

Linking the functions of unrelated proteins using a novel directed evolution domain insertion method

Wayne R. Edwards^{1,2}, Kathy Busse¹, Rudolf K. Allemann² and D. Dafydd Jones^{1,*}

¹School of Biosciences and ²School of Chemistry, Cardiff University, Cardiff, UK

Received April 11, 2008; Revised April 29, 2008; Accepted May 22, 2008

ABSTRACT

We have successfully developed a new directed evolution method for generating integral protein fusions comprising of one domain inserted within another. Creating two connections between the insert and accepting parent domain can result in the inter-dependence of the separate protein activities, thus providing a general strategy for constructing molecular switches. Using an engineered transposon termed MuDel, contiguous trinucleotide sequences were removed at random positions from the *bla* gene encoding TEM-1 β -lactamase. The deleted trinucleotide sequence was then replaced by a DNA cassette encoding cytochrome *b*₅₆₂ with differing linking sequences at each terminus and sampling all three reading frames. The result was a variety of chimeric genes encoding novel integral fusion proteins that retained TEM-1 activity. While most of the tolerated insertions were observed in loops, several also occurred close to the termini of α -helices and β -strands. Several variants conferred a switching phenotype on *Escherichia coli*, with bacterial tolerance to ampicillin being dependent on the presence of haem in the growth medium. The magnitude of the switching phenotype ranged from 4- to 128-fold depending on the insertion position within TEM-1 and the linker sequences that join the two domains.

INTRODUCTION

Proteins consisting of multiple domains are commonly observed in nature and their domain arrangement is generally important for function (1). In most cases domain fusions are arranged tandemly in a 'head-to-tail' manner but 9% contain the insertion of one protein within another to produce integral fusion proteins (2). This results in the accepting or 'parent' domain being split by the 'insert' domain. Such an arrangement potentially

allows communication between the separate domains by linked conformational changes, with the integral fusion protein thus acting as a molecular switch. A classic example in nature that illustrates this is the GroEL chaperone where an integral domain arrangement links the peptide binding and ATPase sites to allow the efficient folding of proteins (3).

The ability to build tailored protein switches that change their properties in response to a desired input will allow significant new possibilities for creating novel sensors, modulators and transducers for use in both natural and artificial contexts (4–7). The construction of novel integral protein fusions comprised of normally disparate domains or proteins may provide a general approach for creating such switches (8,9). Although the generation of recombinant fusion proteins is common, they are usually constructed 'end-to-end' so that each domain acts autonomously. Insertion of one protein within another creates two shared links, so decreasing the degrees of freedom between the two domains and intimately linking their structures. This promotes communication between the functional centres so that one domain influences the structure and therefore activity of the other.

The key to success of the domain insertion strategy is the identification of sites within a protein permissible to insertions of whole domains that also retains and links the functions of the individual domains. Such stringent criteria make predicting potential insertion sites difficult. There has been some success with rationally designed protein switches (10–12) but generally a number of different insertion positions were sampled (mostly within loops) and only modest switching magnitudes were observed. The construction of integral fusion proteins was therefore greatly aided by the development of combinatorial approaches to domain insertion that used DNaseI to introduce random breaks into a target DNA segment (13,14). However, it is notoriously difficult to generate single cuts in DNA using DNaseI and digestion with this nonspecific nuclease regularly produces tandem duplications and nested deletions of unpredictable sizes within the parent gene. This leads to frameshifts along with insertions and deletions of greatly varying lengths in the parent

*To whom correspondence should be addressed. Tel: +44 (0) 29 2087 4290; Fax: +44 (0) 29 2087 4116; Email: jonesdd@cf.ac.uk
Present address:

Kathy Busse, Institut für Biochemie der Med. Fakultät, Universität Leipzig, Leipzig, Germany

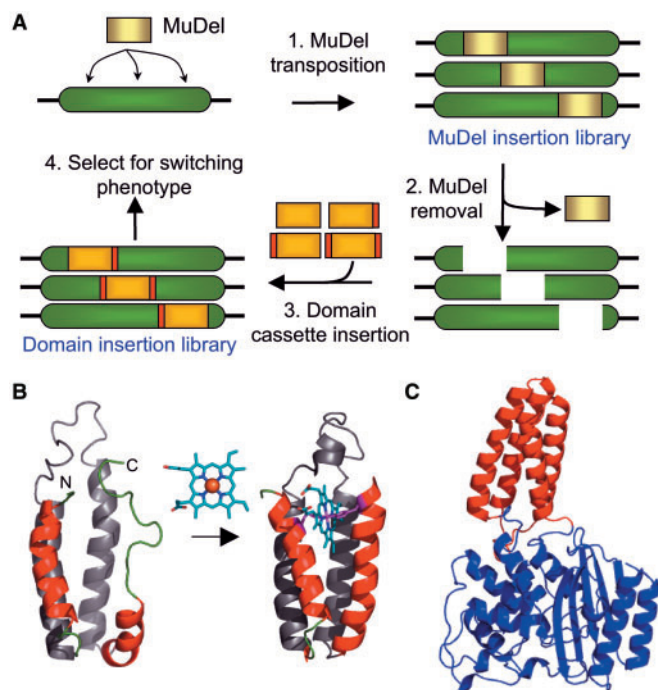


Figure 1. The domain insertion strategy for coupling protein functions. (A) Schematic outline of the transposon-based domain insertion method. *Step 1.* The MuDel insertion library is generated by the *in vitro* transposition of MuDel (yellow) in the target DNA (green). *Step 2.* Restriction endonuclease digestion removes MuDel together with 3 bp of the original target DNA to generate a single break per molecule. *Step 3.* The domain insertion library is created by the ligation of a DNA cassette (orange) encoding the insert domain with various linker sequence combinations (red appendages) into the break within the target DNA. *Step 4.* The domain insertion library is then screened for a switching phenotype. (B) Conformational changes to apo-cyt b [PDB 1apc (18)] on binding haem [PDB 1qpu (17)]. (C) Model of a potential integral fusion protein with cyt b inserted at an arbitrarily chosen site within TEM-1 β -lactamase.

protein, so reducing library quality and increasing the number of variants that need to be sampled.

A proposed transposon-based method (Figure 1A) could be used as an alternative to DNaseI to generate random insertions within a target gene. The engineered transposon termed MuDel has been used previously to generate random trinucleotide deletions throughout a target gene (15,16). Novel functional integral fusion proteins aimed at linking the haem binding properties of cytochrome b_{562} (cyt b) to the β -lactamase activity of TEM-1 could be constructed using this transposon-based approach. Haem is a biologically important co-factor that plays a role in many cellular processes, including oxygen transport, catalysis, electron transfer and sensing. Cyt b is a small four helical bundle electron transfer protein found in the periplasm of *E. coli* that undergoes a major conformational change on haem binding (Figure 1B) (17,18). Haem affinity for cyt b is oxidation state dependent, which potentially provides a way of monitoring cellular redox changes. Cyt b has been used previously to link redox-dependent haem binding events to other functions, such as DNA recognition (19,20). However, coupling haem binding to a reporter protein with an easily detected output would be advantageous.

TEM-1 is one of the main enzymes responsible for conferring bacterial resistance to β -lactam antibiotics such as penicillin (21). TEM-1 has been used previously in domain insertion strategies as a reporter, but acted as the insert protein (11,12,14,22). The use of wild-type TEM-1 as the insert protein has the disadvantage that all switching events in the reporter had to be channelled through the termini, which are relatively remote from the active site. Hence observed switching magnitudes were usually small. The use of circularly permuted versions of TEM-1 for domain insertion has been shown to significantly enhance the switching properties of this reporter (14,23).

Here we show that our method can generate a range of chimeric genes encoding functional integral fusions with cyt b inserted at a variety of different positions within TEM-1. Many of these integral fusion proteins exhibited switching behaviour, with *in vivo* β -lactamase activity regulated positively or negatively in the presence of exogenous haem.

MATERIALS AND METHODS

Construction and digestion of the BLA Δ ¹⁹⁸ library

The transposon insertion library (termed BLA Δ ¹⁹⁸) that contained 198 variants with MuDel inserted randomly within the *bla* gene was a subset of a previously generated library (16). Briefly, transposition was performed using MuA transposase (Finnzymes), MuDel and the *bla* gene resident within the plasmid pNOM as template [see (15) for details]. The transposition reaction products were used to transform *E. coli* DH5 α by electroporation and the cells grown on LB agar containing 25 μ g/ml chloramphenicol (Cam). The selection of clones with MuDel inserted within the *bla* gene of pNOM was performed using a positive selection for Cam resistance followed by a screen for ampicillin (Amp) sensitivity (15). Each of the selected colonies was transferred to 96 well culture plates (Nunc) containing LB and 25 μ g/ml Cam and grown overnight at 37°C. A 50 μ l aliquot of each culture was pooled and plasmid DNA isolated using the Qiagen HiSpeed plasmid midi kit. The plasmid pool was digested with MlyI (NE Biolabs) to remove MuDel (Figure 2A). The cleaved DNA was then treated with APexTM heat-labile alkaline phosphatase (Epicentre Biotechnologies) for 30 min at 37°C. After inactivation of the phosphatase by heating to 70°C for 10 min, the linearized pNOM DNA was separated by agarose gel electrophoresis and purified.

Construction and analysis of the domain insertion library

The *cybC* cassettes constructed as outlined in Supplementary Methods using various combinations of oligonucleotide primers (Supplementary Tables 1 and 2) were ligated into the break in pNOM left by MuDel removal. Ligations were performed using the Quick T4 DNA Ligase kit (NE Biolabs) in a total volume of 20 μ l containing 50 ng linear pNOM and 22.6 ng of a single *cybC* ORF cassette library (molar ratio of 1:3). The ligation products were used to transform *E. coli* DH5 α cells by electroporation. Transformed cells were grown on LB agar supplemented with 25 μ g/ml Amp at 37°C for at least 16 h. Amp-resistant

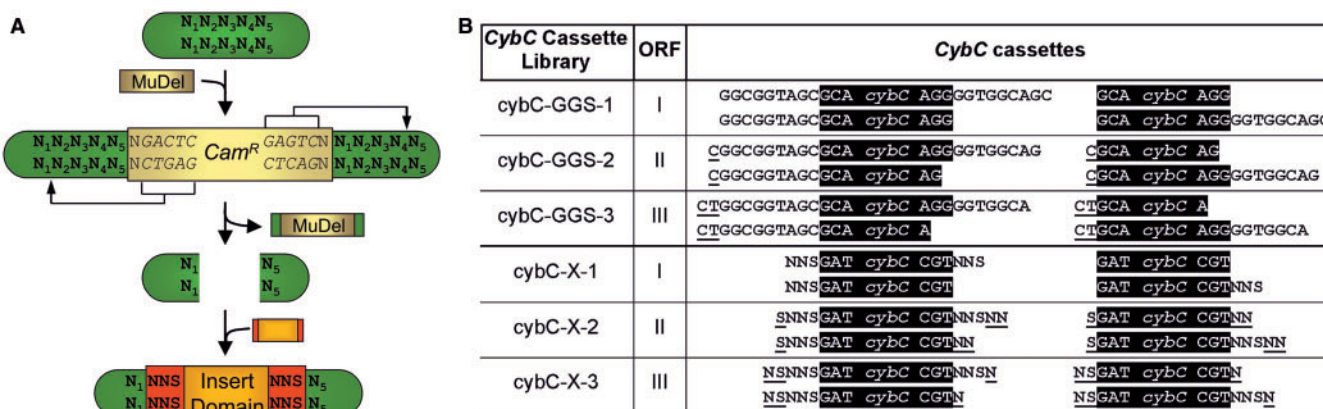


Figure 2. Mechanism of domain insertion. (A) Insertion of MuDel results in the duplication of 5 bp of the target DNA at the insertion point. Digestion with MlyI removes MuDel together with 8 bp of the target DNA (4 bp at each end), which equates to removal of a contiguous 3 bp sequence from the original gene. The DNA cassette encoding the domain insert is then ligated into the gap vacated by MuDel and effectively replaces the trinucleotide sequence deleted from the target gene. Colours as in Figure 1A. (B) Composition of each *cybC* DNA cassette library. Sequence contributed by *cybC* is highlighted in white on a black background with linker sequences shown as appendages. ORF refers to the open reading of the particular cassette library. Additional nucleotides used to maintain reading frame are underlined. N refers to either T, A, G or C, while S refers to C or G.

colonies were screened by colony PCR using the GoTaq DNA polymerase (Promega) system and primers DDJdi010 and DDJdi011 (Supplementary Table 1). Clones deemed to have a *cybC* cassette inserted within *bla* were further analysed by colony PCR using primers DDJdi010 (vector-based) and WREcbr012 (*cybC*-based). Full-length PCR products generated using DDJdi010 and DDJdi011 were sequenced (DNA Sequencing Core, Molecular Biology Unit, Cardiff University).

Analysis of the minimum inhibitory concentration (MIC) of Amp required to prevent cell growth was performed in 96-well flat-bottomed plates containing M9 minimal medium (24) supplemented with 0.001% w/v thiamine, 0.5% v/v glycerol, 0.1% w/v NaHCO₃ and either 10 μM hemin (Fluka) or an equivalent volume of 0.5 M NaOH. Hemin was prepared fresh as a 10 mM stock solution in 0.5 M NaOH and filtered through a 0.45 μm filter to remove aggregated material. Amp was added to the medium as a two-fold serial dilution from 32 000 μg/ml to 8 μg/ml. The wells were filled and covered with an impermeable plate seal to prevent gaseous exchange. Cells were grown under oxygen starvation conditions for about 20–24 h at 37°C. Optical density (OD) was measured at 650 nm using a Molecular Devices Thermomax microplate reader. Cultures with an OD₆₅₀ that was ≤5% of that in the absence of Amp were deemed to exhibit no growth. Western blot analysis of periplasmic expression levels of tested proteins was performed as outlined in Supplementary Methods.

RESULTS

Linking haem binding to β-lactamase activity

Our rationale for constructing integral domain fusions that act as molecular switches is based on the concept of allostery (25). Natural allosteric proteins have distinct regions for activity and its regulation. Switching is achieved by binding of an effector molecule (typically

a small molecule) to the regulation site, causing conformational changes that result in the direct modulation of protein activity. The architecture of integral domain fusion proteins should allow this process to be mimicked by creating an intimate link between two normally unrelated proteins to promote the transmission of conformational changes (Figure 1C). One criterion for the insert domain is that the N- and C-termini need to lie close in space in at least one of its conformations to allow the two sections of the parent protein to be juxtaposed in the integral fusion protein and facilitate the formation of an active protein. As ca 50% of known protein structures have their N- and C-termini proximal (26), many proteins should be amenable to productive domain insertion.

Cyt b and TEM-1 were chosen to act as the regulation and reporter domains, respectively. Cyt b binds haem non-covalently close to its termini resulting in major conformational changes (Figure 1B). This is particularly marked in the C-terminal region, which changes from a dynamic structure into a helix locked into its conformation by the co-ordination of a histidine to the haem iron (17,27). Coupled with the close proximity of the N- and C-termini (Figure 1B), this makes cyt b an ideal insert domain that can act as a regulatory element. TEM-1 has a number of properties that make it an ideal reporter, such as the ability to display switching through changes to catalysis that can be analysed both *in vivo* and *in vitro*. Therefore, by using TEM-1 as the parent domain and cyt b as the insert domain it was anticipated that haem-dependent conformational changes could be transmitted to TEM-1, so modulating enzyme activity of the parent domain.

Construction of domain insertion libraries

Our directed evolution domain insertion approach was implemented as outlined in Figure 1A. Initially, single breaks within the target *bla* gene were introduced using an engineered variant of the Mu transposon (28) termed MuDel (15). MuDel was originally constructed for

randomly removing contiguous trinucleotide sequences to generate single amino acid deletion variants of a protein (15). The deletion of a single amino acid will introduce a suitable gap in the structure of the parent protein to accommodate the inserted domain (10,23). Trinucleotide deletion was achieved by digestion with the type IIS restriction endonuclease MlyI (5'GAGTC(N)₅↓) that removes all the transposon sequence and 3 bp of the target gene to generate a single and defined blunt-ended break in both strands of the DNA (Figure 2A). *In vitro* transposition with MuDel is accurate and efficient with a very low target site preference allowing transposon insertion to occur randomly throughout a gene (15,16,28). As one transposon is inserted per gene, only a single break in the DNA will be introduced. The DNA cassette encoding the insert domain is introduced by ligation to replace the deleted trinucleotide sequence and generate the chimeric gene (Figure 2A).

The nature of the DNA cassette encoding the insert domain is critical; two factors that need to be considered are linker sequence and coding frame. Short linkers may adversely affect the tolerance of the parent protein to a domain insertion, whereas long linkers may result in functional decoupling of the two domains. In this work, two types of linkers were incorporated into the *cybC* cassette encoding cyt b (Figure 2B). The first constituted an extended flexible linker encoding a consensus sequence based on the tripeptide Gly-Gly-Ser. The second linker was encoded by the NNS codon allowing the nature of the single amino acid comprising the linker to vary. The alanine that normally occupies the N-terminus of mature cyt b was removed from this set of libraries to reduce the potential linker flexibility, as the residue does not appear to be part of the first helix in holo-cyt b (17,27). The *cybC* cassette was constructed so that all possible linker variations (no linker and linker at one or both termini) would be sampled (Figure 2B).

Due to the nature of transposon insertion, the single break can occur at three different positions with respect to a single codon (15). The use of only one open reading frame (ORF) for the *cybC* cassette would make 2/3 of the library unproductive due to frameshifts, so reducing the potential diversity of insertion positions sampled. Therefore, two further versions of the *cybC* cassette were constructed with both termini modified to compensate for any potential frameshifts (Figure 2B).

A total of six separate libraries of *cybC* inserts were constructed to represent the different linker sequences and ORFs: *cybC*-GGs-1 to 3 comprised the different tripeptide linker combinations in each reading frame and *cybC*-X-1 to 3 contained the randomized single amino acid linker combinations, also in each reading frame.

A previously constructed library of 198 *bla* gene variants with MuDel inserted at different positions (16), termed BLA Δ ¹⁹⁸, was used as the starting point. MuDel was removed from the plasmid pool representing the 198 variants and each *cybC* cassette library was ligated individually into the different break points in *bla*, followed by transformation of *E. coli*. Cell growth on LB agar supplemented with 25 μ g/ml Amp was used to select clones containing modified *bla* genes encoding active TEM-1.

Tolerance of TEM-1 to insertion of cyt b

A total of 70 colonies derived from the *cybC*-GGs libraries and 42 colonies derived from the *cybC*-X libraries were arbitrarily chosen and screened by PCR to identify chimeric *bla* genes with in-frame *cybC* insertions. We found 42 of the *cybC*-GGs derived clones and 12 of the *cybC*-X derived clones to contain an insert within *bla* (data not shown). Of these, two appeared to contain two *cybC* cassettes inserted within *bla*. The colonies that did not contain an insert generated PCR products close to the size expected for wild-type *bla* (data not shown), probably due to recircularization of the host plasmid. Reconstitution of the *bla* gene with a trinucleotide segment removed can give rise to false positives, as TEM-1 can tolerate deletion of an amino acid at various positions (16). Considering that intramolecular ligation of plasmid DNA is far more efficient than intermolecular ligation, our efforts to minimize recircularization by dephosphorylation were largely successful as 48% of the colonies contained a chimeric *bla*-*cybC* gene. The drop in frequency of positive clones derived from the *cybC*-X libraries could be a consequence of decreased tolerance of TEM-1 to the insertion of cyt b with shorter linkers.

An indication of the distribution of tolerated insertion positions within *bla* was obtained by PCR using one primer that annealed 58 bp upstream of *bla* and a second that bound to the 3' end of the *cybC* insert. Depending on the insertion position within *bla*, the expected PCR fragment size would range from *ca* 380 bp (insertion at 5' end of *bla*) to *ca* 1250 bp (insertion at 3' end of *bla*). The size distribution of the various fragments indicated the insertion of *cybC* occurred throughout *bla* (Supplementary Figure 1).

To confirm both the exact insertion position and nature of any required linker sequence, 20 variants derived from the *cybC*-GGs libraries and all 12 variants derived from the *cybC*-X libraries were sequenced (Supplementary Table 3). Eighteen different insertion positions distributed throughout *bla* were observed (Figure 3) and all but 1 of the variants was unique at the genetic level (Supplementary Table 3). The decision to sample all three reading frames was also vindicated as more than 2/3 of insertion positions (13/18) required frameshift compensation to encode the anticipated integral fusion protein (Figure 3). A total of 24 of the 32 sequenced variants contained frame shift-compensating *cybC* inserts (Supplementary Table 3). A diverse range of linking sequences was also observed, ranging from no linker to linkers at both connection positions (Table 1 and Supplementary Table 3). Two variants (Tc-06 and Tc-13) were found to have two *cybC* cassettes inserted in tandem within a single *bla* site.

Sixteen distinct positions within the primary structure of TEM-1 were found to tolerate insertion of cyt b (Table 1). The tolerated insertion positions were found throughout the protein, but with a higher proportion in the C-terminal half of the protein. Mapping the insertion positions onto the tertiary structure of TEM-1 revealed that most occurred in loops, or close to the terminus of a helix or β -strand (Figure 4 and Table 1). A 7-residue loop (comprising residues 112–118) linking helices H3 and

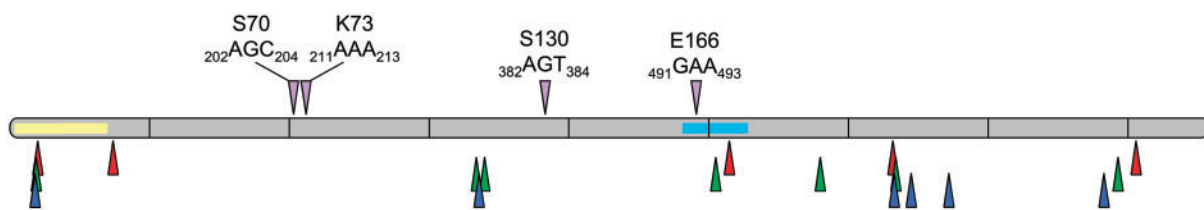


Figure 3. Sequence analysis of *cybC* cassette insertion at positions tolerated within *bla*. Red, green and blue arrowheads signify insert positions of *cybC* cassettes in ORF I, II or III, respectively. The codons for catalytically important residues (S70, K73, S130 and E166) are indicated on the diagram. The regions encoding the signal sequence and Ω loop are shown as yellow and cyan lines, respectively.

Table 1. Sequence and *in vivo* switching properties of integral cyt b-TEM-1 fusion proteins

| Amino Acid(s) replaced ^a | Variant number | Secondary structure ^b | Fusion protein ^c | Amp MIC (μ g/ml) | | |
|-------------------------------------|----------------|----------------------------------|--|-----------------------|-----------------|--------------------------|
| | | | | No haem | 10 μ M haem | Fold-change ^d |
| n/a | pNOM | – | – | 40960 | 40960 | 0 |
| n/a | pPB10 | – | – | 40960 | 40960 | 0 |
| n/a | DH5 α | – | – | < 8 | < 8 | 0 |
| F6L-R7 Δ | Tc-25 | ss | TEM ₅ - <u>E</u> -DL-cytb-YR- <u>IS</u> - ₈ TEM | 64 | 64 | 0 |
| R7 Δ | Tc-11 | ss | TEM ₆ - <u>GS</u> -ADL-cytb-YR- <u>YS</u> - ₈ TEM | 8192 | 8192 | 0 |
| R7 Δ | Tc-26 | ss | TEM ₆ - <u>G</u> -DL-cytb-YR- <u>HP</u> - ₈ TEM | 4096 | 4096 | 0 |
| R7 Δ | Tc-8 | ss | TEM ₆ -ADL-cytb-YR- <u>GGG</u> - ₈ TEM | 8192 | 8192 | 0 |
| R7 Δ | Tc-17 | ss | TEM ₆ - <u>GGG</u> -ADL-cytb-YR- <u>GGG</u> - ₈ TEM | 8192 | 8192 | 0 |
| R7 Δ | Tc-21 | ss | TEM ₆ - <u>S</u> -DL-cytb-YR- ₈ TEM | 512 | 512 | 0 |
| R7 Δ | Tc-27 | ss | TEM ₆ - <u>G</u> -DL-cytb-YR- ₈ TEM | 8192 | 8192 | 0 |
| P27 Δ | Tc-10 | H1 | TEM ₂₆ - <u>GGG</u> -ADL-cytb-YR- <u>GGG</u> - ₂₈ TEM | 32768 | 32768 | 0 |
| T114 Δ | Tc-2 | LP H3-H4 | TEM ₁₁₃ - <u>GGG</u> -ADL-cytb-YR- <u>GGR</u> - ₁₁₅ TEM | 32768 | 32768 | 0 |
| T114S-D115 Δ | Tc-7 | LP H3-H4 | TEM ₁₁₄ - <u>GGG</u> -ADL-cytb-YR- <u>GGN</u> - ₁₁₆ TEM | 32768 | 32768 | 0 |
| D115 Δ | Tc-4 | LP H3-H4 | TEM ₁₁₄ - <u>GGG</u> -ADL-cytb-YR- <u>GGN</u> - ₁₁₆ TEM | 16384 | 16384 | 0 |
| D115 Δ | Tc-5 | LP H3-H4 | TEM ₁₁₄ -ADL-cytb-YR- <u>GGN</u> - ₁₁₆ TEM | 8192 | 8192 | 0 |
| D115 Δ | Tc-6 | LP H3-H4 | TEM ₁₁₄ -ADL-cytb- <u>GGPGGS</u> -cytb-YR- <u>YN</u> - ₁₁₆ TEM | 8192 | 32768 | 0.25 |
| G116 Δ | Tc-12 | LP H3-H4 | TEM ₁₁₅ -ADL-cytb-YR- <u>YS</u> - ₁₁₇ TEM | 32768 | 32768 | 0 |
| G116 Δ | Tc-13 | LP H3-H4 | TEM ₁₁₅ -ADL-cytb-cytb-YR- <u>YS</u> - ₁₁₇ TEM | 2048 | 2048 | 0 |
| G116 Δ | Tc-19 | LP H3-H4 | TEM ₁₁₅ -ADL-cytb-YR- <u>GGG</u> - ₁₁₇ TEM | 32768 | 32768 | 0 |
| E171 Δ | Tc-15 | Ω loop | TEM ₁₇₀ - <u>GGG</u> -ADL-cytb-YR- <u>GGR</u> - ₁₇₂ TEM | 512 | 512 | 0 |
| P174 Δ | Tc-9 | Ω loop | TEM ₁₇₃ - <u>GS</u> -ADL-cytb-YR- ₁₇₅ TEM | 256 | 128 | 2 |
| G196 Δ | Tc-3 | LP H8-H9 | TEM ₁₉₅ -ADL-cytb-YR- <u>YS</u> - ₁₉₇ TEM | 32768 | 4096 | 8 |
| A213 Δ | Tc-1 | LP H9-H10 | TEM ₂₁₂ - <u>GGG</u> -ADL-cytb-YR- <u>GGG</u> - ₂₁₄ TEM | 16384 | 512 | 32 |
| A213G-D214 Δ | Tc-29 | LP H9-H10 | TEM ₂₁₃ - <u>V</u> -DL-cytb-YR- <u>N</u> - ₂₁₅ TEM | 16384 | 128 | 128 |
| A213G-D214 Δ | Tc-32 | LP H9-H10 | TEM ₂₁₃ -DL-cytb-YR- <u>N</u> - ₂₁₅ TEM | 16384 | 256 | 64 |
| D214 Δ | Tc-20 | LP H9-H10 | TEM ₂₁₃ - <u>GGG</u> -ADL-cytb-YR- <u>GGG</u> - ₂₁₅ TEM | 32768 | 16384 | 2 |
| A217G-G218 Δ | Tc-28 | LP H9-H10 | TEM ₂₁₇ - <u>S</u> -DL-cytb-YR- <u>PR</u> - ₂₁₉ TEM | 16384 | 4096 | 4 |
| P226R-A227 Δ | Tc-23 | LP H10-S3 | TEM ₂₂₆ - <u>G</u> -DL-cytb-YR- <u>LP</u> - ₂₂₈ TEM | 4096 | 1024 | 4 |
| P226R-A227 Δ | Tc-30 | LP H10-S3 | TEM ₂₂₆ - <u>G</u> -DL-cytb-YR- <u>T</u> - ₂₂₈ TEM | 16384 | 1024 | 16 |
| T265M-T266 Δ | Tc-22 | S5 | TEM ₂₆₅ - <u>M</u> -DL-cytb-YR- <u>GA</u> - ₂₆₇ TEM | 4096 | 512 | 8 |
| S268 Δ | Tc-24 | LP S5-H11 | TEM ₂₆₇ - <u>R</u> -DL-cytb-YR- ₂₆₉ TEM | 16384 | 8192 | 2 |
| Q269 Δ | Tc-14 | LP S5-H11 | TEM ₂₆₈ - <u>GGG</u> -ADL-cytb-YR- ₂₇₀ TEM | 32768 | 32768 | 0 |
| Q269 Δ | Tc-18 | LP S5-H11 | TEM ₂₆₈ -ADL-cytb-YR- ₂₇₀ TEM | 16384 | 4096 | 4 |
| D273 Δ | Tc-31 | H11 | TEM ₂₇₂ -DL-cytb-YR- <u>H</u> - ₂₇₄ TEM | 8192 | 512 | 32 |

^aResidue numbering according to Ambler (37). Δ signifies residue deleted

^bSecondary structure assignments according to Jelsch *et al.* (36). LP signifies loop and ss signal sequence.

^cLinker sequences are shown in bold and underlined. The sequence segment representing cyt b is coloured red. Only terminal cyt b residues are shown, with the intervening sequence indicated by 'cytb'.

^dFold change calculated by dividing Amp MIC value in the absence of haem by value in the presence of haem.

H4 was particularly tolerant to insertion of cyt b. Eight of the variants, including the two with tandem insertion of two cyt b modules, contained an insertion in this region. Several of the chimeric proteins contained insertions

in regions considered important to TEM-1 catalysis and structure (21), such as within the Ω -loop (residues 163–178) (Figures 3 and 4). We also found an insertion that replaced Gly196. TEM-1 has been shown previously

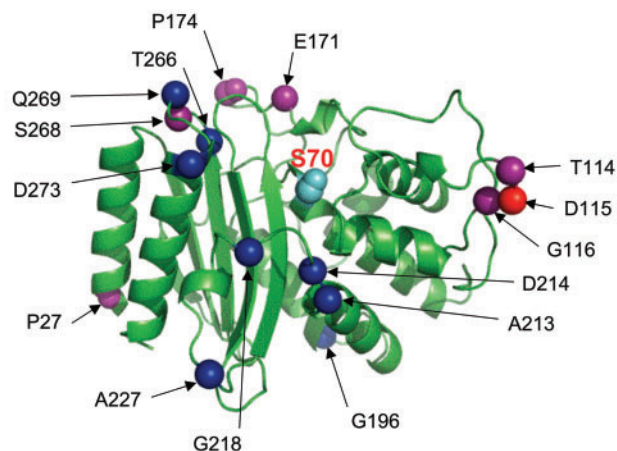


Figure 4. Position of cyt b insertion positions on the tertiary structure of TEM-1 [PDB code 1bt1 (36)]. Residues highlighted as blue or red balls represent insertion positions in which negative or positive modulation, respectively, was observed. Residues highlighted as purple balls represent insertion positions where no significant switching was observed. The catalytic serine (S70) is shown as space-fill. Figure was generated using PyMol (DeLano Scientific LLC).

to tolerate bisection in this region (29). Seven of the variants contained an insertion within the signal sequence of TEM-1 and could not be mapped onto the structure of the mature protein (Table 1). It is reasonable to assume that *in vivo* cleavage of the presequence by leader peptidase would separate the cyt b domain from mature TEM-1, and these clones are therefore not true integral fusion proteins.

Analysis of the linker regions indicated that most of the variants required at least one linker (Table 1). Five variants (Tc-03, Tc-06, Tc-13, Tc-16 and Tc-18) had no linker, which included both variants with the insertion of two tandem cyt b domains (Tc-06 and Tc-13). Analysis of variants derived from the cybC-X-series of libraries showed that glycine was the most common residue (5 times as an N-terminal linker and once as a C-terminal linker). Most amino acids that comprised the linker were small and/or polar residues, but hydrophobic residues such as valine, isoleucine and leucine were also observed. It must be noted that the length (by one residue) and sequence of the linker can vary in order to accommodate the frame compensation versions of the *cybC* insert (ORFs II and III in Figure 2B). Variants derived from cybC-GGS-2 or 3 may also have the terminal arginine residue of cyt b mutated to serine (e.g. Tc-11) or asparagine (e.g. Tc-06), respectively, if no linker sequence is present. Two variants derived from the cybC-GGS insert libraries, Tc-09 and Tc-11, did have an anomalous N-terminal cyt b linker sequence of just Gly-Ser as the first three nucleotides of the 5' *cybC* linker (GGC) were absent (Supplementary Table 3). We do not know how this occurred but one explanation is that truncation arose during the oligonucleotide synthesis.

Influence of cyt b insertion on TEM-1 activity

Insertion of cyt b would normally be considered a major and sometimes detrimental influence on TEM-1 structure

and hence activity. This in turn will influence the MIC of Amp that prevents *E. coli* growth. Furthermore, if the integral fusion protein displays switching properties, the presence of haem in the growth media will also affect the degree of bacterial resistance to Amp. To estimate activity of the various chimeric proteins *in vivo*, the Amp MIC required to inhibit *E. coli* growth both in the absence and presence of haem was measured. Cells containing the original starting vector that contains wild-type *bla* (pNOM) or a vector expressing separate wild-type *bla* and *cybC* genes (pPB10 (30)) or no vector were used as controls.

Many of the clones exhibited resistance to Amp that was only slightly compromised relative to cells producing wild-type TEM-1 via pNOM-borne *bla* (Table 1). Half of the 32 analysed clones conferred an Amp MIC value on *E. coli* of 16 000 µg/ml or greater. Variants that contained cyt b insertions at similar positions in most cases conferred a similar Amp MIC. Differences in linkers or the presence of tandem cyt b inserts could account for the slight variation in Amp MIC values (e.g. compare variants with insertions that replace Asp115 of TEM-1). Insertions within the TEM-1 signal sequence that replace Arg7 did not totally abolish protein export but must have affected the level of active enzyme in the periplasm as these variants conferred an Amp MIC of ≤8000 µg/ml (Table 1). Several variants with cyt b insertions positioned close to the active site still retained significant β-lactamase activity, although this was generally much lower than that of cells expressing wild-type TEM-1. For example, while TEM-1 tolerated an insertion between residues Ile173-Asn175 (Tc-09) and Asn170-Ala172 (Tc-15) in the catalytically important Ω-loop, the activity of these variants was considerably compromised compared to wild-type TEM-1 (Table 1).

Eleven of the 32 analysed integral fusion variants exhibited significant haem-dependent β-lactamase activity (>2-fold change in MIC value in the presence of 10 µM haem). The change in Amp MIC values varied from a 4-fold increase to a 128-fold decrease when haem was present in the culture medium (Table 1). Variants that displayed negative modulation of TEM-1 activity were predominant. The only variant conferring positive modulation (Tc-06) had a tandem cyt b insertion that replaced Asp115 in TEM-1 (Table 1). The largest haem-dependent decreases in TEM-1 activity were observed for variants with cyt b insertions between Glu212 and Glu215 (Figure 4). Three of the four clones containing insertions in this region conferred a 32- to 128-fold change in Amp MIC, depending on the linker (Table 1). The length of linker sequence appeared to be important in defining switching magnitude of the variant. All of the variants that displayed haem-dependent β-lactamase activity save Tc-01, had no or short linkers at both connection points between TEM-1 and cyt b (Table 1). Neither cells producing wild-type TEM-1 (via expression of pNOM-borne *bla* gene) or both TEM-1 and cyt b (via expression of pPB10-borne genes) exhibited any haem-dependent changes in growth, with observed Amp MIC values being 41 000 µg/ml in both the presence and absence of haem (Table 1).

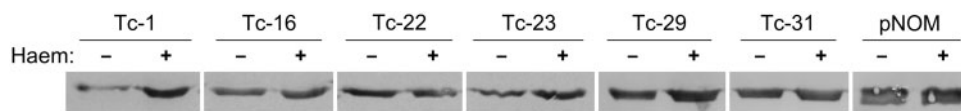


Figure 5. Western blot analysis of the expression level of selected cyt b-TEM-1 integral fusion variants. Cultures from which the periplasmic fraction was isolated were grown in minimal media in the absence (–) or presence (+) of 10 μ M haem and the blots were probed with an antibody against TEM-1.

Integral fusion protein levels in the periplasm

Western blotting was used to assess the level of cyt b-TEM-1 chimeras exported to the periplasm for several clones exhibiting a switching phenotype, to confirm that any change in Amp MIC values was not due to a correlated change in the level of the full-length integral fusion protein. Each of the fusion proteins was detected as a single band on the Western blot, migrating with an apparent molecular weight of *ca* 42–43 kDa, close to that predicted. For three of the fusion proteins (Tc-22, Tc-29 and Tc-31) the intensity of the bands was similar for cells grown in the presence and absence of haem, suggesting that the level of the fusion protein in the periplasmic fraction was unaffected by the presence of haem (Figure 5). Interestingly, for the other three fusion proteins analysed (Tc-1, Tc-16 and Tc-23), the intensity of the band from cells grown in the presence of haem appeared to be more intense than that from cells grown in the absence of haem (Figure 5), with the samples for Tc-1 providing an extreme example of this. Given that each of these fusion proteins exhibited a lower β -lactamase activity in the presence of haem, the switching phenotypes observed are unlikely be attributed to the different levels of fusion protein, but may be of even greater magnitude than the Amp MIC values suggest. Wild-type TEM-1 β -lactamase migrated as a broad band with an apparent molecular weight of *ca* 29 kDa as expected for the parent protein. The intensity of this band was the same in the periplasmic fractions of cells grown in the presence and absence of haem, showing that the level of wild-type β -lactamase was unaffected by haem.

DISCUSSION

While the insertion of a domain could be considered as a destructive modification to the accepting protein due to disruption of its original polypeptide chain continuity, it also provides a means for constructing novel proteins that integrate and therefore couple normally unrelated functions (7–9). However, the challenge lies not only in identifying suitable positions within the parent domain that permit a domain insertion, but also in finding those that allow communication between the functional centres of the individual domains. Here we have shown that our transposon-based method provides a useful alternative to DNaseI-based methods for generating libraries of novel active integral fusion proteins by the replacement of an amino acid in the TEM-1 parent domain with the cyt b insert. Reading frame compensation and the ability to define the nature of domain linkers expanded the degree of useful diversity generated and sampled, leading to the

discovery of integral fusion proteins displaying apparent switching properties.

TEM-1 was found to tolerate the insertion of cyt b at a variety of positions spread throughout the protein (Figures 3 and 4), although these were concentrated to an extent within the C-terminal half of the protein. Although it has been shown previously that TEM-1 can accept insertions of small peptide sequences at a variety of positions (31–33), it was evident from the work presented here that the positions and regions tolerant to domain insertion are limited. For example, 12 of the 32 variants contained an insertion in loops that linked either helices H3 and H4 or H9 and H10, and 7 variants had an insert that replaced Arg7 within the signal sequence. Our previous analysis of single amino acid deletion variants of TEM-1 generated from the same library of MuDel insertions within *bla* (16) confirms the diversity of *cybC* insertion positions available is much more extensive than observed here after selection for retention of β -lactamase activity. This confirms that TEM-1's tolerance to cyt b insertion is restricted to a few positions. As might be expected, most of the insertion positions lay within loop regions (Table 1 and Figure 4). The short loop comprising residues 112–118 was particularly tolerant to insertion, including two tandem cyt b units. Several insertions were observed in organized secondary structure (Table 1 and Figure 4) but were close to the end of the helix or strand. Insertion within the TEM-1 signal sequence that replaced Arg7 did not appear to prevent protein export to the periplasm. However, variants containing this insertion conferred lower Amp MIC values on *E. coli* suggesting export was less efficient.

The very nature of the structural link between the two domains in an integral fusion protein would suggest that in some instances their individual functions would be linked. Haem regulation of TEM-1 was observed for several of the variants, with their *in vivo* activity towards Amp varying depending on the presence of haem in the growth medium (Table 1). Both positive and negative regulation of β -lactamase activity by haem was observed. The observed switching phenotypes displayed a wide range of magnitudes, which depended on both the position of cyt b insertion within TEM-1 and the linker sequence connecting the domains. For some variants, analysis of the insertion point within the context of TEM-1 tertiary structure provided a rationale for the switching properties. Cyt b insertions that replaced Asp214 or residues in the Thr266 to Glu274 region conferred a switching phenotype. These residues lie close to the active site pocket and are situated at the sub-domain interface of TEM-1 (Figure 4). However, the explanation why other variants displayed switching

properties was less clear. For example, Gly196 is relatively distant from the active site (Figure 4) yet replacement of this residue by cyt b causes an 8-fold drop in bacterial resistance to Amp in the presence of haem (Table 1). Mutations in this region have been observed previously to influence TEM-1 activity (34). Western blot analysis of periplasmic levels of the chimeras suggested that switching was not due to a drop in protein concentration in the periplasm on addition of haem (Figure 5). On the contrary, in some cases the concentration of chimeric protein was higher in the presence of haem suggesting that the magnitude of switching may be larger than the Amp MIC values indicate. Although it is unlikely that differential haem-dependent leakage of protein from the periplasm to the cell medium or haem metabolism during cell growth are influencing the level of observed integral fusion protein activity *in vivo*, they cannot be conclusively ruled out as contributory factors of the switching phenotype.

The linker sequence that joins TEM-1 and cyt b appeared to be critical in defining the switching magnitude of the integral fusion protein. For example, the switching magnitude of variants with cyt b replacing Asp214 varied between 32- and 128-fold depending on the nature of the linker sequence. It was apparent that shorter linkers promoted switching properties. Only one variant (Tc-01) with the extended tripeptide linker at both domain connection points exhibited switching behaviour. All the other switching variants contained short or no linkers.

As the majority of variants conferring a switching phenotype displayed negative modulation of TEM-1 activity in the presence of haem, this would suggest that in the present context haem-activated variants would be difficult to construct. One variant, Tc-06, did exhibit a 4-fold positive modulation of β -lactamase activity in the presence of haem (Table 1). This variant is unusual as it contains two tandem cyt b insertions that replace Asp115, with a flexible GGPGGS linker joining the cyt b domains. The lack of switching behaviour for Tc-13 that also contains a tandem insertion of two cyt b but replacing Gly116 (Table 1) could be due to the slight difference in insertion position or the lack of linker that joins the two cyt b units. Tc-13 was also less active *in vivo* than Tc-06 suggesting that this double cyt b insertion was more disruptive to TEM-1.

A number of other reports have identified integral fusion proteins that exhibit switching behaviour in response to ligand binding (10,12–14,35). Most had activities that differed by up to 2- to 8-fold in the presence of their respective ligands. The most successful example thus far was produced by a combinatorial approach using a circular permuted version of TEM-1 as the reporter domain, and a 600-fold difference in activity was reported on analysis *in vitro* (23). This same variant exhibited only a 16-fold change in Amp MIC in the presence and absence of ligand in an *in vivo* assay. Using an *in vivo* assay based on the same principles, we identified cyt b-TEM-1 variants from a relatively small starting library size with the magnitude of the haem-dependent switching phenotype ranging between 32- and 128-fold. A more complete analysis of these particular variants *in vitro* will reveal the true extent of their switching magnitude independent of any potential *in vivo* effects, the influence of redox conditions

on switching and any structural changes on haem binding responsible for regulating β -lactamase activity. This work is currently underway.

In conclusion, we have demonstrated that our transposon-based directed evolution domain insertion method can generate a diverse range of novel integral fusion proteins. Our method combines the ability of rational protein engineering to replace one amino acid with a new insert domain having predefined linkers, with the power of directed evolution to sample many different insertion positions. This method can easily be scaled up to sample many more different insertion positions in TEM-1, and can also be expanded to construct integral fusion proteins that link together a variety of different and normally unrelated proteins to construct novel protein switches.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by the Biotechnology and Biological Sciences Research Council (BB/E007384 and BB/E001084) and the Welsh Assembly Government (HE07POC3007). We would like to thank Dr Paul Barker, University of Cambridge, for supplying pPB10. Funding to pay the Open Access publication charges for this article were waived.

Conflict of interest statement. None declared.

REFERENCES

- Bornberg-Bauer, E., Beaussart, F., Kummerfeld, S.K., Teichmann, S.A. and Weiner, J. III. (2005) The evolution of domain arrangements in proteins and interaction networks. *Cell. Mol. Life Sci.*, **62**, 435–445.
- Aroul-Selvam, R., Hubbard, T. and Sasidharan, R. (2004) Domain insertions in protein structures. *J. Mol. Biol.*, **338**, 633–641.
- Sigler, P.B., Xu, Z., Rye, H.S., Burston, S.G., Fenton, W.A. and Horwich, A.L. (1998) Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.*, **67**, 581–608.
- Ambroggio, X.I. and Kuhlman, B. (2006) Design of protein conformational switches. *Curr. Opin. Struct. Biol.*, **16**, 525–530.
- Buskirk, A.R. and Liu, D.R. (2005) Creating small-molecule-dependent switches to modulate biological functions. *Chem. Biol.*, **12**, 151–161.
- Lim, W.A. (2002) The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.*, **12**, 61–68.
- Villaverde, A. (2003) Allosteric enzymes as biosensors for molecular diagnosis. *FEBS Lett.*, **554**, 169–172.
- Ferraz, R.M., Vera, A., Aris, A. and Villaverde, A. (2006) Insertional protein engineering for analytical molecular sensing. *Microbial Cell Factories*, **5**, 15.
- Ostermeier, M. (2005) Engineering allosteric protein switches by domain insertion. *Protein Eng. Des. Sel.*, **18**, 359–364.
- Baird, G.S., Zacharias, D.A. and Tsien, R.Y. (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl Acad. Sci. USA*, **96**, 11241–11246.
- Betton, J.M., Jacob, J.P., Hofnung, M. and Broome-Smith, J.K. (1997) Creating a bifunctional protein by insertion of β -lactamase into the maltodextrin-binding protein. *Nat. Biotechnol.*, **15**, 1276–1279.

12. Collinet, B., Herve, M., Pecorari, F., Minard, P., Eder, O. and Desmadril, M. (2000) Functionally accepted insertions of proteins within protein domains. *J. Biol. Chem.*, **275**, 17428–17433.
13. Biondi, R.M., Baehler, P.J., Reymond, C.D. and Veron, M. (1998) Random insertion of GFP into the cAMP-dependent protein kinase regulatory subunit from *Dictyostelium discoideum*. *Nucleic Acids Res.*, **26**, 4946–4952.
14. Guntas, G., Mitchell, S.F. and Ostermeier, M. (2004) A molecular switch created by in vitro recombination of nonhomologous genes. *Chem. Biol.*, **11**, 1483–1487.
15. Jones, D.D. (2005) Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 β -lactamase to an amino acid deletion. *Nucleic Acids Res.*, **33**, e80.
16. Simm, A.M., Baldwin, A.J., Busse, K. and Jones, D.D. (2007) Investigating protein structural plasticity by surveying the consequence of an amino acid deletion from TEM-1 β -lactamase. *FEBS Lett.*, **581**, 3904–3908.
17. Arnesano, F., Banci, L., Bertini, I., Faraone-Mennella, J., Rosato, A., Barker, P.D. and Fersht, A.R. (1999) The solution structure of oxidized *Escherichia coli* cytochrome b_{562} . *Biochemistry*, **38**, 8657–8670.
18. Feng, Y., Sligar, S.G. and Wand, A.J. (1994) Solution structure of apo-cytochrome b_{562} . *Nat. Struct. Biol.*, **1**, 30–35.
19. Jones, D.D. and Barker, P.D. (2004) Design and characterisation of an artificial DNA-binding cytochrome. *Chembiochem*, **5**, 964–971.
20. Jones, D.D. and Barker, P.D. (2005) Controlling self-assembly by linking protein folding, DNA binding, and the redox chemistry of Heme. *Angew. Chem. Int. Ed. Engl.*, **44**, 6337–6341.
21. Matagne, A., Lamotte-Brasseur, J. and Frere, J.M. (1998) Catalytic properties of class A β -lactamases: efficiency and diversity. *Biochem. J.*, **330**, 581–598.
22. Guntas, G. and Ostermeier, M. (2004) Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol.*, **336**, 263–273.
23. Guntas, G., Mansell, T.J., Kim, J.R. and Ostermeier, M. (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl Acad. Sci. USA*, **102**, 11224–11229.
24. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. Changeux, J.P. and Edelstein, S.J. (2005) Allosteric mechanisms of signal transduction. *Science*, **308**, 1424–1428.
26. Krishna, M.M. and Englander, S.W. (2005) The N-terminal to C-terminal motif in protein folding and function. *Proc. Natl Acad. Sci. USA*, **102**, 1053–1058.
27. D'Amelio, N., Bonvin, A.M., Czisch, M., Barker, P. and Kaptein, R. (2002) The C terminus of apo-cytochrome b_{562} undergoes fast motions and slow exchange among ordered conformations resembling the folded state. *Biochemistry*, **41**, 5505–5514.
28. 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.
29. Galarneau, A., Primeau, M., Trudeau, L.E. and Michnick, S.W. (2002) β -lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein protein interactions. *Nat. Biotechnol.*, **20**, 619–622.
30. Barker, P.D., Nerou, E.P., Freund, S.M. and Fearnley, I.M. (1995) Conversion of cytochrome b_{562} to c-type cytochromes. *Biochemistry*, **34**, 15191–15203.
31. Hallet, B., Sherratt, D.J. and Hayes, F. (1997) Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. *Nucleic Acids Res.*, **25**, 1866–1867.
32. Hayes, F., Hallet, B. and Cao, Y. (1997) Insertion mutagenesis as a tool in the modification of protein function. Extended substrate specificity conferred by pentapeptide insertions in the omega-loop of TEM-1 β -lactamase. *J. Biol. Chem.*, **272**, 28833–28836.
33. Legendre, D., Soumillon, P. and Fastrez, J. (1999) Engineering a regulatable enzyme for homogeneous immunoassays. *Nat. Biotechnol.*, **17**, 67–72.
34. Baldwin, A.J., Busse, K., Simm, A.J. and Jones, D. (2008) Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). *Nucleic Acids Res.*, doi:10.1093/nar/GKN358.
35. Doi, N. and Yanagawa, H. (1999) Insertional gene fusion technology. *FEBS Lett.*, **457**, 1–4.
36. Jelsch, C., Mourey, L., Masson, J.M. and Samama, J.P. (1993) Crystal structure of *Escherichia coli* TEM-1 β -lactamase at 1.8 Å resolution. *Proteins*, **16**, 364–383.
37. Ambler, R.P., Coulson, A.F., Frere, J.M., Ghuysen, J.M., Joris, B., Forsman, M., Levesque, R.C., Tiraby, G. and Waley, S.G. (1991) A standard numbering scheme for the class A β -lactamases. *Biochem. J.*, **276**, 269–270.