Supplementary Material For:

Transposon-directed base-exchange mutagenesis (TDEM): a novel method for multiple-nucleotide substitutions within a target gene

Yun Cheol Kim[1], Hui Sun Lee[3], Sukjoon Yoon[3], and Sherie L. Morrison[1,2]

[1]Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles, CA, USA, [2]Molecular Biology Institute, David Geffern School of Medicine, University of California Los Angeles, Los Angeles, CA, USA, and [3]Sookmyung Women's University, Department of Biological Sciences, Research Center for Women's Diseases (RCWD), Seoul, Republic of Korea

BioTechniques 46:534-542 (June 2009) doi 10.2144/000113152

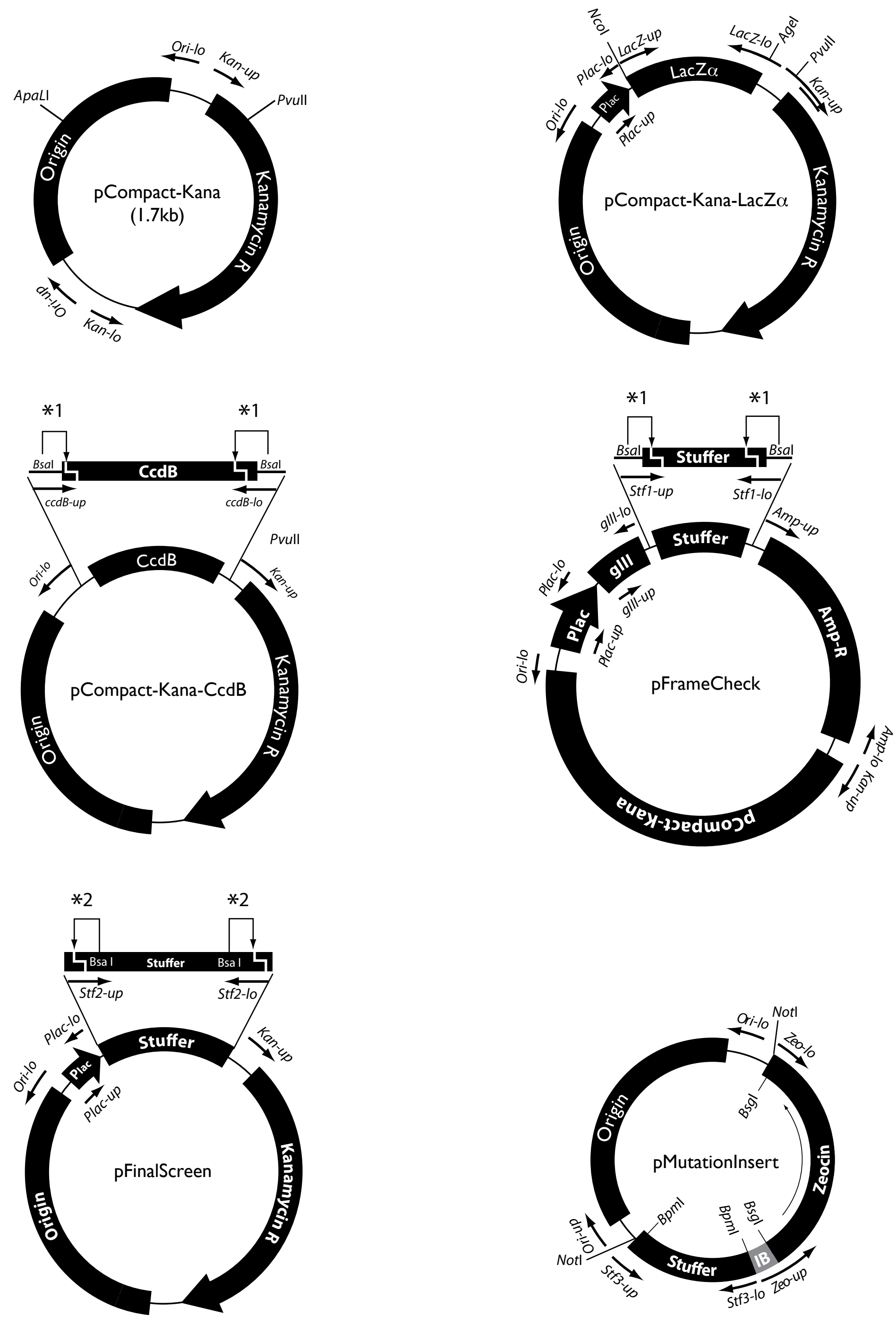
Plasmid constructs and transposons

To construct pCompact-Kana-LacZα (Supplementary Figure 1), the N-terminal 150 bp of β-galactosidase gene (*lacZα*) and the *E. coli* lactose promoter (Plac) were amplified by PCR and ligated (Supplementary Table 1). pCompact-Kana-CcdB (Supplementary Figure 1) was generated by amplifying the *ccdB* gene (300 bp) using primers (CcdB-up and CcdB-lo) that contain *Bsa*I sites (Supplementary Table 1). To construct pFrameCheck (Supplementary Figure 1), the *E. coli* lactose promoter, the gIII signal sequence (40) of the minor capsid protein from the filamentous phage fd, stuffer DNA, and the β-lactamase gene were individually amplified using the indicated primers (Supplementary Table 1) and ligated in the order indicated in Supplementary Figure 1. For constructing pFinalScreen (Supplementary Figure 1), the Plac and stuffer DNA were amplified using PCR with the indicated primers and ligated (Supplementary Table 1). For constructing mutation inserts (MI) (Supplementary Figure 1), the Zeocin resistance gene cassette and the stuffer DNA were amplified and ligated as indicated (Supplementary Figure 1 and Supplementary Table 1). Three incorporating bases (IB) were located in the Zeo-up primers. The IBs in the primers were selected to represent codons for the 20 different amino acids (Supplementary Table 1). The stuffer ligated to the Zeocin resistance gene cassette was ligated to a replication origin (Supplementary Figure 1). Each pMutationInsert containing different IBs was individually constructed. For preparing a mixture of MIs (MI-mix), equal amounts of each vector were mixed and digested with *Not*I (New England Biolabs, Ipswich, MA, USA) followed by gel purification.

For the Tn7 transposon, pGPS2.1 (New England Biolabs) was modified to contain *Not*I sites flanking the right (R) and left (L) consensus sequences of Tn7 using PCR with primer #3379 (5′-TGCGGCCGCACAATAAAGTCTTAAACTGAACAAAA-3′) and #3382 (5′-TGCGGCCGCACAAAATAGTTGGGAACTGGGA-3′). A *Not*I site inside Tn7 was removed. For the Mu transposon, M1-Cam^R (Finnzymes, Espoo, Finland) was used.

Mutagenesis

The plasmids containing the transposon in the target gene were digested with *Not*I to remove the transposon sequence and then gel-purified. The linearized plasmid DNA was ligated with an MI or MI-mix. After electrotransformation of the ligated mixture, the *E. coli* culture was grown in 50 μg/mL kanamycin and 25 μg/mL Zeocin (Invitrogen, Carlsbad, CA) to select clones containing the MI. The plasmid was purified and digested with *Bsg*I (New England Biolabs) followed by treatment with T4 DNA polymerase (New England Biolabs) according to the manufacturer's instruction to generate blunt ends. After agarose gel purification, the vector DNA was self-ligated and electrotransformed into *E. coli* and grown in LB containing 50 μg/mL kanamycin at 37°C overnight with shaking (Excella E24, New Brunswick Scientific, Edison, NJ, USA). The following day, the plasmids were again purified, digested with *Bpm*I (New England Biolabs) followed by the same procedure as for *Bsg*I digestion. For the *lacZα* mutagenesis, *E. coli* strain Top10 was used and for the *ccdB* mutagenesis, DB3.1, which is resistant to the *ccdB*'s cytotoxic activity was used except for the last step of screening for *E. coli* survival in the presence of inactive *ccdB*. An error-prone PCR (epPCR) mutagenesis kit (Stratagene, La Jolla, CA, USA) was used according to the manufacturer's instruction for comparison to TDEM. After 40 cycles with primers [epLacZ-up: 5′-ATGAT**CCATGG**GCACCATGATTACGGATTCACTGGC-3′, and epLacZ-lo: 5′-CGCCA**ACCGGT**TACTGTTGGGAAGGGCGATCGGTGCG-3′ (underlines indicate *Nco*I and *Age*I sites, respectively)], the amplified PCR band was purified from an agarose gel and digested with *Nco*I and *Age*I. The digested DNA was cloned into the pCompact-Kana vector prepared by digesting pCompact-Kana-LacZα with *Nco*I and *Age*I.

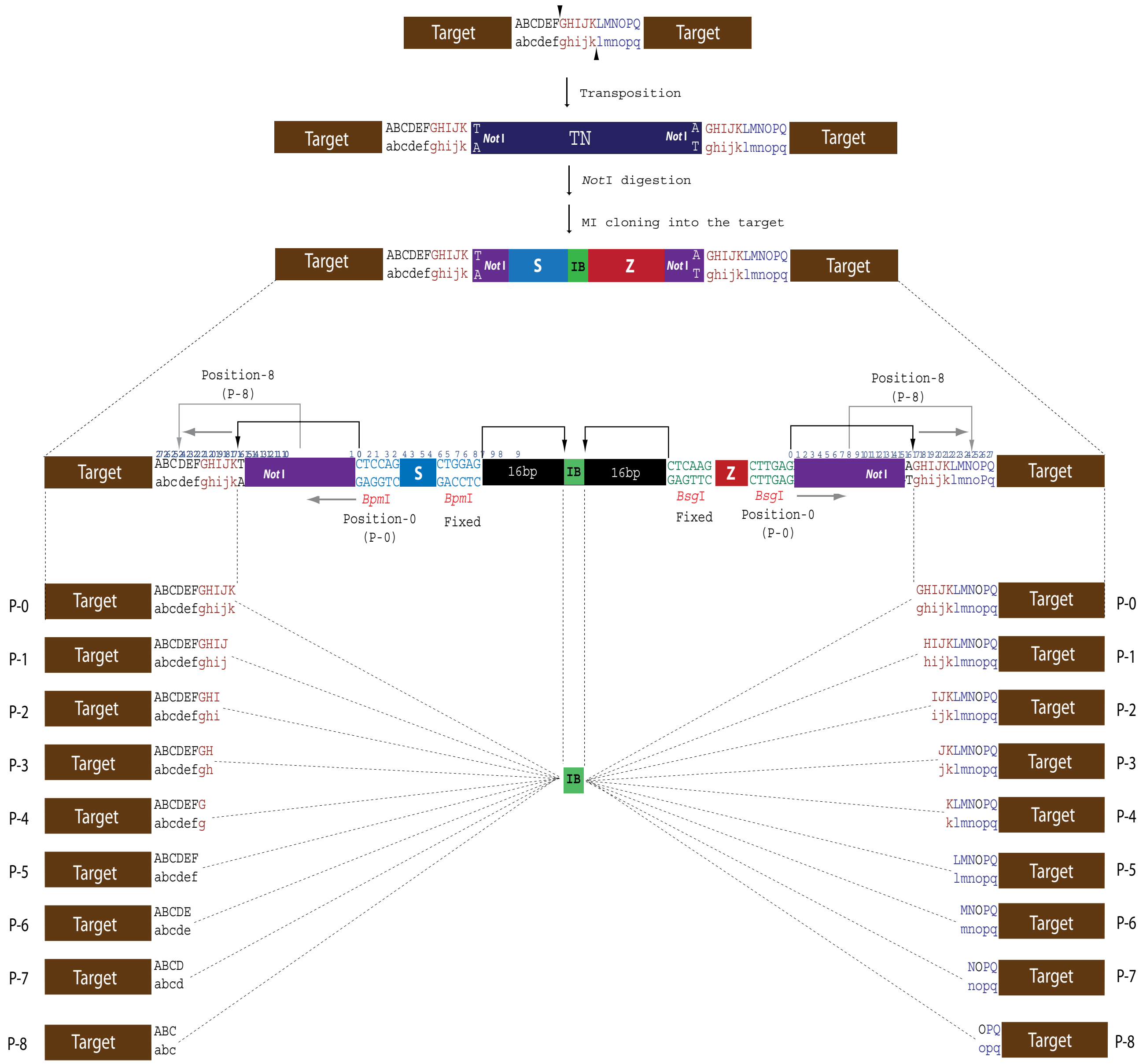


Supplementary Figure 1. Map for the vectors used in this study. The primers used for making each vector are depicted as arrows with its name. The sequences for the primers are in Supplementary Table 1. To construct each vector, PCR products amplified using the primers indicated were ligated and transformed into *E. coli*. *1 and *2 represent two reverse orientations of *Bsa*I sites in the vectors. In pFinalScreen, *Bsa*I cuts sites outside its recognition sequences, however in pFrameCheck, *Bsa*I cut sites inside its recognition sequences. Plac, *E. coli* lactose promoter; Amp-R, β-lactamase; IB, incorporating bases; Kanamycin R, kanamycin resistance gene)

Supplementary Table 1. DNA Fragments and Primer Sequences for PCR Amplification Used for Constructing Vectors

Vector	DNA fragment	Primer	5' Sequence 3'	Template
pCompact-Kana	Origin of replication	Ori-up	GATTTAAAACTTCATTTTTAATTTAAAAG	pUC19
		Ori-lo	CATGTGAGCAAAAGGCCAGCAAAAGGCC	
	Kanamycin resistance gene	Kan-up	CTGGGGCGCCCTC TGGTAAGGTT	pCR Blunt II-TOPO
		Kan-lo	TCAGAAGAACTCGTCAAGAAGGC	
pCompact-Kana-LacZα	lactose promoter	plac-up	GGCAGTGAGCGCAACGCAATTAATGT	*E. coli* genomic DNA
		plac-lo	GTTTCCTGTGTGAAATTGTTATCCGCTCA	
	LacZα	LacZ-up	[1]CCATGGGCACCATGATTACGGATTCACTGGC	pCR Blunt II-TOPO
		LacZ-lo	[2]ACCGGTTACTGTTGGGAAGGGCGATCGGTGCG	
pCompact-Kana-CcdB	CcdB	CcdB-up	G[3]GGTCTCAATGGGAATGCAG TTTAAGGTTTACAC-CTATAAAA	pCR Blunt II-TOPO
		CcdB-lo	CA[3]GGTCTCCGGTTCCTATTCCCCAGAACATCAGGTT	
pFrameCheck	Lactose promoter	lacp-up	GTTCTTTCCTGCGTTATCCCCTGATTCT	pCR Blunt II-TOPO
		lacp-lo	AGCTGTCCTCCTGTGTGAAATTGTTATCCGCTCAC	
	gIII signal sequence	gIII-up	ATGAAAAAACTGCTGTTCG	pBAD/gIII
		gIII-lo	CAGTTTAAACAGCTCGAGCTCCATGGTGCTAT	
	Stuffer DNA	Stf1-up	GA[3]GGTCTCTATGGCCATGATTACGCCAAGCTT	pUC19
		Stf1-lo	GCCTCCACCTCCCAGTTTAAACAGAAGGTCTCCG-GTTCCTGCGGCATC [3]	
	β-lactamase gene	Amp-up	TCCGGAGGTGGAGGCTCCGTTTTTGCTCACCCA-GAAACGCTGGTGAAA	pUC19
		Amp-lo	TTACCAATGCTTAATCAGTGAGG	
pFinalScreen	Stuffer DNA	Stf2-up	ATGGT[3]GAGACCAGACCATGATTACGCCAAGCTTGC	pUC19
		Stf2-lo	TTATTTAGGTTAGAGACCTTATGCGGCATCAGAG-CAGATTGTACTGAGA [3]	
pMutationInsert	Zeocin resistance gene	Zeo-up	*NNNGCCCTACTCTACTGTCGACTGCAC CTGTGCG-GTATTTCACACCGCATACAGGTG [4]	pZero-1
		Zeo-lo	GCGGCCGCCACCTGCACTCACCTAGATCCTTT-TAAATTAAAAATG [5] [4]	
	Stuffer DNA	Stf3-up	[5]GCGGCCGCGCC[6]CTCCAGCAGCCGCACGCGGCG-CATCTCGGGCAGCGTTGGGTCCTG	pET29a
		Stf3-lo	CCAGCTCGTTGAGTTT[6]CTCCAGAAGCGTTAAT-GTCTGGCTTCTGATAAAGCGGGCCATGTTA	

* The NNN sequence was either GCC, CGC, AAC, GAT, TGC, CAG, GAA, GGC, CAT, ATC, CTG, AAA, ATG, TTT, CCA, AGC, ACC, TGG, TAC or GTG. The sequences underlined represent restriction enzyme sites as follow. 1: Nco I, 2: Age I, 3: Bsa I, 4: Bsg I, 5: Not I and 6: Bpm I.

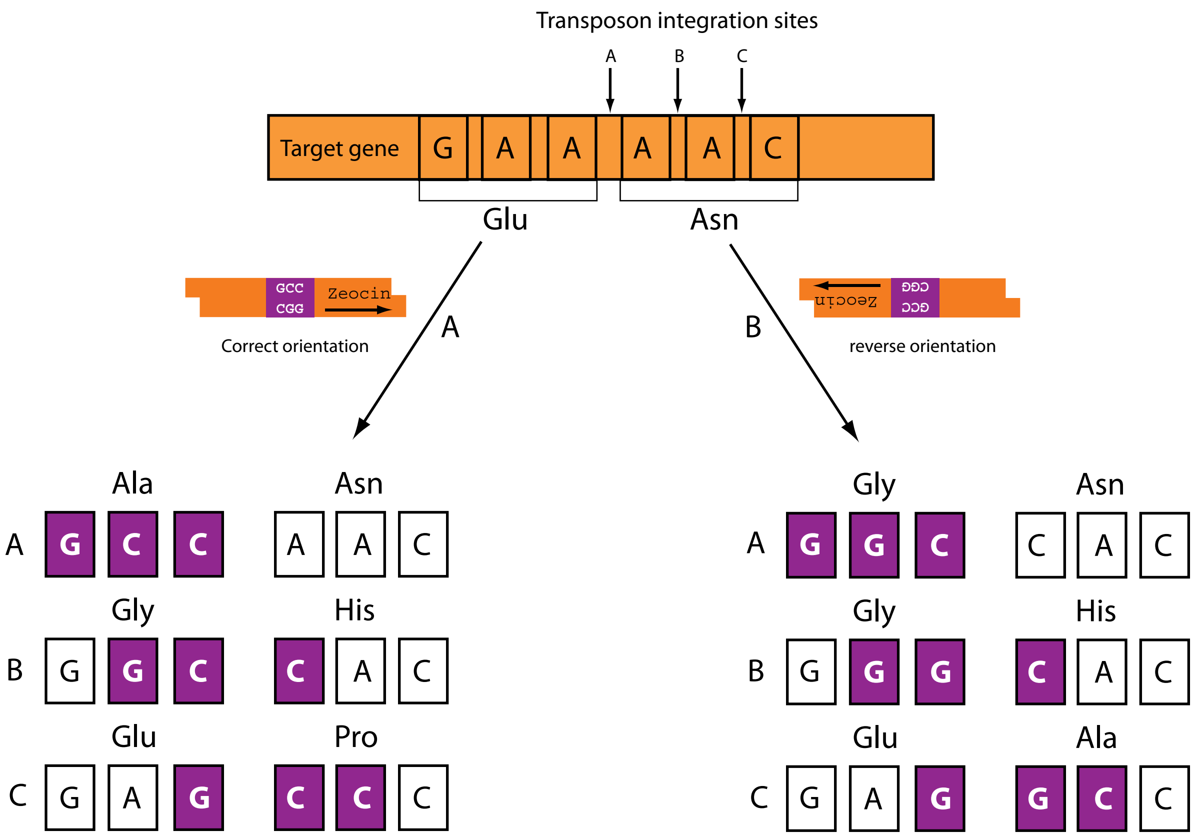


Supplementary Figure 2. Different structures of the mutation insert (MI) will result in different mutations. (A) The transposon integrated into the target has a *Not*I site with an additional A or T base at each end. (B) Following *Not*I digestion, the Mu-transposon is replaced with an MI. The position of the distal *Bpm*I and *Bsg*I sites in the MI can be moved to vary the number of bases to be removed from the target. In diagram C, the *Bpm*I and *Bsg*I sites are located at position 0 at left and right ends. These distal *Bpm*I and *Bsg*I sites can be moved up to eight bases toward each end. By moving these sites, the bases to be removed will be altered, as shown in D, because moving the site 1 base nearer the end results in 1 more base being deleted form the target. Furthermore, the *Bpm*I site and the *Bsg*I site can be repositioned independently and the IB sequence can be any length so that a sequence of any length can be introduced between the proximal *Bpm*I and *Bsg*I sites, which are fixed. Therefore, TDEM with MIs containing the distal *Bpm*I and *Bsg*I sites at different positions and with IBs of different sequences can be used to produce diverse mutant libraries. For example, if both *Bpm*I and *Bsg*I sites are at P-8 and the IB is of 11 bases, TDEM will result in an 11-base substitution. If the *Bpm*I and *Bsg*I position are at P-0 and P-8 respectively and the IB is of three bases, the result of TDEM will be a three base substitution. If the *Bpm*I and *Bsg*I sites are at positions P-5 and P-6, respectively, and the IB is of 6 bases, the result of TDEM will be a 6-base substitution. TN, transposon; S, stuffer DNA; Z, Zeocin resistance gene; IB, incorporating bases.

Supplementary Table 2. The Position and Sequence of Mutations Within *lacZα* Obtained Using Different Mutagenesis Methods

Round of mutagenesis	Phenotype	Frequency[1] (Total no. of colonies)	No. of colonies sequenced	Mutants	Substitution	Occurrence[2]
[4]1st round of TDEM with GCC-MI into LacZα	White	52% (1045)	8	N19P	AAC→CCC	2
				Q24 (+2)[3]	CAA→CGCCA	1
				R27 (-1)	CGC→C-C	1
				G21 (+1)	GGC→GGCC	2
				L25 (-1)	CTT→C-T	1
				P32 (+2)	CCC→CCGCC	1
	Pale blue	2% (1045)	2	L8A	CTG→GCC	1
				P32R	CCCCCT→CGCCCT	1
2nd round of TDEM with GCC-MI into N19P	Dark blue	1.2% (2531)	15	P19A	CCC→GCC	15
	Pale blue	0.4% (2531)	3	P19G	CCC→GGC	1
				Q24G	CAA→GGC	2
2nd round of TDEM with MI-mix into N19P	Dark blue	2.2% (2320)	20	P19A	CCC→GCC	2
				P19D	CCC→GAT	2
				P19H	CCC→CAT	1
				P19I	CCC→ATC	3
				P19N	GAGCCC→GAAAAC	2
				P19Q	CCC→CAG	1
				P19R	CCC→CGC	1
				P19S	GAGCCC→GAATCC	3
				P19T	CCC→ACC	2
				E18D+P19T	GAGCCC→GACACC	2
				E18D+P19H	GAGCCC→GACCAC	1
	Pale blue	5.2% (2320)	5	P19G	CCC→GGC	2
				Q24G	CAA→GGC	1
				T44K+D45N	ACCGAT→AAAAAT	2
epPCR into N19P	Dark blue	0.6% (856)	5	P19A	CCC→GCC	1
				P19H	CCC→CAC	1
				P19L	CCC→CTC	1
				P19R	CCC→CGC	1
				P19S	CCC→TCT	1
	Pale blue	1.2% (856)	NA	NA	NA	NA

[1]The percent of colonies showing indicated phenotype in total colonies observed.
[2]The number of occurrence in colonies sequenced.
[3]The number in parenthesis indicates the number of extrabases added (+) or deleted (-). NA: Not tested. 4: 46 % of colonies showed dark blue phenotype.



Supplementary Figure 3. An example of amino acids changes following TDEM with GCC-MI. The amino acid changes are dependent on the integration site and orientation of the MI. The MI will substitute three bases, which will alter one or two codons. Furthermore, depending on the orientation (A or B) of the MI in the target gene, the sequence found in the mutant will be different. Therefore, if the GCC-MI is used, mutants containing GGC (B) as well as mutants containing GCC (A) will result. In this example, if GCC-MI is used for TDEM, five different mutants will result depending on the frame and orientation of the MI.

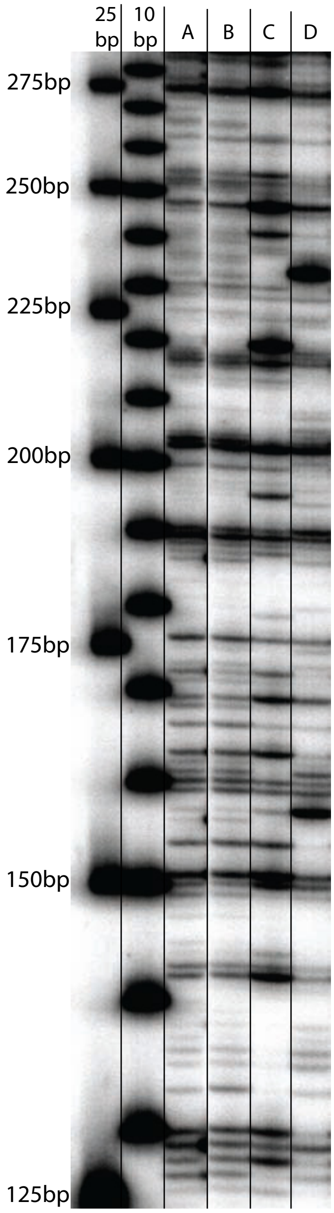
Supplementary Table 3. The Mutants of HIV Protease Generated by TDEM Using MI-mix

No.	Mutant	[1]Occur-rence
1	K20N+E21Q	1
2	D29N	1
3	P39H	1
4	G40X	1
5	G48W	1
6	G51X[2]	1
7	G52P	1
8	G52F	1
9	R57N	1
10	R57A	1
11	Q58A	2
12	Q58D	1
13	Y59D	1
14	Y59A	1
15	Y59I	1
16	L63A	1
17	L63H+I64V	1
18	L63H+I64L	1
19	I64S	1
20	C67D	1
21	C67X	2
22	C67X+G68S	1
23	C67X+G68K	1
24	C67W+G68R	3
25	C67W+G68L	1
26	C67W+G68K	1
27	C67W+G68X	2
28	G68Q	1
29	A71G	1
30	A71G+I72V	1
31	A71D+I72V	1
32	A71V+I72V	1
33	A71G+I72L	3
34	A71E+I72L	1
35	I72V	1
36	I72S	1
37	G73C	2
38	I72M+G73R	1
39	I72N+G73R	1
40	I72K+G73S	2
41	I72L+G73R	1
42	G73D+T74P	2
43	G73E+T74P	1
44	G73D+T74A	1
45	T74P	3
46	T74S	1
47	T74L	2
48	T74V	1
49	T74N	1
50	T74R+V75L	1
51	V77D	1
52	V77E+G78R	1
53	V77G+G78R	1
54	V77G+G78X	1
55	T80P	1
56	I85N+G86R	1
57	L89V	1
58	L90F+T91S	1
59	L90C+T91P	1
60	L90V	1

From 73 clones, 60 different mutations were isolated with 13 mutations present more than once. Among these 60 mutations, 47% contained two consecutive amino acid changes, 53% contained a single amino-acid substitution and 9% contained a stop codon.

[1]occurrence designates the number of times this mutation is represented in the 73 sequenes.

[2]X indicates a stop codon



Supplementary Figure 4. Analysis of the sites in the *ccdB* gene attacked by Mu-transposon. Transposition was performed into the *ccdB* gene using different buffers (A, B, C, and D). The transposon was removed using *Not*I digestion and the linearized plasmid opened at the site of the integration was labeled with ^{32}P. Labeled plasmid was then digested with *Pvu*II, which cuts outside of the *ccdB* gene, resulting in DNA fragments of different sizes. The fragments were analyzed using a 6% polyacrylamide gel containing 5 M urea. 25bp, 25-bp ladder (Invitrogen); 10bp, 10-bp ladder (Invitrogen); A/B/C/D, transposition using buffer A/B/C/D, respectively (see "Materials and methods" section).