Dynamics of Protein-DNA Interactions in a Mu Transpososome

Ilana Michal Goldhaber-Gordon

A.B. Biology
Harvard University, 1995

Submitted to the Department of Biology in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biology
at the
Massachusetts Institute of Technology

September 2002

Signature of author: _____
Department of Biology
August 6, 2002

Certified by: _____
Tania A. Baker
Professor of Biology
Thesis Supervisor

Accepted by: _____
Alan D. Grossman
Professor of Biology
Chairman, graduate committee

Dynamics of Protein-DNA Interactions in a Mu Transpososome

Ilana Goldhaber-Gordon

Submitted to the Department of Biology on August 9, 2002 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biology.

ABSTRACT

Transposition allows movement of a defined stretch of DNA, a transposon, from one DNA location to another. This process is required for the life-cycles of many viruses, from bacteriophage Mu to HIV; it is spreading antibiotic resistances between bacterial populations; and it is responsible for spontaneous mutations in all the kingdoms of life. Transposition is mediated by a protein, the transposase, encoded by the transposon. DNA sequence signals at the two ends of the transposon activate assembly of a transpososome: a complex that include multiple copies of transposase plus both transposon ends. Transpososome assembly, in turn, activates the DNA cleavage and joining reactions required for transposition. This thesis explores aspects of interactions between one transposase, MuA, and the ends of its transposon DNA, the genome of bacteriophage Mu. The first chapter provides an overview of Mu transposition, with special emphasis on the transpososome. The second chapter shows that in the absence of two proper transposon end sequences, an unrelated sequence can substitute for one of the two ends. This leads to some models about the process of transpososome assembly. The third chapter describes and analyzes 18 specific sequences that substituted for one transposon end. The fourth chapter shows that sequence specificity contributes primarily to the initial stage of transposition, transpososome assembly, rather than to the DNA cleavage and joining steps. The fifth chapter shows that the very last nucleotide of the transposon helps maintain the distinction between the two reactions, DNA cleavage and DNA joining.

Thesis supervisor: Tania A. Baker
Title: Professor of Biology

Dedication

This thesis is dedicated to the memory of **Esther Pasal HaLevy**, z"l. Her strength made her the center of her family, into which she welcomed me warmly whenever I was in Israel, and I leaned on her heavily. She was nurturing and accepting, no matter which way I grew. Her black eyes and beautiful English, her *kishuim* (squash), *levivot* (latkes) and *ugat shmarim* (chocolate yeast cake), her comfortable laugh, her strong observations and easy chatter, are now contained only in her name. Her sudden death tears a piece of my soul.

And to the memory of my teacher, **Dr. Roger Coleson** z"l, who would have been proud.

Acknowledgments

I am excited to be thanking many of the people who have touched me over the past six years. I have grown in this time, scientifically and personally, and looking back over graduate school it seems every day brought me a little closer to the person I aspire to become. Each of the people I describe here has taught me something, big or small, that I have kept with me.

Advisor

It is a great privilege to be thanking my advisor, **Tania Baker**. I've grown as a scientist by watching Tania: her creativity in experimental design, her talent for weaving a story from a set of experiments, and her impressive stores of scientific knowledge. Tania's talent in another person might be intimidating, and a barrier to the student-adviser relationship, but her accepting and friendly attitude make it not so. For example, sometimes I need to ask a question whose answer I feel I ought to know. Ironically, Tania, whose opinion matters to me the most, feels like the safest person to ask. Writing manuscripts with Tania was a particular pleasure for me, and helped shaped both the way I write about science and the way I do science. She has a keen sense for editing. She also never tries to take control of my writing, but rather allows me to work her comments into my style. As a result, I was able to learn from her how to think about and present my results honestly but compellingly. I also feel fortunate to have been in the lab through both of Tania's pregnancies and the first 3+ years of her daughter's life, and to have watched and talked to her about her experiences as a mother and scientist.

Collaborators

It is a privilege to be thanking **Mike Early**, my scientific partner for the past year and a half. Mike was a technician in Tania's lab during that time, and I am grateful to Tania for suggesting that he and I work together. When he joined the lab, Tania warned me that most new lab people (technicians or grad students) require a lot of training time before they can give back. But it wasn't so with Mike. I would show him an experiment once, and soon he was doing it better than I could. Working with Mike was a pleasure on many levels. He's a smart and careful thinker, and I loved talking to him to daily about our experiments. He never seemed clouded by pride or competitiveness, and so with him it was science for the pleasure of science. During my pregnancies he took over any tasks that involved potential mutagens. I am selfishly glad that Mike and I left the lab at the same time, because it would have become lonely without him. I hope we will stay in touch, because I will miss exchanging insights with him about science and politics and life.

In addition to Mike, I have had the pleasure of collaborating with several excellent people. Most important was **Gary Gerfen,** an assistant professor at Albert Einstein College of Medicine. I met Gary through his post-doctoral adviser, JoAnne Stubbe, just as he was about to

start his own research program in New York. I was eager to learn about Electron Paramagnetic Resonanace, and Gary, a physical chemist by training, was a patient and clear teacher. He was also a gracious host, and the two weeks I spent working with him in his lab were the two most fun weeks of graduate school. I wish I had succeeded in creating MuA tetramers at sufficiently high concentrations to conduct the interesting experiments. But I do not regret any of the time I spent on the project, because of the physical chemistry I learned and because of the opportunity to interact with Gary. Conversations with **Jack Peisach** deepened the experience – a copper oil jug, a gift from Jack, still decorates my lab bench.

Matthew Gray is a dear friend, originally my husband David's but forever now both of ours, as is his wife **Carrie Heitman Gray.** As David says (with only a tinge of jealousy) "Matthew is a wiz at the computer". He is also a generous person, and offered to help me write a perlscript to analyze the data for my second paper (I use the word "help" loosely – he wrote the program, I just explained what I needed.) David and I will miss Matthew and Carrie terribly: their warmth, their frankness, their curiosity and generosity.

Of everyone in Tania's lab, my interests overlapped most with those of **Tanya Williams**. Tanya's experimental prowess is impressive, and it was helpful to have her working on problems close to mine. We've spent many fruitful hours discussing the intricacies of MuA. In keeping with Tanya's nature as a rigorous experimenter, she chose to follow-up on an unexplained set of bands, and this allowed her to make a direct contribution to my first paper.

Lab members

In addition to official collaborators, each member of Tania's lab has contributed to my growth as a scientist. I am fortunate to have shared a bay with **Briana Burton.** Briana is usually the first person I turn to to discuss my data, because I can trust her to understand obscure points, and to think about everything carefully and critically. I have also always admired her conscientiousness as a member of our community, and I know I will miss seeing her every day. **Leslie Roldan** has been a good friend from day one, and I hope she will be always. I enjoy Leslie's gusto and sincerity, in everything from Mu transposition to Penn and Teller magic shows. Throughout graduate school, but especially in the beginning, Leslie was always a good person to answer questions about my experiments. **Glenn Sanders** and I shared a bay for my first two years, and it often seemed that he, more than anyone, took an interest in my progress. He kept nudging me to define my directions, which was exactly what I needed. He was also the best resource in the lab for technical information, which he enjoyed sharing. I very much miss arguing with **Yong-In Kim** about whether single-turnover-kinetics would be necessary to observe the steps of an unfolding reaction, or about the distribution of products in my pseudo-site reaction. Yong-In added warmth to the lab, for me and for many others. I will miss chatting with **Igor Levchenko** about crystalization conditions, or about the absurdities of American politeness. As everyone knows, Igor is an impressive experimentalist; he is also generous with his reagents and his smile. **Eileen O'donahugh** left the lab very shortly after I joined, and, boy, was I sad when she left. In the short time that we overlapped she helped orient me and make me comfortable. **Mike Yamauchi** was our MuB guru, and I was always asking him questions about how his protein affects mine. **Tanya Sokolsky's** constant smile makes her a comfortable person to turn-to for advice, and her insights are always very helpful. She suggested a reorganization for my first paper (chapter 2) which greatly clarified the paper's arguments. **Elizabeth Courtenay** impresses me with her experimental insights and her knowledge of hard-core biochemistry, and I wish I had overlapped with her in the lab longer. Her comments, especially on the work for chapter 4, were invaluable.

Sheila Farrin graciously backed me up in the feat of long-distance writing. She photocopied my notebooks and picked up my thesis from the printer, all with an electronic smile. I am grateful for her good-nature. I also thank her for her excitement over my pregnancy. **Erica**

Beade was gracious about mailing documents and books to me in California, just as she was kind and gracious through-out the time I've known her. I will miss chatting with Erica in the front office.

Matt Gilpatrick, Meredith Seidel, Erica Jackson and **Amanda Carrite** each in turn ran the lab, and all Baker lab members owe them a tremendous thanks. It was particularly fun for me to work with Matt, and I hope he will be happy in his new position.

When **Julia Flynn** and **Samia Siddiqui** joined the lab, they created a new atmosphere, which **Saskia Nehrer** furthered. Suddenly there were more people going to lunch together, talking science in the tea room, or picking-up coffee for one another. I am grateful to have overlapped with them. I am also impressed by Samia's perseverence through a hard trial, and wish her well.

Mentors

Conversations with **Frank Solomon** have influenced my thinking in every area of importance to me. I couldn't count the times I sat in Frank's office, absorbed in some scientific, personal, or inter-personal problem. Frank listens intently, asking questions designed to both comfort and enlighten, and I can see that he is thinking about what I'm saying. Then in a flash, he cuts through the layers of confusion and shows me the heart of the matter. Frank's biochemistry lectures renewed my fascination with the mechanics of life, at a time when my interest was flagging. Later, he helped me rediscover the personal satisfaction I take from doing science, which had become obscured by a desire to prove my capacities as a scientist. This mentorship then expanded outwards, into other areas of life. He taught me that compassion and weakness are not the same, so that self-sacrifice is not always the moral option. He taught me that a mentor should not try to shield a student from the student's own weaknesses, but should strengthen the student to confront them. Frank helped me see that sometimes there's nothing I can do to help my friends, but that I can often help people who are not particularly my friends. Although scientists need to simplify life into categories, Frank also has the gift of appreciating complexity, and he tries to share that gift with me. In politics, he showed me that knee-jerk liberalism is no better than blind conservatism, and in literature he encourages me to appreciate the subtle. In science and in life, he helped enable me towards defining my personal standards of greatness.

Each member of my thesis committee has had a valuable impact on me and has taught me important things. Back in my first year, Tania recommended that I take **Steve Bell's** cell cycle course, and this was an excellent recommendation. "Cell cycle" turned out to be euphemistic: the course was actually about the G1 to S transition. I have never had so much fun thinking deeply and discussing spiritedly one small piece of nature. Later, Steve's comments on my research manuscripts were outstandingly helpful, and I appreciate his generosity in reading and rereading them and thinking them through carefully. I also thank him for his experimental ideas, and in particular for suggesting the cloning method that I used in my second paper (chapter 3).

I was strongly impacted by **Bob Sauer's** teaching in the graduate biochemistry course, in which I was first a student and then a TA. I particularly remember one review session Bob taught, in which the power of reaction coordinate diagrams suddenly became clear to me. In addition to thanking him for insightful comments at my thesis committee meetings, I thank Bob for meeting with me privately over the years to discuss my data.

Back in my first year, Tania recommended that I take **Steve Bell's** cell cycle course, and this was an excellent recommendation. "Cell cycle" turned out to be euphemistic: the course was actually about the G1 to S transition. I have never had so much fun thinking deeply and discussing spiritedly one small piece of nature. Later, Steve's comments on my research manuscripts were outstandingly helpful, and I appreciate his generosity in reading and rereading

them and thinking them through carefully. I also thank him for his experimental ideas, and in particular for suggesting the cloning method that I used in my second paper (chapter 3).

I have enjoyed many political discussions with **Jonathon King,** and I admire his involvement in environmentalist politics. He has made me reconsider many of my positions on political issues, and though sometimes I end up agreeing with him and sometimes not I always appreciate the challenge to think things through. Jon has also taught me to present my work to appeal to as broad an audience as possible; a skill that will serve me well.

I am deeply honored that **Kiyoshi Mizuuchi** agreed to serve on my thesis committee. Kiyoshi was Tania's post-doc adviser, and he is the long-time leader in the field of transposition. As a scientist he is remarkably creative, insightful, and deep, and his experiments pave the entire field. In addition, his modesty, and his love of truth rather than glory, help set a collegial tone throughout the field. I have always appreciated at conferences that he allowed me to engage him in conversation, and from each conversation I returned with weeks' worth of problems to think about and, sometimes, a new insight into the entire field of biochemistry.

I thank **Sylvia Sanders** for her compassionate and honest advice about careers and family, and for the atmosphere she helped create at Motility and Morphology meetings. I thank **Alan Grossman** for talking to me about my career, and more generally about careers in science. One conversation in particular came at a critical juncture, and helped me decide in favor of academic science. I enjoyed serving as a TA for **Gene Brown's** undergraduate biochemistry course. I learned at least as much from Gene's lectures as the students did, and I appreciate his continued interest in my progress years after my involvement in the course. I thank **Dave Bartel** and **Barbara Imperialli** for very helpful conversations about my work.

MIT Colleagues

First and foremost, I thank my dear friend **Janet Lindow**. Pounding that three-mile loop between the Harvard to the Longfellow bridges, Janet and I shared scientific ideas, and also shared our frustrations, excitement, and thoughts about life. Her unique perspective has broadened mine. Later, **Rob Britton** joined us on many runs, and I will miss Rob's comfortable, frank and open attitude. I will also miss trips to the gym with Janet and **Philina Lee**. It's refreshing to be with powerful scientists who encourage rather than compete with each other.

The Grossman lab, of which Janet, Rob and Phil were members, is an exceptional group of people. I will miss conversations in the tea room with all of them, especially **Kate Bacon, Bill Burkholder, Natalia Comella, Elke Kuester-Schoeke, Katherine Lemon** and **Ryan McQuade. Bill** and **JoAnne Burkholder** have become friends outside of work, too, and I'm so glad they are moving to Palo Alto. Thank you to **Elke** for very helpful comments on my recent paper.

Though I've developed many friendships at MIT outside the Baker and Grossman labs, two in particular I am certain will be lasting. Conversations with **Melissa Hunter-Ensor** always offer me new insights into human relationships. Her love of science and her love of her children have made her a model mother-scientist for me. **Arie Bergruun's** cheerful frankness, his generosity, and his clear-headed intelligence make him a special friend.

My classmates offered a supportive community. **Aurora Burds, Dianne Keen, Janet Lindow** and others were always organizing us to celebrate a wedding or birth, or offer condolences for a death. I especially appreciate my friendships with **Fang Li Chen** and with **Anya Goodman.**

Coffee breaks with **Kate Compton Abruzzi** were always fun and interesting. She and I shared many common experiences, and her perspective was helpful to me. I also very much enjoyed conversation with **Milan DeVries**, whose values are in many respects close to my own. Milan is a creative and caring person and scientist. I thank **Ernest Fraenkel** for extremely

helpful comments on my second paper, and for years of friendship. And I am so glad to have known **Virgina Cornish,** an inspiring scientist. I hope to stay in touch with all of you.

Beyond MIT

Boston is a great place to be a scientist, because its scientific community is so large and broad. I enjoyed the recombination super group meetings that included our lab,**Al Engelman**'s lab and **Margie Oetenger's** lab. Interavting with Margie is always fun and enlightening, because she goes straight to the heart of the matter. Interacting with Al was particularly helpful to me. We had several detailed conversations, about my work and about related topics, and I am grateful for his insight. I've also benefited from maintaining contact with my undergraduate adviser, **Greg Verdine.**

Tania is extremely generous about sending her students to conferences, allowing us to feel connected to an exciting community of scientists. I particularly want to mention **Dave Roth,** with whom I've enjoyed many deep conversations. Dave is an exceptional person for, among other things, his persistence, his modesty, his ambition and his friendliness, and I feel honored to have become friends with him. I also want to especially mention **Keren Swinger,** my fellow warrior for the cause of high-concentration tetramers. If anyone ever tells you MuA is a well-behaved protein, don't believe them! But the best part, for me, of trying to concentrate these tetramers is that it gave me an excuse to get to know Keren, a kind person and a lover of science. I've also very much enjoyed interacting with Keren's adviser. **Phoebe Rice.** It seems that everyone in the field admires Phoebe, and if you've met her you probably understand why.

I've enjoyed talking to many people, far too many to list. But I'd like to especially mention **Katsuhiko Yanagihara,** and also some of his colleagues at the NIH: **I-Hung Shih, Eric Greene** and **Michiyo Mizuuchi.** I'd also particularly like to mention **Dave Sherrat,** who explains site-specific recombination remarkably clearly. I've also enjoyed getting to know **David Haniford, Nacny Craig, George Chaconas, Igor Goryshin** and **Pat Higgins.** I am very grateful to **Sergei Savliev** for his generous sharing of experimental advice.

Past mentors

I'd like to take a little space to thank five scientists who greatly influenced my development. My high-school math teacher, **Shelly Strimling,** made mathematics fun and comfortable, and my high-school biology teacher, **Dave Newman,** specifically excited me to do biology. **Brian Davis** furthered that excitement, and made academic research attractive. **Louise Anderson** had a profound impact on me; in her lab I learned to design an experiment and analyze the data, and also how to tackle a whole new field that I new nothing about.

Most importantly, my father, **Robert Gordon,** taught me to graph a function almost as soon as I could count. He took me and my siblings to the planetarium, he talked to us about his own research, and he and my mother shared with us the beauty of nature through family hiking trips. These early exposures to science and nature cannot be overestimated for their impact on us children. My father has always done his work for the love of God and nature -- never out of pursuit of recognition and glory. I aspire to be that type of scientist.

Friends and Family

Now I've thanked the people who've been most important to my scientific development, but these pages feel empty without my at least mentioning some who've touched my science indirectly, through their friendship. **Philip, Miriam,** and **Tani Bloom Schwartz,** I miss you guys terribly. Thank you for welcoming me so warmly into your home and family. **Rachel Rosner,** I have never felt so instantly connected to a person as I have with you. **Barbara Goldoftas,** you have been an important mentor to me for many years, but now our relationship has deepened into something more mutual and very special. **Pamela Weinfeld,** we have a

friendship bond that will hold for a lifetime. **Melissa Klein,** I am so grateful to you for reaching out to rekindle a wonderful friendship. **Shona Mukherji**, thank you for being there to welcome me to California with the type of friendship that comes from shared values and mutual understanding. And most of all, **Ruti, David, Oren** and **Aliza Abusch-Magder**: there's no one like you. I feel truly blessed with our friendship.

I am also blessed by two wonderful families. To the family I married into: **Fred** and **Suzan Goldhaber** and **Sara** and **Jeremy Goldhaber-Fiebert**; you are incredibly loving, generous and accepting. How could I have gotten so lucky? And to the family that I was born into, **Robert, Evvy, Tamar** and **Dov Gordon**, thank you for 29 years of love and acceptance, for clearly defined values and also room for growth, for knowing how to laugh at yourselves and for enjoying each other more than anything else. To our parents especially, thank you for encouraging us to be thoughtful and moral people, to love thinking through intellectual problems, and to enjoy life.

This past year and a half has seen a lot of changes for our family, some joyous and some difficult. My grandmother, **Sarah Rochel Landau Rabinowitz z"l**, passed away with a few weeks of her beloved niece, Esther HaLevy. My grandmother and Esther were in many ways like sisters. They each had an overwhelming force of will, and they could use it to move continents if they felt it would hold their families together. This thesis is dedicated to Esther, because the suddenness of her death, and the physical distance between us, meant that I never parted from her. But I am remembering my grandmother as well.

And finally the most important person: thank you to **David Goldhaber-Gordon.** Really, I should have put David earlier, because as a scientist he is often my most helpful colleague. I am sometimes astounded that in a field far from his own, David provides insight into my data that no one else could provide. Plus, he lends me his superior ability to solve differential equations, as well as his political intuitions. But though David's input as a scientist has been crucial for me, of course his acknowledgement belongs here, at the end, as he is the most important friend and family member. He supports me and works with me and we share the joy of each other's successes. We are now embarking on a great adventure together, and I couldn't feel more confident in our partnership.

Chapter 1. Introduction
The Mu Transpososome and Friends: a Means of Grabbing Control

DNA transposition occurs in a protein-DNA complex called a transpososome. Within this complex, multiple copies of a transposase protein hold together the two ends of the transposon DNA. Transpososome structure and the processes of transpososome assembly and disassembly are crucial for controlling the outcome of transposition, on many levels. Some transpososome features are shared across the transposon family, and reflect common needs for control. Other features are particular to each individual transposon, reflecting that transposon's biological niche. In this chapter I provide a mechanistic overview of Mu transposition, emphasizing the importance of the transpososome, and comparing the Mu transpososome to that of other family members.

Background information.

Transposition is the movement of a transposon DNA sequence from one DNA location to another. The smallest transposons, known as insertion sequences, can be as short as 700bp (Chandler, 2002). For these small DNA elements, transposition is a means of spreading laterally through a bacterial population. Near the other end of the size-spectrum is bacteriophage Mu, whose entire 40kb phage genome is a transposon. Mu transposes once each time it infects a new cell, to integrate into the host's chromosome. It also transposes multiple times during lytic growth, creating a new copy of its genome with each transposition event (Pato, 1989).

There are three families of transposons. "DNA transposons" encode a transposase protein that provides the active site for the cutting and joining reactions of transposition. The transposase is active in the context of a transpososome: a complex that includes, at a minimum, two copies of transposase and the two outer ends of the transposon DNA. Another family, the non-LTR retrotransposons, transpose via an RNA intermediate and are reverse transcribed directly into their relocation site (Finnegan, 1997). This is the most abundant family in the human genome: DNA derived from non-LTR transposons makes up 34% of all the DNA in the human genome, whereas DNA derived from other transposons comprises a still-substantial 11% (Lander et al., 2001). The third family, LTR retrotransposons, combines aspects of each of the other two. They transpose via an RNA intermediate, but the reverse-transcribed DNA must integrate into its relocation site through the cutting and joining actions of an integrase protein, closely related to transposases (Rice and Baker, 2001). The integrase and the retroviral cDNA form a complex, analogous to a transpososome, to mediate integration (Carteau et al., 1999; Chen and Engelman, 2001; Ellison et al., 1995; Kukolj and Skalka, 1995; McCord et al., 1999; Murphy and Goff, 1992; van Gent et al., 1993; Wei et al., 1998). HIV and other retroviruses belong to this last family. Because the retroviral integrases are related to transposases, our understanding of the mechanics of HIV integration is founded on studies of bacterial transposition, especially Mu transposition.

At a minimum, a transpososome must perform two covalent modifications of DNA: cleavage of the transposon's 3' strand away from the "donor" DNA that flanks it, and a "strand transfer" reaction to insert the cleaved strand into the new DNA location, the target DNA (fig 1). In addition, some transpososomes also cleave the transposon's 5' strand away from the flanking DNA, causing full excision of the transposon from its original donor DNA. This full excision is called cut & paste transposition, and is characteristic of many transposons including Tn5, Tn7, Tn10, the P-elements of fruit-flies, and phage Mu when it is integrating into the chromosome of a new host cell. If the 5' strand is not cleaved, the transposon undergoes replicative transposition (Craig, 2002). The initial transposition product is a branched structure, with the transposon's 3' strands attached to the target DNA and the 5' strands still attached to the donor DNA. Selective replication across the transposon resolves this structure. Phage Mu uses replicative transposition during lytic growth, creating 50-200 copies of its genome before lysing the cell. It is the better-studied of the two pathways of Mu transposition, and is the subject of my thesis work (Pato, 1989). Other transposons that perform replicative transposition include members of the Tn3 family, such as γδ.

The Mu transpososome performs two reactions during replicative transposition, twice each: 3' cleavage and strand-transfer for each end of the Mu DNA. Remarkably, the transposase contains only one active site for the two distinct reactions, and a single transposase subunit of the transpososome performs both reactions for one DNA end (Namgoong and Harshey, 1998; Williams et al., 1999). The transposases from Tn5 and Tn10 perform 3' cleavage, strand transfer, plus additional reactions to cleave the transposon's 5' strand, again all with a single active site (Bolland and Kleckner, 1996; Naumann and Reznikoff, 2000).

Transposase has a complex task. At a chemical level, it must organize the distinct substrates for cleavage and for strand transfer, so that each reaction is performed with the proper substrates. At a higher level, the transposase protein ensures that all of the transposon DNA, and only the transposon DNA, relocates together. The transposase must also select an appropriate target DNA. At a higher level still, transposase may be involved in controlling the timing of transposition. All this control is made possible by strict requirements for transpososome assembly and architecture.

The core Mu transpososome contains a tetramer of the transposase, MuA (fig 2). Yet MuA is monomeric in solution; the MuA monomers assemble into a tetramer only by binding to DNA recognition sites located near the ends of the Mu genome. The recognition sites share a 22 base-pair consensus sequence, and they are named by their position relative to the ends of the Mu DNA: L1, L2 and L3 on the left end, and R1, etc. on the right (fig 2). The DNA cleavage site, the actual site of cleavage and strand transfer, is located five base-pairs beyond the R1 or L1 recognition site, precisely at the end of the Mu DNA.

My thesis focuses on protein-DNA interactions within the Mu transpososome, and the impact of these interactions on the control of transposition. In chapters 2 through 4 I discuss how interactions between MuA and its recognition sites impact transpososome assembly and later steps of transposition. In chapter 5, I discuss how, in the context of a transpososome, interactions between MuA and the very ends of the Mu DNA (the DNA cleavage sites) allow for two distinct reactions, cleavage and strand transfer, to be mediated correctly by a single active site. In the current introductory chapter, I provide a mechanistic overview of transpososome organization and function, emphasizing in the later parts how transpososome formation is the basis for control mechanisms.

MuA – the Transposase

Transposases are related by the two reactions that they all perform – 3' donor cleavage and strand transfer – and they differ most critically by the DNA sequences that activate them. Crystal structures have been solved for the core catalytic domains of MuA transposase (Rice and Mizuuchi, 1995), Tn5 transposase (Davies et al., 1999), HIV integrase (Dyda et al., 1994), ASV integrase (Bujacz et al., 1995), RSV integrase (Yang et al., 2000), and SIV integrase (Chen et al., 2000b). The catalytic domains of all of these proteins share a common fold and are easily superimposable (fig 3). In contrast, the DNA-binding domains and regulatory domains vary dramatically from protein to protein (Rice and Baker, 2001). This modular organization may have permitted rapid evolution of new transposons, as an existing transposase might have swapped its DNA binding domain with that from an unrelated protein and then begun to transpose a new set of DNA sequences. Consistent with this idea, in chapter 4 of my thesis I show that sequence-specific contacts between MuA transposase and its N-terminal DNA binding domain do not contribute directly to DNA cleavage and strand transfer, the common reactions for transposases. Rather, those contacts are most important when the transpososome is first being assembled, and, as discussed below, the details of the assembly process varies with the individual transposon.

The catalytic core.

The transposase catalytic core contains both α helices and β sheets, in an arrangement similar to that of RNAseH, and to the nuclease RuvC that resolves Holliday junctions (Rice et al., 1996; Rice and Baker, 2001; Yang and Steitz, 1995). This structural relationship emphasizes that transposases are, fundamentally, nucleases. The cleavage reaction, like most nucleolytic cleavages, is simply a hydrolysis (Engelman et al., 1991). The strand transfer reaction is a one-step transesterification (Mizuuchi and Adzuma, 1991). Chemically, a transesterification is similar to a hydrolysis, the only difference being the chemical context of the oxygen atom that serves as nucleophile (fig 1). Consistent with the transesterification function, the catalytic core of transposases is more distantly related in structure to proteins that perform other types of

phosphoryl exchange reactions: the kinase family and the 3'-5' exonuclease domain of the klenow fragment (Rice et al., 1996).

The catalytic core of a transposase or retroviral integrase contains three conserved, acidic residues, the "DDE motif" (figs 3 and 4). Mutations in these residues severely reduce the efficiencies of both cleavage and strand transfer, but have only subtle effects on transpososome assembly (Baker and Luo, 1994; Kim et al., 1995; Krementsova et al., 1998). Based on these findings, the DDE residues were predicted to coordinate one or two divalent metal ions at the heart of the active site; several integrase structures, and the most recent structure of Tn5 transposase, now support a two metal-ion model (Bujacz et al., 1997; Bujacz et al., 1996; Goldgur et al., 1998; Lovell et al., 2002). Most likely, the metal ions serve to activate a nucleophile for attack on a DNA phosphate, and also to stabilize build-up of charge on the phosphate in the transition state. There are only a few direct experiments exploring the catalytic role of divalent metal ions in transposition (Junop and Haniford, 1996), but studies of other proteins, such as the 3'-5' exonuclease domain of the klenow fragment, have greatly influenced the way we think about transposition (Steitz and Steitz, 1993; Yang and Steitz, 1995). Over the course of each reaction, the metal ions quite likely serve to activate the nucleophile (water or 3'OH), stabilize the build-up of negative charge on the phosphorus atom in a penta-valent transition state, and finally stabilize the 3' hydroxyl leaving group.

Other domains of MuA involved in catalysis.

MuA transposase is divided into three proteolytically defined domains (fig 4) (Nakayama et al., 1987). The structurally-conserved catalytic core is contained within domain II, and domain II also contains a smaller sub-domain that is not present in other family members. This smaller domain is a positively charged β barrel, suggesting that it might be involved in DNA interactions (Rice and Mizuuchi, 1995). Several mutations within that barrel effect the transition between the cleavage and strand transfer reactions (Krementsova et al., 1998; Namgoong et al., 1998).

Domain IIIα of MuA has an unusual property on its own. When expressed as a peptide, residues 575-600 have non-sequence-specific nuclease activity, linearizing a circular plasmid (Wu and Chaconas, 1995). I know of no other peptide that small that can induce the making and breaking of covalent bonds. The closest is a 20 residue peptide from RecA, that can catalyze a DNA strand exchange, but of course that does not require changes in covalent interactions (Voloshin et al., 1996). Yet it is unclear to what extent the catalytic potential of the IIIα peptide is utilized during transposition, since a triple mutation that abolishes the nuclease activity of the IIIα peptide does not abolish transposition activity in the context of the full protein. Rather, those mutations inhibit transpososome assembly and make transposition more dependent on stimulation by the MuB protein, discussed below. The effects of those mutations are similar in detail to the effects of mutations in the DNA cleavage sites, suggesting that domain IIIα interacts with the cleavage site (Naigamwalla et al., 1998). Perhaps domain IIIα distorts the DNA in such a way as to increase its susceptibility to cleavage – whether it be cleavage by the active site of MuA, or, when the domain is expressed as a lone peptide, cleavage by free radicals in solution.

DNA binding.

Transposases contain DNA-binding domains: domains that bind a recognition sequence in a mode similar to transcription factors. These contrast with the catalytic domain, whose contacts near the cleavage site are less sequence-specific and more invasive, distorting the DNA on a path towards catalysis (Davies et al., 2000). Other proteins use this multi-domain approach as well: site-specific recombinases and the modular restriction enzyme *FokI* are good example (Wah et al., 1997; Yang and Mizuuchi, 1997). Contrast these proteins with non-modular restriction enzymes such as *EcoRV* and *BamHI,* in whom regions responsible for DNA-recognition and regions responsible for homodimerization are all interspersed with a catalytic structure. As a

result of this interspersion of functions within one domain, cleavage occurs internal to a palindromic recognition sequence (Pingoud and Jeltsch, 1997). Such an arrangement would be poorly suited for a transposase, whose cleavage site defines the end of the transposon. Transposases' modular organization permits the cleavage sites to be separated from the recognition sites, so that the cleavage sites are at the transposon's termini and the recognition sites are situated several base-pairs within the transposon.

The DNA-binding domains of transposases are usually N-terminal, but their structures can vary greatly. Even two very closely related transposases, those from the C. Elegans transposons Tc1 and Tc3, have no sequence homology between their DNA binding domains or between the DNA sites that they recognize. This ensures that Tc3 transposase does not assemble a transpososome on Tc1 DNA, for example (van Pouderoyen et al., 1997).

The N-terminus of MuA contains three independently folding helix-turn-helix (HTH) motifs, each recognizing a distinct DNA sequence (Clubb et al., 1994; Clubb et al., 1997; Schumacher et al., 1997). The first HTH motif, Iα, recognizes an enhancer-like sequence called the IAS (Internal Activating Sequence), located ~950bp from the left end of Mu (see below) (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989). HTH motifs Iβ and Iγ each recognize half of the 22 base-pair recognition site sequence from the ends of the Mu DNA, and together these recognize the full recognition sequence (Schumacher et al., 1997). The transposase of Tc3 also contains a DNA-binding HTH motif, plus an additional DNA binding region (van Pouderoyen et al., 1997). By contrast, the transposase from Tn5 contains a single, four-helical DNA binding domain (Davies et al., 2000).

MuB ATPase, a necessary accessory

A mechanistic understanding of MuA requires some knowledge of MuB, a second phage-encoded protein required for robust transposition. MuB is known as the immunity protein, because it causes DNA regions that already contain a copy of the Mu DNA to be immune from selection as transposition targets (Maxwell et al., 1987). This target immunity prevents Mu from inserting into itself.

Others transposons also display target immunity. Bacterial Tn7, for example, encodes a protein TnsC that functions analogously to MuB (Stellwagen and Craig, 1997). The barrier-to-autointegration protein (BAF) plays a role in HIV integration that is conceptually similar to MuB's role in Mu transposition (Lee and Craigie, 1998). Tn3 transposase displays immunity, but the identity of its immunity protein is a mystery (Robinson et al., 1977). However, many transposons do not display immunity. Tn10, for example, can integrate into itself, causing chromosomal rearrangements (Kleckner et al., 1979).

MuB does not actively inhibit insertion into immune targets, but rather stimulates strand transfer into non-immune targets. More specifically, MuB stimulates MuA to use a target DNA that is already bound by MuB. Since this MuB-stimulation is a considerable advantage, those molecules *not* bound by MuB become immune as transposition targets when MuB is present (Baker et al., 1991). Thus to generate proper immunity, MuB must avoid binding to Mu DNA. In an interesting activity loop, interactions with MuA itself determine MuB's binding preferences (fig 5). MuB's affinity for DNA is highest when it is bound to ATP, and MuA stimulates MuB to hydrolyze ATP and thereby release the DNA. As a result, MuB is cleared from DNA molecules that contain a nearby MuA binding site, where MuA is often present (Adzuma and Mizuuchi, 1988). *In vivo,* DNA within 5kb of a Mu DNA end sequence is fully immune, and immunity drops off between 5 and 25kb. These results and others are consistent with immunity being mediated by random collisions between MuA and MuB caused by DNA supercoiling (Manna and Higgins, 1999).

It seems counterintuitive that interactions between the same two proteins, MuA and MuB, should in different contexts have almost opposite effects (fig 5). In one case, MuA clears MuB from the DNA, making that DNA immune as a transposition target. In the other case, MuA

select as a target only MuB-bound DNA. How is this possible? According to one model, the answer may lie in the fact that MuA is only active as a transposase when it is assembled into a transpososome. Before assembling into a transpososome, MuA is unable to commit to a target DNA molecule (personal communication, T. L. Williams and T. A. Baker). At that stage, MuA-MuB interactions clear MuB from the donor DNA, but cannot stimulate strand transfer. By the time the transpososome is assembled and competent to engage a target DNA, presumably MuB is sufficiently cleared from the immune DNA. MuB binds DNA as a polymer, and the polymer does not need to be 100% cleared to generate immunity. It has been suggested that DNA close to MuA will contain incomplete MuB polymers, those farther away will contain longer polymers, and polymer length could influence transposition activity (Greene and Mizuuchi, 2002).

MuB actively stimulates strand transfer by MuA – even versions of MuB that cannot bind DNA stimulate strand transfer, showing that MuB's influence on MuA is more complex than simply brining MuA into close proximity with an appropriate target (Baker et al., 1991; Surette and Chaconas, 1991). One possibility, also discussed below, is that MuB stimulates engagement of target DNA in the MuA active site. However, *in vitro* MuA is able to do strand transfer at a reduced rate in the absence of MuB, so MuB is not absolutely required for productive binding of target.

In addition to directly stimulating strand transfer, MuB also stimulates transpososome assembly (Mizuuchi et al., 1992), and MuB and target DNA under some conditions stimulate donor cleavage (Mizuuchi et al., 1992; Naigamwalla et al., 1998). It is not clear whether MuB stimulation of transpososome assembly or of target cleavage is relevant to target immunity. Transpososomes that are pre-assembled and/or pre-cleaved can respect target immunity (Baker et al., 1991; Surette and Chaconas, 1991).

Organization of the transpososome

DNA binding and catalysis in trans.

Transpososome structure helps ensure that a transposon's two ends, potentially tens of kilobases apart, always transpose together. The core Mu transpososome consists of a tetramer of MuA (Lavoie et al., 1991; Mizuuchi et al., 1992), some of whose subunits are bound simultaneously to both ends of the transposon (Aldaz et al., 1996; Savilahti and Mizuuchi, 1996) (fig 2). Since both transposon ends are required to form a transpososome, transposition of a single end is strongly disfavored. A complex containing both transposon ends is also called a "synaptic complex". In chapter 2 I show that the energetic drive to form a synaptic complex is so strong that if Mu DNA sequences are artificially limited, MuA will synapse a non-Mu sequence with a Mu sequence. Other transpososomes, such as that of Tn5, contain transposase dimers rather than tetramers (Bhasin et al., 2000), but they also form a synaptic complex that simultaneously binds both transposon ends (fig 6) (Davies et al., 2000).

How is simultaneous binding of two DNA ends achieved? The key resides in the modular structure of the transposase. The transposase's DNA binding domain engages a recognition site near one transposon end, while the catalytic domain of the same subunit engages the cleavage site at the opposite transposon end (figs 2 and 6). This "*trans*" binding was first detected in a series of elegant biochemical experiments (Aldaz et al., 1996; Namgoong and Harshey, 1998; Naumann and Reznikoff, 2000; Savilahti and Mizuuchi, 1996; Williams et al., 1999), and dramatically substantiated by the crystal structure of a synaptic complex from Tn5 (fig 6) (Davies et al., 2000). In the structure, a dimer of Tn5 transposase synapses a pair of DNA fragments containing sequences derived from the ends of the Tn5 transposon. On each fragment, one transposase subunit contacts the recognition site and the other subunit contacts seven base-pairs surrounding the cleavage site. Protein-protein contacts are minimal and occur along a single helix. Thus, the structure suggests that transposases participate in a novel form of cooperative DNA binding. Classically, cooperative binding can be mediated through direct

protein-protein contacts and/or through structural changes in the DNA (unstacking, bending, etc.). Subunits of the Tn5 transpososome also appear to cooperate to pay "bridging energy". Transpososome assembly involves an entropic cost to constrain two DNA molecules in one complex. Since at least two transposase subunits bridge the two DNA ends, these subunits are cooperating to bring the ends together.

The significance of this cooperativity, and of the extent of *trans* binding, is highlighted by the contrast with the lambda family of site-specific recombinases (Yang and Mizuuchi, 1997). The recombinases, like transposases, function in synaptic complex. The complexes contain symmetrical recombinase tetramers, and each recombinase subunit primarily contacts one DNA segment. However, the structure of Flp recombinase shows each subunit contributing a catalytic tyrosine residue to its neighbor's active site. Thus the tyrosine residue of Flp is contributed *in trans*, similar to the DDE residues of a transposase. Yet unlike the situation with transposase, the *trans* DNA contacts are minimal: a single helix pokes into a predominantly *cis* active site, allowing access of the *trans* tyrosine to the DNA (Chen et al., 2000a). In fact, the structure of a related recombinase, Cre, shows an entirely *cis* active site (Guo et al., 1997). Thus, even among synaptic complexes, transpososomes are unusual in the extent to which they use cooperative bridging energy.

MuA functions as a tetramer.

At the core of the Mu transpososome is a MuA tetramer (Lavoie et al., 1991; Mizuuchi et al., 1992). It is the two MuA subunits bound proximal to the cleavage sites, at the R1 or L1 sites, that bridge the transposon ends, contributing DDE residues *in trans* (Aldaz et al., 1996; Savilahti and Mizuuchi, 1996) (fig 2). The two distal subunits may contain mutant DDE residues without reducing transposition efficiency (Namgoong and Harshey, 1998; Williams et al., 1999). We do not know whether the two distal subunits (L2 and R2) bridge the two transposon ends in other ways.

More generally, we do not know how the distal subunits contribute to the overall process of transposition. The core Tn5 synaptic complex contains only a transposase dimer, and it not only mediates 3' donor cleavage and strand transfer but 5' donor cleavage as well (Bhasin et al., 2000). Thus a tetramer is not intrinsically required for transposition.

The requirement for the extra subunits may provide additional opportunities for controlling the timing and location of transposition. In general, the most heavily controlled step of transposition is transpososome assembly, and the distal subunits are required for assembly (Allison and Chaconas, 1992; Baker and Mizuuchi, 1992). The distal subunits can contribute to interactions with both the IAS sequence and MuB, each of which is involved in a regulatory mechanism for assembly (Mizuuchi et al., 1995).

The distal subunits can also be the sole providers of domains IIβ and IIIα in an active transpososome (Krementsova et al., 1998; Namgoong et al., 1998). Since those domains are implicated in catalysis, the distal subunits may play a role in post-assembly steps as well. Tn5 transposase includes an inserted region, not present in MuA, that interacts with DNA near the cleavage site (fig 6). It is possible that either MuA domain IIβ or IIIα functions analogously to the Tn5 insertion, and that that function is provided by the domain from a distal subunit (Davies et al., 2000; Rice and Baker, 2001).

Direct coupling of reactivity at the two transposon ends.

Two experiments suggest that reactivity at the two Mu ends is directly coupled, beyond the fact that the two ends cooperate during transpososome assembly. The interpretation of the first experiment (Mizuuchi et al., 1995) is complicated, and can only be made in retrospect thanks to later data to illuminate the original experiment. In this experiment, MuB was necessary to see strand transfer. Transpososomes were assembled from mixtures of (i) full-length MuA subunits with mutations in the DDE residues, and (ii) DDE wild-type subunits, C-terminally truncated to prevent MuB interactions. These mixed transpososomes produced only "single-ended" strand

transfer products: products in which only one transposon end had been transferred to a target. The failure of the other end to transpose indicates incorporation of a DDE mutant at an R1 or L1 position, the position that contributes the DDE residues. But the protein that was DDE active was defective for MuB interactions. The conclusion is that MuB stimulated the strand transfer activity of the wild-type DDE subunit, through interactions with the DDE mutant subunit. Since both subunits were bound at cleavage-site proximal positions (R1 or L1), one was bound to each Mu end, indicating coupling between subunits bound at opposite ends. A later, more straightforward experiment also indicates coupling between the subunits bound at the two Mu ends: in the absence of MuB, failure to cleave one transposon end (due to the presence of a DDE mutant subunit) inhibits strand transfer at the opposite end (Williams et al., 1999).

How is coupling between ends mediated? In the crystal structure of a Tn5 transpososome, direct interactions between subunits are minimal (Davies et al., 2000). If the same is true in the Mu transpososome, it is unlikely that activity is coupled through protein-protein contacts. However, strand transfer of the two Mu ends occurs into sites only five base-pairs apart on the target DNA, into opposite strands of the target. The proximity of the two target sites suggests that the two MuA active sites may cooperate to engage a target DNA. If so, perhaps MuB stimulates this simultaneous engagement of the target DNA in both active sites, which would explain how MuB stimulates strand transfer of one end through interactions at the other end. This model is consistent with the fact that direct coupling has been detected for strand transfer but not for cleavage, in studies of Mu (Williams et al., 1999) and in studies of other transposons (Chen and Engelman, 2001; Haniford and Kleckner, 1994). Some older studies of Moloney retroviral integrase and of MuA report coupling of cleavage, but in retrospect they seem to be describing cooperation in assembling a synaptic complex, rather than direct communication between active sites (Murphy and Goff, 1992; Surette et al., 1991).

Tn10 transposition exhibits end-coupling even *in vivo*; a cleavage site mutation at one end of Tn10 DNA inhibits strand transfer at the opposite end *in vivo* (Haniford and Kleckner, 1994). Thus for Tn10 transposition, direct coupling of reactivity may help ensure that the two transposon ends always relocate together. Studies of end-coupling in Mu have all been *in vitro*. In Mu experiments, the presence of MuB overcomes end-coupling, permitting the synaptic partner of an uncleaved end to transpose (Williams et al., 1999). Likewise, in chapter 3 of my thesis I show that, at least in the presence of MuB, a base-pair insertion near the cleavage site at one end does not interfere with strand transfer at the opposite end (Goldhaber-Gordon et al., 2002a). Thus it is not clear if direct coupling is an important control mechanism of Mu transposition *in vivo*, where MuB is present. The more important mechanism is probably cooperativity during transpososome assembly. Regardless, the coupling that is unmasked *in vitro* is a useful tool in elucidating the molecular mechanisms that underlie transposition – for example, the possible requirement for simultaneous engagement of the target in both active sites.

Moving substrates in and out of an active site.

Transposases are remarkable for their ability to perform multiple reactions with a single active site. Admittedly, the reactions are chemically similar; each is a nucleophilic substitution of a DNA phosphate (fig 1). But the reactions' substrates differ from each other in critical ways. The nucleophile during 3' cleavage is a water molecule; during strand transfer, it is the last 3' hydroxyl on the transposon DNA (Engelman et al., 1991; Mizuuchi and Adzuma, 1991). The phosphate under attack during 3' cleavage is part of the donor DNA, and is located immediately 5' of the transposon end sequences. During strand transfer the phosphate is in the target DNA. It seems fair to say that these two reactions have no more in common than do the three reactions catalyzed by DNA Polymerase I, yet Pol I has separate active sites for each of its two exonuclease activities and a third site for its polymerase activity (Patel et al., 2001).

The conundrum of the transposase can be appreciated when its reactions are compared to those performed by site-specific recombinases, either the lambda family or the gamma-delta

family. These two families also use a single active site to mediate multiple DNA cleavage and joining reactions, and the outcome of site-specific recombination is similar to that of transposition. But site-specific recombinases do not hydrolyze their DNA substrates. The cleavage reaction forms a covalent bond between a DNA phosphate and an active site residue on the protein. The DNA-joining reaction is a reversal of the cleavage reaction: an exposed DNA hydroxyl attacks the protein-DNA bond, becoming joined to the DNA phosphate from that bond. Since site-specific recombination occurs between sets of identical sequences, the reversal is perfect at the level of the cleavage site DNA. In contrast, a transposition target sequence need have no relationship to the transposon sequence (Haapa-Paananen et al., 2002), and even if it did, strand transfer is in not at all a reversal of 3' cleavage: one is a transesterification, the other a hydrolysis (Engelman et al., 1991). This leaves the puzzling question: how does one transposase active site perform two distinct reactions?

A study of thio-substituted transposition substrates has lead to an elegant model for the organization of substrates in a transposase active site. The most informative experiments were with Tn10 transposase, though corroborating experiments were also performed on MuA (Kennedy et al., 2000). Tn10 is a cut & paste transposon, which means both the 3' and 5' strands at the ends of the transposon DNA are cleaved from the flanking DNA by the transpososome. To accomplish 5' cleavage, the transposase first forms a hairpin at the end of the transposon DNA, and then opens the hairpin. This leads to a total of four reactions performed in one active site: (i) hydrolysis of the transposon's 3' strand, (ii) hairpin formation, during which the transposon's 3' OH attacks the transposon's 5' strand. After this step, both strands are now severed from the flanking DNA, (iii) a hydrolysis to resolve the hairpin, again freeing the transposon's 3'OH and (iv) strand transfer, in which the transposon's 3'OH attacks a target DNA (Bolland and Kleckner, 1996; Kennedy et al., 1998). When these reactions were performed with thio-substituted substrates, it was found that the preferred stereoisomers for the reactive phosphate were as follows: Rp for 3' cleavage, Sp for hairpin formation, Rp for hairpin resolution, and Rp for strand transfer. By assuming a two-metal-ion mechanism (see above), the researchers could use this stereselectivity to model placement of DNA substrates relative to the DDE-coordinated metal ions in the transposase active site. Since each reaction causes an inversion of phosphate chirality, the alternating requirements for Rp, Sp, and Rp in the first three steps suggests that the configuration of the active site does not change dramatically between steps.

The researchers proposed that the 3' oxygen at the transposon end is stationary throughout the four reactions, held in place by divalent metal ions (Kennedy et al., 2000). In that position, it alternates chemical roles: playing the leaving group for 3' cleavage and hairpin resolution, and playing the nucleophile for hairpin formation and strand transfer. All other DNA strands—the donor DNA flanking the transposon, the 5' strand of the transposon, and the target DNA -- move in and out relative to the 3' oxygen. The anomalous requirement for an Rp configuration on the target DNA during strand transfer probably reflects a major change in the active site upon engagement of target DNA, but not necessarily a change that would require movement of the transposon's 3'OH. Most likely, the target DNA binds in a unique binding pocket, distinct from the pocket for the other DNA substrates, and a conformational change is required to bring the target-binding pocket to the active site.

The model suggests that the transposon's terminal 3' oxygen must be engaged at a specific location in the active site for the transposase to be active. As a result, every reaction performed in the active site directly involves that terminal oxygen. This model explains why 5' cleavage of the Tn10 DNA occurs via a hairpin intermediate involving the 3' strand (rather than direct hydrolysis of the 5' strand). It also explains how transposase avoids inappropriate hydrolysis of a target DNA, in which a water molecule replaces the transposon's 3' OH in attacking the target (Kennedy et al., 2000). In support of the model, I show in chapter 5 that if

the terminal 3'OH of the Mu DNA is artificially removed (through use of DNA fragments terminating in a dideoxy nucleotide), MuA can cleave a target DNA. Presumably, the 3' nucleotide of this dideoxy Mu DNA is properly engaged in the active site, and water substitutes for the missing 3'OH. I also show that the 3' terminal nucleotide of the Mu DNA is required for rapid cleavage of the target, consistent with steady binding in the active site.

The model also helps delineate new questions. For example, what triggers the movement of substrates in and out of the active site? Where does the target DNA bind, and how dramatic is the shift that moves the target DNA into position for strand transfer?

Assembly of the Mu transpososome

The transposase family shares certain features of transpososome assembly. Transpososome assembly involves synapses of two distant transposon ends, accomplished through multimerization of the transposase. The transposon ends are identified by inverted repeats: long recognition sequences located a set, short distance from a short cleavage sequence. Binding to the recognition sequences triggers initiation of assembly. But beyond such broad features, the details of each assembly process are unique to the particular transposon. Transpososome assembly is the first irreversible step of transposition (Mizuuchi et al., 1992), and therefore a critical point for control. Each assembly pathway has evolved to accommodate, or perhaps define, the particular regulatory needs of the transposon. In this section I will give an overview of the specifics of the Mu assembly process.

Transpososome assembly ensures that only the ends of the Mu DNA are selected as transposition substrates. This is achieved through the MuA recognition sites, three of which are located near each end of the Mu DNA (fig 2) (Baker and Mizuuchi, 1992). In chapters 2, 3 and 4 I discuss the role of the recognition sites in transpososome assembly. The recognition sites share a 22 base-pair consensus sequence -- long enough to be unique in a host genome (Craigie et al., 1984). Recognition of this unusually long site depends on a bipartite DNA binding domain, domains Iβ and Iγ of MuA (Schumacher et al., 1997). MuA binds specifically to its recognition sites, though initially exchanging rapidly between sites. Only under the proper conditions, the recognition sites allosterically activate assembly of a MuA tetramer that has a negligible k_{off} (Baker and Mizuuchi, 1992; Goldhaber-Gordon et al., 2002b). Nuclease protection assays reveal that only three of the six recognition sites – R1, R2 and L1 -- are bound stably in the transpososome, such that their protection pattern persists even after addition of a competitor DNA (Kuo et al., 1991; Lavoie et al., 1991; Mizuuchi et al., 1991). Nonetheless, under the most stringent reaction conditions all six sites are required for transpososome assembly (Allison and Chaconas, 1992), and under near-physiological conditions or *in vivo* five of the six sites are required (Allison and Chaconas, 1992; Baker and Mizuuchi, 1992; Groenen et al., 1985; Lavoie et al., 1991). Figure 2 and its legend gives more details about the recognition sites.

The cleavage sites, which mark the very ends of the Mu DNA, also have a required sequence: (T/A)CA↓. Transpososome assembly involves engagement of the cleavage sites in *trans* (see above), and cleavage site mutations can cause the assembly process to stall at an unstable, inactive intermediate (Watson and Chaconas, 1996). Mutant complexes that do succeed in completing assembly and performing DNA cleavage often fall apart after the cleavage (Coros and Chaconas, 2001; Lee and Harshey, 2001; Namgoong et al., 1994; Surette et al., 1991). Nonetheless, because the cleavage site sequence is short, the recognition sites are the primary identifying feature on the Mu ends. In fact, in chapter 3 I show that even if the cleavage site is displaced by ±1 base-pair from the R1 site, transposition can occur *in vitro.*

As I discuss in chapter 2, the MuA recognition sites modulate transpososome assembly in two distinct ways. They align the MuA subunits in proper position with respect to each other and to the DNA cleavage sites (Baker and Mizuuchi, 1992; Namgoong et al., 1994). They also allosterically activate individual MuA subunits to assemble cooperatively into a stable complex (Baker and Mizuuchi, 1992). In chapter 4 I show that once the transpososome is assembled, the

MuA recognition sequence does not have a major impact on transposition rate. Thus assembly is the primary step for determining that the DNA being transposed is, indeed, the Mu DNA.

The Internal Activating Sequence (IAS) helps control the timing of transpososome assembly *in vivo*. This DNA sequence is located 950 base-pairs from the Mu left end, and is unrelated in sequence to the end recognition sites. The IAS contains operator sequences that are recognized by both the most N-terminal domain of MuA (domain Iα) and by the Mu repressor. Interactions between MuA and the IAS stimulate transpososome assembly. When Mu repressor is present in the cell, it inhibits transposition by competing with MuA for binding to the IAS. Binding of Mu repressor at the IAS also turns-off transcription from the "early promoter" that regulates genes for lytic growth (including the genes for MuA and MuB). Thus Mu repressor maintains phage Mu as a lysogen, both through transcriptional control and through direct competition with transposase (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989).

The molecular basis of IAS stimulation of transpososome assembly is an active area of research, though not one that I discuss in the later chapters of my thesis. Under certain conditions, MuA-IAS interactions are absolutely required for transpososome assembly (Mizuuchi et al., 1995; Yang et al., 1995). Yet once a transpososome is assembled, the IAS is no longer required for transposition, and its interactions with the transpososome are detectably reduced (Mizuuchi et al., 1992; Surette and Chaconas, 1992; Watson and Chaconas, 1996). Consistent with this observation, kinetic studies suggest that the IAS behaves as a catalyst for assembly. However, the molecular or structural basis for this catalysis is not understood. The IAS functions after the initial synapses event, so it is not simply providing a scaffold for assembly (Mizuuchi and Mizuuchi, 2001). The orientation of the IAS relative to the Mu ends matters, and specific subunits in the MuA tetramer interact with specific operator sequences in the IAS (Allison and Chaconas, 1992; Jiang and Harshey, 2001; Jiang et al., 1999). Both of these results suggest that the IAS forms a structurally precise complex with MuA during transpososome assembly.

Additional proteins stimulate transpososome assembly. As mentioned above, MuB is one. *In vivo* experiments have implicated DNA gyrase in synapses of the two Mu ends (Pato and Banerjee, 1996). The DNA bending proteins IHF and HU are also involved in assembly (Craigie et al., 1985; Surette and Chaconas, 1989). IHF binds between two operator sequences within the IAS, and reduces a requirement for supercoiling in the donor DNA (Surette et al., 1989). HU may perform multiple roles during transpososome assembly (Lavoie and Chaconas, 1994), but the best understood involves the left end of the Mu DNA. On the right end of the Mu DNA, all three MuA recognition sites are adjacent to or nearly adjacent to one another (fig 2). The left end is different: L1, like R1, is positioned 5 base-pairs from the cleavage site, but L2 is separated from L1 by ~100 base-pairs. HU binds specifically between these sites, and probably helps bring L2 into close proximity with L1 (Lavoie and Chaconas, 1993).

Some of the unique aspects of Mu transpososome assembly are reminiscent of other assembly pathways. For example, the action-at-a-distance of the enhancer-like IAS, and the related requirement for DNA-bending proteins and/or supercoiling, are familiar from many systems. Examples of systems that use enhancer-like DNA elements and/or DNA bending are pre-transcriptional promoter complexes, the Hin inversion complex, the lambda intasome that promotes lambda integration or excision, and the Tn10 transpososome (Azaro, 2002; Finkel and Johnson, 1993; Hallet and Sherratt, 1997). The roles of IHF and DNA supercoiling during assembly of a Tn10 transpososome have been particularly well characterized (Crellin and Chalmers, 2001; Chalmers et al 1998).

Nonetheless, the details of Mu assembly are unique to Mu. The function of the IAS is particular to the biology of phage Mu, and other transposons do not contain internal activator sequences. The two ends of the Mu DNA are asymmetric, and recent results show that the asymmetry impacts the order in which components join an assembling complex (personal

communication, K. Kobrym, M. A. Watson, R. Allison and G. Chaconas). Yet some other transposons, such as Tn3, contain perfectly symmetrical ends (Wishart et al., 1985). The role of MuB in assembly may not be unique to Mu, as the TnsC protein could play a similar role during assembly of a Tn7 transpososome, but MuB analogs are certainly not ubiquitous among transposons (Stellwagen and Craig, 1997). Even the multimeric state of the transpososome – tetramer vs. dimer – is particular to the individual transposon. And just as the Mu transpososome has its special features which presumably relate to the particulars of its biology, other transposons, too, have evolved particular controls for their assembly pathways. These are beyond the scope of this chapter.

Disassembly of the transpososome

MuA is not an enzyme, because it is not a true catalyst. This was first suggested by elegant *in vivo* studies showing that MuA is used stoichiometrically during transposition (Pato and Reich, 1984). Later *in vitro* studies confirmed that MuA is not used catalytically, because it remains stably bound in the transpososome after it completes transposition. In fact, the stability of the transpososome increases during the initial stages of transposition. The pre-cleavage Stable Synaptic Complex (SSC or type 0 complex) is less stable than the Cleaved Donor Complex (CDC or type 1 complex), which is less stable than the Strand Transfer Complex (STC or type 2), as measured by resistance to heat or to urea (Mizuuchi et al., 1992; Surette et al., 1987). Though Mu transpososomes are among the best studied, highly stable complexes appear to be characteristic of transpososomes in general (Sakai et al., 1995).

Transpososome stability could have evolved for a number of reasons. As discussed above, stable transpososomes help ensure coordinated transposition of the two DNA ends. Stable transpososomes also help ensure that the transposase remains present to complete strand transfer once having performed donor cleavage. This is particularly important during the cut & paste transposition of elements like Tn7, Tn10 and perhaps Mu during integration: if the transpososome were to fall apart after completion of both 3' and 5' cleavage, the transposon would be lost from the cell. The increase in stability between CDC and STC ensures that strand transfer, an otherwise isothermic reaction, only moves forward. And finally, active disassembly of the transpososome signals for initiation of specific replication of the Mu DNA, the final step of Mu transposition (Jones et al., 1998; Kruklitis and Nakai, 1994; Kruklitis et al., 1996; Nakai and Kruklitis, 1995). Thus disassembly is another opportunity for control of transposition.

The Mu transpososome is disassembled by the chaperone ClpX, a member of the Clp/Hsp100 family of ATPases (Levchenko et al., 1995; Mhammedi-Alaoui et al., 1994). ClpX uses the energy of ATP hydrolysis to unfold protein substrates. In many contexts, ClpX works together with the protease ClpP, feeding ClpP unfolded substrates for degradation (Gottesman et al., 1993; Wojtkowiak et al., 1993). However, *clpP* null cells can support Mu replicative transposition and *clpX* null cells cannot, suggesting that transpososome disassembly requires unfolding but not degradation of MuA subunits (Mhammedi-Alaoui et al., 1994). MuA subunits released from a transpososome by ClpX are detectably unfolded, and transpososomes in which only a single MuA subunit possesses the ClpX recognition sequence are disassembled by ClpX, suggesting that unfolding of one subunit can destabilize an entire transpososome (Burton et al., 2001).

ClpX and MuB interact with overlapping sequences in the C-terminus of MuA. This arrangement may hint at yet another method of transpososome-dependent regulation. It seems likely that competition between MuB and ClpX may regulate the activities of both (Levchenko et al., 1997).

To my knowledge, no specific chaperone or protease has been shown to be required for Mu integration, or for transposition of any other transposon. Yet other transposases do form stable transpososomes that do not turn-over catalytically *in vitro* (Sakai et al., 1995). It is possible that other transpososomes can be disassembled by any of several proteins, and hence no

single one is required. It is also possible that the stringency of the requirement for ClpX during Mu replicative transposition is due to the intense frequency of that reaction. During lytic growth, the Mu DNA transposes 50-200 times in one hour; typical transposition frequencies for non-phage bacterial transposons are on the order of 1 for 10^{-3} to 10^{-10} cell generations (Arthur et al., 1984; Shen et al., 1987). Perhaps a slower rate of transposition permits greater flexibility in the disassembly method.

***In vitro* studies of retroviral integrases indicate the value of the transpososome in controlling transposition.**

In vivo, integration of a retroviral cDNA into the chromosome of a new host cell occurs within a "Pre-Integration Complex", or a "PIC". This is a large complex, containing many protein components including the integrase and also the viral cDNA (Bowerman et al., 1989). So far, it has not been possible to reconstitute *in vitro* an integration reaction that fully resembles the *in vivo* process, presumably due to the complexity of the PIC. For example, several studies indicate that retroviral integrases can form synaptic complexes, and must do so for activity *in vivo* (Carteau et al., 1999; Chen and Engelman, 2001; Kukolj and Skalka, 1995; McCord et al., 1999; Murphy and Goff, 1992; Wei et al., 1998). Yet *in vitro*, integrases can bind to a DNA fragment containing a single end sequence, and transpose that fragment by itself (Bushman and Craigie, 1991). MuA and other transposases can also transpose DNA fragments containing individual end sequences, but to do so they pair two such fragments in a synaptic complex, mimicking the two ends of the transposon (Bhasin et al., 2000; Sakai et al., 1995; Savilahti et al., 1995).

In vitro, integration reactions are poorly controlled in many ways, and the contrast between integration and transposition reactions can give us appreciation for the importance of the transpososome as a means of control. For example, integrases can perform an *in vitro* "disintegration" reaction, which appears to be a direct reversal of the strand transfer reaction (Chow et al., 1992). Attempts to get MuA to do disintegration have failed, perhaps in part because the STC (Strand Transfer Complex) is the most stable form of the transpososome, and therefore the lowest energetic state available to a MuA-DNA complex (personal communication, P. Rice and K. Mizuuchi). As another example, integrases are remarkably promiscuous in their choices of nucleophile. They can cleave a "target" DNA molecule using many different alcohols as attacking groups, and they can do so in the absence of an obvious donor DNA (Katzman and Sudol, 1996; Skinner et al., 2001). As discussed in chapter 5, the requirement for a transpososome, which includes both Mu DNA ends, inhibits MuA from cleaving a non-specific target DNA. Thus the requirement for a synaptic complex contributes to a smooth transposition process.

Conclusion

In theory, each reaction underlying transposition could be catalyzed by an enzyme. Many types of restriction enzymes, homing endonucleases, and other enzymes catalyze site-specific cleavage reactions. Enzymes that catalyze true strand transfer reactions are less common, but the reactions catalyzed by topoisomerases and ligases are each similar to strand transfers. Yet rather than a series of enzymes, transposition is performed by a stable transpososome, providing a platform for a well-ordered, regulated, multi-step process.

To illustrate further: the transpososome that mediates Tn7 transposition contains four different types of protein subunits. One of them, TnsA, is responsible for cleavage of the transposon's 5' strand (Sarnovsky et al., 1996). The crystal structure of TnsA shows it to be closely related to type II restriction enzymes, suggesting that TnsA evolved from a restriction enzyme (Hickman et al., 2000). Yet today TnsA functions only as part of a stable transpososome, in complex with a DDE family transposase TnsB (Sarnovsky et al., 1996). Thus 5' cleavage of Tn7 occurs only in the framework of 3' cleavage and strand transfer, ensuring complete transposition of the transposon.

Figure Legends. Schematics are not drawn to scale.

Figure 1. Transposases in general, and MuA in particular, mediate two types of reactions involving the 3' oxygens at the ends of the transposon DNA. The cleavage reaction is a hydrolysis, and the stand transfer reaction is a transesterification. The fate of the DNA strands during these reactions is summarized on the left, and the phosphoryl groups that are directly involved in the chemistry are depicted on the right. The transposon, or Mu DNA, is shown in grey; non-specific sequences are in black.

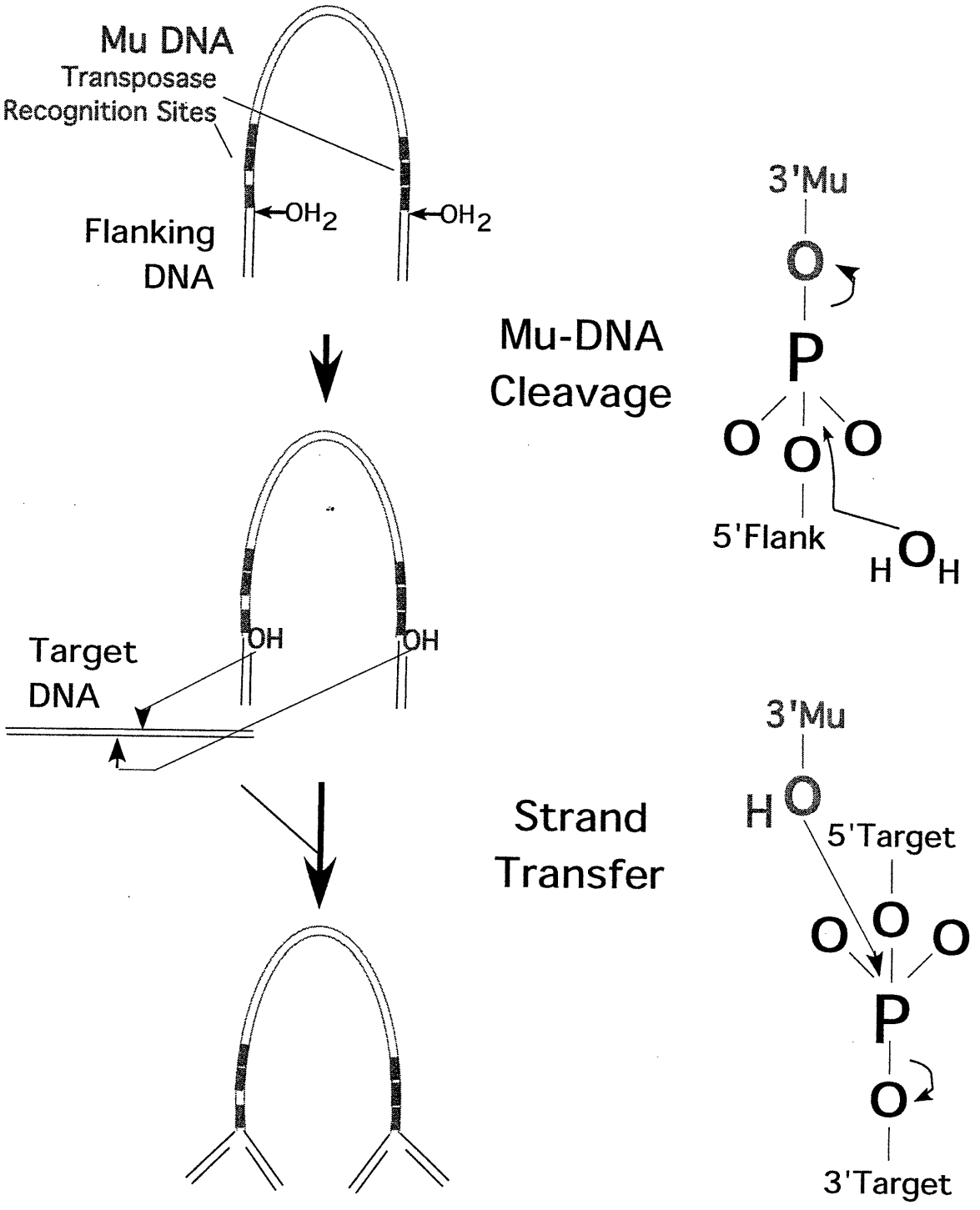
Figure 2. The core Mu transpososome consists of a MuA tetramer, whose subunits are bound to MuA recognition sites near the ends of the Mu DNA. One MuA subunit is shown in white, to highlight bridging of the two transposon ends. The DNA cleavage sites are distinct from the recognition sites, and the 3bp cleavage sequences are spelled-out in the figure. The recognition sites are depicted as blocks, and their names are written next to each block. The L2 site is depicted as a small block, because it lacks ~half of the 22 base-pair recognition consensus. The relative binding affinities for the sites are (roughly): L1=L3=R3>R2>R1>L2 (Craigie et al., 1984). The R1 and R2 sites are positioned in inverted orientation relative to the other four sites, and the inversion is a requirement for transposition (Craigie and Mizuuchi, 1986).

Figure 3. The catalytic domains of transposases and retroviral integrases share a common fold, similar to the folds of RNaseH and RuvC. The individual structures shown here are: a subunit of Tn5 transposase taken from the synaptic-complex structure (PDB accession code 1F3I) (Davies et al., 2000); the RNase H domain of HIV reverse transcriptase (RT), taken from an RT–DNA complex structure (PDB accession code 1RTD) (Huang et al., 1998); the core fragment of MuA transposase (PDB accession code 1BCM) (Rice and Mizuuchi, 1995); and the catalytic and C-terminal domains of Rous Sarcoma Virus (RSV) integrase (PDB accession code 1C0M) (Yang et al., 2000). The figure itself is taken from a review article (Rice and Baker, 2001). Common structural features are shown in rainbow colors, the acidic catalytic residues are shown in ball-and-stick format, and divalent metal ions are shown in pink.

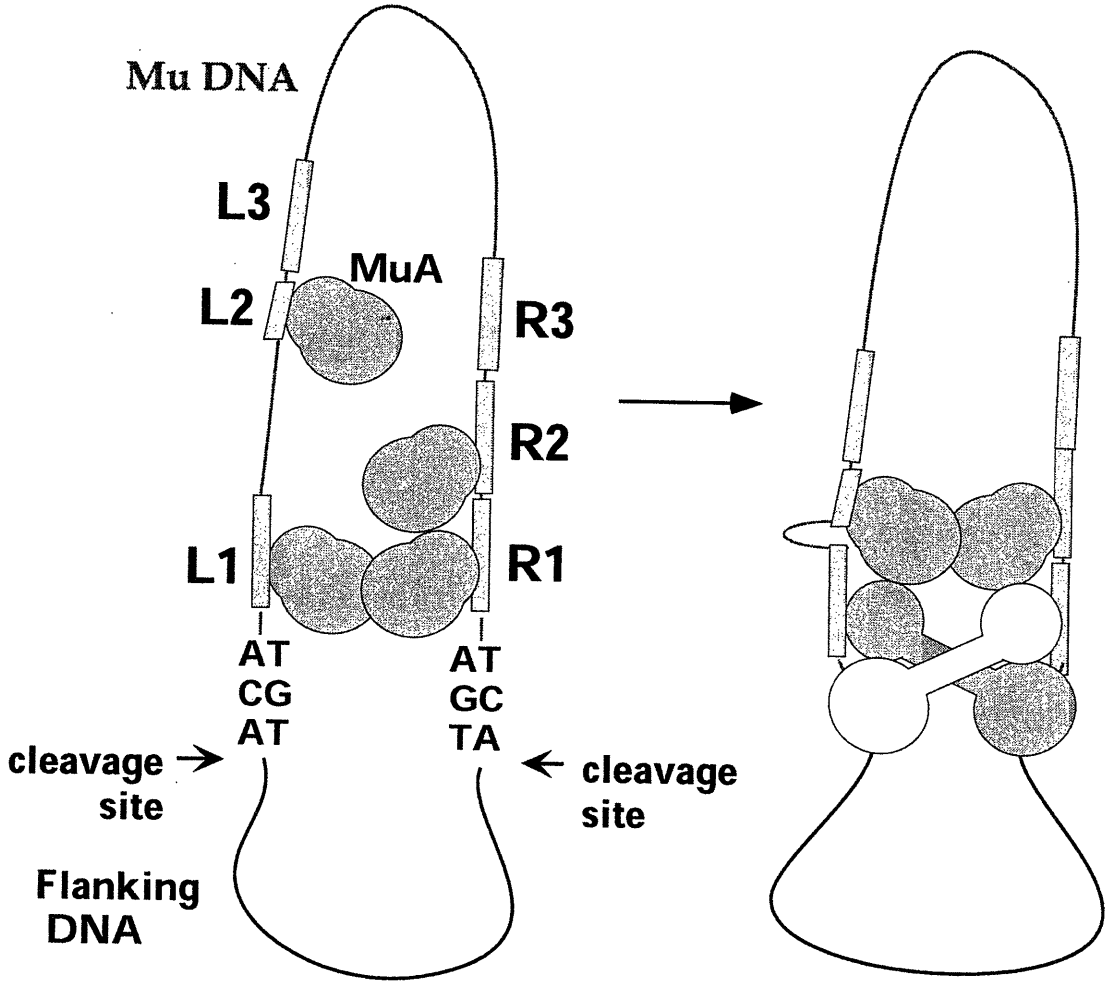
Figure 4. Limited proteolysis divides MuA into three structural domains, and functional analyses further divide each domain. The numbers at the bottom of the figure are amino-acid sequence numbers. Domain Iα binds the Internal Activating Sequence, an enhancer-like DNA sequence that stimulates transpososome assembly. Domains Iβ and Iγ each bind half of the end-DNA recognition sequence. Domain IIα is the catalytic core, whose structure is shown in figure 3. Domain IIβ may also be involved in catalysis, though it is not present in other family members. It is also part of the Mu structure shown in figure 3. Domain IIIα possesses cryptic nuclease activity and may interact with cleavage site DNA. Domain IIIβ interacts with the proteins MuB and ClpX.

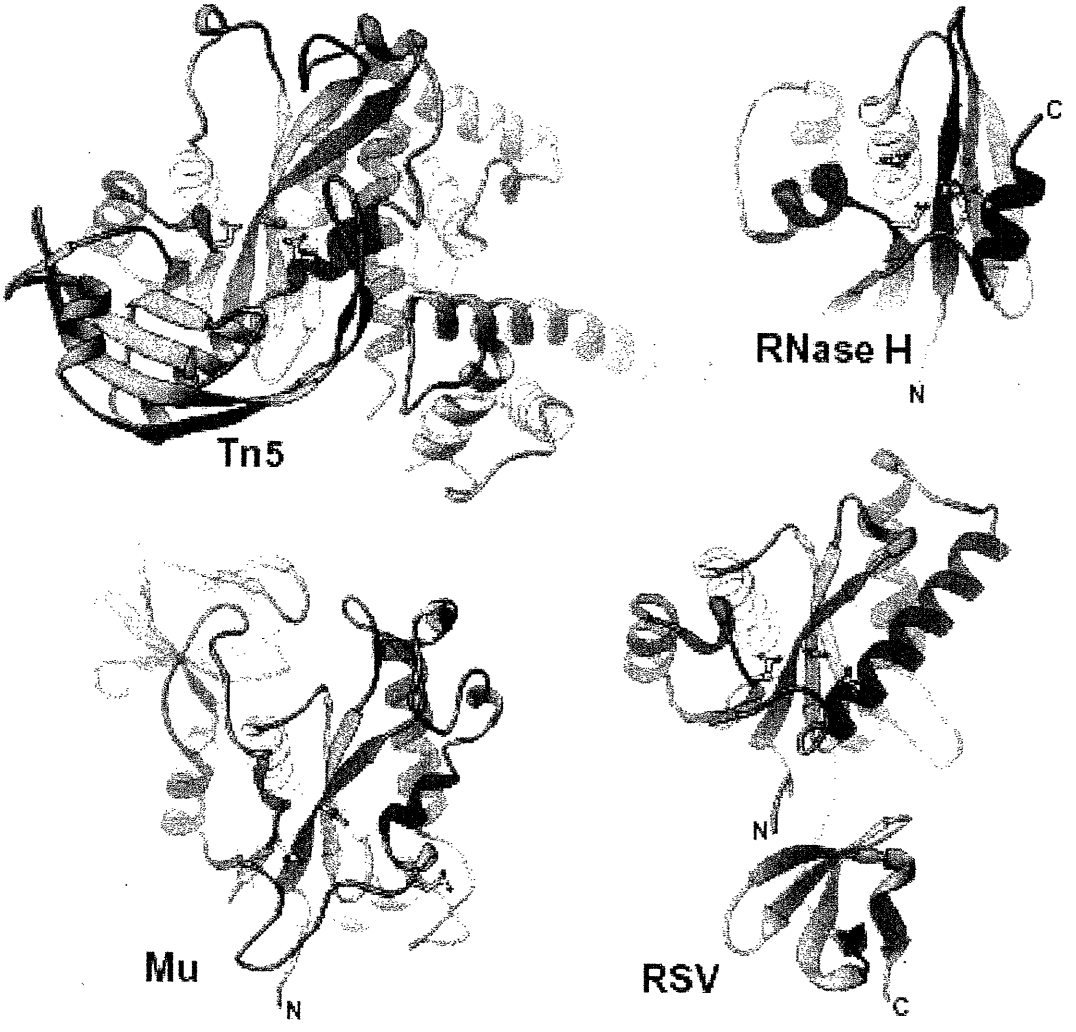
Figure 5. MuA establishes target immunity, by clearing MuB from nearby DNA sequences. MuB then stimulates MuA to select as a target DNA those molecules to which MuB is still bound.

Figure 6. This is the structure of a synaptic complex from Tn5 transposase (Davies et al., 2000). The structure represents a post-cleavage complex: neither the 3' nor the 5' strand of the transposon DNA fragments is attached to a flanking sequence. The DNA at the ends of the fragments is distorted into a hairpin-like structure, though the hairpin is not covalently closed. A region which is present in Tn5 transposase but not in MuA or in retroviral integrases is shown in dark grey and labelled "insert". This figure is taken from a review article (Rice and Baker, 2001) and modified slightly.

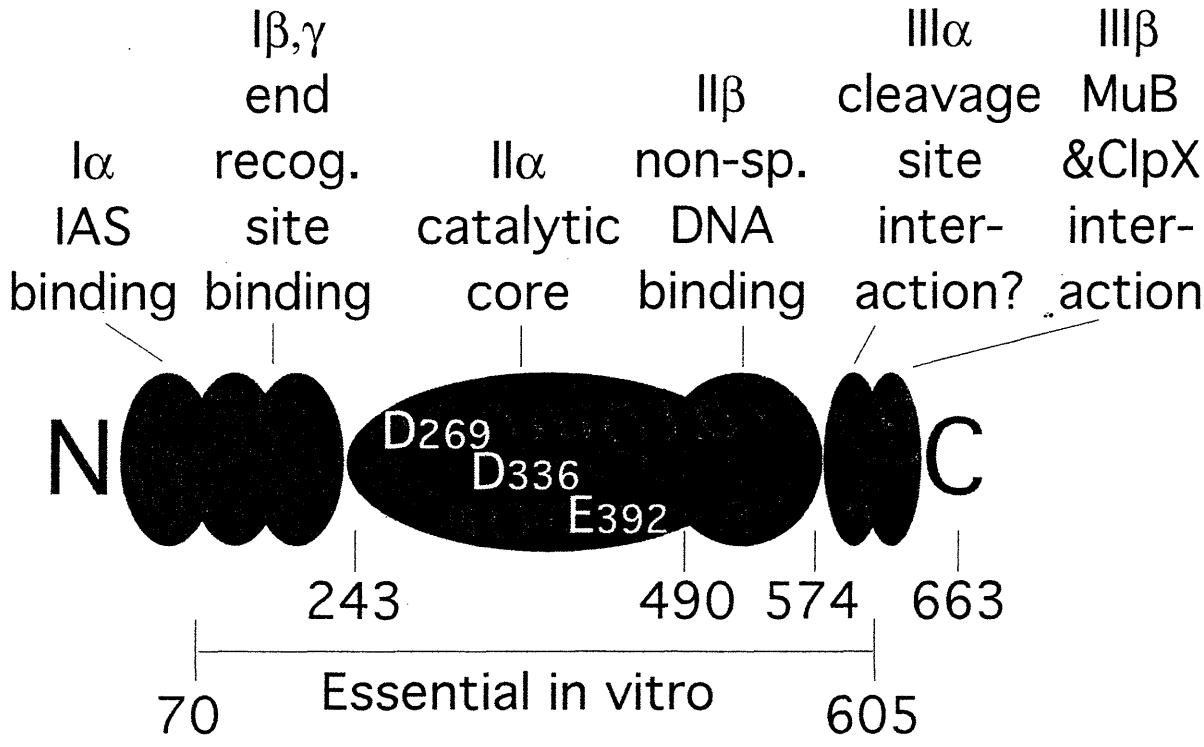


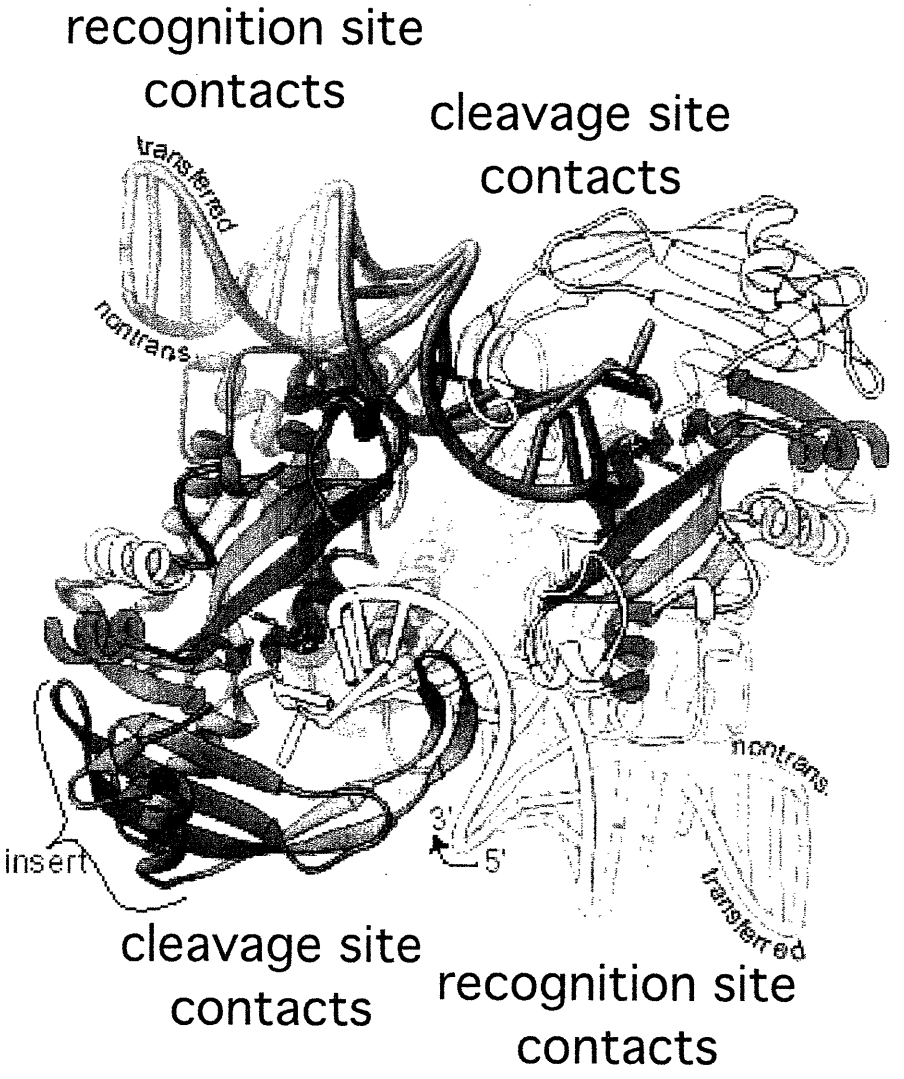
Introduction Figure 1
page 23


Donor DNA
Mu DNA
L3
MuA
L2
R3
R2
L1
R1
AT
CG
AT
AT
GC
TA
cleavage site
cleavage site
Flanking DNA



Introduction
Figure 3
page 25


Iβ,γ
end
recog.
site
binding
IIβ
non-sp.
DNA
binding
IIIα
cleavage
site
inter-
action?
IIIβ
MuB
&ClpX
inter-
action
Iα
IAS
binding
IIα
catalytic
core
N
C
D269
D336
E392
243
490
574
663
70
Essential in vitro
605



Introduction
Figure 6
page 28

Introduction

Transposons are found in all the biological kingdoms, and some perform specialized functions. For example, the machinery that initiates V(D)J recombination likely evolved from a transposon (Agrawal et al., 1998; Hiom et al., 1998), and the cDNA of HIV and other retroviruses integrate into host cell DNA through mechanisms nearly identical to transposition (Haren et al., 1999). The genome of bacteriophage Mu is a transposon, that uses transposition both to integrate into the DNA of a new host cell and to replicate before lysis. Like most DNA rearrangements, transposition is a complex, multi-step process, requiring numerous DNA sequence elements. Studies of bacteriophage Mu have been central to our understanding of both the fundamental mechanisms and the complexities of DNA transposition.

Phage Mu encodes a transposase, MuA, that transfers the Mu genome from one DNA location (the transposition donor) to a new location (the transposition target) (Chaconas, 1999; Mizuuchi, 1992). During transposition, transposase performs two principle reactions: DNA cleavage and DNA strand transfer. During cleavage, the donor DNA is nicked twice, once at each 3' end of the Mu genome. During strand transfer, the cleaved transposon ends are inserted into neighboring sites on the two target strands.

Little or no specific sequence information is needed on the target DNA (Mizuuchi and Mizuuchi, 1993), but the Mu DNA provides many sequence cues for transposition (Fig. 1). For example, the last two nucleotides at either 3' end of the Mu DNA, the cleavage sites, have the sequence 5'CA. Also near each end of the Mu DNA are three recognition sites, distinct from the cleavage sites, which share a 22 base-pair consensus sequence. The recognition sites are referred to as R1, R2 and R3 on the right end and L1, L2 and L3 on the left end (Fig. 1) (Craigie et al., 1984). The recognition sites are bound specifically by the N-terminal domain of MuA (Nakayama et al., 1987), whereas the cleavage sites must be engaged by the protein's active site, contained in a different region of the protein (Rice and Mizuuchi, 1995). Both the recognition sequences and the 5'CA cleavage sequences are required for transposition (Baker and Mizuuchi, 1992; Coros and Chaconas, 2001; Groenen and van de Putte, 1986).

Both the cleavage and strand transfer reactions occur within a stable MuA/DNA complex called a transpososome (Fig. 1). Because the transpososome binds both transposon ends simultaneously, it is also referred to as a synaptic complex. The complex contains three MuA subunits tightly bound to recognition sites (the R1, R2 and L1 sites) plus a fourth subunit tightly bound in the complex but weakly bound to the L2 recognition site (Kuo et al., 1991; Lavoie et al., 1991; Mizuuchi et al., 1991). At least two of the four subunits individually bridge the transposon ends: the subunit bound to the right-end's R1 site engages the left cleavage site, and the subunit bound to the left-end's L1 site engages the right cleavage site (Aldaz et al., 1996; Namgoong and Harshey, 1998; Savilahti and Mizuuchi, 1996; Williams et al., 1999). This crisscross structure helps coordinate reactions at the two DNA ends, ensuring that the transposon moves as a single unit. Given this intertwined structure, and also the intimate involvement of multiple DNA sites in the complex, it is unknown to what extent protein-protein versus protein-DNA interactions contribute to the stability and functionality of the synaptic complex.

In vitro, transpososomes can also assemble on ~50 base-pair DNA fragments, containing the R2, R1 and cleavage site sequences. Two of these "donor fragments" are synapsed by a MuA tetramer, mimicking synapsis of the two ends of a transposon (Fig. 2a) (Savilahti et al., 1995). The fragments are then cleaved at the proper cleavage site, and transferred together to a target DNA. If the fragments are synthesized to appear pre-cleaved, cleavage by MuA is unnecessary prior to strand transfer (Fig. 2b). Because the fragments are small relative to the target, the resulting transposition product comigrates with linear target. Transpososomes formed on donor fragments, like those formed on larger DNA molecules, can resist competition from additional recognition sites for hours or perhaps days. In contrast, monomeric MuA has a rapid on and off

rates from its recognition sites (unpublished results, and (Savilahti et al., 1995; Surette et al., 1987)).

How does transposase change from a form that is monomeric, with a rapid dissociation rate from DNA, into a stable, active synaptic complex? This process involves interactions between at least four multi-domain subunits and multiple DNA sequences, and little is understood about the role of each component in the complex. For example, although the recognition sites are essential, their precise role during assembly is uncertain.

Here, we use a promiscuous activity of MuA to elucidate the functions of MuA recognition sites during transposition. We find that MuA recognition sites can activate MuA to transpose non-Mu DNA. Two covalently linked MuA recognition sites are the strongest activators of transposition, but unlinked sites can also serve in this capacity. Our results suggest that MuA recognition sites perform at least two complementary functions during transposition. Covalently linked sites promote transpososome assembly by spatially constraining two of the four MuA subunits. In addition, each individual recognition site activates transposition, probably by inducing conformational changes in the transposase.

Experimental Procedures

Proteins. In some experiments (Figs. 7b, 7c and 8) the MuA truncation 77-663 was substituted for full-length MuA. This was done because, in the absence of the Internal Activating Sequence (a DNA sequence on the Mu phage that activates transpososome assembly), the truncation protein is hyperactive compared to the full-length protein. The products discussed in Fig. 8 were seen previously using full-length protein (Baker et al., 1994). MuA, $MuA_{77\text{-}663}$ (Baker et al., 1993) and MuB (Yamauchi and Baker, 1998) were prepared as described. Restriction enzymes were purchased from New England Biolabs.

DNA. ϕX174 RF1 was purchased either from Gibco-BRL or New England Biolabs. Oligonucleotides (fragment donors) were purchased either from MIT/HHMI biopolymers lab or from GeneLink, and were purified by denaturing PAGE, except for those less than 30 nucleotides which were purified by reverse phase cartridges. Most oligonucleotide sequences are described in Fig. 2b. However, over the course of this study, we used several different "unjoinable fragments", that differed at 3 non-essential nucleotide positions and/or in the length of the 5' overhang on the non-transferred strand. We saw no qualitative difference in the behavior of these various fragments (data not shown). The unjoinable fragment used in Fig. 3a, was: agtgaagcggcgcacgaaaaacgcgaaagcgtttcacgaaaaacgcgaaagcg/cgctttcgcgtttttcgtgaaacgcttt cgcgtttttcgtgcgccgcttc. The fragment used in Fig. 3b, was: gcatgaagcggcgcacgaaaaacgcgaaagcg tttcacgataaatgcgaaaac/gttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgcttc. All other figures used the unjoinable fragment listed in Fig. 2b. The full sequence of the cleaved strand of the uncleaved fragment is: cgttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgcttcactagacgcttggcgtaatcgggcgtaatgc. The experiment in Figs. 4c and 6b (squares) used the following DNA fragment, to maintain the total fragment DNA at 1440 nM:
gccggtatctttccagcactgggccggtatctttccagcactggcg/cgccagtgctggaaagataccggcccagtgctggaaagataccg gc. The experiment in Fig. 6b (triangles) contained the following fragment to maintain total fragment DNA at 1440 nM: gccggtatctttccagcactgg/ccagtgctggaaagataccggc

Transposition reactions. Unless indicated otherwise, reactions were conducted in a 25 µl volume containing 25 mM Tris-HCl (pH 8 at room temperature), 140 mM NaCl,10 mM $MgCl_2$,1 mM DTT, 0.1 mg/ml BSA, 15% glycerol, 12% DMSO, 0.1 % triton, 2mM ATP, 340 nM MuB and variable amounts of Mu DNA fragments and MuA. Proteins were prepared by dilution of concentrated stock into the following buffers: MuA — 600 mM NaCl, 25 mM Hepes-KOH, 0.1 mM EDTA, 10% glycerol, 1 mM DTT; MuB — 1M NaCl, 25 mM Hepes, 0.1 mM EDTA, 20% glycerol, 1 mM DTT. The reactions were incubated at 30°C for 20 to 60 minutes unless otherwise indicated. They were stopped by addition of 0.2 volumes of a stop solution (~0.1% bromphenol blue, 8% SDS, 50 mM EDTA, 30% glycerol), and electrophoresed through a 0.9%

HGT agarose gel, in 1x TAB buffer (40 mM Tris pH 8, 3.6 mM EDTA, 27 mM sodium acetate). The gels were then analyzed by one of the following methods: (1) Stained in 1 μg/ml ethidium bromide and photographed with a Polaroid camera, (2) Stained in a 1:10,000 dilution Vistra Green in TAB buffer and visualized on a molecular dynamics Flourimager 595, (3) For radioactive samples, the gels were pressed and dried on DEAE paper, and exposed to a molecular dynamics phosphor imager cassette. The plasmid assays, described in Fig. 8, were performed in two steps. The first incubation was as described (Baker et al., 1994) (reactions contained neither DMSO nor triton), except that the reaction volume was 50 μl and incubation time was two hours. For the second step, either 1 μl restriction enzyme or 1 μl 50% glycerol was added to reactions, and they were incubated for another 2 hours at 37°C. The second incubation period increased the amount of intramolecular ϕX174 products.

2D gels. In the first dimension, electrophoreses was as described above. A single lane was then excised and positioned horizontally across the top of a 0.8% HGT agarose gel in alkaline buffer (30 mM NaOH, 1 mM EDTA) for electrophoreses in the second dimension. The gel was analyzed by southern blotting: DNA was transferred to an NEN Genescreen hybridization membrane, and probed with labeled DNA fragments that had been randomly primed off of ϕX174 RFI DNA.

Gel mobility shift assays. Complexes were assembled under standard transposition conditions, in the absence of $MgCl_2$, target DNA and MuB. They were electrophoresed through a 2% MetaPhor (FMC) agarose gel in 0.5x TBE buffer (44.5 mM Tris 8.5, 44.5 mM Borate, 1 mM EDTA).

Results

A non-Mu DNA molecule serves as a transposition donor.

In transposition reactions containing limiting concentrations of Mu-derived DNA, non-Mu DNA molecules were seen to transpose (Fig. 3). There were three principle requirements for this promiscuous activity: (1) MuA protein, (2) a large, non-Mu DNA molecule to serve as a transposition donor. We focused on the 5386 base-pair ϕX174 RFI, but two molecules unrelated to ϕX174, pBR322 or pUC19, worked about equally well (data not shown), and (3) MuA recognition sites supplied *in trans* on another DNA molecule.

In our initial experiments, the MuA recognition sites were supplied on a fragment that is itself defective in the strand transfer step (called "unjoinable", see Fig. 2b). This unjoinable fragment has two MuA recognition sites (R1 and R2) and an incomplete cleavage site. The unjoinable fragment and the non-Mu DNA complemented each other. The fragment supported transpososome assembly, but failed to transpose due to its defective cleavage site (see below, and (Aldaz et al., 1996)). The non-Mu DNA cannot on its own support transpososome assembly (Baker and Mizuuchi, 1992), but it did transpose.

Incubation of the unjoinable Mu fragment, MuA, the MuB protein (see below), and a target DNA (ϕX174 RFI) resulted in the formation of two types of transposition products. One family of products electrophoresed between the nicked and linear target DNAs (Fig. 3a, lanes 1 and 2, labeled "intramolecular"). The second product ran slightly above dimeric target (Fig. 3a, lane 1, labeled "intermolecular"). Both were true recombination products, rather than non-covalent rearrangements trapped by proteins: the samples in Fig. 3a were treated with 1% SDS, and the products were also unaffected by treatment with proteinase K or by phenol extraction (data not shown). The Mu fragment was a critical component of the reaction (lane 5).

Formation of the slower-migrating product (labeled "intermolecular") required the MuB protein (Fig. 3a, lane 2). MuB is a DNA binding protein that stimulates MuA and controls MuA's target selection (Maxwell et al., 1987). For example, when transposition of Mu donor plasmids is assayed in the absence of MuB, the products are usually intramolecular: the Mu sequences define the donor sites, and another site on the same plasmid serves as target site. In the presence of MuB, transposition is usually intermolecular: the Mu plasmid serves as donor, a separate non-

Mu DNA molecule serves as target, and the two molecules become joined to each other through transposition. Here, the slow electrophoretic mobility of the MuB-dependent product (Fig. 3a, labeled "intermolecular") strongly suggested it contained two ϕX174 molecules. This product did not contain Mu DNA (data not shown, results of radiolabeling the Mu donor fragment). Therefore, this MuB-dependent product was most likely the result of an aberrant intermolecular reaction, in which two molecules of ϕX174 were joined together. The MuB-independent product (labeled "intramolecular") could result from intramolecular recombination, in which a single ϕX174 molecule provided both donor and target sites.

Two-dimensional gel electrophoresis confirmed the structures of the new products (Fig. 3b). Transposition reactions were performed with MuA, MuB, ϕX174 and a Mu fragment. The DNA products were resolved on an agarose gel similar to that shown in Fig. 3a. A lane was then excised from the gel (shown lying horizontally in Fig. 3b), and electrophoresed through alkaline (denaturing) agarose. In the denaturing dimension, the intermolecular product yielded a slow-migrating species and a species that co-migrated with single-stranded circular ϕX174 DNA. These are the expected components of a figure-eight-shaped transposition product, in which two full-length, double-stranded ϕX174 molecules are joined by a single strand. The MuB-independent products also yielded a slow-migrating component in the second dimension. This long single-stranded species could result from an intramolecular transposition in which a strand was joined to its complement. Intramolecular transposition products should also yield smaller components of varying mobilities, depending on the exact insertion site. Since these smaller products do not migrate as a distinct band, they are difficult to distinguish on the 2D gel.

Together, these data show that a Mu DNA fragment permits a non-Mu DNA molecule (in this case ϕX174) to participate as a donor DNA during transposition.

Hybrid transpososomes synapse one Mu fragment and a ϕX174 molecule.

It is likely that Mu fragments promote non-Mu transposition, at least in part, by promoting assembly of synaptic complexes. In support of this argument, unjoinable Mu fragments are known to support assembly of MuA tetramers that are stable to high salt and to competitor DNA, a diagnostic of a synaptic complex (Aldaz et al., 1996; Williams et al., 1999). ϕX174 DNA alone does not support assembly of detectable levels of MuA tetramer (Baker and Mizuuchi, 1992).

We considered two models for synaptic complexes that might transpose non-Mu DNA. In the "two-fragment model", two Mu fragments form a synaptic complex, similar to the one shown in Fig. 2a. A ϕX174 molecule associates with this pre-assembled synaptic complex, contributing a few nucleotides to serve as the cleavage site. In the alternative "hybrid complex" model, synaptic complexes form between a single Mu fragment and a site on ϕX174 (Fig. 4a). This model suggests that ϕX174 DNA functions both structurally and chemically as one of the two transposon ends.

The two-fragment model posits that synapsing two Mu DNA fragments is a prerequisite for non-Mu transposition. By contrast, the hybrid-complex model predicts that synapsis of two fragments competes with assembly of hybrid transpososomes. In support of the hybrid complex model, pre-incubating MuA with an unjoinable Mu fragment inhibited ϕX174 transposition (Fig. 4b, lane 4). The reactions shown here were primarily intermolecular, but in reactions without MuB, intramolecular transposition was also inhibited by preincubation with the Mu fragment (data not shown). Pre-incubating MuA alone (lane 1) or with ϕX174 (lane 3) did not significantly inhibit subsequent reactions. Most likely, then, functional complexes are hybrids of a single Mu fragment and a ϕX174 molecule.

The hybrid-complex model also predicts that high concentrations of Mu fragments should always inhibit non-Mu transposition, due to assembly of complexes on pairs of fragments. To test this prediction, we performed transposition assays in which we varied the concentration of a Mu fragment (in this case R1-R1, see Fig. 2b), but maintained the total DNA concentration

constant using a DNA fragment of unrelated sequence but identical length. The yield of transposition products peaked sharply at 8 nM Mu fragment (Fig. 4c), substantially below the MuA concentration of 50 nM. These results further support the hybrid-complex model. This experiment also showed that it is the Mu recognition sequence, as opposed to a general feature of short DNA fragments, that both activates and inhibits non-Mu transposition.

Two additional observations support the hybrid complex model. First, stable complexes of unjoinable fragment, ϕX174 DNA and MuA had been seen previously (Aldaz et al., 1996). ϕX174 was assumed to be bound as target DNA in such complexes, but we now suggest that it is also bound as a donor partner. Second, in chapter 3 I demonstrate that the regions of ϕX174 that participate in transposition, when viewed as a group, bear weak but significant resemblance to a Mu DNA end (Goldhaber-Gordon et al., 2002a).

The cleavage site is optional on the Mu fragment.

Recall that the Mu transposon ends include two types of specific DNA sequences: recognition sites and cleavage sites. The unjoinable and R1-R1 fragments discussed above are, respectively, partially or entirely lacking a cleavage site sequence (Fig. 2b). We found that donor fragments with functional, uncleaved or precleaved cleavage sites also activated non-Mu transposition (Fig. 5, lanes 2-7). These fragments produced the figure-eight intermolecular product described above, plus a novel intermolecular product. This novel product was probably due to transposition of a donor fragment together with the non-Mu DNA (see figure legend for details).

Precise, quantitative comparisons of the activities of different fragments was difficult, due to differences in the types of products each produced. But as a rough estimate, the uncleaved and precleaved fragments seemed to be less active in non-Mu transposition assays than were either the unjoinable or the R1-R1 fragment. Of all the fragments tested, the R1-R1 fragment seemed the most potent activator of non-Mu transposition (Fig. 5, lanes 11-13). This suggests that a cleavage site on the Mu DNA is at best neutral, and possibly inhibitory, to non-Mu transposition.

Fragments containing a single recognition site support non-Mu transposition.

We next asked whether the fragments needed to contain two recognition sites (positioned as the natural R1 and R2 sites) to activate non-Mu transposition. We expected that the two sites contribute to transpososome stability by correctly positioning two of the four MuA subunits with respect to each other. To determine whether this contribution is an essential one, we constructed fragments containing a single R1 site (see Fig. 2b). These R1 fragments were 22 base-pairs long, containing half the sequence of the 46 base-pair R1-R1 fragment. This minimal substrate was sufficient to permit non-Mu transposition (Fig. 6a). To observe activity with this fragment, we did require longer incubation times, and higher fragment concentrations, than with the fragment with two sites (data not shown and Fig. 6b). Nonetheless, the intermolecular ϕX174 transposition product was clearly visible, and present only in reactions containing the Mu fragment (Fig. 6a, see figure legend). We also confirmed that a non-specific sequence could not substitute, by varying the concentration of the R1 fragment with a fragment of identical length but unrelated sequence (Fig. 6b). These results point to a powerful role of the recognition sequence in activating non-Mu transposition. They suggest a specific stimulatory effect of the recognition sequence on MuA transposase, independent of the physical positioning of subunits provided by two linked recognition sites.

Hybrid complexes are stable to competitor DNA.

Transpososomes that synapse two native Mu DNA ends are extremely stable (Surette et al., 1987). Several lines of evidence suggest that both the recognition sequences and the cleavage sequence contribute to this stability through interactions with MuA (Baker and Mizuuchi, 1992; Namgoong et al., 1994). However, it is unclear to what extent stability correlates with functionality, and we wondered whether hybrid transpososomes are long-lived or transient. Our interest in this question was part of two larger questions. (i) To understand the physical basis of

transpososome stability: do non-Mu sequences permit assembly of long-lived transpososomes? (ii) To understand the energetics of transposition: does a transpososome need to be long-lived to be functional?

Both DNA mobility-shift assays and transposition assays were used to probe the stability of hybrid transpososomes. DNA mobility-shift (or band-shift) assays were necessary because some of the relevant complexes are inactive for transposition. Activity assays were necessary to distinguish active from inactive complexes and to assess the longevity of complexes that were not stable to electrophoresis.

To do these experiments, we first established a transposition assay using non-Mu donor fragments. In place of normal MuA recognition sites, these fragments contained two copies of a sequence derived from ϕX174 (Fig. 2b). The chosen ϕX174 sequence has some resemblance to a MuA recognition sequence, and was selected through a functional assay for its ability to be transposed by MuA (Goldhaber-Gordon et al., 2002a). MuA transposed these fragments, albeit very poorly and only when MuA was present at 5 - 10 times our standard concentration of 50 nM (data not shown). Transposition of this non-Mu fragment was stimulated by a fragment containing bona-fide MuA recognition sites (data not shown).

Band-shift assays allow us to distinguish three types of MuA-DNA complexes (Fig. 7b) (Savilahti et al., 1995): (i) One MuA monomer bound to one DNA fragment. These complexes are unstable to competitor DNA or heparin, (ii) MuA bound to DNA in a 2:1 ratio, also unstable to competitor DNA or heparin. These complexes could be either MuA tetramers in a synaptic complex, or dimeric MuA bound to a single fragment, (iii) MuA bound to DNA in a 2:1 ratio, but stable to competitor DNA and/or heparin. This third class is likely to be synaptic complexes.

The band-shifts revealed a difference in the stability of complexes formed on an unjoinable Mu fragment versus an R1-R1 fragment (Fig. 7b). Recall that the R1-R1 fragment lacks the entire cleavage site sequence, while the unjoinable fragment only lacks a single "A" from the cleavage site. There are also minor differences in the recognition sequences of these two fragments (see Fig. 2b). Complexes formed on the unjoinable fragment were stable either to competitor DNA (lane 6) or to heparin (lane 5), whereas those formed on the R1-R1 fragment were stable to competitor DNA (lane 12), but not to heparin (lane 11). The R1-R1 results are the first evidence we have seen that heparin directly destabilizes MuA complexes, as opposed to providing a "sink" for protein has dissociated from DNA. These results also suggest that cleavage site DNA helps stabilize synaptic complexes.

The non-Mu (ϕX174 derived) fragment did not by itself support the formation of competitor-stable complexes (Fig. 7b, lane 3). Yet in the presence of the unjoinable Mu fragment, complexes with the non-Mu fragment were stable to either heparin or competitor DNA (lanes 8 and 9). Thus, the unjoinable fragment has a stabilizing influence on these hybrid complexes. The R1-R1 fragment was not sufficient to stabilize complexes on the non-Mu fragment, as assessed by band-shift assays (lanes 14 and 15). However, the next experiment shows that some R1-R1/non-Mu hybrid complexes were stable to competitor DNA, though not sufficiently stable to be detected by gel electrophoresis.

We next used transposition activity to probe complex stability (Fig. 7c). In this experiment, detectable transposition of non-Mu fragments depended on complementation with Mu fragments (data not shown). Complexes were assembled with a mixture of unlabeled Mu fragments and labeled non-Mu fragments, in the absence of magnesium (to prevent transposition). In a second, twenty minute incubation, we added an excess of competitor DNA to challenge the preformed complexes. Finally, in a third incubation we added target DNA and magnesium to initiate transposition of the non-Mu fragments.

Hybrid transpososomes survived the challenge with competitor DNA, whether the unlabeled Mu partner was the R1-R1 or the unjoinable fragment. Control experiments confirmed that the competitor DNA was sufficient to abolish transposition if complexes were not pre-

assembled in advance of adding the competitor (data not shown). If complexes were pre-assembled, the transposition efficiency was essentially the same irrespective of whether the competitor was added before or together with the magnesium and the target (i.e. second or third incubation). Thus, a functional assay reveals that transpososomes containing a non-Mu donor DNA are long-lived, with a half-life greater than twenty minutes.

Finally, we used a similar strategy to assess the longevity of hybrid complexes with full-sized ϕX174 donor (Fig. 7d). This time, the DNA present in the assembly step was ϕX174 RFI and a Mu fragment. Pre-formed complexes were challenged with excess R1-R1 fragment, and then transposition was initiated with magnesium. Substantial intermolecular product appeared, despite the added challenge fragment and even if the original Mu fragment contained only a single R1 site (Fig. 7d, lane 8). Therefore, active complexes formed between ϕX174 and a fragment with a single R1 site are long-lived. These results show that despite all the deficiencies of their component parts, these hybrid complexes are indeed stable transpososomes.

Mini-Mu plasmids can activate non-Mu transposition.

Armed with the knowledge that non-Mu DNA can transpose, we revisited an unexplained family of products seen in previous studies and discovered the products to be the result of intramolecular transposition of ϕX174. These products were observed in reactions containing mixtures of wild-type MuA and MuA with active site mutations (specifically, the D269N and E392Q substitutions, MuADE/NQ). This mutant protein cannot perform donor cleavage or strand transfer, but it is efficiently incorporated into stable synaptic complexes. The activity of transpososomes containing both wild-type and mutant subunits depends on the placement of subunits. For example, some mixed complexes are fully active; some are not active at all; and some are able to complete cleavage and strand transfer of only one of the two Mu DNA ends (Baker et al., 1994; Baker and Luo, 1994; Kim et al., 1995).

In transposition reactions containing a mini-Mu donor plasmid, ϕX174 RFI DNA, MuB, and a 1:1 mixture of MuA and MuADE/NQ, the most abundant products were the result of mini-Mu donor transposing into ϕX174 target (labeled Mu-ϕX interST in Fig. 8). But an additional family of products appeared that ran between relaxed mini-Mu and relaxed ϕX174 (Fig. 8, lanes 3 and 4, labeled target ϕX-intraST and marked with a bracket). These unexplained products did not contain sequences from the mini-Mu plasmid, as determined by Southern blot analysis (data not shown) and by their insensitivity to BglI, a restriction enzyme that cleaves the Mu plasmid but not ϕX174 (Fig. 8, lane 4). The products did contain ϕX174 DNA, as again determined by Southern blotting (data not shown) and restriction analysis (Fig. 8, lane 5). Although the mini-Mu was not covalently joined to the final products, it was an essential component of the reaction (Fig. 8, lane 6), reminiscent of the role of the Mu donor fragments in the reactions described above.

These ϕX-only products were most likely generated by hybrid complexes with the following structure: the two donor sites were provided by a mini-Mu plasmid and a ϕX174 molecule, the target site was another site on the same ϕX174 molecule, and a MuADE/NQ subunit was incorporated in a position that blocked joining of the mini-Mu molecule to ϕX174 (Fig. 8b). The resulting products were intramolecular transposition products of ϕX174, similar to those characterized in Fig. 3.

Although the products of these reactions were intramolecular, the reactions required the MuB protein (data not shown). This suggests a role for MuB in bringing together two large DNA molecules in one transpososome. Ordinarily, the two large DNAs are a donor and a target respectively. In this special case, both DNAs participated as donors, and one (the non-Mu) was additionally a target.

The mutant version of MuA was needed to see intramolecular ϕX174 products (Fig. 8a, lane 2). However, it is unclear whether the mutant protein stimulated non-Mu transposition, or

simply permitted detection of a reaction whose products are normally obscured by other products. If all of the MuA subunits have functional active sites, products of the analogous reaction should include a mini-Mu molecule. The products would have a mobility similar, though not identical, to standard transposition products. Thus, it is possible that non-Mu transposition occurs as a side reaction in most transposition experiments. Regardless of which explanation is correct, these experiments revealed non-Mu transposition under reaction conditions that are quite different from the fragment assays described above. Thus, multiple reaction conditions reveal the power of Mu DNA to stimulate transposition.

Discussion

We find that Mu DNA can activate MuA to transpose non-Mu DNA. Similar promiscuous activity has been observed previously *in vivo*, with a donor plasmid containing a single transposon end sequence. This "single-end" transposon improvises a second "end" from other sequences on the same plasmid, and transposes with a frequency 100-fold above background. Though this single-end transposon is ~1000 fold less active than a transposon with two proper ends, a plasmid without any end-sequence is indistinguishable from transposase-free controls (Groenen et al., 1986). The transposons Tn3, Tn1721 and Tn21 have also been shown to perform "single-ended" transposition *in vivo* (Arthur et al., 1984; Avila et al., 1984; Motsch and Schmitt, 1984), and *in vivo* use of cryptic recombination signal sequences is well documented for Rag-1/2(Lewis et al., 1997).

Of course, for all these transposable elements the dominant pathway requires pairing of two bona-fide transposon ends. Promiscuous transposition may have evolved as a default pathway, for the rare times when a transpososome begins to assemble on a single end sequence. If a single end were to transpose without first pairing with some other DNA sequence, it would cause chromosomal rearrangements deleterious to the host. Promiscuous transposition, in which a nearby site is synapsed with a transposon end and the two transpose together, would be less harmful. For transposons that remain integrated at one location for many host generations, preserving the host genome is almost as important as preserving their own.

Alternatively, promiscuous transposition may be irrelevant in the wild, as the presence of one transposon end usually means another end is nearby. Nonetheless, the promiscuous activities exposed in artificial settings reflect an important aspect of transposases' natural activity. For example, the Mu synaptic complex is extremely stable, which helps ensure that transposition does not abort before both ends have transposed (Surette et al., 1987). Some of the protein-protein and protein-DNA interactions that stabilize native synaptic complexes presumably contribute to synapsing promiscuous sites. The transposase is designed to tenaciously bind two pieces of DNA, and promiscuous activities are likely a consequence of this tenacious binding.

DNA recognition sites help organize MuA subunits.

A DNA fragment containing two MuA recognition sites strongly activates non-Mu transposition. A shorter fragment with only one recognition site is far less effective, even if protein and DNA concentrations are well above the K_d for simple binding. These results suggest that one important function of the MuA recognition sites, both for standard Mu transposition and for non-Mu transposition, is to spatially constrain two MuA subunits. By bringing two monomers into close proximity to each other, the DNA sites can contribute greatly to driving transpososome assembly.

Other studies also highlight the importance of Mu DNA in spatially organizing the MuA monomers to form an active tetramer. First, under stringent reaction conditions tetramer assembly requires at least two recognition sites at each transposon end (Baker and Mizuuchi, 1992; Namgoong et al., 1994). Transpososome assembly is also sensitive to the spacing between these two required recognition sites (Namgoong et al., 1994). Second, again under stringent conditions, the sites at the two ends must be inverted relative to each other. This orientation requirement suggests an additional level of organization, in which the two ends together establish

the orientation of all four MuA subunits (Baker and Mizuuchi, 1992; Craigie and Mizuuchi, 1986). Third, the accompanying paper suggests that the recognition sites help position MuA subunits relative to the cleavage sites.

On the full Mu transposon, the R1 and R2 sites bring together and position two MuA subunits. The L1 and L2 sites probably function similarly. Although L2 is 80 base-pairs from L1, the DNA-bending protein HU brings L1 and L2 together for assembly (Lavoie et al., 1996).

MuA recognition sites are allosteric activators of transposition.

A fragment with just one MuA recognition site is sufficient to activate transposition of non-Mu DNA. This fragment can pair with ϕX174 RFI DNA to form a MuA transpososome, which is active for transposition even after challenge with a competitor DNA. These results are consistent with and extend a previous study, showing that under permissive reaction conditions a plasmid with a single MuA recognition site could support assembly of MuA tetramers detectable by protein crosslinking (Baker and Mizuuchi, 1992).

A single recognition site, absent the nearby cleavage site, is unlikely to constrain the position of more than one MuA subunit. Chemical and nuclease protection patterns strongly suggest that a MuA monomer contacts the length of an entire recognition site (Zou et al., 1991). Multiple subunits could make overlapping contacts, but an extensive overlap seems unlikely. Thus our single-site R1 fragments are unlikely to have provided much energy towards organizing the subunits of the transpososome. The ability of these R1 fragments to activate transposition points to an additional, more specific role of the individual recognition sites. The simplest explanation is that interactions with the recognition sequence induce a conformational change in MuA necessary for transposition. Most likely, this conformational change allosterically activates transpososome assembly. Although a single recognition site may directly affect only one MuA molecule, allosteric changes in that molecule can cooperatively recruit other MuA molecules to the complex -- some bound to non-Mu DNA. It is also possible that the recognition sites induce conformational changes that are important for a post-assembly step, such as DNA cleavage or strand transfer.

The role of the cleavage site.

Previous studies of MuA indicated a role for the cleavage site in assembling stable synaptic complexes (Namgoong et al., 1994; Watson and Chaconas, 1996). Results of this study support that conclusion (Fig. 7b). Why, then, is the cleavage site not an important feature of activator fragments for transposition of non-Mu DNA? We can offer two thoughts on this topic. (i) Participation in super-stable complexes may not be ideal for an activator fragment, as complexes between two fragments compete with formation of hybrid complexes. (ii) Chapter 3 describes 18 of the sites from ϕX174 that transposed as part of hybrid transpososomes (Goldhaber-Gordon et al., 2002a). The cleavage site sequence is a strong feature among these sites. Thus, the two halves of the complexes in our studies complement each other. The Mu DNA fragment provides the recognition sites, to align two subunits and promote allosteric changes in MuA. The non-Mu DNA provides the cleavage site sequence, which serves as the actual site of transposition and may also provide stabilizing contacts to the complex.

Sequence-specific activation of other proteins

The MuA recognition sites perform a complex set of functions. They provide a structural buttress for the transpososome, arranging multiple protein subunits with precision. They also individually activate MuA subunits, probably by inducing conformational changes that are transmitted from the protein's DNA binding domain to the other domains involved in forming a functional transpososome.

There are many other contexts in which DNA sequences provide an assembly-board for protein-protein interactions. For example, DnaA sites scaffold the initiation machinery at the OriC origin of replication (Messer et al., 2001). Many bacterial repressors multimerize through cooperative binding to neighboring recognition sites, and eukaryotic transcription usually

involves an even more elaborate set of protein complexes which fully assemble only when constrained by DNA (Ptashne and Gann, 1997). In some of these cases, DNA sequences direct the function of a protein complex simply by providing spatial organization for that complex, without additionally inducing allostery. For example, the commonly used yeast two-hybrid assay relies on the fact that the Gal-4 transcription factor uses DNA simply to position itself at a gene promoter. Likewise, the lambda repressor has an experimentally-alterable linker connecting its DNA binding domain to its cooperativity (protein-protein interaction) domain, making it unlikely that DNA allosterically induces cooperative binding of repressor (Astromoff and Ptashne, 1995).

Nonetheless, DNA promoted allostery may be wide-spread as a means of directing a protein activity to a specific DNA sequence. For example, some restriction enzymes cleave the DNA at a fixed distance from a recognition site. The specificity of these enzymes probably relies on allosteric changes transmitted from the DNA-binding domain to the catalytic domain (Wah et al., 1997). Some transcriptional regulators, for example Ets-1 and the glucocorticoid receptor, are also effected by DNA promoted allostery (Lefstin and Yamamoto, 1998). Binding of Tn5 transposase to its recognition sequence promotes protein dimerization by altering a helix connected to a C-terminal protein interaction domain (Davies et al., 2000). In addition, upon forming a synaptic complex Tn5 transposase makes significant changes in its DNA contacts, both within the recognition site and around the cleavage site, suggesting changes in the protein structure (Bhasin et al., 2000).

Transposases are a particularly interesting example of DNA promoted allostery, as potentially are site-specific recombinases, DNA repair complexes such as MutSHL, and certain restriction enzymes, like *FokI*, that may multimerize only on DNA (Bitinaite et al., 1998; Hallet and Sherratt, 1997; Sixma, 2001). These proteins all function in catalytic pathways, but their activities require pre-assembly into multimeric protein-DNA complexes. Thus, DNA promoted conformational changes could function at multiple points along the pathway, including the assembly step and subsequent reaction steps. It will be interesting to learn whether the allosteric changes induced by the MuA recognition sites directly affect the cleavage and strand transfer reactions, or are primarily required for assembly of a functional complex.

Figure legends

Fig. 1. Model of transpososome assembly. Individual MuA monomers bind specifically to six recognition sites, three at either end of the Mu transposon. Four of these monomers form a stable tetramer, synapsing the ends of the transposon and engaging the cleavage sites. One monomer is shown in white, to highlight the crisscross structure. (Components not drawn to scale, in this and all schematic figures.)

Fig. 2. Mu donor fragments. *A. Fragments containing 50 base-pairs from the right end of Mu support transpososome assembly.* Mimicking transposition of a complete transposon, MuA can synapse two fragments and join them to opposite strands of a target DNA. The DFT product results from successful transposition of both fragments; the SFT product results if one fragment fails to transpose. *B. Sequences of fragments used in this study.* The fragments are shown with the transferred strand on top, but that strand is listed 3' to 5'. The important differences among most of these oligonucleotides is the length of the transferred (top) strand on the 3' end. The darker box highlights the sequence of the native R1 site, lighter box the native R2 site, and bold lettering the cleavage site. Some experiments used fragments with the same names but slightly different sequence than those shown here; see experimental procedures. Additional fragments are also listed in experimental procedures, as is the complete sequence of the uncleaved fragment.

Fig. 3. Transposition of a non-Mu DNA molecule (ϕX174 RF1). *A. Ethidium bromide stain of an agarose gel showing non-Mu transposition products.* Transposition of ϕX174 RFI DNA was activated by an unjoinable (UJ) Mu fragment. On this gel, intramolecular transposition products migrated near linear ϕX174. But with TBE running buffer, as opposed to the TAB buffer used in the gel shown here, the same products very clearly separated from linear DNA (data not shown).

A reaction with a pre-cleaved (PC) fragment is shown for comparison. The DFT product of that reaction migrates with linear ϕX174. Fragment and MuA concentrations were each 50 nM. *B. Two-dimensional gel electrophoreses supported the identification of the intra- and inter-molecular transposition products.* Transposition products were formed with an unjoinable fragment and ϕX174 DNA. Electrophoreses in the first dimension was on a native agarose gel as in Fig. 2a; the second dimension was alkaline (denaturing) agarose, and was visualized by southern blotting. Here, an ethidium bromide stain of the first dimension is shown positioned horizontally across the autoradiogram of the second dimension. The ethidium stain is not of the actual gel-slice that was used for the second dimension, but it is the same DNA sample. Circular and linear strands are labeled based on standards (specifically, yield from supercoiled and nicked DNA). "Dimer" is labeled based on its slow migration.

Fig. 4. Non-Mu transposition is due to hybrid synaptic complexes. *A.* Non-Mu transposition is probably due to synapsis of a fragment containing MuA recognition sites and a ϕX174 molecule. *B. In support of this model, preincubation of MuA with the unjoinable fragment inhibited formation of intermolecular transposition products.* In a 1 hour incubation step, 100 nM MuA was incubated without DNA (lane 1), with ϕX174 DNA and 50 nM unjoinable fragment (lane 2), just with ϕX174 (lane 3), or just with unjoinable fragment (lane 4). The missing DNA components were added to the reactions for a subsequent 20 minute incubation. Only the preincubation with the unjoinable fragment caused significant inhibition of transposition (lane 4). For this figure and all that follow, unless noted otherwise, products were visualized by Vistra-green staining of an agarose gel. *C. Graph of product concentration vs. fragment concentration shows a fragment optimum.* The total DNA concentration was maintained constant at 1440 nM, by balancing the Mu fragment with a fragment of identical length but unrelated sequence. The Mu fragment was R1-R1 (see Fig. 2b), and the incubation time was four hours, to permit this reaction to be compared to the one shown in Fig. 6b. The graph shows % of DNA converted to "intermolecular" product as a function of [Mu fragment].

Fig. 5. Cleavage site is not a necessary feature of the Mu fragment. 250, 50 or 5 nM fragment was incubated with MuB, ϕX174, and 50 nM MuA. Depending on which Mu fragment was used, we saw three types of products: 1) The figure-eight intermolecular product described above, 2) The product labeled with a tailed-circle, which contained (data not shown) a ϕX174 dimer (as determined by two dimensional gel electrophoresis) and a Mu fragment (as determined by radiolabeling the Mu fragment). This tailed-circle product has some linear character, as defined by its relative migration in gels with different running buffers and by its sensitivity to exonuclease V (data not shown). This product probably results from pairwise transfer of a fragment and ϕX174, and/or from reuse of a figure-eight product as a target in a second round of transposition, and 3) The standard product of transposition of a fragment pair (DFT), which runs with linear product. (Aside note: at fragment concentrations above the MuA concentration, DFT products are visibly reduced (lane 2). This is probably due to failure of MuA to saturate the recognition sites. Inhibition of non-Mu transposition is visible at much lower fragment concentrations.)

Fig. 6. Single-site fragments can activate transposition. *A.* 1440, 480 or 160 nM R1 fragment was incubated with MuB, ϕX174 and 50 nM MuA for two hours. The longer incubation time was necessary to see products with the short fragment. (Aside note: some linear target is visible, due to contaminating nuclease activity in the MuB prep which becomes apparent with long incubation times. The background levels of visible ϕX174 dimer are also high, a result of nuclease nicking of supercoiled dimer which exists in the ϕX174 prep but is normally obscured by relaxed monomeric ϕX174.) *B. Linked sites are better activators than unlinked sites.* Varying concentrations of R1-R1 fragments (squares), or R1 fragments (triangles) were incubated with MuB, ϕX174 and 50 nM MuA for 4 hours. A DNA fragment of appropriate length (46 bp for the

R1-R1 experiment, 22 bp for R1) and unrelated sequence was used to maintain total [DNA] at 1440 nM. The graph shows % of DNA converted to "intermolecular" product as a function of [Mu fragment].

Fig. 7 Hybrid transpososomes are stable to competitor DNA. *A. Summary of experimental design.* Transpososome complexes were assembled, challenged with an excess competitor DNA fragment, and assayed for stability. *B. Band shift assays directly address complex stability.* In a first incubation, complexes were formed by incubating 50 nM labeled fragment and 50 nM unlabeled fragment with 150 nM MuA for 40 minutes. In some cases the labeled and unlabeled were the same fragment -- both were included to keep the DNA concentration constant between experiments. During a second 60 minute incubation, preformed complexes were challenged with 0.2 μg heparin (h: lanes 2, 5, 8, 11, 14) or 800 nM cold R1-R1 fragment (d: lanes 3, 6, 9, 12 and 15). Complex stability was then confirmed by gel electrophoresis and autoradiogram. MuB, target DNA and $MgCl_2$ were not included at any point in this experiment. *C. Transposition assays reveal that hybrid transpososomes are long-lived.* In this autoradiogram, the only visible product results from strand transfer of a non-Mu fragment into a ϕX174 target. Complexes were assembled by incubating 200 nM MuA with 50 nM labeled non-Mu fragment and 50 nM unlabeled Mu fragment, for 4 hours. The absence of divalent metal at this step prevented transposition (assembly on precleaved fragments does not require divalent metal). In a second twenty minute incubation, excess competitor DNA (500 nM unlabeled, unjoinable fragment) was added to the reactions shown in lanes 1 or 2, to challenge the stability of preformed complexes. In a third incubation, target DNA and $MgCl_2$ were added, as well as competitor DNA to reactions shown in lanes 3 and 4. Numbers above the gel refer to the incubation step for which competitor DNA was added (see Fig. 7a). Lanes 1 and 3: during the first incubation the unlabeled fragment was unjoinable; lanes 2 and 4: it was R1-R1 fragment. *D. Transposition assays reveal stability of transpososomes formed on ϕX174 RFI DNA.* In a first incubation, 150nM MuA was incubated for one hour with 570 nM MuB, ϕX174, and Mu fragment, in 10 mM $CaCl_2$ to permit assembly but prevent transposition. Either the R1-R1 fragment was present at 8 nM (lanes 1-4), or the R1 fragment was present at 1.4 μM (lanes 5-8). In a second incubation, excess competitor DNA (500 nM R1-R1 fragment) was added to the reactions shown in lanes 4 or 8, to challenge the stability of preformed complexes. In a third incubation, $MgCl_2$ was added, along with 300 nM additional MuB, to initiate transposition. Numbers above the gel refer to the incubation step for which excess competitor DNA was added (see Fig. 7a). Lanes 1 and 5 confirm that no transposition occurred in the absence of magnesium. Lanes 3 and 7 confirm that the competitor DNA competed successfully.

Fig. 8. Non-Mu transposition can be activated by mini-Mu plasmids. *A. Agarose gel of plasmid reactions.* Reactions were first incubated at 30°C to allow transposition and then incubated at 37°C (with or without an appropriate restriction enzyme). A bracket marks the products of intramolecular transposition of ϕX174. BglI (lane 4) uniquely cleaves the Mu plasmid, while *XhoI* (lane 5) uniquely cleaves ϕX174. DNA species are labeled as follows: Mu-ϕX interST = intermolecular transposition products with mini-Mu donor and ϕX174 target; ϕX intraST = intramolecular transposition product containing only ϕX DNA; subscripts: N = nicked, L = linear, S = supercoiled; Mu = mini-Mu plasmid pMK586; ϕX = ϕX174 RFI; A(wt) = wildtype Mu transposase; A(DE/NQ) = defective active site version of Mu transposase. Lane 6 is taken from a different gel than lanes1-5. *B. Model of transpososome that produced the intramolecular strand transfer products (labeled ϕX-intra ST in Fig. 8a).* The probable reason for mini-Mu plasmid failing to join to target is inclusion of active-site mutant MuA.

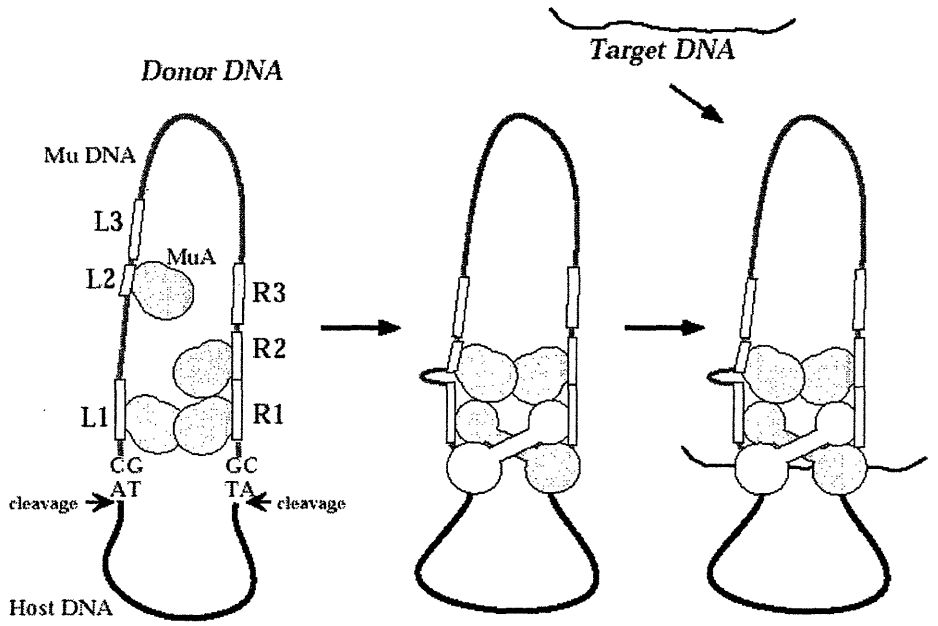


Figure 1, Activation by MuA Recognition Sites

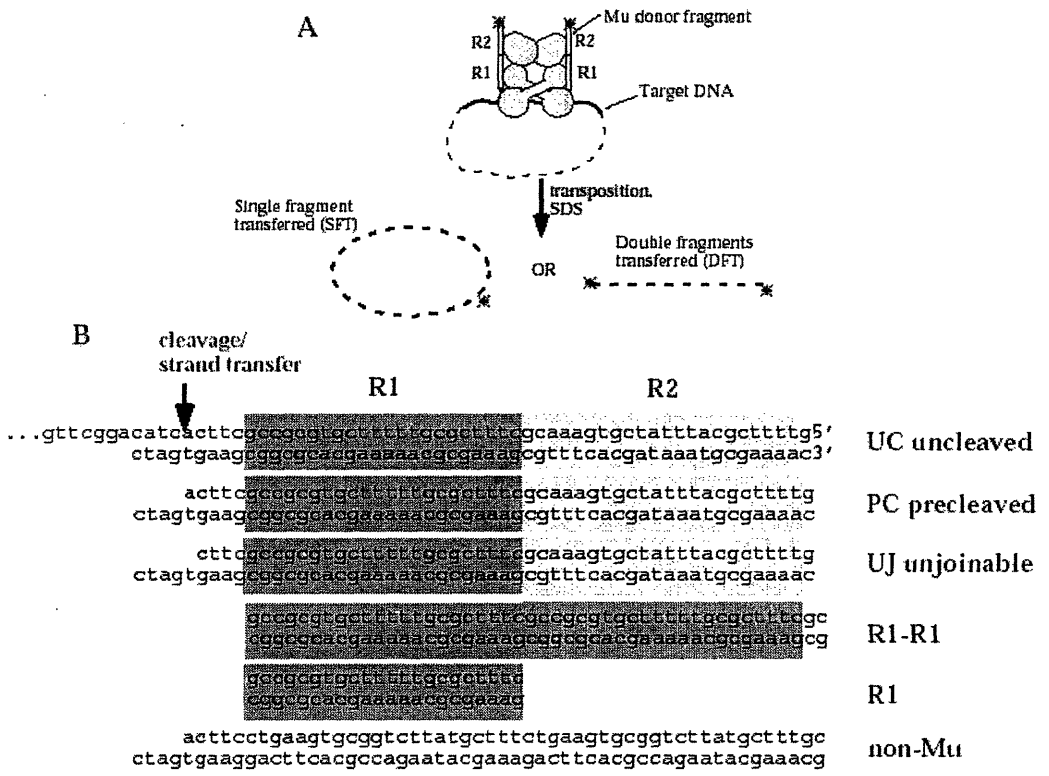


Figure 2, Activation by MuA Recognition Sites

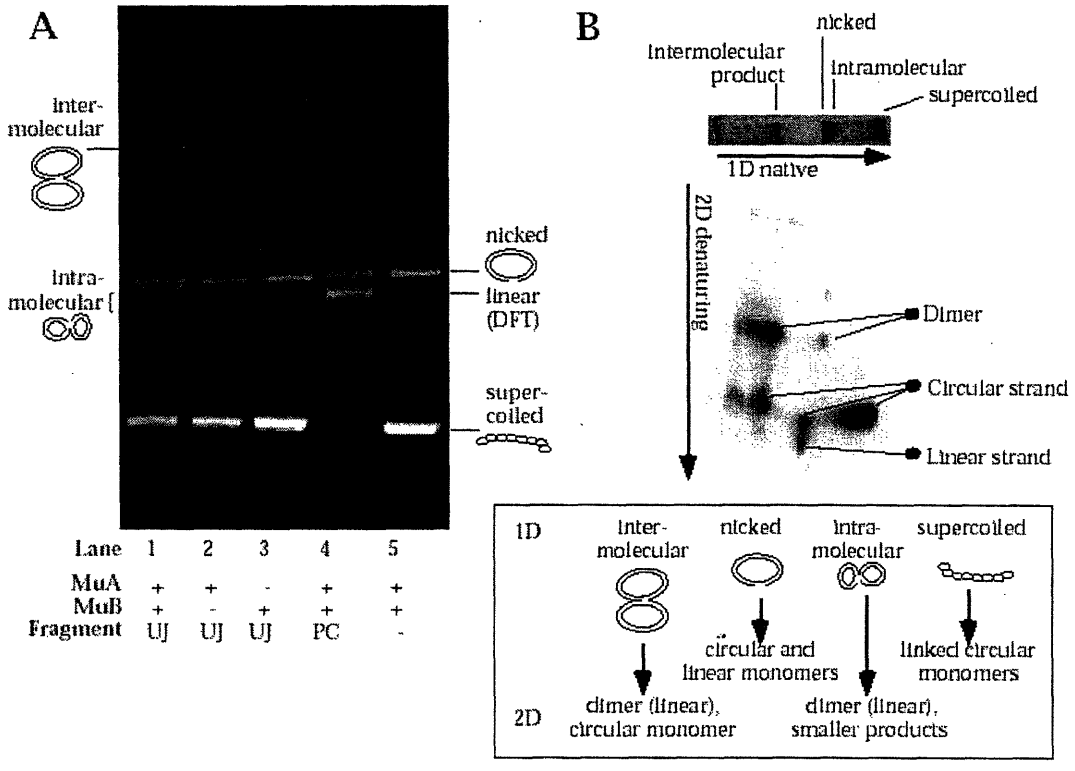


Figure 3, Activation by MuA Recognition Sites

Figure 4, Activation by MuA Recognition Sites

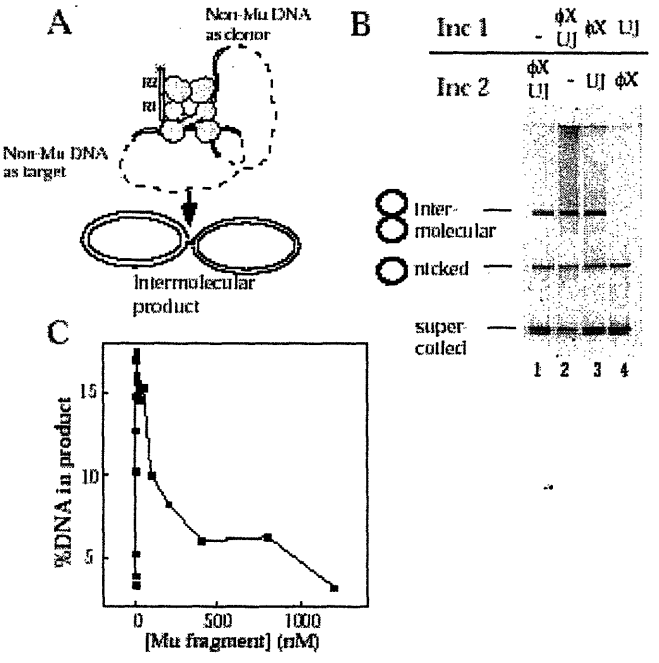


Figure 5, Activation by MuA Recognition Sites

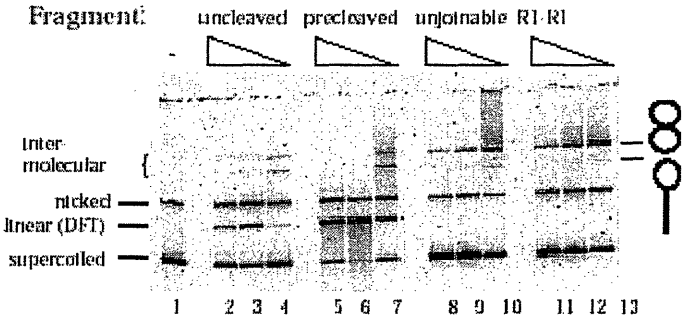


Figure 6, Activation by MuA Recognition Sites

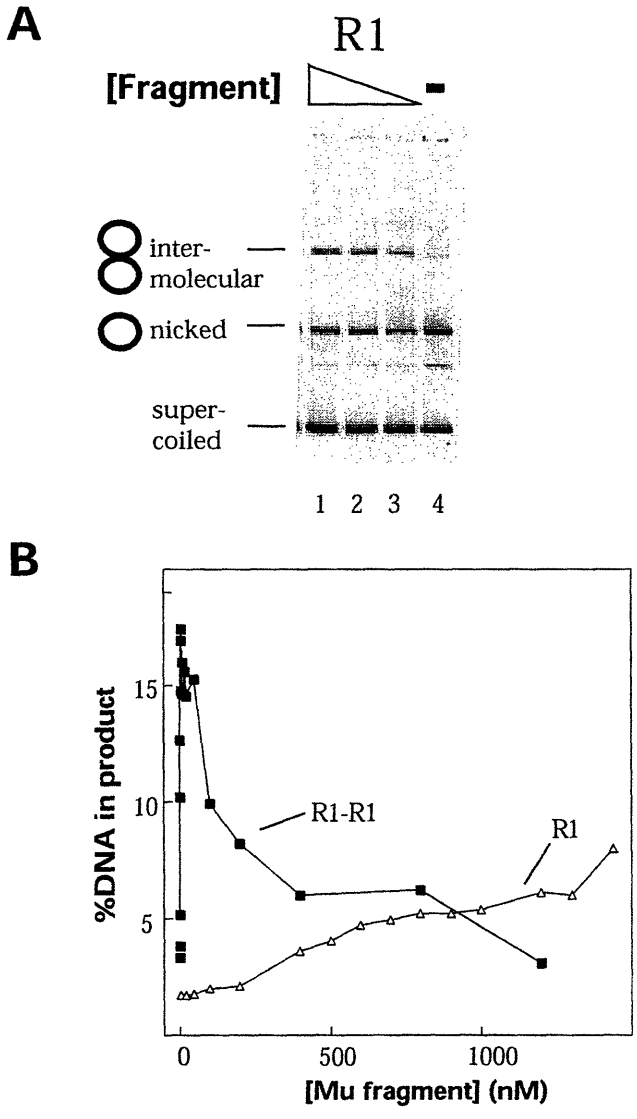
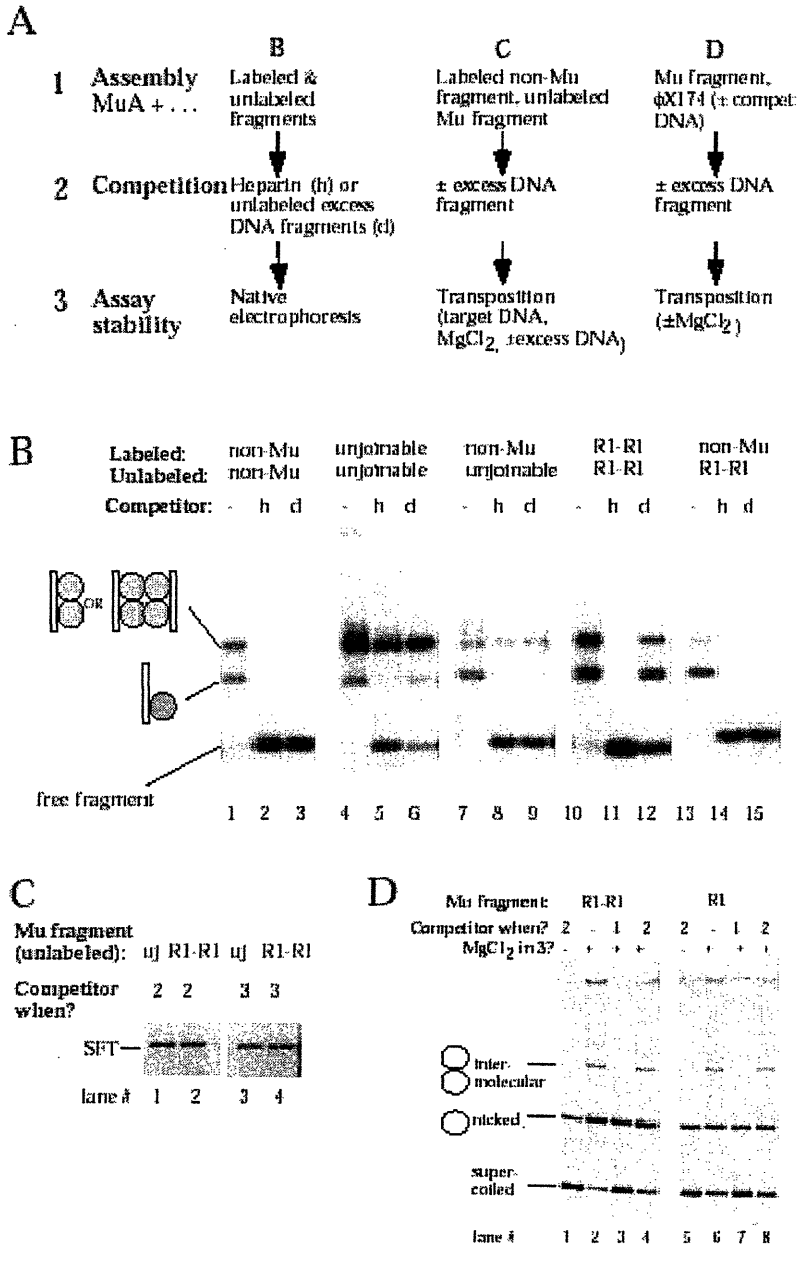


Figure 7, Activation by MuA Recognition Sites



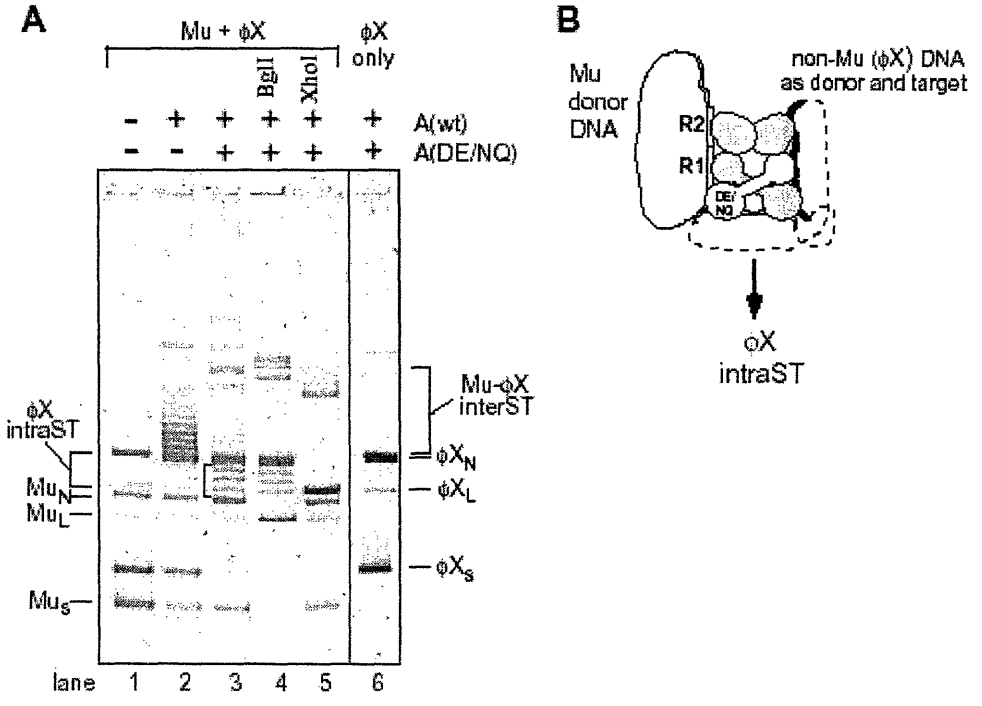


Figure 8, Activation by MuA Recognition Sites
Page 49

Chapter 3. Sequence and Positional Requirements for DNA Sites in a Mu Transpososome

Transposition of bacteriophage Mu uses two DNA cleavage sites and six transposase recognition sites, with each recognition site divided into two half sites. The recognition sites can activate transposition of non-Mu DNA sequences, if a complete set of Mu sequences is not available. In this chapter I present analyses of 18 sequences from a non-Mu DNA molecule, selected in a functional assay for the ability to be transposed by MuA transposase. These sequences are remarkably diverse. Nonetheless, when viewed as a group they resemble a Mu DNA end, with a cleavage site and a single recognition site. Analysis of these "pseudo-Mu ends" indicates that most positions in the cleavage and recognition sites contribute sequence-specific information that helps drive transposition, though only the strongest contributors are apparent from past mutagenesis data. The sequence analysis also suggests variability in the alignment of recognition half-sites. Transposition assays of specifically designed DNA substrates support the conclusion that the transposition machinery is flexible enough to permit variability in half-site spacing, and also perhaps variability in the placement of the recognition site with respect to the cleavage site. This variability causes only local perturbations in the protein-DNA complex, as indicated by experiments in which altered and unaltered DNA substrates are paired.

Specific contributions by colleagues:
This chapter appeared in the Journal of Biological Chemistry on March 8, 2002, Vol. 277, No. 10, pp. 7703-7712. The authors were Ilana Goldhaber-Gordon, Michael H. Early, Matthew K. Gray and Tania A. Baker. Michael Early contributed figure 5a and some of the experiments for figure 5b. Matthew Gray, in collaboration with me, wrote the perlscript that is described in the text and that was used to generate the data for Table I.

Introduction

Some of the most fundamental of biological processes occur within nucleoprotein complexes. These processes include recombination, replication, transcription, and RNA splicing. The nucleoprotein complexes are often large and elaborate, containing features that permit sophisticated regulation. As a result, dissecting the chemistry of interactions within a complex is a challenging task, but one that is critical to understanding biological function.

Transposition of bacteriophage Mu, like that of other transposons, occurs within complexes called transpososomes. Transpososomes mediate at least two sequential chemical reactions, on pathway towards transferring the transposon DNA from one site to another. The two reactions are: (i) DNA cleavage, in which a nick is introduced precisely at the end of the transposon, on the 3' strand, and (ii) DNA strand transfer, in which the nicked strand is joined to a separate DNA molecule called the target (Chaconas, 1999; Haren et al., 1999). The reaction sites on the transposon DNA, or "donor DNA", are defined by specific DNA sequences, but Mu target sites are not very sequence-specific (Mizuuchi and Mizuuchi, 1993).

Transpososomes contain multiple subunits of a transposase protein, bound to DNA sequences from both of the transposon's ends (Haren et al., 1999). These protein-DNA complexes are also called "synaptic complexes", because they bring together the two ends of the transposon DNA. The phage Mu transposase, MuA, is monomeric in solution, but forms a tetramer upon binding to specific DNA recognition sites near the transposon ends (Mizuuchi et al., 1992; Surette et al., 1987).

Each end of the Mu transposon has three MuA recognition sites: R1, R2 and R3 on the right end and L1, L2 and L3 on the left, not all of which are essential for transposition (Allison and Chaconas, 1992; Baker and Mizuuchi, 1992; Craigie et al., 1984; Groenen et al., 1985; Lavoie et al., 1991). Five of the six sites are related by a degenerate 22 base-pair consensus sequence. The exception is L2, which is missing roughly half of the 22 base-pair sequence. In the transpososome tetramer, three of the MuA subunits are bound specifically to the R1, L1 and R2 sites. The fourth subunit probably binds the L2 site, but this contact is not as stable as that of the other three subunits and can be competed away by excess DNA (Lavoie et al., 1991) (Mizuuchi et al., 1991). Additional subunits, not essential for catalysis, may be loosely associated with the R3 and L3 sites. *In vitro,* MuA can transpose "donor fragment" molecules—short DNA molecules carrying only the R2, R1 and cleavage sites (Fig. 1a) (Savilahti et al., 1995). These fragments support assembly of MuA tetramers, which synapse pairs of fragments.

MuA specifically binds its DNA recognition sites through its N-terminal domain (Fig. 1b) (Nakayama et al., 1987). This proteolytically defined domain contains three independently folding subdomains, Iα, Iβ, and Iγ, each of which has a helix-turn-helix (HTH) motif (Clubb et al., 1994; Clubb et al., 1997; Schumacher et al., 1997). Domain Iα binds to an internal enhancer-like site that is not essential *in vitro*. Domains Iβ and Iγ interact with the recognition sites and are essential. Each DNA recognition site can be thought of as two half-sites: the outer half-site is bound by domain Iγ and the inner half by Iβ. This bipartite DNA binding has at least two consequences. It permits the recognition sequences to be unusually long, and therefore rare, and it allows for nanomolar binding affinities to the combined site (Schumacher et al., 1997). It is also possible that the bipartite structure achieves some additional, as yet undiscovered function.

The cleaved phosphodiester bond at the end of the transposon is separated from the outermost recognition site (R1 or L1) by five-base-pairs, the last two base-pairs being 5'CA (often described by the 5'TG complement). Thus, Mu has separate cleavage sites and recognition

sites. Whereas the recognition sites are bound by domains Iβ and Iγ of MuA, the cleavage sites are contacted by the protein's active site which is in domain IIα.

Within the synaptic complex, individual MuA subunits bind a cleavage site and a recognition site on opposite DNA ends. Specifically, the MuA subunit bound to the R1 site engages the left-end cleavage site, and the subunit bound at L1 engages the right-end cleavage site (Fig. 1b) (Aldaz et al., 1996; Namgoong and Harshey, 1998; Savilahti and Mizuuchi, 1996; Williams et al., 1999). This intertwined structure suggests that the various MuA domains and subunits function interdependently in a transpososome.

There are many unanswered questions about interactions between the protein and DNA components of the transpososome. For example, there is no direct structural information about the geometry of the protein-DNA interactions. On another level, although considerable mutagenesis has been performed on the MuA recognition sites (Burlingame et al., 1986; Groenen and van de Putte, 1986), the sites are long and our understanding of sequence requirements is incomplete.

In chapter 2, I described transposition of DNA molecules that have no defined MuA recognition sites (Goldhaber-Gordon et al., 2002b). This unconventional activity depends on MuA recognition sites provided in *trans* on a donor fragment or a donor plasmid, so that "hybrid" synaptic complexes can assemble between the Mu DNA and a site on the non-Mu DNA (Fig. 1c). Here, we describe the sequences of 18 sites on ϕX174 RF1 that were transposed by MuA. These 18 sequences are extremely diverse. Nonetheless, looking at the most commonly represented nucleotide at each position, we find that as a group the 18 sequences resemble a Mu end with a "TG" cleavage site and a single MuA recognition site. We deduce from these 18 sequences the fundamental sequence requirements for transposition. Additionally, we draw conclusions about the arrangement of the individual DNA sites that contribute to the transpososome. Our conclusions, initially based on sequence analysis, are supported and extended through analysis of the transposition activity of defined DNA fragments. Our data point to localized regions of flexibility in the Mu transpososome.

Experimental Procedures

Proteins and DNA. MuA (Baker et al., 1993) and MuB (Yamauchi and Baker, 1998) were prepared as described. ϕX174 RF1 was purchased from New England Biolabs. Oligonucleotides (fragment donors or PCR primers) were purchased either from MIT/HHMI biopolymers lab or from GeneLink, and fragment donors were purified by denaturing PAGE. Donor fragment sequences were as follows: **unjoinable Mu fragment:**gttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgcttc/ctagtgaagcggcgcacgaaaaacgcgaaagcgttcacgataaatgcgaaaacg,**wild-type:** gttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgcttca/ctagtgaagcggcgcacgaa aaacgcgaaagcgtttcacgataaatgcgaaaac, **fragment HSI:** gttttcgcattttacgtgaacgctttcgcgtttttcgtgcgccgcttca /ctagtgaagcggcgcacgaaaaaacgcgaaagcgtttcacgataaaatgcgaaaac, **fragment HSD:** gttttcgcatttacgtg aaacgctttcgcgtttttcgtgcgccgcttca/ctagtgaagcggcgcacgaaaaacgcgaaagcgtttcacgtaaatgcgaaaac, **fragment HSN:** gttttcgcatttatcgtgaaacgctttcgcgtttttgctcccggcttca/ctagtgaagccgggagcaaaaaacgcgaaagcgtttcacgataa a tgcgaaaac, **fragment 17a:** cgtttcgtattctggcgtgaagtctttcgtattctggcgtgaagtccttca/ctagtgaaggacttcacgccagaata cgaaagacttcacgccagaatacgaaacg, **fragment 17b:** cgtttcgtattctggcgtgaagtctttgtcttctggcgtgaagtccttca/

ctagtgaaggacttcacgccagaagacaaaagacttcacgccagaatacgaaacg, **fragment 17c:**
cgtttgtaatcctggcgtgaagtctttgt aatcctggcgtgaagtccttca
/ctagtgaaggacttcacgccaggattacaaagacttcacgccaggattacaaacg, **fragment 17d:**
cgtttaagattctggcgtgaagtctttaagattctggcgtgaagtccttca/ctagtgaaggacttcacgccagaatcttaaagacttcacgcca
gaatcttaaacg, **fragment CSI:**
gttttcgcattatcgtgaaacgctttcgcgtttttcgtgcgccgccttca/ctagtgaacggcgcacgaaaaacg
cgaaagcgtttcacgataaatgcgaaaac, **fragment CSD:** gttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgttca/
ctagtgaaggacttcacgccagaagacaaaagacttcacgccagaatacgaaacg

Isolation of pseudo-ends. The short target had the sequence: 5'ccagtgctggaaagataccggcaggcc cgcgcacaccggccccagatatatgagcagccaacgacccggccgggaggg, annealed to a precise complement. This sequence was derived from an insertional hotspot in pBR322 (Mizuuchi and Mizuuchi, 1993), but was modified to remove any 3 base-pair sequence that resembled parts of a MuA recognition site. This short target [240 nM] was incubated with 15 ng/μl MuB and 0.2 mM ATP at 30°C for 20 minutes, under standard transposition buffer conditions (Goldhaber-Gordon et al., 2002b). If MuB was not included in this step, we saw no transposition into the short target (data not shown). ATP-γ-S was then added to 2 mM, and the reaction was incubated for another 20 minutes at 30°C. If ATP–γ–S was not added, transposition into the short target was reduced (data not shown). We then added ϕX174 RF1 to 10 ng/μl, unjoinable fragment to 5 nM, and MuA to 50 nM, and continued incubating for two hours. Reactions were stopped with SDS/EDTA, phenol/chloroform extracted, and ethanol precipitated.

For the two sequences obtained without cloning, ~2.5 μg of DNA were used in a 50 μl thermal cycling reaction, using a "touchdown" protocol. Annealing temperature was reduced by 1° every 2 cycles for 26 cycles, starting at 60°C, and then amplification continued 12 more cycles with annealing temperature 47°C. The primers were ccagtgctggaaagatac (anneals to the short target) and either gtagaaatgccacaagcc (to obtain sequence 17) or ggcttgtggcatttctac (to obtain sequence 18). The products were run on a 1.5% MetaPhor (FMC Bioproducts) preparative agarose gel. In each case a ~700 bp band was excised and purified with a Qiagen gel extraction kit. It was then amplified in a second round of PCR (20 cycles, 49°C annealing temp), and sequenced with an Amersham Pharmacia cycle sequencing kit.

For the sequences obtained by cloning, the transposition products were purified away from unused fragments (both the unjoinable Mu fragment and the target fragment) on a Superose 6 HR column before being subjected to the "touchdown" PCR. This time, annealing temperature was decreased by 0.5°C every cycle for the first 20 cycles, followed by 12 cycles at annealing temperature of 51°C. Products were purified with a Qiagen PCR Purification kit. The products were then simultaneously digested with *EcoRI* and an appropriate second enzyme, cloned into a pUC19 vector, and transformed into DH5α cells. Clones were sequenced at the MIT/HHMI biopolymers sequencing facility.

The target primer contained an *EcoRI* site: tagaattccagtgctggaaagataccggc. The ϕX174 primer was one of the following, containing one of the following restriction sites: *AatII* sites: tgtgtgactattgacgtccttcccc or cgtacggggaaggacgtcaatagtc; *PstI* sites: gattggcgtatccaacctgcagag or gaagcgataaaactctgcaggttgg; *BamHI* sites: taggatccatgcctttcccatcttggct or taggatccaagatgggaaaggtcatgcg or taggatcctgctatgaggcttgtggca or taggatccacaagcctcaatagcaggttt. Sequences 1, 2 and 3 (Fig. 2) were obtained from faulty priming of the ϕX174 primer to an incorrect sequence on ϕX174 that happened to be similar to the primer sequence and happened to be near the transposition joint.

Transposition reactions. Unless otherwise indicated, transposition reactions were done essentially as described (Goldhaber-Gordon et al., 2002b), except that protein and DNA concentrations were as follows: 200 nM donor fragment, 500 ng ϕX174, 200 ng MuB. MuA concentrations were as indicated in Fig. legends. These conditions were optimized for transposition of fragments carrying pseudo-MuA-recognition sites.

Results

Individual pseudo-ends have identifiable cleavage sites.

Using MuA transposition assays, we identified 18 non-Mu sequences that could function analogously to the Mu DNA end sequences. These 18 "pseudo-Mu ends" were located on ϕX174 RFI DNA, but the results presented in chapter 2 suggest that any large DNA molecule could have yielded a similar group of pseudo-ends (Goldhaber-Gordon et al., 2002b). In addition to the large (5386 bps) ϕX174 DNA, the transposition assays used to isolate pseudo-Mu ends required two types of DNA fragments (Fig. 1c). An 80 base-pair fragment served as a transposition target. The sequence of this fragment was based on a preferred target-site from plasmid pBR322 (Mizuuchi and Mizuuchi, 1993), but the original plasmid sequence was modified slightly to remove any minor sequence similarity to Mu DNA. We also needed a source of native MuA recognition sites, to permit transposition of the non-Mu DNA. For this purpose we used a DNA fragment that contained two MuA recognition sites (R2 and R1) but lacked the pA3' from the cleavage site, and therefore could not participate in the covalent chemistry of transposition. I showed in chapter 2 that this fragment, called an "unjoinable fragment", efficiently activates non-Mu transposition (Goldhaber-Gordon et al., 2002b).

The 80-mer target fragment was incubated with MuB, a Mu protein which controls selection of transposition targets (Maxwell et al., 1987). Subsequently, MuA transposase, the unjoinable Mu DNA fragment, and ϕX174 RFI DNA were added. DNA products in which the ϕX174 DNA became covalently joined to the target fragment were then amplified with a primer to the target and one of several primers designed to arbitrary locations on ϕX174 (Fig. 1c). This method yielded two PCR products that were sufficiently abundant to be directly sequenced. Sixteen additional sites were identified after cloning and sequencing.

The sequences obtained are listed in Fig. 2, beginning with the 5' complement of the nucleotide that was joined to target DNA. In sixteen cases, that complement was a "T", and in the remaining two cases it was an "A". Furthermore, in 8 cases (out of the 18) the following nucleotide was a "G." Thus, these sequences revealed that functional pseudo-Mu DNA ends have cleavage sites that are close in sequence to the 5'TG of the natural cleavage sites.

Several additional lines of evidence confirm that these sequences are indeed pseudo-Mu ends; that is, that despite low similarity to the ends of the Mu genome, they served as transposition donor sites (not all data shown). (i) In control experiments, a 5' label on the target 80-mer revealed some target fragment joined to ϕX174 DNA. (ii) To be functional, the 80mer target needed to be pre-incubated with MuB. Since DNA molecules bound by MuB are preferred by MuA as transposition targets, these data support the claim that ϕX174 DNA was transposing into the 80-mer, rather than vice versa. (iii) For the two sequences analyzed directly (without cloning), we could clearly read on the sequencing gel roughly 100 bp of ϕX174 sequence. Then the ϕX174 sequence ended abruptly, and we could read multiple sequences from the 80-mer target, overlapping with each other on the sequencing gel. This sequence pattern suggests a family of transposition products, corresponding to a single site on ϕX174 recombining with multiple sites on the small target DNA. (iv) To control for faulty priming of the target primer on ϕX174 DNA, we rejected clones that did not include target sequence beyond the target primer.

(v) We also cloned and sequenced DNA that had been treated as in our experiment, except that MuA and B were omitted from the transposition reactions. We were reassured that these controls did not yield sequences that we would have considered to be transposition products. (vi) Finally, as described below, further sequence analysis affirms that these 18 sequences, as a group, resemble a Mu DNA end.

The pseudo-Mu ends have weak similarity to a MuA recognition site.

A striking feature of the sequences in Fig. 2 is that most of them do *not* strongly resemble a MuA recognition site. All of the 16 pseudo-ends isolated by cloning (rather than direct sequencing) were located within 400 base-pairs of the PCR primer that isolated them. All but three were located within 250 base-pairs of their respective primers. This close proximity to an arbitrary set of primers suggested that use of a different set of PCR primers would have resulted in isolation of a different group of sequences, indicating minimal requirements for DNA to function as a transposition substrate in these experiments.

We developed a scoring system to address the resemblance of pseudo-ends to a consensus sequence derived from native MuA recognition sites (right side of Fig. 2). Sites were scored 1 point for a match to the consensus at a position where a nucleotide is specified, and half a point if only purine or pyrimidine is specified. For analysis of a single 22 base-pair recognition site, without the cleavage site, the highest possible score is 17.5.

For the 18 pseudo-ends, the mean score was 6.50 points for the putative outer-site (i.e. nucleotide positions 6-27, equivalent to R1 or L1) (Table 1, row 1). This is indeed low, compared to, for example, the 13.5 points found for the natural R1 site. However, the mean score for all 10,772 possible sites on the two strands of ϕX174 is 5.41, and the standard deviation for randomly chosen groups of 18 such sites is 0.42. Thus, the pseudo-ends, as a group, are 2.6 standard deviations better than a typical group of 18 (Table 1, last column; [6.5-5.41]/0.41 = 2.6). A random group of 18 sequences will score this well 0.5% of the time.

We next conducted the same analysis for the putative inner site (i.e. nucleotide positions 28-49, equivalent to R2 or L2). The results of this calculation gave a mean score of 4.61 (Table 1, row 10). For the group of 18, this is 1.9 standard deviations *lower* than the expected score of 5.41 for a randomly chosen group. The lower-than-mean score is puzzling, but perhaps not significant—a random group should score this poorly 3% of the time.

Analysis of pseudo-ends suggests alternative alignments.

The poor scores for the 18 pseudo-ends could be due to variability in the correct sequence alignment. For example, there might be flexibility in the placement of the recognition sites relative to the cleavage site. In addition, recall that the recognition sites consist of two half-sites, bound respectively by the Iγ and Iβ subdomains of MuA (Fig. 1b). Spacing between half recognition sites could also vary. It would be interesting if the pseudo-end sequences showed variability in their alignments, because this variability might point to areas of flexibility within the Mu transpososome.

With this in mind, we reanalyzed the pseudo-end sequences, allowing variability in the spacing between pairs of sites. The computer program (a perlscript) used for this analysis can be made available on request. The point system used by the program to score sites was the same as described above.

Allowing variability between the two halves of the outer recognition site greatly improved the point score of the pseudo-end group. With flexibility of ±1 nucleotide between half sites, the group of 18 pseudo ends scored more than 4 standard deviations better than random groups of 18 (Table 1, row 6). On the order of 10^{-4}% of random groups of 18 sites would score

this well. Note that if MuA could not permit variability in the spacing of the two half-sites, an analysis that did permit variability would worsen the sites' mean score compared to a random group. If we accept the hypothesis that these 18 sequences were not randomly selected, then the fact that allowing flexibility improved the group's score from 2.6 standard deviations (just significant) to 4.2 (highly significant) strongly suggests that MuA bound some of these sites in an unorthodox register.

Allowing variability in the spacing between cleavage site and outer recognition site did not improve the relative mean score for the pseudo-end population (Table 1, rows 2&3). On the other hand, this variability did not significantly decrease the score either. We will return to this point below.

Analyzing the ϕX174 sequences adjacent or nearly adjacent to those shown in Fig. 2 failed to reveal a set of inner recognition sites (equivalent to R2) (Table 1, rows 10-13, and data not shown). Note that we would not have detected a set of inner sites located at arbitrary distances from the outer sites. Most likely, in our experiments the transpososomes did contain four DNA-bound MuA subunits – two bound to ϕX174 and two bound to a Mu fragment – but the second ϕX174-bound subunit relied on cues from the other three subunits to position it in the transpososome.

Transposition of diverse sequences

It is worth noting that some individual pseudo-ends continued to score poorly, even allowing variable alignments (data not shown). For example, with variability of ±1 nucleotide between half-sites, sequences 2 and 3 continued to score only 3.5 and 4 points, respectively. Several others scored 5 or 6 points. Also note that not a single nucleotide was absolutely required among the 18 pseudo-ends; even at position one, two pseudo-ends contained an A rather than a T. We suggest that almost any sequence can transpose at some frequency, but that similarity to a Mu end determines the transposition frequency.

To test this assertion, we constructed precleaved donor fragments designed to *not* resemble Mu sequences except at the cleavage site. A radiolabel on the fragments showed that they did transpose, though at just barely detectable levels after an hour incubation. By comparison, a fragment with recognition sites taken from sequence 17 in Fig. 2 produced a similar amount of transposition product in only 7 minutes. A wild-type fragment, however, produced ten times as much product in two minutes or less. (Data not shown. Fragment sequences listed in experimental procedures. The "sequence 17" fragment is used again for Fig. 4 below, and also in chapters 2 and 4.

Phasing flexibility is permitted between half-sites.

Table 1 suggested that, under our experimental conditions, MuA can recognize half-sites with non-native alignments. We performed two experiments to further test this hypothesis. First, we designed precleaved donor fragments with base-pair insertions or deletions between half sites. This experiment is the conceptually obvious one, but its results are only suggestive, as explained below. Second, we manipulated the sequence of a pseudo-recognition site whose half-sites seemed to be one base-pair out-of-phase with each other. This second experiment provided more conclusive evidence in favor of flexibility.

In both sets of experiments, each donor fragment was assayed at multiple MuA concentrations. This was done because high MuA concentrations were necessary to see activity for the less-active fragments, but the high MuA made quantitation difficult for the more-active fragments. Also in both sets of experiments, transposition assays were done as single time-points

rather than time courses. As a result, we cannot say whether variability in transposition efficiencies reflected varying rate constants or varying end-points.

Fragment HSI ("half-site insertion"; Fig. 3a&b) contained an added "A/T" at roughly the boundary between the halves of each of R1 and R2. This fragment transposed at approximately 70% the efficiency of wild-type DNA (Fig. 3c, lanes 4-6). Fragment HSD (half-site deletion) had an A/T removed at each half-site boundary (Fig. 3a&b), and it transposed at ~50% the efficiency of a wild-type fragment (Fig. 3c). These results showed that fragments with altered half-site spacing can transpose, albeit with slightly reduced efficiencies. However, these results do not indicate whether MuA recognized fragments HSI and HSD with a rigid or a flexible alignment. Did the protein bind the optimal sequence, perhaps stretching or pinching to accommodate the altered spacing? Or did the protein recognize the rigid alignment, sacrificing sequence-specific contacts at the inner half of R1 and all of R2? As shown in detail in Fig. 3a, because the recognition sites contain stretches of repeating "A" a single base-pair insertion or deletion does not entirely remove specific contacts from the inner half-site.

We designed another fragment (fragment HSN -- "half-site null"), in which the inner half of the R1 site was exchanged for an "anti-consensus sequence", designed to contain the nucleotides most underrepresented at each position in the consensus (Schumacher et al., 1997) (Fig. 3a&b). This fragment transposed at only ~20% the efficiency of the wild type DNA – considerably worse than either fragment with altered spacing (Fig. 3c). The difference in efficiency between HSN and either HSI or HSD suggests that MuA did bind flexibly to fragments HSI and HSD. But we hasten to add that this experiment is not fully conclusive, since we cannot evaluate the stringency of the substitutions in fragment HSN.

Note that this experiment did conclusively show that, for the strand transfer reaction, the natural spacing is the best of the three spacings tested here, and is probably the best possible spacing altogether. This is true whether MuA bound fragments HSI and HSD with a rigid or a flexible alignment. If the latter, the reduced transposition efficiencies of 70% for fragment HSI and 50% for HSD would be due to an energetic cost for components of the transpososome to assume a less-than ideal geometry.

We next used a pseudo recognition site to test the possibility of flexible half-site spacing. Sequence 17 in Fig. 2 scored relatively well as a MuA recognition site; 8.5 out of 17.5 possible points. However, a deletion of one nucleotide near the half-site boundary would dramatically improve the sequence match, producing a new score of 13, compared to 13.5 points for the native R1 site (Fig. 4a). Does MuA bind this particular pseudo-recognition site with its domain Iβ one nucleotide out of register, in order to maximize sequence-specific contacts? We could test this hypothesis with base-pair mutations at positions that contribute favorably to the flexible alignment but not to the rigid alignment. We constructed fragment 17a, containing tandem copies of the recognition sequence from pseudo-end 17. Fragments 17b, 17c and 17d were variants of 17a, designed to remove some of the advantage of flexible binding (Fig. 4b). Note that the cleavage site sequence on these fragments was taken from the natural Mu right end, rather than from sequence 17.

The changes made to generate fragment 17b were particularly conservative. Fragment 17b differs from the original sequence (fragment 17a) at just two positions: positions 21 and 24. Each of the two mutations improves the fragment's rigid alignment with the MuA consensus sequence, but weakens the flexible alignment. Consideration of published mutagenesis data and of the preferences seen among the sequences listed in Fig. 2 also indicates that the two changes

in fragment 17b should weaken the flexible alignment but not the rigid alignment (see Table 2 and Fig. 6, below).

In five independent experiments, fragment 17a transposed at least twice as efficiently as any of the other three (Fig. 4c). These results show that the mutations in fragments 17b, c and d were at positions that, in the original sequence, had communicated sequence-specific information to the transpososome. Given that these positions already did not match the consensus sequence in a rigid alignment, the simplest explanation is that these nucleotides function in the original sequence via contacts involving flexible binding.

Although fragment 17a transposed more efficiently than its variants, it was substantially less efficient than a wild-type fragment substrate (data not shown.) This could be due in part to a penalty for changing the relative spacing of the two half-sites, as discussed above. In addition, although fragment 17a is a good match to the consensus sequence when aligned flexibly, it is still not a perfect match.

Spacing between recognition site and cleavage site is at least as stringent as between recognition half sites.

The statistical analysis in Table 1 did not indicate variability in spacing between the R1 and cleavage sites, but neither did it rule out this possibility. To directly investigate the effect of cleavage site position on transposition, we constructed two fragments with altered R1-to-cleavage site spacing. The native Mu ends contain five nucleotides between outer recognition site and cleavage site: fragment CSI (cleavage-site insertion) contained six nucleotides and fragment CSD (cleavage-site deletion) contained four nucleotides. The fragments were precleaved, so these experiments again probed the strand transfer but not the cleavage reaction. In addition, to prevent an improper cleavage reaction this set of experiments were performed in calcium rather than magnesium. Calcium supports strand transfer but not cleavage (Savilahti et al., 1995).

Fragments CSI and CSD transposed at approximately 45% and 10% the efficiency of wild-type, respectively (Fig. 5a&b). Thus, our results indicate a tolerance for an improperly spaced cleavage site. However, alterations in cleavage site spacing were at least as deleterious as alterations in half-site spacing (Fig. 5a&b).

Changes in cleavage site spacing alter the distribution of transposition products.

Inspection of Fig. 5a reveals that fragments CSI and CSD produced an increase in single fragment transfer (SFT) products as compared to double fragment transfers (DFT). This change in product ratio is distinct from the decrease in total amount of product discussed above. We expect that both SFTs and DFTs require synapsis of two fragments, but that SFTs are produced when one of the two fragments fails to transpose. Even wild-type fragments produce SFTs, though it is not clear what causes the transposition failures that lead to wild-type SFTs. The additional SFTs caused by alterations in cleavage site spacing is likely due to architectural imperfections in the transpososome – for example, failure to engage the DNA cleavage site in the protein's active site.

Are the effects of altered spacing transmitted across the transpososome, such that if one fragment in a complex is imperfect its synapsis partner also suffers? This question interests us because it probes the degree of rigidity in the transpososome structure, by asking whether the transpososome accommodates these DNA imperfections through local versus global changes. To address the question, we performed transposition reactions containing pairwise mixtures of the cleavage-site-insertion, the half-site-insertion, and the wild-type fragments (Fig. 5c). In each case, one type of fragment was radiolabeled, and the other was unlabeled and present at ten-fold

higher concentration. We found that the transposition frequencies of the labeled fragments were independent of the identity of the fragment in excess (Fig. 5c, white bars). Thus, under these conditions, the synaptic partners do not influence each other's transposition efficiencies.

Unlike the total transposition efficiency, the percentage of the total product that comprised SFTs depended only on the identity of the unlabeled, excess fragment (Fig. 5c, dark bars). For example, whenever CSI was the unlabeled partner, we saw the highest percentage SFT (~70%). Note that only radiolabeled products are visible in these experiments. Therefore, the dependence of SFTs on the unlabeled fragment indicates that labeled and unlabeled fragments paired-up indiscriminately. Taken together, these data suggest that the alterations in sequence described here are accommodated through localized, rather than global, perturbations in transpososome structure. This experiment also shows that the structure of an active transpososome can accommodate some asymmetry between its two halves.

Discussion

Functional pseudo-Mu DNA ends reveal sequence requirements for transposition.

The Mu transpososome engages multiple DNA sites, including several 22bp recognition sites and two separate cleavage sites (Craigie et al., 1984). Defining the relevance of each nucleotide position in this array of sites is a tremendous task. Several labs have conducted mutagenesis and chemical protection assays of the Mu DNA ends, and these have gone a long way towards elucidating the sequence requirements for transposition (Burlingame et al., 1986; Coros and Chaconas, 2001; Groenen and van de Putte, 1986; Lee and Harshey, 2001; Zou et al., 1991). The present study asked the protein to select its favorite among disfavored sites, and thus could detect subtle sequence preferences that designed mutagenesis would have been unlikely to find.

We sequenced 18 sites on a non-Mu DNA molecule (ϕX174) that had served as pseudo-Mu end sequences. To distill the information contained in those 18 sequences into an easily comprehended picture, we used a Pictogram analysis developed by Chris Burge (Burge et al., 1998). Figure 6 includes three Pictograms, each of which represents the 18 pseudo-end sequences in a slightly different way (see figure legend for details.) The results of the Pictograms generally agree with past mutagenesis and chemical protection assays, with some exceptions (Table 2). One study (Groenen and van de Putte, 1986) found no effect on transposition from individual mutations at any of 11 positions. Yet in our study, MuA favored matches to its consensus at seven positions that did not appear important in the mutagenesis, as well as at four positions that were not tested by mutagenesis. To illustrate the point further, the top letters of the Pictogram are those that occur most frequently in the 18 pseudo-ends; these top letters correspond well to the consensus sequence derived from the natural sites (written above the Pictogram). This agreement suggests that most base-pairs in the recognition sequence contribute favorable energy for transposition, though in some cases the contribution may be too subtle to detect by point mutations.

The Pictogram represents only the sequences immediately adjacent to the cleavage sites; we found no further sequence similarity in the DNA adjacent to that. Nonetheless, we assume that the observed transposition activity is due to MuA tetramers (ref). The absence of an identifiable second pseudo-recognition site suggests that the position and/or sequence requirements for an inner recognition site (R2/L2) are less rigid than for an outer site (R1/L1).

Cleavage site sequences are highly selected.

The pseudo-ends are highly selected for their cleavage sites (Fig. 6). Among transposons in general, cleavage sites tend to have an identical sequence on the two transposon ends. It is

tempting to speculate that the cleavage site sequence is so constrained because it is intimately engaged in the protein's active site and perhaps directly involved in catalysis. However, the same would be true of the cleavage sites of modular restriction enzymes, for example the type II enzymes *FokI* and *SfiI*. Yet these enzymes have tightly constrained recognition sites rather than cleavage sites (Kim et al., 1988; Williams and Halford, 2001), arguing that there is no intrinsic mechanistic reason why the cleavage site should bear the tightest constraints. The cleavage site of a transposon defines which base-pairs will transpose and which will stay behind. Transposases may have evolved the ability to strongly discriminate their correct cleavage sites, so as to protect the integrity of the transposon ends from one transposition event to the next.

The particular sequence TG/CA marks the cleavage sites of many distantly related transposable elements, including retroviruses and bacterial insertion sequences (Esposito and Craigie, 1998). It is not clear why this is so. The most favored cleavage site of the *SfiI* restriction enzyme is also CA, but *SfiI* cleaves between the C and the A whereas transposases cleave after the A (Williams and Halford, 2001). Analysis of naked DNA by computational modeling and other methods reveals unique structural flexibility for the dinucleotide TG/CA (Packer et al., 2000). It is uncertain whether this intrinsic DNA structure is relevant to enzymatic cleavage reactions.

Flexibility and rigidity within a MuA transpososome

Within a transpososome, MuA subunits use three distinct domains to contact three DNA sites on the transposon ends (Clubb et al., 1994; Clubb et al., 1997; Rice and Mizuuchi, 1995; Schumacher et al., 1997). The cleavage site is contacted by the catalytic domain, and the two halves of the recognition site are contacted by the two subdomains of the bipartite domain Iβ-γ. How flexible are these various binding modules in their relative positioning? Statistical analyses of the 18 pseudo-ends, together with transposition assays of specifically designed Mu end fragments, provide a partial answer.

Why does cleavage site spacing matter? Within a transpososome, a cleavage site and its adjacent recognition site are bound by separate MuA subunits (Aldaz et al., 1996; Namgoong and Harshey, 1998; Savilahti and Mizuuchi, 1996; Williams et al., 1999), whereas the two half-recognition sites are bound by tandem domains of a single subunit (Lavoie et al., 1991; Mizuuchi et al., 1991; Schumacher et al., 1997; Zou et al., 1991). One might therefore have expected higher tolerance for insertions near the cleavage site than within the recognition sites, but in fact the opposite is true (Table 1 and Fig. 5).

The constraints on cleavage-site position may indicate extensive intersubunit interactions. In addition, the crystal structure of the synaptic complex of a related transposase, from Tn5, shows that the catalytic domain and DNA-binding domain of opposite subunits lie immediately adjacent to each other on the DNA (Davies et al., 2000). The subunits' close proximity to each other probably limits the ability to shorten the DNA. Finally, it is possible that the catalytic domain of MuA makes some contacts in the outer-half recognition site, overlapping the contacts made by domain Iγ of the opposite subunit. If so, moving the cleavage site may disrupt the contacts to the catalytic domain, which could account for activity losses.

Altered DNA spacing is accommodated locally in the transpososome and affects a late step of transposition. Mixing heterologous dsDNA fragments revealed that the spacing of sites on one fragment does not effect its synaptic partner (Fig. 5c). This suggests that the transpososome accommodates altered spacing through localized changes, rather than through global changes in protein structure that may be transmitted between subunits.

Two lines of evidence suggest that, with the precleaved fragments used in this study, positional-shifts affect a step that occurs after initial transpososome assembly. The results are most striking for the fragments altered in cleavage-site spacing, and they contrast with a similar study of the restriction enzyme *SfiI,* for which altering the spacer between its two recognition sites does harm complex assembly (Embleton et al., 1999). First, base-pair insertions or deletions in MuA fragments cause an increase in the relative number of single fragment transfers (SFTs) compared to double fragment transfers (DFTs) (Fig. 5a&c), and an assembly defect should not specifically increase complexes that perform SFT. Second, in the fragment-mixing experiments (Fig. 5c), altered and wild-type fragments seem to pair-up indiscriminately in synaptic complexes, as judged by the extent of SFTs in these reactions. This result implies that the altered fragments are defective at a post-synaptic step. Perhaps the altered fragments tend to fail in engaging the cleavage site in the protein's active site. Consistent with this idea, I showed in chapter 2 that MuA can form competitor-stable complexes on DNA fragments that lack a cleavage site, indicating that the initial commitment to a pair of DNA molecules does not require engagement of the cleavage site.

Flexibility between half-sites, at a cost. Analyses of 18 pseudo-Mu DNA ends suggest phasing flexibility of ±1 base-pair between recognition half-sites (Table 1). Transposition assays of various DNA fragments supported this assertion (Figs. 3&4), but also implied that the natural spacing is optimal. Others have found that adding or removing an A/T base-pair from an A/T stretch in the middle of R1 dramatically reduces *in vivo* transposition, consistent with the natural spacing being optimal (Burlingame et al., 1986).

Other protein families contain bipartite DNA binding domains, similar to domain Iβ-γ of MuA: for example, the POU, cut, myb and paired families of transcription factors (Harada et al., 1994; Herr and Cleary, 1995; Ogata et al., 1994; Xu et al., 1999). Pax6, a member of the paired family, has a rigid 15 residue tether connecting the two domains that comprise its bipartite DNA-binding domain. The tether interacts with DNA, directly contributing to binding specificity (Xu et al., 1999). By contrast, the structures of POU proteins Oct-1 and Pit-1 both suggest flexible tethers (Jacobson et al., 1997; Klemm et al., 1994). Consistent with the structural data, a number of POU proteins can bind and regulate DNA sites with variable arrangements of half-sites (Herr and Cleary, 1995; Klemm et al., 1994; Li et al., 1993; Scully et al., 2000). Yet even flexible POU proteins usually have an optimal spacing, analogous to the situation described here for MuA. Optimal spacing can be determined in part by the length of the amino acid tether, but also by cooperative interactions between the protein's two DNA binding domains (Klemm and Pabo, 1996).

NMR structures have been solved for the isolated MuA domains Iβ (Schumacher et al., 1997) and Iγ (Clubb et al., 1997). In the domain Iγ structure, the inter-domain tether is unstructured. The tether is only 13 amino-acids, whereas the tethers in POU proteins range from 15 to 56 amino-acids. The two MuA domains have been modeled docked on DNA (Schumacher et al., 1997), and the authors of this model suggest that upon binding DNA the tether becomes structured and interacts directly with the DNA. This suggestion was made to explain the pattern of nucleotides that are protected from chemical digestion when MuA is bound to its full recognition site (Zou et al., 1991). The flexibility in half-site spacing described in this study is perhaps most consistent with an unstructured tether. However, it is possible that the DNA-bound tether is structured, but that the transpososome overall retains enough flexibility to accommodate one nucleotide more or less – perhaps through changes in DNA structure rather than protein structure.

Biological role for flexible binding. Many transcription factors exploit their bipartite DNA binding properties for biological regulation (Herr and Cleary, 1995; Lannoy et al., 1998; Scully et al., 2000). There is no evidence that bipartite binding plays as significant a role in Mu transposition, but flexible binding might play a role in interactions between MuA and the L2 site. L2 is considered to be a half-site, because it only contains the sequences recognized by domain Iβ of MuA (Fig. 2), and because its chemical protection pattern is less extensive than the other recognition sites (Zou et al., 1991). However, positions 11, 12 and 13 of L2 read "TCA" (Fig. 2), which matches the consensus sequence through flexible alignment to positions 10, 11 and 12. Thus, it is possible that MuA obtains favorable binding energy through transient contacts with these three base-pairs.

Multiple roles of the recognition sites in transposition.

In chapter 2, I discussed two roles for the MuA recognition sites: positioning two MuA subunits close together, and allosterically activating transposase (Baker and Mizuuchi, 1992; Goldhaber-Gordon et al., 2002b). Here we showed that the MuA recognition sites also help define the cleavage sites, since moving the cleavage site reduces transposition efficiency. Among transposons in general, each transposon has a defined spacing between cleavage site and outer recognition site: for example, 5 bps on Tn10 (Kleckner, 1996), 8 bps on Tn7 (Craig, 1996). In contrast, both the position and the number of recognition sites may differ on each of a transposon's two ends. Given the complexity in length and number of transposase recognition sites, it makes sense that they serve multiple functions during transposition.

Figure Legends

Fig. 1. Introductory schematics. Not drawn to scale. *A. Fragment transposition assays.* Donor fragments can be paired in a synaptic complex and transposed by MuA. The fragments usually include two MuA recognition sites (e.g. R1 and R2) and a cleavage site. If both fragments are properly transferred to the target DNA, the target appears to become linear, producing a DFT (Double Fragment Transfer) product. If one of the two fragments fails transfer, a supercoiled target relaxes, producing an SFT (Single Fragment Transfer) product. Either way, if the fragments are radiolabeled the label is incorporated in the DNA product, providing an easy assay for transposition efficiency. In the schematic, DNA fragments are drawn parallel to each other, for simplicity of presentation. *B. DNA binding by MuA transposase.* MuA transposase contains a bipartite DNA binding domain. Its subdomains, Iβ and Iγ, recognize the inner and outer halves respectively of a recognition site. In addition, a separate catalytic domain engages the cleavage site on the synaptic partner. The cleavage site is shown in bold. Not shown are the reciprocal set of interactions which are mediated by a second subunit. The DNA molecules are drawn anti-parallel, as they appear in the crystal structure of the related Tn5 transposase. The placement of the protein domains on the DNA in this drawing is not intended to convey information about actual modes of binding, as the structures were all solved in the absence of DNA. *C. Strategy used to isolate non-Mu donor sequences from a "hybrid" transpososome.* Synaptic complexes were formed between a DNA fragment containing the R1 and R2 recognition sites and a larger non-Mu DNA molecule (ϕX174 RF1). A separate 80 bp DNA fragment served as transposition target. Transposition joints were specifically amplified with a pair of PCR primers, one to the target and one to ϕX174.

Fig. 2. Sequences of 18 sites on a non-Mu DNA molecule (ϕX174 RFI) that were transposed by MuA. The sequences are listed 5' to 3', beginning with the complement of the nucleotide that was joined to the target. Rows 19 – 26 list the natural MuA recognition sites. The top and bottom rows of the table list a consensus sequence derived from the 6 natural MuA recognition sites and

the 6 natural D108 sites. In this and all figures, black squares indicate base-pairs that match this consensus. The numbers on the right side of the table (second-to-last column) are the point scores for the recognition sites (positions 6-27), as described in the text. The numbers in the last row identify sites that were isolated multiple times. We know that multiple isolates represent multiple transposition events, as opposed to sibling pairs of an identical clone, because each transposition was into a unique target site.

Fig. 3. Evidence for flexibility between recognition half sites. *A. Summary of fragment sequences.* Shown are nucleotides 5-32 on the non-transferred strands of specially designed donor fragments. This includes the cleavage site and the first recognition site—not shown is the second recognition site and the 5' overhang. (The complete fragment sequences are described in experimental procedures.) Fragment names are: "HSI"—half-site insertion, "HSD"—half-site deletion, "HSN"—Half-site null. The capital "A" denotes an inserted base-pair. The point scores, on the right side of the figure, were calculated as for figure 2. *B. Schematic of the same fragments.* Light gray box – outer half-site sequence (R1 or R2). Dark gray box – inner half site sequence (R1 or R2.) White box – spacer nucleotide. Hatched box – anti-consensus sequence. *C. Autoradiogram showing transposition of radiolabeled fragments into ϕX174 target.* In these experiments, ϕX174 served primarily as target, not donor. Smears at high MuA concentration result from reuse of individual target DNA molecules in multiple transposition reactions. [fragment] = 50nM. **[MuA] = 50, 100, 150 nM.**.

Fig. 4. Further evidence for flexibility between recognition half sites. *A. Rigid and flexible alignments of the pseudo-recognition site from sequence 17.* The arrow highlights the alignment shift. *B. Summary of fragment sequences.* Donor fragments contained recognition sites based on sequence 17 in table 2 and cleavage sites from the natural Mu right end. Capital letters represent base-pairs that were mutated from the original sequence 17. The fragment sequences are shown aligned with the MuA consensus sequence in two ways: a rigid alignment, and a flexible alignment based on a one-base-pair shift after position 14. This figure does not show entire fragments, which include two copies of the same pseudo-recognition site. The complete fragment sequences are listed in experimental procedures. *C. Autoradiogram showing transposition of fragments 17a,b,c and d.* [MuA] = 400, 800, 1600 nM.

Fig. 5. Analysis of effects of base-pair insertions or deletions. *A. Comparison of cleavage site vs. half-site phasing.* The autoradiogram shows the products of transposition of precleaved donor fragments into ϕX174 target. New fragments are CSI (cleavage site insertion) and CSD (cleavage site deletion), which contained a base-pair insertion and a deletion, respectively, between positions four and five. Fragments HSI (half-site insertion) and HSD (half-site deletion) are as described in Fig. 3. Transposition reactions were performed in 10mM $CaCl_2$, as opposed to 10mM $MgCl_2$, to prevent donor cleavage. (The previous experiments, shown in Figs. 3 and 4, were performed in $MgCl_2$.) [Fragment] = 125nM. [MuA] = 100nM. *B.* The bar-graph summarizes results of four experiments, including the one shown in Fig. 5a. Error bars represent one standard deviation. *C. Fragment mixing experiment reveals autonomy of the two transpososome halves.* Transposition reactions were performed with 50 nM unlabeled fragment, 5 nM labeled fragment, 150 nM MuA, and 10 mM $CaCl_2$ in place of $MgCl_2$. White bars represent the percentage of radiolabel that became incorporated in strand-transfer products. Dark bars represent the percentage of the total that comprised SFTs (single fragment transfers). Bars represent the mean of three independent experiments. Error bars represent one standard deviation.

Fig. 6. Pictogram analysis of 18 pseudo-end sequences. The sequences are listed individually in Fig. 2. The size and position of each letter in the Pictogram represent the frequency of appearance of that nucleotide at that position on the non-transferred strand of the pseudo-ends. The consensus sequence, listed above the pictogram, is derived from the 6 natural Mu recognition sites and the 6 natural D108 recognition sites. Underlined letters in the consensus are positions where the pictogram and the consensus match. *A*. Analysis of 18 pseudo-ends, without manipulation of the sequences. *B.* Analysis of 18 pseudo-ends, of which sequences 1 and 5 each had a nucleotide inserted at position 15, and sequence 17 had a nucleotide removed at position 15. These changes resulted in an improvement of ≥ 5 nucleotides (or 3.5 points) per sequence. We chose to manipulate the alignment of only those three sequences because similar changes to other sequences gave improvements of no more than 3 nucleotides (or 2.5 points). *C.* For this analysis, sequences 1, 5 and 17 were manipulated as in Fig. 6b. In addition, sequences that were isolated multiple times were counted that many times in the analysis (see Fig. 2).

Table 1. Statistical analyses of 18 pseudo-ends indicate flexibility between recognition half-sites.

	1. cleavage to first site (nts)[a]	*2.outer to inner half of outer site (nts)[b]*	*3. mean score for outer site (points)[c]*	*4. mean score for random (points)[d]*	*5. std dev for random group of 18 (points)[e]*	*6. # of std devs from random (points)[e]*
1	5	1	6.50	5.41	0.42	2.6
2	4 to 5	1	7.14	6.28	0.38	2.3
3	5 to 6	1	7.00	6.28	0.38	1.9
4	5	1 to 2	7.56	6.00	0.41	3.8
5	5	0 to 1	7.31	6.00	0.40	3.3
6	5	0 to 2	7.97	6.33	0.39	4.2
7	5	0 to 4	8.33	6.77	0.38	4.1
8	5	0 to 6	8.44	7.06	0.38	3.6
9	5	0 to 10	8.53	7.37	0.38	3.1
	outer to inner site	*outer to inner half of inner site*	*mean score for inner site*			
1	0	1	4.61	5.41	0.42	-1.9
1	0 to 1	1	5.31	6.28	0.38	-2.6
1	0 to 2	1	6.28	6.78	0.36	-1.4
1	0	0 to 2	5.53	6.33	0.39	-2.1

[a&b] Describes spacing parameters. We defined the outer half site as the consensus YGTTTCAYT and the inner half site as RAARYRCGAAAC. The precise boundary between the two half-sites is not known; the two might even overlap slightly.

[c]Mean point score for 18 sites listed in figure 2.

[d]Mean point score for all 10,772 sites on both strands of ϕX174

[e]Standard deviation for the mean scores of randomly chosen groups of 18 sites, calculated as the standard deviation for the group of 10,772 individual sites, divided by $\sqrt{18}$.

[f]The number of standard deviations between the mean for the 18 pseudo-ends and the mean for the random population. (column 6 = (column 3+column 4/column 5)).

Table 2. Comparison of pseudo-end sequences to previously published data.

	consensus[a]	psuedo-ends[b]	mutagenesis[c]	protection/interference[d]	other
1	T	strongly favors T, then A	Left end favors T>>C>A>>G Right end favors T>>A>>C>>G		
2	G	favors G>A>T>C	Favors G>A>C>T		
3	N	favors T, then A	Left end favors T Right end favors A		Natural left end contains T, right end A
4	N	no bias	ND		
5	N	no bias	ND		
6	Y	favors Y	R2 C to T no effect		
7	G	strongly favors G (then A)	L1 G to A down	strong purine inteference	
8	T	favors T	R1 G to A no effect		
9	T	no bias	R1 C to A no effect		
10	T	strongly favors T	ND		
11	C	strongly favors C	L1 C to T down	strong purine interference	
12	A	strongly favors A	ND		
13	Y	favors Y	R2 or L1 C to T no effect	weak purine interference	
14	T	favors C or G	R1 or R2 G to A no effect		"C" is most favored here, yet no natural site has "C"
15	N	no bias	ND		
16	R	favors T or A bias against C	L1 G to A no effect		"R" in consensus perhaps reflects steric clash with "C"
17	A	no bias	ND	protection but no intereference	
18	A	favors A	ND	protection but no intereference	
19	R	no bias	L1 G to A no effect		
20	Y	favors Y	R1 C to T no effect		
21	R	strongly favors A, then G	R2 G to A no effect	weak purine interference	Of native sites, only L1 and L3 have A at this position.
22	C	favors C	R1, R2, L1, L2 C to T down	protection but no intereference	
23	G	favors G	R1, R2, L1 G to A down	strong purine interference	
24	A	favors A	ND		
25	A	no bias	ND		
26	A	favors A	L1 A to G no effect	protection but no intereference	
27	R	no bias	L1 A to G no effect		

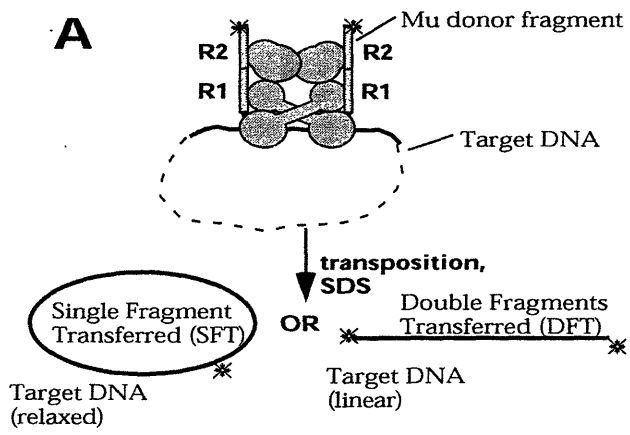
[a] See legend to Fig. 2.

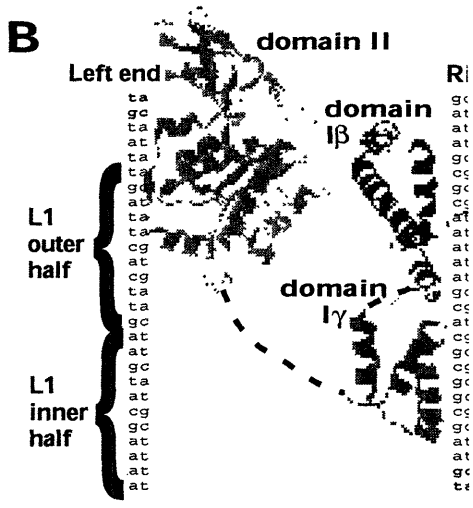
[b] Summarizes Fig. 6b&c, taking into account the consensus sequence in determining what constitutes "favoring".

[c] Positions 1-3, *in vitro* and *in vivo* data (Coros and Chaconas, 2001; Lee and Harshey, 2001);position 5-27, *in vivo* data (Groenen and van de Putte, 1986).

[d] Effects of DMS methylation (Zou et al., 1991). Interference--methylation interferes with MuA binding. Protection--MuA protects the nucleotide from methylation.

Figure 1, DNA Site in a Mu Transpososome





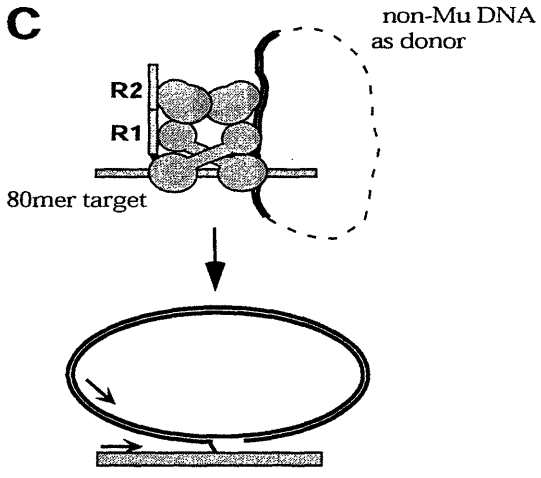


Figure 2, DNA sites in a Mu transpososome

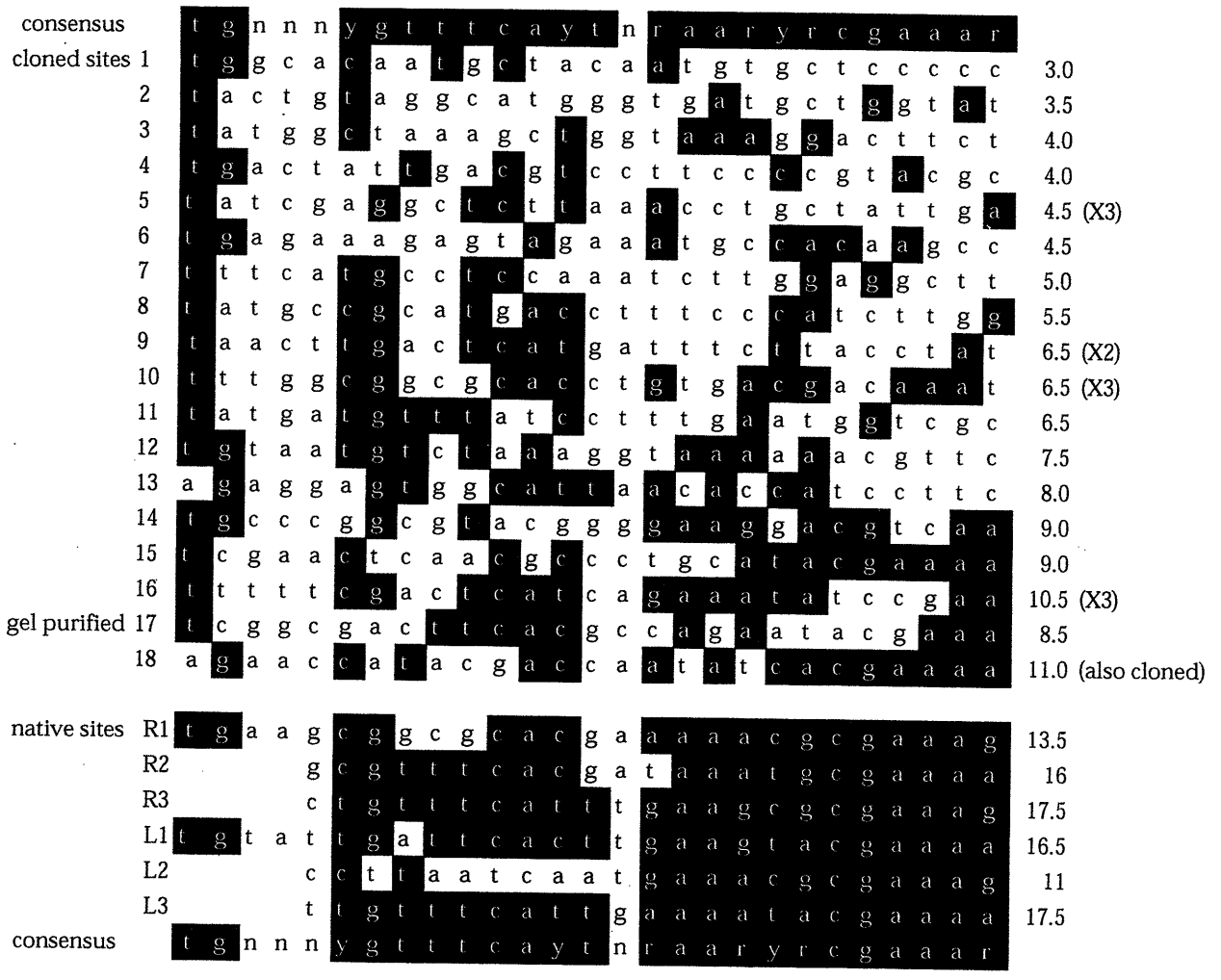
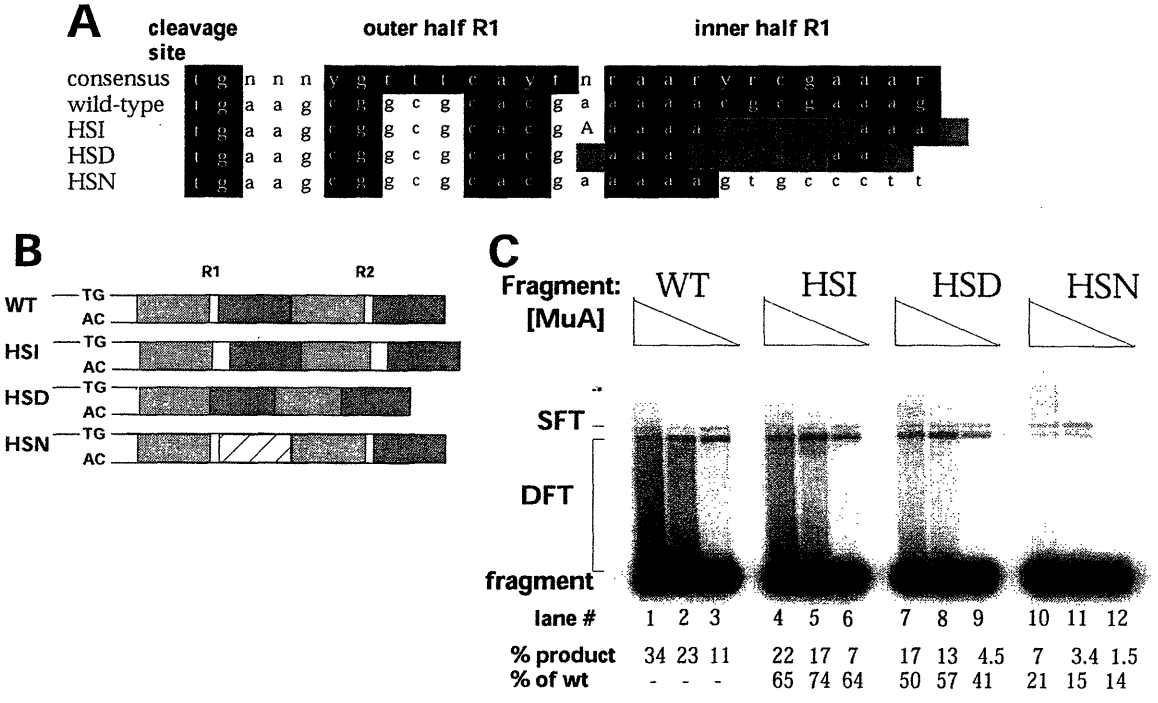


Figure 3, DNA Site in a Mu Transpososome



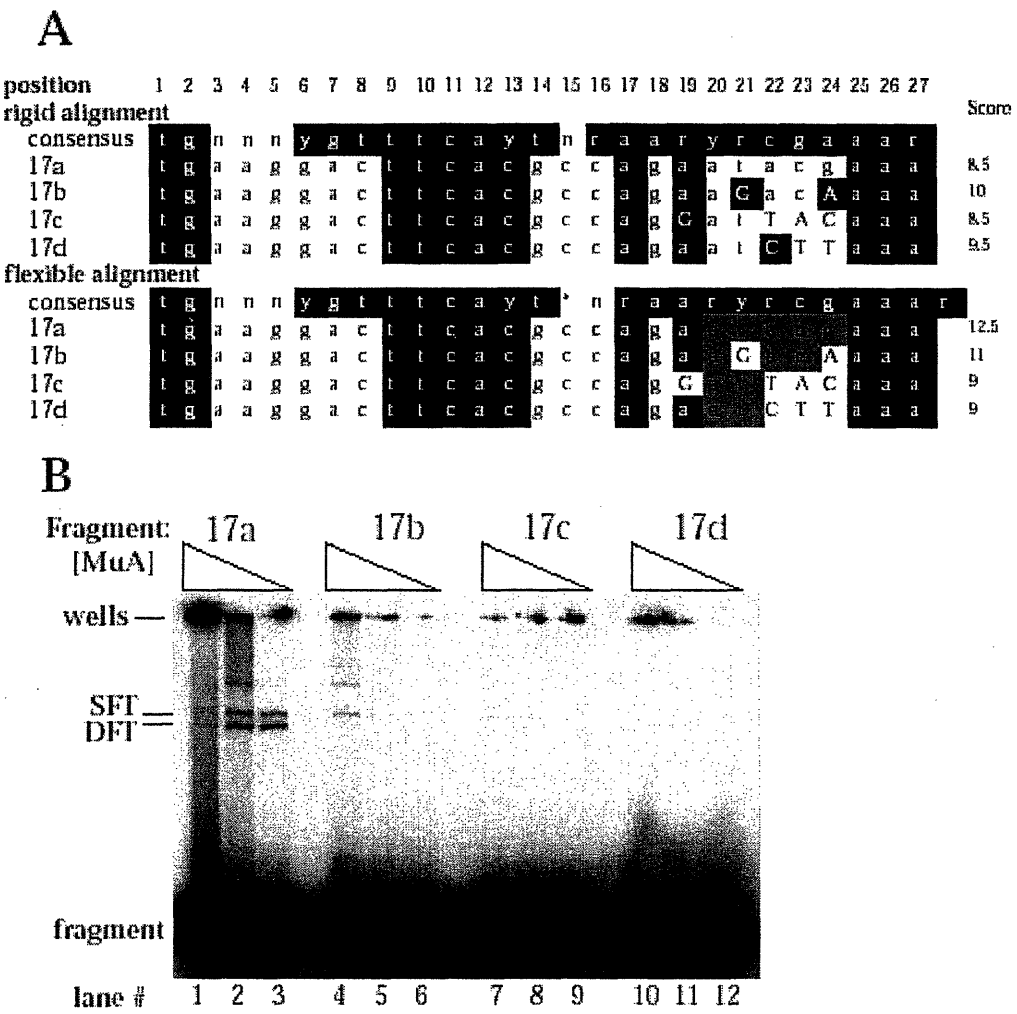


Figure 4, DNA Sites in a Mu Transpososome
Page 70

Figure 5, DNA Site in a Mu Transpososome

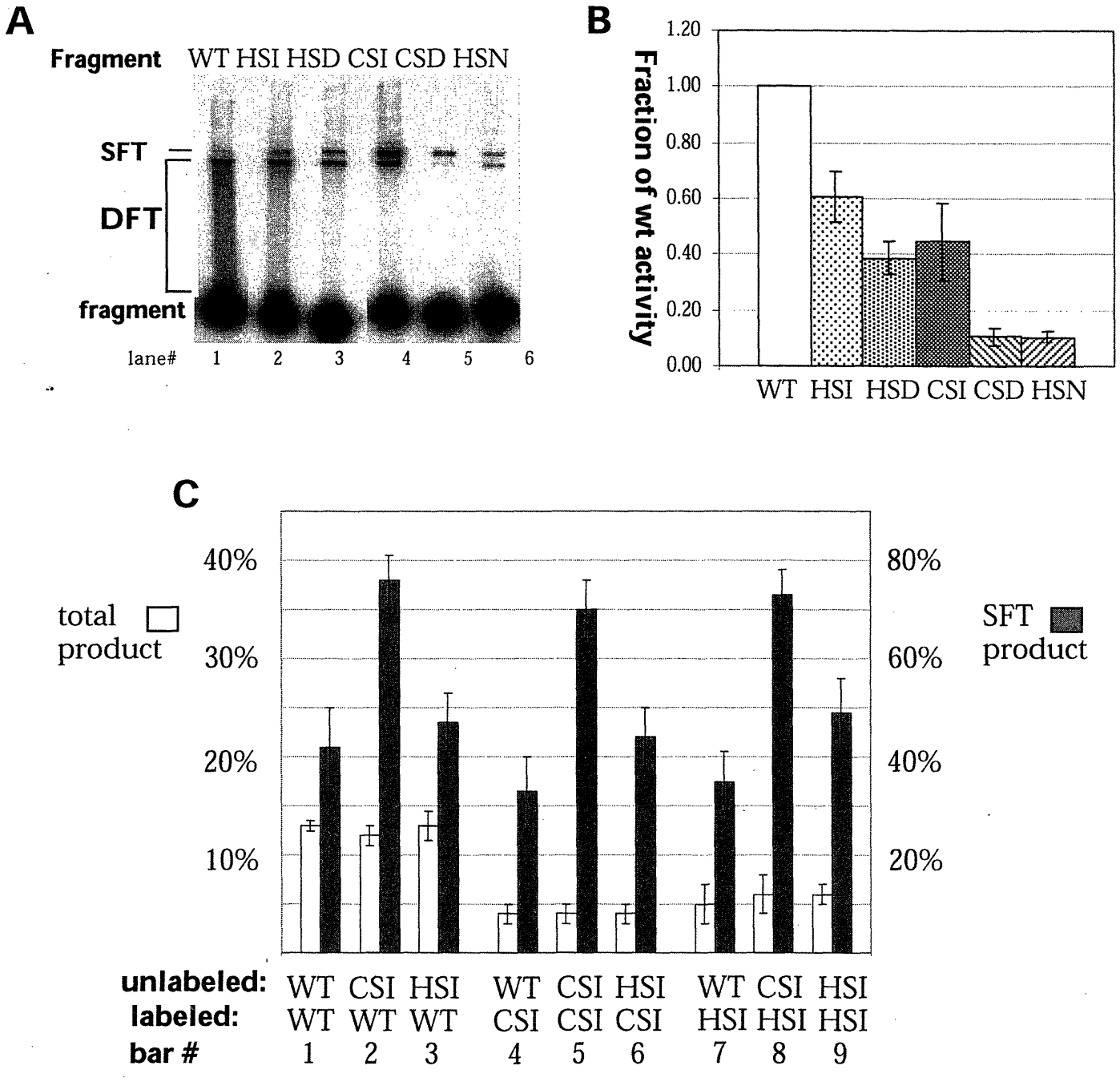


Figure 6, DNA Site in a Mu Transpososome

Chapter 4. DNA Recognition Sequences Are Not Critical for the Cleavage and Strand Transfer Steps of Transposition

Confronted with thousands of potential DNA substrates, how does a site-specific enzyme restrict itself to the correct DNA sequence? For several well-studied restriction enzymes, it has been shown that site-recognition is inextricably linked to catalysis. MuA transposase, like a restriction enzyme, performs a site-specific DNA cleavage reaction. MuA also performs a DNA strand exchange, or "strand transfer" reaction. However, unlike the best-studied restriction enzymes, transposase performs its reactions in ultra-stable, multimeric protein-DNA complexes, that assemble only when the proper DNA sequence cues are present. Thus the commitment step for transposition is not a cleavage event, but rather assembly of a stable complex. In this chapter I present kinetic analyses of transposition of DNA fragment substrates containing either wild-type transposon sequences or heavily mutated sequences. The mutations are restricted to the recognition sites, which are contacted by a DNA-binding domain of transposase, rather than the cleavage site, which is contacted by the catalytic domain. The mutations severely reduce both the extent and the rate of assembly of active complexes. Yet if complexes are pre-assembled, the first-order rate constants for either cleavage or strand transfer are unaffected by the mutations. I conclude that sequence-specific contacts contribute energetically to complex assembly, but not to subsequent steps of transposition.

Specific contributions by colleagues:
The experiments presented in this chapter are the result of a close and fruitful collaboration with Michael Early, who was for the past 18 months a technician in Tania Baker's lab. The original idea for this project was mine. I nurtured the idea for several years, and eventually I developed the assays to test it and conducted preliminary experiments for it. The experiments that actually appear here were almost all done by Mike Early, in close (usually daily) consultation with me. Mike and I discussed the contents of this chapter in detail before we wrote anything, and then he wrote the first drafts of the results section and he prepared the data figures.

Introduction

Enzymes that act on specific DNA sequences face a unique challenge. Surrounded by incorrect sequences -- potential substrates chemically similar to their own -- they must select their correct substrate. Decades of careful experimentation have suggested how this problem is solved by some enzymes, most notably the type II restriction enzymes EcoRV, EcoRI and BamHI. In brief, these enzymes use specific binding energy to lower the activation barrier for DNA cleavage, so that site-specific binding is inextricably linked to catalysis (Jen-Jacobson, 1997; Pingoud and Jeltsch, 1997). Like restriction enzymes, members of the transposase protein family perform site-specific modifications of DNA. But the molecular basis of the transposases' site-specificity is less well understood.

MuA transposase transposes the genome of bacteriophage Mu. To do so, MuA cleaves the 3' end of the Mu DNA away from its surrounding sequence, and then in a one-step "strand transfer" reaction it inserts the cleaved 3' end into a new DNA molecule called the target. All told, transposition requires 4 reactions of MuA: one cleavage and one strand transfer at each end of the Mu DNA (Chaconas, 1999; Mizuuchi, 1992). A single MuA subunit performs both reactions for one end, using a single active site (Namgoong and Harshey, 1998; Williams et al., 1999).

MuA has evolved complex strategies to ensure that it always transposes the entire Mu genome, and only the Mu genome. For example, MuA contains several DNA-recognition domains and a separate catalytic domain (Nakayama et al., 1987). The DNA-recognition domains bind selectively to sites internal to the Mu DNA, ensuring correct placement of MuA on the DNA. The catalytic domain engages the very end of the Mu DNA, cleaving precisely at the junction between the Mu DNA and the non-specific sequence that flanks it (Figure 1a). The Mu genome actually has three MuA recognition sites close to each of its ends, each of which can be bound by a separate MuA monomer. The six sites share a 22 base-pair consensus sequence, more than long enough to be unique within a host genome (Craigie et al., 1984). These long sites are bound by independently folding, tandem domains of MuA, domains Iβ and Iγ (Clubb et al., 1997; Schumacher et al., 1997). Iβ specifically binds the inner half of the recognition sequence, and Iγ binds the outer half (Schumacher et al., 1997). (A third DNA-binding domain, Iα, binds an enhancer-like sequence, which is dispensable in vitro and will not be discussed further in this paper (Leung et al., 1989; Mizuuchi and Mizuuchi, 1989).) DNA cleavage occurs 5 base-pairs outside of the first recognition site, and is marked by the cleavage site sequence 5'(T/A)CA↓.

MuA functions as part of a large protein-DNA complex called a transpososome or a synaptic complex (Craigie and Mizuuchi, 1987; Surette et al., 1987). At the core of the transpososome are four MuA subunits (Lavoie et al., 1991; Mizuuchi et al., 1992), at least two of which simultaneously bind both ends of the Mu DNA (Figure 2a) (Aldaz et al., 1996; Savilahti and Mizuuchi, 1996). Additional proteins and DNA sequences interact transiently with the core transpososome. For example, the MuB protein is a DNA-binding ATPase that interacts directly with MuA. MuB is best-known for its role in assisting MuA to find a target DNA, but it also stimulates other steps of transposition (Adzuma and Mizuuchi, 1991; Baker et al., 1991; Naigamwalla et al., 1998; Surette et al., 1991). Additional monomers of MuA may also associate themselves with the core transpososome, stimulating its activities (Wu and Chaconas, 1997).

The core transpososome is remarkably stable, resisting long incubations with competitor DNA or high levels of heat or urea. Yet in the absence of its recognition sequences, MuA is

monomeric, inactive, exchanges between DNA sites with a rapid k_{off}. Thus the recognition sequences stimulate multimerization of MuA and cause four MuA subunits to commit to each other and to their bound recognition sites (Goldhaber-Gordon et al., 2002b; Kuo et al., 1991; Lavoie et al., 1991; Mizuuchi et al., 1991; Mizuuchi et al., 1992; Surette et al., 1987).

Returning to our original topic, MuA transposase, and transposases in general, differ fundamentally from classic restriction enzymes. Transposases perform at least two types of DNA modifications, not one. They have distinct DNA recognition domains and catalytic domain, whereas these two functions are interwoven in a prototypical restriction enzyme. And, perhaps most importantly, transposases commit to a DNA sequence by assembling a stable transpososome on that sequence; DNA cleavage occurs only after this commitment. These differences between transposases and classic restriction enzymes raise questions about how a transposase uses the free-energy of specific DNA binding. Many restriction enzymes utilize this energy to lower the transition state for cleavage. Does a transposase also utilize its recognition sequences to enable cleavage or strand-transfer? Or do the recognition sequences directly contribute only to the commitment step, namely, assembly of a stable synaptic complex? Previous mutagenesis, sequence selection, and nuclease protection studies have been invaluable in delineating the contributions of individual base-pairs within the MuA recognition sites to Mu transposition overall (Figure 1b) (Goldhaber-Gordon et al., 2002a; Groenen and van de Putte, 1986; Zou et al., 1991). In the current study, we assess the contributions of entire halves of a recognition site to individual steps of transposition.

MuA is particularly tractable to this type of mechanistic study. A "fragment assay", in which short (~50 base-pair) fragments substitute for the large Mu DNA, is easy to manipulate and quantify. The fragment substrates usually contain the first 50 base-pairs from the right end of Mu: this includes the cleavage site and the two outermost recognition sites (referred to as R1 and R2; R1 is the site closest to the cleavage site). MuA pairs-up the fragments in synaptic complexes, mimicking two ends of a transposon, and the fragments are cleaved and transferred to a target DNA (Figure 1c). The fragments can also be synthesized to simulate pre-cleaved ends. Transposition assays can be staged into three discrete steps, each with unique metal-ion requirements: transpososome assembly, DNA cleavage, and DNA strand transfer. If uncleaved Mu fragments are used, the presence of calcium in the reaction mix permits transpososome assembly. Adding magnesium initiates cleavage, and in the absence of MuB and a target DNA this reaction does not go on to strand transfer. If the fragments are precleaved, they can assemble into transpososomes in the absence of divalent metal ions, and strand transfer may be initiated by the addition of target DNA, MuB, and either magnesium or calcium (Savilahti et al., 1995).

In order to study the contribution of the MuA DNA recognition sequence to individual steps of transposition, we have constructed Mu fragments with multiple mutations in the R1 recognition site (Figure 1b). By following the kinetics of each reaction step performed with these fragments, we assess the contribution of the mutated regions to that step. We find that the mutant fragments are severely compromised in their ability to assemble active transpososomes. Yet once assembled, the mutants can be cleaved with a rate constant equal to or higher than the rate constant for wild-type fragments. Similar results were obtained for the strand transfer reaction. We conclude that the recognition sequences contribute directly and substantially only to the assembly stage of transposition, the commitment step for the whole process.

Results

Mutant fragments are severely compromised as transposase substrates.

Our goal was to massively disrupt sequence specific contacts between MuA transposase and the transposon end sequences, and dissect the impact on individual steps of transposition. We designed three fragments with mutations that would severely compromise, but not eliminate, transposition (Figure 1b). For two of the fragments, mutations were only in the R1 site, as previous studies indicate R1 to be more important than R2 in the overall transposition process (Goldhaber-Gordon et al., 2002a). The "outer mutant" is most severely disrupted close to the cleavage site, in the region that is contacted by MuA domain Iγ. This was the most active of the three mutants (see below). The "inner mutant" is also disrupted in the R1 recognition site, but in the cleavage-site distal half of R1. This region is normally contacted by domain Iβ of MuA. The R1-R2 mutant contains mild mutations throughout both the R1 and R2 sites. Not all fragments were used in all experiments.

Each of the three mutants was, as planned, a poor transposition substrate. Under our standard reaction conditions, incubation of MuA with any of these mutants did not produce cleaved product even after a 24 hour incubation (data not shown.) Fortunately, the mutant DNA molecules could participate in transposition if the reaction was separated into stages, so that the buffer conditions could be optimized for each stage (Figure 2).

Mutant fragments are specifically compromised in assembling active complexes.

To assess the mutants' ability to assemble into active transpososomes, we used reactions with two stages: an assembly stage, during which time points were taken, and a cleavage stage, to assay for active complexes (Figure 2a). During the assembly stage, transpososomes were assembled in a large reaction volume in the presence of 10mM calcium. At various time points we withdrew aliquots and added magnesium to 70mM to initiate fragment cleavage, and cleavage was allowed to go to completion (see below, Figure 4). The high Mg^{2+} concentration during the cleavage stage prevented further complex formation, as shown in Figure 2b by the lack of cleavage after 0 minutes of assembly. Thus, the final extent of cleavage in this second stage reflected the success of transpososome assembly during the first stage.

Of the two mutants tested in this assay, both assembled very slowly into active complexes. After more than 24 hours, only 15% of the outer mutant and 5% of the inner mutant had assembled into active complexes (Figures 2b and c, and the average of 3 experiments). By that time, the assembly reactions on these mutant fragments had slowed considerably but did not yet appear to reach full completion (Figure 2d). By contrast, assembly with wild-type sequences was essentially complete within two hours, with an average of 70% of fragments assembled into active complexes (Figures 2b and c).

Although the mutant reactions had not fully reached completion after 27 hours, they had slowed sufficiently to see that they would never reach the levels of active complexes possible with wild-type fragments (Figures 2c and d). We anticipated that the assembly reaction would be slow with these mutants; it was less certain that the reactions would also have a lower end-point. Adding additional MuA to these reactions prior to initiation of cleavage did not significantly increase the yield of cleaved product, so the low end-point is not simply due to loss of MuA activity (data not shown). More likely, the end-point is low because the mutant fragments become trapped in dead-end complexes. These hypothetical complexes could be inactive transpososomes, or they could be MuA-stimulated precipitates.

Since the accessory protein MuB would be present during an *in vivo* assembly, we also performed our assembly assays in the presence of this protein. MuB is known to directly

stimulate both transpososome assembly and strand transfer (Baker et al., 1991; Mizuuchi et al., 1992; Surette and Chaconas, 1991), and under some conditions it stimulates cleavage (Mizuuchi et al., 1992; Naigamwalla et al., 1998). In our assembly reactions, addition of MuB and ATP stimulated the rate of assembly on wild-type sequences, but had no obvious effect on mutant fragments (data not shown). Thus with or without MuB, DNA sequence strongly contributes to the rate and efficiency of transpososome assembly.

Recognition site sequence does not contribute directly to DNA cleavage.

Classically, substrate specificity directly impacts enzyme catalysis. But MuA is not a catalyst, because it performs its reactions in stable transpososomes that remain associated with their DNA substrates after the reactions are complete (Surette et al., 1987). So we wondered whether recognition site sequence directly impacts the cleavage rate, distinct from the assembly rate. To answer this question, we assembled transpososomes for 18 hours, and began taking time points after initiating cleavage (Figure 3a).

The cleavage data for all fragments fit well to a first-order rate equation (Figure 3b). Remarkably, the first-order rate constants for each of the three mutant fragments were equal to or even slightly higher than that for the wild-type fragment (Table 1). This can be seen visually by normalizing each data set to its asymptote, and thereby comparing the relative rates for each fragment type (Figure 3c).

To confirm our assumption of first-order or pseudo-first-order kinetics, we diluted the assembly mixture an additional ten-fold when initiating cleavage. This resulted in a ten-fold dilution of MuA and DNA during the cleavage stage, but no change in concentrations of other reaction components. The dilution did not change the rate constant, whether for wild-type fragments (Figure 3d) or for outer mutant fragments (data not shown), confirming the first-order fit.

The cleavage reactions reached completion with only a small percentage of mutant fragments cleaved (Table 1), fully consistent with the results of the assembly assays. Given the lengthy assembly and rapid cleavage, we find it most likely that the loss of activity occurred during the assembly stage and not during the cleavage stage, as discussed above. To test this assumption, we assayed cleaved complexes for their ability to participate in strand transfer after an extended incubation in cleavage buffer. If complexes are destabilized or inactivated by the cleavage buffer, the extended incubation should reduce strand transfer efficiency. Complexes were assembled on uncleaved fragments, either wild-type or outer mutant, and then allowed to complete cleavage. MuB and target DNA were then added to half the reaction mix; the other half sat for an additional three hours in cleavage buffer before the addition of MuB and target. The additional three hour incubation in the cleavage buffer prior to initiation of strand transfer did not significantly change the extent of strand transfer (data not shown). We conclude that the cleavage buffer does not especially cause transpososome instability. Together, this experiment and the dilution experiment (Figure 3d) strongly support our model of first-order or pseudo-first-order kinetics, confirming the similarity in rate constants for mutant and wild-type cleavage reactions (Table I).

Recognition site sequence does not contribute directly to strand transfer.

Since MuA transposase mediates two DNA modifications, we also investigated the contribution of recognition site sequence to the second modification, strand transfer. This was done using "pre-cleaved" transposon fragments, synthesized to terminate at the cleavage site. Assembly was conducted in the absence of divalent metal ions, because pre-cleaved fragments do not require divalent metal ions to assemble into active complexes, and because calcium (used above for

assembly of uncleaved complexes) supports strand transfer. Strand transfer was initiated by diluting the assembly reaction 1 in 10 into transposition buffer containing 70mM magnesium, target DNA, MuB and ATP (Figure 4a). The dilution and the high magnesium concentration prevented further assembly (data not shown). The strand transfer reactions are harder to quantify than the cleavage reactions, and only the outer mutant gave data of sufficient quality for analysis.

The strand transfer data also fit well to a first-order equation (Figures 4b and c). In an experiment with wild-type fragments, diluting the assembly mixture an additional ten fold upon initiation of strand transfer did not change the first-order rate constant for strand transfer (data not shown). This dilution experiment supports the first-order analysis.

The rate constant for the mutant fragment was indistinguishable, within error, from wild-type. As with cleavage rates, this can be seen visually by normalizing each data set to the asymptote for that set, and directly comparing the relative rates of strand transfer (Figure 4c). Thus substantial mutations in the outer half of the R1 recognition site do not directly impact a fragment's ability to participate in strand transfer. The strand transfer rate constant was 0.4 min^{-1}, roughly four times the cleavage rate constant (Table I).

Delimiting the rate determining steps under our reaction conditions.

Differences in metal ion requirements allowed us to separate transposition into its three primary stages: assembly, cleavage and strand transfer. Each of those stages includes many smaller reaction steps, for example binding of substrate, discharging and binding of metal ions, conformational changes in the protein and the DNA, and the actual covalent modifications of the DNA. Many of these smaller steps could follow first-order kinetics. Since most of our conclusions rely on determined first-order rate constants, we were curious to know which steps were rate determining in these reactions.

This analysis proved simplest for the strand transfer reactions, because MuB and target concentrations played a role in the rate of strand transfer (Figure 5). In reactions with wild-type fragments, diluting MuB ten-fold compared to the standard conditions greatly decreased the strand transfer rate. Diluting target by the same amount had no effect, but when *both* MuB and target were diluted, the rate decrease was even greater than if MuB alone was diluted. One function of MuB during transposition is to bind and deliver a target DNA molecule to the transpososome (Adzuma and Mizuuchi, 1988). Our results suggest that interactions between the transpososome and a MuB-target DNA complex were rate limiting in the strand transfer experiments. The same step may well be rate limiting for strand transfer *in vivo,* since MuB stimulation is required for robust transposition *in vivo* (Chaconas et al., 1984; Coelho et al., 1982).

We also performed several experiments to delimit the rate determining step for the cleavage reactions. Magnesium association was a possible rate determining step, since magnesium was present in great excess over the transpososomes and therefore its association was pseudo-first-order. If magnesium association was rate limiting, the apparent first-order rate constant would include a term for the magnesium concentration, and doubling the magnesium concentration would double the rate constant. In fact, if we doubled the magnesium chloride concentration during cleavage (from 70mM to 140mM), the rate constant increased, but only by 50%. Increasing [$MgCl_2$] to 210mM caused another 25% increase in rate constant (Table I). Similar increases in sodium chloride concentration did not alter the reaction rates (data not shown). Thus, the rate determining step does involve magnesium ions, but not in a simple way.

Dissociation of a Ca^{2+} ion, to be replaced by Mg^{2+}, could also be rate limiting. If so, we would expect that removing calcium prior to initiating cleavage ought to enhance the reaction

rate. To test this possibility, we added a 10 minute incubation with 15mM EDTA before initiating cleavage. This allowed 10 minutes for calcium to dissociate before we added magnesium and could begin taking time points. But rather than enhancing the cleavage rate, the incubation with EDTA slowed the cleavage rate constant by roughly a factor of 2, even if calcium was subsequently added back to its original concentration at the start of cleavage (data not shown). The final extent of cleavage was not substantially reduced, so we do not think that the EDTA incubation caused massive disassembly of the transpososomes. It may have caused temporary rearrangements in the transpososomes that were slow to repair themselves upon resupply of divalent metal ion. At any rate, these results suggest that calcium dissociation is not rate limiting for cleavage, but they are not conclusive because we do not understand why EDTA slowed the reaction.

We also asked whether the presence of MuB would influence cleavage rates. Generally, cleavage experiments in this study did not include MuB. Adding MuB and ATP to the cleavage stage, or to the assembly stage prior to a cleavage time-course, did not change the rate of cleavage (data not shown).

It is likely that the rate determining step for cleavage is a conformational change in the transpososome, necessary to engage the DNA cleavage site precisely in the protein's active site. In support of this, we found that mutations near the cleavage site (either mutation of position 2 from G•C to A•T, or a mismatch mutation at position 1 from T•A to A•A, Figure 1b) *do* significantly slow cleavage rates (Figure 6 and Table 1). Unlike the recognition sites, the cleavage sites are intimately engaged in the MuA active site (see chapter 5 of my thesis). Presumably, the wrong cleavage site nucleotide can disrupt the local structure of the active site and inhibit the cleavage reaction. Thus our results suggest that the rate determining step in the cleavage experiments involves both the active site and the cleavage site.

All told, we cannot say whether the rate determining step under these conditions is the same as under physiological conditions. In fact, cleavage reactions can be more rapid than reported here if the transposon sequences are contained within a circular plasmid (Mizuuchi et al., 1992). Thus it is possible that recognition site mutations do slow the physiological rate determining step of cleavage, and that that slowing is masked by a still slower step *in vitro*. However, if so the slowing must be mild, since a dramatic slowing of any step of cleavage would have slowed the entire cleavage reaction, and would have been detected here.

With either wild-type or mutant fragments, most complexes that are active for cleavage are also active for strand transfer.

Since the recognition site mutations did not directly influence the cleavage or strand transfer steps of transposition, we wondered if they function at the transition between the two. Uncleaved wild-type or outer-site mutant fragments were assembled into transpososomes and assayed for cleavage. Then MuB and target DNA were added for an additional incubation, and we assayed the extent of strand transfer. With each fragment, ~70% of cleaved complexes went on to participate in strand transfer. Thus for mutant and wild-type fragments alike, the majority of active complexes are fully active for all steps of transposition. Altogether, we conclude that specific binding energy between MuA and its recognition sequence does not directly contribute to its covalent modifications of DNA: not to the cleavage step, nor to the strand transfer step, and not to the transition between the two.

Discussion

MuA recognition sites regulate assembly.

Once MuA transposase and the Mu DNA assemble into a transpososome, they remain in that complex until actively removed (Levchenko et al., 1995; Mizuuchi et al., 1992; Surette et al., 1987). Thus assembling the transpososome is a commitment step in transposition, and is subject to many levels of regulation. In particular, the MuA recognition sites, located at the ends of the Mu DNA, are essential for assembly, ensuring that only the correct DNA, i.e. the Mu ends, is transposed (Baker and Mizuuchi, 1992). The current study presents the first kinetic analysis of transpososome assembly on mutant transposon end sequences. We find that mutations in the R1 recognition site dramatically reduce assembly rates. Presumably, assembly on a fully randomized sequence (no intact recognition sequence) would be poorer still.

The effect is primarily not at the level of simple binding. MuA has high non-specific DNA binding activity, and the mutant DNA fragments were fully occupied by MuA under our experimental conditions (unpublished results). Rather, past studies indicate that the recognition sites stimulates assembly by at least two mechanisms. Interactions with the recognition sequence induce allosteric changes in MuA that activate assembly, and the recognition sites help align the MuA subunits with respect to each other and with respect to the cleavage site DNA (Baker and Mizuuchi, 1992; Goldhaber-Gordon et al., 2002b; Namgoong et al., 1994).

MuA recognition site sequence is not critical during post-assembly events.

Mutations in the recognition site sequences did not reduce the relative rates of cleavage or of strand transfer, nor did they reduce the ability to transition from cleavage to strand transfer. The mutations did reduce the final extent of these reactions; as compared to fragments with wild-type sequences, fewer mutant fragments were cleaved or transferred overall. This result is most easily interpreted as an assembly defect – i.e. fewer of the mutant fragments were assembled into stable, active complexes, therefore fewer of them participated in later reaction steps. But for those that did assemble into active complexes, their ability to continue on with cleavage or strand transfer was unimpaired by their sequence modifications, as indicated by the first order rate constants for cleavage and for strand transfer. These results suggest that, once a transpososome is assembled, the fine structure of the MuA DNA binding domain does not greatly impact the catalytic domain.

Cleavage constants for two of the mutant fragments (outer mutant and R1-R2 mutant) were slightly higher than for wild-type. (The error on the third mutant was too great to detect this subtle difference). Precleaved R1-R2 mutant fragments are known to form transpososomes that are stable to competitor DNA but not to other chemical challenges (Goldhaber-Gordon et al., 2002b), and most likely the same is true for all the mutants used in this study. It is possible that the mutations raise the free energy level of the pre-cleavage complexes (the "SSC" or "type 0" complexes) more than they do the transition state for cleavage, which could explain the enhanced rate of cleavage .

Comparison to other proteins.

Like MuA transposase, restriction enzymes also have high affinity for non-specific DNA sequences. Though their affinity for cognate sequences can be several orders of magnitude higher than for non-specific sequences, this difference does not account for the far greater rate enhancements for cleavage of cognate vs. non-cognate sites (Engler et al., 1997; Maxwell and Halford, 1982). However, in several cases it has been shown that the *mode* of binding is unique at the cognate site. For example, EcoRV and EcoRI each severely distorts the DNA at their cognate sites, and this distortion is necessary to activate cleavage (Kim et al., 1990; Winkler et al., 1993). BamHI only mildly distorts its DNA site, but the protein itself is altered by cognate binding (Newman et al., 1995). For all three of these restriction enzymes, it is believed that the

free-energy of interactions between the enzyme and its cognate site lowers the activation barrier to cleavage. Interactions at non-specific sites do not have the same effect (Jen-Jacobson, 1997; Pingoud and Jeltsch, 1997; Vipond and Halford, 1993).

Unlike these restriction enzymes, MuA is a modular protein (Nakayama et al., 1987). This modularity makes it easy to believe that the fine structure of MuA's N-terminal DNA-binding domain does not directly influence the catalytic domain. Realize, however, that the protein's modularity does not *require* separation of function between domains. For example, the N-terminal DNA-binding domain undoubtedly influences other domain's inter-subunit interactions, since recognition-site binding triggers transpososome assembly (Baker and Mizuuchi, 1992; Goldhaber-Gordon et al., 2002b).

Like MuA, many other proteins (i) have modular structures, (ii) assemble into multimeric complexes only when bound to specific DNA sequences, and (iii) when in the multimeric form perform reactions on nearby sequences. MuA itself is part of a large family of transposases that share many properties, including the need to function within multimeric synaptic complexes (Rice and Baker, 2001). Two families of site-specific recombinases, the lambda integrase family and the gamma-delta resolvase family, are also prime examples. Another example is a restriction enzyme: *FokI* is modular, is monomeric in solution, and appears to function as a dimer (Bitinaite et al., 1998). Any of these other proteins could be like MuA, in that their recognition sequences may contribute directly only to complex assembly and not to subsequent reaction steps. Chimeras of FokI's catalytic domain and another protein's DNA-binding domain are able to cleave DNA efficiently (Huang et al., 1996; Kim et al., 1996; Kim and Chandrasegaran, 1994; Kim et al., 1997; Kim et al., 1998), showing that the fine structure of the DNA-binding domain is not directly critical for cleavage. Similarly, chimeras between two related site-specific recombinases, Gin invertase and ISX*c5* resolvase, can perform recombination (Schneider et al., 2000). Lambda integrase can function as a topoisomerase, catalyzing DNA cleavage and joining reactions, in the absence of its specific recognition sequences (att sites) (Kikuchi and Nash, 1979). This suggests that the energy of specific binding is more important for establishing the structure of the lambda integration complex than it is for performing covalent modification of DNA.

In general, site-specificity requires that there be a structural distinction between the specific protein-DNA complex and non-specific protein-DNA complexes. In some cases, that structural difference lowers the activation barrier for cleavage of the specific DNA site. In the case of Mu transpososomes, those structural differences allow assembly of an active complex

Materials and Methods

Proteins and DNA

MuA (Baker et al., 1993) and MuB (Yamauchi and Baker, 1998) were prepared as described. Target DNA (ϕX174 RFI) was purchased from New England Biolabs, and DNA fragments were synthesized by MIT/HHMI biopolymers lab and gel purified. The full sequence of the wild-type fragment was: (non-cleaved strand) ctagtgaagcggcgcacgaaaaacgcgaaagcgtttcacgataaatgcgaaaac/ (cleaved strand)
gttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgcttcactagacgcttggcgtaatcgggcgtaatgc/ (precleaved strand) gttttcgcatttatcgtgaaacgctttcgcgtttttcgtgcgccgcttca. Mutant fragments were identical to wild-type fragments, except for the changes indicated in Figure 1b and its figure legend.

Transposition reactions

Buffer conditions for all transposition reactions (assembly, cleavage, or strand transfer) were as described (Goldhaber-Gordon et al., 2002b), unless otherwise specified. All DNA fragments

used in the study were 5' labeled on the cleaved or transferred strand with T4 Polynucleotide Kinase. During assembly, MuA was at 200 nM and DNA fragments were at 50 nM; these were diluted 2-fold for cleavage, or 10-fold for strand transfer. During strand transfer, MuB was at 690nM and target DNA was at 2.8nM (or 10ng/μl). (In dilution experiments, the dilutions described in the figures were in addition to the dilutions mentioned here. Thus a ten-fold transpososome dilution for cleavage (Figure 4D) was really a 20-fold dilution, since the standard dilution was 2-fold.) Reactions were performed at room temperature (22±1°C), except the assembly steps in preparation for cleavage and strand transfer experiments were performed at 30°C to enhance the assembly rate.

Gel analyses

Assembly and cleavage experiments were analyzed on 8% acrylamide gels containing 0.05% SDS and 0.5x TBE buffer. Samples were heated to 40°C before loading on gel. After electrophoresis, gels were transferred to Whatmann paper, dried, and exposed to a Molecular Dynamics phosphorimaging cassette. Strand transfer reactions were analyzed on agarose gels, as described (Goldhaber-Gordon et al., 2002b).

Figure Legends

Figure 1. Introductory schematics. Not drawn to scale. (A) *Transpososome assembly.* The recognition sites are named "R" for right end or "L" for left end, and are represented here with boxes. Mu DNA is in grey, other DNA in black. One MuA subunit is shown in white, to highlight the crisscross structure of the transpososome. (B) *Summary of the fragments used in this study.* These are the sequences of the uncleaved strands (so that they could be listed 5' to 3' beginning at the end of the Mu DNA.) We list here only the cleavage site and R1 site sequences. Each fragment also contained a second recognition site, not shown in the figure: the natural R2 sequence for the wild-type, outer mutant, inner mutant and cleavage-site mutant fragments, and a repeat of the mutant sequence for the R1-R2 mutant. The fragments also contained DNA sequences beyond the cleavage site (identical for all four fragments). (Complete fragment sequences are listed in Materials and Methods.)

The consensus sequence (top line) is derived from the 6 naturally occurring MuA recognition sites, plus 6 sites from the closely related phage D108 (Craigie et al., 1984). In the consensus sequence, highlighted boxes are positions at which base-specific contacts were suggested by a sequence selection study (Goldhaber-Gordon et al., 2002a); darker highlights indicate stronger selection. W stands for T or A. In the fragment sequences, black boxes are positions that do not match the consensus. Note that the natural R1 site contains several black boxes: compared to the other natural sites, R1 has relatively weak affinity for MuA (Craigie et al., 1984).

The R1-R2 mutant sequence is based on a sequence from the genome of an unrelated phage, ϕX174. This sequence was shown to transpose at a low level when complemented by limiting concentrations of Mu end fragments, apparently due to fortuitous sequence similarity to the MuA recognition sequence. The grey boxes in the R1-R2 mutant sequence represent positions that match the consensus sequence with a phase shift of +1 near the center of the recognition site; several experiments suggest that this fragment is contacted with this phase shift (Goldhaber-Gordon et al., 2002a). Mutations in the outer and inner mutants were chosen based on an "anti-consensus" sequence – i.e. the nucleotides that appear most infrequently in the six natural sites. This anti-consensus binds poorly to the isolated MuA DNA binding domains Iβ and Iγ □□□□□□□□□□□1□□□11□□.

(C) *Transpososome assembly on DNA fragments.*

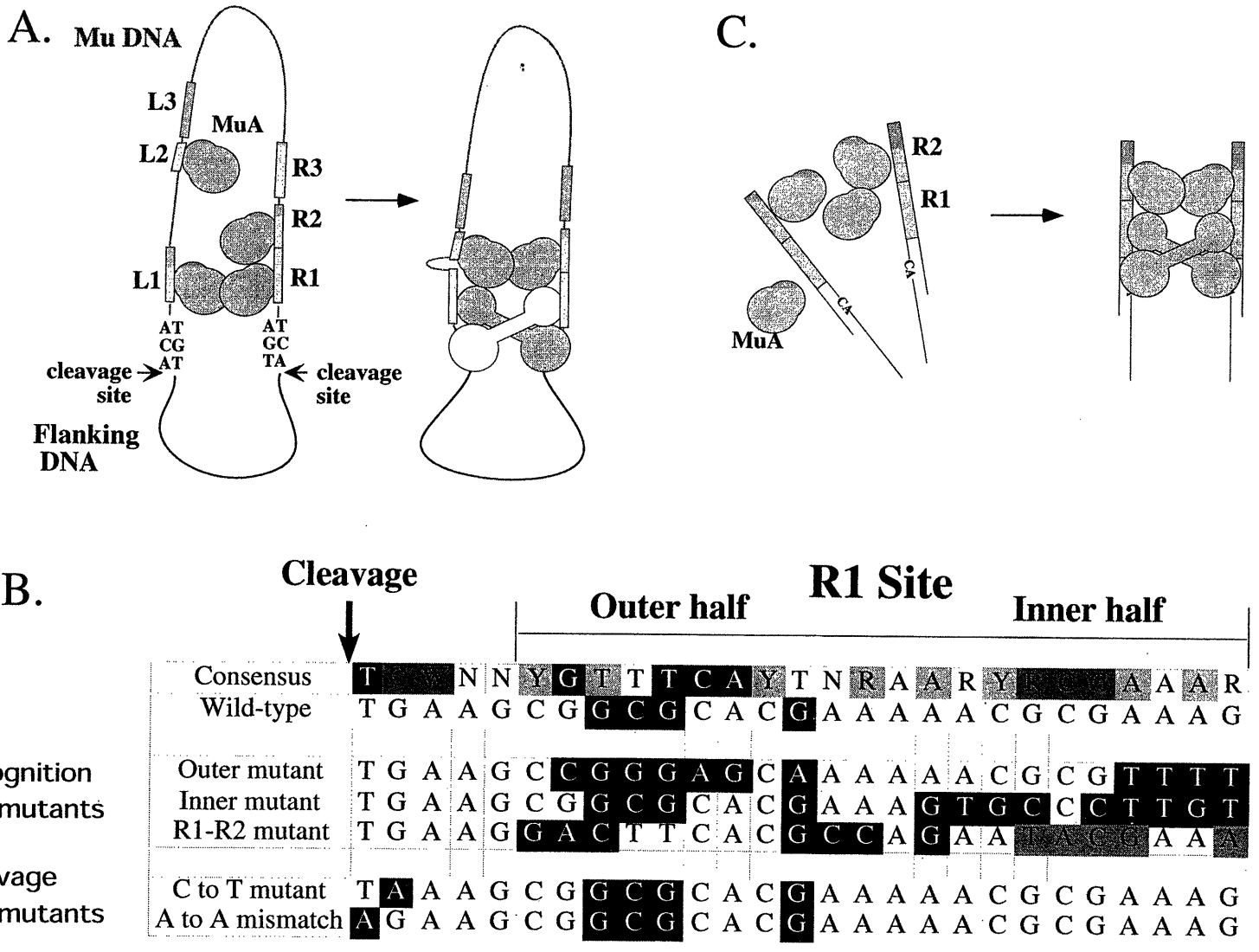
Figure 2. Mutant fragments assemble slowly into active complexes. (A) *Summary of experimental design.* This is a functional assay for assembly of active complexes. The read-out is extent of cleavage. (B) *Acrylamide gel from an assembly experiment, showing bands of cleaved and uncleaved fragments.* (C) *Graph of assembly rates.* This is a graph from a single assembly experiment. The data are fit to a second-order rate equation. (D) *The relative assembly rate is slow with mutant fragments.* This graph shows normalized results from three independent experiments for the wild-type and outer mutant, and two experiments for the inner mutant. Each data point was normalized to the asymptote for that data set, again using a second order rate equation. The normalization gives a sense of relative assembly rates, given that reactions with mutant fragments appear to reach completion with few active complexes assembled.

Figure 3. Mutant fragments are cleaved with rate constants similar to wild-type. (A) *Summary of experimental design.* (B) *Graph of one data set for each fragment type.* Data are fit to a first order equation. Table 1 summarizes results from multiple data sets. (C) *Normalization reveals that mutant fragments are cleaved with relative rates similar to wild-type.* This is the same data as is shown in Figure B, but each point was normalized to the asymptote for its set. The normalization shows visually what the rate constants, listed in Table 1, show numerically. The wild-type data is highlighted with a dark line. (D) *Dilution of pre-assembled transpososomes does not change the cleavage rate, supporting that the data fit first-order kinetics.* For simplicity only the experiment with wild-type fragments is shown, but the outer mutant fragment gave the same results.

Figure 4. Mutant fragments are transferred with strand-transfer rate constants similar to wild-type. (A) *Summary of experimental design.* (B) *Graph of one data set for each fragment type.* To limit the effects of scatter in the data points, each time point was analyzed on two separate agarose gels, and the data from both gels are included on the graph. Data are fit to a first-order equation. (C) *Normalization reveals that mutant fragments are cleaved with relative rates similar to wild-type.* This is the same data as is shown in Figure B, but each point was normalized to the asymptote for its set. The normalization shows visually what the rate constants show numerically.

Figure 5. The strand transfer reaction rate is dependent on [MuB] and [target DNA]. For simplicity we did not include a data set for the standard conditions, but the experiment in which target alone was diluted superimposes excellently on a standard curve. By contrast, diluting MuB (but not target) or target and MuB does cause substantial rate reductions. The data is again normalized to the asymptote for each data set, because diluting MuB also reduces the final extent of the reaction. Thus the absolute rates are even more dependent on MuB than are the relative rates represented here. For all experiments shown here, transpososome dilutions were standard (see Figure 4a).

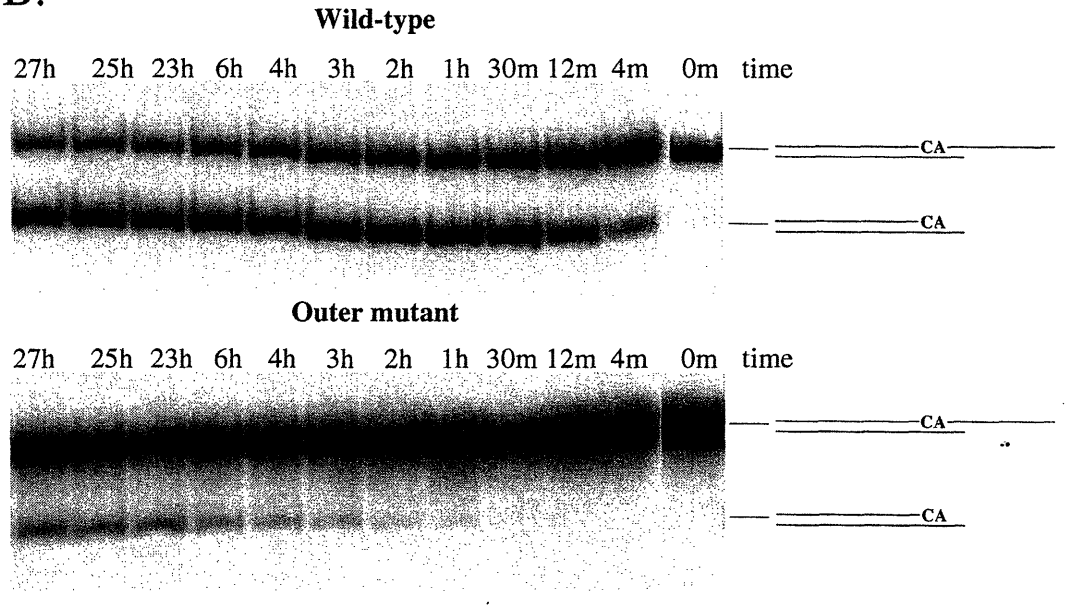
Figure 6. Mutations near the cleavage site slow the relative cleavage rate. These experiments follow the experimental design summarized in Figure 3A. The C to T mutant is a base-pair substitution at the second position from the Mu DNA end; A to A mismatch is a single nucleotide substitution on the uncleaved strand at the first position. The results are summarized in more detail in Table 1.



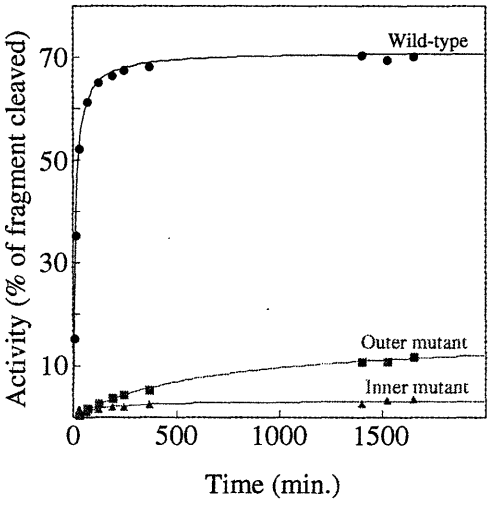
Recognition Site Contribute Only to Assembly
Figure 1
page 84

A.

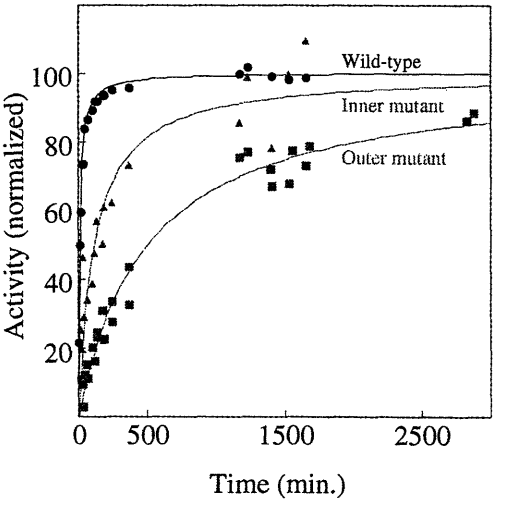
B.



C.



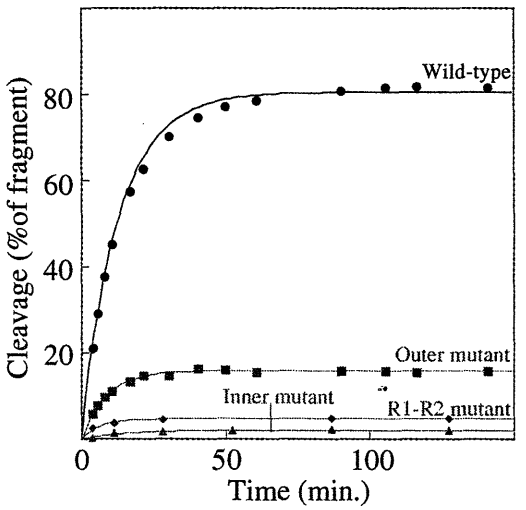
D.



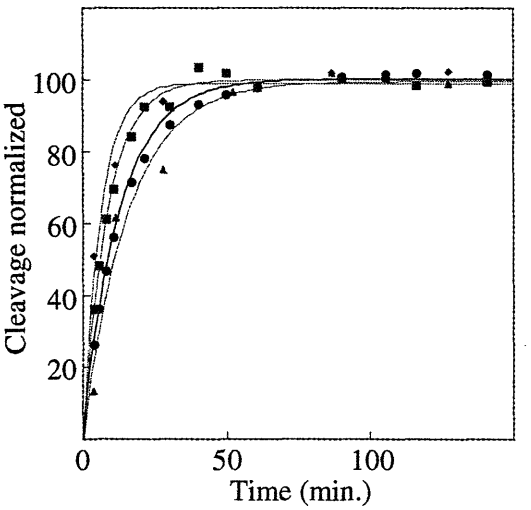
Recognition Sites Contribute Only to Assembly
Figure 2
page 85

A.

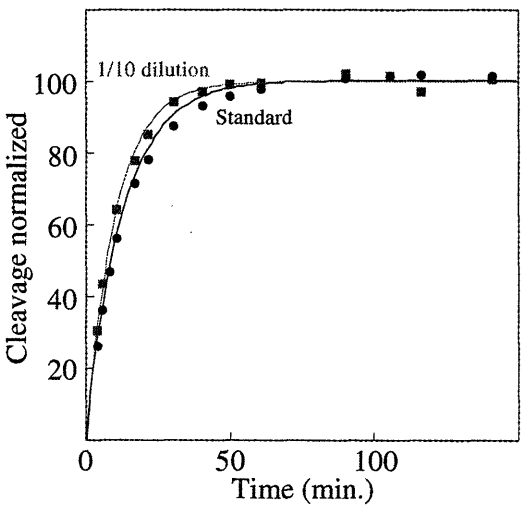
B. Cleavage



C. Normalized Cleavage

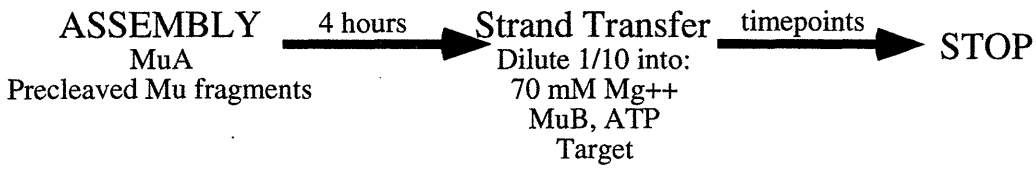


D. Dilution

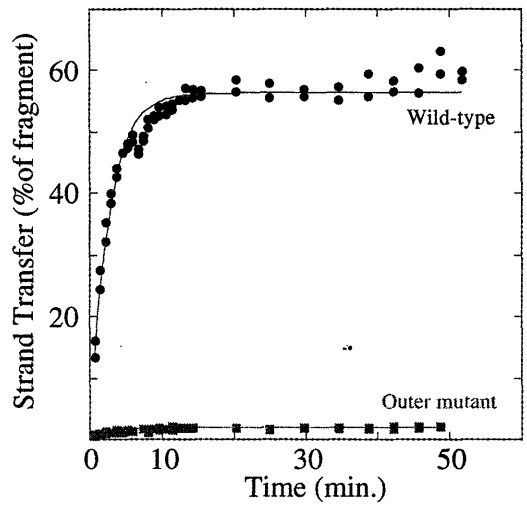


Recognition Sites Contribute Only to Assembly
Figure 3
page 86

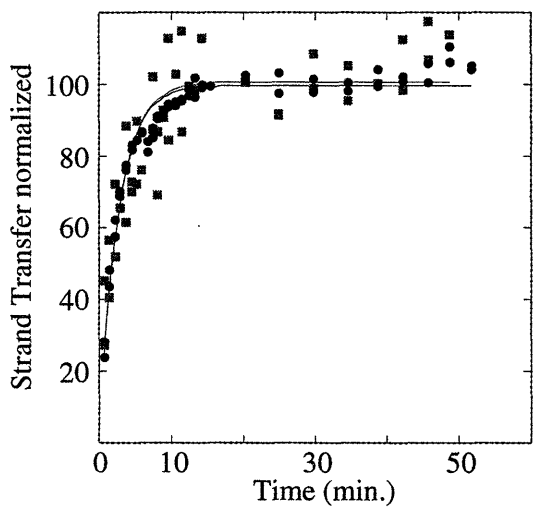
A.



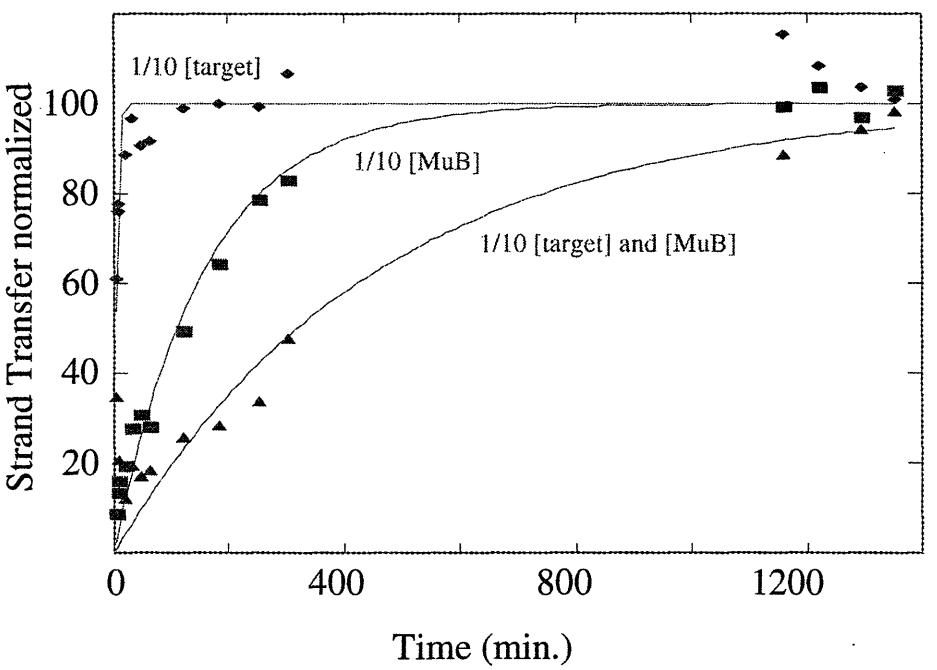
B.



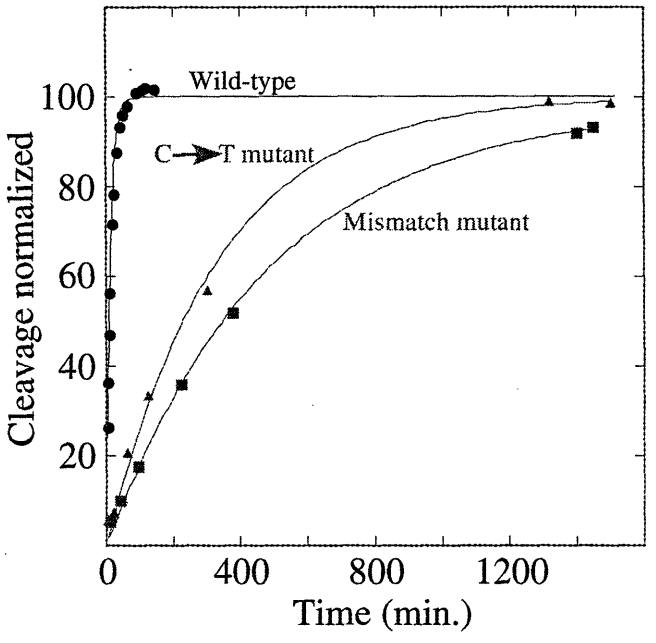
C.



Recognition Sites Contribute Only to Assembly
Figure 4
page 87



Recognition Sites Contribute Only to Assembly
Figure 5
page 88



Recognition Sites Contribute Only to Assembly
Figure 6
page 89

Chapter 5. DNA Activation Prevents MuA Transposase From Functioning as a Non-specific Endonuclease

A single transposase active site both hydrolyzes an end of the bacteriophage Mu DNA and joins that end to a target DNA. A long-term goal is to understand how these distinct reactions are orchestrated by one active site. We find that Mu DNA substrates that terminate in a dideoxy nucleotide allow transposase to hydrolyze a target DNA, in a reaction that combines aspects of the two natural reactions. Ordinarily, the presence of a 3'OH on the hydrolyzed ends of the Mu DNA causes DNA joining to be favored over target hydrolysis. We also find that the ribose of the last nucleotide of the Mu DNA activates transposase. This helps ensure the presence of the 3'OH in the transposase active site.

Introduction

Transposases, and their cousins retroviral integrases, are unusual proteins in that they use a single active site to perform multiple, distinct reactions. The overall function of these proteins is to transport a transposon (or a retroviral cDNA) from one DNA location to another (figure 1). To do so, the transposase must cleave the transposon ends away from the DNA that flanks it, and insert the cleaved ends into the new DNA location, the transposition target (Chaconas, 1999; Mizuuchi, 1992). Both the cleavage and insertion reactions occur in the same active site, and it is not clear how the distinct specificities of the two reactions are maintained.

To initiate transposition, multimers of transposase assemble into complexes called transpososomes, which hold together both ends of the transposon DNA. For transposition of the genome of bacteriophage Mu, the transposase is the MuA protein, and it assembles into a homotetramer when bound to specific recognition sequences near the ends of the Mu DNA (Lavoie et al., 1991; Mizuuchi et al., 1992). Short DNA fragments containing two transposase recognition sites ("end fragments") can also trigger transpososome assembly. The transpososome pairs two end fragments, mimicking pairing of the two ends of the transposon DNA (Savilahti et al., 1995). The cleavage site is five base-pairs beyond the recognition sites, and is marked by the sequence (T/A)CA↓ (figure 2a). Since the cleavage reaction defines the ends of the transposon, the terminal 3' nucleotide at the end of the Mu DNA is an A.

Though the two types of reactions performed by transposase are distinct from each other, they are chemically related (figure 1). Both are one-step substitutions of the oxygen groups on a DNA phosphate (Engelman et al., 1991; Mizuuchi and Adzuma, 1991). The major differences between the two reactions are (i) the identity of the nucleophile and (ii) the identity of the phosphate at the reaction's center. The first reaction, transposon cleavage, is a hydrolysis. A water molecule serves as nucleophile, and it attacks the phosphate 3' of the cleavage site A, freeing a 3'OH at the end of the Mu DNA. The second reaction is a DNA strand transfer. The 3'OH now serves as nucleophile, attacking a phosphate in the target DNA. The 3'OH at the end of the Mu DNA is thus joined to the target DNA. Unlike the Mu DNA, the target DNA has little sequence specificity. Any DNA molecule can serve as target, and although sites with the consensus sequence 5'-C(C/T)(G/C)(A/G)G-3' are most preferred, other sites are frequently selected (Haapa-Paananen et al., 2002).

During strand transfer, the two Mu ends attack phosphates that are precisely five base-pairs apart from each other, on opposite strands of the target (Allet, 1979; Kahmann and Kamp, 1979; Mizuuchi and Mizuuchi, 1993). The result is similar to a staggered double strand break in the target, except that the transposon DNA is inserted in the break. In fact, if the nucleophile at the strand transfer step were a water molecule, rather than a 3' hydroxyl, the target DNA would be cleaved on both strands (see figure 3a below.)

At the heart of the transposase active site are three conserved acidic residues, the DDE motif, essential for both cleavage and strand transfer, (Baker and Luo, 1994; Kim et al., 1995; Krementsova et al., 1998). A single transposase subunit contributes this DDE active site for cleavage and strand transfer of one transposon end (Namgoong and Harshey, 1998; Williams et al., 1999). Though there are many other examples of proteins that catalyze multiple reactions, transposases are unusual for doing so with a single active site.

What prevents transposase from hydrolyzing its target DNA, as it does the ends of the Mu DNA? Such non-specific endonuclease activity would be counterproductive biologically, as it would cause breaks in the host DNA while failing to promote transposition of the Mu DNA.

Understanding the mechanism for avoiding target hydrolysis will clarify the molecular mechanisms of the productive reactions, Mu cleavage and strand transfer.

Here we report that MuA transposase *can* produce double strand breaks in a target DNA. This target cleavage activity required transposon DNA fragments that terminated in a dideoxy A, so that they lacked the 3' hydroxyl which ordinarily serves as nucleophile for strand transfer. We also found that the 3' terminal nucleotide on the Mu DNA activates transposase. This finding provides insight into why strand transfer is ordinarily favored over target cleavage: the required presence of the terminal nucleotide in the transposase active site creates a great advantage for the attached 3'OH to serve as nucleophile.

Results

Dideoxy end-fragments support target cleavage by MuA.

We constructed transposon end fragments with pre-cleaved Mu sequences, but these new fragments terminated in a dideoxy A (figure 2a). If not for the dideoxy A these fragments would be good strand transfer substrates, but the missing 3'OH should prevent strand transfer. Incubation of these "ddA" fragments with transposase and a target DNA converted the circular target to a linear form. The product was susceptible to digestion with exonuclease V or lambda exonuclease (data not shown) and its electrophoretic mobility was the same as linear target's in any of several running buffers (figure 2b and data not shown).

Strand transfer of normal (non-dideoxy) transposon fragments also produces a product with similar mobility to linear target (figure 2b, lane 2). But the "ST" (strand transfer) fragments become covalently attached to the product, so that when the fragments are radiolabled the label is incorporated in the product. In contrast, radiolabeling ddA fragments did not result in transfer of the label to the target, confirming the ddA fragments' inability to complete strand transfer (data not shown). This is also reflected by a slightly faster electrophoretic mobility of the ddA products compared to the strand transfer products (figure 2B, compare ST and ddA).

This new reaction carried the DNA sequence specificities of a Mu strand transfer reaction. It required a complete Mu DNA end fragment; even a fragment one nucleotide short on the 3' strand (no-A fragment, figure 2A) did not support the reaction under the conditions shown (figure 2b lane 4). There were also no obvious sequence requirements for the target DNA: pUC19 (figure 2b) or ϕX174 RFI (figure 2c) were each suitable substrates. Moreover, the target was cleaved at many different sites, since digesting the linear product with a restriction enzyme that cleaves once in the target sequence produced a smear of products (data not shown). Together, these results show that transposase can introduce a sequence independent, double strand break in a DNA molecule, in a reaction that is similar to the normal strand transfer reaction.

Target cleavage requires the MuA active site.

The DDE residues coordinate divalent metal ions in the transposase active site (Lovell et al., 2002). Mutations in these residues abolished target cleavage (figure 2c) without perturbing transpososome assembly (data not shown). These results strongly suggest that target cleavage occurs in the same active site as Mu-DNA cleavage and strand transfer. Target cleavage required addition of magnesium or manganese to the reactions mix (data not shown).

Strand transfer substrates do not support target cleavage

Target cleavage shares features with each of the two natural transposase reactions: Mu-DNA cleavage and strand transfer. Like Mu-DNA cleavage, the nucleophile for target cleavage must be either a water molecule or another small molecule such as glycerol. But whereas Mu-DNA cleavage nicks only the 3' strand, target cleavage breaks both DNA strands. Thus during

target cleavage, as during strand transfer, both target strands are engaged in transposase active sites. Based on these observations, we propose that target cleavage is a frustrated attempt at strand transfer. The ddA fragments arrange the MuA active site to perform strand transfer, and in the absence of the 3'OH nucleophile a water molecule substitutes.

If this model is correct, the nucleophile for strand transfer should ordinarily outcompete the nucleophile for target cleavage, so that strand transfer substrates should not support target cleavage as a strong alternative reaction. To test the model, we designed an assay to sensitively distinguish the product of strand transfer from the product of target cleavage. The assay relied on the fact that target cleavage products should contain complementary 5' overhangs, excellent ligation substrates (figure 3a). To test for ligation efficiency, products of reactions with either strand transfer substrates (ST fragments) or target cleavage substrates (ddA fragments) were 5' radiolabled, and then lightly treated with T4 DNA ligase. The linear products of reactions with target cleavage substrates were ligated closed (figure 3b), confirming that these products contain complementary ends. By contrast, products of reactions with strand transfer substrates did not ligate closed. Thus no cleavage product was detected in strand transfer reactions, to within a 35-fold detection limit. These results indicate that the ability to perform strand transfer precludes hydrolysis of the target DNA. Together with results discussed above, these results strongly suggest that target cleavage is a variant of strand transfer. It occurs only when the transposase active-site is arranged for strand transfer, but is unable to complete the normal reaction due to the missing 3'OH nucleophile on the Mu DNA.

Strand transfer is faster than target cleavage.

How does the terminal 3'OH, nucleophile for strand transfer, prevent target cleavage? It could be that strand transfer is simply a faster reaction than target cleavage, and in t each transpososome strand transfer occurs before target cleavage has a chanve. Because the Mu DNA is a structural component of the transpososome, its terminal 3'OH is present at a high local concentration in or near the transposase active site. An exogenous nucleophile must bind the active site from solution, and the extra binding step for the exogenous nucleophile could represent a considerable disadvantage for target cleavage. If so, we would expect that even in the absence of the terminal 3'OH, target cleavage would occur significantly more slowly than strand transfer.

To test this model, we compared the rates of strand transfer and target cleavage. Transpososomes were assembled in the absence of magnesium and target DNA, using either ST fragments or ddA fragments. After addition of target DNA and magnesium, aliquots were removed regularly to assay the extent of appearance of linear product. The target cleavage reaction occurred with an initial rate ~7 times slower than the initial strand transfer rate (average of 4 experiments). This result supports the idea that the high local concentration of the terminal 3'OH causes strand transfer to be favored over target cleavage. Yet it leaves open the possibility that additional factors also favor strand transfer over target cleavage. In the previous section we showed that, in reactions with strand transfer substrates, the strand transfer product was at least 35 times more abundant than target cleavage product (figure 3b). This is significantly higher than the ~7-fold intrinsic rate difference, indicating that the presence of the 3'OH slows target cleavage beyond its independently slow rate. A number of models could explain this result, but one likely explanation is that the 3'OH directly occludes proper binding of an exogenous nucleophile.

The Mu DNA cleavage site strongly stimulates target cleavage.

Transpososomes can assemble on transposon fragments that lack the terminal adenosine (no-A fragments, figure 2a). But like ddA fragments, no-A fragments cannot participate in strand transfer, presumably because their recessed 3'OH is unable to engage as a nucleophile in the active site (Goldhaber-Gordon et al., 2002). We might have expected that no-A fragments would support target cleavage, but figure 2b suggests otherwise. Thus, the terminal adenosine, present in ddA fragments but absent in no-A fragments, must help activate transposase to perform target cleavage.

To explore this activation further, we conducted long time courses with truncated end fragments (figure 4a). Eventually the no-A fragment could be seen to support target cleavage, but the reaction was ~12 times slower than the dideoxy supported reaction (see legend to figure 4). A fragment missing the entire DNA cleavage site (5 base-pairs) did not support target cleavage even after a three day incubation (figure 4a; the small amount of linear DNA observed after >20 hour incubation was also present in reactions that lacked a Mu DNA fragment, and is presumably due to minor nuclease contaminant). These results show that: (1) the terminal A activates transposase, but is not absolutely required for activity, and (2) the terminal 5 base-pairs further activate transposase. The activation seen by the terminal A is not simply a result of improved transpososome stability. Over a 6 hour period, the number of transpososomes was similar whether the Mu fragment was ddA or no-A (figure 4b).

Since no-A fragments form stable transpososomes, but these transpososomes are relatively inactive for target cleavage, the terminal A must directly activate catalysis. Based on this observation, we propose that the terminal A is an important structural component of the active site. Moreover, since transposon cleavage occurs in the strand transfer active site, we propose that the terminal A also activates strand transfer.

Specifically the ribose of the terminal A activates transposase.

To dissect the requirements for activation by the terminal nucleotide, we constructed dideoxy-T fragments, in which the dT•ddA base-pair was replaced with dA•ddT (figure 2a), and assessed their ability to support target cleavage. We expected that this substitution would inhibit transpososome assembly (Coros and Chaconas, 2001; Lee and Harshey, 2001), so transpososomes were preassembled in a long incubation in the absence of target DNA. After this long preassembly, ddT fragments catalyzed target cleavage at a rate similar to catalysis by ddA fragments (figure 4a. See also the figure legend.) This result suggests that an adenine base per se is not required to activate transposase.

A likely possibility is that the terminal ribose is the critical feature of the terminal nucleotide. To test this idea, we constructed end fragments that terminated in an abasic site – that is, they contained the terminal ribose and its 3' OH but were missing the adenine base. These abasic fragments were good substrates for strand transfer (figure 4c). Their strand transfer rate was slightly slower than that of normal ST fragments (~0.5x the initial rate, average of 3 experiments), a rate difference that is far less than the ~12 fold decrease seen in target cleavage rates due to removal of the entire nucleotide (figure 4a, compare ddA to no-A fragments). Thus we conclude that a terminal base is not critical for activating strand transfer, suggesting that the ribose (or the attached 5' phosphate) is the critical activating feature of the terminal nucleotide.

Discussion

The terminal nucleotide activates transposase, contributing to avoidance of target cleavage.

Our results show that the terminal nucleotide of the Mu DNA – the A of the cleavage site – activates transposase during a target cleavage reaction. Given that target cleavage requires the DDE active site residues, we suggest that the terminal nucleotide is also important in structuring

the transposase active site for its too native reactions: Mu-DNA cleavage and strand transfer. This conclusion seems natural, because the 3' oxygen of the terminal A is a direct participant in both native reactions. Indeed, in the one co-crystal structure available for a transposase and a transposon DNA the transposon's terminal nucleotides are deeply engaged near the active site (Davies et al., 2000). It is impossible to test experimentally how full removal of the terminal nucleotide might affect the rates of the native reactions, because the terminal oxygen is a direct participant in these reactions. Thus target cleavage experiments revealed a role for the terminal nucleotide in catalysis that was previously not testable.

This nucleotide-dependent activation helps explain how target cleavage is avoided in nature. We have shown that MuA transposase is capable of catalyzing sequence-independent double strand cleavages, but the presence of the 3'OH at the end of the Mu DNA prevents it from doing so. We suggest that this is due to a high local concentration of the 3'OH in the transposase active site, providing a competitive advantage for its selection as nucleophile. In addition, the 3'OH may directly block an exogenous molecule from binding in the position of nucleophile. Either way, the 3' hydroxyl's presence is ensured by a structural contribution of the terminal nucleotide in establishing the active site.

The terminal nucleotide is at the heart of a model for transposition.

Based on experiments with transposon Tn10, a model has been proposed for the movement of DNA substrates in and out of a transposase active site (Kennedy et al., 2000). According to this model, the transposon's terminal nucleotide is held in one position in the active site throughout the transposition process. The water nucleophile, the DNA flanking the transposon, and the target DNA move in and out of the active site as they are needed, positioning themselves with respect to the terminal nucleotide. This model is appealing for several reasons, including the fact that the terminal nucleotide is the only substrate common to the multiple reactions normally performed by a transposase: transposon cleavage and strand transfer, as well as to two additional, related reactions that are performed by Tn10 transposase but not by MuA (Kennedy et al., 1998). Our data fit well with this model. We suggest that the terminal nucleotide remains in one position throughout the reactions because it is required to establish the functional conformation of the active site.

Retroviral integrases are promiscuous *in vitro*.

Unlike transposases, retroviral integrases have long been known to catalyze sequence-independent cleavage reactions: both hydrolysis and alcoholysis. Retroviral integrases are structurally and functionally related to transposases (Rice and Baker, 2001). Integrases ordinarily cleave several nucleotides from the 3' end of retroviral cDNA and then transfer the cleaved strand to a target DNA. However, in the absence of a retroviral DNA substrate, some integrases are able to cleave other DNA molecules. This non-specific cleavage can occur at many positions on the DNA molecule, distinguishing it from retroviral cleavage which occurs at a precise position near the end of the retroviral DNA (Katzman and Sudol, 1996).

Non-specific cleavage by integrases differs from the target cleavage reaction described here, in that it does not require retroviral DNA to activate it. This difference is consistent with other known differences between the *in vitro* activities of transposases and integrases. Transposases are only active in the context of a transpososome which includes two DNA ends, and transpososome assembly is highly dependent on DNA sequence (Baker and Mizuuchi, 1992). By contrast, integrases appear to function *in vitro* on single DNA ends (Vink and Plasterk, 1993). *In vivo,* integrases, like transposaes, function in synaptic complexes and, presumably, do not cause random hydrolysis of their host DNA. But *in vitro*, the success of a

transposase or integrase at avoiding target cleavage depends on how active the protein is in the absence of proper DNA end sequences, when small molecules are the only nucleophiles present. If the protein absolutely requires its DNA end sequences for reactivity, the 3'OH at the DNA ends will inhibit target cleavage. This observation highlights the regulatory importance of the transpososome, a DNA complex whose existence depends on transposon end sequences.

Acknowledgments

We thank Stephen Bell, Briana Burton, Elke Kuester-Schoeck and Sergei Saveliev for comments on the manuscript and/or helpful discussions. We thank Richard Cook and the people at MIT/HHMI Biopolymers Lab for their good nature in constructing unusual oligos.

Experimental Procedures

Protein and DNA: MuA and MuA mutants were purified as described (Baker et al., 1993). ϕX174 RFI target DNA was purchased from New England Biolabs (NEB), as were all non-MuA proteins, and pUC19 target was purchased from MCBI Fermentas. DNA fragments were purchased from MIT/HHMI biopolymers lab, and gel purified. The abasic phosphoamidite was purchased from Glen Research. The purity of the abasic fragment was confirmed by Mass Spec (data not shown). Transposon fragment sequences were:
gttttcgcatttatcgtgaaacgctttcgcgttttcgtgcgccgcttca/ctagtgaagcggcgcacgaaaaacgcgaaagcgtttcacgataa atgcgaaaac, the underlined nucleotides varying appropriately. The "target" fragment used for figure 4b was: gccggtatctttccagcactgggccggtatctttccagcactggcg/
cgccagtgctggaaagataccggcccagtgctggaaagataccggc

Dideoxy fragments: The dideoxy fragments were synthesized one nucleotide short on the 3' ends of the "transferred" strand. After annealing to a full-length compliment, the duplex was treated with klenow (exo$^-$) and either ddATP or ddTTP. The DNA was boiled and cooled to room temperature, to inactivate klenow and to reanneal the fragments. In control experiments, phenol extraction of the klenow reaction did not improve target cleavage efficiency (data not shown). In most of the experiments described, the transposon fragments were not radiolabeled. However, we radiolabeled ddA fragments from three independent preparations to analyze on a sequencing gel. Among the three preparations, 63%, 65% and 67% of the fragment had been successfully converted to ddA (as judged by length). Therefore, for the strand transfer reaction shown in figure 3c we used only 65% precleaved fragment and 35% no-A fragment, mimicking the conditions in reactions with ddA fragments.

The "non-transferred" fragment strand contained a 4-nucleotide 5' overhang, but the klenow treatment caused it to be shortened by 2 nts. Any fragment that we compared to a dideoxy fragment was treated with klenow, in the absence of NTPs, which caused the same shortening. The transposition reactions themselves were supplemented with dideoxy nucleotide to maintain the chemical composition across an experiment (final concentration of 1.2 μM nucleotide).

Transposition reactions. Reactions were performed as described (Goldhaber-Gordon et al., 2002), except that target concentrations were 5 nM, fragment concentrations were 50 nM, and MuA concentrations were 50 nM. The target DNA was ϕX174 RFI, except for figure 2b for which it was pUC19. Reactions shown in figure 2c contained 0.2 μg/μl tRNA, to inhibit nucleases that contaminated some of the mutant protein preps. In control experiments with wild-type MuA, the presence of tRNA did not affect target cleavage efficiency (data not shown). Reactions were analyzed on 0.7% HGT agarose gels (FMC bioproducts) in 1x TBE buffer, which were stained with VistraGreen and visualized with a Molecular Dynamics Flourimager

595. For both strand transfer and target cleavage experiments, a single band that migrated close to linear target was counted as product.

Ligation reaction. Transposition reactions were performed for three hours in a 200 µl volume. Reactions were stopped with SDS and EDTA, and the DNA was purified on a Qiagen PCR Purification column. The DNA products were then treated with Calf Intestinal Phosphatase for 1 hour, again purified on a Qiagen column, and treated with T4 Polynucleotide Kinase and ATPγ^{32}-P for one hour. The products were phenol extracted and purified on an Amersham G50 spin column. Ligation buffer was added directly to the DNA, along with 400 units T4 DNA ligase, and the reactions were incubated for 20 minutes at room temperature. Reactions were stopped with SDS and EDTA and analyzed by gel electrophoresis and autoradiography.

Transpososome analyses. To analyze the number of transpososomes present in a reaction, fragments were 5' radiolabeled on the transferred strand. Transpososomes were formed as described in the figure legend, and loaded directly onto a 2% MetaPhor agarose gel in 0.5x TBE buffer. The gel and the running buffer each contained 1µg/ml heparin. This method was used to generate figure 4b, and also to test the activity of MuA mutants used in figure 2c.

Figure legends.

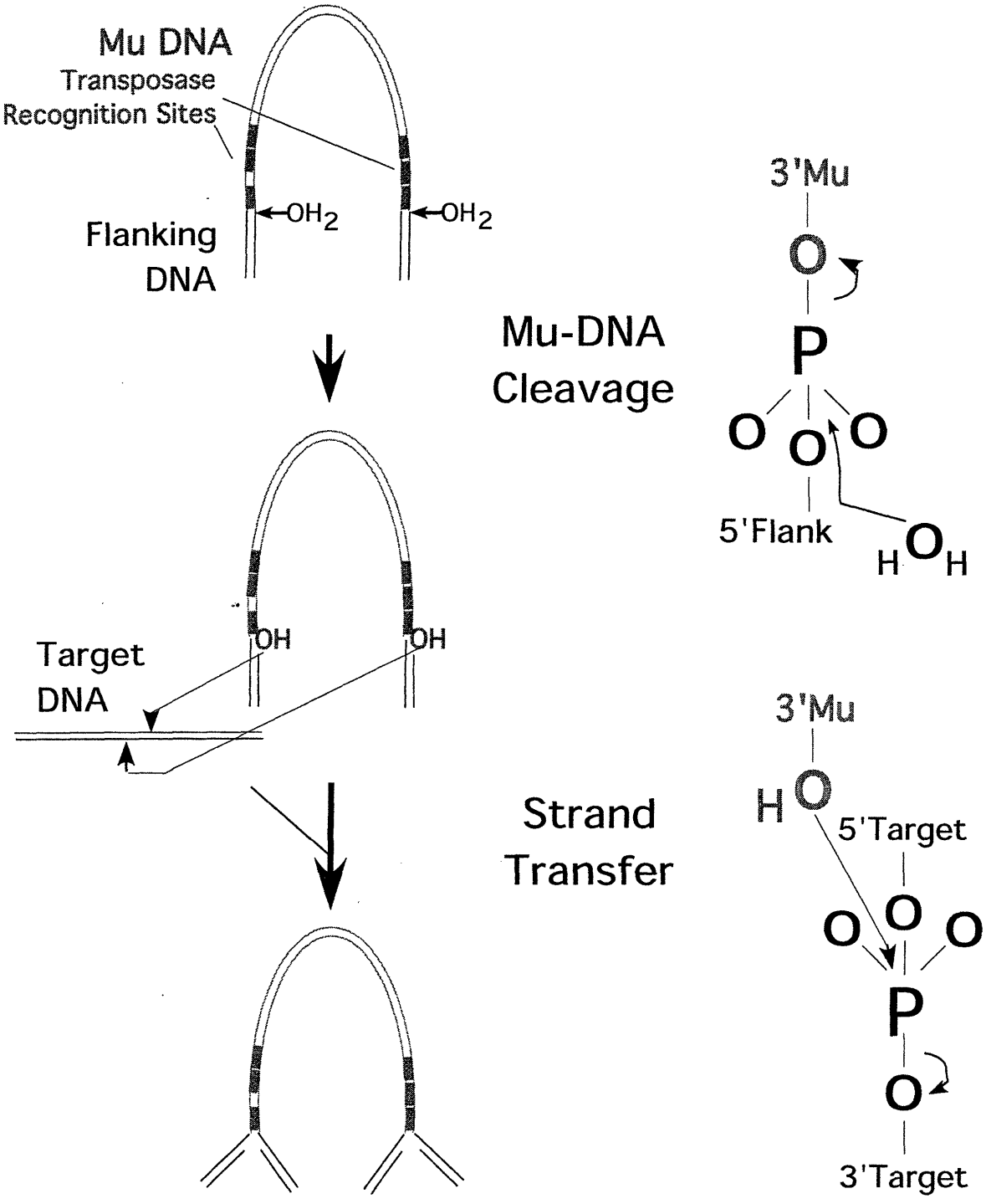
Figure 1. *Schematics of transposition* (all figures not drawn to scale). MuA transposase performs two distinct but related reactions that involve the Mu DNA: Mu-DNA cleavage and strand transfer. On the left is a representation of the DNA during these reactions; on the right is a representation of the phosphates and nucleophiles that directly participate. Grey – Mu DNA; Black – other DNA.

Figure 2. ddA transposon fragments induce MuA transposase to cleave a target DNA. A. *Schematics of fragments used in this study.* DNA fragments were constructed using the sequence of the last 50 bps from the right end of the Mu genome. Grey boxes represent MuA recognition sites (specifically R1 and R2 sites). The pentagon represents an abasic nucleotide. B. *The linear product of target cleavage is easily visualized with a DNA stain.* Transposase, transposon fragments and pUC19 target DNA were incubated for 20 hours in transposition buffer. Lane 1 contains linear target, generated by restriction digest, as a size marker. ST fragments produced a strand transfer product, which electrophoreses close to linear target. ddA fragments produced a product that migrates precisely with linear target. C. *Target cleavage requires the MuA active site.* ddA fragments and ϕX174 RFI target DNA were incubated with wild-type or mutant versions of transposase for 3 hours (shown here) or 20 hours (data not shown). In either case, no linear product was visible with the mutants. We used twice the standard transposase concentrations in order to maximize the likelihood of seeing product. The difference in mobility here compared to figure 2C is because it is a different target DNA.

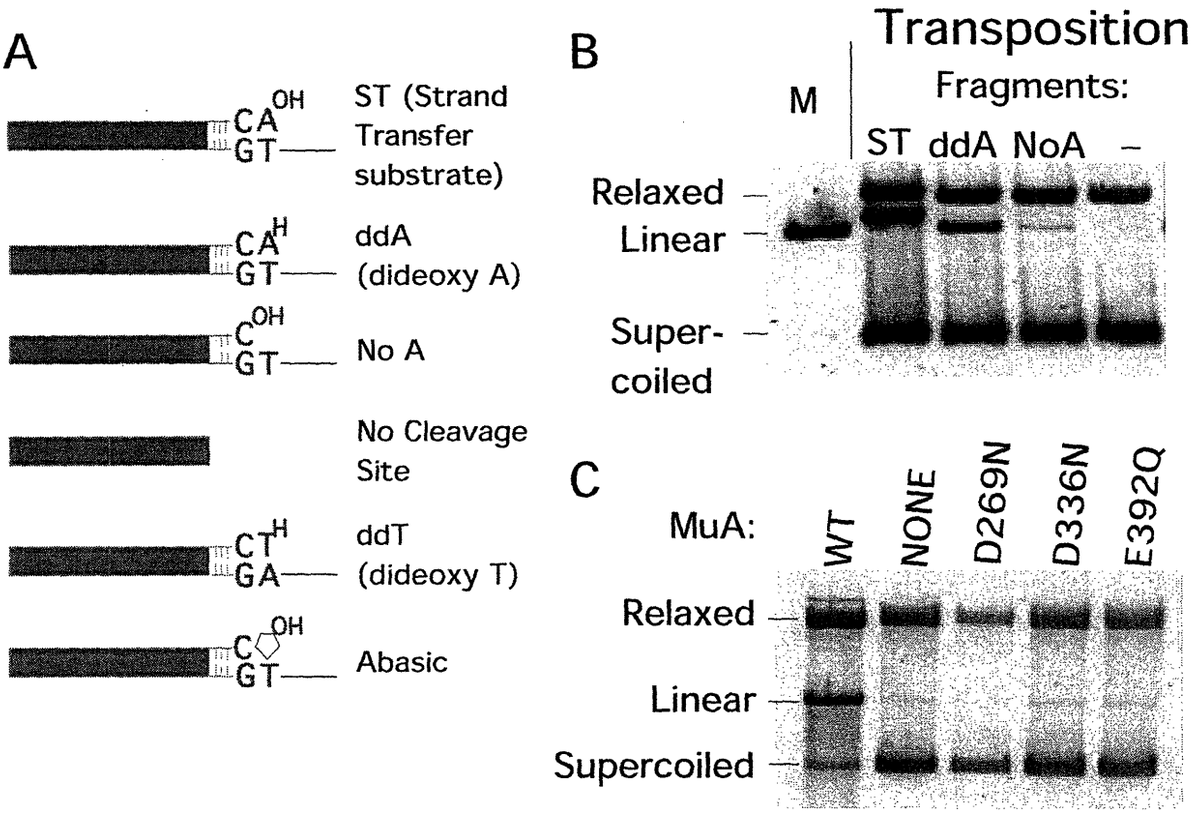
Figure 3. Strand transfer competes with target cleavage. A. *Schematic illustrating the two types of products.* Strand transfer occurs into sites 5 bps apart, on opposite strands of the target DNA. If target cleavage uses the same mechanism, the product will contain complementary 5' overhangs. B. *Strand transfer substrates do not support target cleavage.* Because the target cleavage products contain 5' overhangs and the strand transfer products do not, target cleavage products are more readily ligatable. Only ddA fragments produced ligatable product; ST fragments did not. C. *Strand transfer is faster than target cleavage.* Transpososomes were preassembled in the absence of target DNA and magnesium, and time points were taken starting with the addition of these two components.

Figure 4. The terminal nucleotide of the transposon activates transposase. A. *The DNA cleavage site, and especially the terminal nucleotide, are important features of a target*

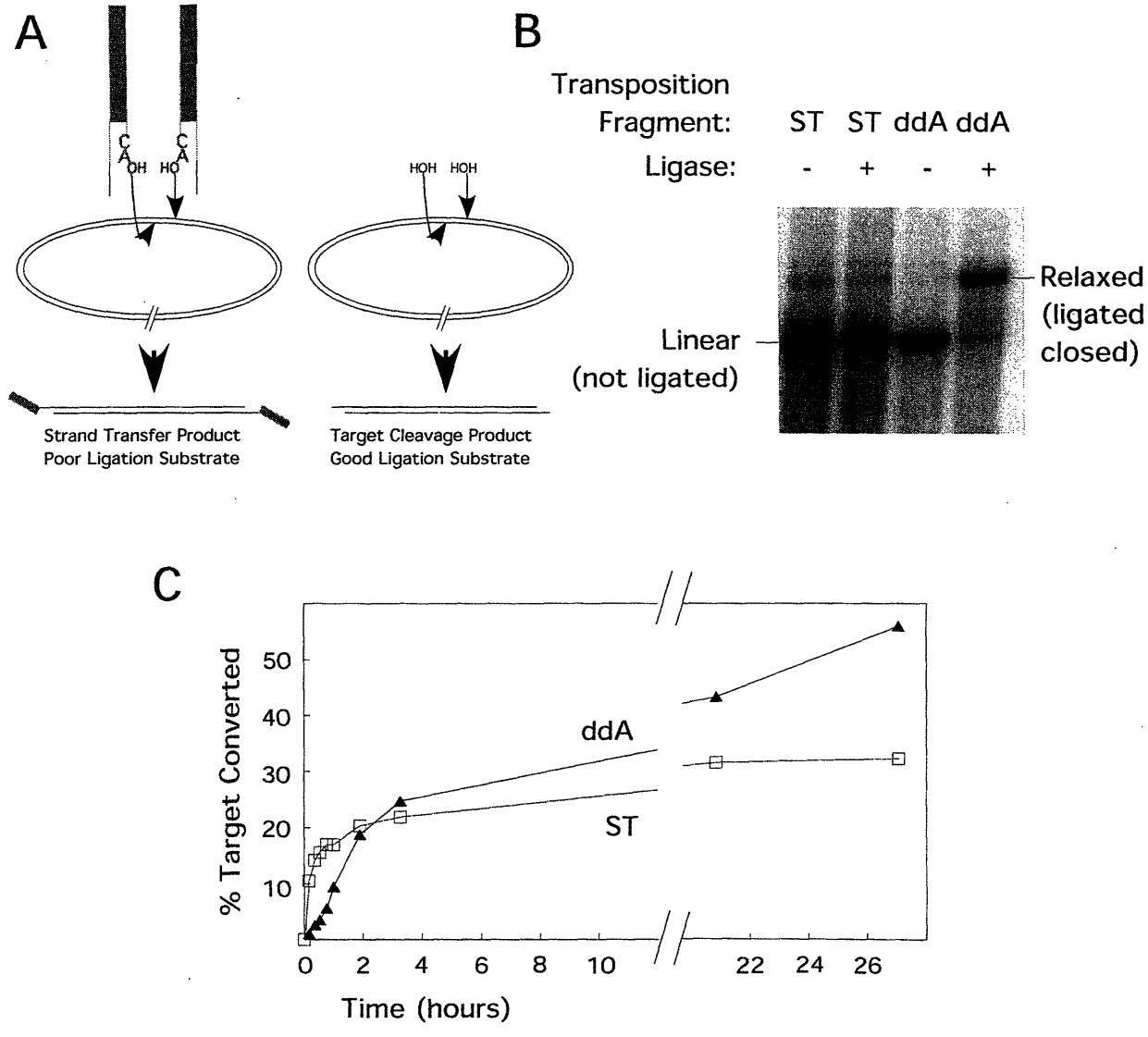
cleavage-activating fragment. However, the terminal nucleotide does not need to be an adenosine. Transpososomes were assembled overnight, in the absence of target DNA (magnesium was included). The next day, an additional 50 nM MuA was added, along with target DNA, to initiate the time-course. This resulted in an intial rate for the ddA reaction that was ~20x the rate of the No-A reaction (shown here). Failure to include the additional MuA resulted in an initial rate of the ddA reaction that was only ~4x that of the no-A reaction, because the ddA rate was enhanced and the no-A rate was not. Independent of this effect, reaction rates with the ddT fragment were more variable than with other fragment used in this study, presumably due to its deficiency during transpososome assembly. As the average of five independent experiments, the rate of the ddT reaction was not significantly different from the rate of the ddA reaction. (Aside note: on this long time scale both the ddA fragment and the no-A fragment participated in some strand-transfer, but the target cleavage product was at least 20 fold more abundant than the strand-transfer product. In the case of the ddA fragment, this minimal strand transfer activity is probably due to contaminating deoxy nucleotide in the dideoxy stock used to create the fragment. The ddT fragment was not tested for strand transfer, and the no-cleavage-site fragment showed no strand transfer activity.) B. *The no-A fragment forms stable transpososomes, despite its failure to support vigorous target cleavage.* Transpososomes were assembled overnight, under conditions identical to those used for figure 4a, except that the fragments were radiolabeled. The next day, along with the additional MuA, a short "target" fragment (sequence unrelated to Mu DNA) was added to the same base-pair concentration as the target DNA used for figure 4a. Shown here: the presence of transpososomes was confirmed by agarose gel prior to addition of the target fragment, and again at 1 hour and ~6 hours after addition of the fragment. The gels contained heparin to remove unstable protein-DNA complexes. In the absence of heparin more complexes were visible, but there was still no distinction between complexes with ddA fragments and those with no-A fragments (data not shown). C. *The adenine base of the terminal nucleotide is not a critical feature for strand transfer.* Transpososomes were preassembled as for figure 3c. The "abasic fragment" terminates on the transferred strand with a deoxy nucleotide that does not contain a base.



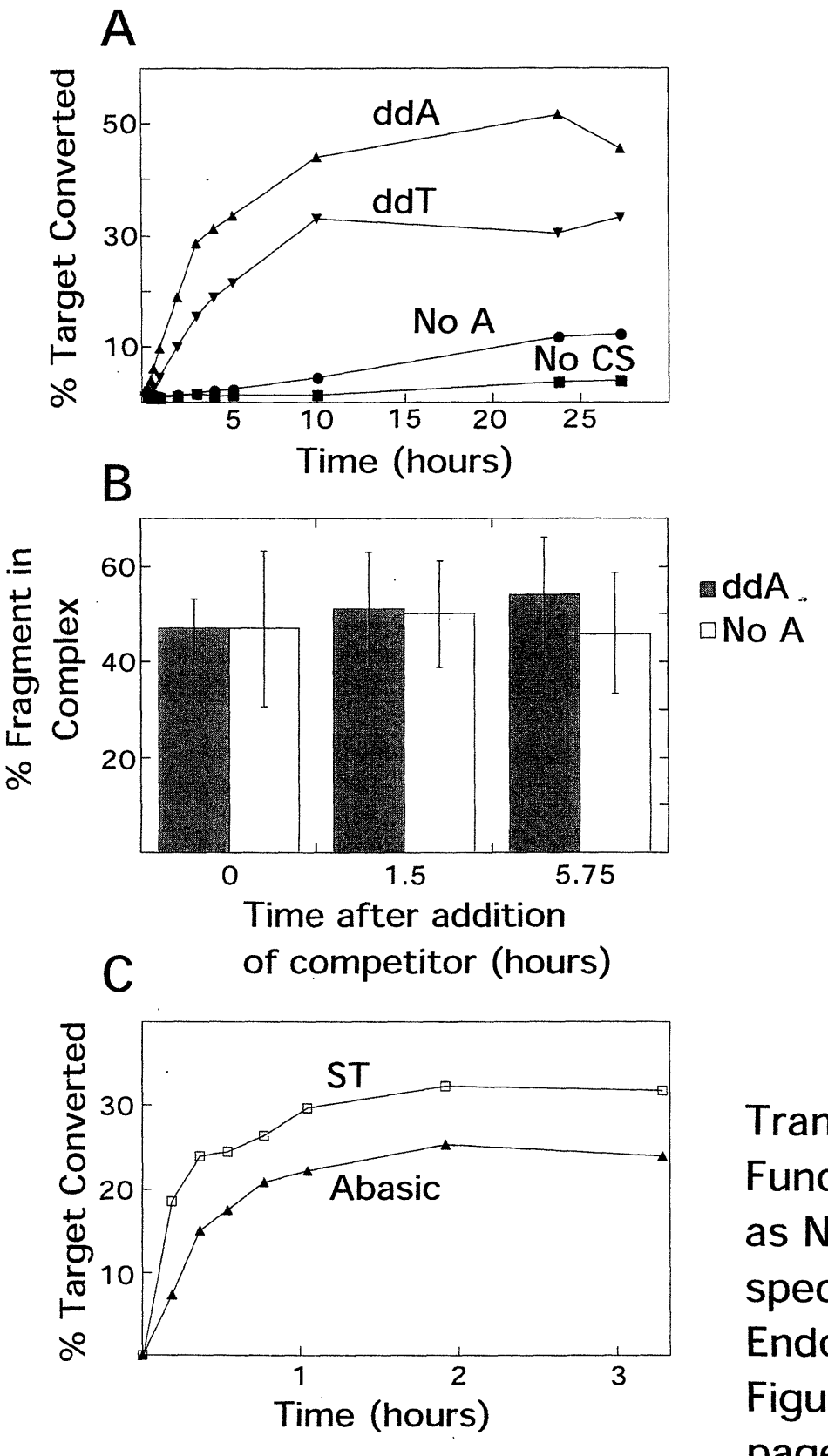
Transposase Function as Non-specific Endonuclease
Figure 1
page 99



Transpoase Function as Non-specific Endonuclease
Figure 2
page 100



Transposase Function as Non-specific Endonuclease
Figure 3
page 101



Transposase Function as Non-specific Endonuclease
Figure 4
page 102

References

Adzuma, K. and Mizuuchi, K. (1988) Target immunity of Mu transposition reflects a differential distribution of Mu B protein. *Cell*, **53**, 257-266.

Adzuma, K. and Mizuuchi, K. (1991) Steady-state kinetic analysis of ATP hydrolysis by the B protein of bacteriophage mu. Involvement of protein oligomerization in the ATPase cycle. *J Biol Chem*, **266**, 6159-6167.

Agrawal, A., Eastman, Q.M. and Schatz, D.G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*, **394**, 744-751.

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

Allet, B. (1979) Mu insertion duplicates a 5 base pair sequence at the host inserted site. *Cell*, **16**, 123-129.

Allison, R.G. and Chaconas, G. (1992) Role of the A protein-binding sites in the in vitro transposition of mu DNA. A complex circuit of interactions involving the mu ends and the transpositional enhancer. *J Biol Chem*, **267**, 19963-19970.

Arthur, A., Nimmo, E., Hettle, S. and Sherratt, D. (1984) Transposition and transposition immunity of transposon Tn3 derivatives having different ends. *Embo J*, **3**, 1723-1729.

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

Avila, P., de la Cruz, F., Ward, E. and Grinsted, J. (1984) Plasmids containing one inverted repeat of Tn21 can fuse with other plasmids in the presence of Tn21 transposase. *Mol Gen Genet*, **195**, 288-293.

Azaro, M.A.a.L.A. (2002) λ Integrase and the λ Integrase Family. In Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A. M. (ed.), *Mobile DNA II*. ASM Press, Washington D. C., pp. 119-148.

Baker, T.A., Kremenstova, E. and Luo, L. (1994) Complete transposition requires four active monomers in the mu transposase tetramer. *Genes Dev*, **8**, 2416-2428.

Baker, T.A. and Luo, L. (1994) Identification of residues in the Mu transposase essential for catalysis. *Proc Natl Acad Sci U S A*, **91**, 6654-6658.

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

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

Baker, T.A., Mizuuchi, M., Savilahti, H. and Mizuuchi, K. (1993) Division of labor among monomers within the Mu transposase tetramer. *Cell*, **74**, 723-733.

Bhasin, A., Goryshin, I.Y., Steiniger-White, M., York, D. and Reznikoff, W.S. (2000) Characterization of a Tn5 pre-cleavage synaptic complex. *J Mol Biol*, **302**, 49-63.

Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, **95**, 10570-10575.

Bolland, S. and Kleckner, N. (1996) The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell*, **84**, 223-233.

Bowerman, B., Brown, P.O., Bishop, J.M. and Varmus, H.E. (1989) A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev*, **3**, 469-478.

Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrake, M., Katz, R.A. and Skalka, A.M. (1997) Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J Biol Chem*, **272**, 18161-18168.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol*, **253**, 333-346.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R.A. and Skalka, A.M. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure*, **4**, 89-96.

Burge, C.B., Padgett, R.A. and Sharp, P.A. (1998) Evolutionary fates and origins of U12-type introns. *Mol Cell*, **2**, 773-785.

Burlingame, R.P., Obukowicz, M.G., Lynn, D.L. and Howe, M.M. (1986) Isolation of point mutations in bacteriophage Mu attachment regions cloned in a lambda::mini-Mu phage. *Proc Natl Acad Sci U S A*, **83**, 6012-6016.

Burton, B.M., Williams, T.L. and Baker, T.A. (2001) ClpX-mediated remodeling of mu transpososomes: selective unfolding of subunits destabilizes the entire complex. *Mol Cell*, **8**, 449-454.

Bushman, F.D. and Craigie, R. (1991) Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc Natl Acad Sci U S A*, **88**, 1339-1343.

Carteau, S., Gorelick, R.J. and Bushman, F.D. (1999) Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J Virol*, **73**, 6670-6679.

Chaconas, G. (1999) Studies on a "jumping gene machine": higher-order nucleoprotein complexes in Mu DNA transposition. *Biochem Cell Biol*, **77**, 487-491.

Chaconas, G., Gloor, G., Miller, J.L., Kennedy, D.L., Giddens, E.B. and Nagainis, C.R. (1984) Transposition of bacteriophage mu DNA: expression of the A and B proteins from lambda pL and analysis of infecting mu DNA. *Cold Spring Harb Symp Quant Biol*, **49**, 279-284.

Chandler, M.a.M., J. (2002) Insertion Sequences Revisited. In Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A. M. (ed.), *Mobile DNA II*. ASM Press, Washington D. C., pp. 305-366.

Chen, H. and Engelman, A. (2001) Asymmetric processing of human immunodeficiency virus type 1 cDNA in vivo: implications for functional end coupling during the chemical steps of DNA transposition. *Mol Cell Biol*, **21**, 6758-6767.

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

Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., Emini, E.A., Hazuda, D. and Kuo, L.C. (2000b) X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50-293)--an initial glance of the viral DNA binding platform. *J Mol Biol*, **296**, 521-533.

Chow, S.A., Vincent, K.A., Ellison, V. and Brown, P.O. (1992) Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science*, **255**, 723-726.

Clubb, R.T., Omichinski, J.G., Savilahti, H., Mizuuchi, K., Gronenborn, A.M. and Clore, G.M. (1994) A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase. *Structure*, **2**, 1041-1048.

Clubb, R.T., Schumacher, S., Mizuuchi, K., Gronenborn, A.M. and Clore, G.M. (1997) Solution structure of the I gamma subdomain of the Mu end DNA-binding domain of phage Mu transposase. *J Mol Biol*, **273**, 19-25.

Coelho, A., Maynard-Smith, S. and Symonds, N. (1982) Abnormal cointegrate structures mediated by gene B mutants of phage Mu: their implications with regard to gene function. *Mol Gen Genet*, **185**, 356-362.

Coros, C.J. and Chaconas, G. (2001) Effect of mutations in the Mu-host junction region on transpososome assembly. *J Mol Biol*, **310**, 299-309.

Craig, N.L. (1996) Transposon Tn7. *Curr Top Microbiol Immunol*, **204**, 27-48.

Craig, N.L. (2002) Mobile DNA: an Introduction. In Craig, N.L., Craigie, R., Gellert, M., Lambowitz, A. M. (ed.), *Mobile DNA II*. ASM Press, Washington D. C., pp. 3-11.

Craigie, R., Arndt-Jovin, D.J. and Mizuuchi, K. (1985) A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. *Proc Natl Acad Sci U S A*, **82**, 7570-7574.

Craigie, R. and Mizuuchi, K. (1986) Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. *Cell*, **45**, 793-800.

Craigie, R. and Mizuuchi, K. (1987) Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell*, **51**, 493-501.

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

Davies, D.R., Braam, L.M., Reznikoff, W.S. and Rayment, I. (1999) The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-A resolution. *J Biol Chem*, **274**, 11904-11913.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S. and Rayment, I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*, **289**, 77-85.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981-1986.

Ellison, V., Gerton, J., Vincent, K.A. and Brown, P.O. (1995) An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. *J Biol Chem*, **270**, 3320-3326.

Embleton, M.L., Williams, S.A., Watson, M.A. and Halford, S.E. (1999) Specificity from the synapsis of DNA elements by the Sfi I endonuclease. *J Mol Biol*, **289**, 785-797.

Engelman, A., Mizuuchi, K. and Craigie, R. (1991) HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*, **67**, 1211-1221.

Engler, L.E., Welch, K.K. and Jen-Jacobson, L. (1997) Specific binding by EcoRV endonuclease to its DNA recognition site GATATC. *J Mol Biol*, **269**, 82-101.

Esposito, D. and Craigie, R. (1998) Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *Embo J*, **17**, 5832-5843.

Finkel, S.E. and Johnson, R.C. (1993) The Fis protein: it's not just for DNA inversion anymore. *Mol Microbiol*, **7**, 1023.

Finnegan, D.J. (1997) Transposable elements: how non-LTR retrotransposons do it. *Curr Biol*, **7**, R245-248.

Goldgur, Y., Dyda, F., Hickman, A.B., Jenkins, T.M., Craigie, R. and Davies, D.R. (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *Proc Natl Acad Sci U S A*, **95**, 9150-9154.

Goldhaber-Gordon, I., Early, M.H., Gray, M.K. and Baker, T.A. (2002a) Sequence and positional requirements for DNA sites in a mu transpososome. *J Biol Chem*, **277**, 7703-7712.

Goldhaber-Gordon, I., Williams, T.L. and Baker, T.A. (2002b) DNA recognition sites activate MuA transposase to perform transposition of non-Mu DNA. *J Biol Chem*, **277**, 7694-7702.

Gottesman, S., Clark, W.P., de Crecy-Lagard, V. and Maurizi, M.R. (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of Escherichia coli. Sequence and in vivo activities. *J Biol Chem*, **268**, 22618-22626.

Greene, E.C. and Mizuuchi, K. (2002) Dynamics of a protein polymer: the assembly and disassembly pathways of the MuB transposition target complex. *Embo J*, **21**, 1477-1486.

Groenen, M.A., Kokke, M. and van de Putte, P. (1986) Transposition of mini-Mu containing only one of the ends of bacteriophage Mu. *Embo J*, **5**, 3687-3690.

Groenen, M.A., Timmers, E. and van de Putte, P. (1985) DNA sequences at the ends of the genome of bacteriophage Mu essential for transposition. *Proc Natl Acad Sci U S A*, **82**, 2087-2091.

Groenen, M.A. and van de Putte, P. (1986) Analysis of the ends of bacteriophage Mu using site-directed mutagenesis. *J Mol Biol*, **189**, 597-602.

Guo, F., Gopaul, D.N. and van Duyne, G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40-46.

Haapa-Paananen, S., Rita, H. and Savilahti, H. (2002) DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. *J Biol Chem*, **277**, 2843-2851.

Hallet, B. and Sherratt, D.J. (1997) Transposition and site-specific recombination: adapting DNA cut-and- paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev*, **21**, 157-178.

Haniford, D. and Kleckner, N. (1994) Tn 10 transposition in vivo: temporal separation of cleavages at the two transposon ends and roles of terminal basepairs subsequent to interaction of ends. *Embo J*, **13**, 3401-3411.

Harada, R., Dufort, D., Denis-Larose, C. and Nepveu, A. (1994) Conserved cut repeats in the human cut homeodomain protein function as DNA binding domains. *J Biol Chem*, **269**, 2062-2067.

Haren, L., Ton-Hoang, B. and Chandler, M. (1999) Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol*, **53**, 245-281.

Herr, W. and Cleary, M.A. (1995) The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev*, **9**, 1679-1693.

Hickman, A.B., Li, Y., Mathew, S.V., May, E.W., Craig, N.L. and Dyda, F. (2000) Unexpected structural diversity in DNA recombination: the restriction endonuclease connection. *Mol Cell*, **5**, 1025-1034.

Hiom, K., Melek, M. and Gellert, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell*, **94**, 463-470.

Huang, B., Schaeffer, C.J., Li, Q. and Tsai, M.D. (1996) Splase: a new class IIS zinc-finger restriction endonuclease with specificity for Sp1 binding sites. *J Protein Chem*, **15**, 481-489.

Huang, H., Chopra, R., Verdine, G.L. and Harrison, S.C. (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **282**, 1669-1675.

Jacobson, E.M., Li, P., Leon-del-Rio, A., Rosenfeld, M.G. and Aggarwal, A.K. (1997) Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev*, **11**, 198-212.

Jen-Jacobson, L. (1997) Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers*, **44**, 153-180.

Jiang, H. and Harshey, R.M. (2001) The Mu enhancer is functionally asymmetric both in cis and in trans. Topological selectivity of Mu transposition is enhancer-independent. *J Biol Chem*, **276**, 4373-4381.

Jiang, H., Yang, J.Y. and Harshey, R.M. (1999) Criss-crossed interactions between the enhancer and the att sites of phage Mu during DNA transposition. *Embo J*, **18**, 3845-3855.

Jones, J.M., Welty, D.J. and Nakai, H. (1998) Versatile action of Escherichia coli ClpXP as protease or molecular chaperone for bacteriophage Mu transposition. *J Biol Chem*, **273**, 459-465.

Junop, M.S. and Haniford, D.B. (1996) Multiple roles for divalent metal ions in DNA transposition: distinct stages of Tn10 transposition have different Mg2+ requirements. *Embo J*, **15**, 2547-2555.

Kahmann, R. and Kamp, D. (1979) Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. *Nature*, **280**, 247-250.

Katzman, M. and Sudol, M. (1996) Nonspecific alcoholysis, a novel endonuclease activity of human immunodeficiency virus type 1 and other retroviral integrases. *J Virol*, **70**, 2598-2604.

Kennedy, A.K., Guhathakurta, A., Kleckner, N. and Haniford, D.B. (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell*, **95**, 125-134.

Kennedy, A.K., Haniford, D.B. and Mizuuchi, K. (2000) Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. *Cell*, **101**, 295-305.

Kikuchi, Y. and Nash, H.A. (1979) Nicking-closing activity associated with bacteriophage lambda int gene product. *Proc Natl Acad Sci U S A*, **76**, 3760-3764.

Kim, K., Namgoong, S.Y., Jayaram, M. and Harshey, R.M. (1995) Step-arrest mutants of phage Mu transposase. Implications in DNA- protein assembly, Mu end cleavage, and strand transfer. *J Biol Chem*, **270**, 1472-1479.

Kim, S.C., Podhajska, A.J. and Szybalski, W. (1988) Cleaving DNA at any predetermined site with adapter-primers and class-IIS restriction enzymes. *Science*, **240**, 504-506.

Kim, Y.C., Grable, J.C., Love, R., Greene, P.J. and Rosenberg, J.M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307-1309.

Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A*, **93**, 1156-1160.

Kim, Y.G. and Chandrasegaran, S. (1994) Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A*, **91**, 883-887.

Kim, Y.G., Shi, Y., Berg, J.M. and Chandrasegaran, S. (1997) Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene*, **203**, 43-49.

Kim, Y.G., Smith, J., Durgesha, M. and Chandrasegaran, S. (1998) Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol Chem*, **379**, 489-495.

Kleckner, N., Chalmers, R. M., Kwon, D., Sakai, J., Bollard, S. (1996) Tn10. In Saedler, H., Gierl, A. (ed.), *Transposable Elements*. Springer-Verlag, New York, pp. 49-82.

Kleckner, N., Reichardt, K. and Botstein, D. (1979) Inversions and deletions of the Salmonella chromosome generated by the translocatable tetracycline resistance element Tn10. *J Mol Biol*, **127**, 89-115.

Klemm, J.D. and Pabo, C.O. (1996) Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev*, **10**, 27-36.

Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell*, **77**, 21-32.

Krementsova, E., Giffin, M.J., Pincus, D. and Baker, T.A. (1998) Mutational analysis of the Mu transposase. Contributions of two distinct regions of domain II to recombination. *J Biol Chem*, **273**, 31358-31365.

Kruklitis, R. and Nakai, H. (1994) Participation of the bacteriophage Mu A protein and host factors in the initiation of Mu DNA synthesis in vitro. *J Biol Chem*, **269**, 16469-16477.

Kruklitis, R., Welty, D.J. and Nakai, H. (1996) ClpX protein of Escherichia coli activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. *Embo J*, **15**, 935-944.

Kukolj, G. and Skalka, A.M. (1995) Enhanced and coordinated processing of synapsed viral DNA ends by retroviral integrases in vitro. *Genes Dev*, **9**, 2556-2567.

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

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

Lannoy, V.J., Burglin, T.R., Rousseau, G.G. and Lemaigre, F.P. (1998) Isoforms of hepatocyte nuclear factor-6 differ in DNA-binding properties, contain a bifunctional homeodomain, and define the new ONECUT class of homeodomain proteins. *J Biol Chem*, **273**, 13552-13562.

Lavoie, B.D. and Chaconas, G. (1993) Site-specific HU binding in the Mu transpososome: conversion of a sequence-independent DNA-binding protein into a chemical nuclease. *Genes Dev*, **7**, 2510-2519.

Lavoie, B.D. and Chaconas, G. (1994) A second high affinity HU binding site in the phage Mu transpososome. *J Biol Chem*, **269**, 15571-15576.

Lavoie, B.D., Chan, B.S., Allison, R.G. and Chaconas, G. (1991) 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. *Embo J*, **10**, 3051-3059.

Lavoie, B.D., Shaw, G.S., Millner, A. and Chaconas, G. (1996) Anatomy of a flexer-DNA complex inside a higher-order transposition intermediate. *Cell*, **85**, 761-771.

Lee, I. and Harshey, R.M. (2001) Importance of the conserved CA dinucleotide at Mu termini. *J Mol Biol*, **314**, 433-444.

Lee, M.S. and Craigie, R. (1998) A previously unidentified host protein protects retroviral DNA from autointegration. *Proc Natl Acad Sci U S A*, **95**, 1528-1533.

Lefstin, J.A. and Yamamoto, K.R. (1998) Allosteric effects of DNA on transcriptional regulators. *Nature*, **392**, 885-888.

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

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

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

Lewis, S.M., Agard, E., Suh, S. and Czyzyk, L. (1997) Cryptic signals and the fidelity of V(D)J joining. *Mol Cell Biol*, **17**, 3125-3136.

Li, P., He, X., Gerrero, M.R., Mok, M., Aggarwal, A. and Rosenfeld, M.G. (1993) Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. *Genes Dev*, **7**, 2483-2496.

Lovell, S., Goryshin, I.Y., Reznikoff, W.R. and Rayment, I. (2002) Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat Struct Biol*, **9**, 278-281.

Manna, D. and Higgins, N.P. (1999) Phage Mu transposition immunity reflects supercoil domain structure of the chromosome. *Mol Microbiol*, **32**, 595-606.

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

Maxwell, A. and Halford, S.E. (1982) The SalGI restriction endonuclease. Enzyme specificity. *Biochem J*, **203**, 93-98.

McCord, M., Chiu, R., Vora, A.C. and Grandgenett, D.P. (1999) Retrovirus DNA termini bound by integrase communicate in trans for full-site integration in vitro. *Virology*, **259**, 392-401.

Messer, W., Blaesing, F., Jakimowicz, D., Krause, M., Majka, J., Nardmann, J., Schaper, S., Seitz, H., Speck, C., Weigel, C., Wegrzyn, G., Welzeck, M. and Zakrzewska-Czerwinska, J. (2001) Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. *Biochimie*, **83**, 5-12.

Mhammedi-Alaoui, A., Pato, M., Gama, M.J. and Toussaint, A. (1994) A new component of bacteriophage Mu replicative transposition machinery: the Escherichia coli ClpX protein. *Mol Microbiol*, **11**, 1109-1116.

Mizuuchi, K. (1992) Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu Rev Biochem*, **61**, 1011-1051.

Mizuuchi, K. and Adzuma, K. (1991) 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.

Mizuuchi, M., Baker, T.A. and Mizuuchi, K. (1991) DNase protection analysis of the stable synaptic complexes involved in Mu transposition. *Proc Natl Acad Sci U S A*, **88**, 9031-9035.

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

Mizuuchi, M., Baker, T.A. and Mizuuchi, K. (1995) Assembly of phage Mu transpososomes: cooperative transitions assisted by protein and DNA scaffolds. *Cell*, **83**, 375-385.

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

Mizuuchi, M. and Mizuuchi, K. (1993) Target site selection in transposition of phage Mu. *Cold Spring Harb Symp Quant Biol*, **58**, 515-523.

Mizuuchi, M. and Mizuuchi, K. (2001) Conformational isomerization in phage Mu transpososome assembly: effects of the transpositional enhancer and of MuB. *Embo J*, **20**, 6927-6935.

Motsch, S. and Schmitt, R. (1984) Replicon fusion mediated by a single-ended derivative of transposon Tn1721. *Mol Gen Genet*, **195**, 281-287.

Murphy, J.E. and Goff, S.P. (1992) A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase in vivo. *J Virol*, **66**, 5092-5095.

Naigamwalla, D.Z., Coros, C.J., Wu, Z. and Chaconas, G. (1998) Mutations in domain III alpha of the Mu transposase: evidence suggesting an active site component which interacts with the Mu-host junction. *J Mol Biol*, **282**, 265-274.

Nakai, H. and Kruklitis, R. (1995) Disassembly of the bacteriophage Mu transposase for the initiation of Mu DNA replication. *J Biol Chem*, **270**, 19591-19598.

Nakayama, C., Teplow, D.B. and Harshey, R.M. (1987) Structural domains in phage Mu transposase: identification of the site- specific DNA-binding domain. *Proc Natl Acad Sci USA*, **84**, 1809-1813.

Namgoong, S.Y. and Harshey, R.M. (1998) The same two monomers within a MuA tetramer provide the DDE domains for the strand cleavage and strand transfer steps of transposition. *Embo J*, **17**, 3775-3785.

Namgoong, S.Y., Jayaram, M., Kim, K. and Harshey, R.M. (1994) DNA-protein cooperativity in the assembly and stabilization of mu strand transfer complex. Relevance of DNA phasing and att site cleavage. *J Mol Biol*, **238**, 514-527.

Namgoong, S.Y., Kim, K., Saxena, P., Yang, J.Y., Jayaram, M., Giedroc, D.P. and Harshey, R.M. (1998) Mutational analysis of domain II beta of bacteriophage Mu transposase: domains II alpha and II beta belong to different catalytic complementation groups. *J Mol Biol*, **275**, 221-232.

Naumann, T.A. and Reznikoff, W.S. (2000) Trans catalysis in Tn5 transposition. *Proc Natl Acad Sci USA*, **97**, 8944-8949.

Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656-663.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, **79**, 639-648.

Packer, M.J., Dauncey, M.P. and Hunter, C.A. (2000) Sequence-dependent DNA structure: dinucleotide conformational maps. *J Mol Biol*, **295**, 71-83.

Patel, P.H., Suzuki, M., Adman, E., Shinkai, A. and Loeb, L.A. (2001) Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J Mol Biol*, **308**, 823-837.

Pato, M. (1989) Bacteriophage Mu. In Berg, D.E.a.H., M. M. (ed.), *Mobile DNA*. ASM Press, Washington D. C., pp. 23-52.

Pato, M.L. and Banerjee, M. (1996) The Mu strong gyrase-binding site promotes efficient synapsis of the prophage termini. *Mol Microbiol*, **22**, 283-292.
Pato, M.L. and Reich, C. (1984) Stoichiometric use of the transposase of bacteriophage Mu. *Cell*, **36**, 197-202.
Pingoud, A. and Jeltsch, A. (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur J Biochem*, **246**, 1-22.
Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature*, **386**, 569-577.
Rice, P., Craigie, R. and Davies, D.R. (1996) Retroviral integrases and their cousins. *Curr Opin Struct Biol*, **6**, 76-83.
Rice, P. and Mizuuchi, K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell*, **82**, 209-220.
Rice, P.A. and Baker, T.A. (2001) Comparative architecture of transposase and integrase complexes. *Nat Struct Biol*, **8**, 302-307.
Robinson, M.K., Bennett, P.M. and Richmond, M.H. (1977) Inhibition of TnA translocation by TnA. *J Bacteriol*, **129**, 407-414.
Sakai, J., Chalmers, R.M. and Kleckner, N. (1995) Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn10 transposition. *Embo J*, **14**, 4374-4383.
Sarnovsky, R.J., May, E.W. and Craig, N.L. (1996) The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *Embo J*, **15**, 6348-6361.
Savilahti, H. and Mizuuchi, K. (1996) Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. *Cell*, **85**, 271-280.
Savilahti, H., Rice, P.A. and Mizuuchi, K. (1995) The phage Mu transpososome core: DNA requirements for assembly and function. *Embo J*, **14**, 4893-4903.
Schneider, F., Schwikardi, M., Muskhelishvili, G. and Droge, P. (2000) A DNA-binding domain swap converts the invertase gin into a resolvase. *J Mol Biol*, **295**, 767-775.
Schumacher, S., Clubb, R.T., Cai, M., Mizuuchi, K., Clore, G.M. and Gronenborn, A.M. (1997) Solution structure of the Mu end DNA-binding ibeta subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. *Embo J*, **16**, 7532-7541.
Scully, K.M., Jacobson, E.M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D.W., Hooshmand, F., Aggarwal, A.K. and Rosenfeld, M.G. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science*, **290**, 1127-1131.
Shen, M.M., Raleigh, E.A. and Kleckner, N. (1987) Physical analysis of Tn10- and IS10-promoted transpositions and rearrangements. *Genetics*, **116**, 359-369.
Sixma, T.K. (2001) DNA mismatch repair: MutS structures bound to mismatches. *Curr Opin Struct Biol*, **11**, 47-52.
Skinner, L.M., Sudol, M., Harper, A.L. and Katzman, M. (2001) Nucleophile selection for the endonuclease activities of human, ovine, and avian retroviral integrases. *J Biol Chem*, **276**, 114-124.
Steitz, T.A. and Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A*, **90**, 6498-6502.
Stellwagen, A.E. and Craig, N.L. (1997) Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *Embo J*, **16**, 6823-6834.

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

Surette, M.G. and Chaconas, G. (1989) A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction is Escherichia coli integration host factor. *J Biol Chem*, **264**, 3028-3034.

Surette, M.G. and Chaconas, G. (1991) Stimulation of the Mu DNA strand cleavage and intramolecular strand transfer reactions by the Mu B protein is independent of stable binding of the Mu B protein to DNA. *J Biol Chem*, **266**, 17306-17313.

Surette, M.G. and Chaconas, G. (1992) The Mu transpositional enhancer can function in trans: requirement of the enhancer for synapsis but not strand cleavage. *Cell*, **68**, 1101-1108.

Surette, M.G., Harkness, T. and Chaconas, G. (1991) Stimulation of the Mu A protein-mediated strand cleavage reaction by the Mu B protein, and the requirement of DNA nicking for stable type 1 transpososome formation. In vitro transposition characteristics of mini- Mu plasmids carrying terminal base pair mutations. *J Biol Chem*, **266**, 3118-3124.

Surette, M.G., Lavoie, B.D. and Chaconas, G. (1989) 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). *Embo J*, **8**, 3483-3489.

van Gent, D.C., Vink, C., Groeneger, A.A. and Plasterk, R.H. (1993) Complementation between HIV integrase proteins mutated in different domains. *Embo J*, **12**, 3261-3267.

van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H. and Sixma, T.K. (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of C.elegans in complex with transposon DNA. *Embo J*, **16**, 6044-6054.

Vink, C. and Plasterk, R.H. (1993) The human immunodeficiency virus integrase protein. *Trends Genet*, **9**, 433-438.

Vipond, I.B. and Halford, S.E. (1993) Structure-function correlation for the EcoRV restriction enzyme: from non-specific binding to specific DNA cleavage. *Mol Microbiol*, **9**, 225-231.

Voloshin, O.N., Wang, L. and Camerini-Otero, R.D. (1996) Homologous DNA pairing promoted by a 20-amino acid peptide derived from RecA. *Science*, **272**, 868-872.

Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97-100.

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

Wei, S.Q., Mizuuchi, K. and Craigie, R. (1998) Footprints on the viral DNA ends in moloney murine leukemia virus preintegration complexes reflect a specific association with integrase. *Proc Natl Acad Sci U S A*, **95**, 10535-10540.

Williams, S.A. and Halford, S.E. (2001) SfiI endonuclease activity is strongly influenced by the non-specific sequence in the middle of its recognition site. *Nucleic Acids Res*, **29**, 1476-1483.

Williams, T.L., Jackson, E.L., Carritte, A. and Baker, T.A. (1999) Organization and dynamics of the Mu transpososome: recombination by communication between two active sites. *Genes Dev*, **13**, 2725-2737.

Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. and Wilson, K.S. (1993) The crystal structure of EcoRV

endonuclease and of its complexes with cognate and non-cognate DNA fragments. *Embo J*, **12**, 1781-1795.

Wishart, W.L., Broach, J.R. and Ohtsubo, E. (1985) ATP-dependent specific binding of Tn3 transposase to Tn3 inverted repeats. *Nature*, **314**, 556-558.

Wojtkowiak, D., Georgopoulos, C. and Zylicz, M. (1993) Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of Escherichia coli. *J Biol Chem*, **268**, 22609-22617.

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

Wu, Z. and Chaconas, G. (1997) The Mu transposase tetramer is inactive in unassisted strand transfer: an auto-allosteric effect of Mu A promotes the reaction in the absence of Mu B. *J Mol Biol*, **267**, 132-141.

Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999) Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev*, **13**, 1263-1275.

Yamauchi, M. and Baker, T.A. (1998) An ATP-ADP switch in MuB controls progression of the Mu transposition pathway. *Embo J*, **17**, 5509-5518.

Yang, J.Y., Kim, K., Jayaram, M. and Harshey, R.M. (1995) A domain sharing model for active site assembly within the Mu A tetramer during transposition: the enhancer may specify domain contributions. *Embo J*, **14**, 2374-2384.

Yang, W. and Mizuuchi, K. (1997) Site-specific recombination in plane view. *Structure*, **5**, 1401-1406.

Yang, W. and Steitz, T.A. (1995) Recombining the structures of HIV integrase, RuvC and RNase H. *Structure*, **3**, 131-134.

Yang, Z.N., Mueser, T.C., Bushman, F.D. and Hyde, C.C. (2000) Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J Mol Biol*, **296**, 535-548.

Zou, A.H., Leung, P.C. and Harshey, R.M. (1991) Transposase contacts with mu DNA ends. *J Biol Chem*, **266**, 20476-20482.