encoded in Tn7-like elements include antibiotic resistance, non-ribosomal peptide synthetases, metal resistance, and a CRISPR system (35). The basic transposition machinery of Tn7 and how it directs transposition into the *attTn7* site is well understood and has been the subject of multiple reviews (10, 25, 36). Tn7 encodes five proteins required for two transposition pathways, TnsA, TnsB, TnsC, TnsD, and TnsE (TnsABCDE) (Figure 2.1B). TnsABC are required for both Tn7 transposition pathways, but only allow transposition when they function with one of two target site selecting proteins, TnsD or TnsE (49, 50). Transposition with TnsABC + TnsD directs transposition into the single *attTn7* site located downstream of *glmS* (Figures 2.1A-C) (14). Transposition with TnsABC + TnsE occurs preferentially into mobile plasmids through the ability of the TnsE protein to recognize features found enriched during DNA replication on the lagging-strand template (37, 33).

Tn7-like elements are the only elements characterized to date in which the transposase is comprised of two proteins, TnsA and TnsB, which function together to carry out the chemistry that underlies transposition (28, 44). It has been shown recently that the interaction between these two proteins is important for regulating transposition (9). TnsB belongs to a large family of proteins utilized by many bacterial transposons and retroviruses sometimes referred to as the DDE recombinase (*rve*)(pfam00665)(11, 38). TnsB recognizes DNA sequences found in the left and right ends of the element and catalyzes breaking events at the 3' ends of the element that are joined directly to target DNA (Figure 2.1C) (2, 29). TnsA is an endonuclease that makes breaks in the flanking DNA at the 5' ends, thereby allowing the element to be completely excised from

the donor DNA (18, 28). The ability to liberate both the 3' and 5' ends and join the 3' ends of the element directly to the target DNA allows Tn7 (and presumably other heteromeric transposase elements) to carry out cut-and-paste transposition in a distinct way from elements with a single transposase that form hairpin structures during excision (6, 20, 43). Staggered joining events with the target DNA during transposition results in gaps at the ends of the element that are filled by DNA polymerase forming a target site duplication that is characteristic of each element (5 bp in the case of Tn7) (Figure 2.1C) (3). As explained below, this feature of transposition is useful for determining which ends of the element were involved in a concerted insertion event.

TnsC is a AAA protein that communicates to the TnsAB transposase that an appropriate insertion site has been recognized (25, 48). There are a number of elements that utilize AAA proteins for regulating transposition; one that is particularly well-characterized is MuB from bacteriophage Mu (5, 52). Work with MuB highlights an additional important role for these proteins in helping to engage the target DNA for transposition (7, 27). Transposon Tn5090/Tn5053 also encodes a protein with ATP binding motifs, TniB, which likely helps regulate transposition (Figure 2.1B) (21, 39). Work with Tn7 indicates that TnsC is recruited to the *attTn7* site by TnsD (4, 23). TnsC is important for another regulatory role in which it inhibits transposition into a plasmid or a region of the chromosome that already has a copy of the element in a process called target immunity (45, 46).

TnsD recognizes a specific sequence within the 3’ coding region of the essential glucosamine-6-phosphate synthase (*glmS*) gene (4, 23, 30). However the actual point of insertion is approximately 25 bp downstream allowing recognition of a conserved sequence without disruption of an essential host gene. Moreover the region within the *glmS* gene that is recognized by TnsD encodes the active site of GlmS, and therefore provides a site that is highly

conserved between organisms. The ability of TnsD to bind DNA involves, in part, a CCCH zinc-finger motif, but the capacity for DNA recognition may reside throughout the protein (30). As described above, transposition of more than one element into *attTn7* is inhibited by target immunity, but distantly related Tn7-like elements can be found to accumulate "serially" in *attTn7*. In Tn7, the proteins encoded by the distantly related Tn7-like elements no longer recognize each other's components that are required for the immunity process, and are therefore capable of recognizing and inserting into the *attTn7* (Figure 2.1D) (34). Subsequent loss of essential transposition functions in elements integrated in *attTn7* helps explain the formation of genomic islands at this site where in some cases little or no evidence of the original elements can be found.

A second target site selection pathway that contributes to the ability of Tn7-like elements to generate genomic islands is the TnsE pathway. Unlike the TnsD-mediated pathway that recognizes a specific sequence, TnsABC + TnsE direct transposition by recognizing features of DNA replication found with the lagging-strand template; 3' recessed end structures and the sliding-clamp protein of DNA polymerase (called DnaN)(33, 37). TnsE-mediated transposition directs transposition preferentially into actively replicating mobile plasmids (Figure 2.1A) (37, 51). This ability to insert into mobile plasmids enables horizontal transfer of the element providing a mechanism for Tn7-like elements to accumulate genes as they travel between bacterial hosts.

2.3 New Heteromeric Transposase Elements form Genome Islands

Recently elements have been described with homologs of TnsA, TnsB, and TnsC that lack full-length homologs of TnsD or TnsE (31, 41). In these cases, it appears that the basic structure of the heteromeric transposase with a regulator protein has been adapted to new

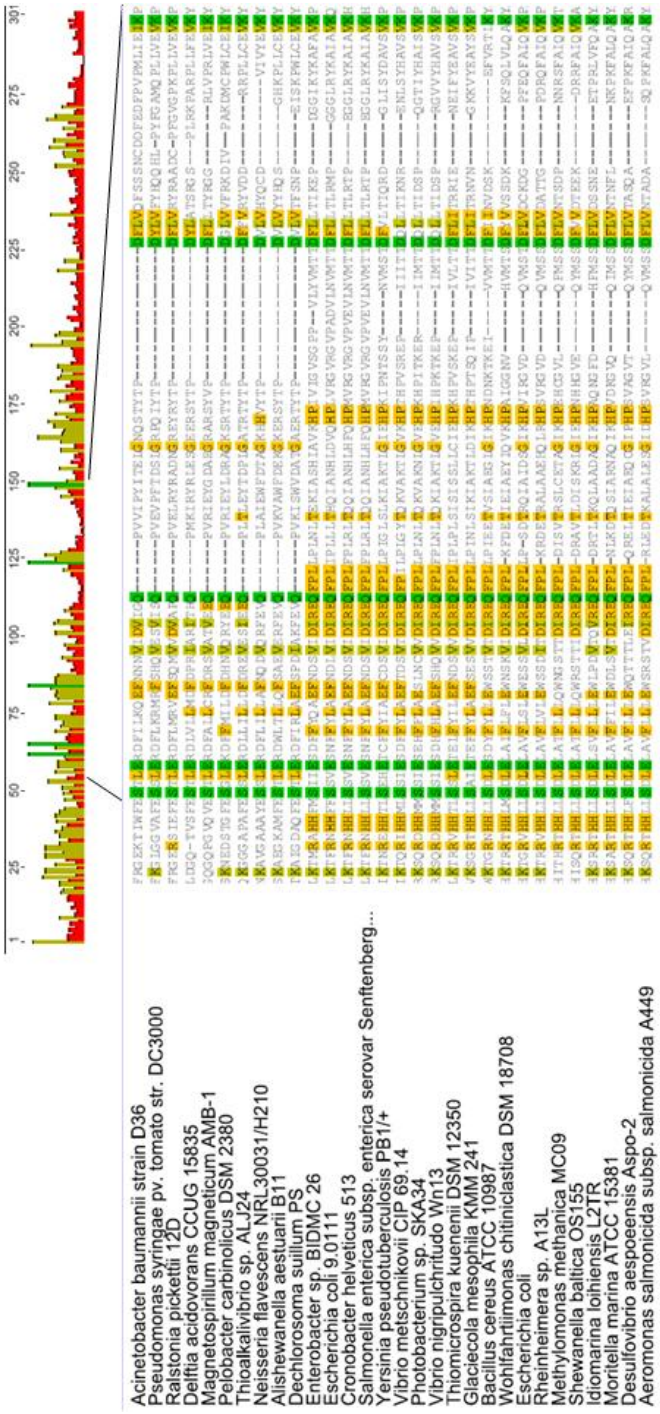

Figure 2.2

Alignment of TnsA homologs reveals four conserved residues
A consensus representation of the ~250 amino acid TnsA protein is represented by the bars at the top, where the green color and taller bar indicates greater conservation. Alignment of residues ~50-170 from all homologs is shown. Sites where the exact residue is conserved across all homologs is indicated by a green color.

targeting pathways in elements that is refered to more broadly as heteromeric transposase elements. TnsA proteins can be identified in genome databases by a conserved motif found in the protein (*Tn7_Tnp_TnsA_N*) (pfam08722)(18, 38). The number of transposons identified using this conserved motif suggests that a large and diverse group of heteromeric transposase elements exists (data not shown). To gather a manageable collection of putative elements for analysis, I performed BLAST searches with TnsA from the canonical Tn7 from *E. coli,* and candidate heteromeric elements that appeared highly diverged from Tn7 in *A. baumannii* and *Salmonella enterica,* (31, 41). Ten candidates from each of these three groups were chosen for further analyses where the amino acid sequence of TnsA was used to infer a phylogeny of the elements (Figure 2.2 and 2.3A). All of the examples had the three catalytic residues found in TnsA from Tn7 (18), but two other TnsA residues were also found conserved, S60 and Q82, both of which are within the active site (Figure 2.2). S60 was previously identified as liganded to a Cl^- in the TnsA crystal structure (18). The exact role of Q82 is unknown, but is thought to coordinate water molecules in the array around one of the divalent metals (Fred Dyda, personal communication). Given the high conservation of both S60 and Q82, they are likely important for transposition. These TnsA-encoding elements also all had homologs of TnsB and TnsC supporting the idea that bona fide transposons were identified. Elements in the clade with Tn7 all encoded homologs of the five Tns proteins found in Tn7 and were inserted downstream of *glmS*, although in two cases, *I. loihinesis* and *B. cereus*, there were a few genes between *glmS* and the element. As explained below, the other clades in the TnsA phylogenetic tree define other heteromeric transposase elements that are capable of forming genomic islands at specific sites that are not at *attTn7* using unexplored targeting pathways.

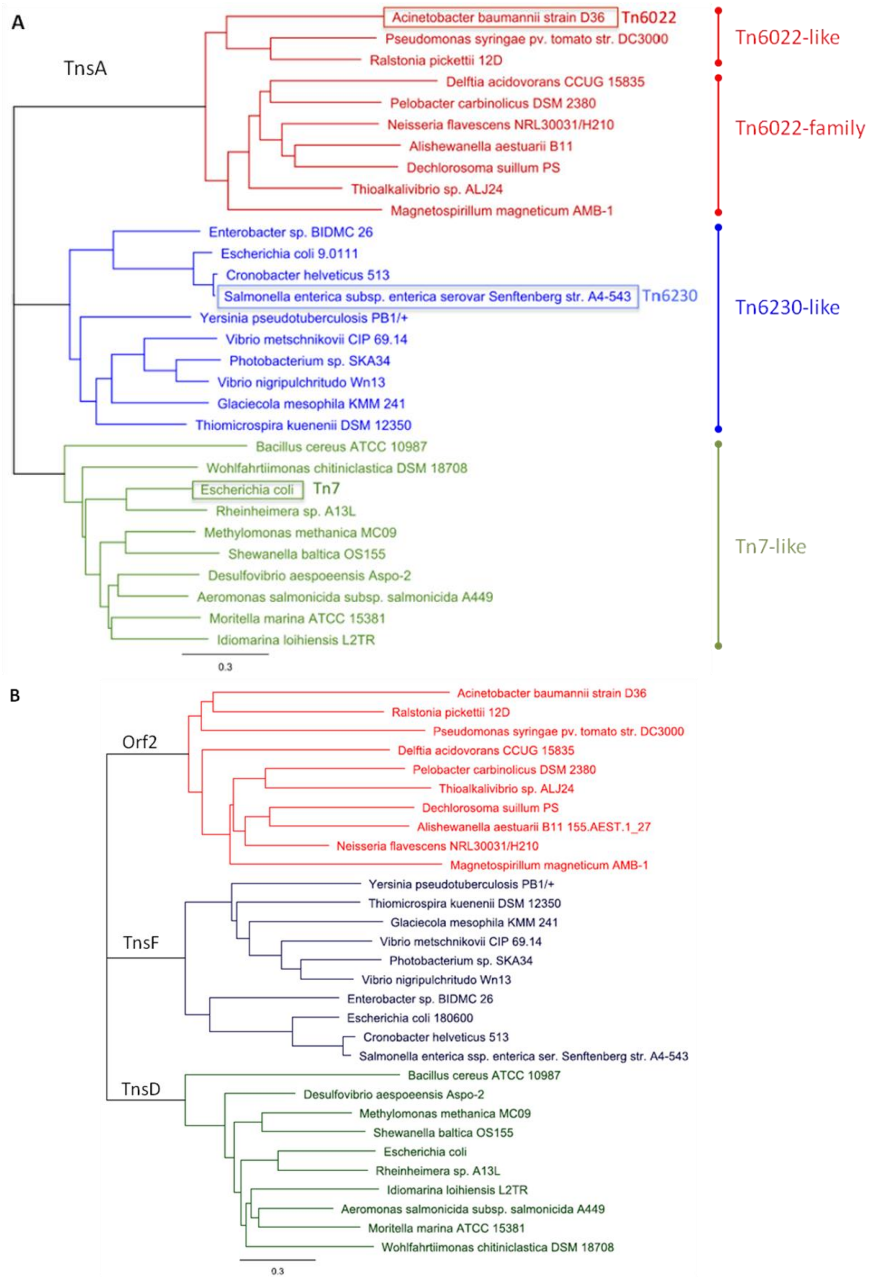

Figure 2.3

Neighbor Joining trees of selected protein homologs from Tn7, Tn6230, and Tn6022.Protein sequences were aligned using MUSCLE and a tree was drawn using Genieous tree builder. Homologs from Tn7 are colored in green, Tn6230 in blue, and Tn6022 in red. **A)** Homologs to TnsA from Tn7 (*Escherichia coli*), Tn6230 (*Salmonella enterica* subsp. enterica ser. Senftenberg), and Tn6022 (*Acinetobacter baumannii* D36) identified using BLAST searches. From each of the three TnsAs, 10 proteins that differed by greater than 10% were selected. See supplemental files for accession numbers and references. **B)** Homologs to TnsD, TnsF, and Orf2 from elements identified by TnsA similarity. From each element, the open reading frames were identified by pBLAST.

2.4 Genomic island formation in the *yhiN* gene with Tn6230-like elements

A bioinformatics screen for mobile elements in sequenced strains of *Salmonella* revealed an element with homologs of the *tnsABC* genes from Tn7-like elements that did not encode the target site selecting protein TnsE or a close homolog of TnsD (31). This 37 kb element that was subsequently named Tn6230 encodes over 30 genes including ones encoding heavy metal resistance and many of unknown function. Tn6230 lacks full-length homologs of the Tn7 target site selecting proteins, but encodes a protein with ~24% amino acid identity to TnsD found over a portion (~60%) of the protein including the zinc finger motif characterized in TnsD from Tn7; the protein alignment shown in Figure 2.4 showing the shared N-terminal and C-terminal domains of the two proteins. The N-terminus contains the homologous CCCH motif, highlighted in yellow in Figure 2.4. Of further interest, the Tn6230 protein was also found to be distantly related (20% identity) to the TniQ protein of Tn5090/Tn5053 elements including the TniQ superfamily motif (pfam06527) (21, 39). Where the TniQ proteins of Tn5090/5053 appear to be very similar to each other, this protein from Tn6230 appears to be more closely related to TnsD than the TniQ proteins (not shown). Together these data support a possible mechanistic link between different types of elements. Irefer to this protein as TnsF because it is distinct from TniQ and TnsD. Tn6230-like elements have a preferred insertion site, similar conceptually to the *attTn7* site, but in an unrelated sequence (31). Insertions of a nearly identical element were found in this same position in multiple different species (31). The Tn6230-like elements in Figure 2.3 (blue) encode homologs of the Tn6230 TnsABC proteins and a protein homologous to TnsF (blue, Figure 2.3B). The TnsF-like protein in the Tn6230-like elements studied here have all been shown to have a CCCH zinc finger motif (Figure 2.7B). Unlike *attTn7* where insertions occur downstream of *glmS* (Figure 2.1C), the attachment site for Tn6230 is within the coding

sequence of the *yhiN* gene, albeit only two to five codons inside the 3' end of the gene (36). Table 2.1 shows how closely related the upstream DNA sequence is to the same upstream region from the representative organism, confirming that all Tn7-like elements are downstream of *glmS*. This table also shows that while the Tn6230 element is immediately downstream of the gene, seven of the Tn6230-like elements are within the coding sequence of *yhiN*, but the other two in non-*yhiN* sequences. The *yhiN* gene is a nonessential putative FAD/NAD(P) binding oxidoreductase in *E. coli*. It is unknown if insertion into the very C-terminal encoding portion of the *yhiN* gene inactivates or otherwise alters the activity of the protein. Most (7/10) of the Tn6230-like elements were located at the predicated 3' terminus of *yhiN*. Interestingly, the element in *S. eneterica* Senftenberg appears to have inserted 147 bp downstream of the *yhiN* gene suggesting a low level of "wobble" may exist in targeting this site.

Elements nearly identical to Tn6230 are also found in mobile plasmids in a number of bacterial species (31), but these sites lack any obvious homology to the chromosomal insertion site in the *yhiN* gene (i.e. the putative *attTn6230* site). An unresolved question is how insertions are directed into mobile plasmids as no TnsE homolog is found in Tn6230. One possibility is that the TnsF protein from Tn6230 has a dual function, either recognizing a specific DNA sequence thereby inserting at the attachment site, or recognizing another structure or protein to target insertions into mobile plasmids. Consistent with this hypothesis, transposons with the *tniQ* gene have been shown to target transposition into active resolution sites used by certain types of resolvase in transposons and on plasmids (21). While none of the non-*yhiN* insertions appear to be *res* sites there may be other features beyond specific nucleotides sequence that are recognized by TniQ homologs.

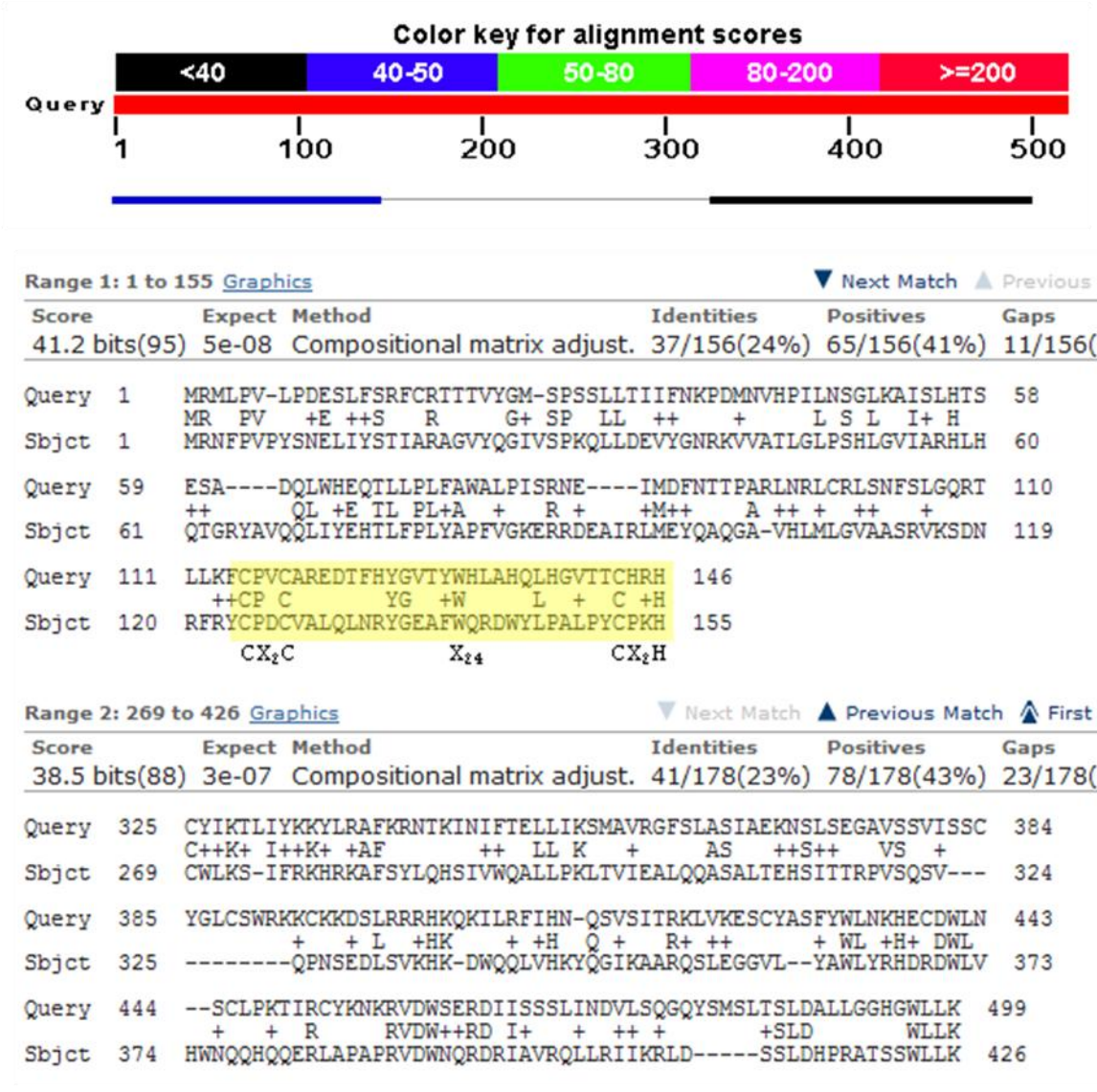

Figure 2.4

Alignment of TnsD and TnsF
An alignment of the protein sequences of TnsD from *E. coli* and TnsF from *S. enterica* ssp. enterica ser. Senftenberg str. A4-543 was done in BLAST using the align two or more sequences function. The CCCH Zinc finger domain characterized in TnsD is highlighted in yellow.

In Tn6230-like elements, insertions into non- *yhiN* sites raises the question as to whether these represent pseudo-attTn6230 sites, ie a site with sequence homology to the *yhiN att6230* site and therefore utilized at a lower frequency. A similar phenomenon is observed in situations

where the Tn7 attachment site is inaccessible, a low frequency of TnsD-mediated insertions is observed into pseudo-*attTn7* sites that have sequence similarity to the *attTn7* located in *glmS* (22). To identify if these non-*yhiN* sites were putative pseudo-*attTn6230* sites, I used Geneious to align the DNA sequence upstream of the element and in cases where no alignment was found, used the NCBI BLAST tool to determine the most closely related sequence (Table 2.1). Significant homology of the DNA sequence from the upstream region of these two elements to the *yhiN* gene was not noticeable. This raises the possibility that TnsF or another protein in the element is capable of targeting insertions into these alternate sites. In the original survey of Tn6230-like elements, 2 were identified in plasmids, further supporting the role of a protein in these elements targeting non-*yhiN* sites. However, this analysis should be expanded using a greater number of elements to better understand how target sites are recognized with Tn6230 and possibly more advanced bioinformatics tools.

2.5 Island formation in the *comM* gene, via Tn6022-like elements

A putative heteromeric transposase element appears to be important in the multi-drug resistant nosocomial pathogen *A. baumannii*. It was first noted that strain *A. baumannii* AYE contained a single large 86 kb element within the *comM* gene with ~90 genes, where half are likely associated with antibiotic or biocide resistance (13). The identification of a five base-pair duplication suggested that transposition accounted for the insertion of this pathogenicity island. Later bioinformatics work suggested that this element was common to many strains and that some of the transposition proteins seemed to be related to those used by Tn7 (41). I refer to these elements found in *comM* as Tn6022-like based on a minimal element that has a core set of six genes, TnsA, TnsB, TnsC, Orf2, Orf3, and Orf4 which are likely involved in transposition.

TABLE 2.1

	Tn insertion (Identities)	**note**
Neisseria flavescens NRL30031/H210	Delta-aminolevulinic acid dehydratase	*comM* not found in contig
Pelobacter carbinocilus DSM 2380	middle *comM* (51%)	
Pseudomonas syringae pv. Tomato str. DC3000	middle *comM* (53%)	
Alishewanella aestuarii B11	hypothetical protein	*comM* not found in contig
Dechlorosoma suillum PS	CheY chemotaxis	*comM* 166kb from tnsA
Delftia acidovorans CCUG 15835	middle *comM* (52%)	
Magnetospirillum magneticum AMB-1	Y4CA hypothetical protein	*comM* 806kb from tnsA
Acinetobacter baumannii D36	middle *comM* (100%)	
Thioalkivibrio sp. ALJ24	middle *comM* (55%)	
Ralstonia pickettii 12D	middle *comM* (52%)	
Vibrio nigripuchritudo Wn13	C-terminus *yhiN* (60%)	
Enterobacter sp. BIDMC26	phosodiesterase	*yhiN* not found in contig
Escherichia coli	C-terminus *yhiN* (88%)	
Salmonella enterica subsp. Enterica serovar Senftenberg str. A4-543	*yhiN** (100%)	
Cronobacter helveticus 513	DDE transposase	*yhiN* not found in contig
Photobacterium sp. SKA34	C-terminus *yhiN* (60%)	
Thiomicrospira kuenenii DSM 12350	C-terminus *yhiN* (44%)	
Yersinia pseudotuberculosis PB1/+	C-terminus *yhiN* (79%)	
Glaciecola mesophila KMM 241	C-terminus *yhiN* (58%)	
Vibrio metschnikovii CIP69.14	C-terminus *yhiN* (60)	
Bacillus cereus ATCC10987	downstream *glmS** (42%)	
Wohlfahrtiimonas chitiniclastica DSM 18708	downstream *glmS* (54%)	
Escherichia coli O28ac:NM	downstream *glmS* (100%)	
Rheinheimera sp. A13L	downstream *glmS* (70%)	
Methylomonas methanica MC09	downstream *glmS* (61%)	
Shewanella baltica OS155	downstream *glmS* (69%)	
Idiomarina loihensis L2TR	downstream *glmS** (68%)	
Moritella marina ATCC15381 MP-1 G33	downstream *glmS* (71%)	
Desulfovibrio aespoeensis Aspo-2	downstream *glmS* (26%)	
Aeromonas salmonicida subsp. Salmonicida A449	downstream *glmS* (74%)	

The location of the transposon relative to the gene it inserts into is indicated by "middle", "C-terminal", or "downstream"

Bold typeface indicates the element found in this organism was used as the reference

* indicates the element was downstream or had multiple open reading frames between the recognized insertion site and the transposon

Most *A. baumannii* strains appear to have the Tn6022-like element at the same position in the *comM* gene flanked by the same 5 bp target site duplication (ACCGC). The ComM protein is a predicted Mg chelatase-like protein originally named for a role in competence in *Haemophilus influenzae* Rd (15). Interestingly the drug susceptible *A. baumannii* strains AB307-0294 and D1279779 do not contain any Tn6022-like elements indicating the element is not ancestral to all *A. baumannii* (1, 12). Another possibility, however remote given the necessity to use the exact ends of the element to hop out leaving a perfect in frame empty site, is that the element is lost from these strains. While there is no experimental evidence to date that this is a single targeted site like *attTn7*, the finding that elements that have homology to Tn6022 are found in the gene homologous to *comM* in highly divergent bacteria makes it virtually certain that this is an *attTn6022* site for these elements (17). While a previous analysis did not recognize TnsA in Tn6022-like elements, our analyses indicate they have significant homology and the conserved motif (described above). Our small sample of 10 elements that are in the Tn6022 clade (red) in Figure 2.3A share TnsB, TnsC, and Orf2, therefore I operationally term these Tn6022-family elements. Out of these ten elements, only 6 were inserted into an obvious *comM* homolog in their respective host strains, which included *A. baumannii* D36, *Pelobacter carbinolicus* DSM 2380, *Pseudomonas syringae* pv. Tomato str. DC3000, *Delftia acidovorans* CCUG 15835, *Thioalkivibrio sp.* ALJ24, and *Ralstonia pickettii* 12D. It is worth noting that these are closely related species, all within the proteobacteria. The DNA sequences flanking the four other Tn6022-family elements did not share any obvious sequence similarity to either *comM* or each other (Table 2.2). The upstream sequence was analyzed as described in the methods. From this, I was able to identify a homolog of the *comM* gene in all of the genomes containing Tn6022-like elements available in NCBI, but the Tn6022-family element in these four was not

proximal to *comM*: in all of these organisms, *comM* was at least 100 kilo base pairs from the 5' end of the element. Insertion events of Tn6022-family elements at locations that are not in *comM* could be explained by other targeting pathways much in the same way that Tn7-like elements have a TnsD pathway to target insertions into *attTn7,* and a TnsE pathway that does not recognize a specific DNA sequence. Another possibility could be that Tn6022-family elements are targeting a pseudo-site that has sequence homology to the *comM attTn6022* site and therefore utilized at a lower frequency as described in more detail above. I gathered bioinformatics evidence that insertion events that were not in *comM* did not appear to be pseudo-sites given that no obvious homology could be found by aligning the DNA sequence upstream of the inserted elements to the *comM* sequence. It is possible that the TnsABC proteins of Tn6022-family elements do allow a low level of transposition without the requirement of a target site selecting protein, as is found with specific mutations isolated with TnsA, TnsB and TnsC with Tn7 (26, 47). In order to address this possibility, additional experiments are required.

2.6 Identification of Putative Target Site Selecting Proteins in Transposons Containing Heteromeric Transposases

TnsA was used to assemble the three transposon families in Figure 2.3, something that is supported by the finding that homologs of TnsB and TnsC could also be identified in the same gene order. Interestingly, other proteins that are putatively involved in targeting transposition events can also be found in the Tn6022-family elements. Here I discuss the conservation of additional genes found in the Tn6022-family elements that may be involved in targeting transposition.

The TnsD targeting protein in Tn7 targets insertions into a site downstream of *glmS*, and while it has not yet been examined experimentally I suspect that the TnsF protein of the Tn6230-

like elements is playing a similar role to TnsD. This idea is supported by the finding that TnsF shows homology and synteny to the TnsD target site selecting protein found in Tn7-like elements. It is remains unstudied as to what the target selecting protein(s) could be (if one exists outside the core machinery) in the Tn6022-family elements. Bioinformatic analysis of four open reading frames shared among elements similar to Tn6022 (Tn6019, Tn6021) were purported to have no significant homology to TnsD or TnsE from Tn7 (17). However, given the location of two of these open reading frames immediately downstream of the TnsC-like protein, they remain good candidates for target site selecting proteins for Tn6022 (Figure 2.1B). As indicated below, a relatively small subset of proteins seem to be conserved in the Tn6022-family elements and even fewer that do not have conserved homologs (see below). It seems likely that one or more of three conserved proteins that have previously been identified as conserved across what I term Tn6022-family elements Orf2, Orf3, and Orf4 (17), could function like target site selection proteins for these elements. Previous analysis did not compare the putative target site selecting proteins from Tn6022 to TnsF from Tn6230, which may share similarity than to TnsD or TnsE from Tn7 based on the tree shown in Figure 2.3A. From the comparison of Orf2 and Orf3 proteins to TnsF (Table 2.2), I show that they are not similar to TnsF, but comprise their own protein family. Because Orf2 and Orf3 do not have significant homology to TnsD, TnsE, or TnsF, another method must be used to determine putative targeting proteins. One way I can identify putative targeting proteins is by finding conserved open reading frames in elements where the elements are found in the same insertion site (a phylogenetic tree of TnsD and putative targeting proteins was generated and shown in Figure 2.3B).

TABLE 2.2

	First open reading frame					Second reading frame, if present				
	TnsD	TnsE	TnsF	Orf2	Orf3	TnsD	TnsE	TnsF	Orf2	Orf3
Acinetobacter baumannii D36*	26/7	29/13	26/9	100/100	20/31	25/18	63/2	38/35	20/21	100/100
Ralstonia pickettii 12D *	45/2	50/4	29/22	28/96	33/16	27/23	39/7	36/19	28/41	33/97
Delftia acidovorans CCUG 15835*	27/17	67/3	33/10	29/43	35/7	100/3	no sig	63/1	56/13	26/86
Pseudomonas syringae pv. Tomato str. DC3000*	36/22	57/18	28/20	26/75	17/22	27/40	20/19	29/10	29/19	31/97
Pelobacter carbinocilus DSM 2380 *	28/14	22/22	30/17	26/58	30/26					
Magnetospirillum magneticum AMB-1	33/23	25/13	26/27	24/87	39/32					
Neisseria flavescens-1.0	29/23	25/12	23/31	25/43	23/55					
Alishewanella aestuarii B11	32/34	20/13	28/22	28/37	50/29					
Dechlorosoma suillum PS	34/11	23/8	22/26	24/43	26/34					
Thioalkivibrio sp. AlJ24*	30/13	27/30	40/17	28/47	35/43					
Enterobacter sp. BIDMC26	22/36	54/2	35/59	57/16	24/9					
Escherichia coli 9.0111*	23/61	38/13	66/98	24/51	38/19					
Cronobacter helveticus 513	22/55	37/3	94/95	29/24	38/14					
Salmonella enterica Senftenberg str. A4-543 *	24/61	no sig	100/100	29/8	38/35					
Yersinia pseudotuberculosis PB1/+ *	27/32	32/11	28/85	67/2	31/17					
Vibrio metschnikovii CIP69.14 *	22/64	28/12	32/98	25/25	24/43					
Photobacterium sp. SKA34*	30/47	no sig	32/98	34/14	40/17					
Vibrio nigripulchritudo Wn13*	25/40	28/10	31/99	46/9	24/19					
Glaciecola mesophila KMM241*	25/42	50/8	31/98	39/4	25/37					
Thiomicrospira kuenenii DSM12350*	24/69	44/10	32/96	39/30	26/6					
Bacillus cereus ATCC10987*	29/92	37/21	23/77	18/17	27/22	21/39	19/56	28/7	53/26	no sig
Wohlfahrtiimonas chitiniclastica DSM 18708*	37/99	no sig	23/50	30/32	25/39	no sig	29/97	38/10	40/2	31/4
Escherichia coli *	100/100	50/4	24/61	26/12	25/22	50/5	100/100	no sig	29/18	63/2
Rheinheimera sp. A13L *	50/99	30/3	24/48	26/27	27/25	28/20	47/98	50/5	32/17	no sig
Methylomonas methanica MC09 *	38/99	43/1	25/46	32/21	29/11	24/6	27/97	30/8	31/3	60/2
Shewanella baltica OS155 *	39/99	27/14	26/43	29/48	56/15	50/2	26/95	29/9	30/6	63/5
Desulfovibrio aespoeensis Aspo-2*	45/96	30/11	30/63	18/62	63/51	70/16	30/97	37/13	41/13	29/16
Aeromonas salmonicida subsp. Salmonicida A449*	43/97	29/11	27/59	24/22	44/12	30/15	30/98	39/10	33/3	29/5
Idiomarina loihensis L2TR *	39/99	33/7	23/45	31/11	39/15	29/5	30/98	67/14	no sig	no sig
Moritella marina ATCC15381 MP-1 G33 *	38/99	58/7	27/45	27/32	80/3	45/10	29/80	36/13	32/27	35/3

* indicates the element from the organism is found in or adjacent to the putative attachment site for that transposon family

Bold typeface indicates the element found in this organism was used as the reference

The numbers refer to percent identity /query coverage, based on the following proteins: TnsD (508aa), TnsE (538aa,), TnsF (518aa) , Orf2 (380aa) , Orf3(483aa)

2.6.1 Conservation of putative target site-selecting proteins

It was of interest to determine whether there were any elements that encode proteins with homology to proteins in more than one heteromeric transposase family. This requires a two step analysis, first to determine conservation of proteins within each family, and then to determine whether there were any proteins that had conservation to proteins of other families. Given that I know that all of the elements had TnsABC, the best proteins for this analysis that are conserved across elements are the putative target site selecting proteins. This section takes a close look at the sequence similarities between the proteins within and between each of the three families or clades inferred from TnsA in figure 2.3A.

First, the parameters used to indicate 'homology' should be defined. The two parameters used here are pair-wise sequence identity (or identity) and the percentage of the total sequence that is represented in the analysis (query coverage). While the total length of the aligned segments of each protein is essential in determining the homology, because all proteins have a about the same length (ranging from 380-508 amino acids), I can use the query coverage as an approximation for the total length. Effectively, the greater the query coverage (generally, greater than 30%), and the higher the sequence identity the more confidently the proteins are related, however analysis remains subjective (42). As noted above, the conservation of the order and orientation of the open reading frames between the transposons that are being compared (or, synteny) greatly suggests that they are bona fide homologs of the respective putative target site selecting proteins, even if they do not appear homologous according to identity and query coverage. The results of this pair-wise analysis are shown in Table 2.1, where the first or second open reading frame following the TnsC-like protein of each element is compared to that from the representative transposon. Proteins that I believe are confidently related are in red numbering,

proteins that do not have a high level of similarity are in light grey, and in instances where there appear to be slight homology are in orange. For ease of viewing, proteins considered to be homologous are boxed in green.

As indicated above, clear TnsD and TnsE homologs a were found in all of the elements in the same clade as Tn7. The most diverged element in this clade was in *B. cereus* based on TnsABC, TnsD and TnsE (Figure 2.3A and Table 2.2). The *B. cereus* strain analyzed here has two Tn7-like elements based on TnsA, TnsB, and TnsC homology, where the one that resides closer to *glmS* lacks a full-length TnsE, and the element containing TnsD and TnsE described above contains the full set of five Tn7-like proteins. Tandem insertions previously have been described (Figure 2.1D and (24, 35)), but the one found *B. cereus* is the only one I identified in the subset of elements studied in this chapter. As compared to each other, the TnsA proteins in these two elements from *B. cereus* are more similar to each other than to any of the other TnsA-like proteins in the Tn7 family (not shown).

The Tn7-like elements analyzed here had TnsA proteins that clustered together (green, Figure 2.3A) and closely related TnsD and TnsE proteins (Table 2.2) that had extensive synteny. This, along with the observation that all of the Tn7-like elements were found in the *attTn7* would suggest that the TnsD-like proteins have similar functions in their respective organisms. While the homology of TnsE-like proteins was also high among Tn7-like elements, and had synteny to their respective TnsD-like proteins, further work is needed to establish if they have the same function in transposition.

Like the TnsD and TnsE proteins found in the Tn7-like elements discussed above, unambiguous TnsF homologs were found in the clade with Tn6230. These *tnsF* genes are located within Tn6230 in the same location following *tnsABC*, as *tnsD* within the Tn7-like

elements. Furthermore, the high level of conservation within the TnsF-like proteins would suggest they all carry out the same function. This function may be related to the localization of these elements in *yhiN* in multiple species which is supported by the conservation between TnsF and TnsD. The future study of one or a few of these proteins *in vivo*, can establish that TnsF is a true targeting protein, and whether the TnsF-like proteins of the more diverged elements maintain targeting capabilities.

We call the elements found in the *comM* gene of *A. baumanii* Tn6022-like because they have nearly identical *orf2*, *orf3*, and *orf4* genes. I refer to the other elements in this same clade in Figure 2.3 (red) as Tn6022-family elements, which all contain TnsABC and believable relative of the *orf2* gene. While all contain recognizable motifs from Orf2 as identified in Tn6022, the level of homology is often quite low suggesting the last shared ancestor of these elements has not yet been identified, and was found a long time ago. The protein that most closely matched Orf2 from *A. baumannii* was the Orf2-like protein from *Ralstonia pickettii* 12D, which had a low percent identity, despite a large amount of overlapping protein sequence. The remaining eight elements contained an Orf2-like protein that was even less homologous to *A. baumannii* Orf2 (Table 2.2). However the gene encoding Orf2-like proteins in each element had synteny with Orf2, suggesting they are bona-fide homologs of Orf2. Because these Orf2-like proteins were poorly conserved, I determined whether they were more closely related to other putative targeting proteins. Each of the Orf2-like proteins was compared to the established targeting proteins, TnsD, and putative targeting proteins TnsF and Orf3 (Table 2.2). While there is no homology of Orf2 to TnsD or TnsF, half of the Orf2-like proteins studied had limited homology to Orf3 from Tn6022 (described below). This raises the question of whether these two proteins could be related, but without further analysis, this cannot be determined.

Only three Tn6230-family elements contained a homolog to Orf3, which were conserved in both sequence and in organization within the element, making it possible they have a similar function in each transposon. A putative CCCH Zinc finger motif was identified in proteins showing limited homology to Orf2 (Figure 2.6C) or Orf3 (Figure 2.6D). In order to determine how widespread this protein is, and whether these proteins are DNA binding proteins a greater number Orf2 and Orf3 proteins should be analyzed.

In five examples, there was a protein that had homology to both Orf2 and Orf3 but no other putative targeting proteins (Table 2.1). An intriguing possibility is that this protein represents an early form of targeting protein that later duplicated to give the Orf 2 and Orf3 proteins found in the Tn6022-like elements. To better understand evolution of this element and which proteins are involved in target site selection, further analysis of a greater number of Tn6022-family elements is merited.

Previous analysis of Tn6022-like proteins suggested Orf4 is highly conserved across Tn6022-like elements (17). All of the Tn6022-like elements I examined in *A. baumannii* in the *comM* gene encode Orf4, but the role of the Orf4 protein in transposition remains unstudied. Future work investigating a role for one of the proteins found in *A. baumannii* D36, *R. picketti* 12D, or *P. syringae* pv. tomato str. DC3000 will be interesting given that there is limited homology with Orf4 in these strains (data not shown). The TnsA proteins from these three elements form a sub-clade (Figure 2.3), and given they appear to have homologs of TnsABC, Orf2, Orf3, and Orf4, and are all found in *comM*, they can also be considered Tn6022-like elements (as indicated in Figure 2.3). The other seven elements form a second subclade in the TnsA tree (Figure 2.3), and may or may not contain an Orf4-like protein. Given that *orf4* is not immediately adjacent to *orf3*, and the right end of the element was not identified in these other

seven elements of the Tn6022-family TnsA tree, our analysis did not include Orf4 in these other elements.

As indicated above, each of the TnsD-like proteins from the Tn7-like elements were compared to TnsD from Tn7 and TnsF from Tn6230. When each of the TnsD-like proteins was compared to TnsD, the homology was much greater than when compared to TnsF. However, there was still moderate homology of each of these TnsD-like proteins to TnsF (on average, ~25% identity over ~55% of TnsF) (orange, Table 2.2). The inverse comparison was also done: each of the TnsF-like proteins was compared to both TnsD and TnsF. Likewise, although TnsF-like proteins were more similar to TnsF from Tn6230 based on homology, a handful of the TnsF-like proteins had moderate homology to TnsD from Tn7. The similarity shared by TnsD and TnsF, albeit weak, may suggest a common ancestor.

In addition to the common CCCH zinc finger motif in TnsD and TnsF (Figure 2.6 A, B) as indicated above, in all but one of the Tn6022-like elements a putative CCCH motif was also observed in both Orf2- and Orf3-like proteins of Tn6022 (Figure 2.6 C, D). In Orf2, the spacing between the central two Cysteines in the C-X_2-C-X_{20-21}-C-X_2-H motif is closer to the established domain in TnsD and may represent a domain similar to, but distinct from the recognized TnsD zinc-finger domain. The Orf3-like proteins have multiple C-X_2-C motifs separated by variable amino acids at approximately the same region as the zinc finger domain identified in TnsD, which may represent zinc finger motifs. However, the closest putative zinc finger domain is the C-X_2-C-X_{11-12}-C-X_2-H, shared by three of the four Orf3-like proteins. The Orf3-like protein of *Ralstonia pickettii*, may be non-functional because it lacks the conserved Histidine. While this is a putative domain in all of these proteins, structural analysis or further study will enable us to confirm that this is a zinc finger domain.

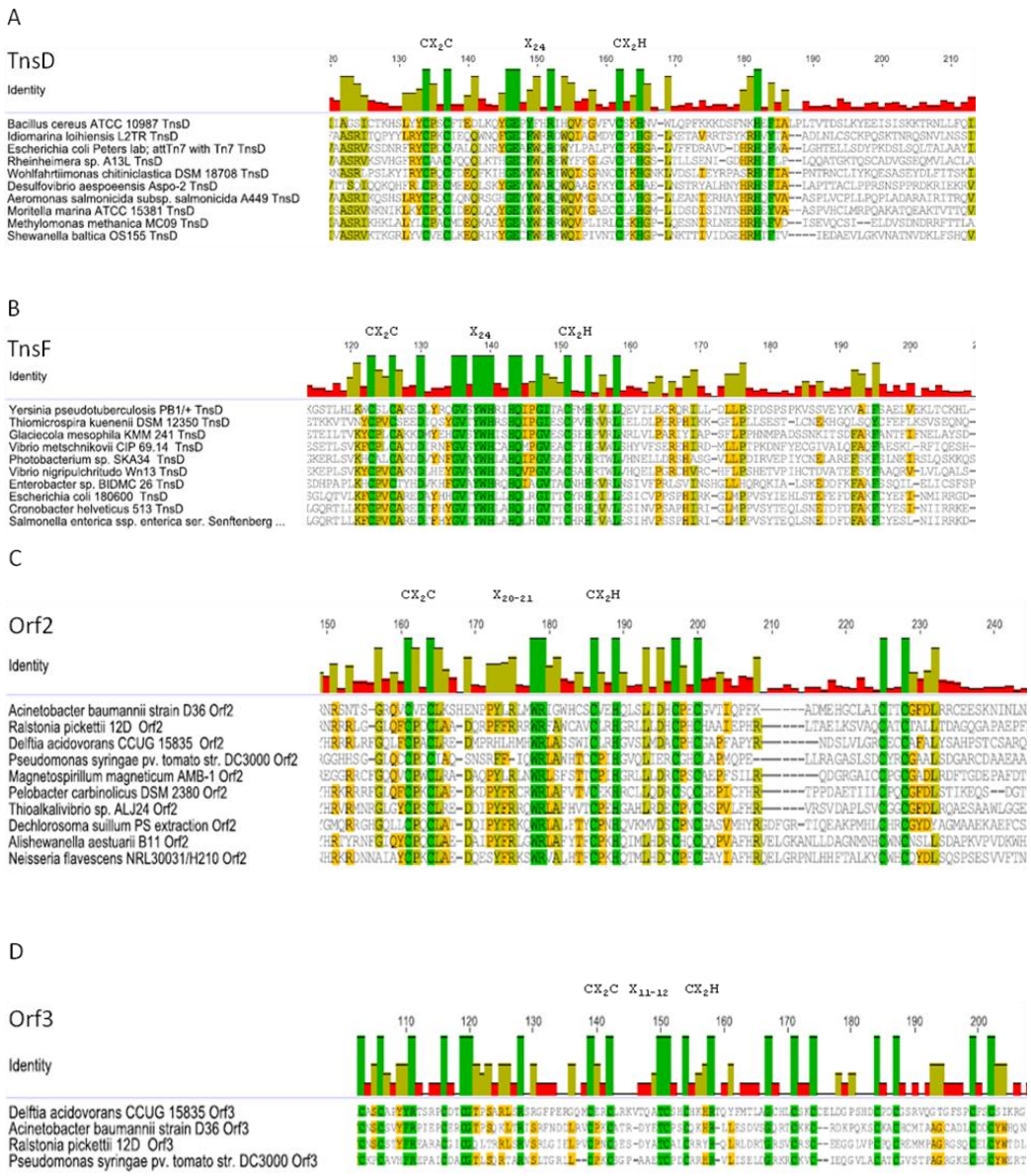

Figure 2.5

Putative CCCH Zinc finger domain in TnsD-like, TnsF-like, Orf2-like, and Orf3-like proteins TnsD-like, TnsF-like, Orf2-like, and Orf3-like proteins from the hosts indicated are aligned using the MUSCLE plugin in the Geneious software. The region representing the putative CCCH Zing finger domain is shown. Conserved residues are indicated above the alignment, and highlighted according to BLOSUM62 matrix, conservation indicated by varying intensities of yellow : green represents a high level of conservation.

2.7 STRING Analysis Links Attachment Sites and Putative Target Site Selecting Proteins

As another tool to help identify candidate attachment sites of Tn6230- and Tn6022-clade elements, I used a search tool: STRING (Search Tool for the Retrieval of Interacting Genes/ Proteins), which is hosted by EMBL (European Molecular Biology Laboratory) using each of the putative targeting proteins as query. This database combines multiple types of preexisting protein-protein association data (for example, mention of two proteins in the same publication, high throughput experiments, or genomic context) to predict interactions. Because here I am not assuming the protein generated from the gene containing the putative attachment site to physically interact or share a functional pathway, I limited the evidence to computational analysis using whole genome sequences. This will reveal putative associated proteins to the input protein, or in our case the expected targeting proteins. For each of the putative targeting proteins, all of the predicted linked proteins are represented by a network in which the proteins are indicated by circles, and the proximal proteins are indicated by lines (Figure 2.7). The computation took into account two types of genomic context: 1) the occurrence of adjacent genes or multiple genes in a close neighborhood across a few species (neighborhood, green lines) and 2) the joint presence or absence of two proteins across multiple species (co-occurrence, blue lines). The protein sequence from predicted protein associations that did not have a gene name were submitted to pBLAST for identification, and the correct gene name is given in the figure.

Given that Tn7 has been well studied, I can confirm the STRING-predicted protein-protein associations, that is, I expect, and find, that TnsD reveals TnsA, TnsB, TnsC and TnsE to be protein associations (Figure 2.7). Furthermore, for all putative targeting proteins (TnsF, Orf2, Orf3), I find that the core machinery proteins (TnsA, TnsB, TnsC) are revealed to be associated proteins. However, in an attempt to identify an association with a gene location, I used the

known association of Tn7 with *glmS* to determine if STRING could identify this "association". Expectedly, when TnsD was input into the STRING search tool, *glmS* was revealed as a protein that may be associated with TnsD. An important caveat in this type of analysis is revealed by the finding that using TnsE in the STRING program also identifies *glmU*, *glmS*, and *yhfN*. While *yhfN* does not appear near *glmUS*, it could indicate a pseudo-attTn7 site, however to tease apart the relationship of *yhfN* in Tn7 transposition, more analysis is needed. Therefore although candidates can be suggested using STRINGS caution is warranted in this type of analysis. When a search for TnsF was used in the program, it revealed the *yhiN* gene to be often found close to the TnsF, which I have identified as a preferred insertion site in earlier work and in the results presented above, further supporting this site as a putative *attTn6230*.

While Orf 3 came up in a search for Orf2, and vice versa, Orf4 only came up with Orf3. The inverse, in a search for Orf4, only Orf 3 but not Orf2 came up. Protein sequences from genes labeled AB57_#### were all aligned to the known TnsABC, Orf234, and ComM sequences and appropriately labeled. Those that did not match were searched in NCBI, and identified as "hypothetical protein". Given the number of hypothetical proteins, it would suggest that the Tn6022-like proteins have not been well characterized. Unfortunately, ComM was not identified as a protein associated with either Orf2, Orf3, or Orf4. This 'non-association' can be explained by the spread of Tn6022-family elements in different non-*comM* sites. Another more likely explanation is that in cases where Tn6022 elements are found in *comM*, the gene is effectively halved, such that based on the evidences used, *comM* would not be in the same run of genes or co-occur with Tn6022-clade elements in the putative attachment site.

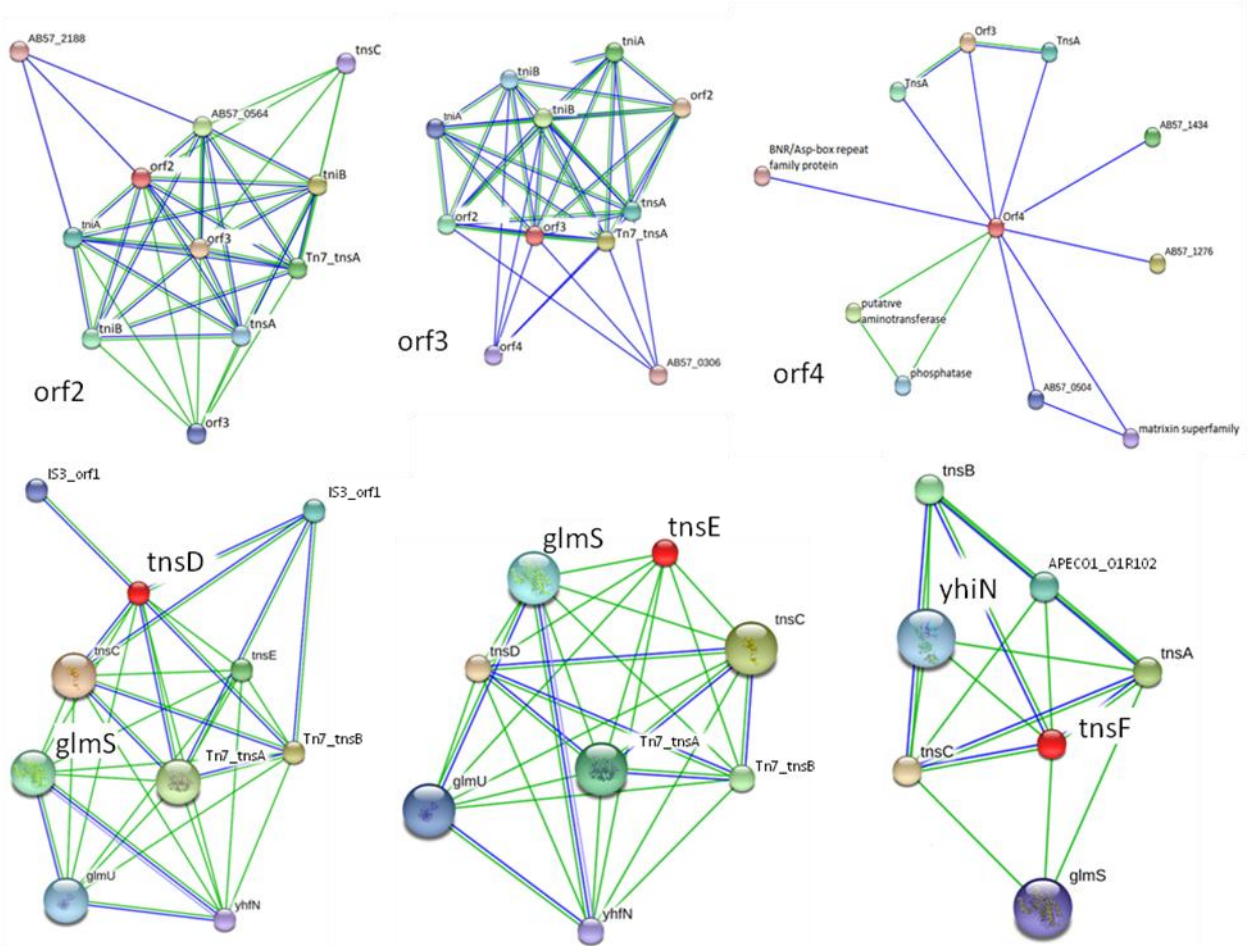

Figure 2.6

Protein Network Association in STRING of TnsD, TnsE, TnsF, Orf 2, or Orf 3
One protein sequence, belonging to any of the five putative target site selecting proteins, was used as input (drawn as a red sphere). The number of interacting proteins was limited to 10 (default settings). A network view of the interacting proteins is shown. The evidence types supporting the interactions were limited to neighborhood (green lines) and co-occurrence (blue lines).

2.8 Conclusions

Based on a review of the literature, previous findings in the laboratory, and bioinformatics analysis I argue that genomic islands can form at least at three distinct loci in bacteria by processes catalyzed by heteromeric transposase elements. These elements have in common TnsA, TnsB, and TnsC like proteins, but distinct types of confirmed or putative target

selecting proteins. It is unknown if there is anything special about the heteromeric transposase that encourages elements to evolve modular transposition pathways, but it seems likely that additional target site selection proteins recognizing different attachment sites will be discovered. In the current examples it is unclear if there are any special benefits to the transposon by targeting the *comM* and *yhiN* genes. It will be interesting to determine how the Tn6022-clade elements maximize recombination within the element in *comM* and the role this plays in the spread of drug resistance in *A. baumannii*. Finally, it will be important to determine any targeting mechanisms that facilitate horizontal transfer with the Tn6022- and Tn6230-like elements.

2.9 Methods

2.9.1 STRING Analysis

The search tool STRING (Search Tool for the Retrieval of Interacting Genes/ Proteins), can be found at http://string-db.org/, was used to help identify candidate attachment sites (19). This tool is hosted by EMBL (European Molecular Biology Laboratory) using the amino acid sequence for each of the putative targeting proteins as query. The prediction methods were limited to "neighborhood" and "co-expression" on the output page. The amino acid sequence for proteins that did not have gene nomenclature was submitted to the BLAST (Basic Local Assignment Search Tool) hosted by NCBI (National Center for Biotechnology Information) at http://blast.ncbi.nlm.nih.gov/Blast.cgi and the resulting name was pasted onto the graphic. Any proteins that were labeled "hypothetical protein" were aligned to the known proteins that are related to the respective element. For Tn6022, this was TnsABC, Orf2, Orf3, Orf4, and ComM; for Tn6230, this was TnsABCF and YhiN.

2.9.2 Upstream DNA analysis

Upon identification of TnsA, the 1 kb of DNA upstream was first aligned to the same amount of upstream DNA of the representative element in each transposon family. This same 1kb upstream DNA was translated to open reading frames and the resulting amino acid sequences were aligned to this representative. Finally, if these sequences did not align, the DNA sequence was submitted to NCBI BLAST to find the most homologous sequence. Percentage identity and query coverage for all of this upstream DNA is shown in Table 2.1.

2.10 Acknowledgments

I thank Peters Lab members, Mike Petassi and Bennett Kapili for helpful discussions and advice on the chapter. This work was funded by the National Science Foundation (MCB -1244227).

REFERENCES

1. Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, et al. 2008. Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. *J Bacteriol*. 190(24):8053–64
2. Arciszewska LK, Craig NL. 1991. Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. *Nucleic Acids Res*. 19(18):5021–29
3. Bainton R, Gamas P, Craig NL. 1991. Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. *Cell*. 65(5):805–16
4. Bainton RJ, Kubo KM, Feng J, Craig NL. 1993. Tn7 transposition: Target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified *in vitro* system. *Cell*. 72(6):931–43
5. Baker TA, Mizuuchi M, Mizuuchi K. 1991. MuB protein allosterically activates strand transfer by the transposase of phage Mu. *Cell*. 65(6):1003–13
6. Bhasin A, Goryshin IY, Reznikoff WS. 1999. Hairpin formation in Tn5 transposition. *J. Biol. Chem.* 274(52):37021–29
7. Chaconas G, Gloor G, Miller JL. 1985. Amplification and purification of the bacteriophage Mu encoded B transposition protein. *J. Biol. Chem.* 260(5):2662–69
8. Cheetham BF, Katz ME. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. *Molecular Microbiology*. 18(2):201–8
9. Choi KY, Li Y, Sarnovsky R, Craig NL. 2013. Direct interaction between the TnsA and TnsB subunits controls the heteromeric Tn7 transposase. *Proc Natl Acad Sci*. 110(22):E2038–E2045

10. Craig, N. L. 2002. Tn7. In *Mobile DNA II* (Craig Nancy, L., Craigie, R., Gellert, M., and Lambowitz-Alan, M., eds) pp. 423-456, ASM Press, Washington, DC
11. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*. 266(5193):1981–86
12. Farrugia DN, Elbourne LDH, Hassan KA, Eijkelkamp BA, Tetu SG, et al. 2013. The complete genome and phenome of a community-acquired *Acinetobacter baumannii*. *PLoS ONE*. 8(3):e58628
13. Fournier P-E, Vallenet D, Barbe V, Audic S, Ogata H, et al. 2006. Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. *PLoS Genet*. 2(1):e7
14. Gringauz E, Orle KA, Waddell CS, Craig NL. 1988. Recognition of *Escherichia coli attTn7* by transposon Tn7: lack of specific sequence requirements at the point of Tn7 insertion. *J. Bacteriol*. 170(6):2832–40
15. Gwinn ML, Ramanathan R, Smith HO, Tomb J-F. 1998. A new transformation-deficient mutant of *Haemophilus influenzae* rd with normal DNA uptake. *J Bacteriol*. 180(3):746–48
16. Hacker J, Kaper JB. 2000. Pathogenicity islands and the evolution of microbes. *Annual Rev Microbiol*. 54(1):641–79
17. Hamidian M, Hall RM. 2011. AbaR4 replaces AbaR3 in a carbapenem-resistant *Acinetobacter baumannii* isolate belonging to global clone 1 from an Australian hospital. *J. Antimicrob. Chemother.* 66(11):2484–91
18. Hickman AB, Li Y, Mathew SV, May EW, Craig NL, Dyda F. 2000. Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol Cell*. 5(6):1025–34

19. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, et al. 2009. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 37(Database issue):D412–D416

20. Kennedy AK, Guhathakurta A, Kleckner N, Haniford DB. 1998. Tn10 transposition via a DNA hairpin intermediate. *Cell*. 95(1):125–34

21. Kholodii GY, Mindlin S z., Bass I a., Yurieva O v., Minakhina S v., Nikiforov V g. 1995. Four genes, two ends, and a res region are involved in transposition of Tn5053: a paradigm for a novel family of transposons carrying either a Mer operon or an integron. *Molec. Microbiol*. 17(6):1189–1200

22. Kubo KM, Craig NL. 1990. Bacterial transposon Tn7 utilizes two different classes of target sites. *J. Bacteriol*. 172(5):2774–78

23. Kuduvalli PN. 2001. Target DNA structure plays a critical role in Tn7 transposition. *EMBO J*. 20(4):924–32

24. Li Z. 2012. A molecular mechanism allowing transposon Tn7 to target active DNA replication. Cornell University

25. Li Z, Craig NL, Peters JE. 2013. Transposon Tn7. *Bacterial Integrative Mobile Genetic Elements*, p.1-32. Austin, TX: Landes Biosci.

26. Lu F, Craig NL. 2000. Isolation and characterization of Tn7 transposase gain-of-function mutants: a model for transposase activation. *EMBO J*. 19(13):3446–57

27. Maxwell A, Craigie R, Mizuuchi K. 1987. B protein of bacteriophage Mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. *Proc Natl Acad Sci U S A*. 84(3):699–703

28. May EW, Craig NL. 1996. Switching from cut-and-paste to replicative Tn7 transposition. *Science*. 272(5260):401–4

29. McKown RL, Waddell CS, Arciszewska LK, Craig NL. 1987. Identification of a transposon Tn7-dependent DNA-binding activity that recognizes the ends of Tn7. *Proc Natl Acad Sci*. 84(22):7807–11

30. Mitra R, McKenzie GJ, Yi L, Lee CA, Craig NL. 2010. Characterization of the TnsD-*attTn7* complex that promotes site-specific insertion of Tn7. *Mobile DNA*. 1(1):18

31. Moreno Switt AI, den Bakker HC, Cummings CA, Rodriguez-Rivera LD, Govoni G, et al. 2012. Identification and characterization of novel Salmonella mobile elements involved in the dissemination of genes linked to virulence and transmission. *PLoS ONE*. 7(7):e41247

32. Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 405(6784):299–304

33. Parks AR, Li Z, Shi Q, Owens RM, Jin MM, Peters JE. 2009. Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell*. 138(4):685–95

34. Parks AR, Peters JE. 2006. Transposon Tn7 is widespread in diverse bacteria and forms genomic islands. *J Bacteriol*. 189(5):2170–73

35. Parks AR, Peters JE. 2009. Tn7 elements: engendering diversity from chromosomes to episomes. *Plasmid*. 61(1):1–14

36. Peters JE. 2014. Tn7. *Microbiol Spectrum*. 2(5): MDNA3-0010-2014

37. Peters JE, Craig NL. 2001. Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev*. 15(6):737–47

38. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. 2012. The Pfam protein families database. *Nucleic Acids Res*. 40(Database issue):D290–D301

39. Radstrom P, Skold O, Swedberg G, Flensburg J, Roy PH, Sundstrom L. 1994. Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. *J Bacteriol*. 176(11):3257–68

40. Reiter WD, Palm P, Yeats S. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res*. 17(5):1907–14

41. Rose A. 2010. TnAbaR1: a novel Tn7-related transposon in *Acinetobacter baumannii* that contributes to the accumulation and dissemination of large repertoires of resistance genes. *Biosci Horiz*. 3(1):40–48

42. Rost B. 1999. Twilight zone of protein sequence alignments. *Protein Eng.* 12(2):85–94

43. Roth DB, Menetski JP, Nakajima PB, Bosma MJ, Gellert M. 1992. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in SCID mouse thymocytes. *Cell*. 70(6):983–91

44. Sarnovsky RJ, May EW, Craig NL. 1996. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J*. 15(22):6348

45. Skelding Z, Queen-Baker J, Craig NL. 2003. Alternative interactions between the Tn7 transposase and the Tn7 target DNA binding protein regulate target immunity and transposition. *EMBO J*. 22(21):5904–17

46. Stellwagen AE, Craig NL. 1997. Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J*. 16(22):6823–34

47. Stellwagen AE, Craig NL. 1997. Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. *Genetics*. 145(3):573–85

48. Stellwagen AE, Craig NL. 1998. Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. *Trends Biochem Sci*. 23(12):486–90

49. Waddell CS, Craig NL. 1988. Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. *Genes Dev*. 2(2):137–49

50. Waddell CS, Craig NL. 1989. Tn7 transposition: recognition of the *attTn7* target sequence. *Proc Natl Acad Sci*. 86(11):3958–62

51. Wolkow CA, DeBoy RT, Craig NL. 1996. Conjugating plasmids are preferred targets for Tn7. *Genes Dev*. 10(17):2145–57

52. Yamauchi M, Baker TA. 1998. An ATP–ADP switch in MuB controls progression of the Mu transposition pathway. *EMBO J*. 17(18):5509–18

CHAPTER 3

TN7 TRANSPOSITION MEDIATED BY TNSE IS HOST SPECIFIC THROUGH INTERACTIONS WITH THE BETA CLAMP OF DNA REPLICATION*

3.1 Summary

The bacterial transposon Tn7 utilizes the element encoded TnsABC+E to direct transposition into actively replicating DNA. These insertions require TnsE interaction with the DnaN protein, an interaction that is thought to be species specific. Previous work by Zaoping was able to functionally replace the normal host DnaN with homologs from two foreign organisms, *Idiomarina loihiensis*, and *Shewanella baltica* by providing these DnaN proteins *in trans* in a strain with a temperature sensitive DnaN mutation (*dnaN159*) (17). In this chapter, I am able to completely replace the host DnaN with these same two foreign clamps in the absence of any other *dnaN* allele. Using the *dnaN159* strain with the foreign clamps and TnsE proteins from these same two organisms, the number of papillae (which suggests transposition) doubled in the presence of TnsE and DnaN from the same host as compared to a control strain lacking TnsE. To unequivocally show that actual transposition events were being monitored in the assay, I genetically mapped the miniTn7 element used to indicate transposition in this strain. In strains where the *dnaN159* mutant was replaced by *I. loihiensis* or *S. baltica*, the mutation frequency increased in the absence of the accessory DNA polymerase II, likely due to the involvement of Pol IV and Pol V in the absence of Pol II, which I confirmed to be real mutations.

3.2 Introduction

Transposons are mobile genetic elements that are capable of movement between positions within a genome. The Tn7 transposon utilizes a cut and paste mechanism to move in

*Figure 3.1A has been published previously in Li Z. 2012. A molecular mechanism allowing transposon Tn7 to target active DNA replication. Cornell University

which the element is completely excised from one location and inserted into a new "target" site (reviewed in (18, 24)). Tn7 transposition is activated upon recognition of one of its preferred target sites. One preferred target is located on the chromosome, the Tn7 attachment site (*attTn7*), and the other preferred target is certain types of DNA replication, but especially replication associated with conjugal plasmid transfer. These two target types could be expected to facilitate both vertical and horizontal transmission.

Tn7 encodes five proteins, TnsA, TnsB, TnsC, TnsD, and TnsE that allow for targeting and insertion into both of its two target sites (*attTn7* and conjugal plasmids). The core machinery shared by both pathways is TnsABC, but transposition does not occur with the core machinery alone. The TnsA and TnsB proteins form a heteromer that is responsible for the DNA breaking and joining activities in transposition (6). TnsB recognizes specific sequences in the cis-acting left and right ends of the element (1). These ends are large, containing multiple TnsB binding sites with varying affinities for TnsB binding. The left end has four overlapping TnsB binding sites, whereas the right end has three widely spaced sites for TnsB binding. TnsC is an ATPase that is recruited by each of the two target site selecting proteins, TnsD or TnsE, to signal the TnsAB transposase (10).

The TnsABC+D pathway targets insertions into *attTn7*, which is downstream of the *glmS* gene (33). TnsD recognizes a specific highly conserved sequence within the *glmS* gene which encodes the active site of the GlmS protein. After recognition of this site, TnsD directs transposition into the chromosomal *attTn7* site, a single site located about 25 basepairs downstream of where TnsD binds. Insertion into this site has no obvious fitness costs to the host, making it a "safe haven", where the Tn7 element is propagated with the host.

Analysis of the DNA sequences that are used for insertion with TnsE-mediated transposition indicates that TnsE has no obvious sequence that is recognized as a target (38). TnsE recognizes features that are enriched for during lagging strand replication, 3' recessed DNA ends and β-clamps (a subunit of DNA Polymerase III). These features are commonly found on conjugal plasmids in recipient cells during mobilization and filamentous bacteriophage during replication (see next two paragraphs), both of which stimulate TnsABC+E transposition.

When conjugation is initiated, the relaxase protein creates a nick in the origin of transfer (oriT) of the conjugal plasmid located in the donor cell (reviewed in (35)). The relaxase binds to the 5' end of the leading strand, and the protein-DNA complex is unwound from the plasmid. The nicked strand is then transferred into the recipient cell in a 5'-terminus to 3'-terminus direction. Host-mediated DNA replication synthesizes the complementary strands in both donor and recipient cells during the process of transfer. In the donor cell, the 3' end resulting from the nick is used to initiate continuous replication, as found with processing on a leading-strand template. However, in the recipient cell, DNA replication is continuously re-primed in a process more similar to events found on the lagging-strand template.

Filamentous phage has a single-stranded DNA genome protected by a protein coat during transfer. During the process of infection, the phage genome is injected into the cell cytoplasm. Upon entry into the cytoplasm, the ssDNA (called the "plus" strand, originally named for its role in Sanger sequencing) is converted into dsDNA by host-mediated replication. This newly made "minus" strand forms the template off of which more plus strands are synthesized. As they are synthesized, the new plus strands are separated from the DNA replication machinery, encapsidated by phage proteins, and assembled into progeny phage. Similarities in infection and replication of the M13 phage to the single stranded DNA bacteriophage PhiX (8) suggests that

the M13 plus strand is replicated in a continuous manner (similar to processing events on the leading strand template during replication), and the minus strand is replicated in a discontinuous manner (similar to processing events found during replication of the lagging strand template).

Transposition assays in which TnsABC+E directs a 'mini-Tn7' element containing the left and right ends and a selectable marker but no transposition machinery have been used to determine orientation of these elements into various substrates. From these assays, TnsABC+E mediated insertions into conjugal plasmids and M13 bacteriophage have been shown to occur almost exclusively in a singular orientation (9, 26). The cut and paste nature of Tn7 transposition leaves a 5 base pair duplication at the ends of the element, an indicator of Tn7 transposition. The process of transposition into a new site requires a nucleoprotein complex formed by the transposition proteins, the Tn7 ends, and the target DNA substrate. The differences between the left and right ends have been used to show that Tn7 has a strong orientation bias, where it inserts into the *attTn7* site and conjugal plasmids in a specific orientation. In the absence of conjugal plasmids, at a very low frequency TnsE-mediated transposition can also insert into the chromosome by recognizing aspects of DNA replication. Chromosomal replication is initiated bidirectionally from a single site in bacteria, the origin of replication (*oriC* in bacteria). In these transposition assays where a mini-Tn7 element is directed into the DNA by TnsABC+E, mini-Tn7 elements have been found in the chromosome in one orientation in each replichore of the chromosome, which corresponds to the direction of replication. The orientation-bias in conjugal plasmids and the chromosome is likely due to recognition of specific features found on these DNA molecules during replication. Two features have been shown to be important for TnsE-mediated transposition: the sliding clamp protein and 3' recessed DNA ends.

The sliding clamp is a protein that acts as a processivity factor in DNA replication, encircling the template DNA and freely sliding along it. The sliding clamp has homologs in all three domains of life where the replicative DNA polymerases alone are unable to replicate long stretches of DNA because they frequently dissociate from the DNA. An association of the sliding clamp with DNA polymerases tethers the polymerases to the template DNA. In bacteria, this protein (called the β-clamp) is encoded by the *dnaN* gene that is encoded in a highly conserved region of the genome near the origin of replication. The crystal structure of the *E. coli* β-clamp showed that two β-subunits are aligned head-to-tail forming a ring, where the two sides of the ring are not symmetrical; one is referred to the "carboxy-terminal face" (or C-terminal face) because the C-termini ends of the from each monomer are found on this face (29). The closed ring structure of the β-clamp necessitates a loading-mechanism which opens the β-clamp for assembly, and subsequently closes it around the DNA. The protein that catalyzes this reaction, the clamp-loader, specifically targets the β-clamps to sites where DNA synthesis is initiated and loads the β-clamp in the correct orientation for interaction with the DNA polymerases. Interactions of the β-clamp with polymerases and other proteins required for DNA replication and repair are mediated through the conserved hydrophobic cleft on the C-terminal face (14). The corresponding clamp binding motif of β-clamp interacting proteins is responsible, in part, for interactions between the β-clamp and the DNA polymerases. Initial bioinformatic analysis identified a clamp binding motif within sequences of DNA polymerases across diverse bacteria (7, 36). One residue on the rim of the clamp has also been shown to play an important role in regulating the interaction of polymerases with the DNA (12). The other face (N-terminal) of the clamp has not been implicated in functional interactions.

TnsE was shown to have a putative β-binding motif that resembles the consensus motif found in other bacterial host proteins (7, 21, 36). This TnsE- β clamp interaction was confirmed by yeast-two hybrid, electron-mobility-shift assay, far western blot, and surface plasmon resonance analysis (21). Additionally, Tn7 transposition could be reconstituted *in vitro* using a target plasmid containing a 20bp single-stranded gap in the duplex DNA and only showed insertions at a specific site adjacent to the 3' end of the ssDNA gap when preloaded with the β-clamp. Work in other labs indicates that the clamp is loaded onto DNAs in a single orientation relative to the 3'OH (13), and remains at the 3' OH end through interactions with single strand DNA (11). A direct interaction between the β-clamp and TnsE and a targeted insertion into a plasmid loaded with the β-clamp suggest that the β-clamp plays an important role in targeting insertions *in vivo*. Given that the β-clamp is more highly represented on the lagging strand template, this would suggest that TnsE is targeting to this strand via the β-clamp, which leads to insertions in one orientation corresponding to lagging-strand template replication.

TnsE interacts preferentially with DNA structures that contain 3' recessed ends but not structures with only 5' recessed ends (26). Given that TnsE gain-of-activity mutants identified in random screen were found to display a specific increase in the ability to bind 3' recessed ends, this interaction appears important for TnsE function (26). The availability of both 3' recessed ends and β-clamps in replicating cells would indicate that these targets are recognized *in vivo*.

Previous bioinformatics work in the Peters lab found that there are Tn7-like elements (elements which contain obvious homologs of TnsABCDE proteins) in widespread bacteria. Naturally, the question of whether Tn7-like elements could move between distantly related bacteria was examined by testing if TnsE from *Idiomarina loihiensis*, *Shewanella baltica*, or *Bacillus cereus* could mediate transposition using TnsABC from *E. coli*. When Tn7

transposition was monitored using TnsE from these foreign hosts with the core machinery from the canonical Tn7 element isolated originally from *E. coli* transposition could not be detected. Additional experiments tested whether functionally replacing β-clamps from these same three foreign hosts would now allow for transposition. In order to do this, a strain carrying an allele of the β-clamp that confers temperature sensitivity, *dnaN159* was utilized. By using the *dnaN159* strain β-clamps from various hosts could be expressed from a plasmid moving them into the strain at the permissive temperature, and their function could be assessed at the non-permissive temperature for the *dnaN* allele. At a high temperature, only the β-clamps from various hosts are expected to be functional in the cell, therefore in the absence of any exogenous β-clamp, the cell will not survive. In order to determine whether these foreign clamps would allow for cell survival, Zaoping complemented the *dnaN159* temperature sensitive phenotype with expression of the foreign clamps *in trans*. However, only *dnaN* from bacteria that are from the same Phyla (*I. loihiensis* or *S. baltica* but not *B. cereus*) allowed for survival. To test whether the foreign TnsEs could now mediate transposition in strains containing the foreign β-clamps, I constructed strains containing TnsABC from *E.coli* and TnsE from the foreign clamps. I then used these strains in an assay which uses a 'mini-Tn7' element containing lactose utilization genes which lack a promoter, making the strain phenotypically Lac-. If transposition occurs, the element moving downstream of a promoter causes the cell to become Lac+. However, other genetic rearrangements also allow a significant number of Lac+ events that are not related to movement of the miniTn7 element as seen in strains that lack essential transposition functions. In strains which contained TnsE and β-clamps from the same host the number of Lac+ colonies doubled as compared to a control lacking TnsE.

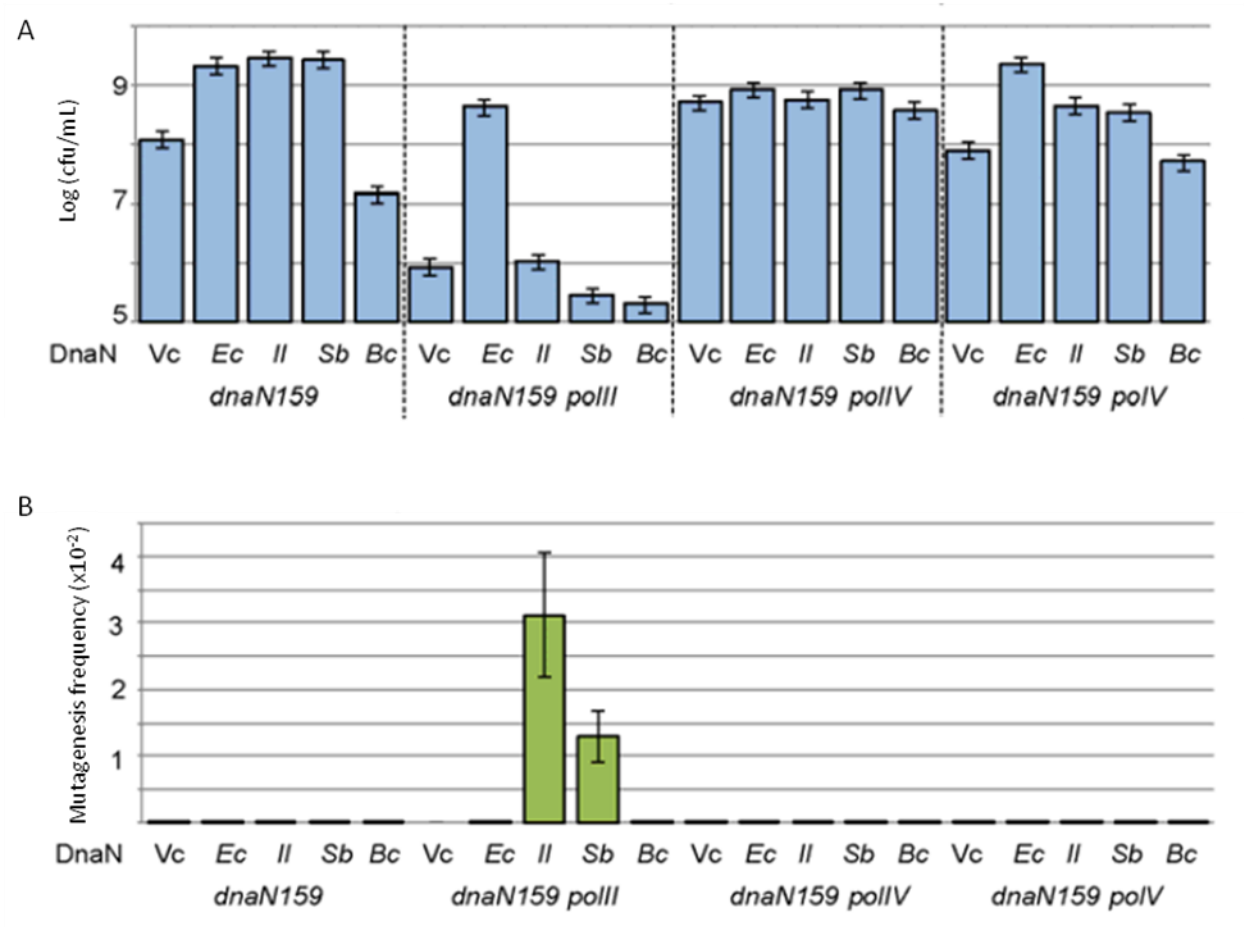

Figure 3.1

Individual accessory polymerase knockouts in *dnaN159* strains
Results were measured at 37°C as described in the Methods. Error bars indicate the standard deviation of mean. The four possible expressed DnaN are Ec *E. coli*; Il, *I. loihiensis*; Sb, *S. baltica*; Bc, *B. cereus*. Vc indicates vector control. **A)** After overnight growth, cultures were plated on LB agar, and the number of colony-forming units was counted **B)** After overnight growth, cultures were plated on media containing rifampicin and the number of colonies per total cell number was calculated.

I suggest that the domains used to interact are likely conserved based on mutations of the β-clamp interacting domain of TnsE that affect interactions with the β-clamp (21), which would lead us to believe that TnsE would be able to interact with β-clamps from different organisms. However, as indicated above, in strains where TnsE and β-clamps were from different hosts there

was no increase in papillae: only when these two proteins were from the same host show was an increase found, suggesting that TnsE can interact with a region of the β-clamp outside the hydrophobic pocket. An interaction with the rim of the β-clamp, which is outside of the pocket, has also been observed in one host protein, Polymerase IV (12), an accessory polymerase with a high mutation rate. Pol IV interacts with the rim of the clamp allowing it to bind to the clamp even when the normal polymerase is replicating DNA. While Pol IV is also capable of interaction with the hydrophobic pocket on the β-clamp, the hydrophobic cleft, its ability to interact with at least one residue outside of this region appear to be unique. However, proteins interacting with residues outside of the pocket have not yet been determined. The ability of Pol IV to interact with the rim is suggested to be used to interact with the clamp allowing a fast response when Pol III encounters damaged DNA over which it cannot replicate. If Pol IV is bound to the rim of the clamp during normal replication, as soon as Pol III stalls it is released from the DNA, Pol IV which is already at the site of damage is available to interact with the hydrophobic cleft and replicate over the damaged DNA.

In the experiments described above which utilizes a miniTn7 element as a proxy for transposition, containing the lactose utilization genes to a site downstream of a promoter, there are many Lac+ events even in the absence of TnsE. Previous studies have found that many Lac+ events occur in the absence of movement of the Tn7 element by unrelated genetic events, which are considered background (26). Additionally, at a very low frequency, (such as levels 10,000-fold less than normal TnsD-mediated transposition or 20-fold less than TnsE-mediated transposition when TnsE is expressed at a low level) some events associated with wild type TnsABC can be detected (25). I was interested in how either Lac+ events as a result of unrelated genetic events or the low frequency of Lac+ events associated with TnsABC may affect the

assay. I believe that the foreign clamps may exacerbate background events due to unrelated genetic events. One possibility for the increase in events is that error prone polymerases may have access to the replication fork because of mis-management of the proteins on the foreign clamps.

In order to address the possibility that the Lac+ events are not a result of transposition, I show that transposition in the strain containing TnsE and DnaN from the same host can only be accounted for by real transposition events. I also find that the percentage of transposition events in a strain containing the empty TnsE vector or strains expressing TnsE and DnaN from different hosts would not significantly affect the assay.

3.3 Results

3.3.1 TnsE and DnaN from the same host are required for transposition with E. coli core machinery

In order to determine if transposition can occur in strains expressing TnsE and β-clamps from other species, a promoter-capture (or papillation) assay was used. This assay uses a "miniTn7" element that encodes the genes for lactose utilization but lacks the requisite lac promoter. This miniTn7 element is located in a region of the chromosome where it is not transcribed, yielding a cell that is phenotypically Lac-. If the transposition machinery is provided *in trans*, the element can move to a new site in the chromosome, which may be downstream of an actively transcribed promoter, generating a cell that is phenotypically Lac+. The conversion of phenotypically Lac- to Lac+ cells can be monitored on indicator media, where red papillae arise from an otherwise lawn of white cells. The number of red papillae can indicate that transposition occurred, however, Lac+ papillae can result from events independent of transposition. For example, one type of change from Lac- to Lac+ that has previously been

identified is movement of an IS2 element endogenous to *E. coli* to a position where the IS2 promoter can read into the miniTn7 element (25). In previous work it was shown that the vast majority of the background events found in strains that did not have a functional target site selecting protein were the result of various type of genetic rearrangements unrelated to movement of the Tn7 element (25). Alternatively, it was also shown Tn7 can transpose at a very low frequency from TnsABC alone, in the absence of TnsD or TnsE (25).

Zaoping constructed strains to test transposition with TnsABC from canonical Tn7 and TnsE and β-clamp from foreign hosts (*I. loihiensis* or *S. baltica*) (17). When TnsE and β-clamps were derived from the same host, approximately twice as many Lac+ events were found as compared to strains containing foreign clamps and no TnsE. The level of Lac+ events in a strain containing foreign clamps and no TnsE was approximately the same as the level of Lac+ found when the foreign clamps and TnsE were from different hosts. Therefore, the simplest explanation was that transposition was occurring only in strains that had TnsE and foreign clamps from the same host. In fact, approximately 40% of the Lac+ papillae from the strain with the TnsE and β-clamp from the same host were confirmed to be actual transposition events. However, for an overabundance of caution to rule out the unexpected result where there was an increase in genetic mutations or rearrangements in the presence of the foreign clamps, I did genetic testing to confirm the background events were a result of previously events.

In order to explicitly show that the Lac+ papillae in the strains that did not have TnsE and β-clamp from the same host were not a result of actual transposition events, I did genetic and PCR testing (Figure 3.2). First, to analyze the Lac+ events in the strains containing TnsE and DnaN from different hosts, the papillation assay with the same TnsE and DnaN combinations was repeated. Then, red microcolonies were isolated, and the KanR determinant was moved to a

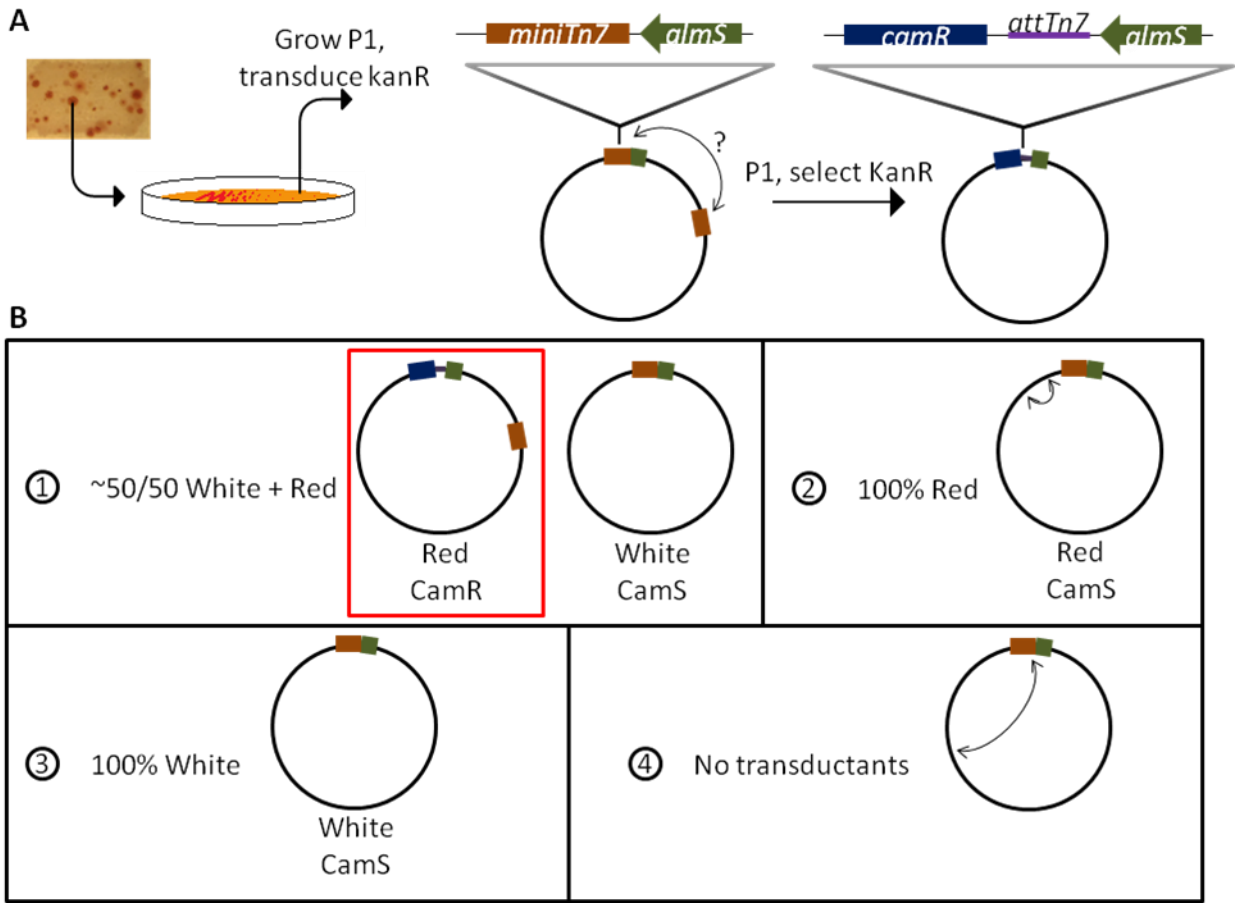

Figure 3.2

Genetic mapping of transposition events in Lac+ papillae
A) Workflow describing genetic mapping of Lac+ papillae. Red microcolonies were restreaked for single colonies. A P1 lysate was made from the restreaked colonies and transduced into a reporter strain containing a chloramphenicol resistance gene linked to *attTn7* selecting for Kanamycyn resistance. **B)** Four possible cell types after P1 transduction. 1) 50% white colonies are expected to be chloramphenicol sensitive, and 50% red colonies are expected to be chloramphenicol resistant 2) 100% red colonies are expected to be chloramphenicol resistant which are a result of small chromosomal rearrangements upstream of the element 3) 100% white colonies are expected to be chloramphenicol sensitive 4) No transduction could be a result of large chromosomal rearrangements

clean genetic background by P1 transduction. This clean background was a strain containing chloramphenicol resistance gene linked to *attTn7* (Figure 3.2A). These colonies were screened to determine if the Lac+ marker also moved (the colonies were either red (Lac+) or white (Lac-) on indicator media), or removed the CamR marker (the colonies were either chloramphenicol

sensitive or resistant). There were three combinations of Lac phenotypes observed: 50% red and 50% white colonies, 100% red colonies, or 100% white colonies after P1 transduction to the clean background.

For transductants that had 50% red colonies and 50% white colonies, the red colonies were either chloramphenicol resistant (indicating possible movement of the miniTn7) or chloramphenicol sensitive (indicating small upstream chromosomal rearrangement). In cases where 100% of the colonies were red, they are expected to be chloramphenicol sensitive. The third case was where 100% of the colonies were white and found to be chloramphenicol sensitive, indicating no movement of the miniTn7 element. The fourth scenario, where no transduction occurred, cannot be discounted, as it could be a result of large chromosomal rearrangements but may also be a result of experimental error. Colonies where the genetics suggested that transposition may have occurred were screened by PCR and the majority of those ruled out transposition of the element.

I found that transposition occurred in the *dnaN159* strain with the TnsABC core machinery plus TnsE in *E.coli* at very low levels, but when TnsE and DnaN came from the same foreign host, the transposition greatly increased. That is, results from the P1-mediated genetic transduction indicate that less than 13% of the strains containing DnaN from *I. loihiensis* and no TnsE were actual transposition events. While not confirmed by DNA sequencing, the strain containing DnaN from *E. coli* and TnsE from *I. loihiensis* had between 11-13% actual transposition events, and the strain with DnaN from *S. baltica* and TnsE from *I. loihiensis* had between 4-6% transposition (Figure 3.2). This supplements previous results from Zaoping that in the strains with TnsE and DnaN from the same hosts, 40% of the Lac+ colonies were a result of actual transposition events. Together, these results indicate that TnsE homologs would only

allow significant transposition over background levels in the assay when the β-clamp from the same host is also being used in the cell, or TnsE is, in fact, species-specific. These results rule out the concern that the level of papillation events that are due to TnsABC are not skewing the results.

3.3.2 DnaN from foreign hosts allows E. coli growth

The β-clamp is made up of two DnaN monomers, therefore having two varieties in the cell (β159 and foreign clamps) may lead to the formation of β-clamp heterodimers. Zaoping took a lot of care to show that the foreign clamps are capable of restoring wildtype phenotypes in the *dnaN159* mutation- especially temperature sensitivity (17). While it is known that the β159 homodimer is incapable of DNA replication at a higher temperature (4), the ability of the β159 molecule at permissive temperatures to complex with wild-type or foreign β-clamps remains to be studied.

To investigate the possibility that foreign β-clamps are sufficient to allow viable cells, the chromosomal copy of *dnaN* was deleted, so that the only β-clamp available in the cell was provided *in trans* from the plasmid. The chromosomal *dnaN* was knocked out by way of Kan-I-SceI (described in Methods, Figure 3.4B). Briefly, to facilitate deletion of *dnaN*, a strain in which DnaN-6xHis was expressed from an alternate locus and also had the capability of expressing the lamda red recombination proteins (Exo, Bet, and Gam) from the chromosome was used. An intermediate strain was generated using lambda red recombination to swap the *dnaN* gene with a DNA fragment containing a Kanamycin resistance gene and I-SceI recognition site. To generate this strain, a DNA fragment that encoded the KanR gene containing tails of DNA homologous to the DNA flanking *dnaN* was added to a strain expressing the lambda red recombination machinery, and selected for resistance to Kanamycin. This intermediate strain

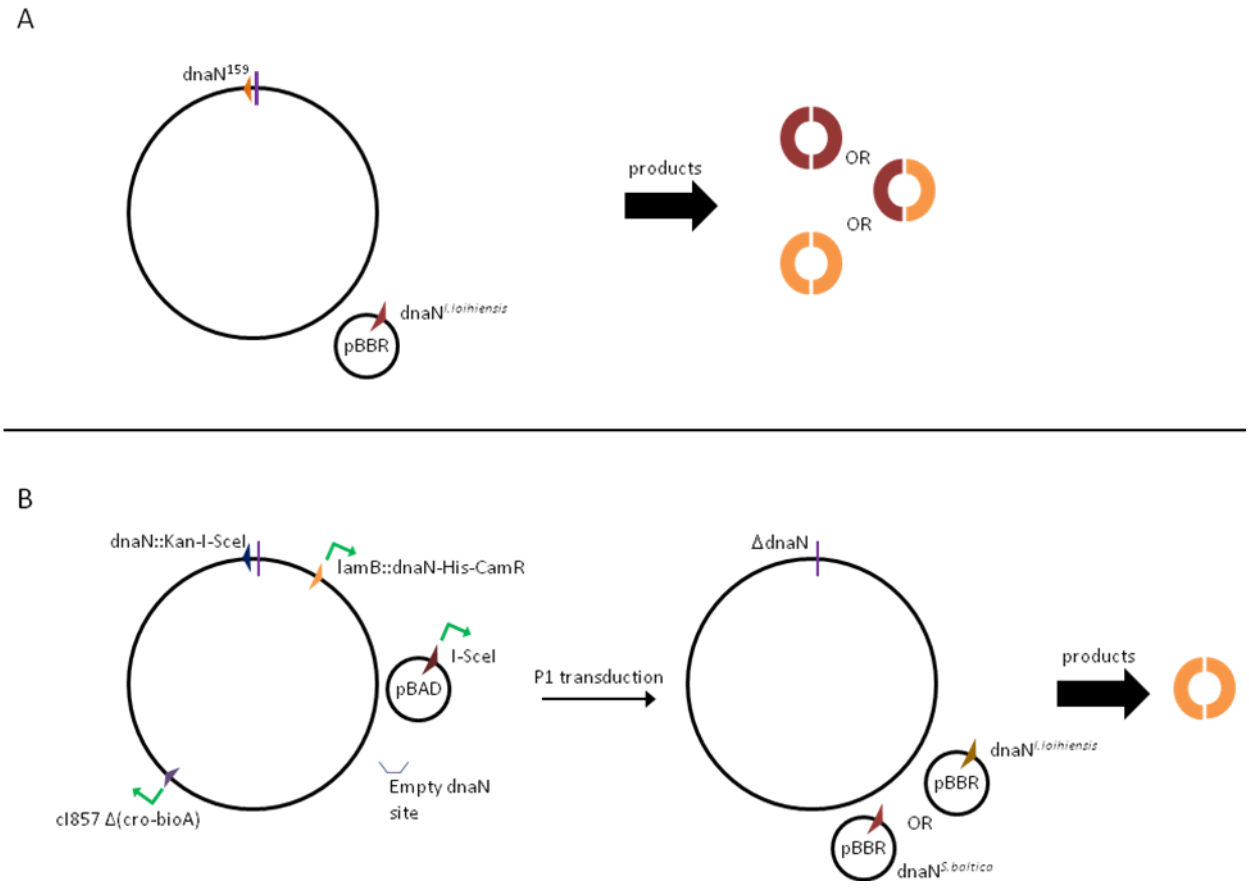

Figure 3.3

Strain development of the *dnaN*⁻ strain containing foreign clamps expressed *in trans*.
A) Original strains were able to express *dnaN159* at 30^{O}C, therefore the possibility that DnaN^{159} homodimers, DnaN^{Il} homodimers, and DnaN $^{159+Il}$ heterodimers are formed. **B)** Strains were made by replacing wild-type *dnaN* with a Kan-I-SceI element using a strain containing temperature inducible exo-beta-gam (cI857 Δ(*cro-bioA*)), selecting for KanR cells. The fragment was deleted by inducing I-SceI from a plasmid, and screening for Kanamycin sensitivity. Primers flanking the deletion site were used to amplify and sequence the deletion.

was then transformed with a DNA fragment containing the "empty site" and a plasmid with an inducible I-SceI endonuclease. The "empty site" was comprised of DNA flanking the chromosomal *dnaN* without any *dnaN* sequence. I-SceI cleavage of the chromosome allows lambda-red recombination between these double stranded breaks and the linear double stranded DNA fragment. The resulting strain was confirmed to be deleted for *dnaN*, and the deletion was

moved by P1 transduction of the linked TnaA::Tn10-TetR gene into our transposition assay strain containing *dnaN* from various hosts provided *in trans*. In separate experiments, the KanR-I-SceI cut site was moved into the *dnaN* locus in the transposition assay strain expressing *dnaN* from foreign hosts *in trans*. In both cases, moving the empty site or the Kan-I-SceI cut site into the *dnaN* locus of the transposition assay strain could be done when the β-clamp from *I. loihiensis*, *S. baltica*, or *E. coli* was expressed from a vector, but not when the cells contained an empty vector or the β-clamp from *B. cereus*. These data confirm that *dnaN* from heterologous hosts can be recognized in the cell by the host machinery to support growth of *E. coli*.

3.3. 3 Is the mutation frequency in dnaN159 strains expressing DnaN from Idiomarina loihiensis the result of real mutations?

The accessory DNA polymerases may be able to interact with these foreign β-clamps and therefore may be able to act at the replication fork when Pol III disassociates from the DNA. Previous work by Zaoping showed that other polymerases are required for DNA replication: in the absence of Pol II in strains containing *dnaN* from foreign hosts, the cell density after overnight at a semi-permissive temperature was three orders of magnitude less than the wild-type, an effect that was not observed in the absence of either Pol IV or Pol V (at 37^{o}C; between the 'permissive' temperature of 30^{o}C and the 'non-permissive' temperature of 42^{o}C) (Figure 3.1A, (17)). One explanation for dependency on Pol II is in the absence of Pol II, these other two accessory polymerases are not impeded from replicating over damaged DNA. While Pol II has proofreading capabilities, these other two polymerases are highly mutagenic. In support of this, it was found that a strain expressing foreign clamps that lacked Pol II had a greater mutation frequency, as measured by spontaneous mutations of the *rpoB* gene (a subunit of the RNA polymerase holoenzyme to which rifampicin binds) to generate rifampicin resistance cells

(Figure 3.1B, (17)). Frequency is measured by the number of rifampicin resistant colonies over the total number of colonies from the same culture. Strains that lacked either Pol IV or Pol V did not show this same increase in spontaneous mutation frequency (Figure 3.1B, (17)).

I set out to determine whether the increased mutation frequency in the Pol II knockout *dnaN159* strain were true mutations. These experiments required repeating the spontaneous mutation frequency assay in which an overnight culture is spread on agar media containing rifampicin and the gene responsible for growth, *rpoB*, sequenced from the resulting colonies. I was able to show that the cells able to grow on this medium had true mutations of the *rpoB* gene in the active site (Figure 3.3).

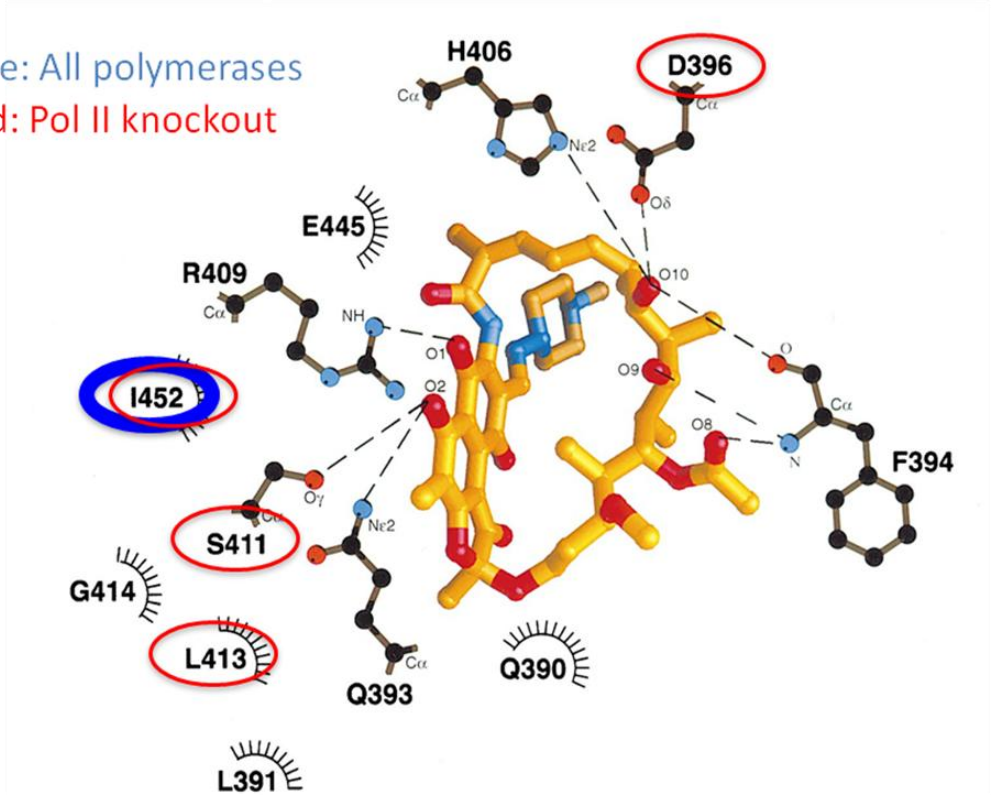

Figure 3.4

RpoB mutants mapped onto a depiction of the active site structure of RpoB bound to rifampicin (5). Mutant amino acids circled in red or blue. These were identified as Rifampicin resistant colonies from the *dnaN159* strain containing pBBR-*dnaN-I.loihensis*, with or without polB. The thickness of the circle indicates the number of mutants identified at that residue.

In the absence of Pol II, I can make two arguments that Pol IV is the main agent in causing the increased cell mortality. The first argument is based on deletion of another polymerase, Pol V, where the expection is that in the absence of Pol V, there is an increase in the usage of the other accessory polymerases, including Pol IV. Therefore, the observation by Zaoping that when Pol V is knocked out in strains containing foreign clamps, there is an order of magnitude drop in survival rate from an overnight culture, would suggest that the other polymerases, and likely Pol IV, are responsible for the drop in survival rate. The other argument I can make for Pol IV leading to increased mortality is that a knockout of Pol IV leads to *increased* cell survival in a strain expressing *dnaN* from *B. cereus* or the empty vector as compared to a wild-type control containing these vectors. These results jibe with findings that Pol IV overexpression in a *dnaN159* strain leads to impeded cell growth, which is thought to be the result of the ability for Pol IV to switch more efficiently with Pol III than with other polymerases, and once in this position, prevents other polymerases from accessing the fork (12). In this scenario, both Pol II and Pol IV may compete for binding the β clamp, such that in the absence of one polymerase, the other would not be blocked from interacting with the β159 protein. In the case of a Pol IV knockout, the Pol II would have access to β159 and be able to 'stabilize' it at higher temperatures, leading to increased cell survival. This suggests that in cells lacking Pol IV the cell's recourse is to utilize Pol II and Pol V to replicate over damaged DNA, neither of which have a negative effect as severe as Pol IV.

Strains containing *dnaN159* and the complete set of polymerases were used to assess TnsE-mediated transposition in the presence of foreign β clamps. Given that Pol IV has been shown to switch with Pol III at a greater efficiency than the other polymerases, it is possible that it leads to an increase in the mutation frequency. If Pol IV does lead to a higher mutation rate, it

may also be responsible for the increased number of background events which are observed as Lac^+ papillae in our transposition assay strain. While I showed that these background events are not transposition events (section 3.3.1), it remained uncertain which accessory polymerase was causing this effect.

While Pol IV and Pol V lack proofreading capabilities, Pol II has an exonuclease domain which is capable of removing an incorrect base, so any incorporated error can be removed and the polymerase can 'try again'. In the presence of a DNA mismatch, if Pol II inserted the wrong base, the error would be rectified. However, the other two polymerases do not have this function and any mismatched bases are not fixed by the polymerase. One commonly used assay to determine mutation frequency in different bacterial strains is the rifampicin resistance assay, in which the number of colonies capable of growing on rifampicin plates is divided by the total number of viable cells to yield the mutation frequency. Additional analysis by sequencing the *rpoB* gene can determine the capacity of polymerases to replicate over different types of DNA damage, resulting in different types of DNA mismatches, and therefore a difference in the overall mutation spectrum for each polymerase (base substitutions, frameshifts, deletions, insertions) Previous studies have determined differences in the types of mutations in a wild-type strain as compared to a strain which lacks the exonuclease domain of Pol II. More specifically, while there was a set of mutations of *rpoB* identified in the wild-type strain, an additional site was found in the Pol II exonuclease knockout strain (28). This contrasts with another polymerase, Pol IV, which when overexpressed, has a different distribution of mutations in *rpoB*, where three hotspots were identified (37).

I attempted to determine which polymerase is responsible for the increased mutation frequency in the Pol II knockout strain by sequencing the *rpoB* gene from colonies growing on

media containing rifampicin. Although there are limitations on the types of mutations identified by this experiment because the RpoB protein is essential, other researchers have used this experiment to identify differences in the types of mutations when various polymerases are overexpressed (28, 37). From this, I determined that the *pol II* deletion strain had a different spectrum of mutations in *rpoB* than the wild-type (Figure 3.3). Despite this difference, it did not produce any of the same hot spots as previously observed (28, 37). These differences may be a result of capabilities of Pol IV and Pol V to weakly interact with either *dnaN159* or the foreign clamps in the absence of Pol II.

3.4 Discussion

3.4.1 Tn7 species specificity: the TnsE-β clamp interaction

Diverse bacterial hosts that contain Tn7-like elements located in attTn7 have been isolated from a broad range of environments (22, 23). The ability to modulate between inserting into conjugal plasmids and into attTn7 would presumably enable the Tn7 element to transfer horizontally and then into the chromosome. Insertions into conjugal plasmids are mediated by TnsE, whereas chromosomal insertions are mediated by TnsD. While the proteins responsible for targeting insertions into the chromosomal site have been well characterized, less is known about TnsE-mediated insertions. Previous work has pointed to the β-clamp of DNA replication and 3' recessed ends as interacting partners for TnsE (21, 26). The β-clamp is diverged, despite its requirement for cell viability.

The β-clamp is a protein encircling DNA capable of sliding along the DNA, which is essential to coordinating events on the DNA during and after replication, including events related to DNA replication, recombination, and repair. The β-clamp limits access to the DNA by providing one surface to which these proteins bind, forcing interacting proteins to act in a

progressive manner. For example, the clamp loader complex must first load the clamp onto the DNA before the β-clamp can interact with any of the DNA polymerases. The surface of the clamp to which these proteins bind is conserved across the entire face, but more importantly contains a highly conserved hydrophobic cleft. More than one β-clamp binding motif has been found that interact with this cleft.

Given the essentiality of the β-clamp, similar ring-shaped structures that coordinate processes on the DNA are found in other domains of life, including bacteriophage T4 (19). Despite being highly diverged, the clamp loader from *Staphylococcus aureus*, a Firmicute, is capable of binding to and loading β-clamps from either *E. coli* (a Proteobacterium) or *Streptococcus pyogenes* (a Firmicute) onto DNA *in vitro*. However, these foreign (*E. coli* or *S. pyogenes*) β-clamps are unable to reconstitute a functional DNA polymerase holoenzyme with the normal DNA polymerase from *Staphylococcus aureus in vitro* (2, 16). That is, after reconstitution of the holoenzyme, replication of the DNA was not observed. This is thought to be due to a strong interaction between the clamp loader and the clamp such that the clamp loader cannot release from the clamp, or a weak interaction between the polymerase and the clamp such that the polymerase readily falls off the clamp. In the first scenario, it is possible that the clamp loader recognizes additional residues on the clamp, possibly in less-conserved regions of the clamp. In the second scenario, the heterologous polymerase may normally require interaction with residues that are not found on the foreign clamp. Both possibilities are in concordance with the result that clamp interacting proteins may recognize more than the hydrophobic cleft, a result that has been noted in PolIV, which is thought to have independent interactions with the rim of the clamp (12, 34). Additionally, other residues important in binding a domain outside of the hydrophobic pocket may be specific to certain proteins within the cell. Previous results have

shown that the β-clamp from *I. loihiensis* and *S. baltica* but not *B. cereus* are able to complement the phenotypes of the *dnaN159* allele (17), suggesting they are recognized by the clamp-loader complex and the normal DNA polymerases. Because we do not know the nature of the inability for *B. cereus* to complement this phenotype, it remains to be seen whether regions other than the conserved pocket are used to form important contacts to interacting proteins.

TnsE has a variant of the β-clamp binding motif that is conserved (21), which would suggest that it is capable of interacting with the conserved hydrophobic pocket of the β-clamp. However, only TnsEs and β-clamps from the same host has an increased papillae number. The species-specific interaction was drawn into question because of the high level of papillae in strains lacking TnsE or containing TnsE and DnaN from different hosts. I showed that only a small fraction of these putative events are real transposition events (Figure 3.1), which complements the previous result of a higher percentage of actual transposition events in the strain containing TnsE and DnaN from the same host (personal communication, Zaoping Li). The percentage of actual transposition events in the four strains studied in combination with the number of papillae indicates that there is nearly four times as much transposition in strains containing TnsE and DnaN from the same host a result that is consistent with the increase in papillation coming form actual transposition events. This confirms that significant transposition cannot be found to explain the Lac+ papillae in strains containing TnsE and DnaN from different hosts. These results confirm a species-specific interaction between TnsE and the β-clamp. Because both the clamp binding domain of TnsE and the hydrophobic pocket of DnaN are conserved, the species specificity would suggest that TnsE interacts with regions of the clamp outside of the binding pocket, that are not conserved. This is supported by the recent finding that other proteins, such as Pol IV that interacts with a residue on the rim of the clamp which is not

well conserved (12). It is also supported by the finding that only overexpression of TnsE induces the SOS response (21), likely because it does not interact exclusively with the hydrophobic cleft. To support this, the β-clamp containing a G174A mutation located near the hydrophobic cleft but not the G174/G66E double mutant which has an additional mutation at the rim of the clamp structure stimulates TnsE-mediated expression (21).

The species specificity maintained by the TnsE-DnaN interaction confirms other observations that Tn7-like elements seem to have host-specificity, in that similar Tn7-like elements are found in related bacteria (17). The TnsD specificity for the highly conserved *attTn7* site (rather than a host protein partner) and ability to mediate insertions across domains of life would suggest that it does not harbor any species specificity.

The resulting model is that the TnsE from a Tn7 element in the chromosome is expressed such that when a conjugal plasmid enters this "recipient cell", TnsE mediates insertions into the conjugal plasmid. After movement of the plasmid to a new cell, the element is transferred to the attTn7 site by TnsD-mediated transposition. At that point, if this new host is evolutionarily close to the previous host, the process will begin anew. If, on the other hand, the new host is highly diverged, given the inability for TnsE to interact with the β-clamp, the Tn7 element would remain in the attTn7, unable to move again. Subsequent mutations in TnsE could adapt to the new bacterial strains, allowing transposition into conjugal plasmids in these strains, but would then not be able to mediate transposition in the original strain.

3.4.2 Polymerase interactions with foreign clamps

Recent work has shown that the DNA polymerase, Pol IV, can interact with unique contacts on the β-clamp (i.e. the rim or N-terminal face) to compete with each other for interaction with the clamp binding motif (12, 30, 31, 32). Experiments utilizing mutations of Pol

IV that disrupted specific interactions with the β-clamp showed that it relied on both the interactions with the hydrophobic cleft and at least one residue on the rim of the β clamp to allow it access to the replication fork (12). The work done by Heltzel et al. 2012, utilized the *dnaN159* strain, which was also used in the work shown in this chapter (12). To study the interactions of various polymerases with β159 in vitro, temperature sensitivity or UV sensitivity can be monitored. However even at a permissive temperature, increased expression (~28 fold greater than steady state conditions) of Pol IV led to growth impairment of *dnaN159* strains (12). This is thought to be due to a greater availability of the β159 protein, where Pol IV is actively competing for the clamp, leading to poor growth. In contrast, increased levels (~7 fold higher than background) of Pol IV confer UV sensitivity in *dnaN159* mutant background (12). In an *in vitro* system, Pol IV interacts with the rim of the β-clamp during normal DNA synthesis by DNA Pol III, and is capable of interacting with the β-clamp binding motif when the Pol III is released from the β-clamp, such that the hydrophobic pocket can interact with other polymerases (15). All of these data together suggest that Pol IV is likely the polymerase responsible for the negative phenotypes observed in a *dnaN159* strain, and may do so by gaining inappropriate access to the replication fork. Furthermore, over-expression of Pol IV leads to severe growth impediments in strains containing the β159 protiein (12). From this we can begin to ask the question: do polymerase interactions with foreign clamps mirror those of β159? This model can be confirmed in future work where the accessory polymerases are deleted as single, double, or triple knockouts in the presence of only the foreign clamps and testing if these strains have decreased cell viability and increased mutation frequency.

Because I can completely delete the *dnaN159* allele in the presence of foreign clamps, I suspect that the foreign clamps form a homodimer even in the *dnaN159* strain. Therefore, in

experiments studying the effect of polymerase knockouts on the mutation frequency in the presence of foreign clamps, the polymerases are able to interact with these foreign clamps, causing the increased mutation frequency. In the absence of Pol II, one of the other two accessory polymerases, Pol IV and/or Pol V, are expected to interact with the β-clamp. Given that these two polymerases have a high rate of mismatch incorporation, when mutation frequency was monitored, an increase in mutation frequency was observed. By sequencing *rpoB* mutants, I was able to confirm that mutations were real events.

Additional work showed that these mutations were not similar to any previously established pattern caused by an accessory polymerase. One possibility for these observations is that Pol IV is unable to bind the rim of the foreign clamps via the secondary binding motif because they are too diverged. This binding which would allow Pol IV to be bound to the clamp during normal replication and 'swing in' to position when Pol III is unable to replicate over damaged DNA, may no longer be available. Pol V has not been shown to have this ability to interact with a residue on the rim of the clamp, therefore, in the absence of Pol II in these strains, it may have an equal chance at accessing the β-clamp as Pol IV.

3.4.3 Homodimer formation

How polymerases use the β-clamp in strains containing both the *dnaN159* allele and foreign clamps could not be confidently assessed given the possibility that the cells have both the β159 protein and the foreign clamps. Despite previous experiments that were done at a higher temperature, these experiments do not rule out the possibility of heterodimer formation (17). The β159 protein has two mutations, G66E located at the rim of the clamp structure, and G174A located on the conserved C-terminal face, near the hydrophobic cleft. β159 is capable of supporting growth at 30^{o}C, but at high temperatures the strains cannot survive. Based on the

inability for the purified protein to restore holoenzyme activity to PolIII at high temperatures in vitro (4), it is possible this is what leads to the inability for the cells to survive at the high temperatures *in vivo*. The high temperatures are thought to disrupt the protein structure, diminishing the capacity of the β-clamp of interacting with the normal polymerase used in DNA replication, DNA Polymerase III. Even β159 purified from cells grown at 30^{o}C has reduced ability to interact with the rest of the Pol III holoenzyme to synthesize DNA (4), possibly due to an impaired interaction with the α- subunit of this polymerase (30). However there was one possibility that the β159 protein may have been stabilized by the formation of heterodimers with foreign clamps. While the rifampicin mutation results would suggest absence of the *dnaN159* (presence of β159 in a Pol II knockout strains grown on rifampicin would leave mutational hotspots in *rpoB* characteristic of Pol IV, resulting from an increase in the access of Pol IV to the β-clamp), heterodimer formation could not be completely ruled out.

In order to address this issue, a complete *dnaN* knockout (lacking any chromosomal *dnaN* allele), with the foreign clamps expressed *in trans*, would be viable. I was able to generate this strain multiple ways, and show that DnaN from *I. loihiensis* and *S. baltica* are able sustain growth of these strains.

3.5 Methods

3.5.1. Media and chemicals

Minimal media, LB broth, and LB agar were prepared as described by Miller (20). Minimal Media was supplemented with 0.2% appropriate carbon sources (glucose, maltose). LB was supplemented with 0.2% glucose to suppress transcription of the Plac promoter. Ability to utilize lactose was evaluated on MacConkey lactose agar (Difco) and arabinose was added to 0.2% to induce expression of the P_{BAD} promoter. Isosensitest agar (Oxoid) was prepared for

growth in the presence of trimethoprim. Antibiotics were used at the following concentrations: ampicillin (Amp), 100 μg/ml; chloramphenicol (Cm), 30 μg/ml; kanamycin (Km), trimethorpim (Tp), 100 μg/ml; and rifampicin (Rif), 50 μg/ml.

3.5.2 Bacterial strains and plasmids

An W3110 Δ(*argF-lac*)*U169* {λ*c*I*857* Δ(*cro-bioA*)} derivative strain containing lamB::*dnaN*-camR (AN13) was used to construct *dnaN* deletion. Deletions were generated by an initial amplification of a Kanamycin cassette containing an I-SceI cut site with tails complementary to sequence flanking the target gene. This amplified PCR product was cleaned following kit instructions (Promega). Strain AN13 was grown to mid-log at 30^{o}C, incubated at 42^{o}C for 15 minutes to induce lambda red recombination machinery, and immediately chilled. The strain was then prepared for electroporation as previously described (39), briefly the cells were rinsed three times and resuspended with ice cold ultrapure water (HPLC grade), the DNA fragment (the Kan-I-SceI deletion fragment) was added, the cells were transferred to 0.2mM cuvettes and electroporated according to electroporator protocol (BioRad). LB media was added to the cell slurry, incubated at 30^{o}C for one hour, and the entire cell mass was plated on LB plates containing kanamycin to select for Kan^{+} recombinants. After PCR screen of successful recombinants, *tnaA*::Tn10-TetR was introduced by P1 phage grown on AP200 (NLC28 *tnaA*::Tn10-tet). This locus is 0.16 minutes from *dnaN*, so after selecting transductants on Tet, and screened on Kan, the tnaA::Tn10-TetR is linked to the *dnaN*::Kan-I-SceI (AF651). The deletion was generated as indicated above, simultaneously transforming in two DNAs: 1. a PCR product of the deleted fragment with tails complementary to sequence flanking the target gene, and 2. an a pBAD24 expression vector containing an arabinose inducible I-SceI restriction enzyme. Immediately upon electroporation, LB containing 0.2% arabinose was added to the

growth medium to induce the I-SceI enzyme. Transformants were selected for on LB+ 0.2% arabinose + 100ug/mL ampicillin. Resulting colonies were streaked onto LB+ kanamycin +ampicillin or LB+ ampicillin only. Transformants that were unable to grow on media containing kanamycin were then identified as having the *dnaN* deletion using PCR to screen for the expected size fragment. DNA sequencing confirmed that unexpected errors were not introduced during PCR amplification. P1 phage was grown on this strain (AF659-665) and transduced into JP1386 expressing *dnaN* in *trans* from either *E. coli* (AF807-810), *I.loihensis* (AF859-864), or *S. baltica* (AF865-870), selecting for resistance to tetracycline (which is now linked to the *dnaN* deletion). The final step was to introduce the mTn7::lacZYA'::Kan cassette which is used in the papillation assay (described below). P1 phage was grown on a strain containing this cassette (JP1776), transduced into the final recipient strain (either AF809, AF862, or AF868), selected for Kanamycin resistance, and screened for Tet. A second confirmation of the deletion was obtained by PCR screen for the expected fragment size, as described below; the resulting strain is: JP1386 Δ*dnaN* tnaA::Tn10-TetR mTn7::lacZYA'::Kan $pB^{Ec,\ Il,\ or\ Sb}$.

3.5.3 Papillation assay

To determine the frequency of transposition in the background strains, first the papillation assay previously known as promoter capture assay was repeated (3)(27). In this assay, a mini Tn7 element encoding lactose utilization genes flanked by the cis-acting ends of Tn7 from *E. coli* and which lack a promoter was present in the *att*Tn*7* of the background strain. In this site, the lac genes are not transcribed, therefore the strain is unable to utilize lactose (Lac^-), so when patched in this assay it presents as a white lawn on MacConkey's lactose indicator media. Transposition to a new position downstream of a promoter in an orientation that would allow for transcription of the Lac genes, the cells will be phenotypically Lac^+, presenting as red

microcolonies on the lawn of Lac^- cells. The number of Lac^+ papillae provides an indicator of transposition levels. The transposition proteins are provided *in trans*, where TnsABC are provided on *pCW15*, and TnsE homologs are provided on *pBAD24*. A negative control (strain lacking TnsE) is always included to indicate "background events" which are likely a result of random chromosomal rearrangements and other mutations upstream of the mTn7 element (25).

These background events need to be confirmed that they are not the result of actual transposition events. In order to do this, I used a combination of genetic and PCR confirmation. Genetic testing involved streaking out the red papillae, making a P1 lysate from these strains, transducing into a clean genetic background, screening these transductants for ability to utilize lactose and loss of chloramphenicol resistance, and PCR amplification of the attachment site. The clean background used had a chloramphenicol resistance gene linked to (78kb from) the attachment site. If transposition did not occur, resulting P1 transductants will remain white when streaked on indicator media and lose the chloramphenicol resistance. However, if transposition did occur to a region far from the attachment site, resulting transductants will be red when streaked on indicator media, and still be resistant to chloramphenicol. In greater detail, two red papillae were streak purified from the lawn onto MacConkey's agar, grown at 30^oC overnight and inoculated into LB+0.2% maltose. A P1 phage lysate was made from these strains, and transduced into a strain carrying attTn7::cat::FRT. The transductants were selected on MacConkey's agar containing Kanamycyin (to select for transductants that crossed in DNA containing the mTn7::lacZYA'::Kan element). The resulting colonies were further screened for chloramphenicol sensitivity by streaking on both LB and LB+30ug/mL Chloramphenicol plates, where sensitivity indicates movement of the mTn7::lacZYA'::Kan element into or nearby the attachment site. P1 events that resulted in red colonies were further mapped by amplification of

one end of Tn7 and the flanking chromosomal glmS, as described below. This amplification acted as a final confirmation in scenarios where no movement of the element was suggested by genetic testing, and as a mechanism to distinguish between actual transposition and small genetic rearrangements. Running the amplified product on a gel showed a band, indicating presence of the mTn7 element in the attachment site, or no band, indicating inability to amplify the element from the attachment site, which was attributed to an actual transposition event.

3.5.4 Spontaneous mutation analysis

Spontaneous mutations were first generated by an assay adapted from Sutton et al. 2005 (32), which was originally used to monitor rifampicin resistance. Each strain background to be tested was grown overnight at 30^{o}C on minimal media agar + 0.2% glucose containing the appropriate antibiotics. The *rpoB* genes of resistant colonies were amplified by inoculating one colony into minimal media broth +0.2% maltose at 30^{o}C, shaking overnight. Cultures were plated onto LB+50 μg/ml rifampicin, and grown overnight. Colonies were re-streaked onto LB +rifampicin, picked into LB, and grown at 30^{o}C overnight, shaking. DNA was extracted using phenol chloroform-CTAB (20). Briefly, cells were lysed with .05% SDS, and 0.1ug/mL proteinase K. NaCl was added to a final concentration of 0.7M . Debris was removed with the addition of CTAB in a NaCl solution to a final concentration of 1% CTAB/1.4MNaCl. Mixtures of Chloroform-Isoamyl Alcohol and Phenol-Chloroform-Isoamyl Alcohol were added sequentially to remove CTAB-protein complexes. DNA was precipitated with isopropanol, washed with 70% ethanol and resuspended in sterile water. The DNA fragments resulting from the PCR reactions were sequenced to distinguish actual mutations of this gene, responsible for sensitivity to rifampicin.

3.5.5 Polymerase chain reaction conditions

All PCR reactions were run in a 25-ul reaction mix containing 0.2uM each primer, 0.15mM deoxynucleoside triphosphosphate, 40mM $MgCl_2$, 1X PCR Buffer, and 0.2uL Taq DNA Polymerase per reaction. The primers and PCR conditions varied as follows. All PCR products were run on a 1% agarose gel and visualized after staining with ethidium bromide.

Amplification of the *dnaN* region was used to confirm *dnaN* deletion. A 1,599 base pair region of the chromosome flanking and including *dnaN* was amplified. The forward primer was 5'-GCATTGCAGGAAAAACTGGT-3', and the reverse primer was 5'-CTCATGGCGAATGACGCGAC-3'. Touchdown PCR conditions were used: an initial denaturation at 95^{o}C for 2 minutes, followed by 11 cycles of denaturation at 95^{o}C for 30s, annealing at 65^{o}C for 30s with a 1^{o}C decrease every step, extension at 72^{o}C for 30s, 25 cycles of denaturation at 95^{o}C for 30s, annealing at 55^{o}C for 30s, extension at 72^{o}C for 30s and a final extension at 72^{o}C for 2 min.

Amplification of the left end of Tn7 and DNA flanking the insertion site was as previously described (25). The forward primer was 5'-GCGTGGCATCCACTAAAGCATTCA -3' and the reverse primer was 5'-ACTTTATTGTCATAGTTTAGATCTATTTTG -3'. The PCR conditions were an initial denaturation at 95^{o}C for 2 minutes, followed by 30 cycles of denaturation at 95^{o}C for 30s annealing at 60^{o}C for 30s and extension at 72^{o}C for 30s, with a final extension at 72^{o}C for 2 min.

Amplification of *rpoB* was modified from experiments previously described (32). A 988-base pair region of the *rpoB* gene (which encompasses the active site that interacts with rifampicin) was amplified by PCR. The forward primer was 5'-TCGAAGGTTCCGGTATCCTGAGC-3' and the reverse primer was 5'-GGATACATCTCGTCTTCGTTAAC-3'. The PCR conditions were an initial denaturation at

95°C for 2 minutes, followed by 30 cycles of denaturation at 95°C for 30s annealing at 49°C for 30s and extension at 72°C for 30s, with a final extension at 72°C for 2 min. PCR products that gave a visible band were cleaned using the Promega PCR purification kit, and submitted for sequencing.

3.6 Acknowledgements

I thank Mark Sutton (Department of Biochemistry, School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York) for *E. coli dnaN* mutant strains. I also thank Aravind Natarajan, Julian Harrison, Mike Petassi, and Bennett Kapili for assistance with experiments.

REFERENCES

1. Arciszewska LK, Craig NL. 1991. Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. *Nucleic Acids Res*. 19(18):5021–29

2. Bruck I, Georgescu RE, O'Donnell M. 2005. Conserved interactions in the *Staphylococcus aureus* DNA PolC chromosome replication machine. *J. Biol. Chem.* 280(18):18152–62

3. Bruck I, O'Donnell M. 2000. The DNA replication machine of a gram-positive organism. *J. Biol. Chem.* 275(37):28971–83

4. Burgers PM, Kornberg A, Sakakibara Y. 1981. The *dnaN* gene codes for the beta subunit of DNA polymerase III holoenzyme of *Escherichia coli*. *Proc Natl Acad Sci* 78(9):5391–95

5. Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, et al. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*. 104(6):901–12

6. Choi KY, Li Y, Sarnovsky R, Craig NL. 2013. Direct interaction between the TnsA and TnsB subunits controls the heteromeric Tn7 transposase. *Proc Natl Acad Sci*. 110(22):E2038–45

7. Dalrymple BP, Kongsuwan K, Wijffels G, Dixon NE, Jennings PA. 2001. A universal protein–protein interaction motif in the eubacterial DNA replication and repair systems. *Proc Natl Acad Sci*. 98(20):11627–32

8. Eisenberg S, Denhardt DT. 1974. Structure of nascent ϕx174 replicative form: evidence for discontinuous DNA replication. *Proc Natl Acad Sci*. 71(3):984–88

9. Finn JA, Parks AR, Peters JE. 2007. Transposon Tn7 directs transposition into the genome of filamentous bacteriophage M13 using the element-encoded TnsE protein. *J Bacteriol*. 189(24):9122–25

10. Gamas P, Craig NL. 1992. Purification and characterization of TnsC, a Tn7 transposition protein that binds ATP and DNA. *Nucleic Acids Res*. 20(10):2525–32

11. Georgescu RE, Kim S-S, Yurieva O, Kuriyan J, Kong X-P, O'Donnell M. 2008. Structure of a sliding clamp on DNA. *Cell*. 132(1):43–54

12. Heltzel JMH, Maul RW, Wolff DW, Sutton MD. 2012. *Escherichia coli* DNA polymerase IV (Pol IV), but not pol II, dynamically switches with a stalled Pol III* replicase. *J Bacteriol*. 194(14):3589–3600

13. Johnson A, O'Donnell M. 2005. Cellular DNA replicases: components and dynamics at the replication fork. *Annu. Rev. Biochem.* 74:283–315

14. Jónsson ZO, Hindges R, Hübscher U. 1998. Regulation of DNA replication and repair proteins through interaction with the front side of proliferating cell nuclear antigen. *EMBO J*. 17(8):2412–25

15. Kath JE, Jergic S, Heltzel JMH, Jacob DT, Dixon NE, et al. 2014. Polymerase exchange on single DNA molecules reveals processivity clamp control of translesion synthesis. *Proc Natl Acad Sci* 111(21):7647–52

16. Klemperer N, Zhang D, Skangalis M, O'Donnell M. 2000. Cross-utilization of the β sliding clamp by replicative polymerases of evolutionary divergent organisms. *J. Biol. Chem.* 275(34):26136–43

17. Li Z. 2012. *A Molecular Mechanism Allowing Transposon Tn7 to Target Active DNA Replication*. Cornell University

18. Li Z, Craig NL, Peters JE. 2013. Transposon Tn7. *Bacterial Integrative Mobile Genetic Elements*, p. 1

19. López de Saro FJ. 2009. Regulation of interactions with sliding clamps during DNA replication and repair. *Curr Genomics*. 10(3):206–15

20. Miller JH. 1992. *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia Coli and Related Bacteria*. CSHL Press. 478 pp.

21. Parks AR, Li Z, Shi Q, Owens RM, Jin MM, Peters JE. 2009. Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell*. 138(4):685–95

22. Parks AR, Peters JE. 2006. Transposon Tn7 is widespread in diverse bacteria and forms genomic islands. *J Bacteriol*. 189(5):2170–73

23. Parks AR, Peters JE. 2009. Tn7 elements: engendering diversity from chromosomes to episomes. *Plasmid*. 61(1):1–14

24. Peters JE. 2014. Tn7. *Microbiol Spectrum*. 2(5):

25. Peters JE, Craig NL. 2000. Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Molecular cell*. 6(3):573–82

26. Peters JE, Craig NL. 2001. Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev*. 15(6):737–47

27. Pham P, Rangarajan S, Woodgate R, Goodman MF. 2001. Roles of DNA polymerases V and II in SOS-induced error-prone and error-free repair in escherichia coli. *Proc. Natl. Acad. Sci.* 98(15):8350–54

28. Rangarajan S, Gudmundsson G, Qiu Z, Foster PL, Goodman MF. 1997. *Escherichia coli* DNA polymerase II catalyzes chromosomal and episomal DNA synthesis *in vivo*. *Proc Natl Acad Sci* 94(3):946–51

29. Stukenberg PT, Studwell-Vaughan PS, O'Donnell M. 1991. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* 266(17):11328–34

30. Sutton MD. 2004. The *Escherichia coli dnaN159* mutant displays altered DNA polymerase usage and chronic SOS induction. *J Bacteriol*. 186(20):6738–48

31. Sutton MD, Duzen JM. 2006. Specific amino acid residues in the β sliding clamp establish a DNA polymerase usage hierarchy in *Escherichia coli*. *DNA Repair*. 5(3):312–23

32. Sutton MD, Duzen JM, Maul RW. 2005. Mutant forms of the *Escherichia coli* β sliding clamp that distinguish between its roles in replication and DNA polymerase V-dependent translesion DNA synthesis. *Mol Microbiol.* 55(6):1751–66

33. Waddell CS, Craig NL. 1989. Tn7 transposition: recognition of the *attTn7* target sequence. *Proc Natl Acad Sci*. 86(11):3958–62

34. Wagner J, Etienne H, Fuchs RP, Cordonnier A, Burnouf D. 2009. Distinct β-clamp interactions govern the activities of the y family PolIV DNA polymerase. *Mol Microbiol*. 74(5):1143–51

35. Waters VL, Guiney DG. 1993. Processes at the nick region link conjugation, t-DNA transfer and rolling circle replication. *Mol. Microbiol.* 9(6):1123–30

36. Wijffels G, Johnson WM, Oakley AJ, Turner K, Epa VC, et al. 2011. Binding inhibitors of the bacterial sliding clamp by design. *J. Med. Chem.* 54(13):4831–38

37. Wolff E, Kim M, Hu K, Yang H, Miller JH. 2004. Polymerases leave fingerprints: analysis of the mutational spectrum in *Escherichia coli rpoB* to assess the role of polymerase IV in spontaneous mutation. *J Bacteriol*. 186(9):2900–2905

38. Wolkow CA, DeBoy RT, Craig NL. 1996. Conjugating plasmids are preferred targets for Tn7. *Genes Dev*. 10(17):2145–57

39. Yu D, Ellis HM, Lee E-C, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci*. 97(11):5978–83

CHAPTER 4

METHODS FOR STUDYING TNS PROTEINS IN VITRO AND MAPPING TRANSPOSITION INSERTION EVENTS VIA HIGH THROUGHPUT ANALYSIS*

4.1 Summary

Tn7 transposition uses two distinct mechanisms to control targeting and frequency of insertions into either a specific site within the genome or random sequences within the DNA. The ability of one protein, TnsE to mediate insertions preferentially into the lagging-strand during DNA replication is due to an interaction with components (the β-clamp and 3' recessed ends) that are enriched on this strand. Another protein has been found to change the nature of targeting to this strand, SeqA, a protein involved in replication initiation in *Escherichia coli*. SeqA binds to and organizes hemimethylated GATC sites throughout the chromosome, sequestering the origin of replication by oligomerization and forming a filament that tracks behind the replication fork. The TnsE mediated insertions map to different regions in the chromosome in a *seqA* knockout mutant as compared to the wild type strain. In order to understand how the different domains of SeqA affect transposition *in vivo*, determining the frequency and targeting of TnsE-mediated transposition in strains expressing SeqA mutants is imperative. The number of potential SeqA mutants that are of interest merits new ways to study how SeqA interacts with TnsE *in vitro* and how the SeqA mutants affect TnsE-mediated transposition *in vivo*. In this chapter I expand upon previously established methods to increase the number of mutants that can be studied in one experiment. SeqA has been shown to protect TnsE from trypsin digestion, and in order to easily determine whether multiple mutant proteins can interact with TnsE I set up an *in vitro* assay. *In*

* Figure 4.1 has been published in Li Z. 2012. A molecular mechanism allowing transposition Tn7 to target active DNA replication. Cornell University

vivo experiments involved establishing a method to map thousands of insertions to determine changes in the targeting profile and subtle attributes of transposition that cannot be resolved with the existing technologies. These methods will be useful moving forward to study the interaction of other proteins, mutants, and DNA structures that may be involved in TnsE-mediated transposition in a moderate to high-throughput manner.

4.2 Introduction

Transposon Tn7 is capable of moving into both a highly conserved specific site in the genome and seemingly random sites in conjugal plasmids. As discussed in previous chapters, Tn7-like elements use a core machinery encoded by three proteins, TnsA, TnsB, and TnsC, which interact with a target-site selecting protein, TnsD or TnsE, to carry out transposition. The TnsAB proteins are responsible for recognition and cleavage of the left and right ends of the element (48). Each end is comprised of multiple TnsB binding sites to which TnsB binds with varying affinities (2, 32). These binding sites are configured differently between the two ends; where the right end contains four overlapping sites, the left end contains three widely spaced sites (1). This configuration allows Tn7 to direct transposition into specific sites with left-to-right orientation specificity (5 16, 19, 27, 31, 41). The TnsD protein recognizes a specific DNA sequence contained within the C-terminal coding region of the *glmS* gene, targeting insertions into a specific site the Tn7 attachment site (*attTn7*) (33, 57). By directing insertions into *attTn7*, at a position not within an open reading frame, TnsD-mediated transposition does not negatively affect the host, enabling Tn7 to remain in this "safe haven".

Tn7 also has a mechanism to target mobile genetic elements (conjugal plasmids and bacteriophage) that are capable of transfer between cells, something that presumably would maximize horizontal transfer of the element between bacteria. The TnsABC+E pathway is

stimulated by and directs insertions into (>95% of all insertion events) conjugal plasmids (59). The ability to modulate between inserting into conjugal plasmids and into *attTn7* would presumably enable the Tn7 element to transfer horizontally and then into the chromosome of the new host. In support of this model, diverse bacterial hosts have been isolated from a broad range of environments that contain Tn7-like elements located in *attTn7* (16, 38).

Previous studies have demonstrated that TnsE interacts with components that are enriched on the lagging-strand during DNA replication, the β-clamp and 3' recessed ends (reviewed in (15)), to mediate transposition preferentially into this strand (37, 41). Evidence for these interactions comes from work with conjugal plasmids, which increase TnsE-mediated insertions 100 fold (59). Additionally, the organization of the left and right ends of the element within the target DNA can be used to identify insertion orientation in these sites. In a conjugal plasmid insertions occur in one orientation, which is thought to be related to discontinuous plasmid replication in recipient cells (41, 42, 59). At a low frequency TnsE mediates transposition into the chromosome, as has been found in the laboratory in strains lacking a conjugal plasmid (41).

Identification of the TnsE-mediated transposition insertions around the chromosome in wild type cells show two regions with a greater number of insertions (Figure 4.1, (25)). One region where a greater number (a "hostpot") of insertions is observed (~40%) occurs where chromosomal DNA replication terminates [Figure 4.1, (25, 40)]. These insertions may be attributed to stalled replication forks where a vacant β-clamp and single stranded DNA gap is available for interaction with TnsE. Another region where a moderate number (more like a "warm spot") of insertions is observed (~15%) is within a region around the origin [Figure 4.1, (25)]. This region comprises an approximately 250kb region around the origin. This relatively

recent finding poses the question: what causes this "warm spot" of insertions in the region around the origin? The cause of these has not yet been elucidated. The remaining insertions are observed to occur randomly around the chromosome, where there is a bias for insertion orientation that corresponds with replication (40). That is, as the replication forks proceed bi-directionally around the chromosome, each half of the chromosome comprises a replichore. TnsE-mediated insertions are found in one orientation in each replichore (40). These insertions are thought to be in the lagging strand during replication, attributed to the discontinuous nature of replication of this strand where TnsE can interact with empty β-clamps and 3' recessed ends that are more abundant on this strand (37, 41). A low frequency (~20%) of the TnsE-mediated insertions that occur randomly around the chromosome are in the opposite orientation (the leading strand template during replication) (36, 40). These very low number of insertions can be attributed to stalled replication forks which are naturally thought to occur during leading strand replication. When replication is restarted downstream of these stalled replication forks (21, 30, 43, 44, 46, 60), a gapped single stranded DNA and the stalled β-clamp remain, which are both associated with TnsE-mediated transposition. Therefore these events could represent a unifying molecular target that could be found on the leading strand, such that insertion events are found in the opposite direction than the majority of the insertions not in the terminus or near the origin of replication.

As discussed in chapter 1, the lagging strand may be inherently more vulnerable during replication. Transposable elements, such as Tn7, appear to capitalize on this vulnerability by targeting components of this strand during replication. Other transposable elements, such as some IS elements (short DNA sequences that act as simple transposable elements, encoding their own transposase and lacking accessory genes) have been shown to capitalize on cell activities,

such as replication, to control transposition (45, 61). One cellular mechanism the transposon "eavesdrops" on cell activity is based on the methylation state of the cell. Immediately after DNA synthesis, the nascent DNA strand is un-methylated, forming a hemi-methylated duplex strand. Dam methylase is the enzyme which catalyzes this reaction, adding a methyl group to the N-6 position of the adenine in the sequence 5'-GATC-3' in newly replicated DNA. Dam, with some lag (~1/3 cell replication cycle), follows the replication fork during DNA replication remethylating the DNA (29). Tn5 (61) and Tn10 (45) transposition are affected by the hemimethylation state of the GATC sites in the transposase promoter and the transposase binding sequence. In a *dam*- strain, the frequency of Tn5 and Tn10 transposition increases (45, 61); similarly, TnsE-mediated Tn7 transposition is stimulated in a *dam*- strain [(14, 36), and here]. However, TnsE-mediated transposition was unaffected by mutating the GATC sites within Tn7 (14). This would suggest that Dam has another role in transposition, as the frequency of TnsE-mediated transposition increases in a *dam*- strain. However, targeting to the lagging or leading strand in this strain has not yet been identified.

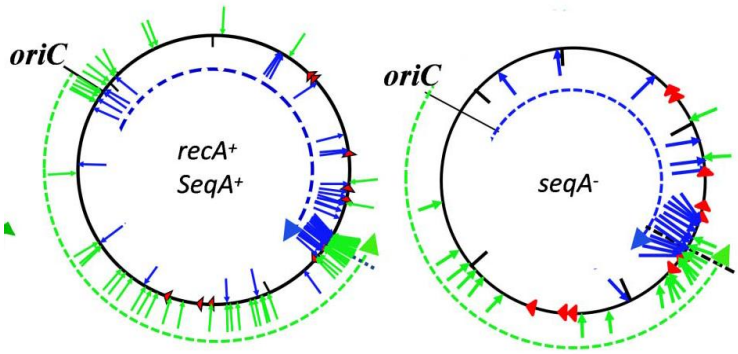

Figure 4.1

Location and orientation of independent TnsE-mediated transposition events in the chromosome of *wild type (wt), and seqA- backgrounds.* Positioning of the arrow inside vs. outside indicates opposite left-to-right orientation. Dashed line indicates the direction of DNA replication forks. All transposition events were isolated from lambda hop assays with transposition proteins expressed from pCW15 for TnsABC and pJP104 for TnsE. Figure from (25)

Given that TnsE-mediated transposition is higher in a *dam*- strain than wild-type, and this is not related to the methylation of GATC sites of the element, how else could Dam be related to TnsE-mediated transposition? By looking at the roles of Dam in normal cell activity, we may be able to identify potential roles for Dam in TnsE mediated transposition. Unpublished previous work from multiple members of the lab hinted at a role of Dam in TnsE-mediated transposition that may be highly indirect. That is, the Dam methylation of DNA may lead to differences the binding of another protein to the DNA, the SeqA protein, a protein involved in replication initiation control. Follow up studies showed an interaction of TnsE with SeqA: TnsE-His6 adhered to an affinity column was able to bind SeqA from a cell lysate (25). This in addition to protease-footprinting of TnsE and SeqA indicated a physical interaction between SeqA and TnsE, where SeqA protects one or more trypsin sensitive sites within TnsE ((25), Guarne personal communication).

To further explore this relationship between Dam, SeqA, and TnsE, we must first examine the nature of the interactions between Dam and SeqA. DNA methylation is one mechanism used to synchronize replication initiation (7). DNA replication (reviewed in (62)) initiates at the chromosomal origin, *oriC* in bacteria. The *oriC* of *E. coli* contains multiple sequence-specific regions (DnaA boxes) to which the DnaA protein binds forming a filament on the DNA. DnaA binding unwinds the AT-rich regions of the origin and forms an open bubble on which the DnaB replicative helicase is loaded by interactions between DnaA and a third protein, DnaC. The enzymes required for DNA replication [helicase (DnaB), primase (DnaG), and DNA Pol III (Pol III Holoenzyme)] are loaded onto this open complex and synthesis of a new DNA strands begins.

Immediately after replication of the origin, the new strand is unmethylated forming a hemimethylated duplex strand. SeqA binds tightly to this hemimethylated DNA, sequestering some of the low affinity DnaA boxes and preventing Dam remethylation for about a third of a cell cycle. Dam re-methylates the nascent strand at the origin and around the chromosome, which then forms a fully methylated duplex strand. SeqA binds weakly to the fully methylated GATC sites, and other proteins such as DnaA are able to bind to the low affinity DnaA boxes in the origin, enabling replication to restart. SeqA was identified as the protein responsible for negative regulation of replication initiation by sequestering GmATC sites at the origin (29). *oriC* remains hemi-methylated for one third of the cell cycle (roughly 8-10 minutes) (6, 10, 34) , whereas chromosomal GATC sites stay hemimethylated for about two minutes (10, 29). Binding of SeqA to the transient sites at the origin prevents re-initiation of the chromosome too quickly, as it prevents DnaA binding and slows the rate of Dam methylation (10). SeqA additionally binds GmATC sites found around the chromosome after passage of the replication fork (47). The binding of sites around the chromosome has been related to active DNA replication, in that fluorescently tagged SeqA molecules were shown to form foci that migrates with the replication fork (9, 23, 24, 35).

The direct interaction of Dam with the DNA is impeded by direct physical hindrance as a result of SeqA binding and indirect hindrance from changes in the DNA structure upon SeqA binding (20). In both the *seqA* and *dam* knockouts, a large stimulation of transposition is observed compared to wild-type (25). Given that mutations of the actual GATC sites (such that these sites would not be affected by the methylation state of the cell) in the Tn7 element does not affect TnsE-mediated transposition (14), it points to another role for either Dam or SeqA in transposition.

When transposition events in a *seqA*- strain are mapped, the pattern of insertions differ compared to those observed in a wildtype strain. In the *seqA*- strain the pattern that is revealed is a hotspot (~50% of insertions) near the terminus, and a paucity (<2.5%) of insertions within approximately 250 kb around the origin [Figure 4.1, (25)]. Compared to a wild-type strain, where ~40% of the insertions occur into the terminus and 15% occur into the region around the origin of replication (25), this would suggest differential targeting by TnsE of the region around the origin versus the rest of the chromosome.

If TnsE is able to interact with SeqA, and there are differences in how TnsE targets transposition in a *seqA* + versus a *seqA*- strain, it suggests potential differences in how SeqA is interacting with the different regions of the chromosome. In fact, there are significant differences in the DNA-protein structure formed as SeqA interacts with the GmATC sites at different points in the chromosome. SeqA interaction at the GmATC sites in the origin initially act as a dimer, binding the unevenly distributed sites and likely causing intervening DNA to loop out (20, 50). Subsequent oligomerization of SeqA dimers may form a left-handed helical filament which releases some of the GmATC sites, relying on the overall structure to organize the DNA, in which the DNA is wrapped around the protein filament (20). This filament is thought to track along behind the replication fork, with SeqA monomers binding GmATC at the head of the filament, and being released at the origin-proximal side, like a conveyor belt (9, 23, 24, 35). If TnsE recognizes some site on the SeqA protein that enables TnsE to remove SeqA and access the lagging strand, the differences between the SeqA structures formed at the origin versus those tracking behind the replication fork may be differently recognized by TnsE, such that there are differences in where TnsE can target insertions in the chromosome. In the absence

of SeqA, there may be significant changes to the newly replicated DNA around *oriC* that make interactions of TnsE within this large region not possible.

SeqA is comprised of two functional domains separated by a linker: the N-terminal dimerization domain and a C-terminal DNA binding domain. Various SeqA mutants have been shown to have different activities on the DNA, depending on the location of the mutation; where a deletion of the linker domain limits SeqA mutant binding to widely spaced GmATC sites, a deletion of the two N-terminal residues prevents DNA binding altogether (11). *In vitro* analysis of TnsE interactions with a subset of the *seqA* mutants will enable us to determine if either the filamented or oligomeric form of SeqA is recognized, and therefore whether these structures are important in TnsE-mediated targeting to a region around *oriC* or other sites in the chromosome. *In vivo* analysis of TnsE-mediated transposition events in strains containing SeqA mutants will further inform the targeting of TnsE.

Thus far, interactions between SeqA or Dam and TnsE have been studied using preexisting tools described above. These include tools to determine transposition frequency, transposition mapping, and purified protein interactions. Unfortunately, these tools are not sufficient for identifying the extent of SeqA mutants I wanted to observe. *In vivo* studies, such as transposition assays, are difficult to assess the effect of SeqA, as cells appear to be very attuned to SeqA levels, where both a *seqA* null and *seqA* over-expressing strains have been shown to affect cell replication (4, 8, 49). While the effect of asynchronous replication on TnsE-mediated transposition has been studied and shown to not be the causal agent for the transposition increase (25), anecdotally *seqA* knockouts quickly pick up suppressor mutants, which makes transposition in these strains difficult to study. The second technique used to study the effect of SeqA on transposition is where transposition events are mapped, a process which

requires a multi-step process resulting in a small set (~50) of insertion sites mapped. The third assay previously used, protein-protein interactions, are very useful in identifying SeqA-TnsE, and how potential mutants affect this interaction, however this requires purification of the protein. If a large number of mutants are of interest, both transposition mapping and purification of each mutants is time consuming and expensive, so a faster, more high-throughput method is required.

Therefore, to best determine the effect of *seqA* on TnsE-mediated transposition, I aimed to develop tools that would enable us to better study *seqA* mutants. One "tool" to study the effect of various *seqA* mutants on TnsE-mediated transposition included moving mutant alleles into the chromosomal *seqA* locus, so they are expressed from the native promoter. As a second mechanism to control the relative *seqA* expression levels, I cloned the *seqA* mutants to a rhamnose expression vector. Both moving alleles into the chromosome and expressing the proteins *in trans* can be used to control the amount of SeqA in the cell without passing through a *seqA* knockout strain. I hope these techniques will limit the number of suppressor mutations that develop in the absence of SeqA. Using the strain with the rhamnose expression vector, I could both express *seqA* mutants and use the cell lysate of these mutant expressing strains in a far-western blot assay to show positive interaction of SeqA and TnsE. Thirdly, given the number of mutants I wanted to study, I develop a new method to determine the insertion-specificity which may be used in a high-throughput manner from our transposition assays. I first confirmed this technique using classical sequencing methodologies, and here suggest a workflow resulting in Illumina sequencing. In the high-throughput sequencing assay, I use a *dam*- strain in transposition assays to determine whether the effect of a *dam*- knockout on the transposition

insertion pattern is similar to that of either *seqA* or wildtype, which may better inform our understanding of how these proteins interact with the chromosome.

4.3 Results

The frequency of TnsE-mediated transposition increased ~10 fold in a *seqA* null strain as compared to wild-type (25), therefore one possibility is that SeqA protects DNA from TnsE-mediated insertions. When these insertion events are mapped, no insertions occurred into a region around *oriC* in the *seqA* null strain, leading to the question: does TnsE recognize some difference in SeqA complexes formed around *oriC* versus those which form after passage of the replication fork?

SeqA is capable of forming different multimeric structures both complexed on DNA and in solution. On DNA, depending on the SeqA binding site (GmATC) distance, SeqA is capable of forming different multimeric structures. SeqA binding closely spaced GmATC sites at the *oriC* forms oligomers, causing excess DNA to loop out. On the other hand, SeqA binding GmATC sites around the chromosome forms a filament which tracks along behind the replication fork (9, 23, 24, 35). Both oligomers and filament structures are sensitive to disruption by *seqA* mutations. The SeqA N-terminus (amino acids 1-59) and C-terminus (amino acids 71-181) form distinct domains, where the amino acids between them (residues 60-70) comprise a linker domain. The N-terminal domain is required for multimerization, the C-terminal domain is required for binding to GmATC sites (18), and the linker domain is required for binding GmATC sites separated by more than two full turns of the DNA (20). Mutations of specific residues in these domains disrupts activity in a predictable way: mutations in the N-terminal domain can disrupt dimerization and filamentation activities and mutations of the C-terminal domain disrupt DNA binding ability. While mutations of specific residues in the linker

domain have no observed effect on DNA interactions, deletions of the linker domain limit binding of widely spaced GmATC sites (such as those found following the replication fork).

By studying the direct interaction of TnsE and a subset of SeqA mutants that disrupt different activities of SeqA, I may able to determine whether TnsE is recognizing a specific domain of SeqA, a complex, or a DNA structure formed by SeqA binding. Given the number of SeqA mutants that are of interest, and the possibility of suppressor mutations arising in a *seqA* knockout, the first course of action is to develop tools that will enable us to study the effect of SeqA mutants on TnsE-mediated transposition in a high-throughput fashion.

4.3.1 Physical interaction of TnsE and SeqA

Direct protein-protein interaction of TnsE and wild-type SeqA was observed in a trypsin assay, where SeqA was able to protect TnsE from trypsin digestion ((25) and results repeated by the Guarne lab, personal communication). On the other hand, SeqA linker domain mutants (both deletion mutants and mutants of multiple residues) were unable to protect TnsE. These results suggest that the linker domain of SeqA is responsible for the protection of TnsE, however, whether this domain directly interacts with TnsE has not yet been studied.

As noted above, SeqA forms multiple types of structures on the DNA depending on how far apart the GmATC sites are from each other. The length of the linker domain limits the ability of SeqA to bind GATC sites that are widely spaced. To this regard, the linker mutants $SeqA^{\Delta 2}$(Δ45-59) and $SeqA^{\Delta 4}$(Δ41-59) will restrict DNA binding to GATC sites that are very close together, such as those found at the origin, whereas the linker mutant $SeqA^{M2}$ (E48G/V49D/V51A/A52S/I56S/V57D/E58K/A59G) will allow for binding widely spaced GmATC sites, but has been shown to be unable to protect TnsE from trypsin digestion (personal communication, Alba Guarne). Mutations in other domains prevent formation of different types

of structures (dimers, oligomers, and filaments). For example, a DNA binding mutant: SeqA$^{N150A/N152A}$ prevents the formation of DNA-SeqA superstructures, however dimers and oligomers will still form in the cytosol. Two other mutants will be important in identifying domains important for TnsE-mediated transposition: the oligomerization mutan, SeqA^{A25R}, does not form the oligomer-looped DNA superstructure at *oriC*, and the filamentation mutant, SeqA$^{R70S/R73S}$, which does not form the filaments that track along behind the replication fork.

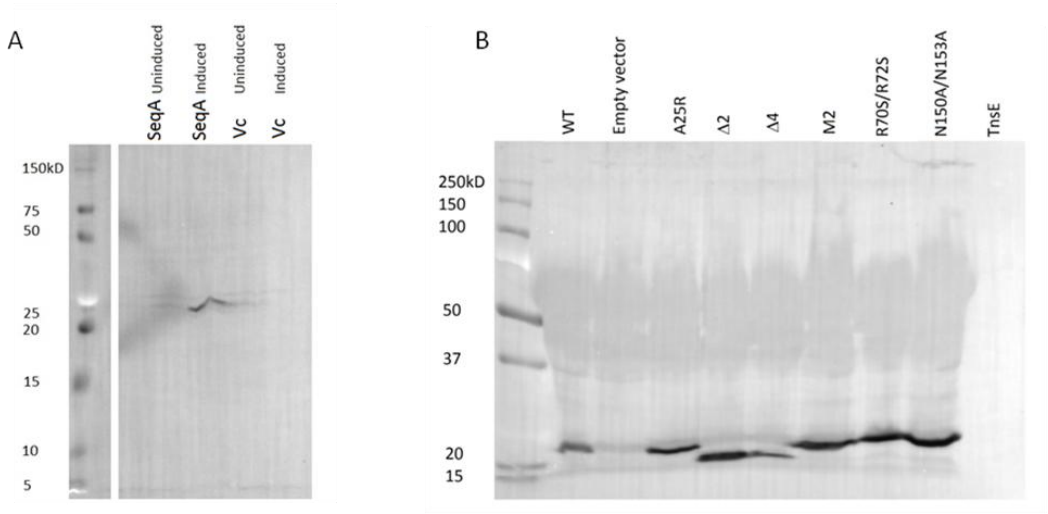

Figure 4.2

Expression of SeqA from a rhamnose vector
SeqA is expressed after 6h induction with 0.2% rhamnose from a BW27784 protein expression strain. Proteins run on a 15% poly-acrylamide gel, transferred to PVDF membrane, and incubated with mouse-anti-flag antibody (1:2000); secondary goat anti-mouse (1:4000). A) WT SeqA expressed after 6h, but not from the empty vector or no inducer. B) SeqA mutants expressed after 6 hours.

Each of these mutant alleles and the wild-type gene were cloned into a rhamnose expression vector with an N-terminal Flag tag (see methods). The proteins were expressed for 6 hours with rhamnose and lysed by sonication. Expression of the proteins was confirmed by running cell lysate on an SDS-PAGE gel and transferred to a PVDF membrane for western blotting. An anti-flag primary antibody was used to detect SeqA expression. Expression of all

proteins could be confirmed with the band found at the appropriate size relative to the marker lane (Figure 4.2). No signal was found in the empty vector lane an no cross-reactivity was found with TnsE. Two of the SeqA mutants, $SeqA^{\Delta 2}$ and $SeqA^{\Delta 4}$, are deleted for 15 and 18 amino acids, respectively, and are expected to give a shorter product. More careful experiments using quantitative Western blots will be needed to determine if there are changes in the stability of the mutants relative to the wild type protein. Careful experiments will need to be conducted to determine if there are differences in stability or expression.

After showing that I could express SeqA utilizing the rhamnose expression vectors, I wanted to determine if the tagged SeqA would interact with TnsE. Since I believe TnsE to interact with SeqA, I used TnsE to show that the expressed tagged SeqA protein is still capable of interacting. To determine whether I could visualize SeqA interaction with TnsE, I did a Slot blot analysis where an increasing concentration of His-tagged TnsE or BSA (acting as a negative control) was blotted onto a PVDF membrane. The blot was then incubated with a lysate prepared from wild-type SeqA over-expressing cells (discussed above), and an anti-flag primary antibody was used to detect SeqA binding. Band intensities on the slot blot correspond levels of SeqA binding: a higher concentration of TnsE gives a stronger band intensities (Figure 4.3). A negative control where slot blot analysis of the highest concentrations of TnsE incubated with a lysate prepared from cells containing the empty expression vector showed no positive signal (Appendix 1). The positive interaction of SeqA and TnsE complements the physical interaction inferred by trypsin digest of purified SeqA and TnsE (discussed above, (25)). I show that I can both express SeqA and obtain functional protein which interacts with TnsE.

In order to determine which domain(s) of SeqA interacts with TnsE, I used vectors containing mutant *seqA* genes and generated additional mutants through site-directed

mutagenesis. These genes were subcloned into a rhamnose expression vector, and cell lysates containing over-expressed mutant SeqA proteins were produced. These SeqA mutations have all variously been shown to disrupt binding DNA or assembling into larger structures. I used the same slot-blot assay done for wild-type SeqA, where stronger band intensity indicates binding of SeqA to the available TnsE on the membrane (Appendix 1). Given that more experiments will need to be done to confirm these interactions and these experiments merit repetition, in this chapter the experiments are shown to indicate the capability for us to test multiple mutants simultaneously rather than a bona fide result.

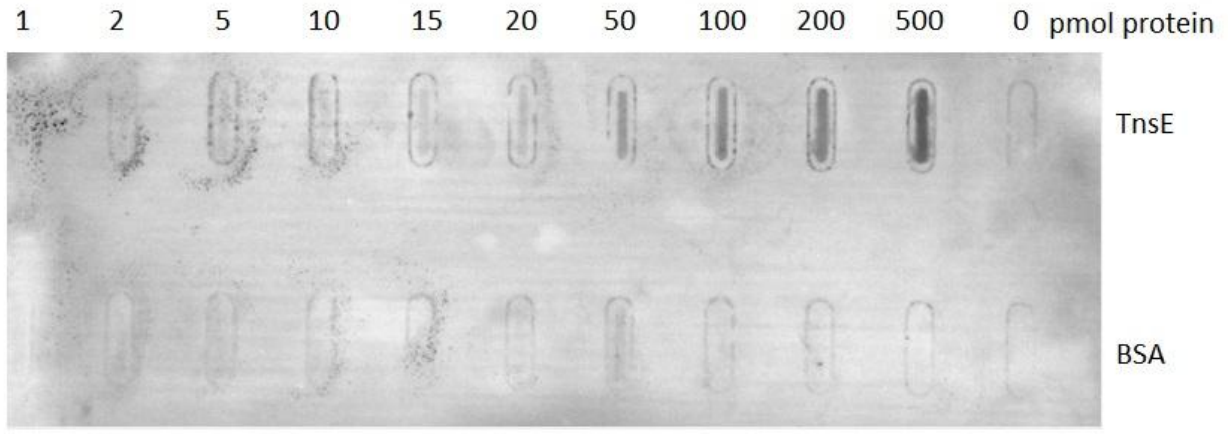

Figure 4.3

Slot Blot of TnsE and BSA overlayed with SeqA
SeqA is expressed after 6h induction with 0.2% rhamnose from a BW27784 protein expression strain. His-tagged TnsE was purified from a BL21 fur-/slyD- strain using Nickel column, and ion exchange. BSA was purchased from Fisher Scientific. TnsE and BSA were blotted onto a PVDF membrane, and incubated overnight with lysate from a N-terminal flag tagged SeqA, and subsequently with mouse-anti-flag antibody (1:2000); secondary goat anti-mouse (1:4000).

There are important differences between the slot blot assay and previously established assays such as the trypsin digest assay, such as the protocol: where this assay uses whole cell lysate and trypsin footprinting assay uses purified protein. Because the slot blot assay does not rely on purifying proteins, it can be used for testing interaction of multiple SeqA mutants (or

other expressed proteins) with TnsE within the same experiment. Other assays that are typically used to show interaction of proteins, such as immune precipitation, trypsin digest assay, and surface Plasmon resonance amongst others are time-consuming, cost-preventative, and require purification of each protein and would make them poor assays to screen for SeqA mutants with greater affinity to the protein of interest, in this case, TnsE.

4.3.2 Introducing SeqA mutants into the chromosome

The TnsE-mediated transposition pattern is different between WT and *seqA* null strains (Figure 4.1, (25)). This pattern may be related to the ability of TnsE to interact with complexes formed by SeqA binding to GmATC sites. These complexes are disrupted by mutations in different domains of SeqA. By studying the transposition targeting in strains expressing only *seqA* mutants, I may be able to further address how TnsE interacts with SeqA *in vivo*. Unfortunately, over-expression of SeqA has a negative effect on cell growth (4). If over-expression of the SeqA mutants has a similarly adverse effect on cell replication, I cannot use this to accurately determine TnsE-mediated transposition. Furthermore, the possibility of obtaining suppressor mutants may affect both transposition frequency and where these events are located. Therefore, I introduced the *seqA* mutants into the chromosome, so they could be transcribed by the *seqA* native promoter (at "wild-type" levels). Generating a strain which has the *seqA* mutants in the chromosome was done by engineering a cassette containing kanamycin resistance gene flanked by DNA containing an I-SceI cut site and sequence homologous to DNA flanking *seqA* that could then be recombined via lambda red recombination with the chromosomal *seqA* locus led to a Kan-I-SceI cut site intermediate in the chromosomal *seqA* locus. To move the final *seqA* mutants into the chromosome, a similar cassette was made containing individual mutants by PCR and moved into the chromosome. Briefly, the

intermediate strain was transformed with a DNA fragment containing a *seqA* mutant flanked by sequence homologous to the DNA flanking the chromosomal *seqA* and a plasmid with an inducible I-SceI endonuclease. Induction of the I-SceI endonuclease led to double strand breaks which recombined with the linear double stranded *seqA* mutant fragment. The resulting strain was confirmed to contain the mutant *seqA* by amplification of the locus by PCR and submission for sequencing to confirm the presence of the mutant at the locus. This procedure appeared to be very efficient, 40-100% of the colonies amplified had the correct size amplification product, out of which only two were sequenced. In this section, I show that this method can be used for high efficiency cross-in of *seqA* mutants.

4.3.3 Mapping Insertions using Next Generation Sequencing: Methods Development

Experiments used to map TnsE-mediated transposition events in all of the strains requires first obtaining transposition events, done by introducing a miniTn7 element containing a kanamycin resistance gene via a replication and integration defective lambda phage (lambda hop (methods)). Only when transposition machinery is provided (TnsABC+ TnsD or TnsE) will the miniTn7 element be integrated into the chromosome, allowing for growth on media containing kanamycin. To determine insertion site, each colony (corresponding to a unique transposition event) is isolated and the insertion site mapped by two rounds of PCR, first, arbitrary PCR and then, nested PCR (methods). In order to compare transposition frequency of the *seqA* mutants to previous studies of WT and *seqA* null strains, at least 50 insertions must typically be mapped. Given the number of *seqA* mutants, mapping transposition in these strains by the classical method seemed daunting. Novel technologies where multiple transposition insertion sites could be mapped at once (multiplexing) seemed like an obvious choice. I opted to sequence using Illumina 150bp single-end sequencing for our final run. In order to confirm Illumina sequencing

accuracy in mapping transposition events, I first mapped transposition events with Sanger sequencing (Figure 4.4).

As noted above, early genetic screens of mutants that affected TnsE-mediated transposition indicated that a *dam* null strain is specifically stimulated for Tn7 transposition, a result that was confirmed in this lab ((36, 25) Figure 4.4). Given that the transposition frequency

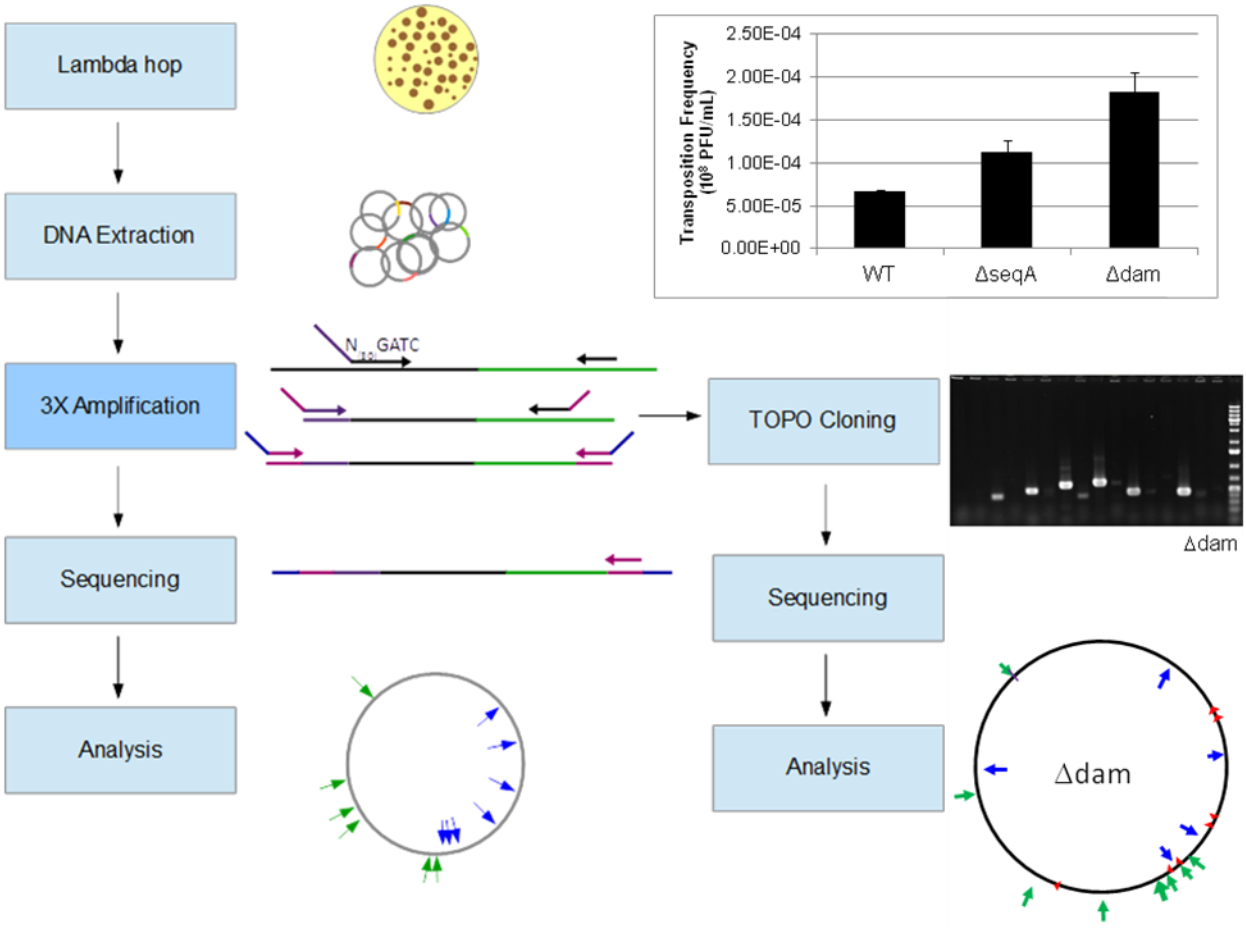

Figure 4.4

Flowchart of mapping TnsABC+E transposition insertion events
Lambda hops were done as previously shown (Peters, 2001?). After 1 day incubation, colonies were collected and DNA was extracted. Confirmation of mapping events in a Ddam strain. Two rounds of PCR were done, the first with arbitrary PCR primers, and the second with nested primers containing adapters for Illumina sequencing (not shown). The resultant PCR mix was cloned into a TOPO TA cloning vector, and amplified using primers complementary to the plasmid sequence (shown).

in this mutant has been well established, and that transposition events have never been mapped in this background, it was an obvious option for the development of identifying Tn7 insertion sites via high throughput sequencing.

In order to generate a library of transposition events, colonies from an entire plate of TnsE-mediated transposition events, comprising approximately 500-1000 colonies, in a *dam*$^-$ strain was collected. The DNA was extracted, and two rounds of PCR were done as previously described (40). A third round of amplification was done using primers that contained adapter sequences to allow the DNA fragments to bind to the Illumina flow cell. At this point, I wanted to ensure that the transposition, collection, and PCR assays would yield transposition events and not random amplification events. This 'control' experiment was used as a confirmation before transitioning to Illumina sequencing. In order to do this, I cloned and sequenced a subset of the fragments in the insertion library. Briefly, the insertion library was cloned into the pCR TOPO 2.1 vector by TA cloning. Thirty clones revealing different size inserts were submitted for sequencing where each clone represented a single insertion event. The result from this mapping assay indicate that approximately half of the sequences were mis-priming events by the primers, an effect which can likely be resolved using hot-start PCR. It is hard to reach too many conclusions with this size of a data set. However, the results seem consistent with the findings with the *seqA*+ and *recA*+ background. Out of the 14 mapped events in the *dam*$^-$ background, 1 insertion into the region around oriC was found, and 7 into the region around the termination sites. The remaining insertions had a bias to one strand in each replichore, which is believed to be related to lagging-strand replication. While I only mapped a few insertions, I can use this data to compare to previously mapped insertion events in *seqA* and wild type backgrounds. This pattern of insertions appears to more closely match insertions in a wild type strain rather than the

seqA null strain, where I observed 1/14 (or 7%) of the insertions are within a region around *oriC*. This limited result suggests that there may be secondary effects of knocking out *seqA* or *dam* that are unrelated to each other which lead to differences in the DNA available for TnsE-mediated transposition. More importantly, these results showed us that insertion events could be amplified with the Illumina primers, resulting in approximately 50% yield of mapped insertion events.

This same method was used to send samples for Illumina sequencing, with two important exceptions. The first, hot start PCR was done to reduce mispriming events, and the second, a separate set of primers was used that incorporated Illumina barcodes (Figure 4.4). Unlike the method described above where the entire pCR TOPO 2.1 insert was sequenced, Illumina sequencing requires DNA fragments ranging from 300-600bp. The correct size smear from an agarose gel was cleaned, and the size was confirmed by Fragment analyzer to be within the size limits for Illumina sequencing. Our sample was added at 1% DNA concentration to the PhiX control lane of the Illumina sequencer. The resulting sequences were analyzed through the BioHPC computing facility where the primers which amplified the insertion event were used to determine which sequences were of interest to us. Analysis of Illumina results revealed no returned transposon sequence, suggesting an error in methodology and the need for addtional troubleshooting.

Moving forward, methods could be adjusted by generating a library with the correct primer set, and submitting it to the same TA cloning as was done for method development prior to Illumina sample submission. Two to three colonies carrying different size inserts and representing independent insertion events should be sequenced to confirm correct addition of the required components (Illumina barcode, adapter, and transposon sequence) for Illumina sequencing and analysis.

4.4 Discussion

Transposon Tn7 has great control over where and when it directs transposition, utilizing different types of target sites. Our current understanding of Tn7 has been facilitated by two *in vivo* transposition assays, one which can be used as a screen for mutations that increase or decrease transposition frequency, and the other which is used to assess the effect of mutants on the frequency and targeting of transposition. These assays have been supplemented by *in vitro* reactions, which have been used to establish how the Tn7 proteins interact with each other and host proteins (reviewed in (13, 39, 26)). While the combination of *in vitro* and *in vivo* tools have been useful in determining the basics of Tn7 transposition, the adaptability of the element requires new tools that can allow for a greater understanding of the molecular role of transposon proteins. In this chapter I make three enhancements to our current techniques that may contribute to a better understanding of Tn7 transposition.

4.4.1 In vitro identification of protein domains that interact with TnsE

I extended the previous result that TnsE physically interacts with SeqA ((25), Alba Guarne personal communication) to show that I can express a functional SeqA with an N-terminal eight amino acid tag. This expression vector was used to express the SeqA mutants with no leaky production of the proteins (data not shown). The expressed SeqA was shown to interact with TnsE using a far-western blot which was subsequently used to determine interaction of SeqA mutants with TnsE. Previous *in vitro* interaction studies used to determine the interaction of the proteins encoded by Tn7 have used purified components.

Many expressed proteins require a large protein tag to aid in purification or visualization *in vivo*, which may affect the protein viability. This has been shown to be the case for C-terminally tagged SeqA with Green Fluorescent Protein variants (17), where the expressed SeqA

was unable to initiate replication simultaneously at all origins present within a single cell. The inability to synchronize replication is likely due to disrupted protein structure, and will therefore not be representative of the wild-type protein in *in vitro* assays. By using a small tag specifically on the N-terminus, I believe that the expressed SeqA protein will be functional, but this will only be determined with further experimentation.

The procedure of purification often requires a multi-step process (12), which requires the use of expensive columns and is time-consuming. The assay used here to determine SeqA-TnsE interaction, the slot-blot, is a high-throughput way to identify which host proteins, and what domains of those proteins interact with TnsE. In this assay, the TnsE is blotted onto a membrane, and incubated with a crude cell lysate containing the expressed proteins, then the blots are pooled and developed. This enabled us to distinguish how 8 different SeqA mutants interacted with TnsE in one experiment, but can easily be scaled up to 24 mutants with lower resolution (plus/minus). Previous work has established that TnsE interacts with the β-clamp (37), however, the residues of this clamp that are necessary for TnsE binding have not been established. In view of the TnsE-β-clamp interaction, it will be interesting to use this same procedure to test various β-clamp mutants in interacting with TnsE. Another interesting analysis will be to determine whether SeqA from *E. coli* can interact with TnsE from different hosts. This experiment derives from work showing that TnsE homologs from *I. loihiensis* and *S. baltica* with TnsABC from *E. coli* can promote transposition of a miniTn7 element only when the clamp from the same host is present (25). Given that the linker domain, which is poorly conserved between SeqA proteins from different bacteria, has been shown to be involved in protecting TnsE from trypsin digestion (Alba Guarne personal communication), it is possible that this domain is highly specific in protecting TnsE. That is, the linker domain of SeqA from *I.*

loihinensis or *S. baltica* may not be able to interact with TnsE from *E. coli*. The expression and interaction experiments discussed here may be able to easily and definitively determine if a cross-species interaction of SeqA and TnsE exists.

Looking forward, one can also use this expression system in *in vivo* assays. In addition to the role SeqA plays in sequestration of the origin immediately after replication, it has roles in organizing the nascent duplex DNA following the replication fork (9, 24, 58) and regulation of gene expression (28, 52, 53, 54). Given the diverse roles of SeqA in cellular function, transposition experiments in a *seqA* knockout strain are difficult to attribute to the direct interaction between SeqA and TnsE. Furthermore, because of the many functions that are disrupted in a *seqA* knockout, these strains anecdotally quickly pick up suppressor mutations. By being able to regulate *seqA* expression, one can use the vectors established in this chapter to express SeqA *in trans* while it is deleted it from the chromosome. Subsequently, one could do the final transposition assay in these strains when SeqA is not induced, minimizing suppressor mutants and the effect the absence of SeqA has on cell activities.

4.4.2 Expression of mutant proteins from native promoter

Early experiments identified SeqA played a role in regulating replication initiation. DNA replication begins by the binding of the initiator protein DnaA to specific sites within *oriC* which leads to melting of the DNA duplex. Upon replication, the hemimethylated GATC sites in the origin are bound by SeqA which then comes together to form oligomers, thereby sequestering the DNA. SeqA binding to DNA prevents DnaA binding at some of its low affinity DnaA boxes, which therefore limits replication initiation until remethylation of the DNA by Dam (51, 63, 55). Dam methylase tracks behind the replication forks, methylating the nascent strand. SeqA binds hemimethylated GATC sites much more strongly than fully methylated DNA (50). Therefore,

after methylation of the DNA, SeqA releases, allowing DnaA to bind, and a new cell cycle to begin. By balancing these positive (DnaA, Dam) and negative (SeqA) effectors, the cell is capable of firing all origins at the same time, and ensuring there are a set number of chromosomes in each daughter cell. Deletion of the *seqA* gene removes the negative effector, leading to premature initiation and asynchronous replication (8, 51). On the other hand, excess SeqA extends the period of hemimethylation of GATC sites at the origin which likely leads to the delay in replication initiation (4). The downstream effects are a delay in nucleoid segregation and cell division in these SeqA over-expressing cells (4, 56).

SeqA mutants have been identified that disrupt the actions of SeqA on the DNA, including dimerization mutants, oligomerization mutants, and DNA binding mutants (20). The effect these mutants have on cell activities remains to be determined. It follows that the effect these mutants have on TnsE-mediated transposition has also not yet been elucidated. However, classical systems using a *seqA* knockout in combination with a *seqA* over-expression *in trans* is concerning given the evidence above that the cells appear highly attuned to cellular SeqA levels. Because we do not know how over-expression of these SeqA mutants will affect replication initiation and cell division, expression of the mutants from the native promoter seemed pertinent. I was able to obtain cells which contained the mutant *seqA* in the native locus.

Moving forward, one can move the *seqA* mutants in the *seqA* locus into our transposition assay strain. From there, one can determine how each of the mutants affects transposition frequency, and where the insertions are targeted. For example, one may expect that in a *seqA* mutant that lacks DNA binding ability, where none of the GATC sites around the chromosome would be bound, the effect would be similar to that of a *seqA* null strain. That is, the

transposition frequency would increase and where the insertions are found in the chromosome would look the same as a *seqA* null strain.

The experiments that can result from introducing the mutant gene into the chromosome extend beyond *seqA*. Other proteins that interact with Tns proteins may have use in this type of assay. One such example is the β-clamp, which has been shown to interact with TnsE (37). β-clamp mutant proteins have been identified that support cell viability, but have only been expressed from a plasmid (3). While there is no evidence that over-expression of the β-clamp affects cell growth, it is possible that excess β-clamps in the cytosol may affect TnsE-mediated transposition. Using the same techniques, future work can involve moving the β-clamp mutants into the chromosomal locus, which will enable us to determine which domains of the clamp are important in TnsE-mediated insertions *in vivo*, complementing *in vitro* experiments proposed above.

4.4.3 High-throughput mapping of transposon insertion sites

There have been many ways in which transposons have been mapped to specific sites within the genome. TnsE-mediated transposition events have been mapped by amplification of the end of the transposon and flanking DNA, sequencing each event individually. Here, I show that I can increase the number of mutants mapped in one experiment, with the possibility to increase the number of events to the thousands. With this technology, there is no bound to the possibilities for studying the effect different protein, pathways, and target DNAs have on TnsE-mediated transposition, an area that remains elusive. Here, I will discuss one clear example of experiments that would benefit from this high-throughput mapping technology.

<u>SeqA mutants</u>

I discussed above a set of 6 *seqA* mutants that may have an effect on TnsE-mediated transposition by direct interaction with the TnsE protein. Some of them have been shown to have reduced protection of TnsE in a trypsin digest assay, where others have not yet been studied. These mutants, amongst other *seqA* mutants that have different effects on disrupting the SeqA-DNA superstructure that is formed upon SeqA binding to GmATC sites, will benefit from high-throughput mapping of TnsE-mediated transposition events in strains expressing these mutants.

TnsE-mediated insertion events in a wild-type compared to a *seqA* strain differed (discussed above, Zaoping, thesis), indicating that some feature of the SeqA-bound DNA may be recognized by TnsE in the WT strain. A major difference was in a region around the origin, where in this region, the *seqA* null strain had no insertions and the wildtype strain had approximately 5% of insertions. Given the structural differences of SeqA binding at the origin and replication fork, I can begin to draw a model of how TnsE may interact with the different DNA structures. Although the binding of SeqA at *oriC*, is within a much smaller region (~250bp) than the region where I find TnsE-mediated insertions (~250 kb), there have been established differences between the SeqA-DNA structures that form in *oriC* and structures that form behind the replication fork that may help explain the differences in TnsE-mediated insertions around the chromosome. At *oriC*, multiple SeqA dimers collate, resulting in oligomerization of SeqA complexes which forms looped DNA. Fifteen percent of TnsE-mediated insertion events in the WT background were in a region around *oriC*, whereas in the *seqA* strain, fewer than 2.5% were found in this region. If I only consider this difference, the DNA loops formed by SeqA binding DNA at the origin may be a target for TnsE-mediated transposition. In a *seqA* null strain, these loops would presumably not be formed, which, if these

are targets for TnsE, would explain the decrease in targeting events to this region in a *seqA*- background. Another possibility is that TnsE is capable of interacting with the oligomerized SeqA complex. Given its presence at the origin, TnsE-mediated insertion events may be targeted to regions where this complex has formed.

However, I must consider chromosomal GATC sites. While clusters of GATC sites on the chromosome have been found (22), these sites are generally more widely spaced, such that SeqA forms long tracks of multimers (9, 23, 24, 35), tightly coiling the DNA. With this in mind, DNA availability following the replication fork may be inaccessible given its interaction with SeqA. These coils may occlude the TnsE targets on the lagging strand in such a way that prevents TnsE binding and subsequent transposition. Another process on the lagging strand template that must be coordinated with SeqA filament formation includes okazaki fragment repair. The limited time between replication of the lagging strand template and repair of this strand is when the lagging strand is presumably the most vulnerable, as numerous single stranded gaps are found along the DNA, as discussed in Chapter 1. By forming a filament on that strand, SeqA 'protects' this strand during the interim between replication and repair. Given that in the presence of SeqA (the wild-type strain), TnsE-mediated insertions are found throughout the chromosome with an orientation bias related to lagging-strand replication, TnsE must target some point during this process. Previous work has shown that TnsE digestion by trypsin is protected by the linker domain of SeqA. This would suggest that TnsE is capable of interacting with the SeqA filaments.

Different mutants, as discussed above, have different effects on the structures formed by SeqA binding to the DNA at different regions of the chromosome. To determine whether TnsE interacts with SeqA *in vivo*, and whether this interaction is specific to the different SeqA-DNA

structures that are formed by binding to differently spaced GmATC sites, transposition in strains containing these different *seqA* mutants is imperative. Given the number of *seqA* mutants I suggest, and the possibility of other mutants that may differently interact with TnsE, using a high-throughput assay to map TnsE-mediated insertions in these strains would be efficient and cost-effective.

4.5 Methods and Materials

4.5.1 Western Blot

Cell lysate from strains expressing the indicated proteins (below) was run on a 15% poly-acrylamide gel and transferred to a PVDF (Immobilon-P, Millipore) membrane according to manufacturers recommendations. To prevent non-specific binding of the antibody, the membrane was blocked for 1 hour at room temperature in blocking buffer. Subsequently, the membrane was probed with an anti-Flag antibody (1:5000 dilution) in TBS-T + 5% non-fat dry milk powder overnight, washed with PBS-T, and incubated with the secondary goat anti-mouse AP antibody (1:2500 dilution) in the same TBS-T + 5% non-fat dry milk powder overnight solution for 40 minutes. The blots were developed with NBT/ BCIP in DMF and imaged in the BioRad Imager.

4.5.2 Slot Blot

A PVDF (Immobilon-P, Millipore) membrane was prepared according to manufacturers recommendations. TnsE or BSA was spotted on the membrane at concentrations ranging from 1 to 500 pmol. The membrane was blocked with interaction buffer in the absence of glycerol. Cell lysate from strains expressing the indicated proteins (below) was poured over blots and incubated for 4h at 4^{o}C. Blots were washed 4 times with PBS and subsequently twice with

PBS+100mM KCl. To develop the blots, they were blocked, probed, and developed as described for a Western Blot (above).

4.5.3 Protein Expression and Purification

His-6-tagged TnsE proteins from different hosts were purified as previously described (41). SeqA proteins were cloned into a pACYC184 backbone vector with a rhamnose inducible promoter and N-terminal FLAG tag upstream of the protein. After overnight growth the strains were subcultured and the proteins induced with 0.2% rhamnose for 6 hours at 30^{o}C. To collect the cells, the culture was centrifuged at 4,000rpm for 10 minutes. In order to make a cell lysate, the cells were resuspended in interaction buffer and sonicated 5 times at 30 sec. Sonicated cells were centrifuged at 18,000 rpm and filtered through a 0.45uM filter to collect only small cell components. These cell lysates were used fresh (made the same day) in the assays described above.

4.5.4 Transposition Assay

Transposition frequency was monitored by lambda hop assay. Core transposition machinery (TnsABC) was expressed from pCW15, TnsE was expressed from pJP104. The empty vector, pTA106 served as the respective negative control. Strains were constructed by transduction of a MG1655 derivative, JP1386 (F- araD139 Δ(argF-lac)U169 rspL150 relA1 flbB5301 fruA25 deoC1 ptsF25 valR Δara714) with P1 phage grown on various preexisting *dam* or *seqA* deletion strains (Miller 1992, Peters 2007). Briefly the lambda hop assay introduced a mini Tn7 element containing kanamycin resistance via an integration and replication defective lambda vector defective λ vector (λKK1 780 hisG9424::Tn10 del16 del17::attTn7::miniTn7::KanR). The resulting colonies growing on LB plates containing kanamycin were counted, indicating the

number of transposition events. Transposition frequency was determined by dividing the number of kanamycin resistant colonies by the number of plaque forming units of the λ phage.

4.5.5 Insertion Mapping

All kanamycin resistant colonies on a singular plate were collected and DNA was extracted using an Invitrogen PureLink Genomic DNA Extraction Kit, as per protocol. DNA was diluted to 20ng/uL with purified HPLC H_2O. Transposition events were mapped by Illumina Nextera sequencing using a modification of the methods described by (41). The intial arbitrary PCR was performed exactly as described previously, using a hot start. The second PCR using a primer that is nested in the transposon and a primer that anneals to the tail of the arbitrary primer. Both of these primers contained tails to the Illumina sequencing primer (5'- TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG ATA ATC CTT AAA AAC TCC ATT TCC ACC CCT and 5'- GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGG CCA CGC GTC GAC TAG TAC). The PCR reactions were equally pooled, and the third PCR added Illumina barcodes and adapters (5' AAT GAT ACG GCG ACC ACC GAG ATC TAC AC [i5] TCG GCG GCA GCG TC' and 5' CAA GCA GAA GAC GGC ATA CGA GAT [i7] GTC TCG TGG GCT CGG') . Final PCR conditions were: 20 mM Tris pH 8.8 (@25 C), 10 mM $(NH_4)_2S0_4$, 10 mM KCl, 0.1% Triton X-100, 250uM dNTP, 0.5uM Illumina i7 barcoded primer, 0.5uM Illumina i5 barcoded primer, 0.4ul/50uL reaction PFU polymerase, 5uL/50uL reaction pooled DNA. Hotstart PCR was set up, with final incubation steps of 98^{o}C for 3 minutes, followed by 8 cycles of 98^{o}C for 30 sec, 63^{o}C for 30 sec, 72^{o}C for 30 sec, and a final step at 72^{o}C for 5 min. PCR reactions were run on a 1% agarose gel and the smear corresponding to 250-600bp was cut and purified using the Progmega Gel Purification kit. DNA concentration was diluted to 2ng/uL and submitted to the Cornell Life Sciences core facility for sequencing on the Nextera sequencer.

4.5.6 Solutions and Buffers

PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 1 mM $CaCl_2$-$2H_2O$, 0.5 mM $MgCl_2$-$6H_2O$), TBS (50mM Tris-Cl, pH 7.5, 150 mM NaCl), Blocking buffer (50mM Tris, 150mM NaCl, 0.1% Tween-20, 5% non-fat dry milk powder), Interaction buffer (20mM HEPES pH7.5, 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1% Triton-X, 1% BSA, 5% Glycerol), -T (0.1% Tween added to the indicated buffers), NBT/ BCIP (50mg/mL NBT in 100%DMF, 50mg/mL BCIP in 70% DMF)

4.6 Acknowledgements

I'd like to thank Peter Schweitzer for kindly sharing the i7 and i5 barcoded illumina sequencing primers, and for assistance with Illumina sequencing and troubleshooting. In addition, I thank Marco Straus for purified TnsE proteins and assistance with protocols.

REFERENCES

1. Arciszewska LK, Craig NL. 1991. Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. *Nucleic Acids Res*. 19(18):5021–29

2. Arciszewska LK, Drake D, Craig NL. 1989. Transposon Tn7: *cis*-acting sequences in transposition and transposition immunity. *J Mol Biol*. 207(1):35–52

3. Babu VMP, Sutton MD. 2014. A DnaN plasmid shuffle strain for rapid *in vivo* analysis of mutant *Escherichia coli* β clamps provides insight into the role of clamp in UmuDC-mediated cold sensitivity. *PLoS One*. 9(6):

4. Bach T, Krekling MA, Skarstad K. 2003. Excess SeqA prolongs sequestration of *oriC* and delays nucleoid segregation and cell division. *EMBO J*. 22(2):315–23

5. Bainton R, Gamas P, Craig NL. 1991. Tn7 transposition in vitro proceeds through an excised transposon intermediate generated by staggered breaks in DNA . *Cell*. 65(5):805–16

6. Boye E. 1991. The hemimethylated replication origin of *Escherichia coli* can be initiated in vitro. *J Bacteriol*. 173(14):4537–39

7. Boye E, Løbner-Olesen A. 1990. The role of Dam methyltransferase in the control of DNA replication in *E. coli*. *Cell*. 62(5):981–89

8. Boye E, Stokke T, Kleckner N, Skarstad K. 1996. Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. *Proc. Natl. Acad. Sci.* 93(22):12206–11

9. Brendler T, Sawitzke J, Sergueev K, Austin S. 2000. A case for sliding SeqA tracts at anchored replication forks during *Escherichia coli* chromosome replication and segregation. *EMBO J*. 19(22):6249–58

10. Campbell JL, Kleckner N. 1990. *E. coli oriC* and the *dnaA* gene promoter are sequestered from Dam methyltransferase following the passage of the chromosomal replication fork. *Cell*. 62(5):967–79

11. Chung YS, Brendler T, Austin S, Guarne A. 2009. Structural insights into the cooperative binding of SeqA to a tandem GATC repeat. *Nucleic Acids Res*. 37(10):3143–52

12. Chung YS, Guarné A. 2012. Iterative optimization of DNA duplexes for crystallization of SeqA-DNA complexes. *J Vis Exp*

13. Craig, N. L. 2002. Tn7. In *Mobile DNA II* (Craig Nancy, L., Craigie, R., Gellert, M., and Lambowitz-Alan, M., eds) pp. 423-456, ASM Press, Washington, DC

14. DeBoy, R. T. 1997. *Transposon Tn7: target DNAs can modulate Tn7 insertion*. Ph.D. Thesis. thesis. Johns Hopkins University

15. Duderstadt KE, Reyes-Lamothe R, Oijen AM van, Sherratt DJ. 2014. Replication-fork dynamics. *Cold Spring Harb Perspect Biol*. 6(1):a010157

16. Finn JA, Parks AR, Peters JE. 2007. Transposon Tn7 directs transposition into the genome of filamentous bacteriophage M13 using the element-encoded TnsE protein. *J Bacteriol*. 189(24):9122–25

17. Fossum-Raunehaug S, Helgesen E, Stokke C, Skarstad K. 2014. *Escherichia coli* SeqA structures relocalize abruptly upon termination of origin sequestration during multifork DNA replication. *PLoS ONE*. 9(10):e110575

18. Fujikawa N, Kurumizaka H, Yamazoe M, Hiraga S, Yokoyama S. 2003. Identification of functional domains of the *Escherichia coli* SeqA protein. *Biochem. Biophys. Res. Commun.* 300(3):699–705

19. Gringauz E, Orle KA, Waddell CS, Craig NL. 1988. Recognition of *Escherichia coli attTn7* by transposon Tn7: lack of specific sequence requirements at the point of Tn7 insertion. *J Bacteriol*. 170(6):2832–40

20. Guarné A, Brendler T, Zhao Q, Ghirlando R, Austin S, Yang W. 2005. Crystal structure of a SeqA–N filament: implications for DNA replication and chromosome organization. *EMBO J*. 24(8):1502–11

21. Heller RC, Marians KJ. 2006. Replication fork reactivation downstream of a blocked nascent leading strand. *Nature*. 439(7076):557–62

22. Hénaut A, Rouxel T, Gleizes A, Moszer I, Danchin A. 1996. Uneven distribution of GATC motifs in the *Escherichia coli* chromosome, its plasmids and its phages. *J Mol Biol*. 257(3):574–85

23. Hiraga S, Ichinose C, Niki H, Yamazoe M. 1998. Cell cycle-dependent duplication and bidirectional migration of SeqA-associated DNA -protein complexes in *E. coli*. *Mol. Cell*. 1(3):381–87

24. Hiraga S, Ichinose C, Onogi T, Niki H, Yamazoe M. 2000. Bidirectional migration of SeqA-bound hemimethylated DNA clusters and pairing of *oriC* copies in *Escherichia coli*. *Genes Cells*. 5(5):327–41

25. Li Z. 2012. *A Molecular Mechanism Allowing Transposon Tn7 to Target Active DNA Replication*. Cornell University

26. Li Z, Craig NL, Peters JE. 2013. Transposon Tn7. *Bacterial Integrative Mobile Genetic Elements*, p. 1

27. Lichtenstein C, Brenner S. 1982. Unique insertion site of Tn7 in the *E. coli* chromosome. *Nature*. 297(5867):601–3

28. Løbner-Olesen A, Marinus MG, Hansen FG. 2003. Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis. *Proc Natl Acad Sci*. 100(8):4672–77

29. Lu M, Campbell JL, Boye E, Kleckner N. 1994. SeqA: a negative modulator of replication initiation in *E. coli*. *Cell*. 77(3):413–26

30. McInerney P, Johnson A, Katz F, O'Donnell M. 2007. Characterization of a triple DNA polymerase replisome. *Mol Cell*. 27(4):527–38

31. McKown RL, Orle KA, Chen T, Craig NL. 1988. Sequence requirements of *Escherichia coli attTn7*, a specific site of transposon Tn7 insertion. *J Bacteriol*. 170(1):352–58

32. McKown RL, Waddell CS, Arciszewska LK, Craig NL. 1987. Identification of a transposon Tn7-dependent DNA -binding activity that recognizes the ends of Tn7. *Proc Natl Acad Sci*. 84(22):7807–11

33. Mitra R, McKenzie GJ, Yi L, Lee CA, Craig NL. 2010. Characterization of the TnsD-*attTn7* complex that promotes site-specific insertion of Tn7. *Mobile DNA*. 1(1):18

34. Ogden GB, Pratt MJ, Schaechter M. 1988. The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. *Cell*. 54(1):127–35

35. Onogi T, Niki H, Yamazoe M, Hiraga S. 1999. The assembly and migration of seqa–gfp fusion in living cells of *Escherichia coli*. *Mol Microbiol*. 31(6):1775–82

36. Parks A. 2008. *Tn7 transposes into replicating DNA using an interaction with the processivity factor, facilitating genome evolution*

37. Parks AR, Li Z, Shi Q, Owens RM, Jin MM, Peters JE. 2009. Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell*. 138(4):685–95

38. Parks AR, Peters JE. 2009. Tn7 elements: engendering diversity from chromosomes to episomes. *Plasmid*. 61(1):1–14

39. Peters JE. 2014. Tn7. *Microbiology Spectrum*. 2(5):

40. Peters JE, Craig NL. 2000. Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Mol Cell*. 6(3):573–82

41. Peters JE, Craig NL. 2001. Tn7 recognizes transposition target structures associated with DNA replication using the DNA -binding protein TnsE. *Genes Devel*. 15(6):737–47

42. Peters JE, Craig NL. 2001. Tn7: smarter than we thought. *Nature Review Mol Cell Biol*. 2(11):806–14

43. Pomerantz RT, O'Donnell M. 2008. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature*. 456(7223):762–66

44. Pomerantz RT, O'Donnell M. 2010. What happens when replication and transcription complexes collide? *Cell Cycle*. 9(13):2537–43

45. Roberts D, Hoopes BC, McClure WR, Kleckner N. 1985. IS10 transposition is regulated by DNA adenine methylation. *Cell*. 43(1):117–30

46. Rupp WD, Howard-flanders P. 1968. Discontinuities in the DNA synthesized in an excision-defective strain of *Escherichia coli* following ultraviolet irradiation. *J Mol Biol*. 31(2):291–304

47. Sánchez-Romero MA, Busby SJW, Dyer NP, Ott S, Millard AD, Grainger DC. 2010. Dynamic distribution of SeqA protein across the chromosome of E*scherichia coli* K-12. *mBio*. 1(1):e00012–10

48. Sarnovsky RJ, May EW, Craig NL. 1996. The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J*. 15(22):6348

49. Shakibai N, Ishidate K, Reshetnyak E, Gunji S, Kohiyama M, Rothfield L. 1998. High-affinity binding of hemimethylated *oriC* by *Escherichia coli* membranes is mediated by a

multiprotein system that includes SeqA and a newly identified factor, SeqB. *Proc Natl Acad Sci*. 95(19):11117–21

50. Skarstad K, Lueder G, Lurz R, Speck C, Messer W. 2000. The *Escherichia coli* SeqA protein binds specifically and co-operatively to two sites in hemimethylated and fully methylated *oriC*. *Mol Microbiol*. 36(6):1319–26

51. Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N. 1995. *E. coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell*. 82(6):927–36

52. Slominska M, Wegrzyn A, Konopa G, Skarstad K, Wegrzyn G. 2001. SeqA, the *Escherichia coli* origin sequestration protein, is also a specific transcription factor. *Mol Microbiol*. 40(6):1371–79

53. Słomińska M, Konopa G, Barańska S, Wegrzyn G, Wegrzyn A. 2003. Interplay between DnaA and SeqA proteins during regulation of bacteriophage lambda pr promoter activity. *J Mol Biol.* 329(1):59–68

54. Słomińska M, Konopa G, Ostrowska J, Kedzierska B, Wegrzyn G, Wegrzyn A. 2003. SeqA-mediated stimulation of a promoter activity by facilitating functions of a transcription activator. *Mol. Microbiol.* 47(6):1669–79

55. Torheim NK, Skarstad K. 1999. *Escherichia coli* SeqA protein affects DNA topology and inhibits open complex formation at *oriC*. *EMBO J.* 18(17):4882–88

56. Von Freiesleben U, Krekling MA, Hansen FG, Løbner-Olesen A. 2000. The eclipse period of *Escherichia coli*. *EMBO J*. 19(22):6240–48

57. Waddell CS, Craig NL. 1989. Tn7 transposition: recognition of the *attTn7* target sequence. *Proc Natl Acad Scis*. 86(11):3958–62

58. Weitao T, Nordström K, Dasgupta S. 1999. Mutual suppression of MukB and SeqA phenotypes might arise from their opposing influences on the *Escherichia coli* nucleoid structure. *Mol Microbiol*. 34(1):157–68

59. Wolkow CA, DeBoy RT, Craig NL. 1996. Conjugating plasmids are preferred targets for Tn7. *Genes Dev*. 10(17):2145–57

60. Yeeles JTP, Marians KJ. 2011. The *Escherichia coli* replisome is inherently DNA damage tolerant. *Science*. 334(6053):235–38

61. Yin JCP, Krebs MP, Reznikoff WS. 1988. Effect of Dam methylation on Tn5 transposition. *J Mol Biol*. 199(1):35–45

62. Zakrzewska-Czerwińska J, Jakimowicz D, Zawilak-Pawlik A, Messer W. 2007. Regulation of the initiation of chromosomal replication in bacteria. *FEMS Microbiol. Rev.* 31(4):378–87

63. Zyskind JW, Smith DW. 1986. The bacterial origin of replication, *oriC*. *Cell*. 46(4):489–90

CHAPTER 5

CONCLUSION

Mobile genetic elements need to find a suitable home to ensure they are protected from the environment. This home has certain requirements, for example the mobile genetic element cannot decrease the viability of the host. In chapter 1, I look at how one possible "home" (ie the lagging strand template during replication) for these mobile genetic elements is vulnerable in certain situations, such as during replication, when the replication forks stall or collapse. We discuss multiple types of mobile genetic elements that recognize different features of the lagging strand, and suggest that Tn7 among other elements can recognize this strand. However, Tn7 is also capable of forming genomic islands in bacterial chromosomes at *attTn7* resulting from the TnsD target site selection pathway. Tn7 transposition requires use of its atypical heteromeric transposase comprised of TnsAB, and in chapter 2, I find that there are at least two other related elements with heteromeric transposases. Tn7-like elements have TnsD-like proteins, where these other two elements appear to have a protein related to TnsD which are likely involved in site selection in their respective organisms. The observation that Tn7-like elements are all found at a single position in the *glmS* gene has been known however, interestingly I find that these other two heteromeric transpsoase elements also appear to target insertions into specific sites. We find that Tn6230 and the majority of related elements are found at a single position in the *yhiN* gene in various bacteria. The other element that shares a heteromeric transposase, Tn6022 found in *Acinetobacter baumannii*,and related elements appear to insert into the *comM* gene. Although Tn6230-like elements appear to have one conserved protein, the TnsF-like protein, that shares features with TnsD, the Tn6022-family elements have one to three poorly conserved

proteins. Despite the diversity of conserved proteins between these three families of elements, as a group heteromeric transposases seem to share a basic mechanism for targeting to these sites adapted to the spread of mobile elements.

While chapter 2 focused on specific sites of the chromosome that are recognized by one target site selecting protein, Tn7 encodes another protein that is responsible for direction transposition into the lagging strand template. The interaction of TnsE with the β-clamp and 3' recessed ends has been well studied, but the ability for TnsE to recognize the β-clamps from different hosts was contentious because of the nature of the experiments previously used. In chapter 3, I confirmed that the observed effect of having the TnsE and β-clamp from the same host led to an increase in papillae was ultimately due to actual transposition events. I also showed that the β-clamps from *I. loihiensis* and *S. baltica* can work with the replication components in *E. coli* in the absence of any *dnaN* allele, indicating that the previous work studying the effect of the foreign clamps in a temperature sensitive strain were not a result of heterodimer formation between the *dnaN159* allele and the foreign clamps. Although these foreign clamps are capable of supporting growth, they may have a greater interaction with the accessory polymerases during replication which may lead to an increase in the mutation rate. The effect of a greater mutation frequency based on rifampicin resistance had been previously observed, and I was able to show that these were actual mutations.

Another protein that has been found to interact with TnsE is SeqA. While the frequency and targeting in the absence of SeqA has previously been studied, the nature of the TnsE-SeqA interaction remains elusive. In order to advance what the mechanism may be, I expanded upon previously existing *in vitro* and *in vivo* techniques. I developed a vector that was used to express SeqA and a set of six additional mutants with an N-terminal tag. Using this vector, not only can

we express additional SeqA mutants, but a host of other proteins and their mutants to study the interaction with TnsE. I also expanded upon the *in vivo* experiments which involved establishing a high-throughput sequencing method to map TnsE-mediated insertion events. This sequencing technique can ideally map transposition events in about 100 different strains at a time, allowing our imaginations to run wild with possibilities. With this technique, we can start to use mapping events as a screen for exciting mutants that affect TnsE-mediated transposition.

APPENDIX 1

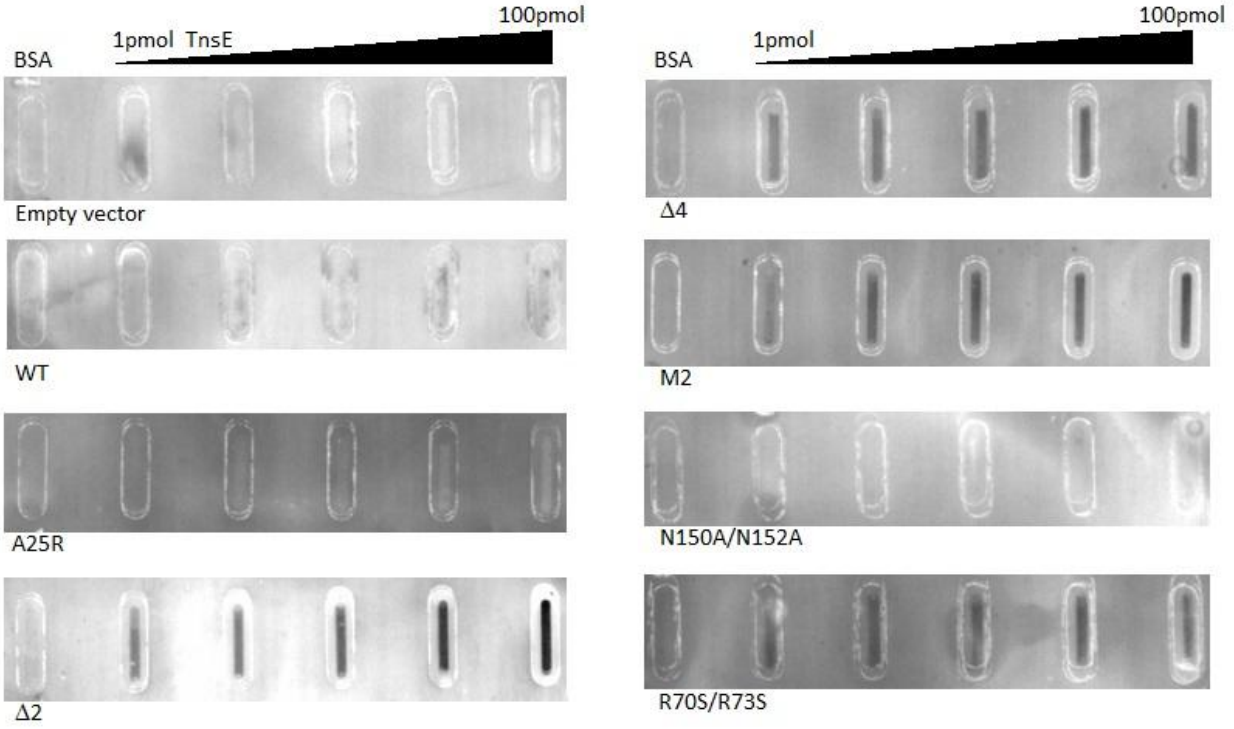

Figure 4.4

Interaction of SeqA mutants with TnsE. SeqA mutants are as follows: RS (R70S/R73S), AR (A25R), NA (N150A/N152A), M2 (48 GDRASSPASDKG59), D4 (41-59), D2 (45-59). A) Slot Blot of TnsE and BSA overlayed with various SeqA mutants from *E. coli* SeqA is expressed after 6h induction with 0.2% rhamnose from a BW27784 protein expression strain. His-tagged TnsEs were purified from a BL21 fur-/slyD- strain using Nickel column, and subsequently ion exchange chromatography. BSA was purchased from Fisher Scientific. Slot blot was incubated overnight with lysate from a N-terminal flag tagged SeqA. B) Western blot of N-terminal flag tagged SeqA. Both blots were incubated with mouse-anti-flag antibody (1:2000); secondary goat anti-mouse (1:4000).