

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Dissecting the Hermes Transposase: Residues Important for Target DNA Binding and Phosphorylation

Permalink

<https://escholarship.org/uc/item/6sn4w1f0>

Author

Knapp, Joshua Allen

Publication Date

2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Dissecting the *Hermes* Transposase: Residues Important
for Target DNA Binding and Phosphorylation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry and Molecular Biology

by

Joshua Allen Knapp

December 2011

Dissertation Committee:
Dr. Peter W. Atkinson, Chairperson
Dr. Julia Bailey-Serres
Dr. Howard Judelson

Copyright by
Joshua Allen Knapp
2011

The Dissertation of Joshua Allen Knapp is approved:

Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor Dr. Peter Atkinson. He has been an outstanding mentor over these years and has helped me grow as a scientist and provided me with amazing research opportunities. I would like to thank my committee members Dr. Howard Judelson and Dr. Julia Bailey-Serres for all of their guidance and valuable input that has enhanced the quality of my dissertation research. I would also like to thank my undergraduate advisor Dr. Martin Case who taught me how to write scientifically.

Members of the Atkinson lab have become a family for me – one of them literally. I met my wife and the love of my life, Jennifer Wright, in the Atkinson lab. She has been a rock for me, providing me with love and support in so many ways. I feel blessed to have her my life. I would not be the scientist that I am with out Rob Hice. He has taught me more about science than anyone else in my academic career and I am thankful to have him in my life as both a teacher and a friend. I would like to thank Dr. Peter Arensburger who I've been able to get to know both him and his wonderful family over these years. He never let me get away with anything during lab meeting!

I would like to also thank my family. My mother has always been there to love and support me and has always believed in me even when I felt like I was failing. My dad has always been there when I needed him, always believed in me and always reminded me of how far I come from where I started out in life. I would also like to thank my father and mother in-law, Charlie and Kathy Inman, for all of their support and encouragement. They have always been there to take my mind off the stress.

ABSTRACT OF THE DISSERTATION

Dissecting the *Hermes* Transposase: Residues Important
for Target DNA Binding and Phosphorylation

by

Joshua Allen Knapp

Doctor of Philosophy, Graduate Program Biochemistry and Molecular Biology
University of California, Riverside, December 2011
Dr. Peter Atkinson, Chairperson

The developing fields of gene therapy and genetic transformation are constantly searching for new autonomously replicating transposable elements with the ability to stably integrate into the genome of a wide range of diverse hosts. The *Hermes* transposase has become a particularly attractive element because of its ability to transform a wide range of hosts, the diverse distribution of *hAT* elements in nature, and the extensive genetic, molecular, and biochemical data available. In order to develop *Hermes* as an efficient molecular biology tool this dissertation explored (1) how the transposase chooses its target DNA for integration and (2) regulatory mechanisms which may effect transposition activity.

The work presented in this dissertation identifies, for the first time, residues in *Hermes* that are important for target DNA binding. These residues are part of a model that I call the “rim of the wheel” based on the *Hermes* crystal structure. Data collected

from a comprehensive analysis designed to tease apart different steps of transposition suggests that these residues may contribute to a bi-functional domain responsible for both excising the transposon and binding to target DNA. The work presented here adds to our understanding of transposase integration biochemistry and can serve as a model for other elements.

In order for *Hermes* to be used as a biological tool for gene therapy or genetic studies in medically and agriculturally important insects it is important that *Hermes* has its greatest chance for successful integration into its host. The second project presented here goal of explored whether *Hermes* was subject to phosphorylation and if so whether this affected transposition. Here I report for the first time, direct evidence of phosphorylation of a eukaryotic class II transposase. Indirectly two other residues in *Hermes* were found to be important for transposition, as mutating them to residues with different chemistries decreased *Hermes* transposition frequency. This work has expanded our still growing understanding of essential residues that are functionally important for *Hermes* transposition.

Table of Contents

Chapter One:	
Introduction.....	1
1.1 General Introduction to Transposable Elements.....	1
1.2 Public Health and Socioeconmical Impact of Mosquito-Borne Diseases.....	3
1.3 Transposable Elements for Genetic Manipulation of Mosquitoes.....	4
1.4 Hermes as an Insect Gene Vector.....	6
1.5 Hermes Crystal Structure: The Structural and Functional Subdomains...7	7
1.6 Hermes Synaptic Complex.....	9
1.7 Hermes Cleavage: Link Between Transposable Elements and V(D)J Recombination.10	10
1.8 Hermes Target Site Binding.....	12
1.9 Hermes Non-Specific Binding.....	14
1.10 Role of Phosphorylation of Transposase Function and Activity.....	16
1.11 Dissertation Objectives.....	18
1.12 References.....	20
Chapter Two: The “Rim of the Wheel” Model: Residues Involved in Hermes Target DNA Binding.....	29
2.1 Abstract.....	29
2.2 Introduction.....	30
2.3 Results.....	36
2.4 Discussion.....	44
2.5 Conclusions.....	55
2.6 Future Experiments.....	57
2.7 Materials and Methods.....	58
2.8 References.....	72
Chapter Three: <i>Hermes</i> Phosphorylation: Identifying Residues that are Subject to Phosphorylation Using Bioinformatics and Mass Spectrometry.....	96
3.1 Abstract.....	96
3.2 Introduction.....	96
3.3 Results.....	100
3.4 Discussion.....	106
3.5 Conclusions.....	109
3.6 Future Experiments.....	110
3.7 Materials and Methods.....	111
3.8 References.....	117

Chapter Four: Summary and Conclusions.....	129
4.1 Summary.....	129
4.2 The <i>Hermes</i> “Rim of the Wheel” Model.....	129
4.3 <i>Hermes</i> is Phosphorylated in <i>Drosophila</i> S2 Cells.....	134
4.4 Conclusions.....	138
4.5 References.....	140

List of Figures

Figure 1.1 The <i>Hermes</i> transposase crystal structure	27
Figure 1.2 <i>Hermes</i> cleavage reaction showing hairpin formation on the flanking DNA	28
Figure 2.1 The transposition mechanism of <i>Hermes</i>	75
Figure 2.2 <i>Hermes</i> transposase hexamer crystal structure	76
Figure 2.3 The <i>Hermes</i> transposase octameric structure	77
Figure 2.4 <i>Hermes</i> transposase crystal structure overlay	78
Figure 2.5 “Rim of the wheel” residues	79
Figure 2.6 Schematic of the five-plasmid interplasmid transposition assay	80
Figure 2.7 Weblogo consensus target site duplication for WT <i>Hermes</i> and K300A	82
Figure 2.8 Weblogo consensus target site duplication for K292A and R306A	83
Figure 2.9 Schematic representation of the transformation assay in <i>Drosophila</i>	85
Figure 2.10 Graphical representation of the <i>Drosophila</i> transformation frequency of WT <i>Hermes</i> and the “rim of the wheel” mutants	86
Figure 2.11 The yeast excision frequency of WT <i>Hermes</i> , “rim of the wheel” mutants, and negative control pGalS	87
Figure 2.12 The yeast integration frequency of WT <i>Hermes</i> , ROW mutants, and the negative control pGalS	88
Figure 2.13 Coomassie blue stained SDS-PAGE His-tagged WT and mutant protein elution profiles	89
Figure 2.14 Autoradiograph of the strand transfer assay	90
Figure 2.15 Autoradiograph of the coupled cleavage and strand transfer assay	91

Figure 2.16 Autoradiograph of binding of WT and “rim of the wheel” mutant <i>Hermes</i> transposase to transposon left end 1-30	92
Figure 2.17 Autoradiograph of binding of WT and “rim of the wheel” mutant <i>Hermes</i> transposase to transposon left end 193	93
Figure 2.18 Position of the “rim of the wheel” mutants in the <i>Hermes</i> transposase octamer structure	94
Figure 3.1 NetPhosK results	118
Figure 3.2 Monomer from <i>Hermes</i> transposase crystal structure showing the Ser309 and Ser310 positions	119
Figure 3.3 Ser309 and Ser310 bind the <i>Hermes</i> transposon left end	120
Figure 3.4 Coomassie blue stained SDS PAGE of the S2 nuclear extract and immunoprecipitation pull-down <i>Hermes</i> transposase	121
Figure 3.5 Peptide fragments results from mass spectrometry of immunoprecipitation pull-down <i>Hermes</i> transposase from S2 cells	122
Figure 3.5 Peptide fragments results from mass spectrometry of immunoprecipitation pull-down <i>Hermes</i> transposase from S2 cells	122
Figure 3.6 Location of Ser233, Ser468 and Ser469 in the <i>Hermes</i> transposase octamer crystal structure	123
Figure 3.7 Trypsin and Chymotrypsin digest fragments	124
Figure 3.8 Weblogo consensus target site duplication for phosphorylation mutant <i>Hermes</i> transposases	125

Figure 3.9 Graphical representation of the <i>Drosophila</i> transformation frequency of WT <i>Hermes</i> and the SS309/10 mutants	126
Figure 3.10 The yeast excision frequency of WT <i>Hermes</i> , phosphorylation mutants, and negative control pGalS	127
Figure 3.11 The yeast integration frequency of WT <i>Hermes</i> , phosphorylation mutants, and the negative control pGalS	128

List of Tables

Table 2.1 Interplasmid transposition rates in injected <i>Drosophila</i> pre-blastoderm embryos for ROW mutants	81
Table 2.2 Integration sites of WT <i>Hermes</i> and K300A recovered from <i>Drosophila</i> interplasmid transposition assays	84
Table 2.3 K300A target site nucleotide frequency	85
Table 2.4 The <i>Drosophila</i> germline transformation frequency of WT <i>Hermes</i> and the “rim of the wheel” mutants	86
Table 3.1 The <i>Drosophila</i> germline transformation frequency of WT <i>Hermes</i> and the each of the SS309/10 mutant transposases	124
Table 3.2 Interplasmid transposition rates in injected <i>Drosophila</i> embryos for the phosphorylation mutants	125

Chapter One

Introduction

1.1 General Introduction to Transposable Elements

Transposable elements (TEs) are discrete pieces of DNA with the ability to move both within and between genomes. TEs have been found in every prokaryotic and eukaryotic genome to date and have played an important role in genome evolution serving as a source of genetic variation (Craig, N.L., 2002). They were first described by Barbara McClintock over 50 years ago in her experiments with maize that showed stress-induced genome instability resulted from discrete DNA segments moving between chromosomes (McClintock, 1950). Since then, our understanding of transposable elements has expanded greatly. Often described as “junk” or “selfish” DNA with no functional purpose other than to propagate themselves, or at the very least to remain present in the genome, we now know that a small number of TEs have a very important functional role in the genome providing a source of genetic variability (Biemont and Viera, 2006, Kazazian, Jr., 2004). For example, in the human genome they have contributed to thousands of proteins (Britten, 2006, Volff, 2006). While it is apparent that the vast majority of TEs have no obvious functional role or are inactive remnants in the genome, so far as we know, some organisms have adapted them for important cellular functions through a process known as molecular domestication. This is a process by which the host genome co-opts the DNA-binding properties or in fewer cases the enzymatic activity of the transposase protein and has been shown to be responsible for

the creation of proteins involved in such cellular processes as cellular proliferation, apoptosis, cell cycle progression, chromosome segregation, chromatin modification, transcriptional regulation and variable, diverse, and joining (V(D)J) recombination in the formation of antibodies (Sinzelle *et al.*, 2009, Chatterji *et al.*, 2004).

There are two classes of TEs (Finnegan, 1990). Class I elements are retrotransposons that copy themselves prior to insertion at a new genomic location using an RNA intermediate in a “copy-and-paste” mechanism. The “copy-and-paste” mechanism proceeds by first transcribing the transposon into an RNA intermediate and then reverse transcribing and reintegrating into the genome. Retrotransposons can be divided into two groups distinguished by the presence or absence of long terminal repeats (LTRs). In the human genome, LTR elements are endogenous retroviruses and account for ~8% of the entire genome (Cordaux and Batzer, 2009). Non-LTR elements remain active and account for about one third of the entire human genome (Cordaux and Batzer, 2009). These elements will not be discussed further.

Class II TEs are DNA transposons and are classified by the fact that they move using a DNA intermediate by a so-called “cut-and-paste” mechanism. In contrast to retrotransposition, class II transposition is typically non-replicative. Class II TEs are separated into subclasses based on the generation of single- or double-stranded DNA cuts during transposition (Sinzelle *et al.*, 2009). Furthermore, subclass 1 TEs are flanked by terminal inverted repeats (TIRs) (Sinzelle *et al.*, 2009). Transposons within the TIR order are distinguished by their TIR sequence motifs, and length of their target site

duplications (TSDs), which result from duplication of the host DNA flanking both ends of the transposon (Sinzelle *et al.*, 2009). Subclass 2 DNA transposons transposition requires replication and does not introduce double strand breaks (Sinzelle *et al.*, 2009).

Both classes of TEs contain nonautonomous and autonomous elements. Autonomous TEs encode all of the enzymes required for transposition whereas nonautonomous elements depend entirely on the enzymatic machinery of the autonomous relatives in order to mobilize (Sinzelle *et al.*, 2009).

1.2 Public Health and the Socioeconomical Impact of Mosquito-Borne Diseases

Mosquitoes cause immense human suffering with mosquito-borne diseases inflicting over one million deaths annually (WHO, 2010). Malaria is arguably the most important mosquito-borne disease due to both its impact on public health and socioeconomics in poverty stricken regions. While malaria is prevalent in Asia, Latin America, the Middle East, and parts of Europe, most cases and deaths occur in sub-Saharan Africa (WHO, 2010). In heavily impacted areas malaria can account for a staggering 40% of public health expenditures and can constitute 60% of outpatient visits alone (WHO, 2010).

Current efforts to combat mosquito-borne diseases include vaccines and medications, long lasting insecticide-treated bed nets, and insecticides. Unfortunately, mosquito nets are expensive and hard to distribute to some areas, and insecticides can inflict irreparable damage to the environment and provide little protection once mosquitoes grow resistant. Anti-malarial medications are available but can be expensive,

cause adverse side effects and are usually quickly countered by insect resistance.

Because of the resurgence of some mosquito-borne diseases, like malaria, and the emergence of diseases spread by mosquitoes into new geographical regions, such as West Nile Virus, it is important to develop tools for gene identification and ways to genetically modify these organisms. This will aid the understanding of host-pathogen interactions, which are currently difficult to do because of the lack of genetic tools and difficulty of mosquito transformation (Robinson *et al.*, 2004).

1.3 Transposable Elements for Genetic Manipulation of Mosquitoes

To combat mosquito-borne diseases such as malaria, dengue, and yellow fever, genetic tools must be developed. The ability of transposable elements to replicate and spread through a population in a non-Mendelian fashion makes them a valuable tool for genetically altering a mosquito and studying host-pathogen interactions as well as finding, isolating, and analyzing important insect genes (Atkinson, 2002; Robinson *et al.* 2004; Sinkins and Gould, 2006). Autonomous members of the Class II transposable elements are strong candidates because of their reasonable transformation frequency in mosquitoes as well as their ability to stably integrate into the germ line (Atkinson *et al.*, 2001, Robinson *et al.*, 2004). Germ-line transformation is an important quality in order to introduce a gene or create a knockout mutant that can be used to study how that effects a population.

At least four Class II TEs have been used to transform several mosquitoes of agricultural or human medical importance; they include *Hermes*, *Mos1*, *piggyBac*, and

Minos (O'Brochta *et al.*, 2003). *Hermes* has a wide host range and has been used to transform the yellow fever mosquito, *Aedes aegypti* and the West Nile virus mosquito, *Culex quinquefasciatus*, at rates of <10% and 12% respectively (Jasinskiene *et al.*, 1998, Allen *et al.* 2001). All germ-line integrations in both species have included non-canonical breaks of the flanking DNA sequence from the donor plasmid (Jasinskiene *et al.*, 2000). *Mos1* (*mariner*) from *D. mauritiana* has been used to transform *Ae. aegypti* at a rate of 4% and three out of four events were canonical cut-and-paste integrations (Coates *et al.*, 1998). The *piggyBac* element has been used to transform a larger number of mosquitoes including, *Ae. aegypti*, *Anopheles stephensi*, *Anopheles albimanus*, and the malaria vector *Anopheles gambiae* (Lobo *et al.*, 2002, Ito *et al.*, 2002, Perera *et al.*, 2002, Grossman, 2001). The transformation rate for these mosquitoes varies from 1% in *A. gambiae* to 40% in *A. albimanus* with all of the events being perfect cut-and-paste integrations into the TTAA target site (O'Brochta *et al.*, 2003). The *Minos* transposon from *D. hydei* has not been as extensively tested but has been used to transform *A. stephensi* (Catteruccia *et al.*, 2000). The results showed an average of 20% survival of injected embryos and 10% survival to adult. All integrations were perfect cut-and-paste events (Catteruccia *et al.*, 2000).

As these studies highlight, it is impossible to predict the transformation rate and efficiency for a given transposable element in heterologous hosts. It is also impossible to predict how the element will behave in the host during and after integration into the genome. For example, *Hermes* transposed through an imperfect integration in *A. aegypti* and *C. quinquefasciatus* that resulted in the presence of flanking DNA (O'Brochta *et al.*,

2003). In *D. melanogaster* the transformation efficiency is much higher (~20%) and accompanied by perfect cut-and-paste integration of the donor plasmid with the creation of the 8 bp target site duplications at the insertion site (Guimond *et al.* 2003). Additionally, in both species remobilization of the element was higher in the soma than in the germline. In order to create efficient, predictable genetic transformation systems for mosquitoes it is important to understand how the transposase functions at both a biochemical and genetic level.

1.4 *Hermes* as an Insect Gene Vector

Transposons of the *hAT* superfamily are widespread in plants and animals and the ability to transpose in species divergent from their hosts appears to be a common feature (Arensburger *et al.* 2011, Kempken and Windhofer, 2001). The *hAT* superfamily of TEs derives its name from three well-known representatives, *hobo* from *D. melanogaster*, *Ac* from maize, and *Tam3* from snapdragon (Atkinson *et al.*, 1993). In general *hAT* elements have short TIRs, generate 8 bp TSDs upon transposition and encode a single transposase protein that performs the DNA binding, cleavage, and rejoining reactions involved in transposition (Rubin *et al.*, 2001). Another well-characterized *hAT* representative, the *Hermes* element, is an active Class II TE isolated from the genome of the housefly, *Musca domestica* (Warren *et al.*, 1994). *Hermes* has been used to transform up to 15 different insect species including several non-drosophilid insects and is the first non-drosophilid TE to be converted into a gene vector (Sarkar *et al.*, 1997). The element is 2,749 nucleotides in length and contains imperfect 17 bp TIRs. The transposon contains

a single open reading frame encoding a 612 amino acid 70 kDa *Hermes* transposase capable of binding, excising and integrating the element in a targeted location (O'Brochta *et al.*, 1996). Upon integration of the element into a target site it creates an 8 bp insertion site duplication consistent with the length and sequence consensus for other *hAT* elements. The *Hermes* element has a consensus TSD of 5'-nTnnnAn-3' (Sarkar *et al.*, 1997).

1.5 *Hermes* Crystal Structure: The Structural and Functional Subdomains

The crystal structure of a truncated *Hermes* transposase (*Hermes*₇₉₋₆₁₂) was solved and published in 2005 and provided great insight into how it may function and assemble into a catalytically active higher ordered structure (Hickman *et al.*, 2005, Perez *et al.*, 2005). The *Hermes* transposase is organized into three domains, the N-terminal domain (residues 1-150); a catalytic domain with a retroviral-like fold; and a large α -helical insertion domain (residues 265-552) which interrupts the catalytic domain, Figure 1.1a (Hickman *et al.*, 2005). Three catalytically essential residues predicted to be the catalytic triad, D180, D248 and E572 are in close proximity and suitably arranged to coordinate the catalytically required Mg^{2+} ion, Figure 1.1b (Hickman *et al.*, 2005). The crystal structure also revealed that a highly conserved tryptophan residue, W319, projected into the active site (Hickman *et al.*, 2005). In the crystal structure of Tn5, a tryptophan projected into the active site provides an important stacking interaction stabilizing a flipped out base at the -2 position during hairpin formation (Davies *et al.*, 2000). It has been demonstrated that a *Hermes* mutant W319A was defective for either strand cleavage

or subsequent hairpin formation in *in vitro* plasmid cleavage assays. However, when the *Hermes* mutant W319A was provided pre-cleaved transposon ends it was able to catalyze the formation of single- and double-ended joined products (Hickman *et al.*, 2005). These data, taken together with the position of W319 in the crystal structure active site, strongly suggests a role for the W319 in binding and stabilizing the flanking-end DNA hairpin during the cleavage stage of transposition.

The N-terminal domain of *Hermes* (residues 1-150) contains a DNA binding domain, a nuclear localization signal, (residues 1-38), and a zinc-chelating BED finger domain, residues 25-78 (Michel and Atkinson, 2003). The absence of this domain in the truncated *Hermes*₇₉₋₆₁₂ has been shown to abolish non-specific binding while retaining transposition activity and TIR binding (Hickman *et al.*, 2005). BED domains are found in a wide variety of DNA binding proteins important for cellular regulation (Aravind, 2000).

The most highly conserved region among the transposases of *hAT* elements is a motif CPTRWNS from amino acids 316-322. The C-terminal region, residues 527-604 also show strong homology among *hAT* elements and this region has been shown to be important for multimerization of *hAT* elements (Michel *et al.*, 2003). Residues 551-569 in *Hermes* are required for multimerization and transposition although there are regions responsible for oligomerization that occur along the entire transposase (Hickman *et al.*, 2005, Michel *et al.*, 2003).

1.6 The *Hermes* Synaptic Complex

The *Hermes* transposase encoded by the *Hermes* element performs the DNA binding, cleavage, and rejoining reactions involved in transposition. The process of DNA transposition by class II TEs is initialized by the production of transposase that binds to specific motifs within the TIRs of the transposon (Auge-Gouillou *et al.*, 2005). Next, the protein forms a catalytically active synaptic complex in which the left and right ends of the transposon are thought to be brought together through oligomerization of the transposase protein (Carpentier *et al.*, 2010; Michel *et al.*, 2003). Synaptic complex formation facilitates the cleavage of the element from the donor DNA, relocation of the transposon, and integration into a specific target site (Gueguen, *et al.*, 2005).

The *Hermes* crystal structure was solved lacking bound transposon ends or DNA while the *Tn5* crystal structure was solved in its synaptic complex form bound to pre-cleaved DNA (Davies *et al.*, 2000; Rice and Baker, 2001). *Tn5* uses a DDE motif and has already been used as a model to identify and assign a role for the *Hermes* W319 in stabilizing a flipped out base during transposition (Hickman *et al.* 2005). Therefore it is useful to treat the *Tn5* crystal structure as a blueprint for understanding how *Hermes* forms a catalytically active synaptic complex.

The *Tn5* transposase requires dimerization of two *Tn5* monomers to form the synaptic complex (Davies *et al.* 2000; Steiniger-White *et al.*, 2000). Currently, it is not understood whether the monomer, dimer, or hexamer is the form of *Hermes* responsible for forming the catalytically active synaptic complex. Previous biochemical work using

yeast two-hybrid assays identified several areas of the *Hermes* transposase important for mediating its multimerization (Michel *et al.* 2003). In size exclusion chromatography experiments, *Hermes*₇₉₋₆₁₂ formed two oligomeric species, one eluting at a position consistent with a hexamer and a smaller species consistent with a heterotetramer that was later shown to be a degraded hexamer (Hickman *et al.* 2005). Sedimentation equilibrium studies were used to confirm that the larger species was a hexamer and *in vitro* activity assays revealed that only the hexamer was the active species. To take this a step further, a *Hermes* triple mutant (R369A, F503A, F504A) was shown to abolish *Hermes* hexamer formation *in vitro* (Hickman *et al.* 2005). Size exclusion chromatography was used to show that the triple mutant was only able to form dimers while monomers were also present. The *Hermes*₇₉₋₆₁₂ triple mutant demonstrated robust *in vitro* target joining and hairpin formation. However, in a *Saccharomyces cerevisiae in vivo* activity assay the triple mutant was 20 times less active than wild-type *Hermes* transposase. Thus, a *Hermes* dimer may be sufficient to form a basic catalytic unit able to capture oligonucleotides *in vitro* while a hexamer may be required to interact with chromosomal DNA.

1.7 Hermes Cleavage: Link Between Transposable Elements and V(D)J Recombination

Hermes transposition involves excision of the transposon from the donor site by double-strand breaks and hairpin formation on the flanking donor-site DNA, a process catalyzed by the formation of the *Hermes* synaptic complex. This mechanism of hairpin

formation is characteristic of V(D)J recombination, the process underlying combinatorial formation of antigen receptor genes, and has not been observed with other transposon families, thereby linking *hAT* elements and V(D)J recombination (Zhou *et al.* 2004, Lu *et al.* 2006).

The first step of transposition involves the *Hermes* transposase locating and binding to each of the specific TIRs of the transposon ends. Once the transposon ends are bound, the *Hermes* transposases are believed to oligomerize forming a catalytically active synaptic complex. Once the synaptic complex is formed, the ends are positioned into the transposase active site and cleaved. A divalent cation, usually Mg^{2+} or Mn^{2+} , positioned in the catalytic site uses a water molecule as the attacking nucleophile to initiate nicking on the 5' end of the transposon. The exposed 3'-OH of the nicked donor then undergoes nucleophilic attack on the 3'-OH of the transposon end, covalently joining the bottom flanking DNA at the 3' end of the transposon to form a hairpin (Figure 1.2). This is followed by release of the transposon from the donor DNA. The 3'-OH of the transposon is then free to undergo nucleophilic attack on the target strand of DNA once the target sequence has been located by the transposase. The reaction does not require an external energy source, as there is no change in the net number of high-energy chemical bonds.

This cleavage mechanism, by which there is hairpin formation on the flanking donor DNA, is unique to *hAT* elements and Rag1/2 recombinases and has not been observed with any other TE to date. Whereas hairpin formation on the donor DNA is a mechanistic characteristic that links *hAT* elements and V(D)J, there is a small variation between the two. *Hermes* excision from the donor is accompanied by a single nucleotide

of flanking DNA at its 5' end, whereas Rag1 excision produces a flush 5' terminus (Zhou *et al.*, 2004). It is presumed that this overhanging nucleotide is excised and resolved during DNA repair, after the integration of the transposon into a new target site.

1.8 *Hermes* Target Site Binding

Once the *Hermes* transposase has excised the transposon from the donor DNA the next step during transposition is for the transposase to find a target site. While insertions on a large scale are random, there is a preferred target site consensus for *Hermes* (Gangadharan *et al.*, 2010). The target site is 8 bp long and contains a T and an A in the 2nd and 7th positions respectively, nTnnnnAn, where n is any nucleotide (Sarkar *et al.*, 1997). Most transposable elements studied exhibit sequence bias in choosing their integration site that can be due to a combination of factors such as host protein interactions, interactions with other accessory proteins, and accessibility of the target DNA. The *in vivo* genomic integration preference of *Hermes* in *S. cerevisiae* is profoundly influenced by the three dimensional structure and accessibility of the DNA. NextGen sequencing of both *in vivo* and *in vitro* genomic insertions were able to analyze $\sim 2 \times 10^5$ unique *de novo* transposon insertion events and found that *Hermes* transposase prefers to integrate in nucleosome-free regions and targets intergenic regions of the genome (Gangadharan *et al.* 2010). Additionally, it was found that *Hermes* preferentially inserts into AT-rich regions near the 5' end of genes (Gangadharan *et al.*, 2010). Insertions into GC-rich regions found just outside of ORF's were underrepresented *in vivo* but were the preferred target areas *in vitro* due to the greater flexibility of GC-rich

stretches of DNA (Gangdharan *et al.*, 2010). This is consistent with a study that found *Hermes* preferentially integrated into areas that were of high DNA flexibility (Chowdhury *et al.*, 2004). Increased flexibility reduces the energy required for bending of the target DNA.

It is not yet known which residues in the *Hermes* transposase are responsible for binding to the target DNA. Studies with Tn5 transposase have shown that the residues responsible for binding to target DNA were also responsible for binding to the flanking sequence of the transposon ends (Gradman *et al.*, 2008). Furthermore, bias detected in the sequences flanking the preferred insertion sites in Tn5 and Tn10 are likely due to interactions with the transposase. In both of these transposases it was discovered that nicks on the boundaries of a proposed 9 bp insertion site enhanced targeting to that site (Gradman *et al.*, 2008). It was hypothesized by these authors that the nicks lead to an increase in DNA flexibility, reducing the energy needed to create DNA bends during insertion into the target (Gradman *et al.*, 2008). These were reminiscent of the results found in *Hermes* in which the flexibility of the DNA played a significant role in target site selectivity. Similarly, with *Hermes* there exists a regional sequence bias on either side of the target site in nucleosome-free regions up to 200 bp in length in *S. cerevisiae* integrations (Gangdharan *et al.*, 2010). The regional bias consisted of a distinct T-rich top strand bias 5' to the target site and a top strand A-rich bias 3' to the target site consistent with *Hermes* ability to integrate in either orientation (Gangdharan *et al.*, 2010).

Understanding which residues or regions in the *Hermes* transposase that are

responsible for binding to the target site and whether they are target site specific or responsible for long-range targeting is important for understanding *Hermes* integration biochemistry. It also elucidates both the beginning and end points of transposition, steps that up to this point in time have been very difficult to discern. With the 2005 publication of the *Hermes* transposase crystal structure, a model has been proposed for regions that may be responsible for target site DNA binding (Hickman *et al.*, 2005). The model is based on the computer generated hexameric structure and has subsequently been bolstered by an unpublished but crystallographically observed octamer structure that was co-crystallized bound with 16 bp of *Hermes* transposon left ends (F. Dyda and A. Hickman, personal communication). Rendering the hexamer structure in a surface charge representation presented a negatively charged trench between each of the monomers at interface 3 – the computer generated interface. This model also puts the active sites at ~35 Å apart, close to the 28 Å that would be needed for the ends of the transposon to be integrated 8 bp apart.

1.9 *Hermes* Non-Specific Binding

Non-specific binding is a property exhibited by most DNA binding proteins and is important for target localization for many recombinases, transcription factors, and restriction enzymes (Halford and Marko, 2004). The role of transposase non-specific binding is not very well characterized or understood and its role has only been investigated for a few transposases (Steiniger *et al.*, 2006). Understanding non-specific binding by transposases is important for complete comprehension of their reaction

mechanism as this ability can affect reaction rates, target recognition and the activity of the transposase. Research performed on Tn5 showed that non-specific binding was important for stabilizing the transposase, significantly increasing the lifetime of transposase 95 times longer than in the absence of non-specific DNA (Steiniger *et al.*, 2006). Additionally, it was demonstrated that Tn5 transposase can utilize non-specific DNA to facilitate localization of an intramolecular transposase recognition end sequence (analogous to the *Hermes* TIR) over distances less than 464 bp. Finally, it was shown that Tn5 synaptic complex formation was inhibited in the presence of increasing concentrations of non-specific DNA (Steiniger *et al.* 2006).

Little is known of the role of *Hermes* non-specific binding. The N-terminal domain of *Hermes* transposase (residues 1-150) is the site-specific DNA binding domain, presumably responsible for recognizing the transposon TIRs, and contains a zinc-chelating BED finger domain, (residues 25-78) (Michel and Atkinson, 2003). The absence of this domain in a truncated *Hermes*₇₉₋₆₁₂ has been shown to abolish non-specific binding while retaining transposition activity and ITR binding (Hickman *et al.* 2005). The BED domain is found in a wide variety of DNA binding proteins which suggests a general DNA binding role for this domain and could explain why the truncated *Hermes* transposase no longer bound non-specific DNA (Aravind, 2000).

Understanding the role of *Hermes* transposase non-specific binding during transposition is essential to answering fundamental questions about how it affects reaction rates, target site recognition, and synaptic complex formation.

1.10 Role of Phosphorylation on Transposase Function and Activity

Cell cycle control and response to DNA damage involves the phosphorylation of proteins as a regulatory mechanism to modulate protein activity (Zhou and Elledge, 2000). When cells experience DNA damage, checkpoint pathways become activated to halt the cell cycle in order to allow repair of the DNA damage (Zhou and Elledge, 2000). DNA transposition is regulated to minimize the level of DNA damage that may result in lethal effects to the host cell. One way this can be achieved is through phosphorylation of the transposase or interaction with host factors. Previous studies have shown that eukaryotic transposases can be sensitive to phosphorylation and that they interact with host machinery involved with DNA damage repair (Beall *et al.*, 20002; Izsvak *et al.*, 2004; Walisko *et al.*, 2006). A major pathway of DNA repair that has been shown to be activated in response during transposition is nonhomologous end joining (NHEJ) (Heemst *et al.*, 2004; Izsvak *et al.*, 2004; Walisko *et al.*, 2006). There are several kinase families involved in the activation of this pathway. One such family, the ATM-family DNA damage checkpoint protein kinases, has been found to be activated during transposition of several transposases as well as during V(D)J recombination, which shares several mechanistic similarities with *Hermes* transposition (Beall and Rio, 1996; Grawunder *et al.*, 1998; Lin *et al.*, 1999). In fact it has been shown that in mice, where NHEJ genes have been knocked out, there is severe combined immunodeficiency caused by the inability to carry out V(D)J recombination (Zhu *et al.*, 1996).

Because the host cell tries to inactivate the transposase whereas the transposase strives to remain active, the transposase must evolve a way around the inactivation mechanism. Transposase phosphorylation may be involved in this regulation. There are several outcomes to the fate of the transposase that can occur due to phosphorylation. (i) The transposase can evolve to require phosphorylation in order to function properly. (ii) Another is that the transposase can succumb to the effects of the phosphorylation and be inactivated. (iii) Yet another is to evolve interactions with host factor proteins that can shut down signaling mechanisms that would, if activated, phosphorylate the transposase.

The *Sleeping Beauty (SB)* transposase, a once extinct transposase that was resurrected into an active element from zebrafish, has been shown to interact with and down-regulate Miz-1, a transcriptional regulator of genes involved in cell cycle regulation including cyclin D1 (Walisko, *et al.*, 2006). Down regulation of cyclin D1 results in a slow down of cell cycle in the G1 phase that increases *SB*'s efficiency of transposition, presumably because transposon-induced DNA damage can be efficiently repaired by NHEJ. This is an example of the transposase evolving a mechanism that allows it to bypass the host cells response machinery in order to maximize its chance for a successful transposition event. It should be pointed out that this response was not due to the presence of transposon DNA but to the *SB* transposase itself, indicating that this response was not associated with transposition induced DNA-damage. In another study involving the *Drosophila P-element* it was demonstrated that the NHEJ pathway was used for DNA repair at the donor site following transposition (Beall *et al.*, 2002).

The ATM-family of protein kinases, which are activated during NHEJ, regulate cell cycle progression and DNA damage responses through a cascade of protein phosphorylation (Zhou and Elledge, 2000). Because there are several kinases involved in this pathway and they have a minimum site of S/TQ, it is conceivable that these kinases could also act to phosphorylate the transposase as a means of regulation in response to double strand breaks (Kim *et al.*, 1999). This also allows the cell to down-regulate the transposase while it repairs the DNA damage so that the transposase cannot cause any more damage. To date, the study of phosphorylation and its effects on transposition are limited and most studies deal with the DNA damage repair mechanisms rather than the phosphorylation effects on the recombinase proteins themselves. In the most detailed case involving the *Drosophila P-element*, Beall *et al.* (2002), mutated eight of 10 possible consensus phosphorylation sites for the ATP family DNA damage checkpoint kinases from serine or threonine to alanine. They found that one of these mutants, S129A, actually resulted in an increase in the transposition rate of the mutant over wild type in *in vivo* recombination assays and *Drosophila* germline transformation. The other seven out of eight residues showed reduced activity when individually substituted to alanine, suggesting a role of phosphorylation in regulating the *P-element* transposase.

1.11 Dissertation Objectives

The developing fields of gene therapy and genetic transformation are constantly searching for new autonomously replicating TEs with the ability to stably integrate into the genome of a wide range of diverse hosts. The *Hermes* transposase has become a

particularly attractive element because of its ability to transform a wide range of hosts, the diverse distribution of *hAT* elements in nature, and the extensive genetic, molecular, and biochemical data available. However, despite this wealth of information the residues responsible for binding to target DNA remain unknown. Additionally, it is not known whether Hermes is subject to phosphorylation and, if so, what affect that has on its transposition activity. In order for *Hermes* to be used as a biological tool for gene therapy or genetic studies in medically and agriculturally important insects it is important to understand how the transposase chooses its target DNA for integration and the role, if any, phosphorylation plays in *Hermes* transposition.

The goals of my research were to:

1. Locate those residues and regions of the *Hermes* transposase that were responsible for binding to target DNA. These regions have not been identified for any eukaryotic transposase and the work here reports the location of residues of the *Hermes* transposase that appear to fulfill this role.
2. Determine whether the *Hermes* transposase is phosphorylated and, if so, whether this plays a role in regulating the frequency of *Hermes* excision and integration.

1.12 References

- Allen, M.L., O'Brochta, D.A., Atkinson, P.W., Levesque, C.S. (2001) Stable germ-line transformation of *Culex quiquefaciatus* (Dipter: Culicidae). *J. Med. Entomol.* **38**: 701-10.
- Aravind, L. (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* **25**: 421-423.
- Arensburger, P., Hice, R., Zhou, L., Smith, R.C., Tom, A.C., Wright J.A., Knapp, J., O'Brachta, D.A., Craig, N.L., Atkinson, P.W. (2011) Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics.* **188**: 45-57.
- Arensburger, P.A., O'Brachta, D.A., Orsetti, J. and Atkinson, P.W. (2005) *Herves*, a new active *hAT* transposable element from the African malaria mosquito *Anopheles gambiae*. *Genetics.* **169**: 697-708.
- Atkinson, P.W. (2002) Genetic engineering in insects of agricultural importance. *Insect Biochem. Mol. Biol.* **32**: 1237-42.
- Atkinson, P.W., Pinkerton, A.C., O'Brochta, D.A. (2001) Genetic transformation systems in insects. *Ann. Rev. of Entomol.* **46**: 317-346.
- Atkinson, P.W., Warren, W.D., O'Brochta, D.A. (1993) The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci.* **90**: 9693-7.
- Auge-Gouillou, C., Brillet, B., Germon, S., Hamelin, M.H., Bigot, Y. (2005) *Mariner MOSI* transposase dimerizes prior to ITR binding. *J. Mol. Biol.* **351**: 117-130.
- Auge-Gouillou, C., Brillet, B., Hamelin, M.H., Bigot, Y. (2005) Assembly of the *mariner MOSI* synaptic complex. *Molec. and Cell. Biol.* **25**: 2861-2870.
- Beall, E.L., Mahoney, M.B., Rio, D.C. (2002) Identification and analysis of a hyperactive mutant form of *Drosophila P-element* transposase. *Genetics.* **162**: 217-227.
- Beall, E.L., and Rio, D.C. (1996) *Drosophila* IRBP/Ku p70 corresponds to the mutagen-sensitive mus309 gene and is involved in P-element excision in vivo. *Genes Dev.* **10**: 921-33.

- Biemont, C. and Vieira, C. (2006) Junk DNA as an evolutionary force. *Nature*. **443**: 521-524.
- Britten, R. (2006) Transposable elements have contributed to thousands of human proteins. *Proc. Natl. Acad. Sci.* **103**: 1783-1803.
- Carpentier, G., Jaillet, J., Pflieger, A., Adet, J., Renault, S., Auge-Gouillou, C. (2010) Transposase-transposase interactions in MOS1 complexes: a biochemical approach. *J. Mol. Biol.* **405**: 892-908.
- Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C., and Crisanti, A. (2000) Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature*. **405**: 959-962.
- Chatterji, M., Tsai, C.L., Schatz, D.G. (2004) New concepts in the regulation of an ancient reaction: transposition by RAG1/RAG2. *Immunol. Rev.* **200**:261-71.
- Chowdhury, M.H., Julian, A.M., Coates, C.J., Cote, G.L. (2004) Detection of differences in oligonucleotide-influenced aggregation of colloidal gold nanoparticles using absorption spectroscopy. *J Biomed Opt.* **9**: 1347-57.
- Coates, C.J., Jasinskiene, N., Miyashiro, L., James, A.A. (1998) *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes Aegypti*. *Proc Natl Acad. Sci.* **95**: 3748-51.
- Cordaux, R. and Batzer, M.A (2009). The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **10**: 691-703.
- Craig, N.L. (2002) Mobile DNA: an introduction. In: Craig, R., Gellert, M., Lambowitz, A.M. (Eds), *Mobile DNA II*. ASM Press, Washinton, DC. 3-11.
- Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., Rayment, I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*. **289**: 77-85.
- Finnegan, DJ (1990) Transposable elements and DNA transposition in eukaryotes. *Curr. Op. Cell. Biol.* **2**: 471-477.
- Gangdharan, S., Mularoni, L., Fain-Thorton, J., Wheelan, S.J., Craig, N.L. (2010) DNA transposon *Hermes* inserts into DNA in nucleosome-free regions in vivo. *Proc. Natl. Acad. Sci.* **107**: 21966-21972.

- Germon, S., Bouchet, N., Casteret, S., Carpentier, G., Adet, J., Bigot, Y., Auge-Gouillou, C. (2009) *Mariner MOSI* transposase optimization by rational mutagenesis. *Genetica*. **137**: 265-276.
- Gradman, R.J. Reznikoff, W.S. (2008) Tn5 Synaptic complex formation: role of transposase residue W450. *J. Bacteriology*. **190**: 1484-1487.
- Gradman, R.J., Ptacin, J.L., Bhasin, A., Reznikoff, W.S. and Goryshin, I.Y. (2008) A bifunctional DNA binding region in Tn5 transposase. *Molec. Microbiol.* **67**: 528-540.
- Grawunder, U., West, R.B., Lieber, M.R. (1998) Antigen receptor gene rearrangement. *Curr. Opin. in Immunol.* **10**: 172-180.
- Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., Benedict, M.Q. (2001) Germline transformation of the malaria vector, *Anopheles gambiae*, with the *piggyBac* transposable element. *Insect Mol. Biol.* **10**: 597-604.
- Gueguen, E. Rousseau P., Duval-Valentin, G., Chandler, M. (2005) The transpososome: control of transposition at the level of catalysis. *TRENDS in Microbiol.* **13**: 543-549.
- Guimond, N., Bideshi, D.K., Pinkerton, A.C., Atkinson, P.W., O'Brachta, D.A. (2003) Patterns of *Hermes* transposition in *Drosophila melanogaster*. *Mol. Genet. and Genom.* **268**: 779-790.
- Heemst, D.V., Brugmans, L. Verkaik, N.S., van Gent., D.C. (2004) End-joining of blunt DNA double-strand breaks in mammalian fibroblasts is precise and requires DNA-PK and XRCC4. *DNA Repair.* **3**: 43-50.
- Hill, C.A., Kafatos, F.C., Stanfield, S.K., Collins, F.H. (2005) Arthropod-borne diseases: vector control in the genomics era. *Nat. Microbiol.* **3**: 262-268.
- Hickman, A.B., Chandler, M., Dyda, F. (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit. Rev. in Biochem. and Mol. Biol.* 1-20.
- Hickman, A.B., Perez, Z.N., Zhou, L., Primrose, M., Ghirlando, R., Hishaw., J.E., Craig, N.L., Dyda, F. (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.* **12**: 715-721.

- Ito, J., Ghosh, A., Moreira, L., Wimmer, E.A., Jacobs-Lorena, M. (2002) Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature*. **417**: 452-55.
- Izsvak, Z., Stume, E.E., Fiedler, D., Katzer, A., Jeggo, P.A. and Ivics, Z. (2004) Healing the wounds inflicted by *Sleeping Beauty* transposition by double-strand break repair in mammalian somatic cells. *Molec. Cell*. **13**: 279-290.
- Jasinskiene, N., Coates, C.J. and James, A.A. (2000) Structure of *Hermes* integrations in the germline of the yellow fever mosquito, *Aedes aegypti*. *Insect Molec. Biol.* **9**: 11-18.
- Jasinskiene, N., Coates, C.J., Benedict, M.Q., Cornel, A.J., Rafferty, C.S., James, A.A., and Collins, F.H. (1998) Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the *Hermes* element from the housefly. *Proc. Natl. Acad. Sci.* **95**: 3743-3747.
- Kazazian Jr., H.H. (2004) Mobile Elements: Drivers of genome evolution. *Science*. **303**: 1626-1632.
- Kempken, F., Windhofer, F. (2001) The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. *Chromosoma*. **110**: 1-9.
- Kim, S.T., Lim, D.S., Canman, C.E., Kastan, M.B. (1999) Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**: 37538-43.
- Lin, J.M., Landree, M.A., Roth, D.B. (1999) V(D)J recombination catalyzed by mutant RAG proteins lacking consensus DNA-PK phosphorylation sites. *Molec. Immunol.* **36**: 1263-1269.
- Lobo, N.F., Hua-Van, A., Li, X., Nolen, B.M., Fraser, M.J. (2002) Germ line transformation of the yellow fever mosquito, *Aedes aegypti*, mediated by transpositional insertion of a *piggyBac* vector. *Insect Mol. Biol.* **11**: 133-39.
- Lu, C.P., Sandoval, H., Brandt, V.L., Rice, P.A., Roth, D.B. (2006) Amino acid residues in Rag1 crucial for DNA hairpin formation. *Nat. Struct. & Molec. Biol.* **13**: 1010-1015.
- Halford, S.E. & J.F. Marko (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acid Res.* **32**: 3040-3052.

- McClintock, B. (1950) The origin and behavior of mutale loci in maize. *Proc. Natl. Acad. Sci.* **36**: 344-55.
- Michel, K., Atkinson, P.W. (2003) Nuclear localization of the *Hermes* transposase depends on basic amino acid residues at the N-terminus of the protein. *J. Cell. Biochem.* **89**: 778-790.
- Michel, K., O'Brochta, D.A., Atkinson, P.W. (2003) The C-terminus of the *Hermes* transposase contains a protein multimerization domain. *Insect Biochem. Molec. Biol.* **33**: 959-970.
- O'Brochta, D.A., Sethuraman, N., Wilson, R., Hice, R.H., Pinkerton, A.C., Levesque, C.S., Bideshi, D.K., Jasinskiene, N., Coates, C.J., James, A.A., Lehane, M.J. and Atkinson, P.W. (2003) Gene vector and transposable element behavior in mosquitoes. *J. of Exp. Biol.* **206**: 3823-3834.
- O'Brochta, D.A., Warren, W.D., Saville, K.J., Atkinson, P.W. (1996) *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica*. *Genetics.* **142**: 907-914.
- Perera, O.P., Harrell, R.A., Handler, A.M. (2002) Germ-line transformation of the South American malaria vector, *Anopheles albimanus*, with a *piggyBac/EGFP* transposon vector is routine and highly efficient. *Insect Mol. Biol.* **11**: 291-97.
- Perez, Z.N., Musingarimi, P., Craig, N.L., Dyda, F., Hickman, A.B. (2005) Purification, crystallization and preliminary crystallographic analysis of the *Hermes* transposase. *Acta Crystallographica.* **F61**: 587-590.
- Rice, P.A., Baker, T.A. (2001) Comparative architecture of transposase and integrase complexes. *Nat. Struct. Biol.* **8**: 302-307.
- Robinson, A.S., Gerald, F., Atkinson, P.W. (2004) Insect transgenesis and its potential role in agriculture and human health. *Insect Biochem. Molec. Biol.* **34**: 113-120.
- Rubin, E., Lithwick, G., Levy, A.A. (2001) Structure and evolution of the *hAT* transposon superfamily. *Genetics.* **158**: 949-957.
- Sinkins, S.P., Gould, F., (2006) Gene drive systems for insect disease vectors. *Nat. Genet.* **7**: 427-435.
- Sinzelle, L., Izsvak, Z., Ivics Z. (2009) Molecular domestication of transposable

- elements: From detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* **66**: 1073-1093.
- Sarkar, A., Coates, C.J., Whyard, S., Willhoeft, U., Atkinson, P.W., O'Brochta, D.A. (1997) The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica.* **99**: 15-29.
- Sarkar, A., Yardley, K., Atkinson, P.W., James, A.A., O'Brochta, D.A. (1997) Transposition of the *Hermes* element in embryos of the vector mosquito, *Aedes aegypti*. *Insect Biochem. Molec. Biol.* **27**: 369-363.
- Steiniger, M., Adams, C.D., Marko, J.F. and Reznikoff, W.S. (2006) Defining characteristics of Tn5 transposase non-specific DNA binding. *Nucleic Acid Res.* **34**: 2820-2832.
- Steiniger-White, M., Reznikoff, W.S. (2000) The C-terminal α -helix of Tn5 transposase is required for synaptic complex formation. *J. Biol. Chem.* **275**: 23127-23133.
- Tu, Z., Coates, C. (2004) Mosquito transposable elements. *Insect Biochem. Molec. Biol.* **34**: 631-644.
- Volff, J.N. (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays.* **28**: 913-922.
- Walisko, O., Izsvak, Z., Szabo, K., Kaufman, C.D., Herold, S., Ivics, Z. (2006) *Sleeping Beauty* transposase modulates cell-cycle progression through interaction with Miz-1. *Proc. Natl. Acad. Sci.* **103**: 4062-4067.
- Warren, W.D., Atkinson, P.W. and O' Brochta, D.A. (1994) The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (hat) element family. *Genet. Res.* **64**: 87-97
- Wu, M., Li, L., Sun, Z. (2007) Transposable element fragments in protein-coding regions and their contributions to human functional proteins. *Gene.* **401**: 165-171.
- Yant, S.R., Huang, Y., Akache, B. and Kay, M. (2007) Site-directed transposon integration in human cells. *Nucleic Acid Res.* **35**: 1-13.
- Zhou, B.B., Elledge, S.J. (2000) The DNA damage response: putting checkpoints in perspective. *Nature.* **408**: 433-39.

- Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F. and Craig, N.L. (2004) Transposition of *hAT* elements links transposable elements and V(D)J recombination. *Nature*. **432**: 955-1001.
- Zhu, C., Bogue, M.A., Lim, D.S., Hasty, P., Roth, D.B. (1996) Ku86-deficient mice exhibit severe combined immunodeficiency and defective processing of V(D)J recombination intermediates. *Cell*. **86**: 379-89.

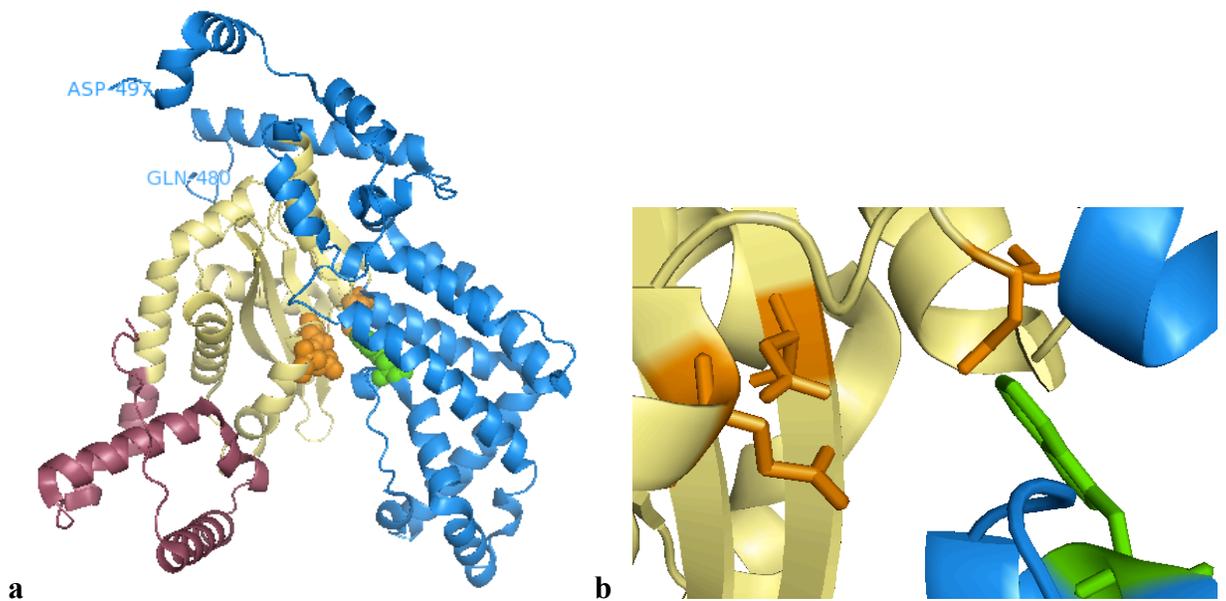


Figure 1.1: (Above Left) The *Hermes* transposase crystal structure. (a) The N-terminal domain is shown here in red (residues 1-150), the catalytic domain in yellow, and the all-alpha helical insertion domain is shown in blue. (b) The catalytically essential DDE (Asp180, Asp248, Glu572) motif residues are shown in orange and the highly conserved tryptophan 319 is shown in green. (Above Right) A closer view of the *Hermes* active site. Color assignment remains the same (Hickman *et al.*, 2005).

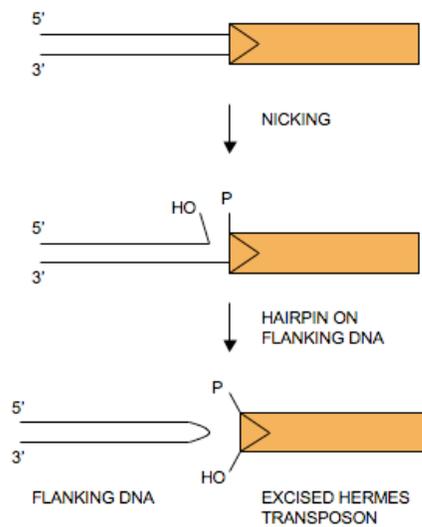


Figure 1.2: *Hermes* cleavage reaction showing hairpin formation on the flanking DNA. The *Hermes* transposon is shown in orange.

Chapter Two

The “Rim of the Wheel” Model: Residues Involved in *Hermes* Target DNA Binding

2.1 Abstract

The *Hermes* transposable element is a class II DNA transposon isolated from the housefly *Musca domestica* and transposes by a “cut and paste” mechanism. The *Hermes* transposase has become a model in transposase research because of the extensive genetic, molecular, and biochemical data available. However, the transposase residues responsible for binding to target DNA remain unknown. Here I present, for the first time, a model for *Hermes* target DNA binding which I coined the “rim of the wheel” model. This model was based on the *Hermes* hexamer transposase crystal structure and supported by the crystallization of the *Hermes* transposase octamer structure bound with 16 bp *Hermes* transposon left ends. There were four positively charged residues in the *Hermes* transposase hypothesized to bind target DNA based on the “rim of the wheel” model. These residues were individually mutated to non-charged alanine residues. One of these mutants displayed a change in its target site duplication and three of the four mutants displayed a decrease in transposition activity. The work presented here adds to our understanding of transposase integration biochemistry and can serve as a model for other elements.

2.2 Introduction

Transposons are pieces of DNA with the ability to move themselves both within and between genomes. They have been an enormous force in increasing genetic diversity and large proportions of many eukaryotic genomes originated as transposons (Craig, 2002). Autonomous DNA transposons encode the transposase protein that mediates the movement of the transposon through a so-called cut-and-paste mechanism involving a double-strand DNA intermediate. DD[E/D] transposase proteins are members of the retroviral superfamily of proteins that include retroviral integrases and Rag1 involved in V(D)J recombination, which all share structurally conserved catalytic cores, chemical mechanisms, and an insertion domain with an RNaseH-like fold that draws three catalytically essential residues, DDE or DDD, into close proximity (Hickman *et al.*, 2010, Nesmelova *et al.*, 2010). The *hAT* element, *Hermes*, is a member of this transposon family and encodes an autonomous transposase that performs three functions (i) the excision of the transposon from donor DNA via double strand breaks between the donor DNA and the transposon ends, (ii) location of a target site sequence and (iii) the joining of the exposed transposon ends to the target DNA at the reinsertion site.

Transposition is a highly coordinated event that involves many protein-DNA and protein-protein interactions. Several of these protein-DNA and protein-protein interactions are known or are in the process of being understood for the *Hermes* transposase, but others are yet to be determined. One interaction that is uncharacterized is that between the *Hermes* transposase and its target DNA. Understanding which residues in *Hermes* are important for locating and binding of the transposase to the target

site DNA is fundamental for understanding transposition. This knowledge will expand out understanding of transposase integration biochemistry, which may facilitate their use in directed manipulation and modification of genomes.

During the first step of transposition the *Hermes* transposase locates and binds to the left and right terminal inverted repeats (TIRs) of the *Hermes* transposon (Figure 2.1) (O'Brochta *et al.*, 1996). It is thought that this is preceded by the formation of a synaptic complex; the oligomerization of the *Hermes* transposases bound to the left and right transposon ends (Michel *et al.*, 2003). The formation of the synaptic complex is followed by a nick one nucleotide into the donor strand flanking the 5' end of the each TIR by the Mg²⁺ cofactors positioned in the transposase catalytic core that uses a water molecule as the attacking nucleophile (Zhou *et al.*, 2005). The 3'-OH of the upper transfer strand then attacks the phosphodiester bond of the transfer strand and in the process covalently joins the upper and lower flanking DNA to form a hairpin and release the transposon bearing a free 3'-OH (Figure 1.4) (Zhou *et al.* 2005). The 3'-OH groups on the ends of the transposon are nucleophiles for attack on the target strand of DNA 8 bp apart, once the target sequence has been located by the transposase. Hairpin formation on the flanking donor DNA is unique to *hAT* elements and RAG1 recombinase and has not been observed with any other transposase (Zhou *et al.*, 2005). Whereas hairpin formation on the donor DNA is a mechanistic characteristic that links *hAT* elements and V(D)J recombination, which is responsible for generating immunoglobulins and T-cell receptors in vertebrates, there is a small variation between the two; *Hermes* excision from the donor is accompanied by a single nucleotide of flanking DNA at its 5' ends whereas

Rag1-catalyzed excision produces a flush 5' terminus (Zhou *et al.*, 2005). It is presumed that this overhanging nucleotide is excised and resolved during DNA repair after integration of the transposon into a new target site.

Hermes has been shown to be effective in integrating into a wide variety of different host genomes with varied frequencies (Sarkar *et al.*, 1997). Oftentimes transposition frequency in one system does not reflect the frequency in another. For example, *Hermes* transposition in *Aedes aegypti* embryos into a plasmid target are precise and contain the usual 8 bp TSD, whereas germline integrations are non-canonical, containing substantial donor plasmid flanking DNA and do not have TSD's (Jasinskiene *et al.*, 2000). The differences in transposition in different systems may be attributed to a combination of factors such as host protein interactions, interactions with other accessory proteins, interaction with endogenous elements, and accessibility of the target DNA. While insertions on a large scale are random, there is a preferred 8 bp target site consensus for *Hermes*. This consensus is nTnnnnAn, where n is any nucleotide (Gangdharan *et al.*, 2010; Sarkar *et al.* 1997). Most transposases studied exhibit sequence bias in their integration site. Indeed in *Hermes* there exists a regional sequence bias on either side of the target site in nucleosome-free regions up to 200 bp in length in *S. cerevisiae* integrations with a distinct T-rich upper strand bias 5' to the target site and an upper strand A-rich bias 3' to the target site consistent with *Hermes* ability to integrate in either orientation (Gangdharan *et al.*, 2010).

The *Hermes* transposase crystal structure was solved lacking bound transposon ends or DNA (Hickman *et al.*, 2005). The structure was created using a truncated

Hermes, *Hermes*₇₉₋₆₁₂, and was solved as a heterotetramer formed by two heterodimer asymmetric units of *Hermes*₇₉₋₆₁₂, each bound to an N-terminal domain fragment. Domain swapping of two helices between residues 497 and 516 of two heterodimers created interface 2, which was a crystallographically observed interaction that contributed to the heterotetramer. Because *in vitro* activity assays revealed that a hexamer form of *Hermes* transposase was the active complex Hickman *et al.* (2005) used the two-fold symmetry axes of the heterotetramer to generate a computer model of the hexamer structure (Figure 2.2.a). The resulting model was a spiral of six *Hermes*₇₉₋₆₁₂ monomers formed from a trimer of dimers with alternating observed and modeled interfaces, interfaces 2 and 3, respectfully (Figure 2.2.a). Although the exact higher oligomerization of active *Hermes* is unknown, Hickman *et al.* (2005) showed that *Hermes* monomers are transpositionally inactive, whereas dimers formed when interface 2 has been abolished, constitute a basic catalytic unit capable of robust activity *in vitro* but has 20-fold lower activity *in vivo* (Hickman *et al.*, 2005).

It was hypothesized that the positively charged channels could be important for DNA binding (Hickman *et al.*, 2005). Building upon this idea I investigated the hexameric structure and found that on the rim of the circular wheel-like structure there resided a positively charged trench that knitted together the area between the active sites 35 Å apart; close to the 27 Å required to bind to an 8 bp target site. The residues that contributed to the positively charged trench were K292, K299, K300, and R306. These residues were also predicted to bind DNA when we used the DISPLAR bioinformatics server. This program predicts DNA binding residues based on the three dimensional

structure of the protein. I therefore came up with a model coined the “rim of the wheel” (ROW) which posits that the ends of the *Hermes* transposon bind in the positively charged channel directed towards the outside of the “wheel”. This positions the ends in the active sites approximately 35 Å apart where they could be integrated into the target DNA that we hypothesize is bound by the positively charged trench on the rim of the “wheel”.

Recently, a *Hermes* transposase was crystallized by Fred Dyda and Alison Hickman bound to double stranded *Hermes* transposon left end (LE) containing the first 15 bp of the upper strand and 16 bp of the lower strand (Figure 2.3) (F. Dyda and A. Hickman, personal communication). This new structure has provided support for the ROW model as well as insight into the oligomerization of active *Hermes* transposase and residues responsible for LE binding. This study has also crystallographically confirmed the role of W319 in stabilizing a flipped out base of the transposon LE (F. Dyda and A. Hickman, personal communication). One of the most striking features of this *Hermes* transposase structure is that it crystallized as an octamer that was shown to be active. Whether the hexamer or octamer is the naturally occurring active *Hermes* transposase is yet to be determined (Hickman *et al.*, 2005, F. Dyda and A. Hickman, personal communication). Another unexpected feature of the DNA-bound octamer is that it did not undergo a large degree of conformational change relative to the hexamer, although the area hypothesized to function in binding the target site DNA increased in size. In fact the residues with the largest degree of conformational change were between residues 297-317, with some residues moving up to 12.7 Å (Figure 2.4). Several residues within this

area comprise the positively charged trench that is part of the ROW model (Figure 2.5). As can be seen in Figure 2.3, the bound *Hermes* transposon LEs run along the positively charged channel of the transposase that was speculated by Hickman *et al.* (2005) to bind DNA. This positions the 3'-OH bearing ends of each DNA into the positively charged trench on the ROW. This is close to the ~ 27 Å needed for the ends to undergo nucleophilic attack on an 8 bp target site. The conformational flexibility required to reconcile this difference may occur upon target site binding. It should be noted that the other two transposases for which a DNA-bound crystal structures are available, *Tn5* and *Mu*, also have active site distances that are inconsistent with the size of their target sites by ~ 10 Å and ~ 13 Å, respectively (Davies *et al.*, 2000; Yuan *et al.*, 2005). The relationship of the positively charged trench of *Hermes* transposase with respect to target site binding has the possibility to provide important information about the mechanism by which the *Hermes* transposase integrates. Knowing which residues function in DNA target sequence binding may also allow for understanding of *Hermes* integration frequencies in different systems which is important if *Hermes* is to be used as a genetic tool to modify or manipulate genomes.

It is not yet known which residues in the *Hermes* transposase are responsible for binding to the target DNA. Studies with *Tn5* transposase have shown that the residues responsible for binding to target DNA were also responsible for binding to the flanking sequence of the transposon ends (Gradman *et al.*, 2008). *Tn5* has been a useful model for *Hermes* in understanding transposition chemistry, as both of these transposases contain an insertion domain that disrupts the RNaseH-like domain in the same location. While

the two insertion domains are structurally unrelated, it has been shown that the Tn5 insertion domain β -strands are important for binding the DNA hairpin that forms on the Tn5 transposon ends (Hickman *et al.*, 2010). The binding is facilitated by the Tn5's Trp298 which stabilizes a flipped-out base during excision; the same type of interaction that was found to be facilitated by the *Hermes* Trp319 as mentioned above. Therefore, it is conceivable that *Hermes* makes extensive contact with the donor DNA during excision and that the same residues responsible for interacting with donor DNA during excision may be important for target DNA or flanking DNA during integration.

The goal of this study is to determine which residues of the *Hermes* transposase are important for binding to the target site DNA during transposition. Using the crystal structure and bioinformatics program, DISPLAR, I identified several residues in *Hermes* hypothesized to be important for target site binding (Tjong and Zhou. 2007). Four mutations were made in residues comprising the positively charged trench and were tested in several different experiments designed to tease apart different aspects of the transposition pathway in order to characterize their role and function during *Hermes* transposition.

2.3 Results

2.3.1 Somatic tissue activity of the ROW mutants: Interplasmid transposition assay in injected Drosophila embryos.

To test the somatic transposition activity of each of the ROW mutants, a five-plasmid interplasmid transposition assay was performed in *Drosophila* embryos using the

pKH70new helper plasmid into which each of the mutants was cloned (Figure 2.6). The pKH70new helper plasmid is a modified version of the pkhsp70 plasmid used in interplasmid transposition as reported in Arensburger *et al.* (2005). Plasmids encoding wild-type (WT) *Hermes* and each of the mutant transposases were co-injected with the pHDG1 *Hermes* donor plasmid containing a gentamicin (Gent) selectable marker and LacZ gene flanked by the *Hermes* transposon TIRs and the pGDV1 target plasmid carrying chloramphenicol (Cam) resistance. These were injected with the *piggyBac* (*PB*) donor plasmid carrying the Gent selectable marker and EGFP gene flanked by the *PB* transposon TIRs and a *PB* helper plasmid encoding the *PB* transposase (Figure 2.6). *PB* functions as an internal control because the transposition frequencies of WT and mutants can be compared to each other relative to the transposition frequency of *PB* in each of the respective injections. Each set of injections was performed three times into 0-2 hr old *Drosophila* embryos. Following heat shock to activate expression of the transposase genes and 24 hr incubation, target plasmids were recovered. Plasmids conferring events were sequenced to determine the location and target site into which the donor transposons were inserted. Table 2.1 shows interplasmid transposition frequencies in injected *Drosophila* embryos for each of the ROW mutants and WT *Hermes* transposases. Each of the samples was co-injected with the *piggyBac* (*PB*) transposon donor and helper as an internal control used to normalize the data between different injection sets and samples that arise due to inherent variability in the assay. The frequency was calculated by dividing the number of independent events by the donor titer in each of the injection sets. The “Fold Difference” is the difference in the transposition rate between the *Hermes*

transposition frequency and the transposition frequency of the *PB* internal control with which it was co-injected. As can be seen from the Table 2.1, K299A has a 108-fold lower difference in transposition frequency than its corresponding co-injected *PB* frequency. Although each of the *Hermes* samples has a lower transposition frequency, K299A is the only mutant with a statistically significant decrease in transposition frequency compared to the internal *PB* control.

2.3.1.1. K300A has an altered target site preference.

At least one event was recovered for each of the ROW mutants of the *Hermes* transposase. If feasible, as an initial analysis, eight events (pGDV1 plasmids with Gent and Cam resistance) were sequenced from the *Hermes* LE. A consensus TSD sequence was then compiled using the Weblogo program (Figures 2.7 and 2.8) (Crooks *et al.*, 2004). If the initial consensus TSD sequence looked different from the WT *Hermes* transposon target site then up to 21 plasmids were sequenced. This number is required to determine whether the target site variation is statistically significant using the Fisher's exact test (Table 2.2). HmK300A was the only mutant that showed a statistically supported change in its TSD. As can be seen from the consensus, K300A has less of a preference for A in the 7th position of the TSD compared to WT *Hermes* transposase (Figure 2.7 and Table 2.3). Using the Fisher's exact test the P-value for this base was calculated to be 6.8×10^{-6} . In this test 0.05 is considered to be statistically significant and 0.01 to be very significant.

2.3.2. Germ line tissue activity of ROW mutants in *Drosophila*.

To assess the transposition dependent germline transformation frequency of the mutants in *Drosophila*, each of the ROW pKH70new helper plasmids and the *w+* gene donor plasmid were injected into pre-blastoderm CSW embryos (Figure 2.9). CSW are a strain of *Drosophila* that does not carry the *w+* gene and therefore have a white-eye phenotype. After a 16 hr recovery period the embryos were heat shocked at 37° C for 1 hr and then collected 24 hrs later. After eclosion each of the G0 flies were crossed with three non-transformed CSW virgin flies. Transformation was detected by the presence of red, orange, or yellow eye pigment in the G1 progeny. The G0 from which the red, orange or yellow G1 progeny came from was considered transgenic. The transposition dependent transformation frequency for each sample was calculated by dividing the number of transgenics (“# Transg”) by the number of fertile crosses (“Fertile”). As can be seen from Table 2.4, K292A, K299A and K300A mutants had significant reduction in their ability to transform *Drosophila* by transposition compared to WT *Hermes*, with K292A having a 30-fold decrease in transformation capability, and K300A having a 6.5-fold decrease (Figure 2.10). K299A was not able to produce transgenics whereas R306A had a transformation rate that is nearly WT.

2.3.3 Yeast Excision Frequency of ROW Mutants.

The excision capability of WT *Hermes* transposase, each ROW mutant transposase, and a negative control, pGalS, were tested *in vivo* in yeast using a mini-*Hermes* transposon from a donor plasmid. The mini-*Hermes* donor plasmid contains 711

of the 1,100 bp and TIR of the *Hermes* transposon LE and 512 of the 1,400 bp and TIR of the *Hermes* transposon right end (RE) flanking the *Hermes* ORF. Yeast assays, both excision and integration, were performed on each of the mutants in a high throughput manner. Each of the ROW mutants were cloned into the pGal vector and transformed into the yeast strain BY4727 with the SG15 excision plasmid. The SG15 plasmid contains a mini-*Hermes* transposon containing the essential TIRs interrupts a URA3 gene. *Hermes* transposase expression is under the control of a galactose promoter. After induction on galactose-containing medium, if an excision event occurs the yeast are able to grow in the absence of uracil. Transformed yeast grown on selection plates containing synthetic complete (SC) medium without tryptophan (-T) and histidine (-H) containing 2% Dextrose (Dex) were diluted and then plated on SC-T-H + 2% Galactose (Gal) followed by five day induction. Excision events were selected by growth on SC-T-H-U + 2% Dex for two days and the donor titer was calculated by making a five-fold serial dilution and plating on SC-T-H + 2% Dex. The excision frequency was calculated by dividing the number of SC-T-H-U + 2% Dex by number of SC-T-H + 2% Dex. This was done 10 times for each sample. As can be seen from Figure 2.11, both K292A and K299A mutant transposases had very low excision frequencies compared to WT *Hermes*. K300A and R306A had excision frequencies that were lower but not statistically different from WT *Hermes*. K292A in particular had an excision frequency that was close to the negative control, pGalS, which is a pGal vector without an ORF had an approximately 1,900-fold lower excision frequency than WT *Hermes*. K299A had a 70-fold higher

excision frequency than that of K292A and pGalS, but it was still 27-fold lower than WT *Hermes*.

2.3.4 Yeast Integration Frequency of ROW mutants.

Because the ROW residues are hypothesized to be involved in target site selection it is possible that a ROW mutant has the ability to excise a transposon but is unable to bind a target sequence and perform integration. To test this hypothesis and the general ability for each of the ROW mutants to perform integration *in vivo*, each of the ROW mutant transposases along with WT *Hermes* and pGalS were individually tested in a yeast integration assay. In this assay the pGal mutants, under the control of a galactose promoter, were transformed into the BY4727 yeast strain with donor plasmid SG30. After three days growth on SC-T-U + 2% Dex to select for transformants, each sample was re-plated on SC-T-U + 2% Gal. After a five day induction, 10 colonies of each mutant, WT *Hermes* and pGalS were plated on SC+FOA + 2% Dex and SC+FOA+CloNat + 2% Dex plates in order to calculate the number of excision events and select for integration events, respectively. The integration frequency is a calculation of the number of integration events divided by the number of excisions that occurred: number of SC+FOA+CloNat/ number of SC+FOA. As can be seen from Figure 2.12, the integration frequency of all samples was quite variable and the integration frequencies of K292A, K300A, and R306A were not statistically significant compared to WT *Hermes*. However, K299A had a 64-fold lower integration frequency than that of WT *Hermes*.

2.3.5 Strand transfer activity of ROW mutants is lower than WT in the presence of Mn^{2+} and Mg^{2+} .

E. coli produced His-tagged transposase was purified for each of the ROW mutants and WT *Hermes* (Figure 2.13). The purified protein was tested for its ability to integrate a pre-cleaved radiolabelled *Hermes* LE base pairs 1-30 into the pUC19 plasmid. The assay was performed in the presence of the divalent cations Mn^{2+} (1 mM) and Mg^{2+} (10 mM). Although Mg^{2+} is the natural cation, higher concentrations are needed to detect strand transfer events *in vitro* compared with Mn^{2+} . Because of conformational changes that are caused by the binding differences between Mg^{2+} and Mn^{2+} in the active site of DD[E/D] transposases it has been shown to have robust integration activity in the presence of Mn^{2+} and can often times suppress defective transposition mutations (Asante-Appiah et al., 1997, Junop and Haniford, 1996, Sarnovsky et al., 1996). As can be seen in Figure 2.14, K292A and K299A had noticeably lower LE integration activity than WT in the presence of 1 mM Mn^{2+} while K300A and R306A had close to WT activity. In the presence of Mg^{2+} only K292A seemed to have decreased activity while K299A, K300A, and R306A had near WT strand transfer activity.

2.3.6 Coupled cleavage and strand transfer activity of ROW mutants is lower than WT in the presence of Mn^{2+} and Mg^{2+} .

Purified ROW mutant transposases were tested for their ability to cleave and then integrate a 193 bp radiolabelled fragment containing 91 bp of *Hermes* LE and 102 bp of *Musca domestica* genomic flanking sequence. As can be seen in Figure 2.15, in the

presence of 1 mM Mn^{2+} both K292A and K299A had lower activity for integration into pUC19 compared to WT while K300A and R306A had near WT activity. However, in the presence of 10 mM Mg^{2+} , which is the natural divalent cation, K299A had undetectable strand transfer activity. K292A, K300A and R306A all had lower cleavage and strand transfer activity than WT. It should be noted however that WT *Hermes* transposase also had lower activity for cleavage and strand transfer in the presence of 10 mM Mg^{2+} than in Mn^{2+} .

2.3.7 Cleavage and strand transfer activity levels of ROW are not due to a change in their ability to bind to LE.

Because random mutations can have consequences on a proteins stability and because residues that are important for target DNA binding may also be responsible for excision it was important to make sure that the results from the *in vitro* assays were due to the functional importance of the mutated residues. Electromobility shift assays (EMSAs) were performed with both radiolabelled Hm LE 1-30 and Hm LE 193 to test WT and ROW mutant transposase binding activity. *E. coli* expressed and purified ROW mutant and WT transposase protein was incubated with either 1 pmol of Hm LE 1-30 or 1 pmol of Hm LE 193. As seen in Figure 2.16 and 2.17 there was no difference in the ability of any of the mutants to bind to either Hm LE 1-30 or Hm LE 193 compared to WT. Additionally, based on similar electrophoretic mobility of the DNA-protein complexes the mutant transposases appear to be forming the same higher ordered structures upon binding as the WT protein.

2.4 Discussion

2.4.1 *Hermes* Rim-of-the-Wheel Model

To determine which residues in *Hermes* transposase are responsible for locating and binding to the target site DNA we first used a bioinformatics program, DISPLAR, that predicts DNA binding residues based on the 3-D crystal structure of the protein (Tjong and Zhou, 2007). This analysis was used in conjunction with the hexameric crystal structure in my development of the ROW model to describe *Hermes* transposase target DNA binding. Investigation of the hexameric structure found that on the rim of the circular wheel-like structure there resided a positively charged trench that knitted together the area between the active sites 35 Å apart; close to the 27 Å required to bind to an 8 bp target site (Figure 2.2.b). The residues that contributed to the positively charged trench were K292, K299, K300, and R306 and were also identified using the DISPLAR server. I therefore came up with a model coined the “rim of the wheel” which posits that the ends of the *Hermes* transposon bind in the positively charged channel directed towards the outside of the “wheel”. This positions the transposon ends in the active sites, approximately 35 Å apart, where they could be integrated into the target DNA that we hypothesize is bound by the positively charged trench on the rim of the “wheel”. It is thought that this discrepancy may be reconciled by a conformational change during target site binding. The discrepancy between the distance of the ends and the size of target site that they bind and integrate into is not unique to *Hermes* as this is a feature seen with *Tn5* and Mu crystal structures (Davies *et al.*, 2000, Yuan *et al.*, 2005).

Each of the charged ROW residues were individually mutated to alanine to make each of the following mutants: K292A, K299A, K300A, and R306A. The mutants were all tested in various assays which allowed us to dissect the effects of each mutation on the different steps of transposition: 1) binding of transposon ends, 2) excision of the transposon end from donor backbone DNA, 3) binding of target site and integration of the transposon end. This was important, as it is conceivable that because the residue may be functioning in binding to the target DNA it very well may be that the mutant *Hermes* can bind and excise the transposon end from the donor DNA site yet not be able to bind and integrate that transposon end into target DNA. An as of yet unpublished *Hermes* transposase octamer crystal structure helped validate several aspects of our model (F. Dyda and A. Hickman, personal communication). The octamer was co-crystallized with the 15 bp (upper strand) and 16 bp (lower strand) of the *Hermes* LE by Fred Dyda and Alison Hickman who graciously allowed us access to the unpublished crystal structure (Figure 2.3). As hypothesized, the ends lead into the active sites from the center of the wheel making several important DNA-protein contacts along the positive channel. While the distance between active sites at interface 3 did not change, the area of the negatively charged trench increased as the interface appears to have “opened up” due to conformational changes as a result of LE binding. Most of the *Hermes* protein from the original published crystal structure overlaps perfectly with the octamer but there is one region in particular that appears to have undergone rather dramatic conformational changes in accommodating the binding of LE DNA. This region stretches from residue 294 to residue 317 and has movement up to 12.9 Å for some residues (Figure 2.4). This

region contains the four ROW mutants believed to be important for target site binding and also contains residues that make contact with the LE DNA itself.

Perhaps the most striking feature of the *Hermes* structure is its size. Data to this point support the octamer as the active species (F. Dyda and A. Hickman, personal communication). Its size, 560 kDa, is several orders larger than histones and may help explain why *Hermes* has a proclivity to integrate into nucleosome free regions in the yeast genome (Gangdharan *et al.*, 2010). Besides the fact that nucleosome free regions are inherently more accessible to DNA binding proteins, it would be difficult for such a large protein complex to interact with DNA wrapped around histones. Other fundamental questions arise from the octamer structure such as how the circular protein complex initially binds to the DNA. Because the ends of the DNA project from the center to the outside of the circle it is difficult to imagine how the remaining portion of the transposon – if we were to imagine that the left ends in the crystal structure were connected by some 2 kb of transposon DNA – would physically reside. One may hypothesize that the *Hermes* transposase first binds to a transposon end still bound to donor DNA and then either cleaves and assembles into an octamer or first forms the synaptic complex containing both of the transposon ends paired together in the octameric structure followed by cleavage. Because the *Hermes* cleavage mechanism is so closely related to V(D)J recombination it is hypothesized that the *Hermes* protein forms a synaptic complex before cleavage of the transposon from the donor DNA (Zhou *et al.*, 2005). It has been shown that monomer and dimer forms of *Hermes* are insufficient to perform transposition *in vivo* but further experimentation is needed to validate a step-wise

binding, synaptic complex formation and cleavage model that confirms the whether the octamer or hexamer is the active form (Hickman *et al.*, 2005).

2.4.2 Altered target site and activity of K300A

The *Drosophila* somatic tissue activity of each of the mutants versus WT in interplasmid transposition assays showed that although there was not a difference in activity between K300A and WT or *PB* there was a alteration in the target site preference: K300A had less of a preference for A in the 7th position compared to WT (Figure 2.7). This was seen in the target site consensus and was confirmed statistically by the Fisher's exact test. The effects of this mutation were difficult to tease apart as were the effects of the altered target site. The germline tissue transformation activity of K300A in *Drosophila* was 6.5-fold lower than WT, 4.4% versus 28.8% (Table 2.4 and Figure 2.10). The nucleotide most frequently found at the 7th position of the K300A target sites was A, nine times, followed by G which was found five times (Table 2.3). In fact only four of the 21 pGDV1 hot spots for WT integrations were hit by K300A (Table 2.2). There did not seem to be much of an effect by this mutation in yeast in either excision or integration assays compared to WT. Although the excision and integration frequencies were lower they were not statistically significant (Figure 2.11 and 2.12). With a more robust and less restricted target site selection ability one may hypothesize that K300A would be hyperactive for integration because it has more possibilities for binding to a target site, increasing the likelihood of integrating after it has excised. Gangharan *et al.* (2010) showed that there was a strong regional bias in the target site of

Hermes in vivo in yeast. They also found that the target site had long stretches of T on the 5' end of the target up to 200 bp and that there are long stretches of A on the 3' end of the target up to 200 bp when integrating into the yeast genome (Gangdharan *et al.* 2010). Stretches of T and A are also found to be higher in nucleosome free regions and upstream of genes which accounts for *Hermes* proclivity to integrate immediately upstream to ORF's and in nucleosome free regions. Gangdharan *et al.* (2010) also found that when given naked genomic DNA targeting was essentially random and that G/C rich regions seemed to be the more frequently targeted area, a result the researchers hypothesized was due to the higher flexibility of those regions versus T/A rich areas. The regional effects associated with *Hermes* targeting *in vivo* of the yeast genome may account for why we did not see a significant change in the excision or integration frequencies in yeast; although K300A has an altered target site in pGDV1 in *Drosophila* somatic tissue it still is restricted to the regional nucleotide composition of the yeast genome in accessible regions, such as nucleosome free areas which are T/A rich. Therefore, even though the number of sites that it can now target has increased the nucleotide composition that is accessible in yeast has not. These restrictive regional affects were not present in assays in *Drosophila* with pGDV1.

The *in vitro* activity of purified K300A protein showed a pattern for activity similar to its activity in yeast when compared to WT. While the altered target site preference of K300A would imply that it should have increased activity *in vitro* when supplied with naked plasmid DNA, that was not the case. As Figure 2.14 shows, the strand transfer activity of K300A was near that of WT. The coupled cleavage and strand

transfer activity of K300A showed the same result (Figure 2.15). That led us to question whether binding of the LE and LE containing flank could be a contributing factor to the lowered activity of K300A. In Gradman *et al.* (2008), the researchers found that *Tn5* has a bi-functional DNA binding region and that some amino acids that contact donor DNA also contact target DNA. The EMSA from incubation with Hm LE 1-30 show that K300A has the same affinity for LE as WT (Figure 2.16). Additionally, when K300A was incubated with Hm LE 193, which contains 101 bp of flanking genomic DNA, there was no difference from WT in the proteins affinity to bind to the fragment (Figure 2.17). Thus, the lowered activity of K300A compared to WT during transposition is not due to a change in the mutant transposases affinity for the LE nor to flanking donor DNA and this residue is therefore unlikely to be involved in binding to donor DNA. We can conclude in this case that an altered or less restricted target site does not equal higher transposition frequencies. The position of K300 in the octamer crystal structure shows that it is the only ROW residue that projects outward from the positively charged trench as opposed to in towards the trench as K292 and K299 do (Figure 2.18). This structural aspect supports its role in binding to the target DNA as it is properly positioned to contact DNA wrapping around the “rim” of the “wheel”. The results imply that K300 may be responsible for locating or binding to the target DNA during integration.

2.4.3 Activity of K292A

In *Drosophila* somatic tissue K292A had near WT transposition frequency into the pGDV1 plasmid in interplasmid transposition assays. Eleven of the events were

sequenced and there was no change in the target site consensus (Figure 2.8). The interplasmid transposition assays are subject to high variability due to several factors that effect their efficiency such as varying desiccation times due to humidity levels, difference in injection needle size, efficiency of plasmid recovery and electroporation. While the frequency is something that can be quite variable the assay has proven to be an effective method for looking at the target site selectivity of *Hermes* in *Drosophila* somatic tissue (Sarkar *et al.*, 1997). Although K292A did not vary much compared to both WT and the internal *PB* control transposition frequencies, it did vary significantly in *Drosophila* germline transformation frequency. As Table 2.4 shows only one transgenic K292A line was found out of 104 fertile crosses with a transformation frequency of only 0.96% compared to a WT transformation frequency of 28.8%; a 30-fold difference. This implies that K292 is an essential residue in transposition with severe consequences in the efficiency of *Hermes* to integrate into the germline of *Drosophila* when K292 is mutated to alanine. These results were consistent in the yeast excision assay but not in the integration assay Figure 2.11 and 2.12. Both of these results have been demonstrated before: (1) that elements can be active in yeast but not in *Drosophila* (Wright *et al.*, submitted), and (2) an element can have a lower excision frequency than integration in yeast (personal communication, Nancy Craig). The former is possibly due to interactions with host factors that are present in *Drosophila* but not yeast while the latter is hypothesized to be a consequence of imprecise excision events that make double strand repair of the excision difficult or impossible, the result being that the yeast is unable to repair the *URA3* gene and therefore can not grow in the absence of uracil and gives a

lower excision frequency than integration. K292A also had lowered activity in strand transfer and coupled cleavage and strand transfer activity, giving support to its role as an important residue for both excision and integration (Figure 2.14 and 2.15). As is true with all of the ROW mutants, K292A had the same affinity for the Hm LE 1-30 and Hm LE 193 that WT, demonstrated in EMSAs (Figure 2.16 and 2.17). This showed that the activity of K292A was not due to an altered affinity or ability to bind the *Hermes* left end and flanking DNA. This also demonstrated that the lower activity of this mutant was not due to a change in the stability of the protein. The fact that K292A did not have an altered target site does not exclude it from being an important residue for target site selection, as evidenced from the fact that mutation of K292 had such a dramatic effect on integration and excision strongly implies that it plays a critical role. One explanation for these results may be that K292 is important for binding target site and integration but that target site selection may be a concerted effort that depends on several residues and that not just one residue is important. Additionally, as evident from the biochemical analysis and results in yeast assays K292 may be important for excision as this appeared to be compromised in coupled cleavage and strand transfer but not integration as it was able insert the pre-cleaved transposon LE. As opposed to K300, K299 is positioned pointing in towards the positively charged trench (Figure 2.18). If the flanking DNA of the transposon ends were present in the octameric structure they would run straight out from the “wheel” and could possibly make contact with the K292 residue. These structural aspects of the crystal structure help support the data presented here.

2.4.4 Activity of K299A

Mutation of K299 had the most dramatic effect on activity in all of the assays used. Since only one event was recovered from interplasmid transposition assays in *Drosophila* no consensus sequence could be made to determine whether this residue had an effect on the target site sequence selection. Interestingly, the only event recovered was to nucleotide position 2154(+) in pGDV1, which is a hot spot for WT *Hermes* integrations (Sarkar *et al.* 1997). It should be noted also that K299A had a 108-fold lower transposition rate than both WT *Hermes* and the *PB* internal control (Table 2.1). This was the only mutant to have a statistically significant difference in interplasmid transposition frequency from *PB*. Additionally, in *Drosophila* germline K299A was apparently dead as no transgenic lines were created with this mutant (Table 2.4 and Figure 2.10). In addition to having no germline activity and a significantly lower interplasmid transposition frequency, K299A had just above negative control excision and integration frequency in yeast with a 27- and 64-fold lower frequency than WT *Hermes* in each assay respectfully (Figure 2.11 and 2.12). When provided Hm LE 1-30, K299A had strand transfer activity approximately the same as K292A but slightly lower than WT in the presence of 1 mM Mn^{2+} (Figure 2.14). However, in the presence of 10 mM Mg^{2+} K299A had higher strand transfer activity than K292A, though not as high as WT. This difference in activity between Mn^{2+} and Mg^{2+} has been demonstrated with other transposases and is due to conformational changes associated with differences in metal ion binding in the active site which give the transposase more robust activity in the presence of Mn^{2+} (Asante-Appiah *et al.*, 1997, Junop and Haniford, 1996, Sarnovsky *et*

al., 1996). The lowered activity of K299A was not due to an alteration in its ability to bind either LE 1-30 or LE 193 containing flanking sequence (Figure 2.16 and 2.17). However, there was a diminished ability for K299A to both cleave then integrate, as demonstrated in Figure 2.15. In the presence of Mn^{2+} the ability to perform coupled cleavage and strand transfer was about the same as K292A but much lower than WT. In the presence of 10 mM Mg^{2+} , which is the natural divalent cation, this activity was almost completely gone. This may be due to the fact that in the presence of Mn^{2+} *Hermes* has more robust activity than in the presence of Mg^{2+} (Zhou et al, 2005). Again, the fact that K299A binds the left end with the same affinity as WT demonstrates that the results were not a consequence of altered protein stability (Figure 2.16 and 2.17). It does appear, as demonstrated by the *in vitro* results, that the mutation of K299 does have an effect on the ability for *Hermes* to excise. This is supported by the fact that the strand transfer activity of this mutant was almost the same as WT *Hermes* transposase. Although K299 is positioned next to K300 it points into the positively charged trench just as K292 does (Figure 2.18). Again, if the flanking DNA of the transposon ends were present in the octameric structure they would run straight out from the “wheel” and could possibly make contact with the K299 residue. Therefore, it appears that three of the ROW mutants function in a bi-functional domain responsible for both excising the transposon (K292 and K299) and binding to target DNA (K300). This type of bi-functionality within a target DNA binding domain has been seen before with Tn5 (Gradman *et al.*, 2008).

2.4.5 Activity of R306A

Only four events were recovered in interplasmid transposition assays of R306A in somatic tissue in *Drosophila*. The consensus target site sequence for R306A was the same as WT but there were not enough events to use the Fisher's exact test to statistically determine whether there was a change or not (Figure 2.8). The transformation frequency in germline tissue of *Drosophila* was about half that of WT (Table 2.4 and Figure 2.10). In yeast excision and integration R306A had a lower but not statistically significant difference in frequency compared to WT (Figure 2.11 and 2.12). These results were also seen *in vitro* in strand transfer and couple cleavage and strand transfer activity in which R306A had slightly lower activity compared to WT in the presence of both Mn^{2+} and Mg^{2+} (Figure 2.14 and 2.15). The EMSA results show that left end binding and left end with flanking DNA binding affinity was not affected by the mutation (Figure 2.16 and 2.17). In conclusion the data does not support a role for R306 in target site sequence binding. Additionally, when comparing the position of R306 in the 2005 crystal structure to the octamer crystal structure it is apparent that due to conformational changes associated with LE binding the R306 residue moves out of the positively charged trench and no longer appears to be positioned in such a way that would support its role as a target DNA binding residue based on the ROW model (Figure 2.5 and 2.18) (Hickman *et al.*, 2005, F. Dyda and A. Hickman, personal communication).

2.5 Conclusions

Based on the fact that K300A had an altered TSD compared to WT *Hermes* and the functional data showing that the transposition frequency of K292A and K299A were compromised in several assays it appears that we have indeed targeted an important region in the *Hermes* transposase. The ROW model was based on the computer generated hexamer transposase crystal structure and bioinformatic analysis. This model was further supported with the *Hermes* octamer recently crystallized at NIH in which the *Hermes* LEs were binding in the positively charged channels and the 3'-OH of the ends were sitting in the active sites and positioned into the positively charged trench comprised of K292, K299 and K200 (F. Dyda and A. Hickman, personal communication). When K300 was mutated to alanine we did in fact have a change in the TSD in *Drosophila* somatic tissue. This altered target site binding does not appear to have a statistically significant change in the K300A somatic activity but does seem to have an effect on its germline activity for reasons that are still unknown. A general trend towards a slightly lowered activity in yeast was seen in both K300A's ability to excise and integrate but this difference was not statistically different from WT *Hermes*. It would be interesting to sequence the yeast integration events to determine whether the altered TSD also occurred in this system. Additionally, purified K300A protein showed slightly lower strand transfer activity in the presence of both Mn^{2+} and Mg^{2+} but quantification of these needs to be done to determine the extent of that difference compared to WT *Hermes*. Additionally, when the mutate was challenged to both cleave and then integrate the *Hermes* LE *in vitro* it had lower than *Hermes* coupled cleave and

strand transfer activity in the presence of both metal ions tested but especially in the presence of Mg^{2+} . This may have to do with the consequences of the conformational change associated with Mg^{2+} binding on the mutant's ability to cleave. This type of compromised activity in coupled cleavage and strand transfer assays was seen for all mutants and may be an effect of the concentration of Mg^{2+} . It may be possible to increase their activity under higher concentrations of Mg^{2+} but further experimentation is needed to determine why its activity in these assays is so compromised but not in others.

The K292A and K299A mutants each had compromised activity in most of the assays they were tested in. The data presented suggests that these residues are essential for *Hermes* transposition and that they may be involved in excision of the *Hermes* transposon. Since integration cannot occur without excision if these residues are playing a major role in excision that can be the reason that their transposition frequencies were compromised in *Drosophila* germline and in the yeast excision assay. K299A had the most compromised in every assay tested. The *in vitro* data with purified protein for both of these suggests that they may be playing a major role in excision since when they are supplied with pre-cleaved ends they had only slightly lowered integration frequencies compared to WT *Hermes*. But, when they had to cleave the LE in coupled cleavage and strand transfer assays their activity was severally diminished, especially in the presence of Mg^{2+} . More biochemical testing is needed to determine the exact role of these residues in transposition. As mentioned earlier, R306 does not appear to be an essential residue for *Hermes* transposition as it had near WT transposition in almost every assay tested but coupled cleavage and strand transfer. Again, more testing would need to be performed to

determine why in the presence of Mg^{2+} these mutants have such compromised activity in this assay versus WT. This residue comprised the positively charged trench in the 2005 crystal structure but in the octamer it appears that this residue has moved out of the positively charged trench due to conformational changes associated with LE binding (Hickman *et al.*, 2005; F. Dyda and A. Hickman, personal communication). The results presented here for K292, K299, and K300 support the ROW model. It data suggest that a region of the *Hermes* transposase has been identified that works as a bi-functional domain where K292 and K299 excise the transposon and K300 binds to target DNA.

2.6 Future experiments

The ROW transposase mutants presented here have begun to elucidate a region that appears to be important for target DNA binding and possibly cleavage of the *Hermes* transposon ends. In order to fully comprehend which residues in the *Hermes* transposase are directly responsible for binding to target DNA other positively charged residues in this region stretching for 292-318 should be mutated to alanine. As a first test, these transposase mutant proteins should be purified and their activity tested in biochemical assays for strand transfer, coupled cleavage and strand transfer, and EMSAs. Additionally, they should be tested in *in vitro* interplasmid transposition assays where events can be isolated and the TSDs can be sequenced and checked for differences compared to WT *Hermes*. Additionally, because target DNA binding presumably requires several residues it would be practical to make several mutations in the transposase at a time since mutation of only one residue may not have an effect.

A unique feature of the *hAT* superfamily of transposable elements is the fact that the *Buster* members of this family have a different 8 bp TSD, 5'-nnnTAnnn-3'. This is a feature that can be used in mutational studies in understanding Hermes transposition as the residues in the 292-318 region of the *Hermes* transposase can be compared to the residues in the *Busters* in a primary sequence alignment. Depending on the degree of sequence similarity, the residues in the *Hermes* transposase can then be mutated to match that of the *Busters* or the stretch this region can be swapped out of *Hermes* and replaced with either the corresponding *AeBuster* or *TcBuster* regions. Both of these *Buster* elements are active and would make great candidates. The *Hermes* transposases bearing the “*Buster-like*” mutations can then be tested for their activity in biochemical assays as mentioned above.

2.7 Materials and Methods

Strand-transfer reaction with pre-cleaved *Hermes*-L ends

Pre-cleaved *Hermes* LE primers were designed with an extra T-nucleotide on the 5' end of the top strand to mimic cleaved substrate 5'-TCAGAGAACAACAACAAGTGGCTTATTTTGATACTTATGCG-3' (top) and 5'-CGCATAAGTATCAAATAAGCCACTTGTTGTTGTTCTCTG-3' (bottom). Ten pmol of the bottom strand was radiolabeled at its 5' end with T4 polynucleotide kinase (Fermentas) and [γ -³²P] ATP (GE Healthcare Life Sciences). After cleaning with a Micro Bio-Spin P-30 Tris Chromatography Column (Bio-Rad) the *Hermes* LE substrates were made by annealing the oligonucleotides to form the double stranded oligo Hm LE 1-30.

The *Hermes* LE 1-30 was incubated with 140 nM *Hermes* transposase and 250 ng of pUC19 target plasmid in 25 mM MOPS pH 7.0, 1 mM MnCl₂ or 10 mM MgCl₂, 50 mM NaCl, 5% glycerol, 2 mM DTT, and 100 ng/μl BSA in a total volume of 20 μl. Reactions were incubated at 37° C for 1 hr. The reactions were stopped by addition of SDS and EDTA to 1% SDS and 20 mM EDTA and incubated at 65° C for 15 min. 5 μl of the DNA was then run on a 1% TBE agarose gel, dried and then exposed to film (Kodak).

Site-directed mutagenesis of *Hermes*

The site-directed mutagenesis was performed on pKH70new *Hermes* to obtain individual mutants (K292A, K299A, K300A, and R306A) using QuikChange site-directed mutagenesis kit (Stratagene). pKH70new is an altered form of the pkhsp70 encoding the *Hermes* transposase driven by the heat shock 70 promoter (Arensburger *et al.* 2005). The primers used were PAGE purified from Sigma:

HmLys292A For:

5'-GCCTATTCTTGCTTGCGCAAATATTGTAATAATTTCAAGAAAGCC-3'

Hm Lys292A Rev:

5'-GGCTTTCTTGAAATATTTTACAATATTTTGCGCAAGCAAGAATAGGC-3'

HmLys299A For:

5'-GCAAAAATATTGTAATAATTTTCGCGAAAGCCAATCTGCAGCACAGACTTCG-3'

Hm Lys292A Rev:

5'-CGAAGTCTGTGCTGCAGATTGGCTTTCGCGAAATATTTTACAATATTTTGC-3'

HmLys300A For:

5'-GCAAAAATATTGTAATAATTTCAAGGCAGCCAATCTGCAGCACAGACTTCG-3'

Hm Lys300A Rev:

5'-CGAAGTCTGTGCTGCAGATTGGCTTGCCTTGAAATATTTTACAATATTTTGC-3'

HmArg306A For:

5'CCAATCTGCAGCACGCACTTCGAAGTTCTTTAAAAAG-3'

HmArg306A Rev:

5'CTTTTAAAGAACTTCGAAG**TGCGTGCTGCAGATTGG**-3'

The codon changes of either a K or R residue to an A are underlined and in bold. All of the constructs were sequenced to verify that the desired mutations were present and that the complete *Hermes* transposase ORF was intact and that no undesired mutations or deletions had occurred during the engineering.

Wild-type and mutant *Hermes* transposase expression and purification

Each of the mutant transposase ORFs were cloned into the *PBHTH* plasmid which contains the WT *Hermes* transposase in a *PBAD* (Invitrogen) construct containing an N-terminal His-tag using *BglII* and *AfeI* sites.

Each of the constructs were individually transformed into LMG cells and grown overnight at 30° C on LB plates containing 100 mg/ml carbenicillin. A colony was picked the next day and inoculated into 20 ml LB medium containing 100 mg/ml carbenicillin and grown overnight at 30° C with shaking. The following day the overnight culture was diluted 1:100 with fresh LB medium + carbenicillin and was grown at 30° C with shaking until the OD 600 nm = 0.6. The culture was then put at 16° C and protein expression was induced with the addition of 0.1% arabinose for 16 hr. After induction each of the cells were washed with centrifugation at 4° C with TSG (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 10% v/v glycerol), and frozen in liquid nitrogen. The frozen cells were resuspended in 10 ml TSG at 4° C and lysed using a French Press, psi = 1250. All subsequent steps were performed at 4° C unless otherwise noted. The lysates were cleared by centrifugation and then filtered using a 0.45 µm sterile filter (Nalgene)

before being loaded onto a pre-equilibrated Ni²⁺ Sepharose column (Amersham) and washed with ten column volumes of TSG with 10 mM imidazole, six column volumes of TSG with 50 mM imidazole and six column volumes of 100 mM imidazole. Each of the *Hermes* protein samples were eluted with five column volumes of TSG with 200 mM imidazole, dialysed against TSG, and stored at -80°C.

Fly stocks

A strain of *Drosophila melanogaster*, Canton-S White (CSW), was used for interplasmid transposition assays and transformation. The flies were raised on fruit fly media and supplemented with dry active yeast. The laboratory strain is maintained in the Atkinson laboratory at University of California, Riverside.

Plasmid constructs for yeast excision and integration assays

Each of the ROW mutant transposase constructs were made through the excision of the transposase from pKH70new using sticky-ended *SpeI-XhoI* and then each fragment was ligated into the vector pGal*Hermes*. pGal*Hermes* was derived from the p414GalS plasmid with the addition of *Hermes* transposase between base pairs 3245 and 3246 (Mumberg *et al.* 1994).

***piggyBac* Donor Plasmid**

The *PBacGoEGFP* donor plasmid used as the internal control for normalization in the interplasmid transposition assays was constructed by PCR of the left and right ends of

PB from the *PBac*[3xP3-EGFPafm] plasmid and included the TSD of the flanking sequence. The left and right end fragments were digested and inserted into *PBluescript* II KS+ to create *PBSPBacLR*. *PBSPBacLR* was linearized by digestion with *XbaI* to incorporate a *Nhe* I linearized pGENToriEGFP plasmid to create the *PBacGoEGFP* donor plasmid. This construction creates a *PBacGoEGFP* donor plasmid that contains the EGFP and GENT ORFs flanked by the left and right ends of the *piggyBac* transposon (Smith 2007).

***piggyBac* Helper Plasmid**

The *piggyBac* helper is phsp70-Bac (formerly *PBhs*Δ*Sa*) (Handler *et al.* 1998). The phsp70-Bac plasmid contains the ampicillin resistance gene in its backbone. The transposase helper plasmid expresses the *piggyBac* transposase whose expression is under the control of the *D. melanogaster* hsp70 promoter.

Target Plasmid

The plasmid pGDV1 functions as the interplasmid transposition assays target plasmid. pGDV1 is a *Bacillus subtilis* low copy plasmid (Sakar *et al.* 1997) that is unable to replicate within *E. coli* without the addition of an *E. coli* origin of replication. The origin of replication is provided by donor plasmid during a transposition event.

***Hermes* Donor Plasmid**

The pHDG1 donor plasmid was constructed with the full-length left and right ends flanked by the 8 bp TSD from the pHERKS3 plasmid. pHERKS3 is formally known as p*Hermes*KSacO α (Sarkar *et al.*, 1997). pHDG1 does not contain any of the *Hermes* ORF. Instead the left and right ends flank an *E. coli* origin of replication and a gentamycin resistance gene from pGentOri α .

***Hermes* Helper Plasmid**

pKh70new is an altered form of the pkhsp70 (Arensburger *et al.* 2005). The plasmid was designed with a *SpeI* and *XhoI* site on either side of the ORF to allow for easier cloning. The plasmid has a kanamycin resistance gene in its backbone and the *Hermes* transposase ORF under the control of the hsp70 promoter.

Interplasmid Transposition Assay Mix

Because of the inherent variability involved with microinjection of *Drosophila* it is necessary to have an internal control that can be used to normalize the data and eliminate the background noise. In order to compare the transposition frequency across several injections a five-plasmid interplasmid transposition assay was developed (Smith 2007). The five-plasmid assay involves the co-injection of 250 ng/ μ l each of the *piggyBac* donor and helper plasmids along with 250 ng/ μ l each of the *Hermes* donor and helper plasmids and 1 μ g/ μ l of pGDV1 target. Thus, the transposition frequencies of WT and mutant *Hermes* can be compared to each other relative to the transposition frequency

of *piggyBac* in each of the respective injections. The transposition events of *Hermes* can be distinguished from *piggyBac* because the *Hermes* donor plasmid contains the LacZ gene flanked by *Hermes* TIR's while the *piggyBac* donor contains the EGFP gene flanked by *piggyBac* TIRs. Therefore, *Hermes* events express an azure pigment phenotype while the *piggyBac* events glow green under fluorescence.

Transformation Injection Mix

The injection mix for transformation of *Drosophila* contains 250 ng/ μ l of the pKH70new containing either WT *Hermes* or one of the four ROW mutants and 250 ng/ μ l of the *PBSHermes w+* plasmid, which contains the white-eye gene flanked by the *Hermes* left and right TIR's.

Microinjection of *Drosophila melanogaster*

The microinjection protocol applies for both interplasmid transposition and transformation experiments except when specified. The *D. melanogaster* were placed on pineapple agarose plates to induce oviposition and pre-blastoderm embryos were collected <45 min. of oviposition. The pre-blastoderm embryos were dechorionated with 60% bleach and washed with dH₂O followed by lining up of the embryos on glass slides and covering them in halocarbon oil to prevent desiccation. Using a Flaming/Brown Micropipette Puller, borosilicate glass capillaries 0.7 mm in diameter were pulled and filled with either the interplasmid transposition injection mix or transformation injection mix depending on the experiment. An Eppendorf Femtojet was used for all injection to

deliver the plasmid mix into each embryo. Post-injection each slide of embryos was placed in a humidity chamber under 100% oxygen for 15-20 hrs. All embryos were then heat shocked for 1 hour at 37° C. The embryos were then processed by either the interplasmid assay protocol or transformation protocol as described below.

***Drosophila melanogaster* Interplasmid Transposition Assay**

After the embryos were heat shocked at 37° C for one hour as described above they were allowed to rest at room temperature for one hour before processing. Each of the live embryos were collected from the glass slides and placed into 100µl grind buffer containing 0.5% SDS, 80 mM NaCl, 160 mM sucrose, 60 mM EDTA and 120 mM Tris-HCl, pH 9.0. The embryos were then ground up into a lysate and incubated for 30 min at 65° C. Next, 8 M KAc was added to a final concentration of 1M and the embryos were placed on ice for one hour. The embryo lysate was spun for 10 min at >14k rpm at room temperature followed by transfer of the supernatant to a new microcentrifuge tube. Two volumes of 100% ethanol was added to the supernatant, vortexed and put at -20°C overnight. The next day the solution was spun at >14k rpm and to pellet the recovered DNA. The supernatant was removed and the pellet was washed twice with 70% ethanol with spinning at >14k rpm for 5 min. The supernatant was removed and the pellet was allowed to dry before being resuspended in 1ul of dH₂O for every 10 embryos that were collected. The sample was electroporated into DH10b *E. coli* (Gibco-BRL). Of these cells 1/200th were spread on LB Amp/Gent plates containing IPTG and X-GAL and incubated for 24 hrs at 37° C to calculate the donor titer. The remaining cells were

streaked onto LB Cam/Gent plates containing IPTG and X-GAL and incubated at 37° C for 3 days. The *Hermes* transposition frequency was calculated by counting the number of azure colonies on Cam/Gent plates and dividing that by the number of azure colonies from the Amp/Gent plates (multiplied by 200). The *piggyBac* transposition frequency was calculated by counting the number of EGFP colonies on the Cam/Gent plates and dividing by the number of opaque colonies from the Amp/Gent plates (multiplied by 200). All of the events were grown to miniprep volume and tested first by plating an aliquot onto LB Amp plates. The cultures that failed to grow were miniprep'd and digested. The *Hermes* samples that digested as events were sequenced at the University of California, Riverside Genomic Core Facility to determine the integration site into the pGDV1 target plasmid using the HL216 plasmid. Those events that did not pass each of the tests were not counted as events when calculating the transposition frequency.

***Drosophila melanogaster* Transformation Injections**

After the embryos were heat shocked at 37° C for one hour as described above they were allowed to recover at room temperature for 24 hrs to ensure that the embryos collected were alive and healthy. Each live embryo was put into a vial containing fruit fly media until they eclosed – about 2 weeks. G0 flies were collected <8 hrs post eclosion and separated into individual vials so that they could not mate with each other. Each individual was crossed with 3 virgin flies from non-injected fly stocks and were discarded once their progeny began to pupate. All G1 progeny were screened for the presence of the white-eye gene, which has a phenotype ranging from yellow to red eye

color. Even if only one G1 fly had the white-eye phenotype the G0 was considered transgenic. To calculate the transformation frequency for each mutant the number of transgenic G0's was divided by the number of fertile G0's.

Sequencing of Transposition Events

All transposition events were sequenced at the University of California, Riverside Genomics Core Facility and processed to analyze insertion sites and TSD. Sequencing of transposition events was performed using Sanger sequencing on an Applied Biosystems 3730xl machine.

Yeast Excision Assay for Determining Excision Frequency

To measure the excision frequency of the ROW mutants, WT *Hermes* transposase, and the negative control pGalS the donor plasmid pSG15 was used. pSG15 contains a URA3 gene interrupted by a mini-*Hermes* transposon that contains 711 bp of the *Hermes* LE and TIR and 512 bp of the *Hermes* right end and TIR.

Each of the rim-of-the-wheel mutants, WT *Hermes*, and the negative control p414GalS, which is an empty p414Gal vector, were transfected into yeast strain BY4727 (MAT α his3 Δ 200, leu2 Δ 0, lys2 Δ 0, met15 Δ 0 trp1 Δ 63 ura3 Δ 0) using the PEG/Lithium acetate method (Mumberg *et al.* 1994). Colonies were then streaked on Synthetic Complete media lacking Tryptophan and Histidine (SC-T-H) plates containing 2% dextrose (Dex) and incubated at 30°C for 3 days to select for the pGal (contains Trp) and pSG15 (contains His) co-transfected plasmids. To measure the integration frequency a single colony from each plate was resuspended in water, serially diluted 1000 fold and

streaked onto SC-T-H plates containing 2% galactose (Gal) and grown for 5 days at 30°C to induce expression of *Hermes* transposase. As a negative control each of the samples was also plated on SC-T-H plates containing 2% Dex, which should not induce protein expression. After 5 days 10 colonies from each plate, both induced and non-induced, were individually resuspended in 500 µl water, 50 µl were then serially diluted 10⁵ fold. Next, 100 µl of the serial dilution was plated on SC-T-H + 2% Dex to determine the total number of donor plasmids and the entire original resuspended colony was plated on SC-T-H-U + 2% Dex to determine the number of excised donor plasmids. The frequency of transposition is the ratio of the number of colonies on SC-T-H-U divided by the number of colonies on the SC-T-H.

Yeast Integration Assay for Determining Integration Frequency

In order to select for integrations a yeast ARS CEN donor plasmid, pSG30, was used which contains a URA3 marker, and *Hermes*-NatMX transposon. The NatMX cassette confers resistance to antibiotic clonNat and is flanked by 711 bp of the *Hermes* LE and TIR and 512 bp of the *Hermes* right end and TIR (Gandharen *et al.*, 2010).

Each of the ROW mutants, WT *Hermes*, and the negative control pGalS were transfected into yeast strain BY4727 (MAT α his3 Δ 200, leu2 Δ 0, lys2 Δ 0, met15 Δ 0 trp1 Δ 63 ura3 Δ 0) using the PEG/Lithium acetate method. Colonies were then streaked on SC-T-U plates containing 2% Dex and incubated at 30° C for 3 days to select for the pGal (contains Tryp gene) and pSG30 (contains Ura) co-transfected plasmids. To measure the integration frequency a single colony from each plate was resuspended in water, serially

diluted 1000 fold and streaked onto SC-T-U plates containing 2% Gal and grown for 5 days at 30° C to induce expression of *Hermes* transposase. As a negative control each of the samples were also plated onto SC-T-U plates containing 2% Dex which should not induce protein expression. After 5 days 10 colonies from each plate were individually resuspended in 500ul water, 50ul were then serially diluted 10⁴ fold. Next, 100ul of the serial dilution was plated on SC+5-FOA (1 mg/mL) + 2% Dex to determine the total number of plasmid-free cells and the entire original resuspended colony was plated on SC+5-FOA+Clon- NAT (100 µg/mL) + 2% Dex to determine the number of integrants. The frequency of transposition is the ratio of the number of colonies on SC+5-FOA+Clon- NAT divided by the number of colonies on the SC+5-FOA.

Electrophoretic Mobility Shift Assays (EMSA)

Each of the probes were created by labeling 1 pmol of double-stranded DNA with T4 polynucleotide kinase (Fermentas) and [γ -³²P] ATP (GE Healthcare Life Sciences) and purified using Micro Bio-Spin P-30 Tris Chromatography Column's (Bio-Rad). For each EMSA 20 fmol of labeled probe was incubated with 140 nM of His-tagged *Hermes* protein (see purification method) in a 10 µl reaction containing 15 mM HEPES, 2 mM DTT, 2 µg BSA, 0.5 µg poly(dI-dC):poly(dI-dC), 0.4 µg T3 single-stranded oligo. Each protein sample was first incubated with buffer for 15 min. at room temperature followed by addition of the labeled probe and incubation at room temperature for 20 min. Samples were then run on a 5% TBE polyacrylamide gel (Bio-Rad) at 100 volts and 4°C. The gels were dried and exposed to film (Kodak).

Probes:

Hm LE 1-30 (Sigma):

5'-TCAGAGAACAACAACAAGTGGCTTATTTTGATACTTATGCG-3' (top) and

5'-CGCATAAGTATCAAAATAAGCCACTTGTTGTTGTTCTCTG-3' (bottom) were annealed prior to radiolabeling.

Hm LE 193:

Hm LE 193 was cloned from pHERKS3. This fragment is 193 bp in length and contains 90 bp of the *Hermes* left end and 103 bp of flanking genomic DNA from the housefly *Musca domestica* which was cloned into the pHERKS3 plasmid (Sarkar *et al.*). The fragment was excised from a 5% TBE polyacrylamide gel (Bio-Rad) and recovered from the gel.

Hm RE 47-76 (Sigma):

(76) 5'-CAAAAGGCTTGACACCCAAAACACTTGTGC-3' (47) top and

5'-TCACAAGTGTTTTGGGTGTCAAGCCTTTTG-3' bottom strand were annealed prior to radiolabeling.

Coupled Cleavage and Strand Transfer Assay

The coupled cleavage and strand transfer reaction was performed using a 193 bp fragment of DNA containing 91 bp of the *Hermes* left end and 102 bp of flanking genomic DNA from the housefly *Musca domestica* which was cloned into the pHERKS3 plasmid (Sarkar *et al.*, 1997). The fragment was excised from a 5% TBE polyacrylamide gel (Bio-Rad) and recovered from the gel. Next 1 pmol of the fragment was radiolabeled

on the 5' end of both strands with T4 polynucleotide kinase (Fermentas) and [γ - 32 P] ATP (GE Healthcare Life Sciences) and purified using Micro Bio-Spin P-30 Tris Chromatography Column's (Bio-Rad). For each reaction 20 fmol of labeled probe was incubated with 140 nM *Hermes* transposase and 250 ng of pUC19 target plasmid in 25 mM MOPS pH 7.0, 1 mM MnCl₂ or 10 mM MgCl₂, 50 mM NaCl, 5% glycerol, 2 mM DTT, and 100 ng/ μ l BSA in a total volume of 20 μ l. Reactions were incubated at 37°C for 1h. The reactions were stopped by addition of SDS and EDTA to 1% SDS and 20 mM EDTA and incubated at 65°C for 15 min. Five microliters of the DNA was then run on a 1% TBE agarose gel, dried and then exposed to film (Kodak).

2.8 References

- Asante-Appiah, E., Skalka, A.M. (1997) A metal-induced conformational change and activation of HIV-1 integrase. *Jour. Of Biol. Chem.* **272**: 16196-16205
- Arensburger, P.A., O'Brachta, D.A., Orsetti, J. and Atkinson, P.W. (2005) *Herves*, a new active *hAT* transposable element from the African malaria mosquito *Anopheles gambiae*. *Genetics*. **169**: 697-708.
- Craig, N.L. (2002) Mobile DNA: an introduction. In: Craig, R., Gellert, M., Lambowitz, A.M. (Eds), *Mobile DNA II*. ASM Press, Washinton, DC. 3-11.
- Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004) Weblogo: A sequence logo generator. *Genome Research*. **14**: 1188-90.
- Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., Rayment, I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*. **289**: 77-85.
- Gangdharan, S., Mularoni, L., Fain-Thorton, J., Wheelan, S.J., Craig, N.L. (2010) DNA transposon *Hermes* inserts into DNA in nucleosome-free regions in vivo. *Proc. Natl. Acad. Sci.* **107**: 21966-21972.
- Gradman, R.J., Ptacin, J.L., Bhasin, A., Reznikoff, W.S. and Goryshin, I.Y. (2008) A bifunctional DNA binding region in Tn5 transposase. *Molec. Microbiol.* **67**: 528-540.
- Handler, A.M., McCombs, S.D., Fraser, M.J. and Saul, S.H. (1998) The lepidopteran transposon vector piggyBac, mediates germ-line transformation in the Mediterranean fruit fly. *Proc Natl Acad Sci.* **95**:7520-7525
- Hickman, A.B., Chandler, M., Dyda, F. (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Critical Reviews in Biochem. and Mol. Biol.* 1-20.
- Hickman, A.B., Perez, Z.N., Zhou, L., Primrose, M., Ghirlando, R., Hishaw., J.E., Craig, N.L., Dyda, F. (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.* **12**: 715-721.
- Jasinskiene, N., Coates, C.J. and James, A.A. (2000) Structure of *Hermes* integrations in the germline of the yellow fever mosquito, *Aedes aegypti*. *Insect Molec. Biol.* **9**: 11-18.

- Junop, M.S., Haniford, D.B. (1996) Multiple roles for divalent metal ions in DNA transposition: distinct stages of Tn10 transposition have different Mg²⁺ requirements. *EMBO Jour.* **15**: 2547-2555.
- Michel, K., O'Brochta, D.A., Atkinson, P.W. (2003) The C-terminus of the *Hermes* transposase contains a protein multimerization domain. *Insect Biochem. Molec. Biol.* **33**: 959-970.
- Mumberg, D., Muller, R., Funk M. (1994) Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucl. Acids Res.* **22**: 5767-68.
- Nesmelova, I.V. and Hackett, P.B. (2010) DDE transposases: Structural similarity and diversity. *Advanced Drug Delivery Reviews.* **62**: 1187-95.
- O'Brochta, D.A., Warren, W.D., Saville, K.J., Atkinson, P.W. (1996) *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica*. *Genetics.* **142**: 907-914.
- Sarkar, A., Coates, C.J., Whyard, S., Willhoeft, U., Atkinson, P.W., O'Brochta, D.A. (1997) The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica.* **99**: 15-29.
- Smith, R.C. (2007) Germ line expression of genetic markers and transposases in the yellow fever mosquito *Aedes aegypti*. University of California, Riverside
- Sarnovsky, R.J., May, E.W., Craig, N.L. (1996) The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO Jour.* **15**: 6348-6361.
- Tjong, H., Zhou, H.-X. (2007) DISPLAR: an accurate method for predicting DNA-binding sites on protein surfaces. *Nucl. Acids Res.* **35**: 1465-77.
- Wright, J.W., Smith, R.C., Xie K., Craig, N.L. and Atkinson, P.W. Tni:hyPBse7 transposase behavior in *Drosophila melanogaster* and *Aedes aegypti*. *Insect. Biochem. Mol. Bio.* (Submitted).
- Yuan, J.F., Beniac, D.R., Chaconas, G. Ottensmeyer, F.P. (2005) 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. *Genes Dev.* **19**: 840-852.

Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F. and Craig, N.L. (2004) Transposition of *hAT* elements links transposable elements and V(D)J recombination. *Nature*. **432**: 955-1001.

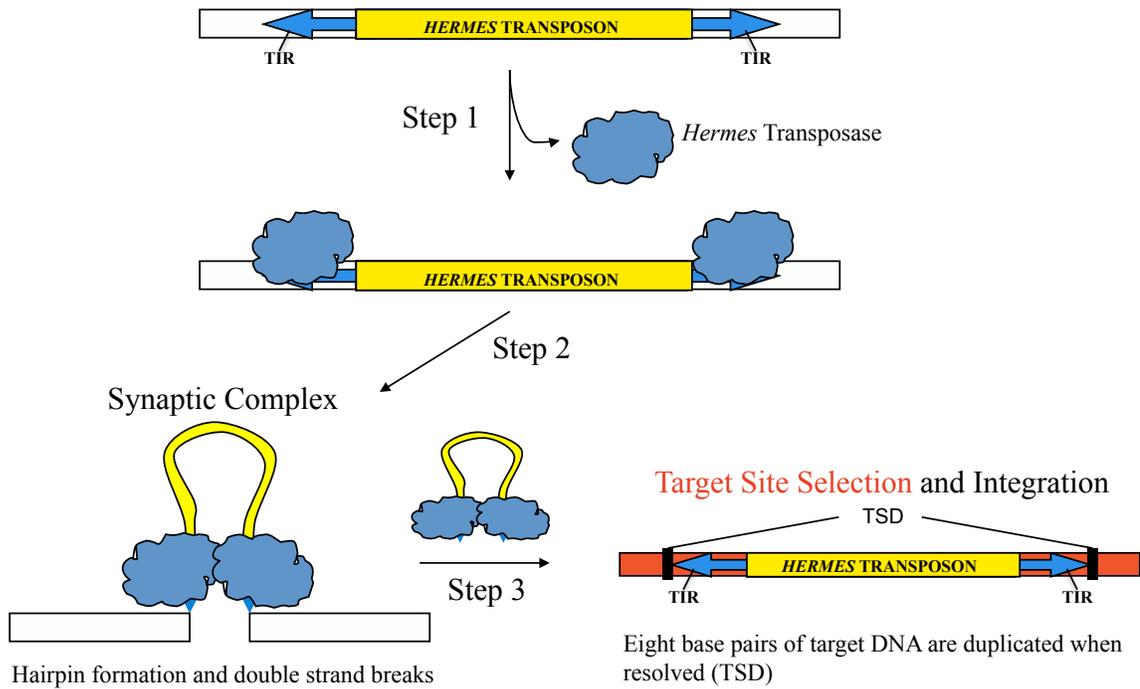


Figure 2.1: The transposition mechanism of *Hermes*. After translation of the *Hermes* transposase, the protein binds to the terminal inverted repeats (TIRs) of the *Hermes* transposon. It is then thought that the *Hermes* transposases undergo oligomerization to form the synaptic complex which is then followed by excision of the transposon and binding of a target site and integration of the transposon in a new location. The target sites are duplicated when resolved to create the target site duplication (TSD).

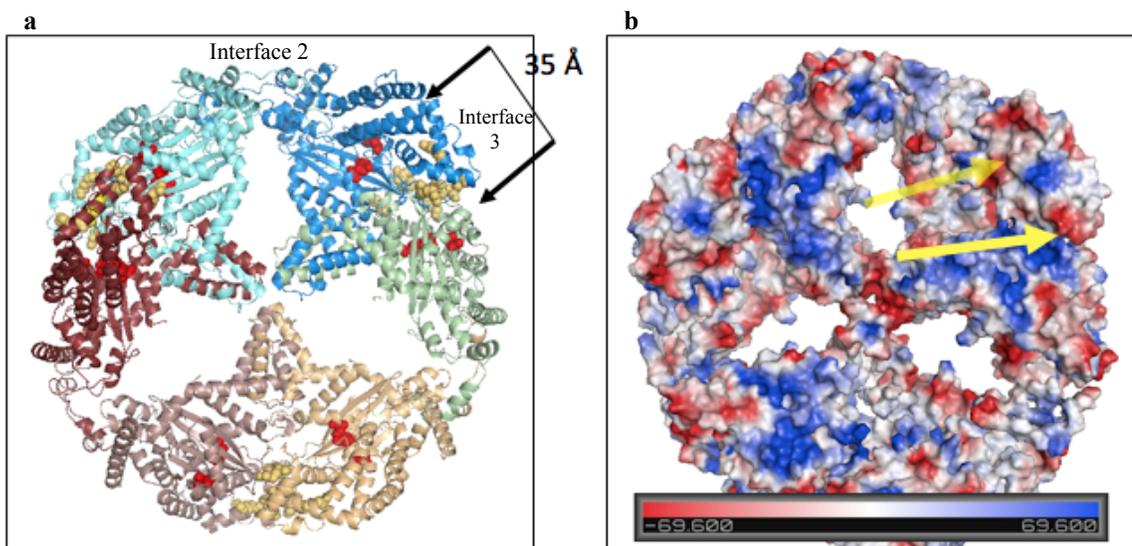


Figure 2.2: *Hermes* transposase hexamer crystal structure. (a) The 2005 *Hermes* hexamer structure showing the active sites in red and the “rim of the wheel” (ROW) residues in yellow (Hickman et al., 2005). The active sites are 35 Å apart. (b) A surface charge representation of *Hermes* showing the positively charged channel (in blue) leading away from the center of the hexamer into the active sites and towards the “rim-of-the wheel” (Hickman et al., 2005). These were hypothesized to be the binding site for the left and right transposon ends (in yellow) which would put the attacking nucleophile of the excised transposon ends into the positively charged trench comprised of the ROW residues. The ROW residues are hypothesized to be responsible for binding the target DNA.

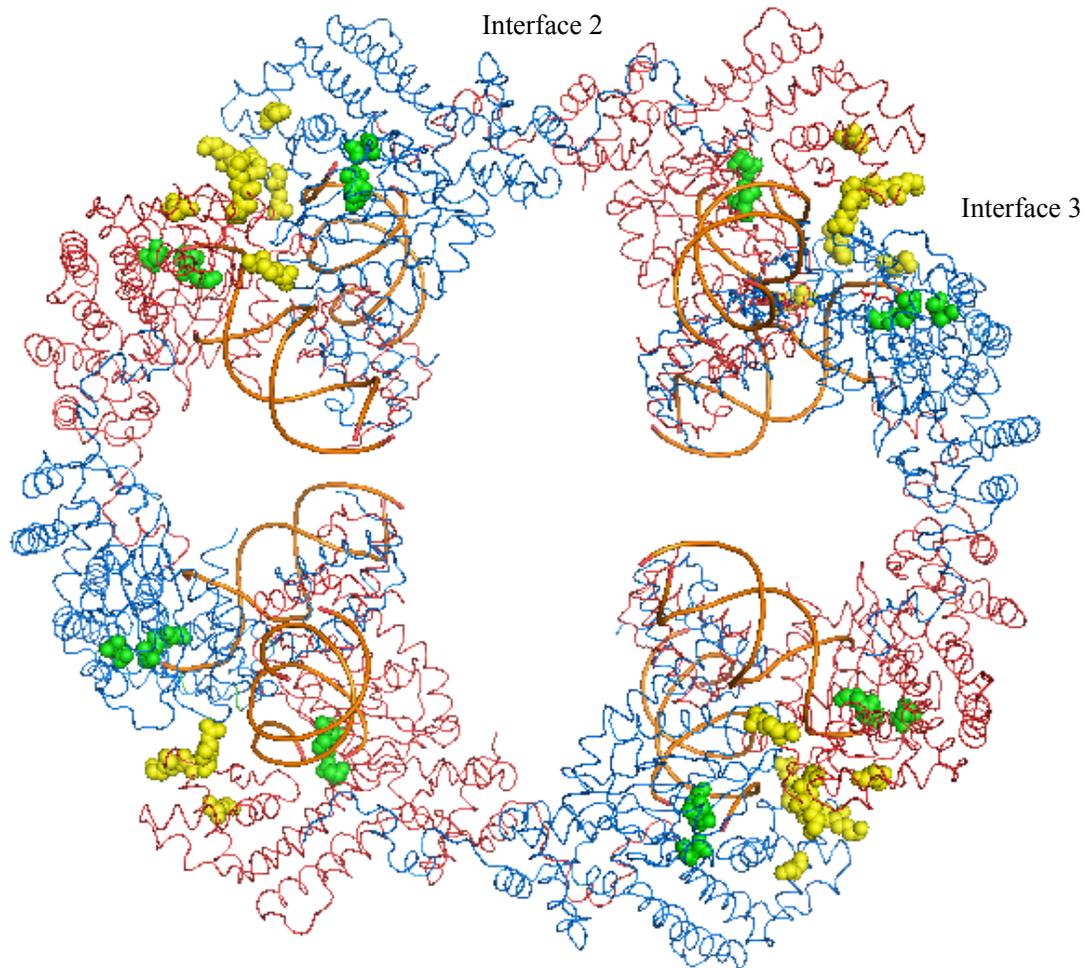


Figure 2.3: The *Hermes* transposase octameric structure. The *Hermes* transposase octamer structure crystallized by Fred Dyda and Alison Hickman at NIH in 2011 (F. Dyda and A. Hickman, personal communication) . Each of the four transposase dimers are bound to the *Hermes* transposon left ends, in orange. The monomer in each dimer pair is shown in red and blue. The DDE active site residues are shown in green and the ROW residues are shown in yellow. The octamer was shown to be active, (F. Dyda and A. Hickman, personal communication).

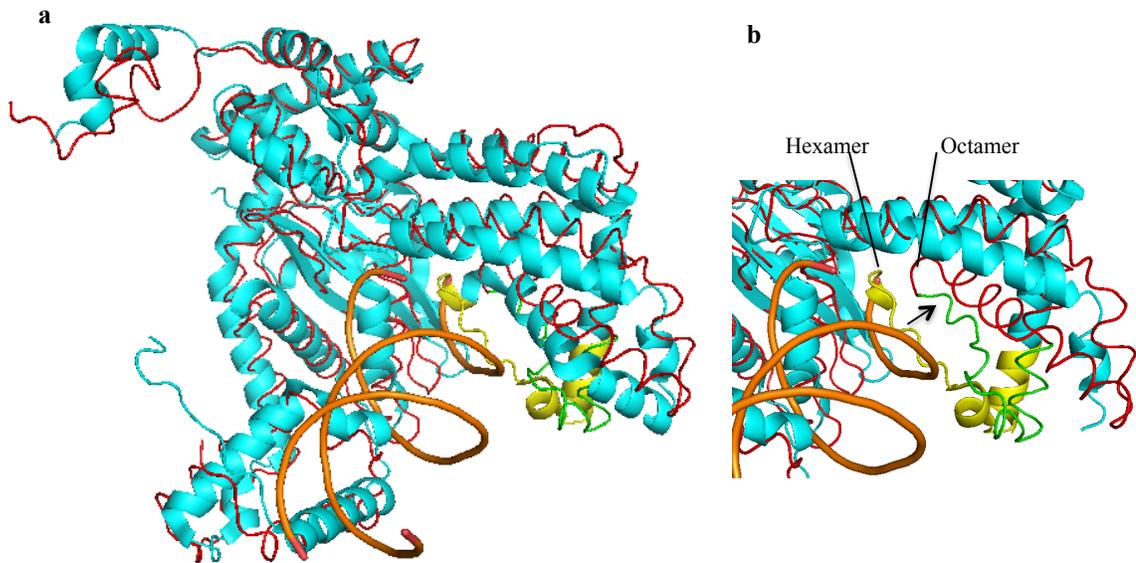


Figure 2.4: *Hermes* transposase crystal structure overlay. (a) An overlay of the *Hermes* transposase monomer from the 2005 crystal structure shown in blue with a *Hermes* transposase monomer from the octamer crystal structure bound to the *Hermes* transposon left end fragment, shown in orange (Hickman *et al.*, 2005; F. Dyda and A. Hickman, personal communication). (b) The region from 294-317, shown in yellow in the hexamer structure and green in the octamer, did not overlap very closely as the region had undergone a conformational change between 3.8-12.9 Å to facilitate transposon left end binding.

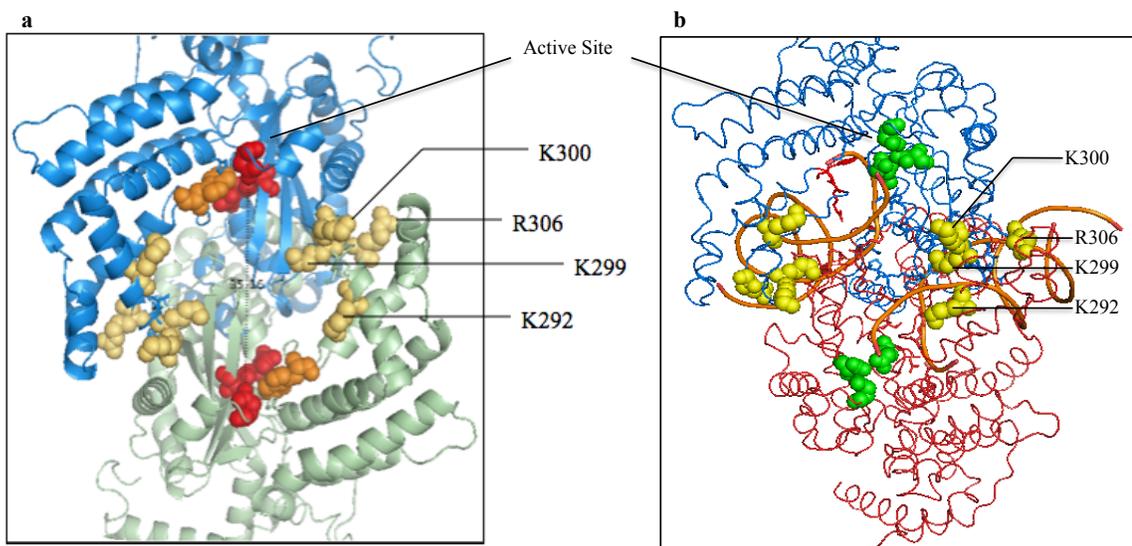


Figure 2.5: “Rim of the wheel” residues. (a) The side view of the *Hermes* transposase hexamer showing the “rim of the wheel” (ROW) residues in yellow and the active sites in red (Hickman *et al.*, 2005). (b) The side view of the *Hermes* transposase octamer showing the ROW residues in yellow and the active sites in green (F. Dyda and A. Hickman, personal communication). The back bone of the bound transposon left ends are shown in orange and project into the positively charged trench which is made up of the ROW residues. R306 has moved out of the trench during the conformational change and does not appear to be in a position to bind to the target DNA any longer.

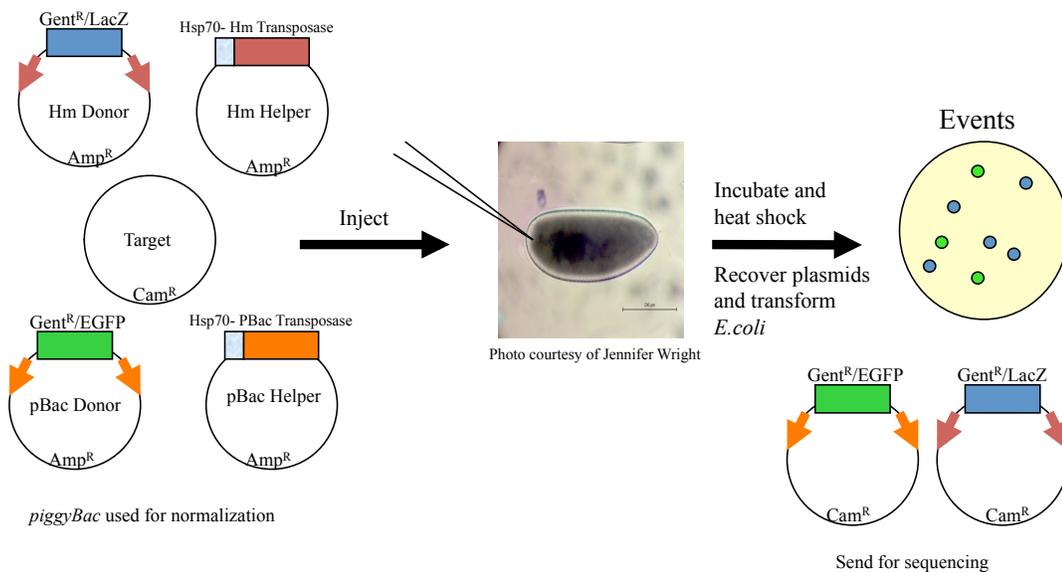
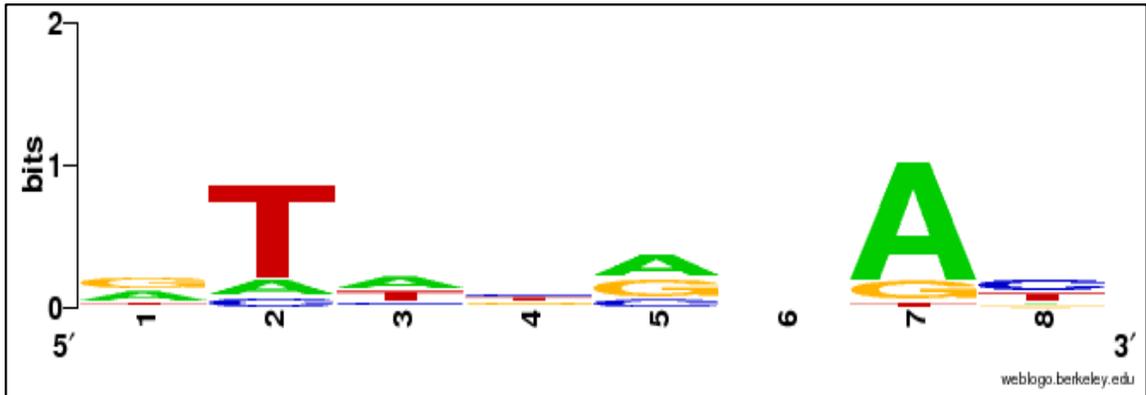


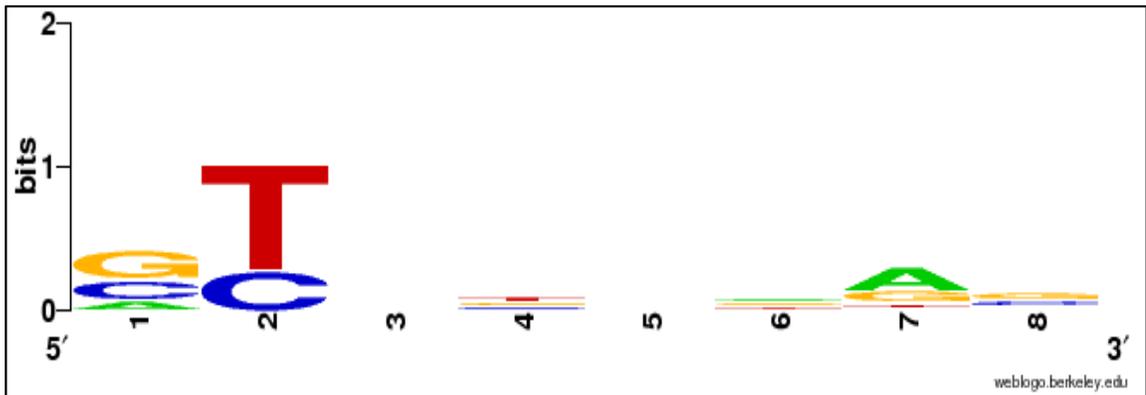
Figure 2.6: Schematic of the five-plasmid interplasmid transposition assay. The assay was performed in injected *Drosophila* embryos. The injected embryos are < 2 hrs old. They are allowed to recover for 16 hrs before heat shock. Heat shock drives expression of the *Hermes* and *piggyBac* transposase expression. Both the *Hermes* and *piggyBac* transposon donor plasmids carry gentamicin (Gent) selectable markers and either the LacZ (in *Hermes*) or EGFP (*piggyBac*) selectable markers flanked by the corresponding transposon left ends, shown as arrows that match the respective transposase. After heat shock, the plasmids are recovered from the embryos and transformed into *E.coli*. *Hermes* events are distinguished by their blue colony color on Gent/Cam antibiotic containing plates while *piggyBac* events are Gent/Cam resistant and show green fluorescence.

No. Injections	<i>Hermes</i> Transposase	<i>Hermes</i> DT	<i>pBac</i> DT	<i>Hermes</i> Events	<i>pBac</i> Events	<i>Hermes</i> Frequency	<i>pBac</i> Frequency	Fold Difference	Stdev <i>Hermes</i>	Stdev <i>pBac</i>
3	WT <i>Hermes</i>	163,000	152,000	67	57	4.1×10^{-4}	3.8×10^{-4}	+ 1.05	2.0×10^{-4}	3.4×10^{-4}
3	K292A	11,000	29,000	19	119	1.8×10^{-3}	4.1×10^{-3}	- 2.6	1.0×10^{-3}	2.6×10^{-3}
3	K299A	262,400	323,600	1	132	3.8×10^{-6}	4.1×10^{-4}	- 108	4.0×10^{-6}	3.6×10^{-4}
3	K300A	96,800	143,800	41	137	4.2×10^{-4}	9.5×10^{-4}	- 3.0	4.6×10^{-4}	1.4×10^{-3}
3	R306A	55,600	64,800	5	45	8.9×10^{-5}	6.9×10^{-4}	- 7.8	1.6×10^{-4}	3.6×10^{-5}

Table 2.1: Interplasmid transposition rates in injected *Drosophila* pre-blastoderm embryos for ROW mutants. Each of the samples was co-injected with the *piggyBac* (PB) transposon donor and helper plasmids as an internal control. Each event represents an independent event. The frequency was calculated by dividing the number of events by the donor titer (DT) in each of the injection sets. The DT is the number of donor plasmids that were available. K299A has a ~108 fold difference in transposition frequency versus WT *Hermes* and its PB internal control.



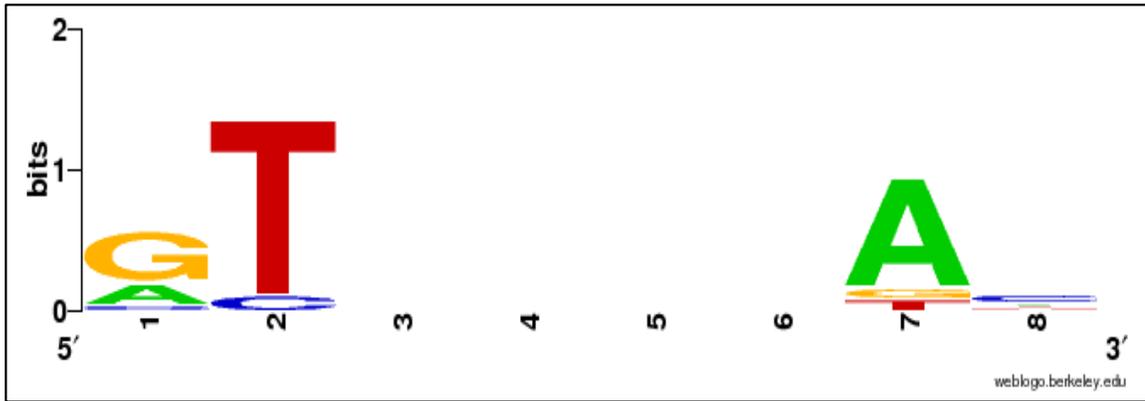
WT *Hermes*



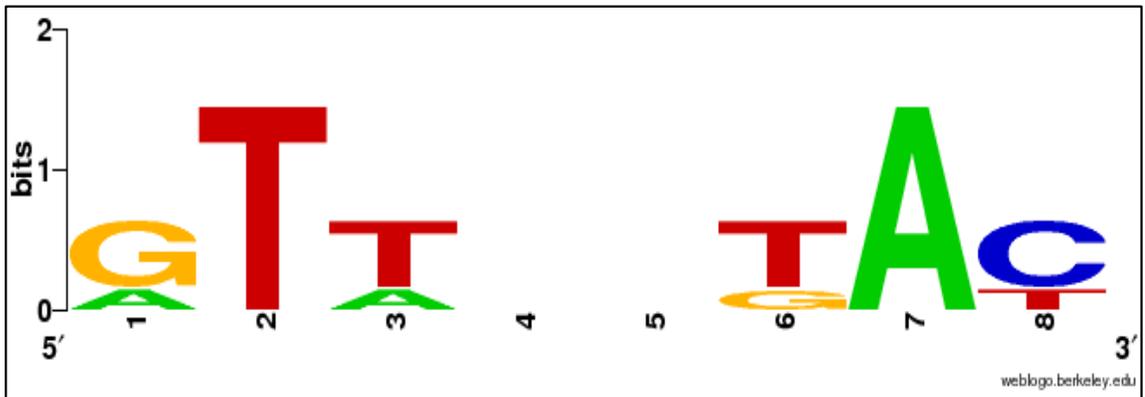
K300A Mutant *Hermes*

Figure 2.7: Weblogo consensus target site duplication for WT *Hermes* and K300A.

The K300A weblogo consensus target site duplication was based on 18 independent events. The WT *Hermes* was based on 19 independent events. K300A has less of a preference for A in the 7th position compared to WT *Hermes*. The P-value for this change was calculated using the Fisher's exact test. In this test 0.05 is considered significant and 0.01 is very significant. The WT *Hermes* integration sites are from Sarkar *et al.*, 1997.



K292A Mutant *Hermes*



R306A Mutant *Hermes*

Figure 2.8: Weblogo consensus target site duplication for K292A and R306A. The K292A consensus sequence is based on 11 sequenced independent events. The R306A consensus TSD is based on the 4 independent sequenced events.

<i>Hermes</i> Wild Type	Target site	Number of Integrations	Site	K300A	Target site	Number of Integrations	Site
	GTATGCAC	14	2154 (+)		GTACCGAG	2	2010 (-)
	GTCTGAAC	12	736 (+)		GTCTGAAC	2	736 (-)
	GTTCCGAC	10	2303 (+)		GTACAGAG	2	318 (+)
	GTTGGGAT	4	423 (+)		GTTCCGAC	1	2303 (+)
	GTACCAAC	4	2271 (+)		ACTGGAAT	1	789 (-)
	TTTGGGAT	3	339 (+)		CCCGATTG	1	834 (+)
	ATAGACGT	3	429 (+)		GTATGTAC	1	2395 (-)
	TTCGTGAC	3	660 (+)		CTATGTGT	1	2210 (-)
	ATTTACAC	3	810 (+)		CCTGTAGG	1	2297 (-)
	TTACACAT	3	812 (+)		GTGTTAGC	1	2214 (+)
	GTACAGAG	2	318 (+)		GTTGATTA	1	782 (+)
	TTACAAAC	2	350 (+)		GCGTGTAG	1	751 (+)
	ATTTACAC	2	576 (+)		CTAAAGCA	1	369 (-)
	GAAGGCAC	2	890 (+)		CCTGCAGG	1	1981 (+)
	CCCCGGGT	2	2004 (+)		ATCCCCGG	1	2002 (+)
	GTACCGAG	2	2010 (+)		CTCTAGAG	1	1993 (+)
	TCCCATAC	2	2070 (-)		ATAGCAAC	1	1904 (+)
	GTATGGGA	2	2070 (+)				
	GTGCATAC	2	2154 (-)				
	TTATAGAA	2	2162 (+)				
	AAATGCAT	2	2171 (-)				
	GTCGGAAC	2	2303 (-)				

Table 2.2: Integration sites of WT *Hermes* and K300A recovered from *Drosophila* interplasmid transposition assays. Each of the sequences were used only once regardless of how many integrations occurred at the site to create the consensus sequence using weblogo and to construct Table 2.3. The WT *Hermes* integration sites are from Sarkar *et al.*, 1997.

Nucleotide position								
	1	2	3	4	5	6	7	8
G	8	0	2	6	5	5	5	8
A	3	0	6	1	5	6	9	2
T	0	12	5	6	2	5	2	2
C	6	5	4	4	4	1	1	5

Table 2.3: K300A target site nucleotide frequency. Nucleotide frequency at each position in the recovered K300A unique target sites from *Drosophila* interplasmid transposition assays. Each site was counted only once and were input into weblogo to create the consensus integration site for K300A.

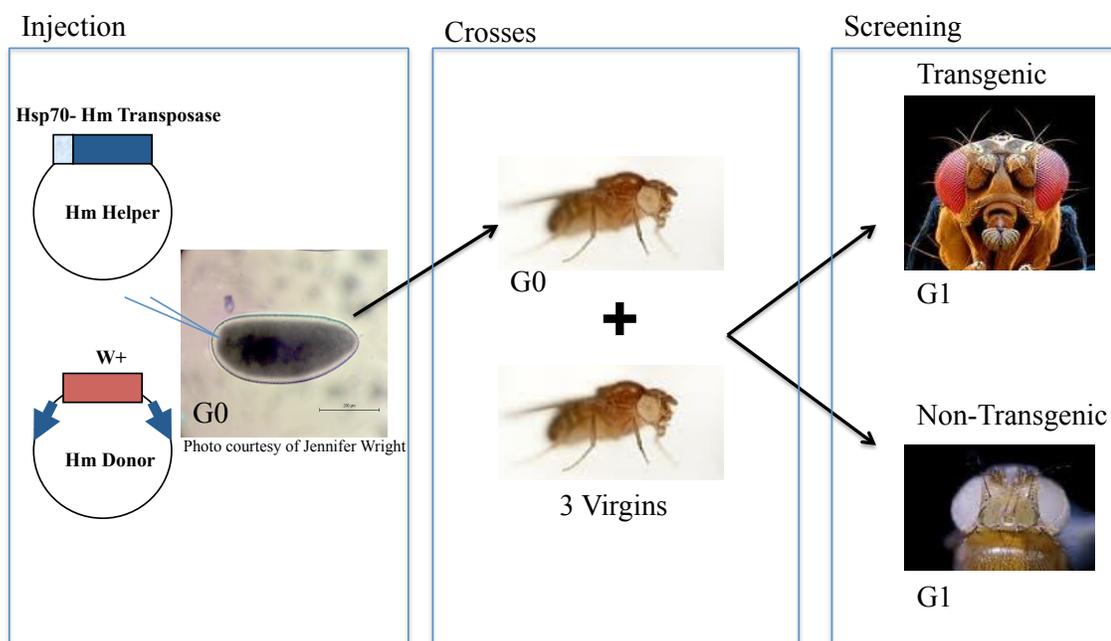


Figure 2.9: Schematic representation of the transformation assay in *Drosophila*.

The G1 progeny are screened for red eye pigment color. The G0 of any G1 progeny showing red eye color are considered transgenic.

Images from:

http://www.google.com/imgres?imgurl=http://4.bp.blogspot.com/_3FTO6EjRbe4/SwZK87iAy3I/AAAAAAAXFI/rqQuvQGfMM/s640/Z340499-Drosophila_fly_head_SEM-SPL.jpg

Trials	<i>Hermes</i> Transposase	Injected	Survived	Eclosed	Crossed	Fertile	Sterile	No. Transgenic	Sterility Rate	Transformation Rate
3	WT <i>Hermes</i>	679	254	77	69	52	17	15	24.6%	28.8%
3	K292A	667	308	164	140	104	36	1	25.7%	0.96%
3	K299A	440	146	102	68	57	11	0	16%	0%
3	K300A	752	237	109	100	68	32	3	32%	4.4%
3	R306A	605	258	123	88	67	21	17	23.9%	25.4%

Table 2.4: The *Drosophila* germline transformation frequency of WT *Hermes* and the “rim of the wheel” mutants. The *Drosophila* transformation frequency K292A, K299A, and K300A mutant transposases were each lower than that of WT *Hermes* transposase. The R306A mutant transposase had a transformation frequency that was close to WT.

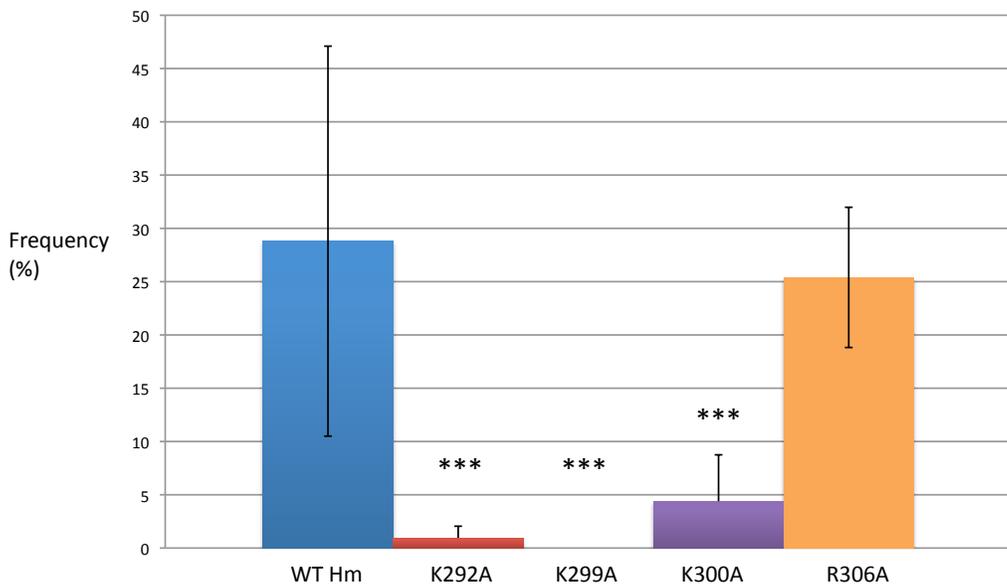


Figure 2.10: Graphical representation of the *Drosophila* transformation frequency of WT *Hermes* and the “rim of the wheel” mutants. The K292A, K299A and K300A mutant transposases had statistically significant decreases in their transformation efficiency. The error bars represent the standard deviation for each *Hermes* transposase. The asterisks signify a statistically significant difference compared to WT *Hermes*.

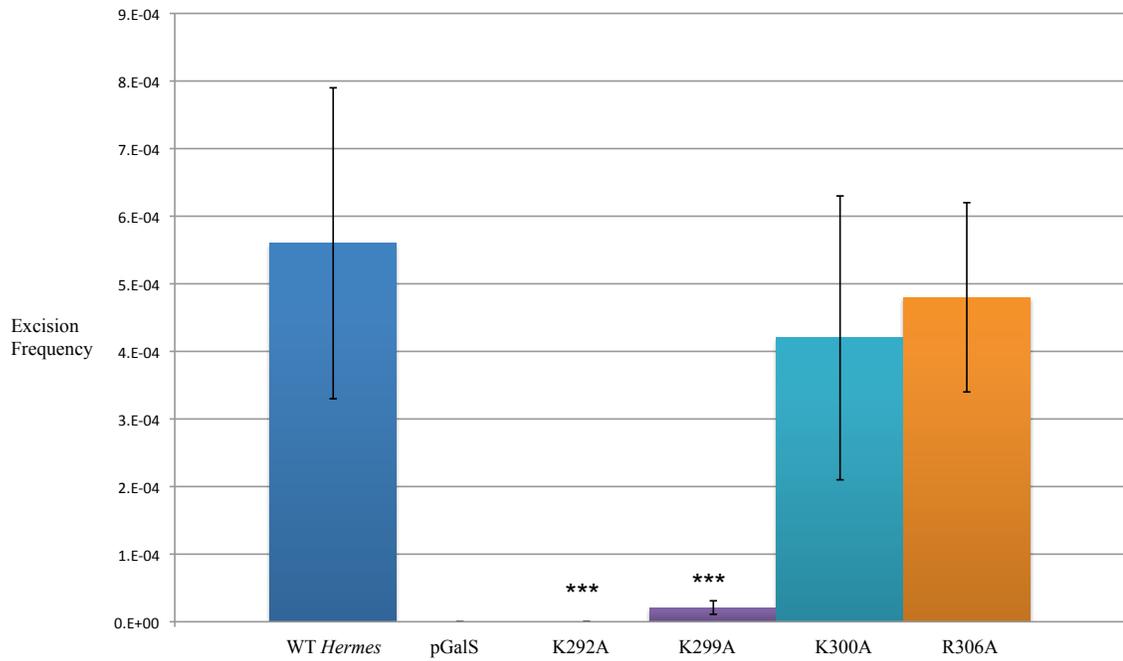


Figure 2.11: The yeast excision frequency of WT *Hermes*, “rim of the wheel” mutants, and negative control pGalS. The error bars represent the standard deviation for each of the *Hermes* transposases. The asterisks signify a statistically significant difference compared to WT *Hermes*.

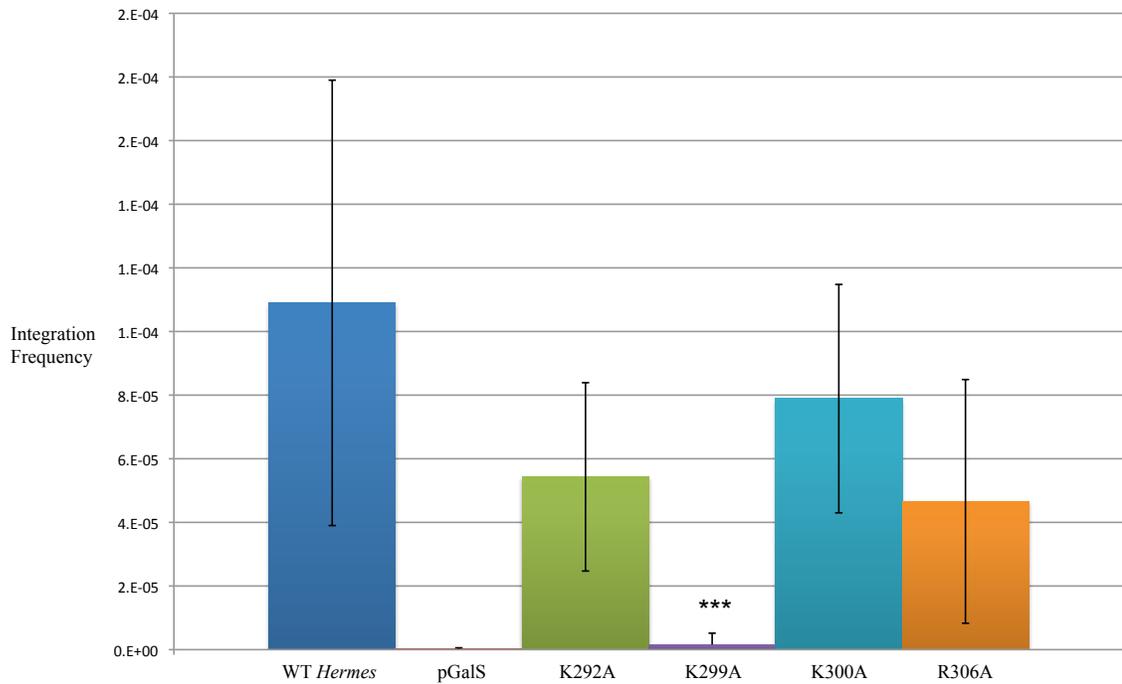


Figure 2.12: The yeast integration frequency of WT *Hermes*, ROW mutants, and the negative control pGalS. The error bars represent the standard deviation for each of the *Hermes* transposases. The asterisks signify a statistically significant difference compared to WT *Hermes*.

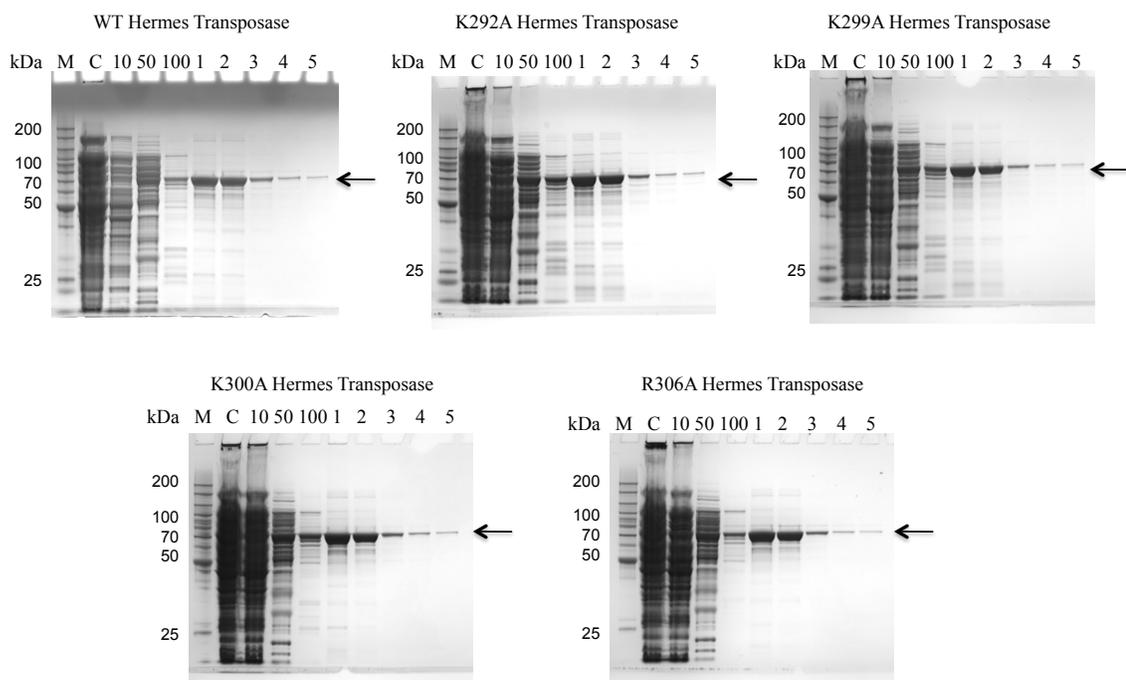


Figure 2.13: Coomassie blue stained SDS-PAGE His-tagged WT and mutant protein elution profiles. The protein elution profiles for the His-tagged WT *Hermes* and “rim of the wheel” *Hermes* transposase mutants. A sample of the crude (C) elution, and 10, 50, and 100 mM imidazole are shown. Five elution fractions were collected and was performed with 200 mM imidazole. The 70 kDa *Hermes* transposase is indicated with an arrow.

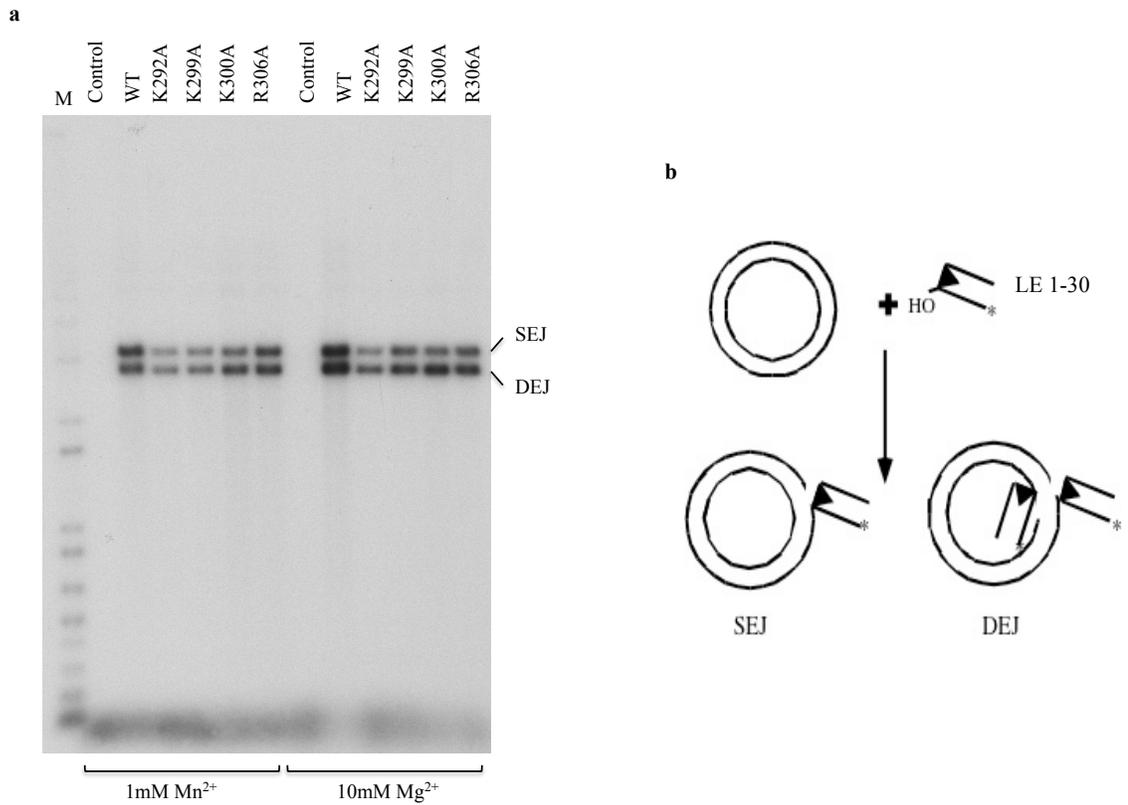


Figure 2.14: Autoradiograph of the strand transfer assay. (a) The DNA target joining products from *Hermes* transposase incubated with radiolabelled pre-cleaved Hm LE 1-30 with an overhanging nucleotide on the 5' end of the top strand to mimic the natural cleavage product. The formation of the single end joining products (SEJ) and double end joining products (DEJ) are visualized on a 1% native agarose gel. **(b)** Schematic showing the formation of the target joining SEJ and DEJ products. Diagram from Arensburger et al., 2011.

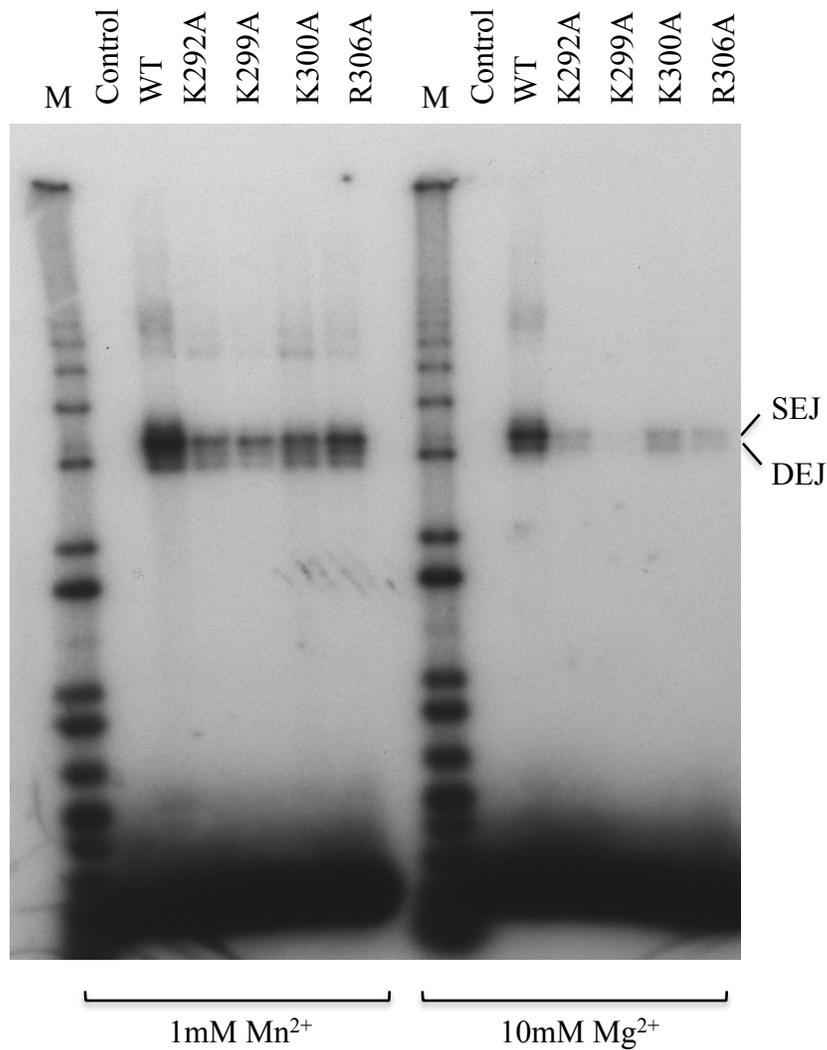


Figure 2.15: Autoradiograph of the coupled cleavage and strand transfer assay. The coupled cleavage and strand transfer products produced when *Hermes* transposase is incubated with radiolabelled Hm LE 193 which has 91 bp of *Hermes* transposon left end and terminal inverted repeat and 102 bp of flanking genomic DNA. Single end joining products (SEJ) and double end joining products (DEJ) are displayed on a 1% native agarose gel.

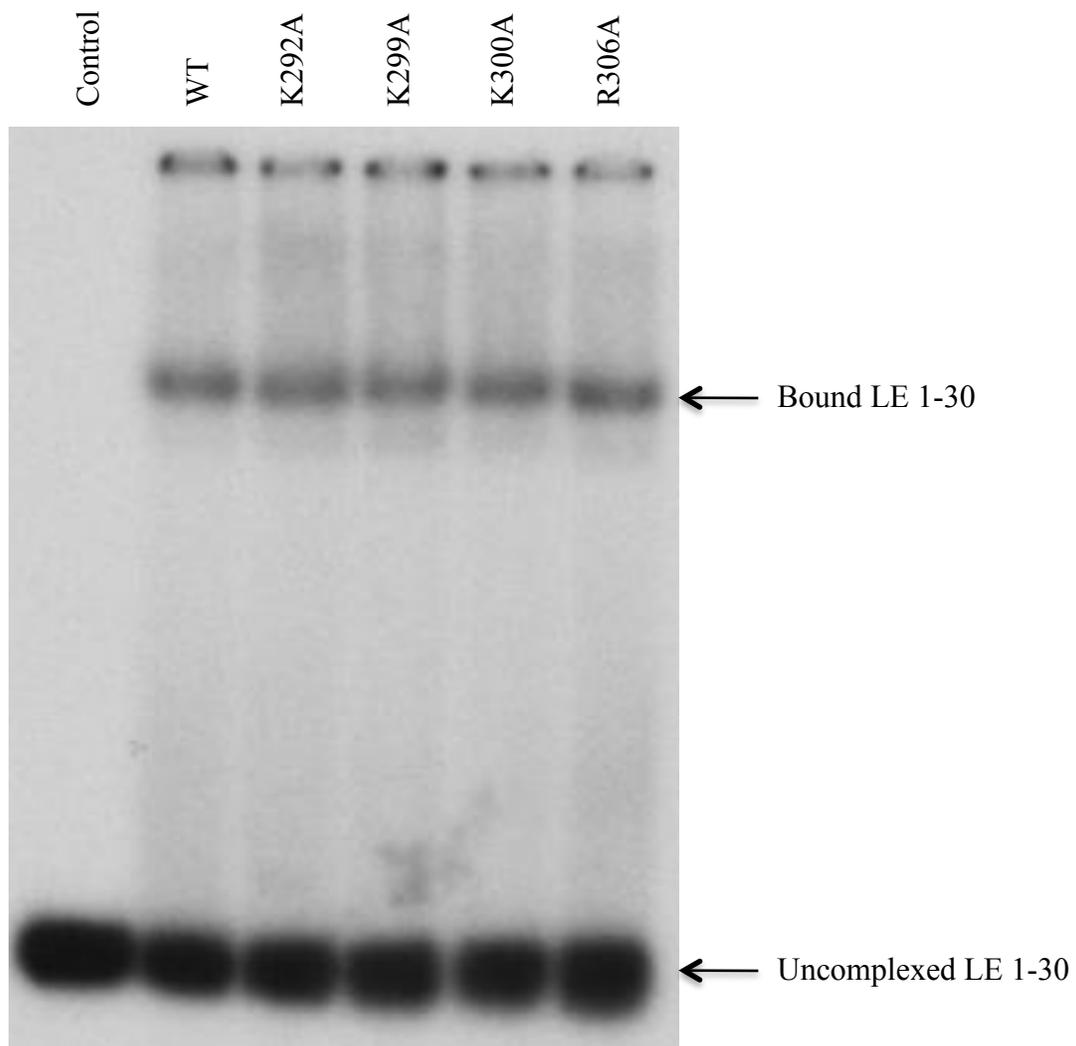


Figure 2.16: Autoradiograph of binding of WT and “rim of the wheel” mutant *Hermes* transposase to transposon left end 1-30. EMSA with *Hermes* transposase incubated with radiolabelled Hm LE 1-30. There is no noticeable difference in the affinity for the transposon left end between the ROW mutant transposases and WT *Hermes* transposase. The control did not contain any transposase protein. Samples were run on 5% TBE PAGE gel.

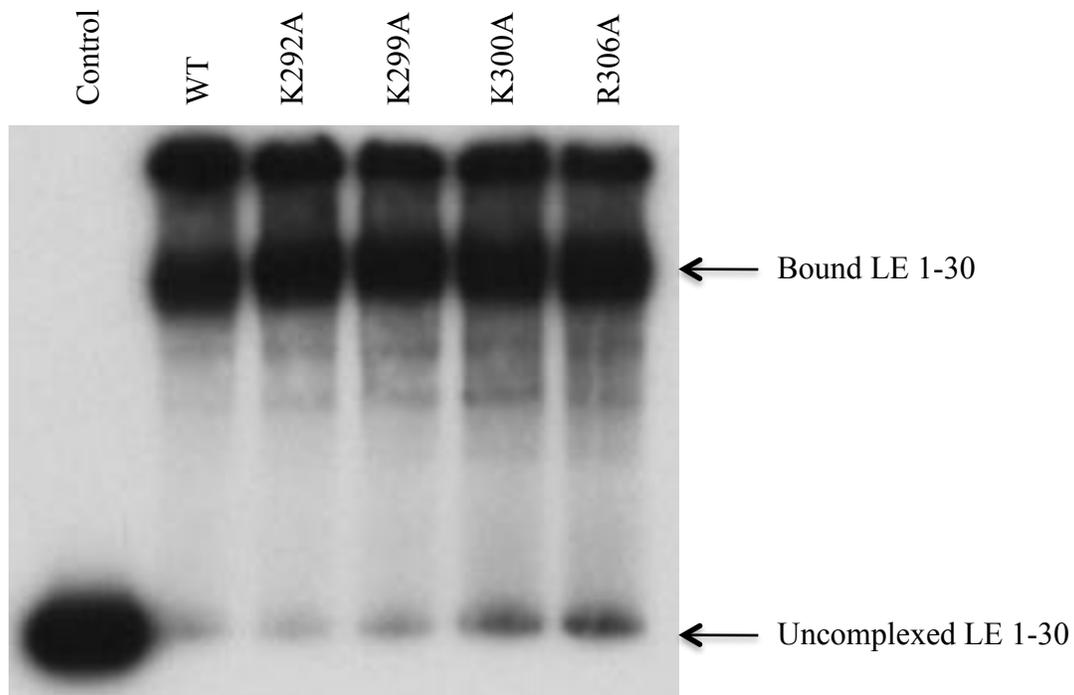


Figure 2.16: Autoradiograph of binding of WT and “rim of the wheel” mutant *Hermes* transposase to transposon left end 193. EMSA with *Hermes* transposase incubated with radiolabelled Hm LE 193 containing 91 bp of *Hermes* transposon left end and 101 bp of flanking genomic DNA. There is no noticeable difference in the affinity for the transposon left end between the ROW mutant transposases and WT *Hermes* transposase. The control did not contain any transposase protein. Samples were run on 5% TBE PAGE gel.

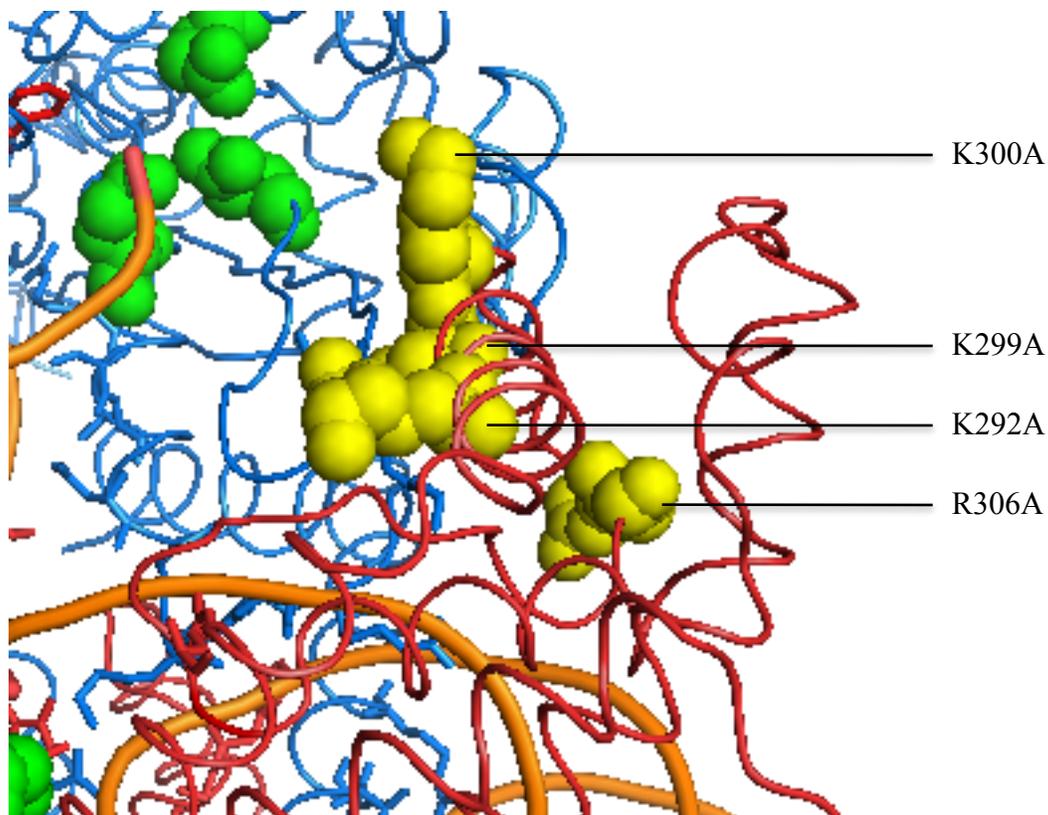


Figure 2.18: Position of the “rim of the wheel” mutants in the Hermes transposase octamer structure. A close up of the ROW residues in the *Hermes* octamer structure showing their position after *Hermes* has bound to left end fragments (F. Dyda and A. Hickman personal communication). The K292 and K299 residues are pointing into the positively charged trench while K300 is projecting out of the positively charged trench perpendicular to the wheel. The R306 residue has moved out and away from the positively charged trench due to a conformational change that occurred to facilitate left end binding.

Chapter Three

***Hermes* Phosphorylation: Identifying Residues That are Subject to Phosphorylation Using Bioinformatics and Mass Spectrometry**

3.1 Abstract

The *Hermes* transposition mechanism creates double strand breaks upon excision from a donor site and integration into a new location. Because double strand breaks are detrimental to the viability of the cell, checkpoint pathways are used to detect DNA damage and repair the breaks. To minimize the damage that can be caused by transposon mobility, the transposition activity must be regulated by the cell. One way that this can be achieved is through the phosphorylation of the transposase to inactivate or decrease its activity. I have used bioinformatics and mass spectrometry to investigate whether the *Hermes* transposase is phosphorylated and, if so, whether phosphorylation has a regulatory effect on *Hermes* transposase activity. I found direct evidence of phosphorylation of the *Hermes* transposase. This type of direct evidence has not been observed before with any other transposase and may play a role in regulating *Hermes* activity. Additionally, in directly two serine residues in the *Hermes* transposase that are demonstrated to be important for transposition.

3.2 Introduction

The *Hermes* transposase moves via a “cut-and-paste” mechanism in which the transposase protein binds to the ends of the *Hermes* transposon and makes double strand breaks (DSBs) between the transposon ends and the flanking DNA (Zhou *et al.*, 2005).

The transposase then moves the transposon to a different genomic location where the ends of the transposon undergo nucleophilic attack on the target DNA creating 8 bp single strand gaps that must be repaired. The flanking DNA of the excision site contains hairpins that must be resolved before the DSBs can be repaired (Zhou *et al.*, 2005).

When cells experience DNA damage, checkpoint pathways become activated in response and halt the cell cycle in order to repair the DNA damage (Zhou and Elledge, 2000). DNA transposition is regulated to minimize the level of DNA damage that may result in lethal effects to the host cell. One way this can be achieved is through phosphorylation of the transposase or interaction with host factors. Previous studies have shown that eukaryotic transposases can be sensitive to phosphorylation and that they interact with host machinery involved with DNA damage repair (Beall *et al.*, 2002; Izsvak *et al.*, 2004; Walisko *et al.*, 2006).

Non-homologous end-joining (NHEJ) is a major DNA repair pathway that has been shown to be activated in response to transposition (Heemst *et al.*, 2004, Izsvak *et al.*, 2004, Walisko *et al.*, 2006). There are several kinase families involved in the activation of the pathway including the ATM-family of DNA damage checkpoint protein kinases. Furthermore, this family of kinases has been found to be activated during transposition of the *Sleeping Beauty* (SB), Tn5, and *P*-element transposases as well as during V(D)J recombination which shares several mechanistic similarities with *Hermes* transposition (Beall and Rio, 1996; Grawunder *et al.*; 1998, Heemst *et al.*; 2004, Lin *et al.*, 1999).

One ATM-family kinase involved in NHEJ repair of immunoglobulin V(D)J recombination intermediates is DNA-PK, which consists of a heterotrimeric complex containing a catalytic subunit DNA-PKcs, and regulatory component composed of the Ku 70- and 80-kD proteins (Beall et al., 2002). Ku 70 and Ku 80 and the NHEJ pathways are conserved in yeast, *Caenorhabditis elegans*, and *Drosophila* but the DNA-PKcs subunit gene is absent from these species (Beall et al., 2002). However, in *Drosophila* there are two ATM-related genes, *mei-41* and *dATM*. The function of *dATM* is unknown but *mei-41* has been shown to function in meiotic recombination and DNA damage checkpoint control (Pastwa and Blasiak, 2003). In yeast, NHEJ is a minor pathway for DSB repair while the major pathway is based on DNA homology (Pastwa and Blasiak, 2003).

Although the DNA-PKcs is apparently absent in these eukaryotes, NHEJ still occurs presumably through the use of some other ATM family kinase. For example, in a study with the *Drosophila* *P*-element it was demonstrated that NHEJ was used for DNA repair at the donor site following transposition in *Drosophila* (Beall et al., 2002). Additionally, in *Drosophila* when eight of the 10 possible phosphorylation sites for ATM family of DNA checkpoint kinases were mutated from serine or threonine to non-phosphorylatable alanines in the *P*-element there was one mutant, S129A, which resulted in an increase in transposition frequency and germline transformation efficiency, suggesting that phosphorylation of this residue could negatively effect transposition. Additionally, three of these residues negatively affected transposition when mutated to alanine, suggesting that phosphorylation could have both negative and positive effects on

transposition, and supports a role for phosphorylation in regulating the *P*-element transposase (Beall et al. 2002). This is an example of one of several ways that transposases may circumvent the effects of phosphorylation.

Because the host cell strives to inactivate the transposase and the transposase strives remain active, the transposase must evolve a way around the effects of phosphorylation if this is involved in negative regulation. One way for transposases to do this is to evolve interactions with host factor proteins that can shut down signaling mechanisms that would, if activated, phosphorylate the transposase. The *Sleeping Beauty* (*SB*) transposase has been shown to interact with and down regulate Miz-1, a transcriptional regulator of genes involved in cell cycle regulation including cyclin D1 (Walisko, *et al.*, 2006). Down regulation of cyclin D1 results in a slow down of cell cycle in the G1 phase that increases *SB*'s efficiency of transposition, presumably because transposon-induced DNA damage can be efficiently repaired by NHEJ. It should be pointed out that this response was not due to the presence of transposon DNA but to the *SB* transposase itself, indicating that this response was not associated with transposition induced DNA-damage.

To understand whether phosphorylation plays a role in the activity or regulation of the *Hermes* transposase, several mutants were created based on two separate methods. The first series of mutations were based on a bioinformatic analysis of the *Hermes* transposase sequence to predict sites in the protein that are targets for phosphorylation of eukaryotic kinases. Using this approach two serine residues were found to be hotspots

for phosphorylation, S309 and S310. These residues were simultaneously mutated to non-phosphorylatable alanine residues, phosphorylation-mimicking aspartic acids, and chemically similar threonine residues and tested for their transposition ability. The second approach was a mass spectrometry (MS) analysis of a *Hermes* transposase isolated from the *Drosophila* S2 cell line to physically detect phosphorylation. The analysis yielded two and possibly three residues that were found to be phosphorylated, S233 and S468 and/or S469. These residues were mutated to alanine to determine the effect on transposition.

3.3 Results

Hermes residues Ser309 and Ser310 are bioinformatically hotspots for phosphorylation.

Using the bioinformatics program NetPhosK, *Hermes* transposase residues 309 and 310 were predicted to be hotspots for phosphorylation, with 309 being a potential target for up to four different kinases (Figure 3.3.1) (Blom et al., 2004). When these residues were computationally changed to alanine and the protein sequence was re-entered into the prediction program they were no longer targets of the kinases.

Previous studies on the insect transposases *P*-element and *Mos1* suggested that eukaryotic transposases are susceptible to phosphorylation and host machinery plays an important part in regulation of transposase activity (Beall et al., 2002; Germon et al., 2009). The position of S309 and S310 in the 2005 *Hermes* crystal structure suggests that they could have an important function as they are on an α -helix that leads into the active site (Figure 3.3.2) (Hickman et al., 2005). Additionally, in the octamer crystal structure

solved at NIH it was shown that these residues interact with the *Hermes* LE through hydrogen bonding (Figure 3.3) (F. Dyda and A. Hickman, personal communication). Phosphorylation of either or both of these residues would put bulky negatively charged phosphates on the helix potentially blocking access to the active site and disrupting transposon end binding by the *Hermes* transposase. Because both residues were potential targets for a number of kinases we decided to make three double mutants at the 309 and 310 positions. The first mutant is a non-phosphorylatable mutant where the serine residues were changed to alanine, SS309/10AA. The second mutant was a phosphorylation substitute where the serine residues were changed to threonine, which is nearly chemically identical to serine, SS309/10TT. And the third mutant was a serine to aspartic acid pseudo-phosphorylation mimic, SS309/10DD.

Hermes mutants SS309/10AA and SS309/10DD show lowered transformation efficiency in Drosophila.

The non-phosphorylatable SS309/10AA mutant was tested for its ability to mediate genetic transformation of *Drosophila* transformation efficiency. Transformation and screening were performed in the same manner as described in Chapter 2. The transformation efficiency of SS309/10AA was 7.7 times lower than WT in *Drosophila* (Table 3.1). The phosphorylation mimic SS309/10DD was also very impaired for transformation as it had a transformation frequency ~18 times lower than WT (Table 3.1). This was not unexpected as it was also severely impaired in interplasmid transposition assays. SS309/10TT had near WT transformation activity (Figure 3.9).

Mass spectrometry of S2 cell expressed Hermes protein shows phosphorylation on Ser233, Ser468 and/or Ser469.

The S2 cell line, DEV-8, transfected with a pMT-V5 HisC plasmid containing the *Hermes* transposase ORF under the control of the metallothionein gene promoter was used to produce the *Hermes* protein. Nuclear extracts were produced in the presence of 1X phosphatase inhibitor cocktail containing four inhibitors with broad specificity for serine/threonine and tyrosine phosphatases and run on a 10% SDS-PAGE gel to confirm the production of the *Hermes* transposase (Figure 3.4). The transposase was then immunoprecipitated with a polyclonal rabbit antibody to the *Hermes* transposase and submitted to mass spectrometry (MS). The sample was digested with the trypsin proteolytic enzyme prior to MS. The MS results were analyzed using the Mascot program which takes the experimental mass values of each peptide fragment detected and compares that to the calculated peptide mass. Each of the matches are scored according to how closely the experimental and calculated masses match. Each match is given an ion score based on how close they are. Ion scores greater than 30 are considered strong matches. As Figure 3.5 shows, at least two and possibly three serine residues were phosphorylated, S233 and S468 and/or S469. We were unable to determine whether S468 or S469 were the phosphorylated residue and there was some evidence that possibly both were phosphorylated. S233 had three of five ion scores greater than 30. S233 is a potential phosphorylation target by DNA-PK, which is involved in the DNA damage pathway. S468 and S469 are potential targets by the PKA and RSK kinases. As can be seen in Figure 3.6, S233 resides on the surface of *Hermes* on the first α -helix of the

catalytic RNase H-like fold on the backside of the *Hermes* transposase. While there is no DNA-PK catalytic subunit or ortholog in *Drosophila* that can account for the phosphorylation of this residue, the DNA-binding subunits of DNA-PK, Ku 70 and Ku 80, are present (Sekelsky et al., 2000). However, even without the catalytic subunits *Drosophila* cells have a strong NHEJ activity for the repair of DNA double strand breaks. It is therefore possible that some other kinase acting in place of DNA PK that may be responsible for S233 phosphorylation.

As seen in Figure 3.7, immunoprecipitated *Hermes* transposase was submitted to MS twice and digested using either trypsin and chymotrypsin. During the first submission trypsin was used to digest *Hermes* protein resulting in the detection of seven phosphorylated fragments of residues 467 - 496 with significant ion scores (Figure 3.5). Two of the 467 – 496 fragments were that were detected were doubly phosphorylated but their ion scores were not considered to be significant. Immunoprecipitated *Hermes* protein was submitted a second time and digested using chymotrypsin. Although peptide fragments of residues 470 – 502 were detected, none of the fragments showed phosphorylation. We therefore were able to narrow down which residues were the phosphorylated residues on the 467 – 496 peptide fragment. Because the exact residue is unknown but two fragments detected by MS showed double phosphorylation, we decided to make a double non-phosphorylatable mutant to test, SS468/69AA. The residues are found in the all α -helical domain in an area of low electron density that are unresolved in the octamer structure due to heat oscillations (Figure 3.6) (F. Dyda and A. Hickman, personal communication). In the 2005 crystal structure they lie right on the boundary of

the unresolved region (Hickman *et al.*, 2005). These residues reside in an area that has been shown to be important for *Hermes* oligomerization (Michel and Atkinson, 2003). There are two kinases that target these residues, PKA and RSK.

Non-phosphorable Hermes mutant activity in interplasmid transposition assays in injected Drosophila

Because of the variability inherent in the interplasmid transposition assay due to factors such as desiccation of the embryos, differences in needle size, success of plasmid recovery and electroporation efficiency, it is important to have an internal control to allow for the normalization of the data. Therefore, we used the five-plasmid interplasmid transposition assay as described in Chapter 2. As seen in Table 3.2, in injected *Drosophila* embryos SS309/10AA did not differ significantly from the internal *PB* control. Both transposases had nearly a three-fold higher transposition frequency compared to WT and the WT internal *PB* control. Sequencing of 21 SS309/10AA events did not reveal a difference in the TSD compared to WT *Hermes* (Figure 3.8). This result was seen for both SS309/10AA and for SS468/69AA in which there was no statistically significant difference in the interplasmid transposition frequency or their target sites (Table 3.2). The SS309/10DD mutant however had severely impaired somatic tissue activity in *Drosophila*, as not one event was recovered (Table 3.2).

Hermes phosphorylation mutants in yeast excision assays

The excision frequency of each of the mutants was tested *in vivo* in yeast using the same technique described in Chapter 2 and compared to WT *Hermes* and the negative control pGalS. The yeast assays, both excision and integration, provide a high throughput manner to test the activity of each of the mutants in another eukaryotic system. As seen in Figure 3.10, the only mutant that displayed a statistically significant difference in their excision frequency compared to WT *Hermes* was SS309/10AA. The S233A mutant showed an excision frequency that was higher than WT *Hermes* but was not statistically significant, whereas the SS468/69AA mutant showed a trend toward a lower but not statistically significant excision frequency (Figure 3.10).

Hermes phosphorylation mutants in yeast integration assays

The ability of WT *Hermes*, each of the mutants, and a negative control (pGalS) were tested for their integration frequency *in vivo* in yeast assays using the same technique described in Chapter 2. As can be seen from Figure 3.11, the only mutants with a statistically significant change in their integration frequency were SS309/10AA and SS309/10DD. This result was also seen in transformation of *Drosophila* and supports the fact that these two mutants are impaired in their transposition frequencies for reasons that are yet unknown. The SS309/10AA mutant has a 17-fold difference in its integration frequency compared to WT in yeast. This is identical to the fold difference of SS309/10DD compared to WT *Hermes*. Again, the SS468/69AA mutant had an integration frequency that was lower than WT *Hermes* but not statistically significant

while the S233A had a higher but not statistically significant integration frequency (Figure 3.11).

3.4 Discussion

Non-phosphorylatable SS309/10AA has lower transposition frequency in yeast and the Drosophila germline

Phosphorylation of proteins is an important mechanism by which cells modulate the activity of proteins and enzymes as well as turn on or off pathways important in cell maintenance or in response to environmental stimuli and signals. One important pathway that is activated and regulated by kinases is the DNA double strand break (DSB) response. Because *Hermes* creates DSBs during transposition, we wanted to test whether *Hermes* was phosphorylated and if so what kind of effect that had on the function of *Hermes*. Using the bioinformatics program NetPhosK we identified two residues, S309 and S310, which were predicted hotspots for phosphorylation (Blom et al. 2004). Using site-directed mutagenesis both of the residues were changed to non-phosphorylatable alanines, phosphorylation mimicking aspartic acids, and chemically similar threonines to test their effects on transposition. As seen in Figure 3.3.2, these residues reside on an α -helix that leads into the active site and make hydrogen bonding contact with the *Hermes* left end in the octamer crystal structure (F. Dyda and A. Hickman, personal communication). As is seen in Table 3.1, when the residues are mutated to non-phosphorylatable alanines the ability for *Hermes* to transform *Drosophila* is 7.7-fold lower than WT *Hermes*. The transposition impairment for this mutant was also seen *in vivo* in

yeast as the integration and excision frequency for SS309/10AA was 17- and 5-fold lower, respectively, than WT *Hermes*. There was no difference in the SS309/10AA mutants interplasmid transposition frequency compared to either the internal *PB* control or WT *Hermes*. This may have to do with the variability of the assay as discussed earlier or may be germline specific. Twenty-one of the events were sequenced and there is no evident difference in the TSD of this mutant compared to WT *Hermes* (Figure 3.8).

Hermes mutant SS309/10DD has a severe decrease in transposition.

The SS309/10DD mutant showed severe impairment in its transposition frequency, as it was nearly inactive in all of the assays that it was tested in. The SS309/10DD was designed to mimic the phosphorylation of the 309 and 310 residues. There were no events recovered from interplasmid transposition assays although the mutant did retain transformation capability. The transformation frequency for the SS309/10DD mutant was almost 18-fold lower than WT *Hermes*. Additionally, *in vivo* in yeast this mutant had an integration frequency that was 17-fold lower than WT *Hermes*. As can be seen in Figure 3.3.2, serine 309 and 310 are located on an α -helix that leads into the active site and are important for making hydrogen bonding interactions with the transposon left end in the octamer crystal structure. Phosphorylation of the serine residues can have important consequences on the ability for the *Hermes* transposase to bind to DNA. By mutating the serines to aspartic acid we made a pseudo-phosphorylated *Hermes* transposase that allowed us to test the effect of phosphorylation of these residues on transposition. Although these residues were not phosphorylated in MS analysis, the

data implies that these residues are important for transposition and that if these residues are in fact phosphorylated then it can be assumed to negatively impact the transposition ability of *Hermes*.

Non-phosphorylable mutants S233A and SS468/69AA appear to have no change in their transposition activity

Although S233 and S468 and/or S469 were found to be phosphorylated through MS, when the residues were mutated to alanine they were not found to display any statistically significant difference in any of the assays in which they were tested. Although S233A had a higher excision and integration frequency than WT *Hermes*, as can be seen in Figure 3.10 and 3, S233A did not show a statistically significant frequency in either assay in yeast. Serine 233 was an interesting residue because it is a target for DNA-PK. Although the catalytic subunit for this kinase is absent in *Drosophila* and yeast, both of these systems are able to perform NHEJ and it is thought that there are other ATM family kinases responsible for phosphorylation of proteins in this pathway (Beall *et al.*, 2002). The results therefore imply that phosphorylation of this residue is not essential for activity of the *Hermes* transposase. That does not exclude the possibility that phosphorylation of this residue does not negatively impact the activity. Further testing is needed to determine whether a pseudo-phosphorylated mutation at serine 233 has an impact on its transposition activity.

Although the SS468/69AA non-phosphorylable mutant showed slightly decreased excision and integration frequencies in yeast, these results were not statistically

significant (Figure 3.10 and 3). Additionally, in *Drosophila* somatic tissue in interplasmid transposition assays, even though SS468/9AA had a higher transposition frequency than WT *Hermes* it was not statistically significant compared to its internal *PB* control (Table 3.2). These residues are in an area of low electron density in the *Hermes* octamer structure and were not able to be resolved. Although the results imply that these residues are not dependent on phosphorylation for transposition activity further testing will be needed with pseudo-phosphorylated mutants to determine whether their phosphorylated state has an impact on transposition.

3.5 Conclusion

As the MS data show, in *Drosophila* somatic tissue S233 and S468 and/or S469 are phosphorylated. This is the first time direct evidence of phosphorylation has been detected for a transposase. The S233 is particularly interesting since it is a target of DNA-PK, a component of the NHEJ pathway that is activated in response to double strand breaks. The S309 and S310 were mutated because were predicted to be hotspots for phosphorylation and because of their position in the protein structure suggested that they could be important. Through MS analysis phosphorylation of these residues were detected but mutation of these residues to alanine or aspartic acid had a negative consequences on *Hermes* transposition activity. This is hypothesized to be a result of the fact that in the *Hermes* octamer crystal structure these residues are important for forming hydrogen bonds with the *Hermes* LE. While we did not find a statistically significant difference in the mobility properties of either S233A or SS468/69AA, these residues

showed trends towards higher and lower excision and integration frequencies respectively. While it does not appear that phosphorylation of either of these residues are necessary for activity it can not be ruled out that phosphorylation does not have a negative or positive consequence on the proteins activity and pseudo-phosphorylated mutants at these positions will need to be tested.

3.6 Future Experiments

The S233 and S468 and/or S469 *Hermes* transposase residues were detected to be phosphorylated in *Drosophila* S2 cells. The mutant transposases, S233A and SS468/69AA were only tested for their transposition activity in yeast, whereas only SS468/49AA was tested in interplasmid transposition assays in injected *Drosophila* embryos. In neither of these assays were the mutant transposases found to have activity that significantly differed from WT *Hermes*. Because the phosphorylation was detected in somatic tissue further testing of these mutant transposases should be performed to test their activity in *Drosophila* somatic tissue in interplasmid transposition assays and for their germline transformation activity. Additionally, phosphorylation-mimicking mutant transposases S233D and SS468/69DD should be cloned and tested to determine the affects on transposition activity of phosphorylation of these residues. Because S468 and S469 are located on a region of the *Hermes* transposase that is important for creating interface 2 in the octameric structure, if a SS468/69DD mutant is defective for transposition size exclusion chromatography of this transposase should be performed to determine whether oligomerization is being affected.

Because S309 and S310 are important for transposition, presumably because they make hydrogen bonding interactions with the *Hermes* LE, they should be individually mutated to alanine residues to determine the extent of each of their roles in transposition. The mutant transposases proteins should then be purified and tested biochemically in strand transfer, coupled cleavage and strand transfer, and EMSAs to determine their function. The same should be performed with the SS309/10AA, SS309/10TT, and SS309/10DD mutant transposases as a comparison.

3.7 Materials and Methods

In silico analysis

Hermes phosphorylation predictions were performed using the NetPhosK server.

NetPhosK 1.0 (<http://www.cbs.dtu.dk/services/NetPhosK>) makes kinase specific predictions of eukaryotic protein phosphorylation sites for the following kinases: PKA, PKC, PKG, CKII, Cdc2, Cam-II, ATM, DNA PK, Cdk5, p38 MAPK, GSK3, CKI, PKB, RSK, INSR, EGFR and Src (Blom et al. 2004).

S2 cell line

The *Hermes* ORF was PCR amplified from pCRS*Hermes* and cloned into the plasmid pMT-V5 HisC (Invitrogen) after digestion with *Eco* RV and *Xho* I (NEB) (Laver, 2006). The plasmids were co-transfected with pCoHYGRO (Invitrogen) into the *Drosophila* S2 cell line using the CellFECTIN reagent (Invitrogen) to produce the *Hermes* DEV-8 cell line. Cells were grown at 25° C in Schneider's *Drosophila* Medium

(Invitrogen) supplemented with heat inactivated fetal bovine serum and antibiotics. Transfected cells were selected for with the addition of hygromycin to the medium.

Nuclear extracts

Nuclear extracts of S2 cells were produced from 30 ml cultures in which *Hermes* protein production was induced with 250 μ M CuSO₄. Cells were harvested and washed with PBS and then resuspended in buffer containing 10 mM HEPES pH 7.9, 150 mM NaCl, 1 mM EDTA, 4 mM DTT, 0.6% Triton X-100, 0.5% protease inhibitors cocktail P8340 (Sigma), and 1X Halt phosphatase inhibitor cocktail (Thermo) and incubated on ice for 5 min. The cells were then spun 1500xg for 10 min at 4° C. The supernatant was removed and the nuclei were extracted with the addition of 300 μ l of buffer containing 20 mM HEPES pH 7.9, 420 mM NaCl, 0.2 mM EDTA, 1.2 mM MgCl₂, 1 mM DTT, 25% glycerol 1% protease inhibitor cocktail (Sigma), and 1X Halt phosphatase inhibitor cocktail (Thermo). The cells were incubated on ice for 20 min and then spun at 15000 rpm at 4° C. The supernatant containing the extraction was removed and an aliquot was run on a 10% Bis-Tris Criterion Pre-cast polyacrylamide gel (Figure).

***Hermes* immunoprecipitation pull-down**

The *Hermes* transposase was isolated from the S2 cell nuclear extract by immunoprecipitation with a polyclonal rabbit antibody to the *Hermes* transposase. The nuclear extract was first incubated with 30 μ l of 50% slurry of Protein A Sepharose CL-4B beads (Pharmacia) at 4° C for 30 min to get rid of any non-specific proteins that might

bind to the beads. Next, the sample was spun for 3 min at 14k rpm to pellet the beads and the supernatant was carefully removed and put into a new tube. 20 ng of *Hermes* antibody was added and incubated for 6 hrs at 4° C. This was followed by an overnight incubation with 30 µl of 50% slurry of Protein A Sepharose CL-4B beads (Pharmacia) at 4° C. The beads were spun for 7 sec at 14 rpm and washed twice with 1 mL of cold wash buffer (0.1% Triton X-100, 50 mM Tris-HCl pH 7.4, 300 mM NaCl, 5 mM EDTA, 0.02% sodium azide) with spinning for 7 sec at 14 rpm each time to pellet the beads. The final wash was 1 mL of cold 1X PBS. The supernatant was discarded and a ~3 µl aliquot of the beads was run on 10% PAGE (Figure). The remainder of the 30 µl beads were submitted to the UCR High Resolution Mass Spectrometry Facility.

Mass Spectrometry

All samples submitted to the UCR High Resolution Mass Spectrometry Facility were analyzed with a Waters GCT (2008) high-resolution mass spectrometer using EI/MS/MS. The samples were digested with either chymotrypsin or trypsin proteolytic enzymes prior to loading onto the spectrophotometer. The MS data was analyzed using the Matrix Science servers Mascot program. This program takes the experimental mass values for each peptide fragment that was detected and compares it to the calculated peptide mass obtained from applying cleavage rules to the protein sequence and inputting the proteolytic enzyme used. If the protein is known then the closest matches to each detected peptide fragment can be pulled from the database. Each of the matches are scored according to how closely the experimental fragment mass matches to the

calculated mass. These are called ion scores. Ion scores of greater than 30 are considered strong matches.

Site-directed mutagenesis

Site-directed mutagenesis was performed with the QuikChange site-directed mutagenesis kit (Stratagene) and pKH70new *Hermes* as template plasmid using the same protocol described in Chapter 2. The following primers were used:

SS309/10AA F:

GCACAGACTTCGAGCTGCTTTAAAAAGTGAGTG

SS309/10AA R:

CACTCACTTTTTAAAGCAGCTCGAACTCTGTGC

SS309/10TT F:

GCAGCACAGACTTCGAACTACTTTAAAAAGTGAGTG

SS309/10TT R:

CACTCACTTTTTAAAGTAGTTCGAAGTCTGTGCTGC

HmP SS310DD F:

GCAGCACAGACTTCGAGATGATTTAAAAAGTGAGTG

HmP SS310DD R:

CACTCACTTTTTAAATCATCTCGAAGTCTGTGCTGC

S233A F:

GAAGCTTAAAGCCATTTTTGCACAATTCAACGTCGAAGAC

S233A R:

GTCTTCGACGTTGAATTGTGCAAAAATGGCTTTAAGCTTC

SS468/69AA F:

GGAATTAATAAACCGCATGGCTGCCTTTAACGAATTATCCGCAAC

SS468/69AA R:

GTTGCGGATAATTCGTAAAGGCAGCCATGCGGTTTATTAATTCC

Interplasmid transposition assay mix

The five-plasmid interplasmid transposition assay mix for each of the mutants was made the same way as described in Chapter 2.

Transformation Injection Mix

A transformation injection mix was made for each of the phosphorylation mutants and were made in the same manner as described in Chapter 2.

Yeast excision assay

The *in vivo* yeast excision assay with each mutant was performed the in the manner as described in Chapter 2.

Yeast integration assay

The *in vivo* yeast integration assay with each mutant was performed the in the manner as described in Chapter 2.

Microinjection of *Drosophila*

The microinjection of *Drosophila* for both interplasmid transposition assays and transformation assays was performed in the same manner as described in Chapter 2.

3.8 References

- Beall, E.L., Mahoney, M.B., Rio, D.C. (2002) Identification and analysis of a hyperactive mutant form of *Drosophila P-element* transposase. *Genetics*. **162**: 217-227.
- Beall, E.L., and Rio, D.C. (1996) *Drosophila* IRBP/Ku p70 corresponds to the mutagen-sensitive mus309 gene and is involved in P-element excision in vivo. *Genes Dev*. **10**: 921-33.
- Blom N., Sicheritz-Ponten T., Gupta R., Gammeltoft S., Brunak S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. **4**: 1633-49.
- Germon, S., Bouchet, N., Casteret, S., Carpentier, G., Adet, J., Bigot, Y., Auge-Gouillou, C. (2009) *Mariner MOS1* transposase optimization by rational mutagenesis. *Genetica*. **137**: 265-276.
- Grawunder, U., West, R.B., Lieber, M.R. (1998) Antigen receptor gene rearrangement. *Current Opinion in Immunology*. **10**: 172-180.
- Izsvak, Z., Stume, E.E., Fiedler, D., Katzer, A., Jeggo, P.A. and Ivics, Z. (2004) Healing the wounds inflicted by *Sleeping Beauty* transposition by double-strand break repair in mammalian somatic cells. *Molecular Cell*. **13**: 279-290.
- Heemst, D.V., Brugmans, L. Verkaik, N.S., van Gent., D.C. (2004) End-joining of blunt DNA double-strand breaks in mammalian fibroblasts is precise and requires DNA-PK and XRCC4. *DNA Repair*. **3**: 43-50.
- Lin, J.M., Landree, M.A., Roth, D.B. (1999) V(D)J recombination catalyzed by mutant RAG proteins lacking consensus DNA-PK phosphorylation sites. *Molecular Immunology*. **36**: 1263-1269.
- Hickman, A.B., Perez, Z.N., Zhou, L., Primrose, M., Ghirlando, R., Hishaw., J.E., Craig, N.L., Dyda, F. (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol*. **12**: 715-721.
- Michel, K., O'Brochta, D.A., Atkinson, P.W. (2003) The C-terminus of the *Hermes*

transposase contains a protein multimerization domain. *Insect Biochem. Molec. Biol.* **33**: 959-970.

Pastwa, E., Blasiak, J. (2003) Non-homologous DNA end joining. *Acta Biochim. Pol.* **50**: 891-908.

Walisko, O., Izsvak, Z., Szabo, K., Kaufman, C.D., Herold, S., Ivics, Z. (2006) *Sleeping Beauty* transposase modulates cell-cycle progression through interaction with Miz-1. *Proc. Natl. Acad. Sci.* **103**: 4062-4067.

Zhou, B.B., Elledge, S.J. (2000) The DNA damage response: putting checkpoints in perspective. *Nature.* **408**: 433-39.

Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F. and Craig, N.L. (2004) Transposition of *hAT* elements links transposable elements and V(D)J recombination. *Nature.* **432**: 955-1001.

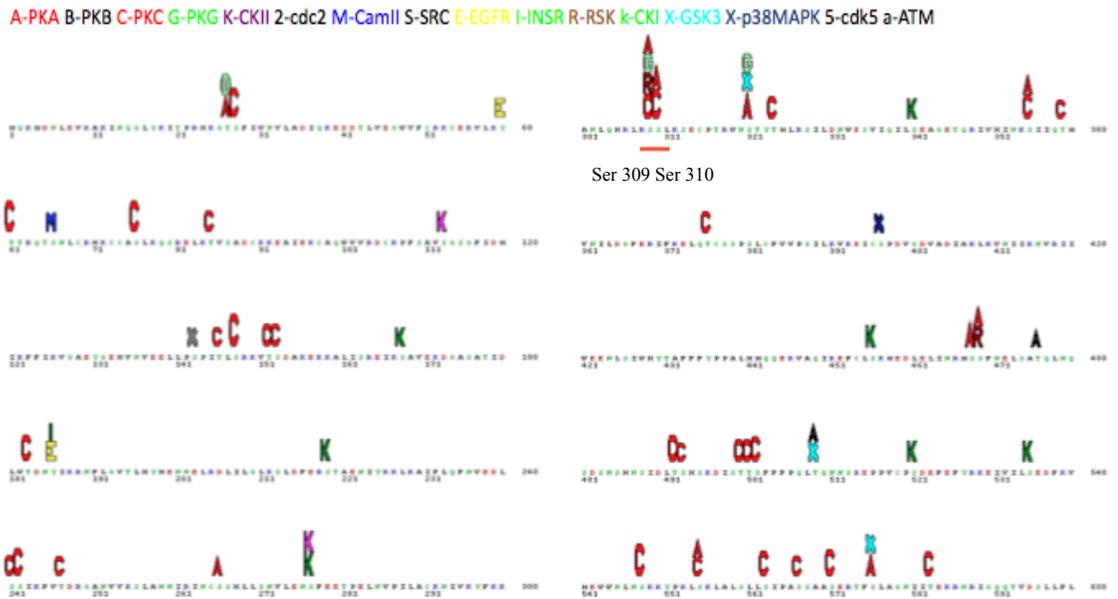


Figure 3.1: NetPhosK results. The *Hermes* transposase primary sequence was entered in to the NetPhosK server which predicts potential phosphorylation sites. The *Hermes* transposase residues ,Ser309 and 310, were predicted as being hotspots for phosphorylation. The kinases that are used in the prediction are shown at the top. Ser309 and 310 are underlined in red.

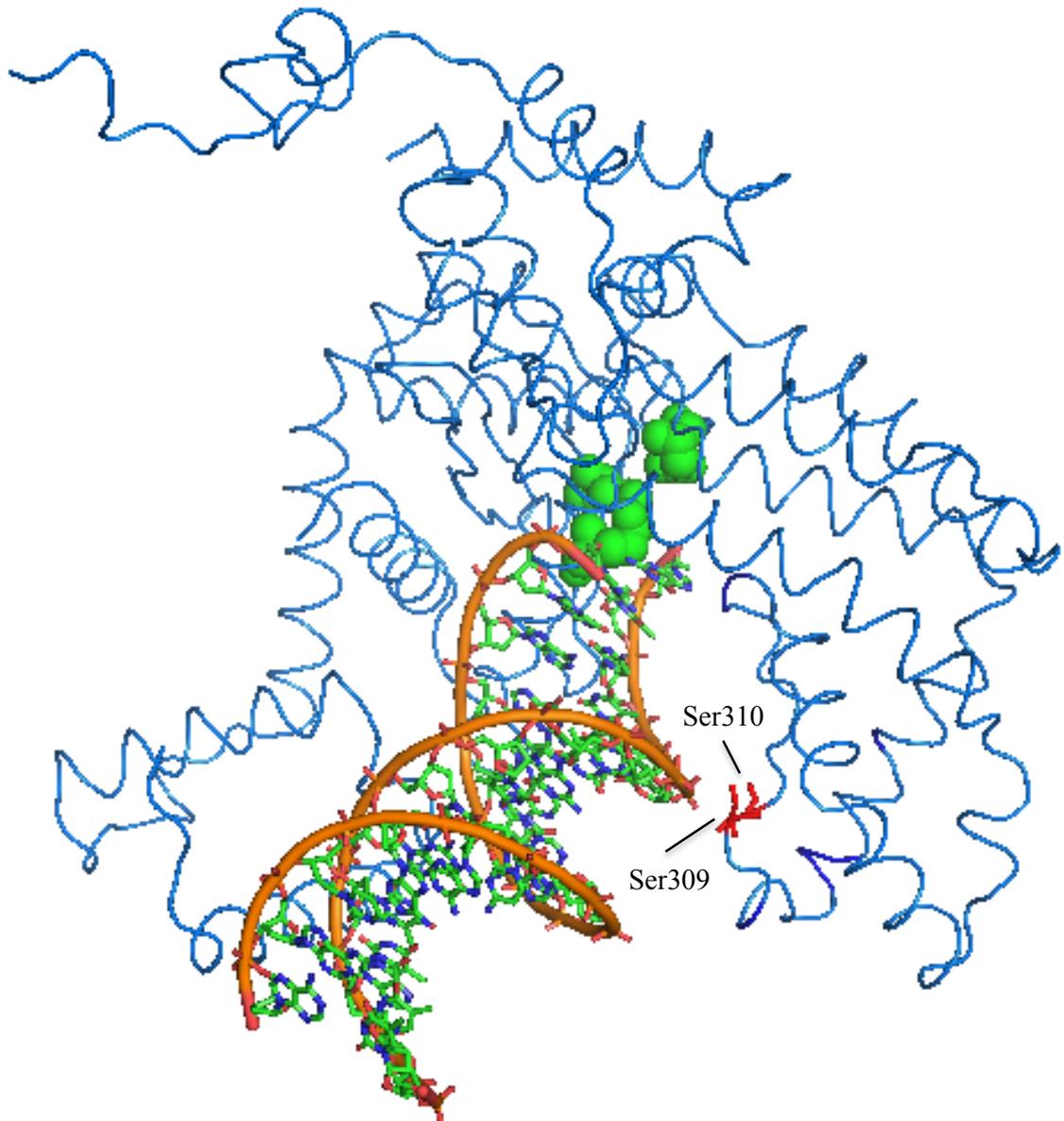


Figure 3.2: Monomer from *Hermes* transposase crystal structure showing the Ser309 and Ser310 positions. A monomer from the *Hermes* octamer crystal structure that was co-crystallized with the first 16 bp of the *Hermes* left end showing the position of Ser309 and Ser310 in red. The active site residues are shown in green.

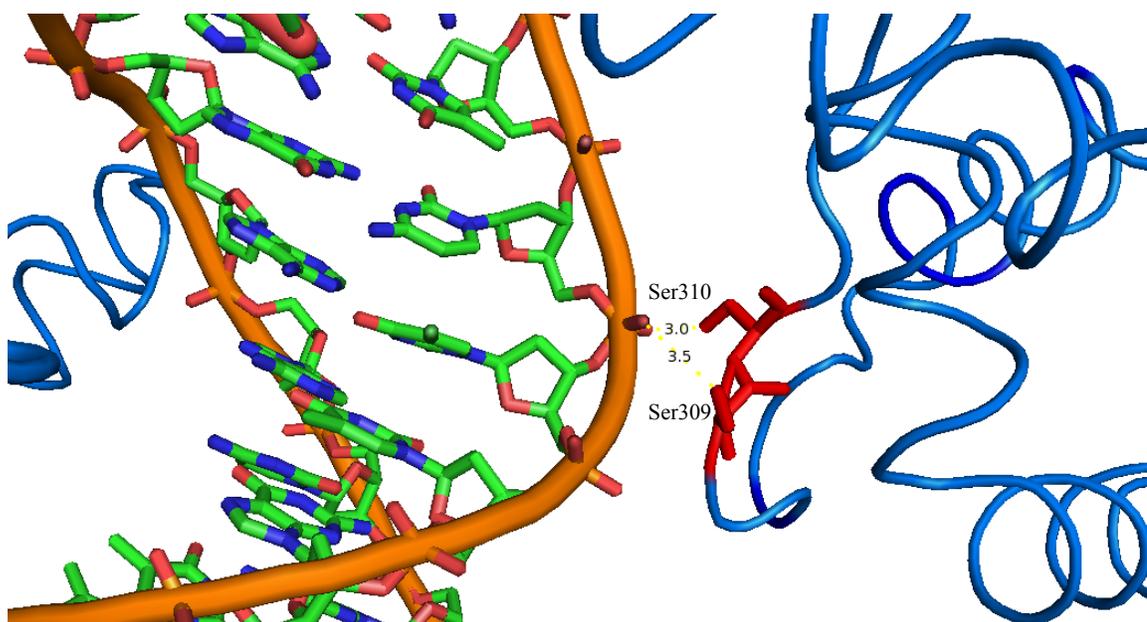


Figure 3.3: Ser309 and Ser310 bind the Hermes transposon left end. The hydrogen bonds of *Hermes* transposase Ser309 and Ser310 residues with the *Hermes* transposon left end and their distances, shown in Å.

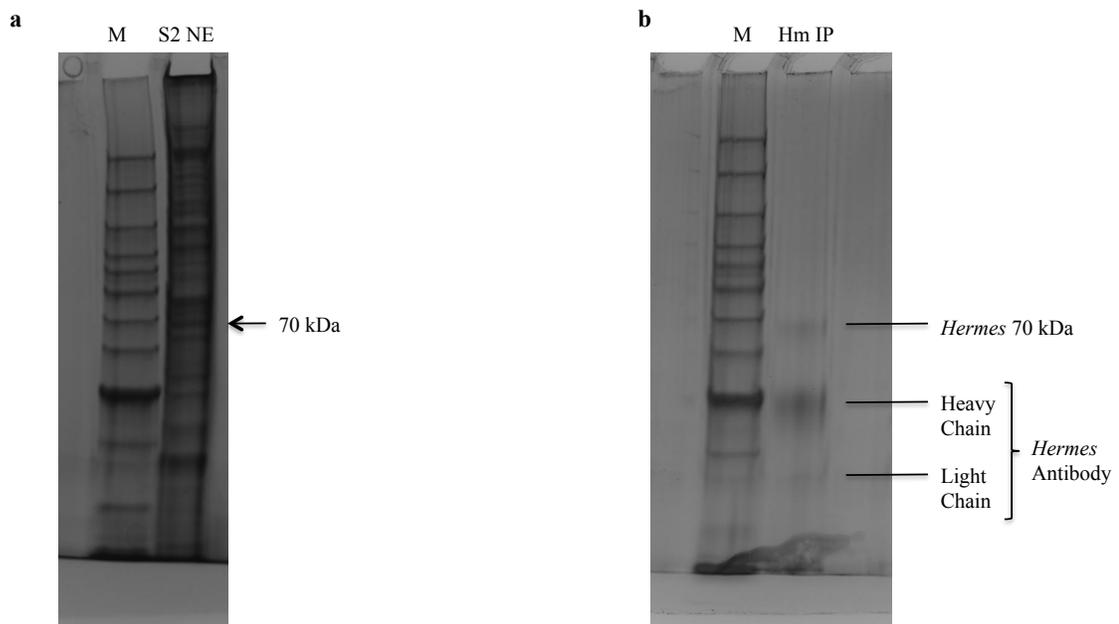


Figure 3.4: Coomassie blue stained SDS PAGE of the S2 nuclear extract and immunoprecipitation pull-down *Hermes* transposase. (a) SDS-PAGE evaluation of S2 cell nuclear extract with the *Hermes* transposase indicated at 70 kDa. (b) SDS-PAGE evaluation of the immunoprecipitated *Hermes* transposase. The heavy and light chains of the *Hermes* antibody are also indicated. An aliquot was run on a gel to confirm the presence of the *Hermes* transposase prior to its submission for mass spectrometry. In both gels the marker is indicated (M).

Observed	Mr (expt)	Mr (calc)	Delta	Miss	Sequence	
820.8588	3279.4061	3278.4745	0.9315	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	(Ions score 35)
824.6101	3294.4113	3294.4695	-0.0582	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Oxidation (M) (Ions score 80)
824.8558	3295.3941	3294.4695	0.9246	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Oxidation (M) (Ions score 110)
840.6044	3358.3885	3358.4409	-0.0524	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Phospho (ST) (Ions score 67)
1120.4769	3358.4089	3358.4409	-0.0320	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Phospho (ST) (Ions score 155)
1120.4773	3358.4101	3358.4409	-0.0308	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Phospho (ST) (Ions score 183)
1120.7935	3359.3587	3358.4409	0.9178	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Phospho (ST) (Ions score 46)
840.8497	3359.3697	3358.4409	0.9288	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Phospho (ST) (Ions score 12)
1125.7964	3374.3674	3374.4358	-0.0684	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Oxidation (M); Phospho (ST) (Ions score 101)
844.6005	3374.3729	3374.4358	-0.0629	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Oxidation (M); Phospho (ST) (Ions score 65)
1126.1228	3375.3466	3374.4358	0.9108	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	Oxidation (M); Phospho (ST) (Ions score 48)
860.5935	3438.3449	3438.4072	-0.0623	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	2 Phospho (ST) (Ions score 8)
1147.1223	3438.3451	3438.4072	-0.0621	0	R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D	2 Phospho (ST) (Ions score 18)

Ser468 ↑ Ser469

230 - 244	593.2588	1776.7546	1776.8335	-44	0	K.AIFSQFNVEDLSSIK.F	Phospho (ST) (Ions score 24)
230 - 244	593.2627	1776.7663	1776.8335	-38	0	K.AIFSQFNVEDLSSIK.F	Phospho (ST) (Ions score 74)
230 - 244	593.2631	1776.7675	1776.8335	-37	0	K.AIFSQFNVEDLSSIK.F	Phospho (ST) (Ions score 44)
230 - 244	593.2650	1776.7732	1776.8335	-34	0	K.AIFSQFNVEDLSSIK.F	Phospho (ST) (Ions score 60)
230 - 244	889.4048	1776.7950	1776.8335	-22	0	K.AIFSQFNVEDLSSIK.F	Phospho (ST) (Ions score 1)

↑
 Ser233

Figure 3.5: Peptide fragments results from mass spectrometry of immunoprecipitation pull-down *Hermes* transposase from S2 cells. Peptide fragments shown to be phosphorylated after trypsin digest are shown based on Mascot results of the mass spectrometry of S2 cell isolated *Hermes* transposase. Ion scores, shown in blue on the right, are a measurement of how well the MS signal for each experimental peptide fragment matches the calculated fragment. Scores >30 are considered strong matches. The phosphorylated residue(s) is identified with the arrow.

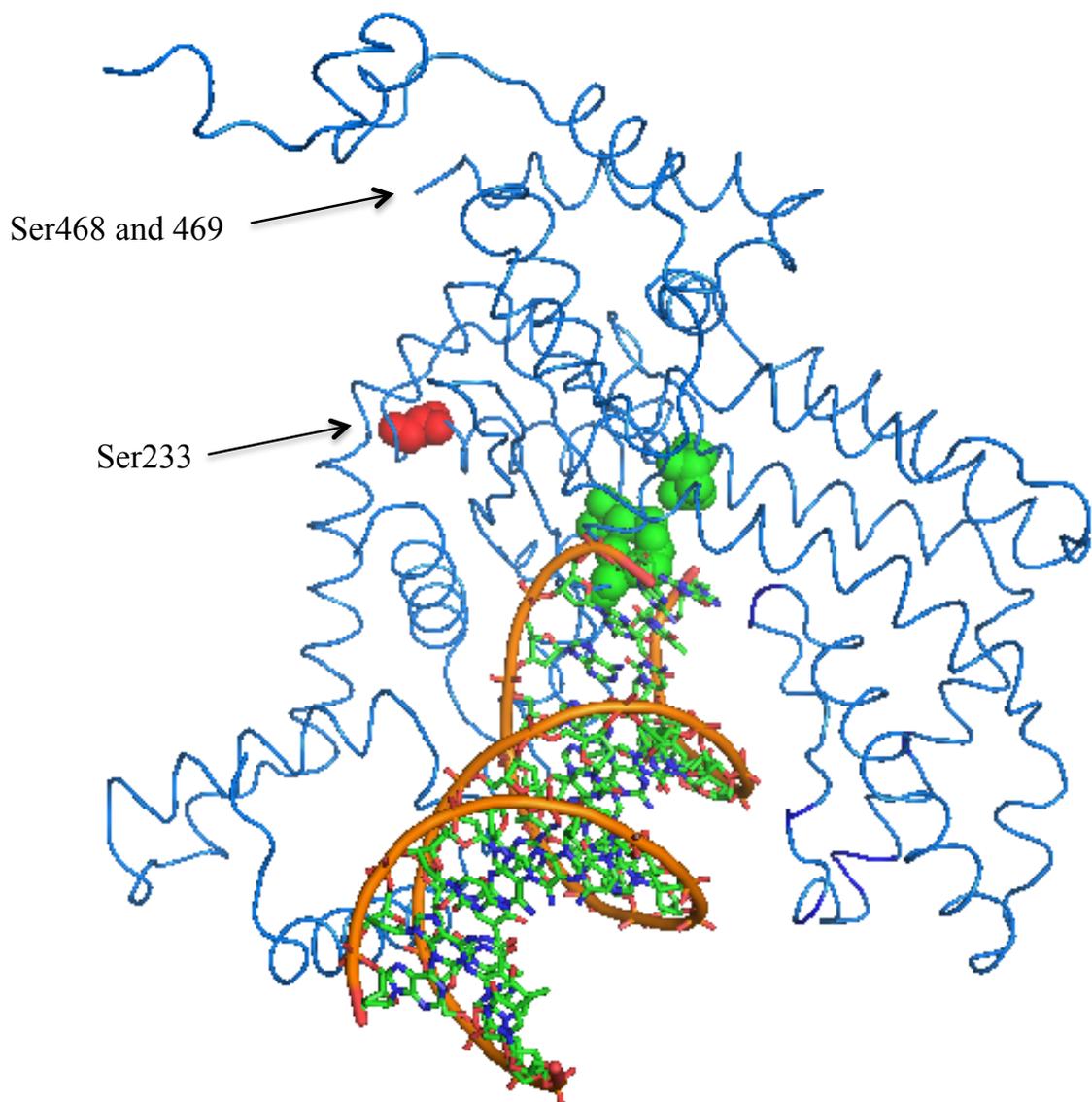


Figure 3.6: Location of Ser233, Ser468 and Ser469 in the Hermes transposase octamer crystal structure. A monomer from the *Hermes* transposase octamer crystal structure that was co-crystallized with the first 16 bp of the *Hermes* transposon left end showing the position of Ser233 in red. The active site is shown in green. The positions of Ser468 and Ser469 are indicated with the arrow. These residues are not resolved in the octamer structure.

+P 468/69 Peptide Fragment: Trypsin Digest
R.MSSFNELSATQLNQSDSNHNSIDLTSHSK.D

-P 468/69 Peptide Fragment: Chymotrypsin Digest
F.NELSATQLNQSDSNHNSIDLTSHSKDISTTSF.F

Figure 3.7: Trypsin and Chymotrypsin digest fragments. The first immunoprecipitated *Hermes* sample submitted to MS was digested with the trypsin proteolytic enzyme which cleaves after positively charged residues, resulted in the top peptide fragment and was positive for phosphorylation. A second immunoprecipitated *Hermes* sample was submitted and digested with chymotrypsin, which cleaves after tyrosine, tryptophan, and phenylalanine. Under these conditions the peptide fragment without the serine 468 and 469 residues and was not phosphorylated. This led us to conclude that the one or both of the residues were being phosphorylated. The regions that overlap between the two fragments are shown in red.

Trials	Hermes Transposase	Injected	Survived	Eclosed	Crossed	Fertile	Sterile	No. Transgenic	Sterility Rate	Transformation Frequency
3	WT <i>Hermes</i>	679	254	77	69	52	17	15	24.6%	28.8%
3	SS309/10AA	692	329	153	131	107	24	4	18.3%	3.7%
3	SS309/10TT	715	344	214	198	181	17	28	8.6%	15.5%
3	SS309/10DD	602	289	178	144	131	13	1	9.0%	0.76%

Table 3.1: The *Drosophila* germline transformation frequency of WT *Hermes* and the each of the SS309/10 mutant transposases. The SS309/10AA and SS309/10DD mutant *Hermes* transposases each had *Drosophila* germline transformation frequencies that were much lower than WT *Hermes* transposase. The transformation frequency was calculated by dividing the number of transgenic by the number of fertile crosses.

No. Injections	Hermes Transposase	Hermes DT	pBac DT	Hermes Events	pBac Events	Hermes Freq.	pBac Freq.	Fold Difference	Stdev Hermes	Stdev pBac
3	WT <i>Hermes</i>	163,000	152,000	67	57	4.1×10^{-4}	3.8×10^{-4}	+ 1.05	2.0×10^{-4}	3.4×10^{-4}
3	SS309/10AA	96,400	115,800	105	121	1.1×10^{-3}	1.0×10^{-3}	+ 1.1	1.7×10^{-4}	5.2×10^{-3}
1	SS309/10TT	246,000	230,000	8	23	3.2×10^{-5}	1.0×10^{-4}	- 3.1	-	-
3	SS309/10DD	470,000	568,800	0	228	0	4.0×10^{-4}	-	-	3.0×10^{-4}
1	SS468/9AA	38,900	62,000	76	102	1.95×10^{-3}	1.6×10^{-3}	+ 1.5	-	-

Table 3.2: Interplasmid transposition rates in injected *Drosophila* embryos for the phosphorylation mutants. Each of the samples was co-injected with the *piggyBac* (PB) transposon donor and helper plasmids as an internal control. Each event represents an independent event. The frequency was calculated by dividing the number of events by the donor titer (DT) in each of the injection sets. The DT is the number of donor plasmids that were available.

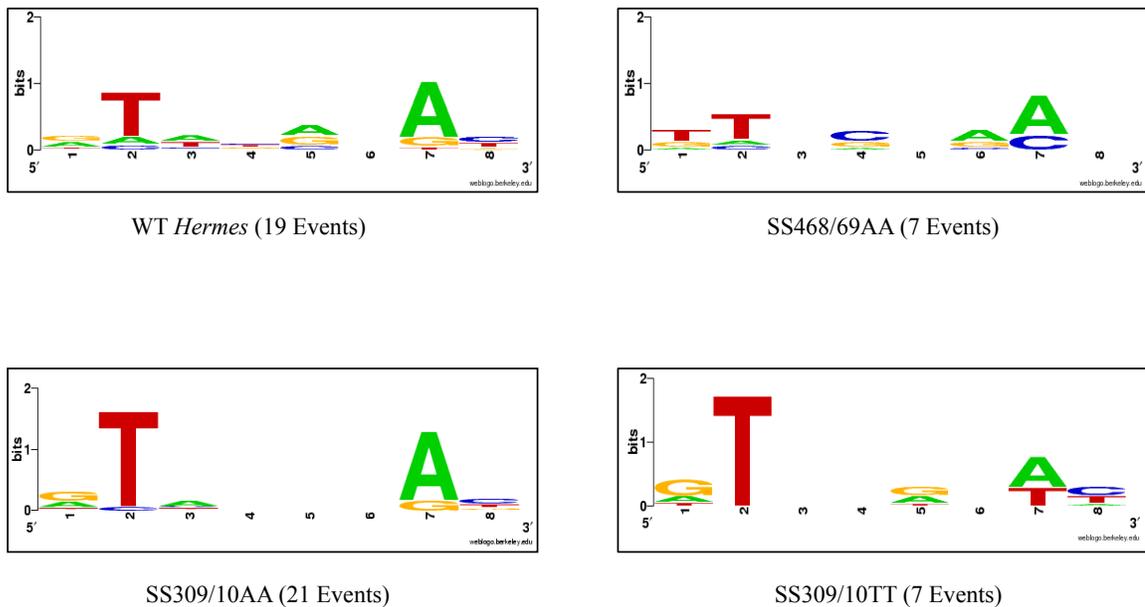


Figure 3.8: Weblogo consensus target site duplication for phosphorylation mutant *Hermes* transposases. The number of independent events that were sequenced to create the consensus are indicated above.

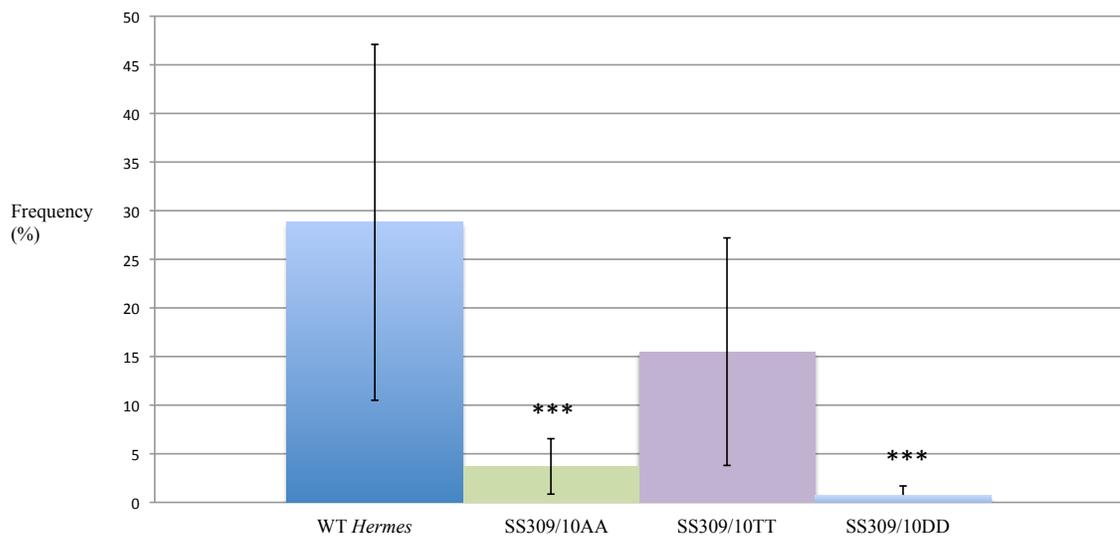


Figure 3.9: Graphical representation of the *Drosophila* transformation frequency of WT *Hermes* and the SS309/10 mutants. The SS309/10AA and SS309/10DD mutant transposases had statistically significant decreases in their transformation efficiency. The error bars represent the standard deviation for each *Hermes* transposase. The asterisks signify a statistically significant difference compared to WT *Hermes*. These were supported by calculating the P-value using the Fisher's exact test which gave a p-value for SS309/10AA, SS309/10TT, and SS309/10DD of 1.2×10^{-2} , 0.13, and 6.0×10^{-3} respectively. In this statistical analysis 0.05 is considered significant.

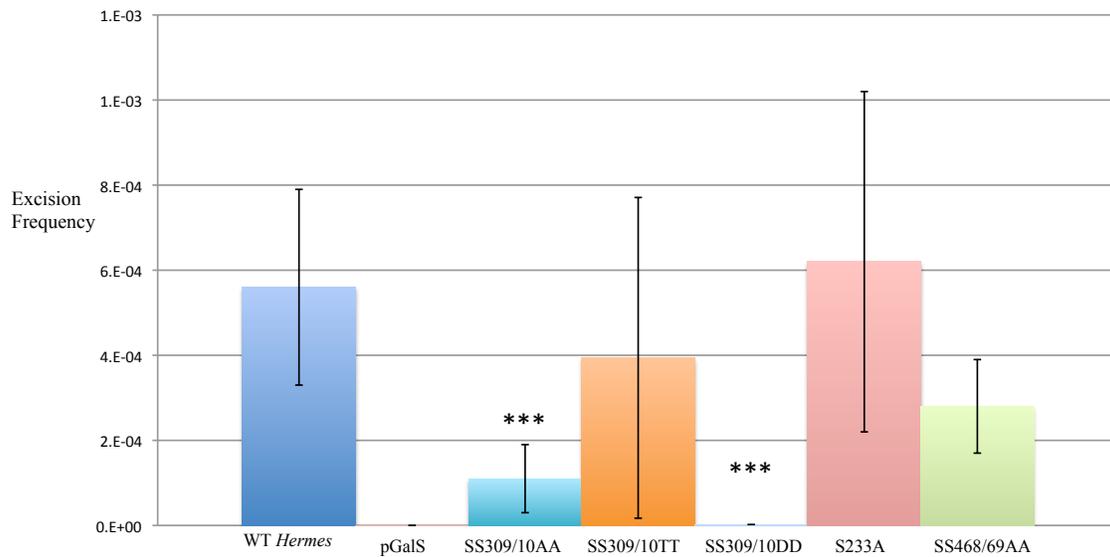


Figure 3.10: The yeast excision frequency of WT *Hermes*, phosphorylation mutants, and negative control pGalS. The error bars represent the standard deviation for each of the *Hermes* transposases. The asterisks signify a statistically significant difference compared to WT *Hermes*. Using the Fisher's exact test the p-value was calculated for each of the above. SS309/10AA, SS309/10TT, SS309/10DD, S233A, and SS468/69AA had p-values of 4.3×10^{-9} , 0.013, 2.7×10^{-13} , 0.76, 1.5×10^{-4} , respectively.

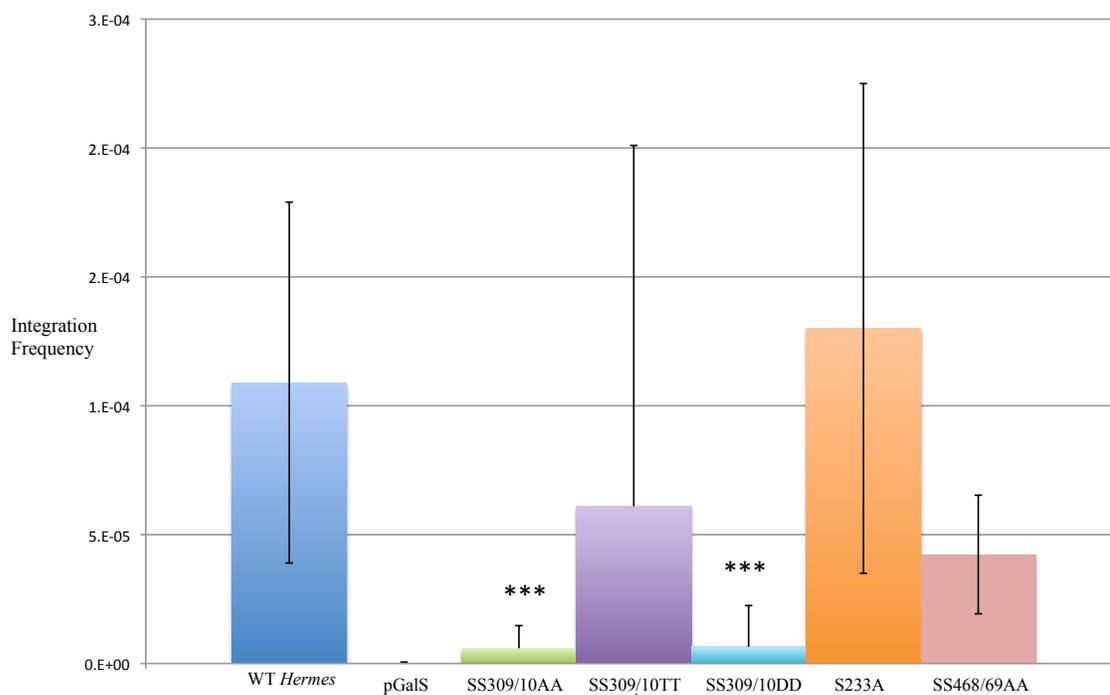


Figure 3.11: The yeast integration frequency of WT *Hermes*, phosphorylation mutants, and the negative control pGalS. The error bars represent the standard deviation for each of the *Hermes* transposases. The asterisks signify a statistically significant difference compared to WT *Hermes*. SS309/10AA, SS309/10TT, SS309/10DD, S233A, and SS468/69AA had p-values of 3.6×10^{-6} , 0.02, 4×10^{-6} , 0.86, 2.2×10^{-3} , respectively.

Chapter Four

Summary and Conclusions

4.1 Summary

This dissertation provides, for the first time, a model for *Hermes* target DNA binding which we coined the “rim of the wheel” model. This model was supported by data showing that mutation of three of the four residues hypothesized to be involved in target DNA binding affected transposition activity of *Hermes* and that the TSD for one of the mutated residues changed. This model was also supported by the crystallization of the octamer structure in which *Hermes* transposon LE fragments bound in a manner consistent with my hypothesis. In addition this dissertation provides direct evidence of phosphorylation of a eukaryotic class II transposase. This type of direct evidence has not been observed before with any other transposase and may play a role in regulating *Hermes* activity.

4.2 The *Hermes* “Rim of the Wheel” Model

The publication of the *Hermes* hexamer crystal structure in 2005 allowed the opportunity to develop a model for target DNA binding (Hickman *et al.*, 2005). Although the structure was crystallized without transposon end DNA, the surface charge landscape showed positively charged channels that led into the active sites. It was hypothesized that the positively charged channels could be important for DNA binding (Hickman *et al.*, 2005). Building upon this idea I investigated the hexameric structure

and found that on the rim of the circular wheel-like structure there resided a positively charged trench that knitted together the area between the active sites 35 Å apart; close to the 27 Å required to bind to an 8 bp target site. The residues that contributed to the positively charged trench were K292, K299, K300, and R306. These residues were also predicted to bind DNA when we used the DISPLAR bioinformatics server. This program predicts DNA binding residues based on the three dimensional structure of the protein. I therefore came up with a model coined the “rim of the wheel” which posits that the ends of the *Hermes* transposon bind in the positively charged channel directed towards the outside of the “wheel”. This positions the ends in the active sites approximately 35 Å apart where they could be integrated into the target DNA that we hypothesize is bound by the positively charged trench on the rim of the “wheel”. The rim of the wheel model for target DNA binding was supported by the *Hermes* octamer structure that was recently solved by Fred Dyda and Alison Hickman (F. Dyda and A. Hickman, personal communication). The octamer structure was co-crystallized with a *Hermes* LE fragment bound in each of the active sites. The LE’s are bound in the positively charged channel that leads from the middle of the “wheel” and end in the active sites and position the ends on the rim of the “wheel”. Furthermore, three of the four residues that were identified with the hexamer structure still comprised the positively charged trench; K292, K299, and K300. In the octamer structure R306 has moved out of the positively charged trench due to conformational changes to accommodate LE DNA binding.

Each of the four residues comprising the positively charged trench were individually mutated to alanine and examined for their effects on target DNA selection

and transposition activity. While only one mutant, K300A, actually exhibited a change in its TSD, two other mutants K292A and K299A had decreased activity in *Drosophila* germline tissue, yeast excision, and coupled cleavage and strand transfer. The R306A mutant had close to WT *Hermes* activity in almost every assay in was tested in.

As mentioned previously, K300A exhibited a statistically supported change in its TSD in *Drosophila* somatic tissue, having less of a preference for adenine in the seventh position of its 8 bp target site than WT *Hermes*. While an altered TSD did not affect its somatic activity the mutation affected *Hermes* germline activity for reasons that remain unknown. This mutation did not appear to have an effect on *Hermes* transposition activity in yeast as K300A was able to both excise and integrate. Purified K300A protein was tested biochemically for its ability to integrate into, and cleave, DNA and for its affinity for the *Hermes* LE with and without flanking DNA. *In vitro* K300A showed slightly lower strand transfer activity in the presence of both Mn^{2+} and Mg^{2+} but this difference was minimal compared to WT *Hermes*. Additionally, when this mutant was tested for its ability to both cleave and then integrate the *Hermes* LE *in vitro* it had lower than WT *Hermes* coupled cleavage and strand transfer activity in the presence of both metal ions tested but especially in the presence of Mg^{2+} . This may have to do with the consequences of the conformational change associated with Mg^{2+} binding on the mutant's ability to cleave. This type of compromised activity in coupled cleavage and strand transfer assays was seen for all mutants and may be an affect of the concentration of Mg^{2+} . The decreased coupled cleavage and strand transfer activity of K300A was not due to this mutant's affinity for the first 30 bp of the transposon LE nor to 92 bp of LE

and 101 bp of flanking genomic DNA as K300A had WT levels of binding in gel shift assays.

While K300A was the only mutant to show a change in its TSD, both K292A and K299A mutants had decreased activity compared to WT *Hermes* in *Drosophila* transformation, yeast excision, and coupled cleavage and strand transfer suggesting that these residues are essential for *Hermes* transposition. The biochemical analysis and results in yeast assays of these residues suggested that they may be important for excision as this appeared to be compromised in coupled cleavage and strand transfer but they could both integrate when supplied with pre-cleaved transposon LE. Integration cannot occur without excision, therefore if these residues are playing a major role in excision as opposed to target DNA binding that may explain their lower transposition frequencies in *Drosophila* germline and in the yeast excision assay. The integration frequency in yeast of K292A was not statistically significant compared to WT *Hermes*, even though it had negative control levels of excision in yeast. This may be due to imprecise excision events by K292A that make it difficult to repair of the double strand breaks giving a false negative excision frequency. The coupled cleavage and strand transfer activity of K292A was lower in the presence of Mg^{2+} than Mn^{2+} although the strand transfer activity for this mutant was close to WT. The *in vitro* results for the couple cleavage and strand transfer of K292A were not due to a change in the affinity of the protein to the first 30 bp of the transposon LE nor 92 bp of LE and 101 bp of flanking genomic DNA as they had the same gel shift profile as WT *Hermes*. These results may suggest a role for K292 in the excision chemistry during transposition.

K299A had the most compromised activity in every assay tested except for strand transfer. In the coupled cleavage and strand transfer assay K299A had decreased activity in the presence of Mn^{2+} and almost no activity in the presence of Mg^{2+} . Just as with all the other mutants, the activity of K299A *in vitro* was not due to its affinity to the *Hermes* transposon LE or LE containing flanking DNA. More biochemical testing is needed to determine the exact role of these residues in transposition but the data provided here suggests that K292 and K299 may both be involved in excision of the *Hermes* transposon. These two residues in the octameric structure are positioned facing in toward the positively charged trench and the ends of the LE. If the flanking DNA of the transposon ends were present in the octameric structure they would run straight out from the “wheel” and could possibly make contact with the K292 and K299 residues. These structural aspects of the crystal structure help support the data presented here.

As mentioned earlier R306 does not appear to be an essential residue for *Hermes* transposition as it had near WT transposition levels in almost every assay tested but coupled cleavage and strand transfer. Again, more testing would need to be performed to determine why in the presence of Mg^{2+} these mutants have such reduced activity in this assay versus WT. This residue comprised the positively charged trench in the 2005 crystal structure but in the octamer it appears that this residue has moved out of the positively charged trench due to conformational changes associated with left end binding (Hickman *et al.*, 2005, F. Dyda and A. Hickman, personal communication).

The results presented here taken together support the “rim of the wheel” model. It appears that a region of the *Hermes* transposase has been identified that works as a bi-

functional domain responsible for both excising the transposon (K292 and K299) and binding to target DNA (K300). This type of bi-functionality within a target DNA binding domain has been seen before with Tn5 and the work presented here adds to the growing work on our understanding of transposase integration biochemistry (Gradman *et al.*, 2008).

4.3 *Hermes* is Phosphorylated in *Drosophila* S2 Cells

For the first time we have directly detected phosphorylation of a transposase protein using mass spectrometry (MS). We employed two methods of analysis in our investigation of phosphorylation of *Hermes*, a bioinformatics approach and MS. In our original investigation using the bioinformatics program, NetPhosk, two residues were identified as being targets of phosphorylation by up to four different kinases. These residues, S309 and S310 were doubly mutated to non-phosphorylatable alanine residues, chemically identical threonine residues and phosphorylation mimicking aspartic acids. These mutants, SS309/10AA, SS309/10TT, and SS309/10DD were tested for their *in vivo* activity in *Drosophila* and yeast. In the *Hermes* octameric structure these serine residues make hydrogen bonds to the LE. Because of the importance of their function and the position of these residues, my hypothesis was that mutation of these residues to alanine would compromise the transposition activity of *Hermes* because they are not available to make hydrogen bonding interactions with the LE.

The somatic activity of SS309/10AA was not statistically different from WT *Hermes* when compared to the activity of the internal *PB* transposition frequency. The

activity of this mutant in *Drosophila* germline tissue was very reduced as it had a roughly 10-fold decrease in its transformation rate compared to WT *Hermes*. This trend was also seen in its excision and integration frequencies in yeast. This was not a surprising result as it was hypothesized that the activity of this mutant would be compromised because the serine residues are not there to bind to the transposon end. Presumably total activity would not be lost because binding of the end is preformed by several residues.

Building upon this model, it was hypothesized that SS309/10TT would have the same level of activity as WT because it has the same chemical properties and could still bind to the LE. While there was a general trend towards lower activity in each of the assays that this mutant was tested in no statistically significant difference from wild-type activity was observed. This decrease may be due to the size difference of the threonine residue compared to serine causing there to be a slight difference in the ability of this mutant to bind to the transposon end. Finally, my model indicated that the addition of two negatively charged residues, which mimic phosphorylation, would abolish the binding capability of *Hermes* due to the repulsiveness between these residues and the negatively charged backbone of the transposon end to which the serine residues usually bind to. Surprisingly, this was not the result of the assays with SS309/10DD. While the activity of this mutant was severely compromised in *Drosophila* somatic tissue as no events were recovered, the mutant retained some activity in the germline. Additionally, while in yeast excision this mutant had close to negative control activity levels, in yeast integration assays SS309/10DD had the same level of activity as SS309/10AA.

The fact that mutating S309 and S310 to alanine had such a dramatic effect on transposition demonstrated that these residues are important for transposition. Furthermore, although neither of these residues were found to be phosphorylated in MS analysis from *Drosophila* S2 cells, if these residues are phosphorylated it can be assumed that it would have a severe effect on transposition as demonstrated by the results of SS309/10DD.

In our second approach to investigating *Hermes* phosphorylation we used MS analysis of *Hermes* transposase isolated from S2 NE. As the MS data showed, in *Drosophila* somatic tissue S233 and S468 and/or 469 were phosphorylated. This is the first time that phosphorylation has been directly detected for a transposase. Although the *P*-element was found to be sensitive for phosphorylation, this was determined indirectly through analysis of the transposition activity of non-phosphorylatable and phosphorylation mimicking mutations (Beall *et al.*, 2002). S233 was mutated to a non-phosphorylatable alanine and a double mutant for S468 and S469 was also made: S233A and SS468/69AA respectively. These mutants were tested in yeast excision and integration assays that allowed for a high throughput method of analysis of activity.

Although S233A had a higher excision and integration frequency than WT *Hermes* it did not show a statistically significant frequency in either assay. Interestingly, S233 is a target for DNA-PK, a major component of the NHEJ pathway for double strand break repair. Although the catalytic subunit for this kinase is absent in *Drosophila* and yeast, both of these systems are able to perform NHEJ and it is thought that there are other ATM family kinases responsible for phosphorylation of proteins in this pathway

(Beall *et al.*, 2002). The results therefore imply that phosphorylation of this residue is not essential for activity of the *Hermes* transposase. That does not exclude the possibility that phosphorylation of this residue does not negatively impact the activity. Further testing is needed to determine whether a pseudo-phosphorylated mutation at S233 has an impact on its transposition activity.

Although the SS468/69AA non-phosphorylatable mutant showed a trend of slightly decreased excision and integration frequencies in yeast, these results were not statistically significant. Additionally, in *Drosophila* somatic tissue in interplasmid transposition assays, even though SS468/9AA had a higher transposition frequency than WT *Hermes* it was not statistically significant compared to the internal *PB* control. These residues are in an area of low electron density in the *Hermes* octamer structure and were not able to be resolved. Additionally, these residues are in a region that is believed to be important for oligomerization as the alpha helix immediately before this area makes contact with the adjacent monomer, creating interface 2 in the octameric structure. Although the results imply that these residues are not dependent on phosphorylation for transposition activity further testing will be needed with pseudo-phosphorylated mutants to determine whether their phosphorylated state has an impact on transposition and oligomerization.

Additionally, since these residues were detected to be phosphorylated in S2 cells which are a *Drosophila* somatic tissue cell line they need to be tested in *Drosophila* somatic tissue and germline.

The work presented on *Hermes* phosphorylation demonstrated for the first time direct evidence for the phosphorylation of a eukaryotic class II transposase protein. The

effect of phosphorylation will have to be determined in order to understand the consequences of post-translational modification of *Hermes* transposition *in vivo*. Additionally, we identified two residues in *Hermes*, S309 and S310, which are important for transposition and adds to our functional understanding of essential residues in the *Hermes* transposase.

4.4 Conclusions

The developing fields of gene therapy and genetic transformation are constantly searching for new autonomously replicating TEs with the ability to stably integrate into the genome of a wide range of diverse hosts. The *Hermes* transposase has become a particularly attractive element because of its ability to transform a wide range of hosts, the diverse distribution of *hAT* elements in nature, and the extensive genetic, molecular, and biochemical data available (Arensburger *et al.*, 2011, Hickman *et al.*, 2005, Michel and Atkinson, 2003, Michel *et al.*, 2003, Sarkar *et al.*, 1997, Zhou *et al.*, 2004). In order to develop *Hermes* as an efficient molecular biology tool it is important to understand (1) how the transposase chooses its target DNA for integration and (2) regulatory mechanisms which may effect transposition activity.

The work presented in this dissertation identifies, for the first time, residues in *Hermes* that are important for target DNA binding. These residues are part of a model that I call the “rim of the wheel” based on the *Hermes* crystal structure. Data collected from a comprehensive analysis that teased apart the different steps of transposition suggests that these residues may contribute to a bi-functional domain responsible for both

excising the transposon (K292 and K299) and binding to target DNA (K300). Bifunctionality within a target DNA binding domain has also been demonstrated with the Tn5 transposase (Gradman *et al.*, 2008). The work presented here adds our understanding of transposase integration biochemistry and can serve as a model for other elements.

In order for *Hermes* to be used as a biological tool for gene therapy or genetic studies in medically and agriculturally important insects it is important that *Hermes* has its greatest chance for successful integration into its host. The other goal of this dissertation was to understand whether *Hermes* was subject to phosphorylation and whether this affected transposition. The *Hermes* transposase was found to be phosphorylated at two residues, S233 and S468 and perhaps also at S469. Direct evidence for phosphorylation of a transposase has not been reported before highlighting the significance of this data. Whether phosphorylation of these residues had an effect on their transposition was not determined and should be further investigated. Indirectly two other serine residues in *Hermes* were found to be important for transposition, as mutating them to residues with different chemistries decreased *Hermes* transposition frequency. This work has expanded our still growing understanding of essential residues that are functionally important for *Hermes* transposition.

4.5 References

- Arensburger, P., Hice, R., Zhou, L., Smith, R.C., Tom, A.C., Wright J.A., Knapp, J., O'Brachta, D.A., Craig, N.L., Atkinson, P.W. (2011) Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics*. **188**: 45-57.
- Beall, E.L., Mahoney, M.B., Rio, D.C. (2002) Identification and analysis of a hyperactive mutant form of *Drosophila P-element* transposase. *Genetics*. **162**: 217-227.
- Gradman, R.J., Ptacin, J.L., Bhasin, A., Reznikoff, W.S. and Goryshin, I.Y. (2008) A bifunctional DNA binding region in Tn5 transposase. *Molec. Microbiol.* **67**: 528-540.
- Hickman, A.B., Perez, Z.N., Zhou, L., Primrose, M., Ghirlando, R., Hishaw., J.E., Craig, N.L., Dyda, F. (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.* **12**: 715-721.
- Michel, K., Atkinson, P.W. (2003) Nuclear localization of the *Hermes* transposase depends on basic amino acid residues at the N-terminus of the protein. *J. Cell. Biochem.* **89**: 778-790.
- Michel, K., O'Brochta, D.A., Atkinson, P.W. (2003) The C-terminus of the *Hermes* transposase contains a protein multimerization domain. *Insect Biochem. Molec. Biol.* **33**: 959-970.
- Sarkar, A., Coates, C.J., Whyard, S., Willhoeft, U., Atkinson, P.W., O'Brochta, D.A. (1997) The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica*. **99**: 15-29.
- Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F. and Craig, N.L. (2004) Transposition of *hAT* elements links transposable elements and V(D)J recombination. *Nature*. **432**: 955-1001.