

Production of combinatorial libraries of fused genes by sequential transposition reactions

Todd A. Naumann, Igor Y. Goryshin and William S. Reznikoff*

Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA

Received June 13, 2002; Revised and Accepted September 9, 2002

ABSTRACT

The use of *in vivo* and *in vitro* transposition reactions to perform non-combinatorial manipulation of DNAs in molecular biology is widespread. In this work we describe a technique that utilizes two sequential, directed transposition reactions in order to carry out combinatorial DNA manipulations. The methodology relies on the use of two different mutant Tn5 transposase proteins that have different transposon end recognition specificities. We demonstrate that the technique can be used to create large libraries of random fusions between two genes. These transpositional fusions are defined by insertion of a 32 bp linker sequence. We applied the technique to a model system, chloramphenicol acetyl transferase, to create functional fusions from N- and C-terminally truncated, non-functional genes. Comparative structural analysis suggests that both sides of the linker are inserted into disordered regions in functional proteins.

INTRODUCTION

Transposons have long been utilized as tools for genetic studies. More recently, the advent of high efficiency *in vitro* transposition systems has led to the development of transposon tools that are useful for many molecular biology and genetic applications (1–5). Transposon Tn5 is a bacterial composite transposon that encodes a protein, transposase (Tnp), that is responsible for its movement (6). The transposition reaction requires only Tnp and specialized 19 bp inverted repeats that act as specific end-binding sequences for Tnp. Wild-type Tnp has evolved to be inefficient, and purified wild-type Tnp has no detectable *in vitro* activity. However, two highly efficient *in vitro* transposition systems have been developed for Tn5 (7,8). Each transposition system relies on a different combination of mutated Tnp and specific end-binding sequence.

A double mutant of Tnp, Tnp EK/LP, is able to promote transposition of DNAs that are bordered by a naturally occurring 19 bp end sequence termed outside end (OE). Tnp EK/LP has even higher activity with DNAs that are bordered

by two mutated end sequences termed mosaic end (ME). Hereafter, Tnp EK/LP will be referred to as Tnp-ME.

A second mutant Tnp, Tnp sC7v2.0, contains a combination of four mutations that allow this protein to efficiently transpose DNAs bordered by a naturally occurring 19 bp binding site termed inside end (IE). IE sequences contain two modification signals for *dam* methylase that result in the addition of four methyl groups into the major groove. This methylation inhibits binding of wild-type Tnp (and Tnp-ME), but increases activity of Tnp sC7v2.0 (hereafter referred to as Tnp-IE) (8,9).

Recently, non-transposition-based methods for generating libraries of randomly fused genes have been reported. Ostermeir *et al.* (10) and Lutz *et al.* (11) first described a method termed ITCHY that involves the incremental truncation of two genes by use of *ExoIII* nuclease followed by S1 nuclease treatment, polymerization, and ligation of fragments to form random fusions. A second method has also been reported that utilizes random cleavage of DNA followed by a series of digestion and ligation reactions to create random fusions (12).

In this work, we describe a methodology that allows for the creation of random gene fusions via the use of two simple transposition reactions. By taking advantage of the end specificity difference of Tnp-ME and Tnp-IE in separate, sequential transposition reactions, we have created two similar methods that are able to create large libraries (>10⁷ independent members) of random fusions between any two DNA sequences. Both methods result in DNA fusions that are defined by the presence of a 32 bp DNA linker. In contrast to earlier reported methods that require multiple enzymatic steps separated by DNA purifications, the methods reported here rely on direct transformation of *in vitro* transposition reactions. We further report the use of this method to create functional chloramphenicol acetyl transferase (CAT) enzymes by random fusion of two non-functional, truncated genes.

MATERIALS AND METHODS

Construction of plasmid vectors

Vectors for method 1. A plasmid containing a kanamycin-resistance gene bordered by one IE and one linker end was constructed as follows. Plasmid pGT4 (8), a pUC19-based plasmid containing a kanamycin gene bordered by two IEs, was digested to completion with *EcoRI* and partially digested

*To whom correspondence should be addressed. Tel: +1 608 262 3608; Fax: +1 608 265 2603; Email: reznikoff@biochem.wisc.edu

with *Bcl*I. The linker sequence was created by annealing of two oligonucleotides (5'-GTGTCTCTTGATCAGATGTGTA-TAAGAGTCAGATCT-3' and 5'-CTGACTCTTATACACA-TCTGATCAAGAGACAG-3') followed by ligation into the digested plasmid to form pNG1. This plasmid was used to prepare pre-cleaved transposon substrate (see below) for the first transposition reaction.

Plasmid pNG2 was constructed as a plasmid containing a CAT gene with a C-terminal truncation and a single ME sequence. It was constructed by modification of pGRCAM2, a pUC19-based plasmid with a CAT gene bordered by two MEs. The plasmid was digested with *Nco*I and blunt ended by use of T4 polymerase and dNTPs. The plasmid was then digested with *Afl*III to delete the C-terminal part of the CAT gene and one ME. The vector fragment was then re-ligated to form pNG2. This plasmid was the target in the first transposition reaction.

Plasmid pNG3 was constructed to contain a CAT gene with an N-terminal truncation. Plasmid pACYC184 (New England Biolabs) was cleaved at two locations by digestion with *Pvu*II and re-ligated. This plasmid is the target in the second transposition reaction.

Vectors for method 2. Plasmid DNA containing a kanamycin-resistance gene bordered by two linker ends was constructed by modification of the IE end of pNG1. Plasmid pNG1 was digested to completion with *Bgl*III and partially digested with *Psh*AI. Two oligonucleotides (5'-CTGTCTCTTGATCAGATGTGTAAGAGTCAGATCT-3' and 5'-CTGACTCT-TATACACATCTGATCAAGAGACAG-3') were annealed and ligated into the digested plasmid to form pNG101. This plasmid was used to prepare pre-cleaved transposon substrate for the first transposition reaction. Due to low transposition efficiency in the second transposition reaction using linker A, a plasmid that contained a single base pair change on each linker end was constructed to form pNG201.

Plasmid pGN3 was constructed as a plasmid containing a conditional origin of replication (π -dependent origin) and a CAT gene with an N-terminal truncation by a three-step process. Plasmid pMF36 (13), which contains the π -dependent origin, was digested with *Sal*I and partially digested with *Dra*I to form pGN1. pGN1 was then digested with *Eco*RI, filled in with T4 polymerase, and ligated to a *Bsa*AI/*Bsa*AI fragment from pACYC 184 (New England Biolabs) to form pGN2. pGN2 was then digested with *Nco*I and *Aat*II, filled in with T4 polymerase and ligated to form pGN3. Method 2 also used pNG3, the same plasmid as described in method 1.

Preparation of Tnps and DNA substrates

Tnp-ME and Tnp-IE were purified using the IMPACT system (New England Biolabs) as described previously (14).

Pre-cleaved transposons were prepared as follows. The appropriate plasmid was purified from *Escherichia coli* K12 (DH5 α) using a Qiafilter midi kit (Qiagen). Purified plasmid was digested with *Psh*AI releasing the pre-cleaved transposon from the plasmid. The pre-cleaved transposon band was then purified by electrophoresis and gel extraction using the Qiaquick gel purification kit (Qiagen).

Plasmid substrates were purified using a Qiafilter midi kit. All plasmids were purified from strain DH5 α except pGN3,

which requires a specialized strain for replication of the plasmid (EC100Dpir-116; Epicentre).

Gene fusion protocol

In all transposition reactions, DNAs were combined with the appropriate Tnp in transposition reaction buffer (0.1 M potassium glutamate, 25 mM Tris-acetate, pH 7.5, 10 mM Mg²⁺ acetate, 50 μ g/ml bovine serum albumin, 0.5 mM β -mercaptoethanol, 2 mM spermidine, 100 μ g/ml tRNA; final concentrations) in a total volume of 20 μ l. In the first transposition reaction (both methods), pre-cleaved transposon and target plasmid were added to a final concentration of 20 nM each. Tnp-IE was added to a final concentration of 200 nM and the reactions were incubated at 37°C for 3 h. Products were then subjected to treatment with SDS (0.5%) and heat (68°C for 5 min). This removes Tnp from transposition products and results in a higher transformation efficiency (data not shown). Treated reaction products were then dialyzed against water and transformed into electro-competent *E. coli* K12 cells (DH5 α ; efficiency = 8.0×10^8 c.f.u./ μ g plasmid). Cells were grown for 1 h at 37°C in 1 ml of Luria-Bertani medium (LB; Difco) to allow expression of antibiotic-resistance genes. The culture was then used to inoculate 50 ml of LB containing appropriate antibiotics and grown at 37°C for 16 h. A small aliquot of culture following the initial 1 h growth was removed and plated onto agar media to estimate library size ($>10^5$).

In the second reaction (both methods), purified plasmid DNA from the insert library and the second target plasmid were added to a concentration of 80 nM each. Tnp-ME was then added to a concentration of 400 nM. Reactions were allowed to incubate at 37°C for 1 h and then treated with SDS and heat, dialyzed, and transformed as above. For direct selection of functional CAT fusions, cultures were plated on LB agar containing 20 μ g/ml chloramphenicol. For creation of fusion libraries via method 2, cells were outgrown in 1 ml of LB without antibiotics for 1 h and subsequently inoculated into 50 ml of LB and allowed to grow overnight in the presence of double antibiotic selection.

RESULTS

The process described in this work creates libraries of random gene fusions via two sequential transposition reactions. The system utilizes a specialized end sequence that consists of two specific end-binding sites that are facing in opposite directions (Fig. 1). In the resulting gene fusions this double end acts as a linker sequence, forming the junction between two genes. A simple linker would consist of the two required specific end-binding sequences, IE and OE (or IE and ME), in inverted orientation with no intervening DNA. In an effort to reduce the size of the amino acid linker in the resulting fusion proteins, the internal transposon ends have been overlapped to reduce the size of the DNA linker from 38 to 32 bp.

Linker A was designed by combining the IE and OE sequences. The overlap region does not match the sequence of either IE or OE. The 19 bp read from the top left are the IE sequence with position 18 changed from C to G (IE'). The 19 bp read from the bottom right match the OE sequence with positions 17 and 18 (linker positions 16 and 15) changed from AG to TC (OE').

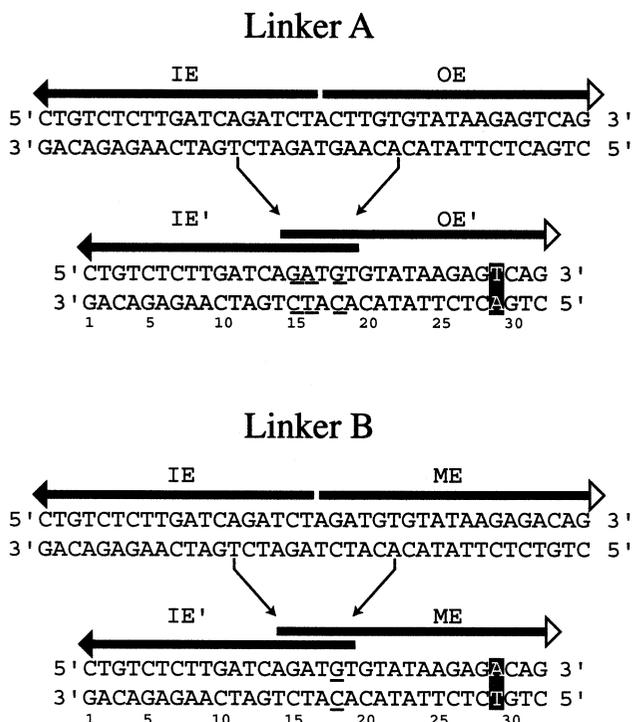


Figure 1. Design of linker sequences. A hypothetical linker would contain either IE and OE or IE and ME sequences facing in opposite directions. Linkers used in this work were shortened by overlapping the ends of the two sequences. For linker A, positions 15 and 16 match the sequence for IE, but are mutated for OE (OE'). Position 18 corresponds to the sequence of OE, but is altered for IE (IE'). In the overlap region of linker B, position 18 is still altered relative to IE, while the ME end is not changed. The two linkers differ only at position 29 (shaded).

Linker B was designed by combining the IE and ME sequences. The overlap region matches the ME sequence and contains the same IE mutation at position 18 as in linker A (IE'). This resulting linker differs from linker A at position 29 only.

Two different methodologies were developed for creating the random gene fusions (see below). In method 1, the IE end of the linker A is randomly fused to gene A in the first transposition reaction, while the OE end of linker A is randomly fused to gene B in the second transposition reaction. Both linkers contain a single transparent reading frame in the IE to OE (ME) direction. Therefore one in three in-frame fusions (or one in nine total gene fusions) can encode for fusion proteins and all functional proteins will contain the same linker epitope. In the second methodology, two fusions are created. One reads through the linker in the same IE to ME orientation while the second fusion reads in the opposite or ME to IE direction. The reverse orientation of linker A contains one transparent reading frame and linker B contains two transparent reading frames. For fusions that contain linker B in the ME to IE direction, two of three in-frame fusions (or two in nine total gene fusions) can encode for fusion proteins, and two different linker epitopes are possible.

Transpositional fusion by method 1

In method 1, the gene that contains the promoter end of the fusion is cloned into a specialized plasmid that contains a

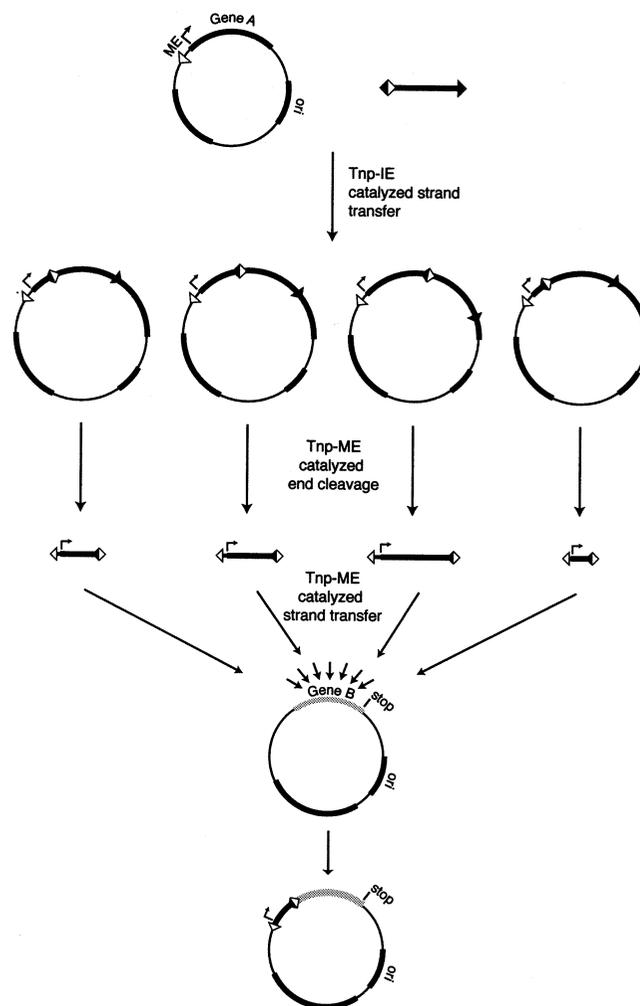


Figure 2. Method 1 for creating transitional gene fusions. A pre-cleaved transposon bordered by one IE and one linker end is randomly inserted into the target plasmid by Tnp-IE. Reaction products are transformed and selected by double antibiotic resistance to create an insert library. The insert library is then incubated with the second target plasmid and Tnp-ME. Tnp-ME cleaves transposons defined by ME and the linker end from the insert library plasmid and inserts them into the target to create the random gene fusions.

single ME sequence (Fig. 2). In the first transposition reaction, purified plasmid DNA containing a gene of interest (gene A) and a single ME sequence is mixed with a pre-cleaved transposon that consists of an antibiotic-resistance gene flanked by IE on one end and linker A on the other. The addition of Tnp-IE to these DNAs results in the insertion of the transposon into random locations in the target plasmid. Reaction products are then transformed into electrocompetent *E.coli* cells. Transformed cells are grown for 1 h in the absence of antibiotics and are then added to 50 ml of liquid media containing drug selection for both the plasmid and transposon insert and allowed to grow overnight in mixed liquid culture to produce the initial library of random linker inserts. Plasmid DNA from this culture is then purified and used as the transposon donor for the second transposition reaction.

In the second transposition reaction, the insert library is mixed with a second plasmid that contains the gene that will

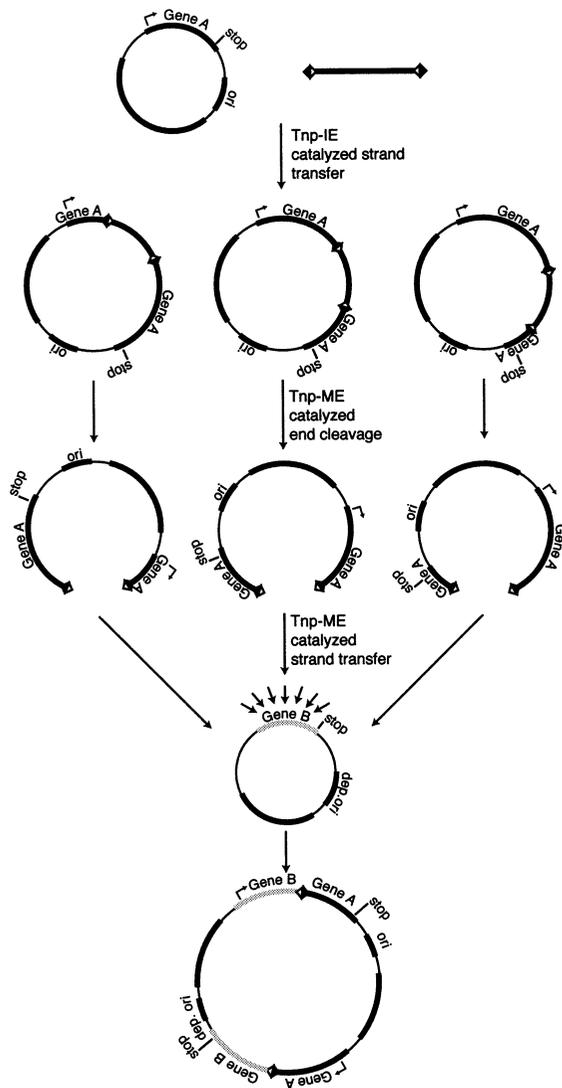


Figure 3. Method 2 for creating transpositional fusions. A pre-cleaved transposon bordered by two linker ends is randomly inserted into a plasmid containing one of the genes by Tnp-IE. Reaction products are transformed and selected by double antibiotic resistance to make the insert library. The insert library is then incubated with the second plasmid, containing the other gene, as well as Tnp-ME. Tnp-ME cleaves the plasmids at the MEs of the linkers and inserts them randomly into the target plasmid.

form the C-terminus of the gene fusion. Incubation of these DNAs with Tnp-ME results in the cleavage of transposons from the insert library and their insertion into random locations in the target plasmid to create the library of random fusions. It should be noted that in the insert library, half of all inserts are in the incorrect orientation and do not serve as substrate in the second reaction.

Transpositional fusion by method 2

Method 2 also uses three DNA molecules (a pre-cleaved transposon and two plasmids with one gene of interest each). In this method, the pre-cleaved transposon is defined by two linker sequences and neither of the two plasmids contain Tnp-binding sequences. The additional requirement of method 2 is that in this methodology one of the two plasmids contains a

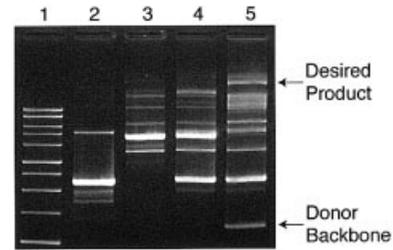


Figure 4. Gel analysis of method 2. In the second transposition reaction Tnp-ME cleaves the insert library plasmids at the ME ends of the two linkers, releasing the antibiotic-resistance gene (donor backbone). The original plasmid bordered by linker ends is then converted into strand transfer products. Insert into the second target plasmid results in the production of gene fusions (desired product) [lane 1, 1 kb DNA ladder (Promega); lane 2, second target plasmid; lane 3, insert library from first transposition reaction; lane 4, negative control, no Tnp-ME; lane 5, library from second transposition reaction with Tnp-ME].

conditional origin of replication. This is due to the fact that the final library of fused genes consists of one large fusion plasmid that is formed by joining of the two initial plasmids. If both of the initial plasmids contain non-dependent origins of replication, the resultant fusion plasmid would have two origins and may not replicate properly.

An illustration of the two transposition reactions is shown in Figure 3. In the first reaction, the pre-cleaved transposon with two linker ends is inserted randomly into a plasmid containing the first gene of interest (gene A) by Tnp-IE. As in method 1, the reaction products are transformed into *E.coli* and grown overnight in mixed liquid culture to produce an insert library containing random inserts into the plasmid. Unlike method 1, the orientation of insertion is not important, as the double ends exist on both ends of the transposon.

In the second transposition reaction, the library of inserts is combined with a target plasmid containing the second gene of interest and Tnp-ME. Tnp-ME binds to the MEs and promotes blunt-ended cleavage of both ends, releasing the plasmid from the antibiotic-resistance gene. The first plasmid, which is now a transposon, can then strand transfer into the second plasmid at a random location to create the final fusion plasmid.

A gel analysis of a typical second reaction is shown in Figure 4. Lane 2 represents the target plasmid and lane 3 is the initial insert library. These libraries typically contains $>10^5$ independent survivals as determined by plating of dilutions of the transformation mixture onto agar plates in the presence of both antibiotics. This number is sufficient to include all possible insertions sites. Lane 4 is a negative control, consisting of equal amounts of both the second target plasmid and the initial insert library, incubated without Tnp-ME. Lane 5 shows the reaction products when the insert library, target plasmid and Tnp-ME are incubated for 1 h at 37°C under appropriate buffer conditions. Tnp-ME-catalyzed double-end cleavage at the MEs results in the disappearance of the insert library and production of a donor backbone band (antibiotic-resistance gene from the initial transposon insert) that is the result of Tnp-ME-promoted cleavage at the two MEs. The released transposons defined by two MEs are then converted into strand transfer products. The desired product band is indicated by the arrow.

The resulting reaction products can be directly transformed into electrocompetent *E.coli* cells and grown in mixed liquid

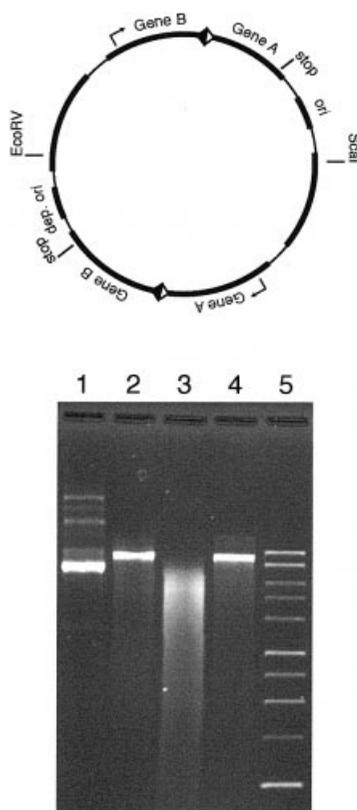


Figure 5. Gel analysis of transpositional fusion library. After overnight growth in the presence of antibiotics, plasmid DNAs were isolated. Products that result from fusion of the plasmids have unique *ScaI* and *EcoRV* sites. Therefore, digestion with either restriction endonuclease results in linearization of the plasmids. Double digest of an individual fusion plasmid results in the creation of two linear DNAs (data not shown). However, since each of the restriction sites is from a different original plasmid, the size of the two bands should vary from fusion to fusion. Hence, digestion of the library with *EcoRV* and *ScaI* results in a smear of DNA bands of varying size [lane 1, plasmid fusion library; lane 2, *ScaI* digest; lane 3, *ScaI* + *EcoRV* digest; lane 4, *EcoRV* digest; lane 5, 1 kb DNA ladder (Promega)].

culture under double-drug selection to generate a library of desirable products. Plating of aliquots from the reactions has revealed that the libraries contain $>10^7$ independent gene fusions. Reaction products can also be directly subjected to selection for functional fusions as described in method 1.

An analysis of a fusion library created by this method is shown in Figure 5. In this method, all plasmids contain the same DNA, differing only in the position at which the two initial plasmids were fused by the transposition reactions. Plasmids prepared from overnight liquid culture contain one major band, indicating that the double-drug selection is efficiently choosing the desired products from the second transposition reaction without the need for a prior purification step (lane 1). Fused plasmids in the given example should contain a unique *EcoRV* site originating from one original plasmid and a single *ScaI* site that originated from the second plasmid. Digestion of the library with either of these enzymes results in linearized plasmid (lanes 2 and 4). Double digestion of an individual plasmid results in two linear DNA bands the sum of which are equal to the total size of the fused plasmids (data not shown). However, since each of the restriction sites



Figure 6. Restoration of CAT activity from random transpositional fusion of truncated genes. CAT Δ C, encoding the first 171 amino acids of the CAT gene, was randomly fused to CAT Δ N, lacking the first 38 amino acids, by double transposition. Plasmid DNAs were recovered from colonies growing on chloramphenicol plates. DNA sequencing was performed to determine the amino acid sequence of functional proteins. Bars represent the end points of the amino acid duplication or deletion for each individual fusion sequenced. Seven of eight fusions contained amino acid duplications and the remaining fusion a short deletion of sequence. Labels on the right indicate which method and linker were used.

is from a different original plasmid, the size of the two bands should vary from fusion to fusion. Hence, digestion of the library with *EcoRV* and *ScaI* results in a smear of DNA bands of varying size (lane 3).

Restoration of chloramphenicol acetyl transferase from N- and C-terminally truncated genes

We devised a test system in order to demonstrate the ability of each of the methods to generate functional fusion proteins. We tested the ability of the double transposition fusion system to create functional fusions of CAT from two truncated CAT genes where one gene encodes a C-terminally truncated CAT enzyme and the other gene encodes for an N-terminally truncated CAT enzyme.

Method 1 was tested as shown in Figure 2. A plasmid containing a CAT gene fragment lacking the coding sequence for the 47 terminal amino acids was cloned next to a ME sequence. Pre-cleaved transposons with a single linker end (linker A) were inserted into the linker by Tnp-IE to make the first insert library. After transformation and outgrowth with appropriate drug selection, the plasmid library was purified. In the second reaction, plasmid DNA from the insert library was mixed with a target plasmid that contained a partial CAT gene that lacks both a promoter and the coding sequence for the first 38 amino acids. Tnp-ME was added to cleave N-terminal fragments from the CAT gene in the insert library and insert them randomly into the target plasmid. The reaction products were transformed and plated on agar plates containing 20 μ g/ml chloramphenicol to select for functional CAT fusions.

Transpositional fusion of the CAT gene fragments was also carried out by method 2 using a pre-cleaved transposon with two linker A ends. In the first reaction, the double linker pre-cleaved transposon was inserted into a target plasmid encoding the N-terminal truncation of the CAT gene by use of Tnp-IE to generate the insert library. The library of inserts was then mixed with the second plasmid that contained the

C-terminal truncation of the CAT gene and Tnp-ME. The reaction products were then transformed into *E.coli* and plated on agar containing chloramphenicol to select for individual functional fusions. Although successful, the use of linker A resulted in a low transposition frequency in the second reaction. Therefore, linker B was tested in the same way. Due to the order in which the two plasmids were used, the linker orientation for fused genes was in the opposite direction (ME to IE) as those created by method 1.

For each method, plasmid DNA from a few chloramphenicol-resistant colonies was purified and sequenced in order to confirm the presence of the transposon linker sequence and also to determine the amino acid sequence of the functional fusions. The results are summarized in Figure 6.

Of the eight functional CAT fusions sequenced, none are 'perfect' fusions (no loss or duplication of amino acids) and only one is formed by deletion of amino acids. The majority of functional fusions characterized are formed by duplication of a fragment of CAT. These duplications range from 5 to 53 amino acids in length.

Structural information for the CAT protein used in this study, which originates from transposon Tn9, does not exist. However, X-ray crystallographic structure information is available for the related type III Chloramphenicol Acetyltransferase (CATIII) (15). We used BLAST to align the primary amino acid sequences of CAT and CATIII. The alignment revealed that the two proteins have 46% amino acid identity with only a single gap of one amino acid. We used this alignment to map the location of the linker insertion in functional CAT fusions to the CATIII structure. This analysis predicts that, of the 16 linker-CAT junctions, 15 are in disordered regions, while the remaining junction is predicted to be inserted into a β -sheet one amino acid from the terminus. Although these results come from only one protein fusion example, the analysis strongly suggests that functional fusions are formed when the linker sequence does not insert into a location that interrupts either α -helices or β -sheets.

DISCUSSION

In this work we report the use of sequential, *in vitro* transposition reactions to generate a library of random fusions between two genes. The method is capable of producing large libraries that have been shown by double digest to have a high level of variability. We further demonstrate the ability of the system to produce functional CAT proteins from splicing of inactive, truncated genes. By utilizing metabolic auxotrophs, the demonstrated methodology can be directly adapted for creating and selecting functional hybrids of metabolic enzymes.

Of the two methods described, each has specific strengths. In method 1, the main advantage is that in each transposition reaction, a combination of one linker end and one normal transposon end are used while in method 2, both reactions depend on the use of two linker ends. Any constructed linker that is paired with a normal end has a higher probability of functioning than it would if paired with an identical linker end. This was demonstrated by our observation that the efficiency of transposition with linker A, while very high in both reactions with method 1, was associated with noticeably reduced transposition efficiency in the second reaction of

method 2. Method 2 has three main advantages that make it preferable in cases where desired linker ends function in tandem. The first advantage is no special constructs are required for the first reaction. Secondly, the presence of two linker ends on the initial transposon means that there is no dependence on insert orientation in the first transposition reaction. Finally, in the second transposition reaction of method 2 all of the transposons are equal in length. This removes any size-dependent biases that may be present in method 1.

One aspect of the described methods that was not addressed above is the 'randomness' of target site selection. All biochemical reactions such as these have biases and it is clear from previous work that target sequence biases exist for Tn5 transposition. However, the critical question for the described methodology is whether there is an apparent exclusion of any sequence from being a target. Published experiments indicate that all sequences should be targeted at some frequency (16,17). In other words, in order to generate libraries that contain all possible fusions by this or any method, the library size will have to be somewhat larger than would be predicted by an ideal method that is absent of all biases.

Proposed improvements to constructs

The main drawback of this system of making double transpositional gene fusions is the amount of undesirable DNA fusions that are contained in the library. When used with genes of reasonable size and with a direct selection for activity, as demonstrated by restoration of CAT from two truncated genes, this can be ignored due to the large library size. However, if it is necessary to use a screen to isolate functioning fusions from the library, this can be a considerable drawback.

The number of transpositional fusions that result in protein fusion is limited by both the chances of maintaining the correct reading frame (one in three) and also the correct orientation of transposition insert in the second reaction (one in two). Taking these two fundamental non-idealities into consideration, a maximum of one out of six (16.7%) of all transpositional fusions would be true protein fusions.

The percentage of transpositional gene fusions that are able to produce fusion proteins is also limited by both the presence of stop codons in the linker and non-essential DNA other than the gene of interest in the plasmid. The first non-ideality can be overcome by changing a single base pair in the sequence of linker B in the ME to IE direction. Such a change might alter the efficiency of the transposition reaction and result in a useless linker. However, there are eight different changes that can be tried in order to remove the single stop codon that exists in the new linker in the ME to IE direction and it is likely that at least one of these alterations will not be detrimental. The second non-ideality, insertion into DNA outside of the gene of interest, can be overcome by the construction of minimal plasmids. If each of the two gene-encoding plasmids only contains an antibiotic-resistance gene, DNA necessary for replication of the plasmid, and the gene of interest, then only inserts into the gene of interest would produce transposition products that are able to replicate and still encode proper drug resistance. These two changes would bring the actual 'useful' fusion yield nearer the theoretical limit of 16.7%.

Transpositional fusions that result in the production of a fused protein can also be directly selected from the library if the C-terminal gene is fused to another gene that encodes a protein that can be directly selected for. The use of this technique would not only select expressed fusion proteins for screening, but also eliminates translated fusions that result in proteins that are poorly expressed or have low solubility (18). The successful use of a CAT gene for the purpose of selection of non-homologous gene fusions has previously been reported (12).

CONCLUSIONS

The test system that we designed demonstrated the ability of using this sequential, double transposition method for creating functional gene fusions. The real usefulness of the method can only be demonstrated by fusion of two non-homologous proteins to form a functional hybrid. For example, the method could be employed to join the domains of two different proteins to create a new protein that consists of a functional domain(s) from each of the two original proteins.

ACKNOWLEDGEMENTS

We thank Carol Pfeffer for help preparing the manuscript. This research work was supported by NSF grant MCB 84089, NIH grant GM 50692 and a University of Wisconsin Industrial Relations grant 1680.

REFERENCES

- Gwinn,M.L., Stellwagen,A.E., Craig,N.L., Tomb,J.F. and Smith,H.O. (1997) *In vitro* Tn7 mutagenesis of *Haemophilus influenzae* Rd and characterization of the role of atpA in transformation. *J. Bacteriol.*, **179**, 7315–7320.
- Haapa,S., Taira,S., Heikkinen,E. and Savilahti,H. (1999) An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res.*, **27**, 2777–2784.
- Akerley,B.J., Rubin,E.J., Camilli,A., Lampe,D.J., Robertson,H.M. and Mekalanos,J.J. (1998) Systematic identification of essential genes by *in vitro* Mariner mutagenesis. *Proc. Natl Acad. Sci. USA*, **95**, 8927–8932.
- Devine,S.E. and Boeke,J.D. (1994) Efficient integration of artificial transposons into plasmid targets *in vitro*: a useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res.*, **22**, 3765–3772.
- Goryshin,I.Y., Jendrisak,J., Hoffman,L.M., Meis,R. and Reznikoff,W.S. (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nat. Biotechnol.*, **18**, 97–100.
- Reznikoff,W.S., Bhasin,A., Davies,D.R., Goryshin,I.Y., Mahnke,L.A., Naumann,T., Rayment,I., Steiniger-White,M. and Twining,S.S. (1999) Tn5: A molecular window on transposition. *Biochem. Biophys. Res. Commun.*, **266**, 729–734.
- Goryshin,I.Y. and Reznikoff,W.S. (1998) Tn5 *in vitro* transposition. *J. Biol. Chem.*, **273**, 7367–7374.
- Naumann,T.A. and Reznikoff,W.S. (2002) Tn5 transposase with an altered specificity for transposon ends. *J. Bacteriol.*, **184**, 233–240.
- Yin,J.C., Krebs,M.P. and Reznikoff,W.S. (1988) Effect of dam methylation on Tn5 transposition. *J. Mol. Biol.*, **199**, 35–45.
- Ostermeier,M., Shim,J.H. and Benkovic,S.J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.*, **17**, 1205–1209.
- Lutz,S., Ostermeier,M. and Benkovic,S.J. (2001) Rapid generation of incremental truncation libraries for protein engineering using alpha-phosphothioate nucleotides. *Nucleic Acids Res.*, **29**, e16.
- Sieber,V., Martinez,C.A. and Arnold,F.H. (2001) Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.*, **19**, 456–460.
- Filutowicz,M., McEachern,M.J. and Helinski,D.R. (1986) Positive and negative roles of an initiator protein at an origin of replication. *Proc. Natl Acad. Sci. USA*, **83**, 9645–9649.
- Bhasin,A., Goryshin,I.Y. and Reznikoff,W.S. (1999) Hairpin formation in Tn5 transposition. *J. Biol. Chem.*, **274**, 37021–37029.
- Leslie,A.G., Liddell,J.M. and Shaw,W.V. (1986) Crystallization of a type III chloramphenicol acetyl transferase. *J. Mol. Biol.*, **188**, 283–285.
- Goryshin,I.Y., Miller,J.A., Kil,Y.V., Lanzov,V.A. and Reznikoff,W.S. (1998) Tn5/IS50 target recognition. *Proc. Natl Acad. Sci. USA*, **95**, 10716–10721.
- Shevchenko,Y., Bouffard,G.G., Butterfield,Y.S.N., Blakesley,R.W., Hartley,J.L., Young,A.C., Marra,M.A., Jones,S.J.M., Touchman,J.W. and Green,E.D. (2002) Systematic sequencing of cDNA clones using the transposon Tn5. *Nucleic Acids Res.*, **30**, 2469–2477.
- Maxwell,K.L., Mittermaier,A.K., Forman-Kay,J.D. and Davidson,A.R. (1999) A simple *in vivo* assay for increased protein solubility. *Protein Sci.*, **8**, 1908–1911.