ELSEVIER

Novel methods for directed evolution of enzymes: quality, not quantity

Stefan Lutz* and Wayne M Patrick

In the past decade methods of directed molecular evolution have proven revolutionary in protein engineering. An increasing number of powerful new combinatorial techniques have joined rational design methods as effective tools for the manipulation and tailoring of biocatalysts. More and more, research in this maturing field is focusing on the quality and comprehensiveness of library construction and analysis. Additionally, in-depth studies have begun to highlight the underlying evolutionary mechanisms, limitations, and consequences of the various methodologies. Together, these investigations are creating a framework for future engineering projects.

Addresses
Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, Georgia 30322, USA
*e-mail: sal2@emory.edu

Current Opinion in Biotechnology 2004, **15**:291–297

This review comes from a themed issue on
Protein technologies and commercial enzymes
Edited by Karl-Erich Jaeger

Available online 5th June 2004

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2004.05.004

Abbreviations
epPCR error-prone polymerase chain reaction
IVC *in vitro* compartmentalization

Introduction

The laboratory approach to Darwinian evolution of biological macromolecules, often referred to as directed molecular evolution, presents a powerful means for exploring and altering enzyme substrate specificity and catalysis. In contrast to rational design, which generally concentrates on a small number of variants, evolutionary methods for protein engineering rely on the generation of vast molecular diversity by random mutagenesis and recombination. From among the thousands, millions or even billions of progeny, effective search engines such as genetic complementation and high-throughput screening identify candidates with improvements in a desired phenotype. Multiple iterations of this two-step process have led to the identification of numerous proteins with novel catalytic, biophysical, and molecular recognition properties (reviewed in [1]).

Methods for constructing diverse molecular libraries continue to accumulate. More important than the motivation to circumvent patent laws, these new approaches aim to generate more comprehensive, less biased libraries. This emerging focus on library quality is also reflected in new computational methods. Increasingly sophisticated algorithms offer valuable assistance in the design and analysis of directed evolution experiments.

New and exciting trends in directed evolution extend beyond library construction. Although the maxim that 'you get what you select for' [2] remains unavoidable, methods for identifying the properties that we actually desire are improving. Joining the ranks of genetic selection, panning for binding, and high-throughput screening, elegant *in vitro* selection systems for true catalysis now provide a powerful tool for library deconvolution.

Finally, several groups have started to address the fundamentals of directed molecular evolution. These detailed library analyses characterizing evolutionary intermediates and comparing existing methodologies provide valuable lessons for shaping the next generation of techniques and experimental design strategies.

In this review we will discuss each of these aspects, emphasizing their general applicability to, and implications for, the field of directed molecular evolution.

Methods for creating molecular diversity

Oligonucleotide-directed randomization, the error-prone polymerase chain reaction (epPCR), and *in vitro* recombination remain the cornerstones of directed molecular evolution (reviewed in [3,4]). However, several groups have recently reported interesting variations and modifications to these protocols (summarized in Table 1).

Oligonucleotide-directed randomization

Synthesizing partially or fully randomized oligonucleotides provides a means to generate diversity at specific locations in a gene. However, as discussed by Hughes *et al.* [5•], the degeneracy of the genetic code ensures that increasing the number of randomized codons decreases the efficiency of randomization. They provide a novel solution, in which a set of primers containing 20 MAX (for 'maximum efficiency') codons — those used preferentially to encode each amino acid in *Escherichia coli* — are hybridized to a template containing fully randomized (NNN or NNK, where N represents any nucleotide and K represents G or T) codons. Replacement of the template strand is then effected by PCR. This MAX

Table 1

Summary of new technologies for generating molecular diversity.

Methodology	Target	Library size	Outcome	Advantage/disadvantage	Ref.
Oligonucleotide-directed randomization					
MAX randomization	Zinc-finger protein	1×10^2	More uniform codon distribution; bias from genetic code removed	Reduces target library size	[5•]
Codon shuffling (hexamer assembly)	TEM-1 β-lactamase	1×10^3	Severely truncated, divergent sequences retain near-WT activity	Randomization not constrained by length of parental gene	[7]
Whole gene randomization					
Sequence Saturation Mutagenesis(SeSaM)	Green fluorescent protein	nr	100 clones sequenced; 88% of point mutations attributable to SeSaM	Unbiased or controllable mutation spectrum	[8]
Homology-dependent recombination					
Degenerate homoduplex recombination (DHR)	Mammalian epidermal growth factors	3×10^6	Mouse–human hybrid has 123-fold greater mitogenic potency	All shuffled variants equally likely	[10••]
Synthetic shuffling	15 *Bacillus* subtilisins	1.5×10^3	20% of variants active; improved pH and thermostability profiles	All shuffled variants equally likely	[11•]
Assembly of designed oligonucleotides (ADO)	*B. subtilis* lipases LipA and LipB	3×10^3	Twofold improvements in conversion and/or enantioselectivity	All shuffled variants equally likely	[12•]
Recombination-dependent exponential amplification PCR (RDA-PCR)	Green and yellow fluorescent proteins	nr	All clones analyzed have fluorescence similar to parents	Guarantees odd number of crossovers	[14]
Mutagenic and unidirectional reassembly (MURA)	*Serratia* phospholipase PlaA	3×10^3	Lipase activity increased 1000-fold; no decrease in phospholipase activity	Combines epPCR, DNA shuffling and unidirectional random truncation	[15]
Recombined extension on truncated templates (RETT)	Two *Serratia* chitinases	8×10^2	1.5-fold increase in activity; epPCR improved thermostability	Removes sequence bias of DNase I digestion	[16]
Homology-independent recombination					
Sequence-independent site-directed chimeragenesis (SISDC)	TEM-1 and PSE-4β-lactamases	1.5×10^3	Of 256 possible variants, 14 retain up to 30% WT activity	Number and locations of crossovers specified	[20•]
Structure-based combinatorial protein engineering (SCOPE)	DNA polymerases β and X	6.5×10^5	Hybrids with up to five crossovers complement auxotroph	Some variability in linkers between specified subdomain building blocks	[21]
Enhanced crossover SCRATCHY	Rat and humanGSTs	1×10^5	Average of 1.4 crossovers per gene; 1% retain parental activity	Increases number of crossovers; some variability in crossover location	[24•]

Abbreviations: WT, wild-type; GST, glutathione-*S*-tranferase; nr, not reported.

randomization necessitates the synthesis of at least 20 oligonucleotides per experiment, and does not allow for the randomization of multiple adjacent codons. However, by removing superfluous redundancy it offers the considerable advantage of decreasing library sizes required at the DNA level, while maintaining diversity amongst the encoded proteins. A method for the inverse outcome, protecting one specified codon while the remainder of the gene is targeted for randomization, has also been described [6].

In a second example, palindromic hexamer DNA duplexes representing the 20 proteinogenic amino acids as codon pairs have been used to replace large sections of the TEM-1 β-lactamase gene [7]. Remarkably, by ligating hexamer pools with a minimal segment of the β-lactamase gene — with no selection for the size of the resulting randomized constructs — several soluble, functional daughter sequences were obtained.

Whole gene randomization

In addition to oligonucleotide-directed methods for random mutagenesis at selected positions (*vide supra*), whole gene epPCR approaches have also been modified and improved. A method of 'sequence saturation mutagenesis' [8] has been devised to address the polymerase-induced biases inherent in traditional protocols [9]. It makes use of universal bases such as inosine to afford control over the mutational spectra obtained, although lacks the technical simplicity of epPCR. Polymerases

such as Stratagene's second-generation Mutazyme® (http://www.stratagene.com/) — which in combination with *Taq* DNA polymerase displays a uniform mutational spectrum — would appear to provide a simpler, more generally applicable method for constructing unbiased epPCR libraries.

Homology-dependent recombination

The most dramatic results in the field of directed evolution are united in their utilization of *in vitro* recombination. One intuitive and significant advance came with the near-simultaneous descriptions of 'degenerate homoduplex recombination' [10••], 'synthetic shuffling' [11•] and 'assembly of designed oligonucleotides' [12•]. In each, overlapping oligonucleotides encoding all of the degeneracy found in two or more parental genes are used as substrates in a polymerase-mediated reassembly reaction. The result is a library in which all shuffled variants are equally likely, regardless of whether or not parental diversity is tightly linked. The statistical basis of the advantages of this approach has also been discussed [13].

Other variant methods for homology-dependent recombination include 'recombination-dependent exponential amplification PCR' (RDA-PCR) [14], 'mutagenic and unidirectional reassembly' (MURA) [15], and 'recombined extension on truncated templates' (RETT) [16]. The former is closely related to staggered extension process (StEP) PCR [17], with the use of gene-specific primers guaranteeing an odd number of crossovers (but most commonly only one). MURA consists of a combination of incremental truncation (ITCHY [18,19]) and DNA shuffling mediated by a unidirectional primer; the result is a library of shuffled sequences that have been randomly truncated from one terminus. Finally, RETT seeks to avoid possible sequence biases arising from the use of DNase I digestion to generate shuffling blocks by instead using unidirectional, random priming from a single-stranded template. Template switching during reassembly results in recombination.

Homology-independent recombination

Homology-independent or structure-guided recombination has also seen several recent developments. 'Sequence-independent site-directed chimeragenesis' (SISDC) [20•] has expanded the concept of library construction by shuffling semi-rationally designed peptide building blocks. Although somewhat similar to 'structure-based combinatorial protein engineering' (SCOPE) [21], which uses visual inspection of protein structures to guide building block design, SISDC instead relies on the computational predictor SCHEMA (*vide infra*) to define peptide fragments from the parental proteins that can be interchanged with minimal structural interference. The subsequent combinatorial reassembly of these peptide fragments through clever use of restriction sites resulted in chimeras of the desired diversity, including hybrids that retained activity. However, these methods are (at least currently) limited to proteins with known three-dimensional structures.

An attractive approach to homology-independent recombination remains the construction of multiple-crossover libraries in which crossover positions are randomly distributed. SCRATCHY, the first method described for creating such libraries [22], yielded a minority of clones containing greater than one crossover. Moreover, a preselection for catalytically active, single-crossover parents was necessary to identify in-frame clones. This step inevitably eliminated numerous in-frame single-crossover clones that may have proven beneficial in the shuffled multiple-crossover SCRATCHY library. A new vector for the purpose of reading frame selection was subsequently constructed [23]. In addition, a method to increase the number of crossovers in SCRATCHY libraries was reported by Georgiou, Benkovic and coworkers [24•]. Single-crossover ITCHY libraries were divided into arbitrary sections and amplified with skewed sets of primers. The goal was to enrich for crossovers within broad sections of the chimera, rather than at strictly defined locations, and also to allow for some control over crossover number. The amplification products were used in the subsequent DNA shuffling step, where the increased concentration of heterologous fragments led to higher crossover numbers in the reassembled progeny.

Computational methods in directed evolution

In parallel with experimental techniques for directed molecular evolution, computational predictive frameworks have made dramatic advances in recent years (reviewed in [25]). Although the field of *de novo* protein design has enjoyed some spectacular successes [26,27], this review will focus on computational approaches developed to complement directed evolution experiments.

Computational methods in the context of directed evolution can rationalize published experimental results and guide future engineering studies. Potentially, the integration of computational data in library design will also allow researchers to focus on smaller — but more targeted and comprehensive — libraries, making previously untenable screening methods viable. In turn, this should generate a more detailed picture of the sequence space under investigation and provide insight into the structure–function relationships of proteins.

Recent reports demonstrate the increasingly successful use of iterations between modeling and experimental testing to evolve more reliable *in silico* predictors for protein engineering. Conceptually, these algorithms can be divided into two groups: simulating DNA

recombination; and predicting the functional consequences of amino acid and peptide fragment substitution.

DNA models

Algorithms to model recombination have relied on the hydrogen-bonding requirement between complementary strands and the energetics of nearest-neighbor interactions [25]. By integrating various contributing factors to DNA shuffling, such as fragmentation conditions, thermodynamic constraints and kinetic effects, Maheshri and Schaffer have improved the model for *in vitro* recombination [28]. Their predictions were largely confirmed by an accompanying set of experimental data. However, the main drawback of previous algorithms still remains: modeling the processes of library generation at the nucleic acid level provides no information on the functions of translated protein products.

Protein models

Computational methods have also been employed to simulate directed evolution by quantifying the impact of replacement, insertion, and deletion of building blocks in proteins. Scanning the known three-dimensional structures of protein engineering candidates for local networks of interactions, the SCHEMA algorithm [29] can identify recombination points between proteins that minimize the disruption of favorable steric and electrostatic interactions. Although the approach has shown some promise in guiding the generation of hybrid β-lactamases [20•,30], it only sums interactions within defined primary sequence windows and does not account for equally important interactions between distal residues.

Alternatively, the second-order mean-field identification of residue–residue clashes in hybrid proteins (SIRCH) procedure [31•] can identify and quantify potential conflicts (in terms of energy) between residues from parental sequences in the context of their tertiary structures. Relying on the proteins' total energy, as calculated from individual rotamer energies at each position, the approach successfully predicted local and long-distance clashes on two model systems. Although the calculated severity of unfavorable interactions was not always consistent with the experimental findings, overall predictions of preferred crossover points showed a promising correlation with laboratory data. Two other approaches, dubbed 'residue correlation analysis' (RCA) [32] and FamClash [33], borrow from bioinformatics by using multiple sequence alignments to recognize protein residues that vary in a coordinated fashion. Tested on dihydrofolate reductase, the resulting correlation maps identify potential clashes between regions of the parental structures. An interesting aspect of these results was the discovery of distant residues (>8 Å apart) that showed correlated amino acid substitution patterns. Networks of coupled motions, affecting the dynamic and even the functional properties of proteins, may explain this phenomenon [34].

In vitro selection for catalysis

For a successful directed evolution experiment, generating a large library is optional, but being able to recover the targeted progeny is mandatory. The development of novel and improved methods for library analysis is therefore as important as new approaches for creating library diversity. Among the variety of selection and screening techniques [35], particularly exciting advances have been reported in three areas of *in vitro* selection.

Blackburn, Plückthun and co-workers recently demonstrated the ability of phage display to select for catalytic activity, describing a selection based on the complete catalytic cycle of phosphate monoester hydrolysis [36••]. In their system, catalytic antibodies were identified that could hydrolyze an aryl phosphate. By design, the hydrolysis product spontaneously rearranged into a quinone-methide, which in turn reacted with nucleophiles at or near the active site of the catalyst, thereby capturing the phage particle.

Although phage display remains the most common technique for *in vitro* selection, water-in-oil *in vitro* compartmentalization (IVC) [37] has emerged as a promising and robust alternative for selecting or screening for catalytic function. Recent implementations have included selection for a *Hae*III methyltransferase with altered sequence specificity [38] and the description of water-in-oil-in-water double emulsions that are readily amenable to screening via fluorescence-activated cell sorting [39]. IVC has also been used with DNA display libraries, in which biotinylated DNA templates encode streptavidin fusion proteins [40]. Improvements in this method make it a useful complement to phage display for identifying peptide ligands [41]. Finally, Griffiths and Tawfik adapted their IVC system to identify functional phosphotriesterase variants from a microbead-displayed library [42•]. In an outcome similar to most genetic selections, all 35 variants analyzed had different sequences, and in spite of activities ranging from <1% to 180% that of the wild-type enzyme, the parental sequence was not observed. This observation emphasizes the strong analogy between compartmentalization in the cell and by IVC.

In an extension of the ribosome display and mRNA–protein fusion systems [43], McGregor and co-workers described CIS display [44•]. Their approach exploits the ability of the RepA protein to bind exclusively to the DNA from which it has been expressed. A library variant fused to the N terminus of RepA therefore remains tightly associated with its encoding DNA and the stability issues of the RNA-based techniques are avoided. The fidelity of this *cis*-activity obviates the requirement for any compartmentalization and facilitates transcription and translation

entirely *in vitro*, potentially giving access to very large libraries.

Concluding remarks: quality over quantity

The most heartening developments in the field over the past year have been ideological, not methodological. Inspection of Table 1 makes apparent the trend away from simply constructing the largest libraries possible towards identifying variants from smaller, higher quality libraries. As discussed above, improvements in predictive algorithms will facilitate increasingly reliable *in silico* library prescreening. Further — and in contrast to the approach of the 'blind watchmaker' that typifies most results to date — two noteworthy papers indicate a new willingness to dissect more fundamental aspects of directed evolution in carefully defined systems.

Rothman and Kirsch have investigated the evolution of tyrosine aminotransferase (TATase) from a narrow-specificity aspartate aminotransferase (AATase) ancestor [45••]. Their study compared directly the outcomes of rational redesign [46] and DNA shuffling. Summarizing the similarities and differences between the two enzymes using Venn diagrams provided a particularly elegant way of investigating 'what makes a TATase'. As expected, mutations in the set $AAT - TAT$ (i.e. residues conserved in AATases but not in TATases) were heavily over-represented in evolved variants. The results also provided an illustration of how directed evolution can be employed to elucidate pathways of natural evolution. In this case they provided further evidence for a model in which the evolution of one specialized enzyme from another, be it in nature or in the laboratory, proceeds via broad-specificity intermediates [47,48].

Finally, Matsumura and co-workers applied the theory and terminology of classical genetics in presenting the first rigorous side-by-side comparison of methods for generating molecular diversity [49•]. In addition to their discussion of how sexual and asexual evolutionary approaches lead to distinct progeny, their data demonstrated that even small changes in the screening protocol could alter the experimental outcome.

In summary, we expect that the powerful combination of *in silico* and *in vitro* methods for directed evolution will further accelerate the successful manipulation and tailoring of biocatalysts. Equally exciting, these tools will herald an era in which the untapped potential of directed evolution is applied to fundamental questions of macromolecular structure–function relationships and evolutionary biology.

Update

Liu and co-workers have recently described a modified version of their nonhomologous random recombination (NRR) method for nucleic acid evolution [50], enabling its use for protein-encoding genes [51]. Their novel approach utilized DNA hairpins during gene fragment ligation, affording some control over the reassembly process. Protein NRR on a chorismate mutase model system yielded functional hybrids containing dramatically reordered and recombined parental sequences.

Acknowledgements

The authors wish to thank Monica Gerth for her critical reading of this manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Brakmann S: **Discovery of superior enzymes by directed molecular evolution**. *ChemBioChem* 2001, **2**:865-871.

2. Zhao H, Arnold FH: **Combinatorial protein design: strategies for screening protein libraries**. *Curr Opin Struct Biol* 1997, **7**:480-485.

3. Neylon C: **Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution**. *Nucleic Acids Res* 2004, **32**:1448-1459.

4. Arnold FH, Georgiou G (Ed): *Directed Evolution Library Construction: Methods and Protocols* Totowa, New Jersey: Humana Press; 2003.

5. • Hughes MD, Nagel DA, Santos AF, Sutherland AJ, Hine AV: **Removing the redundancy from randomised gene libraries**. *J Mol Biol* 2003, **331**:973-979.

A description of an oligonucleotide-based randomization procedure for removing library biases that arise from degeneracy in the genetic code. This 'MAX' randomization also has the advantage of decreasing target library sizes at the DNA level, with no loss of diversity at the protein level.

6. Umeno D, Hiraga K, Arnold FH: **Method to protect a targeted amino acid residue during random mutagenesis**. *Nucleic Acids Res* 2003, **31**:e91.

7. Chopra S, Ranganathan A: **Protein evolution by 'codon shuffling': a novel method for generating highly variant mutant libraries by assembly of hexamer DNA duplexes**. *Chem Biol* 2003, **10**:917-926.

8. Wong TS, Tee KL, Hauer B, Schwaneberg U: **Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution**. *Nucleic Acids Res* 2004, **32**:e26.

9. Cadwell RC, Joyce GF: **Randomization of genes by PCR mutagenesis**. *PCR Methods Appl* 1992, **2**:28-33.

10. •• Coco WM, Encell LP, Levinson WE, Crist MJ, Loomis AK, Licato LL, Arensdorf JJ, Sica N, Pienkos PT, Monticello DJ: **Growth factor engineering by degenerate homoduplex gene family recombination**. *Nat Biotechnol* 2002, **20**:1246-1250.

A novel method for synthetic recombination *in vitro*, in which all polymorphisms from a variety of mammalian epidermal growth factor genes were encoded in three 'top-strand' and two 'bottom-strand' oligonucleotides. Regions of overlap between top and bottom strands facilitated partial homoduplex formation, allowing top strand synthesis to be completed by gap filling and ligation. The result was completely unbiased randomization, independent of the tightness of linkage between parental polymorphisms.

11. • Ness JE, Kim S, Gottman A, Pak R, Krebber A, Borchert TV, Govindarajan S, Mundorff EC, Minshull J: **Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently**. *Nat Biotechnol* 2002, **20**:1251-1255.

Designed, overlapping oligonucleotides encoding all the genetic diversity in 15 *Bacillus* subtilisin genes were assembled in a modified DNA shuffling reaction. As with [10••] and [12•], the synthetic shuffling protocol makes all shuffled progeny equally likely regardless of linkage, effectively increasing library diversity.

12. Zha D, Eipper A, Reetz MT: **Assembly of designed**
• **oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution**. *ChemBioChem* 2003, **4**:34-39.
A near-identical approach to [11•], with the addition of some new (non-parental) diversity as a consequence of the oligonucleotide design strategy.

13. Patrick WM, Firth AE, Blackburn JM: **User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries**. *Protein Eng* 2003, **16**:451-457.

14. Ikeuchi A, Kawarasaki Y, Shinbata T, Yamane T: **Chimeric gene library construction by a simple and highly versatile method using recombination-dependent exponential amplification**. *Biotechnol Prog* 2003, **19**:1460-1467.

15. Song JK, Chung B, Oh YH, Rhee JS: **Construction of DNA-shuffled and incrementally truncated libraries by a mutagenic and unidirectional reassembly method: changing from a substrate specificity of phospholipase to that of lipase**. *Appl Environ Microbiol* 2002, **68**:6146-6151.

16. Lee SH, Ryu EJ, Kang MJ, Wang E-S, Piao Z, Choi YJ, Jung KH, Jeon JYJ, Shin YC: **A new approach to directed gene evolution by recombined extension on truncated templates (RETT)**. *J Mol Catal, B Enzym* 2003, **26**:119-129.

17. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH: **Molecular evolution by staggered extension process (StEP) *in vitro* recombination**. *Nat Biotechnol* 1998, **16**:258-261.

18. Ostermeier M, Shim JH, Benkovic SJ: **A combinatorial approach to hybrid enzymes independent of DNA homology**. *Nat Biotechnol* 1999, **17**:1205-1209.

19. Lutz S, Ostermeier M, Benkovic SJ: **Rapid generation of incremental truncation libraries for protein engineering using α-phosphothioate nucleotides**. *Nucleic Acids Res* 2001, **29**:e16.

20. Hiraga K, Arnold FH: **General method for sequence-independent**
• **site-directed chimeragenesis**. *J Mol Biol* 2003, **330**:287-296.
Computer-aided combinatorial engineering of β-lactamases, using the authors' SCHEMA algorithm [29]. The seamless fusion of predicted protein fragments from TEM-1 and PSE-4 via type II restriction endonucleases resulted in homology-independent recombination with high numbers of crossovers. Functional hybrids were identified.

21. O'Maille PE, Bakhtina M, Tsai MD: **Structure-based combinatorial protein engineering (SCOPE)**. *J Mol Biol* 2002, **321**:677-691.

22. Lutz S, Ostermeier M, Moore GL, Maranas CD, Benkovic SJ: **Creating multiple-crossover DNA libraries independent of sequence identity**. *Proc Natl Acad Sci USA* 2001, **98**:11248-11253.

23. Lutz S, Fast W, Benkovic SJ: **A universal, vector-based system for nucleic acid reading-frame selection**. *Protein Eng* 2002, **15**:1025-1030.

24. Kawarasaki Y, Griswold KE, Stevenson JD, Selzer T, Benkovic SJ,
• Iverson BL, Georgiou G: **Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers**. *Nucleic Acids Res* 2003, **31**:e126.
A simple but effective method for addressing the original SCRATCHY protocol's limited crossover numbers. Portions of parental ITCHY libraries containing crossovers are amplified before shuffling, raising the probability of incorporating such fragments during subsequent reassembly and increasing the mean number of crossovers per sequence.

25. Moore GL, Maranas CD: **Computational challenges in combinatorial library design for protein engineering**. *AIChE J* 2004, **50**:262-272.

26. Dwyer MA, Looger LL, Hellinga HW: **Computational design of a Zn^{2+} receptor that controls bacterial gene expression**. *Proc Natl Acad Sci USA* 2003, **100**:11255-11260.

27. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D: **Design of a novel globular protein fold with atomic-level accuracy**. *Science* 2003, **302**:1364-1368.

28. Maheshri N, Schaffer DV: **Computational and experimental analysis of DNA shuffling**. *Proc Natl Acad Sci USA* 2003, **100**:3071-3076.

29. Voigt CA, Martinez C, Wang ZG, Mayo SL, Arnold FH: **Protein building blocks preserved by recombination**. *Nat Struct Biol* 2002, **9**:553-558.

30. Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, Wang ZG, Arnold FH: **Library analysis of SCHEMA-guided protein recombination**. *Protein Sci* 2003, **12**:1686-1693.

31. Moore GL, Maranas CD: **Identifying residue-residue clashes in**
• **protein hybrids by using a second-order mean-field approach**. *Proc Natl Acad Sci USA* 2003, **100**:5091-5096.
A detailed predictive framework offering valuable guidance for the identification and quantification of residue-residue clashes in protein hybrids. Calculations on two model systems, dihydrofolate reductase and glycinamide ribonucleotide formyltransferase, demonstrated the accuracy of the simulations.

32. Saraf MC, Moore GL, Maranas CD: **Using multiple sequence correlation analysis to characterize functionally important protein regions**. *Protein Eng* 2003, **16**:397-406.

33. Saraf MC, Horswill AR, Benkovic SJ, Maranas CD: **FamClash: a method for ranking the activity of engineered enzymes**. *Proc Natl Acad Sci USA* 2004, **101**:4142-4147.

34. Suel GM, Lockless SW, Wall MA, Ranganathan R: **Evolutionarily conserved networks of residues mediate allosteric communication in proteins**. *Nat Struct Biol* 2003, **10**:59-69.

35. Arnold FH, Georgiou G (Ed): *Directed Enzyme Evolution: Screening and Selection Methods* Totowa, New Jersey: Humana Press; 2003.

36. Cesaro-Tadic S, Lagos D, Honegger A, Rickard JH, Partridge LJ,
•• Blackburn GM, Plückthun A: **Turnover-based *in vitro* selection and evolution of biocatalysts from a fully synthetic antibody library**. *Nat Biotechnol* 2003, **21**:679-685.
An excellent example of functional selection based on single-turnover chemistry. The authors sought novel phosphatases from a phage-displayed antibody library. When incubated with immobilized 2-difluoromethylphenyl phosphate as the substrate, proteins that cleaved the phosphate ester bond generated a reactive intermediate, which immediately captured the catalyst and its associated phage particle. Phage recovery led to the identification of several novel functional phosphatases.

37. Tawfik DS, Griffiths AD: **Man-made cell-like compartments for molecular evolution**. *Nat Biotechnol* 1998, **16**:652-656.

38. Cohen HM, Tawfik DS, Griffiths AD: **Promiscuous methylation of non-canonical DNA sites by *Hae*III methyltransferase**. *Nucleic Acids Res* 2002, **30**:3880-3885.

39. Bernath K, Hai M, Mastrobattista E, Griffiths AD, Magdassi S, Tawfik DS: ***In vitro* compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting**. *Anal Biochem* 2004, **325**:151-157.

40. Doi N, Yanagawa H: **STABLE: protein-DNA fusion system for screening of combinatorial protein libraries *in vitro***. *FEBS Lett* 1999, **457**:227-230.

41. Yonezawa M, Doi N, Kawahashi Y, Higashinakagawa T, Yanagawa H: **DNA display for *in vitro* selection of diverse peptide libraries**. *Nucleic Acids Res* 2003, **31**:e118.

42. Griffiths AD, Tawfik DS: **Directed evolution of an extremely fast**
• **phosphotriesterase by *in vitro* compartmentalization**. *EMBO J* 2003, **22**:24-35.
In vitro compartmentalization provides the opportunity to select not only for substrate binding, product formation and rate acceleration, but also for turnover. Testing their system on bacterial phosphotriesterases, the authors identified variants with up to 63-fold increased k_{cat}.

43. Plückthun A, Schaffitzel C, Hanes J, Jermutus L: ***In vitro* selection and evolution of proteins**. *Adv Protein Chem* 2000, **55**:367-403.

44. Odegrip R, Coomber D, Eldridge B, Hederer R, Kuhlman PA,
• Ullman C, FitzGerald K, McGregor D: **CIS display: *In vitro* selection of peptides from libraries of protein-DNA complexes**. *Proc Natl Acad Sci USA* 2004, **101**:2806-2810.
An elegant and potentially very powerful new *in vitro* selection system. On fusing the genes for a peptide library to *repA*, *in vitro* transcription/translation results in a translation product that can capture its own DNA template before polymerase and ribosome dissociation. Formed exclusively in *cis*, the tightly bound protein–DNA complex comprises

a stable link of genotype and phenotype, ready for functional selection.

45. Rothman SC, Kirsch JF: **How does an enzyme evolved *in vitro* compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase**. *J Mol Biol* 2003, **327**:593-608.
•• Detailed comparisons of rational design with DNA shuffling and of natural evolution with directed evolution. The novel use of set theory in the analysis of the parental and evolved enzymes proved particularly insightful.

46. Onuffer JJ, Kirsch JF: **Redesign of the substrate specificity of *Escherichia coli* aspartate aminotransferase to that of *Escherichia coli* tyrosine aminotransferase by homology modeling and site-directed mutagenesis**. *Protein Sci* 1995, **4**:1750-1757.

47. Matsumura I, Ellington AD: ***In vitro* evolution of β-glucuronidase into a β-galactosidase proceeds through non-specific intermediates**. *J Mol Biol* 2001, **305**:331-339.

48. Patrick WM, Weisner J, Blackburn JM: **Site-directed mutagenesis of Tyr354 in *Geobacillus stearothermophilus* alanine racemase identifies a role in controlling substrate specificity and a possible role in the evolution of antibiotic resistance**. *ChemBioChem* 2002, **3**:789-792.

49. Rowe LA, Geddie ML, Alexander OB, Matsumura I: **A comparison of directed evolution approaches using the β-glucuronidase model system**. *J Mol Biol* 2003, **332**:851-860.
• The authors used their well-defined model system to compare the results of DNA shuffling and epPCR approaches, under selection pressures of varying but controlled stringency. Their analysis offered a glimpse of the fundamental processes determining the outcomes of directed evolution experiments.

50. Bittker JA, Le BV, Liu DR: **Nucleic acid evolution and minimization by nonhomologous random recombination**. *Nat Biotechnol* 2002, **20**:1024-1029.

51. Bittker JA, Le BV, Liu JM, Liu DR: **Directed evolution of protein enzymes using nonhomologous random recombination**. *Proc Natl Acad Sci USA* 2004, **101**:7011-7016.