Published in final edited form as:
Methods Enzymol. 2013 ; 523: 369–388. doi:10.1016/B978-0-12-394292-0.00017-5.

Protein Switch Engineering by Domain Insertion

Manu Kanwar, **R. Clay Wright**, **Amol Date**, **Jennifer Tullman**, and **Marc Ostermeier**[1]
Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, USA

Abstract

The switch-like regulation of protein activity by molecular signals is abundant in native proteins. The ability to engineer proteins with novel regulation has applications in bio-sensors, selective protein therapeutics, and basic research. One approach to building proteins with novel switch properties is creating combinatorial libraries of gene fusions between genes encoding proteins that have the prerequisite input and output functions of the desired switch. These libraries are then subjected to selections and/or screens to identify those rare gene fusions that encode functional switches. Combinatorial libraries in which an insert gene is inserted randomly into an acceptor gene have been useful for creating switches, particularly when combined with circular permutation of the insert gene. Methods for creating random domain insertion libraries are described. Three methods for creating a diverse set of insertion sites in the acceptor gene are presented and compared: DNase I digestion, S1 nuclease digestion, and multiplex inverse PCR. A PCR-based method for creating a library of circular permutations of the insert gene is also presented.

1. INTRODUCTION

The ability of proteins to be regulated by molecular signals is a hall-mark of biological systems. Such proteins function as switches in which recognition of an input signal (e.g., small molecule or protein) regulates an output function (e.g., enzyme activity or DNA affinity). The ability to create novel protein switches would have applications in sensing and therapeutics. One approach to building new switches is to reengineer existing natural switches. A more challenging approach, but perhaps more powerful in the long term, is to build new switches from existing protein domains with the prerequisite input and output functions of the desired switches. Domain insertion, the insertion of one protein domain into another, has been used to couple the functions of two proteins to create protein switches (Ostermeier, 2005). The design challenge lies in the fusion of the two protein domains such that communication is established between the two functions.

One approach is using evolution as the design algorithm. The first step in this algorithm is creating protein diversity, from which selections or screens can be used to identify the rare variants with switching properties. One approach to creating this diversity is creating combinatorial libraries in which an insert gene is randomly inserted into the acceptor gene. The insert gene can encode either the output function (Guntas, Mansell, Kim, & Ostermeier,

[1]Corresponding author: oster@jhu.edu.

2005; Guntas, Mitchell, & Ostermeier, 2004; Guntas & Ostermeier, 2004; Tullman, Guntas, Dumont, & Ostermeier, 2011) or the input function (Edwards, Busse, Allemann, & Jones, 2008; Wright, Wright, Eshleman, & Ostermeier, 2011) of the switch.

For example, insertion of β-lactamase (BLA) into sugar-binding proteins can create switch proteins in which BLA activity is modulated by the respective sugar (Guntas et al., 2005, 2004; Guntas & Ostermeier, 2004; Tullman et al., 2011). We have extensively created and characterized switches comprising fusions of BLA and maltose-binding protein (MBP). These studies have shown that switches can be created that provide >100-fold change in activity upon effector binding (Guntas et al., 2005) and that switching is reversible (Guntas et al., 2004). The switches are modular, allowing mutations that increase effector affinity (Kim & Ostermeier, 2006) or provide the switch with the ability to be activated by new effectors (Guntas et al., 2005). The MBP–BLA switch genes function in *E. coli* cells and confer a new phenotype: maltose-dependent resistance to β-lactam antibiotics (Guntas et al., 2005, 2004). Many of the MBP–BLA switches function as heterotropic allosteric enzymes, and NMR and crystallographic studies of one switch are consistent with the expectation that the individual domain structures of RG13 are substantially conserved from MBP and BLA (Ke et al., 2012; Wright, Majumdar, Tolman, & Ostermeier, 2010). More recently, we have identified MBP–BLA switch genes that do not encode allosteric proteins but rather encode a protein whose cellular accumulation increases in the presence of the effector, thereby conferring an effector-dependent switching phenotype to cells (Heins, Choi, Sohka, & Ostermeier, 2011; Sohka et al., 2009). Finally, our MBP–BLA switch work has shown how domain fusion can result in emergent properties. One of our MBP–BLA switches is negatively, allosterically regulated by Zn^{2+}, which is unexpected since neither MBP nor BLA has significant affinity for Zn^{2+} (Ke et al., 2012; Liang, Kim, Boock, Mansell, & Ostermeier, 2007).

Recently, we have also used domain insertion to create prodrug-activating enzymes that are turned on in the presence of a protein cancer marker and render cancer cells selectively sensitive to the prodrug in a cancer marker-dependent fashion (Wright et al., 2011). From random domain insertion libraries comprising the CH1 domain of *p*300 (which binds the cancer marker HIF-1α) inserted into yeast cytosine deaminase, which can activate the prodrug 5-fluorocytosine, we selected two switch genes that conferred on 5-fluorocytosine an increased sensitivity to *E. coli* cells and human colon and breast cancer cell lines in a HIF-1α-dependent manner. Evidence suggested that the better performing switch functioned primarily through a mechanism in which HIF-1α caused the switch protein to accumulate to higher levels in the cell. This strategy offers a platform for the development of inherently selective protein therapeutics for cancer and other diseases, although significant delivery challenges will need to be overcome for success to be achieved.

This chapter covers the methodology for creating random domain insertion libraries using nonspecific endonucleases or PCR-based methods (Fig. 17.1) but does not cover random insertions using transposons (Edwards et al., 2008). Although the methods are presented within the context of switch construction, they are applicable to any study in which random insertion of one DNA segment into another is useful (e.g., structure–function studies, studies of promoter or chromosomal structure, and studies of oligomeric or repeat proteins).

2. CREATION OF RANDOM DOUBLE-STRANDED BREAKS IN PLASMIDS CONTAINING THE ACCEPTOR DNA

The acceptor gene is cloned into an expression vector of choice. This plasmid is then targeted for the creation of a random double-stranded break by endonuclease digestion with DNase I (Section 2.1) or S1 nuclease (Section 2.2), or for the amplification of the desired linear double-stranded plasmid DNA with "breaks" at select locations by inverse PCR (Ochman, Gerber, & Hartl, 1988) in a multiplex fashion. The three methods (DNase I digestion, S1 nuclease digestion, and multiplex inverse PCR) have different strengths and weaknesses (Table 17.1). Because the DNase I and S1 nuclease methods create a lot of "junk" library members (e.g., insertions outside the acceptor gene, insertions out of frame or backward), these methods should be used only with a powerful selection or screen that can handle large libraries. Targeting the insertion using multiplex inverse PCR makes for a highly focused library enriched in the types of fusions that might be switches. The drawbacks are the cost of the sets of primers and the trade-off between increased diversity and the large number of PCR reactions to perform. Methods that create short deletions or tandem duplications at the insertion site can be beneficial, as this added diversity will adjust the distance and geometric orientation between the two fused domains (i.e., tandem duplications can serve as pseudo-natural "linkers" between the two domains).

2.1. Creation of random breaks using DNase I

1. Prepare approximately 100 μg of plasmid DNA containing the gene encoding for the acceptor protein using the Qiagen DNA Miniprep Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. We have found that the Miniprep kits have better yield with less genomic DNA contamination than the Midiprep or Maxiprep kits.
2. Prepare 50 μL of working stock solution of DNase I (1 U/μL) (New England Biolabs, Ipswich, MA) in 25 m*M* Tris–HCl pH 7.5 containing 50% glycerol and store it at −20 °C.
3. Prepare 1.2 mL of diluent solution in 50 m*M* Tris–HCl pH 7.5 containing 1 m*M* $MnCl_2$ and 6 μL of 100 × BSA (New England Biolabs, Ipswich, MA). The diluent solution must be freshly prepared and kept at room temperature.
4. Determine the DNase I concentration that will result in the highest yield of single cut double-stranded DNA for a 5 μg plasmid digestion for 8 min at room temperature in a total reaction volume of 100 μL. In brief, 5 μg of plasmid DNA in 95 μL of diluent solution is incubated with varying concentrations of DNase I (5 μL) for 8 min at room temperature. The reaction is stopped by the addition of 1.2 μL of 1.0 *M* EDTA and the mixture incubated at 75 °C for 10 min to heat inactivate the DNase I. The reaction product is visualized on a 0.8% agarose, lithium borate (LB) (Faster Better Media, Hunt Valley, MD) gel to determine the optimal DNase I dilution. The optimal DNase I concentration is the one that produces a sharp band for linearized plasmid (no smearing). Achievement of this outcome usually requires

that 30–50% of the plasmid remain supercoiled while 30% of the supercoiled plasmid DNA is converted to the desired linear DNA product.

5. Once the optimal DNase I dilution is determined, prepare 1140 μL of plasmid DNA solution (60 μg plasmid DNA in 1140 μL of diluent solution) and dispense in 12 microtubes each containing 95 μL (i.e., 5 μg DNA) of plasmid DNA solution. Incubate for 10 min at room temperature.

6. Add 5 μL of the optimal DNase I dilution (determined earlier) to the first microtube containing 95 μL plasmid DNA solution, and flick to mix. Add 5 μL of the optimal DNase I dilution to the next tube after 30 s and repeat the step until DNase I has been added to each of the 12 microtubes. Add 1.2 μL of 1.0 *M* EDTA to each tube after it has incubated for 8 min. Incubate all 12 tubes at 75 °C for 10 min to heat inactivate the DNase I.

7. Combine the contents of all 12 tubes and purify the DNA using 25 μg DNA clean and concentrator (Zymo Research Corp., Irvine, CA). Use at least three spin columns. Elute the DNA in 100 μL of 0.1× elution buffer prewarmed to 50 °C per column and combine the eluate from each spin column. The prepared DNA can be stored at −20 °C.

2.2. Creation of random double-stranded breaks using S1 nuclease

Unlike DNase I digestion, S1 nuclease digestion of supercoiled DNA essentially halts after the creation of linear dsDNA with blunt ends; hence, libraries are easier to construct. S1 nuclease is typically known as a single strand-specific-nuclease. However, S1 nuclease has been shown to digest supercoiled plasmids especially in regions containing inverted repeats or forming cruciform structures (Lilley, 1980; Panayotatos & Wells, 1981). Although inverted repeats are common in the origins of replication of plasmids, plasmids digested at these sites are generally excluded from the combinatorial library upon transformation into *E. coli* due to the inability of the plasmid to be replicated when the origin is disrupted by an inserted sequence. We established that S1 nuclease also creates dsDNA breaks elsewhere in the plasmid as well, and that these breaks appear fairly random, although we have not extensively studied the distribution of cut sites (Tullman et al., 2011). If, however, inverted repeat sequences occur in the gene sequence of the acceptor protein domain, one should choose another method to create the random double-stranded breaks. We have used a plasmid containing the p15a origin of replication to successfully create ribose-activated BLAs using S1 nuclease digestions to create sites for insertion into the acceptor gene encoding ribose-binding proteins (Tullman et al., 2011).

1. Prepare 50 μg of plasmid containing the acceptor gene sequence via standard procedures (e.g., Qiagen Qiaquick Miniprep Kit, Qiagen, Valencia, CA). It is important that the DNA be of high quality (RNA and genomic DNA-free). We have found that the Miniprep kits have better yield with less genomic DNA contamination than the Midiprep or Maxiprep kits.

2. Aliquot 2 μg of plasmid into each of 12 tubes and incubate in 50 m*M* sodium acetate, 280 m*M* NaCl, and 4.5 m*M* $ZnSO_4$ at pH 4.5 (at 25 °C) with 10 units of S1 nuclease (Promega, Madison, WI) in a final volume of 25 μL for 20 min at 37 °C.

3. Stop the reaction by pooling the 12 tubes and adding the DNA-binding buffer from the DNA clean and concentrator 25 μg kit (Zymo Research Corp.). Continue the purification of the DNA using the binding column and wash buffer as per the manufacturer's instructions and elute in 50 μL of nuclease-free DI water prewarmed to 50 °C. The prepared DNA can be stored at −20 °C.

2.3. Creation of random "breaks" using multiplex inverse PCR

For multiplex inverse PCR (Fig. 17.2A), divergent abutting primers that amplify the entire plasmid need to be designed for every desired insertion site (e.g., at every codon in the acceptor gene or a subset of these codons). These primers are used to "open up" the plasmid at desired locations for insertion. Mixing and matching of nearby primers can be used to open up the plasmid such that deletions or tandem duplications of portions of the acceptor gene occur (Fig. 17.2B). For example, a pair of divergent primers that exclude the codons between them can be used to create deletions. Using primers that have three overlapping bases (one codon overlap) will create a one amino acid repeat on either side of the insert gene in the final library, primers with six base overlap (two codon overlap) will create two amino acid repeat and so on. However, the number of amino acid tandem duplications that can be obtained is limited. The yield of the correct product for PCR reactions using primers corresponding to more than two codons of overlap can be significantly reduced due to the primer dimerization.

This method not only defines the location of insertion of target DNA, but also ensures a library with all in-frame fusions (although the insert gene will still be inserted backward 50% of the time). Additionally, the insertion of DNA is targeted only in the gene of interest, and therefore, unproductive insertions in the backbone of the plasmid that do not produce combinatorial fusions of the target genes can be eliminated. The highly focused library that is produced is ideal for creating protein switches for which selection or screening methods are limited and cumbersome. However, there are some limitations and disadvantages of the PCR-based method that need to be considered while selecting the best method for construction of a DNA library. Two oligos are required for every insertion location, and a separate PCR reaction needs to be performed for every desired insertion site. The design of all the primers and execution of all the reactions is time consuming and may require optimization.

2.3.1 Multiplex inverse PCR—PCR reactions can be executed in a 96-well PCR plate using a 96-well plate thermocycler. For the PCR reactions to be performed effectively in parallel, it is crucial that all the primers have similar melting temperatures.

1. Prepare a master mix that consists of polymerase buffer, dNTPs, polymerase, DNA template, and water for the required reactions. For 100 PCR reactions each of 50 μL total volume, the following master mix can be prepared:

   10× olymerase buffer—1000 μL

   10 m*M* dNTP—150 μL

   DNA template—1000 ng (10 ng each reaction)

Phusion HF polymerase—100 units (New England Biolabs, Ipswitch, MA)

Add water to a final volume of 5 mL.

2. Dispense 50 μL of the master mix into each well of the 96-well PCR plate. Add 1 μL of 10 μ*M* forward and reverse primers to each well (final concentration 0.2 μ*M*).
3. Perform the thermocycling reaction. A typical cycling protocol for a 4 kb plasmid with primers designed to have approximately a 60 °C melting temperature is as follows:

   Initial incubation at 98 °C for 3 min

   30 cycles of

   98°C for 60 s

   60°C for 60 s

   72°C for 30 s/kb

   After 30 cycles, incubate at 72 °C for 5 min and hold at 4 °C.

   Note that longer extension times are required than in typical PCR reactions since the whole plasmid is amplified. Extension times should be at least 30 s per kb of PCR product.
4. Electrophorese the PCR reactions on an agarose gel to confirm successful amplification of the desired product. We typically use 0.8% agarose, LB (Faster Better Media) gel and load 5 μL of DNA or less per lane. Perform electrophoresis at 275 V for about 10 min (adjust depending on the size of the plasmid).
5. For some reactions, optimization of the amount of template DNA, the extension time, the annealing temperature and/or the number of cycles may be necessary to obtain amplification of the desired product. In general, try a change in annealing temperature first, followed by a change in the extension time, and finally the amount of template DNA. The addition of DMSO and/or betaine to a final concentration of 1–5% and 1 *M*, respectively, may help decrease nonspecific products.

2.3.2 Purification of PCR products

1. Combine all successful PCR reactions.
2. Prepare LB or TAE deep-well 0.8% agarose gel and load up to 10 μg of DNA per lane in the deep wells of the gel.
3. Electrophorese at 95 V for about 50 min to ensure good separation between linearized PCR product and supercoiled plasmid. Isolate the desired DNA from the gel using Qiaquick gel extraction kit (Qiagen) according to the manufacturer's instructions.
4. Determine the concentration of the recovered DNA using UV/Vis spectrophotometry.

3. REPAIR, PURIFICATION AND DEPHOSPHORYLATION OF ACCEPTOR DNA

This step is necessary for DNase I- and S1 nuclease-digested DNA only. For acceptor DNA prepared by multiplex inverse PCR, skip to Section 4.

3.1. Evaluate quality and determine quantity of linear DNA

1. Determine the total DNA concentration of the DNase I or S1 nuclease digested DNA using a UV/Vis spectrophotometer.
2. Analyze 200–500 ng of the recovered DNA via agarose gel electrophoresis on a 0.8% agarose, LB gel. The use of LB rather than TAE or TBE facilitates the separation of the linear and nicked circle DNA. On an LB gel, linear runs the fastest, followed by nicked circles, and then supercoiled.
3. Quantify the fraction of linear DNA using gel-imaging software and determine the concentration of linear plasmid DNA. This is a particularly important step for linear DNA prepared by DNase I digestion. With DNase I digested DNA, we have had good success with digestions in which ~30% of the plasmid has been converted to linear DNA. If the sample is over digested, the resulting combinatorial library will be dominated by members in which large segments of the acceptor plasmid are deleted. In the case in which plasmid is over digested, repeat the supercoiled plasmid DNA I digestion using less DNase I.

3.2. Repair the DNA (blunt the ends and seal any nicks that might be present in the DNA)

1. Repair the digested DNA (to create blunt ends) using T4 DNA ligase and T4 DNA polymerase (New England Biolabs, Ipswich, MA). The amount of ligase and polymerase employed is calculated based on the amount of linear DNA determined earlier. Incubate multiple tubes each containing not more than 75 μL of total reaction mixture containing DNA, T4 DNA ligase (160 cohesive end units/μg linear DNA), T4 DNA polymerase (1 unit/μg linear DNA), 200 μ*M* dNTPs, 1× T4 ligase buffer, and 1× BSA in water at 12 °C for 20 min.
2. Add EDTA to 10 m*M* and incubate at 75 °C for 15 min to stop the reaction.
3. Allow the solution to cool to room temperature. Purify and isolate the repaired DNA using DNA clean and concentrator 25 μg kit (Zymo Research Corp.) as described in Section 2.2 in step 3.

3.3. Isolate the repaired, linear DNA

1. Make a 0.8% agarose, LB (Faster Better Media) gel. The use of LB rather than TAE or TBE buffer facilitates the separation of the linear and nicked circle DNA.
2. Load 1.5 μg of DNA or less per lane. If the DNA has been frozen, heat to 65 °C for approximately 2 min before loading the gel to ensure better separation on the gel.
3. Run the gel at 275 V for about 30 min (adjust depending on the size of the plasmid), until the DNA is separated into three distinct bands. The slowest

migrating band is supercoiled DNA, the fastest is linear DNA, and the band in the middle is nicked circles.

4. Extract the band containing the linear DNA and purify using the Qiagen gel extraction kit (Qiagen) per the manufacturer's instructions.

3.4. Dephosphorylate the acceptor DNA

1. For every microgram of DNA, add 20 units of Antarctic phosphatase (New England Biolabs, Ipswitch, MA) in 1× Antarctic phosphatase buffer. Other phosphatases such as calf intestinal or shrimp alkaline can be used; however, Antarctic phosphatase is the only one of these that can be easily heat inactivated, and the buffer does not interfere with the subsequent ligation reaction. This property is advantageous as it avoids any DNA loss during a purification step at this stage.
2. Incubate for 1 h at 37 °C.
3. Heat to 65 °C for 10 min to inactivate the phosphatase. The dephosphorylated vector does not require purification and can be directly used in the ligation step.

4. PREPARATION OF INSERT DNA

We have successfully created protein switches with circularly permuted inserts (Guntas et al., 2005, 2004; Tullman et al., 2011) and non-circularly permuted inserts (Guntas & Ostermeier, 2004; Wright et al., 2011). The protocol for a noncircularly permuted insert is the easiest (Section 4.1), as it is just a simple PCR reaction. However, our switches with the largest differences between their on and off states contain circularly permuted insert proteins (Guntas et al., 2005). Although we have previously used protocols for random circular permutation of the insert gene with DNase I and S1 nuclease (Guntas et al., 2005, 2004; Wright et al., 2011), the protocols are very challenging. We have recently switched to a parallel PCR method to create the circularly permuted library of insert genes (Fig. 17.1B) and present this recommended protocol here (Section 4.2). This method has the additional advantage that circular permutations in the library can be researcher-defined. For example, the loci for circular permutation can be limited to regions that are solvent accessible, flexible, and not functionally relevant (Lo et al., 2012)—sites that are most likely to tolerate circular permutation. Successful domain insertion and circular permutation generally requires proximity between the N- and C-termini of the insert domain or longer linkers to span the added distance. Although there is no requirement for the protein to have known structure in order to construct the library, the choice of appropriate linker length is greatly informed by structural information. We have not attempted switch construction from proteins of unknown structure.

4.1. Creation of noncircularly permuted insert DNA

1 Design primers to anneal to the 5′ and 3′ ends of the insert DNA with approximately 15–24 base pairs and 55 °C melting temperatures. To the extent that the termini of the insert protein are not proximal, longer linkers will be

required. To add linker amino acids between the insert and acceptor proteins, add the corresponding nucleotides to the 5′ ends of the primers.

2 Phosphorylated primers can be ordered from most manufacturers or unphosphorylated primers can be phosphorylated using T4 kinase (New England Biolabs, Ipswich, MA) following the manufacturer's directions.

3 Perform PCR using Phusion HF (New England Biolabs, Ipswich, MA) or another high fidelity polymerase according to the manufacturer's directions using a DNA containing the desired insert gene as the template. A sample protocol is provided.

10 μL 5× HF Buffer (or GC buffer if high GC content)

2.5 μL each primer (0.5 μ*M* final concentration)

1 μL 10 m*M* dNTPs

0.5 μL Phusion HF polymerase

1 μL template DNA (10 ng)[1]

32.5 μL nuclease-free water.

Temperature cycling:

1. 98 °C for 30 s
2. 30 cycles of 98 °C for 5 s/55 °C for 15 s (or other appropriate annealing temperature)/72 °C for 15 s (15–30 s per kb)
3. 72 °C for 10 min
4. 4 °C hold.

4 Electrophorese the PCR product on an agarose gel and extract and purify the band that appears at the correct size using Qiaquick gel extraction kit (Qiagen) according to the manufacturer's instructions.

5 Determine the concentration of the recovered DNA using a UV/Vis spectrophotometer.

4.2. Creation of circularly permutation libraries as insert DNA

Circular permutation of the insert protein can drastically increase the diversity of domain insertion libraries thereby increasing the chance of isolating a hybrid enzyme with the desired properties. Below we describe in general the creation of circular permutation libraries by parallel PCR using primers corresponding to the beginning and end of the desired circular permutant.

4.2.1 Design criteria for circular permutation—The length and composition of the linker spanning the distance between the original N- and C-termini must be rationally

[1]It is beneficial to use template DNA with a different resistance marker than that of your linearized acceptor plasmid to avoid carrying the template plasmid forward into the library.

designed. However, one is not necessarily limited to a single linker choice as several linker variants may be included in the library design since the linkers are encoded by the PCR primers used to create the template construct.

1. Linker length—Determine the distance between the alpha carbons of the N- and C-terminal residues in the insert protein using Pymol (Schrödinger, 2010) or similar molecular visualization software. Depending on the secondary structure of the linker residues, one amino acid will span a distance of 1.5–3 Å.
2. Linker composition—In most cases, flexible linkers rich in glycine, serine, and threonine are used to span the original termini in a manner that does not constrain the termini location or distort the domain's structure. Linkers of more diverse composition or higher rigidity can also be used.
3. Choice of the set of new termini—In the case of small domains, every residue may be queried; however, larger domains may require some limitation to minimize both library size and primer costs. Residues that are solvent accessible, flexible, loosely packed, and between secondary structure elements are most favorable for successful circular permutation (Lo et al., 2012). However, a variety of studies have shown that sites for successful circular permutation can lie within secondary structure elements (Meister, Kanwar, & Ostermeier, 2009).

4.2.2 Creation of template constructs and primer pairs

1. Create template construct containing an end-to-end fusion of the desired insert domain spanned by the designed linker(s) using standard cloning techniques (Fig. 17.1B).
2. Mix each template construct (i.e., one could use a set of templates with different circular permutation linkers between the N- and C-termini) in equimolar quantities to be used as a template for each PCR reaction (Section 4.2.3).
3. For each codon or select codons within the insert gene, design a forward primer starting at the first base of the codon and a reverse primer starting at the last base of the previous codon and extending in the 3′ direction until the individual melting temperatures are between 60 and 65 °C, and other standard primer design criteria are optimized. We have written a MATLAB script to design such primer pairs, which is freely available.
4. For each primer pair, make a primer mix containing 10 μ*M* each (forward and reverse) primer. Duplications or deletions at the new termini can be made by shifting the reverse primer ahead or back one codon, respectively, analogous to how the multiplex inverse PCR can create similar diversity.

4.2.3 Parallel PCR reactions

1. Make the following template mixture and aliquot 9 μL to each well:

   Water—4 μL/reaction

   Betaine (5.5 *M* stock)—4 μL/reaction

DMSO—0.6 μL/reaction

Template (10 ng/μL stock)—0.4 μL/reaction.

2. To each well, add 1 μL of the 10 μ*M* primer pair mix.
3. Immediately before putting the PCR plate into a thermocycler preheated to 98 °C, add 10 μL of Phusion HF Master Mix (New England Biolabs, Ipswich, MA) to each well.
4. After an initial 30 s 98 °C incubation, perform the following cycle 30×

   98°C—10 s

   63°C—20 s

   72°C—15 s/kb.
5. Incubate at 72 °C for 5 min.
6. The success of each reaction can be verified by electrophoresing 5 μL in an agarose gel.

4.2.4 Purification and preparation of insert for cloning

1. Combine 12 μL from each reaction in a polypropylene tube and mix.
2. Purify this library mixture by phenol/chloroform extraction followed by ethanol precipitation (Sambrook & Russell, 2001).
3. Electrophorese the DNA on an agarose gel and isolate the desired DNA from the gel using Qiaquick gel extraction kit (Qiagen) according to the manufacturer's instructions.
4. Phosphorylate the PCR product library mixture using NEB's Quick Blunting Kit (E1201), per the manufacturer's instructions. It is important that T4 polymerase be included in this reaction to chew back 3′ ends leaving free 5′ends, on which the kinase acts much more readily than on blunt ends. Using T4 polynucleotide kinase alone, even in the presence of crowding factors, does not yield enough 5′ phosphorylated DNA for efficient library creation. The products of this reaction, after heat treatment at 70 °C for 10 min, may be directly added to the ligation reaction (Section 5.1).

5. LIGATION, TRANSFORMATION, RECOVERY, AND STORAGE OF THE LIBRARY

5.1. Ligation of the insert and acceptor DNA

The following reaction may be scaled up after initial transformation tests to yield a library of the desired number of transformants. The main parameter to optimize to obtain better ligations is the insert:vector ratio.

1. Mix the following:

   Acceptor DNA—50 ng

Insert DNA—threefold molar excess over acceptor DNA

PEG-8000 (30% w/v)—5 μL

T4 ligase buffer—2 μL

Water—up to 20 μL

T4 DNA Ligase (2,000,000 cohesive end units/mL)—0.5 μL.

2. Place ligation mixture in a thermocycler and cycle between 10 and 30 °C every 30 s for 30 min (Lund, Duch, & Pedersen, 1996).
3. Dilute ligation 10-fold to prevent PEG coprecipitation and purify over a clean and concentrator column (Zymo Research Corp.) and elute in 20 μL of water preheated to 50 °C. Optionally, concentrate the DNA by vacuum centrifugation to a final volume of 2 μL.

5.2. Transformation

1. Add the 2 μL of ligation product to 40 μL of high efficiency (>1×10^9 transformants/μg) DH10B cells (Life Technologies, Grand Island, NY) or any other highly electrocompetent strain.
2. Incubate on ice for 15 s and electroporate in a 0.2-cm cuvette (e.g., in a GenePulser II electroporator (Biorad Laboratories, Hercules, CA) set to 25 μF capacitance, 200 L/500H Ω resistance, and 2.5 kV).
3. Add 1 mL SOC (2% w/v bacto-tryptone, 0.5% w/v yeast extract, 8.56 m*M* NaCl, 2.5 m*M* KCl, 10 m*M* $MgCl_2$, 10 m*M* $MgSO_4$, 20 m*M* glucose) that has been prewarmed to 37 °C and mix.
4. Dispense into a 15 mL centrifuge tube and incubate with shaking at 37°C for 1 h. Multiple electroporations can be combined at this step.
5. Make four serial dilutions of 10 μL of the cells into 90 μL of SOC and plate 50 μL of each dilution on 10 cm LB plates containing the appropriate selective antibiotic. Place the remaining cells on large LB plates (e.g., 245×245 mm bioassay dish; Nalgene-Nunc) containing the appropriate selective antibiotic. If multiple electroporations are performed, one can either plate aliquots of 1 mL on multiple plates or concentrate the cells by centrifugation as above followed by resuspension in 1 mL SOC and plating on a single plate. Although single isolates are not required at this point, overcrowding on a plate can lead to bias within one's library. The colonies that grow on the 10 cm plates are used to determine the number of transformants (total transformants = (number of colonies on the 10 cm plate)/(volume of transformed cells plated on 10 cm plate) × (volume of cells plated on large plates)).
6. Incubate overnight at 37 °C.

5.3. Recovery and storage of library

1. Recover cells from 245×245 mm bioassay dish by adding 2×15 mL storage media (18 mL LB, 9 mL 50% glycerol, 3 mL 20% w/v glucose) to the top of the plate, scrape cells from media using a cell spreader, and then pipette cells into a 50 mL polypropylene centrifuge tube.
2. Spin cells in a centrifuge at 2000×*g* at 4 °C for 20 min.
3. Decant supernatant and add 2 mL of storage media.
4. Resuspend the pelleted cells by gentle shaking.
5. Store in aliquots in 1.5 mL cryovials at −80 °C.

6. CHARACTERIZATION OF THE LIBRARY

The colonies appearing on the small plate of the library can be used to characterize the library. There are many ways to access the library quality, from analyzing plasmid size, to restriction digestions, to PCR screens of individual colonies. The appropriate and most useful method(s) depend on the nature of the library created. The following restriction digest assay on purified plasmid from 10 colonies is straightforward and provides information on both the fraction of the library that received an insert and the diversity of insertion site. Choose a restriction enzyme (or pair of enzymes) that digests once within the insert DNA and once within the acceptor DNA. Digest the prepared DNA from each of the 10 colonies following manufacturer's instructions. Analyze the size of the digestion products via agarose gel electrophoresis.

Acknowledgments

This work was supported by the National Institute of General Medicine at the National Institutes of Health (R01 GM066972), the National Science Foundation (CBET-0828724), and the Defense Threat Reduction Agency (HDTRA1-09-1-0016).

References

Edwards WR, Busse K, Allemann RK, Jones DD. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. Nucleic Acids Research. 2008; 36:e78. [PubMed: 18559359]

Guntas G, Mansell TJ, Kim JR, Ostermeier M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:11224–11229. [PubMed: 16061816]

Guntas G, Mitchell SF, Ostermeier M. A molecular switch created by in vitro recombination of nonhomologous genes. Chemistry & Biology. 2004; 11:1483–1487. [PubMed: 15555998]

Guntas G, Ostermeier M. Creation of an allosteric enzyme by domain insertion. Journal of Molecular Biology. 2004; 336:263–273. [PubMed: 14741221]

Heins RA, Choi JH, Sohka T, Ostermeier M. In vitro recombination of non-homologous genes can result in gene fusions that confer a switching phenotype to cells. PLoS One. 2011; 6(11):e27302. [PubMed: 22096548]

Ke W, Laurent AH, Armstrong MD, Chen Y, Smith WE, Liang J, et al. Structure of an engineered β-lactamase maltose binding protein fusion protein: Insights into heterotropic allosteric regulation. PLoS One. 2012; 7:e39168. [PubMed: 22720063]

Kim JR, Ostermeier M. Modulation of effector affinity by hinge region mutations also modulates switching activity in an engineered allosteric TEM1 β-lactamase. Archives of Biochemistry and Biophysics. 2006; 446:44–51. [PubMed: 16384549]

Liang J, Kim JR, Boock JT, Mansell TJ, Ostermeier M. Ligand binding and allostery can emerge simultaneously. Protein Science. 2007; 16:929–937. [PubMed: 17400921]

Lilley DM. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. Proceedings of the National Academy of Sciences of the United States of America. 1980; 77:6468–6472. [PubMed: 6256738]

Lo WC, Dai T, Liu YY, Wang LF, Hwang JK, Lyu PC. Deciphering the preference and predicting the viability of circular permutations in proteins. PLoS One. 2012; 7:e31791. [PubMed: 22359629]

Lund AH, Duch M, Pedersen FS. Increased cloning efficiency by temperature-cycle ligation. Nucleic Acids Research. 1996; 24:800–801. [PubMed: 8604328]

Meister, GE.; Kanwar, M.; Ostermeier, M. Circular permutation of proteins. In: Lutz, S.; Bornscheuer, U., editors. Protein engineering handbook. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2009.

Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. Genetics. 1988; 120:621–623. [PubMed: 2852134]

Ostermeier M. Engineering allosteric protein switches by domain insertion. Protein Engineering, Design & Selection. 2005; 18:359–364.

Panayotatos N, Wells RD. Cruciform structures in supercoiled DNA. Nature. 1981; 289:466–470. [PubMed: 7464915]

Sambrook, J.; Russell, D. Molecular cloning: A laboratory manual. 3. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.

Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.

Sohka T, Heins RA, Phelan RM, Greisler JM, Townsend CA, Ostermeier M. An externally-tunable bacterial band-pass filter. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:10135–10140. [PubMed: 19502423]

Tullman J, Guntas G, Dumont M, Ostermeier M. Protein switches identified from diverse insertion libraries created using S1 nuclease digestion of supercoiled-form plasmid DNA. Biotechnology and Bioengineering. 2011; 108:2535–2543. [PubMed: 21618478]

Wright CM, Majumdar A, Tolman JR, Ostermeier M. NMR characterization of an engineered domain fusion between maltose binding protein and TEM1 β-lactamase provides insight into its structure and allosteric mechanism. Proteins. 2010; 78:1423–1430. [PubMed: 20034108]

Wright CM, Wright RC, Eshleman JR, Ostermeier M. A protein therapeutic modality founded on molecular regulation. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:16206–16211. [PubMed: 21930952]

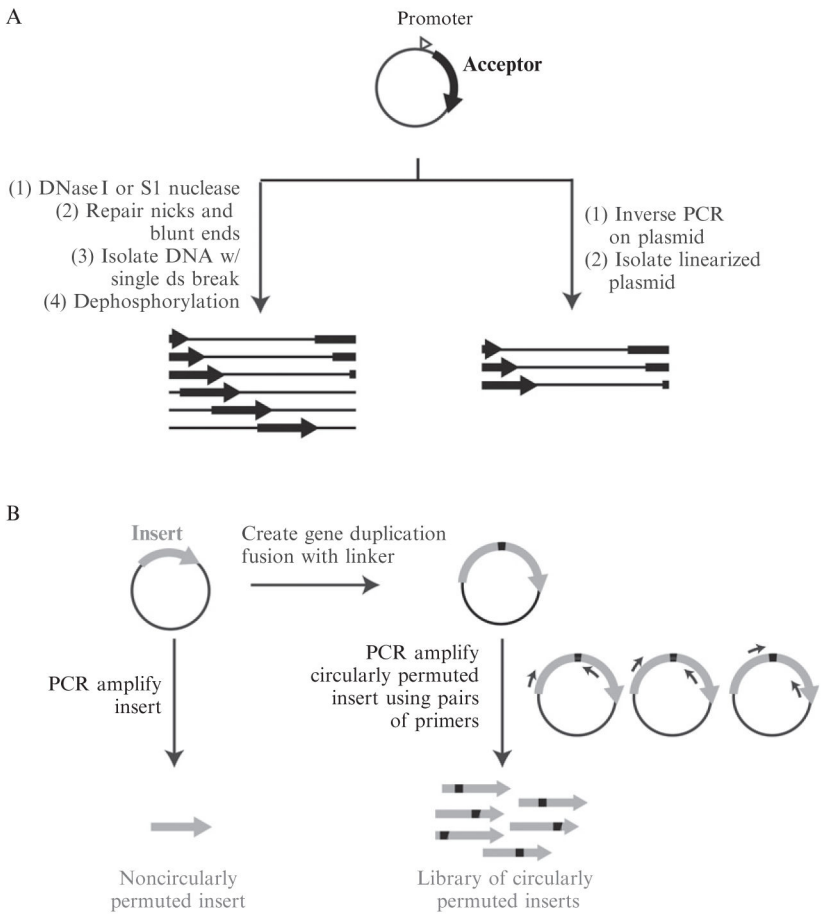

Figure 17.1.

Overview of methods for random domain insertion with optional circular permutation of insert. (A) Methods for making a linear plasmid DNA containing the acceptor gene in which the site of linearization is created randomly throughout the plasmid by DNase I or S1 nuclease or targeted to occur in the acceptor gene by multiplex inverse PCR. (B) Methods for preparing the insert DNA either without circular permutation or with circular permutation via a PCR on a gene duplication fusion.

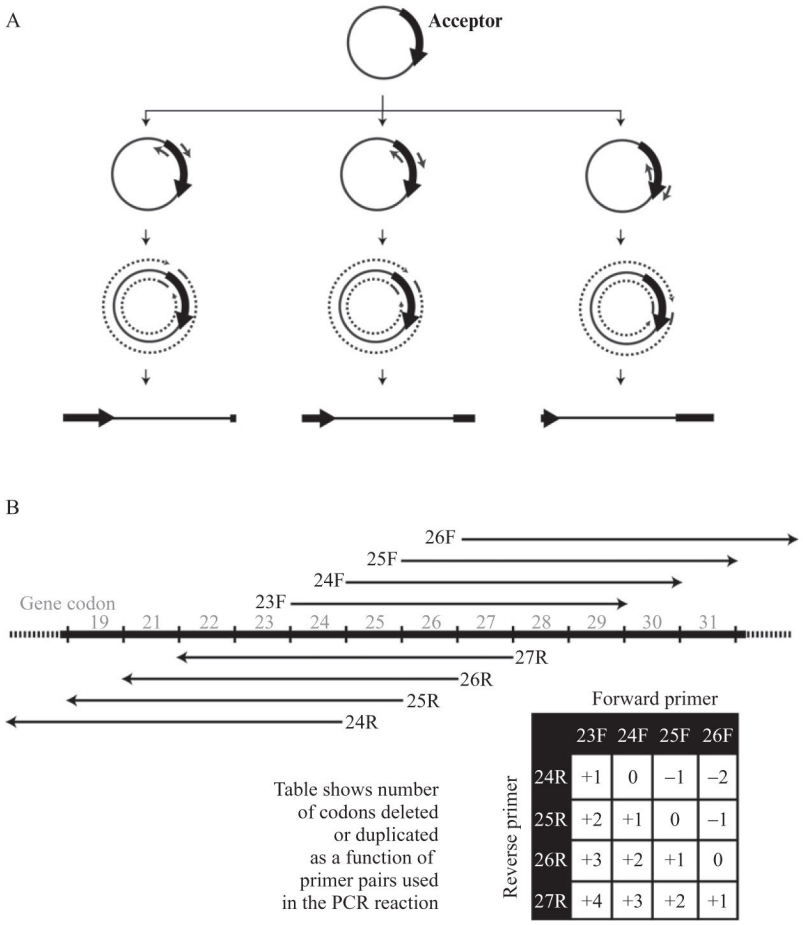

Figure 17.2.

Schematic depiction of multiplex inverse PCR. (A) Basic mechanism. PCR is performed with abutting sets of primers designed to amplify in opposite directions around the plasmid resulting in a set a linear plasmid dsDNA molecules whose site of linearization is defined by the primers. (B) Details of primer design and method for creation of deletions and duplications at the site of linearization. As an example, amplification using a set of four forward and four reverse primers is shown in the region around codons 19–31 of a gene. Different combinations of forward and reverse primers results in codon deletions or tandem duplications as indicated in the table.

Table 17.1

Comparison of methods for creating random double-strand breaks in acceptor DNA in order to create domain insertion libraries

	DNase I	S1 nuclease	Multiplex inverse PCR
Pros	• More random than S1 nuclease • Some tandem duplications produced	• Easier library creation than DNase I • Unlike DNase I, not prone to large deletions	• Creates focused library • The distribution of direct insertions, tandem duplications, and deletions is user-defined • Easier library creation • Allows for simultaneous creation of linker libraries at insertion site • Shorter series of steps
Cons	• Library construction is challenging • Prone to large deletions • Large fraction of library inserted outside gene or out of frame	• May not be as random as DNase I • Prone to insertions at sequences that produce stem-loop structures, such as inverted repeats • May not produce tandem duplications at insertion site • Large fraction of library inserted outside gene or out of frame	• Requires large sets of primers and many PCR reactions • Library is less diverse • Creation of large tandem duplications problematic
References	Guntas et al. (2005, 2004), Guntas and Ostermeier (2004), and Wright et al. (2011)	Tullman et al. (2011) and Wright et al. (2011)	This work