

**THE MECHANISMS, APPLICATIONS, AND TARGET SITE
SELECTION OF BACTERIOPHAGE MU MINIMAL *IN*
VITRO DNA TRANSPOSITION REACTION**

by
Saija Haapa-Paananen

Institute of Biotechnology and
Department of Biosciences, Division of Genetics,
Helsinki Graduate School in Biotechnology and Molecular Biology,
Faculty of Science, University of Helsinki

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in auditorium 2 at the Viikki Infocenter (Viikinkaari 11, Helsinki) on 25 January 2002, at 12 o'clock noon.

Helsinki 2002

Supervisor

Docent Harri Savilahti,
Institute of Biotechnology,
University of Helsinki

Reviewers

Prof. Dennis Bamford
Department of Biosciences,
and Institute of Biotechnology
University of Helsinki

and

Prof. Alan H. Schulman
MTT Agrifood Research Finland, Jokioinen,
and Institute of Biotechnology,
University of Helsinki

Opponent

Michael Chandler, Director
Laboratoire de Microbiologie et Génétique Moléculaire du CNRS,
France

ISBN 952-10-0258-1
ISBN 952-10-0259-X (PDF, <http://ethesis.helsinki.fi/>)
ISSN 1239-9469
Yliopistopaino Oy
Helsinki 2002

TABLE OF CONTENTS

| | |
|---|----|
| ORIGINAL PUBLICATIONS | 5 |
| ABBREVIATIONS | 6 |
| A. ABSTRACT | 7 |
| B. INTRODUCTION | 8 |
| 1. TRANSPOSITION IS A WIDESPREAD PHENOMENON | 8 |
| 2. TRANSPOSABLE ELEMENTS CAN BE DIVIDED INTO TWO MAJOR GROUPS | 8 |
| 2.1. Some elements transpose via an RNA intermediate | 8 |
| 2.1.1. Retroviruses | 9 |
| 2.1.2. Retrotransposons | 9 |
| 2.2. DNA transposons move via a DNA intermediate | 9 |
| 3. MECHANISMS OF DNA TRANSPOSITION | 10 |
| 3.1. Unity in transposition mechanisms | 10 |
| 3.2. Similarity in the transposase proteins | 11 |
| 3.3. Retroviral integration | 12 |
| 3.4. Cut and paste transposition | 12 |
| 3.4.1. Tn7 | 12 |
| 3.4.2. Tn10 (IS10) | 13 |
| 3.5. Replicative transposition of Bacteriophage Mu | 13 |
| 3.5.1. DNA sequences involved in the Mu transposition | 14 |
| 3.5.2. The proteins involved in Mu transposition | 15 |
| 3.5.3. Three-site synapsis (LER) | 16 |
| 3.5.4. Stable synaptic complex formation | 17 |
| 3.5.5. Donor cleavage | 17 |
| 3.5.6. Strand transfer | 18 |
| 3.5.7. Processing of the strand transfer complex | 18 |
| 3.5.8. Division of labour within the tetramer of MuA | 19 |
| 3.5.9. Relaxed requirements in vitro | 19 |
| 4. TARGET SITE SELECTION IN TRANSPOSITION | 20 |
| 4.1. Several protein factors can act in the target site selection | 20 |
| 4.2. Regional target site selection | 22 |
| 4.3. A consensus sequence can often be found | 22 |
| 4.4. Certain target DNA structures can also be preferred | 22 |
| 4.5. Bacteriophage Mu target site selection | 22 |
| 5. TRANSPOSONS AS MOLECULAR TOOLS | 23 |
| 5.1. Why transposon tools? | 23 |
| 5.2. Different ways to use transposon tools | 24 |
| 5.2.1. In vivo tools | 24 |
| 5.2.2. Simplified in vitro tools | 25 |
| C. AIMS OF THE PRESENT STUDY | 27 |
| D. MATERIALS AND METHODS | 28 |

| | |
|---|----|
| E. RESULTS AND DISCUSSION | 29 |
| 1. THE MINIMAL IN VITRO TRANSPOSITION REACTION RECAPITULATES THE HALLMARKS OF IN VIVO TRANSPOSITION | 29 |
| 1.1. One-ended versus two-ended integration (I) | 29 |
| 1.2. Efficiency (I, II) | 30 |
| 1.3. Accuracy (I, II) | 30 |
| 1.4. Target site selection (I, II) | 30 |
| 2. APPLICATIONS OF THE MINIMAL MU IN VITRO TRANSPOSITION METHODOLOGY | 31 |
| 2.1. Genetic footprinting (I) | 31 |
| 2.2. Sequencing (II) | 31 |
| 3. TARGET SITE SELECTION (III) | 32 |
| 3.1. A new quantitative nucleotide-level assay | 32 |
| 3.2. The nucleotide consensus | 33 |
| 3.3. Dinucleotide preferences | 34 |
| 3.4. An apparent symmetry in the target site selection | 34 |
| 3.5. Insertion site preferences can be predicted | 35 |
| 3.6. The effect of the sequences surrounding the target pentamer core | 35 |
| 3.7. The structural properties of the target DNA | 35 |
| 3.8. The extent of the "surrounding sequence effect" | 36 |
| 3.9. A general picture of the target site selection | 37 |
| F. CONCLUSIONS AND FUTURE PERSPECTIVES | 38 |
| G. ACKNOWLEDGEMENTS | 39 |
| H. REFERENCES | 40 |

ORIGINAL PUBLICATIONS

This thesis is based on the following articles, which are referred to in the text by their Roman numerals.

- I Haapa S., Taira S., Heikkinen E., and Savilahti H. (1999) An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res.*, **27**, 2777-2784. A published erratum appears in *Nucleic Acids Res.*, **27**, no. 15. The printing error in Figure 1A has been corrected for the electronic version.
- II Haapa S., Suomalainen S., Eerikäinen S., Airaksinen M., Paulin L., and Savilahti H. (1999) An efficient DNA sequencing strategy based on the bacteriophage Mu *in vitro* DNA transposition reaction. *Genome Res.*, **9**, 308-315.
- III Haapa-Paananen S., Rita H., and Savilahti H. DNA transposition of bacteriophage Mu: a quantitative analysis of target site selection *in vitro*. *J. Biol. Chem.* In press (JBC Papers in Press. Published on November 7, 2001 as Manuscript M108044200).

ABBREVIATIONS

| | |
|------|--|
| A | adenine |
| ATP | adenosine triphosphate |
| bp | base pair(s) |
| C | cytosine |
| CDC | cleaved donor complex |
| cDNA | complementary deoxyribonucleic acid |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| ds | double-stranded |
| G | guanine |
| HPLC | high-pressure liquid chromatography |
| HU | accessory protein (hydroxyurea) |
| IHF | integration host factor |
| IR | inverted repeat |
| IS | insertion sequence |
| kb | kilobase pairs |
| LTR | long terminal repeat |
| MRF | Mu replication factors |
| MuA | bacteriophage Mu transposase protein A |
| MuB | bacteriophage Mu protein B |
| nt | nucleotide(s) |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| Pu | purine(s) |
| Py | pyrimidine(s) |
| R | purines: adenine or guanine |
| RE | restriction enzyme |
| RNA | ribonucleic acid |
| ss | single stranded |
| SSC | stable synaptic complex |
| STC | strand transfer complex |
| T | thymidine |
| TCC | target capture complex |
| Tn | transposon |
| Y | pyrimidines: cytosine or thymidine |

A. ABSTRACT

Transposons are mobile genetic elements that are able to translocate from one site to another within the genomes of their host organisms by a special form of DNA recombination called transposition. Transposons have been utilised as essential tools in genetics for many years. However, many of the current transposon-based methods suffer from limitations such as: (a) transposition is performed *in vivo* as a multistep process with limitations in the host ranges; (b) accuracy and/or efficiency of the transposition reaction is not optimal; (c) a limited set of target sites is used.

Bacteriophage Mu uses DNA transposition to replicate its genome and is one of the best-characterised mobile genetic elements. Mu was also the first mobile element for which a cell-free *in vitro* transposition reaction was established. The *in vitro* transposition reaction of bacteriophage Mu has been further simplified, and under certain reaction conditions only three macromolecular components are required: MuA transposase, short Mu right-end sequences and target DNA. We describe here a new Mu transposon tool, which is based on a minimal one-step *in vitro* transposition reaction and circumvents the above limitations. This Mu transposon-based tool works efficiently and accurately, and displays low target site selectivity with some preferences. Since the reaction is performed *in vitro*, host-range limitations are avoided. Firstly, the utility of this system in functional genetic analysis is demonstrated using restriction analysis and PCR-based genetic footprinting strategies. Secondly, we describe an efficient and simple DNA sequencing

strategy based on the minimal bacteriophage Mu *in vitro* transposition reaction. This proof of concept has led to a commercial product for DNA sequencing.

The features behind the preferences in the integration site selection are also further characterised. The Mu transpositional DNA recombination machinery selects target sites by assembling a protein-DNA complex that interacts with the target DNA and reacts whenever it locates a favourable sequence composition. Preferential usage of different target pentamers was studied with a minimal Mu *in vitro* system, and quantitatively compiled consensus sequences for the most preferred and the least preferred sites were generated. The favoured and disfavoured sites were also analysed as base pair steps, which dictate the sequence-related DNA structure. This study demonstrated for the first time the impact of the surrounding sequences on the target site selection by MuA. Symmetrical patterns in several DNA structural parameters spanning a ~23-24-bp region around the target pentamer were detected. As a result, a new, more structurally oriented perspective on the target site selection is obtained. Additionally, this study provides a new means to predict insertion site preferences.

B. INTRODUCTION

1. TRANSPOSITION IS A WIDESPREAD PHENOMENON

Transposable elements are discrete segments of genetic material that are capable of moving from one locus to another in their host genome or even between different genomes. The phenomenon was first realised by Barbara McClintock in the 1940s, when she discovered the mobile “controlling elements” in maize (McClintock, 1987). Because of this ability to translocate between genomic locations, transposable elements (i.e. transposons) are also referred to as mobile elements or “jumping genes”. Generally, transposons move by a special form of recombination called transposition, which does not require sequence homology in the target locus. As a by-product of their movement, transposable elements cause insertional mutations and mediate various types of other genome rearrangements, such as deletions and inversions. Transposable elements play an important role in evolution by creating a source of genetic variation and by introducing new genetic material into host genomes. Before their discovery, genomes were regarded as stable, but the growing knowledge of different mobile elements has been a major factor in changing the thinking towards genomic fluidity (reviewed by Berg and Howe, 1989; Leach, 1996; Saedler and Gierl, 1996; Sherratt, 1995).

Transposable elements are ubiquitous residents of all organisms studied in detail. Different mobile elements have been found in prokaryotes as well as in eukaryotes. For instance, over 50 % of the maize genome and 45 % of the human genome are comprised of trans-

posable elements (review by Hurst and Werren, 2001). Transposition reactions are utilised in a variety of different processes, such as in the spreading of antibiotic resistance-coding genes in bacteria, and in the life cycles of certain bacteriophages (e.g. Mu). Also, the integration of retroviruses, such as HIV, and the intracellular movement of retroviral-like elements follow transposition reaction chemistry. Recently, a similar reaction mechanism was discovered in V(D)J recombination, which is behind the rearrangement of gene segments during the maturation of B and T cell lymphocytes in mammals. V(D)J recombination generates the diversity of the immune system, which is needed to defend against foreign antigens invading our bodies. Because of the similarities in reaction mechanisms, it has been suggested that V(D)J recombination is of an ancestral transpositional origin (Agrawal et al., 1998; Hiom et al., 1998).

2. TRANSPOSABLE ELEMENTS CAN BE DIVIDED INTO TWO MAJOR GROUPS

Transposable elements have adopted different strategies for their movement from one genomic site to another. Two major groups can be distinguished according to the genetic material utilised as an intermediate in the reactions.

2.1. Some elements transpose via an RNA intermediate

Some transposable elements exist in one stage of their life cycle as an RNA intermediate. They are often referred to as retroelements, as distinct from elements that utilise solely DNA intermediates (Figure 1). Retroelements include retrotransposons and several

other groups, but retroviruses also use similar strategies in their life cycles.

2.1.1. Retroviruses

Common to retroviruses is that they require a reverse transcriptase to transcribe the RNA intermediate into a cDNA copy. Infectious retrovirus particles contain RNA, which is reverse-transcribed after infection into a double-stranded DNA copy. The dsDNA is processed and joined to the host genome by a retroviral enzyme, the integrase (reviewed by Hindmarsh and Leis, 1999). New infectious retroviruses can then be produced from this integrated provirus by the synthesis of new RNA copies and the packaging of these RNAs into nucleocapsid particles prior to the release of mature retroviruses (reviewed by Coffin et al., 1997; Polard and Chandler, 1995).

2.1.2. Retrotransposons

Retrotransposons (reviewed by Boeke and Stoye, 1997) are closely related to retroviruses in their many conserved genes and similar life cycles. The major difference generally is that although retrotransposons make nucleoprotein capsids, these do not mature into infectious viruses and thus do not exit the host cell. The retrotransposon life cycle is described in Figure 1. Retrotransposons can further be subcategorized by the presence or lack of long terminal repeats (LTRs) at their terminal ends. The LTR-retrotransposons code for retrovirus-like proteins, but lack an envelope protein that is needed for the exit from the cell. A diverse group of elements, which are not flanked by long terminal repeats, are called non-LTR retrotransposons. Some retrotransposons have accumulated mutations or lack necessary proteins, preventing

RETROTRANSPOSITION

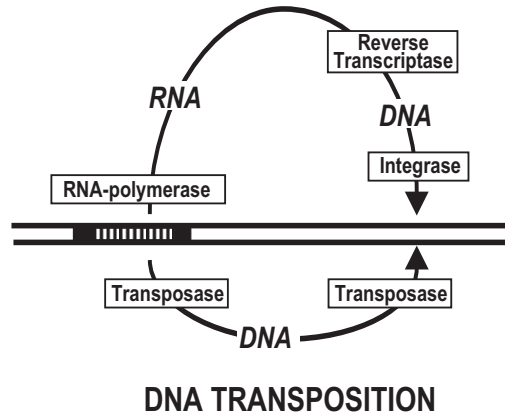


Figure 1. Transposable elements can be divided into two major groups according to the transposition intermediates used. The transposable element is shown as a hatched box within a larger DNA-fragment (horizontal double line). The enzymes involved in the reactions are enclosed in boxes. Retrotransposition, as shown above, involves transcription of the integrated DNA copy of the element into RNA by the host RNA polymerase. This is followed by reverse transcription of the RNA into a double-stranded cDNA copy by the element-encoded reverse transcriptase (RT). The dsDNA copy is then integrated into a new target DNA location by the element-encoded integrase (IN). The retrotransposition produces new copies of the element and is therefore called replicative. DNA transposition, shown below, is catalysed by an element-encoded transposase (see text for more details). DNA transposition can be either replicative or non-replicative (Adapted from Polard and Chandler, 1995).

them from transposing unless suitable proteins are provided *in trans*.

2.2. DNA transposons move via a DNA intermediate

A diverse group of transposable elements relies solely on DNA interme-

diates without an RNA phase. These are generally called DNA transposons and they vary in size and complexity from small simple insertion sequences (ISs) to longer composite transposons made of ISs, and further to more complex bacteriophages. Several DNA transposable elements, as well as RNA elements, are known to occur in high copy numbers in plants (reviewed by Fedoroff, 2000; Flavell et al., 1994). The diversity of DNA transposable elements is reflected in the over 500 insertion sequences found in bacteria (Mahillon and Chandler, 1998), and the number is constantly growing due to ongoing sequencing projects. Different DNA transposons can be classified by similarities in their structure, transposition reaction mechanism, and nucleotide or protein sequence conservation. However, little is still known about the details of the reaction mechanisms of many elements.

Generally, DNA transposons have distinct and unique sequences at the ends of the element. These sequences are recognised by an element-encoded recombinase protein, the transposase. The transposase needs to recognise the element ends accurately to avoid shortening and thus destruction of the element. Variable numbers of transposases, accessory proteins and antibiotic resistance genes, among others, can be encoded between the ends. Similarly to retrotransposons, some DNA transposons have lost necessary functions (i.e. defective elements) and are able to transpose only if those functions are provided *in trans*. Host proteins may also be needed for effective transposition.

3. MECHANISMS OF DNA TRANSPOSITION

3.1. Unity in transposition mechanisms

DNA transposons can be defined as pieces of DNA residing in the genome, which are capable of translocating themselves into new genomic locations. Although the reaction mechanisms of different DNA transposons seemed complex and varied when discovered, common themes have emerged during the past few years (Craig, 1995). DNA transposons as well as retroviruses share two critical chemical reaction steps, namely, phosphodiester bond cleavage at the ends of the element (donor DNA) and joining of the element ends into a new target DNA site (see Figure 2).

These two critical reaction steps, the donor cleavage and strand transfer, are catalysed by one or more transposase proteins that bind in a sequence-specific manner to the transposon ends and assemble in a protein-DNA complex called a transpososome (for reviews see Haniford and Chaconas, 1992; Mizuuchi, 1992; Craig, 1995; Mizuuchi, 1997). Firstly, a pair of site-specific endonucleolytic phosphodiester bond cleavages is made at the flanking host-transposon DNA boundary, exposing the donor 3'-hydroxyl (OH) ends (donor cleavage step). Secondly, the donor 3'-OH ends are joined in a pair of DNA strand transfer reactions to the target DNA 5'-ends, which are made by a concerted staggered cut (strand transfer step). These two steps can be considered as a special form of site-specific recombination, resulting in the formation of the transposition intermediate, the strand transfer product.

These two steps are referred to here as the transposition reaction, although also a third step takes place in host cells where the strand transfer product is repaired and /or replicated by the host proteins (see Figure 2 and more details below). This repair step results in duplication of the target site. The length of the target site duplication depends on the staggered cut made during strand transfer.

Depending on the element, either one or both strands of donor DNA can be cut by the element-encoded transposase(s), or integrase in the case of retroviruses. Two different outcomes can follow, depending on the cleavage step: either a simple insertion or a cointegrate formation (see Figure 2).

3.2. Similarity in the transposase proteins

In addition to the mechanistic parallels described above, similarities are also seen at the protein level. The transposase proteins and retroviral integrases share structurally related catalytic domains. In particular, a three acidic residue DDE-motif, which is essential for catalytic activity, is conserved (reviewed by Haren et al., 1999). The growing number of solved crystal structures of various integrases and transposases have provided a detailed picture of the structural homologies (reviewed by Rice and Baker, 2001; MuA core domain, Rice and Mizuuchi, 1995; ASV integrase, Bujacz et al., 1995; Tn5 transposase synaptic complex, Davies et al., 2000; HIV-1 integrase, Dyda et

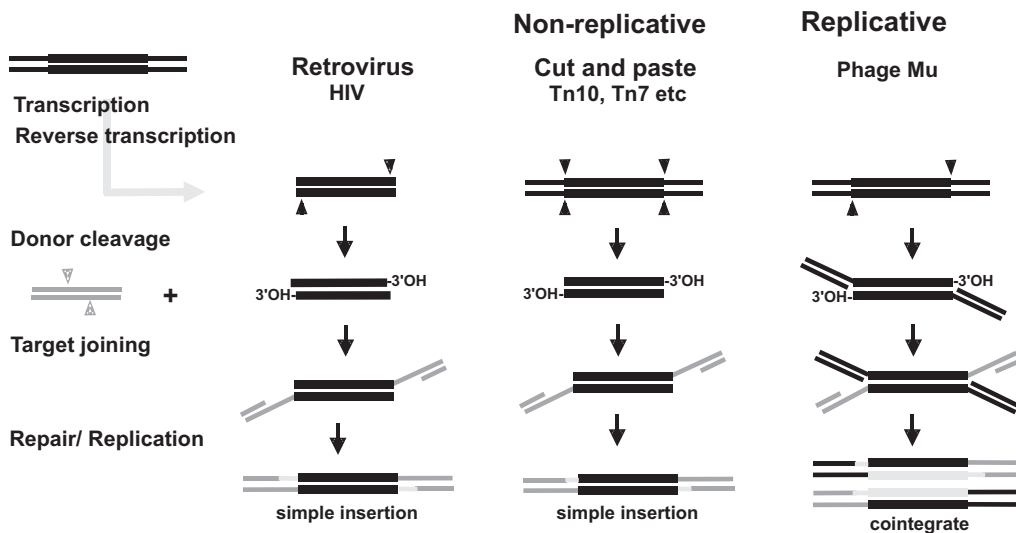


Figure 2. Unity can be seen in the transposition reaction mechanisms. The mobile element is depicted by a bold black double-line and the target DNA by a dark grey double-line. Light grey denotes areas that are repaired/replicated by host cell factors. (Adapted from Craig 1995).

al., 1994, Goldgur et al., 1998; RSV integrase, Yang et al. 2000).

3.3. Retroviral integration

During retrovirus integration (reviewed by Hindmarsh and Leis, 1999), the viral RNA genome is first converted to double-stranded DNA by the virus-encoded reverse transcriptase. A few nucleotides from each 3'-end of this dsDNA are then cleaved by the retroviral integrase and a simple insert is produced as a result of strand transfer (Figure 2). Because only one DNA strand is cleaved, a few unpaired nucleotides remain attached to the so-called non-transferred strand. Host proteins presumably remove or repair these unpaired nucleotides as well as the single stranded gaps that result from the staggered cut made during strand transfer. Because an integrated retrovirus (or provirus) is transcribed from the DNA copy to RNA and further reverse-transcribed to DNA before new integration (as illustrated in Figure 1 for retrotransposition), the number of retrovirus DNA copies increases. Therefore, retrovirus integration is always replicative.

3.4. Cut and paste transposition

Some transposons are excised totally from the surrounding DNA and thus can freely insert into new target DNA. These elements are jointly referred as the "cut and paste" transposons. The strand transfer product of the cut and paste – type reaction is a simple insert with single stranded gaps around the element (Figure 2). These gaps are then repaired by host proteins. The cut and paste transposition mechanism results in the movement of an element from one place to another without preserving the element in its original location. Thus, cut and paste transposition is also

referred to as non-replicative transposition. However, non-replicative transposons can also become replicated indirectly if the host cell replication machinery replicates through the genomic region, thus copying the transposon before excision.

Different transposons have evolved varying strategies for cutting both DNA strands (reviewed by Turlan and Chandler, 2000). Transposition mechanisms of Tn7 and Tn10 are briefly considered here as examples.

3.4.1. Tn7

Tn7 is a relatively large (14-kb) bacterial transposon (reviewed by Craig 1991, 1996). It is exceptional in that it has separate transposases for cutting the 3'- and 5'-ends (Sarnovsky et al., 1996). The Tn7 encodes five separate transposase genes: *TnsA*, *B*, *C*, *D* and *E* (Waddell and Craig, 1988). The TnsA protein mediates DNA cleavage reactions at the 5'-ends of Tn7, while TnsB cuts the 3'-ends and catalyses strand transfer using the newly made 3'-OH ends as nucleophiles (May and Craig 1996, Sarnovsky et al., 1996). TnsB is a member of the retroviral integrase superfamily, with the common DDE-motif (Sarnovsky et al., 1996). Rather surprisingly, structural studies revealed that the TnsA resembles type II restriction endonucleases (Hickman et al., 2000). Minimally, TnsA together with TnsB can mediate DNA breakage and intramolecular joining alone in certain special conditions (Biery et al., 2000a). Normally TnsC, an ATP-dependent regulatory protein, is needed for transposition, as it activates the TnsAB transposase and mediates contacts to the target selecting proteins TnsD and TnsE (Bainton et al., 1993; Stellwagen and Craig, 1997a; 1998; 2001; Lu and

Craig, 2000; see the target site selection chapter below).

3.4.2. *Tn10* (IS10)

Tn10 is a composite tetracycline resistance-causing transposon with inverted repeats of insertion sequence IS10 at its ends (Kleckner et al. 1975). The *Tn10* possesses only one transposase protein containing the common catalytic DDE-motif (Bolland and Kleckner, 1996; Kennedy and Haniford 1996), which supposedly functions in divalent metal ion binding (Allingham et al. 1999). The *Tn10* uses non-replicative cut and paste-type transposition involving double-strand breaks at the transposon termini (Morisato and Kleckner 1984, Bender and Kleckner 1986, Benjamin and Kleckner 1992). The first strand is cut with a mechanism that seems to be conserved among different DNA transposons. After the hydrolytic cleavage of the first strand, the exposed 3'-OH can attack the second (non-transferred) strand and join the 3'-OH to the scissile phosphate to form a transposon end hairpin. The *Tn10* transposase can then open the hairpin, regenerating a 3'-OH residue (Kennedy et al. 1998, Allingham et al. 2001). These steps yield double-strand breaks, which free the mobile element from the flanking host DNA. The target DNA enters the reaction only after the transposon excision (Junop and Haniford 1997, Sakai and Kleckner 1997, Sakai et al. 2000). The catalytic DDE-motif is needed also for the target capture (Junop and Haniford 1997).

Bacterial transposon *Tn5* also uses a similar hairpin-mechanism to cut the two DNA strands (Bhasin et al., 1999). Still another variation of the theme is shown by IS911, which excises through a circle-form intermediate (Polard and

Chandler, 1995; Ton-Hoang et al., 1997, 1998, 1999).

3.5. Replicative transposition of Bacteriophage Mu

Bacteriophage Mu is a temperate phage, which was discovered in *Escherichia coli* by Larry Taylor (1963) because of its ability to cause mutations. Hence it was given the name Mu, for mutator. Bacteriophage Mu is exceptional in that it utilises DNA transposition reaction in two ways: (1) it integrates its DNA into the host genome non-replicatively following infection, thus producing simple insertions during lysogenisation (Harshey, 1984; Chaconas et al. 1983); (2) it replicates itself by multiple rounds of replicative transposition during lytic growth (Chaconas et al. 1981). Little is known about the mechanism of the first integration of infecting bacteriophage Mu, which leads to a simple insertion. Simple inserts have been proposed as the result of the repair of the so-called Shapiro-intermediate (Shapiro 1979) of the replicative pathway; the 5'-ends may be nucleolytically cut and repaired, resulting in a simple insertion (Harshey 1987, Craigie and Mizuuchi 1985a, Lavoie and Chaconas 1996).

The Mu transposition has been studied for years *in vivo*, but the development of a cell-free *in vitro* system (Mizuuchi, 1983) has allowed more detailed analysis. This has made the transposition reaction mechanism of bacteriophage Mu one of the most extensively studied (reviewed by Mizuuchi 1992; Baker 1995; Chaconas et al. 1996; Craigie 1996; Lavoie and Chaconas 1996).

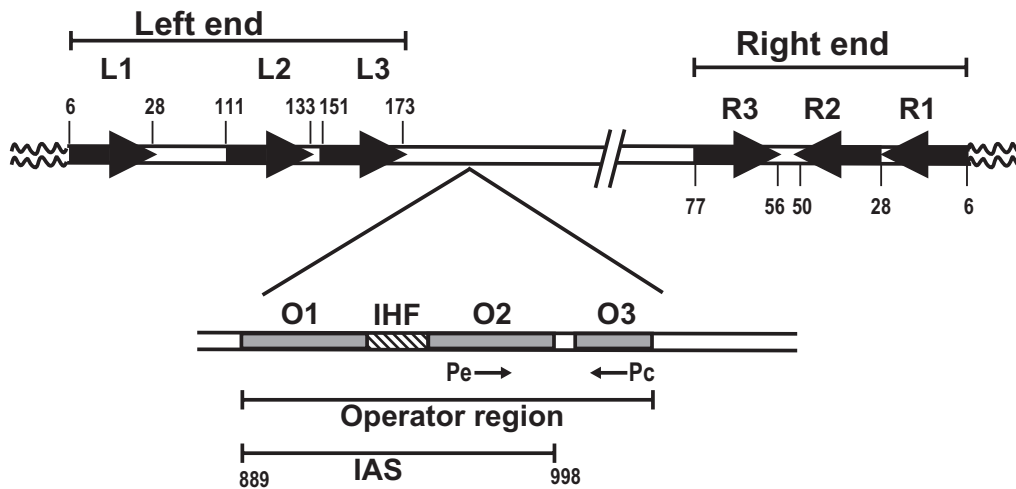


Figure 3. The organisation of the MuA protein binding sites and the internal activator sequence (IAS) in the Mu genome. The MuA-binding sites, L1-L3 and R1-R3, are shown by bold arrows. The direction of the arrow marks the orientation of the 22-bp MuA-binding consensus sequence. Numbering starts from the 5'-ends, the wavy line marks the flanking host DNA. The IAS/operator region is shown as an expanded internal fragment with the IHF binding site shown as a hatched rectangle. Two promoters are also shown; Pe directs the expression of a rightward transcript that includes the *A* and *B* genes. Pc directs the expression of the Mu repressor *c* gene. The figure is not drawn to scale. The left and right-ends are drawn according to Craigie et al., 1984; the IAS/ operator part is adapted from Baker, 1995.

3.5.1. DNA sequences involved in the Mu transposition

Bacteriophage Mu has a linear dsDNA genome of 37-Kb in length (Howe, 1987). In addition to the Mu sequences, heterogeneous host DNA is present in the phage Mu-ends due to "headfull" DNA packaging (Bukhari et al., 1976). Actual Mu DNA has three transposase binding sites at each end of the Mu genome, named L1, L2 and L3 at the left-end and R1, R2 and R3 at the right-end (Craigie et al., 1984, see Figure 3 for a description of their orientations and positions). These six sites share a 22-nucleotide consensus sequence, with no obvious internal symmetry (Craigie et al. 1984; Mizuuchi 1992). The R1 and R2 are close to each other and are ori-

ented similarly, while R3 is oriented in a different direction. The left-end sites are oriented similarly, but L1 is separated from L2 and L3 (Craigie et al., 1984). All six binding sites are not equally important for transposition. *In vitro* studies with the mini-Mu donor plasmid have shown that only R1, R2 and L1 are protected by the MuA transposase in the CDC and STC complexes (see below) in a Dnase I protection assay (Lavoie et al., 1991; Mizuuchi et al. 1991).

The terminal 5'-CA-3' nucleotides of the donor ends, especially the last adenine, are essential for *in vivo* and *in vitro* transposition reactions (Burlingame et al., 1986; Surette et al., 1991; Coros and Chaconas, 2001). Mu-

tations in the terminal nucleotides affect formation of the transposition complex, donor cleavage, and strand transfer *in vitro*; the extent depends on the type and position of the mutation (Coros and Chaconas, 2001).

In addition to the terminal transposase binding sites, an internal activator sequence (i.e. IAS, or transpositional enhancer) is also needed for efficient transposition *in vivo* and in certain conditions *in vitro* (Mizuuchi and Mizuuchi, 1989; Surette et al. 1989; Leung et al., 1989). IAS is close to the left end of the Mu genome and it contains a binding site for IHF (Figure 3). The enhancer overlaps with two transcriptional operators, which are also the binding site for a Mu phage encoded repressor. When the repressor is produced, it presumably prevents the synaptic complex assembly by blocking the enhancer, and also negatively regulates the transcription of Mu A and

B genes (Craigie et al., 1984; Mizuuchi and Mizuuchi, 1989).

3.5.2. The proteins involved in Mu transposition

Bacteriophage Mu encodes two proteins involved in transposition; the transposase MuA (Craigie and Mizuuchi, 1985b) and an accessory activator of transposition, MuB. MuA is a 663 amino acid, 75-kDa protein, which can be divided into separate domains by susceptibility to proteolytic cleavage (Figure 4). The N-terminal domain contains two sequence-specific DNA-binding domains, which are used to bind to the operator sequences in the IAS and the MuA-binding sites within the left and right ends of the Mu genome (Nakayama et al., 1987; Mizuuchi and Mizuuchi, 1989; Leung et al., 1989). The C-terminal domain IIIb is responsible for interaction with MuB (Baker et al., 1991; Leung and Harshey 1991) and it has been suggested that domain IIIa inter-

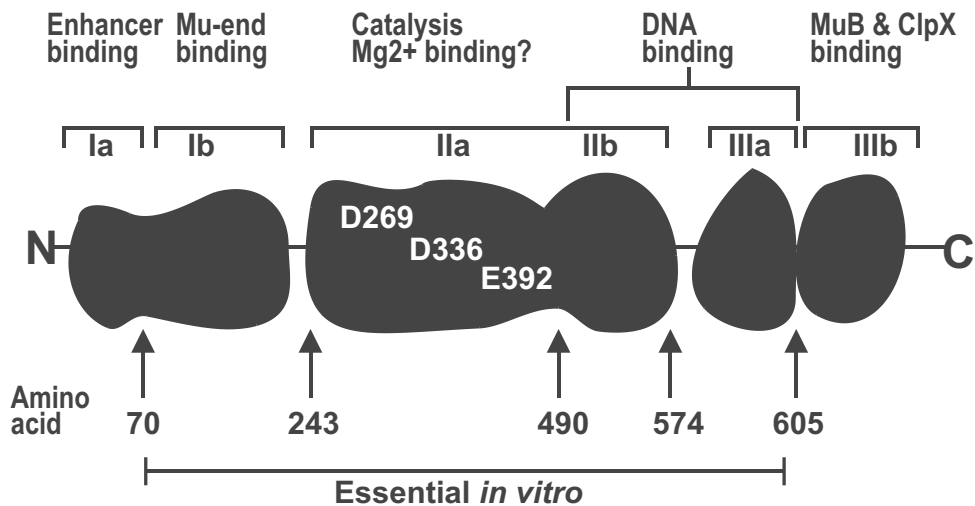


Figure 4. Illustration of the MuA structure with different functions assigned to various domains. Two protease hypersensitive sites at residue 243 and 574 divide MuA into three main domains (Adapted from Baker and Luo, 1994; Kremensova et al., 1998).

acts with the Mu-host junction (Naigamwalla et al., 1998). The central core (domain II) contains three acidic residues (D269, D336 and E392) essential for DNA cleavage and strand transfer (Baker and Luo, 1994; Kremmentsova et al., 1998). This catalytic motif is conserved at the amino acid level but especially at the structural level between the MuA protein, different transposases and retroviral integrases (Rice and Mizuuchi, 1995; Rice and Baker, 2001).

Bacteriophage Mu also encodes an activator protein, MuB (Baker et al., 1991; Surette et al., 1991). The MuB protein stimulates intermolecular strand transfer in the presence of ATP (Maxwell et al., 1987) by binding to targets without Mu-end sequences (Adzuma and Mizuuchi, 1988; Darzins et al., 1988). Phage Mu thus avoids destructive self-insertion because of differential distribution of the MuB between “immune” (Mu-end sequence-containing) and “non-immune” DNA molecules, a phenomenon called target immunity (Adzuma and Mizuuchi, 1988, 1989; Darzins et al., 1988). It has been suggested that MuB may have a differential role in non-replicative transposition during infection and replicative transposition; MuB is not required for the first integration although it seems to enhance the process (Roldan and Baker, 2001).

Other transposition accessory proteins include the host-encoded DNA bending proteins HU (i.e. hydroxyurea; Craigie et al., 1985) and IHF (integration host factor) in the case of *Escherichia coli*. HU promotes assembly by binding and bending DNA between the L1 and L2 sites (Lavoie and Chaconas, 1993; Lavoie et al., 1996). HU also has a second binding site in the transposition complex, which is possibly recog-

nised through an altered DNA structure (Lavoie and Chaconas, 1994).

The second host-encoded protein, IHF, binds the transpositional enhancer (IAS) sequence during initial assembly (Surette and Chaconas, 1989; Surette et al., 1989). Highly supercoiled plasmid donor DNA is required for efficient transposition and is important for the IAS function *in vitro*, but within physiological supercoiling levels IHF acts as a “supercoiling relief factor” enabling transposition (Surette and Chaconas, 1989). High concentrations of IHF can substitute for HU, both of which are DNA flexer proteins (Surette and Chaconas, 1989).

3.5.3. Three-site synapsis (LER)

Bacteriophage Mu transposes through a series of protein-DNA complexes called transpososomes, which consist of the two ends of the bacteriophage Mu genome, synapsed by the MuA transposase protein. MuA binds initially as a monomer to the MuA binding sites at the ends of the donor DNA, but functions as a tetramer of MuA (Lavoie et al., 1991; Baker and Mizuuchi, 1992; Baker et al. 1993). The first complex known to form during *in vitro* transposition is a three-site synaptic LER-complex between the enhancer, left- and right-ends of Mu (Watson and Chaconas, 1996). Host HU and IHF proteins help in the formation and stabilisation of productive LER. The N-terminal domain of MuA, which binds to the enhancer, is essential for LER-formation. The LER-complex is a transient intermediate that is rapidly converted to a stable synaptic complex (Watson and Chaconas, 1996). Figure 5 shows the formation of LER and the following transpososome complexes *in vitro* with supercoiled plasmid substrates.

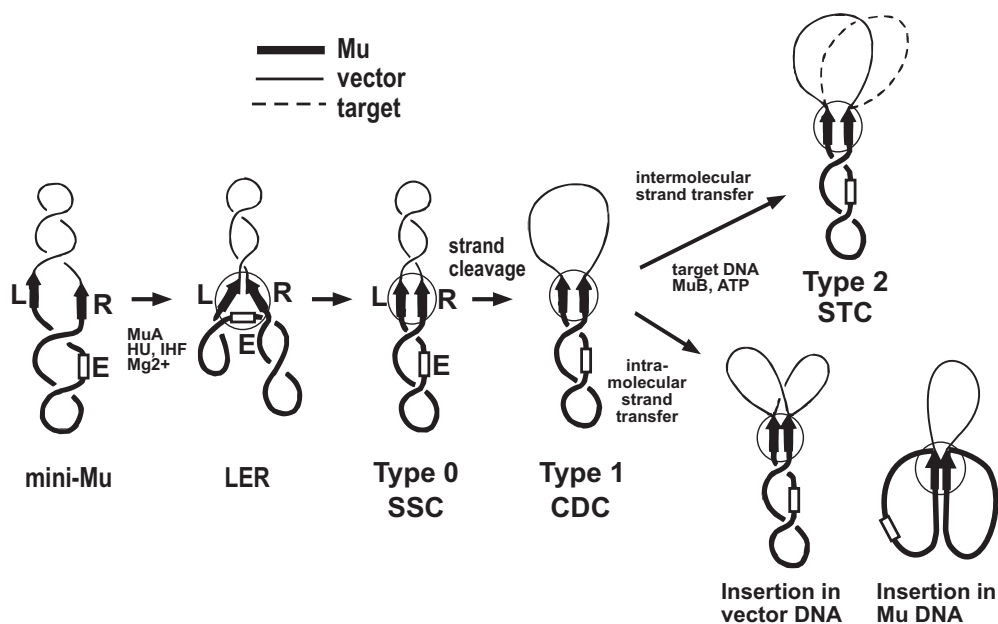


Figure 5. The replicative bacteriophage Mu DNA transposition pathway. Mu *in vitro* transposition reaction with supercoiled Mini-Mu donor plasmid, enhancer sequence, right- and left-end binding sites for MuA. In addition to MuA, also HU, IHF, Mg²⁺, MuB, ATP and target plasmid are required for the reaction. See the text for more details. (Drawn according to Wu and Chaconas, 1997).

3.5.4. Stable synaptic complex formation

The transient LER is converted to a stable synaptic complex (SSC, Mizuuchi et al., 1992) also called a type 0 complex. In the SSC the two ends of Mu are synapsed stably with a tetramer of MuA, but no cleavage has yet occurred. The IAS is not a stable component of SSC and it is required only for assembly (Mizuuchi et al. 1992). The SSC accumulates in the presence of Ca²⁺, whereas either Mg²⁺ or Mn²⁺ is needed for the following steps (Mizuuchi et al., 1992). Formation of SSC is a critical control point in the Mu transposition pathway before any chemical reaction steps are made (Mizuuchi et al., 1992). High control is achieved by meeting

several requirements: the properly oriented Mu-ends, accessory DNA sites, suitable divalent cation, supercoiled DNA topology and probably extensive structural transitions in the DNA-protein transpososome with high transition energy requirements.

3.5.5. Donor cleavage

Mu transposition includes two critical chemical steps similar to those described above for other DNA transposons. In the first chemical step, a pair of site-specific endonucleolytic cleavages is made at the flanking host – transposon DNA boundary, exposing the Mu 3'-OH ends. In the presence of Mg²⁺, the SSC is converted to a cleaved donor complex (CDC, Craigie and Mizuuchi, 1987), also called a type 1

transpososome complex (Surette et al., 1987). In the CDC, the cleaved Mu-ends are held together non-covalently ready to complete strand transfer if MuB, ATP and target DNA is provided (Figure 5). Also, donor DNA pre-cleaved with restriction enzymes can accomplish strand transfer upon incubation with MuA, MuB, ATP, and target DNA (Craigie and Mizuuchi, 1987). Thus, the cleavage and strand transfer are not energetically coupled.

Donor cleavage occurs by a single-step mechanism involving direct hydrolysis by water without covalent intermediates (Mizuuchi et al., 1999). In donor cleavage, a water molecule acts as the nucleophile attacking the scissile phosphodiester bond at the transposon 3'-end.

3.5.6. Strand transfer

The second chemical step consists of a pair of DNA strand transfer reactions in which the Mu donor 3'-OH ends are joined to the target DNA 5'-ends, which are made by a concerted 5-bp staggered cut. The transpososome complex is called a strand transfer complex (STC) or a type 2 transpososome (Craigie and Mizuuchi, 1985a; Surette et al., 1987). The strand transfer reaction results in a branched molecule (i.e. the Shapiro-intermediate; Shapiro, 1979), in which the donor DNA is covalently linked to the target DNA via 5-nt single stranded gaps.

Strand transfer occurs via a one-step transesterification mechanism (Mizuuchi and Adzuma, 1991). The 3'-OH ends of the donor DNA act as the nucleophiles attacking the target DNA phosphates, which are in staggered positions, thus accomplishing a simultaneous one-step cleavage and joining. Similar polynucleotidyl transfer reaction

mechanisms are shared among several DNA transposition, retrovirus, and V(D)J recombination reactions (Kennedy et al., 2000; Mizuuchi, 1997).

MuB, together with ATP, stimulates the strand transfer by recruiting a suitable target and delivering it to the transpososome (Baker et al., 1991; Yamauchi and Baker, 1998). The MuB bound target DNA can also be captured earlier along the reaction pathway during the LER, SSC or CDC complexes, thus before the actual strand transfer (Naigamwalla and Chaconas, 1997). MuB also stimulates donor cleavage, especially if a terminal nucleotide has been mutated in one transposon end or the transposon is surrounded by unfavourable flanking sequences (Surette et al., 1991; Wu and Chaconas, 1992).

3.5.7. Processing of the strand transfer complex

The strand transfer product is a branched molecule with single stranded gaps and the MuA tetramer still intact. The ST-product usually serves as a template for Mu DNA replication, which leads to the formation of cointegrates. Alternatively, simple inserts can be formed possibly via repair of the ST-product (Craigie and Mizuuchi, 1985a). The stably bound MuA at the strand transfer product poses an impediment for assembly of the replication fork by host replication factors (Kruklitis and Nakai, 1994). Therefore, during transition from the transpososome to a replisome, the STC is destabilised and disassembled by *Escherichia coli* host proteins ClpX and other Mu replication factors (MRFs) (Kruklitis et al., 1996; Jones et al., 1998; Burton et al., 2001; Nakai et al., 2001). MuA apparently has a critical function at the beginning of this

transition process (Nakai and Krukltis, 1995). The host proteins process the single stranded gaps around the ST-product; thus the final transposition product is flanked by the 5-bp target site duplications (Allet, 1979; Kahmann and Kamp, 1979), a hallmark of genuine transposition.

3.5.8. Division of labour within the tetramer of MuA

There have been vigorous and even conflicting studies concerning the sub-unit arrangement of MuA tetramer supplying the active site DDE-motif for donor cleavage and strand transfer. The catalysis is known to take place in *trans*; that is, a MuA monomer bound to one Mu-end catalyses reactions on the other Mu-end (Aldaz et al., 1996; Savilahti and Mizuuchi, 1996). It was originally suggested that complete transposition required four active monomers (Baker et al., 1993, 1994). However, more recent studies have shown that only two monomers bound to the R1 and L1 sites (or the R1 sites of two R-end donors) are sufficient to provide the DDE-motifs for active site and catalyse complete transposition in *trans* (Namgoong and Harshey, 1998; Williams et al., 1999). The active sites thus work co-operatively, providing a way to control reciprocal two-ended integration. This is also shown by the ability of an uncleaved Mu-end to inhibit strand transfer on the other Mu-end (Williams et al., 1999).

3.5.9. Relaxed requirements in vitro

Bacteriophage Mu was the first transposon for which a cell-free *in vitro* transposition system was described (Mizuuchi, 1983). The system enables more detailed analyses in versatile experimental settings. The original *in vitro*

system used a donor DNA carrying the two Mu-ends in a proper orientation (i.e. in the same orientation as in the phage DNA) within a plasmid, extracts containing the A and B gene products of Mu, and *E. coli* extract with host factors (Mizuuchi, 1983). Mu transposition has been analysed extensively since, and the requirements for transposition have been further defined and minimised.

High superhelicity of the donor DNA plasmid eliminates the requirement of IHF (Surette et al., 1989; Surette and Chaconas, 1989), thus allowing HU to substitute for IHF. Addition of glycerol and reduction of the salt concentration relax requirements for formation of transpososomes (Mizuuchi and Mizuuchi, 1989). Even more drastic relaxation of the topological requirements can be obtained with the use of dimethyl sulfoxide i.e. DMSO (Craigie and Mizuuchi, 1986; Mizuuchi and Mizuuchi, 1989; Baker and Mizuuchi, 1992). The requirement of supercoiled donor DNA *in vitro* (Craigie et al., 1985) can be avoided by using DMSO (Craigie and Mizuuchi, 1986). In this case transposition can also proceed with a Mini-Mu donor plasmid that carries the Mu-ends in the wrong orientation (Mizuuchi and Mizuuchi, 1989; Baker and Mizuuchi, 1992). DMSO also enables transposition without HU, the IAS enhancer sequence, or without the N-terminal domain of MuA, which binds IAS (Mizuuchi and Mizuuchi, 1989). Addition of DMSO thus seems to relax the conditions particularly during the initial assembly of the transpososome.

Transposition reactions can be further simplified by the use of a pre-cut donor DNA fragment. The pre-cut ends of the donor are transferred to the target DNA bypassing the donor cleavage

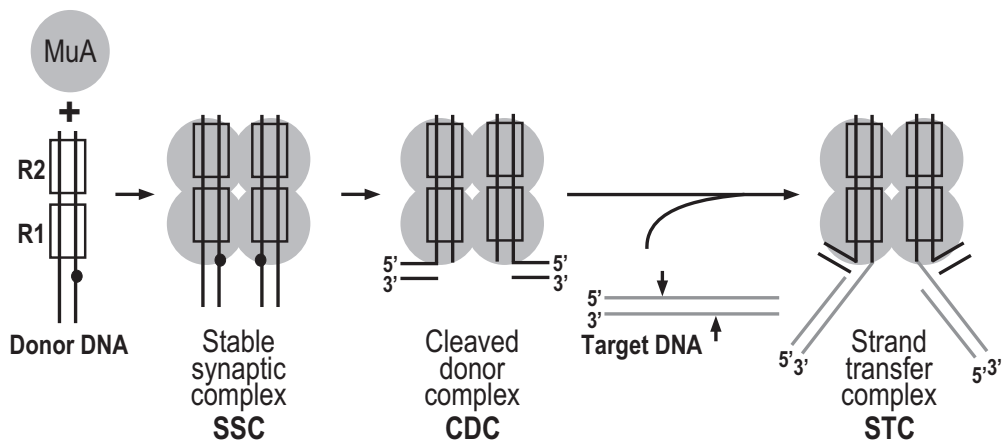


Figure 6. The Mu *in vitro* transposition reaction with minimal components: a short Mu right-end donor with the R1 and R2 binding sites, MuA transposase, linear target DNA (in grey) in modified *in vitro* reaction conditions with DMSO. (Adapted from Savilahti et al., 1995).

step. The strand transfer is most efficient if the normal left and right-ends are replaced by a pair of two pre-cut right-ends as the linear donor fragment (Craigie and Mizuuchi, 1987; Namgoong et al., 1994). The STC transpososomes thus formed contain two copies of the Mu right-ends, with the R1 and R2 sites bound by a tetramer of MuA (Mizuuchi et al., 1991). Therefore, the asymmetric left/right -end conformation is not necessary for the proper transpososome assembly and stability. Further, in modified *in vitro* reaction conditions including DMSO, a pair of short dsDNA fragments containing only the R1 and R2 binding-sites for MuA can form SSC, CDC and STC transpososomes in the presence of only MuA protein (Figure 6, Savilahti et al., 1995).

4. TARGET SITE SELECTION IN TRANSPOSITION

The choice of the integration site can have significant consequences not only for the host organism but also for the

transposon itself. Insertions into essential host genes can mean destruction of the host as well as the transposon. Therefore, transposons have evolved various strategies to avoid essential genes, find safe-havens, limit transposition efficiency or restrict transposition to times of emergency. Most elements show some level of target selectivity, ranging from totally sequence-specific to nearly random selection. Examples of target site selection strategies of different transposons are discussed in the following paragraphs and summarised in Table 1.

4.1. Several protein factors can act in the target site selection

In addition to the transposase, other transposon-coded or host proteins may also influence selection of a target site. For example, the bacterial transposon Tn7 shows an elaborate target site selection arsenal (Craig, 1991; 1996). TnsABC together with TnsD protein promotes a high frequency insertion into

Table 1. Summary of the consensus sequences recognised and other features of the target site selection within different transposable elements.

| TRANSPOSON | TARGET SITE DUPLICATION | CONSENSUS | OTHER FEATURES | REFERENCES |
|---------------------------|-------------------------|--------------------------------|--|--|
| ASV (avian sarcoma virus) | 4-6-bp | - | Crusiforms | Katz et al., 1998 |
| HIV-1 | 4-6-bp | No specific seq. | DNA bending, crusiforms | Bor et al., 1996; Muller and Varmus, 1994; Katz et al. 1998 |
| IS231A | 10-12-bp | GGG(N)5CCC | S-shaped structure, AT-periodicity | Hallet et al., 1994 |
| IS903 | 9-bp | 21-bp motif | Effect of flanking sequences and DNA structure | Hu and Derbyshire, 1998; Hu et al., 2001 |
| Mu phage | 5-bp | N-Y-G/C-R-N | Regional preferences and target immunity | Mizuuchi and Mizuuchi, 1993; Adzuma and Mizuuchi, 1988, 1989 |
| P-element | 8-bp | 14-bp motif | Palindromic pattern with structural features | Liao et al., 2000 |
| Tc1 (Tc3) | 2-bp | CAYA-TA-TRTG | | Van Luenen and Plasterk, 1994; Van Luenen et al., 1994; Ketting et al., 1997 |
| Tn3 | 5-bp | 19-bp motif | Suggested role for DNA secondary structure, target immunity | Davies and Hutchinson, 1995; Lee et al., 1983 |
| Tn5/IS50 | 9-bp | A-GNTYWRANC-T | Clusters of insertions | Goryshin et al., 1998 |
| Tn7 | 5-bp | attTn7 or many unrelated sites | Specific attTn7-site by TnsD + TnsABC, or many sites by TnsE + TnsABC, see text. Target immunity | Kubo and Craig, 1990; DeBoy and Craig, 1996, 2000; Kudu-valli et al., 2001; Rao et al., 2000; Peters and Craig 2000, 2001; Wolkow et al., 1996; Hauer and Shapiro, 1984; Stellwagen and Craig, 1997b |
| Tn10 | 9-bp | NGCTNAGCN | Effect of flanking sequences and structure | Bender and Kleckner, 1992; Halling and Kleckner, 1982; Pribil and Haniford, 2000 |

a specific *attTn7*-site in the *Escherichia coli* genome (Craig, 1991; 1996; DeBoy and Craig, 2000). TnsD binds to the *attTn7*-site sequence-specifically and induces DNA distortion that can be recognised by TnsC, which in turn activates the TnsAB transposition machinery (Kuduvalli et al., 2001; Rao et al., 2000). If the *attTn7*-site is unavailable, the TnsABC and TnsE proteins mediate a low-frequency insertion into various sites (Kubo and Craig, 1990; Peters and

Craig, 2000). The TnsE protein has been shown to direct transposition close to regions where DNA replication terminates or proximal to the DNA double-strand breaks (Peters and Craig, 2000, 2001). Also, conjugating plasmids are favoured targets of the TnsE-mediated pathway (Wolkow et al., 1996). Thus the Tn7 transposon has evolved two clever strategies: to either insert into a safe haven, *attTn7*, or to adventurously spread into unknown re-

gions with the chance to survive and potentially spread into new host cells.

4.2. Regional target site selection

Transposons can select target sites at various, and often intertwined levels, ranging from the broader “regional level” to the actual sequence level. The regional target site preference means that many sites within limited regions of chromosomes are utilised favourably, whereas other regions are disfavoured (Kleckner, 1981). Regional target preferences can guide transposons to certain safe regions of their host genome, such as transcriptionally silenced regions with an altered chromatin structure or regions upstream of genes transcribed by the RNA polymerase III (yeast Ty-retrotransposons, Boeke and Devine, 1998). The TnsE-mediated target site selection of Tn7, described above, can also serve as an example of the regional selection. Some transposons, such as Mu (Adzuma and Mizuuchi, 1988, 1989), Tn7 (DeBoy and Craig, 1996; Hauer and Shapiro, 1984; Stellwagen and Craig, 1997b), and the Tn3 family (Lee et al., 1983), avoid insertion into themselves and thus show target immunity (see below for more details) that can also be regarded as a form of regional selection.

4.3. A consensus sequence can often be found

Most commonly transposons insert into non-random positions with variable levels of sequence preference detectable within the duplicated target sequence (Craig, 1997). For example, the bacterial transposon Tn10/IS10 prefers the sequence 5'-NGCTNAGCN-3' (Halling and Kleckner, 1982; Bender and Kleckner, 1992), while the Tc1/*mariner* family elements always insert into a TA

(Van Luenen et al., 1994; Van Luenen and Plasterk, 1994). In some cases, preferences can also be detected in the region surrounding the duplicated sequence. For example, Tc1 prefers the TA embedded in CAYA-TA-TRTG (Ketting et al., 1997) and Tn5/IS50 prefers a sequence of A-GNTYWRANC-T within a cluster of similar sequences (Goryshin et al., 1998).

4.4. Certain target DNA structures can also be preferred

DNA is far from a monotonous rod; it is a dynamic, flexible molecule with sequence-dependent DNA structure governed by the limitations of the backbone (review by Alleman and Egli, 1997). It is therefore natural that some transposons seem to recognise certain structures or DNA sequence-related structure. For example, IS231A recognises a consensus within an apparent S-shaped DNA structure (Hallet et al., 1994). Similarly, DNA bent around nucleosomes and cruciform DNA are favoured as retroviral integration target sites *in vitro* (Muller and Varmus, 1994; Katz et al., 1998). A triple-helical DNA structure can be recognised by the Tn7 TnsC protein (Rao et al., 2000). A sequence-related DNA structure has been considered important in the Tn10 target site flanking sequences (Bender and Kleckner, 1992; Pribil and Haniford, 2000) and in P element insertion sites (Liao et al., 2000). Generally, target recognition can be thought of as fitting the target DNA to the protein active site through optimisation of the contacts made.

4.5. Bacteriophage Mu target site selection

Bacteriophage Mu was originally thought to transpose essentially ran-

domly into a number of target sites in the host genome (Bukhari and Zipser, 1972; Daniell et al., 1972; Couturier, 1976). However, some degree of non-randomness in targeting was revealed by closer examination *in vivo* (Raibaud et al., 1979; Silhavy et al., 1979; Emr and Silhavy, 1980; Castilho and Casabadian, 1991). Preference has been observed, for example, for the control region of *lacZ* when it is not being transcribed (Wang and Higgins, 1994) and near the *bgl* promoter around an A/T-rich segment (Manna et al., 2001).

Some preference at the 5-bp target site duplication was revealed for the first time by the phage genome -host DNA junction sequence determination (Kamp and Kahmann, 1980). Analysis of junction sequences of mini-Mu insertion sites further revealed a similar consensus sequence (5'-N-Y-G/C-R-N-3', where N is any nt, Y is C or T, and R is A or G) in *in vivo* and *in vitro* transposition (Mizuuchi and Mizuuchi, 1993). An identical consensus was obtained from *in vitro* reactions containing MuA and HU, either with or without MuB.

A regional target site preference has been associated with DNA stretches exhibiting high affinity for the MuB protein (Mizuuchi and Mizuuchi, 1993). As already mentioned, MuB is also responsible for the target immunity process, which helps Mu to avoid insertion into itself and thus self-destruction. In the presence of ATP, MuB binds to potential "non-immune" target DNAs that do not contain Mu-end sequences (Adzuma and Mizuuchi, 1988) and stimulates intermolecular strand transfer (Maxwell et al., 1987). The MuA protein, bound to Mu-end sequences, can interact with the MuB protein bound on the same "immune" DNA molecule and catalyse the dissociation of MuB

via a process that requires ATP hydrolysis (Adzuma and Mizuuchi, 1988, 1989). Target immunity is cis-acting under normal conditions, so Mu-end sequences in the same DNA molecule confer gradually decaying immunity to distances up to 20-25 kb (Adzuma and Mizuuchi, 1989; Darzins et al. 1988; Manna and Higgins, 1999). Looping (or the domain structure of chromosome) between DNA sites bound by MuA and MuB has been proposed as a mechanism for the target immunity (Adzuma and Mizuuchi, 1989; Manna and Higgins, 1999). Thus MuB has multiple roles in target immunity, target capture and regional target selection. It has also been asserted that different forms of MuA (monomers vs. tetramers in transpososomes) can react with an ATP-ADP switch controlled MuB in different ways (Yamauchi and Baker, 1998).

Transposition can also be performed without MuB, and the target can be captured into transpososomes by MuA. However, mini-Mu donors tend to use sites on the donor plasmid itself as targets (intramolecular targets) and the reaction is less efficient with MuA alone (Maxwell et al., 1987). Nevertheless, in certain *in vitro* conditions the Mu transposition reaction can be performed efficiently with only MuA, a short Mu right-end donor DNA, and target DNA as the minimal macromolecular components (Savilahti et al., 1995).

5. TRANSPOSONS AS MOLECULAR TOOLS

5.1. Why transposon tools?

Since their discovery over fifty years ago, transposable elements have been utilised widely in bacteria, but also in eukaryotes for various molecular ge-

netic tasks *in vivo* and during recent years in cell-free systems *in vitro* (reviewed by Hamer et al., 2001; Hayes and Hallet, 2000). Transposons catalyse their own insertion into variable target sites, thus performing the duties of a restriction enzyme/ endonuclease and a DNA ligase in a single package. The ability of transposon insertions to cause mutations and potentially alter the expression of genes has been exploited widely to analyse the functions of genes and the proteins they encode. One advantage of transposon tools is that pools of mutants can be produced easily. This contrasts with other mutagenesis strategies using standard molecular cloning methods, which are often laborious multistep procedures and demand the construction of each mutant one by one.

Modern molecular biology methods have enabled the relatively easy construction of “design”-transposons; a multitude of phenotypic or physical markers can be engineered between the transposon end sequences. For example, different selectable marker genes, predetermined restriction sites, reporter genes, promoters and replication origins can be included within transposons (Berg and Berg, 1995). This allows a wide utilisation of transposons for different tasks, such as mutagenesis, induction of RNA synthesis by promoter transposons, monitoring transcription or translation by reporter transposons, identifying open reading frames, characterising operon organisation, generating rearrangements (deletions or fusions), and DNA sequencing (Berg and Berg, 1995).

5.2. Different ways to use transposon tools

Different transposon tools utilise trans-

posable elements and transposases from various sources. The actual transposition reaction can take place either *in vivo* or *in vitro*. *In vivo* transposition requires either mobilisation of an endogenous transposon residing in the host chromosome or introduction of a suitable transposon and transposase into the host cell by transformation or bacterial mating. The use of transposons as molecular tools *in vivo* can suffer from the restricted host ranges of different elements. This limitation can be avoided in some cases by shuttle mutagenesis: transposition takes place within the surrogate host *Escherichia coli*; the mutagenised DNA is then isolated and transformed into the original host (Hamer et al., 2001).

5.2.1. *In vivo* tools

Different transposons have been utilised successfully *in vivo* for many purposes, some examples of which are summarised in the first part of Table 2. For example, several DNA sequencing strategies have been developed on the basis of *in vivo* transposition (Adachi et al., 1987; Strathman et al., 1991; Kasai et al., 1992; Phadnis et al., 1989; Berg et al., 1993). Important genes (or parts of genes) required for growth under certain selective conditions can be mapped by PCR-based genetic footprinting. This strategy has been utilised for the analysis of gene functions in yeast *Saccharomyces cerevisiae* by the Ty1 transposable element (Smith et al., 1995) and by the Mariner-based transposon in *Pseudomonas aeruginosa* (Wong and Mekalanos, 2000). Transposon facilitated insertional mutagenesis has been utilised to tag proteins with e.g. protease recognition sites (Ehrmann et al., 1997), antibody-binding epitopes (Sedgwick et al., 1991) or

reporter genes such as gene coding for the green fluorescent protein (Ross-Macdonald et al., 1997). *In vivo* transposition has been applied successfully to the functional analysis of genes, proteins and whole genomes (i.e. functional genomics) in diverse organisms including several bacteria, plants, fungi and yeast (reviewed by Hamer et al., 2001).

5.2.2. Simplified *in vitro* tools

Complications associated with host-range limitations can be avoided if the transposition reactions are performed *in vitro* and the resulting integrants are transformed into the host cell. The *in vitro* systems are based on the use of purified, or partially purified, components of the transposition machines in defined cell-free conditions. This allows easy manipulation of the components; in particular the donor fragment can be modified to include selection markers, tags, and epitopes. An *in vitro* transposition system has already been established for several different transposons. Some of the most useful and/or simple *in vitro* systems are based on e.g. Tn5, Tn7, Tn552, Ty1, Mariner, and Mu-phage transposition (Tn5, Goryshin and Reznikoff, 1998; Tn7, Waddell and Craig, 1988, Biery et al., 2000b; Tn552, Leschziner et al., 1998; Ty1, Eichinger and Boeke, 1988; Mariner, Lampe et al., 1996; Mu, Mizuuchi, 1983, Haapa et al., 1999a). The *in vitro* systems have been used for various tasks, examples of which are summarised in Table 2. In addition to larger insertional mutations, transposons can be utilised to make smaller, for example, five amino acid (i.e. pentapeptide or 15-bp) insertions into the target proteins (reviewed by Hayes and Hallet, 2000). The *in vitro* conditions enable the manipulation of

the mutant pool *en masse*. This is exploited in the pentapeptide scanning mutagenesis, where most of the transposon is cut away from the insertion site by the use of restriction enzymes and ligase (Mu, Taira et al., 1999, Laurent et al., 2000; Tn7, Biery et al., 2000b; Tn4430 with transposition *in vivo*, Hallet et al., 1997).

Several *in vitro* transposition systems are currently available commercially for different purposes. These systems are based on transposition reactions catalysed by the Ty1 virus-like particles (Primer Island system from Perkin-Elmer), a hyperactive mutant transposase of Tn5 (EZ::TN™ systems from Epicentre), TnsABC transposases of Tn7 (GPS™: Genome Priming System from New England Biolabs) or the MuA transposase of phage Mu (TGS: Template Generation System and MGS: Mutation Generation System from Finnzymes).

Also, a new combination of *in vitro* and *in vivo* systems is produced by transforming, usually by electroporation, *in vitro* pre-assembled transposon-transposase complexes (i.e. transpososomes) into the host cells where transposition takes place *in vivo* (Tn5, Goryshin et al., 2000; Mu, Lamberg et al., in press). This transpososome approach is free of the host-range limitations, because it only requires a procedure to introduce the transposomes into the host cell. The transpososome strategy has been shown to work with several bacterial species (the Tn5 strategy with *Escherichia coli*, *Salmonella typhimurium*, *Proteus vulgaris*, *Pseudomonas sp.*, Goryshin et al., 2000; Hoffman et al., 2000; and the Mu strategy with *E. coli*, *S. typhimurium*, *Erwinia carotovora*, *Yersinia enterocolitica*, Lamberg et al., in press) and

in yeast (Goryshin et al., 2000).

Table 2. Summary of versatile applications of different *in vivo* and *in vitro* transposition systems.

| TRANSPOS. REACTION | TRANSPOSON | USE(S) OF SYSTEM | REFERENCES |
|-----------------------|-----------------------------------|---|---|
| <i>In vivo</i> | $\gamma\delta$ | DNA sequencing | Strathmann et al., 1991 |
| | Ty1 | Genetic footprinting | Smith et al., 1995 |
| | Tn5-based | Insertion of protease cleavage sites, Transposome insertional mutagenesis and sequencing | Ehrmann et al., 1997; Hoffman et al., 2000; Goryshin et al., 2000; Kasai et al., 1992; Phadnis et al., 1989 |
| | Tc1/Mariner-like, Sleeping beauty | Genetic footprinting, non-homologous integration of transgenes into mice. Chromosomal transposition and mutagenesis | Wong and Mekalanos, 2000; Yant et al., 2000; Horie et al., 2001 |
| | Mu | DNA sequencing, <i>in vivo</i> cloning, probe for transcriptional control sequences and coding regions, formation of gene fusions | Adachi et al., 1987; Groisman and Casadaban, 1986, 1987; Casadaban and Cohen, 1979; Groisman et al., 1984, 1991; Casadaban and Chou, 1984 |
| | Tn1000 | Mutagenesis, mapping of epitopes. Recombinant DNA manipulation, localisation of transcriptional regulators | Sedgewick et al., 1991; Morgan et al., 1996 |
| | Tn3-based | Making short in-frame insertions. Defining membrane topology. Transposon tagging, gene disruption, analysing protein production, localisation and function. | Hoekstra et al., 1991; Das and Xie, 1998; Ross-Macdonald et al., 1997, 1999a, 1999b |
| | Tn4001 | Genome-scale analysis by mutagenesis | Hutchison et al., 1999 |
| | Tn4430 | Pentapeptide scanning mutagenesis | Hallet et al., 1997 |
| | <i>In vitro</i> | Ty1 | DNA mapping, sequencing and genetic analysis, genome scanning/ functional mapping of essential genes, construction of gene targeting vectors for generation of transgenic animals |
| Tn552 Tn552'phoA | | DNA sequencing and mutagenesis, generation of fusion libraries and identification of genes encoding certain proteins | Griffin et al., 1999; Braunstein et al., 2000 |
| Tn5 | | Generation of nested deletions and inversions, genome scale mutagenesis | York et al., 1998; Gehring et al., 2000 |
| Tn7 | | Genome mutagenesis and gene function analysis, translational and transcriptional target gene fusions, pentapeptide insertions | Gwinn et al., 1997; Biery et al., 2000b |
| Tn10 | | Genome mutagenesis and gene function analysis | Sun et al., 2000 |
| Mariner | | Identification of essential genes by genomic footprinting, genome scale mutagenesis | Akerley et al., 1998; Pelicic et al., 2000; Gehring et al., 2000 |
| MoMLV | | Genetic footprinting of a cloned gene | Singh et al., 1997 |
| Mu | | DNA sequencing, construction of gene-targeting vectors for generation of transgenic animals, genetic footprinting, pentapeptide scanning mutagenesis | Haapa et al., 1999b; Vilen et al., 2001; Haapa et al., 1999a; Laurent et al., 2000; Taira et al., 1999 |

C. AIMS OF THE PRESENT STUDY

A cell-free *in vitro* reaction was developed for the bacteriophage Mu transposition reaction some years ago (Mizuuchi 1983). This reaction has subsequently been simplified further, and in certain *in vitro* conditions only MuA transposase, short Mu right-end donor DNA, and target DNA are needed as minimal macromolecular components to perform the transposition reaction (Savilahti et al., 1995). We believed that this minimal *in vitro* system could be developed into a simple and efficient *in vitro* transposon tool. For that purpose, it was necessary to analyse several details of the transposition reaction mechanism prevailing in minimal *in vitro* conditions. Target site selection is an important aspect of transposition with regard to applications, but it also includes more general scientific interests.

The aim of this study was specifically to:

- 1) construct mini-Mu donor DNA fragments and to examine the bacteriophage Mu *in vitro* transposition reaction mechanism (the frequency of one-ended vs. two-ended integration, reaction efficiency, target site duplication accuracy and target site selection) as well as to test the suitability of the minimal Mu *in vitro* transposition assay for applicational purposes.
- 2) design a new sequencing template generation methodology based on the minimal Mu *in vitro* transposition assay and to test the suitability of the methodology for a sequencing project of substantial size.
- 3) analyse the bacteriophage Mu target site selection at nucleotide level in conditions under which the effect of MuA transposase can be analysed alone. A quantitative assay had to be developed for that purpose.

D. MATERIALS AND METHODS

The bacterial strains used in cloning as well as the plasmids and oligonucleotides are described in the original publications. The transposon donor DNA fragments are described in detail in the original publications and summarised in Table 3. The experimental methods used in this study are described in the original publications and summarised in Table 4.

Table 3. Transposon donor DNA fragments used in this study.

| Donor name | Marker | Length | 5'-overhang | Specifications | Described in |
|---------------|---|----------|-------------|---------------------------------------|--------------|
| supF-Mu | supF amber suppressor tRNA | 0.4 Kb | 4 nt | Precut with BglIII, purified fragment | I |
| cat-Mu | Chloramphenicol acetyltransferase gene, cat | 1.2 Kb | 4 nt | Precut with BglIII, purified fragment | I, II |
| R-end (blunt) | None | 50 bp | 0 nt | Oligo-construct, precut | III |
| R-end | None | 50/51 bp | 1 nt | Oligo-construct, precut | III |

Table 4. Methods used in this study.

| Method | Described and used in | | |
|--|-----------------------|----|-----|
| Autoradiography | I | | III |
| DNA sequencing | | | |
| Automated | I | II | |
| Manual | | | III |
| Molecular cloning techniques | I | | III |
| 5'-radiolabeling | I | | III |
| Plasmid DNA isolation | I | II | III |
| PCR | I | | III |
| Computer-aided sequence analysis | I | II | III |
| Computer-aided sequence-related structure analysis | | | III |
| Oligonucleotide gel purification | | | III |
| Transposition reaction based assays | | | |
| Minimal in vitro mini-Mu assay | I | II | |
| Minimal in vitro oligo-donor assay | | | III |
| Genetic footprinting | I | | |
| Sequencing template generation | | II | |
| PCR based target site selection assay | I | | |
| Quantitative target site selection assay | | | III |
| Urea-PAGE-analysis | I | II | III |
| HPLC | I | | III |
| Restriction analysis | I | II | III |
| Bacterial transformations | I | II | III |
| Electroelution | I | | |
| Electrophoresis techniques | I | II | III |
| Statistical data analysis | | II | III |

E. RESULTS AND DISCUSSION

1. THE MINIMAL *IN VITRO* TRANSPOSITION REACTION RECAPITULATES THE HALLMARKS OF *IN VIVO* TRANSPOSITION

The essential features of the genuine productive *in vivo* Mu transposition reaction are: 1) two-ended integration of the element; 2) accuracy of the reaction and 5-bp target site duplication; 3) relatively random target site selection. We analysed the details of the minimal Mu *in vitro* transposition reaction, focusing on these above features.

In the minimal *in vitro* reaction conditions (Savilahti et al., 1995), a short Mu right-end double stranded DNA fragment was used as the donor DNA. For applicational and other purposes it is practical to include at least a selectable marker between the two right-ends of Mu. Two such mini-Mu donor DNA constructs were designed and constructed (supF-Mu and cat-Mu) with either a *supF* amber suppressor tRNA or chloramphenicol acetyltransferase gene (*cat*) as a marker (I). Transposition reactions can be further simplified by using pre-cut donors, whose active 3'-ends are already cleaved and thus ready for the strand transfer step (Craigie and Mizuuchi, 1987). Therefore, the supF-Mu and cat-Mu donors were cut out of the plasmid backbone with BglII, which liberates the reactive 3'-OH ends and leaves the 4-nt flanks on the non-transferred strands. The donor fragments were further purified using anion exchange chromatography.

1.1. One-ended versus two-ended integration (I)

The frequency of one-ended versus

two-ended integration was studied by using electrophoresis and autoradiography (I; Figure 2) to analyse the transposition reaction products resulting from reactions with the radioactively labelled pre-cut donor (supF-Mu) and the supercoiled plasmid (pBC SK+) target. The reaction products were additionally digested with Avall (which cuts once within the donor fragment) to distinguish between the two-ended and one-ended integrations. The amounts of different reaction products can be quantified easily. To summarise, in the minimal *in vitro* transposition reaction conditions used, the major products result from the two-ended integration of a single donor into a target (I; Figure 2). One-ended integrations can be deliberately generated by using a mixture of wild-type MuA and an active site mutant MuA_{E392Q}, which is defective in the transposition reaction (Baker and Luo, 1994).

Multiple transposons can also integrate into a single target (I; Figure 2, 10 nM of donor), because the target immunity is not effective in the *in vitro* conditions when MuB is not present. However, multiple integrations into a single target can easily be avoided by using a molar excess of the target DNA in relation to the donor. In our reaction conditions, the majority of the transposition products result from single integrations (I; Figure 2). Thus the *in vitro* reaction product profile is to some extent modifiable and can be suited to applicational use. For instance, it is desirable to have a single transposon inserted via both ends into a target in order to be able to read the DNA sequence outwards from the transposon end specific primers (as is done in publication II).

1.2. Efficiency (I, II)

Efficiency of integration into target plasmids can be scored by transforming products of the *in vitro* transposition reactions into competent *E. coli* cells and selecting for the mini-Mu and plasmid markers. The recovery of mini-Mu transposon insertions was about 4-11.8 % of the total plasmids recovered, which carried the target plasmid marker (I; table 1, 10 nM donor). Approximately 10^3 transposon marker-carrying colonies were obtained easily when 1 μ l of transposition reaction (i.e. 10 fmol of donor DNA) was transformed into standard competent cells (competence status 10^6 - 10^7 cfu/ μ g plasmid DNA). For sequencing purposes, we further lowered the amount of the donor in relation to the target in order to avoid multiple insertions into a single target plasmid. We recovered 540 transposon marker-bearing colonies by electroporation (competence status 10^7 cfu/ μ g plasmid DNA) when only one tenth of the phenol extracted transposition reaction was transformed (i.e. 2 fmol of donor DNA; II). With electroporation and/or better competent cells (10^9 cfu/ μ g plasmid DNA) even up to 10^7 - 10^8 transposon-carrying colonies per μ g of DNA can be acquired (I). Thus the minimal *in vitro* transposition reaction efficiency is more than sufficient for applicational purposes.

1.3. Accuracy (I, II)

Accuracy of the reaction is demanded in two reaction steps: in cutting the 3'-ends of the transposon and in the concerted target site cleavage/joining. The mobile element avoids degrading itself by recognising and precisely cleaving its own 3'-ends. Accuracy in generating the target site duplication is impor-

tant for applications, because in sequencing, the obtained DNA sequences are joined together into a mini-contig by the target site duplications, which are repeated at each transposon end. Also, in the pentapeptide scanning mutagenesis, accurate 5-bp target site duplications are required to generate the final 15-bp insertion. We have analysed transposon insertion sites from a total of 125 individual clones by sequencing (54 clones in I; 71 clones in II). All of the sequenced clones contained an accurately duplicated 5-bp target site flanking the transposon end sequences and thus showed the hallmark target site duplication (Allet, 1979; Kahmann and Kamp, 1979).

1.4. Target site selection (I, II)

Bacteriophage Mu shows only moderate target site preference in comparison with other transposons. A consensus sequence of duplication (5'-N-Y-G/C-R-N-3') has been found in both *in vivo* and *in vitro* transposition reactions by determining the junction sequences of mini-Mu insertion sites (Mizuuchi and Mizuuchi, 1993). Target site selection is important with regard to applicational purposes as well. A fairly even distribution of transposon insertion sites, except within the essential origin of replication, was seen within a 3.4-kb target plasmid pBC SK+ when the junction sequences of 52 individual clones were sequenced (I; Figure 3A). Similarly, a relatively even distribution of insertions was also seen during the sequencing of a 10.3-kb mouse genomic insert cloned within a 2.9-kb plasmid (II; Figures 2 and 3, Table 2). In addition, no orientation dependence was seen within the transposon insertions into the target plasmid (I, II).

Additionally, the fine-scale distribution of integrated supF-Mu transposons was analysed with a PCR-based assay. This assay, with one 5'-labelled target-specific and one transposon-specific primer, revealed that nearly all phosphodiester bonds were used as targets, but with different frequencies (I; Figure 3B). As well as supercoiled plasmid targets, open circular and linear DNA can also serve as targets. The target DNA conformation does not affect the transposition reaction efficiency or the integration site selection (I; data not shown). Nor does the type of donor DNA, neither the supF-Mu nor a 50-bp Mu right-end oligo donor, affect the integration site selection (I; data not shown).

These experiments showed that although some preferred sites are observed, MuA directs integrations relatively evenly within longer DNA segments. This is in contrast to *in vitro* reactions where MuB is present. A regional target site preference associated with DNA sites exhibiting high affinity for MuB has been detected (Mizuuchi and Mizuuchi, 1993).

2. APPLICATIONS OF THE MINIMAL MU *IN VITRO* TRANSPOSITION METHODOLOGY

Since the generation of mutants by the minimal Mu *in vitro* transposition is highly efficient, accurate and relatively random, we believe that this system can easily be used for versatile purposes as a tool for molecular biology. This was demonstrated by two different examples.

2.1. Genetic footprinting (I)

Functional DNA regions (e.g. genes, origins of replication, promoters) can be

localised by insertional mutagenesis. The extensive pools of insertion mutants, which are needed for the functional analysis, can be easily generated by Mu transposition. We designed an assay based on the comparison of two pools of insertion mutants: an unselected pool and one selected for a function. Both pools were selected for the transposon (cat-Mu) marker. As a test case, we selected for functions essential for plasmid replication (selection with ampicillin). The DNA regions of interest were localised with two different assays: either a simple restriction digestion followed by Southern blotting (I; Figure 4) or a PCR-based footprinting assay (I; Figure 5). The regions essential for the plasmid replication cannot tolerate transposon insertions. In the digestion-based assay, these essential areas are seen as underrepresented, transposon-containing restriction fragments as compared to the unselected pool (I; Figure 4). In the PCR-based detection assay, a clear footprint can be seen at the replication origin (I; Figure 5). Both detection assays proved feasible for functional mapping; however, the PCR-based assay gives a clearly more detailed nucleotide level resolution.

2.2. Sequencing (II)

As a second application, we designed a system for generating templates for DNA sequencing based on the minimal Mu *in vitro* transposition (II). To test the system, we determined a 10,288-bp sequence from a mouse genomic locus containing the *Kcc2*-gene. The corresponding protein, KCC2, is a K⁺/Cl⁻ transporter that is abundantly expressed in the mouse brain and assumed to be involved in extruding chloride ions in mature neurons (Payne et

al., 1996). The cat-Mu mini-transposon was integrated *in vitro* into the target plasmid, the reaction products were electro-transformed into *E. coli*, and the integrants were selected for resistance to chloramphenicol and ampicillin (from the target plasmid). We screened with restriction digestions 95 clones, of which 83 % had a transposon integrated into the insert portion of the plasmid. Additionally, colony PCR can be used for screening, or optionally, the integrants can be shotgun-sequenced without screening. For 71 clones, the DNA sequence was determined starting from the unique primers within the transposon ends. Totally, 10,288-bp worth of sequence was retrieved with a redundancy of 6.6 (II; Table 1). The mean interval distance between insertions was 146-bp; the minimum distance was 0-bp (two integrations into same site) and the maximum distance 570-bp (II; Table 2). On average, about 500-bp of sequence was determined with a single primer; thus a single transposon insertion can yield about 1-kb worth of sequence information.

The minimal *in vitro* transposition system proved to be well suited for sequencing template generation in several ways. The system is efficient and easy to use. The requirements for transposition reaction are simple and easily modifiable. A single protein, MuA, and only a few nanograms of the artificial donor DNA are needed for the one-hour *in vitro* reaction. The artificial cat-Mu transposon provided an easily selectable marker for the integrant retrieval, although other selectable marker containing donors can also be designed. Only two putative double-integrants, which are unsuitable for sequencing, were observed during screening (~2 % of all, II; Table 1). And

finally, the target site selection by MuA was random enough for retrieving suitable templates for sequencing (II; Figure 2, Tables 2 and 3). The integrants collected during a sequencing project can additionally be exploited as insertion mutants for functional studies.

Based on these studies on the minimal Mu *in vitro* transposition, a commercial product from Finnzymes has become available for transposon-facilitated sequencing (TGS, Template Generation System). The Mu-based sequencing system is now in standard use in the Sequencing facility at the Institute of Biotechnology, University of Helsinki.

3. TARGET SITE SELECTION (III)

In the two publications described above (I, II), the target site selection and accuracy of the target site duplication were analysed with an emphasis on the applicational use. The results showed a relatively even selection of integration sites with some preferences. We were interested in further analysing these preferences. The minimal Mu *in vitro* transposition assay enables dissection of the target site selection in controlled reaction conditions in isolation from regional level selection factors. We believe that while regional target selection can be modulated by contacts between the transpososomes and accessory factors (such as MuB) bound to or embedded in the DNA structure, the final DNA-protein interactions between the target site and the transpososome must act in the sequence level selection.

3.1. A new quantitative nucleotide-level assay

The sequence level target site selection is difficult to study *in vivo* because

multiple selection factors can act simultaneously. Therefore, we set up a direct nucleotide level *in vitro* assay in which the target choice is influenced by the MuA transposase alone (III). The assay (III; Figure 2) gives the opportunity to quantitatively analyse how frequently different sequences are used as the target sites. Our aim was to examine the preference of all 1024 different 5-bp sequences. To do so, 20 different target DNA fragments with a single 5'-end labelled were made (III; Figure 3A and B) and together with a short right-end donor DNA used as substrates in the transposition reactions. The transposition reaction products were analysed as populations of 5'-labelled DNA molecules by denaturing polyacrylamide gel electrophoresis followed by autoradiography. The quantification results were adjusted for differences in the labelling. The data set collected contained at least one observation from 806 different pentamers, which corresponds to 79 % of all possible 1024 pentamers (III; Figure 3C, D and E). The data was notably skewed when arranged by intensity (III; Figure 3D); only one tenth of the 1540 data points account for half of the total intensity sum. Thus relatively few sites are clearly preferred, whereas most sites are acceptable.

In our assay the transposition reaction products are detected directly without potentially artifactual amplification steps. However, all quantitative assays inevitably include some level of error. Based on the data shown in Figure 2C (III), we estimated that about 3.8 % of the total variation (see also chapter 3.6) is caused by poor gel quality and/or pipetting accuracy.

3.2. The nucleotide consensus

A consensus sequence of duplication (5'-N-Y-G/C-R-N-3') has previously been observed in both *in vivo* and *in vitro* transposition reactions either with or without MuB (Mizuuchi and Mizuuchi, 1993). Our data set enables the calculation of the target consensus in several novel ways. A target site consensus sequence can be compiled by tabulating the band intensities (i.e. the insertion frequencies of sites) with the corresponding nucleotide sequence (III; Figure 4A). In addition, consensus sequences for the most favoured, but also for the disfavoured target sites can be calculated. The consensus for the most preferred 10 % of pentamers was 5'-C-Y-G/C-R-G-3' (III; Figure 4B) with statistically significant differences in the nucleotide usage in all five positions. A consensus for the disfavoured sites has usually not been analysed or it has been assumed to be the opposite of the consensus of the preferred sites (Davies and Hutchison, 1995). From our data set we could actually analyse the disfavoured sites as well; the consensus sequence for the least preferable 10 % of pentamers, 5'-N-A/G-A/T-C/T-N-3' (III; Figure 4D), was essentially the opposite of that of the most preferred pentamers. No significant preferences in the nucleotide usage were found for the medium preferable sites (III; Figure 4C).

Thus, the consensus sequence obtained for the preferred sites follows the consensus of the previously published *in vivo* and *in vitro* experiments (Mizuuchi and Mizuuchi, 1993), except for the additional statistically significant preferences at positions 1 and 5 in our data. Cytosines were previously assumed to be preferred *in vivo* at posi-

tions 1 and 5 at the left end and guanines at positions 1 and 5 at the right end (Kamp and Kahmann, 1980). No explanation for these differences exists, but the low number of insertions studied previously, different reaction conditions, and additional protein factors (MuB, HU) in the reactions might affect the results.

3.3. Dinucleotide preferences

Because the target site selection of phage Mu transposition seems to be relatively flexible, recognition could be based on the sequence-related DNA structure rather than on the specific sequence. The DNA structure can be described by 16 base steps (or dinucleotides), which reflect the structural relationships of the adjacent bases. We wanted to determine whether these base steps could describe the target site selection more accurately than simply bases. Some dinucleotides were found to be more preferred than others (III; Figure 5). As expected, preferences were more pronounced when only the most preferred 10 % of pentamers (III; Figures 5C) were analysed as compared to the frequency-type consensus (III; Figure 5A and B). Again, the dinucleotide preference for the least preferable 10 % of pentamers (III; Figure 5D) was the opposite of that of the most preferred pentamers.

Generally, this kind of dinucleotide “Manhattan-model” represents the target site preferences more accurately and with a better resolution than the standard sequence analysis (16 base-steps vs. 4 bases). Some dinucleotides are known to be especially flexible (CA/TG, TA), bistable (GG/CC, CG, AG/CT, TA) or rigid (AA/TT, AT, GA/TC) thus reflecting sequence-related DNA structure (Lilley, 1995). In our data, the flex-

ible or bistable dinucleotides seem preferred at the central positions and either the bistable or rigid steps are favoured at the edges of the 5-bp core target site (III; Figures 5A, B, C).

3.4. An apparent symmetry in the target site selection

The nucleotide and dinucleotide consensus of the preferred target sites (III, Figures 4 and 5) show two-fold inverted symmetry (i.e. palindromy), which reflects the Watson-Crick base pairing rules of the dsDNA. We believe that the apparent symmetry in the target sites arises from functional two-fold symmetry of the transposition reaction. Biochemical data on the Mu transposition complexes has indicated that the active sites of two of the four MuA protomers are used to promote cleavage and strand transfer reactions at the two ends of the element (Williams et al., 1999). The MuA monomers and, in our assay, also the right end donors are functionally symmetrical. Because the transposition complex is most probably symmetrical also in relation to the target DNA, recognition can be made in both orientations without affecting the specific contacts. Therefore, it is impossible to know in which orientation each sequence should be aligned, and the consensus will inevitably look symmetrical, even if the individual targets are asymmetrical. This kind of reasoning may also apply to other transposons for which symmetry or palindromy in the target sites has been observed (Tn10, Bender and Kleckner, 1992; Tn3, Davies and Hutchison, 1995; Tn5, Goryshin et al., 1998; IS231A, Hallet et al., 1994; IS903, Hu et al., 2001; Tc1, Ketting et al., 1997; P element, Liao et al. 2000). It is also noteworthy that a

consensus does not represent the best individual target site but an average.

3.5. Insertion site preferences can be predicted

We believed that the dinucleotide data could be used to make predictions on the attractiveness of each pentamer. This was proven with statistical methods as well as experimentally. The hypothesis was that a sum of base-step values (predictional value) calculated for a pentamer (from Figure 5B, III) would correlate with the observed intensity value. Statistical testing of the hypothesis (III; Figure 6A) using Spearman's rank correlation showed highly significant correlation between the predicted and observed intensity values. The predictional value of the "Manhattan"-model was tested experimentally by designing four different custom targets. The results show that the Mu transposon preferentially inserts where we had predicted and the surrounding sequences are virtually without insertions as designed (III; Figure 6B). Thus, the attractiveness of a pentamer can be predicted with relatively high accuracy by additively calculating the base-step values over the 5-bp target site and new target molecules can be designed for various purposes. For example, target fragments with only one major target site can be designed. This type of a target will be essential in protein-DNA cross-linking experiments for further probing the contacts between the transpososome and the target site, and also for target capture complex crystallisation.

3.6. The effect of the sequences surrounding the target pentamer core

No consensus has previously been detected beyond the duplicated target site (Mizuuchi and Mizuuchi, 1993). This does not mean, however, that the sequences surrounding the target pentamer can not influence target site selection. We had multiple observations in different surroundings of 73 pentamers, which showed clear variations in their intensities (III; Figure 3E). The analysis of variance revealed that 38 % of the total variance was due to the surroundings (i.e. variation between multiple observations within each pentamer) and 62 % to the pentamer itself (i.e. variation of pentamer intensity means). In this approach, the variation caused by the environment is assumed to be equal for all pentamers. Approximately 10 % of the surrounding component variation (i.e. 3.8 % of the total variation) was estimated to originate from variation in gel quality and pipetting accuracy. The calculation is based on the data shown in Figure 2C (III).

3.7. The structural properties of the target DNA

The importance of the structural properties of DNA at the target site and around it was further analysed with a program described by Liao et al. (2000). Four different measures of physical properties of DNA: protein induced deformability (Olson et al., 1998), B-DNA twist (Gorin et al., 1995), A-philicity (Ivanov and Minchenkova, 1995), and propeller twist (El Hassan and Calladine, 1996) as well as GC-content were calculated for aligned 35-bp sequences (n=77). All four physical pa-

rameters studied showed notable symmetrical signals at the 5-bp target site extending to the flanking DNA (III; Figure 7, from second to fifth panel). A random data set lacks the patterns seen in the insertion site data set. Symmetrical alteration in the GC/AT preference was also noted in the target area extending to the flanks (III; Figure 7, upper panel). Because no clear sequence preferences were seen in the surrounding sequences, this suggests that the transposase-DNA contacts could be made with the backbone phosphate groups and/or sugar residues instead of direct base contacts.

Interestingly, the pyrimidine-purine dimers are most easily deformed in protein-DNA complexes and they can act as flexible “hinges”, fitting the DNA to the protein surface (Olson et al., 1998). In our analysis, three peaks were seen in the protein-induced deformability (III; Figure 7, second panel). The periodicities of these peaks (9-10 nt) coincide well with the B-DNA helical pitch (10.5-bp), and thus may be indicative of anisotropic bending. Previously, DNA bending has been proposed for the IS231A target site (Hallet et al., 1994) and the retroviral insertion sites (Muller and Varmus, 1994).

Target recognition probably involves fitting of the target DNA to the protein active site through optimisation of the contacts made. This can be an additive process, suggested for IHF by Rice et al. (1996) and for Tn10 transposase target recognition by Pribil and Haniford (2000). Transpososome binding to the consensus base pairs can facilitate a kink formation, and this in turn would permit additional contacts to the flanking DNA. In Mu target site recognition, the three putative protein-induced deformability areas could act as

hinges in a similar, additive, sequential process. If the duplicated pentamer is surrounded by a poor sequence, which is unable to form suitable hinges, part of the potential contacts to the flanking DNA are omitted. Thus, the transpososome would be more prone to dissociate and locate to a new target site. However, this is hypothetical, and structural information from target capture complexes is required to clarify how Mu transpososome recognises the target DNA.

3.8. The extent of the “surrounding sequence effect”

Because the flanking sequences were found to contribute significantly to the attractiveness of the target pentamers, either there must be protein contacts with a wider area around the duplicated pentamer, or the sequence context must affect the conformation of the target pentamer (e.g. widening or tightening the B-DNA grooves). This demands further studies, but some estimates on the length of this effective surrounding area can be provided by our data. The symmetrical patterns in the structural parameters analysed (III; Figure 7) extended some 9-10 nucleotides to the surrounding areas. Also, parallel results have been seen in experiments in which we tested how close to the DNA edge Mu can transpose (III; data not shown). Together, the observations suggest that the Mu transpososome can make contacts with, or is influenced by, a target covering a region up to a 23-25 -bp. A broad zone of contacts (~24-bp) has been detected with contact-probing experiments between the transpososome and the target DNA in the target capture complex of Tn10 (Pribil and Haniford, 2000). Longer sequence motifs with preferred sequences or puta-

tive sequence features have been detected for IS903, P-element and Tn3 (Hu et al., 2001; Liao et al., 2000; Davies and Hutchinson, 1995). Thus, a broad zone of contacts may be a general feature in transpososome-target recognition.

3.9. A general picture of the target site selection

Regionally preferred sites have been found to be clustered next to, but not within, the area of MuB footprint (Mizuuchi and Mizuuchi, 1993). These results were most convincingly explained via a protein-protein interaction; MuB would bind some regions of the target DNA preferentially, deliver transpososomes to the region, activate the complex for strand transfer, and then allow the complex to diffuse along the target DNA before strand transfer (Mizuuchi and Mizuuchi, 1993). If this was the case, MuA could then choose the exact sequence level target site, which would explain the similar consensus sequence results obtained from experiments either with or without MuB (Mizuuchi and Mizuuchi, 1993; III). Our experiments have provided further proof that MuA is able to find suitable targets and bind target DNA without contacting MuB.

F. CONCLUSIONS AND FUTURE PERSPECTIVES

This study described a new methodology for molecular biology applications based on minimal *in vitro* bacteriophage Mu transposition reaction (I). The described minimal Mu *in vitro* transposition system was shown to work efficiently and with low target site selectivity (I, II). The target site duplication of the genuine *in vivo* transposition reaction was reproduced accurately in the minimal *in vitro* reaction (I, II). In addition, the Mu transposon tool is simple since it is composed of only a few components. Applicability of the Mu system in a functional genetic analysis was demonstrated using restriction analysis and PCR-based genetic footprinting strategies (I). In addition, the described methodology was shown to be easily modifiable and applicable to efficient and easy DNA sequencing template generation (II).

The nucleotide level target site selection by MuA alone was further studied with a new type of quantitative assay (III). The quantified data was analysed for the most favoured and disfavoured target sequences as base pairs, as well as base pair steps that dictate the sequence-related DNA structure. The results revealed a new, more structurally oriented perspective on target site selection, providing a new means to predict preferences of target pentamers. The study further demonstrated for the first time the impact of the surrounding sequences on the target site selection by MuA, and detected symmetrical patterns in several structural parameters of DNA spanning a ~23-24-bp region around the target pentamer. These results suggest a large DNA binding area for the transpososome and potential hinge regions

in the target DNA.

In the future, protein-DNA cross-linking experiments and structural information obtained from target capture complexes will be a necessity for finally uncovering the details of Mu insertion site selection. For these studies it will be necessary to design a target DNA molecule, which has a single major target site (such as is in III, Figure 6B). Additionally, our quantitative nucleotide level assay provides the opportunity to study the regional level target site selection in conditions where MuB is present. The ability to predict potential insertion sites can be attractive for projects utilising Mu transposition as a tool for sequencing or generating insertion mutant pools. It would also be interesting to make computational predictions or molecular models that would also take into account the structural preferences at the target core and in the flanking sequences.

The work conducted during this study has led to a commercial product for DNA sequencing (TGS: Template Generation System, Finnzymes). Recently, a second product for insertional mutagenesis based on Mu transposition has become available (MGS: Mutation Generation System, Finnzymes). The Mu *in vitro* transposon tool has already been applied to the analysis of the HIV type-1 genome, the *Pseudomonas syringae hrpA* gene encoding the Hrp pilus subunit, the construction of gene-targeting vectors for transgenic studies and for the analysis of the potato virus A genome (Laurent et al., 2000; Taira et al., 1999; Vilen et al., 2001; Kekarainen et al., in preparation). In the future, Mu technology is expected to be useful in a variety of novel applications in the functional analysis of genes, genomes and proteins.

G. ACKNOWLEDGEMENTS

This work was carried out in the DNA Recombination Laboratory at the Institute of Biotechnology, Research program in Cellular Biotechnology and at the Department of Biosciences, Division of Genetics, University of Helsinki. I wish to express my sincere gratitude to professor Mart Saarma, the Director of the Institute of Biotechnology, for providing me with excellent facilities for this work. I wish to express my appreciation to the present and former substitute heads of the Department of Genetics, Professor Hannu Saari-Lahti and Pekka Heino, for their help in coping with the PhD bureaucracy.

I wish to express my gratitude to my supervisor, Docent Harri Savilahti. He once selected me as his first PhD student and has given me an interesting subject to study. I wish to thank Harri for providing guidance and the facilities that have made this work possible.

I thank Professors Dennis Bamford and Alan Schulman for reviewing this thesis so quickly and giving insightful suggestions for improvement.

The head of the Helsinki Graduate School in Biotechnology and Molecular Biology, Professor Heikki Rauvala, is acknowledged for providing the major part of the funding for my research, as well as for interesting courses and symposia. Additional financial support obtained from the Finnish Cultural Foundation and from the Jenny and Antti Wihuri Foundation is gratefully acknowledged.

I wish to express my warm gratitude to Dennis Bamford and Teemu Teeri, who have shown the characteristics of experienced mentors. They have followed the progression of my PhD studies over the years in "grad"

school, provided sympathy and good advice.

I also wish to thank my collaborators who have taken part in the completion of the three publications: Suvi Taira, Lars Paulin, Sini Suomalainen, Simo Eerikäinen, Matti Airaksinen and Hannu Rita.

Anna-Helena Saariaho, Eini Heikkinen, Sari Nieminen, Heikki Vilen, Pirjo Rahkola and Hilikka Turakainen, the present members of our group, are all warmly thanked for all the fun times, support and help you have given me during these years. Also, an equally big thanks goes to the former members of our group: Arja Lamberg, Suvi Taira, Auli Saarinen, Juha-Matti Aalto, Anna Kassinen and Mingqiang Qiao. An especially warm hug goes to Ansku and Eini for sharing all the ups and downs, and having discussions about everything from future dreams to gardening and canines.

A special warm hug goes to Pia Pauloff for her valuable friendship, which has lasted almost a quarter of a century. A big, big thanks goes to Pia, Nippe, Anna, Johanna, Maki and Maarit for all the fun & fashion filled "girlie nights", which have reminded me to take life less seriously. For all the fun and relaxing times together in Harjakari, I wish to thank Terhi, Ripa, Arja, Matti, Petteri, Tuke, Tiina, Kirsi, Jarkko and the occasional guest stars.

I am deeply grateful to my parents, Anja and Kauko, for their endless love, support and encouragement. Pai also deserves a tender scratch under her muzzle for giving me exercise and fresh-air whether or not I felt like it.

Finally, all my love goes to Vesku.
Helsinki, December 2001

Saija

H. REFERENCES

- Adachi T., Mizuuchi M., Robinson E.A., Appella E., O'Dea M.H., Gellert M., and Mizuuchi K. (1987) DNA sequence of the *E. coli* *gyrB* gene: application of a new sequencing strategy. *Nucleic Acids Res.*, **15**, 771-784.
- Adzuma K., and Mizuuchi K. (1988) Target immunity of Mu transposition reflects a differential distribution of MuB protein. *Cell*, **53**, 257-266.
- Adzuma K., and Mizuuchi K. (1989) Interaction of proteins located at a distance along DNA: mechanism of target immunity in the Mu DNA strand-transfer reaction. *Cell*, **57**, 41-47.
- 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.
- Akerley B.J., Rubin E.J., Camilli A., Lampe D.J., Robertson H.M., and Mekalanos J.J. (1998) Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc. Natl. Acad. Sci. USA*, **95**, 8927-8932.
- Aldaz H., Schuster E., and Baker T.A. (1996) The interwoven architecture of the Mu transposase couples DNA synthesis to catalysis. *Cell*, **85**, 257-269.
- Alleman, R.K., and Egli, M. (1997) DNA recognition and bending. *Chemistry & Biology*, **4**, 643-650.
- Allet B. (1979) Mu insertion duplicates a 5 base pair sequence at host insertion site. *Cell*, **16**, 123-129.
- Allingham J.S., Pribil P.A., and Haniford D.B. (1999) All three residues of the Tn10 transposase DDE catalytic triad function in divalent metal ion binding. *J. Mol. Biol.*, **289**, 1195-1206.
- Allingham J.S., Wardle S.J., and Haniford D.B. (2001) Determinants for hairpin formation in Tn10 transposition. *EMBO J.*, **20**, 2931-42.
- Bainton R.J., Kubo K.M., Feng J.N., and Craig N.L. (1993) Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. *Cell*, **72**, 931-43.
- Baker T.A. (1995) Bacteriophage Mu: a transposing phage that integrates like retroviruses. *Seminars in virology*, **6**, 53-63.
- Baker T.A., and Luo L. (1994) Identification of residues in the Mu transposase essential for catalysis. *Proc. Natl. Acad. Sci. USA*, **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., 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., 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.
- Bender J., and Kleckner N. (1986) Genetic evidence that Tn10 transposes by a nonreplicative mechanism. *Cell*, **45**, 801-15.
- Bender J., and Kleckner N. (1992) Tn10 insertion specificity is strongly dependent upon sequences immediately adjacent to the target-site consensus sequence. *Proc. Natl. Acad. Sci. USA*, **89**, 7996-8000.
- Benjamin H.W., and Kleckner N. (1992) Excision of Tn10 from the donor site during transposition occurs by flush double-strand cleavages at the transposon termini. *Proc. Natl. Acad. Sci. USA*, **89**, 4648-4652.
- Berg C.M., and Berg D.E. (1995) Transposable elements as tools for molecular analyses in bacteria. In Sherrat D.J. (ed.), *Molecular genetic elements; Frontiers in molecular biology*. Oxford University Press, Oxford, UK, pp. 38-68.
- Berg C.M., Wang G., Strausbaugh L.D., and Berg D.E. (1993) Transposon-facilitated sequencing of DNAs cloned in plasmids. *Methods Enzymol.*, **218**, 279-306.
- Berg, D.E., and Howe, M.M. (1989) *Mobile DNA*. American Society for Microbiology, Washington DC.
- Bhasin A., Goryshin I.Y., and Reznikoff W.S. (1999) Hairpin formation in Tn5 transposition. *J. Biol. Chem.*, **274**, 37021-37029.

- Biery M.C., Lopata M., and Craig N.L. (2000a) A minimal system for Tn7 transposition: the transposon-encoded proteins TnsA and TnsB can execute DNA breakage and joining reactions that generate circularized Tn7 species. *J. Mol. Biol.*, **297**, 25-37.
- Biery M.C., Stewart F.J., Stellwagen A.E., Raleigh E.A., and Craig N.L. (2000b) A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucleic Acids Res.*, **28**, 1067-1077.
- Boeke J.D., and Stoye J.P. (1997) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In Coffin J.M., Hughes S.H. and Varmus H.E. (eds.), *Retroviruses*. Cold Spring Harbour Press, New York, USA, pp. 343-436.
- Boeke, J.D., and Devine, S.E. (1998) Yeast retrotransposons: finding quiet neighborhood. *Cell*, **93**, 1087-1089.
- 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.
- Bor Y.-C., Miller M.D., Bushman F.D., and Orgel L.E. (1996) Target-sequence preferences of HIV-1 integration complexes in vitro. *Virology*, **222**, 283-288.
- Braunstein M., Griffin T.J. IV, Kriakov J.I., Friedman S.T., Grindley N.D.F., and Jacobs W.R. (2000) Identification of genes encoding exported Mycobacterium tuberculosis proteins using a Tn552'phoA in vitro transposition system. *J. Bacteriol.*, **182**, 2732-2740.
- 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.
- Bukhari A.I., Froshauer S., and Botchan M. (1976) Ends of bacteriophage Mu DNA. *Nature*, **264**, 580-583.
- Bukhari, A.I., and Zipser, D. (1972) Random insertion of Mu-1 DNA within a single gene. *Nature New Biol.*, **236**, 240-243.
- 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 λ :mini-Mu phage. *Proc. Natl. Acad. Sci. USA*, **83**, 6012-6016.
- Burton B.M., Williams T.A., and Baker T.A. (2001) ClpX-mediated remodelling of Mu transpososomes: selective unfolding of subunits destabilizes the entire complex. *Mol. Cell*, **8**, 449-454.
- Casadaban M.J., and Chou J. (1984) In vivo formation of gene fusions encoding hybrid beta-galactosidase proteins in one step with a transposable Mu-lac transducing phage. *Proc. Natl. Acad. Sci. USA*, **81**, 535-539.
- Casadaban M.J., and Cohen S.N. (1979) Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: in vivo probe for transcriptional control sequences. *Proc. Natl. Acad. Sci. USA*, **76**, 4530-4533.
- Castilho, B.A., and Casadaban, M.J. (1991) Specificity of mini-Mu bacteriophage insertions in a small plasmid. *J. Bacteriol.*, **173**, 1339-1343.
- Chaconas G., Harshey R.M., Sarvetnick N., and Bukhari A.I. (1981) Predominant end-products of prophage Mu DNA transposition during the lytic cycle are replicon fusions. *J. Mol. Biol.*, **150**, 341-59.
- Chaconas G., Kennedy D.L., and Evans D. (1983) Predominant integration end products of infecting bacteriophage Mu DNA are simple insertions with no preference for integration of either Mu DNA strand. *Virology*, **128**, 48-59.
- Chaconas G., Lavoie B.D., and Watson M.A. (1996) DNA transposition: Jumping gene machine, some assembly required. *Current Biology*, **6**, 817-820.
- Coffin J.M., Hughes S.H., and Varmus H.E. (1997) *Retroviruses*. Cold Spring Harbour Press, New York, USA.
- 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.
- Couturier M. (1976) The integration and excision of the bacteriophage Mu-1. *Cell*, **7**, 155-63.
- Craig N.L. (1991) Tn7: a target site-specific transposon. *Mol. Microbiol.*, **5**, 2569-2573.
- Craig N.L. (1995) Unity in transposition reactions. *Science*, **270**, 253-254.
- Craig N.L. (1996) Transposon Tn7. *Curr. Topics Microbiol. Immunol.*, **204**, 27-48.
- Craig N.L. (1997). Target site selection in trans-

- position. *Annu. Rev. Biochem.*, **66**, 437-74.
- Craigie R. (1996) Quality control in Mu DNA transposition. *Cell*, **85**, 137-140.
- Craigie R., and Mizuuchi K. (1985a) Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. *Cell*, **41**, 867-876.
- Craigie R., and Mizuuchi K. (1985b) Cloning of the A gene of bacteriophage Mu and purification of its product, the Mu transposase. *J. Biol. Chem.*, **260**, 1832-1835.
- 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., 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. USA*, **82**, 7570-7574.
- 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.
- Daniell E., Roberts R., and Abelson J. (1972) Mutations in the lactose operon caused by bacteriophage Mu. *J. Mol. Biol.*, **69**, 1-8.
- Darzins A., Kent N.E., Buckwalter M.S., and Casadaban M.J. (1988) Bacteriophage Mu sites required for transposition immunity. *Proc. Natl. Acad. Sci. USA*, **85**, 6826-6830.
- Das A., and Xie Y.H. (1998) Construction of transposon Tn3phoA: its application in defining the membrane topology of the *Agrobacterium tumefaciens* DNA transfer proteins. *Mol. Microbiol.*, **27**, 405-14.
- Davies C.J., and Hutchinson C.A. III (1995) Insertion site specificity of the transposon Tn3. *Nucleic Acids Res.*, **23**, 507-514.
- Davies D.R., Goryshin I.Y., Reznikoff W.S., and Rayment I. (2000) Three-dimensional structure of Tn5 synaptic complex transposition intermediate. *Science*, **289**, 77-85.
- DeBoy R.T., and Craig N.L. (1996) Tn7 transposition as a probe of cis interactions between widely separated (190 kilobases apart) DNA sites in the *Escherichia coli* chromosome. *J. Bacteriol.*, **178**, 6184-91.
- DeBoy R.T., and Craig N.L. (2000) Target site selection by Tn7: attTn7 transcription and target activity. *J. Bacteriol.*, **182**, 3310-3313.
- Devine S.E., and Boeke J.D. (1994) Efficient integration of artificial transposons into plasmid targets in vitro: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res.*, **22**, 3765-3772.
- Devine S.E., Chisoe S.L., Eby Y., Wilson R.K., and Boeke J.D. (1997) A transposon based strategy for sequencing repetitive DNA in eukaryotic genomes. *Genome Res.*, **7**, 551-563.
- 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.
- Ehrmann M., Bolek P., Mondigler M., Boyd D., and Lange R. (1997) TnTIN and TnTAP: mini-transposons for site-specific proteolysis in vivo. *Proc. Natl. Acad. Sci. USA*, **94**, 13111-5.
- Eichinger D.J., and Boeke J.D. (1988) The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. *Cell*, **54**, 955-966.
- El Hassan M.A., and Calladine C.R. (1996) Propeller-Twisting of Base-pairs and the Conformational Mobility of Dinucleotide Steps in DNA. *J. Mol. Biol.*, **259**, 95-103.
- Emr, S.D., and Silhavy, T.J. (1980) Mutations affecting localisation of an *Escherichia coli* outer membrane protein, the bacteriophage I receptor. *J. Mol. Biol.*, **141**, 63-90.
- Fedoroff N. (2000) Transposons and genome evolution in plants. *Proc. Natl. Acad. Sci. USA*, **97**, 7002-7.
- Flavell A.J., Pearce S.R., and Kumar A. (1994) Plant transposable elements and the genome. *Curr. Opin. Genet. Dev.*, **4**, 838-844.

- Gehring A.M., Nodwell J.R., Beverley S.M., and Losick R. (2000) Genomewide insertional mutagenesis in *Streptomyces coelicolor* reveals additional genes involved in morphological differentiation. *Proc. Natl. Acad. Sci. USA.*, **97**, 9642-7.
- Goldgur Y., Dyda F., Hickman A.B., Jenkins T.M., Craigie R., and Davies D. (1998) Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. *Proc. Natl. Acad. Sci. USA.*, **95**, 9150-9154.
- Gorin A.A., Zhurkin V.B., and Olson W.K. (1995) B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.*, **247**, 34-48.
- Goryshin I.Y., and Reznikoff W.S. (1998) Tn5 in vitro transposition. *J. Biol. Chem.*, **273**, 7367-7374.
- Goryshin I.Y., Jendrisak J., Hoffman L.M., Meis R., and Reznikoff W.S. (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nature Biotech.*, **18**, 97-100.
- Goryshin I.Y., Miller J.A., Kil Y.V., Lanzov V.A., and Reznikoff W.S. (1998) Tn5/IS50 target recognition. *Proc. Natl. Acad. Sci. USA.*, **95**, 10716-10721.
- Griffin T.J. 4th, Parsons L., Leschziner A.E., DeVost J., Derbyshire K.M., and Grindley N.D. (1999) In vitro transposition of Tn552: a tool for DNA sequencing and mutagenesis. *Nucleic Acids Res.*, **27**, 3859-65.
- Groisman E.A., and Casadaban M.J. (1986) Mini-mu bacteriophage with plasmid replicons for in vivo cloning and lac gene fusing. *J. Bacteriol.*, **168**, 357-364.
- Groisman E.A., and Casadaban M.J. (1987) Cloning of genes from members of the family Enterobacteriaceae with mini-Mu bacteriophage containing plasmid replicons. *J. Bacteriol.*, **169**, 687-693.
- Groisman E.A., Castilho B.A., and Casadaban M.J. (1984) In vivo DNA cloning and adjacent gene fusing with mini-Mu-lac bacteriophage containing a plasmid replicon. *Proc. Natl. Acad. Sci. USA.*, **81**, 1480-1483.
- Groisman E.A., Pagratis N., and Casadaban M.J. (1991) Genome mapping and protein coding region identification using bacteriophage Mu. *Gene*, **99**, 1-7.
- Gwinn M.L., Stellwagen A.E., Craig N.L., Tomb J.F., and Smith H.O. (1997) In vitro Tn7 mutagenesis of *Haemophilus influenzae* Rd and characterization of the role of atpA in transformation. *J. Bacteriol.*, **179**, 7315-20.
- Haapa S., Suomalainen S., Eerikainen S., Airaksinen M., Paulin L., and Savilahti H. (1999b) An efficient DNA sequencing strategy based on the bacteriophage Mu in vitro DNA transposition reaction. *Genome Res.*, **9**, 308-315.
- Haapa S., Taira S., Heikkinen E., and Savilahti H. (1999a) An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res.*, **27**, 2777-2784.
- Hallet B., Rezsöházy R., Mahillon J., and Delcour J. (1994) IS231A insertion specificity: consensus sequence and DNA bending at the target site. *Mol. Microbiol.*, **14**, 131-139.
- Hallet B., Sherrat D. J., and Hayes F. (1997) Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in target protein. *Nucleic Acids Res.*, **25**, 1866-1867.
- Halling, S.M., and Kleckner, N. (1982) A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. *Cell*, **28**, 155-163.
- Hamer L., DeZwaan T.M., Montenegro-Chamorro M.V., Frank S.A., and Hamer J.E. (2001) Recent advances in large-scale transposon mutagenesis. *Curr. Opin. Chem. Biol.*, **5**, 67-73.
- Haniford D.B., and Chaconas G. (1992) Mechanistic aspects of DNA transposition. *Curr. Opin. Genet. Dev.*, **2**, 698-704.
- Haren L., Ton-Hoang B., and Chandler M. (1999) Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.*, **53**, 245-281.
- Harshey R.M. (1984) Transposition without duplication of infecting bacteriophage Mu DNA. *Nature*, **311**, 580-581.
- Harshey R.M. (1987) Integration of infecting Mu DNA. In Symonds N., Toussaint A., van de Putte, P., and Howe M.M. (eds.), *Phage Mu*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New

- York, USA, pp. 111-135.
- Hauer B., and Shapiro J.A. (1984) Control of Tn7 transposition. *Mol. Gen. Genet.*, **194**, 149-58.
- Hayes F., and Hallet B. (2000) Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks. *Trends Microbiol.*, **8**, 571-7.
- Hickman A.B., Li Y., Mathew S.V., May E.W., Craig N.L., and Dyda F. (2000) Unexpected diversity in DNA recombination: the restriction endonuclease connection. *Mol. Cell*, **5**, 1025-1034.
- Hindmarsh P., and Leis J. (1999) Retroviral DNA integration. *Microbiol. Mol. Biol. Rev.*, **63**, 836-43.
- 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.
- Hoekstra M.F., Burbee D., Singer J., Mull E., Chiao E., and Heffron F. (1991) A Tn3 derivative that can be used to make short in-frame insertions within genes. *Proc. Natl. Acad. Sci. USA*, **88**, 5457-61.
- Hoffman L.M., Jendrisak J.J., Meis R.J., Goryshin I.Y., and Reznikof S.W. (2000) Transposome insertional mutagenesis and direct sequencing of microbial genomes. *Genetica*, **108**, 19-24.
- Horie K., Kuroiwa A., Ikawa M., Okabe M., Kondoh G., Matsuda Y., and Takeda J. (2001) Efficient chromosomal transposition of a Tc1/mariner-like transposon Sleeping Beauty in mice. *Proc. Natl. Acad. Sci. USA*, **98**, 9191-9196.
- Howe M.M. (1987) Phage Mu: An overview. In Symonds N., Toussaint A., Van De Putte P., and Howe M.M. (eds.), *Phage Mu*. Cold Spring Harbour Laboratory, New York, USA, pp. 25-40.
- Hu W.-Y., and Derbyshire K.M. (1998) Target choice and orientation preference of the insertion sequence IS903. *J. Bacteriol.*, **180**, 3039-3048.
- Hu W.-Y., Thompson W., Lawrence C.E., and Derbyshire K.M. (2001) Anatomy of a preferred target site for the bacterial insertion sequence IS903. *J. Mol. Biol.*, **306**, 403-416.
- Hurst G.D., and Werren J.H. (2001) The role of selfish genetic elements in eukaryotic evolution. *Nat. Rev. Genet.*, **2**, 597-606.
- Hutchison C.A., Peterson S.N., Gill S.R., Cline R.T., White O., Fraser C.M., Smith H.O., and Venter J.C. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science*, **286**, 2165-9.
- Ivanov V.I., and Minchenkova L.E. (1995) The A-form of DNA: in search of biological role. *Mol. Biol.*, **28**, 780-788.
- Jones J., 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. (1997) Factors responsible for target site selection in Tn10 transposition: a role for the DDE motif in target DNA capture. *EMBO J.*, **16**, 2646-2655.
- Kahman R., and Kamp D. (1979) Nucleotide sequences of the attachment sites of bacteriophage Mu. *Nature*, **280**, 247-250.
- Kamp D., and Kahman R. (1980) Two pathways in bacteriophage Mu transposition? *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 329-336.
- Kasai H., Isono S., Kitakawa M., Mineno J., Akiyama H., Kurnit D.M., Berg D.E., and Isono K. (1992) Efficient large-scale sequencing of the Escherichia coli genome: implementation of a transposon- and PCR-based strategy for the analysis of ordered lambda phage clones. *Nucleic Acids Res.*, **20**, 6509-15.
- Katz R.A., Gravuer K., and Skalka A.M. (1998) A Preferred Target DNA Structure for Retroviral Integrase in Vitro. *J. Biol. Chem.*, **273**, 24190-24195.
- Kekarainen T., Savilahti H., and Valkonen J.P.T. Functional genomics on *Potato virus A*: a virus genome-wide map of sites essential for replication. In preparation.
- Kennedy A.K., and Haniford D.B. (1996) Isolation and characterization of IS10 transposase separation of function mutants: identification of amino acid residues in transposase that are important for active site function and the stability of transposition intermediates. *J. Mol. Biol.*, **256**, 533-47.

- 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.
- Ketting R.F., Fisher S.E.J., and Plasterk R.H.A. (1997) Target choice determinants of the Tc1 transposon of *Caenorhabditis elegans*. *Nucleic Acids Res.*, **25**, 4041-4047.
- Kleckner N. (1981) Transposable elements in prokaryotes. *Annu. Rev. Genet.*, **15**, 341-404.
- Kleckner N., Chan R.K., Tye B.K., and Botstein D. (1975) Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. *J. Mol. Biol.*, **97**, 561-75.
- 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.
- Krukltis 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**, 16469-77.
- Krukltis 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.
- Kubo K.M., and Craig N.L. (1990) Bacterial transposon Tn7 utilizes two different classes of target sites. *J. Bacteriol.*, **172**, 2774-2778.
- Kuduvalli P.N., Rao J.E., and Craig N.L. (2001) Target DNA structure plays a critical role in Tn7 transposition. *EMBO J.*, **20**, 924-932.
- Lamberg A., Nieminen S., Qiao M., and Savilahti H. Efficient insertional mutagenesis strategy for bacterial genomes by electroporation of *in vitro* assembled DNA transposition complexes of bacteriophage Mu. *Appl. Environ. Microbiol.* In Press.
- Lampe D.J., Churchill M.E., and Robertson H.M. (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.*, **15**, 5470-9.
- Laurent L.C., Olsen M.N., Crowley R.A., Savilahti H., and Brown P.O. (2000) Functional characterization of the human immunodeficiency virus type 1 genome by genetic footprinting. *J. Virol.*, **74**, 2760-9.
- Lavoie B.D., and Chaconas G. (1993) Site-specific HU binding in the Mu transpososome: conversion of a sequence-independent DNA-binding protein into 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., and Chaconas G. (1996) Transposition of phage Mu DNA. *Curr. Top. Microbiol. Immunol.*, **204**, 83-102.
- 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**, 761-71.
- Leach D.R.F. (1996) *Genetic recombination*. Blackwell Science LTD, Oxford, UK.
- Lee C.H., Bhagwat A., and Heffron F. (1983) Identification of a transposon Tn3 sequence required for transposition immunity. *Proc. Natl. Acad. Sci. USA*, **80**, 6765-9.
- Leschziner A.E., Griffin T.J. IV, and Grindley N.D.F. (1998) Tn522 transposase catalyzes concerted strand transfer in vitro. *Proc. Natl. Acad. Sci. USA*, **95**, 7345-7350.
- Leung P.C., and Harshey R.M. (1991) Two mutations of phage mu transposase that affect strand transfer or interactions with B protein lie in distinct polypeptide domains. *J. Mol. Biol.*, **219**, 189-99.
- 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 en-

- hancer. *Nature*, **338**, 656-658.
- Liao G-C., Rehm E.J., and Rubin G.M. (2000) Insertion site preferences of the P transposable element in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*, **97**, 3347-3351.
- Lilley D.M.J. (1995) *DNA-protein: structural interactions*. Oxford University Press, New York.
- Lu F., and Craig N.L. (2000) Isolation and characterization of Tn7 transposase gain-of-function mutants: a model for transposase activation. *EMBO J.*, **19**, 3446-3457.
- Mahillon J., and Chandler M. (1998) Insertion sequences. *Microbiol. Mol. Biol. Rev.*, **62**, 725-774.
- Manna D., and Higgins P.N. (1999) Phage Mu transposition immunity reflects supercoil domain structure of the chromosome. *Mol. Microbiol.*, **32**, 595-606.
- Manna D., Wang X., and Higgins N.P. (2001) Mu and IS1 transposition exhibit strong orientation bias at the *Escherichia coli bgl* locus. *J. Bacteriol.*, **183**, 3328-3335.
- 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. USA*, **84**, 699-703.
- May E.W., and Craig N.L. (1996) Switching from cut-and-paste to replicative Tn7 transposition. *Science*, **272**, 401-4.
- McClintock B. (1987) *The discovery and characterization of transposable elements: the collected papers of Barbara McClintock*. Garland, New York.
- Mizuuchi K. (1983) In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. *Cell*, **35**, 785-794.
- Mizuuchi K. (1992) Transpositional recombination: Mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.*, **61**, 1011-51.
- Mizuuchi K. (1997) Polynucleotide transfer reactions in site-specific DNA recombination. *Genes to cells*, **2**, 1-12.
- Mizuuchi K., and Adzuma K. (1991) Inversion of the phosphate chirality at the target site of Mu DNA strand transfer: evidence for one-step transesterification mechanism. *Cell*, **66**, 129-140.
- Mizuuchi K., Nobbs T.J., Halford S.E., Adzuma K., and Qin J. (1999) A new method for determining the stereochemistry of DNA cleavage reactions: application to the SfiI and HpaII restriction endonucleases and to the MuA transposase. *Biochemistry*, **38**, 4640-4648.
- 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., Baker T.A., and Mizuuchi K. (1991) DNase protection analysis of the stable synaptic complexes involved in Mu transposition. *Proc. Natl. Acad. Sci. USA*, **88**, 9031-5.
- 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.
- Morgan B.A., Conlon F.L., Manzanares M., Millar J.B., Kanuga N., Sharpe J., Krumlauf R., Smith J.C., and Sedgwick S.G. (1996) Transposon tools for recombinant DNA manipulation: characterization of transcriptional regulators from yeast, *Xenopus*, and mouse. *Proc. Natl. Acad. Sci. USA*, **93**, 2801-6.
- Morisato D., and Kleckner N. (1984) Transposase promotes double strand breaks and single strand joints at Tn10 termini in vivo. *Cell*, **39**, 181-190.
- Muller H-P., and Varmus H.E. (1994) DNA bending creates favoured sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.*, **13**, 4704-4714.
- Naigamwalla D.Z., and Chaconas G. (1997) A New set of Mu DNA transposition intermediates: Alternate pathways of target capture preceding strand transfer. *EMBO J.*, **16**, 5227-5234.
- Naigamwalla D.Z., Coros C.J., Wu Z., and Chaconas G. (1998) Mutations in domain IIIa of Mu transposase: evidence suggesting an active site component which interacts with the Mu-host junction. *J. Mol. Biol.*, **282**, 265-274.
- Nakai H., and Kruklytis R. (1995) Disassembly of the bacteriophage Mu transposase

- for the initiation of Mu DNA replication. *J. Biol. Chem.*, **270**, 19591-8.
- Nakai H., Doseeva V., and Jones J.M. (2001) Handoff from recombinase to replisome: Insights from transposition. *Proc. Natl. Acad. Sci. USA*, **98**, 8247-8254.
- Nakayama C., Teplow D.B., and Harshey R. (1987) Structural domains in bacteriophage 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.
- Olson W.K., Gorin A.A., Lu X.-J., Hock L.M., and Zhurkin V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA*, **95**, 11163-11168.
- Payne J.A., Stevenson T.J., and Donaldson L.F. (1996) Molecular characterization of a putative K-Cl cotransporter in rat brain. *J. Biol. Chem.*, **271**, 16245-16252.
- Pellic V., Morelle S., Lampe D., and Nassif X. (2000) Mutagenesis of *Neisseria meningitidis* by in vitro transposition of Himar1 mariner. *J. Bacteriol.*, **182**, 5391-8.
- Peters J.E., and Craig N.L. (2000) Tn7 transposes proximal to DNA double-strand breaks and into regions where chromosomal DNA replication terminates. *Mol. Cell*, **6**, 573-582.
- Peters J.E., and Craig N.L. (2001) Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev.*, **15**, 737-747.
- Phadnis S.H., Huang H.V., and Berg D.E. (1989) Tn5supF, a 264-base-pair transposon derived from Tn5 for insertion mutagenesis and sequencing DNAs cloned in phage lambda. *Proc. Natl. Acad. Sci. USA*, **86**, 5908-12.
- Polard P., and Chandler M. (1995) Bacterial transposases and retroviral integrases. *Mol. Microbiol.*, **15**, 13-23.
- Pribil, P.A., and Haniford, D.B. (2000) Substrate recognition and induced DNA deformation by transposase at the target-capture stage of Tn10 transposition. *J. Mol. Biol.*, **303**, 145-159.
- Raibaud O., Roa M., Braun-Bretton C., and Schwartz M. (1979) Structure of the malB region in *Escherichia coli* K12. I Genetic map of the malK-lamB operon. *Mol. Gen. Genet.*, **179**, 241-248.
- Rao J.E., Miller P.S., and Craig N.L. (2000) Recognition of triple-helical DNA structures by transposon Tn7. *Proc. Natl. Acad. Sci. USA*, **97**, 3936-41.
- Reich K.A., Chovan L., and Hessler P. (1999) Genome scanning in *Haemophilus influenzae* for identification of essential genes. *J. Bacteriol.*, **181**, 4961-4968.
- 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. *Nature Struct. Biol.*, **8**, 302-307.
- Rice P.A., Yang S.-W., Mizuuchi K., and Nash H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA turn. *Cell*, **87**, 1295-1306.
- Roldan L.A.S., and Baker T.A. (2001) Differential role of the MuB protein in phage Mu integration vs. replication: mechanistic insights into two transposition pathways. *Mol. Microbiol.*, **40**, 141-155.
- Ross-Macdonald P., Coelho P.S., Roemer T., Agarwal S., Kumar A., Jansen R., Cheung K.H., Sheehan A., Symoniatis D., Umansky L., Heidtman M., Nelson F.K., Iwasaki H., Hager K., Gerstein M., Miller P., Roeder G.S., and Snyder M. (1999a) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature*, **402**, 413-8.
- Ross-Macdonald P., Sheehan A., Friddle C., Roeder G.S., and Snyder M. (1999b)

- Transposon mutagenesis for the analysis of protein production, function, and localization. *Methods Enzymol.*, **303**, 512-32.
- Ross-Macdonald P., Sheehan A., Roeder G.S., and Snyder M. (1997) A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*, **94**, 190-5.
- Saedler H., and Gierl A. (1996) Transposable elements. *Curr. Top. Microbiol. Immunol.*, **204**.
- Sakai J., and Kleckner N. (1997) The Tn10 synaptic complex can capture a target DNA only after transposon excision. *Cell*, **89**, 205-214.
- Sakai J.S., Kleckner N., Yang X., and Guhathakurta A. (2000) Tn10 transpososome assembly involves a folded intermediate that must be unfolded for target capture and strand transfer. *EMBO J.*, **19**, 776-785.
- 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**, 6348-61.
- 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.
- Sedgwick S.G., Nguyen T.M., Ellis J.M., Crowne H., and Morris G.E. (1991) Rapid mapping by transposon mutagenesis of epitopes on the muscular dystrophy protein, dystrophin. *Nucleic Acids Res.*, **19**, 5889-94.
- Shapiro J.A. (1979) Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proc. Natl. Acad. Sci. USA*, **76**, 1933-7.
- Sherrat D.J. (1995) *Mobile genetic elements*. Oxford University Press, Oxford, UK.
- Silhavy T.J., Brickman E., Bassford P.J., Casabadan M.J., Shuman H.A., Schwartz M., and Beckwith J.R. (1979) Structure of the malB region in *Escherichia coli* K12. II Genetic map of the malE,F,G operon. *Mol. Gen. Genet.*, **174**, 249-259.
- Singh I.R., Adams C.R., and Brown P.O. (1997) High-resolution functional mapping of cloned gene by genetic footprinting. *Proc. Natl. Acad. Sci. USA*, **94**, 1304-1309.
- Smith V., Botstein D., and Brown P.O. (1995) Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proc. Natl. Acad. Sci. USA*, **92**, 6479-6483.
- Stellwagen A.E., and Craig N.L. (1997a) Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7. *Genetics*, **145**, 573-85.
- Stellwagen A.E., and Craig N.L. (1997b) Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition. *EMBO J.*, **16**, 6823-34.
- Stellwagen A.E., and Craig N.L. (1998) Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. *Trends Biochem. Sci.*, **23**, 486-490.
- Stellwagen A.E., and Craig N.L. (2001) Analysis of gain-of-function mutants of an ATP-dependent regulator of Tn7 transposition. *J. Mol. Biol.*, **305**, 633-42.
- Strathmann M., Hamilton B.A., Mayeda C.A., Simon M., Meyerowitz E.M., and Palazzolo M.J. (1991) Transposon-facilitated DNA sequencing. *Proc. Natl. Acad. Sci. USA*, **88**, 1247-1250.
- Sun Y.-H., Bakshi S., Chalmers R., and Tang C.M. (2000) Functional genomics of *Neisseria meningitidis* pathogenesis. *Nat. Med.*, **6**, 1269 - 1273.
- Surette M.G., and Chaconas G. (1989) A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction in *Escherichia coli* integration host factor. *J. Biol. Chem.*, **264**, 3028-34.
- 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., Harkness T., and Chaconas G. (1991) Stimulation of the MuA protein-mediated strand cleavage reaction by the MuB protein, and the requirement

- of DNA nicking for stable Type 1 transpososome formation. *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.
- Taira S., Tuimala J., Roine E., Nurmiäho-Lassila E.L., Savilahti H., and Romantschuk M. (1999) Mutational analysis of the *Pseudomonas syringae* pv. tomato hrpA gene encoding Hrp pilus subunit. *Mol. Microbiol.*, **34**, 737-44.
- Taylor A.L. (1963) Bacteriophage-induced mutation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **50**, 1043-1051.
- Tong-Hoang B., Betermier M., Polard P., and Chandler M. (1997) Assembly of a strong promoter following IS911 circularization and the role of circles in transposition. *EMBO J.*, **16**, 3357-3371.
- Tong-Hoang B., Polard P., and Chandler M. (1998) Efficient transposition of IS911 circles in vitro. *EMBO J.*, **17**, 1169-1181.
- Tong-Hoang B., Polard P., Haren L., Turlan C., and Chandler M. (1999) IS911 transposon circles give rise to linear forms that can undergo integration in vitro. *Mol. Microbiol.*, **32**, 617-627.
- Turlan C., and Chandler M. (2000) Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.*, **8**, 268-274.
- Waddell C.S., and Craig N.L. (1988) Tn7 transposition: two transposition pathways directed by five Tn7-encoded genes. *Genes Dev.*, **2**, 137-49.
- Van Luenen H.G.A.M., and Plasterk H.A. (1994) Target site choice of the related transposable elements Tc1 and Tc3 of *Caenorhabditis elegans*. *Nucleic Acids Res.*, **22**, 262-269.
- Van Luenen H.G.A.M., Colloms S.D., and Plasterk H.A. (1994) The mechanism of transposition of Tc3 in *C. elegans*. *Cell*, **79**, 293-301.
- Wang X., and Higgins P.N. (1994) 'Muprints' of the lac operon demonstrate physiological control over the randomness of in vivo transposition. *Mol. Microbiol.*, **12**, 665-677.
- 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.
- Westphal C.H., and Leder P. (1997) Transposon-generated 'knock-out' and 'knock-in' gene-targeting constructs for use in mice. *Curr. Biol.*, **7**, 530-3.
- Vilen H., Eerikainen S., Tornberg J., Airaksinen M.S., and Savilahti H. (2001) Construction of gene-targeting vectors: a rapid Mu in vitro DNA transposition-based strategy generating null, potentially hypomorphic, and conditional alleles. *Transgenic Res.*, **10**, 69-80.
- Williams T.L., Jackson E.L., Caritte A., and Baker T.A. (1999) Organization of the Mu transposome: recombination by communication between two active sites. *Genes Dev.*, **13**, 2725-2737.
- Wolkow C.A., DeBoy R.T., and Craig N.L. (1996) Conjugating plasmids are preferred targets for Tn7. *Genes Dev.*, **10**, 2145-57.
- Wong S.M., and Mekalanos J.J. (2000) Genetic footprinting with mariner-based transposition in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. USA*, **97**, 10191-6.
- Wu Z., and Chaconas G. (1992) Flanking host sequences can exert an inhibitory effect on the cleavage step of the in vitro Mu DNA strand transfer reaction. *J. Biol. Chem.*, **267**, 9552-9558.
- Wu Z., and Chaconas G. (1997) The Mu transposase tetramer is inactive in unassisted strand transfer: an autoallosteric effect of Mu A promotes the reaction in the absence of Mu B. *J. Mol. Biol.*, **267**, 132-141.
- 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 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.
- Yant S.R., Meuse L., Chiu W., Ivics Z., Izsvak Z., and Kay M.A. (2000) Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nat. genet.*, **25**, 35-41.

York D., Welch K., Goryshin I.Y., and Reznikoff W.S. (1998) Simple and efficient generation in vitro of nested deletions and inversions: Tn5 intramolecular transposition. *Nucleic Acids Res.*, **26**, 1927-33.