THE UNIVERSITY OF CHICAGO

DNA DYNAMICS AND IRREVERSIBILITY REGULATE MU TRANSPOSITION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY

JAMES RICHARD FULLER

CHICAGO, ILLINOIS

AUGUST 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My work here would not have been possible without the work of countless other scientists. They have built the mountain of knowledge upon which I hope to have dropped my few pebbles of data and ideas. The Mu transposition system, in particular, would not be available to me as a model system if it weren't for several happy concurrences of chance and brilliance that lead to its discovery and subsequent refinement for *in vitro* biochemical experiments. Many of these breakthroughs occurred decades before my birth, which has been both empowering and, at times, a challenging legacy to extend. I'd like to thank the following scientists for paving the Mu path: Larry Taylor, Kiyoshi Mizuuchi, Ahmad Bukhari, George Chaconas, Rasika Harshey, and Harri Savilahti.

More proximally, two people are responsible for setting me up to work on this project successfully. First, Dr. Phoebe Rice, my advisor: Phoebe deserves many thanks for providing me a space in her lab, encouragement and, perhaps most importantly, choosing interesting ideas to work on. The quality of our underlying questions kept me motivated even when exploring them successfully and to our desired precision seemed daunting. Second, Dr. Sherwin Montaño: Sherwin solved the crystal structure of the Mu transpososome as I was joining the Rice lab, a structure that served as the foothold by which we were able to ask more detailed questions about MuA biochemistry. This project would not have existed without that structure. He also designed the SinMu system, which allowed me to probe MuA using mutations with unprecedented effectiveness. Ying Pigli also deserves considerable credit for the progress I have made here. Without her advice on the topics of buffers, cloning, and protein purification, the experiments themselves would never have been actualized.

The Rice lab would not have been nearly as engaging of a workplace were it not for the other members I overlapped with: Ross Keenholtz, Aga Misiura, Caitlin Trejo, Justin Salat, Ignacio Mir Sanchis, and Vishok Srikanth. Biochemistry and molecular biology are accomplished largely by the precise movement and mixing of clear liquids; I am thankful that the Rice lab members added a vibrant human element and humor to the mix. They were aided by the members of the flanking Crosson and Pan labs, whose camaraderie was similarly indispensable.

Outside of lab, my life has been enriched by so many members of the cohort of graduate students in the molecular biosciences cluster. To the six of us who entered the biochemistry program in the fall of 2010 – who can be conveniently recalled by the acronym JAKJAK – I am so proud that we will all make it to our doctorates. To all those who have ever occupied my apartment on a Wednesday evening, thank you so much for being my friend and making graduate school such a meaningful time in my personal life.

Michelle Stein, thank you for showing me how strong a bond between two people can be. You have accompanied me through this time with great love and encouragement. To give back to you as much as I receive is a challenge I look forward to facing every day.

Finally, I would not have made it to graduate school at all without the support of my parents, Richard and Penny Fuller. My mother encouraged my interest in science from a very young age, and my father encouraged me to follow whatever it was I wanted to study regardless of the bumps along the way; together, they prepared me well for this endeavor. I am so very grateful for the life, time, and love they have given me.

ABSTRACT

Transposition moves or copies specifically defined segments of DNA from one location to another. The transposon – the mobile DNA – is delineated by sequences at each end that mark its boundary. In many transposons, these ends are specified by DNA sequences to which copies of the transposase protein bind. To accomplish transposition, multiple copies of the transposase and accessory proteins synapse both ends of the transposon DNA into a transpososome, a large nucleoprotein complex. The transposase can then cleave the host-transposon DNA boundary and facilitate the joining of those ends to a new target site. Many transposons whose behavior has been studied in detail belong to the DDE family of transposition systems. In addition to many transposons, this family also includes the integrase proteins used by retroviruses to insert their viral DNA to their host's genome. Although not related by sequence, the available high resolution structures of transpososome and intasomes from the DDE family show that they have converged on key strategies to ensure the fidelity of their transposition activity.

Mu is an *Escherichia Coli* bacteriophage that replicates its genome using transposition. Its transposase, MuA, is a DDE transposase for which a high resolution transpososome structure has been solved. We have built on this structure and decades of MuA biochemical observations to test broadly-applicable hypotheses about how transposition is regulated. We have made the first precise *in vitro* measurements to test how altered DNA flexibility or conformations can modulate destination site DNA capture and attack. Bent or very flexible DNA is highly preferred as a destination, and continued bending after transposition is necessary to keep the transposon joined to its destination. This has direct implications for how DDE transposases and retroviral integrases select their target sites. We have also begun to de-convolute the subunit rearrangements that occur during Mu transpososome assembly process. Overcoming its

intransigence towards traditional structural techniques, we have shown by crosslinking and SAXS that the MuA protein is constitutively a monomer prior to engaging Mu end binding sites. By combining this observation with the characterization of a mutant that dramatically accelerates assembly, we have proposed a more detailed pathway for the conformational changes that create a catalytically competent transpososome core.

Chapter 1

Introduction to replicative transposition

1.1 Recombination: Changing the connectivity or topology of DNA elements

The information encoded in the sequence of DNA is ultimately responsible for orchestrating the molecular events inside a cell. During processes collectively known as recombination, large contiguous segments of this information can change in topology or location within the genome. Broadly, there are three routes by which this can occur: transposition, site-specific recombination, and homologous recombination. Homologous recombination exchanges DNA segments based on flanking tracts of similar but unspecified sequence, and is exemplified by the process by which double-stranded breaks in the genome are repaired[1] (Figure 1.1A). In contrast, transposition and site-specific recombination act at specific DNA sequences to mediate the movement of DNA segments or changes in their connectivity, respectively[2].

Transposition and site-specific recombination are carried out by transposases and recombinases, proteins that catalyze both the cleavage and joining of the DNA phosphate backbone. Often these proteins recognize and bind to specific sequences at the ends of their mobile DNA element and synapse the ends together into a single complex, which results in their coordinated exchange or transfer. During site-specific recombination, DNA connectivity is exchanged between two recombination sites[3] (Figure 1.1B). Generally, these sites are marked by flanking inverted repeats that are the binding sites for the recombinase. A single site-specific recombination event can either invert a segment of DNA or cause integration or excision of a circular segment, depending on how the cleaved strands at the recombination sites are re-

connected. Transposition, the movement or copying of a DNA segment ("transposon") from one location to another, can occur as a result of two site-specific recombination events that excise a DNA segment as a circular intermediate and then integration it elsewhere. In this case, the choice of new destination ("target") sites is limited to those sites which also contain a recombination site. However, a wide range of other mechanisms have been discovered for the transposition of DNA, which will be discussed in the next section. Unlike site-specific recombination, many other transposition systems have no requirement for homology or specific sequences in their target sites. These systems rely on other methods besides sequence-specific DNA binding to engage their target sites and thus can, in theory, transpose to almost any sequence location[2] (Figure 1.1C). Nevertheless, site-specific recombination and transposition without target site specificity can serve overlapping biological functions in the integration and excision of transposons (including but not limited to so-called "selfish" genetic elements) and viral genomes. Notably, the bacterial Tn7 transposon is capable of switching between both methods to facilitate its survival and spread[4].

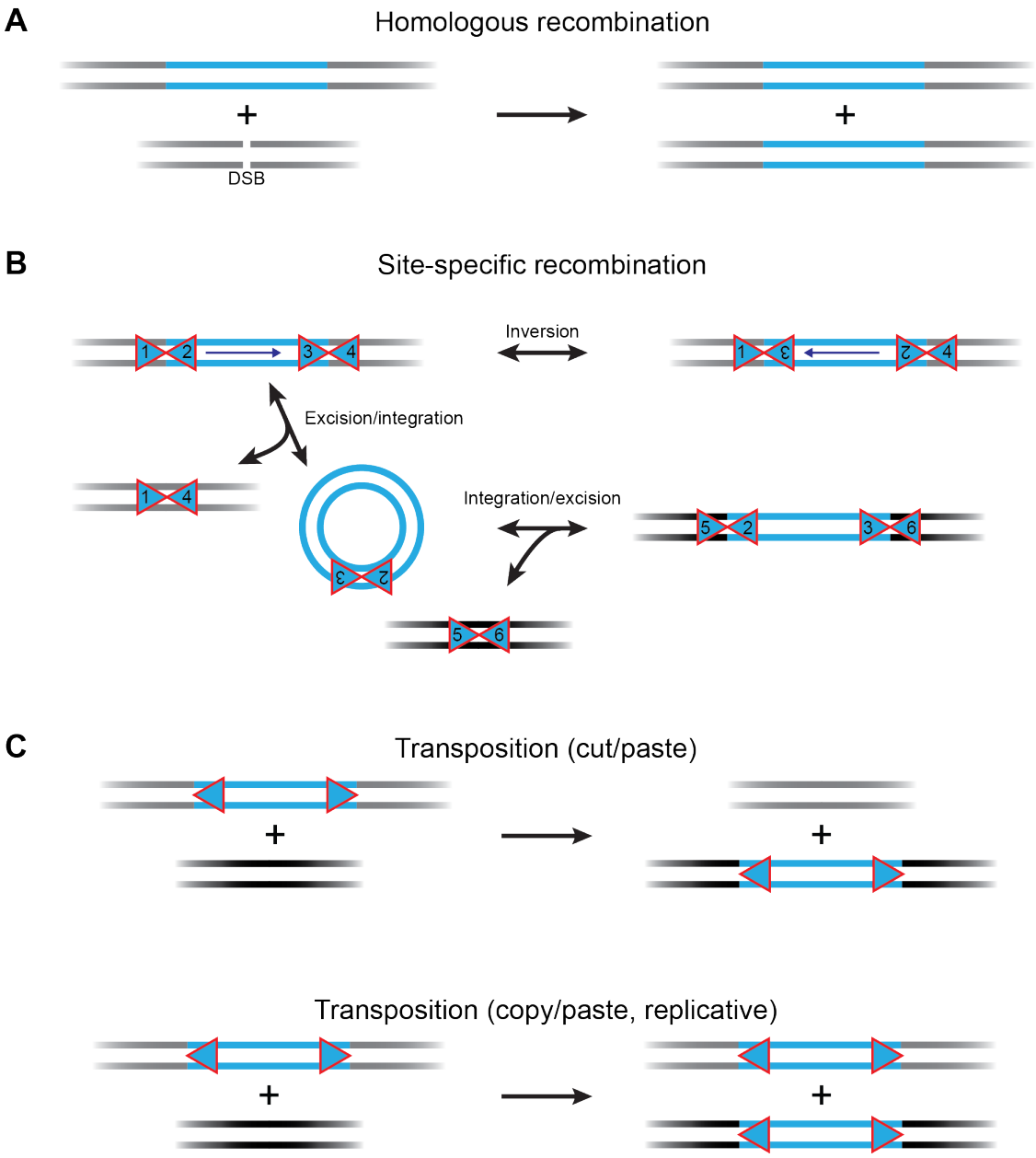

Figure 1.1: Recombination events.
These are the broad classes of events that can change the connectivity or topology of genomic DNA. The DNA undergoing recombination is in blue, with flanking sequences in grey. Red outlined triangles represent specific recombinase or transposase protein binding sites. DNA that is separate from, and non-homologous to, the flanking DNA is in black.
A. Homologous recombination transfers intervening sequences between flanking DNAs of homologous sequence. This is often triggered by the repair of a double strand break (DSB).
B. Site-specific recombination occurs between two sites which are both flanked by recombinase binding sites. This can result in inversion, or integration/excision.

(**Figure 1.1**, continued)
C. Transposition moves or copies segments of DNA between two sites. The ends of the transposon are precisely defined by protein binding sites or other features, but for some transposons the destination is not.

1.2 Families of transposases

In general, one of the more obvious distinctions between transposons is whether the transposon DNA is moved via excision and integration (as in a cut and paste operation), or duplicates itself into its new destination ("replicative" transposition) (Figure 1.1C). Transposons are primarily categorized by their transposases, which dictate the mechanism by which they transpose. There are five families of transposases that have been relatively well characterized, whose names make reference to their catalytic mechanism: (1) DDE, (2) Tyrosine (Y), (3) Serine (S), (4) Tyrosine-HUH (Y2), and (5) Reverse Transcriptase/Endonuclease (RT/En)[2,5] (Figure 1.2). The DDE, Y, and RT/En groups include retrotransposons, which are distinguished by their use of transcription into an RNA intermediate followed by reverse transcription into DNA as a means of copying themselves prior to being inserted into their target site. A sixth family of transposases, the casposases (or Cas1 transposases), has recently been discovered[6]. They are relatives of the Cas1 spacer acquisition proteins from CRISPR systems, and although at least one casposase has been shown to be capable of integration *in vitro*[7], like their CRISPR relatives their mechanism has yet to be fully elucidated.

1.2.1 Tyrosine and serine transposases

The tyrosine (Y) and serine (S) transposases are a superset of the tyrosine and serine site-specific recombinases capable of carrying out transposition. They move their transposon by two

instances of the same strand exchange reaction: first, to excise the transposon into a circular intermediate, and second, to insert the circular transposon into a target site (Figure 1.2A). Given that they excise their DNA, Y- and S-transposition move via "cut and paste." During strand exchange, the transposase forms temporary phospho-tyrosine or phospho-serine linkages, respectively, between transposase protein subunits and each DNA strand involved, resulting in two double strand breaks with short overhangs ("sticky ends"). These phospho-protein intermediates are then reversed by the attack of a DNA strand from the other cleavage site, thus exchanging the connectivity of the DNA between the two sites. Strand exchange is performed sequentially (one strand at a time) by Y-transposases but is concerted in S-transposases.

Although this family includes rigidly site-specific recombinases, some Y- and S-transposases exhibit relaxed target site sequence specificity. This can be achieved by not requiring the presence of recombinase binding sites when selecting a target (i.e., transposase subunits that can bind non-specifically to target DNA and still synapse properly with other subunits that have specifically bound the transposon ends), and not requiring that the sticky ends generated at both recombination sites by cleavage be complementary[8-12]. There are also Y-retrotransposons, which encode a reverse transcriptase (RT) in addition to a Y-transposase[13]. Transcription of the transposon by the host organism followed by reverse transcription by the element-encoded RT results in a double stranded circular cDNA transposon intermediate just as in the Y-transposases, although the exact details of this process have not been fully explored[14]. The result of Y-retrotransposition is thus a copy, leaving the original template transposon intact and in-place.

1.2.2 Tyrosine-HUH transposases

The tyrosine-HUH transposases, also known as Y2 or rolling-circle (RC) transposases, also use a phospho-tyrosine intermediate to break and join the ends of their transposon.[15] Unlike the Y-transposases, however, these transposases act on and move only one strand of their transposon. Rather than recognizing specific double-stranded DNA (dsDNA) sequences to mark the ends of the transposon, many Y2-transposases bind to hairpin sequences that form when the transposon DNA is single stranded (ssDNA), e.g. as a result of the passage of host replication forks[16]. Transposase-mediated strand exchange can excise the single-stranded transposon while repairing the origin strand. Then the transposon can be inserted into ssDNA location (usually another replication fork) by another round of strand exchange. A variation on this process occurs in some Y2 transposons that utilize rolling-circle replication (Figure 1.2B)[17]. Rather than depending on a replication fork, these Y2-transposases form a phospho-tyrosine linkage with the 5' end of the transposon, revealing a 3' end from which replacement-strand synthesis can proceed to copy and displace one strand of the transposon. Strand exchange with another copy of the transposase at the 3' end of the transposon then frees it as an ssDNA circle. Then, a third Y2-transposase can bind to a target site, cleave it, and exchange strands with the freed single stranded transposon. Regardless of which of these two routes are taken, transposition is completed when the host replication machinery synthesizes the opposite strand of the new transposon copy.

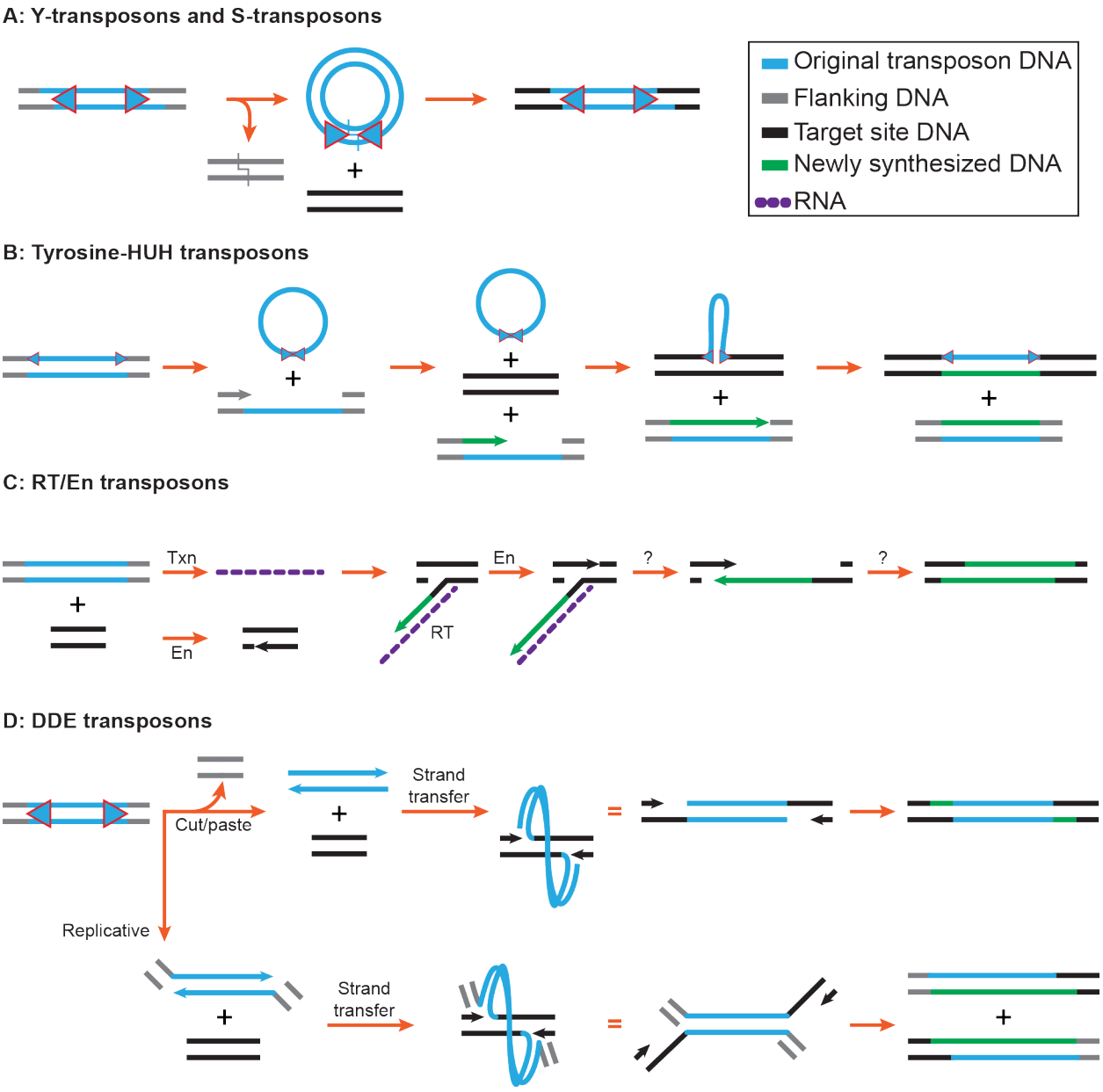

Figure 1.2: Mechanisms of transposition.
Any 3'-hydroxl groups generated during transposition are marked with an arrowhead.
A. Transposons encoding a tyrosine (Y) or serine (S) transposase are moved by strand exchange into a circular intermediate followed by strand exchange into the target site. In some Y-transposons, this circular intermediate results from transcription to RNA followed by reverse transcription to a circular cDNA.
B. Tyrosine-HUH transposases mobilize their element by strand exchange of only a single strand. Host replication is needed to generate a double stranded product and regenerate the original transposon.

(**Figure 1.2**, continued)
C. Some transposons move by combined reverse transcriptase (RT) and endonuclease (En) activities. An RNA intermediate serves as a template for reverse transcription of the transposon into its target site. **D.** DDE transposases can either move or replicate their transposon, depending on if one or both of the flanking DNA strands, respectively, at each end are cleaved. Both cases proceed through a strand transfer mechanism, the products of which are resolved by gap filling polymerases or full replication, respectively.

1.2.3 RT/En transposons

The RT/En family covers retrotransposons that move by the combined action of reverse transcriptase and endonuclease activities (Figure 1.2C), which may exist either in the same transposase polypeptide or as separate proteins. These transposons move via a single stranded RNA transcript of the element. The endonuclease activity is responsible for nicking the target DNA, revealing a 3'OH that can be used as a primer by the reverse transcriptase. RT/En elements often have stronger target site sequence preferences than the other families of transposons, except those that move by site-specific recombination. Target site preferences can be driven by sequence specificity of the endonuclease and/or the need for small patches of homology to anneal the transposon RNA with the target DNA primer. The transposon's reverse transcriptase uses the transposon RNA as a template, copying it into the target location. This process is known as Target Primed Reverse Transcription (TPRT)[18]. The remaining events are not clearly understood: The opposite strand of the target DNA must at some point be cleaved as well, generally downstream of the first site. Then, the new transposon cDNA must be used as a template for synthesis of its other strand, as well as duplicating any target DNA between the two cleavage sites[5]. How the cDNA strand is annealed to the target DNA (without guarantees that they are complimentary) so that second strand synthesis can occur, as well as the protein

responsible for second strand synthesis (which could be the RT itself, or a host factor) are not definitively known. Transposon family is noteworthy because it includes the LINE (*L1*) and SINE (*Alu*) retrotransposons, inactive versions of which constitute about 30% of the human genome[19].

1.2.4 DDE transposases

Finally, the DDE-transposases are possibly the most well understood transposase family. They have also been studied for the longest time – the first transposons described in the literature by Barbara McClintock in 1950[20], the Ac/Ds elements, move via a DDE-transposase[21]. This is also true of the first transposition system to be reconstituted *in vitro*, bacteriophage Mu [22,23], the subject of my work here. These transposases are named because their catalytic domains are all RNase H-like folds in which three amino acids with acidic side chains – usually two aspartates (DD) and one glutamate (E) – are positioned in the active site to coordinate essential divalent metal cofactors[24] (Figure 1.3). Their transposons have terminal inverted repeat(s) to position the DDE-transposase at the ends of the elements. During transposition, the transposase cleaves at least the 3' strand, if not both, at the ends of the transposon. If both strands at each end of the transposon are cleaved, the result is generally a cut/paste movement (also known as integration), while cleaving only the necessary strands will lead to replicative transposition. The resulting transposon 3'ends are then used in a transesterification / phosporyl transfer reaction to join them to the target DNA. This second step is referred to as strand transfer, in contrast to the strand exchange of the tyrosine and serine site specific recombinases and transposases. The joints are generally staggered, which generates single stranded gaps when the intervening sequence is

melted. During integration, these gaps are filled by host factors. During replicative transposition, replication proceeds from the target 3' ends though not only the gaps but also the entire transposon, regenerating the original transposon as well as a copy in the target location (Figure 1.2D).

DDE-transposases also mediate the transposition of some retrotransposons, as well as retroviral integration. These retroelements are generally flanked by much longer inverted terminal repeats sequences, and so the transposons are known as Long Terminal Repeat (LTR) transposons. Their mechanism is identical to that outlined for DDE cut-and-paste transposons above, except that they are first copied out as RNA by transcription and converted into double stranded DNA by element-encoded reverse transcriptase[25,26].

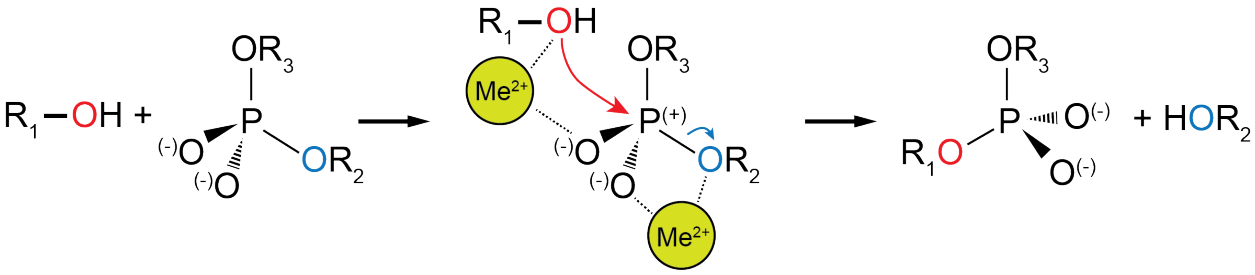

Figure 1.3: DDE active site metal catalysis of phosphoryl transfer.
DDE transposase catalytic domains have RNAse H folds that bind divalent metals. Divalent metals (Me^{2+}) can both help position the attacking hydroxyl group nucleophile and increase the electrophilicity of the phosphate atom.

1.3 Bacteriophage Mu and its transposition system

The transposition system from bacteriophage Mu has served as a critical model system for over five decades[22,27] and will be the subject of the remainder of this work. Bacteriophage

Mu first appeared in scientific literature in 1963 [28], and was so-named because its high activity made it as effective at introducing mutations in *E. coli* as the chemical mutagens in use at the time. Its high activity combined with the fact that its natural host was the highly domesticable *E. coli* made it an ideal model system to begin to understand transposition as we know it today: notably, Mu established that transposons have defined boundaries which allow them to excise while leaving the host DNA sequence intact[29], and that transposition can be replicative rather than only cut-and-paste[30]. Perhaps most notably, observations made with Mu were instrumental for the formulation of the Shapiro model of transposition, which correctly deduced the minimal DNA cleavage and joining reactions used by DDE-transposases[31].

The MuA protein is a DDE-transposase[23,32] and is the transposase responsible for the initial integration and subsequent replicative transposition of the phage Mu genome[22,33,34]. This genome is approximately 36.7 kilobases (kb) in length[30,35]. When packaged into the phage and upon infection/injection, the genome is flanked by excess sequences derived from the (previous) host. Although technically linear at this stage, before integration the genome behaves as a closed supercoiled circle because the phage protein N protects the linear ends and holds them together[36,37]. This is the substrate that MuA integrates into the *E. coli* genome to form the prophage. The initial integration event is not replicative – the infecting DNA molecule itself is inserted into the host genome and the excess flanking sequences are removed[38]. Subsequent transposition events, however, are replicative and are the route by which the phage genome is duplicated for the production of new phage particles during lysis. Having integrated into the host genome, the substrate for replicative transposition is also a closed supercoiled circle, but now the flanking host is not removed. The MuA protein catalyzes both of these events, as the core

chemistry of both is the same (discussed below). Impressively, the activity and efficiency of the Mu transposition system is so high that, by the time the lytic phase is complete, more than half of the DNA recoverable from the cell is Mu, rather than host, sequence[39].

1.4 Chemical mechanism of replicative transposition

The chemistry of replicative transposition catalyzed by MuA and related DDE-transposases involves two successive reactions mediated by the same metal-binding active sites. First, the transposase catalyzes the cleavage of the 3'-ended strand (hereafter the "transferred strands") at each transposon end. The nucleophile for this cleavage is a water molecule, and it results in free 3'-hydroxl (3'OH) groups at each end of the transposon. These 3'OH groups then become the nucleophiles for cleavage of the target DNA, joining the transferred strands to the target DNA and displacing new 3'OH groups in the target DNA. This is known as the strand transfer reaction. By monitoring the chirality of synthetic phosphorothioate target substrates, it has been shown in the Mu system that both the hydrolytic cleavage and strand transfer reactions occur as single nucleophilic substitution (S_N2) reactions, and this discovery has held true throughout the known DDE-transposases and integrases[40,41]. This is in contrast to, e.g., the chemical mechanism of site-specific recombinases that involve two steps and an intermediate where the protein and DNA are covalently linked. For the two chemical steps to proceed, the MuA active site must be occupied by a divalent metal. The relevant *in vivo* metal is probably Mg^{2+}, which yields the highest activity overall, but cleavage will proceed *in vitro* using almost any divalent metal except Ca^{2+} (Mg^{2+}, Mn^{2+}, Zn^{2+}, Co^{2+}), and strand transfer proceeds (to

varying degrees) in the presence all divalent metals tested thus far in the literature[42-44]. MuA residues D269, D294, and E392 comprise the DDE motif responsible for metal binding[23].

After these two reactions, the transposase itself has completed its task and other proteins are responsible for resolving the branched products into contiguous dsDNA, which will be discussed for Mu below. The transposase remains bound to its products until it is degraded. From the strictest definition, then, MuA is technically not a true enzyme or catalyst because it does not turn over. Instead it is more appropriately considered another reactant, just like the DNA it acts on. In addition, the manner in which the products are resolved determines whether the outcome will be integration or replicative transposition – nothing in the chemistry performed by the transposase specifies which will occur. This feature has permitted Mu to be a model system not only for replicative transposition, but also for the integration of some phages and all retroviruses.

1.5 The architecture of MuA transposase and the transposable Mu genome

MuA is a 75 kDa protein whose full sequence includes 663 amino acids. It contains seven domains (Figure 1.4) whose borders have been determined by limited proteolysis[45]. Beginning at the N-terminus, the first three domains are all helix-turn-helix DNA binding domains referred to as, in order, Iα, Iβ, and Iγ. Iβ and Iγ work together to recognize the MuA binding sites that mark the ends of the Mu genome. Iα has binding sites in what is known as the enhancer, a region further into the interior of the Mu genome that will be discussed below. Individual atomic structures of these domains have been solved by NMR[46-48]. MuA domain IIα is the RNAse H catalytic domain that chelates the divalent metals necessary for catalysis. The remaining three domains – IIβ, IIIα, and IIIβ – contain nonspecific DNA binding activity, the purpose of which

will become clearer upon discussion of the transposase crystal structure. An atomic level crystal structure of the MuA catalytic domain (IIα + IIβ) is available[32], and the structure of domain IIIα was captured in the transpososome crystal structure[49]. Domain IIIβis thus the only remaining domain for which no structural information is available. It contains sites of interaction for both the ClpX[50] and MuB proteins[51], which will be discussed in later sections. In the absence of DNA, MuA is resistant to inter-peptide chemical crosslinking, indicating that it exists as a monomer prior to engaging DNA and partaking in the assembly process detailed in the following section.

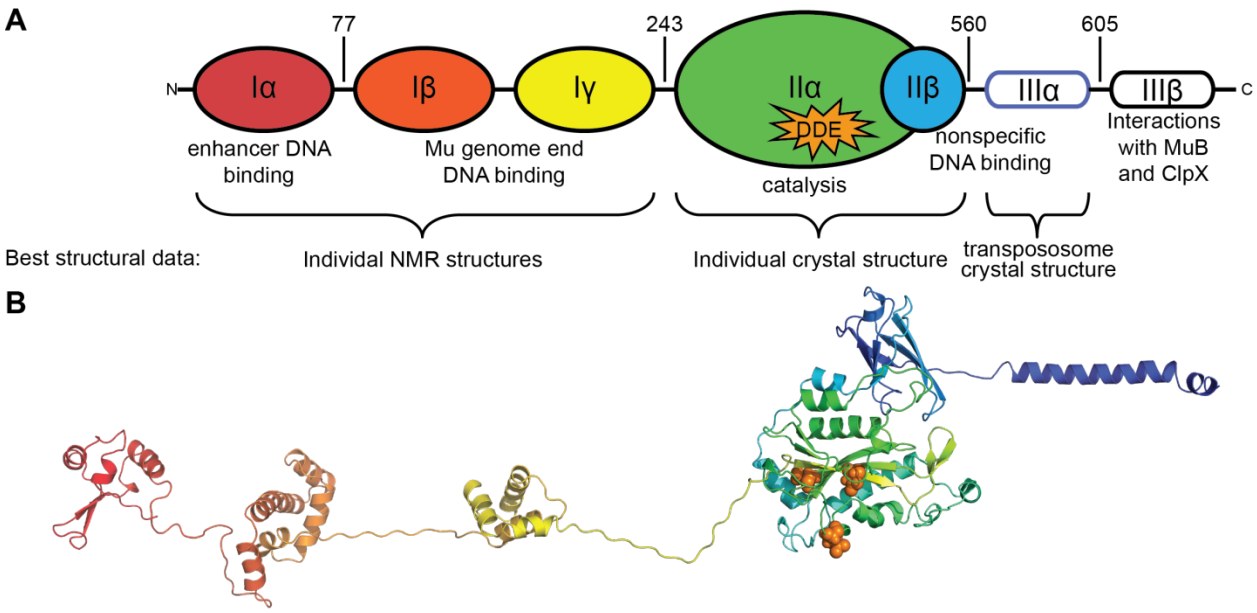

Figure 1.4: MuA protein domain layout and structures.
A. Schematic of the domains in the MuA transposase peptide. Domains are labeled with their name and function. Some domain boundaries often used in the literature are noted.
B. A model for a MuA monomer created from the available high-resolution structural information. Linkers are drawn as extended peptide chains. Colors and positions correspond to the schematic in **A.** The sidechains of the DDE catalytic motif are shown as orange spheres. Note that there is no structural information available for domain IIIβ.

Generated from Protein Data Bank (PBD) accession numbers 1tns, 2ezk, 2ezh, 1bco, and 4fcy (left to right).

The architecture of the Mu genome is also relevant for discussion of transposition (Figure 1.5). Like almost all transposable elements, it contains terminal inverted repeats that are binding sites for its transposase. The Mu genome in particular contains three binding sites for MuA at each end, L1-3 and R1-3 [34]. The copies of MuA that bind to the outermost sites (L1 and R1) will be the copies that actually perform the chemical steps of transposition. The remaining binding sites and features have structural roles that will be discussed in the next section. These sites are neither perfect inverted repeats or symmetric at each end: The series of binding sites at the left end is interrupted by a high-affinity site for the DNA bending protein HU[52], and the L2 site does not bind domain Iγ[53]. At the right end, the R3 binding site is flipped and so faces towards the inside of the genome. An enhancer site, so-called because it enhances transposition efficiency over 100-fold [54,55], lies about 1 kb inside of the left end of the Mu genome[56]. The enhancer contains three binding sites for MuA domain Iα [54,57] as well as a binding site for another DNA bending protein, IHF[58]. Finally, a high affinity gyrase site (strong gyrase site, SGS) lies in approximately the center of the Mu genome[59]. Its role in transposition will be discussed in the following section.

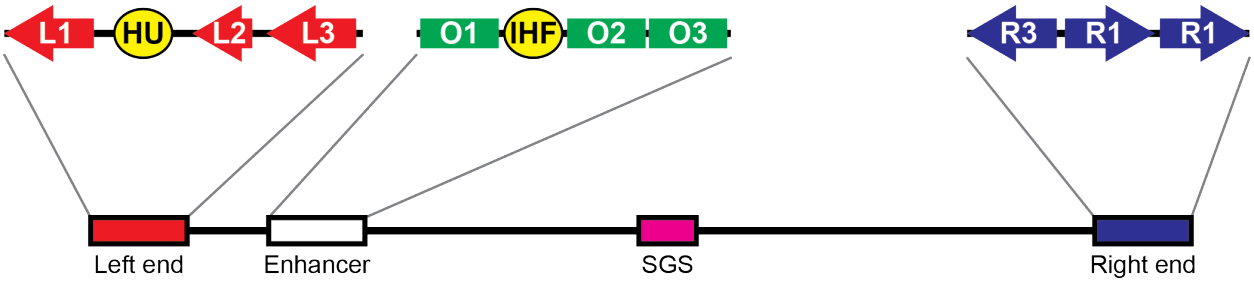

Figure 1.5: Mu genome features relevant to transposition**.**
The Mu genome ends contain three binding sites each for MuA (L1-3, R1-3), although the ends are not symmetrical. An enhancer region about 1 kb inside the left end contains three binding sites for domain Iα (O1-3), which assist in templating transpososome assembly. The DNA

(**Figure 1.5**, continued) bending protein HU binds in the Mu left end, and the DNA bending protein IHF binds in the enhancer. The Mu genome also includes a central strong gyrase site (SGS).

1.6 The Mu transposition pathway

1.6.1 Transpososome assembly

The complex of multiple MuA subunits and Mu DNA is known as the transpososome. This is the form of MuA that is active to perform chemistry on the transposon ends. The minimal transpososome that is competent to catalyze cleavage and strand transfer contains both Mu ends held together by a tetramer of MuA subunits, of which only three are tightly bound to DNA at the L1, R1, and R2 sites[60,61]. However, to get to this point requires the action of MuA subunits at all binding sites. The multiple MuA binding sites at each transposon end and the enhancer are all involved in this complicated assembly process. The purpose of this process is likely to ensure that only a legitimate pair of left and right ends can be the substrates for transposition.

Although they are separated by many kilobases, the SGS site maintains the Mu prophage as a cohesive supercoiled domain, which probably assists in increasing the chance that the left end and enhancer encounter the right end[62,63]. Assembly begins with MuA subunits bound to the R1 and R3 sites (via their Iβ and Iγ domains) engaging the enhancer (via the Iα domains) to trap the two DNA regions together (Figure 1.6, step 1)[64]. This complex can then go on to capture the left end and associated MuA subunits, as long as HU is present (Figure 1.6, step 2). This LER complex has a specific layout that traps five negative supercoiling nodes, and this layout is specified by the enhancer and relative orientation of all the MuA binding sites[65]. Once the MuA

subunit bound to the L1 position is properly in place, the transpososome becomes a very stable complex called the stable synaptic complex (SSC)[66] (Figure 1.6, step 3). As noted above, the SSC can be challenged by high temperature or salt concentration *in vitro*, which removes the L3- and R3-bound subunits entirely and disrupts the DNA-binding of the (formerly) L2-bound subunit. This pared-down tetrameric SSC is nevertheless still competent for Mu end cleavage and strand transfer.

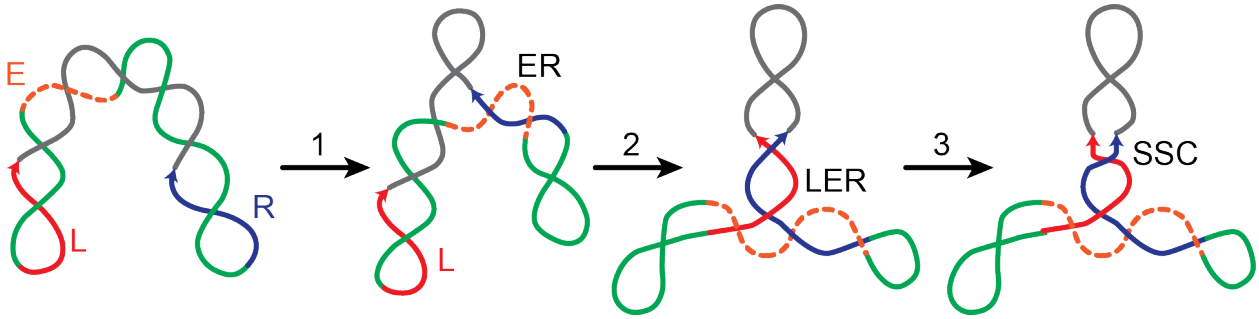

Figure 1.6: Transpososome assembly pathway.
This sequence of events is required for transpososome assembly on the full Mu genome *in vivo*. **(1)** MuA proteins bound to the right end (R, blue) can capture the enhancer (E, orange) via their domain Iα binding to operator sites. **(2)** The ER complex is competent to capture the left end (L, red) as long as that end is in the right conformation due to HU binding. The LER complex has a specific topology that traps 5 supercoiling nodes. **(3)** The LER complex can initiate transposition if both Mu ends can be melted and engaged in transposase active sites, forming a highly stable transpososome (stable synaptic complex, SSC).

1.6.2 Transposition reactions

The transition to the SSC also requires that the R1- and L1-bound MuA subunits be able to melt the DNA around the cleavage sites and engage that DNA in the active site[67,68]. This boundary is marked by a terminal CA dinucleotide sequence on the transferred strand. This CA dinucleotide is conserved across many transposons and retroviruses[69,70], probably because it is

particularly easy to melt[71,72]. Indeed, after the LER is assembled, the rate-limiting step for transposition appears to be melting the region at the ends of the Mu genome[44]. Mutations of the terminal dinucleotide to increase its melting temperature, withholding divalent cations, or mutation of the MuA active site metal binding residues all block the transition from LER to SSC[43,73]. The R1 and L1 MuA subunits engage and catalyze both cleavage and strand transfer reactions in *trans*, which means that the R1 subunit catalytic domain is engaged with the left end DNA, and vice versa[74].

Once the 3'OH group of the terminal adenosine has been released by hydrolysis (Figure 1.6, step 3), the transpososome is called the cleaved donor complex (CDC) and can perform strand transfer into target DNA. The target DNA phosphates that are attacked are 5 bp apart, such that the 5 nt just outside the genome ends on each transferred strand are complementary. Capture of target DNA can occur at any point once the SSC has formed, indicating that this is the point at which the target DNA binding surface assembles[75]. Here too, as in transpososome formation, the reactions at both Mu ends are highly coupled. In the vast majority of strand transfer products, both Mu ends have undergone strand transfer. Furthermore, if strand transfer at only one end is blocked, strand transfer at the functioning end is rapidly reversed[76]. Completion of strand transfer marks the final transition of the transpososome into the strand transfer complex (STC). The STC is especially remarkable for its stability: *in vitro* it can survive incubations at 75°C or in 6M urea[77]. *In vivo*, this stability allows the STC to remain bound to the branched strand transfer products, protecting them until the transpososome is disassembled.

1.6.3 Disassembly and replication

The final steps of Mu transposition disassemble the transpososome and resolve the branched DNA products back into contiguous dsDNA (Figure 1.7A). Disassembly is carried out by the host protein ClpX[50], which is a member of the heat shock protein family of AAA+ ATPase chaperones that convert the energy of ATP hydrolysis into unfolding polypeptides. ClpX forcefully unfolds one of the catalytic subunits (bound to L1 or R1) to destabilize the STC[78]. ClpX is recruited to the transpososome by interacting with multiple peptide sequences in the C-terminal domain IIIβ of MuA[79,80]. MuA monomers resist unfolding by ClpX prior to transpososome formation because ClpX can only associate through the avidity provided by the four or more MuA subunits[81]. Presumably, the STC is a particularly favored substrate because the conformation of domain IIIβ from the catalytic (L1 and R1) subunits changes as a result of target DNA binding.

Losing the L1- or R1-bound subunits destabilizes the transpososome and leaves the remainder of the MuA subunits vulnerable to being removed by another host factor that has yet to be identified. The host replication restart machinery, including PriA, is then recruited to the strand transfer products, possibly by the unknown host factor, or possibly because the products themselves resemble a failed replication fork. The replication restart machinery in turn recruits the DnaB replicative helicase and thus the DNA PolIII holoenzyme replisome[82]. Replication proceeding from the 3'OH ends in the target DNA through the Mu genome copies it into the midst of the target DNA. Like many transposons, Mu generates target site duplications: both copies become flanked by the 5 bp site around which strand transfer occurred. This completes replicative transposition.

As mentioned above, the initial generation of a prophage after Mu infection is an integration, rather than replicative, transposition event. This difference comes about as a result of how the strand transfer products are resolved after destabilization by ClpX, but is possibly one of the least well understand processes in Mu transposition. It must be that, somehow and at some point, the flanking DNA (which, in the case of integration following initial infection, is *not* the current hosts genome) is removed by cleaving the non-transferred strand, and the 5 nt gaps left by the staggered sites of strand transfer are filled by a host polymerase (Figure 1.7B). There is evidence that the removal of the flanking DNA is linked to a cryptic nuclease activity in domain IIIα [83,84], but how this nuclease activity is triggered only in this specific instance is unclear. A genetic screen for mutations that prevent successful integration implicated PriA and homologous recombination rather than the more straightforward gap filling polymerase PolI as the necessary components for filling the 5 nt gaps[85,86]. From this it appears that a more complicated set of steps is required to resolve strand transfer into integration products[87].

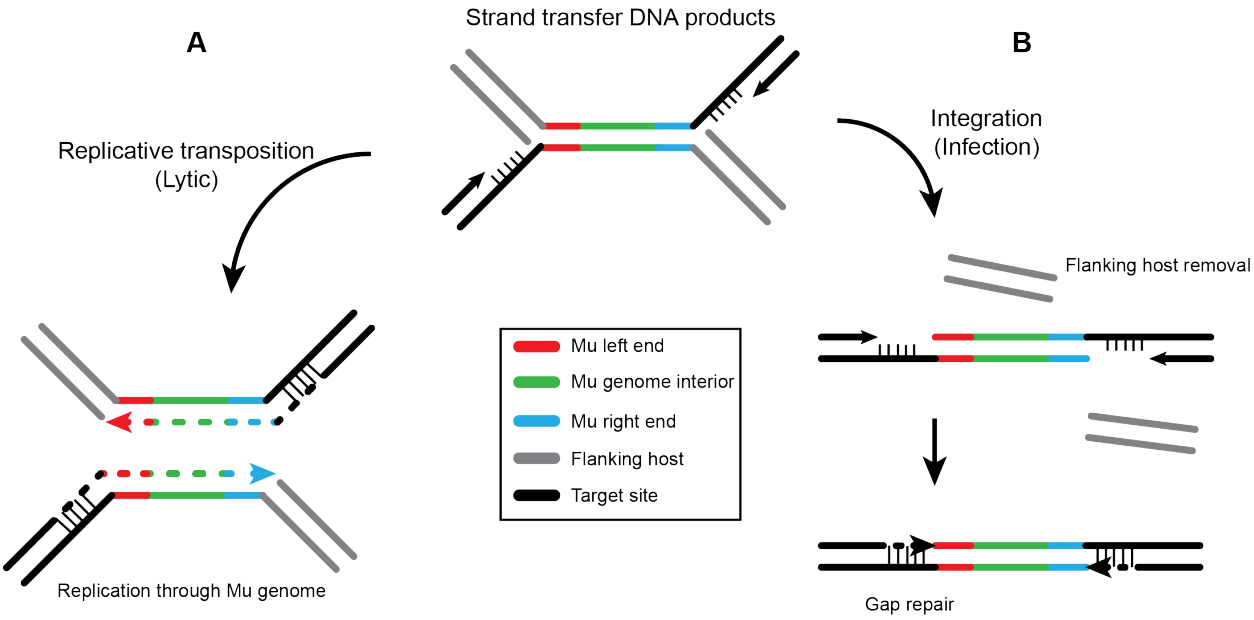

Figure 1.7: Two pathways for resolution of Mu strand transfer products.
A. In replicative transposition, replication proceeds through the Mu genome from each target 3' end. This results in duplication of the Mu genome, as well as the central target site pentamer.
B. Integration occurs when the flanking host is removed by severing the non-transferred strands. The target pentamer gaps are repaired without replicating the Mu genome itself, although the target site is still duplicated. The exact choreography of these integration steps is not known.

1.7 MuB: target selection and genome immunity

Mu has no strict sequence requirements for target site selection and does not display an orientation bias relative to its host *E. coli* genome. However, in the absence of other factors to drive target site selection, *in vivo* transposition sites do show a preferred central motif: 5'-C-Y-(G/C)-R-G-3' (where R is a purine base and Y is a pyrimidine base)[88]. Note that this motif is 5 bp long, which is the size of the strand transfer site stagger / target site duplication generated by Mu, and is two-fold symmetric around the center base pair. Moving father outward, insertion sites are observed to be weakly symmetrical for about 9 to 10 bp on either side the central base

pair, implying (correctly, given the transpososome crystal structure below) that the target is bound by symmetric protein contacts. The base pair steps in the preferred central pentamer correspond those that should be easy to deform – in particular, it avoids 5'-purine-pyrimidine-3' dinucleotides, which are the most stably stacked[71,72]. Thus, rather than specific sequences, Mu target selection is driven by the plasticity of potential target DNAs. The transpososome crystal structure, described below, gives a clear structural basis for the preference: the target DNA is bent by the transpososome during the transposition reaction.

The experiments that identified these target site preferences were performed in the absence of the phage-encoded protein MuB. MuB is an AAA+ ATPase [89] that, in the presence of ATP, greatly accelerates transposition [51,90] and is the primary driver of target site selection *in vivo*[89,91]. The ATP-bound form of MuB behaves like many DNA-binding AAA+ ATPases, such as the DnaA replication initiator[92] or DnaC helicase loader[93]: MuB polymerizes as a helical filament around DNA (Figure 1.8), although it will also form "empty" filaments if no DNA is available[89]. Its DNA-binding activity is non-specific, with only a slight preference for A/T-rich sequences[94,95]. Like many similar AAA+ ATPases, upon hydrolyzing its bound DNA MuB depolymerizes and disengages from DNA[96]. MuB ATPase activity is stimulated by interacting with MuA, and this interaction has been mapped to MuA domain IIIβ[51]. In addition to the two domains that comprise the AAA+ fold, MuB also has an N-terminal domain whose isolated structure is known by NMR. It follows the helix-turn-helix motif, but exhibits little to no DNA binding activity *in vitro*. Instead this domain facilitates MuB aggregation and filament bundling, behaviors which have historically hampered *in vitro* biochemical and structural assays of MuB function[97].

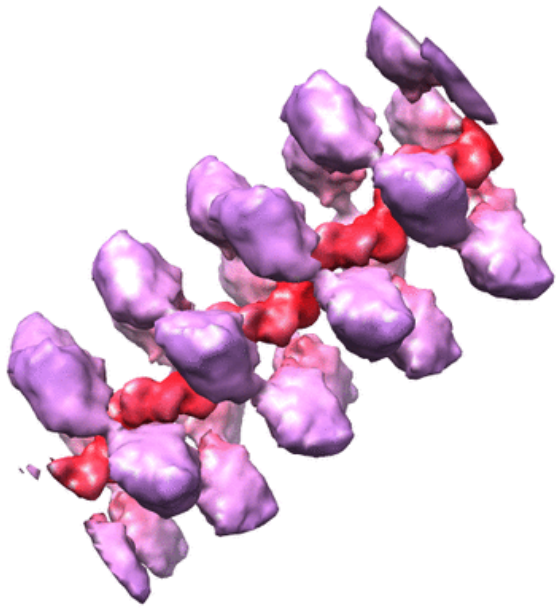

Figure 1.8: Cryo-electron microscopy resconstruction of a MuB filament. This structure is formed by the AAA+ ATPase domains of MuB in the presence of ATP and DNA. The red rod in the center of the filament is a helix with the proper diameter and pitch to be B-form DNA. The purple outer helical density corresponds to MuB.

This volume was segmented and rendered by the Electron Microscopy Data Bank (EMDB) from accession number EMD-2395.

The MuB target selection mechanism selects for target DNA that is distant from the ends of the Mu genome, or on another molecule or plasmid entirely[91]. Although at equilibrium MuB would be bound nonspecifically throughout the cell's genome in the absence of MuA, interaction with MuA monomers that are bound at the ends of the Mu genome but have not yet formed a transpososome clear MuB from DNA[95,98]. This clearance has been described as a "diffusion-ratchet," where DNA-bound MuA monomers encounter MuB by diffusion, bind temporarily, trigger MuB ATP hydrolysis and dissociation from DNA, and then can diffuse farther along to the next MuB copy[98,99]. There is a kinetic/temporal window for this process to occur because formation of a strand transfer-competent transpososome from individual DNA-bound MuA monomers is slow. By the time a transpososome has formed, its subunits will have already

cleared MuB from its vicinity, and so its own transposition-stimulating encounter with MuB will be on distant DNA.

The activity of MuB is not fully explored. In concert with simply mooring the transpososome to the selected DNA, MuB is thought to stimulate the catalytic activity of the transpososome[51]. The mechanism by which this occurs is not yet known. One potential answer lies in the peculiar helical pitch (48 Å) of MuB filaments: their pitch differs from that of B-form DNA (34 Å), but do they enforce their pitch on the DNA they encase. It has been suggested that MuB might transiently enforce its helical pitch on DNA during ATP hydrolysis. This could cause the unwinding of the encased DNA, which might in turn distort the naked DNA adjacent to the MuB filament[89]. Transpososomes could be particularly reactive to this distorted DNA, given their sequence preferences and my results in chapter 2 of this work. However, this activity of MuB has never been observed directly, and it is not clear if the tendency of MuB to bundle and aggregate would even permit transpososomes steric access to the DNA adjacent to a filament.

Finally, it is not clear how MuB (or other factors) generate the transposition immunity (the so called *cis* immunity) of the full Mu genome. Measurements *in vivo* have found that the diffusion ratchet described above is able to clear MuB from about 5 kb on either side of a Mu genome end[99]. While this explains how MuB selects distant targets, it is not enough to explain why the entire 37 kb Mu genome interior is very immune to transposition – the coverage of MuB removal would be insufficient. Chromatin immunoprecipitation with MuB indicates that, rather than being cleared from the Mu genome interior, MuB is particularly densely bound[100]. In this case, it may be that the Mu genome is immune because it is so fully encased in aggregated or

bundled MuB filaments. While the SGS-mediated highly supercoiled state of the Mu genome would probably assist, how this increased MuB density is mediated remains mysterious.

1.8 Mu transposition *in vitro*

Mu has been a powerful model system for decades in no small part because it was the first transposition system to be reconstituted *in vitro*. The first *in vitro* transposition system was published in 1983 by Kiyoshi Mizuuchi[22]. In this paper, Mizuuchi launched decades of discovery when he showed that a plasmid containing the first 3.4 kb of each Mu genome end, oriented relative to each other as in the normal phage genome, was a sufficient substrate for Mu transposition into a second target plasmid. Note that, fortuitously, this chosen size (3.4 kb) is enough to capture the enhancer region as part of the left end segment, although it had yet to be discovered. Transposition required only *E. coli* extracts from strains overexpressing MuA, although in the same publication Mizuuchi was able to observe the dramatic positive effects of also including MuB with ATP. This so-called "mini-Mu" system (Figure 1.9A) paved the way to dissecting the transposition reaction and its intermediates.

An even simpler system for transposition was revealed in a 1995 publication by Harri Savilahti, Phoebe Rice, and Kiyoshi Mizuuchi[44]. They showed that short DNA fragments derived from the phage right end containing only the R1 and R2 binding sites and at least 2 flanking base pairs was sufficient to recapitulate donor cleavage and strand transfer. Furthermore, if the short DNA fragments were designed to mimic the products of Mu end cleavage, transpososomes would assemble without any requirement for divalent metals. These substrates feature the transferred strand terminated at the 5'-CA-3' dinucleotide (thus being "pre-cleaved"), while

including 2-3 nt of overhanging flanking DNA on the non-transferred strand. The resulting transpososomes are competent for strand transfer upon addition of divalent metals, which allows assembly and strand transfer to be separable events *in vitro* (Figure 1.9B).

Addition of DMSO (traditionally at a concentration of 15% volume fraction) to the reaction buffer enhances transposition of both the mini-Mu and linear fragment systems[44,57,101]. In the former, it relaxes the requirement for the Mu ends to be in inverted repeat orientation and for the enhancer. Similarly, the linear fragment system both forms transpososome and performs strand transfer more efficiently in the presence of DMSO. In chapter 2, I explore this last phenomenon in more detail.

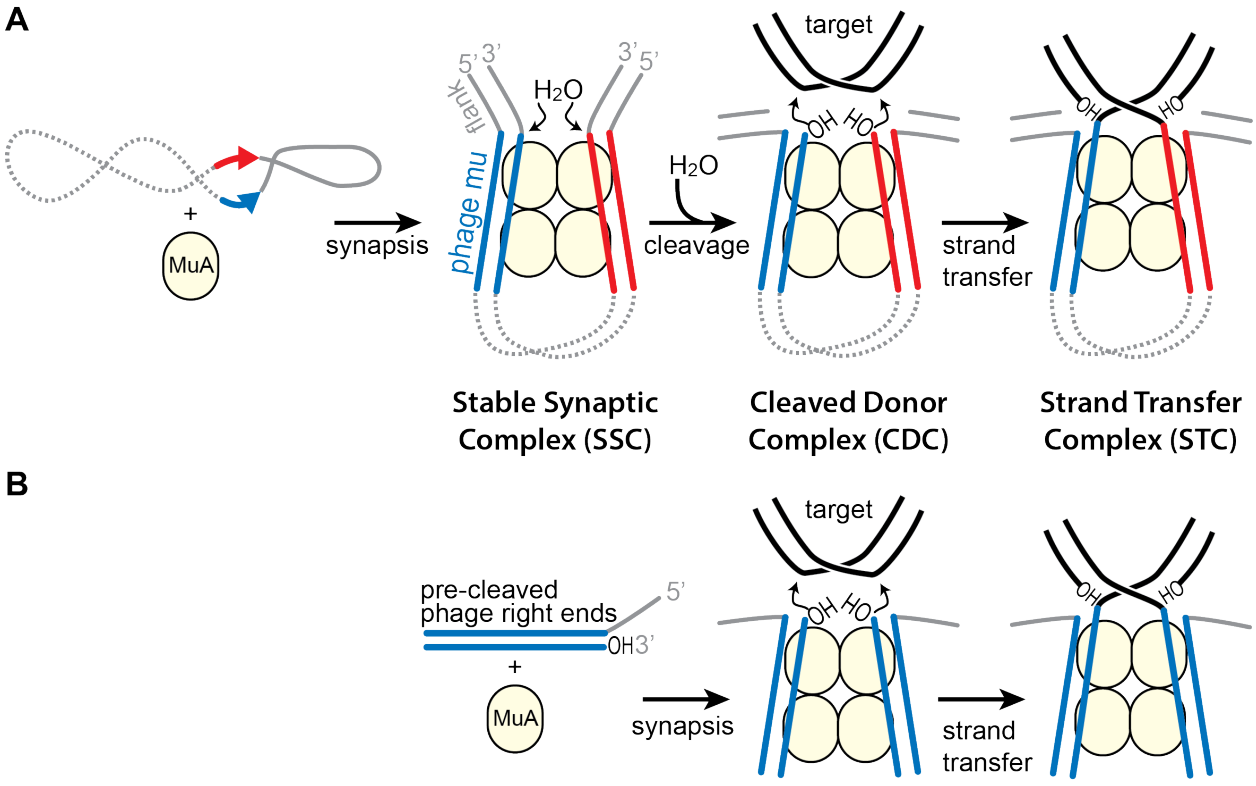

Figure 1.9: *In vitro* Mu transposition systems**.**

(**Figure 1.9**, continued)
A. Transposition *in vitro* can proceed using supercoiled plasmids with the left and right Mu ends (mini-Mu). The target DNA can be another plasmid or short DNA fragment. Although they are not strictly required, transposition is very inefficient unless both the enhancer and MuB + ATP are included along with MuA.
B. In the simplest *in* vitro transposition reactions, short linear DNA fragments mimicking cleaved Mu right ends are used as the donor DNA. The enhancer and MuB are not required here for efficient transposition. Activity in both **A.** and **B.** can be enhanced by addition of DMSO.

The MuA protein itself can also be simplified by truncation for *in vitro* work. The *in vitro* system described above does not require the enhancer element either in *cis* or *trans*. This means that the N-terminal DNA binding domain (Iα) is no longer necessary – indeed, if not just because it becomes a steric hindrance during tetramerization, in the absence of the enhancer this domain is mildly inhibitory – and can be removed. In addition, biochemical or structural dissection of MuA does not require its unfolding by ClpX or stimulation by MuB, and so domain IIIβ too can be removed from *in vitro* constructs. Domain IIIα, however, is necessary for transpososome assembly at the R2-bound subunit positions and must remain included[102]. The minimal fully active *in vitro* MuA construct has thus been defined as stretch from residues 77 to 605, which includes domains Iβ through IIIα (Figure 1.4A).

1.9 Monomer and Transpososome structures

The atomic structures of the MuA DNA binding domains and catalytic domain have permitted attempts to capture the structure of an entire MuA monomer and the assembled transpososome. This was first undertaken using cryo-electron microscopy by the group of George Chaconas (Yuan *et al.*).[103] Their work produced a structure of a full-length MuA monomer at a nominal 16 angstrom resolution, and a structure of the cleaved donor complex at

34 angstroms. They were able to dock the known structures of the individual MuA domains into their monomer density, and predicted a structure in which the three DNA binding domains dock closely with each other and with the catalytic domain. These domains filled most of the density they observed, and so they were only able to speculate as to the positions of domains IIIα and IIIβ.

For their STC structure, Yuan *et al.* used electron spectroscopic imaging (ESI) to locate the path of DNA though the density. This electron microscopy technique measures electrons scattered in-elastically by the phosphate atoms in the DNA backbone. Despite the challenging resolution, they were able to dock four copies of their monomer density in positions consistent with the approximate path of DNA and their MuA monomer model. This docking was undertaken without any rearrangement of the domains from the isolated monomer structure. The Mu end DNA they modeled is thus contorted so it can contact the relevant DNA binding domains. Using this model, they went on to guess that the location of the target DNA binding surface would be in the cleft of their V-shaped density, in agreement with their interpretation of where the donor DNA ends would terminate.

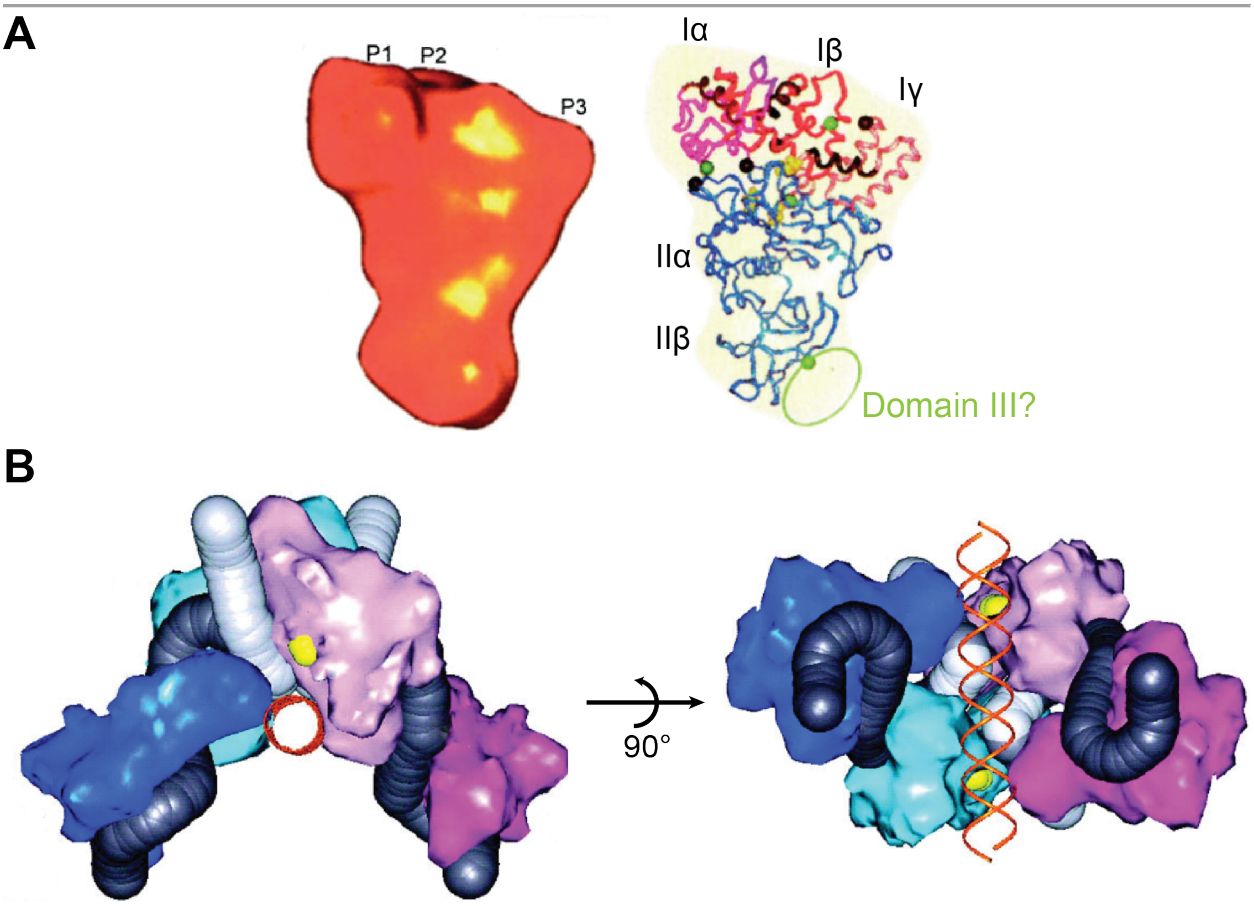

Figure 1.10: Cryo-electron microscropy structures of MuA and the CDC transpososome.
A. The MuA monomer cryo-EM density (left) exhibits three lobes (P1-3) docked atop a larger triangular region with another lobe at its bottom. It can be filled by docking high-resolution structures of the MuA DNA binding domains and catalytic domain (right). Domain III, although present, is difficult to locate. At the time of this analysis shown here, no structural information for domain III was available. The nominal resolution of the data used to dock the structures is 14 angstroms.
B. Four copies of the MuA monomer density from **A** can be docked into the 34 angstrom density generated from CDCs. A model for the path of Mu end DNA (black tube) and flanking host DNA (white tube) was generated from electron scattering data (see text). The position of the active site for the catalytic subunits is shown in yellow. Orange ribbon DNA is a model for where researchers guessed target DNA would bind.

Adapted from Yuan *et al.*[103].

Years later, our understanding of the Mu transpososome was revolutionized when

Sherwin Montano, Ying Pigli, and Phoebe Rice solved and published the crystal structure of the

stand transfer complex[49]. This revealed a picture of the transpososome that has served as the foundation for my work in this thesis. Although they exhibit similar V-shaped architecture, the details of the crystal structure contradict much of the cryo-EM structure. In the crystal structure, the Mu end DNA is relatively linear, and the MuA protein components are elongated to accommodate this. Any one MuA monomer is very extended. The catalytic domain of the R2-bound subunits extends forward to lay on the DNA binding domains of the upstream R1-bound subunit, an interface that I will discuss in detail in chapter 3. Similarly, the catalytic domain from the R1-bound subunit reaches forward in order to contact the junction between the transferred strand and the target DNA. Given that the nominal resolution of the crystal structure is 10-fold that of the cryo-EM structure, I consider that the arrangement of protein and DNA in the former is almost certainly the correct one.

The crystal structure confirms and explains many assertions made about the transpososome in the biochemical literature that preceded it. The significant level of intertwining and the amount of buried surface area between the four subunits, particularly between the two MuAs bound to the same Mu end, explain the impressive stability of the transpososome after assembly. The necessity of domain IIIα at the L2/R2 positions is clear: is it positioned to hold the two MuAs bound to the same Mu end together, and may reach across to also hold each half of the tetramer together. The catalytic domains from the R1-bound subunits are positioned for catalysis *in trans*, that is, they are positioned for catalysis on the opposite Mu end from the one they are bound to. Given this new structural certainty for the roles of the R1- and R2-bound subunits, throughout my work here I will use the term “catalytic subunits” to refer to MuA subunits bound to the R1 sites in *in vitro* transpososome tetramers, and “structural subunits” to

refer to MuA subunits at the R2 positions. Finally, unlike the Mu end DNA, the target DNA is very bent in an almost 180 degree U-turn. This corresponds nicely with the target site preference for deformable base pair steps discussed above.

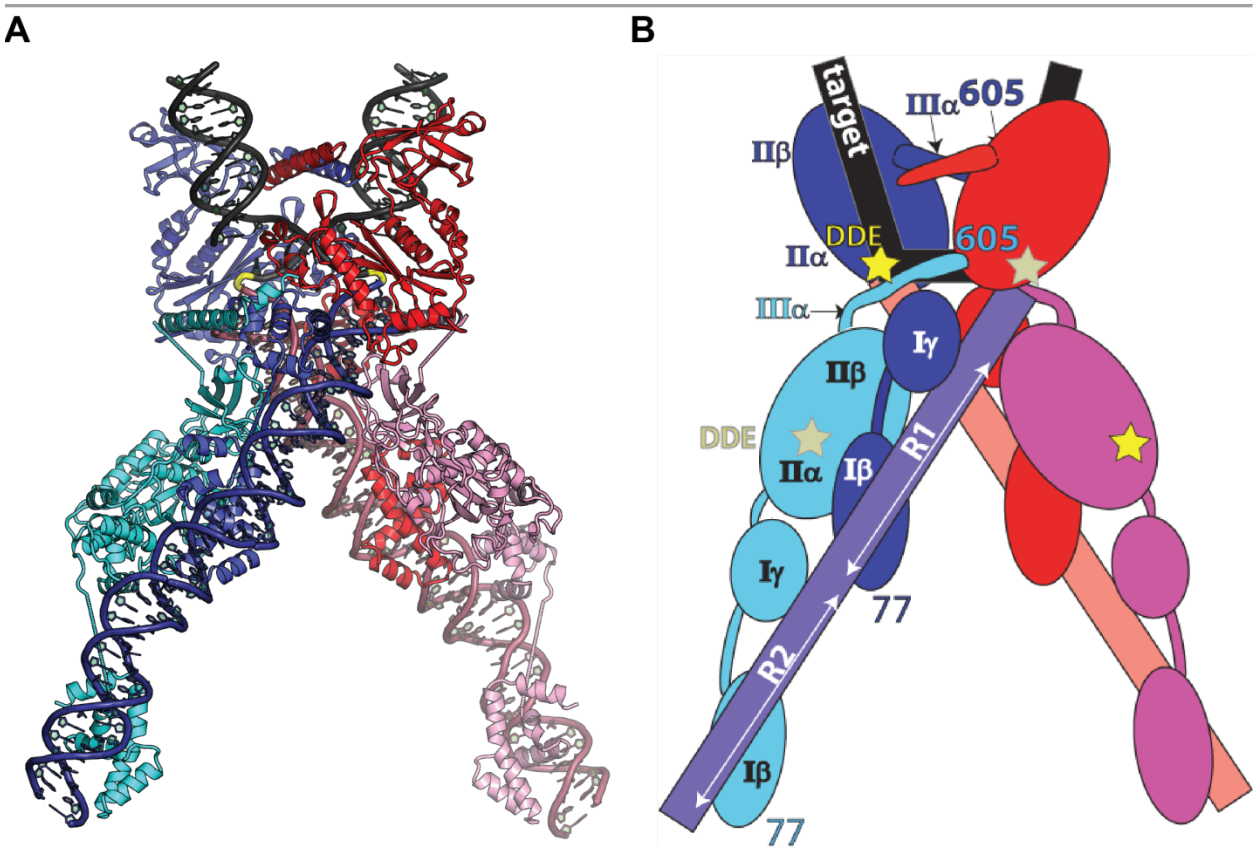

Figure 1.11: The crystal structure of the Mu strand transfer complex.
A. The Mu STC crystal structure. One Mu end and the two MuA subunits bound to it are in shades of blue, and the other in shades of red. The phosphodiester bond formed by strand transfer is highlighted in yellow. The target DNA is black. Generated from PDB ID 4fcy.
B. Schematic diagram of the structure from **A**., with corresponding colors. Adapted from [49].

In the following chapters, I use this crystal structure as a springboard to explore how Mu transposition is regulated. Although Mu has been studied *in vitro* for over three decades, having a structural foothold has granted me the opportunity to probe the system with precision that was previously unobtainable. In the next chapter, I address the question of why Mu and many of its

relatives have all converged on a bend in the target DNA. This question was of particular interest because of its broad application to DDE transposases and retroviral integrases. I propose a model where bending prevents premature strand transfer, allows transpososomes to be recruited to their targeting proteins (e.g., MuB), and prevents reversal of strand transfer once the transpososome has committed to a target. In chapter 3, I report my work towards a better understanding transpososome assembly. In particular, I address why such a thermodynamically stable entity like the transpososome is so slow to form, even in simplified *in vitro* transposition reactions. As part of this, I revisit the MuA monomer structure and investigate the activity of hyperactive MuA mutants. I conclude that isomerization, from two MuA subunits bound at adjacent DNA sites to the docked dimers seen in the crystal structure, is the limiting step to transpososome formation. These two chapters converge on a common theme: that regulation of transposition relies on conformational changes with sizeable energetic barriers, such that they only occur under the proper circumstances and, once committed, are unlikely to reverse.

Chapter 2

Target bending promotes careful transposition and prevents its reversal

The work described in this chapter was recently submitted to the journal 'Molecular Cell': Fuller, J.R.and Rice, P.A. "Target bending promotes careful transposition and prevents its reversal".

The SinMu chimeric transpososome system was engineered by Sherwin Montaño. The creation of a SinMu protein construct encoding the full MuA C-terminal domain, used in this chapter, was done in collaboration with Lorraine Ling, Robert T. Sauer, and Tania Baker from the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA. It first appeared in their publication in the 'Journal of Molecular Biology': "Deciphering the Roles of Multicomponent Recognition Signals by the AAA + Unfoldase ClpX." [81]

2.1 Summary

DDE transposases and retroviral integrases join their elements' ends into target site DNA by concerted strand transfer reactions. Once this step is completed, they remain bound to the products until disassembled by host factors. Four crystal structures of these strand transfer complexes have been solved to date, including that from bacteriophage Mu. In each of them, the target DNA is held in a highly bent conformation. This bend has been hypothesized to be part of the interaction between transposases and their target selection machinery and/or a strategy to resist reversal of the energetically uphill strand transfer chemistry. Using the Mu transposition system, we test these ideas directly *in vitro* by measuring how target capture and strand transfer in both directions are altered by DNAs of increased flexibility or mutations that reduce bending. Our results indicate that bending is a significant energetic barrier to target capture and strand transfer, thus inhibiting the interaction between transpososomes and target DNA. Strand transfer

reversal is not affected by the properties of the target DNA, but is accelerated when bending is compromised and coincides with unbending of the target DNA. Target DNA bending thus plays two roles in DNA transposition and integration: DNA that is flexible or contorted by other means is used preferentially as a target, and the post-strand transfer conformation of DNA (as stabilized by the transposase) prevents accidental catalysis of disintegration.

2.2 Introduction

Transposons are mobile DNA elements that move or copy their DNA sequence from one location to another. They have exhibited a remarkable ability to spread, such that sequences derived from transposons are pervasive in the genomes of prokaryotes and eukaryotes alike [104]. Among transposons whose behavior has been examined *in vitro,* the transposable *Escherichia coli* bacteriophage Mu is one of the most active[22] and well-studied[27]. Its transposase, MuA, belongs to the large DDE family of transposases [23,32]. Members of this family all include a catalytic domain with an RNase-H fold that utilizes divalent metals for catalysis. In addition to the transposases for many common transposons, the DDE family also includes retroviral integrases, which use the same reaction mechanism to integrate viral genomes into their hosts chromatin[24,26].

To catalyze transposition (Figure 2.1A), DDE recombinases like MuA bind specific sequences at each end of their element and synapse them together in a complex known as the transpososome (or, for retroviral integrases, the intasome)[77,105]. The transpososome then hydrolyzes the host-transposon boundary, and catalyzes the attack of the resulting 3' hydroxyl groups from each transposon end into the "target" destination DNA. This critical second

chemical step is referred to as strand transfer. For Mu specifically, the host-phage boundaries are nicked and strand transfer occurs at positions 5 bp apart in the target DNA[106]. Rather than turning over, the Mu transpososome remains bound to the branched strand transfer products until forcefully disassembled by the host ClpX chaperone [50], after which transposition can be completed by the host's own DNA repair and replication machinery.

Transposases and integrases face two challenges when interacting with target DNA sites. First, they must avoid selecting their own DNA as a target, because resolving intra-element strand transfer products results in deletion of segments of the element and/or a double strand break. To this end, many DDE transposases bind to a second DNA binding protein to select distant and non-self target sites. For Mu, this is the MuB protein encoded by the phage itself[89,91], whereas retroviral integrases interact with host nucleosomes to choose a target[107]. However, the mechanism(s) by which transpososomes resist the high local concentration of self-DNA but are activated to attack DNA bound by their targeting protein partner are not well understood.

Once a proper target site has been used for strand transfer the transposase must also avoid catalysis of the reverse reaction (or "disintegration") in the time before host machinery completes transposition. Strand transfer exchanges one pair of 3' hydroxyl groups and phosphodiester bonds for another and so should not result in the net release of chemical bond energy that could drive the reaction in the forward direction. Nevertheless, MuA displays a strong bias towards catalysis of only the forward strand transfer reaction[76,108,109]. It seems likely that this bias stems in some way from product binding energy, as transpososomes remain tightly bound to their products.

Here we show that target DNA bending contributes solutions to both of the above challenges. Four crystal structures of transpososomes and intasomes from the DDE family that include the target DNA are available, including that of the Mu transpososome[49,110-112] (Figure 2.1B). Beyond the shared catalytic domain fold, the specifics of the protein-protein and protein-DNA contacts in these structures are very different. Nevertheless, they have converged on a few high level features. For example, all four are poised for catalysis in *trans*, where a protein subunit bound to one DNA end catalyzes reactions on the opposite end. This helps ensure that chemistry does not happen before synapsis [74]. They have all also converged on a bend in the target DNA following strand transfer. Clues that target DNA bending can play a role in target site selection come from studies of the target site preference for Mu and retroviral integrases. In the absence of target selection partner proteins, target sites with more easily deformable sequence steps are preferred, suggesting that DNA flexibility is interrogated prior to strand transfer and can guide target site selection[88,107,113]. This is further supported by a structure of the prototype foamy virus (PFV) intasome bound-to, but not integrated into, target DNA, which shows the target DNA bent in a very similar overall conformation to that in the strand transfer complex[110]. It has also been suggested that target DNA bending could prevent the reversal of strand transfer by driving the products (the new target 3' hydroxyl group and transposon-host phosphodiester bond) out of the active sites to reduce the conformational strain of the bend[49,110]. Both ideas have yet to be tested directly *in vitro*.

In this work, we measure how target DNA flexibility and bending affect target DNA binding and strand transfer by the Mu transpososome. We show that increasing DNA flexibility has a dramatic positive effect on both, implying that bending occurs as part of both binding and

strand transfer and carries a steep energetic cost. Once completed, we show that reversal of strand transfer is very rare and occurs under extreme conditions that coincide with disruption of the bend. Further, a mutant transpososome with compromised target DNA binding can be rescued by pre-bent DNA and is particularly prone to strand transfer reversal. Our results are the first to biochemically link unbending to reversal, and point to target DNA bending as a key energetic barrier to strand transfer. This barrier would allow DNA deformation generated by other proteins to steer target site selection, and provide a way to channel product binding energy into preventing strand transfer reversal.

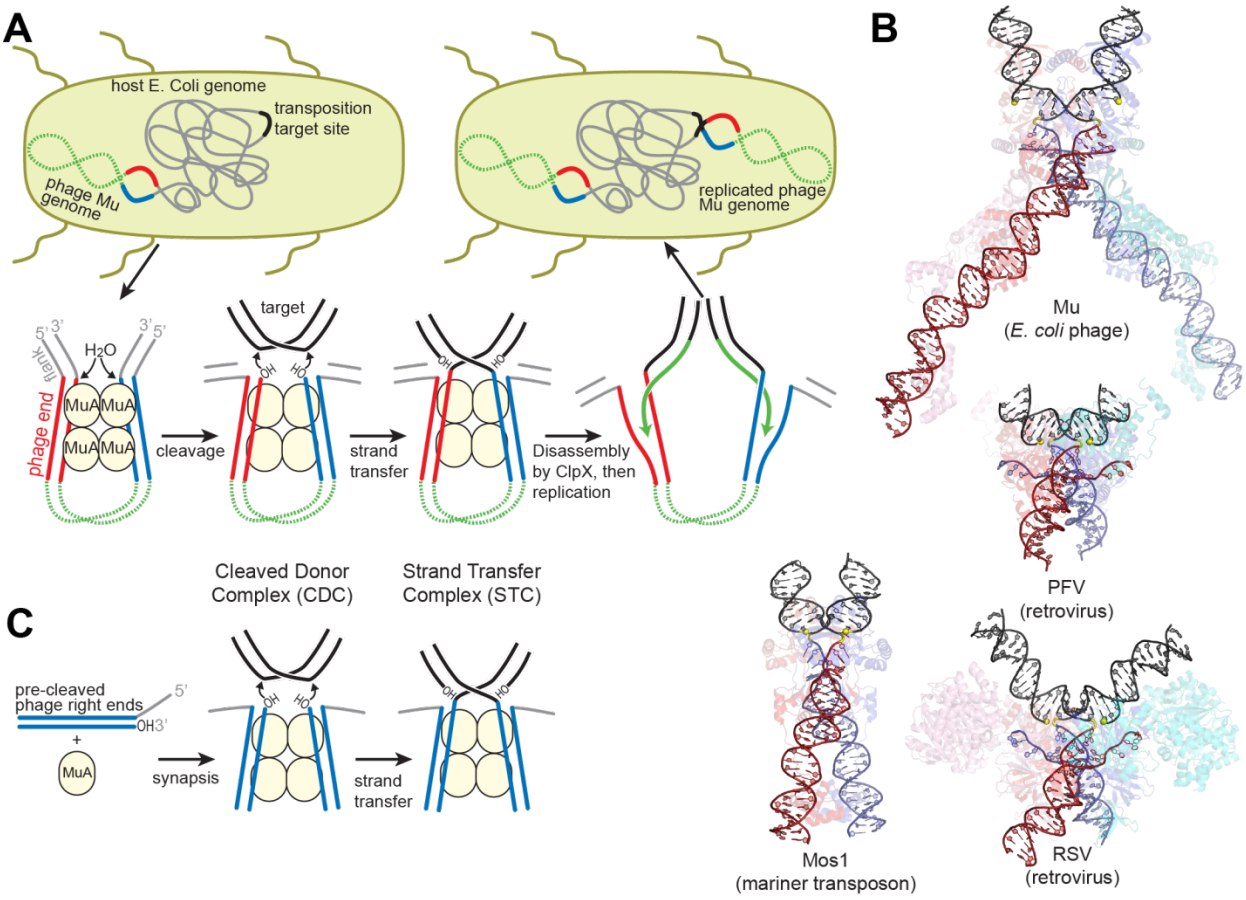

Figure 2.1: Replicative Transposition by Bacteriophage Mu and Available Structures of DDE family members.
A. Diagram of the replicative transposition process. The phage genome (green, dotted) ends (which carry transposase binding sites) are in red and blue. Transposase subunits (light yellow circles) synapse element ends and catalyze their nicking and joining to target DNA (black). The transpososome complex is then disassembled to permit DNA replication (green arrows) through the element, resulting in target site and transposon duplication.
B. Transpososome structures. DNA is colored as in A., with protein components colored according to their bound DNA. Generated from PDB IDs 4FCY (Mu), 3OS0 (PFV), 5HOO (Mos1), and 5EJK (RSV).
C. A diagram of the simplified *in vitro* Mu transposition system used in this study, which uses purified protein and short linear DNA fragments. Colors as in **A**.

2.3 Results

2.3.1 Flexible or bent DNA is a highly reactive transposition target

Mu is an attractive model system because transpososomes can be assembled *in vitro* from MuA protein and short DNA substrates[44] (Figure 2.1C). The sequence of these DNAs is taken from the phage genome right end and includes two MuA protein binding sites (Figure A1.1). The DNAs we use here mimic the products of phage-host junction cleavage, thus removing that earlier reaction step from our analysis. The assembled transpososome without target DNA is referred to as the cleaved-donor complex (CDC), and a transpososome that has bound and attacked target DNA as the strand transfer complex (STC).

We first sought to determine whether target DNA flexibility has an effect on target DNA binding and the kinetics of the strand transfer reaction. We use two methods to increase the flexibility of duplex DNA: DMSO added to the reaction buffer [114,115], and/or a single G:G base-pairing mismatch [116-118] incorporated at the center of the target DNA sequence. In addition to changing the biophysical properties of DNA, both have been used in previous studies as general enhancers the transposition activity of MuA *in vitro* [44,57,101]. Because previous reports suggest that DMSO might interact with earlier transpososome assembly steps, and to eliminate spare Mu ends or MuA protomers that could compete with our intended target DNAs, we first purified the CDC form of the transpososome by gel filtration chromatography. To prevent premature catalysis of strand transfer, CDCs were assembled and purified in a buffer containing EDTA and lacking Mg2+. A 2:1 mixture of MuA protein to Mu end DNA in this buffer results in a gel filtration peak corresponding to the molecular weight of the CDC (Figure 2.2A).

We monitored the effect of DNA flexibility on the kinetics of strand transfer by sampling reactions containing 100 nM each of purified CDC and 32P-labeled target DNA as a function of time. In these experiments, the target DNA strand is labeled at the 5' end and becomes cleaved as a result of strand transfer. As has been observed previously[119], a mismatched base pair directs the vast majority of strand transfer events to occur centered around it, hence the single dominant product (Figure 2.2B). The fully base paired target DNA, which is identical in sequence except for the central nucleotide in one strand, results in two major products and a number of minor products visible at very high contrast (Figure 2.3). It is likely that the major insertion site remains approximately centered even without the mismatched base pair because our 35 bp target DNA is not much larger than the total target DNA binding surface of the transpososome.

Increased target DNA flexibility via the G:G mismatch or DMSO significantly enhances the rate of strand transfer (Figure 2.2B,C). The unmodified reaction requires about 40 minutes to convert 30% of the target DNA to strand transfer product, reaching a peak slope of about 1 nM min^{-1}. In a reaction buffer containing 15% (v/v) DMSO, the reaction requires only 4 minutes to reach this same level and peaks at about 8 nM min^{-1}. The mismatch is even more powerful and lead to the consumption of 30% of the target DNA in about 1 minute (corresponding to a slope of 30 nM min^{-1}).

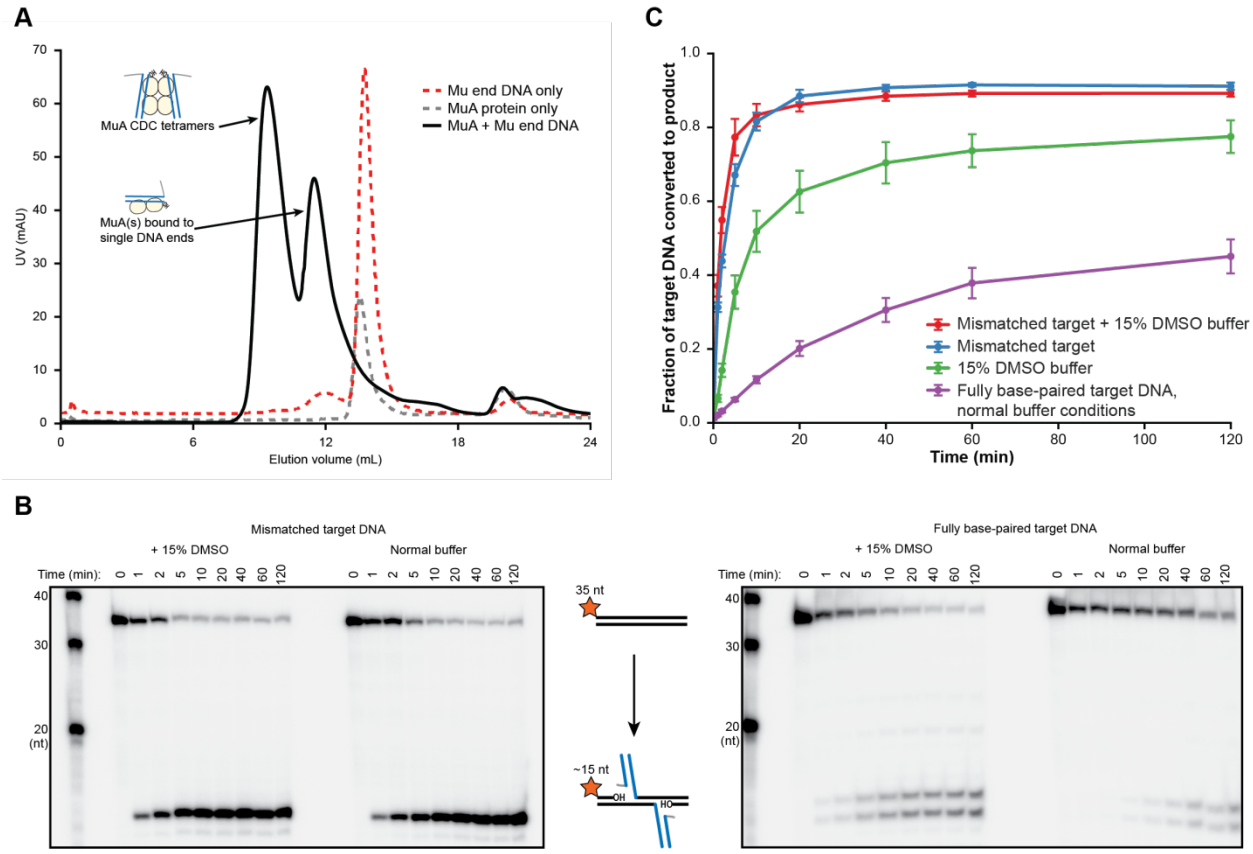

Figure 2.2: A Base-Pairing Mismatch and/or DMSO Accelerate Strand Transfer by Purified Mu Transpososomes.

A. Purification of transpososomes by gel filtration. CDC transpososome tetramers are separable from lower molecular weight complexes (black line). Also shown are samples of MuA protein monomers and phage end DNAs alone (dashed lines). The contents of only the CDC tetramer peak were captured and used for kinetics and binding assays.

B. Strand transfer kinetics visualized by denaturing gel electrophoresis and 5'-32P-labeled target DNA. Where indicated, the 35bp target DNA sequence includes a G:G base-pairing mismatch at the central (18th) position, and reaction buffer was supplemented with 15% (v/v) DMSO. Strand transfer results in cleavage of the labeled strand.

C. Quantification of the strand transfer kinetics experiments described in (B). Y-axis represents the fraction of total lane signal present in product band(s). Error bars represent mean ± the standard error of the mean (SEM), n = 4 independent time courses per condition. Solid lines simply connect points to guide the eye.

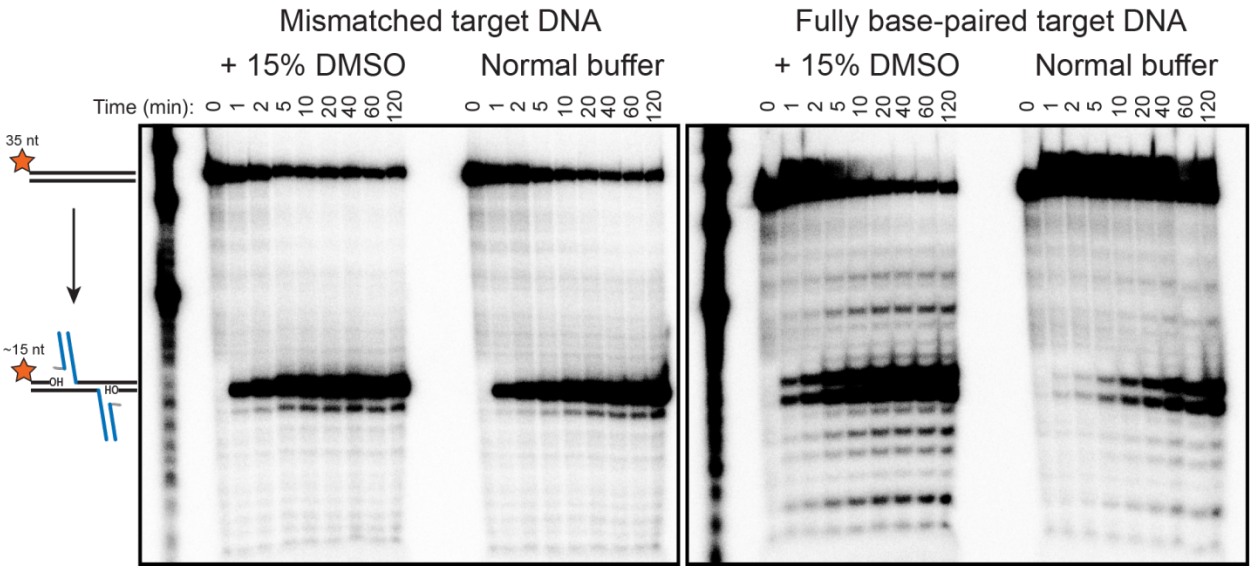

Figure 2.3: Mismatches target transposition.
Pictured are the same gels from **Figure 2.2**, but at very high (and matching) contrast. Target DNA with a base-pairing mismatch results in strand transfer site specificity, in addition to acceleration. On the left, nearly all strand transfer events have occurred centered relative to the mismatched base-pair products. A fully base-paired target DNA of otherwise the same sequence (right) gives strand transfer products of various sizes, resulting from the use of multiple sites along the DNA.

To determine whether enhanced the strand transfer rates result from tighter binding to the more flexible target DNA substrates, we used fluorescence anisotropy to measure the affinity (K_D) of the interaction between purified transpososomes and target DNA (Figure 2.4). To measure these values under conditions where the MuA active site and the ion cloud surrounding the DNA would be as realistic as possible, we performed these experiments in the presence of Mg^{2+}. Rather than by withholding divalent metals, strand transfer was prevented by using Mu end DNAs lacking the terminal 3'OH group on the transferred strand, which is the nucleophile for strand transfer. Binding affinities followed a similar pattern to reaction rates. The K_d of CDCs for mismatched target DNA with or without DMSO was about 15 nM, and for fully base paired target DNA with DMSO, about 29nM. In the case of the mismatched target, good fits to the data

required use of a cooperative binding model with a Hill coefficient of 3.3 (for the mismatch alone) and 2.6 (when combined with DMSO) We suspect this arises from transient transpososome disassembly events that become significant at low nanomolar concentrations. Conversely, we were unable to detect enough binding to normal DNA under normal buffer conditions to confidently fit a binding curve, but can visually estimate that Kd is between 0.5 – 1 µM. Thus, increasing the flexibility of the target DNA can enhance its affinity for transpososomes by at least 33-fold.

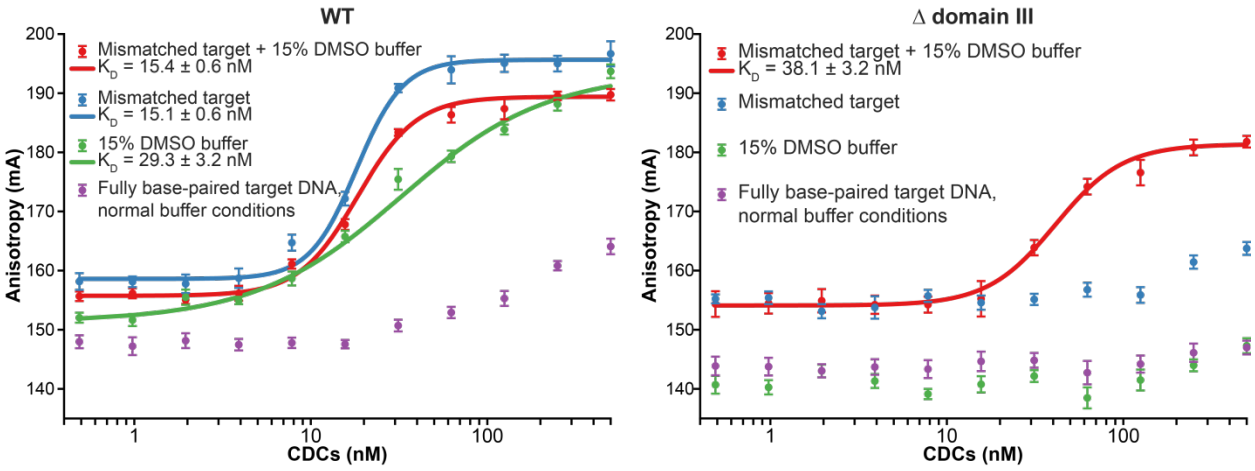

Figure 2.4 Enhanced flexibility triggers tight target DNA binding.
Fluorescence anisotropy measurements of transpososome binding to Atto565-labeled target DNAs. Dots are experimental data, with error bars representing 95% confidence intervals from the product of 5 technical replicates of 3 independent experiments. Solid lines represent fits to the data to obtain the KD values indicated in the legends.
A. Wild-type transpososomes, **B.** Δdomain III transpososomes (see Figure 2.4A).

These results indicate that bending or otherwise deforming the target DNA poses a significant energetic barrier to binding target DNA. We also note that the binding and kinetic data do not exactly correspond. For instance, although the mismatch and DMSO produced

similar enhancements in target DNA affinity (with K_ds differing by only two-fold), the mismatch had a much stronger effect on the strand transfer reaction than DMSO (with initial rates differing by about four-fold). This suggests that there may be additional target DNA conformational changes, after an initial binding and bending step, which are modulated specifically by local flexibility directly at the site of transposition. These may be necessary to properly position the target DNA phosphate backbone in the active site.

MuA construct(s)	Condition	K_D	error	A_{bound}	error	A_{free}	error	n_H	error
WT	**Mismatch + DMSO**	15.43	0.5626	189.4	0.5313	155.7	0.4134	2.610	0.1905
WT	**Mismatch**	15.12	0.5972	195.6	0.8420	158.6	0.5744	3.278	0.3296
WT	**DMSO**	29.34	3.238	193.6	1.003	151.3	0.4893	N/A	N/A
WT / Sin Mu	**Mismatch + DMSO**	19.12	0.9409	184.3	0.6454	153.9	0.5942	2.168	0.2278
WT / Sin Mu	**Mismatch**	10.51	0.3568	197.7	0.4909	159.3	0.5060	2.570	0.1951
WT / Sin Mu	**DMSO**	26.44	4.230	178.2	0.9730	149.1	0.5914	N/A	N/A
Δ domain III / SinMu	**Mismatch + DMSO**	38.12	3.170	181.4	0.8538	154.0	0.7911	2.264	0.4385

Table 2.1: Fitted fluorescence anisotropy binding parameters.
The values of the variable parameters used in modeling CDC – target DNA binding from the fluorescence anisotropy data in **Figure 2.4**, with their calculated uncertainties in the following column. Uncertainties are expressed as 95% confidence intervals. A_{bound} is the anisotropy of the CDC:target DNA complex. A_{free} is the anisotropy of the labeled target DNA alone. n_H is the Hill coefficient that results from fitting the data from target DNAs containing a mismatched base pair to a cooperative binding model.

2.3.2 MuA domain III participates in target DNA binding and bending

The high binding affinity of transpososomes for flexible target DNA indicates that it must be bent in order to make optimal contacts with the transpososome. The crystal structure of the Mu transpososome shows that domain IIIα (the first ~45 residues of domain III) from the catalytic MuA subunits lies on top the target DNA in the middle of the target DNA U-turn. Note that domain IIIβ, which comprises the C-terminal ~60 residues of MuA, was not included in the crystal structure and is not required for transposition *in vitro* but has been included in the constructs used in this work. The position and positive charge of domain IIIα implied that it might be important for binding and bending the target DNA. To test this, we repeated the above kinetic and binding experiments using transpososomes lacking domain III on the catalytic MuA subunits. Although domain III can be removed by truncating MuA constructs at residue 560, transpososome assembly requires domain IIIα to be present on the two MuA subunits not involved in catalysis[102]. In order to form intact transpososomes by placing MuA lacking domain III at only the catalytic positions, we utilized our "SinMu" system[81] (Figure 2.6A). SinMu is a chimeric protein in which the sequence-specific DNA binding domains of MuA have been replaced with that of the unrelated Sin recombinase. It can be placed at the non-catalytic positions in the transpososome with a corresponding substitution in the Mu end DNA sequence. Chimeric "SinMu"-containing transpososomes that include the full MuA C-terminus at all subunit positions behave indistinguishably from wild type transpososomes (Figure 2.5) , and subsequent truncation of the catalytic subunits' domain III does not hinder transpososome formation (Figure 2.6B).

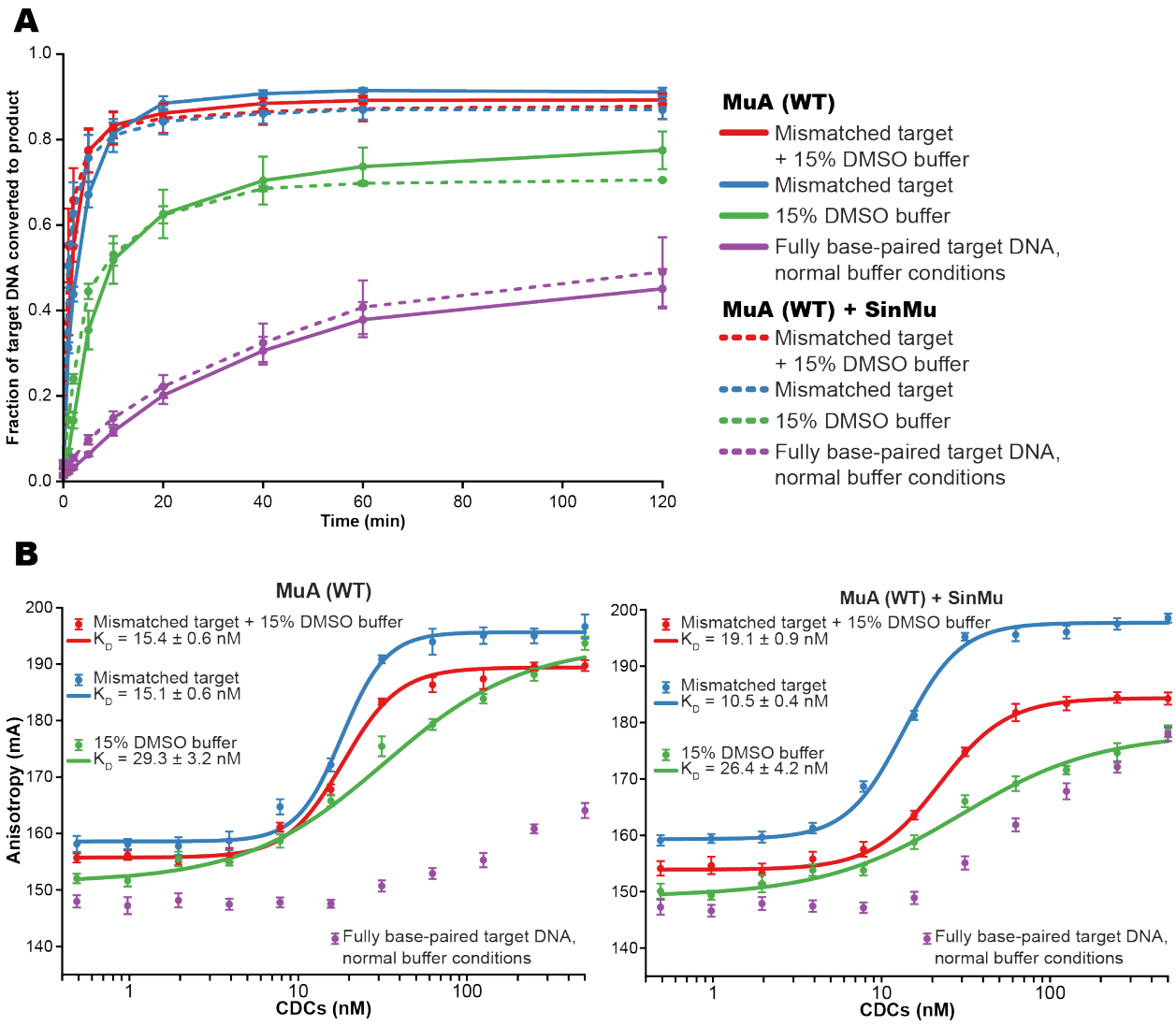

Figure 2.5: The SinMu system does not perturb target DNA interactions**A.** Strand transfer kinetics of WT MuA transpososomes compared to WT MuA + SinMu transpososomes under the same conditions. Data for WT MuA transpososomes is the same as in **Figure 2.2**.
B. Fluorescence anisotropy measurements of target DNA binding by WT MuA transpososomes compared to WT MuA + SinMu transpososomes under the same conditions. Data for WT MuA transpososomes is the same as in **Figure 2.4**.

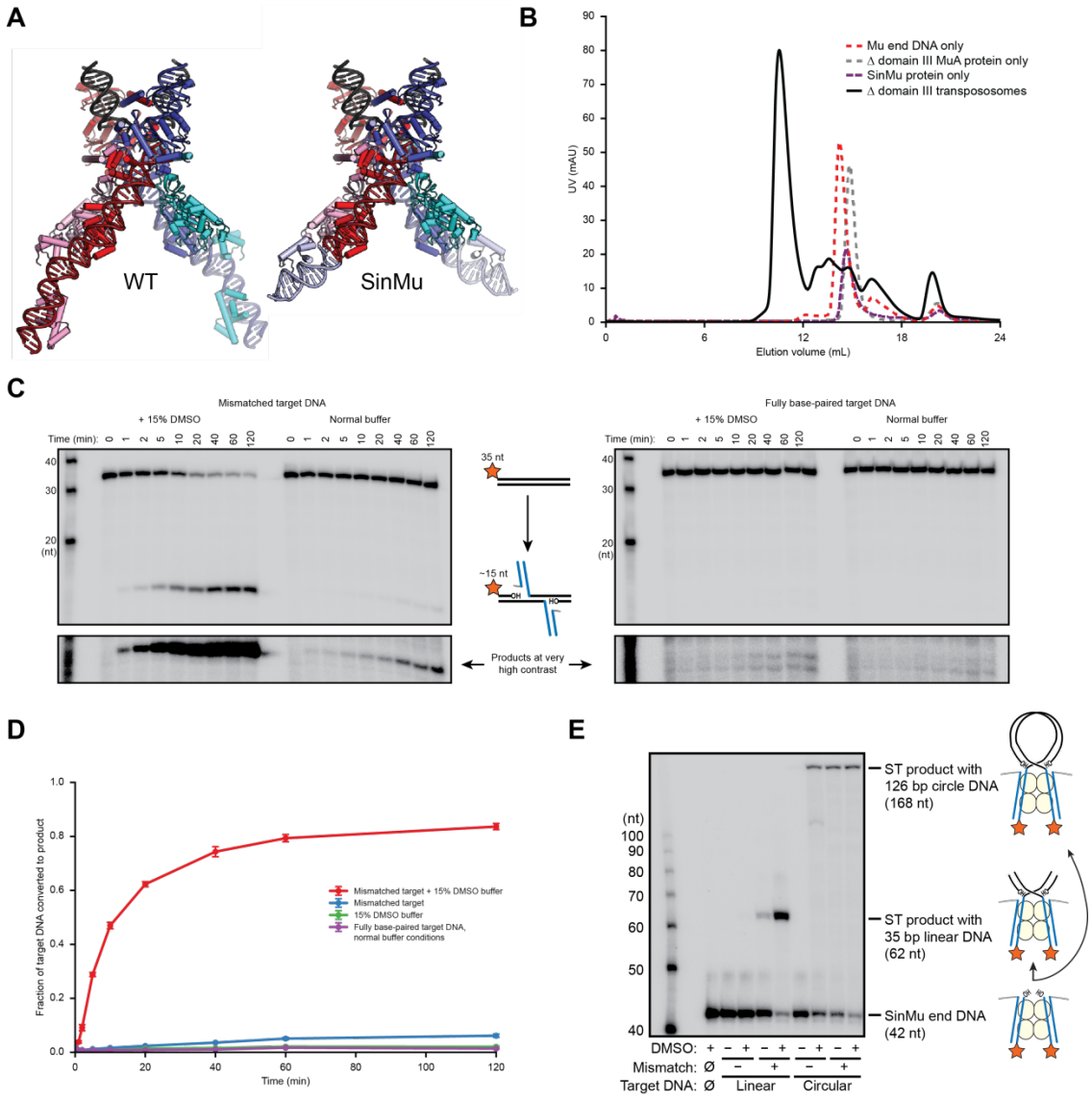

Figure 2.6 Δ Domain III Transpososomes Have Almost no Activity Except in the Presence of a Base-Pair Mismatch and DMSO.

A. Diagram of the SinMu system. The non-catalytic subunits (pink and cyan) are specified by an alternate DNA binding domain and DNA sequence (lavender). This allows Δ domain III MuA constructs to be targeted to only

B. Gel filtration chromatography of Δ domain III transpososomes. Moving to the chimeric SinMu system and truncating the catalytic MuA subunits to remove domain III does not hinder transpososome formation or their subsequent purification by gel filtration.

C. Strand transfer kinetics visualized by denaturing gel electrophoresis and 5'-^{32}P-labeled target DNA, as in **Figure 2.2B**, except using Δ domain III transpososomes. Lower panels are the product band(s) from the upper panels at greatly increased contrast.

(**Figure 2.6**, continued)
D. Quantification of the strand transfer kinetics experiments described in **B.** Y-axis represents the fraction of total lane signal present in product band(s). Error bars represent mean ± SEM, n = 4 time courses per condition.
E. Rescue of the strand transfer activity of truncated transpososomes by circular DNAs. ^{32}P-labeled Mu ends increase in size as a result of strand transfer. Samples were taken 2 hours after transpososomes were mixed with Mg^{2+} and indicated target DNAs.

We found that transpososomes in which the catalytic MuA protomers lack domain III ("Δ domain III transpososomes") display very little strand transfer activity, except under select conditions. Rather than just enhancing the reaction, DMSO and a base-pairing mismatch were virtually required for the truncated transpososomes to catalyze strand transfer into short linear target DNAs (Figure 2.6C, D). The reaction rate is otherwise exceedingly slow (< 1% of input after two hours). Measurements of target DNA capture once by the truncated transpososomes once again follow a similar pattern to strand transfer kinetics: binding is relatively robust when both modifications are present simultaneously ($K_D \approx 38$ nM), but undetectable otherwise (Figure 2.4B). Thus, domain III does indeed provide many of the protein-DNA contacts that are important to capture a target. High target DNA flexibility, however, can make up for the loss of these contacts. This confirms our assertion that protein-DNA contacts are optimized only for bent DNA.

To investigate further the connection between binding and bending, we tested whether the strand transfer activity of the truncated transpososomes could be rescued by providing pre-bent DNA for use as a target, rather than the linear DNA fragments used thus far. We generated pre-bent DNA by creating DNA minicircles (126 bp circularized DNAs). Instead of just being more flexible, minicircles should be naturally and permanently held in a bent state. This should provide an even lower energy barrier to bending. Accordingly, we found that minicircle DNAs

can rescue the strand transfer activity of Δ domain III transpososomes under a number of conditions (Figure 2.6E and Figure 2.7). While Δ domain III transpososomes generate a substantial amount of strand transfer products into linear DNA only in the presence of both the mismatched base pair and DMSO, strand transfer into minicircles requires only one or the other. We take this as further evidence that DMSO and the base-pairing mismatch allow the DNA to be more easily bent, and that target DNA binding and bending by the Mu transpososome are linked.

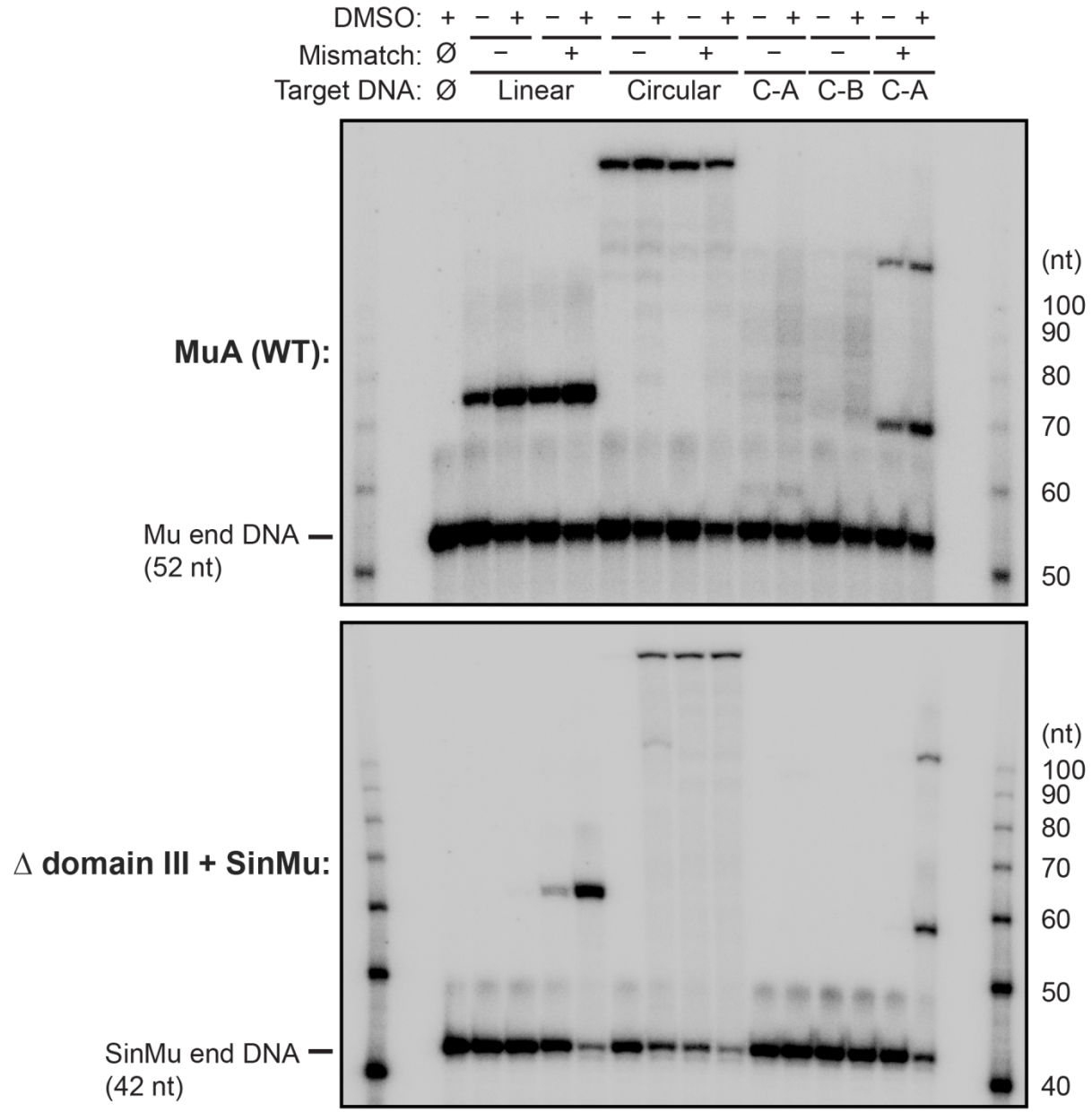

Figure 2.7: Strand transfer rescue conditions for Δ domain III transpososomes.
Labeled Mu end DNA increases in length as a result of strand transfer. The linear and minicircle targets are the same as those used in **Figure 2.6** and throughout this chapter. C_A (65 bp) and C_B (61 bp) are the linear DNA fragments used to construct the minicircles. Note that strand transfer products appear in all lanes (except the control) for WT MuA transpososomes, although C_A and C_B create a ladder of faint products due to their length and Mu's lack of target site specificity. Strand transfer by Δ domain III transpososomes into linear DNA targets requires the combination of a mismatched base-pair and DMSO. However, a minicircle target can substitute for either modification.

2.3.3 Strand transfer reversal only occurs when target DNA binding is severely compromised

In principle, the chemical step of strand transfer is simply a change in phosphodiester bond connectivity. Coupled to the entropically unfavorable action of bringing disparate DNAs together, the energy landscape of strand transfer might be expected to favor its reversal (hereafter referred to as "disintegration"). Nevertheless, our data above show that strand transfer is capable of going nearly to completion (Figure 2.2C), indicating that the transpososome is able to suppress disintegration. Published reports currently differ as to the rate at which Mu transpososomes catalyze disintegration. This has previously been difficult to address because it requires purification of STCs away from excess components (which would otherwise be identical to disintegrated products) coupled to sensitive detection methods. Generating a pure population of STCs by direct assembly on DNA substrates mimicking the strand transfer products or by purification via prolonged gel electrophoresis both result in an off-pathway pseudo-disintegration reaction that has been referred to as a "foldback"[108,109].

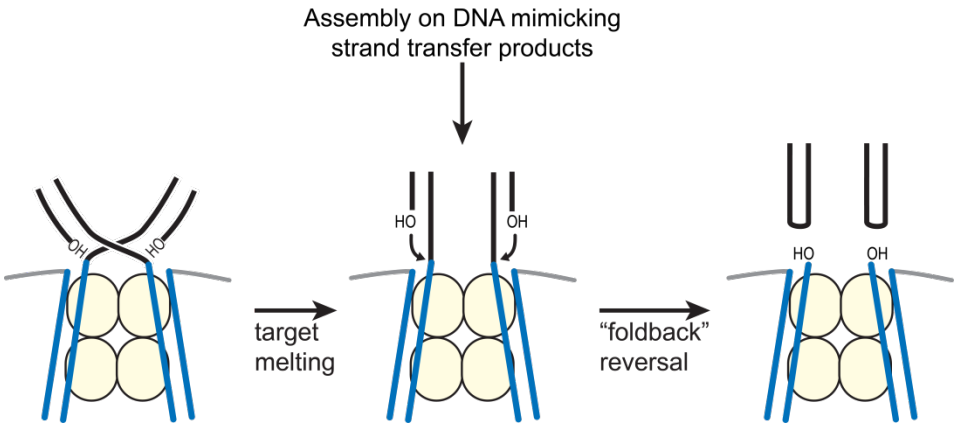

Figure 2.8: Psuedo-disintegration via the "foldback" pathway.

(**Figure 2.8**, continued) Assembly of transpososomes on branched DNA substrates designed to mimic the strand transfer products, or purification of strand transfer complexes by gel electrophoresis, causes the melting of the 5 base pairs between strand transfer sites in the target DNA. Mu transpososomes react with this DNA configuration in such a way that the incorrect 3' hydroxyl group is used to reverse strand transfer, resulting in disintegrated Mu end DNA but with the target site unrepaired and hairpinned.

To assay disintegration here, we rapidly and gently purify strand transfer complexes immobilized on neutravidin-coated magnetic beads via biotinylated Mu end DNA. Unreacted labeled target DNA is washed away and replaced with excess mismatch-containing (and thus high affinity) cold competitor target to trap disintegration events; residual reactants can be accounted for by sampling the reaction immediately after purification. Monitoring these purified strand transfer complexes as a function of time shows that they are impressively robust against disintegration under our normal reaction conditions, with, at most, less than 2% of strand transfer products reverting to re-ligated target DNA after 60 minutes (Figure 2.8A). This is true even for Δ domain III transpososomes, despite removal of the DMSO that was needed for the forward reaction and their compromised target DNA binding and bending ability. The remaining transpososome-target contacts to the bent target DNA must be sufficient to prevent disintegration.

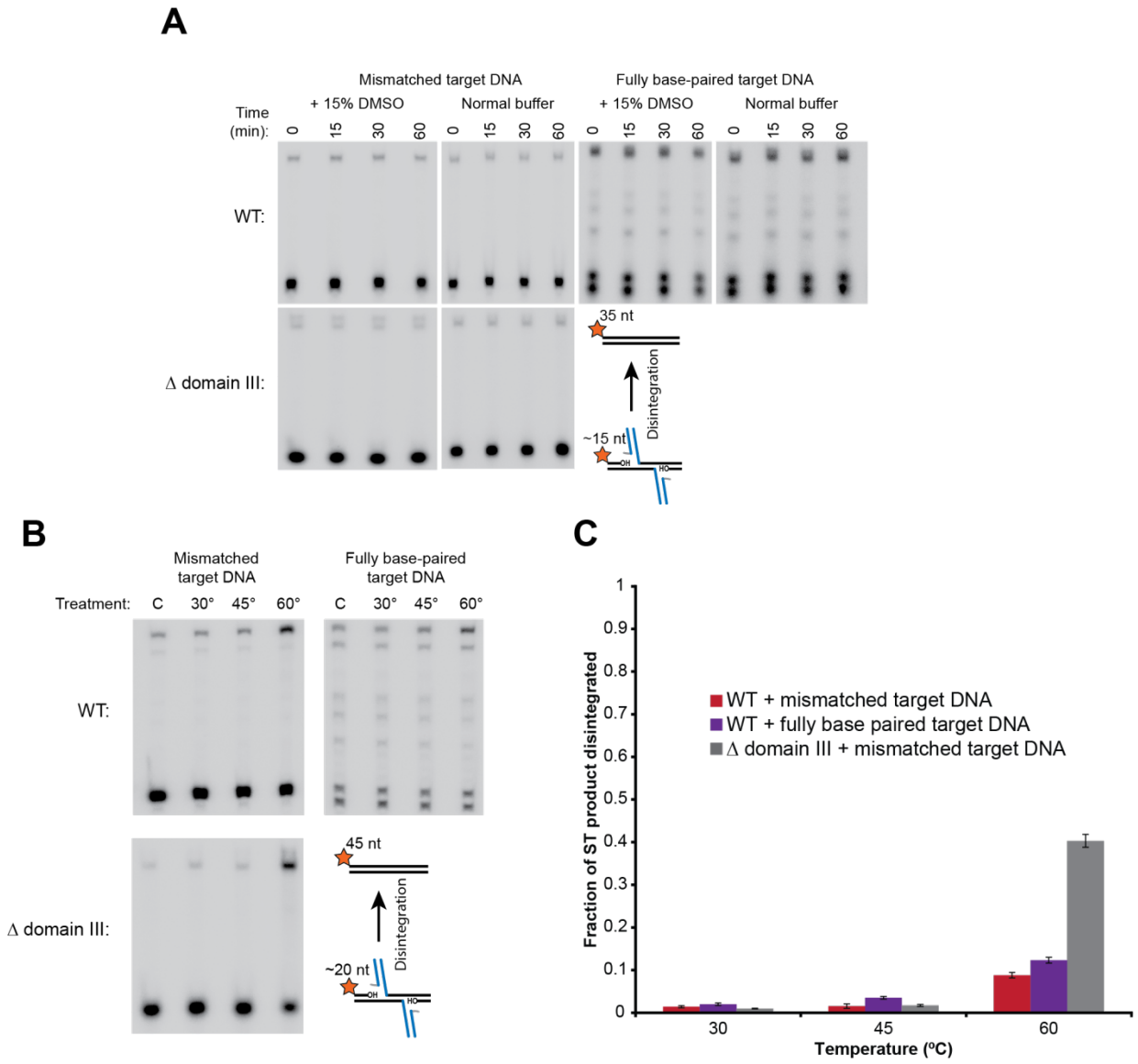

Figure 2.9: Strand transfer complex disintegration.
A. Disintegration under normal reaction conditions visualized by denaturing gel electrophoresis and 5'-^{32}P-labeled target DNA. STCs were rapidly purified by immobilization on magnetic beads and left at 30°C in the same buffer in which they were formed.
B. Disintegration under modified reaction conditions. The same procedure as in (A), but STCs were purified into a modified buffer (see text) and held for 1 hour at the indicated temperatures. Target DNAs in these experiments have 5 bp added to each end to prevent melting during treatment.
C. Quantification of replicates of the experiment described in **B**. The Y-axis represents the fraction of strand transfer product present immediately after purification that became re-ligated after treatment. Error bars represent mean ± SEM, n = 4 independent replicates.

Previous studies have indicated several factors that could stimulate reversal: increased temperature, increased pH, and higher concentrations of glycerol in the reaction buffer[76,108]. We note that the first two modifications might be expected to weaken protein-DNA binding between the transpososome and bent target DNA. Indeed, we found that the combination of slightly increased buffer pH (7.9 instead of 7.4), increased glycerol content (16% instead of 5%), and high temperatures (60°C) triggered disintegration. Under these conditions, transpososomes converted between 10-15% of strand transfer products back into intact target DNA after 1 hour (Figure 2.8B, C). In our assay, this occurs without producing the off-pathway "foldback" products that have complicated previous studies. Unlike the forward strand transfer reaction, disintegration was not highly affected by including a mismatch in the target DNA sequence. This suggests that the energetic barrier to reversal does not involve changes in DNA conformation at the center of the integration site.

Remarkably, when subjected to the same buffer and heat challenge, Δ domain III transpososomes catalyze 4-fold more disintegration than the WT complex, reverting about 40% of strand transfer products back into intact target DNA. Although this could be explained by these transpososomes releasing transiently disintegrated DNA more readily, in light of the lack of difference between normal and mismatched target DNA above, such an explanation seems unlikely. Rather, we favor a model where domain III is important for controlling the conformation of the target DNA. For forward strand transfer, it assists in bending and/or deforming the target DNA at the center of the target site. After strand transfer, it likely binds to

and maintains the U-shaped target DNA observed in the crystal structure (Figure 2.1B). Maintenance of the bend seen in the crystal structure thus appears to resist disintegration.

2.4 Discussion

In this study, we have shown that the conformation of the target DNA has dramatic consequences for transposition by phage Mu. Our model for this process is shown as a free energy reaction coordinate diagram in Figure 2.9. Newly assembled transpososomes bind to potential target DNA low affinity and are slow to utilize bound DNA for strand transfer. This this is because the energy needed to bend the DNA outweighs the resulting protein-DNA binding energy, and bending is necessary to position DNA in the transposase active sites. These properties change sharply, however, in the presence of DNA whose structure has been compromised to favor flexibility or bending. Such DNA is bound tightly and used rapidly as a transposition target because bent DNA can optimize transpososome-target contacts while incurring a smaller energetic penalty for strain in the DNA. That a highly flexible base pair mismatch can precisely localize the transposition sites at single base pair resolution strongly suggests that transpososomes scan potential target DNAs for flexibility, as has been proposed previously. Then, in our model, much of the conformational strain accrued by the bent target DNA is released by strand transfer nicking the target DNA backbone. These nicks can become the foci of structural irregularity, permitting much of the target DNA to assume a lower energy conformation. In combination with protein-DNA contacts with the transpososome, this creates a deep energetic well out of which the strand transfer products rarely reverse. Given the

established structural and behavioral similarities between them, we expect that this model applies broadly to DDE transposases and retroviral integrases.

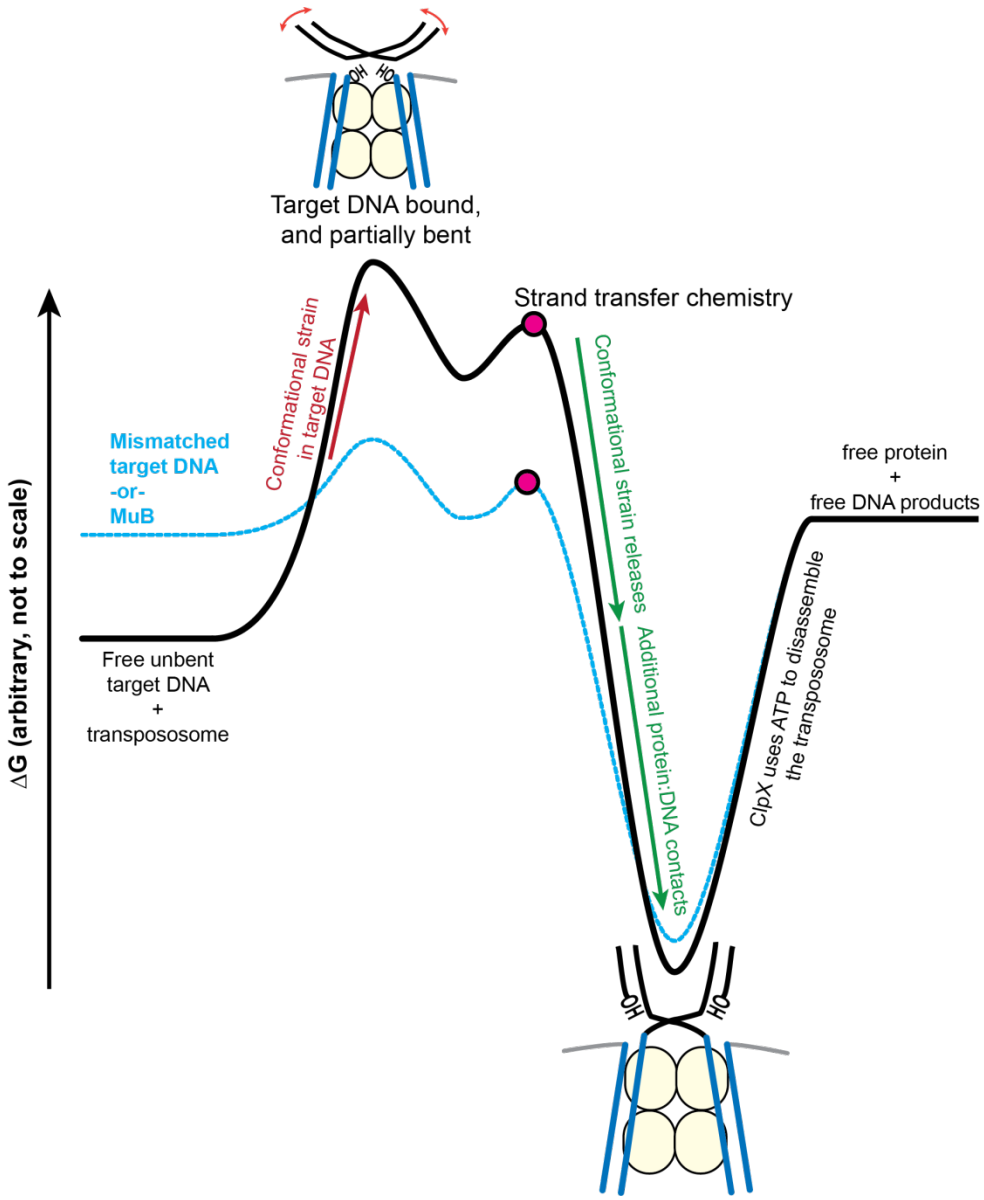

Figure 2.10: Target bending is an energy barrier to transposition.
The events surrounding strand transfer are represented here as a free energy diagram. Interactions with target DNA are unfavorable because the transpososome target binding surface is shaped to accommodate bent DNA. This can be ameliorated by enhancing DNA flexibility *in vitro* or by the actions of transpososome targeting mechanisms like MuB. Strand transfer releases

(**Figure 2.10**, continued) much of the bending strain and optimizes transpososome:target interactions. The strand transfer complex then occupies an energy well out of which it is unlikely to move.

This energy landscape provides solutions to two challenges transpososomes face in their interaction with target DNA. First: why should transposition be slow by default? Why scan through potential target sites? Although many transposons like Mu do not hunt for specific target sequences, target selection *in vivo* is most often not random. In particular, transposition of an element into itself or another copy of itself is commonly avoided. Strand transfer of an element into itself must be avoided because the resolution of such products is destructive to segments of the element and/or its host[91]. Avoiding self would be particularly difficult for a highly reactive transpososome because the local concentration of self-DNA neighboring a transpososome is, by its very nature, high. Many bacterial transposons, including IS21 elements,[120] Tn7,[121] and Mu, encode a separate ATPase protein that directs transposition to non-self targets, while retroviruses seek out the nucleosomes that mark host chromatin[107]. We propose that transposition is slow by default to repress transposition except in the presence of DNA bound by targeting protein(s).

Our results also suggest how transposition targeting mechanisms might activate transposition: by providing the flexibility that the transpososomes are scanning for (blue dashed line in Figure 2.9). In particular, it has been difficult to explain how target DNA coated or encased in targeting protein can be a high-affinity target. We propose that the answer lies in extruded bent DNA, in light of the orders of magnitude increase in affinity that would provide. The structural data available for MuB and IstB (the targeting ATPase from IS21elements) show that they have the ability to bundle and change the helical conformation of the DNA they coat[89,122], which could deform adjacent DNA and/or induce looping. This is supported by the

observation that Mu transposition is focused on DNA immediately adjacent to MuB coated DNA, rather than inside the filaments themselves[123]. These loops could act like our mismatched DNAs and minicircles, triggering binding and rapid transposition. For retroviral integration, nucleosomes provide pre-bent DNA by their very nature. Retroviral integration happens directly on the nucleosome, and a high-resolution structure of this event has been obtained by cryo-electron microscopy[124]. Nucleosome DNA could be a favored target as long as the energy needed to dislodge a DNA loop slightly from the nucleosome surface and into the integrase active sites is less than that required to bend DNA *de novo*. There is biochemical evidence for this energetic compromise, as integration is known to occur predominantly at positions in nucleosomal DNA that are bound least tightly to the histone core[113,124].

Transposons also face a fundamental thermodynamic challenge. Strand transfer is a rearranging of phosphodiester bonds, resulting in products that are energetically quite similar to (or entropically uphill from) the reactants. In theory, a traditional enzyme should bring such a reaction to no more than about 50% completion. Yet we and others have shown that transpososomes are able to catalyze strand transfer overwhelmingly in the forward direction only as they remain bound to their products. What is the structural basis for translating product binding energy into preventing strand transfer reversal? Our work here provides strong evidence for a model where binding and keeping the target DNA in a bent state is the key. We have shown that transpososomes compromised in target DNA bending are prone to reversal, which is the first direct experimental corroboration of this idea of which we are aware. The nicks opened in the target DNA backbone by strand transfer give it added conformational freedom to release the bending strain built up during binding, permitting much of the target DNA to relax. The

structural consequences of this relaxation could eject the strand transfer products from the active site. The four available transpososome STC crystal structures show the result of this change: the newly-created phosphodiester bond and/or target 3'OH are dislodged from the active site[49,110-112]. As long as the transpososome binds tightly to keep this conformation in place, no reverse reaction can occur.

2.5 Methods

2.5.1 Expression and purification of MuA and SinMu proteins

Proteins were expressed in the Rosetta DE3 Escherichia Coli strain (EMD Millipore) from coding sequences cloned into the pET3c plasmid. Transformed cells were grown at 37°C in LB media supplemented with 100 µg/mL ampicillin to OD600 ≈ 0.7, then protein expression was induced by addition of 0.66 mM IPTG and an additional 20 µg/mL ampicillin. At 2 hours after induction, the cells were collected by centrifugation and stored at -80C until lysis. Cells were resuspended in a solution of 25 mM HEPES pH 7.5, 1 mM EDTA, 1 M NaCl, 10% sucrose, 10% glycerol, 5 mM DTT, and Complete protease inhibitor (Roche Diagnostics) and lysed by two passes through a Microfluidics LV1 microfluidizer. Cell debris was removed by centrifugation and the resulting supernatant was fractionated by addition of ammonium sulfate to 30% saturation. The precipitate was collected by centrifugation, and the resulting pellet was resuspended in 20 mM MES pH 5.5, 0.2 M NaCl, 0.5 mM EDTA, 5% glycerol, and 1 mM DTT (Buffer A). This was passed over a HiPrep Heparin FF 16/10 affinity column (GE Healthcare) and eluted with a gradient from 0.2 M (Buffer A) to 2 M (Buffer B) NaCl. Fractions containing the protein were diluted back into Buffer A and the same procedure was used to run a second

pass of the heparin affinity chromatography. Fractions containing the protein were concentrated and further purified by gel filtration chromatography using HiLoad 16/600 Superdex 75 prep-grade column (GE Healthcare) equilibrated with 25 mM HEPES pH 7.5, 0.4 M NaCl, 0.5 mM EDTA, 5% glycerol, and 1 mM DTT. Peak fractions were combined, dialyzed into the same buffer supplemented to 20% glycerol, concentrated, and stored at -80°C. The final concentration of proteins was determined by measuring their absorbance at 280 nm.

We would like to thank Lorraine Ling, Robert T. Sauer, and Tania Baker from the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA for their gift of a SinMu construct including the full MuA C-terminal domain.

2.5.2 DNA substrates

Linear DNA substrates were created by annealing complementary synthetic oligonucleotides of the desired sequence. For annealing, oligonucleotides were mixed in equimolar amounts in a buffer of 10 mM Tris pH 8, 100 mM NaCl, and 1 mM EDTA, heated to 80°C, and then slowly cooled to room temperature. All oligonucleotides were synthesized by Integrated DNA Technologies (IDT). Oligonucleotides modified with biotin or fluorophores were HPLC purified by IDT, and oligonucleotides for 32P labeling were PAGE purified by IDT. Oligonucleotides were resuspended in TE (10 mM Tris pH 8, 1 mM EDTA), and those that were not purified by IDT were desalted using P6 spin columns (BioRad) equilibrated in TE. The concentration of all oligonucleotides was verified by measuring their absorbance at 260 nm. DNA substrates were radiolabeled at the 5' end using γ-32P ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs).

2.5.3 Transposition reactions

Cleaved donor complexes (CDCs) were formed by mixing protein and Mu end DNAs in a 1:1 protein:DNA binding site ratio and incubating at 30°C for at least 1 hour. CDCs were formed in a buffer of 25 mM HEPES pH 7.4, 200 mM NaCl, 0.5 mM EDTA, 5% glycerol, and 0.6 mM Zwittergent 3-12 (EMD Millipore). To increase the yield of CDC tetramers in cases where gel filtration was used, this buffer also contained 15% DMSO. Purification of CDCs by gel filtration, where indicated, was performed using a Superdex 200 Increase 10/300 column (GE Healthcare) equilibrated in the same buffer, but lacking DMSO. After gel filtration the tetramer peak was collected, spin concentrated and used immediately. The strand transfer reaction was initiated by diluting equimolar amounts of CDC and the appropriate target DNA into a buffer identical to the previous except that EDTA was omitted and replaced with 10 mM MgCl2. This buffer contained 15% DMSO where indicated. Strand transfer was carried out at 30°C. Except in cases involving transpososomes attached to magnetic beads, strand transfer reactions were stopped by phenol:chloroform extraction.

2.5.4 Fluorescence anisotropy

CDCs were formed as described above, except with DNAs ending in dideoxy-A and in a buffer containing 10 mM MgCl2, which was used for all subsequent steps (supplemented with 15% DMSO where indicated). Purified CDCs were serially diluted, mixed with 6 mM Atto565-labeled target DNA, and arrayed into a Corning 3575 black polystyrene 384 well microplate. Binding reactions were incubated for 1 hour at room temperature. A Victor X5 plate reader was

used to read the anisotropy of the Atto565 fluorophore, using a 531 nm excitation and 595 nm emission filters and a 1.0 second counting time. The raw parallel and perpendicular photon counts were adjusted to account for the instrument response (G factor and buffer blanks), and the resulting anisotropies were fit to an equilibrium binding model, accounting for receptor depletion, using the optimize.curve_fit function from the Python scipy package.

2.5.5 Minicircle DNA construction

Minicircle DNAs were constructed from two linear dsDNA pieces of approximately equal length. The DNA pieces were mixed in an equimolar ratio along with a twofold molar excess of IHF in T4 Ligase buffer (New England Biolabs) for 30 minutes at room temperature. T4 DNA Ligase (New England Biolabs) was then added and the mixture incubated for 12 hours at room temperature. Proteins were removed by phenol:chloroform extraction followed by P6 column buffer exchange (BioRad) into fresh T4 Ligase buffer. T4 DNA Ligase was added back in and incubated at room temperature for an additional 2 hours to remove any remaining nicks. This reaction mixture was phenol:chloroform extracted and separated by gel electrophoresis on an 8% polyacrylamide TBE gel. The band corresponding to the circular product was identified by UV shadowing, excised, and extracted by shredding the gel slice and soaking in a ~15-fold volume excess of 10 mM Tris pH 8, 250 mM NaCl, and 2 mM EDTA for 8 hours at room temperature. Gel fragments were removed by filtration. To remove any remaining linear fragments or nicked circles, an equal volume of 2x BAL-31 nuclease buffer and BAL-31 nuclease (New England Biolabs) were added and incubated for two hours each at 30°C and 40°C. The nuclease was removed by addition of EGTA to 20 mM followed by

phenol:chloroform extraction and P6 column buffer exchange into a buffer of 10 mM Tris pH 7.5, 100 mM NaCl, 15% glycerol, 2 mM EDTA, and 2 mM EGTA. Minicircles were stored at -20°C.

2.5.6 Strand transfer reversibility

CDCs were formed as described above with DNAs where the transferred strand was biotinylated on the 5'end, then incubated for an additional 30 minutes after addition of Sera-Mag SpeedBeads NeutrAvidin particles (GE Healthcare). Using the magnetic beads, the CDCs were washed to remove unbound protein and DNA, resuspended in reaction buffer (including 10 mM MgCl2 and 15% DMSO), and provided with an approximately two-fold molar excess of the indicated target DNA. The strand transfer reaction proceeded for 2 hours at 30°C. During the final 5 minutes, heparin was then added to a final concentration of 0.05 mg / mL to encourage dissolution of any partially formed complexes. Immobilized strand transfer complexes were then washed four times to remove excess target DNA and exchange into the desired (± DMSO) buffer. Snapshots of the reaction were then taken by diluting aliquots into a 10-fold volume of excess of 97% formamide + 10 mM EDTA that had been pre-heated to 100°C. The first snapshot was always taken immediately after the final wash (t = 0 min) to record the initial baseline levels of reacted / unreacted DNA.

Chapter 3

Dimerization is a slow step in transpososome formation

The hyperactivating MuA mutations described in this chapter were discovered by our collaborators Tiina S. Rasila, Mauno Vihinen, Lars Paulin, and Harri Savilahti at the Institute of Biotechnology, University of Helsinki, Finland.

The work to clone, express, and characterize the hyperactive MuA mutants was done with the help of University of Chicago undergraduate students Justin Salat and Vishok Srikanth.

Determination of intra-molecular MuA crosslinks by mass spectrometry was carried out by Dr. Bradley Evans at the Donald Danforth Plant Science Center, St. Louis, MO.

3.1 Introduction

Recombination of DNA is accomplished through the action of large nucleoprotein complexes. These complexes involve multiple protein binding sites at disparate locations in order to pair the proper DNA sites for exchange of DNA. When these multiple binding sites are contiguous, they are often tandem repeats that bind multiple copies of the same protein. This allows for cooperative binding to increase the accuracy of recognition of critical sites. In addition, many mobile DNA elements employ additional nucleoprotein machinery to ensure that recombination occurs between sites of proper topology. For instance, many transposases and site-specific recombinases are only active as parts of larger nucleoprotein complexes that depend

on secondary protein binding sites, supercoiling, and DNA bending proteins. These complexes read out the relative orientation of potential recombination sites and bring the proper ones together, despite sometimes distant separation in linear sequence space.

The transposition system of the *E. coli* bacteriophage Mu is one of the best studied of these complex recombination systems[39]. It depends on the same transposition chemistry to accomplish both its initial integration into its host genome and subsequent replication prior to lysis[31,33,125]. The Mu genome is 36.7 kb in length, but its ends must be correctly paired for successful transposition. This is expected to become especially challenging because of the proliferation of incorrect end pairings that would result from amplification of the genome during the lytic phase. The Mu transpososome is the nucleoprotein complex that carries out integration and replicative transposition. Its core is made up of four copies of the MuA transposase protein that are bound to the ends of the phage genome[44,126].

The Mu transpososome assembles by a complex series of events. These events involve MuA subunits bound to both ends of the Mu genome, as well as to an "enhancer" region in the genome interior[57]. Assembly also depends on the binding of the DNA bending proteins HU[52] and IHF[58]. These three sites are all synapsed together in the transpososome, but they come together in a particular order: the right end and enhancer capture each other first, and then intertwine with the left end. The assembled transpososome traps a defined number of supercoils between the left end, right end, and enhancer, and so is topologically specific[65]. Presumably, this complex and specific assembly process ensures that transpososomes only form on proper pairs of Mu left and right ends.

After assembly, the actual chemistry of transposition occurs when the transpososome cleaves the ends of the Mu genome, and then catalyzes the transesterification of those ends into the destination "target DNA" site. Every step in the transposition process is marked by an increase in transpososome stability[77]. The core tetrameric transpososome, once fully formed, is resilient to high temperature, salt, and removal of accessory subunits[60,61]. The crystal structure of the Mu strand transfer complex offers an explanation for this behavior (Figure 3.1). It shows a highly intertwined protein-DNA complex. In particular, the dimer of subunits that are bound to the same end bury about 1500 $Å^2$ of protein-protein surface area. The MuA subunit not engaged in catalysis has docked its catalytic domain on the DNA binding domains of the upstream catalytic subunit, and its domain IIIα is stretched across one upstream DNA binding domain. For the remainder of this chapter, unless otherwise specified, I use the term "dimer" to refer specifically to this coupling of two MuA subunits on the same Mu end DNA.

The 1500 $Å^2$ dimer interaction surface is not directly adjacent to any of the DNA components of the transpososome. Given this and the impressive stability of the higher order transpososome complex, a reasonable hypothesis would be that MuA could form dimers in the absence of DNA. However, the literature consensus is that MuA is a monomer. Exposing up to 1.4 mg/mL MuA to a lysine crosslinker fails to generate significant amounts of intermolecular crosslinked peptides[127], and a 16 Å cryo-electron microscopy structure of MuA generates an envelope volume that is well filled by only one set of MuA domains[103]. What, then, prevents MuA from oligomerizing prematurely? The cryo-EM monomer structure suggests that the DNA binding domains are docked against the catalytic domain, but in a different orientation and along a different face of the catalytic domain than their configuration in the dimer. This conformation

would compete with the dimer interface seen in the crystal structure and would need to be unfolded as part of transpososome assembly.

Here, I report new insights into the transition between MuA monomers and a transpososome tetramer. I begin by verifying the MuA monomer structure as suggested by cryo-EM using small-angle x-ray scattering (SAXS) and crosslinking. This work also suggests a new definition of the "minimal" MuA N-terminus. I then explore the behavior of hyperactive MuA mutants which appear to alter the monomer to dimer transition.

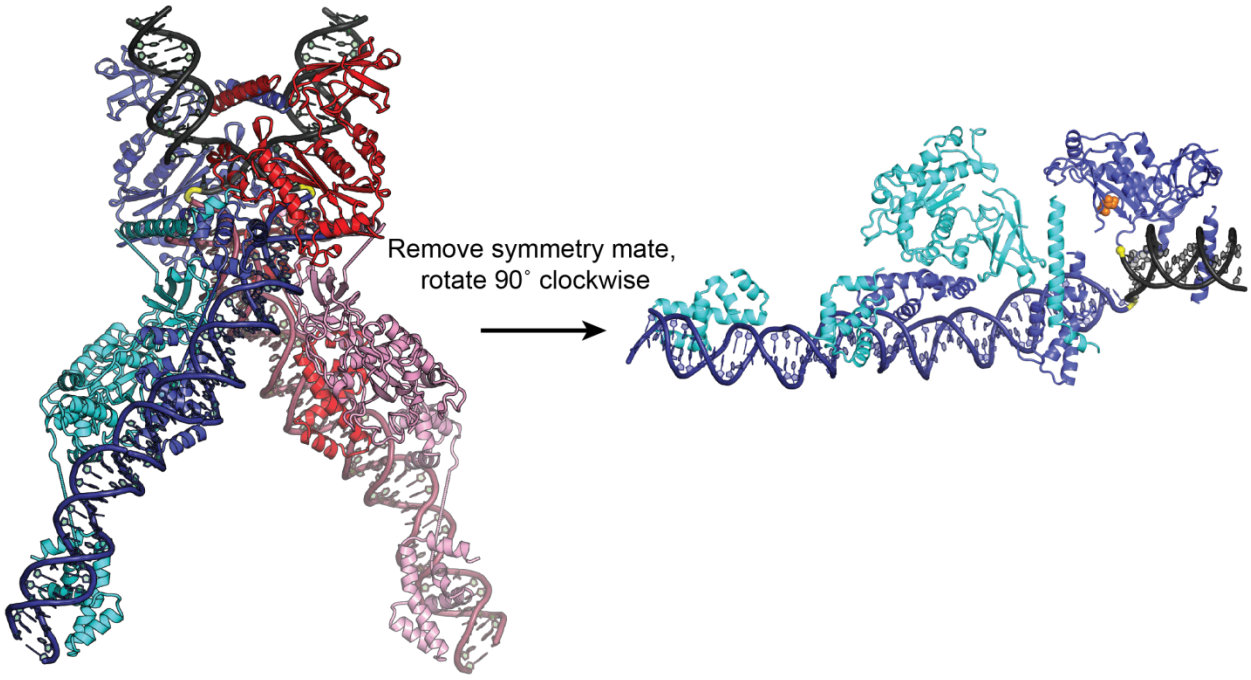

Figure 3.1: Transpososome subunits bound to the same DNA have significant contact area. On the left is the Mu transpososome crystal structure, which is two-fold symmetric. The asymmetric unit of the structure, right, contains a pair of MuA protein subunits bound to a single Mu end DNA. These two subunits bury about 1500 $Å^2$ between them. Each half of the transpososome is in shades of blue or red. The target DNA is in black. Yellow represents the phosphodiester bond between the Mu end DNA and target DNA.

3.2 Results

3.2.1 SAXS analysis of minimal MuA constructs

I first set out to verify the general shape and oligomeric state of MuA. Although previous studies had concluded MuA was a monomer, the crystal structure represents the highest resolution data about MuA to-date. It suggested two other possibilities: that MuA could be either a dimer, or a monomer in which the domains were not tightly associated. The former is supported by the large dimer interface, whereas the latter is suggested by the elongated nature of the domains of any one MuA subunit (Figure 3.1). The MuA construct used in the crystal structure contains residues 77 through 605, the minimal construct necessary to recapitulate transpososome assembly and strand transfer on short linear DNA *in vitro*.

SAXS data represent the rotationally averaged x-ray diffraction pattern from individual particles (in this case, proteins). This is a result of the experiment occurring in solution, where sparse particles are tumbling randomly. SAXS has the benefit of being able to observe proteins in their natural solvated environment. However, without the repeating lattice of a crystal to strengthen diffraction or the cryo-protection and fixed orientation of electron microscopy, it suffers from limited resolution. SAXS data can reliably yield the radius of gyration (R_g) of the scattering particle and, if it is a reasonably consistent shape (i.e., not overly flexible or unfolded), a probability distribution of intramolecular distances. The longest distance in this distribution is D_{max}, which represents the longest inter-atomic distance in the particle.

SAXS data for MuA 77-605 and its real space transform into a distance distribution plot are shown in Figures 3.2 and 3.3. Comparing this against theoretical distributions for both of the

crystallographic MuA dimers shows that MuA alone in solution is considerably more collapsed (Figure 3.4). This is strong evidence for a monomer conformation. However, my attempts to generate candidate structures by flexible fitting of the individual domain structures or by *de novo* generation of a molecular envelope were inconclusive. I suspect that my analysis was foiled by two factors. First, the considerable length of the inter-domain linkers gives them significant conformational freedom. Second, the presence of 3 pieces that can move independently (assuming one domain remains fixed as a reference) pushes the limit of the current capabilities of SAXS software.

Seeking further structural information, we wondered if truncating domain IIIα from the construct would either trigger a structural change or improve the viability of the SAXS reconstruction. In the transpososome dimer, domain IIIα from the structural subunit (left in Figure 3.1) makes extensive contacts with domain Iγ of the downstream catalytic subunit. We hypothesized that this same interaction could occur in a monomer between domain IIIα and Iγ of the same polypedtide. SAXS would reveal if domain IIIα was a critical lynchpin to keeping the DNA binding domain docked to the catalytic domain because the protein would appear as a larger particle after the truncation. This is not the case. SAXS data for a MuA construct further truncated at position 560 results suggests the particle size is only slightly smaller than the 77-605 construct (Figures 3.2B and 3.3).

In an attempt to improve future structural studies of MuA we also characterized an even shorter MuA construct by SAXS: 88-560. The N-terminal border of domain Iβ was defined at residue 77 largely by limited proteolysis[45]. This was carried forward as the N-terminus of the construct used in the NMR structure of domain Iβ[48]. In that structure, residues 77 through 87 are

not part of an alpha helix or beta sheet, but are nevertheless docked on the structure as an extended loop. Evaluating the interface between 77-87 and the rest of the structure, the docking seems to be driven by burying the hydrophobic residues I77 and L82 against the rest of the structure. We were lead to test their importance to the folding and function of domain Iβ for two reasons: (1) These residues were not resolved in the transpososome crystal structure, and (2) the NMR data deposited along with the structure of domain Iβ (PDB ID 2EZK) did not actually contain any restraints for this region. The SAXS data confirmed our suspicions. The smaller R_g indicates that 88-560 construct is slightly more compact than 77-560 (Figure 3.2B), indicating that it is still well folded. That the removal of just ten residues could make a detectable difference in the R_g of a 483 residue protein further suggests that they are located near the periphery of the protein and are not closely docked. This construct is also expressed at higher levels than 77-560, increasing the yield from protein purification.

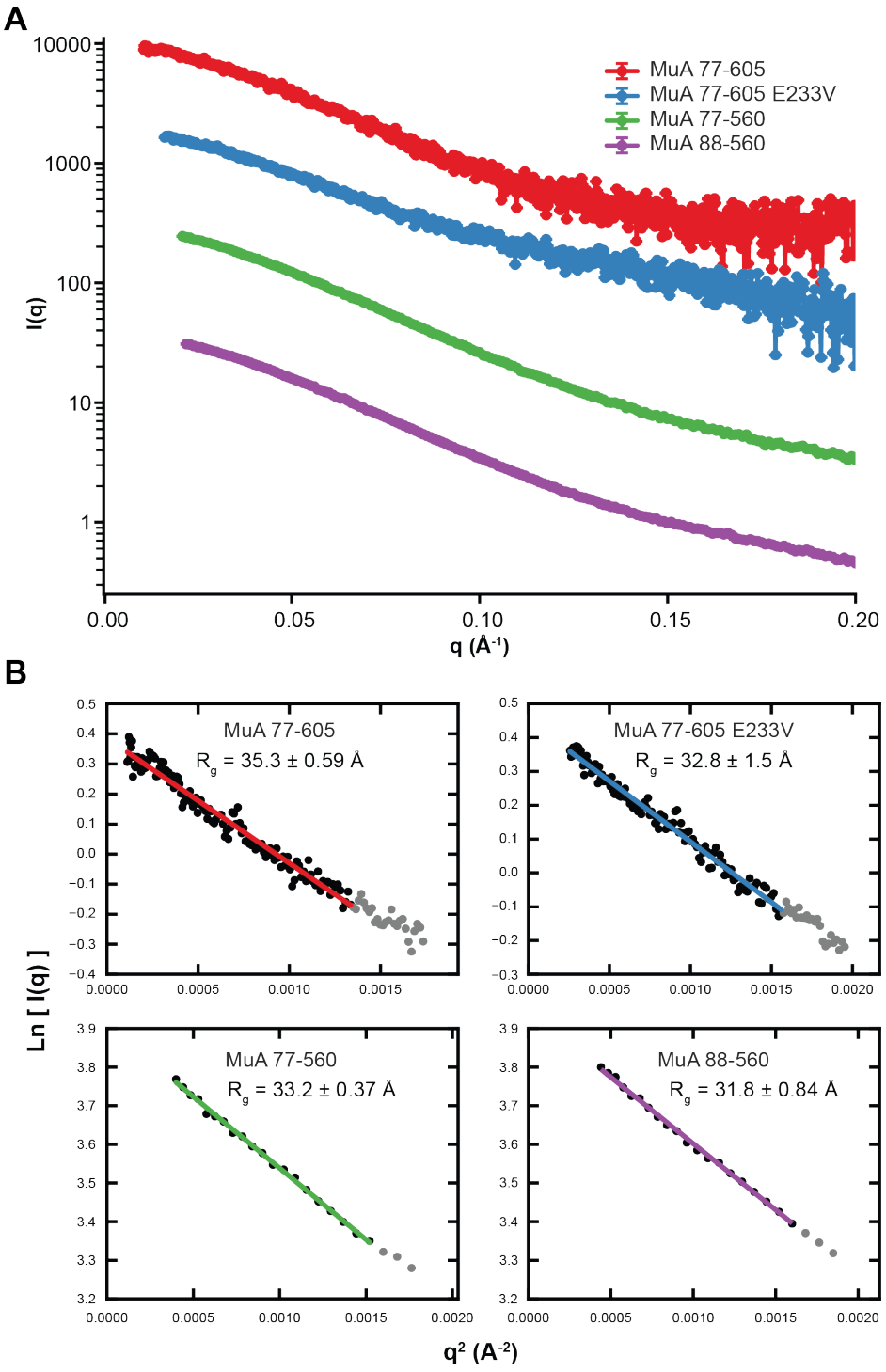

Figure 3.2: SAXS data and gunier analysis for MuA constructs.
A. Small angle x-ray scattering from the four indicated MuA constructs. Data for the 77-605 constructs were collected at a different beamline than the 77-560 and 88-560 constructs, hence the different grid on which q is sampled.

(**Figure 3.2**, continued)
B. Gunier analysis to determine the radius of gyration of the four MuA constructs from **A**. All constructs show excellent linearity in this region, indicating that particle size was uniformly distributed. The R_g for each is indicated in the figure.

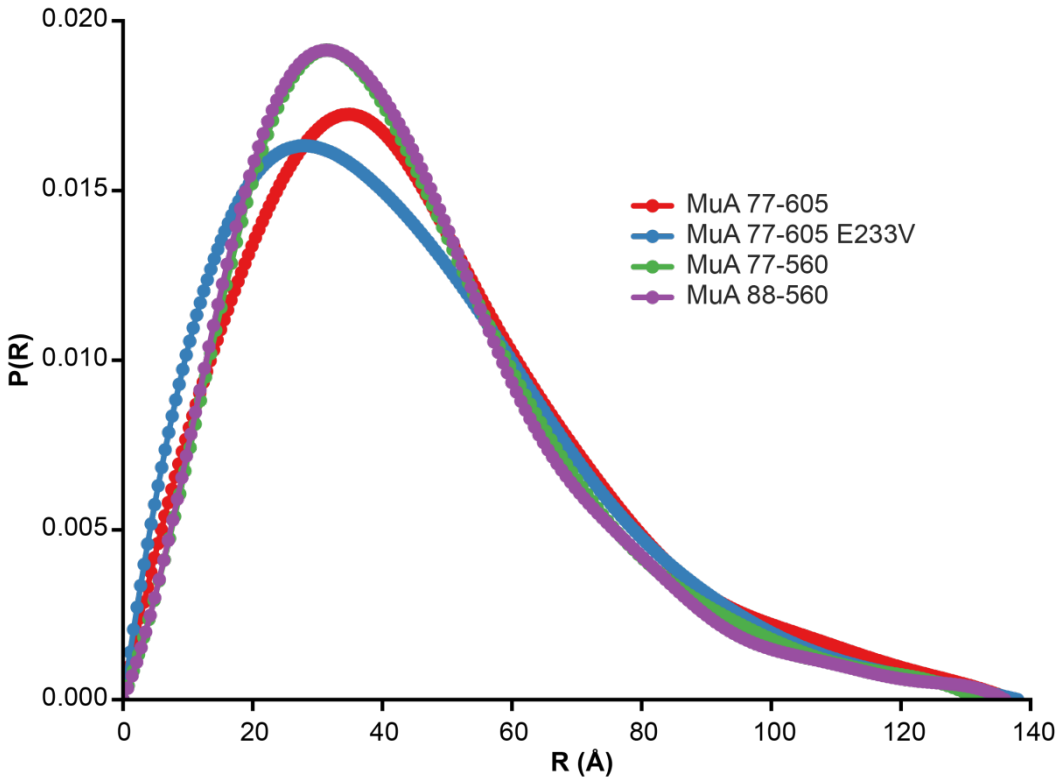

Figure 3.3: Real space distance distributions for MuA constructs.
SAXS data for all constructs can be well fit by a smooth distribution with 130 Å < D_{max} < 140 Å. This indicates that the differences between the constructs do not have a significant impact on their overall particle shape. Distributions are adjusted to unity area.

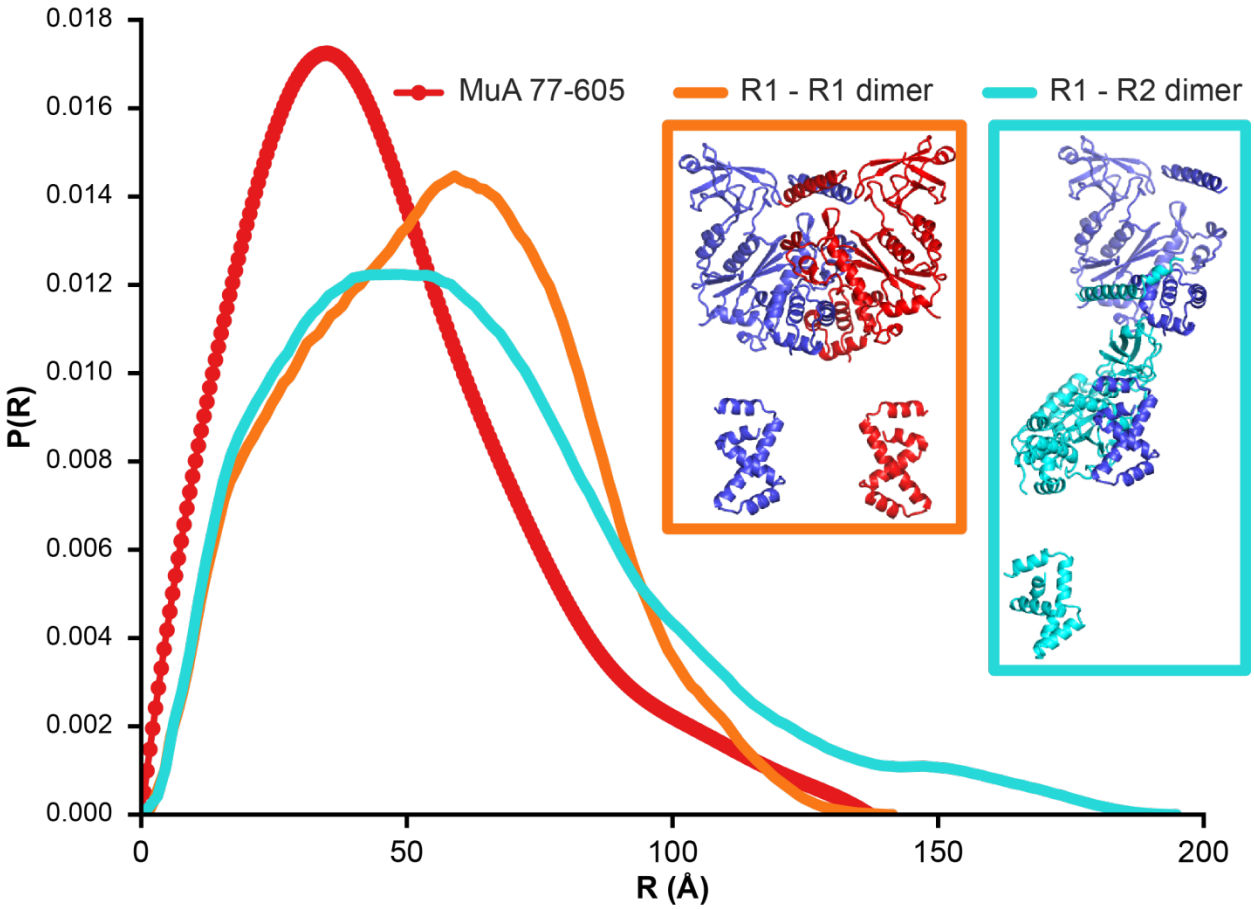

Figure 3.4 Distance distributions of MuA monomer and crystallographic dimers. Experimentally derived P(R) distribution for MuA 77-605, in red, is the same as in Figure 3.3. Orange and cyan are theoretical distributions calculated from the atomic coordinates of both dimer pairs in the transpososome crystal structure (PDB ID 4FCY).

Given the excellent properties of the 88-560 construct in SAXS, we decided to evaluate the possibility of using it in crosslinking studies. Having exhausted crystallography, intra-molecular crosslinking would provide additional restraints for future structural attempts by SAXS or electron microscopy. As a protein where every domain makes contact with DNA, MuA has many surface-exposed lysine residues. We selected a biotinylated lysine-reactive crosslinker designed to allow crosslinking positions to be detected by mass spectrometry[128]. 12 hour

incubation of a 10:1 mixture of crosslinker:MuA 88-560 showed no detectable accumulation of inter-molecular crosslinks (Figure 3.5), consistent with already published literature and the monomeric nature of MuA. A pilot attempt to sequence the crosslinking products resulted in the identification of 9 positions that reacted with the crosslinker. Of these, 5 were successfully crosslinked on both ends to peptides. All but one of the successful crosslinks occurred within the same domain, between two lysines that are separated by approximately the length of the crosslinker molecule (about 12 Å). Of particular interest, though, was 1 inter-domain crosslink between residues K157 (domain Iβ) and K461 (domain IIα). This is evidence that domain these two residues are positioned about 12 Å apart in the MuA monomer structure. This would require that domain Iβ is docked next to domain Iiα in approximately the position predicted in the monomer cryo-EM structure (Figure 3.6).

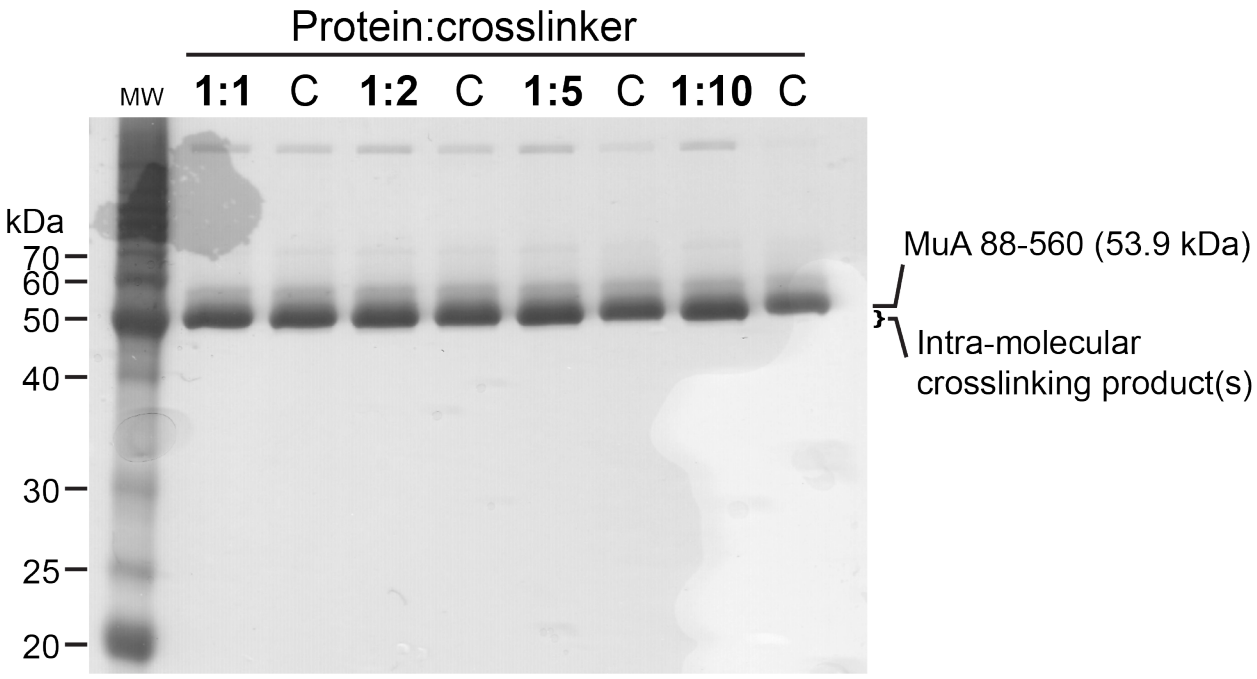

Figure 3.5: MuA 88-560 crosslinking produces only intra-molceular products.

(**Figure 3.5**, continued) The results of 12 hour crosslinking reactions are resolved here by SDS-PAGE. The indicated ratios are molar ratios of MuA 88-560 to lysine reactive crosslinker. Lanes labeled *C* are untreated protein for comparison. The presence of intra-molecular crosslinks yields a slight increase in the migration of the protein.

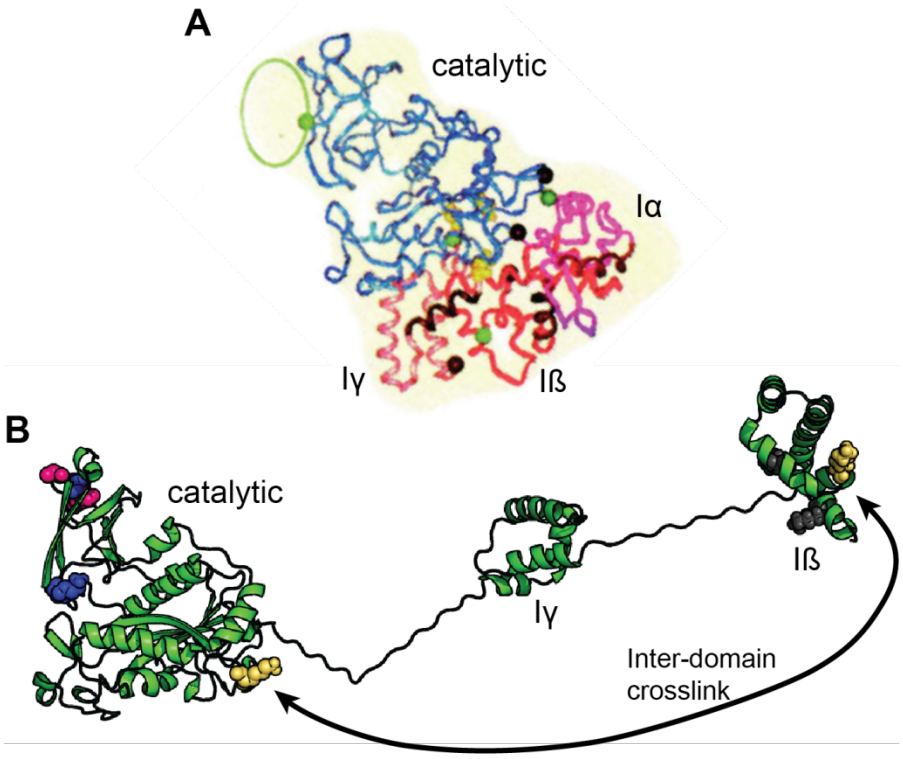

Figure 3.6: MuA monomer crosslinks.
A. MuA monomer structural model proposed by Yuan *et al.* Their model includes all three DNA binding domains and the catalytic domain.
B. Experimentally recovered lysine crosslinking positions mapped onto the domain structures of MuA 88-560. Three crosslinks (sidechains shown as blue, magenta, and black spheres) are intra-domain. One crosslink (yellow spheres, arrow) suggests a domain configuration that agrees with the model in **A**, placing domain Iβ in close proximity with the catalytic domain.

3.2.2 Mutations on the MuA dimer interface increase the efficiency of transpososome formation

Our collaborators from the Savilahti lab at the University of Helsinki, Finland have developed a screen for hyperactive mutations in the MuA sequence. Briefly, these mutations are

generated by error-prone PCR of a mini-Mu plasmid containing the gene for MuA and β-galactosidase[129]. The results are transformed into *E. coli*, and then colonies growing in the presence of X-gal are monitored for the development of blue papillae, which indicate that transposition has occurred. The causative single amino acid change can then be determined by sequencing. Somewhat surprisingly, very few of the hyperactivating mutations this group identified occur in the vicinity of the active site. Instead, they cluster along the dimer interface that has been the subject of this chapter. This includes some of the most powerful mutants, E233V and R478H, highlighted in Figure 3.7. E233V is the most powerful single amino acid change discovered so far. It is on the surface domain Iγ, on the opposite side from the DNA binding surface. When incorporated in a catalytic transpososome subunit (dark blue in Figure 3.3), it lies on the interface between Iγ and the domain IIβ and IIIα of the downstream subunit. We assume this is where is has its effect, because it is solvent exposed on the other subunit.

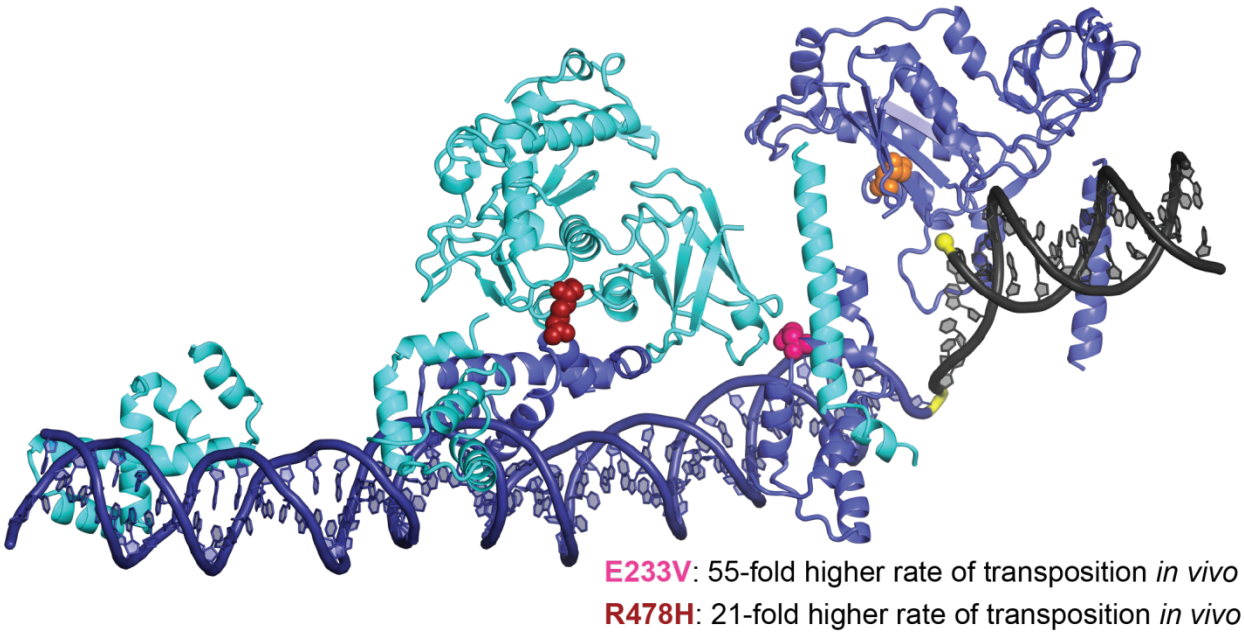

Figure 3.7 The location of powerfully hyperactivating MuA mutations.
The positions of two powerful hyperactivating amino acid substitutions are marked with pink and red spheres We assume these are the relevant positions because on the opposing subunits they are solvent exposed (not shown). For comparison, the catalytic DDE residues of the catalytic subunit (dark blue) are in orange.

To determine whether E233V hyperactivity is maintained even in the *in vitro* short linear DNA model system, we monitored the appearance of strand transfer DNA products as a function of time (Figure 3.8). Note the appearance of detectable levels of strand transfer products after 3 minutes in using E233V but not the wild type. This indicates that the mutation enhances a step in transposition that involves the core transpososome tetramer. We also sought to verify our assumption that E233V exerts its effect when in the dimer interface, i.e. when present on the catalytic subunit as shown in blue in Figure 3.7. This can be determined by taking advantage of the SinMu system (refer to chapters 2.3.2 and A1.1 for additional details) to direct the E233V mutant protein only to the catalytic positions. Here, too, we have found that hyperactivity is maintained (Figure 3.9), demonstrating that the dimer interface is the location at which E233V

and nearby mutations are acting. This effectively rules out a direct catalytic mechanism for E233V.

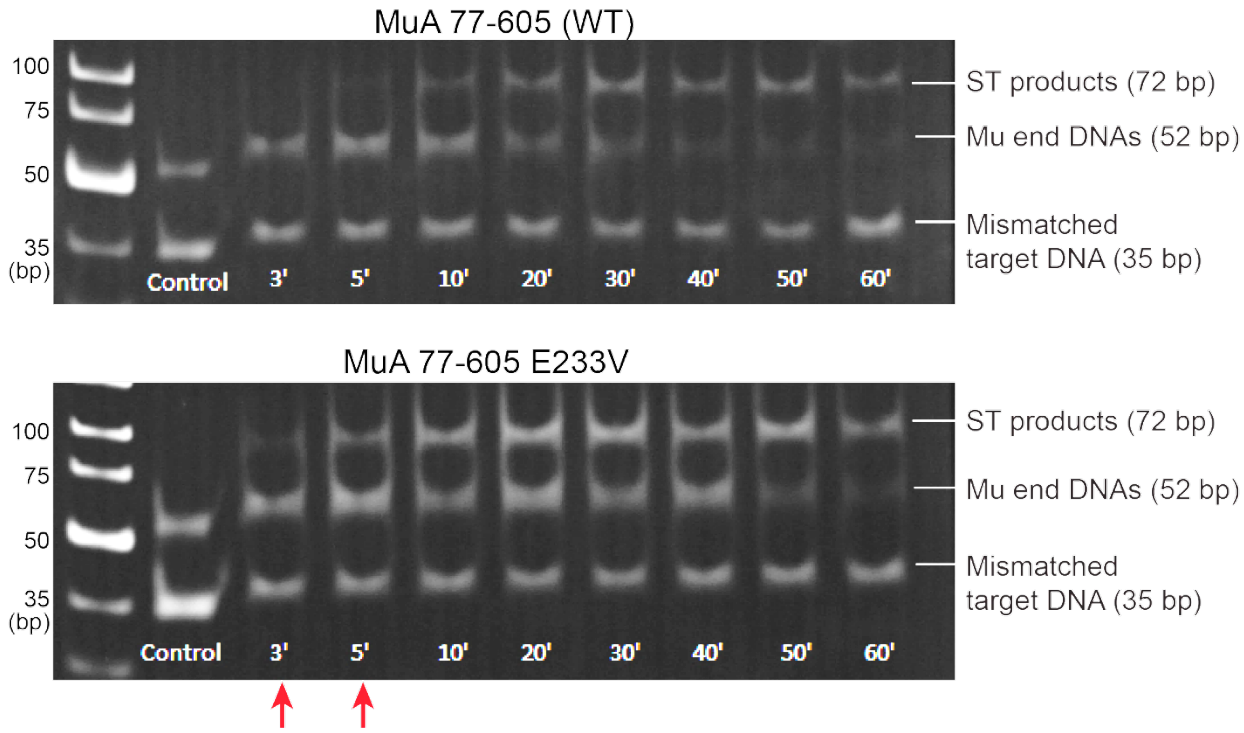

Figure 3.8 MuA E233V reaction kinetics.
The effect of the E233V mutation on transposition by MuA 77-605 constructs. Both DNA fragments and MuA protein were mixed at time = 0 minutes. Aliquots of the reaction were removed at the indicated time points and stripped of protein by phenol:chloroform extraction. The results were visualized by EtBr staining. Red arrows indicate early time points at which the mutant has accumulated much more strand transfer (ST) product.

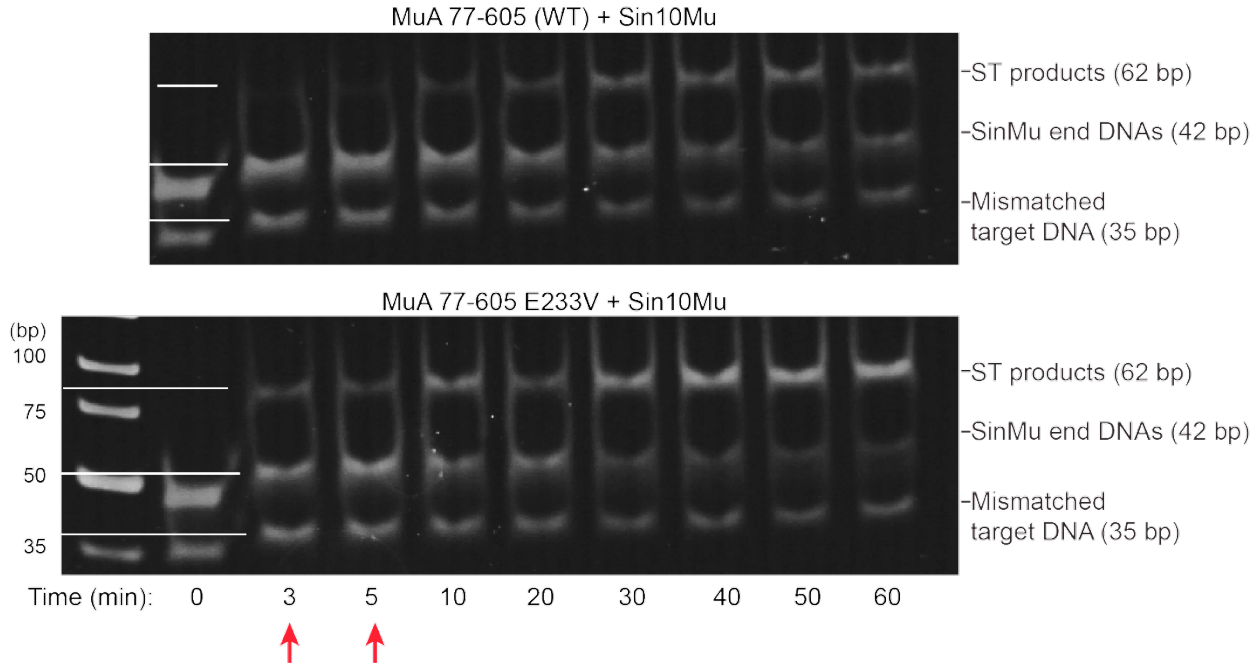

Figure 3.9 E233V hyperactivity in the SinMu system.
The hyperactivity of the E233V mutation persists when targeting specifically to the catalytic subunits positions using the SinMu system. Experiment is the same as described in Figure 3.4. Red arrows indicate early time points at which the mutant has accumulated much more strand transfer (ST) product.

We then began a search for the mechanism of E233V hyperactivity. It was possible that an improved dimer interface made transpososomes more stable, shifting the equilibrium between transpososome tetramers and lower-order complexes. Once formed, the core transpososome tetramer is known to be impressively resilient to disruption by urea, a chaotropic salt.[77] We have found that this property remains unchanged – and not enhanced – by the E233V mutation (Figure 3.10). There is no difference between the wildtype and mutant stability that would account for the sizeable enhancement in activity. It was also possible that the mutation altered the properties of the MuA monomer to destabilize an inhibitory conformation like that we proposed above. This could increase the rate of transpososome formation. However, we were not able to detect

any enlargement of the basic shape of MuA 77-605 constructs bearing the mutation by SAXS (Figures 3.2 and 3.3). Although this would not detect subtle changes in the MuA monomer structure, it does rule out gross changes to how the domains associate in the monomer state. Thus, it appears that hyperactivating dimer interface mutations do not exert their effect by making transpososomes significantly more stable or monomers significantly less stable.

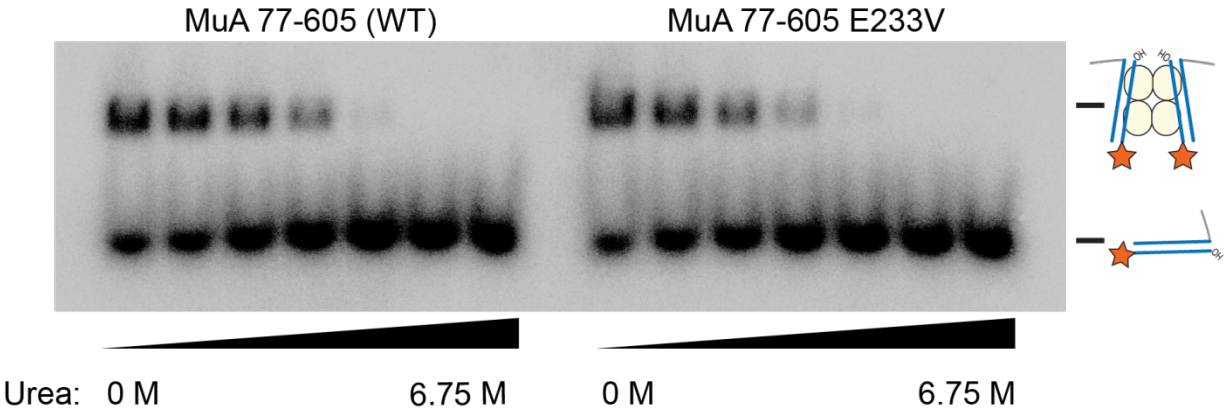

Figure 3.10: WT and E233V CDC stability.
EMSA analysis of the proportion of intact CDC transpososomes after being exposed to increasing concentrations of urea. The bottom band is consistent with the size of the Mu end DNAs alone. The top band represents transpososome tetramers surviving heparin challenge.

The E233V mutant does differ from the wildtype in one respect: the yield of transpososomes that are resilient to heparin. Heparin is an oligosaccharide substituted with negatively charged sulfate groups. It has been shown to remove loosely bound MuA molecules from both DNA and incompletely assembled transpososome. Only the core tetrameric subunits of transpososomes that have reached a conformation where the catalytic subunits are properly positioned for cleavage are resistant to heparin[73] (i.e., only after transitioning from the LER form to the SSC form, see chapter 1.6.2). The equilibrium level of heparin-resistant transpososomes

formed by the mutant, prior to addition of divalent metals, is significantly higher than wild type (Figure 3.11). This same effect can be recapitulated by supplementing the reaction buffer with 15% (v/v) DMSO, a known enhancer of transposition *in vitro*[44]. In the presence of DMSO, the difference in the amount of heparin sensitive transpososomes formed by WT versus E233V transpososomes is greatly reduced. It should be noted that this is how the transpososomes in Figure 3.10 were prepared: formation in DMSO followed by heparin challenge. This suggests that E233V affects transitions within the core transpososome tetramer that are also sensitive to DMSO.

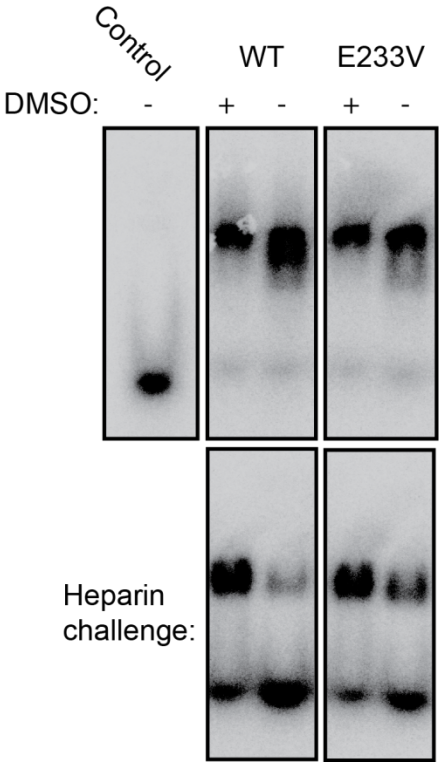

Figure 3.11: WT and E233V response to heparin and DMSO.
EMSA analysis of the relative proportion of Mu end DNA incorporated into transpososomes. This proportion is modulated by including DMSO in the reaction buffer, and/or challenge with

(**Figure 3.11**, continued) heparin immediately before separation. DMSO equalizes the amount of heparin-resistant complexes formed by WT vs. E233V MuA.

3.3 Discussion

Our results here illuminate the behavior of MuA during transpososome formation. First, we have used orthogonal methods to confirm the monomeric state of MuA prior to DNA binding. Our preliminary crosslinking results further show that the model proposed as a result of cryo-EM envelope fitting by Yuan *et al.*[103] is essentially correct: MuA is not only a monomer, but the DNA binding domains are docked closely to the catalytic domain. The docking of the DNA binding domains occurs on a different face of the catalytic domain than the dimer pair interface in the transpososome. We hypothesize that this contributes to keeping MuA monomeric prior to engaging DNA.

Somewhat unexpectedly, this alternative docking is robust to the removal of domain III. We had thought that domain III could hold the DNA binding domains and catalytic domain together in a conformation that would prevent dimerization (see the monomer conformation cartooned in Figure 3.11). This would fulfill the same protein-protein contacts it makes between the transpososome dimer pair. This is either not the case, or domain III is dispensable for keeping the dimer interface from forming. We conclude that the DNA binding domains and catalytic domain alone encode a (as yet unknown) docking interface that precludes the dimer interface.

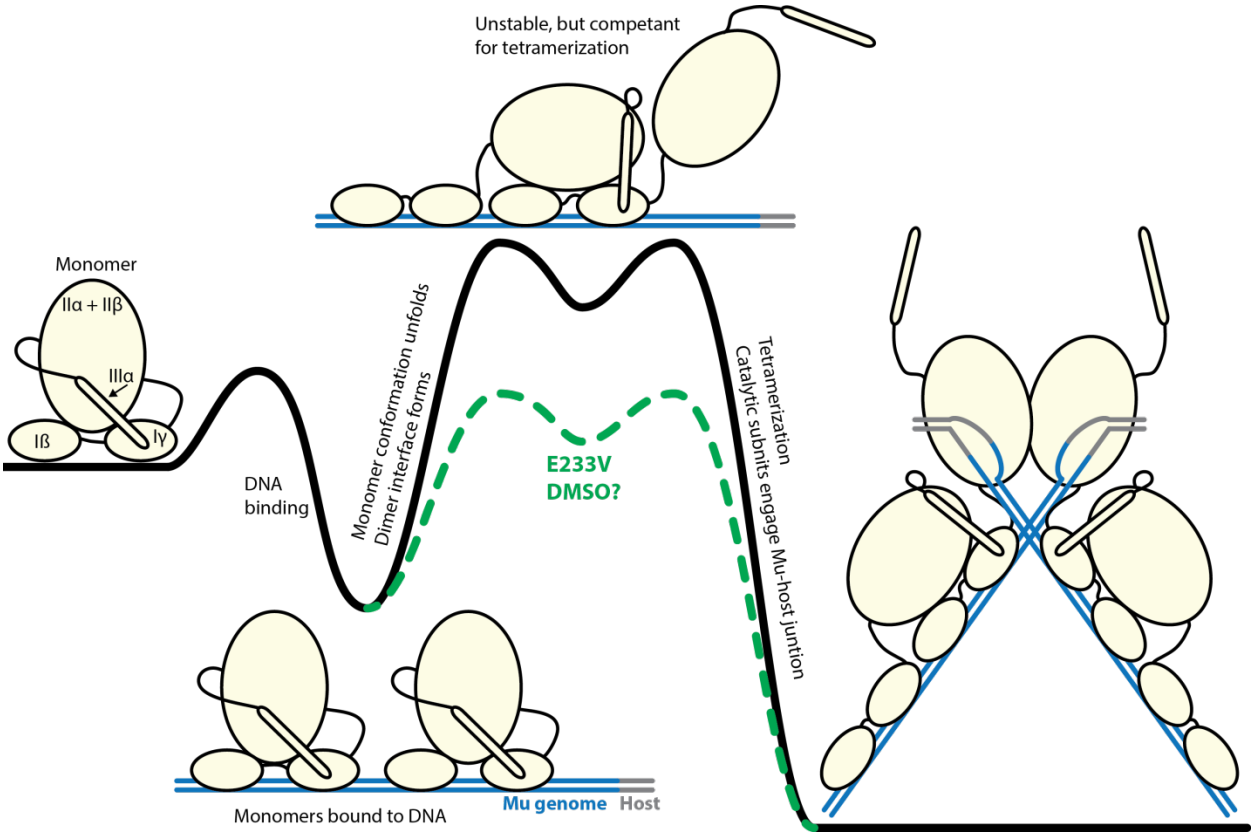

Figure 3.12: Dimerization, a slow step in stable transpososome formation.
My model for transpososome formation, presented as a free-energy diagram. DNA-binding is a favorable process in order to permit MuA to find its binding sites. However, dimerization of adjacent MuA subunits is unstable. If two pairs of dimers encounter each other, however, they can combine to become a stable transpososome complex.

Our current view of the transpososome assembly process – as it relates only to the core tetramer – is shown in Figure 3.11. The structure of the MuA monomer conformation at atomic resolution is still unknown. It does not appear to inhibit DNA binding, as MuA can be readily footprinted and used in EMSA experiments at isolated binding sites. The configuration proposed by Yuan *et al.*[103] is plausible given our data here. It is also supported by bend-phasing experiments that suggest the MuA bound to isolated sites induces a surprisingly strong (70 - 90°) DNA bend[127]. This bending would be required to dock DNA on the Yuan *et al.* monomer

structure if the monomer were to remain static. We remain skeptical of such a significant bend, however, as nothing in the transpososome crystal structure suggests that MuA is capable of this activity. Nevertheless, that multiple hyperactivating mutations localize along the dimer interface suggests that adjacent DNA-bound MuA subunits do not dimerize rapidly. It seems likely that they do not dimerize because the MuA monomer configuration persists after DNA binding. DNA bending may be involved in some fashion during the transition. This could explain why the DNA denaturant DMSO stimulates stable transpososome formation even when the Mu end DNA is pre-cleaved (and thus relatively little melting at the Mu termini is required).

A particularly slow step in transposition occurs between (1) the point at which MuA has formed a nucleoprotein complex synapsing the left and right ends of the genome with the enhancer, and (2) when it becomes a competent to cleave the Mu ends for transposition[61,77]. This transition coincides with the core transpososome tetramer becoming resistant to heat, salt, and heparin, and with melting of the DNA at the Mu-host junction. I envision that this slow step represents the unfolding of the MuA monomer conformation and the formation of the dimer interface. This could be energetically uphill because (1) the contacts from the monomer conformation are broken, but (2) the contacts that hold the two halves of the tetramer together have yet to form.

It is particularly compelling that E233V, a mutation that imparts a 55-fold increase in transposition efficiency, has not been incorporated by Mu and risen to fixation. Given that Mu depends on rapid replicative transposition during lytic growth, why has evolution not optimized the dimer interface? The answer may be that this conformational change is the key regulatory step before the transpososome begins to break and join whatever DNA it has bound. This slow

step would give a transiently synapsed but topologically incorrect transpososome time to dissociate before committing to transposition.

3.4 Materials and Methods

3.4.1 Protein expression and purification

The wildtype MuA 77-605 construct was created as described in [47]. The MuA 77-560 expression plasmid was created by PCR mutagenesis of the 77-605 expression plasmid to insert two stop codons after residue 560. The 88-560 expression plasmid was created by PCR mutagenesis of the 77-560 expression plasmid to delete the coding sequence of residues 77 through 87. Sin10Mu protein was a gift from Sherwin Montaño. All proteins were expressed purified as described in [49]. DNA substrates used are listed in Appendix A1.3. For assays involving strand transfer, the S35 including a base-pairing mismatch was used.

3.4.2 SAXS

SAXS data collection for MuA 77-605 WT and E233V was performed at the Bio-CAT beamline 18ID at the Advanced Photon Source (APS) synchrotron, Argonne National Laboratory. The concentration of protein samples was adjusted to 5 mg/mL in a buffer containing 25 mM HEPES pH 7.0, 150 mM NaCl, 10 mM MgCl2, 1 mM DTT, and 5% glycerol. 0.2 mL of each sample was applied over a Superdex 200 10/300 gel filtration column equilibrated in the same buffer, the output of which was directed to the capillary for x-ray

exposure. 20 frames of exposure immediately preceding the appearance of protein in the eluate were averaged together as a buffer blank. This buffer blank was subtracted from the 15 frames with the highest total scattering during protein elution. These ten curves were then scaled to the highest intensity curve using ALMERGE[130] and averaged together.

Data collection for MuA 88-560 and 77-560 was performed at beamline 12-ID-B, also at the Advanced Photon Source (APS) synchrotron, Argonne National Laboratory. Samples were dialyzed and diluted into the same buffer used above. Exposures were taken at 5 mg/mL, 2.5 mg/mL and 1 mg/mL while being agitated by a syringe pump. Matched buffer blanks were also recorded between each sample and concentration. 20 total exposures for each sample concentration and corresponding buffer blank were averaged together. The concentration series were then scaled to the highest concentration and averaged using ALMERGE, extrapolating to infinite dilution.

The radius of gyration for each protein was determined using AUTORG[131]. P(R) distance distributions were calculated using DATGNOM[131]. The D_{max} used for DATGNOM calculations was determined by hand. This was done by choosing the smallest distance that gave a reciprocal space fit lacking systemic deviations from the experimental data, a distance distribution lacking obvious oscillations, and a distance distribution that smoothly approached 0.0 at D_{max}. Distance distributions for atomic models were calculated explicitly using the atomic coordinates.

3.4.3 Crosslinking

Crosslinking was carried out in a buffer consisting of 50 mM HEPES pH 7.5, 150 mM NaCl, 10 mM MgCl2 and 2 mM DTT. MuA 88-560 was buffer exchanged and diluted to 4

mg/mL (74 µM) in this buffer using a P6 spin column. The crosslinker was suspended to 10 mM in DMSO, and aliquots were diluted to 500 µM in crosslinking buffer. For screening crosslinker:protein ratios, the protein concentration was held constant at 1.6 mg/mL (30 µM) and the crosslinker was added to achieve a 1:1, 1:2, 1:5 or 1:10 protein:crosslinker molar ratio. Crosslinking was allowed to proceed at room temperature for 12 hours before being quenched by the addition of Tris pH 7.5 to 20 mM.

For mass spectrometry analysis of crosslinks, the procedure above was performed on a 50 µL reaction of 2 mg/mL (37 µM) MuA 88-560 and 185 µM crosslinker. This sample was sent to the Donald Danforth Plant Science Center, St. Louis MO where it was analyzed by Dr. Bradley S. Evans in their proteomics facility, according to the published protocol for this crosslinker technology[128].

3.4.4 Transposition kinetics assays

Reactions consisted of 500 nM MuA protein construct and 250 nM each Mu end DNA and target DNA. These components were added simultaneously to initiate the reaction. These were mixed in a buffer of 25 mM HEPES pH 7.0, 150 mM NaCl, 10 mM MgCl2, 15% glycerol, 1 mM DTT and 0.6 mM zwittergent 3-12. Reactions were carried out at room temperature. For each indicated timepoint, an aliquot of the reaction was removed and immediately stopped by phenol:chloroform extraction to remove MuA. Timepoint samples were separated on 8% TBE PAGE by applying 115V for 1 hour, and DNA was visualized by ethidium bromide staining.

3.4.5 Transpososome stability assays

Cleaved donor complex transpososome formation reactions consisted of 2 µM nM MuA protein with 500 nM Mu end DNA in a buffer of 25 mM HEPES pH 7.0, 150 mM NaCl, 15% glycerol, 1 mMT DTT, and 0.6 mM zwittergent 3-12. Note that this buffer lacks Mg^{2+} to prevent strand transfer. The Mu end DNA was labeled at the 5'end of the transferred strand with ^{32}P. Transpososomes were allowed to form for 2 hours at room temperature. Transpososome samples were then diluted 5-fold into a buffer identical to the above, with the various modifications indicated. DMSO was used at 15% (v/v) concentration, and heparin at 50 µg/mL, where applicable. The transpososomes were equilibrated in this second buffer for 1 hour. Samples were then analyzed by electrophoretic mobility shift assay (EMSA), where they were separated over 5% TBE PAGE run for 1.5 hours at 140V. EMSA was performed with the gel and buffers pre-cooled to 4°C. For assaying stability versus urea, the transpososome formation buffer included 15% DMSO to increase transpososome yield. The dilution buffer then lacked DMSO, but included 50 µg/mL heparin and increasing concentrations of urea.

Chapter 4

Conclusions and Future Directions

The results I have presented here illuminate the behavior of MuA during two critical steps in Mu transposition, and suggest how those two steps are regulated. First, the initial formation of a stable, catalytically competent Mu transpososome has long been known to be a slow step in transposition. My data suggest that MuA acts as a compact monomer in solution and does not form dimers or higher order complexes in the absence of DNA, even at very high concentrations. These monomers must unfold and interact with each other at some point, however, in order to form the transpososome. This only occurs once they have bound DNA with the proper spacing and orientation, but even then the process is deliberately slow. I have shown here that the overall kinetics of the transposition process can be dramatically accelerated mutations along the protein-protein interface that forms between adjacent DNA-bound MuAs – mutations that are only one base-pair away in sequence. Later, after transpososome formation, interactions with potential target DNA are also weak. I have shown that the binding affinity between transpososomes and target DNA is rather poor, and this limits the rate of strand transfer. This interaction once again has the potential to be dramatically accelerated: by increasing the flexibility or providing inherently bent DNA, target DNA affinity changes by at least an order of magnitude and the strand transfer rate even more so.

These two steps in transposition have a common theme. The kinetics of both are slow. This must be because the energetic barrier to these transitions is high. Breaking the protein-protein interactions that hold a MuA monomer together or bending DNA over just tens of base pairs is

energetically uphill. However, these two steps are also nearly irreversible once completed, consistent with an even larger favorable free energy release on the other side. Thus, transpososomes are resilient to heat and salt challenge once formed, and strand transfer is almost never reverted despite being energetically unfavorable. Nevertheless, most of the changes I have made to perturb the system and lower the energetic barriers to transpososome formation and target capture/strand transfer do not have deleterious consequences *in vitro*: MuA E233V mutants do not dimerize before binding DNA, and target DNAs with mismatched base-pairs do not cause strand transfer to reverse significantly more often. Why, then, has evolution not optimized MuA to cross these energetic barrier more quickly? I hypothesize that these energetic barriers are important for regulation *in vivo* because crossing them entails a commitment to performing transposition chemistry. Kinetically slow transpososome formation would allow time for transiently and inappropriately DNA-bound MuA to dissociate before being incorporated into a transpososome. Poor target capture and slow strand transfer ensure that the mechanism that Mu uses to select the best target site, MuB, has a window in which to exert its effect. Furthermore, there is strong structural evidence that many transposable elements, including retroviruses, have converged upon DNA bending during target capture and strand transfer. I hypothesize that these transposable elements would behave similarly to Mu in the experiment outlined in chapter 2.

It was the crystal structure of the Mu transpososome that inspired and guided the work in this thesis, and future work on transpososome structures are necessary for further understanding. This work needs to go to both smaller and larger scales. Understanding the assembly of the Mu transpososome would be facilitated by a high resolution structure of the MuA monomer. In particular, this would guide the creation of mutant MuA proteins to better test hypotheses about

conformational rearrangements in the core transpososome tetramer subunits during assembly. I hope that my identification of residue 88 as the new minimal N-terminus for domain Iβ and my observation that the monomer structure is preserved without domain III will be helpful in these endeavors, particularly in crystallization. The molecular weight of the MuA protein makes it unsuitable for NMR, but if crystallographic attempts fail, it may be possible to resolve the full-length monomer by cryo-EM as that technology progresses. Alternatively, as SAXS software improves it may become even more feasible to continue my attempt to combine modeling with restrains derived from crosslinking to achieve an approximate structure.

On a larger scale, there are still things that could be learned from a structure of the transpososome similar to the currently available structure, but at higher resolution. In particular, it would be enlightening to improve our resolution of the area around the active site of the catalytic subunits. This would inform two outstanding questions: what is the path that the flanking host DNA takes to exit the transpososome, and how does the same active site catalyze two successive transesterification reactions? Another potential structural target would be a transpososome in which the MuA constructs include domain IIIβ. This would give us an understanding of how a transpososome interacts with MuB and ClpX differently than a MuA monomer.

Finally, on an even larger scale, a structure of a transpososome assembled on full left and right ends would provide a wealth of information. Such a structure would include all three binding sites on each end, and HU in the left end. There already exists such a wealth of biochemical data on this process, but with no structure upon which to map it. This complex would be of an ideal molecular weight and dimensions for the current state of cryo-EM.

Appendix 1

Protein and DNA substrate sequences

A1.1 Designing experiments and constructs in the SinMu system

The SinMu system was used in chapter 2.3. This system allows MuA constructs bearing different mutations to be targeted to either the structural (R2) or catalytic (R1) positions within an *in vitro* transpososome tetramer. This is accomplished by using two protein constructs to make the transpososome. The construct intended for the catalytic positions is a normal MuA construct, with the desired modifications for that position. For example, in chapter 2.3, this was a MuA construct ending at residue 560 and thus lacking domain III. The construct intended for the structural subunit positions is designed as a chimera. This chimera replaces the DNA binding domains of MuA (domains Iα through Iγ) with the (single) DNA binding domain from an unrelated recombinase, Sin. This is possible because the DNA binding domains from the structural subunits do not make any protein-protein contacts with any other components besides their DNA binding domains. The domains II and III from the structural subunits can, however, be modified as desired (although domain III is required at this position for transpososome assembly). The DNA fragments used in a SinMu experiment must be altered accordingly, to replace the R2 MuA binding site with the sequence of the Sin binding site. The appropriate spacing between the R1 and Sin sites was determined experimentally by Sherwin Montaño. My SinMu DNA substrate will be given in the following section. A model for a SinMu

transpososome based on a combination of structures of the Mu transpososome and DNA-bound Sin is given in Figure 2.6A.

A1.2 SinMu protein constructs

My work in chapter 2 used a construct I will refer to a *Sin15Mu663*. This designates that there are 15 residues (10 residues of a serine-glycine linker, and 5 residues of the original MuA linker) between the Sin DNA binding domain and the MuA catalytic domain, and that the C-terminal residue of the construct is (MuA) 663. All SinMu constructs begin with a 6xHis tag. This tag is dispensable for purification (SinMu constructs can be purified by the same protocol as MuA) but necessary for expression, I suspect because the Sin DNA binding domain is normally a C-terminal, rather than N-terminal, domain. The work in chapter 3 and the crystallographic trials I will describe in Appendix 2 used *Sin10Mu605*, which lacks domain IIIβ and the 5 native MuA residues N-terminal to domain IIα.

Included below are the amino acid sequences for Sin15Mu633 and Sin10Mu605. Yellow highlights the required 6xHis tag, cyan the Sin DNA binding domain, light green the serine-glycine linker, dark green the native Mu linker, and grey MuA domain II and III.

Sin15Mu663:

MHHHHHHGRPLLYSPNAKDPQKRVIYHRVVEMLEEGQAISKIAKEVNITRQTVYRIKHD
NGSGSGSGSGSGQQRTVEHLDAMQWINGDGYLHNVFVRWFNGDVIRPKTWFWQDVK
TRKILGWRCDVSENIDSIRLSFMDVVTRYGIPEDFHITIDNTRGAANKWLTGGAPNRYRF
KVKEDDPKGLFLLMGAKMHWTSVVAGKGWGQAKPVERAFGVGGLEEYVDKHPALA

GAYTGPNPQAKPDNYGDRAVDAELFLKTLAEGVAMFNARTGRETEMCGGKLSFDDVF
EREYARTIVRKPTEEQKRMLLLPAEAVNVSRKGEFTLKVGGSLKGAKNVYYNMALMN
AGVKKVVVRFDPQQLHSTVYCYTLDGRFICEAECLAPVAFNDAAAGREYRRRQKQLKS
ATKAAIKAQKQMDALEVAELLPQIAEPAAPESRIVGIFRPSGNTERVKNQERDDEYETER
DEYLNHSLDILEQNRRKKAI

Sin10Mu605:

MHHHHHHGRPLLYSPNAKDPQKRVIYHRVVEMLEEGQAISKIAKEVNITRQTVYRIKHD
NGSGSGSGSGSGEHLDAMQWINGDGYLHNVFVRWFNGDVIRPKTWFWQDVKTRKILG
WRCDVSENIDSIRLSFMDVVTRYGIPEDFHITIDNTRGAANKWLTGGAPNRYRFKVKED
DPKGLFLLMGAKMHWTSVVAGKGWGQAKPVERAFGVGGLEEYVDKHPALAGAYTGP
NPQAKPDNYGDRAVDAELFLKTLAEGVAMFNARTGRETEMCGGKLSFDDVFEREYAR
TIVRKPTEEQKRMLLLPAEAVNVSRKGEFALKVGGSLKGAKNVYYNMALMNAGVKKV
VVRFDPQQLHSTVYCYTLDGRFICEAECLAPVAFNDAAAGREYRRRQKQLKSATKAAI
KAQKQMDALEVAELLP

A1.3 DNA substrates

The same set of DNA substrates were used consistently throughout the biochemical work presented in this thesis. MuA transpososomes were assembled on using a Mu end DNA fragment identical to that used in the transpososome crystal structure. I refer to this as *KS* because it was first used by Kerren Swinger, a previous graduate student in the Phoebe Rice lab. The SinMu system Mu end fragment is also given below. In both cases, note that the transferred strand is 3

nt shorter than the non-transferred strand. This mimics the flanking host DNA and is required for transpososome assembly.

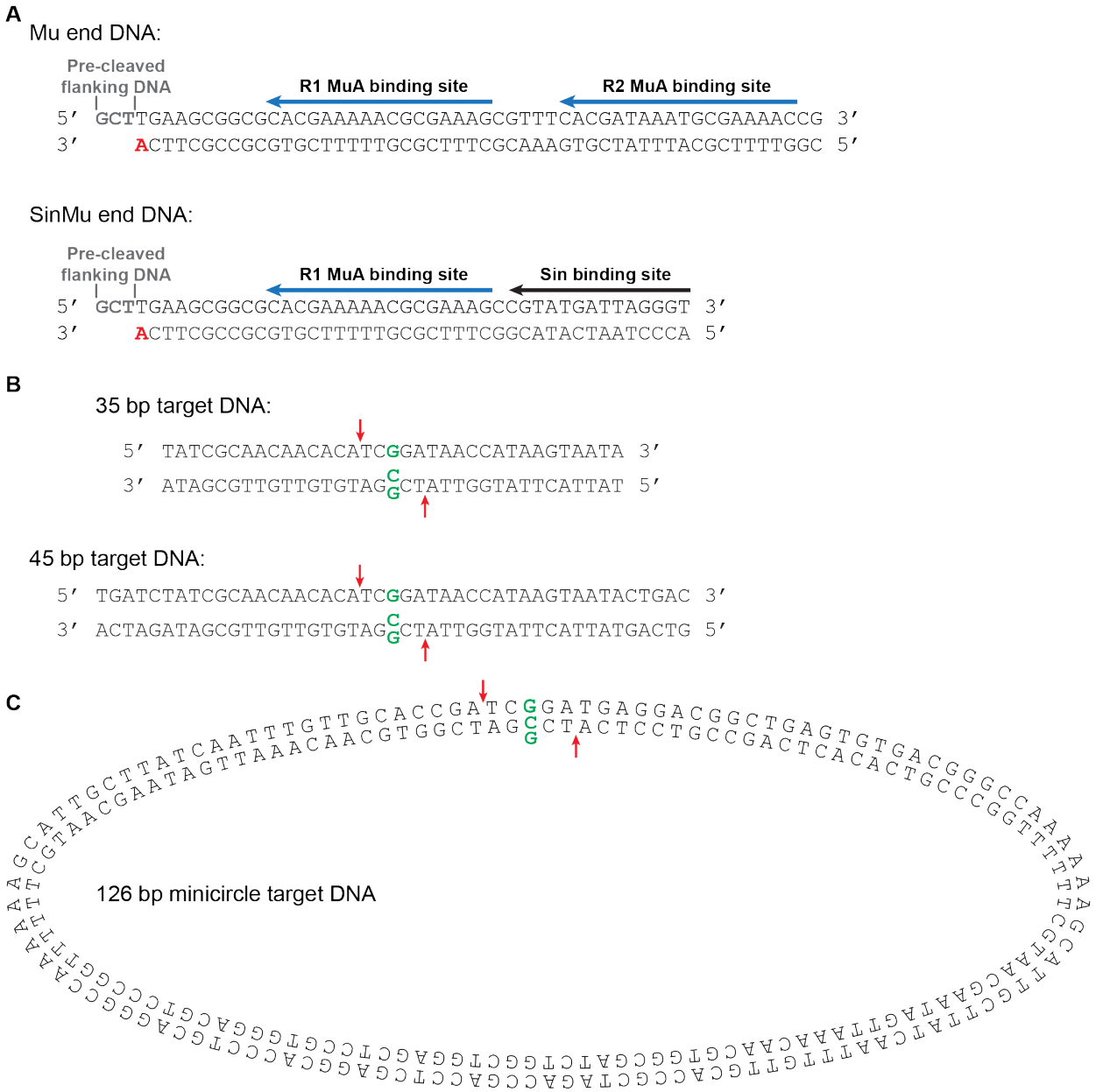

Figure A1.1: Mu end and target DNA fragments.
A. The Mu end DNA fragments used for transposition reactions. The bottom strand becomes joined to the target DNA during strand transfer. The bold red 3’ adenine provides the 3’ hydroxyl

(**Figure A1.1**, continued) group that is the nucleophile for this reaction. In fluorescence anisotropy experiments (chapter 2), this adenine lacked a 3' hydroxyl group to suppress strand transfer. Blue arrows mark the binding sites for MuA; the R2 binding site is replaced with a binding site for the Sin recombinase in the SinMu system.
B. Linear target DNA substrates. The G:G mismatch, when used, was placed in the center of the sequence by replacing a C in the bottom strand with a G, indicated in green. Red arrows mark the position of the predominant strand transfer product when the mismatch is present. The 45 bp target DNA was used in the heated disintegration experiments from chapter 2 to discourage melting of the strand transfer products during heating, and is identical in sequence to the 35 bp target used elsewhere except for 5 bp added at each end. Target DNAs were labeled at the 5' end of the top strand with ^{32}P for all radiographic experiments, and with the Atto565 fluorophore for fluorescence anisotropy.
C. Minicircle target DNA. This is the 126 bp circular target DNA used in chapter 2. The position of the optional G:G mismatched base pair is marked again in green, and the strand transfer attack positions it would produce are marked again with red arrows.

The linear target DNA fragment containing a central G:G mismatched base pair used in this work is also the same one used in the transpososome crystal structure. I will refer to it as *S35*.The mismatch is removed by changing only the bottom strand G to a C for proper base pairing. In chapter 2.5.6, this DNA was lengthened by adding 5 bp to either end to create *S45* in order to deter melting at high temperatures.

KS top (non-transferred)	GCTTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAACCG
KS bottom (transferred)	CGGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA
SinMu top (non-transferred)	GCTTGAAGCGGCGCACGAAAAACGCGAAAGCCGTATGATTAGGGT
SinMu bottom (transferred)	ACCCTAATCATACGGCTTTCGCGTTTTTCGTGCGCCGCTTCA
S35 top	TATCGCAACAACACATCGGATAACCATAAGTAATA
S35 bottom	TATTACTTATGGTTATCGGATGTGTTGTTGCGATA
S35 bottom (no mismatch)	TATTACTTATGGTTATCCGATGTGTTGTTGCGATA
S45 top	TGATCTATCGCAACAACACATCGGATAACCATAAGTAATACTGAC
S45 bottom	GTCAGTATTACTTATGGTTATCGGATGTGTTGTTGCGATAGATCA
S45 bottom (no mismatch)	GTCAGTATTACTTATGGTTATCCGATGTGTTGTTGCGATAGATCA

Table A1.1: Oligonucleotide sequences for the linear DNA fragments used in this work. All sequences are given in the 5' to 3' direction. KS and SinMu entries are pre-cleaved phage end DNAs. S35 and S45 are target DNAs.

The minicircle target DNAs used in chapter 2.3.2 were constructed by the circular ligation of two DNA fragments (designated A and B in TableA1.2) bearing 4 nt overhanging complementary ends. Their sequences are given below:

126Circle top A	AGGCACCCTGCAGGGCCAAAAAAGCATTGCTTATCAATTTGTTGCACCGATCGGATGAGGACGGC
126Circle top B	TGAGTGTGACGGGCCAAAAAAGCATTGCTTATCAATTTGTTGCACCGCTAGACCGACCTCG
126Circle bottom A	TCCTCATCGGATCGGTGCAACAAATTGATAAGCAATGCTTTTTTGGCCCTGCAGGGTGCCTCGAG
126Circle bottom A NoGG	TCCTCATCCGATCGGTGCAACAAATTGATAAGCAATGCTTTTTTGGCCCTGCAGGGTGCCTCGAG
126Circle bottom B	GTCGGTCTAGCGGTGCAACAAATTGATAAGCAATGCTTTTTTGGCCCGTCACACTCAGCCG

Table A1.2: Oligonucleotides used in the construction of DNA minicircles. All sequences are given in the 5' to 3' direction.

Appendix 2

MuA monomer and SinMu transpososome crystallization trials

A2.1 Introduction

The atomic structure of the MuA monomer (protein only), MuA bound to isolated binding sites, and the cleaved donor complex (CDC) form of the transpososome have all yet to be captured. The monomer and DNA-bound monomer forms of MuA are a worthy structural target because so little is known about them (refer to chapter 3 for a more detailed discussion). A structure of the CDC form of the transpososome would serve as an interesting comparison to the currently available strand transfer complex (STC) structure as a way to detect any structure changes that result from target DNA binding or attack. We were also driven to try the CDC because the B-factors for the STC structure are highest in and around the target DNA. We chose the SinMu system for CDC crystallization due to (1) its comparative simplicity (Sin is a single DNA binding domain vs. the two of WT MuA, and the Mu end DNA substrate is also shorter), (2) the observation that the Sin DNA-binding domain can mediate crystal packing contacts, and (3) so that we could remove domain III entirely from the catalytic subunits, which we suspected would not be stably associated with the rest of the transpososome prior to target DNA binding.

A2.2 MuA protein only

I have screened MuA constructs 77-605, 77-560, and 88-560 against the Hampton Research Crystal Screen 1 & 2, Index, and Natrix commercial kits, as well as a wide search through a matrix of $(NH_4)_2SO_4$ vs. PEG 3350. Not a single combination of the above produced

any promising crystallization hits. Although MuA protein is normally very soluble at high salt, when exposed to many crystallization conditions it often precipitates rapidly. If crystallography of MuA monomers is pursued in the future, I recommend it be diluted to 1 mg/mL or less prior to screening. Although this is well below the usual recommended concentration for crystallography, this has worked for other, unrelated proteins in the lab recently and could be explored for MuA.

A2.3 MuA bound to isolated binding sites

I have screened MuA 77-605 bound to a number of DNA fragments derived from the Mu genome R3 binding site. This site is judged to be one of the highest affinity of the six genome end MuA binding sites, given its strong footprinting signal.[34] We attempted to design substrates that, if packed end-to-end in a crystal, would not allow bound MuA proteins to form the dimers seen the transpososome. To form the protein-DNA complexes, protein and DNA were mixed to 30 µM concentration each in a buffer of 20 mM HEPES pH 7.0, 100 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, and 1 mM DTT, and incubated for 1 hour at room temperature. Regrettably, none of these produced any promising crystal hits. I have tried the following DNA fragments across the same screens listed in the previous section. For each pair, the first strand is given as 5' -> 3' and the second 3' -> 5'.

R3blunt:
ATCTGTTTCATTTGAAGCGCGAAAGCT
TAGACAAAGTAAACTTCGCGCTTTCGA

R3cgt:
TGCCTGTTTCATTTGAAGCGCGAAAGGC
 CGGACAAAGTAAACTTCGCGCTTTCCGA

R3cggc:

```
CGGCCTGTTTCATTTGAAGCGCGAAAGGC
  CGGACAAAGTAAACTTCGCGCTTTCCGGC
```

R3cgta-cggc

```
ATGCCTGTTTCATTTGAAGCGCGAAAGGC
  CGGACAAAGTAAACTTCGCGCTTTCCGGC
```

R3cggc-cgta

```
CGGCCTGTTTCATTTGAAGCGCGAAAGGC
  CGGACAAAGTAAACTTCGCGCTTTCCGTA
```

R3t

```
TCTGTTTCATTTGAAGCGCGAAAG
 GACAAAGTAAACTTCGCGCTTTCA
```

R3ta2

```
ATCTGTTTCATTTGAAGCGCGAAAG
  GACAAAGTAAACTTCGCGCTTTCTA
```

R3ta1

```
ATCTGTTTCATTTGAAGCGCGAAAGC
 AGACAAAGTAAACTTCGCGCTTTCGT
```

A2.4 SinMu CDC transpososome

Crystallization trials for CDC SinMu transpososomes were performed using MuA 77-560 and Sin10Mu605 constructs. Transpososomes were formed by mixing 19 µM each 77-560 and Sin10Mu605 and 14 µM Mu end DNA in a buffer of 25 mM HEPES pH 7.5, 120 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 0.6 mM zwittergent 3-12, 12% glycerol, and 10 mM DTT. This was incubated at room temperature for 1 hour. The following DNA constructs were screened against the Hampton Research Index screen. Each DNA fragment is constructed from four constituent oligonucleotides. For each set, the non-transferred strand is given first as 5’ -> 3’ and the

transferred strand as 3’ -> 5’. Note that the transferred strand lacks the terminal adenosine nucleotide, which should prevent unintended strand transfer.

S5MT4_1
5′ GCTTGAAGCGGCGCACGAAAAACGCGAAAGCCGTATGATTAGCGCG
CTTCGCCGCGTGCTTTTTGCGCTTTCGGCATACTAATCGC

S5MT4_2
5′ GCTTGAAGCGGCGCACGAAAAACGCGAAAGCCGTATGATTAGC
CTTCGCCGCGTGCTTTTTGCGCTTTCGGCATACTAATC
S5MT4_3
5′ GCTTGAAGCGGCGCACGAAAAACGCGAAAGCCGTATGATTAG
CTTCGCCGCGTGCTTTTTGCGCTTTCGGCATACTAATCGC

S5MT4_4
5′ GCTTGAAGCGGCGCACGAAAAACGCGAAAGCCGTATGATTAGC
CTTCGCCGCGTGCTTTTTGCGCTTTCGGCATACTAAT

S5MT4_5
5′ GCTTGAAGCGGCGCACGAAAAACGCGAAAGCCGTATGATTAG
CTTCGCCGCGTGCTTTTTGCGCTTTCGGCATACTAATC

S5MT4_3 produced particularly promising crystals in conditions around 1.4 M sodium phosphate / potassium phosphate, pH 7.2. Despite extensive optimization, including microseeding and dehydration, the best of these crystals diffracted x-rays only to about 7Å and so we deemed it not worth pursuing further. S5MT4_2, 3, and 5 also produced microcrystals in conditions containing neutral pH organic acids (sodium malonate, ammonium citrate, sodium citrate) and medium molecular weight PEGs (PEG 3350, PEG 5000MME). An additive screen identified dextran sulfate as a useful agent to encourage transpososome crystallization. High molecular weight dextran (M_r 9,000 to 20,000) in combination with PEG 20,000 in neutral pH buffers also produced small crystals with S5MT4_2, 4 and 5. These crystals remained small despite extensive optimization and did not diffract x-rays to less than about 20 Å resolution.

I also attempted crystallization of CDC SinMu transpososomes that included both strands of flanking host DNA. As detailed in chapter 4, the path of the flanking host DNA as it exits the transpososome is not yet known. We anticipated that a challenge to crystallizing this complex would be melting of the short strand of flanking host DNA that used to be part of the transferred strand. To prevent this, we designed DNA fragments for transpososome formation where the flanking host was capped with a highly stable DNA hairpin. These hairpin Mu ends are given below, 5' -> 3', with the hairpin highlighted in blue. The number in their names denotes the number of base-pairs between the hairpin and the Mu-host boundary.

JFHP_9

GCAGCTTGCGCGAAAGCGCAAGCTGCTGAAGCGGCGCACG

JFHP_11

GCAGCTTGCATGCGAAAGCATGCAAGCTGCTGAAGCGGCGCACG

JFHP_13

GCAGCTTGCATGCGCGAAAGCGCATGCAAGCTGCTGAAGCGGCGCACG

These were annealed with the previous SinMu oligonucleotides to pair each hairpin with the distal DNA ends of S5MT4_1, 3, 4, and 5. Transpososomes were formed on these DNA substrates exactly as above. Screening against the Hampton Research Index conditions set revealed a promising crystal form for the 11- and 13-bp hairpins in combination with the S5MT4_4 ends. The same crystal form (as judged visually) grew across a number of near-neutral pH conditions using either organic acids or $LiSO_4$ or $(NH_4)_2SO_4$ as the crystallization salt and a wide range of PEG molecular weights. After optimization, the largest crystals resulted from the 13 bp hairpin substrate in pH 6.5 – 7.0 organic acid buffers and 7 – 10% PEG 20,000. Only

crystals grown in sodium malonate or sodium citrate organic acids and frozen using glycerol as a cryo-protectant diffracted x-rays at all, and those only to a maximum of about 12 Å.

If further crystallization of transpososomes that include flanking host DNA is attempted, I would recommend screening flanking host DNA lengths of 11 bp or longer. The 9 bp fragments behaved poorly for crystallization across all conditions. If resolving the flanking host is the primary goal, it is my opinion that it would be prudent to abandon attempts for the CDC form of the transpososome and rely instead on the more stable STC. The CDC is still a worthy crystallographic target, but should be kept as simple as possible if pursued further. If this was attempted, I would recommend revisiting the microcrystal hits in organic acids and medium molecular weight PEGs mentioned above.

BIBLIOGRAHPY

1 Krejci, L., Altmannova, V., Spirek, M. & Zhao, X. Homologous recombination and its regulation. *Nucleic acids research* **40**, 5795-5818, doi:10.1093/nar/gks270 (2012).

2 Curcio, M. J. & Derbyshire, K. M. The outs and ins of transposition: from mu to kangaroo. *Nature reviews. Molecular cell biology* **4**, 865-877, doi:10.1038/nrm1241 (2003).

3 Grindley, N. D., Whiteson, K. L. & Rice, P. A. Mechanisms of site-specific recombination. *Annual review of biochemistry* **75**, 567-605, doi:10.1146/annurev.biochem.73.011303.073908 (2006).

4 Peters, J. E. & Craig, N. L. Tn7: smarter than we thought. *Nature reviews. Molecular cell biology* **2**, 806-814, doi:10.1038/35099006 (2001).

5 Craig, N. L. *et al. Mobile DNA III*.

6 Krupovic, M., Makarova, K. S., Forterre, P., Prangishvili, D. & Koonin, E. V. Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. *BMC biology* **12**, 36, doi:10.1186/1741-7007-12-36 (2014).

7 Hickman, A. B. & Dyda, F. The casposon-encoded Cas1 protein from Aciduliprofundum boonei is a DNA integrase that generates target site duplications. *Nucleic acids research* **43**, 10576-10587, doi:10.1093/nar/gkv1180 (2015).

8 Kitts, P. A. & Nash, H. A. Homology-Dependent Interactions in Phage-Lambda Site-Specific Recombination. *Nature* **329**, 346-348, doi:Doi 10.1038/329346a0 (1987).

9 NunesDuby, S. E., Yu, D. & Landy, A. Sensing homology at the strand-swapping step in lambda excisive recombination. *Journal of molecular biology* **272**, 493-508, doi:DOI 10.1006/jmbi.1997.1260 (1997).

10 Caparon, M. G. & Scott, J. R. Excision and Insertion of the Conjugative Transposon Tn916 Involves a Novel Recombination Mechanism. *Cell* **59**, 1027-1034, doi:Doi 10.1016/0092-8674(89)90759-9 (1989).

11 Trieucuot, P., Poyartsalmeron, C., Carlier, C. & Courvalin, P. Sequence Requirements for Target Activity in Site-Specific Recombination Mediated by the Int Protein of Transposon Tn-1545. *Molecular microbiology* **8**, 179-185, doi:DOI 10.1111/j.1365-2958.1993.tb01214.x (1993).

12 Boocock, M. R. & Rice, P. A. A proposed mechanism for IS607-family serine transposases. *Mobile DNA* **4**, doi:Unsp 24

10.1186/1759-8753-4-24 (2013).

13 Goodwin, T. J. D. & Poulter, R. T. M. The DIRS1 group of retrotransposons. *Mol Biol Evol* **18**, 2067-2082 (2001).

14 Cappello, J., Handelsman, K. & Lodish, H. F. Sequence of Dictyostelium Dirs-1 - an Apparent Retrotransposon with Inverted Terminal Repeats and an Internal Circle Junction Sequence. *Cell* **43**, 105-115, doi:Doi 10.1016/0092-8674(85)90016-9 (1985).

15 Garcillan-Barcia, M. D., Bernales, I., Mendiola, M. V. & de la Cruz, F. Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Molecular microbiology* **39**, 494-501 (2001).

16 Barabas, O. *et al.* Mechanism of IS200/IS605 family DNA transposases: Activation and transposon-directed target site selection. *Cell* **132**, 208-220, doi:10.1016/j.cell.2007.12.029 (2008).

17 Mendiola, M. V., Bernales, I. & Delacruz, F. Differential Roles of the Transposon Termini in Is91 Transposition. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 1922-1926, doi:DOI 10.1073/pnas.91.5.1922 (1994).

18 Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse Transcription of R2bm Rna Is Primed by a Nick at the Chromosomal Target Site - a Mechanism for Non-Ltr Retrotransposition. *Cell* **72**, 595-605, doi:Doi 10.1016/0092-8674(93)90078-5 (1993).

19 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

20 Mcclintock, B. The Origin and Behavior of Mutable Loci in Maize. *Proceedings of the National Academy of Sciences of the United States of America* **36**, 344-355, doi:DOI 10.1073/pnas.36.6.344 (1950).

21 Yuan, Y. W. & Wessler, S. R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7884-7889, doi:10.1073/pnas.1104208108 (2011).

22 Mizuuchi, K. In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. *Cell* **35**, 785-794 (1983).

23 Baker, T. A. & Luo, L. Identification of Residues in the Mu-Transposase Essential for Catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 6654-6658, doi:DOI 10.1073/pnas.91.14.6654 (1994).

24 Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. G. & Skalka, A. M. Residues Critical for Retroviral Integrative Recombination in a Region That Is Highly Conserved among

Retroviral Retrotransposon Integrases and Bacterial Insertion-Sequence Transposases. *Mol Cell Biol* **12**, 2331-2338 (1992).

25 Garfinkel, D. J., Boeke, J. D. & Fink, G. R. Ty Element Transposition - Reverse-Transcriptase and Virus-Like Particles. *Cell* **42**, 507-517, doi:Doi 10.1016/0092-8674(85)90108-4 (1985).

26 Fujiwara, T. & Mizuuchi, K. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**, 497-504 (1988).

27 Harshey, R. M. The Mu story: how a maverick phage moved the field forward. *Mobile DNA* **3**, 21, doi:10.1186/1759-8753-3-21 (2012).

28 Taylor, A. L. Bacteriophage-Induced Mutation in Escherichia Coli. *Proceedings of the National Academy of Sciences of the United States of America* **50**, 1043-1051 (1963).

29 Bukhari, A. I. Reversal of Mutator Phage Mu-Integration. *Journal of molecular biology* **96**, 87-&, doi:Doi 10.1016/0022-2836(75)90183-7 (1975).

30 Ljungquist, E. & Bukhari, A. I. State of Prophage Mu DNA Upon Induction. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 3143-3147, doi:DOI 10.1073/pnas.74.8.3143 (1977).

31 Shapiro, J. A. Molecular-Model for the Transposition and Replication of Bacteriophage Mu and Other Transposable Elements. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 1933-1937, doi:DOI 10.1073/pnas.76.4.1933 (1979).

32 Rice, P. & Mizuuchi, K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**, 209-220 (1995).

33 Harshey, R. M. Switch in the transposition products of Mu DNA mediated by proteins: Cointegrates versus simple insertions. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 2012-2016 (1983).

34 Craigie, R., Mizuuchi, M. & Mizuuchi, K. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. *Cell* **39**, 387-394 (1984).

35 Chaconas, G. *et al.* In vitro and in vivo manipulations of bacteriophage Mu DNA: cloning of Mu ends and construction of mini-Mu's carrying selectable markers. *Gene* **13**, 37-46 (1981).

36 Harshey, R. M. & Bukhari, A. I. Infecting bacteriophage mu DNA forms a circular DNA-protein complex. *Journal of molecular biology* **167**, 427-441 (1983).

37 Harshey, R. M. Nonreplicative DNA transposition: integration of infecting bacteriophage mu. *Cold Spring Harbor symposia on quantitative biology* **49**, 273-278 (1984).

38 Au, T. K., Agrawal, P. & Harshey, R. M. Chromosomal integration mechanism of infecting mu virion DNA. *Journal of bacteriology* **188**, 1829-1834, doi:10.1128/JB.188.5.1829-1834.2006 (2006).

39 Harshey, R. M. Transposable Phage Mu. *Microbiology spectrum* **2**, doi:10.1128/microbiolspec.MDNA3-0007-2014 (2014).

40 Mizuuchi, K. & Adzuma, K. Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for a one-step transesterification mechanism. *Cell* **66**, 129-140 (1991).

41 Mizuuchi, K. Polynucleotidyl transfer reactions in transpositional DNA recombination. *The Journal of biological chemistry* **267**, 21273-21276 (1992).

42 Mizuuchi, M., Baker, T. A. & Mizuuchi, K. Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. *Cell* **70**, 303-311 (1992).

43 Wang, Z., Namgoong, S. Y., Zhang, X. & Harshey, R. M. Kinetic and structural probing of the precleavage synaptic complex (type 0) formed during phage Mu transposition. Action of metal ions and reagents specific to single-stranded DNA. *The Journal of biological chemistry* **271**, 9619-9626 (1996).

44 Savilahti, H., Rice, P. A. & Mizuuchi, K. The phage Mu transpososome core: DNA requirements for assembly and function. *The EMBO journal* **14**, 4893-4903 (1995).

45 Nakayama, C., Teplow, D. B. & Harshey, R. M. Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 1809-1813 (1987).

46 Clubb, R. T. *et al.* A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase. *Structure* **2**, 1041-1048 (1994).

47 Clubb, R. T., Schumacher, S., Mizuuchi, K., Gronenborn, A. M. & Clore, G. M. Solution structure of the I gamma subdomain of the Mu end DNA-binding domain of phage Mu transposase. *Journal of molecular biology* **273**, 19-25, doi:10.1006/jmbi.1997.1312 (1997).

48 Schumacher, S. *et al.* Solution structure of the Mu end DNA-binding ibeta subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. *The EMBO journal* **16**, 7532-7541, doi:10.1093/emboj/16.24.7532 (1997).

49 Montano, S. P., Pigli, Y. Z. & Rice, P. A. The mu transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**, 413-417, doi:10.1038/nature11602 (2012).

50 Levchenko, I., Luo, L. & Baker, T. A. Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes & development* **9**, 2399-2408 (1995).

51 Baker, T. A., Mizuuchi, M. & Mizuuchi, K. MuB protein allosterically activates strand transfer by the transposase of phage Mu. *Cell* **65**, 1003-1013 (1991).

52 Craigie, R., Arndt-Jovin, D. J. & Mizuuchi, K. A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 7570-7574 (1985).

53 Zou, A. H., Leung, P. C. & Harshey, R. M. Transposase contacts with mu DNA ends. *The Journal of biological chemistry* **266**, 20476-20482 (1991).

54 Leung, P. C., Teplow, D. B. & Harshey, R. M. Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature* **338**, 656-658, doi:10.1038/338656a0 (1989).

55 Surette, M. G., Lavoie, B. D. & Chaconas, G. Action at a distance in Mu DNA transposition: an enhancer-like element is the site of action of supercoiling relief activity by integration host factor (IHF). *The EMBO journal* **8**, 3483-3489 (1989).

56 Krause, H. M., Rothwell, M. R. & Higgins, N. P. The early promoter of bacteriophage Mu: definition of the site of transcript initiation. *Nucleic acids research* **11**, 5483-5495 (1983).

57 Mizuuchi, M. & Mizuuchi, K. Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. *Cell* **58**, 399-408 (1989).

58 Krause, H. M. & Higgins, N. P. Positive and negative regulation of the Mu operator by Mu repressor and Escherichia coli integration host factor. *The Journal of biological chemistry* **261**, 3744-3752 (1986).

59 Pato, M. L. & Banerjee, M. Genetic analysis of the strong gyrase site (SGS) of bacteriophage Mu: localization of determinants required for promoting Mu replication. *Molecular microbiology* **37**, 800-810 (2000).

60 Lavoie, B. D., Chan, B. S., Allison, R. G. & Chaconas, G. Structural aspects of a higher order nucleoprotein complex: induction of an altered DNA structure at the Mu-host junction of the Mu type 1 transpososome. *The EMBO journal* **10**, 3051-3059 (1991).

61 Mizuuchi, M., Baker, T. A. & Mizuuchi, K. DNase protection analysis of the stable synaptic complexes involved in Mu transposition. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 9031-9035 (1991).

62 Saha, R. P., Lou, Z., Meng, L. & Harshey, R. M. Transposable prophage Mu is organized as a stable chromosomal domain of E. coli. *PLoS genetics* **9**, e1003902, doi:10.1371/journal.pgen.1003902 (2013).

63 Oram, M. & Pato, M. L. Mu-like prophage strong gyrase site sequences: analysis of properties required for promoting efficient mu DNA replication. *Journal of bacteriology* **186**, 4575-4584, doi:10.1128/JB.186.14.4575-4584.2004 (2004).

64 Pathania, S., Jayaram, M. & Harshey, R. M. A unique right end-enhancer complex precedes synapsis of Mu ends: the enhancer is sequestered within the transpososome throughout transposition. *The EMBO journal* **22**, 3725-3736, doi:10.1093/emboj/cdg354 (2003).

65 Pathania, S., Jayaram, M. & Harshey, R. M. Path of DNA within the Mu transpososome. Transposase interactions bridging two Mu ends and the enhancer trap five DNA supercoils. *Cell* **109**, 425-436 (2002).

66 Kobryn, K., Watson, M. A., Allison, R. G. & Chaconas, G. The Mu three-site synapse: a strained assembly platform in which delivery of the L1 transposase binding site triggers catalytic commitment. *Molecular cell* **10**, 659-669 (2002).

67 Lee, I. & Harshey, R. M. The conserved CA/TG motif at Mu termini: T specifies stable transpososome assembly. *Journal of molecular biology* **330**, 261-275 (2003).

68 Watson, M. A. & Chaconas, G. Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. *Cell* **85**, 435-445 (1996).

69 Lee, I. & Harshey, R. M. Patterns of sequence conservation at termini of long terminal repeat (LTR) retrotransposons and DNA transposons in the human genome: lessons from phage Mu. *Nucleic acids research* **31**, 4531-4540 (2003).

70 Li, M., Ivanov, V., Mizuuchi, M., Mizuuchi, K. & Craigie, R. DNA requirements for assembly and stability of HIV-1 intasomes. *Protein science : a publication of the Protein Society* **21**, 249-257, doi:10.1002/pro.2010 (2012).

71 Lilley, D. M. *DNA-protein: structural interactions*. (IRL Press Ltd., 1995).

72 El Hassan, M. A. & Calladine, C. R. Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **355**, 43-100, doi:10.1098/rsta.1997.0002 (1997).

73 Yanagihara, K. & Mizuuchi, K. Progressive structural transitions within Mu transpositional complexes. *Molecular cell* **11**, 215-224 (2003).

74 Savilahti, H. & Mizuuchi, K. Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. *Cell* **85**, 271-280 (1996).

75 Naigamwalla, D. Z. & Chaconas, G. A new set of Mu DNA transposition intermediates: alternate pathways of target capture preceding strand transfer. *The EMBO journal* **16**, 5227-5234, doi:10.1093/emboj/16.17.5227 (1997).

76 Lemberg, K. M., Schweidenback, C. T. & Baker, T. A. The dynamic Mu transpososome: MuB activation prevents disintegration. *Journal of molecular biology* **374**, 1158-1171, doi:10.1016/j.jmb.2007.09.079 (2007).

77 Surette, M. G., Buch, S. J. & Chaconas, G. Transpososomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. *Cell* **49**, 253-262 (1987).

78 Abdelhakim, A. H., Sauer, R. T. & Baker, T. A. The AAA+ ClpX machine unfolds a keystone subunit to remodel the Mu transpososome. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 2437-2442, doi:10.1073/pnas.0910905106 (2010).

79 Abdelhakim, A. H., Oakes, E. C., Sauer, R. T. & Baker, T. A. Unique contacts direct high-priority recognition of the tetrameric Mu transposase-DNA complex by the AAA+ unfoldase ClpX. *Molecular cell* **30**, 39-50, doi:10.1016/j.molcel.2008.02.013 (2008).

80 Levchenko, I., Yamauchi, M. & Baker, T. A. ClpX and MuB interact with overlapping regions of Mu transposase: implications for control of the transposition pathway. *Genes & development* **11**, 1561-1572 (1997).

81 Ling, L., Montano, S. P., Sauer, R. T., Rice, P. A. & Baker, T. A. Deciphering the Roles of Multicomponent Recognition Signals by the AAA+ Unfoldase ClpX. *Journal of molecular biology* **427**, 2966-2982, doi:10.1016/j.jmb.2015.03.008 (2015).

82 North, S. H. & Nakai, H. Host factors that promote transpososome disassembly and the PriA-PriC pathway for restart primosome assembly. *Molecular microbiology* **56**, 1601-1616, doi:10.1111/j.1365-2958.2005.04639.x (2005).

83 Wu, Z. & Chaconas, G. A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage. *The EMBO journal* **14**, 3835-3843 (1995).

84 Choi, W. & Harshey, R. M. DNA repair by the cryptic endonuclease activity of Mu transposase. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 10014-10019, doi:10.1073/pnas.0912615107 (2010).

85 Choi, W., Jang, S. & Harshey, R. M. Mu transpososome and RecBCD nuclease collaborate in the repair of simple Mu insertions. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 14112-14117, doi:10.1073/pnas.1407562111 (2014).

86 Jang, S., Sandler, S. J. & Harshey, R. M. Mu insertions are repaired by the double-strand break repair pathway of Escherichia coli. *PLoS genetics* **8**, e1002642, doi:10.1371/journal.pgen.1002642 (2012).

87 Jang, S. & Harshey, R. M. Repair of transposable phage Mu DNA insertions begins only when the E. coli replisome collides with the transpososome. *Molecular microbiology* **97**, 746-758, doi:10.1111/mmi.13061 (2015).

88 Haapa-Paananen, S., Rita, H. & Savilahti, H. DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. *The Journal of biological chemistry* **277**, 2843-2851, doi:10.1074/jbc.M108044200 (2002).

89 Mizuno, N. *et al.* MuB is an AAA+ ATPase that forms helical filaments to control target selection for DNA transposition. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E2441-2450, doi:10.1073/pnas.1309499110 (2013).

90 Mizuuchi, M. & Mizuuchi, K. Conformational isomerization in phage Mu transpososome assembly: effects of the transpositional enhancer and of MuB. *The EMBO journal* **20**, 6927-6935, doi:10.1093/emboj/20.23.6927 (2001).

91 Maxwell, A., Craigie, R. & Mizuuchi, K. B protein of bacteriophage mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 699-703 (1987).

92 Erzberger, J. P., Mott, M. L. & Berger, J. M. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nature structural & molecular biology* **13**, 676-683, doi:10.1038/nsmb1115 (2006).

93 Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* **135**, 623-634, doi:10.1016/j.cell.2008.09.058 (2008).

94 Greene, E. C. & Mizuuchi, K. Visualizing the assembly and disassembly mechanisms of the MuB transposition targeting complex. *The Journal of biological chemistry* **279**, 16736-16743, doi:10.1074/jbc.M311883200 (2004).

95 Tan, X., Mizuuchi, M. & Mizuuchi, K. DNA transposition target immunity and the determinants of the MuB distribution patterns on DNA. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 13925-13929, doi:10.1073/pnas.0706564104 (2007).

96 Greene, E. C. & Mizuuchi, K. Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. *The EMBO journal* **21**, 1477-1486, doi:10.1093/emboj/21.6.1477 (2002).

97 Dramicanin, M., Lopez-Mendez, B., Boskovic, J., Campos-Olivas, R. & Ramon-Maiques, S. The N-terminal domain of MuB protein has striking structural similarity to DNA-binding domains and mediates MuB filament-filament interactions. *Journal of structural biology* **191**, 100-111, doi:10.1016/j.jsb.2015.07.004 (2015).

98 Greene, E. C. & Mizuuchi, K. Target immunity during Mu DNA transposition. Transpososome assembly and DNA looping enhance MuA-mediated disassembly of the MuB target complex. *Molecular cell* **10**, 1367-1378 (2002).

99 Han, Y. W. & Mizuuchi, K. Phage Mu transposition immunity: protein pattern formation along DNA by a diffusion-ratchet mechanism. *Molecular cell* **39**, 48-58, doi:10.1016/j.molcel.2010.06.013 (2010).

100 Ge, J., Lou, Z. & Harshey, R. M. Immunity of replicating Mu to self-integration: a novel mechanism employing MuB protein. *Mobile DNA* **1**, 8, doi:10.1186/1759-8753-1-8 (2010).

101 Yin, Z. & Harshey, R. M. Enhancer-independent Mu transposition from two topologically distinct synapses. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18884-18889, doi:10.1073/pnas.0506873102 (2005).

102 Aldaz, H., Schuster, E. & Baker, T. A. The interwoven architecture of the Mu transposase couples DNA synapsis to catalysis. *Cell* **85**, 257-269 (1996).

103 Yuan, J. F., Beniac, D. R., Chaconas, G. & Ottensmeyer, F. P. 3D reconstruction of the Mu transposase and the Type 1 transpososome: a structural framework for Mu DNA transposition. *Genes & development* **19**, 840-852, doi:10.1101/gad.1291405 (2005).

104 Aziz, R. K., Breitbart, M. & Edwards, R. A. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic acids research* **38**, 4207-4217, doi:10.1093/nar/gkq140 (2010).

105 Li, M., Mizuuchi, M., Burke, T. R., Jr. & Craigie, R. Retroviral DNA integration: reaction pathway and critical intermediates. *The EMBO journal* **25**, 1295-1304, doi:10.1038/sj.emboj.7601005 (2006).

106 Mizuuchi, K. Mechanism of transposition of bacteriophage Mu: polarity of the strand transfer reaction at the initiation of transposition. *Cell* **39**, 395-404 (1984).

107 Pryciak, P. M. & Varmus, H. E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69**, 769-780 (1992).

108 Au, T. K., Pathania, S. & Harshey, R. M. True reversal of Mu integration. *The EMBO journal* **23**, 3408-3420, doi:10.1038/sj.emboj.7600344 (2004).

109 Mizuuchi, M., Rice, P. A., Wardle, S. J., Haniford, D. B. & Mizuuchi, K. Control of transposase activity within a transpososome by the configuration of the flanking DNA segment of the transposon. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14622-14627, doi:10.1073/pnas.0706556104 (2007).

110 Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326-329, doi:10.1038/nature09517 (2010).

111 Morris, E. R., Grey, H., McKenzie, G., Jones, A. C. & Richardson, J. M. A bend, flip and trap mechanism for transposon integration. *eLife* **5**, doi:10.7554/eLife.15537 (2016).

112 Yin, Z. *et al.* Crystal structure of the Rous sarcoma virus intasome. *Nature* **530**, 362-366, doi:10.1038/nature16950 (2016).

113 Harshey, R. M. & Bukhari, A. I. A mechanism of DNA transposition. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 1090-1094 (1981).

114 Escara, J. F. & Hutton, J. R. Thermal stability and renaturation of DNA in dimethyl sulfoxide solutions: acceleration of the renaturation rate. *Biopolymers* **19**, 1315-1327, doi:10.1002/bip.1980.360190708 (1980).

115 Herrera, J. E. & Chaires, J. B. A premelting conformational transition in poly(dA)-Poly(dT) coupled to daunomycin binding. *Biochemistry* **28**, 1993-2000 (1989).

116 Guest, C. R., Hochstrasser, R. A., Sowers, L. C. & Millar, D. P. Dynamics of mismatched base pairs in DNA. *Biochemistry* **30**, 3271-3279 (1991).

117 Rossetti, G. *et al.* The structural impact of DNA mismatches. *Nucleic acids research* **43**, 4309-4321, doi:10.1093/nar/gkv254 (2015).

118 Marathias, V. M., Jerkovic, B., Arthanari, H. & Bolton, P. H. Flexibility and curvature of duplex DNA containing mismatched sites as a function of temperature. *Biochemistry* **39**, 153-160 (2000).

119 Yanagihara, K. & Mizuuchi, K. Mismatch-targeted transposition of Mu: a new strategy to map genetic polymorphism. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11317-11321, doi:10.1073/pnas.132403399 (2002).

120 Schmid, S., Berger, B. & Haas, D. Target joining of duplicated insertion sequence IS21 is assisted by IstB protein in vitro. *Journal of bacteriology* **181**, 2286-2289 (1999).

121 Peters, J. E. & Craig, N. L. Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes & development* **15**, 737-747, doi:10.1101/gad.870201 (2001).

122 Arias-Palomo, E. & Berger, J. M. An Atypical AAA+ ATPase Assembly Controls Efficient Transposition through DNA Remodeling and Transposase Recruitment. *Cell* **162**, 860-871, doi:10.1016/j.cell.2015.07.037 (2015).

123 Ge, J., Lou, Z., Cui, H., Shang, L. & Harshey, R. M. Analysis of phage Mu DNA transposition by whole-genome Escherichia coli tiling arrays reveals a complex relationship to distribution of target selection protein B, transcription and chromosome architectural elements. *Journal of biosciences* **36**, 587-601 (2011).

124 Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366-369, doi:10.1038/nature14495 (2015).

125 Chaconas, G., Harshey, R. M., Sarvetnick, N. & Bukhari, A. I. Predominant end-products of prophage Mu DNA transposition during the lytic cycle are replicon fusions. *Journal of molecular biology* **150**, 341-359 (1981).

126 Baker, T. A. & Mizuuchi, K. DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes & development* **6**, 2221-2232 (1992).

127 Kuo, C. F., Zou, A. H., Jayaram, M., Getzoff, E. & Harshey, R. DNA-protein complexes during attachment-site synapsis in Mu DNA transposition. *The EMBO journal* **10**, 1585-1591 (1991).

128 Petrotchenko, E. V., Serpa, J. J. & Borchers, C. H. An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Molecular & cellular proteomics : MCP* **10**, M110 001420, doi:10.1074/mcp.M110.001420 (2011).

129 Pajunen, M. I. *et al.* Universal platform for quantitative analysis of DNA transposition. *Mobile DNA* **1**, 24, doi:10.1186/1759-8753-1-24 (2010).

130 Franke, D., Kikhney, A. G. & Svergun, D. I. Automated acquisition and analysis of small angle X-ray scattering data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **689**, 52-59, doi:http://dx.doi.org/10.1016/j.nima.2012.06.008 (2012).

131 Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *Journal of Applied Crystallography* **40**, s223-s228, doi:doi:10.1107/S0021889807002853 (2007).