

Tn5 transposition: a molecular tool for studying protein structure–function

W.S. Reznikoff¹

Department of Biochemistry, University of Wisconsin, Madison, WI 53706, U.S.A.

Abstract

Transposon-based technologies are important genetic tools for global genome analysis and, as discussed in the present paper, in detailed studies of protein structure–function. Various different transposition systems can be used in these studies but this paper uses Tn5-related systems as a model. In particular, the following four different technologies are described in this paper: (i) using transposition to generate nested deletion families, (ii) using transposons to generate functional protein fusions to reporter functions, (iii) mapping protein secondary structures through the generation and analysis of in-frame linker insertions and (iv) using sequential transposition events to generate random gene fusions. The success of these forward genetic technologies requires that the transposition system be efficient and manifest near-random target sequence selection.

Introduction

Transposable elements have been an important tool in genome analysis ever since their discovery some 30 years ago. These genome analysis studies have been performed on the macro-level resulting from the generation of transposon knockout insertion libraries of whole genomes with varying levels of saturation. In the present paper, I examine the use of transposons on the micro-level, as a tool for detailed structure–function studies of individual protein-encoding genes. This micro-level transposon-based gene analysis depends on several principles that are also used in the larger scale genome analysis studies. That is, the procedure involves forward genetic approaches that require near random, near saturating populations of inserts that are easy to select and analyse. I will use technologies developed for derivatives of the transposon Tn5 as examples in my analysis of gene-targeted transposon procedures, keeping in mind that for many of the examples essentially identical procedures have been developed and are available for the experimentalist based on other transposable element systems such as Mu, Ty1 and Tn7 (background in [1–3] and reviewed in [4,5]).

The basic transposition process as exemplified by Tn5 transposon proceeds as follows (see [6]). The element-specific transposase protein binds to the RSs (recognition sequences) that define the element's ends (in the case of Tn5, these are 19 bp in length) and forms a synaptic complex. In the presence of Mg²⁺, the transposase in the synaptic complex catalyses cleavage of the transposon free of adjacent donor DNA. The released synaptic complex then binds target DNA in an essentially random fashion (target binding is not completely random but at differing efficiencies can probably

accommodate all sequences). Strand transfer then occurs, generating an insertion of the transposon bracketed by a 9 bp duplication of the target DNA sequence. There are several important features of this process. Although it is mechanistically a little complicated, it is experimentally very simple, involving but three readily available macromolecules: the transposase, the transposon-containing DNA and the target DNA. The transposon DNA can contain any desired sequence between its 19 bp RS ends (with some generous size limitations). In addition, the transposase catalytic activity is absolutely dependent on, and thus can be controlled by, the presence of a bivalent cation such as Mg²⁺.

Below I will briefly discuss the following four general micro-level technologies that capitalize on the above guidelines: (i) the generation of nested deletion families, (ii) the generation of reporter fusions, (iii) the mapping of protein domains by inserting defined in-frame insertions and (iv) the generation of combinatorial protein fusions. The reader will notice that although one of these technologies is specifically designed to map protein domains, in fact two of the other technologies accomplish a similar goal. All of these technologies can be achieved by designed reverse genetic approaches but I believe that the transposon forward random approaches offer several advantages such as the ability to generate large random libraries, the ability to isolate the most functionally desirable construct and the pure fun of the exploration.

Nested deletion families

Nested deletion families are of use in mapping specific protein functional regions or epitopes and in making large varieties of the deletion proteins. A straightforward transposon-based technology for generating nested deletions involves the generation of multiple transposon inserts in a gene of choice followed by PCR amplification using a primer upstream of the target gene's expression signals and a primer based in the transposon. The transposon needs to be constructed to have stop codons in all three reading frames very close to the end

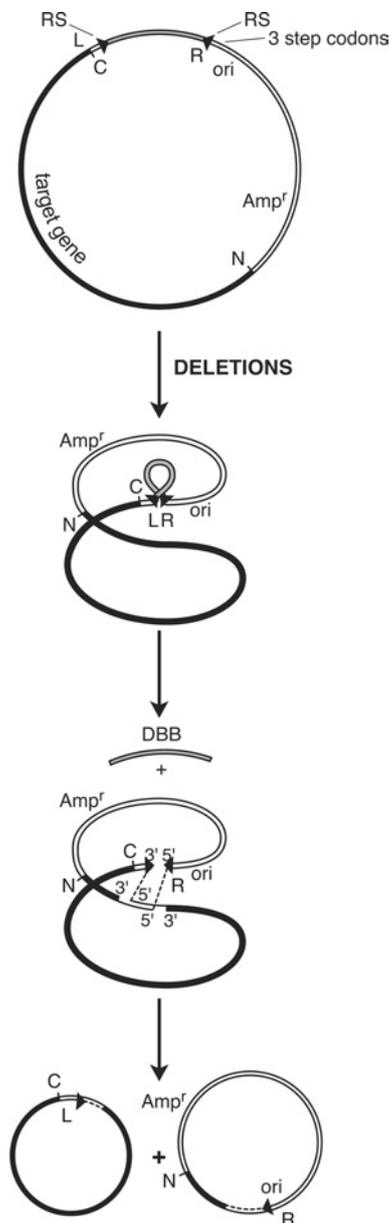
Key words: genome analysis, protein structure–function, recognition sequence, Tn5 transposition, transposon.

Abbreviations used: CFP, cyan fluorescent protein; Kan^R, kanamycin-resistance; RS, recognition sequence; YFP, yellow fluorescent protein.

¹email reznikoff@biochem.wisc.edu

Figure 1 | Nested deletion formation through intramolecular transposition

The most common form of transposition under most experimental conditions is intramolecular transposition and not the usually presented intermolecular transposition. This is because target binding is a concentration-dependent process and usually sequences on other molecules are at a lower concentration than transposon sequences. Intramolecular transposition can be used as a strategy for generating nested deletion families. The transposon is constructed to contain nonsense codons in all three reading frames adjacent to the appropriate transposon RS, a plasmid origin of replication (*ori*), an antibiotic-resistance marker (*Amp^r*), and the target gene (black) with appropriate expression signals. Intramolecular transposition will generate two families of molecules, inversions (not shown) and deletion circles (one pair is shown in this Figure). The circles can be distinguished from inversions by their size and loss of donor DNA genes. Different sized circles carrying the *ori*, the *Amp^r* gene and sequences encoding the N-terminus of a protein (N) with different sized deletions of the target gene's C-terminus are shown. DBB,



donor backbone. L, left end; R, right end. Modified from York, D., Welch, K., Goryshin, I. and Reznikoff, W.S. (1998) Simple and efficient generation *in vitro* of nested deletions and inversions: Tn5 intramolecular transposition. *Nucleic Acids Res.* **26**, 1927–1933, with permission. © 1998 Oxford University Press.

facing the target gene and a selectable marker and an origin of replication within the PCR product. Following insert library generation and PCR amplification, one simply ligates the PCR products to form circles and transforms cells to select the nested deletion family of plasmids [7].

However, my favourite approach for generating nested deletions makes use of the most efficient transposition reaction: intramolecular transposition [8]. This approach (shown in Figure 1) is conceptually complicated but simple operationally. The transposon is constructed to contain inward facing stop codons in all three reading frames at one of its termini, an origin of replication, a selectable marker, and the target gene in that order (and orientation) within the boundaries of the inverted transposon RSs. A standard *in vitro* transposition reaction is performed and most of the transposition events are intramolecular (usually these are throwaways) which generate a mixture of inversion events (of no interest in this case) and deletion circles. The products are introduced into host bacteria, selecting for the appropriate antibiotic resistance, and the sizes of the deletion circle plasmids are determined by agarose gel electrophoresis. Sequence analysis of the resulting deletion junctions demonstrates that the deletions are randomly spread out along the included target gene, thus generating the desired nested deletion family.

The example presented above and in Figure 1 demonstrates the generation of C-terminal nested deletions. A relatively simple redesign of the transposon to have outward facing transcription and translation initiation signals will allow the generation of N-terminal nested deletion families.

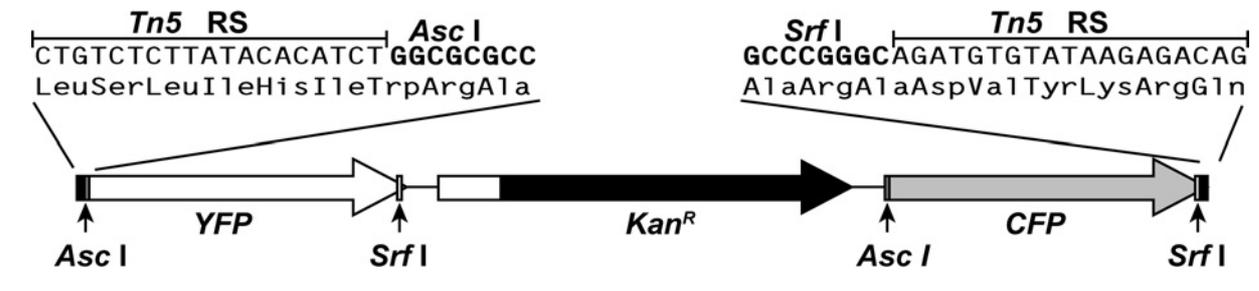
Reporter fusion construction

One of the most advantageous properties of transposition technology in protein analysis is the ability to make random fusions between the target protein and the desired reporter function. The randomness frees the investigator from guessing what would be the best fusion in terms of reporter activity and maintaining target protein function. Rather one generates a library of fusions and screens or selects for the best of both activities. Once the fusion is constructed, it can be used to determine the abundance and cellular location of the target protein, or the fusion protein can act as a biosensor of intracellular events through fluorescence yield changes or as a donor or receiver in energy transfer experiments.

An excellent example of this technology was recently published in a paper by Sheridan and Hughes [9]. These investigators developed a procedure that allows the creation of bi-coloured fusion libraries of fluorescent fusion proteins in which one can choose fusions that have maintained target

Figure 2 | Transposon for generating bicolour GFP (green fluorescent protein)-like functional fusions

The pictured transposon is used to generate Kan^R, YFP-positive inserts into the plasmid-located expressed target gene. In order to generate and find YFP fusions to a target gene in which the target gene maintains function, the fusion-containing plasmids are harvested, digested with SrfI, ligated and transformed, selecting for a plasmid-encoded gene. Then one carries out a selection or screen for target gene function. In this way, one generates functional fusions in which the YFP sequence is inserted at the N-terminus, at the C-terminus or at an internal position typically located at a permissible location between secondary structure motifs. From the original fusion, one can switch to a functional CFP fusion by performing an AscI digestion followed by ligation. Modified from [9], with permission.



protein function. Figure 2 describes one of the transposon reporter constructs that they have developed for this technology. The transposon carries two reporter genes encoding high-yield YFP (yellow fluorescent protein) and CFP (cyan fluorescent protein), a Kan^R (kanamycin-resistance) gene and strategically located restriction enzyme cleavage sites. One selects for Kan^R inserts in the plasmid-encoded target gene and screens for YFP-positive colonies (restricted to 1/6 of the inserts due to orientation and reading frame limitations). Treatment of the plasmid with SrfI followed by ligation results in removal of the Kan^R gene, the CFP gene and the stop codon internal to the transposon. The result is the generation of a target protein (N-terminus)–YFP–target protein (C-terminus) fusion that can be screened for target protein activity. Alternatively, an AscI cleavage followed by ligation generates a similar fusion protein containing the CFP sequence.

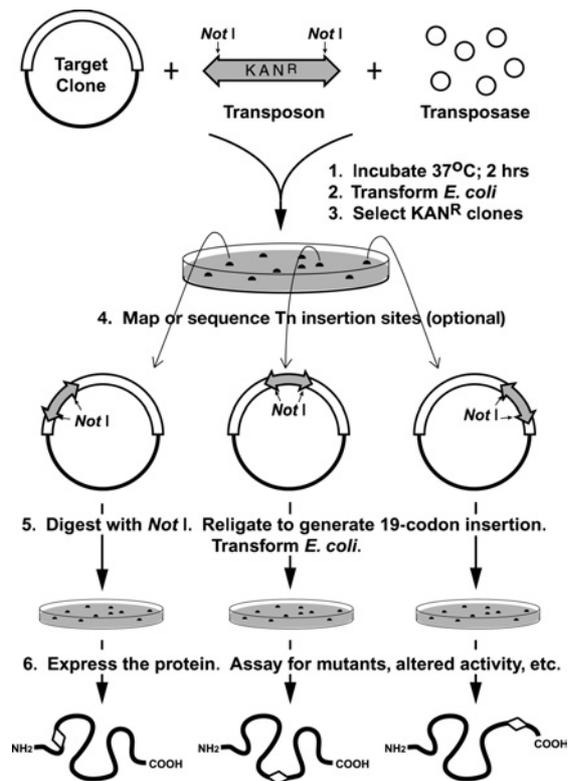
This technology is bound to give exciting constructs. In addition, the technology also maps surface-exposed regions of the target protein located between domains in a manner similar to that found in the domain mapping strategy below.

In-frame linker insertions

An additional method has been developed to map surface-exposed domain boundaries in target proteins. The general method has been developed for several transposon systems with the Tn5-related work being performed primarily in the laboratory of Manoil [10] (also see [11]). The relevant transposon and methodology is described in Figure 3. The transposon carries a Kan^R gene bracketed by two NotI sites. Inserts into a plasmid-encoded target gene are made using the kanamycin selection. The Kan^R gene is then excised by NotI digestion followed by ligation. This leaves behind one of three possible 19 codon (57 nt pair) insertions in the target gene. If full function of the target gene is restored, then the insertion is probably located in a surface-exposed region between two protein domains.

Figure 3 | In-frame linker insertion mutations

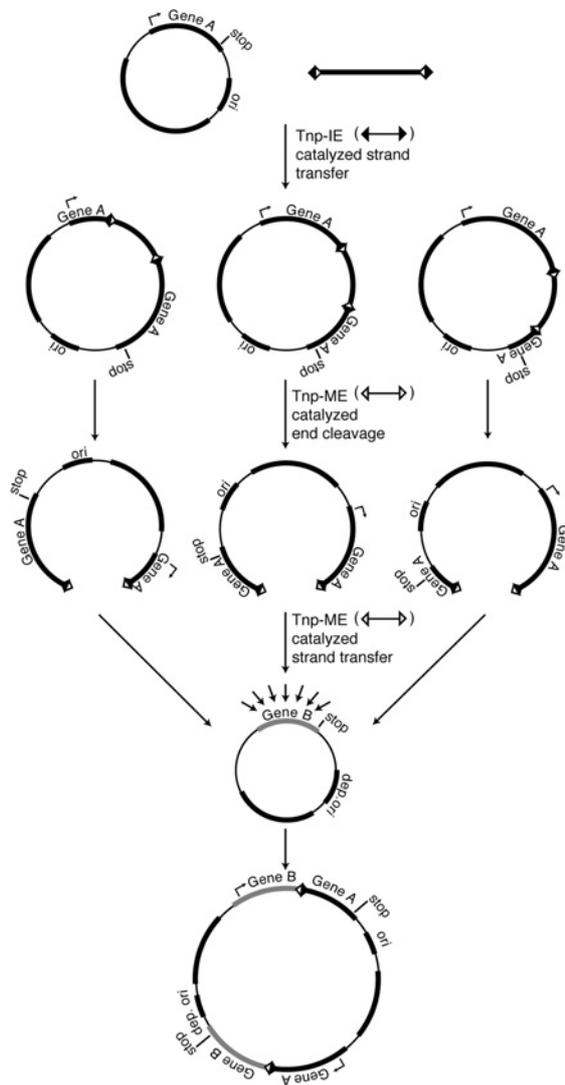
In-frame, 19-residue, linker insertions can be generated by insertion of a specially designed transposon followed by NotI digestion (to excise the bulk of the transposon sequences) and ligation. Reproduced with permission from *Epicentre Forum* 7(3), 2000. © 2000 EPICENTRE Technologies Corporation. All rights reserved.

**Random gene fusions**

The general characteristic of all the above technologies is that one partner of the fusion has a random fusion joint while the other partner has a joint fixed by the end of the transposon. We sought to develop a technology that allows

Figure 4 | Generating random combinatorial gene fusions by two sequential transposition events

This method makes use of the existence of two different Tn5-related transposases that recognize two different 19 bp signal sequences during the first step in transposition. The initial transposon is constructed to have at its ends two back-to-back inverted RSs of the different specificities (white and black triangles). Using black specific transposase, the transposon is inserted in gene A carried on a plasmid. This generates a new transposon defined by the white triangles. The white specific transposase is then used to excise this transposon and insert it randomly into gene B, thus creating A-B and B-A gene fusions. One-sixth of these fusions will be in-frame fusions with potential A+B hybrid characteristics. Tnp-IE, transposase inside end; Tnp-ME, transposase mosaic end. Modified from Naumann, T.A., Goryshin, I.Y. and Reznikoff, W.S. (2002) Production of combinatorial libraries of fused genes by sequential transposition reactions. *Nucleic Acids Res.* **30**, E119, with permission. © 2002 Oxford University Press.



random joints to be formed with both partners, with the idea that we could develop a type of combinatorial fusion genetics [12]. As shown in Figure 4, one manifestation of this

procedure involves double sequential transposition events using a transposon with hybrid back-to-back RSs (32 bp total length) and two sources of transposase, one that recognizes one RS and one that recognizes the second RS. The initial transposon/transposase combination is used to attack gene A encoded in a plasmid. This generates a second transposon defined by the second partner RS which is used with the appropriate second transposase to mobilize the second transposon carrying cleaved gene A into gene B. In this manner, random gene A–gene B and gene B–gene A fusions are generated with a linker between the two sequences. So far, this technology has only been used to regenerate functional chloramphenicol acetyltransferase genes from C-terminal–N-terminal deletions of the CAT^R (chloramphenicol acetyltransferase-resistance) genes. An interesting feature is that the active fusions always contain fusion joints located within unstructured regions between secondary structure determinants and in this way resemble the reporter gene fusion and linker insertion results described above.

Conclusion

The present paper offers some examples of how transposition technology can help dissect protein (and RNA) structure–function. So far, this technology is in its infancy. I believe that additional transposon-based techniques are bound to be developed and that these will be powerful tools for forward genetic approaches in protein structure–function studies.

Thanks to Jerry Jendrisak of Epicentre Technologies (Madison, WI, U.S.A.) and to the members of my laboratory for stimulating conversations and research work on the applied uses of Tn5 transposition, and to the Department of Biochemistry media department for the illustrations. The work performed in my laboratory was supported by grants from the National Institutes of Health (GM 50692) and National Science Foundation (MCB 0084089).

References

- Devine, S.E. and Boeke, J.D. (1994) *Nucleic Acids Res.* **22**, 3765–3772
- Gwinn, M.L., Stellwagen, A.E., Craig, N.L., Tomb, J.F. and Smith, H.O. (1997) *J. Bacteriol.* **179**, 7315–7320
- Haapa, S., Taira, S., Heikkinen, E. and Salviltahti, H. (1999) *Nucleic Acids Res.* **27**, 2777–2784
- Boeke, J.D. (2002) in *Mobile DNA II* (Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M., eds.), pp. 24–37, ASM Press, Washington, DC
- Hayes, F. (2003) *Annu. Rev. Genet.* **37**, 3–29
- Reznikoff, W.S. (2002) in *Mobile DNA II* (Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M., eds.), pp. 403–422, ASM Press, Washington, DC
- Fiandt, M. (2004) *Epicentre Forum* **11**, 20–21
- York, D., Welch, K., Goryshin, I. and Reznikoff, W.S. (1998) *Nucleic Acids Res.* **26**, 1927–1933
- Sheridan, D.L. and Hughes, T.E. (2004) *BMC Biotechnol.* **4**, 17–25
- Bailey, J. and Manoil, C. (2001) *Epicentre Forum* **8**, 1–3
- Hoffman, L.M. and Loomis, K.B. (2000) *Epicentre Forum* **7**, 4–6
- Naumann, T.A., Goryshin, I.Y. and Reznikoff, W.S. (2002) *Nucleic Acids Res.* **30**, E119

Received 11 October 2005