UNIVERSITY OF CALIFORNIA SAN DIEGO

Analysis of the Effects of Genomic Context on IS5 Transposition in *Escherichia coli*

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Jonathan Onstead

Committee in charge:

Professor Milton H. Saier Jr., Chair
Professor James W. Golden, Co-Chair
Professor Kevin D. Corbett

2024

The Thesis of Jonathan Onstead is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Milton Saier, Dr. Zhongge Zhang, and everyone I have worked with in the Saier Lab during my stay there. Without your help this thesis would not have been possible!

The information contained within this thesis, in full, is being prepared for publication: Onstead, Jonathan; Zhang, Zhongge; Saier, Milton H., Jr. The thesis author was the primary researcher and author of this paper.

ABSTRACT OF THE THESIS

Analysis of the Effects of Genomic Context on IS5 Transposition in *Escherichia coli*

by

Jonathan Onstead

Master of Science in Biology

University of California San Diego, 2024

Professor Milton H. Saier Jr., Chair
Professor James W. Golden, Co-Chair

An important transposable element (transposon) within the genome of the *Escherichia coli* K-12 strain BW25113 is IS5, of which there are 11 total copies. This work follows up on past research into IS5 by investigating which one of these 11 copies is the most significant and most impactful on overall IS5 activity and transposition. In addition, it investigates possible environmental and genomic factors influencing the most significant copies, as well as the

mechanism of transposition (“cut and paste” or “copy and paste”) used by these copies. We do so by the use of mutation assays, in which strains to be evaluated are plated on minimal media such that the only way for bacterial colonies to grow is for certain IS5 transpositions to occur, making the number of resulting growing colonies a useful indicator of overall IS5 activity. We find that, by far, the most significant copy of IS5 is the copy at the *nmpC* location, followed by copies at the *lnt* and *wbbL* locations, and that a further two copies have mutations that prevent IS5 transposition. We find that the significance of these locations remains invariant with respect to the operon and minimal media used in the mutation assay. We also examine the *nmpC* copy in more detail and find the reason for its significance is likely due to the presence of a very strong promoter upstream of this gene, and that there is minimal evidence for a bias of IS5 transposases towards acting on this copy. We construct a strain with only the IS5 copy at *nmpC* and establish a “copy and paste” mechanism for IS5 transposition from *nmpC*. Lastly, we investigate the possible roles that two genes that overlap the *ins5a* transposase gene, *ins5b* and *ins5c*, play in regulating IS5 transposition, and find that higher levels of their expression have minimal effect.

INTRODUCTION

Transposon-mediated Regulation

Two basic yet critically important biological processes involve gene regulation and genetic mutation. Gene regulation involves a network of molecular machinery that determines when certain genes are turned "on" or "off" and thus which proteins the organism produces, which in turn can affect the organism's phenotype. On the other hand, genetic mutation involves heritable changes to the genome of an organism that, for instance, can be selected for or against depending on the organism's environment. While these are generally considered two distinct processes, surprisingly, there are times when they converge. One such instance involves the use of transposons, mobile genetic elements that can excise and then insert into other regions within the organism's genome that generate insertional mutations (McClintock, 1950).

These "hopping" events are mediated by transposases, enzymes that are often encoded by genes on the transposons themselves, and can occur in two mechanisms: "cut-and-paste", where the transposon excises itself from the genome and re-inserts elsewhere, and "copy-and-paste" where the transposon can both remain in its original location and insert a copy into another region of the genome (Berg & Howe, 1989; May & Craig, 1996). Previous research indicated that transposition occurring upstream of or within a gene or operon may also are able to play a role in the regulation of that gene or operon, with the regulatory role depending on the gene or operon, the transposon, and the insertional location (Jaurin & Normark, 1983; Hall, 1998; Zhang & Saier, 2009a; Saier *et al*., 2017). We shall refer to this process as "transposon-mediated regulation".

In some cases, such as in the bacterium, *Escherichia coli* (*E. coli*), the insertion of transposons in transposon-mediated regulation may be nonrandom, with certain "hot spot"

regions displaying higher rates of insertion than in any other part of the genome. This includes regions near *glpFK,* an operon responsible for the uptake and metabolism of glycerol (Zhang & Saier, 2009b); *bglGFB,* an operon responsible for the uptake and metabolism of aromatic β-glucosides (Reynolds *et al*., 1981); *flhDC,* an operon containing flagellar regulatory genes (Barker *et al*., 2004); *fucAO,* an operon responsible for metabolism of propanediol (Chen *et al*., 1989); and *nfsA/nfsB,* a system responsible for nitroreduction and furazolidone sensitivity (Whiteway *et al*., 1998). These "hot spots" can be generated by sequences known as superhelical stress-induced duplex destabilization (SIDD) sites, which act by possessing a specific structure that makes it energetically favorable for transposition into them to occur. Many of these SIDD sites in *E. coli* are found in or near genes or operons affected by transposon-mediated regulation, including all the aforementioned operons. Furthermore, rates of transposition into these SIDD regions can increase during times of stress, allowing transposition to generate mutations that might be beneficial to the survival of the organism through this regulation (Humayun *et al*., 2017). In total, nonrandom mutations induced by transposons, influenced by the environment and heightened during times of stress to convey a fitness advantage, provide evidence of a sort of natural "directed mutation" process, which has been well-studied by our group in the past (Zhang & Saier, 2009b; Zhang & Saier, 2011; Zhang & Saier, 2016; Zhang *et al*., 2017; Saier *et al*., 2017; Zhang *et al*., 2024; Kopkowski *et al*., 2024).

Unlike gene regulation through other mechanisms like protein DNA binding, which acts on single regulatory locations for each gene or operon, transposon-mediated regulation involves two key regions: the transposition insertion site near or within the target gene or operon, and the location of origin of the transposon. The genomic context of a transposon's initial location might have many key implications for transposition; for instance, position of the transposon near a

strong promoter could increase the rate of transposition due to higher expression of the transposase. As such, to get a full story on a particular transposon, and any regulatory behavior it may possess, it is imperative to study not only its targets of insertion but also the genomic context of its initial location (Sawers, 2005). However, given this latter factor is an oft-overlooked consideration, there is sometimes surprisingly little literature on key transposons focusing on the impact of this initial location genomic context. Our study aims to fill this gap, at least for one transposon, by confirming the importance of initial location genomic context and by investigating how this context might impact transposition (such as through the aforementioned example with promoters). Specifically, in this study, we shall investigate the *E. coli* K-12 strain BW25113. In this strain, there are two transposons, referred to as insertion sequences or IS elements (Mahillon & Chandler, 1998), involved in most transposon-mediated gene regulation, known as IS1 and IS5. We shall mostly focus on IS5 given its intricacy and numeracy of initial locations and targets.

IS5 and its Copies

IS5 is a 1195 base pair transposon, flanked by two Inverted Repeats (IRs), with three genes: *ins5a*, encoding the 981 bp transposase for IS5 in the forward direction, and *ins5b* (324 bp) and *ins5c* (255 bp) with unknown purpose and relevance in the reverse direction. IS5 has very weak internal promoters along *ins5a* as well as *ins5b/c* which provide a minimal baseline to the transcription of these genes (Schoner & Kahn, 1981; Kröger & Hobom, 1982; Rak & von Reutern, 1984). There are a total of 11 copies of IS5 in the BW25113 strain (Grenier *et al*., 2014) which raises the question of how much the genomic context of each location affects the rate of transposition of IS5, and which copy of IS5 instigates the most transposition (Sawers, 2005). Determining the answers to these questions will have major implications on the study of the

evolutionary history of transposon-mediated regulation as well as increasing the body of knowledge on how the entire transposon-mediated regulatory system works in *E. coli*.

The aims of this work are then to (1) determine how significant each location of IS5 is to the overall IS5 transposition, thus finding the most significant location(s) of IS5, and (2) analyzing the most significant location to determine the mechanism and basis by which the genetic environment surrounding this copy of IS5 gives rise to a high rate of transposition. In so doing, we shall also uncover the transposition mechanism (cut or copy and paste) at this location, as well as the significance of the *ins5b* and *ins5c* genes, if any, to transposition.

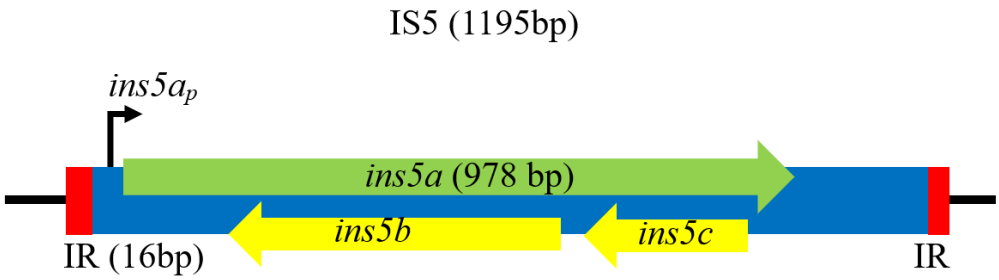


Figure 1.1. Schematic Depiction of the IS5 Transposon.
The transposase gene, *ins5a*, is shown in green, the inverted repeats (IR) are shown in red, and the main body of the transposon is shown in blue. The figure is approximately to scale.

The 11 copies of IS5 in the BW25113 genome include locations at *nmpC* (IS5B), *wbbL* (IS5I), *lnt* (IS5D), *yejO* (IS5K), *mmuP* (IS5A), *ynaI* (IS5F), *yghO* (IS5LO), *yhcE* (IS5R), *yhiS* (IS5T), *yoeG* (IS5H), and *ychE* (IS5U) (Schnetz & Rak, 1992). These genes encode an outer membrane porin, *nmpC* (Hindahl *et al*., 1984), a sugar transferase, *wbbL* (Rubirez *et al*., 1997), a mediator of acylation of apolipoproteins, *lnt* (Gupta & Wu, 1991; Robichon *et al*., 2005), an outer membrane autotransporter, *yejO* (Henderson & Owen, 1999), an *S*-methylmethionine transporter, *mmuP* (Thanbichler *et al*., 1999), a mechanosensitive channel, *ynaI* (Edwards *et al*., 2012), a transcription regulator active during biofilm production, *yghO* (Beloin *et al*., 2004), a gene fragment, *yhcE* (Korea *et al*., 2010), another gene fragment, *yhiS* (Homma *et al*., 2002), a

defective prophage, *yoeG* (Goodall *et al*., 2018), and an inner membrane protein, *ychE* (Daley *et al*., 2005), respectively. None of these genes (except *lnt*) is essential (Gerdes *et al*., 2003; Baba *et al*., 2006), which explains how IS5 can be present within some of them without affecting the growth and health of the bacterium. Of these genes, *ins5a* exists under the nearest promoter in the *nmpC, lnt, wbbL,* and *ychE* locations and against it in the *mmuP, ynaI, yejO, yghO, yhcE,* and *yhiS* locations (the *yoeG* location does not contain a known nearby promoters) (Huerta & Collado-Vides, 2003; Mendoza-Vargas *et al*., 2009; Grenier *et al*., 2014; Goodall *et al*., 2018). Previous work conducted by our group tentatively indicated that the *nmpC* copy may contribute the most to transposase production of any IS5 copy, as most sequenced cDNAs from the IS5 transposase mRNA contained signs of origination from this location (Zhou, 2023). As such, a more specific goal of the current study was to verify this observation and determine if *nmpC* is the most significant initial location of an IS5 copy contributing to the overall IS5 insertional activity, and if so, to determine why this was so.

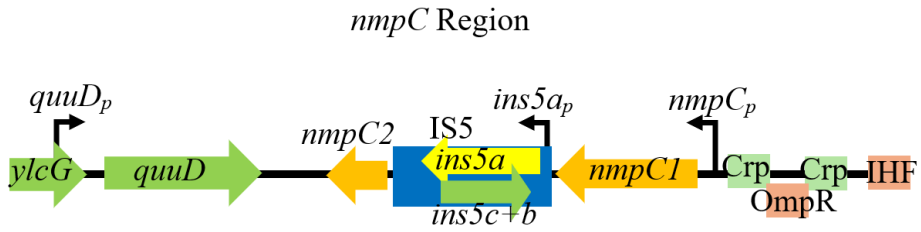


Figure 1.2. Schematic depiction of the *nmpC* gene and the surrounding context.
This schematic portrayal highlights the surrounding genomic context of the *nmpC* gene, including the presence of two promoters pointing through *nmpC* and IS5 in both directions (*nmpCp* and *quuDp*). It also shows regions where protein binding may enhance the promoter (light green, Crp) and where it may repress the promoter (light red, OmpR and IHF). The figure is approximately to scale.

As mentioned before, *nmpC* in the BW25113 strain encodes an inactivated outer membrane porin protein (Hindahl *et al*., 1984), a member of the General Bacterial Porin (GBP)

family (Zhai & Saier, 2002). Inactivation of the gene does not affect *E. coli* growth due to the presence of other genes coding for porins that serve similar functions. In fact, due to this redundancy, the higher sensitivity to certain phages induced by activation of *nmpC* and the surrounding genomic structure, it is thought that *nmpC* might not be native to *E. coli*; rather, it may have been introduced into the genome in the past via a phage (Pugsley & Schnaitman, 1978, Highton *et al*., 1985). The gene itself consists of a promoter, *nmpCp*, which is positively regulated by upstream Crp binding sites and negatively regulated by upstream OmpR and IHF binding sites, indicating, for instance, possible regulation by external pH (Coll *et al*., 1994). The *nmpC* gene consists of two fragments separated by IS5: the first, starting at +44 from the promoter, is 1040 base pairs long before it is interrupted by IS5; the second is a mere 58 base pairs long, for a total length of 1098 base pairs (the entire region, including IS5, is 2297 base pairs in length). The *nmpC* promoter lies along the same direction as *ins5a* and the natural IS5 promoter *insHp*; thus, activation of *nmpC* transcription should also result in IS5 transposase transcription (Blasband *et al*., 1986; Grenier *et al*., 2014). The entire *nmpC* region neighbors the *ybcQ* gene, whose promoter runs counter to that of the *nmpC* promoter (Huerta & Collado-Vides, 2003; Grenier *et al*., 2014). This might be significant for transcription of *ins5b* and *ins5c*, which run in the same direction.

Relevant Target Operons

To test our hypothesis, a systematic methodology for ascertaining general IS5 transposition rates is needed so that the IS5 activity in various test strains can be compared. To do this, we exploited the transposon-mediated regulatory attributes of IS5, specifically in the *bglGFB* operon and tangentially for the *fucAO* and *flhDC* operons. The *bglGFB* operon consists

of three primary genes: *bglG*, an anti-terminator protein that positively regulates IS insertion into the operon, *bglF*, a phosphotransferase Enzyme II that phosphorylates its sugar substrates, and *bglB*, a beta-glucosidase (Prasad & Shaefler, 1974; Schnetz & Rak, 1988). In wild type strains, the operon is inactive (Schnetz 1995), primarily due to binding of the histone-like, H-NS protein to the promoter region of the *hns* gene) (Dole *et al*., 2004; Lam *et al*., 2022) as well as other factors (Tran *et al*., 2022), and consequently, the bacterium cannot usually process beta-glucosides. However, certain mutations, such as the insertion of either IS1 or IS5 upstream of the *bgl* promoter (made possible due to the presence of a SIDD region between positions -70 and -150), or a non-insertional mutation affecting either this H-NS binding site or the *hns* gene itself, can activate the operon (Reynolds *et al*., 1981; Hall, 1998; Humayun *et al*., 2017). Due to this, by plating wild type cultures on a minimal medium plate containing the beta-glucoside salicin as the sole carbon source, the only colonies that should grow will be the ones that have experienced one of these mutations, and we can refer to the colony as a *bgl*+ mutant. By PCR amplifying the region upstream of *bgl* can be used to count the number of IS5 insertions specifically, as non-insertional mutants will leave no bands on a gel, while bands from IS1 insertional mutants will run noticeably further than bands from IS5 insertional mutants due to the smaller size of IS1 (Kopkowski *et al*., 2024). Counting the number of these insertions should therefore be a useful approach for measuring the overall activity of IS5 within a certain strain, with the more active transposition resulting in more colonies on the minimal medium + salicin plates due to higher counts of IS5 insertion into *bgl.*

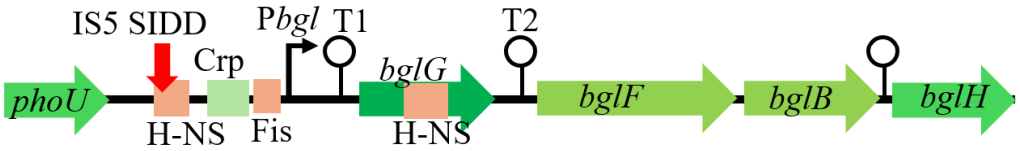


Figure 1.3. Schematic of the *bglGFB* operon.
This schematic depiction shows the overall structure of the *bglGFB* operon that is relevant to the present study. Loops indicate terminators (T1 and T2), light green indicates binding sites that promote *bgl* expression (Crp), and orange indicates binding sites that repress *bgl* expression (most notably H-NS, which creates a DNA loop between its two binding sites (Ref.). The SIDD region where IS5 is likely to insert is also indicated (Ref.). The figure is approximately to scale.

Similarly, like *bglGFB*, the *fucAO* operon is normally inactive in the wild type genetic background, allowing the bacteria to make use of fucose. When activated, such as by IS5 insertion, the bacteria lose the ability to metabolize fucose, but gain the ability to metabolize propanediol (PPD) (Chen *et al*., 1989). As with salicin and the *bgl* operon, one can grow strains in a minimal medium with propanediol as the carbon source and count the resulting colonies to get a reliable indication of IS5 activity (Zhang *et al*., 2024). Lastly, the *flhDC* operon can be a target of a mutation assay. This operon encodes two transcriptional regulatory proteins that function together to promote transcriptional activation of the flagellar synthesis regulon, but it has low activity in wild type *E. coli* (Wang *et al*., 2006). Insertion of IS5 upstream of *flhDC* can activate it and induce "swarming" behavior that can be observed in semi-solid agar media (Barker *et al*., 2004; Zhang *et al*., 2017). Counting the number of swarming colonies is therefore also a good indicator of IS5 activity.

Table 2.1. List of strains used.

Strain	**Description**
BW25113 WT	Main background strain and wild type
IS5:*km @ nmpC*	Knockout of IS5 @ *nmpC*
nmpCp:*km* and Δ*nmpCp*	Knockout of *nmpCp*
nmpC:*km*	Knockout of *nmpC* region
IS5:*km @ wbbL*	Knockout of IS5 @ *wbbL*
IS5:*km @ lnt*	Knockout of IS5 @ *lnt*
IS5:*km @ yejO*	Knockout of IS5 @ *yejO*
IS5:*km @ mmuP*	Knockout of IS5 @ *mmuP*
IS5:*km @ ynaI*	Knockout of IS5 @ *ynaI*
IS5:*km @ yghO*	Knockout of IS5 @ *yghO*
IS5:*km @ yhcE*	Knockout of IS5 @ *yhcE*
IS5:*km @ yhiS*	Knockout of IS5 @ *yhiS*
ΔIS5 @ *nmpC* + *wbbL*, ΔIS5 @ *nmpC* + *wbbL* + *lnt*	Knockouts of multiple IS5 copies
ΔIS5 @ *nmpC* + *wbbL* + *lnt* + *yejO* aka Δquad	Knockout of four key IS5 copies
IS1:*km @ yrhD*	Knockout of IS1 @ *yrhD*
IS1:*km @ yihU*	Knockout of IS1 @ *yihU*
IS1:*km @ nhaR*	Knockout of IS1 @ *nhaR*
IS1:*km @ csdH*	Knockout of IS1 @ *csdH*
IS1:*km @ afuB*	Knockout of IS1 @ *afuB*
IS1:*km @ argF*	Knockout of IS1 @ *argF*
ΔIS5 @ *yrhD* +*yihU*, ΔIS5 @ *yrhD* + *yihU* + *nhaR*, ΔIS5 @ *yrhD* + *yihU* + *nhaR* + *csdH*, ΔIS5 @ *yrhD* + *yihU* + *nhaR* + *csdH* + *afuB*	Knockouts of multiple IS1 copies
ΔIS5 @ *yrhD* + *yihU* + *nhaR* + *csdH* + *afuB* + *argF* aka ΔIS1	Knockout of all IS1 copies
ΔIS1 + IS5:*km @ nmpC*	Knockout of IS5 @ *nmpC* on ΔIS1 background
ΔIS1 + IS5:*km @ wbbL*	Knockout of IS5 @ *wbbL* on ΔIS1 background
ΔIS1 + IS5:*km @ lnt*	Knockout of IS5 @ *lnt* on ΔIS1 background
ΔIS1 + IS5:*km @ yejO*	Knockout of IS5 @ *yejO* on ΔIS1 background
T1 + *km @ nmpC* and T1 @ *nmpC*	Insertion of terminator T1 in front of *nmpCp*
T2 + *km @ nmpC* and T2 @ *nmpC*	Insertion of terminator T2 behind IS5 @ *nmpC*
T1 + T2 @ *nmpC*	Insertion of T1 in front of *nmpCp* and of T2 behind IS5 @ *nmpC*
nmpCp:P*tet*	Replacement of promoter at *nmpC* with P*tet*

Table 2.1. List of strains used, continued.

Strain	Description
RI-*nmpCp*:P*tet*	*nmpCp*:P*tet* with the addition of *tetR*
IS5m @ *nmpC*	Strain with IS5 @ *nmpC* tagged with single base pair mutation A -> T at position 1152 along IS5
ΔIS1 + ΔIS5 @ *wbbL* + *lnt*, ΔIS1 + ΔIS5 @ *wbbL* + *lnt* + *yejO*, ΔIS1 + ΔIS5 @ *wbbL* + *lnt* + *yejO* + *mmuP*, ΔIS1 + ΔIS5 @ *wbbL* + *lnt* + *yejO* + *mmuP* + *ynaI*, ΔIS1 + ΔIS5 @ *wbbL* + *lnt* + *yejO* + *mmuP* + *ynaI* + *yghO*, ΔIS1 + ΔIS5 @ *wbbL* + *lnt* + *yejO* + *mmuP* + *ynaI* + *yghO* + *yhcE*	Strains with multiple IS5 copy deletions on the IS1 background strain
ΔIS1 + ΔIS5 @ *wbbL* + *lnt* + *yejO* + *mmuP* + *ynaI* + *yghO* + *yhcE* + *yhiS* aka ΔIS1+IS5 ex. @ *nmpC*	Knockout of all IS1 and IS5 copies except for the IS5 copy at *nmpC*

Table 2.2. List of oligonucleotides used.

Name	Sequence	Use
Pbgl-F2	tggcgatgagctggataaactgctg	Determining kind of mutation in a *bgl*+ colony via PCR
Pbgl-R2	tcagttcatgactgctcaaggcatac	Determining kind of mutation in a *bgl*+ colony via PCR, sequencing IS5 that have jumped
yoeG-ver-F	ttgattcaacgtcagcgacagatctg	PCR for IS5 @ yoeG
yoeG-ver-R	taaatctggcaagtcacattctgttg	PCR for IS5 @ yoeG, Sequencing IS5 @ yoeG for mutations
ychE-ver-F	tatggcttacagccgtgtgtcgac	PCR for IS5 @ ychE
ychE-ver-R	ttcggtcgggttaaactgcacgccag	PCR for IS5 @ ychE, Sequencing IS5 @ ychE for mutations
IS5-nmpC-P1	gtctactttcgttgattacaaaatca acctgcttgacaaaaatgacttcac tgtgtaggctggagctgcttc	Construction of knockout of IS5 @ *nmpC*
IS5-nmpC-P2	gtaaaccagacctacagcaacgatgt catcagtgcttacaccgagtgctttatt ccggggatccgtcgacctg	Construction of knockout of IS5 @ *nmpC* and *nmpC* region
nmpCp-km-P1	tgaaccttcaaattatagagcactt ataaataacagccgttaatataaatt gtgtaggctggagctgcttc	Construction of knockout of *nmpCp* and *nmpC* region

Table 2.2. List of oligonucleotides used, continued.

Name	Sequence	Use
nmpCp-km-P2	cattagtactgatgcagctacagca gaaattgccactgttaatttttcatatt ccggggatccgtcgacctg	Construction of knockout of *nmpCp*
ybcQ-ver-R	aagcaaggttgtgcttctaaaggaag	Verification of knockout for IS5 @ *nmpC*, sequencing IS5 and IS5m @ *nmpC*, determination of mechanism of IS5 @ *nmpC*
IS5-all-P1	ggaaggtgcgaataagcggggaaat tcttctcggctgactcagtcatttcatg tgtaggctggagctgcttc	Construction of IS5 knockout at any location
IS5-wbbL-P2	cgttatgcttttagtacattatgcctgttc cgagatgaagcgaaatctttacatatg aatatcctccttag	Construction of knockout of IS5 @ *wbbL*
wbbL-ver-R	atcatgaagcatgatgatttgctgac	Verification of knockout for IS5 @ *wbbL*
IS5-lnt-P2	gtagcactaaaagagtgcattttaagaa ttttgcttcggaatggtgcggcgcata tgaatatcctccttag	Construction of knockout of IS5 @ *lnt*
lnt-ver-R	tcagtacaacctagttgcaccacag	Verification of knockout for IS5 @ *lnt*
IS5-yejO-P2	caatctggttctgtttctctttgtcgttcc gcaatatctgttctggtggctacata tgaatatcctccttag	Construction of knockout of IS5 @ *yejO*
yejO-ver-R	ttcataggatgaaagctcaatgcatc	Verification of knockout for IS5 @ *yejO*
IS5-mmuP-P2	cttaatagcgggtggggattaataatca gatacaaaaagggccgccagctcatt ccggggatccgtcgacctg	Construction of knockout of IS5 @ *mmuP*
mmuP-ver-R	tatggacatctaaacatccaaagaag	Verification of knockout for IS5 @ *mmuP*
IS5-ynaI-P2	taccgcgacaattcattcagatcatca atagtcagggaaggaagtagcaacat tccggggatccgtcgacctg	Construction of knockout of IS5 @ *ynaI*
ynaI-ver-R	atttgtaaacagttcagcgatcattg	Verification of knockout for IS5 @ *ynaI*
IS5-yghO-P2	ccgctcatttcgctggtgtttggcgtt aagtatctcgaagcgtatgacctgat tccggggatccgtcgacctg	Construction of knockout of IS5 @ *yghO*
yghO-ver-R	agaccgtggttgttaggtgtgaagtc	Verification of knockout for IS5 @ *yghO*
IS5-yhcE-P2	tgtcagtgtgaggctaacgcttttac cgctatcaattaccacttcttgaccatt ccggggatccgtcgacctg	Construction of knockout of IS5 @ *yhcE*
yhcE-ver-R	tccaccaggaccgtcagattgcctg	Verification of knockout for IS5 @ *yhcE*
IS5-yhiS-P2	tcgggagagtgtcgtaaaggcttta acttgcttataacattgtctttggaatt ccggggatccgtcgacctg	Construction of knockout of IS5 @ *yhiS*
yhiS-ver-R	ttcgaattgatattcagacatttctg	Verification of knockout for IS5 @ *yhiS*

Table 2.2. List of oligonucleotides used, continued.

Name	Sequence	Use
IS1F-P1	ggtgatgctgccaacttactgattta gtgtatgatggtgtttttgaggtgctg tgtaggctggagctgcttc	Construction of IS1 knockout at any location
yrhD-IS1-P2	gagaaatcactaaacgaactgaata tattttctgtgccaatattatctctaatt ccgggggatccgtcgacctg	Construction of knockout of IS1 @ *yrhD*
yrhD-ver-R	tattatgaacaactgtccatgatttcg	Verification of knockout of IS1 @ *yrhD*
yihU-IS1-P2	tatggggcaggattttattgattagc ggcgataataagaacaatacccac aattccggggatccgtcgacctg	Construction of knockout of IS1 @ *yihU*
yihU-ver-R	tgattgttgatgaaggaacggcagag	Verification of knockout of IS1 @ *yihU*
nhaR-IS1-P2	cttgagacattgtttccatatgtacgc gggcgaataaatagaggaatctgat tccggggatccgtcgacctg	Construction of knockout of IS1 @ *nhaR*
nhaR-ver-R	agatagattagttgtacattaccacg	Verification of knockout of IS1 @ *nhaR*
csdH-IS1-P2	ttcacaaggatgttatactcaaagta gacaatcctgatctcaagcgtacgta ttccggggatccgtcgacctg	Construction of knockout of IS1 @ *csdH*
csdH-ver-R	tatgctcacaagtcgtatttccagag	Verification of knockout of IS1 @ *csdH*
afuB-IS1-P2	cgcagggaaccggcgcgcaggct gagtgaggcttcatcgagcgatttat cgattccggggatccgtcgacctg	Construction of knockout of IS1 @ *afuB*
afuB-ver-R	agctgtaaatcagcgccgagaggatc	Verification of knockout of IS1 @ *afuB*
argF-IS1-P2	attatgcttctaaaatggcggagaaca tggcgaggcttgctgccttacttcatt ccggggatccgtcgacctg	Construction of knockout of IS1 @ *argF*
argF-ver-R	aggattatgcttctaaaatggcggag	Verification of knockout of IS1 @ *argF*
Pnmp-T-F	acttacatcttgaaataatcacattgat tagatgaatatttatcgcgcagtgtgta ggctggagctgcttc	Construction of T1 @ *nmpC* and *nmpCp*:*km*
Pnmp-T-R	gtactgatgcagctacagcagaaa ttgccactgttaatttttcatcgtgaaa ggttcatcgcgctcgagacgca	Construction of T1 @ *nmpC*
Ptet-nmp-R	ccattagtactgatgcagctacagcag aaattgccactgttaatttttcatggtac ctttctcctctttaatgaattc	Construction of *nmpCp*:*km*
euuD-T-F	agagatggatatcgttgttaataact ctaattaatatgccaattgtttactgtg taggctggagctgcttc	Construction of T2 @ *nmpC*
euuD-T-R	actatataaaccttctgttatatattac cctttatttttgggggcgtctcaaagg ttcatcgcgctcgagacgca	Construction of T2 @ *nmpC*

Table 2.2. List of oligonucleotides used, continued.

Name	Sequence	Use
IS5nmp-P1	caaggcatttacgggagaaaaaat cggctcaaacatgaagaaatgaaat gtctgtgtaggctggagctgcttc	Construction of IS5m @ *nmpC* used in deletion and amplification of the end-IS5 @ *nmpC* region; the red "T" replaces an "A" from wildtype IS5
IS5nmp-P2	taacagaaggtttatatagttagaagc aaggttgtgcttctaaaggaagtggc atatgaatatcctccttag	Construction of IS5m @ *nmpC* used in deletion and amplification of the end-IS5 @ *nmpC* region, amplification of *km* from pKD4, and fusion PCR
nmP-km-F	caacatataagtggctccctcaagt gtgtaggctggagctgcttc	Amplification of *km* from pKD4 during IS5m @ *nmpC* construction
IS5-int-F	gctgattcaaggcatttacgggaga aaaaatcggctcaaacatgaagaa atgaaatgtc	Fusion PCR of mutated end-IS5 @ *nmpC* region with *km* from pKD4 in IS5m @ *nmpC* construction

Construction of Strains

The knockout strains (for the various IS5 and IS1 copies as well as for *nmpCp* and *nmpC*) were primarily constructed through the lambda Red procedure designed to remove a specific target gene region. In this system, the plasmid pKD46 (containing the lambda Red gene and its regulators) is transformed into the background strain through electroporation. Meanwhile, oligomer primer pairs (listed above in Table 2) are used by PCR to amplify a segment of DNA, from a template pKD4 or pKD13 plasmid, with both ends designed to be homologous to the two sides of the target region. This segment of DNA also consists of a kanamycin resistance gene (*km*) for selection, flanked by FRT regions. When the amplified DNA segment is introduced via electroporation into the background strain, the lambda-red recombinase mediates the replacement of the target region with the amplified segment of DNA, replacing the target gene or sequence with *km* (Datsenko & Wanner, 2000).

The strains inserting T1 and T2 into the *nmpC* region were constructed by the aforementioned lambda Red procedure on a pre-existing template plasmid pKDT Pu which, in addition to containing the *km* gene within the cassette, also contains T1 or T2. Likewise, the strain inserting P*tet* into the *nmpC* region was constructed by the lambda Red procedure using the pre-existing template plasmid pKDT P*tet* which contains P*tet* in the cassette (Klumpp *et al.*, 2009).

The strain with a single base pair change in the IS5 copy at *nmpC* was constructed by first selecting a region (end-IS5 @ *nmpC*) encompassing the last part of IS5 and ~200 bp of the downstream region after IS5. We then used the lambda Red knockout procedure to replace this whole region with *km,* which was removed, leaving behind FRT. Then, we used a primer with the A -> T mutation in place to amplify, via PCR, the same end-IS5 @ *nmpC* region within a wildtype background. Using fusion PCR we combined this segment with *km* from pKD4 and then transformed it into the cells, replacing the FRT scar with the usual wildtype sequence, now with the addition of the A -> T mutation and *km* gene downstream of IS5.

Construction of Compound Strains

To construct a compound strain such as a multiple deletion strain, first, we selected a donor strain with an antibiotic resistance gene (usually *km*) next to the region we wished to transfer. We then made a P1 phage while preparing a recipient strain. If the recipient strain still had *km* from the preceding procedure, transformation with pCP20 should have removed the *km* gene to leave behind FRT. We then conductd a P1 transduction into the recipient strain with the phage we had made (Thomason *et al.*, 2007). The RI-*nmpCp*:P*tet* strain was constructed by transferring RI from an existing strain into a *nmpCp*:P*tet* recipient (Kopkowski *et al.*, 2024).

bgl Mutation Assay

The *bgl* mutation assay is the primary method in this work for deducing the overall IS5 activity/transposition rate/ mutation rate of a strain (Kopkowski *et al*., 2024). The *bgl* mutation assay takes place on M9/Salicin plates, which contain 25 ml media in total at concentrations of 1.5% agar, 1x M9 and 0.5% salicin (a beta-glucoside). The plates are made starting with 1.68% water agar (to account for the future addition of appropriate amounts of 10x M9 and 5% salicin). To ensure consistency, the plates were usually made directly after the agar was autoclaved, and they were vented the next day. On the third day, strains were cultured in a 37 C shaker for 7 hours, after which the OD was measured. With the OD value, we performed dilutions with 1x M9 of the culture until we had a concentration of 10^7 cells per 100 microliters. We then plated 200 microliters, consisting of $2x10^7$ cells, onto each M9/Salicin plate and then placing these plates in a 30 C incubator.

Usually a total of 5 plates per strain were used, but occasionally contamination resulted, in which case the contaminated plate was removed from the incubator and from the study. The mutation assays were run for a total of 10 days (occasionally 12 days), but to account for the possibility of pre-existing *bgl*+ cells in the culture, we only started counting colonies from the second day after the plates began incubating, as these would be genuinely new mutations. As such, we started counting the days of a mutation assay after the second day of incubation. For example, we would label the 10^{th} day of incubation as Day 8 of the mutation assay. During the incubation period, we checked in every other day to count colonies to keep track of the total growth over time. After the mutation assay, we would then perform a PCR upstream of *bgl* to determine what kind of *bgl*+ mutant (IS1, IS5, or non-insertional) each colony was.

flhDC Mutation Assay

The *flhDC* mutation assay took place on motility plates that allowed the bacteria to swim in a noticeable way (Barker *et al*., 2004). These plates were LB agar plates with a concentration of 0.3% agar, making them partially fluid. Plates were used soon after first cooling to maintain consistency. Strains were cultured in a 37 C shaker for 7 hours before the OD was measured, and then the strains were diluted appropriately with 1x M9. Afterwards, three points of 10^7 cells each (for a total of $3x10^7$ per plate) were dotted into three separate corners of each plate. The plates were carefully placed into a 30 C incubator, checked at 18 hours, and removed at 20 hours. Then, the colonies that exhibited the swimming phenotype were counted.

fucAO Mutation Assay

The *fucAO* mutation assay took place on PPD plates, which consist of 1% propanediol (PPD) and 1x M9. As for the *bgl* assay, these plates were prepared soon after the agar had cooled and left to vent on the second day. Strains were cultured in a 37 C shaker for 7 hours before the OD was measured, and then the strains were diluted with 1x M9. Approximately 10^8 colonies were plated onto each PPD plate, which was then placed in the 30 C incubator. The assay was run for a total of 14 days with check-ins every other day to keep track of colony growth.

Statistical Analysis of Mutation Assay Data

Every mutation assay was run a total of three independent trials. Afterwards, we tallied the total number of mutants or IS5 insertions, then divided this number by the number of plates to account for assays with varying numbers of plates. This gave us an average number of mutants

or IS5 insertions per plate for each of the three assays. We could then calculate a mean and standard deviation for this set of three data points, which reflects a mean number of mutants per plate across the assays. If we wished to compare the results of two strains to determine how much of a difference there was between average IS5 insertions per plate, and thus overall IS5 activity within each strain, we used a 2-sample t-test: $t = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n_1} - \frac{s_2^2}{n_2}}}$ where in this case x is the average number of mutants per plate for a strain across all three assays, s is the standard deviation of this distribution, and n is the sample size, which will always be three for the three trials.

This should yield a p-value that would rate the significance of the difference. The typical threshold of significance is $p = 0.05$, but 2-sample t-tests are sensitive to low sample sizes, and our sample sizes were always just three. To account for this, we adjusted the threshold of significance for a study to be $p = 0.005$, putting more stringent constraints so we could be more confident when a true difference was observed. In this case, a p-value above 0.005 does not reflect a significant difference, whereas one less than this should reflect a significant difference.

RESULTS

IS5 is Unable to Jump with Mutations

As mentioned, there are 11 total copies of IS5 in strain BW25113 that we analyzed in this study. Given that we chose to study the genomic context of each region and its impact on IS5 transposition, before we began the study, we ensured that all other conditions were appropriately controlled. A primary example would ensure that each IS5 copy was identical and could transpose with the same ease as any other IS5 copy. To verify this, we sequenced all 11 copies of IS5 and obtained the results shown in Figure 3.1A. While 9 of the copies of IS5 were identical and matched the wild type IS5 sequence as expected, 2 of them, the copies at *ychE* (IS5U) and at *yoeG* (IS5H) had mutations within the coding domains of *ins5a*. While the copy at *yoeG* had only two point mutations, the copy at *ychE* had numerous mutations. Our sequencing data also confirmed the sequences given in the literature for the BW25113 strain (Figure 3.1A, Grenier *et al*., 2014).

With this baseline set, we determined if either of these strains retained the capacity to transpose. To do so, we performed a standard *bgl* mutation assay with a wild type BW25113 strain to accumulate as many *bgl*+ mutants as possible. We then performed PCR amplification of the region upstream of *bgl* to confirm IS5 insertion, and sequenced 25 of these IS5s to see if any of them possessed the mutations that would signify an origin at *ychE* or *yoeG.* Our results, shown in Figure 3.1B, revealed no such mutations, with the only insertions into *bgl* exhibiting the wild type sequence from the nine other locations. We concluded that these mutations make it unlikely for the IS5s at the *ychE* or *yoeG* loci to transpose. The implications of this for the ongoing study were that *ychE* and *yoeG* IS5s do not contribute significantly to IS5 transposition in BW25113, and so we can primarily focus our attention on the nine functional IS5s with identical sequences.

Figure 3.1. Sequencing of IS5 Elements and Insertions.
(A) Sequence of IS5U and IS5H compared with the wild type sequence that the other 9 copies of IS5 also possessed. The sequence focuses on the mutated portions of the ins5a sequence and runs in the direction of ins5a. It starts at position 347 after the main *ins5a* coding sequence begins. (B) Sequences (at position 456 after *ins5a*, with WT (wild type), IS5U, and IS5H as reference) of 25 IS5 insertions into *bgl* after a mutation assay, showing no deviations from the wild type sequence.

The IS5 Copy at *nmpC* is the Most Significant Copy for Insertion into the *bgl* Region

With this background established, we next wanted to confirm the previous results that indicated that IS5 at *nmpC* might be a significant copy for transposition. We began by analyzing the IS5 copy at *nmpC* (IS5B) on its own terms with a *bgl* mutation assay in order to determine how removing this one copy of IS5 impacts the overall activity of IS5. Firstly, we constructed three new strains from the BW25113 wild type background, each knocking out a particular section of the *nmpC* region: IS5:*km* @ *nmpC*, which knocks out only the IS5 region but leaves the rest of the *nmpC* gene, *nmpCp*:*km* and Δ*nmpCp*, which knocks out the promoter of *nmpC* but leaves the IS5 behind, and *nmpC*:*km*, which knocks out the entire region. The schematic diagram for these knockout strains is shown below in Figure 3.2A. We then tested each strain against each other in a *bgl* mutation assay, using the wild type background as a control, and hypothesized that we should observe a significance difference in the mutation and IS5 insertion rates between the wild type and IS5 @ *nmpC* knockout strain.

Figure 3.2B shows the average number of colonies and hence *bgl*+ mutants per plate over the course of the 8 days of the mutation assay. After performing PCR on these colonies upstream of *bgl*, we could count the number of IS5 insertions; Figure 3.2C shows the average number of these per plate over the course of the 8 day mutation assay. Qualitatively, the data indicate a major difference between the wild type and IS5 @ *nmpC* knockout strain, with an approximately five times difference in the average number of colonies appearing and approximately three times difference in the average number of IS5 insertions. The fact that greater than a 50% reduction in IS5 activity was observed indicates that this one copy of IS5 outperforms all eight other active IS5 copies combined.

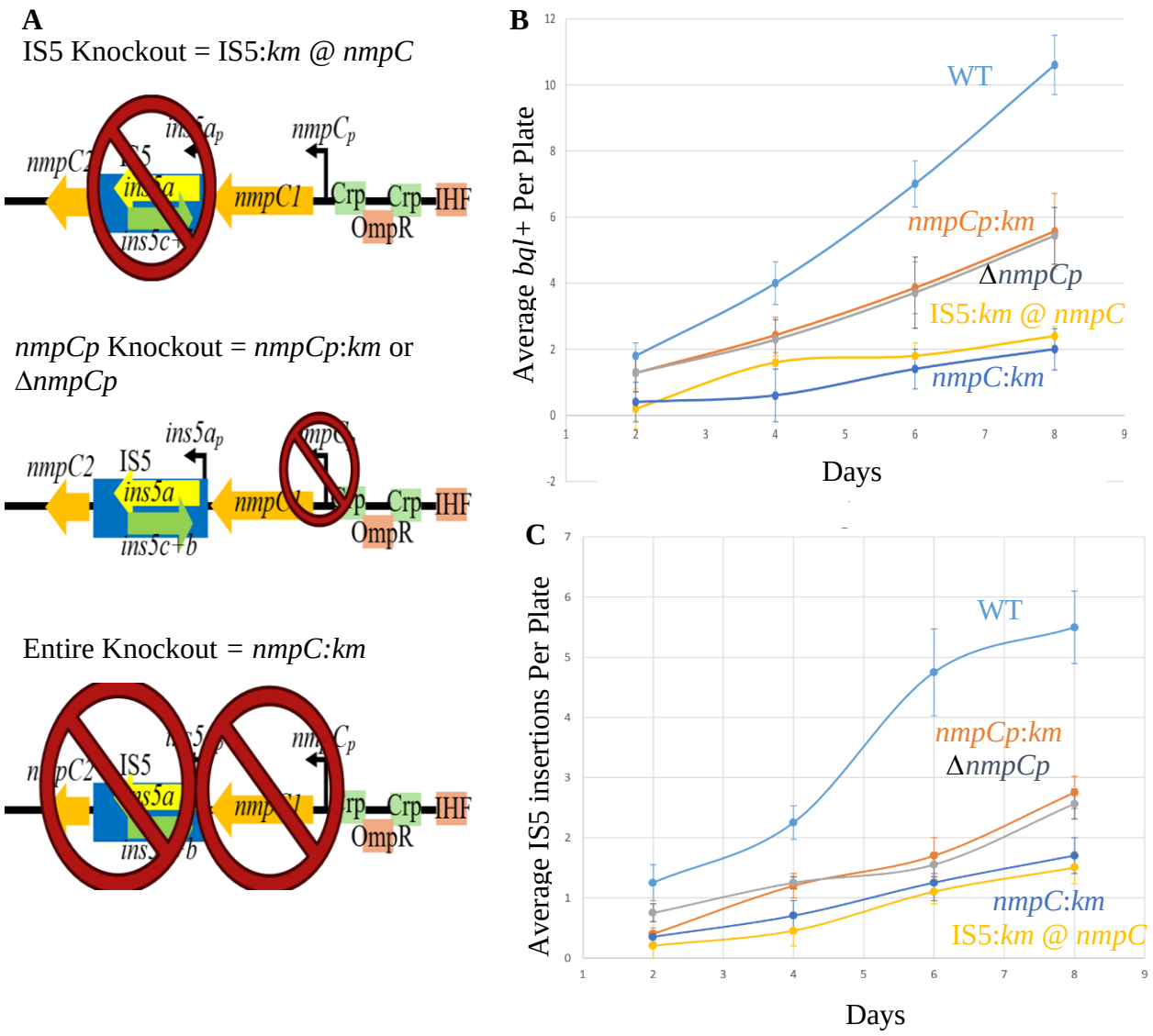


Figure 3.2. Testing the Significance of IS5 @ *nmpC*.
(A) Schematic of the construction of the knockout strains used to test the significance of the IS5 copy at *nmpC*. (B) The average number of colonies per M9/Salicin plate of $2x10^7$ cells over the course of the 8 day mutation assay, with the wild type (WT) as control compared with the IS5 copy at *nmpC* knockout strain (IS5:*km*), *nmpC* knockout strain (*nmpC*:*km*), and *nmpCp* knockout strains (*nmpCp*:*km* and Δ*nmpCp*). (C) The average number of IS5 insertions into *bgl* per plate over the 8 day period mutation assay as counted from PCR of the same colonies shown in B.

Quantitatively, when comparing the average number of *bgl*+ colonies per plate after 8 days for the wild type (10.6 ± 0.9) with the IS5 @ *nmpC* knockout strain (2.4 ± 0.3) we obtained a p-value of 0.00012; when comparing the average IS5 insertions per plate after 8 days for the

wild type (5.5 ± 0.5) with the IS5 @ *nmpC* knocked out (1.5 ± 0.3), we obtained a p-value of 0.00029. Both of these values reflect a significant difference between the mutation assay results, and thus, the overall IS5 activity, for wild type and the IS5 @ *nmpC* knockout strain, and they thus provide evidence to support the hypothesis that the IS5 copy at *nmpC* is the most significant one. Furthermore, we noted insignificant deviation between the *nmpC* and IS5 @ *nmpC* knockout strains, as well as between *nmpCp*:*km* and Δ*nmpCp* strains. This was to be expected as both were included to ensure proper validity of the mutation assays. Interestingly, there are notable deviations from the other strains exhibited by Δ*nmpCp,* but we shall investigate this more thoroughly in an upcoming section.

The IS5 Copy at *nmpC* Is a Significant Copy for *flhDC* and *fucAO Insertional Mutations*

Now that it is established the IS5 copy at *nmpC* is most important for transposition, according to *bgl* mutation assays, we wanted to establish this significance in general, and that this was not just a product of the *bgl* assay in particular. To this extent, we have performed *flhDC* and *fucAO* assays in which we compared the wild type and IS5 @ *nmpC* knockout strains, and hypothesized that there should again be a significant difference between their mutation rates.

Figure 3.3A shows the average number of swarming mutants per 0.3% LB agar plate in the *flhDC* assays, and Figure 3.3B shows the average number of PPD+ colonies per PPD plate in the *fucAO* assays. Qualitatively, the difference was observed for both to a degree of around 2-3 fold. When comparing the average *flhDC* mutants per plate after 20 hours of the wild type (7.7 ± 0.6) with the IS5 @ *nmpC* knockout strain (3.6 ± 0.4) we obtained a p-value of 0.00060. When comparing the average *flhDC* mutants per plate after 20 hours of the wild type (8.2 ± 0.5) with the IS5 @ *nmpC* knockout strain (3.4 ± 0.4) we obtained a p-value of 0.00020.

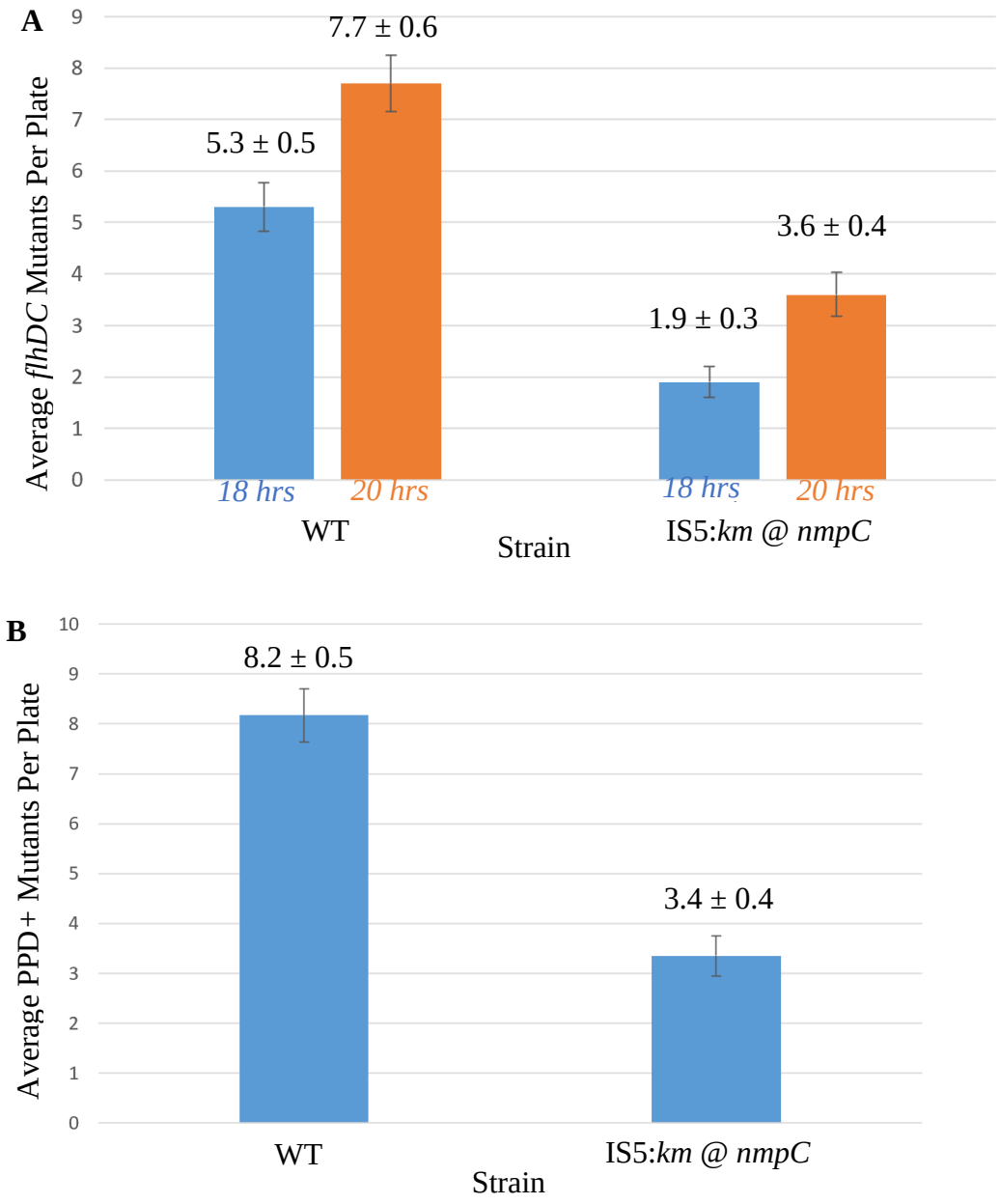


Figure 3.3. Significance of IS5 *@ nmpC* outside of *bgl.*
(A) Results of the *flhDC* assay comparing the mutation rates of wild type (WT) and IS5 *@ nmpC* knockout strain (IS5:*km @ nmpC*) by the average number of *flhDC* mutants per 0.3% LB agar plate of $3x10^7$ cells at 18 and 20 hours. (B) Results of the *fucAO* assay comparing the mutation rates of wild type (WT) and IS5 *@ nmpC* knockout strain (IS5:*km @ nmpC*) by the average number of PPD+ mutants per PPD plate of $2x10^7$ cells at 14 days.

Both results were statistically significant, providing further support for the hypothesis that *nmpC* is the primary contributor for insertion into all three gene regions. Thus, the results are

essentially the same, regardless of the test used to measure its significance. To sum up, the data indicate that removing the IS5 copy at *nmpC* has approximately a 2-3 fold effect on the overall activity of IS5 as the transposing element, which is a surprising detail given that there are not one but eight other active copies within BW25113.

IS5 Copies with Promoters are the Most Significant Copies

While we have shown that the IS5 copy at *nmpC* is most active for transposition, we have not yet answered the question as to whether it is the most significant copy of IS5 in the genome for transposition generally. To do this, we needed to compare all nine IS5 copies at once in a mutation assay and see which one deviates the most from the wild type strain and the others. We therefore constructed a total of 8 more IS5 knockout strains from the BW25113 background, one for each of the active IS5 copies other than the one at *nmpC.* We thus had the strains IS5:*km @ wbbL*, IS5:*km @ lnt*, IS5:*km @ yejO*, IS5:*km @ mmuP*, IS5:*km @ yhcE*, IS5:*km @ ynaI*, IS5:*km @ yghO*, and IS5:*km @ yhiS*. We then tested each strain against each other in a *bgl* mutation assay, using the wild type as a control, and hypothesized that the strain that produces the greatest difference from the wild type strain in terms of IS5 insertion should remain IS5:*km @ nmpC*.

In addition, recall that the *ins5a* gene of IS5 copies at *nmpC, lnt, wbbL,* and *ychE* exist alongside the same direction as a nearby promoter. If these promoters are strong or active, it stands to reason that these copies might produce more transposase than the copies that lack a nearby promoter in the same direction as their *ins5a* genes, since the latter copies should only have the weak internal promoter of IS5 to mediate transcription. Given *ychE* is an inactive copy, this leaves the copies at *wbbL* and *lnt* as potential other candidates for being the most significant copies. Thus, we also hypothesized that out of all 9 knockout strains being tested, the IS5 knockouts at *nmpC, lnt,* and *wbbL* will stand out as being the most significant.

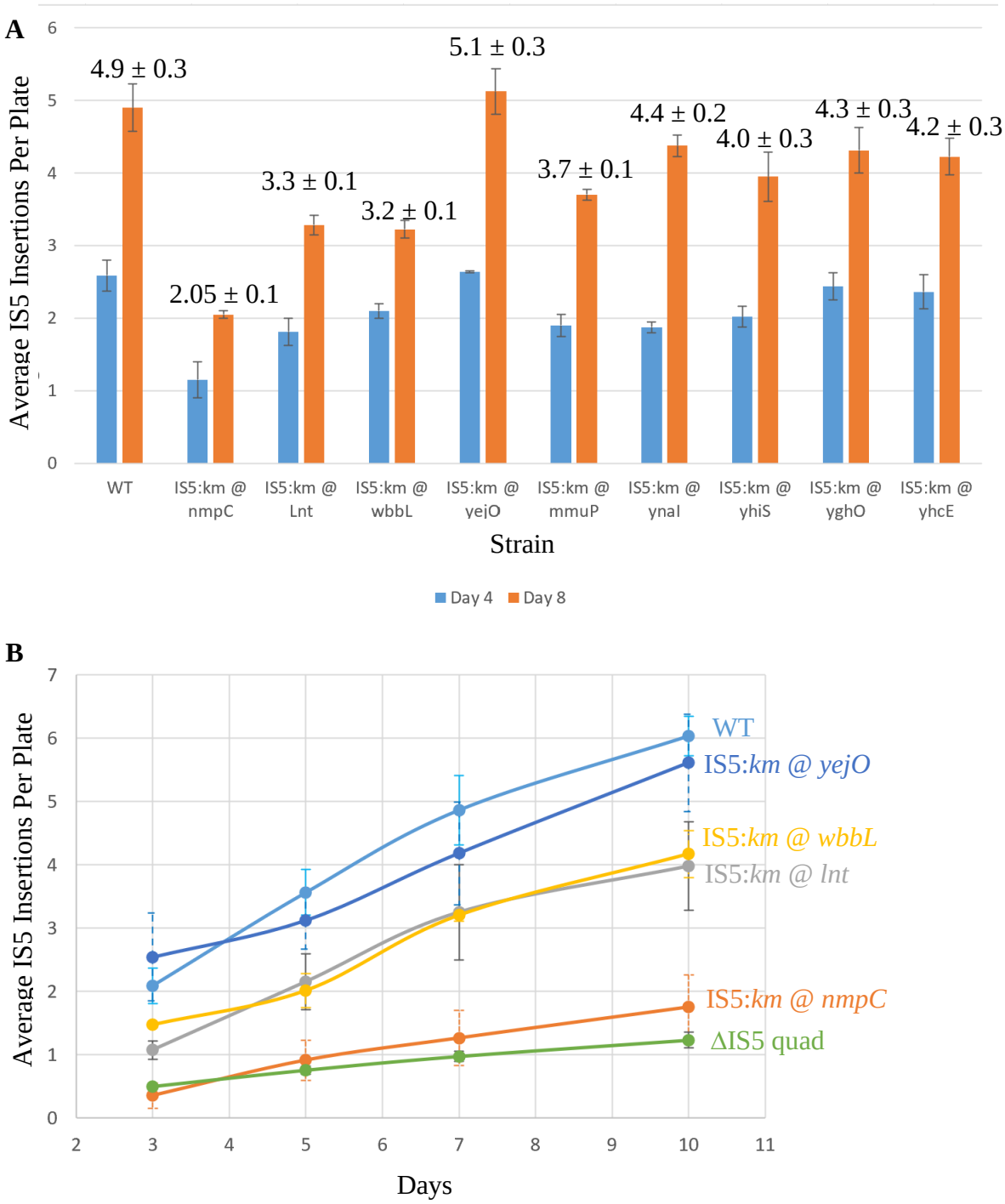


Figure 3.4. Comparison of the Significance to Transposition of all 9 Active IS5 Copies. (A) Results of *bgl* assays comparing the mutation rates of wild type (WT) and all 9 IS5 knockout strains by the average number of IS5 insertions into *bgl* per M9/Salicin plate of $2x10^7$ cells at 4 and 8 days. (B) The average number of colonies per M9/Salicin plate of $2x10^7$ cells over the course of the 10 day mutation assay, with the wild type (WT) as the control compared to the IS5 knockouts at *nmpC*, *wbbL*, *lnt*, *yejO*, and the mutant with all four knocked out (ΔIS5 quad).

Figure 3.4A shows the results of the *bgl* mutation assay comparing all 9 IS5 knockout strains, along with the wild type as a control, in terms of total IS5 insertion into *bgl* (and thus overall IS5 activity) determined via PCR. Qualitatively, it seems that the IS5 copies at *yejO, ynaI, yhiS, yghO,* and *yhcE* are the least significant as they show the smallest difference from the wild type. They are followed by *mmuP* and then *lnt* and *wbbL,* which appear at an intermediate difference from the wild type, with *lnt* and *wbbL* close together at around a 2/3 reduction from the wild type. Lastly, *nmpC* appears removed from the rest of the strains as the highest difference from the wild type, again between a 2 and 3 times reduction from the wild type. Quantitatively, we obtained the following p-values comparing the wild type against each IS5 knockout strain in terms of day 8 average number of IS5 insertions per plate: 0.00012 for IS5:*km @ nmpC,* 0.0014 for IS5:*km @ lnt,* 0.0012 for IS5:*km @ wbbL,* 0.44 for IS5:*km @ yejO,* 0.0036 for IS5:*km @ mmuP,* 0.066 for IS5:*km @ ynaI,* 0.025 for IS5:*km @ yhiS,* 0.089 for IS5:*km @ yghO,* and 0.048 for IS5:*km @ yhcE.* This leaves only IS5:*km @ mmuP, lnt, wbbL,* and *nmpC* beneath the significance threshold of p = 0.005, and so confirms the qualitative observations made above. From this we can conclude that our hypotheses stated above, that the IS5 knockout strain at *nmpC* will have the highest difference from the wild type and that it will be followed by the IS5 knockouts at *wbbL* and *lnt,* are well supported.

We next wanted to focus on the IS5 copies at the *nmpC, lnt,* and *wbbL* loci, specifically as to how they relate to each other. To do this, we first created a new strain ΔIS5 *@ nmpC, lnt, wbbL,* and *yejO,* which we will abbreviate as ΔIS5 quad. We then ran a *bgl* mutation assay comparing the individual knockouts of IS5 at *nmpC, lnt, wbbL,* and *yejO* along with ΔIS5 quad, with the usual addition of the wild type as a control. The goal was not only to determine if the IS5 knockout at *nmpC* was significantly different from the IS5 knockouts at *wbbL* and *lnt,* given

that we had already established its difference from the wild type, but also to test to what extent deleting the three extra IS5 locations compared with just removing IS5 at *nmpC.* We hypothesized that the IS5 @ *nmpC* knockout IS5 insertion rate would be significantly different from that of the IS5 @ *wbbL* or *lnt* knockouts.

Figure 3.4B shows the results of this *bgl* mutation assay. Qualitatively, the results appear consistent with the assays outlined in Figure 3.4A - the IS5 @ *lnt* and *wbbL* knockout strains once again appeared similar at approximately a 2/3 reduction in IS5 insertion compared to the wild type strain. Quantitatively, we obtained a p-value of 0.00087 for the comparison between IS5 knockout at *nmpC* and *lnt,* and likewise a p-value of 0.00074 between *nmpC* and *wbbL.* This observation indicated that there is a significant difference between removing IS5 from the *nmpC* locus versus from either the *lnt* or *wbbL locus* when considering the impact on IS5 activity. This again supports the hypothesis that the IS5 copy at *nmpC* is the most important with respect to transposition. Considering the difference between IS5:*km @ nmpC* and ΔIS5 quad, there was a 2/3-3/4 reduction when comparing the IS5:*km @ nmpC* with ΔIS5 quad. However, quantitatively we obtained a p-value of 0.034, which did not clear the significance threshold of 0.005. The lack of significance in this difference could potentially indicate that removing IS5 @ *nmpC* creates the majority of the impact on IS5 rates such that removing additional IS5s at other locations has little effect. However, it could also be due to the fact that *nmpC* already reduces the IS5 insertion count greatly, making it difficult to collect enough IS5 insertions to produce distinguishing data.

Removing IS1 Does Not Affect the Importance of IS5 @ *nmpC*

For one final test of the significance of the IS5 copy at *nmpC,* we removed all of the copies of IS1 from the genome and then ran a *bgl* mutation assay. Any insertional mutant

resulting from this assay would necessarily be due to IS5 insertion. This further isolates IS5 from any other influences (such as potential interactions with IS1) to yield a more conclusive understanding of the significance of each copy of IS5. First, we designed a new background strain where we removed all six copies of IS1 from the genome, and given there were six mutations introduced (one to knock out each IS1 copy), we refer to this strain as the IS1 sextuple deletion background, or ΔIS1 for short. In this genetic background we removed IS5 at *nmpC, lnt, wbbL,* and *yejO* individually. We then ran a usual *bgl* mutation assay, with both the usual wild type and the ΔIS1 strain, itself as controls, and hypothesized that the results would be consistent with what was observed in Figure 3.4B.

Figure 3.5 shows the results of this *bgl* mutation assay, run for 8 days with these strains, measuring the total number of insertional mutants (which in the ΔIS1 strains, were all IS5 insertions). Qualitatively, the results appear fairly consistent with the assays outlined in Figure 3.4B, when comparing the strains against their IS1 background. However, it seems the drop from ΔIS1 to ΔIS1 + IS5:*km @ nmpC, lnt,* or *wbbL* was less dramatic than what was observed between the wild type and these same knockouts in the wild type background. For instance, the p-value of the difference between ΔIS1 and ΔIS1 + IS5:*km @ lnt* is 0.009932 and for ΔIS1 + IS5:*km @ wbbL* it is 0.009773, neither of which pass statistical significance. We still obtain statistical significance with ΔIS1 + IS5:*km @ nmpC* with a p-value of 0.00045, but the reduction was a factor of less than 2 times, unlike what was observed previously. It is likely, as with the comparison of IS5:*km @ nmpC* with ΔIS5 quad (discussed above), that reducing the total number of colonies available for analysis made accurate data harder to obtain. Thus, we continued to use the wild type as our background throughout for the rest of the study, with the extra step of performing PCR to verify which mutants are due to IS5 insertions subsequently.

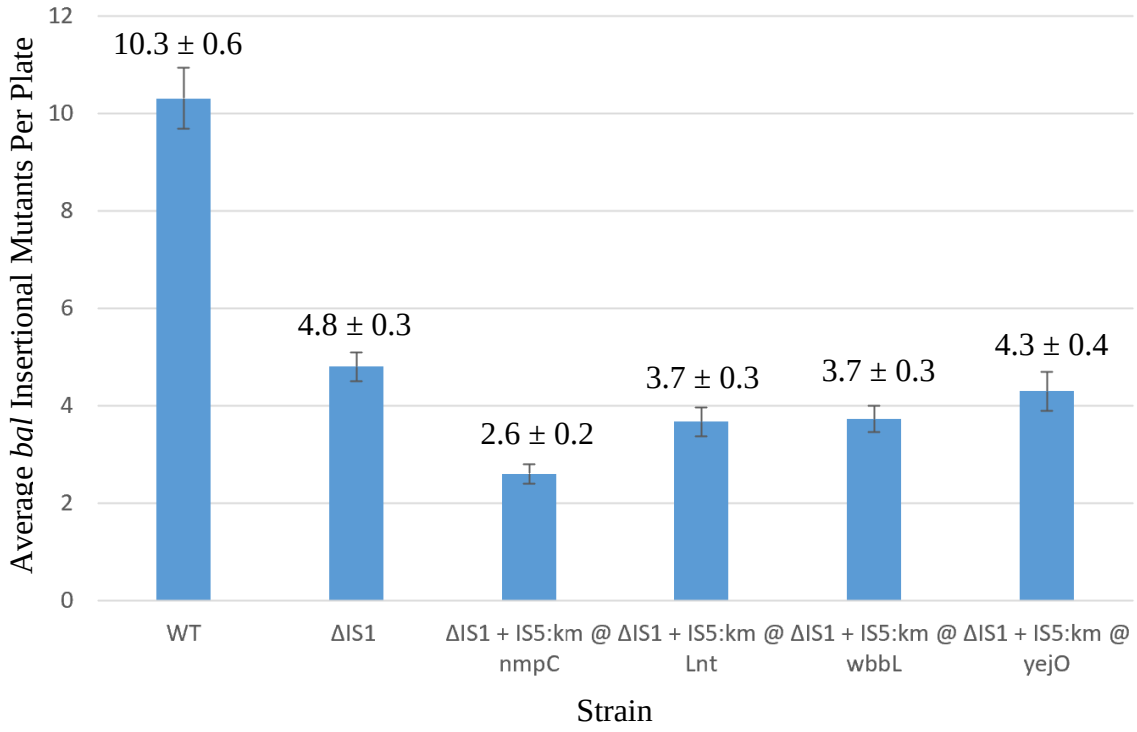


Figure 3.5. Comparison of IS5 Copies in an IS1 Deletion Background.
Results of *bgl* mutation assays comparing the mutation rates of strains by the average number of insertions into *bgl* per M9/Salicin plate of $2x10^7$ cells at 8 days. The strains to which the results were compared were the wild type strain (WT), the mutant with all six copies of IS1 removed (ΔIS1), and the deletion strains of IS5 at *nmpC, lnt, wbbL, and yejO* in this IS1 deletion background.

Inactivation of the *nmpC* Promoter Leads to a Major Reduction in IS5 Activity

To summarize what we have considered above, we now have ample evidence to support the initial hypothesis that the IS5 copy at *nmpC* is the most significant copy in its contribution to overall IS5 transposition activity. Next, we turned our attention towards answering the question of why IS5 at *nmpC*, of all 9 intact copies of IS5, is the most important. As stated before, there is a promoter (*nmpCp*) in the same direction as *ins5a*, and it could be responsible for the greatest number of IS5 transpositions being due to this promoter inducing large amounts of transposase transcription. We hypothesized that this was the primary explanation, in which case promoters, and processes affecting promoter strengths, are most relevant to transposition.

The clearest way to test this hypothesis was simply to block *nmpCp* and determine if doing so reduces the IS5 activity as much as removing IS5 itself from the *nmpC* region. To do this, we once again tested the strain previously constructed *nmpCp*:*km* which knocks out *nmpCp*. In addition to this, we constructed two new strains, T1 @ *nmpC* and T2 @ *nmpC*, which inserts the terminator T1 in front of *nmpCp* as well as the terminator T2 behind the IS5 (See Figure 3.6A). T1 is a strong terminator that should block most transcription mediated by *nmpCp*, and should do so with a cleaner and neater effect than removing *nmpCp* entirely, which could have residual effects on the nearby regions of the genome. We tested T2 behind the IS5 element in order to evaluate the significance of the directionality of the promoter relative to *ins5a*. Recall that there is a nearby promoter *quuDp* that runs in the opposite direction of *ins5a*, and T2 should block it from transcribing the IS5 element. These terminator sequences should leave the internal promoters of IS5 unaffected, so *ins5a* could still be transcribed, but at a diminished rate.

After constructing these strains, we ran a *bgl* mutation assay with all of them for 8 days, along with the usual wild type strain as control and with the IS5 knockout at *nmpC* as the necessary point of comparison. We hypothesized that addition of T1 should have an effect similar to removal of *nmpCp* entirely, and that both should result in a reduction of IS5 insertions into *bgl*, comparable to that of removing IS5 from *nmpC*. The results of this assay are summarized in Figure 3.6B, which presents the average number of IS5 insertions into *bgl* per plate for each strain tested on day 8. Qualitatively, the results appear in good agreement with the hypothesis, with a minor difference between the consequences of removal of *nmpCp* and addition of T1, as well as a minor difference between either of these and removal of IS5 at *nmpC*.

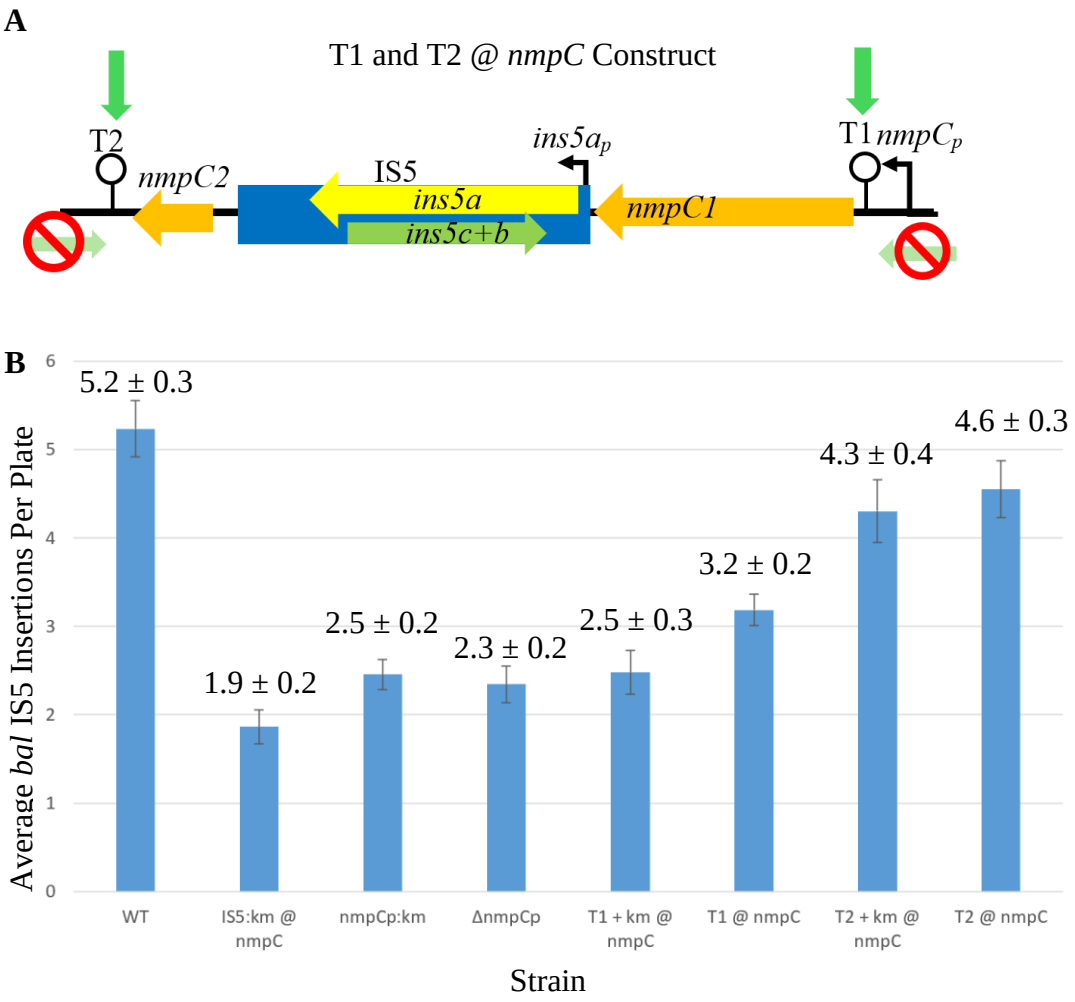


Figure 3.6. Testing the Influence of *nmpCp* on transposition of IS5 @ *nmpC* into *bgl*. (A) Schematic depiction of the use of the T1 and T2 terminators @ *nmpC* strains. (B) The average number of IS5 insertions into *bgl* per M9/Salicin plate of $2x10^7$ cells at the end of the 8 day mutation assay. The tested strains were the wild type (WT) as a control, the IS5 knock-out at *nmpC* strain (IS5:*km* @ *nmpC*), *nmpCp* replaced by *km* or removed, and the T1 and T2 addition strains, both with and without the *km* added as part of the strain construction.

When comparing *nmpCp*:*km* and Δ*nmpCp* with IS5 knocked out at *nmpC*, we obtained p-values of 0.016 and 0.041, respectively, both of which are well above the significance threshold. This suggests that there is an insufficient difference between these results to be significant, which supports the hypothesis that we should see minimal differences in the results obtained when using these two strains. The differences are similar to what was encountered when

comparing the wild type strain to some of the IS5 knockout strains for copies of IS5 away from a same-direction promoter observed in Figure 3.4A. This might suggest that the minor difference between these two could just be the result of the actions of the internal promoter of IS5 that was left intact in the constructs where *nmpCp* was inhibited.

Meanwhile, T1 + *km @ nmpC* and *nmpCp*:*km* gave overlaping results, which supports the prediction that there should be little difference between the two means of blocking the promoter. However, there was a larger deviation between T1 *@ nmpC* (without *km*) and *nmpCp*:*km*, with a p-value of 0.0064. This value is barely above significance. We speculate that this difference might be due to the large nature of *km* inserted near the *nmpC* region, acting in a disruptive way that might enhance the effects of T1 in blocking transcription. More interestingly, there was a difference observed between the wild type and the T2 insertion strains which was somewhat unexpected. When comparing the T2 + *km @ nmpC* and T2 *@ nmpC* strains with the wild type, we obtained p-values of 0.027 and 0.059, neither of which is significant.

Upstream Transcription Has a Negligible Effect on Transposition Activity

Recall that transcription in the *quuDp* strain should be blocked by T2, and also that *ins5b* and *ins5c* lie along the same direction as *quuDp.* Could the minor decrease in IS5 activity with the introduction of T2 be the result of lower expression of *ins5b* and *ins5c,* implying that they have a significant role in IS5 activity? If so, this could also have implications for understanding the genomic context of IS5 as a whole, since it means that both sides of IS5 would have to be studied to get a full picture of the relevant context. To test this possibility, we designed a new strain, T1 + T2 *@ nmpC,* which we designed by transferring T2 into a T1 *@ nmpC* background. With both T1 and T2 flanking the IS5, it is closed off to transcription from both sides. We then

ran a *bgl* mutation assay with T1 + T2 @ *nmpC,* along with the wild type as control, and IS5 knockout at *nmpC,* T1 @ *nmpC,* and T2 @ *nmpC* as primary comparisons. If the hypothesis that increased expression of *ins5b* and *ins5c* leads to greater IS5 activity is true, then the T1 + T2 @ *nmpC* strain should generate significantly fewer IS5 insertions into *bgl* than T1 or T2 @ *nmpC* alone.

The results of this assay are summarized in Figure 3.7, which shows the average number of IS5 insertions into *bgl* per plate for each strain tested on days 4 and 8. Qualitatively, the distribution of the previously tested results agrees with those shown in Figure 3.6B as expected. As for T1 + T2 @ *nmpC,* the results appear to be similar, almost identical to the results for T1 @ *nmpC* alone. Even more surprisingly, Δ*nmpCp* appeared to have lower IS5 insertion rates than either T1 @ *nmpC* or T1 + T2 @ *nmpC.* The differences between T1 @ *nmpC* and T1 + T2 @ *nmpC* are insignificant ($p = 0.53$), and when comparing T2 @ *nmpC* with the wild type we again get a high p-value of 0.16. Thus, there is almost no difference between using just T1 versus both T1 and T2. This in turn suggests that blocking *quuDp* has little impact after all on why *nmpC* is so significant, with the main cause appearing to be *nmpCp* on its own. We also failed to confirm any significant difference between wild type and T2 @ *nmpC,* so the small deviations might just be due to random chance or to minor genomic disruption induced during the construction process.

Finally, we observe a p-value of 0.011 between T1 + T2 @ *nmpC* and the IS5 knockout at *nmpC,* which is above significance, and as previously stated, the difference may be due to the internal promoter given T1 + T2 completely isolates the IS5 from any outside transcriptional effect. We shall investigate this hypothesis in more detail later.

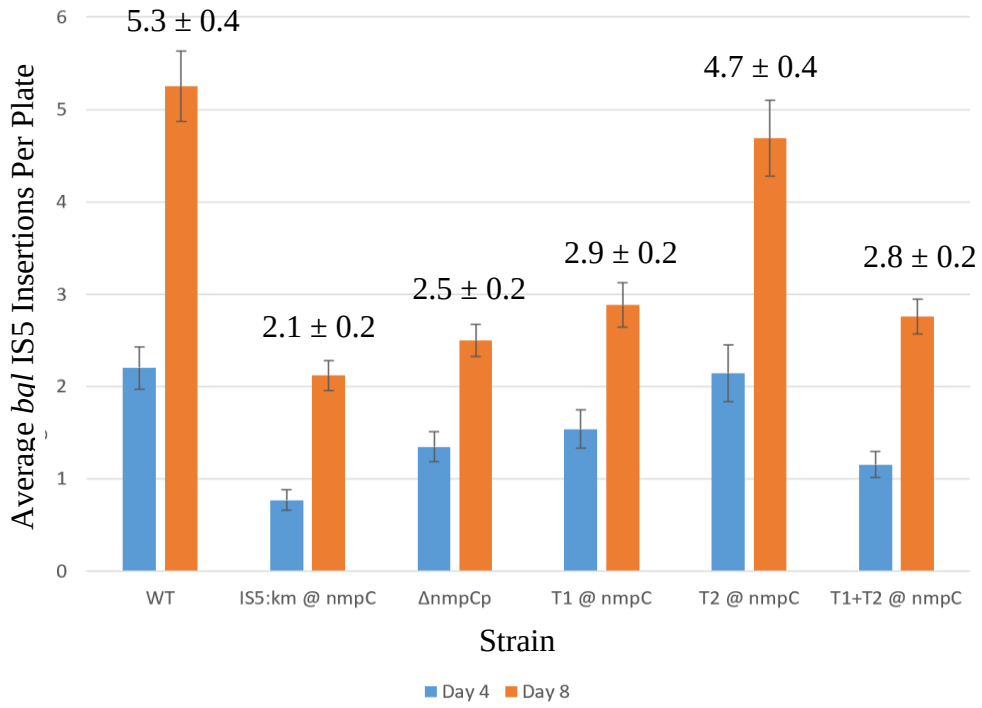


Figure 3.7. Effects of Blocking Transcription at Both Ends of IS5.
The average number of IS5 insertions at *bgl* per M9/Salicin plate with $2x10^7$ cells per plate at days 4 and 8 of the mutation assay. The tested colonies were the wild type (WT) as a control, IS5 knocked out at *nmpC* (IS5:*km @ nmpC*), *nmpCp* removed, the T1 and T2 addition strains, and lastly the T1 + T2 *@ nmpC* strain with both T1 and T2 blocking transcription at either end of IS5.

IS5 Activity Increased by Increasing the Promoter Strength

From the above data, it seemed likely that *nmpCp* is the primary driver giving rise to the significance of the copy of IS5 at *nmpC*. We wondered if there is something special about *nmpCp* that explains the significance of *nmpC*'s IS5 copy, or if any strong promoter in the same direction of *ins5a* would elevate the importance of IS5. To answer this question, we constructed a strain in which not only did we replace *nmpCp* with another strong promoter and test so see if IS5 *@ nmpC* remained the primary donor of IS5 during transposition. We varied the strength of the promoter to see if this varies the significance of IS5 *@ nmpC.* To accomplish this, we settled on P*tet,* the promoter for the *tet* operon that controls resistance to tetracycline, as it is a strong

and controllable promoter. If the gene *tetR* is introduced along with P*tet*, it will produce a *tet* repressor protein that dampens P*tet*, but if we introduce anhydrotetracycline (aTc), it will bind to the repressor and cause it to release P*tet*, so the greater the aTc concentration, the more active the promoter will be (Ehrt *et al.*, 2005; Evans & Mizrahi, 2015). This is detailed in Figure 3.8A.

We ran a mutation assay with the wild type as control and with *nmpCp* replaced by P*tet*, *nmpCp*:P*tet*. In addition, we constructed a strain RI-*nmpCp*:P*tet* that includes the *tetR* gene that repressed P*tet.* For this strain, we prepared four different M9/Salicin plates with varying concentrations of aTc added so that we could test this strain: using plates with no aTc, 5 μM aTc, 15 μM aTc, and 30 μM aTc. The RI-*nmpCp*:P*tet* cells growing on the no aTc plate should have P*tet* maximally repressed while the cells on plates with 30 μM aTc should have it mostly unrepressed. The wild type and *nmpCp*:P*tet* with no *tetR* can be plated onto typical M9/Salicin plates with no aTc added.

The results of this assay are presented in Figure 3.8B, which shows the average number of IS5 insertions into *bgl* per plate for each strain tested on day 8. Qualitatively, the results mirror what we expect, with the increase in P*tet* activity, induced by the higher concentrations of aTc, leading to higher overall IS5 insertional activity. We then analyzed the significance of the difference between the wild type (with *nmpCp*) and *nmpCp*:P*tet* strains and obtained a p-value of 0.0087. The non-significant result highlights the strength of *nmpCp* as a promoter, but the fact that we obtain a higher value for P*tet* confirms that *nmpCp* is not special as a promoter when it comes to increasing IS5 activity. The results thus suggest that one can exert major control over IS5 activity merely by controlling the promoter in the same direction as the *ins5a* of a nearby IS5 copy, and suggests that the only special feature regarding the *nmpC* location of IS5 is that *nmpCp* happens to be a very strong promoter.

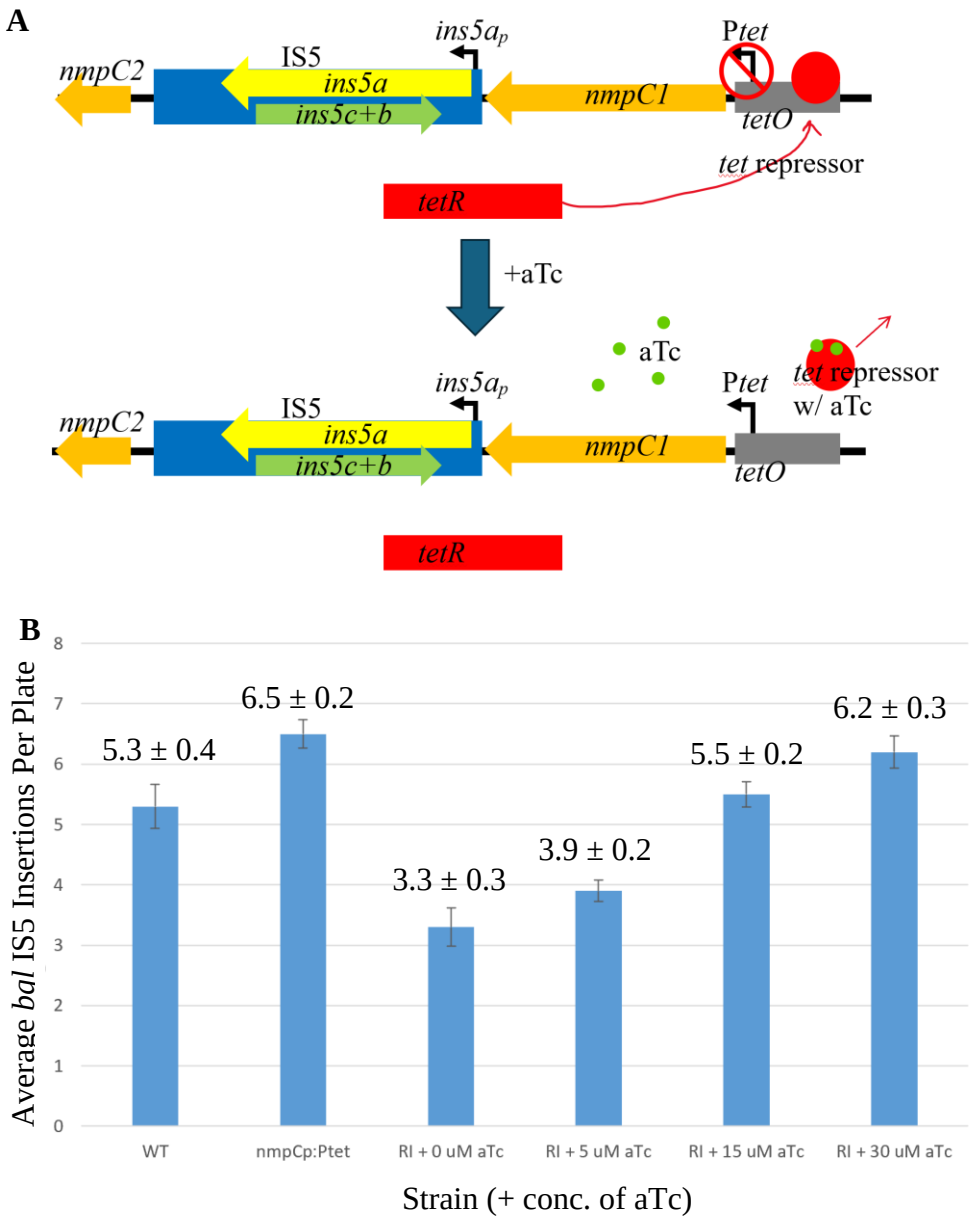


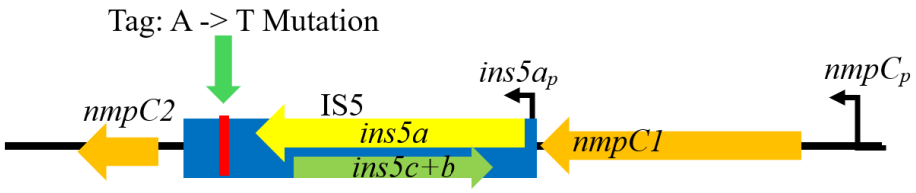
Figure 3.8. Controlling the Promoter Driving *ins5a* at *nmpC*.
(A) Schematic depiction of the mechanism by which the aTc concentration can control the strength of the P*tet* promoter, given the presence of *tetR* controlling transcription of IS5 @ *nmpC*. (B) The average number of IS5 insertions into *bgl* per M9/Salicin plate of $2x10^7$ cells at the end of the 8 day mutation assay period. The tested strains were the wild type (WT) as a control, *nmpCp* replaced by P*tet* (*nmpCp*:P*tet*), and the introduction of *tetR* into a *nmpCp*:P*tet* background (RI). This last strain was tested with four different environments: M9/Salicin plates with 0 μM aTc, 5 μM aTc, 15 μM aTc, and 30 μM aTc. The first two strains were tested on M9/Salicin plates with 0 μM aTc.

A Lack of Evidence of IS5 Transposase Bias Towards the IS5 Copy at *nmpC*

We noted (above) that there was a minor difference in IS5 activity between the strain where IS5 was removed from the *nmpC* locus and where it was blocked on either side by terminators. Could this difference be another factor influencing the significance of this copy of IS5, or is it merely a consequence of the internal promoter of IS5? If it is due to another factor, it is unlikely to be transcriptional in nature due to the transcriptionally closed environment that T1 and T2 place on the IS5 copy at *nmpC*. On the other hand, it could be a post-transcriptional and potentially post-translational factor. In order to provide evidence for our running hypothesis, that the transcriptional rate of IS5 is a direct indicator of its significance for transposition, we must rule out any factor outside of transcription that could also have an influence. The most interesting possibility is: what the transposase does after it is generated; does it go back to the copy of IS5 it was transcribed from, does it target towards a particular copy of IS5, or does it select a random copy of IS5? It is possible that the *nmpC* region possesses features that are ”attractive” towards IS5 transposase, which could bias the transposase to transpose this one copy of IS5 over the others. Thus, not only would the *nmpC* copy of IS5 produce the most transposase, it would also be transposed most frequently.

In order to test whether or not this occurs, we needed to be able to determine when the IS5 we find at *bgl* after a mutation assay has transposed from the *nmpC* region or from one of the other 8 active copies of IS5. To do this, we constructed a strain, IS5m *@ nmpC*, where we made a small base pair change in the noncoding region of IS5 *@ nmpC* to act as a tag (see Figure 3.9A). This way, we could run a mutation assay, collect colonies, and sequence the IS5 we find upstream of *bgl*. If we find the mutation, then we can be confident that the copy that transposed into *bgl* was from the *nmpC* site.

A



B

```
IS5 WT:          GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-1:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-2:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-3:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-4:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-5:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-6:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGTCTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-7:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-8:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-9:  GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-10: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-11: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-12: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-13: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-14: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-15: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-16: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-17: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGTCTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-18: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-19: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
IS5@nmpC-Mut-20: GAGAAAAAATCGGCTCAAACATGAAGAAATGAAATGACTGAGTCAGCCGAGAAGAA
```

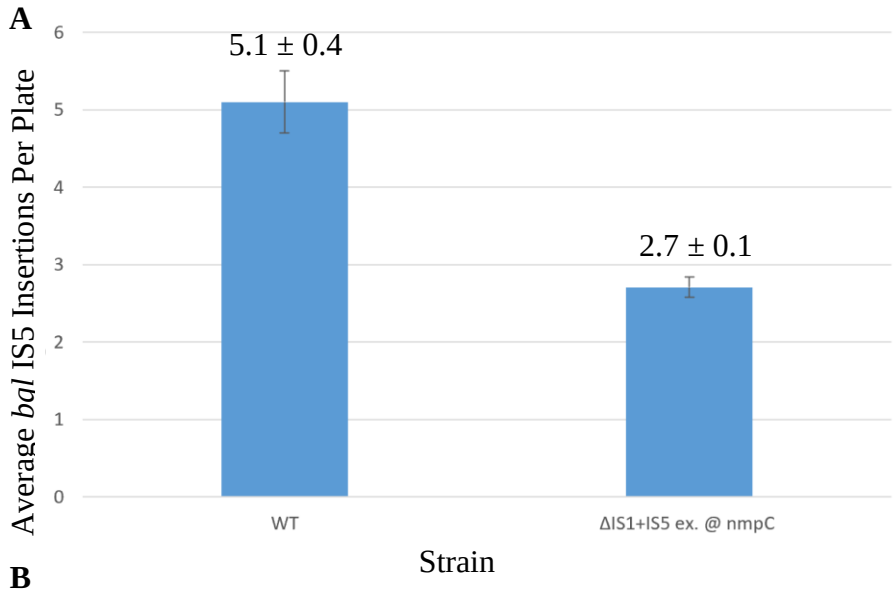
Figure 3.9. Testing Transposase Bias for IS5 @ *nmpC*.
(A) Schematic diagram of the A -> T single base pair mutation we induced into IS5 @ *nmpC* in order to "tag" it. (B) The sequences of 20 IS5 elements derived from the mutation assay with the IS5m @ *nmpC* strain. The red base is at position 1152 bp in IS5 and is the one we mutated; an A reflects the wild type sequence (and so from a copy other than at *nmpC*), while a T reflects an origin of this IS5 at *nmpC*. Only two of the 20 IS5s we sequenced originated from *nmpC*.

We selected a base at 1152 bp within IS5 (in the direction of *ins5a*), given its position after any coding region for *ins5a*, *ins5b*, or *ins5c*, but before the inverted repeat region. We wanted to avoid making any invasive mutations, given what we had observed that the two copies of IS5 with mutations are unable to jump. This particular base is usually an A, but due to the mutation we introduced, this A was changed to a T. We then ran a mutation assay and collected

many colonies, performed PCR to confirm which colonies had the IS5 insertion, and then sequenced the IS5s. The result is given in Figure 3.9B. Only 2 out of 20 sequenced IS5s turned out to be from *nmpC.* This is roughly what one would expect if there was equal probability for the transposase to select any intact copy of IS5 to transpose into *bgl,* and so this experiment provides evidence against the hypothesis that *nmpC* has any features that would bias a transposase favorably. However, this result does not rule out the possibility that other copies of IS5 have such an attractive trait since we only tagged the copy at *nmpC.* We suggest that this influence would need to be minimal in comparison to the established influence of nearby promoters to remain compatible with the data presented in Figure 3.4A.

The Dominant Transposition Mechanism of IS5 out of *nmpC* is "Copy and Paste"

As a last confirmation of our results, we wanted to complement our first knockout experiment of IS5 @ *nmpC* by performing the reverse, and knocking out all IS1 and IS5 copies other than IS5 @ *nmpC.* We constructed a strain starting from the ΔIS1 genetic background where we had deleted all 8 other active copies of IS5, leaving only the IS5 copy at *nmpC.* We refer to this strain as ΔIS1+IS5 ex. @ *nmpC.* We ran a mutation assay comparing the results to the wild type as a control. In addition, we realized that with a strain with only one copy of IS5, we could ascertain the mechanism of the IS5 copy at *nmpC* as to whether it was copy and paste or cut and paste. Thus, after the mutation assay, we used PCR to identify all insertional mutants from ΔIS1+IS5 ex. @ *nmpC* and then performed another PCR, this time at the *nmpC* locus. If a band appeared, it would signify that there was both an IS5 copy there and at *bgl* and signal a copy and paste mechanism of action; if the band did not appear, it would signify that the IS5 had jumped to *bgl* from the *nmpC* locus in a cut and paste mechanism.



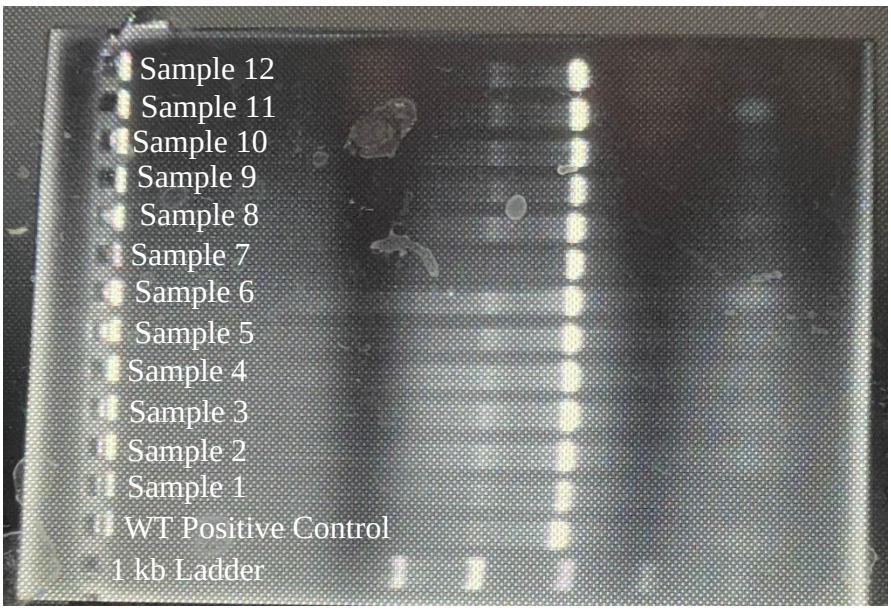


Figure 3.10. Analysis of a Strain with Only One IS5 @ *nmpC*.
(A) The average number of IS5 insertions into *bgl* per M9/Salicin plate of $2x10^7$ cells at the end of the 8 day mutation assay period. The tested strains were the wild type (WT) as a control and the strain where the only IS element capable of transposition to *bgl* is the IS5 copy at *nmpC* (ΔIS1+IS5 ex. @ *nmpC*) (B) A gel showing PCR products amplified from primers targeting both *nmpC* and IS5 such that a band will only show if IS5 is present at *nmpC.* A 1 kb ladder is shown along with a positive control wild type, with the rest being ΔIS1+IS5 ex. @ *nmpC* colonies that also have IS5 inserted into *bgl.*

The results of this assay are summarized by Figure 3.10A, which shows the average amount of IS5 insertions into *bgl* per plate for each strain tested on day 8. It shows a significant ($p = 0.00060$) drop in IS5 insertions between the wild type and ΔIS1+IS5 ex. @ *nmpC*, which is to be expected when removing almost all of the copies of IS5 from the genome. Still, the loss is less than a 2-fold reduction, which confirms our observation from Figure 3.2C that the IS5 copy at *nmpC* accounts for more than half of the total IS5 activity. Meanwhile, PCR analysis, summarized in Figure 3.10B, shows that in a sample of 12 colonies of ΔIS1+IS5 ex. @ *nmpC* with IS5 insertion into *bgl*, all 12 also possessed IS5 at the *nmpC* location. Since the IS5 could only have come from *nmpC* given that the other copies were deleted, this result provides strong evidence that IS5, at least at the *nmpC* location, operates with copy and paste being the dominant transposition mechanism.

DISCUSSION

The IS5 Copy at *nmpC* is the Most Significant Copy in BW25113 for Transposition

Previous research identified the possibility that the IS5 copy at *nmpC* was the most significant contributor to the production of transposase mRNA (Zhou, 2023). From this suggestion, we hypothesized that IS5 at *nmpC* might be the most significant copy of IS5 in terms of its contribution to the overall IS5 activity, which can be measured through its insertion rates into operons such as *bgl* during mutation assays. The data overall seem to confirm this hypothesis from multiple different perspectives. Firstly, we noted a persistent 2-3 times drop in overall IS5 activity when this particular copy of IS5 was removed (Figure 3.2C), and that this drop could be recorded across *bgl*, *flhDC*, or *fucAO* mutation assays (Figure 3.3). We also saw that the *nmpC* copy could stand on its own with an IS5 activity level more than half that of the wild type when all other IS5 copies were removed (Figure 3.10A). This suggested that the *nmpC* copy contributes more to the overall activity of IS5 than all other 8 active IS5 copies combined.

While this result agrees with the hypothesis and stands in good relation to the previous research, there is a discrepancy that could provide an opportunity for future research. In the previous study, the authors noted that the vast majority of the mRNA they collected originated from *nmpC*, such that mRNA from other copies was not fully visible. While IS5 at *nmpC* is undoubtedly the most significant copy of IS5, it is not the only significant copy, as that would exclude the influence we observed from IS5 @ *lnt* and *wbbL* (Figure 3.4). In addition, leaving IS5 @ *nmpC* by itself proved to be insufficient to recapture wild type IS5 activity (Figure 3.10A), giving further evidence transposase transcription is not solely conducted by IS5 at *nmpC*. It thus remains to explain this discrepancy, as it might point to a new factor not yet considered.

Initial Location of Transposons Should be a Major Consideration

Many past studies have focused primarily on where transposons such as IS5 may insert and how the genomic context of those regions will affect transposition into them (Reynolds *et al*., 1981; Chen *et al*., 1989; Whiteway *et al*., 1998; Barker *et al*., 2004; Zhang & Saier, 2009b; Humayun *et al*., 2017). This study aimed to fill in the important gap of considering the genomic context not of possible targets, but of the initial locations of IS5, and how this could have an impact on transposition and thus, on transposon-mediated regulation. Our work revealed that the genomic context of the initial location is just as important to consider as the context of the possible target. If the genomic context of the initial location were not relevant, we would have expected to see each copy of IS5 contribute equally to the overall transposition. However, as we can clearly see in Figure 3.4A, there is a wide diversity in the contribution of each copy, with some (such as *nmpC*) contributing heavily while others (e.g., *yhiS*) exhibiting a minimal contribution. In addition, we saw how when *nmpC* was removed, it resulted in a 2-3 fold drop in transposition overall, and that this result held for several operons IS5 can insert into (Figure 3.3). Thus, studying and manipulating the genomic context of IS5's initial location has a major and global impact on transposition rate and transposon-mediated regulation that applies regardless of which insertion location one might be focusing on.

Post-Transcriptional Impact on Transposition Activity of IS5 @ *nmpC* Negligible

One central investigation on the role genomic context plays in IS5 activity was whether or not the significance of an IS5 copy was due to transcription alone, or if post-transcriptional influences existed. For instance, we hypothesized that there might be something about the *nmpC* copy that biases transposases to act upon it to the exclusion of other copies. We

tested this by marking the *nmpC* copy with a mutation. When sequencing the IS5 that had jumped into *bgl* after a mutation assay, we found no more IS5 originating from *nmpC* than would be expected if every active copy had an equal chance to jump (Figure 3.9). This suggests that, at least for the IS5 copy at *nmpC,* there is no post-transcriptional cause for its significance. These results do not completely rule out post-transcriptional effects; an interesting follow-up study could involve marking each copy of IS5 with a unique identifier and running multiple different types of mutation assays. We could then determine if the distribution is skewed in any way towards one particular copy, and if this skew might depend on the target of that IS5.

Interestingly, the results from Figure 3.9B and from Figure 3.1 seem to be in tension- if a transposase from any copy of IS5 can mediate transposition of any other copy of IS5, then the inability for the two mutant copies of IS5 to produce transposase should not prevent their transposition entirely as was observed in Figure 3.1. We hypothesize that the ability for IS5 to transpose could be very sensitive to mutations, even those outside the inverted repeats. In this case, the carefully selected A -> T mutation we induced might have been non-invasive enough to still permit transposition, while the mutations found in the two mutant copies might have been too disruptive. That said, the cause of this discrepancy would need to be resolved with future research. For instance, we could perform an experiment where we mutate various regions of IS5 to various degrees to identify how each mutation impacts its ability to transpose.

The Promoter is the Most Important Part of the IS5 Initial Location Genomic Context

Of course, once we established that *nmpC* was the most significant location of IS5 for transposition, and showed that studying the initial location genomic context was important, we wanted to know both why *nmpC* was the most significant, and what aspect of the initial

location's genomic context led to the importance of its consideration. The data indicate that the answers to both of these questions converge towards promoters. Firstly, when the strong *nmpCp* was removed or inhibited, it resulted in a major loss of transposition activity, even though the IS5 copy was still present (Figure 3.2C). In addition, if the strains where *nmpCp* had been inhibited were the wild type, the copy of IS5 at *nmpC* would no longer be considered the most significant copy (Figure 3.6B). Removing the promoter seems to "demote" the IS5 copy at *nmpC* almost fully to the level of some of the less significant copies of IS5, such as those at the *yhiS* or *yghO* loci (Figure 3.7). Furthermore, we determined that the significance of the IS5 copy at *nmpC* can be easily replicated so long as there is a strong promoter pointing in the same direction as *ins5a*, which dispels the possibility that *nmpCp* might have an intrinsic ability to increase the nearby IS5 copy's significance (Figure 3.8). Instead, it seems clear that the IS5 copy at *nmpC* is of primary significance due to the fact that a strong promoter happens to be nearby and pointing in the same direction as *ins5a.*

An interesting investigation was into what role *ins5b* and *ins5c* play in transposition, and what their function(s) might be. This question happened to neatly overlap with the ongoing investigation of the impact of genomic context on IS5 transposition, as the genomic context might allow for greater expression of these two mysterious genes, allowing us to evaluate what impact this greater expression has on overall IS5 transposition activity. There were also some interesting signs that *ins5b/c* might play a role as to why IS5 at *nmpC* was so significant, since this was the only copy of IS5 in the genome with a promoter both in the direction of *ins5a* and in the direction of *ins5b/c* (Huerta & Collado-Vides, 2003). However, when we completely isolated IS5 at *nmpC* with respect to transcription from both directions, the effect on overall IS5 activity was about the same as if we had only blocked transcription along the direction of *ins5a* (Figure

3.7). This fact suggests that it is transposase production that seems to have the greatest impact on overall IS5 transposition. While the potentially greater transcription of *ins5b/c* at *nmpC* might not have increased IS5 activity or been partially responsible for the significance of this IS5 copy, it does not completely rule out the possibility that these proteins might play some role. This is because the internal promoters of IS5, while weak, do create trace amounts of these proteins for any copy of IS5 (Schoner & Kahn, 1981; Kröger & Hobom, 1982). Further research, such as potentially investigating the effect of blocking these internal promoters, could potentially resolve the question of whether or not these proteins play a role in transposition.

Overall, our results suggest that the primary factor behind why IS5 at *nmpC* is so important, and potentially what could make any transposon copy transpositionally important, is the amount of transposase produced. This in turn may directly be affected by the presence and strength of promoters in the same direction as the transposase gene. Our current findings may be relevant to transposon research in general, as it suggests the simplest way to increase or moderate transposition is by placing stronger and controllable promoters in front of various copies of the target transposon.

Environmental Impacts on Transposition Rate and Initial Location Genomic Context

The last important consideration we wanted to consider was about how the environment might impact transposition rate depending on the initial location's genomic context. This would put our work more in parallel with the efforts to determine how the environment impacts the ability of IS5 to insert into a particular target (Humayun *et al*., 2017; Zhang *et al*., 2017; Saier *et al*., 2017; Zhang *et al*., 2024; Kopkowski *et al*., 2024). We hypothesized that varying osmolarity and pH might have an impact on the transcriptional and transpositional rates given that *nmpCp* is

itself controlled by OmpR, which is responsive to these environmental conditions. We started by setting up a standard mutation assay where we plate a wild type strain onto M9/Salicin plates with differing levels of osmolarity or pH, but soon it became clear that this experiment was probably not feasible. This is because *E. coli* is very sensitive to differing levels of osmolarity and pH, and will not grow as well under non-optimal conditions. Thus, for instance, it would be difficult to tell whether a reduction in the number of *bgl*+ colonies and IS5 insertions into *bgl* is due to less IS5 activity, or to slower growing conditions due to the non-optimal environment. Nevertheless, we hope to return to this experiment in the future, as we consider it an important next step in establishing the value of the initial transposon location to the analysis of transposon-mediated regulation and directed mutation. It may also be relevant to the question as to why the copy of IS5 within *nmpC* has been retained over a substantial period of evolutionary time.

The information contained within this thesis, in full, is being prepared for publication: Onstead, Jonathan; Zhang, Zhongge; Saier, Milton H., Jr. The thesis author was the primary researcher and author of this paper.

REFERENCES

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology, 2*, 2006.0008. https://doi.org/10.1038/msb4100050

Barker, C. S., Prüss, B. M., & Matsumura, P. (2004). Increased motility of Escherichia coli by insertion sequence element integration into the regulatory region of the flhD operon. *Journal of bacteriology, 186*(22), 7529–7537. https://doi.org/10.1128/JB.186.22.7529-7537.2004

Beloin, C., Valle, J., Latour-Lambert, P., Faure, P., Kzreminski, M., Balestrino, D., Haagensen, J. A., Molin, S., Prensier, G., Arbeille, B., & Ghigo, J. M. (2004). Global impact of mature biofilm lifestyle on Escherichia coli K-12 gene expression. *Molecular microbiology, 51*(3), 659–674. https://doi.org/10.1046/j.1365-2958.2003.03865.x

Berg, D. E., & Howe, M. M. (1989). *Mobile DNA*. American Society for Microbiology.

Blasband, A. J., Marcotte, W. R., Jr, & Schnaitman, C. A. (1986). Structure of the lc and nmpC outer membrane porin protein genes of lambdoid bacteriophage. *The Journal of biological chemistry, 261*(27), 12723–12732.

Chen, Y. M., Lu, Z., & Lin, E. C. (1989). Constitutive activation of the fucAO operon and silencing of the divergently transcribed fucPIK operon by an IS5 element in Escherichia coli mutants selected for growth on L-1,2-propanediol. *Journal of bacteriology, 171*(11), 6097–6105. https://doi.org/10.1128/jb.171.11.6097-6105.1989

Coll, J. L., Heyde, M., & Portalier, R. (1994). Expression of the nmpC gene of Escherichia coli K-12 is modulated by external pH. Identification of cis-acting regulatory sequences involved in this regulation. *Molecular microbiology, 12*(1), 83–93. https://doi.org/10.1111/j.1365-2958.1994.tb00997.x

Daley, D. O., Rapp, M., Granseth, E., Melén, K., Drew, D., & von Heijne, G. (2005). Global topology analysis of the Escherichia coli inner membrane proteome. *Science (New York, N.Y.), 308*(5726), 1321–1323. https://doi.org/10.1126/science.1109730

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America, 97*(12), 6640–6645. https://doi.org/10.1073/pnas.120163297

Dole, S., Nagarajavel, V., & Schnetz, K. (2004). The histone-like nucleoid structuring protein H-NS represses the Escherichia coli bgl operon downstream of the promoter. *Molecular microbiology, 52*(2), 589–600. https://doi.org/10.1111/j.1365-2958.2004.04001.x

Edwards, M. D., Black, S., Rasmussen, T., Rasmussen, A., Stokes, N. R., Stephen, T. L., Miller, S., & Booth, I. R. (2012). Characterization of three novel mechanosensitive channel activities in Escherichia coli. *Channels (Austin, Tex.), 6*(4), 272–281. https://doi.org/10.4161/chan.20998

Ehrt, S., Guo, X.V., Hickey, C.M., Ryou, M., Monteleone, M., Riley, L.W., & Dirk, S. (2005). Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Research, 33*(2), e21. https://doi.org/10.1093/nar/gni013

Evans, J. C., & Mizrahi, V. (2015). The application of tetracyclineregulated gene expression systems in the validation of novel drug targets in Mycobacterium tuberculosis. *Frontiers in microbiology, 6*, 812. https://doi.org/10.3389/fmicb.2015.00812

Gerdes S.Y., Scholle M.D., Campbell J.W., Balázsi G., Ravasz E., Daugherty M.D., Somera A.L ., Kyrpides N.C., Anderson I., Gelfand M.S., Bhattacharya A., Kapatral V., D'Souza M., Baev M.V., Grechkin Y., Mseeh F., Fonstein M.Y., Overbeek R., Barabási A., Oltvai Z.N ., & Osterman A.L. (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *Journal of bacteriology, 185*(19), 5673–5684. https://doi.org/10.1128/JB.185.19.5673-5684.2003

Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., Lund, P. A., Cole, J. A., & Henderson, I. R. (2018). The Essential Genome of Escherichia coli K-12. *mBio, 9*(1), e02096-17. https://doi.org/10.1128/mBio.02096-17

Grenier F, Matteau D, Baby V, Rodrigue S. (2014). Complete Genome Sequence of Escherichia coli BW25113. *Genome Announc 2*(5):genomea.01038-14. https://doi.org/10.1128/genomea.01038-14

Gupta, S. D., & Wu, H. C. (1991). Identification and subcellular localization of apolipoprotein N-acyltransferase in Escherichia coli. *FEMS microbiology letters, 62*(1), 37–41. https://doi.org/10.1016/0378-1097(91)90251-5

Hall B. G. (1998). Activation of the bgl operon by adaptive mutation. *Molecular biology and evolution, 15*(1), 1–5. https://doi.org/10.1093/oxfordjournals.molbev.a025842

Henderson, I. R., & Owen, P. (1999). The major phase-variable outer membrane protein of Escherichia coli structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and oxyR. *Journal of bacteriology, 181*(7), 2132–2141. https://doi.org/10.1128/JB.181.7.2132-2141.1999

Highton PJ, Chang Y, Marcotte WR Jr, Schnaitman CA. (1985). Evidence that the outer membrane protein gene nmpC of Escherichia coli K-12 lies within the defective qsr' prophage. *Journal of Bacteriology, 162*(1), 256-262. https://doi.org/10.1128/jb.162.1.256-262.1985

Hindahl, M. S., Crockford, G. W., Hancock, R. E. (1984). Outer membrane protein NmpC of Escherichia coli: pore-forming properties in black lipid bilayers. *Journal of bacteriology, 159*(3), 1053–1055. https://doi.org/10.1128/jb.159.3.1053-1055.1984

Homma, K., Fukuchi, S., Kawabata, T., Ota, M., & Nishikawa, K. (2002). A systematic investigation identifies a significant number of probable pseudogenes in the Escherichia coli genome. *Gene, 294*(1-2), 25–33. https://doi.org/10.1016/s0378-1119(02)00794-1

Huerta, A. M., & Collado-Vides, J. (2003). Sigma70 promoters in Escherichia coli: specific transcription in dense regions of overlapping promoter-like signals. *Journal of molecular biology, 333*(2), 261–278. https://doi.org/10.1016/j.jmb.2003.07.017

Humayun M.Z., Zhang Z., Butcher A.M., Moshayedi A., Saier M.H. (2017). Hopping into a hot seat: Role of DNA structural features on IS5-mediated gene activation and inactivation under stress. *PLoS ONE, 12*(6):e0180156. https://doi.org/10.1371/journal.pone.0180156

Jaurin, B., & Normark, S. (1983). Insertion of IS2 creates a novel ampC promoter in Escherichia coli. *Cell, 32*(3), 809–816. https://doi.org/10.1016/0092-8674(83)90067-3

Klumpp, S., Zhang, Z., & Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. *Cell, 139*(7), 1366–1375. https://doi.org/10.1016/j.cell.2009.12.001

Kopkowski, P. W., Zhang, Z., Saier, M. H. (2024). The effect of DNA-binding proteins on insertion sequence element transposition upstream of the bgl operon in Escherichia coli. *Frontiers in Microbiology, 15*. https://doi.org/10.3389/fmicb.2024.1388522

Korea, C. G., Badouraly, R., Prevost, M. C., Ghigo, J. M., & Beloin, C. (2010). Escherichia coli K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environmental microbiology, 12*(7), 1957–1977. https://doi.org/10.1111/j.1462-2920.2010.02202.x

Kröger, M., Hobom, G. (1982). Structural analysis of insertion sequence IS5. *Nature 297*, 159–162. https://doi.org/10.1038/297159a0

Lam, K. J. K., Zhang, Z., & Saier, M. H. (2022). Histone-like nucleoid structuring (H-NS) protein silences the beta-glucoside (bgl) utilization operon in Escherichia coli by forming a DNA loop. *Computational and structural biotechnology journal, 20*, 6287–6301. https://doi.org/10.1016/j.csbj.2022.11.027

Mahillon, J., & Chandler, M. (1998). Insertion sequences. *Microbiology and molecular biology reviews : MMBR, 62*(3), 725–774. https://doi.org/10.1128/MMBR.62.3.725-774.1998

May, E.W., Craig, N.L. (1996). Switching from Cut-and-Paste to Replicative Tn7 Transposition. *Science, 272*(5260), 401-404 https://doi.org/10.1126/science.272.5260.401

McClintock, B. (1950). The Origin and Behavior of Mutable Loci in Maize. *Proceedings of the National Academy of Sciences, 36*(6), 344-355. https://doi.org/10.1073/pnas.36.6.344

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juárez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J., & Morett, E. (2009). Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. *PloS one, 4*(10), e7526. https://doi.org/10.1371/journal.pone.0007526

Prasad, I., & Schaefler, S. (1974). Regulation of the beta-glucoside system in Escherchia coli K-12. *Journal of bacteriology, 120*(2), 638–650. https://doi.org/10.1128/jb.120.2.638-650.1974

Pugsley, A. P., & Schnaitman, C. A. (1978). Identification of three genes controlling production of new outer membrane pore proteins in Escherichia coli K-12. *Journal of bacteriology, 135*(3), 1118–1129. https://doi.org/10.1128/jb.135.3.1118-1129.1978

Rak, B., & von Reutern, M. (1984). Insertion element IS5 contains a third gene. *The EMBO journal, 3*(4), 807–811. https://doi.org/10.1002/j.1460-2075.1984.tb01889.x

Reynolds, A. E., Felton, J., & Wright, A. (1981). Insertion of DNA activates the cryptic bgl operon in E. coli K12. *Nature, 293*(5834), 625–629. https://doi.org/10.1038/293625a0

Robichon, C., Vidal-Ingigliardi, D., & Pugsley, A. P. (2005). Depletion of apolipoprotein N-acyltransferase causes mislocalization of outer membrane lipoproteins in Escherichia coli. The *Journal of biological chemistry, 280*(2), 974–983. https://doi.org/10.1074/jbc.M411059200

Rubirés, X., Saigi, F., Piqué, N., Climent, N., Merino, S., Albertí, S., Tomás, J. M., & Regué, M. (1997). A gene (wbbL) from Serratia marcescens N28b (O4) complements the rfb-50 mutation of Escherichia coli K-12 derivatives. *Journal of bacteriology, 179*(23), 7581–7586. https://doi.org/10.1128/jb.179.23.7581-7586.1997

Saier, M. H., Jr, Kukita, C., & Zhang, Z. (2017). Transposon-mediated directed mutation in bacteria and eukaryotes. *Frontiers in bioscience (Landmark edition), 22*(9), 1458–1468. https://doi.org/10.2741/4553

Sawers, G.S. (2005). Transcript analysis of Escherichia coli K-12 insertion element IS5. *FEMS Microbiology Letters, 244*(2) 397–401, https://doi.org/10.1016/j.femsle.2005.02.019

Schnetz K. (1995). Silencing of Escherichia coli bgl promoter by flanking sequence elements. *The EMBO journal, 14*(11), 2545–2550. https://doi.org/10.1002/j.1460-2075.1995.tb07252.x

Schnetz, K., & Rak, B. (1988). Regulation of the bgl operon of Escherichia coli by transcriptional antitermination. *The EMBO journal, 7*(10), 3271–3277. https://doi.org/10.1002/j.1460-2075.1988.tb03194.x

Schnetz K. & Rak B. (1992). IS5: a mobile enhancer of transcription in Escherichia coli. *Proc. Natl. Acad. Sci. USA. 89*(4), 1244–1248. https://doi.org/10.1073/pnas.89.4.1244

Schoner, B., & Kahn, M. (1981). The nucleotide sequence of IS5 from Escherichia coli. *Gene, 14*(3), 165–174. https://doi.org/10.1016/0378-1119(81)90112-8

Thanbichler, M., Neuhierl, B., & Böck, A. (1999). S-methylmethionine metabolism in Escherichia coli. *Journal of bacteriology, 181*(2), 662–665. https://doi.org/10.1128/JB.181.2.662-665.1999

Thomason, L. C., Costantino, N., & Court, D. L. (2007). E. coli genome manipulation by P1 transduction. *Current protocols in molecular biology*, *Chapter 1*, 1.17.1–1.17.8. https://doi.org/10.1002/0471142727.mb0117s79

Tran, D., Zhang, Z., Lam, K. J. K., & Saier, M. H., Jr (2022). Effects of Global and Specific DNA-Binding Proteins on Transcriptional Regulation of the E. coli bgl Operon. *International journal of molecular sciences, 23*(18), 10343. https://doi.org/10.3390/ijms231810343

Wang, S., Fleming, R.T., Westbrook, E.M., Matsumura, P., & McKay, D.B. (2006). Structure of the Escherichia coli FlhDC Complex, a Prokaryotic Heteromeric Regulator of Transcription. *Journal of Molecular Biology 355*(4), 798-808. https://doi.org/10.1016/j.jmb.2005.11.020

Whiteway, J., Koziarz, P., Veall, J., Sandhu, N., Kumar, P., Hoecher, B., & Lambert, I. B. (1998). Oxygen-insensitive nitroreductases: analysis of the roles of nfsA and nfsB in development of resistance to 5-nitrofuran derivatives in Escherichia coli. *Journal of bacteriology, 180*(21), 5529–5539. https://doi.org/10.1128/JB.180.21.5529-5539.1998

Zhai, Y., & Saier, M. H. (2002). The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes. *Protein science : a publication of the Protein Society, 11*(9), 2196–2207. https://doi.org/10.1110/ps.0209002

Zhang, Z., Huo, J., Velo, J., Zhou, H., Flaherty, A., & Saier, M. H., Jr (2024). Comprehensive Characterization of fucAO Operon Activation in Escherichia coli. *International journal of molecular sciences, 25*(7), 3946. https://doi.org/10.3390/ijms25073946

Zhang, Z., Kukita, C., Humayun, M. Z., & Saier, M. H. (2017). Environment-directed activation of the Escherichia coli flhDC operon by transposons. *Microbiology (Reading, England), 163*(4), 554–569. https://doi.org/10.1099/mic.0.000426

Zhang, Z., & Saier, M. H. (2009). A novel mechanism of transposon-mediated gene activation. *PLoS Genetics, 5*(10):e1000689. doi:https://doi.org/10.1371/journal.pgen.1000689

Zhang, Z., & Saier, M. H. (2009). A mechanism of transposon-mediated directed mutation. *Molecular microbiology, 74*(1), 29–43. https://doi.org/10.1111/j.1365-2958.2009.06831.x

Zhang, Z., & Saier, M. H., Jr (2011). Transposon-mediated adaptive and directed mutations and their potential evolutionary benefits. *Journal of molecular microbiology and biotechnology, 21*(1-2), 59–70. https://doi.org/10.1159/000333108

Zhang, Z., & Saier, M. H., Jr (2016). Transposon-mediated activation of the Escherichia coli glpFK operon is inhibited by specific DNA-binding proteins: Implications for stress-induced transposition events. *Mutation research, 793-794*, 22–31. https://doi.org/10.1016/j.mrfmmm.2016.10.003

Zhou, H. (2023). Characterizing the fucAO promoter and elucidating the transcriptional mechanism behind IS5-driven activation of the fucAO operon. *UC San Diego*. ProQuest ID: Zhou_ucsd_0033M_22268. Merritt ID: ark:/13030/m5rs1p0s. Retrieved from https://escholarship.org/uc/item/2m84361v