



# LUND UNIVERSITY

## Flexibility in MuA Transposase Family Protein Structures: Functional Mapping with Scanning Mutagenesis and Sequence Alignment of Protein Homologues

Rasila, Tiina S.; Vihinen, Mauno; Paulin, Lars; Haapa-Paananen, Saija; Savilahti, Harri

Published in:  
PLoS ONE

DOI:  
[10.1371/journal.pone.0037922](https://doi.org/10.1371/journal.pone.0037922)

2012

[Link to publication](#)

### Citation for published version (APA):

Rasila, T. S., Vihinen, M., Paulin, L., Haapa-Paananen, S., & Savilahti, H. (2012). Flexibility in MuA Transposase Family Protein Structures: Functional Mapping with Scanning Mutagenesis and Sequence Alignment of Protein Homologues. *PLoS ONE*, 7(5), Article e37922. <https://doi.org/10.1371/journal.pone.0037922>

Total number of authors:  
5

### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

# Flexibility in MuA Transposase Family Protein Structures: Functional Mapping with Scanning Mutagenesis and Sequence Alignment of Protein Homologues

Tiina S. Rasila<sup>1</sup>, Mauno Vihinen<sup>2,3,4</sup>, Lars Paulin<sup>1</sup>, Saija Haapa-Paananen<sup>1‡</sup>, Harri Savilahti<sup>1,5\*</sup>

**1** Institute of Biotechnology, Viikki Biocenter, University of Helsinki, Helsinki, Finland, **2** Institute of Biomedical Technology, University of Tampere, Tampere, Finland, **3** BioMediTech, Tampere, Finland, **4** Department of Experimental Medical Science, Lund University, Lund, Sweden, **5** Division of Genetics and Physiology, Department of Biology, University of Turku, Turku, Finland

## Abstract

MuA transposase protein is a member of the retroviral integrase superfamily (RISF). It catalyzes DNA cleavage and joining reactions via an initial assembly and subsequent structural transitions of a protein-DNA complex, known as the Mu transpososome, ultimately attaching transposon DNA to non-specific target DNA. The transpososome functions as a molecular DNA-modifying machine and has been used in a wide variety of molecular biology and genetics/genomics applications. To analyze structure-function relationships in MuA action, a comprehensive pentapeptide insertion mutagenesis was carried out for the protein. A total of 233 unique insertion variants were generated, and their activity was analyzed using a quantitative *in vivo* DNA transposition assay. The results were then correlated with the known MuA structures, and the data were evaluated with regard to the protein domain function and transpososome development. To complement the analysis with an evolutionary component, a protein sequence alignment was produced for 44 members of MuA family transposases. Altogether, the results pinpointed those regions, in which insertions can be tolerated, and those where insertions are harmful. Most insertions within the subdomains I $\gamma$ , II $\alpha$ , II $\beta$ , and III $\alpha$  completely destroyed the transposase function, yet insertions into certain loop/linker regions of these subdomains increased the protein activity. Subdomains I $\alpha$  and III $\beta$  were largely insertion-tolerant. The comprehensive structure-function data set will be useful for designing MuA transposase variants with improved properties for biotechnology/genomics applications, and is informative with regard to the function of RISF proteins in general.

**Citation:** Rasila TS, Vihinen M, Paulin L, Haapa-Paananen S, Savilahti H (2012) Flexibility in MuA Transposase Family Protein Structures: Functional Mapping with Scanning Mutagenesis and Sequence Alignment of Protein Homologues. PLoS ONE 7(5): e37922. doi:10.1371/journal.pone.0037922

**Editor:** Shuang-yong Xu, New England Biolabs, Inc., United States of America

**Received:** February 14, 2012; **Accepted:** April 26, 2012; **Published:** May 29, 2012

**Copyright:** © 2012 Rasila et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study has been funded by the Academy of Finland, Finnish National Technology Agency (TEKES), Biocenter Finland, Medical Research Fund of Tampere University Hospital, Sigrid Juselius Foundation, Finnish Cultural Foundation, and Emil Aaltonen Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: harri.savilahti@utu.fi

‡ Current address: VTT Medical Biotechnology, Turku, Finland

## Introduction

Transposable genetic elements constitute a diverse group of discrete DNA segments with a capability of moving within and between genomes [1]. They are abundant in all kingdoms of life and present in virtually every genome examined to date [1,2]. A wealth of data from sequenced genomes has implicated the fundamental importance of mobile DNA in shaping genomes during evolution [3–6]. The increasing knowledge of DNA mobility mechanisms has facilitated the versatile use of transposable elements for research purposes and provided efficient tools for a variety of applications including genome-wide insertional mutagenesis, protein engineering, transgenesis, and gene therapy [7–9].

Many mobile DNA elements transpose via a DNA intermediate and are mobilized by an enzyme called transposase. An important class of such transposases shares a structurally and functionally conserved catalytic core domain. This domain folds into a structure first identified in *Escherichia coli* RNase H1 (thus called an RNase H fold), and it includes three catalytically critical acidic amino acids

known as the DDE motif [10–13]. These DDE-motif transposases belong to a larger group of RNase H fold proteins called a retroviral integrase superfamily (RISF), which also includes retroviral integrases, the Holliday junction resolvase RuvC, the V(D)J recombinase RAG, and Argonaute, the nuclease component of an RNA-induced silencing complex (RISC) [13–17]. In addition, the RNase H fold is included in the carboxy-terminal domains of UvrC (DNA-repair) and Prp8 (RNA-processing) proteins, and therefore they are also classified as RISF proteins [13]. Because of a similar molecular architecture, all RISF proteins are expected to use a common mechanism for nucleic acid cleavage and joining reactions [13]. Accordingly, structural and functional insights gained from any member of the RISF proteins can potentially be extrapolated to the entire superfamily.

Bacteriophage Mu propagates via DNA transposition. Owing to its efficient DNA mobilization capacity *in vivo* [18] and the early development of an *in vitro* system ([19], Figure S1), it has served as an important model system for DNA transposition studies [20]. Mu encodes MuA transposase, a well-characterized member of RISF [12,13,21,22], which catalyzes the critical steps of trans-

position: (i) initial cleavages at the transposon-host boundaries (donor cleavage) and (ii) covalent integration of the transposon into the target DNA (strand transfer). These steps proceed via sequential structural transitions within a nucleoprotein complex, a transpososome [20,23,24], the core of which contains four MuA molecules and two synapsed transposon ends ([25–27], Figure 1). *In vivo*, the critical MuA-catalyzed reaction steps also involve the phage-encoded MuB targeting protein, host-encoded DNA architectural proteins (HU and IHF), certain DNA cofactors (MuA binding sites and transpositional enhancer sequence), as well as stringent DNA topology [28]. The critical reaction steps mimicking Mu transposition into external target DNA (Figure 1) can be reconstituted *in vitro* using MuA transposase, 50 bp Mu R-end DNA segments, and target DNA as the only macromolecular components [27,29]. Such a minimal system has been instrumental for the detailed analyses on the molecular mechanisms of Mu transposition [30–32]. A versatile use of the reaction series with custom-designed substrates has generated a wealth of tools for molecular biology applications [33–38] and produced novel strategies for genetics/genomics research [39–43].

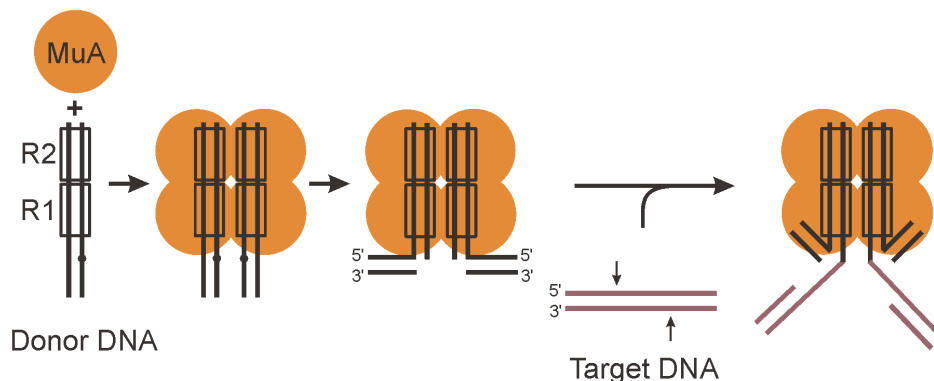
MuA is a 75-kDa protein (663 amino acids) and can be divided into structurally and functionally defined major domains (I, II, III) and subdomains (I $\alpha$ , I $\beta$ , I $\gamma$ ; II $\alpha$ , II $\beta$ ; III $\alpha$ , III $\beta$ ) [44–48]. The N-terminal subdomain I $\alpha$  promotes transpososome assembly via an initial binding to a specific transpositional enhancer sequence [49,50]. The specific DNA binding to transposon ends, crucial for the transpososome assembly, is mediated through amino acid residues located in domains I $\beta$  and I $\gamma$  [46,47]. Subdomain II $\alpha$  contains the critical DDE-motif of acidic residues (D269, D336 and E392), which is involved in the metal ion coordination during the catalysis [51,52]. Subdomains II $\beta$  and III $\alpha$  participate in nonspecific DNA binding, and they appear important during structural transitions [26,52]. Subdomain III $\alpha$  also displays a cryptic endonuclease activity, which is required for the removal of the attached host DNA following the integration of infecting Mu [53,54]. The C-terminal subdomain III $\beta$  is responsible for the interaction with the phage-encoded MuB protein, important in targeting transposition into distal target sites [55–58]. This subdomain is also important in interacting with the host-encoded ClpX protein, a factor which remodels the transpososome for disassembly [59].

While all MuA subdomains are required for efficient phage Mu transposition inside *Escherichia coli*, the terminal subdomains I $\alpha$  and III $\beta$  become dispensable in certain *in vivo* and *in vitro* conditions with appropriately altered DNA substrates and/or suitably modified reaction milieu [60,61].

Databases classify a number of MuA homologues in a variety of bacteria. Their conserved primary sequence and similar domain structure suggest conservation in their function as well. Although it is not known whether these homologues are currently transpositionally active, aligning their sequences should allow the detection of evolutionarily relevant changes within the MuA family of proteins. It is likely that the obtained data will be applicable to other DDE-motif transposase families as well.

High-resolution structures have been determined for nearly all of the individual subdomains of MuA by NMR [45–47] or X-ray crystallography [48]. In addition, a low resolution structure of Mu transpososome has been reconstructed using cryo EM [62]. Very recently, also an X-ray structure of Mu transpososome has been resolved. This informative structure includes target DNA and portrays the transpososome architecture at post-integration stage (P. Rice and S.P. Montano, personal communication). The solved structures generate a platform for future studies and offer a wealth of detailed information for the architectural interpretation of functional data.

To date, studies on the function of MuA within the Mu transpososome have been limited to deletion analyses as well as mutational analyses of single amino acids. Studies on DNA substrate specificity have complemented the analyses and revealed a degree of flexibility within the architectural requirements [63–66]. However, for rational protein engineering and further development of Mu-based technology, more structurally oriented functional studies are needed. Here, we report a comprehensive structure-function analysis to map MuA transposase regions that withstand amino acid insertions versus regions that do not tolerate them. By correlating the activities of insertion mutants with MuA structure, important details of the protein function were revealed. As MuA transposase forms an instrumental component within an expanding tool arsenal for a vast variety of genetics/genomics applications, new data with regard its function are critical.



**Figure 1. Assembly and function of Mu transpososome core.** This pathway is based on *in vitro* studies and utilizes a minimum number of macromolecular components [27]. The *in vivo* assay described earlier [60] and used here (Figure S1) is a close mimic of this minimal-component *in vitro* system with regard the following features: (i) the configuration includes two MuR-ends (with R1 and R2 MuA binding sites), (ii) the phage-encoded MuB protein and (iii) transpositional enhancer are not included. See Figure S1 for a full description of Mu transposition pathway and its comparison to the pathway used in the papillation analysis. The R1 and R2 MuA binding sites are shown as rectangles. MuA is drawn as a tetramer of yellow circles and target DNA is shown in purple. The small arrows on the target DNA indicate the 5-bp staggered locations for strand transfer on the two strands. The dots in the assembled transpososome indicate the Mu end cleavage sites.  
doi:10.1371/journal.pone.0037922.g001

## Materials and Methods

### Escherichia Coli

**Strains, reagents, and DNA techniques.** DH10B [67] was used as a standard cloning host, and DH5 $\alpha$  (Invitrogen) was used for routine plasmid DNA isolation and papillation analysis. Standard bacterial cultures were grown in Luria-Bertani (LB) medium or on LB agar plates [68] supplemented with appropriate antibiotic(s) when required: ampicillin (Ap, 100  $\mu$ g/ml), kanamycin (Km, 25  $\mu$ g/ml), chloramphenicol (Cm, 10  $\mu$ g/ml). Electrocompetent cells for cloning and standard competent cells for papillation analysis were prepared as described in [42] and [69], respectively. Plasmid DNA was prepared using appropriate kits from QIAGEN. Restriction enzymes and DNA modifying enzymes were used as recommended by their supplier (New England Biolabs). Antibiotics and arabinose were from Sigma. Lactose was from BDH and 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (Xgal) from AppliChem.

**Generation of insertion mutant plasmid library.** Mutation Generation System (Finnzymes) was used to generate a pool of plasmids for papillation analysis, each plasmid containing a 15-bp insertion within the *MuA* gene. Within the 15-bp insertion, 10 bp (TGCGGCCGCA) is derived from the ends of the transposon used, and 5 bp from the duplicated target DNA at the insertion site [37]. Translation through the insertion is dependent on the reading frame and the sequence at the insertion site (Table S1). In detail, a two-fold scale-up of a standard *in vitro* transposition reaction was performed in a total volume of 40  $\mu$ l with 200 ng of Entranceposon (M1-Kan<sup>R</sup>) as a donor DNA and 770 ng of plasmid pALH6 [70] as a target DNA. Following incubation at 30°C for 2 h, reaction products were extracted with phenol and subsequently with chloroform, ethanol precipitated, and resuspended in 10  $\mu$ l of water. Individual aliquots (1  $\mu$ l) were used to electrotransform DH10B electrocompetent cells as described [42]. Transposon-containing plasmid clones were selected on LB-Ap-Km plates. A total of  $\sim 6.2 \times 10^4$  colonies were pooled. Plasmid DNA from the pool was isolated, double-digested with NcoI and EcoRI, and subjected to preparative electrophoresis on a 0.8% SeaPlaque GTG agarose gel in TAE buffer [68]. The 3.1-kb DNA fragment pool, corresponding to transposon insertions into the MuA-encoding DNA segment, was isolated using electroelution [68] and consecutive 1-butanol and chloroform extractions. The DNA was ethanol precipitated, resuspended in TE-buffer (10 mM Tris [pH 7.5], 0.5 mM EDTA), and ligated into NcoI-EcoRI double-digested plasmid pTLH1 [60]. The ligation mixture was electrotransformed into DH10B cells as above, and insert-containing plasmid clones were selected on LB-Km-Cm plates. A total of  $\sim 2.7 \times 10^4$  colonies were pooled and grown in LB-Km-Cm medium at 37°C for 2 h, after which plasmid DNA from the pool was isolated. The transposon core sequence was eliminated from the plasmid pool by a cleavage with NotI, followed by preparative electrophoresis on a 0.6% SeaPlaque GTG agarose gel, and isolation of the plasmid backbone as above, and recirculation by ligation at low DNA concentration ( $\sim 2$  ng/ $\mu$ l). The ligation mixture was electrotransformed into DH10B cells as above, and library clones were selected on LB-Ap-Cm plates. A total of  $\sim 2.6 \times 10^4$  colonies were pooled and grown in LB-Ap-Cm medium at 37°C for 2 h. DNA was isolated to yield the final insertion mutation library. For the isolation of individual insertion mutant plasmids, DNA from the mutant library was electrotransformed into DH5 $\alpha$  cells. Individual clones were selected on LB-Ap-Cm plates and grown in LB-Ap-Cm medium for DNA isolation.

**Mapping the insertion sites.** The 15 bp insertion sites in isolated plasmids were roughly mapped by an initial screen using either NcoI-NotI or NotI-EcoRI double digestions for 608 clones. On the basis of this initial screen, a total of 331 clones were subjected to DNA sequence analysis to reveal the exact location of the insertion in each individual clone. The following primers were used for the analyses: HSP492 (5'-ATCA-GACCGCTTCTGCGTTC), HSP493 (5'-GATTAGCG-GATCCTACCTGAC), HSP574 (5'-GCCGGACAAGACCG-TAACTTG), and HSP680 (5'-GCAACAGGTGCCAGACATTC). The use of these primers produced partially overlapping sequence reads, together covering the entire *MuA* gene region. DNA sequence determination was performed at the DNA sequencing facility of the Institute of Biotechnology (University of Helsinki) by using the BigDye terminator cycle sequencing kit v. 3.1 and ABI 3130 XL sequencer, both from Applied Biosystems.

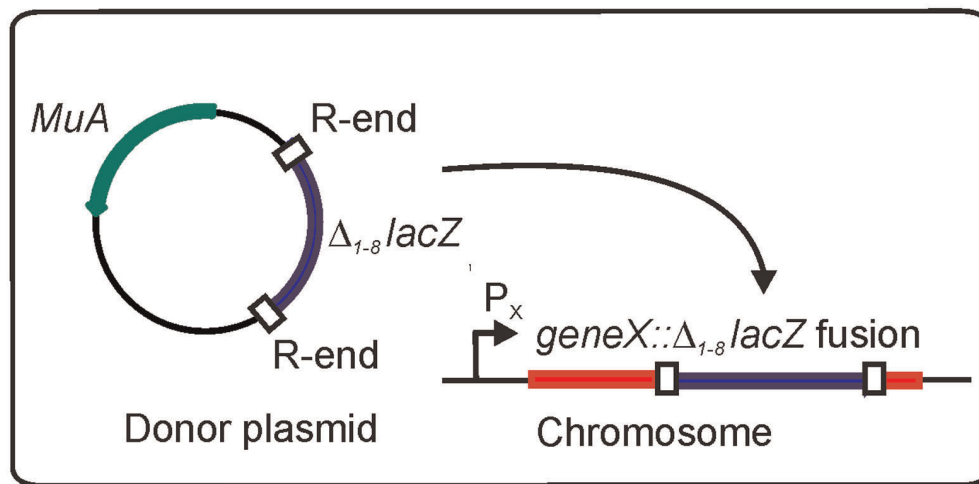
**Papillation assay.** MuA insertion variants were assayed for their transpositional activity using an *in vivo* analysis that is based on transposon mobilization [60]. This quantitative assay scores transposition events as blue microcolonies (papillae) growing on otherwise whitish *E. coli* colonies. It takes advantage of a plasmid, which contains a *lacZ*-containing reporter transposon and a cassette for arabinose-inducible MuA expression. Briefly, each insertion mutant plasmid was transformed into standard competent DH5 $\alpha$  cells (50  $\mu$ l), and the cells were plated ( $\sim 100$  colonies per plate) onto LB agar plates supplemented with 100  $\mu$ g/ml Ap, 20  $\mu$ g/ml Cm, 0.05% lactose, 40  $\mu$ g/ml Xgal, and 0.1% arabinose. The plates were incubated at 30°C for 115 h. For the analysis of each mutant, three representative colonies (diameter  $\sim 5$  mm) were photographed using Olympus ColorView II digital camera attached to Olympus SZX12 stereomicroscope equipped with Zeiss KL1500 LCD cold light source. The number of papillae in each colony was then enumerated by using the digital imaging program Image Pro Plus (Media Cybernetics). The data were used to calculate the mean value and standard deviation (SD) for each protein variant. The mean value of MuA wild type activity, 234 papillae, was employed as a reference for 100% activity.

**Bioinformatics analyses.** Multiple sequence alignment was made with ClustalW [71] (for sequences obtained from Pfam [72] for DDE2 family (PF02914) members of RNaseH clan (CL0219). Partial sequences and those with extensive gaps were omitted. The NMR and X-ray structures for isolated domains of MuA protein (codes: 1tns, 2ezk, 2ezh, 1bcm) were from Protein Data Bank (PDB) [73]. Locations of protein secondary structural elements were identified with DSSP [74]. The structures were visualized with UCSF Chimera [75].

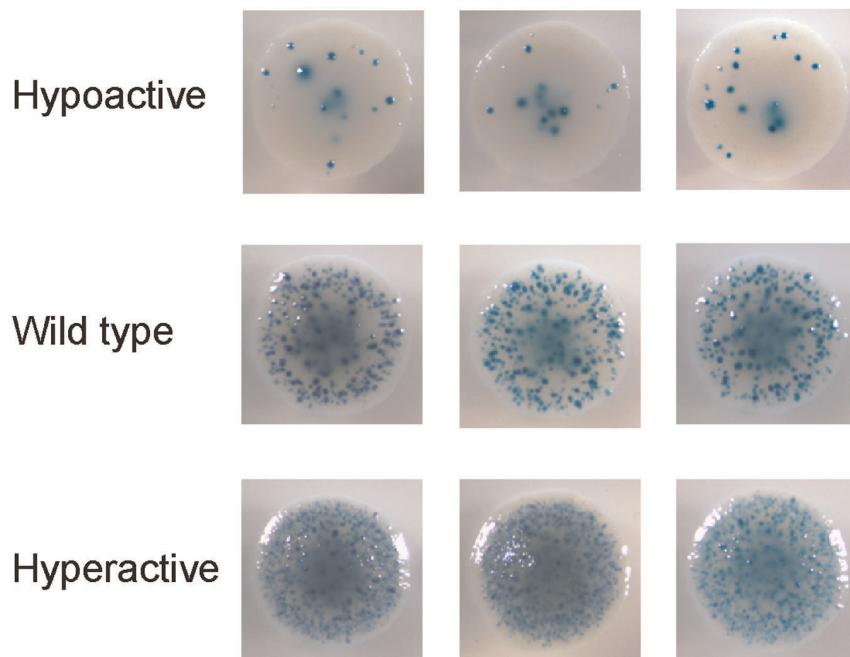
## Results

To gain insights into the structure-function relationships of MuA transposase, we generated 233 unique pentapeptide insertion mutants, and these MuA variants were analyzed for function using a quantitative *in vivo* assay (Figure 2). The activity results were correlated with the MuA domain structures, and the data were evaluated with regard to the protein function. In addition, protein sequences of 44 MuA transposase homologues were aligned to reveal telltale signals of evolutionarily successful insertion/deletion (indel) sites. Combining the quantitative functional data, structural assessment, and alignment-based evaluation generated a comprehensive map of insertion-tolerant versus insertion-intolerant regions in the protein.

A



B



**Figure 2. Papillation assay.** (A) Phenotypically  $Lac^-$  *E. coli* strain is transformed with a plasmid carrying a reporter transposon and encoding arabinose-inducible *MuA* transposase gene. Following expression of *MuA*, the reporter transposon is mobilized. Transposition into an expressed gene (*geneX*) in the correct orientation and reading frame generates a *geneX::lacZ* gene fusion, expressing a protein fusion with a C-terminal  $\beta$ -galactosidase moiety. Such events can be detected as blue papillae in bacterial colonies growing on Xgal-containing indicator plates. This quantitative assay directly measures the activity of the *MuA* variant analyzed [60]. (B) Colonies from papillation assay. Shown are colonies representing one hypoactive *MuA* variant (clone #188, Table S1), wild type *MuA*, and one hyperactive *MuA* variant (clone #170, Table S1). Three representative colonies per variant are shown, indicating a high degree of reproducibility.  
doi:10.1371/journal.pone.0037922.g002

### Generation and Characterization of MuA Insertion Mutant Library

We generated a library of plasmids, in which each clone encoded a randomly positioned five amino acid insertion in

*MuA*. The strategy included a final subcloning step, guaranteeing that the insertions were confined solely to the targeted *MuA*-encoding region. A total of 233 different insertion sites were identified. Their overall distribution as well as their localization

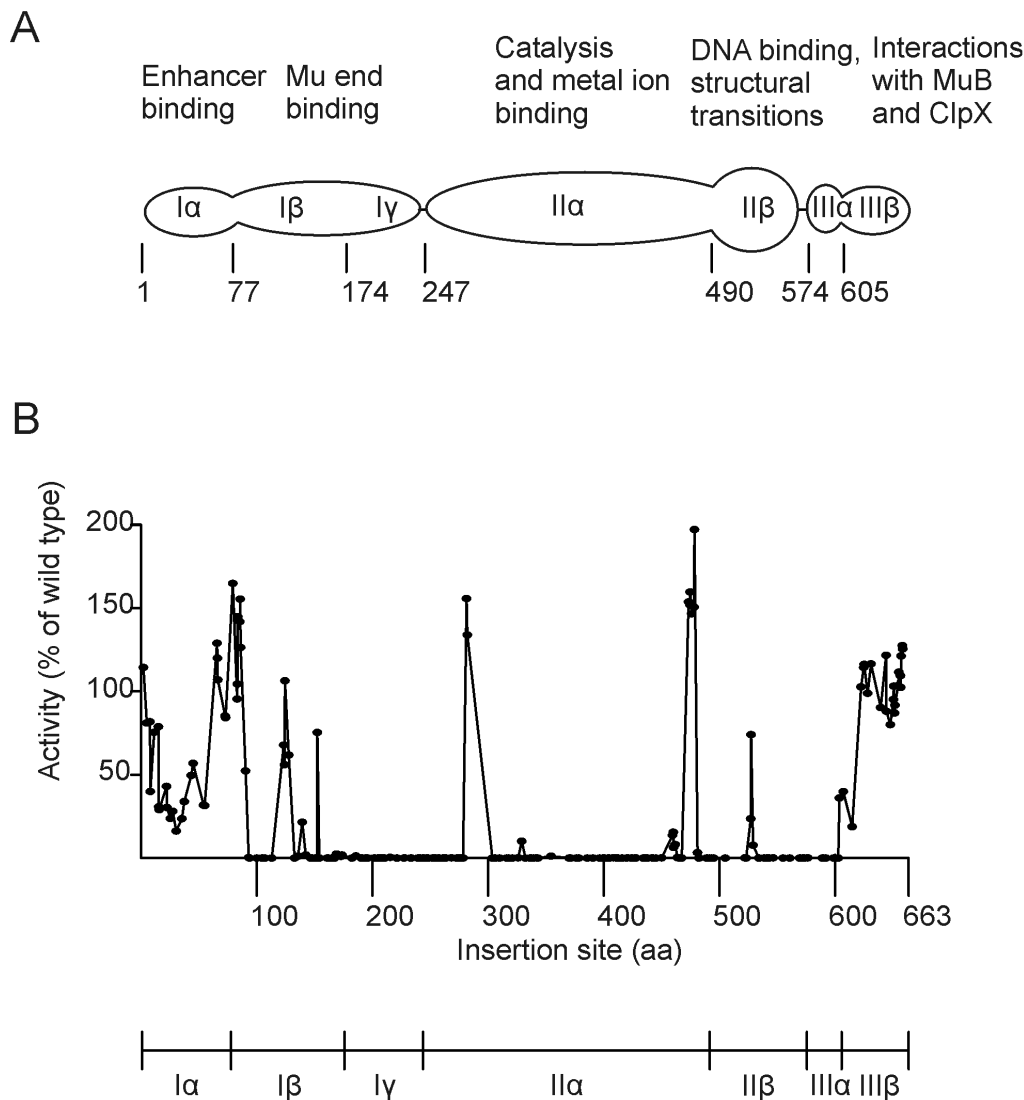


within the secondary structural elements of MuA (Figures 3 and 4) was regarded as being sufficient for a reliable structure-function assessment of the protein. The transpositional activity of each of the MuA variants was monitored using a quantitative *in vivo* assay (Materials and Methods, Figure 2). The assay measures transposition frequency as a number of blue papillae appearing on otherwise whitish bacterial colonies. Only fully productive transposition events generate papillae in this assay, meaning that critical alterations at any stage along the transposition pathway will be reflected in the observed transposition frequency. Most of the insertions reduced the activity of MuA (Figures 3 and 4), as a large number of the mutants (187 of 233) exhibited less than 70% of the wild type activity, and more than half of the mutants (125 of 233) were totally inactive with no papillae produced. Many mutants (36 of 233) showed an activity that can be regarded as a wild type or close to the wild type (activity level

70–130%). Some variants (12 of 233) exhibited activity levels exceeding 130%, the highest score being 197%.

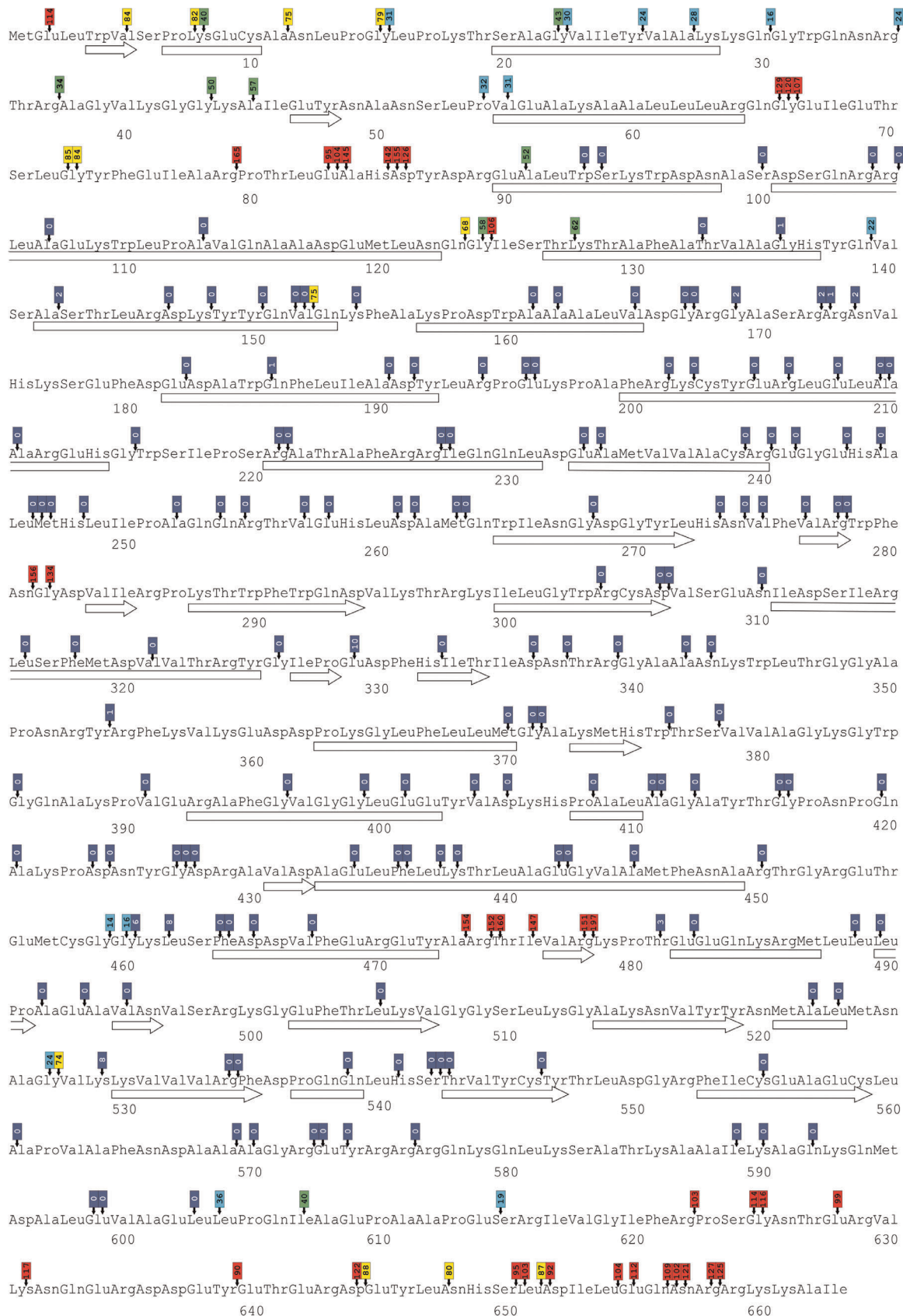
### Insertion Tolerance within MuA Domains

MuA is composed of three major protein domains, each containing two or three subdomains (Figure 3A). Our analysis indicated clear differences in insertion tolerance between the subdomains. In particular, the entire subdomain I $\alpha$  and the main part of subdomain III $\beta$  tolerated insertions relatively well, as none of the insertions into these subdomains totally abolished the MuA activity. Three of the subdomains, I $\beta$ , II $\alpha$  and II $\beta$ , tolerated some insertions but only in certain confined regions. Subdomains I $\gamma$  and III $\alpha$  appeared entirely insertion-intolerant. A five residue insertion forms a substantial change in protein sequence and structure. Accordingly, sites where extra residues can be accommodated indicate regions, which are likely not tightly packed, where the residues do not impede specific structural transitions, and/or



**Figure 3. MuA domain structure and location of pentapeptide insertions with respective transposase activities.** (A) Structural organization of MuA with different functions assigned to various subdomains. The numbers correspond to the amino terminus of each subdomain as specified earlier [47]. (B) Enzymatic activities of MuA variants plotted against the 5 aa insertion site of each respective mutant protein. Activities obtained from the papillation assay are presented as a percentage of the wild type activity. For each protein variant the mean from three replicates is shown.

doi:10.1371/journal.pone.0037922.g003



**Figure 4. Location of insertions within the amino acid sequence and relative transposition activity of each mutant.** Amino acid positions are numbered under the sequence. Secondary structures (determined with DSSP, [74]) are indicated under the corresponding sequence as bars ( $\alpha$ -helices) or arrows ( $\beta$ -strands). Small vertical arrows point to the exact location of each insertion (in each three-letter amino acid code, the first letter represents the first nucleotide of the corresponding codon, etc.), and the attached coloured boxes indicate ranges of transposition activity of each mutant relative to the wild type activity (colors denoting the percentage ranges are shown at the lower-right). The number within the boxes indicate the observed relative activity.  
doi:10.1371/journal.pone.0037922.g004

regions where sequence conservation is not crucial (e.g. for the enzyme activity, ligand binding or protein folding). In the next subsections, the effects of the insertions are discussed in more detail.

**Domain I.** Subdomain I $\alpha$  tolerated insertions relatively well. In particular, insertions into the regions forming the subdomain termini (aa 1–15 and 66–76) either retained the wild type activity or reduced the activity only marginally (with one exception: 40% activity at aa 8, frame 3). However, insertions into the subdomain's central region (aa 16–55) reduced the protein activity noticeably, and only ~20–60% of the wild type activity was retained. Insertions into the extreme N-terminus of subdomain I $\beta$  (aa 80–86) either retained the wild type activity or substantially increased the activity (up to 165%). Insertions into other regions of subdomain I $\beta$  (aa 91–173) almost invariably abolished the activity entirely. However, within particular helix termini and loop sequences insertions preserved some activity (in one case even up to the wild type level: aa 125, frame 1). Subdomain I $\gamma$  was very clearly insertion-intolerant, as the variants had a good coverage of insertions between aa 174–245, and these insertions totally (or in two cases, almost totally) abolished the protein activity.

**Domain II.** In general, subdomains II $\alpha$  and II $\beta$  were for the most part extremely insertion-intolerant with no papillae formed. Yet, a few non-inactivating insertions were located within these subdomains. Two loop regions tolerated insertions particularly well (aa 282 and aa 474–479). In fact, these insertions increased the protein activity, and the highest activity score (197%) was observed for the insertion at the latter loop. Furthermore, a substantial level of activity (14–74%) was retained with insertions at two additional loop regions (aa 460 and aa 527–528).

**Domain III.** Insertions into subdomain III $\alpha$  completely inactivated the protein, except for an insertion into the most C-terminal residue (aa 604, 36% activity). Conversely, subdomain III $\beta$  tolerated insertions relatively well. While insertions into its extreme N-terminus (aa 607–615) decreased the activity to a relatively low level (19–40%), all other insertions produced essentially wild type activities.

### Structural Assessment

For the structural assessment of the insertion data, we used the known secondary and tertiary structures of MuA subdomains (Figures 4 and 5). Subdomain I $\alpha$  involves a winged helix-turn-helix DNA-binding module, which interacts specifically with the Mu transpositional enhancer sequence during the transpososome assembly. As the utilized functional assay does not involve the enhancer [60], insertions into this subdomain were not expected to totally eliminate the protein activity. Fulfilling the predictions, a substantial level of activity was retained even with insertions into secondary structural elements of this subdomain. Subdomains I $\beta$  and I $\gamma$  are both involved in the transposon end sequence recognition by MuA. The helices within subdomain I $\beta$  accepted no insertions, except at their extreme termini, consistent with their critical role in DNA binding. One particular region (aa 123–126) between the two helices anchoring the recognition helix accepted insertions without disturbing the protein function. Subdomain I $\gamma$  was very clearly entirely insertion-intolerant, as both secondary

structural elements and regions between them were not able to accommodate insertions without a total (or almost total) loss of activity. Subdomain II $\alpha$  forms the catalytic core of MuA. This domain did not accept insertions in any of its secondary structural elements, highlighting the importance of the protein fold integrity for the active site function. However, a few loops between secondary structural elements withstood insertions well, and in certain positions insertions actually increased the protein activity. Subdomain II $\beta$  forms a  $\beta$ -barrel structure, and this structure was highly intolerant of insertions. However, one loop region between secondary structural elements was able to accommodate insertions with only a moderate reduction in the protein activity. As the architecture of the entire subdomain III has not been revealed, we were not able to assess the insertion data with the structure of this subdomain. Nevertheless, it appears that subdomain III $\alpha$  is highly intolerant of insertions, whereas subdomain III $\beta$  withstands insertions well.

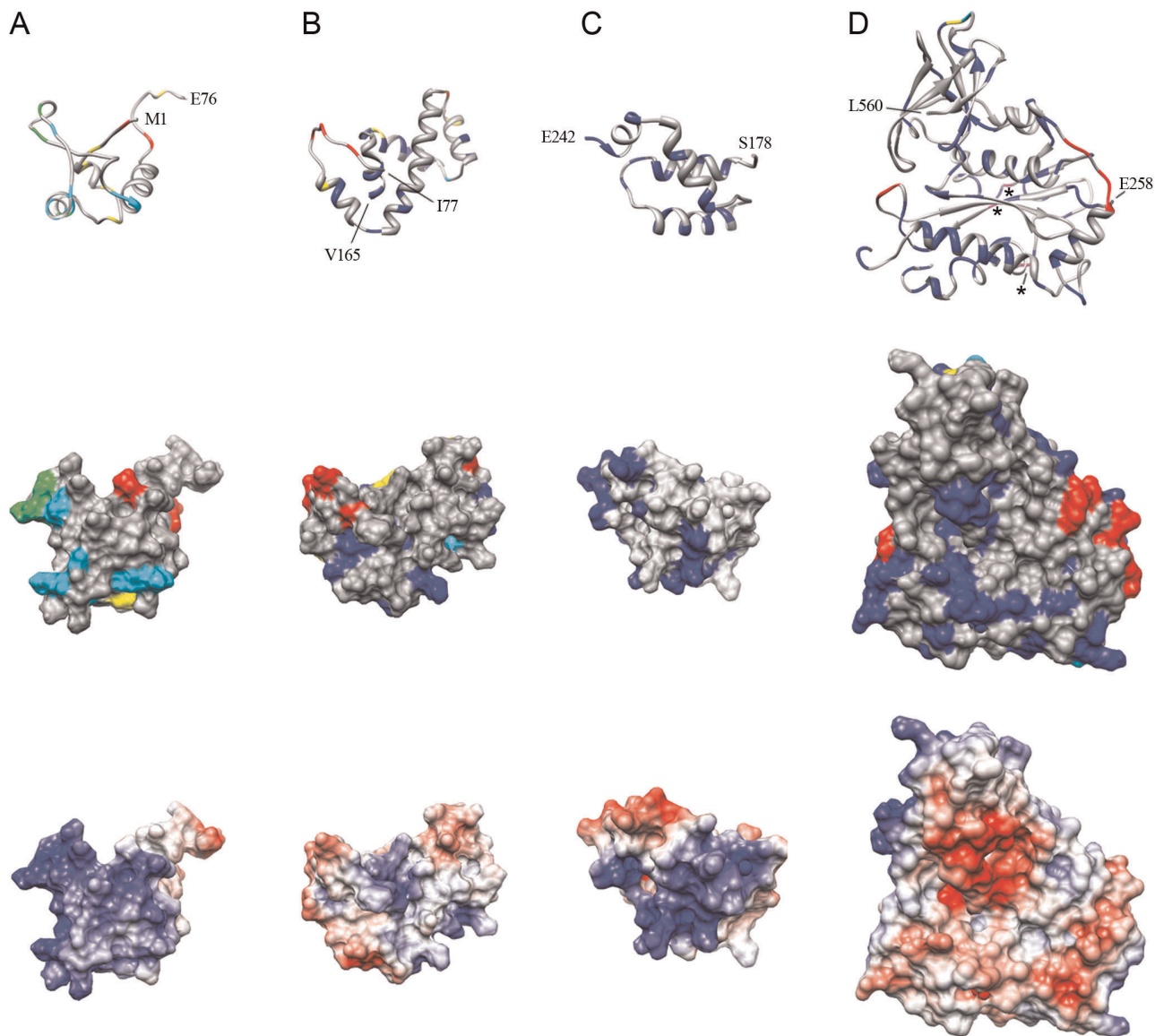
### Alignment of MuA Family Members

During the evolution of a protein family, insertions and deletions are expected to accumulate in regions not critical for the protein function. To pinpoint such regions in MuA, we aligned the amino acid sequences of 44 MuA transposase family members (Figure S2 and Figure 6). Clear differences were observed among the MuA subdomains with regard the accumulation of insertions and deletions. In particular, massive variation was observed within subdomains I $\alpha$  and III $\beta$ . Subdomains I $\beta$ , II $\alpha$  and II $\beta$  exhibited less variation but nevertheless included relatively long (8–10 aa) insertions. Domain I $\gamma$  was even less variable, displaying only a few short (2–4 aa) insertions, and domain III $\alpha$  appeared entirely intolerant of insertions and deletions. In general, a vast majority of length variation was confined to regions between the known secondary structural elements of the protein.

### Correlation of Quantitative Functional Data with Sequence Alignment Data

We next combined the quantitative functional results and the protein alignment data (Figures 6 and 7). In general, a highly congruent pattern between the two data-sets was evident, although some variation could be discerned. In particular, in both analyses subdomains I $\alpha$  and III $\beta$  were extremely tolerant of insertions and deletions, supposedly reflecting a degree of malleability in the functions of these domains as well as their dispensability in assays not involving the IAS or MuB. On the basis of the compiled data, it is clear that the linker sequence between subdomains I $\alpha$  and I $\beta$  can be highly variable, both in length and composition. It also appears that within the subdomains responsible for Mu end binding, I $\beta$  and I $\gamma$ , insertions can be accommodated between secondary structural elements of DNA binding modules. Between subdomains I $\beta$  and I $\gamma$ , there appears to be a region withstanding extensive variation. The compiled data from the catalytic core domain, including subdomains II $\alpha$  and II $\beta$ , indicated four insertion-tolerant regions (roughly: aa 280–290, 350–360, 470–480, and 520–530), as each of these regions were implicated in both of the data sets. Two additional regions of insertion-tolerance were revealed by





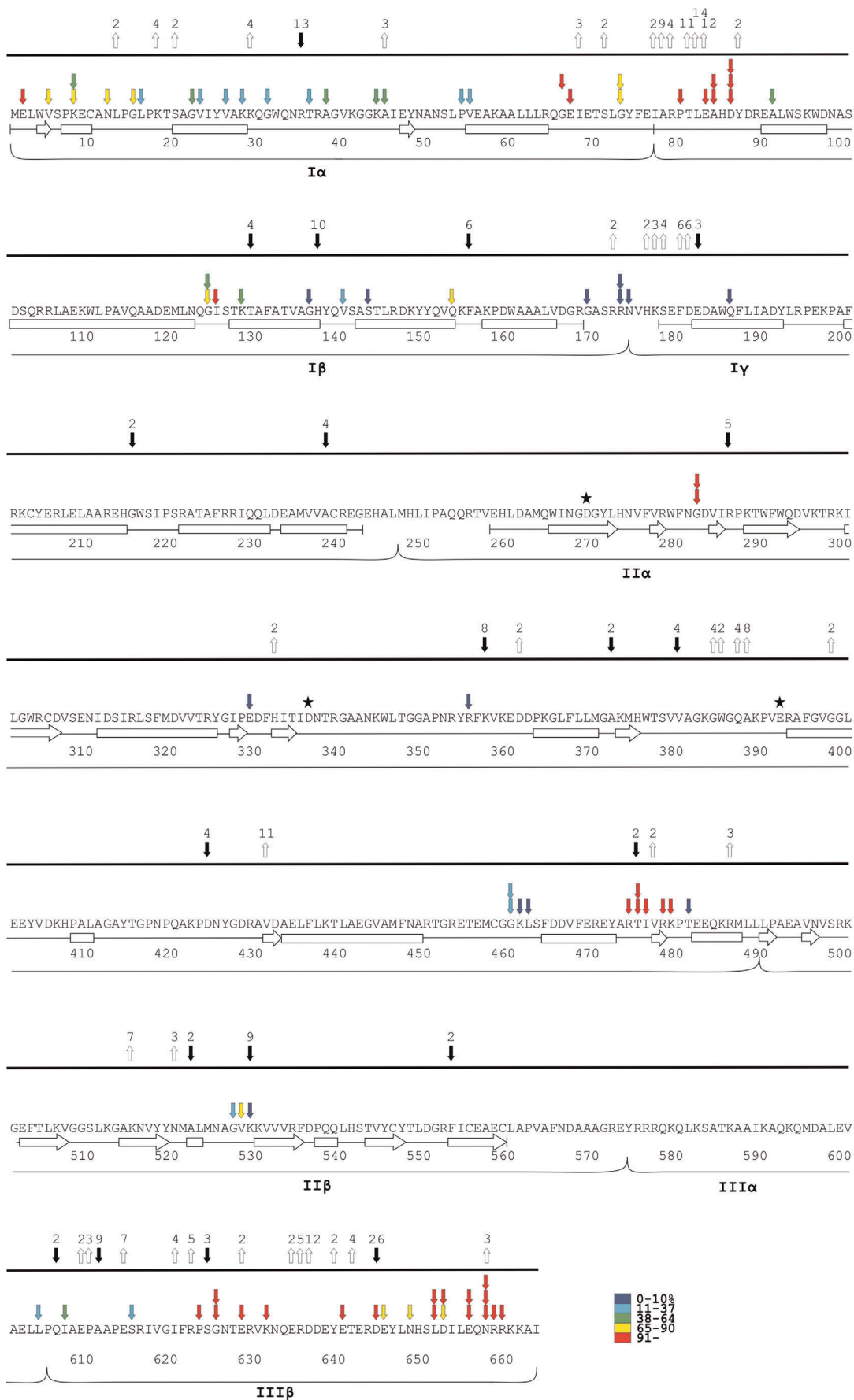
**Figure 5. Insertion sites and corresponding transposition activities displayed in the context of the known domain structures of MuA.** Shown are ribbon and surface models with the insertion mutant data as well as the electrostatic potential of the wild type protein. (A) Subdomain I $\alpha$  (aa 1–76, PDB accession code 1TNS). (B) Subdomain I $\beta$  (aa 77–168, 2EZK). (C) Subdomain I $\gamma$  (aa 178–242, 2EZH). (D) Domain II (aa 258–560, 1BCO). The first and last residues resolved in the structures are labeled with a corresponding number. The amino acid residues corresponding to each insertion site are coloured according to the MuA activity chart and data defined in Figure 4. The stars in panel D ribbon model highlight the DDE-motif residues shown in magenta within the protein chain.  
doi:10.1371/journal.pone.0037922.g005

pentapeptide insertion analysis (around aa 330 and 460). Within the catalytic core, certain sites were implicated in the alignment data set, mostly with alterations of short length. Yet, these regions did not withstand pentapeptide insertions within them or in their close vicinity. As noted earlier, domain III $\alpha$  was intolerant of insertions and deletions in both data sets. Overall, the combined results indicated usefulness of the parallel methodologies and generated a comprehensive global map of insertion-tolerant versus insertion-intolerant regions in MuA.

## Discussion

MuA transposase catalyzes the critical DNA cleavage and joining reactions of bacteriophage Mu transposition in the context

of a large protein-DNA complex, transpososome, ultimately attaching transposon DNA into target DNA. The overall process is elaborate, including initial DNA-binding and transpososome-assembly phases as well as subsequent structural transitions facilitating catalytic steps. Accordingly, MuA needs to function at several levels of complexity, which arguably is reflected in the modular architecture of the protein. Here, we have analyzed the overall transposition process, and defined those structurally and functionally important regions within the protein that are non-modifiable versus those that are modifiable. The analysis has utilized not only the currently available structural data but also sequence comparisons, providing both architectural and evolutionary perspective. Overall, the analyses conform to earlier studies, portraying MuA as a structurally flexible modular protein.



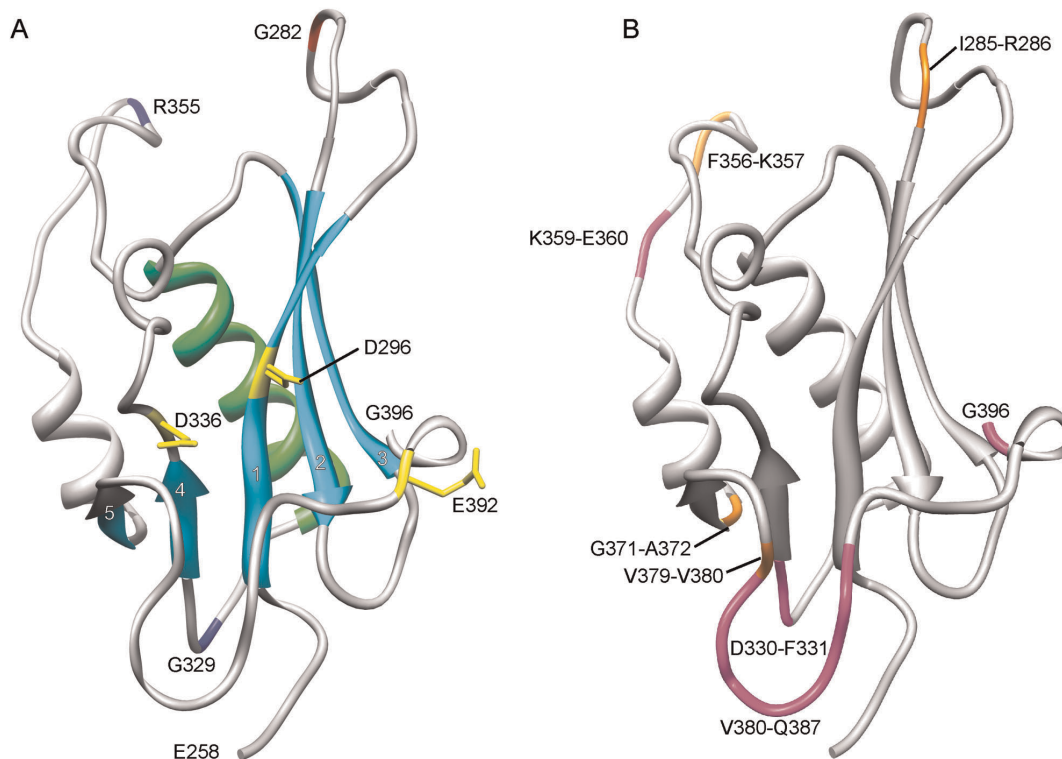
**Figure 6. Identification of insertion-tolerant regions in MuA on the basis of pentapeptide insertion analysis and alignment-based data.** Above the amino acid sequence are shown with arrows the pentapeptide-insertion tolerant sites (activity level  $\geq 1\%$ ) colour-coded as in Figure 4. The percentage range chart is shown at the lower-right. Two or three arrows per site are indicative of insertions involving more than one reading frame. Below the amino acid sequence are shown the secondary structural elements (arrows and rectangles). The elements are connected with line segments indicating the length of each PDB structure. Below the structural elements are shown the subdomains as specified in Figure 3. Above the bolded line, the downward black and upward white arrows represent the alignment-based insertion and deletion (indel) data, respectively (each particular indel precedes the marked amino acid). The maximum indel length at each site is indicated by a number shown above each arrow (data from Figure S1). The stars indicate the DDE-motif residues (D269, D336, E392).  
doi:10.1371/journal.pone.0037922.g006

In addition, a high degree of malleability was revealed in certain inter-domain linker regions as well as in loops connecting secondary structural elements within otherwise insertion-intolerant domain core structures.

The *in vivo* transposition assay used scores only fully productive transposition events. Thus, any mutation-inflicted perturbation at any of the steps along the transposition pathway would be reflected in the frequency of emerging papillae. In this study, we used five amino acid insertions as a probe to dissect the function of MuA. In general, such insertions are expected to locally perturb protein structure, pinpointing structurally and functionally important amino acid regions with a high precision, as shown previously in several studies involving many types of proteins [35,37,76–80]. Indeed, the combined data from all insertion clones allowed a quantitative global assessment of the MuA function. To complement the functional mapping of MuA, we performed an amino acid level comparison of MuA homologues. This comparison of evolutionarily related proteins generated an independent data set paralleling the functional data, but also revealing an extra

set of potential insertion-tolerant regions. Combining the results of the two approaches produced a high-definition structure-function map of MuA.

Five amino acid insertions are substantial in size, moving the originally adjacent residues widely apart. In general, such changes are typically not tolerated in the middle of secondary structural elements or in a tightly packed protein core, and changes at or in the vicinity of functionally essential residues are equally harmful. The insertion-tolerant regions in proteins commonly reside on surfaces, at termini of secondary structural elements, in loops connecting secondary structural elements, or in linkers between protein domains. Inside the protein core, insertions are not critical only in cases where the amino acid chain extension does not severely clash with the rest of the protein structure. A particular insertion can modulate the enzyme activity via structural changes affecting catalytic residues or substrate/cofactor binding residues either directly or indirectly through long-range effects. Other possible reasons for the decreased enzyme activity include reduced protein stability and increased propensity for protein aggregation,



**Figure 7. RNaseH fold of MuA encompassing the amino acid region E258–G396.** (A) The conserved secondary structural elements, the central  $\beta$ -sheet (with numbered strands 1–5) and the adjacent  $\alpha$ -helix, are shown with light blue and light green, respectively. The DDE-motif residues are shown in yellow with exposed sidechains. Pentapeptide-insertion-tolerant sites G282, G329, and R355 are colour-coded as in Figure 4. (B) Mapping of the alignment-based insertions and deletions. The orange-coloured amino acids depict insertion sites, and the accompanying numbers identify the respective amino acids (insertions occur between the specified residues). Maximum length deletions are coloured with magenta and indicated with respective amino acid residue numbers.  
doi:10.1371/journal.pone.0037922.g007

precipitation, or degradation. All of these phenomena typically involve protein misfolding.

The generated MuA variants displayed a broad variety of activities. While more than half of the variants were totally inactive, some of them exhibited activities substantially exceeding the wild type level (Figure 3B). The inactivating mutations clustered in the globular central domains of MuA (from domain I $\beta$  to III $\alpha$ ). Considering the above scheme, it is conceivable that many of the inactivating insertions, indeed, critically distorted the overall architecture of each particular domain. This would then lead to a domain inactive in its primary function, which in turn would result in the overall halting of transposition. However, the domains function in concert with DNA and ultimately assemble the catalytically competent tetrameric transpososome. The assembly and catalysis require protein-DNA and protein-protein interactions as well as structural transitions. Accordingly, it is plausible that some of the inactivating mutations perturb these critical transactions, thus acting at a higher-order level.

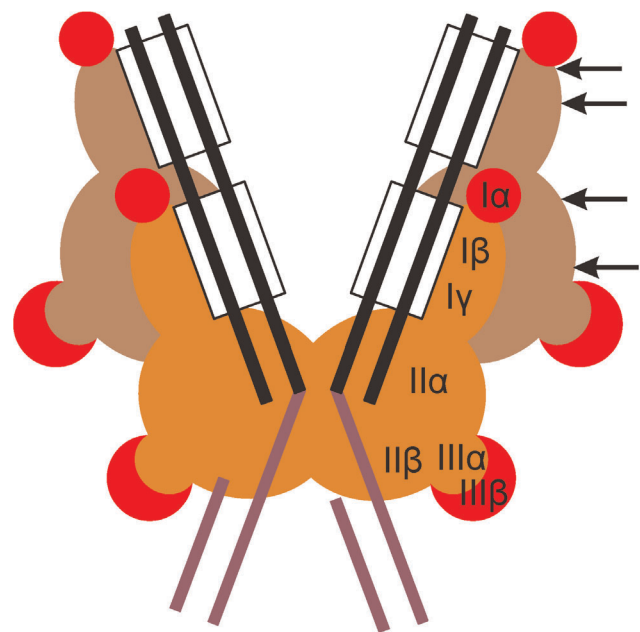
The non-inactivating mutations were mostly confined to the terminal subdomains I $\alpha$  and III $\beta$ , indicating that these subdomains are not critical in the *in vivo* assay used. The data are consistent with our recent MuA deletion variant analysis, indicating dispensability of these subdomains in the same assay [60]. Similar dispensability of subdomains I $\alpha$  and III $\beta$  has also been detected *in vitro* [61]. Also certain surface-exposed loop regions within globular domains and interdomain linker regions tolerated insertions well. Most likely, insertions into these loop regions did not distort the global domain structure, and similarly, did not perturb the higher-order transactions. In fact, a substantial increase in the protein activity suggests that these regions may be critical in structural transitions and elasticity offered by the insertions would be beneficial. Similarly, insertions into the linker regions may have increased the flexibility of the protein leading to advancements in higher-order transactions, from assembly to structural transitions during the progression of transpososome development.

Subdomain I $\alpha$  of MuA is responsible for the IAS binding [49,50]. While the IAS has traditionally been included in Mu transposition assays, it has been shown that it is dispensable in certain reaction conditions *in vitro* [81]. Furthermore, we have recently devised an *in vivo* assay, in which the IAS has been omitted intentionally [60], and this assay was used in the current study. Thus, it was expected that the insertions into subdomain I $\alpha$  would not have had drastic effects on the activity of MuA, and the results were consistent with this prediction. Some level of activity reduction, possibly a result of steric hindrance, could be observed, but nevertheless, none of the insertions into this domain totally abolished the activity. These results are also in concordance with those showing that domain I $\alpha$  can be deleted from the protein without the loss of activity [60,61]. The linker region between domains I $\alpha$  and I $\beta$  (aa 65–89) was highly tolerant of insertions and many insertions actually increased the protein activity. It is plausible that in these cases the transpososome assembly becomes easier, as the insertions into the linker region move the domain I $\alpha$  further away from the Mu end-binding domain, possibly causing less interference. This is in line with the data showing increased activity when the entire subdomain I $\alpha$  is absent [60].

Subdomains I $\beta$  and I $\gamma$  form the DNA-binding module of MuA, which is involved in the recognition and binding into the MuA binding sites located in the transposon ends. Subdomain I $\beta$  did not tolerate insertions within its DNA-binding surface and in structures anchoring the recognition helices, indicating stringency in the architecture. However, certain regions locating more remotely from the recognition helices tolerated insertions, which

indicated flexibility in these structures. Subdomain I $\gamma$  is composed of a helix bundle, and according to our analysis its structural requirements are stringent, as insertions were not tolerated at all. When bound to R1 MuA binding site in MuA transpososome, subdomain I $\gamma$  makes extensive protein-protein contacts (Figure 8), which may also explain the insertion-intolerance. Insertions between the subdomains of the DNA-binding module caused highly deleterious effects. Recent architectural data from the closest structural homologues of the module (DNA binding domains of Mos1 transposase [82] and CENP-B transposon-derived centromere binding protein [83]) indicate that, in contact with DNA, the homologous linker lies deep in the DNA minor groove. Thus, it is highly likely that the linker in MuA behaves similarly, and certainly our insertion data are consistent with this scenario. Recent structural data from the Mu transpososome, although with a relatively low resolution in the region, also suggest minor groove binding for the linker (P. Rice and S.P. Montañó, personal communication).

The catalytic core of MuA contains an RNase H fold (Figure 7) and includes the critical DDE-motif. In general, the folded structure of the core appears highly stringent in its structural constraints. In our analyses neither pentapeptide insertions nor indels of evolutionary origin were allowed in the secondary structural elements of the fold. The rigidity of the core must reflect



**Figure 8. Mu transpososome architecture at post-integration stage and MuA regions allowing pentapeptide insertions.** The overall organization is sketched according to the unpublished crystallographic structure of Mu transpososome (P. Rice and S.P. Montañó, personal communication). The structure constitutes an essential framework for a meaningful interpretation of the functional data (see Discussion). Transposon end segments are shown with black lines, MuA binding sites (R1 and R2) are highlighted with rectangles, and the attached target DNA is shown in magenta. The catalytic MuA protomers are shown in orange and the non-catalytic protomers in brown. Insertion-tolerant subdomains are highlighted with red. The arrows (shown only for one protomer) indicate those linker and loop regions, in which insertions are tolerated well (wild type protein activity retained). The catalytic protomers act *in trans*, i.e. the protomer bound to one end catalyzes DNA cleavage and joining reactions in the other end.

doi:10.1371/journal.pone.0037922.g008



the catalytic requirements and perturbations in the architecture are deleterious. Two loop regions in the catalytic core surface region accommodated insertions with ease reflecting their non-critical role in catalysis. Of particular interest is that insertions in these loops substantially (up to twofold) increased the transpositional activity of MuA. The Mu transpososome crystal structure (P. Rice and S.P. Montaña, personal communication) suggests that mutations in the loop region around aa 282 might allow enhanced flexibility between subdomains II $\alpha$  and II $\beta$ , thus increasing the activity. Further, insertions at aa 474–479 reside in a loop region, which does not make protein-protein or protein-DNA contacts in any of the four MuA protomers within the transpososome. Thus, the activity-enhancing characteristics of insertions in this region may possibly relate to better progression in reaction steps involving structural transitions. Subdomain II $\beta$  folds into a  $\beta$ -barrel structure and is involved in target recognition (Figure 8). In our pentapeptide insertion analysis, this structure appeared rigid, most probably directly reflecting its function in target capture – we observed allowed structural perturbation only in one loop tip (Figure 5D). Consistent with this data, the  $\beta$ -barrel binds target DNA in the Mu transpososome structure, but not with the tips of its loops (P. Rice and S.P. Montaña, personal communication).

Domain III clearly forms two functionally separate units, and this is highlighted also in our study. Subdomain III $\alpha$  is a critical determinant with regard to structural transitions. This is clearly apparent in our study as none of the pentapeptide insertions were tolerated, although some activity was retained with insertion to the subdomain's terminus. In addition, no indels could be discerned in this domain in the evolutionarily relevant sequence comparison. Subdomain III $\beta$  is involved in MuB and ClpX interactions. As MuB is not involved in the papillation assay used, with this regard the domain should be irrelevant. However, as ClpX is encoded by the host *E. coli* strain used in the current study, we expected to see some variation with insertions into this domain. In line with this prediction, deletion of the entire domain has resulted in substantially reduced activity in the same assay [60]. Contrary to our expectations, insertions into subdomain III $\beta$  produced essentially wild type level of activities. The reason for this is currently unknown but may relate to alternative transpososome disassembly pathways (recently discussed in [54]).

MuA functions in the transpososome core as a tetramer synapsing two transposon ends (Figure 8). The protomers within the transpososome are functionally and structurally not equal. Although all of the protomers have a structural role, only two of them function in catalysis. This architecture is based on the initial assembly, during which two of the protomers bind to R1 sites and the other two to R2 sites, and the same arrangement is maintained throughout the transpososome development. Very recent data indicate that the target DNA is contacted only through the residues located in the catalytic protomers (Figure 8), indicating that target capture occurs catalytic-protomer-proximally. The tetrameric nature and non-equal contribution of protomers have consequences, as the tolerated five amino acid insertions need to be tolerated in each of the four protomers within the tetramer.

## References

1. Craig NL, Craigie R, Gellert M, Lambowitz AM, eds (2002) Mobile DNA II. Washington, DC: ASM Press. 1250 p.
2. Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41: 331–368.
3. Kazazian HH Jr. (2004) Mobile elements: Drivers of genome evolution. *Science* 303: 1626–1632.

Which of the protomers are critical with regard to a particular insertion is an interesting question and a subject of future work.

The current study has revealed important determinants with regard to the function of the Mu transpososome. Accordingly, better transposon tools should be possible to design with the aid of the data. We can envision larger additions, even entire domains, into loops that tolerated insertions. In addition, it is plausible that the linker sequences between domains might be modified to yield transposases with enhanced performance characteristics.

## Supporting Information

**Figure S1** A schematic view on the mechanism of Mu DNA transposition. (A) The reaction pathway involving supercoiled plasmid substrates as revealed by a number of *in vitro* studies (reviewed in [28,84]). This pathway recapitulates the essential features of Mu transposition *in vivo*. It involves a supercoiled plasmid carrying Mu left and right end sequences. The left end and right end are in inverted relative orientation, and each end contains three MuA binding sites (L1, L2 and L3 in the left; R1, R2 and R3 in the right). In addition, the transpositional enhancer (E) is included in the plasmid. The transpososome assembly is facilitated via a transient protein-DNA complex LER (gray oval) involving MuA, Mu end sequences and the enhancer. Once assembled, the transpososome (gray circle) function involves helix opening, transposon end cleavage, target capture, and strand transfer. Subsequently, the transpososome is disassembled via action of ClpX. The resulting transposition intermediate is then processed (repaired or replicated) by the use of host functions. (B) The presumed reaction pathway employed by the *in vivo* papillation assay in this study. In comparison to the pathway described in (A), the papillation assay pathway involves only Mu right end sequences (containing R1 and R2 MuA binding sites), and it does not employ transpositional enhancer sequence and MuB protein. Consequently, the pathway does not involve the LER complex, and overall the pathway is less convoluted. (TIF)

**Figure S2** Alignment of 44 MuA family of transposase proteins. (XLSX)

**Table S1** Compilation of data on MuA variants analyzed. (PDF)

## Acknowledgments

We thank Phoebe Rice, Maria I. Pajunen, and Saija Kiljunen for critical reading of the manuscript. Phoebe Rice and Sherwin Montaña are acknowledged for communicating the unpublished Mu transpososome structure and for their help in the structural interpretation of the functional data. We thank Tanja Horko and Jukka Lehtiniemi for their assistance with sequence analysis and figure preparation, respectively.

## Author Contributions

Conceived and designed the experiments: TSR HS. Performed the experiments: TSR SH-P. Analyzed the data: TSR MV HS. Contributed reagents/materials/analysis tools: LP. Wrote the paper: TSR MV HS.

7. Boeke JD (2002) Putting mobile DNA to work: The toolbox. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. *Mobile DNA II*. Washington, DC: ASM Press. pp 24–37.
8. Hayes F (2003) Transposon-based strategies for microbial functional genomics and proteomics. *Annu Rev Genet* 37: 3–29.
9. Ivics Z, Izsvák Z (2010) The expanding universe of transposon technologies for gene and cell engineering. *Mob DNA* 1: 25.
10. Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, et al. (1990) Three-dimensional structure of ribonuclease H from *E. coli*. *Nature* 347: 306–309.
11. Yang W, Hendrickson WA, Crouch RJ, Satow Y (1990) Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* 249: 1398–1405.
12. Rice PA, Baker TA (2001) Comparative architecture of transposase and integrase complexes. *Nat Struct Biol* 8: 302–307.
13. Nowotny M (2009) Retroviral integrase superfamily: The structural perspective. *EMBO Rep* 10: 144–151.
14. Haren L, Ton-Hoang B, Chandler M (1999) Integrating DNA: Transposases and retroviral integrases. *Annu Rev Microbiol* 53: 245–281.
15. Zhou L, Mitra R, Atkinson PW, Hickman AB, Dyda F, et al. (2004) Transposition of *hAT* elements links transposable elements and V(D)J recombination. *Nature* 432: 995–1001.
16. Gellert M (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 71: 101–132.
17. Rivas FV, Tolia NH, Song JJ, Aragon JP, Liu J, et al. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* 12: 340–349.
18. Howe MM, Bade EG (1975) Molecular biology of bacteriophage Mu. *Science* 190: 624–632.
19. Mizuuchi K (1983) *In vitro* transposition of bacteriophage Mu: A biochemical approach to a novel replication reaction. *Cell* 35: 785–794.
20. Mizuuchi K (1992) Transpositional recombination: Mechanistic insights from studies of Mu and other elements. *Annu Rev Biochem* 61: 1011–1051.
21. Rice P, Craigie R, Davies DR (1996) Retroviral integrases and their cousins. *Curr Opin Struct Biol* 6: 76–83.
22. Nesmelova IV, Hackett PB (2010) DDE transposases: Structural similarity and diversity. *Adv Drug Deliv Rev* 62: 1187–1195.
23. Surette MG, Buch SJ, Chaconas G (1987) Transpososomes: Stable protein-DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. *Cell* 49: 253–262.
24. Craigie R, 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.
25. Lavoie BD, Chan BS, Allison RG, 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.
26. Baker TA, Mizuuchi M, Savilahti H, Mizuuchi K (1993) Division of labor among monomers within the Mu transposase tetramer. *Cell* 74: 723–733.
27. Savilahti H, Rice PA, Mizuuchi K (1995) The phage Mu transpososome core: DNA requirements for assembly and function. *EMBO J* 14: 4893–4903.
28. Chaconas G, Harshey RM (2002) Transposition of phage Mu DNA. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. *Mobile DNA II*. Washington, DC: ASM Press. pp 384–402.
29. Haapa S, Taira S, Heikkinen E, 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.
30. Aldaz H, Schuster E, Baker TA (1996) The interwoven architecture of the Mu transposase couples DNA synthesis to catalysis. *Cell* 85: 257–269.
31. Savilahti H, Mizuuchi K (1996) Mu transpositional recombination: Donor DNA cleavage and strand transfer *in trans* by the Mu transposase. *Cell* 85: 271–280.
32. Yang J-Y, Jayaram M, Harshey RM (1996) Positional information within the Mu transposase tetramer: Catalytic contributions of individual monomers. *Cell* 85: 447–455.
33. Haapa S, Suomalainen S, Eerikainen S, Airaksinen M, Paulin L, et al. (1999) An efficient DNA sequencing strategy based on the bacteriophage Mu *in vitro* DNA transposition reaction. *Genome Res* 9: 308–315.
34. Orsini L, Pajunen M, Hanski I, Savilahti H (2007) SNP discovery by mismatch-targeting of Mu transposition. *Nucleic Acids Res* 35: e44.
35. Pajunen M, Turakainen H, Poussu E, Peränen J, Vihinen M, et al. (2007) High-precision mapping of protein-protein interfaces: An integrated genetic strategy combining *en masse* mutagenesis and DNA-level parallel analysis on a yeast two-hybrid platform. *Nucleic Acids Res* 35: e103.
36. Poussu E, Jantti J, Savilahti H (2005) A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: Mapping of the Sec1p binding domain in yeast Mso1p by a Mu *in vitro* transposition-based approach. *Nucleic Acids Res* 33: e104.
37. Poussu E, Vihinen M, Paulin L, Savilahti H (2004) Probing the  $\alpha$ -complementing domain of *E. coli*  $\beta$ -galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu *in vitro* DNA transposition. *Proteins* 54: 681–692.
38. Brady T, Roth SL, Malani N, Wang GP, Berry CC, et al. (2011) A method to sequence and quantify DNA integration for monitoring outcome in gene therapy. *Nucleic Acids Res* 39: e72.
39. Krupović M, Vilen H, Bamford JKH, Kivelä HM, Aalto J-M, et al. (2006) Genome characterization of lipid-containing marine bacteriophage PM2 by transposon insertion mutagenesis. *J Virol* 80: 9270–9278.
40. Pajunen MI, Pulliainen AT, Finne J, Savilahti H (2005) Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes. *Microbiology* 151: 1209–1218.
41. Vilen H, Aalto J-M, Kassinen A, Paulin L, Savilahti H (2003) A direct transposon insertion tool for modification and functional analysis of viral genomes. *J Virol* 77: 123–134.
42. Lamberg A, Nieminen S, Qiao M, Savilahti H (2002) Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of *in vitro*-assembled DNA transposition complexes of bacteriophage Mu. *Appl Environ Microbiol* 68: 705–712.
43. Paatero AO, Turakainen H, Happonen LJ, Olsson C, Palomäki T, et al. (2008) Bacteriophage Mu integration in yeast and mammalian genomes. *Nucleic Acids Res* 36: e148.
44. Nakayama C, Teplow DB, Harshey RM (1987) Structural domains in phage Mu transposase: Identification of the site-specific DNA-binding domain. *Proc Natl Acad Sci U S A* 84: 1809–1813.
45. Clubb RT, Omichinski JG, Savilahti H, Mizuuchi K, Gronenborn AM, et al. (1994) A novel class of winged helix-turn-helix protein: The DNA-binding domain of Mu transposase. *Structure* 2: 1041–1048.
46. Clubb RT, Schumacher S, Mizuuchi K, Gronenborn AM, Clore GM (1997) Solution structure of the I $\gamma$  subdomain of the Mu end DNA-binding domain of phage Mu transposase. *J Mol Biol* 273: 19–25.
47. Schumacher S, Clubb RT, Cai M, Mizuuchi K, Clore GM, et al. (1997) Solution structure of the Mu end DNA-binding I $\beta$  subdomain of phage Mu transposase: Modular DNA recognition by two tethered domains. *EMBO J* 16: 7532–7541.
48. Rice P, Mizuuchi K (1995) Structure of the bacteriophage Mu transposase core: A common structural motif for DNA transposition and retroviral integration. *Cell* 82: 209–220.
49. Leung PC, Teplow DB, Harshey RM (1989) Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature* 338: 656–658.
50. Mizuuchi M, 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.
51. Baker TA, Luo L (1994) Identification of residues in the Mu transposase essential for catalysis. *Proc Natl Acad Sci U S A* 91: 6654–6658.
52. Kremontsova E, Giffin MJ, Pincus D, Baker TA (1998) Mutational analysis of the Mu transposase: Contributions of two distinct regions of domain II to recombination. *J Biol Chem* 273: 31358–31365.
53. Wu Z, Chaconas G (1995) A novel DNA binding and nuclease activity in domain III of Mu transposase: Evidence for a catalytic region involved in donor cleavage. *EMBO J* 14: 3835–3843.
54. Choi W, Harshey RM (2010) DNA repair by the cryptic endonuclease activity of the Mu transposase. *Proc Natl Acad Sci U S A* 107: 10014–10019.
55. Harshey RM, Cuneo SD (1986) Carboxyl-terminal mutants of phage Mu transposase. *J Genet* 65: 159–174.
56. Baker TA, Mizuuchi M, Mizuuchi K (1991) MuB protein allosterically activates strand transfer by the transposase of phage Mu. *Cell* 65: 1003–1013.
57. Leung PC, Harshey RM (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: 189–199.
58. Wu Z, 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.
59. Levchenko I, Luo L, Baker TA (1995) Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev* 9: 2399–2408.
60. Pajunen MI, Rasila TS, Happonen LJ, Lamberg A, Haapa-Paananen S, et al. (2010) Universal platform for quantitative analysis of DNA transposition. *Mob DNA* 1: 24.
61. Kim YC, Morrison SL (2009) N-terminal domain-deleted Mu transposase exhibits increased transposition activity with low target site preference in modified buffers. *J Mol Microbiol Biotechnol* 17: 30–40.
62. Yuan JF, Beniac DR, Chaconas G, Ottensmeyer FP (2005) 3D reconstruction of the Mu transposase and the type 1 transpososome: A structural framework for Mu DNA transposition. *Genes Dev* 19: 840–852.
63. Goldhaber-Gordon I, Williams TL, Baker TA (2002) DNA recognition sites activate MuA transposase to perform transposition of non-Mu DNA. *J Biol Chem* 277: 7694–7702.
64. Goldhaber-Gordon I, Early MH, Gray MK, Baker TA (2002) Sequence and positional requirements for DNA sites in a Mu transpososome. *J Biol Chem* 277: 7703–7712.
65. Saariaho A-H, Savilahti H (2006) Characteristics of MuA transposase-catalyzed processing of model transposon end DNA hairpin substrates. *Nucleic Acids Res* 34: 3139–3149.
66. Saariaho A-H, Lamberg A, Elo S, Savilahti H (2005) Functional comparison of the transposition core machineries of phage Mu and *Haemophilus influenzae* Mu-like prophage Hin-Mu reveals interchangeable components. *Virology* 331: 6–19.
67. Grant SGN, Jessee J, Bloom FR, Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci U S A* 87: 4645–4649.

68. Sambrook J, Russell DW (2001) Molecular cloning: A laboratory manual. Harbor ColdSpring, ed. N.Y.: Cold Spring Harbor Laboratory Press. 2344 p.
69. Hanahan D, Jesse J, Bloom FR (1991) Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol* 204: 63–113.
70. Rasila TS, Pajunen MI, Savilahti H (2009) Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutant strain, and hydroxylamine treatment. *Anal Biochem* 388: 71–80.
71. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
72. Finn RD, Mistry J, Tate J, Coghill P, Heger A, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38 (Database issue). pp D211–D222.
73. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242.
74. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577–2637.
75. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.
76. Petyuk V, McDermott J, Cook M, Sauer B (2004) Functional mapping of Cre recombinase by pentapeptide insertional mutagenesis. *J Biol Chem* 279: 37040–37048.
77. Weber M, Chernov K, Turakainen H, Wohlfahrt G, Pajunen M, et al. (2010) Mso1p regulates membrane fusion through interactions with the putative N-peptide-binding area in Sec1p domain 1. *Mol Biol Cell* 21: 1362–1374.
78. Fransen M, Vastiau I, Brees C, Brys V, Mannaerts GP, et al. (2005) Analysis of human Pex19p's domain structure by pentapeptide scanning mutagenesis. *J Mol Biol* 346: 1275–1286.
79. von Nandelstadh P, Grönholm M, Moza M, Lamberg A, Savilahti H, et al. (2005) Actin-organising properties of the muscular dystrophy protein myotilin. *Exp Cell Res* 310: 131–139.
80. Grönholm M, Muranen T, Toby GG, Utermarck T, Hanemann CO, et al. (2006) A functional association between merlin and HEI10, a cell cycle regulator. *Oncogene* 25: 4389–4398.
81. Baker TA, Mizuuchi K (1992) DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev* 6: 2221–2232.
82. Richardson JM, Colloms SD, Finnegan DJ, Walkinshaw MD (2009) Molecular architecture of the Mos1 paired-end complex: The structural basis of DNA transposition in a eukaryote. *Cell* 138: 1096–1108.
83. Tanaka Y, Nureki O, Kurumizaka H, Fukai S, Kawaguchi S, et al. (2001) Crystal structure of the CENP-B protein-DNA complex: The DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO J* 20: 6612–6618.
84. Akhverdyan VZ, Gak ER, Tokmakova IL, Stoynova NV, Yomantas YAV, et al. (2011) Application of the bacteriophage Mu-driven system for the integration/amplification of target genes in the chromosomes of engineered Gram-negative bacteria—mini review. *Appl Microbiol Biotechnol* 91: 857–871.