

Degenerate Oligonucleotide Gene Shuffling (DOGS) and Random Drift Mutagenesis (RNDM): two complementary techniques for enzyme evolution.

Running Title: DOGS and RNDM as tools for enzyme evolution

Peter L. Bergquist

Biotechnology Research Institute, Macquarie University, Sydney, NSW 2109,
Australia and Department of Molecular Medicine & Pathology, Auckland University
Medical School, Auckland, New Zealand

Rosalind A. Reeves and Moreland D. Gibbs

Biotechnology Research Institute, Macquarie University, Sydney, NSW 2109,
Australia

Corresponding author:

Peter L. Bergquist

Biotechnology Research Institute

Macquarie University

Sydney

NSW 2109

Australia

Phone: +61 2 9850 8614

Fax: +61 2 9850 9748

Email: peter.bergquist@mq.edu.au

Abstract

Improvement of the biochemical characteristics of enzymes has been aided by misincorporation mutagenesis and DNA shuffling. Shuffling techniques can be used on a collection of mutants of the same gene, or related members of a gene family can be shuffled to produce mutants encoding chimeric gene products. One difficulty with current shuffling procedures is the predominance of unshuffled ('parental') molecules in the pool of mutants. We describe a procedure for gene shuffling using degenerate primers that allows control of the relative levels of recombination between the genes that are shuffled and reduces the regeneration of unshuffled parental genes. This procedure has the advantage of avoiding the use of endonucleases for gene fragmentation prior to shuffling and allows the use of random mutagenesis of selected segments of the gene as part of the procedure. We illustrate the use of the technique with a diverse family of β -xylanase genes that possess widely different G+C contents. In addition, we introduce a new method (RNDM) for rapid screening of mutants from libraries where no selection has been imposed on the cells. They are identified only by their retention of activity. The combination of RNDM followed by DOGS allows a comprehensive exploration of a protein's sequence space.

Keywords:

polymerase chain reaction; primer extension; *in vitro* evolution; gene shuffling; flow cytometry; random mutation

Introduction

Proteins have evolved for the survival benefit of an organism and so may not exhibit features essential for *in vitro* applications. They function within the complex processes that make up a living system, and hence some of their properties may be undesirable when removed from this context. Therefore, there is considerable room for improvement of properties for which natural selective pressure was never applied. Protein engineering has the potential to dramatically enhance protein performance in a wide variety of non-natural, but technologically interesting, environments [see recent reviews, refs 1-3]. We have developed two related technologies that can be used to improve the performance and characteristics of proteins such as enzymes, as described below.

Shuffling techniques can be used on a collection of mutants of the same gene, or related members of a gene family can be shuffled to produce mutants encoding chimeric gene products. One difficulty with current shuffling procedures is the predominance of unshuffled ('parental') molecules in the pool of mutants. We have developed a procedure for gene shuffling using degenerate primers that allows control of the relative levels of recombination between the genes that are shuffled and reduces the regeneration of unshuffled parental genes [Degenerate Oligonucleotide Gene Shuffling, DOGS, ref. 4]. This procedure has the advantage of avoiding the use of endonucleases for gene fragmentation prior to shuffling and allows the use of random mutagenesis of selected segments of the gene as part of the procedure. This procedure can be combined with Random Drift Mutagenesis (RNDM) for wider exploration of the sequence space of shuffled genes (see later). The general procedure is shown below in Figure 1.

Figure 1 to go near here

Random Drift Mutagenesis (RNDM) has been developed to determine whether a phenotype that is derived from the interaction of multiple amino acids might require the accumulation and interaction of neutral mutations (neutral in isolation) and adaptive mutations. The procedure uses iterative misincorporation mutagenesis but no screening for adaptive mutations occurs. Instead, is only done for retained ability (whether unchanged, improved or reduced) to catalyze the hydrolysis of a substrate. This procedure is intended to provide high speed screening of mutants for retained activity without tedious assay procedures and to allow a comprehensive examination of sequence space for superior mutants. It is a platform technology applicable to any protein for which there is a colorimetric or fluorescent assay, for example, the precipitation of indigo as the result of hydrolysis of an indoxyl group as the sorting signal combined with flow cytometric analysis (Figure 2). The recombination to allow superior genes to be isolated is performed using DOGS.

Figure 2 to go near here

Random drift mutagenesis has been developed as a front-end complement to DOGS and theoretically gives much greater scope and flexibility in protein evolution. Our approach is to use DOGS for the shuffling step because of the greater control afforded and the high frequency of recombinants compared to regenerated wild type molecules that can be obtained. The procedure can be used with different genes (family shuffling) or a family of mutants can be created from the same gene by

misincorporation mutagenesis and the products of either pathway can be shuffled using the same technology.

The iterative screening of mis-incorporation mutagenesis libraries with low mutation frequencies (1-2 amino acid changes/protein) allows the stepwise accumulation of adaptive mutations. Stepwise accumulation of single mutations allows the sampling of a far larger functional sequence space than if highly mutated clones are screened. Consequently, misincorporation mutagenesis tends to select for adaptive mutations while discarding neutral mutations or mutations which decrease the fitness of an enzyme.

Phenotypes such as thermostability can be readily improved by iterative incorporation of single mutations. RNDM is a technique to determine whether a phenotype that is derived from the interaction of multiple amino acids might require the accumulation and interaction of neutral mutations (neutral in isolation), and adaptive mutations. It uses iterative misincorporation mutagenesis in the normal manner. No screening for adaptive mutations is carried out and screening is done only for retained ability to catalyse the hydrolysis of a substrate. All positive recombinants are combined and used as template for a further round of mutagenesis and so on for as many rounds as are necessary. In this manner, it is assumed that accumulation of multiple adaptive, neutral and harmful (but not inactivating) mutations occurs. Once generated, this library is then screened for recombinants with modified biochemical activity. Recombinants with the desired phenotype are then backcrossed with the parental wild-type gene using DOGS to remove any deleterious or neutral mutations.

Materials, Methods and Techniques

Sources of genes

Family 11 xylanase genes were obtained from the following bacterial strains:

Dictyoglomus thermophilum strain Rt46B.1 *xynB* [5]; *Clostridium stercorearium* *xynB*

[6]; *Bacillus* sp. strain V1-4 [7]; *Caldicellulosiruptor* sp. strain Rt69B.1 *xynD* [8]

Clostridium thermocellum *xynV* [9] and *Streptomyces roseiscleroticus* *xyl3* [10].

Each gene was PCR amplified from genomic DNA using the respective gene-specific primers. RNDM experiments were carried out with a thermophilic β -glucosidase gene from *Caldicellulosiruptor saccharolyticus* [11] because of the availability of suitable substrates.

Methods

As an example of the DOGS procedure, we give examples from the shuffling of the *Dictyoglomus thermophilum* Rt46B.1 xylanase gene reported by us previously [4].

Degenerate-end complementary primer pairs for efficient PCR amplification of gene segments and overlap-extension of adjacent segments

Rose *et al.* [12] have described a strategy that overcomes problems of degenerate methods for primer design called Consensus-Degenerate Hybrid Oligonucleotide Primers (CODEHOP). A modification of the method allows efficient amplification of overlapping segments of related genes, and subsequent overlap-extension of adjacent segments from different genes, resulting in the formation of chimeric gene fragments. The careful design of the primer sequences is the most important step in the DOGS procedure. The DOGS procedure requires the design of perfectly complementary pairs of primers. Each primer has a non-degenerate core flanked by both 5' and 3' prime degenerate ends, referred to here as complementary degenerate-end primers (CDE primers). As with the CODEHOP primers, the 3' degenerate core gives each CDE primer their template-binding specificity, while the non-degenerate region acts as a stabilising clamp in subsequent rounds of the PCR. The 5' degenerate end is not required to contribute to the binding efficiency of the CDE primer during PCR.

However, it plays a pivotal role in allowing efficient binding and subsequent overlap-extension of separate PCR products (gene segments) generated using respectively, the forward or the reverse CDE primers.

The non-degenerate core of individual CDE primers is generally based upon the corresponding coding sequence of one gene, designated the parental gene for shuffling. This results in the formation of chimeric fragments that retain parental sequence at the points of segment overlap.

Example of primer design

An example of the design strategy for making complementary oligonucleotide pairs suitable for the amplification of gene segments from related genes, and for the subsequent overlap extension of segments to generate chimeric genes is given in Figure 3.

Figure 3 to go near here

Features of CDE primers

CDE primers allow efficient and specific amplification of portions (referred to here as gene segments) of related but divergent genes. The 5' degenerate-end of CDE primers ensures that separate PCR products generated with the respective forward or reverse complementary CDE primers. Furthermore, multiple (1 or more) pairs of CDE primers allow the generation of consecutive PCR products (gene segments) with complementary ends suitable for overlap extension and PCR resulting in the generation of recombined segments. CDE primers may also be used in combination with complementary degenerate primers that do not have a non-degenerate core, to generate consecutive PCR products (gene segments) with complementary ends suitable for overlap extension and PCR resulting in the generation of recombined

segments. The mixing of segments amplified from related genes followed by overlap extension and PCR results in the efficient generation of chimeric gene fragments.

The non-degenerate core of each complementary CDE primer set may be (but does not have to be) based upon the gene designated the parental gene. Finally, gene segments amplified from related genes can be mixed in unequal amounts allowing control of the level of incorporation of each segment into resultant chimeric gene fragments. An example of the gene segment amplification and overlap extension with CDE primers is depicted in Figure 3.

Complementary Degenerate-End primer pairs for efficient PCR amplification of gene segments and overlap-extension of adjacent segments

The nucleotide sequences of the xylanase genes were aligned and degenerate primers were designed based on the conserved amino acid motifs found in all of the genes (see Figure 3 for example). The genes could be divided into eight fragments on the basis of alignment of the conserved regions. Degenerate forward and reverse primers were designed which allowed amplification of the DNA in the eight segments when combined, as appropriate, with the nested 5' and 3'-common primers.

All of the segments for each gene were amplified using combinations of adjacent CDE primers and nested 5' and 3'-common primers, then individually gel-purified from the PCR mixtures. PCR conditions for this step were: one cycle of 95°C for 1 min; then 35 cycles 95°C (denaturation), 30 sec; annealing at 35°C, 20 sec; and extension at 72°C, 40 sec with a final incubation at 72°C for 5 min. Platinum *Pfx* polymerase (Life Technologies Pty Ltd, Victoria, Australia) was used for all PCRs. The segments of each gene were mixed in the appropriate ratio to give the desired level of chimerization. For example, using the six candidate genes G1-G6, where G1

is the *D. thermophilum* Rt46B.1 *xynB* gene, and deciding that this sequence should predominate in the shuffled progeny, the pooled PCR segments for each gene were mixed in the ratio of 8.75 G1 to 1:1:1:1:1 to give chimeras with 5/8 segments from Rt46B.1 *xynB*. Fifty to 100 ng of mixed segments were then used as templates for overlap extension [13] using the following conditions: one cycle of 95°C for 1 min.; then 35 cycles 95°C (denaturation), 30 sec; annealing at 35°C, 20 sec; and extension at 72°C, 40 sec. and a final incubation at 72°C for 5 min. The high fidelity Platinum *Pfx* polymerase was used for all overlap extension and PCR to decrease the frequency of PCR misincorporation mutations. A second reason for using a high fidelity polymerase is its lack of terminal transferase activity as addition of 5' nucleotides would interfere with template priming during the overlap extension procedure.

Chimera amplification

Chimeric fragments were regenerated into complete genes by using the overlap-extended products (50-100 ng) as a template for PCR using the common flanking nested 5'- and 3'- primers under the following conditions: one cycle of 95°C for 1 min to activate the enzyme; then 20 cycles consisting of 95°C (denaturation), 30 sec; annealing at 50°C, 20 sec; extension at 72°C, 40 sec. and a final incubation at 72°C for 5 min using Platinum *Pfx* DNA polymerase.

Cloning of shuffled products

DOGS PCR products were digested with the restriction enzymes *Bam*HI and *Hind*III and ligated to pBSII KS- (Stratagene, San Diego, CA) that had been digested with the same restriction enzymes, and treated with Shrimp alkaline phosphatase (Roche Diagnostics Australia, NSW, Australia). The ligated DNA was used to transform *E. coli* strain DH5 α to ampicillin resistance and spread onto plates containing the antibiotic, X-gal and IPTG according to the instructions supplied by the manufacturer

of the vector. Individual colonies were picked and patched in duplicate onto new plates and screened for the expression of xylanase activity by the Congo Red overlay method [14].

Enzyme assays for xylanase activity

Whole cell extracts containing recombinant xylanase was prepared for enzyme assays using the protein extraction reagent BPER II (Pierce Chemical Company, Rockford, IL). Xylanase activity was determined using the method of Lever [15] with Birchwood xylan (Sigma-Aldrich, Sydney, Australia) as substrate. The standard assay reaction mixture consisted of 0.5% (w/v) xylan in 120 mM universal buffer [16] pH 6.5, and enzyme to give a final volume of 0.03 ml. The reaction mixture was incubated at 60°C for 20 min.

Vector Construction for flow cytometry and cloning of β -glucosidase gene

The vector pProEXHTc was cut with *Bgl*III to give two fragments. The larger fragment was isolated and the smaller fragment, carrying the ampicillin resistance marker was discarded. The kanamycin resistance gene was isolated from plasmid pET26B after *Bgl*III digestion and was ligated with the pProEXHTc backbone to give pProEXHTc:kan, enabling selection to be made in 25-30 $\mu\text{g}\cdot\text{ml}^{-1}$ kanamycin. The *Cs. saccharolyticus* β -glucosidase gene was cloned directionally downstream of the pTRC promoter following *Bam*HI-*Hind*III digestion of both vector and gene. The gene encodes a thermophilic β -glucosidase with a T_{opt} under the conditions used of 70°C at pH_{opt} of 6.5 [11].

Flow Cytometry

E. coli cells were sorted on a Becton Dickinson FACS Calibur flow cytometry system (Becton Dickinson, North Ryde, Australia) employing an Argon ion 488 nm laser. X-gal-labelled cells were readily distinguished from unlabelled cells using the FL1

emission filter (515 – 565 nm). ImaGene Green-labelled cells were also distinguished from unlabelled cells using the FL1 emission filter.

Construction of a Misincorporation Mutagenesis Library of the β -glucosidase

Misincorporation mutagenesis PCR was performed using the vector-based M13MG/PPROR primers with biased dNTPs/0.5mM Manganese by standard methods. The resulting PCR product was digested with *Nco*I and *Bam*HI, ligated into pProEXHTC, and then transformed into *E. coli* DH5 α . Randomly selected colonies were picked and patched to duplicate plates, then screened for activity by the overlay method. Selected recombinants were sequenced to establish the mutation frequency.

Mutant Isolation by Flow Cytometry using RNDM, X-gal and ImaGene Green

In preliminary studies we have shown that *Cs. saccharolyticus* β -glucosidase can efficiently cleave the synthetic colorimetric substrate X-gal (5-Bromo-4-chloro-3-indolyl-D-galactopyranoside) and the commercially-available fluorescent substrate C₁₂-fluorescein-digalactoside (ImaGene Green, Molecular Probes Inc. OR) which has a proven application in FACS [17-19].

Mutagenic PCR of the *xynA* gene was carried out so that 2-3 amino acid changes per protein molecule were obtained. The PCR products were ligated into the pPROExHTc:Kan vector and used to transform competent *Escherichia coli* DH5 α .

Outgrowth of the transformation mixture with kanamycin selection followed 50-fold dilution to allow the transformed cells to outgrow the untransformed cell population (to give approximately 1×10^{11} cells). The chromogenic substrate X-gal or the fluorogenic substrate, ImaGene Green were added and the cells heat-shocked at 65°C for 5 minutes, followed by incubation at 37°C for 15 minutes. The method requires the cleaved substrate to be retained within the cells, which are then sorted in flow cytometer and positive cells collected on a 0.2 μ membrane. The cells were washed

off the filter and added to a second mutagenic PCR reaction using AmpliTaq Gold polymerase (Perkin-Elmer) to generate new mutagenic amplification products. The polymerase had an initial 15 min incubation time at 94°C and this incubation facilitates the release of plasmid DNA from the cells.

Results

DOGS

After shuffling using an 8.75:1:1:1:1:1 ratio of input DNAs, chimeric fragments were directly ligated into the pBS KS- vector and transformed into *E. coli* DH5a. Thirty-two transformants were picked at random, sequenced and the resulting data compared to the parental sequences. It was possible to assign the origin of each segment in a recombinant from the sequence data. Figure 4 summarises some results from one round of DOGS of the six xylanase genes.

Figure 4 to go near here.

Some point mutations and one deletion was identified in recombinants that were otherwise wild type (presumably resulting from PCR error in the initial stages where AmpliTaq polymerase was used (it has no proof-reading ability)). What is striking about the procedure is that about 70% of the recombinants are chimeric. The preponderance of Rt46B.1 *xynB* sequences was what would be expected from the input ratios. Another experiment was carried out in which the input ratios for all segments were 1:1:1:1:1:1. A total of twenty chimeras were sequenced and on average only 1 segment per chimera was derived from *D. thermophilum xynB6* (data

not shown). This result indicated that the *xynB6*-derived degenerate clamp of CDE primers did not bias recombination toward reformation of the parental gene.

Correlations between enzyme activity and gene family segment content are best viewed by a comparison of the segment composition of a particular shuffled recombinant and its pH profile from 3 to 10 and its activity at temperatures from 42° to 92°C at pH6.5. The three-dimensional structure of the *D. thermophilum* Rt46B.1 XynB has been solved at 1.8Å [20] and a ribbon representation of the structural features are shown in ref. [4]. Based on site-directed mutagenesis and crystallographic analysis of the active site residues of the closely related *B. circulans* xylanase, it has been proposed that Glu78 and Glu172 act as a nucleophile and an acid-catalyst respectively [21-24]. In ref. 4, the two catalytic glutamic acids are shown and it was found that virtually every segment has some portion sitting in the catalytic groove of the enzyme. The segments which contribute least to the catalytic cleft are 1, 5, 6 and 8, and these segments are the ones most frequently found to be shuffled in the active chimeric genes with 74% of the shuffled segments occurring at these positions. The two glutamic acid residues directly implicated in the catalytic mechanism lie respectively on segment 4, and on the boundary (degenerate primer binding position) of segments 7 and 8. McIntosh *et al.* [25] describe the pK_a of the acid/base catalyst (Glu172) of the *Bacillus circulans* family 11 xylanase and the complex situation likely to be encountered in trying to change the pH optimum of a family 11 xylanase. Their results suggest that a change in the pH optimum of XynB may require quite elaborate structural changes to the enzyme, which might not be compatible with high temperature stability. In addition, acid residues with very high pK_a's like Glu172 are inherently unstable and to increase the pK_a any further may destabilize the enzyme. Although no recombinants with altered pH optima were

found in the mutants screened so far, the small sample of mutants examined probably precluded their identification and we intend to report further after a more extensive examination.

RNDM

ImaGene Green is a derivative of fluorescein di- β -galactopyranoside (FDG) that has been covalently modified to include a 12-carbon lipophilic moiety. Once inside the cell, it is cleaved by β -glucosidase to give a fluorescent product that is retained in the cell because of its lipophilic tail. In early experiments, we stained stationary phase *E. coli* cells with ImaGene Green and examined them in a fluorescence microscope. We found that about 50% of the cells fluoresced and staining with propidium iodide (PI, a standard live/dead test) revealed that the fluorescent cells also fluoresced with PI. DAPI staining showed that all cells still contained DNA, and we concluded that colonies only fluoresce after exposure to ImaGene Green if they are dead. We examined the use of heat-shock to increase the fluorescent signal in the cells and found that all cells could become fluorescent indicating penetration of ImaGene green into the cells, and subsequent hydrolysis by BglA. We tried various methods of getting ImaGene Green into cells and found heating for 5 minutes at 65°C gave the best results (other treatments tested included treatment with citric acid, lactic acid and EDTA). These experiments revealed why overnight cultures (stationary phase with large numbers of dead cells) stained with either X-gal or ImaGene Green, while log phase cultures comprising healthy, rapidly dividing cells do not stain or fluoresce. These results indicated that the DNA is retained in the dead cells, a fact of some interest to us as the RNDM protocol calls for PCR amplification directly from sorted cells.

The ImaGene Green procedure has proved to be extremely sensitive to low amounts of enzymatic activity. The *Cs. saccharolyticus* BglA enzyme is expressed even in the uninduced cells because of incomplete repression by the *tac* promoter. Fluorescence microscopy of DH5 α containing the plasmid pProEXHTc:kan:*bglA* showed the retention of the cleaved product within the cells (not shown). The cells were heat shocked for 5 minutes in the presence of ImaGene Green and then were washed in water. Many cells fluoresced as a result of cleavage of the substrate and they could be sorted by fluorescence-activated cell sorting (FACS) into a BglA-positive and a BglA-negative population. The positive population of cells can be sorted directly onto a filter membrane for DNA isolation and use as a template for the next round of mutagenic PCR (Figures 5 and 6).

Figures 5 and 6 to go near here

How many cells need to be sorted to ensure a representative sampling? From the ligation step with the PCR product, we can expect 10,000-15,000 transformants using competent cells (assumed to give $>1 \times 10^8$ cfu.ug⁻¹ DNA). If we assume that mutagenic PCR with an error frequency such that 25% of the genes are inactivated, we can expect 7,500-15,000 mutant BglA transformants/transformation. The number of cells to be sorted to ensure a representative sampling is:

Sample size $N = \ln(1-P)/\ln(1-1/\text{population size})$. Therefore, for 10,000 unique enzymatically active transformants, we need to collect and sort about 70,000 cells to have a 99.9% chance of having collected samples of all mutants. This number can be sorted in minutes with contemporary flow cytometers.

Discussion

The DOGS procedure that we have described demonstrates that it is possible to shuffle members of a gene family that are not particularly closely related and still generate chimeric molecules at a high enough frequency that comprehensive and time-consuming screens are not necessary. The use of CDE primers has allowed the reliable PCR amplification and shuffling of equivalent gene segments from a diverse range of genes with low overall sequence homology. Although in the example we have given, we used broadly-related xylanase genes of Family 11 for shuffling, the procedure can be utilised with a single gene from which a misincorporation mutagenesis library can be produced and the most promising mutants can take the place of the different gene families used in our example. The individual mutants are PCR amplified using the CDE primers and are mixed in appropriate ratios and subjected to primer extension as in the DOGS procedure described to shuffle the genes and to recombine out deleterious and neutral mutations.

The recombination frequencies can be controlled by altering the segment input ratios so that shuffling of particular fragments can be enhanced or attenuated as required. Accordingly, the procedure allows domain-swapping experiments to be conducted with relative ease, replacing older methods that rely on suitable restriction enzyme sites. It is evident that PCR-induced misincorporation or error-prone mutagenesis can be incorporated as part of the procedure to introduce even more diversity into the products. In this respect, the DOGS method lends itself to random mutagenesis of individual segments to assist in fine-tuning of the encoded gene product. In this case, the CDE primers would be modified so that no degeneracy was available except for the segment to be mutagenized. This would give the investigator control over the

extent and nature of mutagenesis of a particular segment by introducing a DNA polymerase without proofreading activity at the appropriate stage in the procedure. In addition, it is clear from the design of the degenerate primers and the results reported here that altered nucleotide sequences will be generated even using a high fidelity DNA polymerase because of the mismatches designed into the degenerate primers. Even greater misincorporation mutagenesis can be generated by employing a polymerase without proofreading activity in the amplification and primer extension steps. It is apparent that the procedure lends itself to combination with other gene shuffling and combinatorial mutagenesis techniques for the generation of novel proteins with modified characteristics.

The key points we have learned so far with the RNDM procedure is that most enzymatic colorimetric or fluorometric substrates do not enter the cell unless the outer membrane potential is lost. It appears that the blue/white differentiation of colonies seen with X-gal only works because a fraction of the cells have died, allowing the diffusion of the substrate into the cell, where it is cleaved by β -galactosidase to give indoxyl which then dimerizes to give indigo. We have found that the best way to get X-gal or ImaGene Green into the cell for this application to heat-shock briefly at 65°C. ImaGene Green is much more sensitive than X-gal and should allow detection of activity over many orders of magnitude and it should be possible to detect even very low levels of activity if induction with IPTG is used. We believe that the sequence space of a particular protein can be thoroughly explored by first using RNDM and then shuffling the mutants by way of DOGS. We note that in this application the same CDE primers used for family shuffling are not obligatory and they could be positioned to avoid impinging on the active site. We intend to report on the combination of RNDM and DOGS in the near future.

While this technology can be used as it stands with glycosyl hydrolases, it could be further extended by replacing the galactoside side groups of the substrate with other β -1,4-linked sugars or oligosaccharides to allow the evolution of other enzymes such as cellulases and mannanases. 5-chloro-4-indolyl caprylate is available for the determination of hydrolytic reactions of lipases, but more sensitive fluorescent substrates could be envisaged, with the possibility of synthesizing substrates for the determination of transesterification reactions with lipases.

In summary, gene shuffling techniques dramatically increase the opportunities for exploring any protein's sequence space, but the low frequency of recombinant molecules that have to be distinguished from the parental wild type presents a major problem. DOGS offers a flexible technology that can allow saturation mutagenesis of predetermined regions as well as a low frequency of unchanged parental molecules amongst the progeny. RNDM provides an adjacent technology for refining productive sequence space at high throughput without selection for anything except residual activity.

Acknowledgements

This work was supported by grants from the Australian Research Council, the Macquarie University Research grants scheme and the Public Good Science Fund (New Zealand).

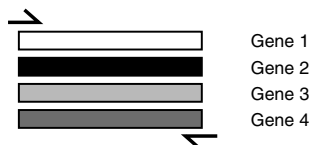
References

- [1]. Bessler C, Schmitt J, Maurer KH, Schmid, RD. 2003; Protein Science 12: 2141-2149.
- [2]. Turner NJ. 2003; Trends Biotech. 21: 474-478.

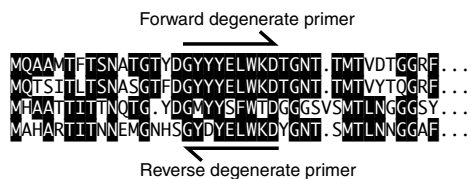
- [3]. Chopra S, Rananathan A. 2003; *Chemistry & Biology* 10: 917-926.
- [4]. Gibbs MD, Nevalainen KMH, Bergquist, PL. 2001; *Gene* 271: 13-20.
- [5]. Morris, DD, Gibbs MD, Chin CW, Koh MH, Wong KKY, Allison RW, Nelson PJ, Bergquist PL. 1998; *Appl. Environ. Microbiol.* 64: 1759 – 1765.
- [6]. Sakka K, Kojima Y, Kondo T, Karita S, Ohmiya K, Shimada K. 1993; *Biosci. Biotechnol. Biochem.* 57: 273 – 277.
- [7]. Yang VW, Zhuang Z, Elegir G, Jeffries TW. 1995; *J. Ind. Microbiol.* 15: 434 – 441.
- [8.] Morris DD, Gibbs MD, Ford M, Thomas J, Bergquist PL. 1999; *Extremophiles* 3: 103 – 111
- [9.] Fernandes AC, Fontes CM, Gilbert HJ, Hazlewood GP, Fernandes TH, Ferreira LM. 1999; *Biochem. J.* 342: 105 – 110.
- [10]. Elegir G, Szakacs G, Jeffries TW. 1994; *Appl. Environ. Microbiol.* 60: 2609 – 2615.
- [11]. Love DR, Fisher R, Bergquist PL. 1988; *Mol. Gen. Genet.* 213: 84-92.
- [12]. Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. 1998; *Nucleic Acids Res.* 26: 1628 - 1635.
- [13]. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989); *Gene* 77: 51 – 59.
- [14]. Teather RM, Wood PJ. 1982; *Appl. Environ. Microbiol.* 43: 777 – 780.
- [15]. Lever M. 1973; *Biochem. Med.* 7: 274-281.
- [16]. Britton HTS, Robinson RA. 1931; *J. Chem. Soc.* 1: 1456-1462.
- [17]. Plovins A, Alvarez, AM, Ibanez M, Molina M Nombela C. 1994; *Appl. Environ. Microbiol.* 60: 4638-4641.
- [18]. Kang Y, Saile E, Schell MA, Denny TP. 1999; *Appl. Environ. Microbiol.* 65: 2356-2362.

- [19]. Alvarez-Barrientos A, Arrayo J, Cantón R, Nombela C, Sánchez-Pérez M. 2000; *Clin. Microbiol. Revs.* 13: 167-195.
- [20]. McCarthy AA, Morris DD, Bergquist PL, Baker EN. 2000; *Acta Crystallogr. D. Biol. Crystallogr.* 56; 136–1375.
- [22]. Wakarchuk WW, Campbell RL, Sung WL, Davoodi J, Yaguchi M. 1994; *Prot. Sci.* 3: 467–475.
- [23]. Wakarchuk WW, Sung WL, Campbell RL, Watson DC, Yaguchi M. 1994; *Prot. Eng.* 7: 1379–1386
- [24]. Miao S, Ziser L, Aebersold R, Withers SG. 1994; *Biochemistry* 33: 7027–7032.
- [25]. Davoodi J, Wakarchuk WW, Campbell RL, Carey PR, Surewicz WK. 1995; *Eur. J. Biochem.* 232: 839–843.

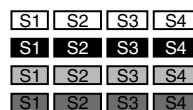
1. Design oligonucleotide primers with 3' ends specific for the N- or C-terminus of each candidate gene. Incorporate common nested 5' ends with suitable restriction sites for directional cloning of PCR products. PCR amplify each gene for use as PCR template.



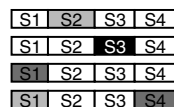
2. Design complementary degenerate primer pairs based upon one or more conserved motifs found in candidate genes.



3. Amplify each of the individual segments (S1-S4) for each gene using the degenerate primers and the common nested primers.



4. Mix segments from each gene to give desired levels of chimerisation. Regenerate full length chimeric genes by overlap extension of segments followed by PCR with primers specific for the common nested ends.



5. Digest and ligate full length fragments into an appropriate cloning vector, transform into expression host and screen individual recombinants for desired properties.

Figure 1: General description of the DOGS procedure

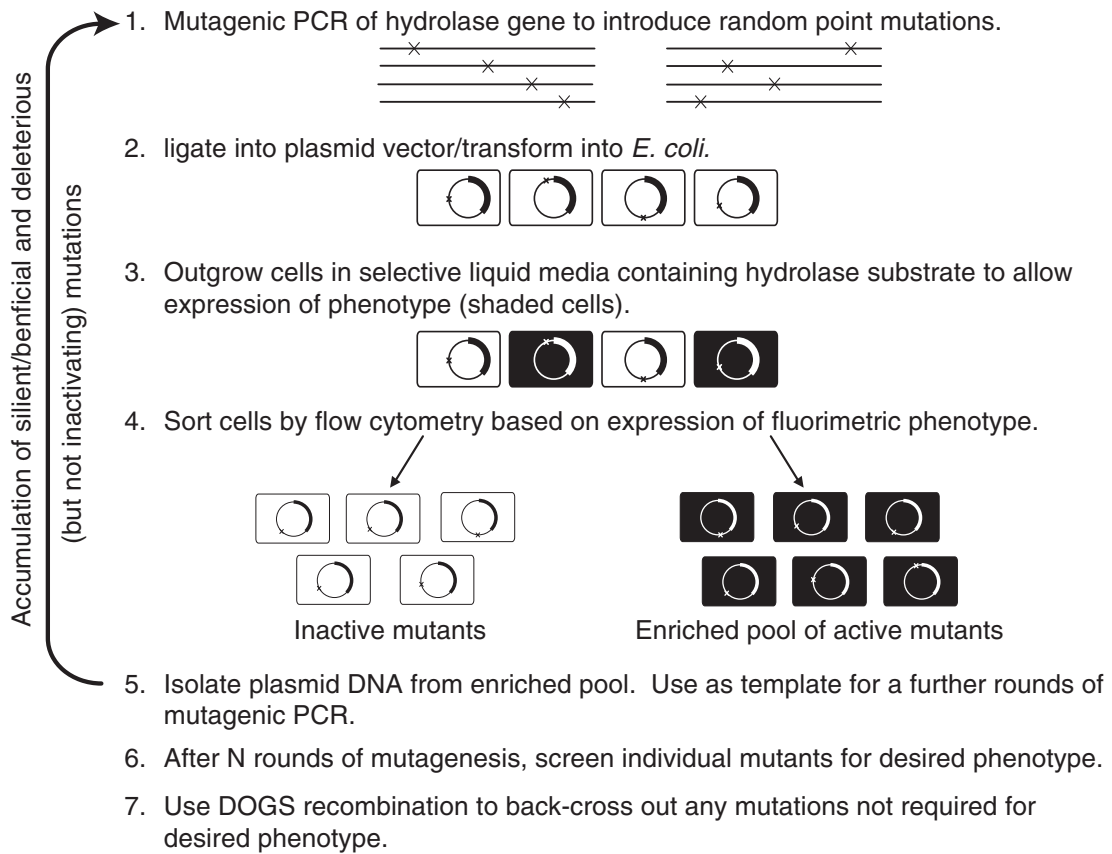


Figure 2: Schematic representation of the Random Drift (RNDM) procedure

A. N-terminal portion of alignment of xylanase protein sequences showing conserved residues.

```

D. thermo XynB6 MQTSITLTSNASGTFDGYYYELWKDTGNT...TMTVYTOGRF...
Ri69B.1 XynD  MQAAVTEISNATGTYDGYYYELWKDTGNT...TMTVDTGGRF...
C. thermo XynV  MRADVLTISNQTGTHGGYFNYWKDTGNG...TMTVKDGGAF...
C. sterco XynA  MLAGRITIDNEIGTHGGYDYELWKDYGNT...TVELNNGGTF...
Bacillus V1-4 XynA MAHARITITNEMGNHSGYDYELWKDYGNT...SMTLNGGAF...
S. rosei Xyl3   MFAATLITITNQTG...YDGYYSFNTDGGGSVSMTLNNGGSI...
    
```

B. Corresponding alignment of conserved DNA sequences showing primer binding positions.

```

SHF
5'-GAAAACCTGCAGTAGATG
XINTF1C 5'-GGBTACDCACTATGAACTATGGAARGA
nondegenerate core
D. thermo xynB6 GAAAACCTGCAGTAGATGCAAACTGTCATAACACTAACAAAGTAATGCAAGCGGTACTTTTTCATGGCTACTACTATGAACTATGGAAGAATACAGGGAATACA...
Ri69B1 xynD   GAAAACCTGCAGTAGATGCAAGCAAGCCATGACATTTACCTCTAATGCAACTGGGACATACGATGGTTACTACTACGAGTTCTGGAAAGACACAGGGAATACT...
C. thermo xynV GAAAACCTGCAGTAGATGCTCGCTGATGTGGTAATTACGTCAAACCAAGCGGTACTCACGGCGGTACCACTTTGAGTACTGGAAAGACACCGGAAATACG...
C. sterco xynA GAAAACCTGCAGTAGATGCTCGCGGGCGAATAATTACGACAATGAGACAGGGACACATGGAGGCTACCACTATGAGTCTGGAAAGACTACGGAAATACG...
Bacillus V1-4 xynA GAAAACCTGCAGTAGATGCTCCATCGGAGAACATTACCAATTAATGAAATGGTAACCATAGCGGTACCACTATGAACTATGGAAGGATATGGAAACACC...
S. rosei xyl3  GAAAACCTGCAGTAGATGCACGGCGCACTACCATACCACCAACCAGACGGC...TACGACGGCATGACTACTGCTTCTGGACCGAGCGCGGCTCC...
nondegenerate core
CCVATGHTGATACTTGATACCTTYCT-5' XINTR1C
nondegenerate core
    
```

C. An example of overlapping segments generated by PCR.

```

Forward strand of segment 1 from D. thermophilum xynB6 amplified using primers SHF + XINTR1C
GAAAACCTGCAGTAGATGCAAACTGTCATAACACTAACAAAGTAATGCAAGCGGTACTTTTTCATGGCTACTACTATGAACTATGGAAGAATACAGGGAATACA...
Region of overlap between segments
Forward strand of segment 2 from Bacillus V1-4 xynA amplified using primers SHF + XINTR1C (XINTR1C not shown on figure)
GGBTACDCACTACGAGTTGTGGAARGATTATGGAACACC...
    
```

Figure 3: An example of the design of CDE primers based upon the DNA coding sequence of conserved amino acids. **A.** Shows the N-terminal portion of an alignment of candidate xylanase protein sequences. **B.** Shows a DNA sequence alignment corresponding to the amino acid alignment. The conserved sequence region upon which a CDE primer pair (XINTF1C and XINTR1C) was designed is defined by the vertical bars. The horizontal bar marks the boundary of the non-degenerate core of each primer. In this example, the non-degenerate core sequence is designed to match the sequence of the selected parental gene (*D. thermophilum xynB6*). The binding position of the SHF primer designed to bind to the common end of all genes is also shown. **C.** Shows the forward strand of gene segments 1 and 2 amplified from *D. thermophilum xynB6* and *Bacillus V1-4* respectively. The segments were generated using the primer combinations SHF + XINTR1C and XINTF1C + XINTR2 (XINTR2 not shown) respectively. Vertical lines mark the region of overlap between the

segments that allows overlap extension between gene segments resulting in the formation of chimeric genes.

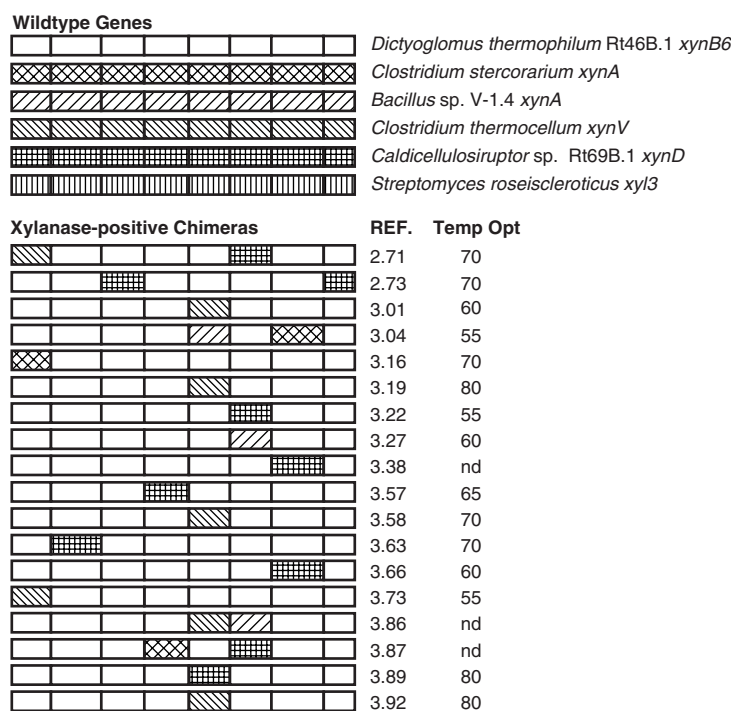


Figure 4: An example of functional chimeric xylanase genes generated using the DOGS procedure. The temperature optimum of each chimera is shown, nd = no data.

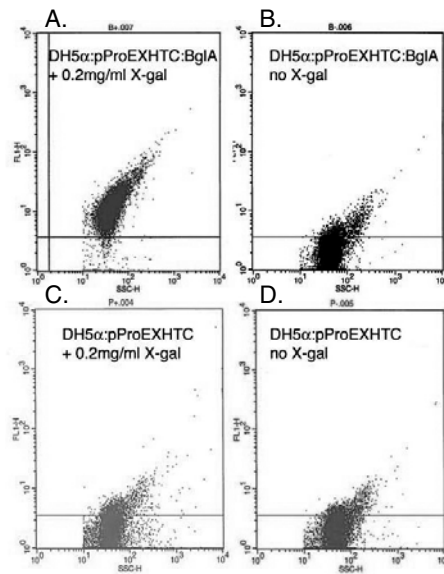


Figure 5: Flow figure of cell sorting by FACS flow cytometry (excitation 488nm, emission FL1 515-565nm) based upon expression of BglA⁺ phenotype and accumulation of insoluble Indigo within *E. coli* DH5-a cells containing the plasmid pProEXHTc:*bglA*. **A.** Sorting of *E. coli* cell populations containing either the plasmid pProEXHTc:*bglA* or pProEXHTc in the presence or absence of the chromogenic substrate X-gal.

Key:

Panel A: Cells expressing BglA in the presence of X-gal

Panel B: Cells expressing BglA in the absence of X-gal

Panel C: Control cells (vector only) in the presence of X-gal

Panel D: Control cells (vector only) in the absence of X-gal

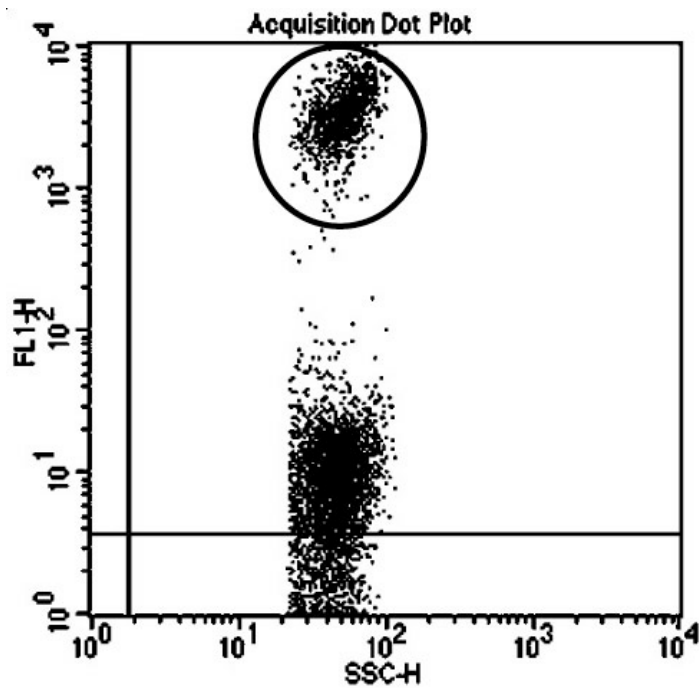


Figure 6: Figure showing cell population sorting by FACS flow cytometry (excitation 488, emission FL1 515-565nm) based upon expression of BglA phenotype and accumulation of ImaGene Green within the BglA⁺ cells. The results show a clear delineation of cell populations expressing the BglA⁺ phenotype (circled) vs BglA⁻ phenotype (not circled) based upon cell fluorescence due to the hydrolysis of ImaGene Green within each cell by BglA.