1866–1867 *Nucleic Acids Research, 1997, Vol. 25, No. 9*

Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein

Bernard Hallet, David J. Sherratt and Finbarr Hayes*

Microbiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received February 4, 1997; Accepted February 28, 1997

ABSTRACT

A new insertion method for probing protein functional organization was developed. The method relies on the random insertion of transposon Tn*4430* and subsequent *in vitro* deletion of the bulk of the transposon after which a 15 bp insertion remains within the target gene. This results in pentapeptide insertions randomly distributed in the target protein. Characterization of 23 pentapeptide insertions in TEM-1 β-lactamase demonstrated the utility of the method. The phenotypes associated with the mutated β-lactamase proteins equated both with the sorts of local peptide structures in which the pentapeptide insertions occurred and their position in the three-dimensional structure of the enzyme.

Linker insertion mutagenesis can provide invaluable insights into protein structure–function relationships. However, most linker insertion mutagenesis procedures are technically demanding and are limited to insertions at preexisting restriction enzyme sites (1). In this study, a simple and efficient transposon-based linker insertion mutagenesis strategy is described.

Transposon Tn*4430*, a Tn*3*-related transposon from *Bacillus thuringiensis*, transposes efficiently in *Escherichia coli* and duplicates 5 bp of host sequences at the insertion point (2, this study). The transposon also contains *Kpn*I restriction enzyme sites 5 bp from the outer ends of its terminal inverted repeats. Following transposition of Tn*4430* to a target gene, a 15-bp in-frame insertion remains if the bulk of the transposon is deleted by restriction with *Kpn*I and religation (Fig. 1). This results in a five amino acid insertion in the protein encoded by the target gene. As the sequence of the insertions is dependent on the 10 bp derived from Tn*4430* and on the sequence of the target site pentanucleotide duplication, the pentapeptide insertions will vary in composition (Fig. 1). No wild-type amino acids are simultaneously mutated.

The pentapeptide insertion mutagenesis strategy utilizes the formation of a cointegrate as a transposition intermediate by Tn*4430*. Plasmid pHT385 consists of the conjugative F derivative pOX38 into which Tn*4430*Ω5, a Tn*4430* derivative encoding kanamycin resistance, has been introduced (2). A donor strain which contains both pHT385 and the plasmid harbouring the target gene is mated with a plasmid-free recipient strain. By selecting for appropriate markers on the target plasmid and the recipient strain, transconjugants containing cointegrates between pHT385 and the target plasmid are isolated. These cointegrates are resolved rapidly *in vivo* resulting in target plasmid derivatives harbouring a copy of Tn*4430*Ω5 (2). Independent mating experiments are performed to isolate different insertions into the target gene. Isolates in which Tn*4430*Ω5 has inserted in the gene of interest are identified by an appropriate genetic screen and/or by restriction enzyme mapping. The bulk of Tn*4430*Ω5 subsequently is deleted by digestion with *Kpn*I and religation.

The utility of the above strategy was tested using the TEM-1 β-lactamase protein encoded by the *bla* gene on pBR322. Thirty independent insertions of Tn*4430*Ω5 were generated in *bla*. The positions of the 15-bp insertions remaining after *Kpn*I deletion of the bulk of the transposon were determined by sequencing. This analysis revealed that Tn*4430*Ω5 insertion was essentially random although five positions in *bla* acted as target sites on more than one occasion resulting in 23 unique insertion positions among the 30 clones characterized. The low insertion specificity of Tn*4430* is typical of members of the Tn*3* family (3).

The effects of the 23 different pentapeptide insertions on β-lactamase activity were assessed by determining the maximum ampicillin concentrations allowing growth conferred by these proteins at 30°C. The wild-type *bla* gene provided resistance to 5500 μg/ml of ampicillin whereas the host strain was sensitive to 10 μg/ml of this antibiotic. The set of mutated proteins was divisible into three categories: (i) five of the mutated β-lactamases conferred high levels of ampicillin resistance (≥2500 μg/ml); (ii) eight of the pentapeptide insertions abolished detectable β-lactamase activity; and (iii) the remaining 10 insertions resulted in intermediate levels of ampicillin resistance (50–1500 μg/ml). The pentapeptide random insertion mutagenesis procedure therefore can generate mutated proteins with a range of activity levels.

As linker insertions potentially confer thermosensitive properties on the mutated proteins (1), the resistance levels conferred by the mutated β-lactamase proteins were also analyzed at 37 and 42°C. All of the high-level and intermediate-level resistance proteins conferred decreased resistance at 37°C which was further reduced to varying levels at 42°C (data not shown). The wild-type β-lactamase protein itself provides less resistance at higher temperatures but the pentapeptide mutagenesis procedure may be an efficient technique for generating proteins with enhanced thermosensitive characteristics.

TEM-1 β-lactamase has an ellipsoidal structure consisting of two globular domains connected by two hinge regions. The depression lying at the interface of the two domains forms the catalytic cavity (4,5). Mapping of the pentapeptide insertions on this structure illustrated the usefulness of the pentapeptide scanning mutagenesis method as the phenotypes conferred by the mutated β-lactamases correlated both with the types of local

*To whom correspondence should be addressed. Tel: +44 1865 275304; Fax: +44 1865 275297; Email: fhayes@worf.molbiol.ox.ac.uk

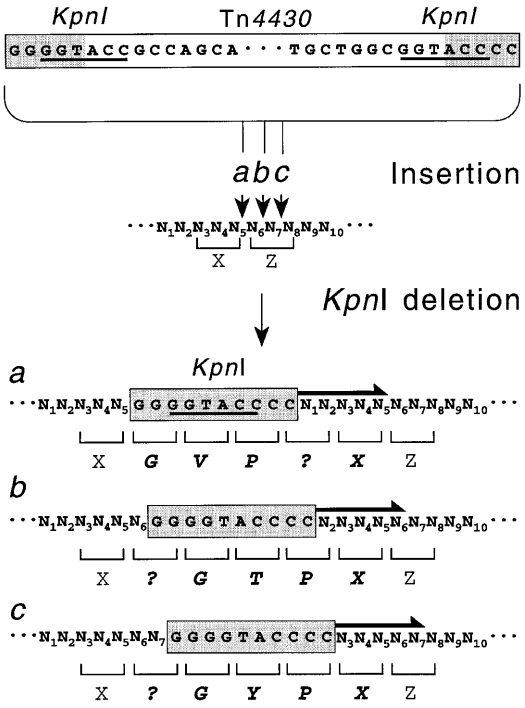

Figure 1. Principle of the pentapeptide scanning mutagenesis method. The open rectangle indicates Tn*4430* sequences. Only the 15 nucleotides at the outer ends of the terminal inverted repeats of Tn*4430* are shown. *Kpn*I restriction sites are underlined. Shaded residues indicate Tn*4430*-derived sequences which form part of the 15 bp insertion after *Kpn*I deletion of the bulk of Tn*4430*. The string of nucleotides denoted N_1N_2 etc. represents the sequence of the target coding region. a, b and c are the three possible insertion positions adjacent to or within a codon Z. Horizontal arrows indicate the 5 nucleotides duplicated during Tn*4430* insertion. The 15-bp insertions at each of the three positions specify a different set of pentapeptide sequences shown in bold italics. Each set contains a duplication of the upstream codon X and a variable codon '?' whose identity depends on N_1 and N_2 (position a), N_6 (position b) or N_6 and N_7 (position c). Only an insertion at position c of a tyrosine codon will generate a 'stop' (TAG) codon.

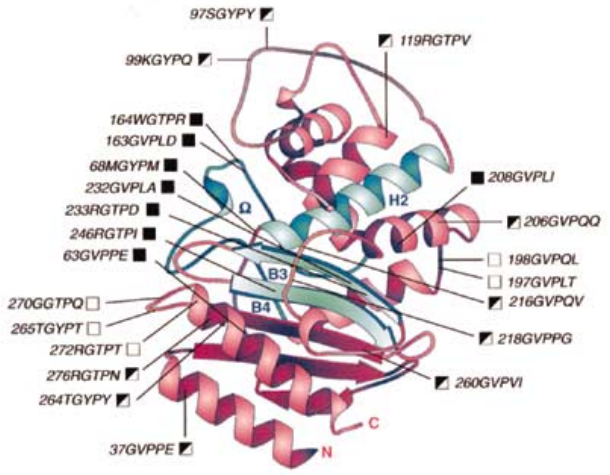

Figure 2. Sites of pentapeptide insertions on the three-dimensional structure of TEM-1 β-lactamase. Numbers correspond to insertion positions using the recommended numbering system (6). The sequences of individual pentapeptide insertions are indicated. Open, half-filled and filled boxes adjacent to insertion numbers indicate pentapeptide insertions which result in high-level, intermediate-level and loss of ampicillin-resistance, respectively. α-helix 2, β-strands 3 and 4, and the Ω-loop which contain catalytically important residues are indicated in blue. The figure was drawn using the MOLSCRIPT program (7).

peptide structures in which the insertions occurred and their location in the three-dimensional organization of the enzyme (Fig. 2). Insertions which abolished resistance to ampicillin were clustered around the catalytic cavity of the enzyme whereas insertions in other regions of the protein conferred a range of resistance levels which presumably reflect a range of structural defects. Pentapeptide insertions in the five mutated β-lactamase proteins which retained high levels of ampicillin resistance mapped to two protruding loops which are distant from the catalytic site (Fig. 2). Insertions which conferred intermediate levels of ampicillin resistance were found either in different regions of secondary structure which are not directly involved in the substrate binding cavity, in one of two hinge regions of the protein, or in a loop whose C-terminus forms the left border of the catalytic site (4,5). Pentapeptide insertions in α-helices and β-strands conferred a higher level of thermosensitivity than insertions in loops.

Pentapeptide scanning mutagenesis is a simple and rapid procedure which requires a single bacterial mating experiment followed by standard *in vitro* plasmid DNA digestion and religation manipulations. The only technical limitation of the method is the necessity for not having a *Kpn*I site within the target coding sequence. However, in many instances, undesirable *Kpn*I restriction enzyme sites could readily be removed by site-directed mutagenesis before employing the pentapeptide insertion procedure. The method promises to be especially useful in probing structure–function relationships in multifunctional proteins since insertions in some regions of a protein may specifically alter different protein activities whereas positions highly tolerant to pentapeptide insertions may pinpoint inter-domain regions. Indeed, application of the method to the XerD site-specific recombinase shows that a range of protein function alterations can readily be obtained (Y.Cao, B.Hallet, D.J.Sherratt and F.Hayes, unpublished data). The pentapeptide scanning mutagenesis procedure has the advantage of providing a unique *Kpn*I restriction enzyme site at the insertion point within the mutated gene. This not only facilitates mapping of the mutations but also potentially allows further manipulations, e.g., expression of protein subdomains, construction of protein fusions, or the introduction of useful peptides such as protein tags or specific epitopes.

ACKNOWLEDGEMENTS

B.H. and F.H. contributed equally to this work which was funded by a Wellcome Trust Research Career Development Fellowship (no. 040822/Z/94/Z/1.4U/AWB/CS) to F.H. B.H. was supported by a European Molecular Biology Organization Long-Term Fellowship. We thank François Cornet, Keith Dyke, Jean-Marie Frère, Rene Rezsöhazy and Andrew Spiers for constructive criticism of the manuscript, Didier Lereclus for kindly providing pHT385 and for helpful discussions, and David Levinger for assistance with the MOLSCRIPT program.

REFERENCES

1 Goff,S.P. and Prasad,V.R. (1991) *Methods Enzymol.* **208**, 586–603.
2 Mahillon,J. and Lereclus,D. (1988) *EMBO J.* **7**, 1515–1526.
3 Davies,C.J. and Hutchinson,C.A. (1995) *Nucleic Acids Res.* **23**, 507–514.
4 Jelsch,C. *et al.* (1993) *Proteins: Struct. Funct. Genet.* **16**, 364–383.
5 Strynadka,N.C.J. *et al.* (1992) *Nature* **359**, 700–705.
6 Ambler,R.P. *et al.* (1991) *Biochem. J.* **276**, 269–272.
7 Kraulis,P. (1991) *J. Appl. Cryst.* **24**, 946–950.