Nucleic Acids Research, 1999, Vol. 27, No. 19 **3859–3865**

In vitro transposition of Tn*552*: a tool for DNA sequencing and mutagenesis

Thomas J. Griffin IV, Linda Parsons[1], Andres E. Leschziner, Justine DeVost[1], Keith M. Derbyshire[1] and Nigel D. F. Grindley*

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114, USA and [1]Wadsworth Center, David Axelrod Institute for Public Health, Albany, NY 12201-2002, USA

Received May 27, 1999; Revised and Accepted July 23, 1999

DDBJ/EMBL/GenBank accession no. AF157643

ABSTRACT

We have explored the potential of the Tn*552 in vitro* transposition reaction as a genetic tool. The reaction is simple (requiring a single protein component), robust and efficient, readily producing insertions into several percent of target DNA. Most importantly, Tn*552* insertions *in vitro* appear to be essentially random. Extensive analyses indicate that the transposon exhibits no significant regional or sequence specificity for target DNA and leaves no discernible 'cold' spots devoid of insertions. The utility of the *in vitro* reaction for DNA sequencing was demonstrated with a cosmid containing the *Mycobacterium smegmatis recBCD* gene cluster. The nucleotide sequence of the entire operon was determined using 71 independent Tn*552* insertions, which generated over 13.5 kb of unique sequence and simultaneously provided a comprehensive collection of insertion mutants. The relatively short ends of Tn*552* make construction of novel transposons a simple process and we describe several useful derivatives. The data presented suggest that Tn*552* transposition is a valuable addition to the arsenal of tools available for molecular biology and genomics.

INTRODUCTION

Transposons have proved to be invaluable and versatile tools for molecular genetics for many years (1,2). These uses have included mutagenesis and gene identification, the production of reporter gene libraries for analysis of gene transcription, translation and protein localization, mapping and DNA sequencing of cloned genes, etc. For all these applications, transposon insertions most often have been obtained *in vivo*, using suicide phage or plasmid delivery vehicles or a 'mating-out' procedure. The development of several *in vitro* transposition systems in recent years has opened up the possibility of replacing the traditional *in vivo* genetic methods with more streamlined and efficient biochemical procedures (3–12).

In vitro transposition systems potentially provide several advantages over the *in vivo* approaches. In many cases efficiency is substantially greater, facilitating the generation of very large numbers of independent insertions in a single reaction. Targets of any topology may be used (including linear DNA fragments) and can be taken from any organism, including ones that lack amenable genetic systems. In addition, target preferences observed *in vivo* that result from host factors or features of gene expression may be eliminated *in vitro*, resulting in a more uniform distribution of insertions.

Ideally, for practical use as a tool for mutagenesis and DNA sequencing, an *in vitro* transposition system must readily yield large numbers of recoverable, independent, random, simple insertions into any DNA target of choice. Thus, efficiency of the insertion reaction and its randomness (with respect to both DNA sequence and regional target specificities) are the most important criteria. Other important attributes are easy preparation of the transposase and the transposon substrate, a simple and robust transposition reaction and versatility of the transposon. In this paper, we show that the insertion reaction recently developed for Tn*552* (13) fulfills these criteria.

Tn*552* is a bacterial transposon, initially identified in *Staphylococcus aureus* (14,15) but also active in *Escherichia coli* (T.J.Griffin and N.D.F.Grindley, unpublished observations). Of the well-characterized transposons, Tn*552* is most similar to Mu (16), encoding a single subunit transposase [TnpA, a 480 residue polypeptide of the D, D(35)E family] plus an accessory protein (TnpB) that is required for transposition *in vivo* (and, by analogy to MuB, may play a role in target selection and immunity). The ends of Tn*552*, which are the only sites required for its transposition, are much simpler than those of Mu, however, consisting of 48 bp terminal inverted repeats (each with two abutting 23 bp transposase binding sites) (13,17). *In vivo* transposition of Tn*552*, like that of Mu, occurs by a replicative process to produce co-integrates. The *in vitro* reaction for insertion of Tn*552* considerably simplifies this process; starting with a linear version of the transposon, efficient strand transfer was obtained with purified TnpA independent of the TnpB protein, producing simple insertions (13). Here we document the efficiency and randomness of the insertion reaction, describe a variety of new Tn*552* derivatives and demonstrate that use of Tn*552* for a cosmid DNA sequencing project can eliminate all downstream subcloning steps,

*To whom correspondence should be addressed. Tel: +1 203 432 8991; Fax: +1 203 432 3104; Email: nigel.grindley@yale.edu

resulting in significant reductions in effort and cost, whilst simultaneously providing a library of precisely mapped insertion mutants.

MATERIALS AND METHODS

Reagents

Restriction enzymes and pUC19, pBR322 and LITMUS29 plasmid DNAs were from New England Biolabs (Beverly, MA). His-tagged Tn*552* transposase was isolated as described by Leschziner *et al.* (13). pJDrec27 was isolated from a cosmid library provided by W. Jacobs.

DNA techniques

Standard procedures were used for DNA manipulations (18). pTG461 contains only the Tn*552kan* transposon (13) and the replication origin from pUC19 (bp 749–1515). The linear transposon substrate was prepared by cleaving pTG461 with *Spe*I; in some cases the fragment was purified, but this was not necessary. Cosmid DNAs were prepared by alkaline lysis of 1.5 ml cultures (18) and were resuspended in 50 µl TE (10 mM Tris, pH 8.0, 1 mM EDTA); for DNA sequencing templates, the protocol included a phenol/chloroform extraction.

Transposition reactions

Twenty microliter reactions contained 20 mM HEPES, pH 7.0, 50% (w/v) glycerol, 125 mM NaCl, 10 mM $MgSO_4$, 100 ng of target DNA (or 1 µl of cosmid DNA), 0.1 pmol of the 1.1 kb Tn*552kan* and 30 ng of TnpA, unless otherwise indicated. The reactions were incubated at 37°C for 30–60 min. HB101 (Promega, Madison, WI) or LE392 competent cells were transformed directly with a portion of this mixture. Alternatively, the reactions were ethanol precipitated, resuspended in ddH_2O and electroporated into DH10B (Gibco BRL, Gaithersburg, MD) or MC1061 (Bio-Rad, Hercules, CA). The frequency of Tn*552kan* insertions was determined by plating cells on ampicillin plates in the presence or absence of kanamycin (all target plasmids encoded ampicillin resistance).

Analysis of Tn*552kan* insertion sites

Cosmids that contained insertions of Tn*552kan* were digested with *Pst*I or *Sph*I and separated on an agarose gel. DNA fragments that received an insertion were identified by the replacement of one fragment by another 1.1 kb larger. A population of insertions into LITMUS29 was analyzed using a PCR-based technique. DNA from a standard reaction was digested with *Ava*I, phenol extracted, ethanol precipitated and resuspended in TE. Insertion products were amplified (20 cycles) in four separate reactions, using a primer that annealed to both transposon ends (AL37, AGGTTAAATCATTCTCATATATCAAG) and one of four LITMUS-specific primers that annealed to each side of the two *Ava*I restriction sites (TNP029, GACCGTACGTGCGCGCG; TNP030, GGAATTCCTGCAGGATATCGTGG; AL35, GATAGGGTTGAGTGTTGTTCC; AL36, GCTATTCTTTTGATTTATAAGGG). ^{32}P-labeled transposon primer (AL37) was added for a final cycle of amplification, to allow detection of the amplified insertion products. These primers were synthesized by the Keck Foundation Biotechnology Resource Laboratory, Yale University. Denatured products were analyzed by electrophoresis on an 8% denaturing acrylamide gel and autoradiography.

DNA sequencing of cosmid clones

One tenth of a cosmid DNA miniprep was used as a template in each cycle sequencing reaction using the dye-deoxy terminator sequencing kit (Applied Biosystems) and the products were analyzed on an automated ABI Prism 377 DNA sequencer. Sequences were determined in both directions from the transposon using left and right end-specific primers (oKD20, GATATATTTTTATCTTGTGC; oKD149, GTTAATTGGTTGTAACACTGGC; synthesized by the Wadsworth Center Molecular Genetics Core Facility), yielding more than 1.1 kb of contiguous sequence for each insert. DNA sequence was compiled and edited using Sequencher (Gene Codes).

RESULTS

The transposition reaction and the transposon delivery system

The *in vitro* reaction for Tn*552* insertion requires that the transposon substrate is in its excised form with the -CA_{OH} 3′-ends exposed. This is readily achieved by digestion of the transposon-containing plasmid with *Spe*I. To circumvent the need to separate the linear transposon from the vector fragment (and from residual undigested plasmid) we have cloned Tn*552kan* (and other Tn*552* derivatives; see below) on a minimal 0.8 kb *ori*-containing fragment from pUC19. Selection of insertion events into a target of choice (and avoidance of background contamination by surviving Kan^R donor plasmids) is achieved by using both kanamycin and a second target-selecting antibiotic (ampicillin for most cosmid and plasmid targets).

To optimize the number of recoverable colonies with single transposon insertions, we have adopted the following standard reaction conditions: 20 µl reactions contain 0.1 pmol of transposon substrate, 30 ng (0.5 pmol) transposase and ~100 ng target DNA in 20 mM HEPES, pH 7.0, 10 mM $MgCl_2$, 125 mM NaCl and 50% glycerol. Cosmid or plasmid DNA prepared by standard alkaline lysis ‘miniprep’ procedures is perfectly satisfactory for the target. The molarity of transposase in the standard reaction is approximately equal to the molarity of the transposase binding sites (since the two transposon ends each contain two binding sites). As shown in Figure 1A, under the conditions used the ratio of transposase protein to DNA binding sites is relatively critical, suggesting that efficiency is maximized by stoichiometric amounts of the protein. A 2-fold variation in the amount of transposase had only a modest effect, but a 5-fold excess of transposase was inhibitory. A time course of the reaction (under the same conditions) showed that the reaction was essentially complete after 30 min (Fig. 1B). The results of both these experiments, obtained using a transformation assay to measure insertion efficiencies, have been confirmed using a ^{32}P-labeled transposon substrate and a gel-based assay (data not shown).

Large numbers of independent Tn*552kan* insertions are readily recovered

For practical applications, efficient and reproducible recovery of single transposon insertions into the target DNA is of prime

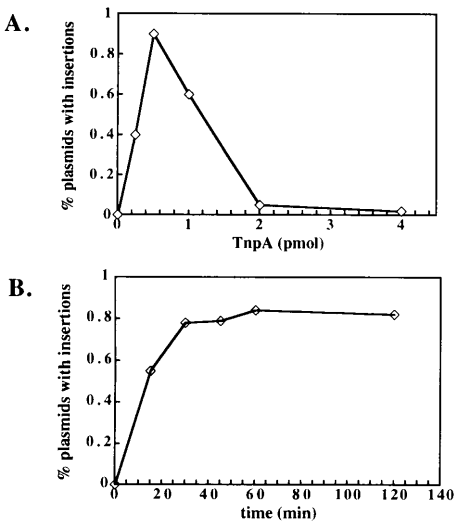

Figure 1. Transposition efficiency as a function of transposase concentration and reaction time. Reactions contained 0.1 pmol Tn*552kan* transposon, 100 ng pUC19 as target and transposase as indicated. *Escherichia coli* was transformed with a sample of each reaction and the proportion of plasmids with Km^R insertions was determined. (**A**) Effect of transposase concentration. (**B**) Effect of incubation time [all reactions contained 0.5 pmol (30 ng) transposase].

importance. In multiple experiments, we found that both small plasmids (pUC19, 2.7 kb or pBR322, 4.3 kb) and six different large (40–50 kb) cosmids consistently served as efficient targets for Tn*552* insertion. DNA from standard reactions was introduced into competent bacterial cells by electroporation or transformation and the proportion of kanamycin-resistant colonies, derived from Tn*552kan* insertions, was determined. Typically, we found that 0.5–3% of the small target plasmids and 2–10% of the cosmids received an insertion. The higher conversion of the latter was expected, since the molarity of cosmid targets was much lower than that of the small plasmids. We consider insertion frequencies of 1–3% to be optimal, resulting in high yields of individual insertions, while keeping double insertions at an acceptable level. Depending on the source and efficiency of electrocompetent cells, single reactions with cosmid targets yielded from 2×10^3 to 6×10^4 kanamycin-resistant colonies, while those with small plasmid targets gave from 5×10^4 to 4×10^5 per reaction.

In vitro insertion of Tn*552* is close to random

Many transposons exhibit distinct target site preferences (19), but for most practical applications, random insertion and, most importantly, an absence of insertion hot spots is highly desirable. In our previous description of the Tn*552 in vitro* reaction, analysis of a relatively small number of insertions into pUC19 indicated no evidence for hot spots or even strong regional preferences in this small target (13).

We have substantially extended the analysis of global insertion site preferences, examining more than 130 kb of target DNA derived from three cosmids. Individual insertions were mapped at low resolution by digestion of the cosmids with a restriction enzyme that did not cleave the transposon and comparing the fragments obtained to those from the corresponding target. If the transposon inserts without strong sequence or regional bias, the number of insertions into a specific fragment should be proportional to the relative size of the fragment; this is the observed result with Tn*552* (Table 1). The numbers of insertions range from ~40 to 200% of the expected numbers and χ^2 analysis showed that the overall distributions did not deviate significantly from expectation. Of the 228 total inserts analyzed, only one appeared not to have resulted from a single concerted insertion. This anomalous product contained one restriction fragment ~2 kb larger than the target fragment and probably contained two copies of the transposon.

Table 1. Distribution of insertions into large target DNAs

	c88-1[a] (*n* = 69)		pJDrec27 (*n* = 90)		cY20H20 (*n* = 69)	
Fragment	Expected[b]	Observed	Expected	Observed	Expected	Observed
1	16	17	25	27	27	28
2	14	11	20	26	15	19
3	12	14	16	12	13	7
4	10	4	13	17	9	12
5	7	6	6	2	8	12
6	7	14	4	3	5	7
7	6	11	3	4	5	6
8	6	3	3	4	4	3
9	4	0	3	1	3	1
10	8[c]	7	2	0	2	3
11	3	0	1	1	2	1
12	3	4	1	0	2	1
13	3	3	1	2		
14	2	1				
15	2	0				
16	1	3				

[a]Target cosmids contained *M.smegmatis* DNA. *n* is the number of insertions analyzed.
[b]Percentage of insertions observed or expected from the target fragment size.
[c]Two fragments of equal size.

High resolution mapping of a very large number of insertions into a small target plasmid also indicated a lack of any strongly preferred sites or silent regions. A pool, containing $>10^7$ independent insertions in the 2.7 kb LITMUS29 plasmid, generated under standard conditions, was mapped by a PCR-based assay analogous to Mu printing (20; see Materials and Methods). Using a single transposon-specific primer (5′-^{32}P-labeled for the final polymerase cycle) in combination with four different LITMUS29-specific primers, insertions throughout the entire target plasmid were detected and mapped. Figure 2 shows an autoradiogram of a denaturing polyacrylamide (DNA sequencing) gel in which lanes 1–4 contain samples of each of the four amplification reactions. Each band corresponds to insertion at a specific site and the band intensity indicates the frequency with which any specific site is used. While the banding patterns are not featureless

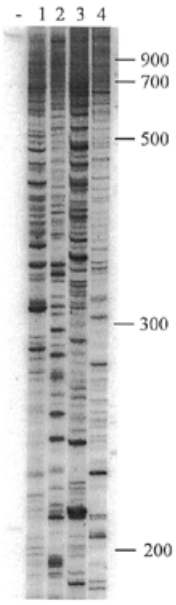

Figure 2. Distribution of Tn*552kan* insertions in LITMUS29 plasmid DNA. Insertions were mapped by PCR (see Materials and Methods), using a ^{32}P-labeled primer (AL37, which anneals to both transposon ends) and one of four LITMUS-specific primers: TNP029, lane 1 (insertions from nt 2750 to 830); TNP030, lane 2 (insertions from nt 2350 to 830); AL35, lane 3 (insertions from nt 1420 to 2300); AL36, lane 4 (insertions from nt 1220 to 120). Labeled products were fractionated by denaturing gel electrophoresis. The autoradiogram displays insertions in the regions of LITMUS29 between coordinates: 2750 (2820/0) and 630 (lane 1, between 200 and 900 nt markers); 2350 and 1650 (lane 2); 1420 and 2120 (lane 3); 1020 and 320 (lane 4). Insertions in the remaining regions were analyzed on separate gels (not shown). –, PCR as in lane 1 but using as the template DNA a reaction that lacked transposase. Numbers on the right indicate fragment sizes in nucleotides.

(they resemble a cleavage pattern created by a non-specific nuclease such as DNase I), it is clear that the insertions occurred with relatively uniform frequency throughout the target, although modest preferences imposed by details of the target sequence can be discerned. Most importantly, there are no dominating hot spots and the few 'silent' gaps are limited to no more than 20 bp in length. Each 100 bp segment has 20 or more distinct and easily detectable insertion sites, indicating that the transposition reaction samples virtually all target phosphodiester bonds.

The *Mycobacterium smegmatis recBCD* locus: an exercise in Tn*552*-facilitated DNA sequencing and mutagenesis

A few transposons have been used as DNA sequencing tools, since scattered insertions provide 'mobile' primer sites, obviating the need for subcloning into phage or plasmid vectors (1,21). We have used a cosmid containing the *M.smegmatis recBCD* locus as a test of the utility of Tn*552 in vitro* transposition as a tool for simultaneously sequencing and mutating an extended DNA sequence.

Southern hybridization, using *Mycobacterium tuberculosis recBCD* sequences as probes, indicated that the 45 kb cosmid, pJDrec27, contained the *M.smegmatis recBCD* locus on four *Bam*HI restriction fragments totaling 12.2 kb. Using ~100 ng of the cosmid target, we performed a single *in vitro* transposition reaction. One tenth of this reaction, electroporated into DH10B, yielded 6×10^3 Kan^R colonies. Restriction analysis of cosmid DNA from 204 colonies indicated that 47 contained Tn*552kan* insertions into the relevant target fragments. This number of insertions is consistent with the relevant size of the region compared with that of the entire cosmid and again demonstrates the random nature of transposon insertion. All 47 of these cosmid DNA preparations were used directly as templates for automated DNA sequencing with two oligonucleotide primers specific for each of the transposon ends (see Materials and Methods). The sequence data generated a single 13.5 kb contig spanning the entire region. More than 90% of the sequence was determined on both strands, each base pair was sequenced on average 3.9 times and the largest interval between adjacent insertions was 1.3 kb (the mean interval was 262 bp). To determine whether this gap was a sampling artifact or a true cold spot, we examined a further 24 random insertions in the 13.5 kb region. In this second sample, three insertions occurred in the 1.3 kb gap (exactly as expected), indicating that it did not represent an insertionally silent target. Figure 3 shows the complete set of 71 insertions and illustrates not only the generally random distribution of insertions throughout the segment, but also the redundancy generated by random sequencing (final redundancy of 6.1), which greatly facilitated the sequence determination.

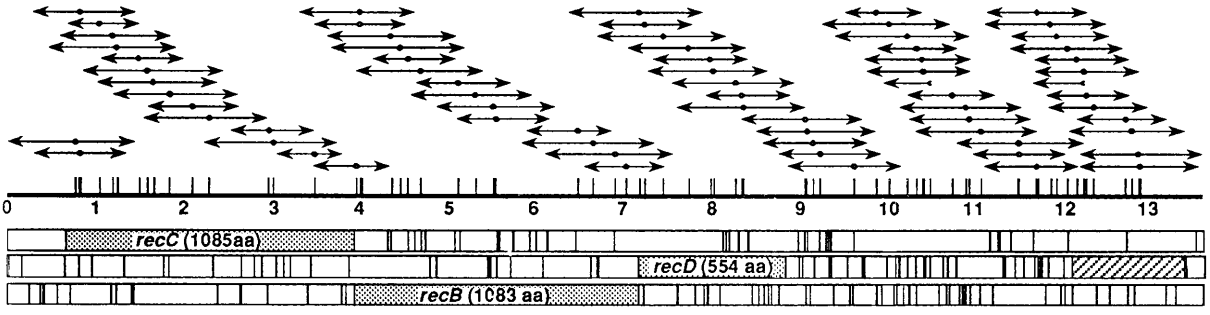

Figure 3. The 13.5 kb contig containing the *M.smegmatis recBCD* operon. The 71 Tn*552* insertions are shown as vertical lines above the 13 kb scale. The four targeted *Bam*HI fragments cover the 12 165 bp region from 743 to 12 908; only insertions into this region were subjected to sequencing. The double-headed arrows above the scale indicate the lengths of the sequence reads from each insertion site; the average read length was 590 bp from the transposon end, giving 1.18 kb of target sequence per insertion. The overall G+C content of the mycobacterial DNA was 68.5%. Shown below the scale is an analysis of coding potential for transcripts running from left to right; vertical lines indicate stop codons. The *recBCD* genes are stippled, the gene for the putative transporter protein is cross-hatched.

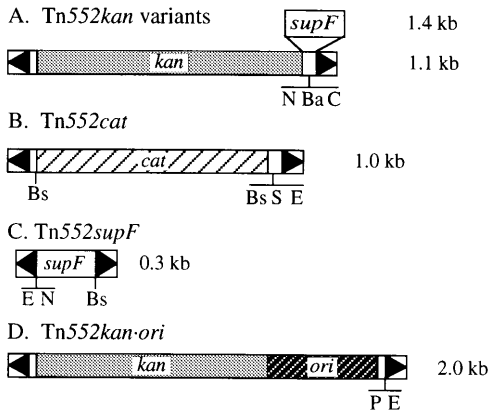

Figure 4. Tn*552*-derived transposons (see text for more information). Relevant features of each transposon are indicated; the IR sequences are shown by filled triangles and the extent of each gene is shown by a differentially shaded rectangle. Some restriction sites useful for cloning or analyses are indicated: Ba, *Bam*HI; Bs, *Bsr*GI; E, *Eco*RI; C, I-*Ceu*I (an intron-encoded endonuclease); N, *Not*I; P, *Pac*I.

Blast analysis (22) confirmed the identity of the three large open reading frames in the contig as homologs of *recBCD* (Fig. 3). Each predicted protein shares extensive identity with the RecBCD products of *M.tuberculosis* and *E.coli*; RecC shares 62 and 28% identity, RecB, 60 and 28% and RecD, 55 and 35%, respectively. In addition, the genes are clearly arranged in an operon with overlapping stop and start codons. The gene organization is identical to that in *M.tuberculosis* (23) but differs from that in *E.coli* (which contains the *ptrA* gene located between *recC* and *recB*) (24–27). Two other open reading frames with significant homologies to the protein database were identified. One, extending from nucleotide 12 090 to 13 378 in the second open reading frame (Fig. 3), was similar to a variety of *E.coli* and *Bacillus subtilis* transporter proteins (30% identity). A second open reading frame, nucleotides 11 025 to 9898, transcribed in the opposite direction to *recCBD* (not shown in Fig. 3), was homologous to the *M.tuberculosis* lipoprotein *lpp* (42% identity over its entire length). This collection of sequenced insertions provides a large selection of mutants that can be used directly to selectively knock-out gene functions in the *M.smegmatis* genome. Since each insertion is precisely mapped, its likely effect on *recBCD* gene expression can be predicted.

Second generation transposons

The Tn*552* substrate is readily manipulated and we have made several derivatives of Tn*552* that have a variety of applications (Fig. 4; the complete sequences are available on request). These include: (i) variants of Tn*552kan* with alternative restriction sites for mapping insertions or for introducing additional DNA segments (two of these also include the *supF* amber suppressor gene which can be used as an alternative selective marker in many strains with known amber mutations); (ii) Tn*552cat* which confers resistance to chloramphenicol and thus provides an alternative selectable marker to kanamycin; (iii) Tn*552supF*; and (iv) a derivative, Tn*552kan·ori*, with the replication origin from pUC19. This transposon can be used to recover clones containing the transposon plus DNA sequences flanking the insertion site (see Discussion).

We have also constructed a transposon, Tn*552′phoA·kan*, that is specifically designed for identifying genes encoding secreted or membrane proteins (M.Braunstein, T.Griffin, J.Kriakov, S.Friedman, N.Grindley and W.Jacobs, manuscript in preparation). This contains a Tn*552* end that has been modified to include an open reading frame, plus an in-frame *phoA* reporter gene that lacks both an initiation codon and a promoter. The ability to engineer a reporter gene of choice into this system enables the investigator to screen/select for insertions into a defined subset of genes. Transposons with alternative reporter genes such as *lac*, *lux* or GFP could readily be constructed.

DISCUSSION

Several properties of Tn*552 in vitro* transposition indicate that it is an excellent system for application as a tool for molecular genetics and DNA sequencing. The *in vitro* reaction is simple and efficient. In addition to the target DNA, it requires just two macromolecular components: a linear DNA substrate consisting of essentially any DNA fragment flanked by the 48 bp Tn*552* terminal inverted repeats and a single, easily prepared protein (10 g of cells yield enough transposase for more than 10^5 reactions; 13). Although the transposon substrate must be prelinearized, this is readily achieved with a simple restriction digestion and the resulting substrate and vector fragments do not need to be separated before use. This digestion has the additional advantage of inactivating the transposon donor plasmid replicon, eliminating the potential for 'false positive' transformants that otherwise result from cells receiving both donor and target plasmids together. Alternatively, the linear substrate could be generated by PCR. The standard 20 µl reaction involves a single 30–60 min incubation of all the components together and has no requirement for a preincubation step. One reaction with just 100 ng of target DNA produces single insertions into 1–3% of target molecules (higher target conversion can be achieved but with a concomitant increase in the frequency of double insertions). Transformation or electroporation of a suitable host strain with the reacted DNA readily yields large numbers of colonies with independent insertions.

Essentially any DNA can be used as the target, ranging from supercoiled plasmids to random linear DNA fragments. Nonreplicating linear fragments are appropriate targets if they are derived from naturally competent organisms such as *B.subtilis*, *Neisseria gonorrhoeae* or *Hemophilus influenzae* that can efficiently recombine linear DNAs into their chromosome (see for example 4). In such cases, *in vitro* transposition can be used as a direct method for generating a gene knock-out library, since the linear insertion products (following repair of the short target gaps at each end of the transposon) can be efficiently assimilated by the combined processes of transformation and homologous recombination. Such a strategy may be particularly useful for those organisms in which good molecular genetics tools have yet to be developed.

Perhaps the most important requirement for general practical utility of a transposition reaction is that insertions occur essentially at random. For applications such as DNA sequencing, it

is particularly important that there should be no hot spots and no transpositionally silent regions greater than 300–500 bp. Ideally, insertions into any segment of a target should occur at a frequency proportional only to the relative size of the segment. Tn*552* appears to fulfill these criteria. Exploring more than 160 kb of target DNA, we have found that the observed numbers of insertions into any particular target segment do not differ significantly from those expected on a random basis. Moreover, we have found no evidence either for hot spots or regions of exclusion. Indeed, from our total sequence database of more than 380 insertions in 100 kb of target DNA (K.M.Derbyshire, unpublished data) we have seen only one example of two independent insertions into the same target site. Additionally, in the *recBCD* cosmid sequencing project (Fig. 3), analysis of random insertions at an average density of 1 per 171 bp had no gaps larger than 0.9 kb, well within random expectations. Examination of more than 380 target site sequences indicates that the target duplications, which are most commonly 6 or 8 bp (but vary from 4 to 9 bp), range from 100% G+C to 100% A+T, consistent with a lack of any strong target sequence preferences.

Tn*552* is an adaptable and easily manipulated transposon. Since the only DNA sequences required *in cis* for Tn*552* activity *in vitro* are the 48 bp ends (13,17), essentially any desired DNA sequence may be inserted between two ends to create novel transposons. The ends of Tn*552* are small enough that almost any derivative can be constructed in a simple manner by PCR, using oligonucleotide primers that include the entire end sequences to amplify the desired gene, and instantly generating a new transposon. We have exploited this to make variants of Tn*552* with alternative selectable markers, unique sites for rare-cutting restriction enzymes (see Fig. 4) and a *phoA* reporter gene. One derivative, Tn*552kan·ori*, which contains the pUC19 origin of plasmid replication, has several potential uses. This transposon lacks cleavage sites for more than 60 commercially available restriction enzymes (for example *Avr*II, *Nhe*I, *Spe*I and *Xba*I, all of which produce a 5′-CTAG sticky end); these can be used, singly or in compatible combinations, to excise the entire transposon plus flanking DNA from any integrated source. The most obvious application for this transposon is to subclone the target DNA in the vicinity of a particular insertion in a large vector such as a cosmid or BAC. However, it could also be used to rescue any DNAs (such as viruses, plasmids or even chromosomal fragments from any source) that cannot normally replicate in *E.coli*. The advantage of this approach is that it requires no prior knowledge of the target DNA.

A relatively small number of *in vitro* transposition reactions have been used as tools for biotechnological applications; these include the bacterial transposons Mu (8,10,28), Tn*7* (4) and Tn*5* (7), Ty1 from yeast (5,6,9,11), MoMLV retroviral integrase (12) and the mariner element *Himar1* (3). Of these, at least Tn*5* and *Himar1* display significant target preferences. Tn*5* appears to favor specific target sequences and significant hot spots exist; for example, almost one sixth of 384 insertions into the 660 bp *cat* gene were in a single site (29). *Himar1* not only targets the 5′-TA-3′ dinucleotide, but also shows additional sequence and regional biases; for example, in an analysis of 65 insertions in a 3 kb target, 40% were in just seven distinct sites and only 6% occurred in the first 1 kb while 51% occurred in the third 1 kb (30). Of the other *in vitro* systems, Mu insertion appears to be both efficient and reasonably random (8,10), while Ty1 insertion, although essentially random, has a rather low efficiency and requires increased amounts of transposon and target DNAs (6). Tn*7* provides an efficient transposition reaction and has the additional advantage that targets with insertions are immune to secondary insertions (31,32). However, Tn*7* has long, complex ends that complicate the construction of novel transposon derivatives; moreover, its transposition requires three protein components and a two-step reaction with a preincubation to facilitate assembly of the transposition machinery. Normally, Tn*7* shows a very high degree of target specificity, however, the version used for mutagenesis or sequencing exploits a mutant of the TnsC protein (33,34) that is reported to mediate transposition with relatively little target specificity (4).

Our experience using Tn*552* for cosmid sequencing indicates that substantial savings of effort and expense can be achieved. First, an entire cosmid (or portion of a cosmid) can be sequenced from a collection of random independent insertions without any further subcloning. Moreover, if mapping of insertions is considered desirable before sequencing (to reduce redundancy and eliminate insertions into vector sequences), the same DNA preparation may be used for both procedures. Second, sequences are obtainable from both ends of an insertion, halving the number of template preparations and generating ‘instant contigs’ of 1 kb or more. Protection against clerical (or other) errors in assembling pairs of sequence files is provided by the short target sequence duplications associated with Tn*552* transposition. These have sufficient complexity to provide a signature of each insertion, ensuring that two sequence files presumed to arise from a single (unrearranged) insertion indeed do so. As has been pointed out before (5), such long contiguous reads (and the ability to map them) are important in dealing with DNA segments containing repetitive DNA sequences. Third, we have found that the use of a single pair of Tn*552*-specific primers not only maximizes uniformity and reproducibility in the sequencing reactions, but also eliminates the uncertainties of primer design and performance that are inherent to a ‘primer walking’ strategy and can be particularly problematical with DNA templates of high G+C content (such as the mycobacterial DNA used here).

One final, yet invaluable, by-product of a transposon-based sequencing strategy, particularly at the level of cosmids (or larger DNA clones such as BACs and YACs), is that one is left with a library of precisely mapped insertion mutations that can be used for further studies of gene action. Indeed, the potential of generating such a defined mutant library provides a strong case for using any transposon system for large scale sequencing projects.

In vitro transposition systems provide a cornucopia of opportunities for the application of molecular biology to genetically refractory systems. The ability to carry out these efficient reactions *in vitro* without the need for replication-proficient systems enables the investigator to randomly mutagenize almost any DNA. The flexibility of these transposon tools lies in the minimum requirements of the transposon itself; only short end sequences are necessary to convert any DNA into a transposon. In this paper, we have demonstrated that the Tn*552* *in vitro* transposition system performs remarkably well as a tool for sequencing and mutagenesis. We believe that it provides a valuable addition to the available systems.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Wadsworth Center Molecular Genetics Core Facility for DNA sequencing and sequencing primers and the Keck Foundation Biotechnology Resource Laboratory at Yale for PCR primers. We thank Dr C. Lawrence of the Wadsworth Center for help in statistical analyses. This research was supported by grants from the Potts Foundation and the National Institutes of Health (AI42308) to K.M.D. and NIH grant GM28470 to N.D.F.G.

REFERENCES

1. Berg,C.E. and Berg,D.E. (1996) In Neidhardt,F.C., Curtis,R., Ingraham,J.L., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd Edn. American Society for Microbiology, Washington, DC, pp. 2588–2612.
2. Kleckner,N., Bender,J. and Gottesman,S. (1991) *Methods Enzymol.*, **204**, 139–180.
3. Akerley,B.J., Rubin,E.J., Camilli,A., Lampe,D.J., Robertson,H.M. and Mekalanos,J.J. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 8927–8932.
4. Gwinn,M.L., Stellwagen,A.E., Craig,N.L., Tomb,J.F. and Smith,H.O. (1997) *J. Bacteriol.*, **179**, 7315–7320.
5. Devine,S.E., Chissoe,S.L., Eby,Y., Wilson,R.K. and Boeke,J.D. (1997) *Genome Res.*, **7**, 551–563.
6. Devine,S.E. and Boeke,J.D. (1994) *Nucleic Acids Res.*, **22**, 3765–3772.
7. York,D., Welch,K., Goryshin,I.Y. and Reznikoff,W.S. (1998) *Nucleic Acids Res.*, **26**, 1927–1933.
8. Haapa,S., Suomalainen,S., Eerikainen,S., Airaksinen,M., Paulin,L. and Savilahti,H. (1999) *Genome Res.*, **9**, 308–315.
9. Garraway,L.A., Tosi,L.R., Wang,Y., Moore,J.B., Dobson,D.E. and Beverley,S.M. (1997) *Gene*, **198**, 27–35.
10. Wei,S.Q., Mizuuchi,K. and Craigie,R. (1997) *EMBO J.*, **16**, 7511–7520.
11. Williams,S., Hayes,L., Elsenboss,L., Williams,A., Andre,C., Abramson,R., Thompson,J.F. and Milos,P.M. (1997) *Gene*, **197**, 101–107.
12. Singh,I.R., Crowley,R.A. and Brown,P.O. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 1304–1309.
13. Leschziner,A.E., Griffin IV,T.J. and Grindley,N.D.F. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 7345–7350.
14. Rowland,S.J. and Dyke,K.G. (1990) *Mol. Microbiol.*, **4**, 961–975.
15. Rowland,S.J. and Dyke,K.G. (1989) *EMBO J.*, **8**, 2761–2773.
16. Mizuuchi,K. (1992) *Annu. Rev. Biochem.*, **61**, 1011–1051.
17. Rowland,S.J., Sherratt,D.J., Stark,W.M. and Boocock,M.R. (1995) *EMBO J.*, **14**, 196–205.
18. Sambrook,J.E., Fritsch,E.F. and Maniatas,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
19. Craig,N.L. (1997) *Annu. Rev. Biochem.*, **66**, 437–474.
20. Wang,X. and Higgins,N.P. (1994) *Mol. Microbiol.*, **12**, 665–677.
21. Kimmel,B., Palazzolo,M.L., Martin,C., Boeke,J.D. and Devine,S.E. (1997) In Birren,B., Green,E., Klapholz,S., Myers,R. and Roskams,J. (eds), *Genome Analysis: A Laboratory Manual*, Vol. 1, *Analyzing DNA*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 455–532.
22. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) *Nucleic Acids Res.*, **25**, 3389–3402.
23. Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry,C.E. *et al*. (1998) *Nature*, **393**, 537–544.
24. Finch,P.W., Storey,A., Brown,K., Hickson,I.D. and Emmerson,P.T. (1986) *Nucleic Acids Res.*, **14**, 8583–8594.
25. Finch,P.W., Storey,A., Chapman,K.E., Brown,K., Hickson,I.D. and Emmerson,P.T. (1986) *Nucleic Acids Res.*, **14**, 8573–8582.
26. Finch,P.W., Wilson,R.E., Brown,K., Hickson,I.D. and Emmerson,P.T. (1986) *Nucleic Acids Res.*, **14**, 7695–7703.
27. Finch,P.W., Wilson,R.E., Brown,K., Hickson,I.D., Tomkinson,A.E. and Emmerson,P.T. (1986) *Nucleic Acids Res.*, **14**, 4437–4451.
28. Haapa,S., Taira,S., Heikkinen,E. and Savilahti,H. (1999) *Nucleic Acids Res.*, **27**, 2777–2784.
29. Goryshin,I.Y., Miller,J.A., Kil,Y.V., Lanzov,V.A. and Reznikoff,W.S. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 10716–10721.
30. Lampe,D.J., Grant,T.E. and Robertson,H.M. (1998) *Genetics*, **149**, 179–187.
31. Arciszewska,L.K., Drake,D. and Craig,N.L. (1989) *J. Mol. Biol.*, **207**, 35–52.
32. Craig,N.L. (1996) *Curr. Topics Microbiol. Immunol.*, **204**, 27–48.
33. Stellwagen,A.E. and Craig,N.L. (1997) *EMBO J.*, **16**, 6823–6834.
34. Stellwagen,A.E. and Craig,N.L. (1997) *Genetics*, **145**, 573–585.