



ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in
Biotechnology

Accelerated protein engineering for chemical biotechnology via homologous recombination

Erik M Nordwald, Andrew Garst, Ryan T Gill and Joel L Kaar

Protein engineering has traditionally relied on random mutagenesis strategies to generate diverse libraries, which require high-throughput screening or selection methods to identify rare variants. Alternatively, approaches to semi-rational library construction can be used to minimize the screening load and enhance the efficiency by which improved mutants may be identified. Such methods are typically limited to characterization of relatively few variants due to the difficulties in generating large rational libraries. New tools from synthetic biology, namely multiplexed DNA synthesis and homologous recombination, provide a promising avenue to rapidly construct large, rational libraries. These technologies also enable incorporation of synthetically encoded features that permit efficient characterization of the fitness of each mutant. Extension of these tools to protein library design could complement rational protein design cycles in an effort to more systematically search complex fitness landscapes. The highly parallelized nature with which such libraries can be generated also has the potential to expand directed protein evolution from single protein targets to protein networks whose concerted activities are required for the biological function of interest.

Addresses

Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80309, United States

Corresponding author: Kaar, Joel L (joel.kaar@colorado.edu)

Current Opinion in Biotechnology, 2013, 24:xx–yy

This review comes from a themed issue on **Chemical biotechnology**

Edited by **Kristala LJ Prather**

0958-1669/\$ – see front matter, © 2013 Elsevier Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.copbio.2013.03.003>

Introduction

Current challenges in protein engineering

Directed evolution provides a valuable tool for engineering novel and improved protein functions for industrial use. Specifically, using conventional evolutionary strategies such as error-prone polymerase chain reaction, DNA shuffling, and site-saturation mutagenesis, enzymes with desired functions have been identified. For example, these methods have been successfully employed to evolve enzymes with improved thermostability and tolerance to pH changes, oxidation, organic solvents, and surfactants [1–4]. However, the size of the sequence

space, namely that involving combinations of mutations, for a protein that can be searched is limited using such strategies. For a single protein (200 residues), the sequence space may range from approximately 10^3 unique single mutations (i.e. linear mutation space) to upwards of 10^{260} combinatorial mutations (Table 1). Conventional evolutionary approaches may instead down select from around only 10^6 to 10^8 mutations, leaving much of the combinatorial space unexplored [5–12]. This issue is compounded when using randomly generated libraries due to large redundancies and nucleotide biases. As a result, the number of screened variants that improve the desired function is typically fewer than 1% [13,14].

In attempt to narrow the relevant sequence space for a protein to be searched experimentally, various computational methods have been developed. Computational methods may specifically aid in narrowing the sequence space of a protein by predicting the effect of mutations on protein folding and molecular recognition [7,15–18]. Such approaches have been successfully applied to the *de novo* design of novel catalytic activities for non-native retro-aldol and Kemp elimination reactions [18,19]. Additionally, computational methods may also aid in predicting beneficial mutations via analysis of existing sequence–activity data derived from empirical directed evolution studies. One such approach developed by Fox and co-workers [20] reported the development of a statistical model (ProSAR) that describes the synergistic effects of mutations based on previously measured fitness values, which can be employed to design focused combinatorial libraries for evolution. This model was employed to improve the activity of a halohydrin dehalogenase by approximately 4000-fold, demonstrating the utility of such models [21]. Nevertheless, the computational methods that most accurately predict protein function based on sequence are those that incorporate experimental data, which are dependent on generation of sequence-to-activity libraries with high diversity.

An alternative approach to evolving proteins is to systematically search the entire linear sequence space after which combinations of beneficial point mutations may be assembled. This approach, in addition to efficient screening and selection methods, requires effective strategies for creating comprehensive, rational libraries that are trackable in high-throughput. Importantly, for a single protein, the reduction of sequence space using this approach enables the use of traditional screening and selection methods. In this review, recent advances in synthetic biology that make the creation of such libraries tractable,

2 Chemical biotechnology

Table 1

Size of the linear and combinatorial mutational search space for the active site of a single protein through that of the entire *E. coli* genome. The number of required oligonucleotide microarrays to cover the entire linear mutational space by recombineering is also shown (based on a 10k oligonucleotide microarray)

Target	Number of residues	Linear mutation space	Combinatorial mutation space	Microarrays required
Active site	5.0×10^0	9.5×10^1	3.2×10^6	0.01
1 protein	2.0×10^2	3.8×10^3	1.6×10^{260}	0.38
Pathway (10 proteins)	2.0×10^3	3.8×10^4	20^{2000}	3.80
Genome-wide enzyme active sites	2.0×10^4	3.8×10^5	$20^{20\ 000}$	38.0
Genome	8.1×10^5	1.5×10^7	$20^{810\ 000}$	1.5×10^3

facilitating accelerated evolution strategies for protein engineering are discussed. The creation of vast, trackable, and rational libraries may also expand the targets of protein engineering beyond the scope of single proteins. In theory, exploitation of these advances permits the engineering of entire gene clusters and metabolic pathways of industrial relevance with unprecedented speed and success.

Solutions from multiplex recombineering

Recombineering as a tool in protein engineering

As a method for producing genetic diversity and creating sequence-to-activity mapping libraries for protein engineering, recombineering presents considerable opportunities. Recombineering involves the use of linear (single or double-stranded) DNA oligonucleotides that contain regions of homology with a target gene to introduce point mutations, insertions, or deletions in chromosomal DNA (Figure 1a). Specifically, in recombineering, a segment of the target gene is replaced by the designed oligonucleotide via seamless insertion of the oligonucleotide between the homologous regions [22,23*,24–28]. The insertion of the oligonucleotide is typically facilitated by the use of Lambda Red, RecET, or other bacteriophage-based recombination systems in *Escherichia coli* [27–29]. Notably, while eliminating the need for *in vitro* cloning techniques, recombineering may be employed to create targeted sequence-to-activity mapping libraries, including comprehensive libraries for alanine scanning, single site saturation, and entire gene saturation mutagenesis with unparalleled genetic diversity. In this way, using recombineering for gene saturation mutagenesis, the effect of all possible single amino acid substitutions across the entire open reading frame of a protein may be mapped to a desired property. Beneficial mutations may subsequently be combined in a rational manner to efficiently explore the complete combinatorial mutation space using libraries that are synthetically tractable (Figure 1b). Furthermore, the efficiency with which the combinatorial space is explored may be, in theory, enhanced via combining the power of recombineering with search algorithms such as ProSAR [20].

The utility of recombineering to rationally design novel protein variants and explore vast combinatorial spaces

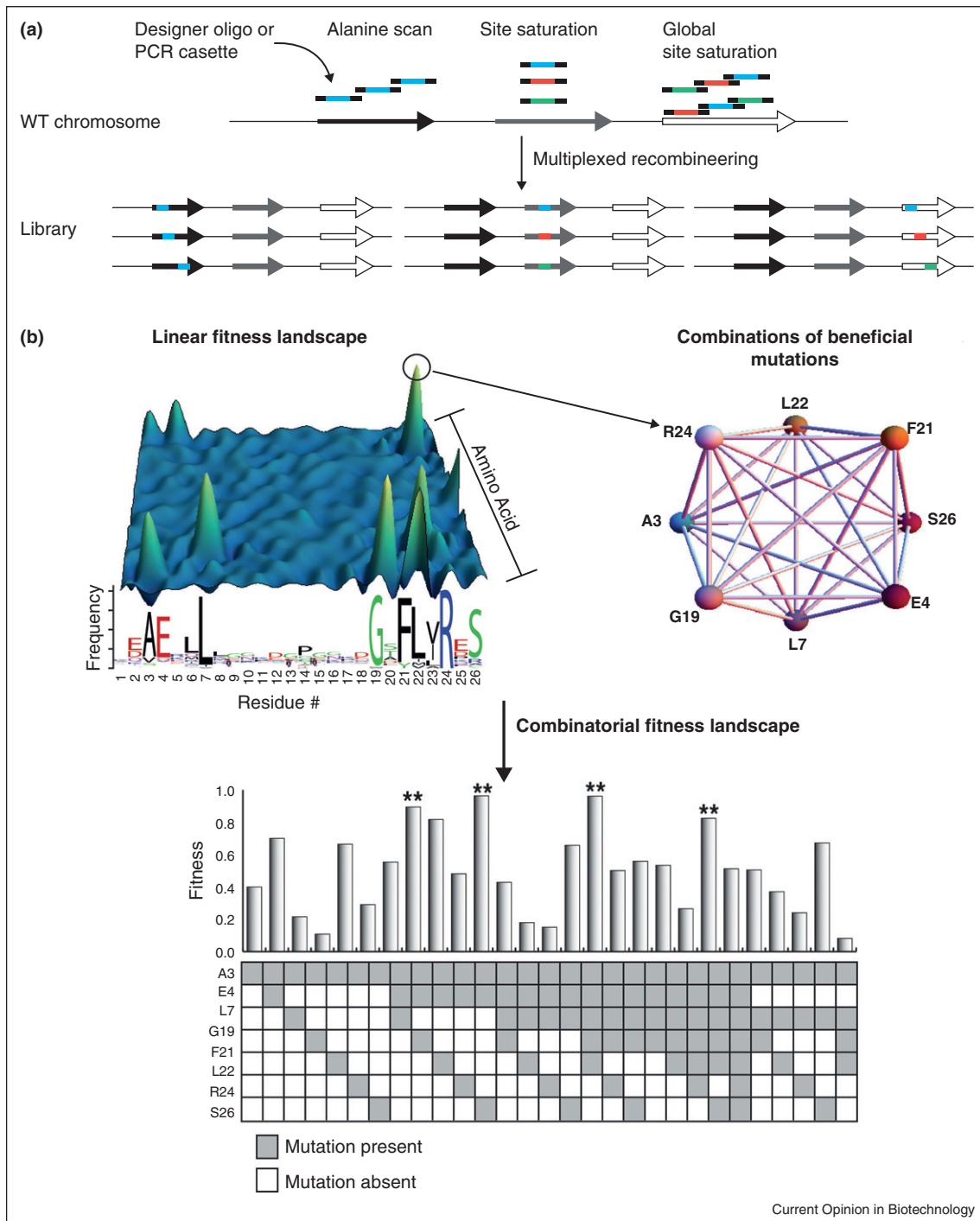
represents a significant advantage over conventional protein engineering methods. This advantage was recently demonstrated using the recombineering-based technique Multiplex Automated Genome Engineering (MAGE). Using MAGE, the expression levels of 24 proteins in *E. coli* were combinatorially optimized in parallel via the introduction of mutations in the ribosome binding sites of the corresponding genes [30]. MAGE, which was used to generate approximately 10^9 genetic variants per day, was specifically employed to engineer lycopene biosynthesis. The resulting variants produced lycopene at higher titers than previously achieved by rational engineering approaches [31]. Additionally, using MAGE, the 20-base pair T7 promoter was introduced in front of multiple genes in parallel to optimize the production of aromatic amino acids [32**]. Such strategies could also be readily extended to the parallel, combinatorial optimization of enzyme active sites or ligand binding surfaces. In this way, MAGE may have considerable utility to explore large, combinatorial sequence spaces of not only single proteins, but also multiple proteins, whole pathways, and potentially entire genomes.

While recombination has been developed primarily in *E. coli* and yeast, expansion of these tools to other organisms will expand the utility of recombineering approaches for protein engineering. The use of homologous recombination in other organisms may permit the engineering of proteins that are difficult to express in bacteria or yeast or that require posttranslational modifications. A recent study utilized standard recombineering techniques in *E. coli* to engineer fluorescent tags on an existing fosmid library of *Caenorhabditis elegans* genes [33*]. The resulting library covered a large number of the open reading frames in *C. elegans*, enabling the function of a wide variety of proteins to notably be probed *in vivo* in *C. elegans*. These tools are also currently being expanded to other bacterial species, including Gram positive bacteria [34–36].

Implications of advances in DNA synthesis and assembly for multiplex recombineering

Continual improvements in DNA synthesis and assembly have widespread implications in the use of recombineering for protein engineering. Specifically, as the cost of

Figure 1



Schematic of the use of recombineering in protein engineering. **(a)** Rational target libraries for various types of scanning mutagenesis, including alanine scanning mutagenesis, single site saturation mutagenesis, and global site saturation mutagenesis, may be rapidly constructed using designed oligonucleotides. **(b)** Once constructed, such libraries may be used to describe the linear fitness landscape of single mutations at every residue such as in the case of global site saturation mutagenesis (top left). The total combinatorial search space may be reduced by combining beneficial mutations from the linear fitness landscape (top right), which may be subsequently used to explore the synergy between positive mutations (bottom).

4 Chemical biotechnology

DNA decreases and the quantity of DNA that can be synthesized increases, larger sequence spaces may be explored by means of recombineering. Current microarray fabrication methods that take advantage of spatially directed synthesis and improved chemistry enable parallel synthesis of approximately 10^4 to 10^6 unique oligonucleotides [37–39]. Strategies for retrieval, amplification, and processing of array-containing oligonucleotides have also been established that enable the use of DNA microarrays for directed library generation [40–42,43^{••}]. As such, oligonucleotides can be tailored for rational targeting of multiple sites within promoters or open reading frames throughout the entire genome of an organism. Thus, while conventional methods for rational library generation have traditionally limited rational library size construction to 10^1 to 10^2 variants, parallel DNA synthesis and assembly methods in conjunction with recombineering facilitate the synthesis of large, targeted libraries on a laboratory timescale (i.e. days to weeks).

With the ability to synthesize large, rational libraries for protein engineering, global site saturation libraries may be constructed for a wide range of feasible targets. As mentioned previously, such targets may span the size of single proteins up to every gene in a modest-sized biosynthetic pathway. If modifications are limited to fewer sites such as those that comprise the active site of an enzyme or promoter regions, the number of feasible protein targets may be expanded to a genome-wide scale. The size of necessary libraries to map all possible single point mutations (i.e. linear mutation space) for a few residues up to every gene in the entire *E. coli* genome is depicted in Table 1. For comparison, the size of the combinatorial mutation space, which remains synthetically intractable even for a single protein, is shown side-by-side. The sheer size of the combinatorial space for even small targets highlights the need for efficient methods for traversing the mutation space of proteins. By limiting the combinatorial space to that of positive mutations found from searching the linear sequence space, the most relevant mutation landscapes may be directly explored.

Ultimately, the ability to optimize the function of pathways that require the concerted activities of many genes to be precisely tuned may have far reaching consequences in synthetic biology. For example, commercially valuable products such as biofuels and secondary metabolites are often produced via heterologous pathways in *E. coli* due to the genetic tractability of this organism [44]. Optimization of such pathways can be a laborious process in which many iterations of directed evolution are required to address particular aspects of the pathway such as gene expression levels or catalytic activities of the relevant proteins. This notion is underscored in recent work involving the optimization of the production of artemisinin, an antimalarial precursor, in *E. coli* [45]. Parallel optimization of such multi-genic systems using an

approach similar to MAGE would enable rapid engineering of metabolic pathways or other complex genetic traits of interest, and allow for identification of genotypes associated with improved activities.

Multiplex tracking of sequence-to-activity relationships via recombineering

In conventional directed evolution methods, beneficial mutations must be identified postselection via sequencing the entire gene. This step often represents a rate, and cost, limiting step in such methods due to the need to isolate and sequence many clones individually. Consequently, the overall throughput of evolutionary methods and, in turn, the resulting number of beneficial mutations that may be identified is usually of the order of only 10^1 to 10^2 . This limitation is particularly inhibitory in the simultaneous evolution of multiple proteins and whole pathways in parallel, where the number of genes to be sequenced may become increasingly large. Through recombineering, this bottleneck may be circumvented via incorporation of unique DNA barcodes into the design of oligonucleotides that are used to introduce the desired mutation. By barcoding each variant in the library, beneficial mutations may be rapidly identified via sequencing of short strands of DNA only. In this way, the use of barcoded libraries enables mutations to be identified more efficiently and cheaply than with conventional libraries. More broadly, using barcoded recombineering libraries in combination with high-throughput sequencing technologies may open new avenues for evolving multiple proteins and entire biosynthetic networks inside of cells with unmatched throughput. The creation of DNA barcoded libraries for recombineering has been facilitated by the development of novel algorithms for the design of large numbers of orthogonal DNA barcode sequences such as DeLOB [46].

The use of molecular barcoding in conjunction with recombineering has recently been established as a means to track genetic modifications *in vivo*. In one demonstration from the Gill group, Warner *et al.* [43^{••}] successfully implemented DNA barcodes via recombineering to map the simultaneous effect of thousands of mutations that modulate the expression of the entire *E. coli* genome to growth under various stress conditions. Using this approach, named Trackable Recursive Multiplex Recombineering (TRMR), specific genes that confer tolerance to various toxic compounds such as β -glucoside, D-fucose, methylglyoxal, and valine were rapidly identified [43^{••}]. As an extension of this work, Sandoval and co-workers [47[•]] exploited the use of DNA barcoded libraries while also combining TRMR with a recursive, multiplex recombineering approach, similar to MAGE. The authors rapidly identified mutations in the *E. coli* genome that permit growth under different stress conditions such as the presence of cellulosic hydrolysate. Additionally, although not for recombineering, barcoded DNA libraries have also

previously been exploited in other organisms, including yeast, highlighting the potential utility of barcoded recombineering techniques for scanning the effect of mutations in proteins in alternative host systems [48,49].

Conclusions

New directions enabled by recombineering technology

Although still dependent on the use of appropriate screens or selections, advancements in recombineering enable the engineering of proteins with novel and improved functions. One area in which recombineering may have utility that is of marked interest in many communities is the evolution of enzyme activity and stability in non-natural solvent environments. Of particular interest to our group and others is enhancing the catalytic efficiency of cellulases in ionic liquids due to the remarkable miscibility of cellulose in many ionic liquids [50–54]. The long-term goal of such efforts is to exploit ionic liquids as solvents for clean and efficient biocatalytic hydrolysis of cellulose while eliminating the need for biomass pretreatment. Ultimately, this goal will necessitate the simultaneous engineering of multiple cellulase enzymes (i.e. endoglucanases, exoglucanases, β -glucosidases) in parallel that are required for cellulose hydrolysis, which is possible via multiplex recombineering. Additionally, recombineering provides a novel avenue for evolving proteins with an expanded genetic code. Specifically, using recombineering, comprehensive sequence-to-activity mapping libraries that include the mutation of residues to non-natural amino acids — for which orthogonal aminoacyl-tRNA synthetase and tRNA have been developed — may be created. While also broadening the sequence search space, the inclusion of non-natural amino acids permits the sampling of enzyme variants with bioorthogonal chemical moieties, which have unique physicochemical properties. Such moieties may include, among others, ketone, aldehyde, azide, alkyne, and halogenated groups, which are non-existent in naturally occurring amino acids. Recombineering-based approaches (i.e. MAGE) have previously been employed to engineer amber stop codons across the entire *E. coli* genome [55], which provide the basis for non-natural amino acid scanning mutagenesis.

Acknowledgements

JLK is a recipient of a US Army Research Office Young Investigator Award. EMN is supported by a Graduate Seed Scholar Fellowship from the Colorado Center for Biorefining and Biofuels.

References and recommended reading

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

- of special interest
- of outstanding interest

1. Chen KQ, Arnold FH: **Tuning the activity of an enzyme for unusual environments — sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide.** *Proc Natl Acad Sci U S A* 1993, **90**:5618–5622.
2. Chirumamilla RR, Muralidhar R, Marchant R, Nigam P: **Improving the quality of industrially important enzymes by directed evolution.** *Mol Cell Biochem* 2001, **224**:159–168.
3. Eijssink VGH, Gaseidnes S, Borchert TV, van den Burg B: **Directed evolution of enzyme stability.** *Biomol Eng* 2005, **22**:21–30.
4. Ó'Fágáin C: **Enzyme stabilization — recent experimental progress.** *Enzyme Microb Technol* 2003, **33**:137–149.
5. Cramer A, Raillard SA, Bermudez E, Stemmer WPC: **DNA shuffling of a family of genes from diverse species accelerates directed evolution.** *Nature* 1998, **391**:288–291.
6. Kazlauskas RJ, Bornscheuer UT: **Finding better protein engineering strategies.** *Nat Chem Biol* 2009, **5**:526–529.
7. Marchisio MA, Stelling J: **Computational design tools for synthetic biology.** *Curr Opin Biotechnol* 2009, **20**:479–485.
8. 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.
9. Sen S, Dasu VV, Mandal B: **Developments in directed evolution for improving enzyme functions.** *Appl Biochem Biotechnol* 2007, **143**:212–223.
10. Weinreich DM, Delaney NF, DePristo MA, Hartl DL: **Darwinian evolution can follow only very few mutational paths to fitter proteins.** *Science* 2006, **312**:111–114.
11. Yuan L, Kurek I, English J, Keenan R: **Laboratory-directed protein evolution.** *Microbiol Mol Biol Rev* 2005, **69**:373–392.
12. Seifert A, Vomund S, Grohmann K, Kriening S, Urlacher VB, Laschat S, Pleiss J: **Rational design of a minimal and highly enriched cyp102a1 mutant library with improved regio-, stereo- and chemoselectivity.** *ChemBiochem* 2009, **10**:853–861.
13. Firth AE, Patrick WM: **GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries.** *Nucleic Acids Res* 2008, **36**:W281–W285.
14. Romero PA, Stone E, Lamb C, Chantranupong L, Krause A, Miklos A, Hughes RA, Fechtel B, Ellington AD, Arnold FH, Georgiou G: **SCHEMA designed variants of human arginase I & II reveal sequence elements important to stability and catalysis.** *ACS Synth Biol* 2012, **1**:221–228.
15. Pleiss J: **Protein design in metabolic engineering and synthetic biology.** *Curr Opin Biotechnol* 2011, **22**:611–617.
16. Shen HB, Chou KC: **Ensemble classifier for protein fold pattern recognition.** *Bioinformatics* 2006, **22**:1717–1722.
17. Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, Delbecq S et al.: **A de novo protein binding pair by computational design and directed evolution.** *Mol Cell* 2011, **42**:250–260.
18. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S et al.: **Kemp elimination catalysts by computational enzyme design.** *Nature* 2008, **453**:190–195.
19. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF III, Hilvert D et al.: **De novo computational design of retro-aldol enzymes.** *Science* 2008, **319**:1387–1391.
20. Fox R, Roy A, Govindarajan S, Minshull J, Gustafsson C, Jones JT, Emig R: **Optimizing the search algorithm for protein engineering by directed evolution.** *Protein Eng* 2003, **16**:589–597.
21. Fox RJ, Davis SC, Mundorff EC, Newman LM, Gavrilovic V, Ma SK, Chung LM, Ching C, Tam S, Muley S, Grate J et al.: **Improving catalytic function by PROSAR-driven enzyme evolution.** *Nat Biotechnol* 2007, **25**:338–344.
22. Lajoie MJ, Gregg CJ, Mosberg JA, Washington GC, Church GM: **Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering.** *Nucleic Acids Res* 2012, **40**:e170.

6 Chemical biotechnology

23. Mosberg JA, Gregg CJ, Lajoie MJ, Wang HH, Church GM: **Improving lambda Red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases.** *PLoS One* 2012, **7**:e44638.
In this work, lambda Red recombination and mutation insertion efficiency were improved by inactivating the nuclease ExoVII, which provides the basis to improve the utility of recombineering.
24. Mosberg JA, Lajoie MJ, Church GM: **Lambda Red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate.** *Genetics* 2010, **186**:791-799.
25. Murphy KC: **Phage recombinases and their applications.** *Adv Virus Res* 2012, **83**:367-414.
26. Wang HH, Church GM: **Multiplexed genome engineering and genotyping methods: applications for synthetic biology and metabolic engineering.** *Methods Enzymol* 2011, **498**:409-426.
27. Ellis HM, Yu DG, DiTizio T, Court DL: **High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides.** *Proc Natl Acad Sci U S A* 2001, **98**:6742-6746.
28. Fu J, Bian XY, Hu SB, Wang HL, Huang F, Seibert PM, Plaza A, Xia LQ, Muller R, Stewart AF, Zhang YM: **Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting.** *Nat Biotechnol* 2012, **30**:440-446.
29. Alper H, Jin YS, Moxley JF, Stephanopoulos G: **Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*.** *Metab Eng* 2005, **7**:155-164.
30. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM: **Programming cells by multiplex genome engineering and accelerated evolution.** *Nature* 2009, **460**:894-898.
31. Datsenko KA, Wanner BL: **One-step inactivation of chromosomal genes in *Escherichia coli* k-12 using PCR products.** *Proc Natl Acad Sci U S A* 2000, **97**:6640-6645.
32. Wang HH, Kim H, Cong L, Jeong J, Bang D, Church GM: **Genome-scale promoter engineering by coselection mage.** *Nat Methods* 2012, **9**:591-593.
The authors report the use of an automated approach to introduce T7 promoters upstream of a 12 genes involved in the biosynthesis of aromatic amino acids. The stochastic nature of the recombineering approach generated a combinatorial set of promoter insertion sites that were screened using a colorimetric assay for indigo production. The authors suggest that the approach can also be extended to investigate epistatic interactions in gene networks.
33. Sarov M, Murray JI, Schanze K, Pozniakovski A, Niu W, Angermann K, Hasse S, Rupprecht M, Vinis E, Tinney M, Preston E et al.: **A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*.** *Cell* 2012, **150**:855-866.
In this work, the authors developed a method for investigating the temporal expression of genes during development in *C. elegans* by combining recombineering with transgene design and high-throughput sequencing. Gene expression was monitored via the incorporation of a fluorescent tag at the C-terminus of various genes through a parallel recombineering-based approach.
34. van Kessel JC, Marinelli LJ, Hatfull GF: **Innovation recombineering mycobacteria and their phages.** *Nat Rev Microbiol* 2008, **6**:851-857.
35. van Pijkeren JP, Britton RA: **High efficiency recombineering in lactic acid bacteria.** *Nucleic Acids Res* 2012, **40**:e76.
36. van Pijkeren JP, Neoh KM, Sirias D, Findley AS, Britton RA: **Exploring optimization parameters to increase ssDNA recombineering in *Lactococcus lactis* and *Lactobacillus reuteri*.** *Bioengineered* 2012, **3**:209-217.
37. Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, Cerrina F: **Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array.** *Nat Biotechnol* 1999, **17**:974-978.
38. Gao X, LeProust E, Zhang H, Srivannavit O, Gulari E, Yu P, Nishiguchi C, Xiang Q, Zhou X: **A flexible light-directed DNA chip synthesis gated by deprotection using solution photogenerated acids.** *Nucleic Acids Res* 2001, **29**:4744-4750.
39. LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, Namsaraev E, Caruthers MH: **Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process.** *Nucleic Acids Res* 2010, **38**:2522-2540.
40. Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church G: **Accurate multiplex gene synthesis from programmable DNA microchips.** *Nature* 2004, **432**:1050-1054.
41. Kosuri S, Eroshenko N, LeProust EM, Super M, Way J, Li JB, Church GM: **Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips.** *Nat Biotechnol* 2010, **28**:1295-1299.
42. Matzas M, Stahler PF, Kefer N, Siebelt N, Boisguerin V, Leonard JT, Keller A, Stahler CF, Haberle P, Gharizadeh B, Babrzadeh F et al.: **High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing.** *Nat Biotechnol* 2010, **28**:1291-1294.
43. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LBA, Gill RT: **Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides.** *Nat Biotechnol* 2010, **28**:856-862.
This paper reports the use of recombineering to map the expression of 8154 genes in *E. coli* to specific traits associated with growth by turning the expression of these genes on or off. Notably, using TRMR, a single research assistant was able to construct the entire mutational library used in this work in only a week.
44. Keasling JD: **Manufacturing molecules through metabolic engineering.** *Science* 2010, **330**:1355-1358.
45. Keasling JD: **Synthetic biology and the development of tools for metabolic engineering.** *Metab Eng* 2012, **14**:189-195.
46. Xu QK, Schlabach MR, Hannon GJ, Elledge SJ: **Design of 240,000 orthogonal 25mer DNA barcode probes.** *Proc Natl Acad Sci U S A* 2009, **106**:2289-2294.
47. Sandoval NR, Kim JYH, Glebes TY, Reeder PJ, Aucoin HR, Warner JR, Gill RT: **Strategy for directing combinatorial genome engineering in *Escherichia coli*.** *Proc Natl Acad Sci U S A* 2012, **109**:10540-10545.
This paper describes the sequential use of TRMR and recursive multiplex recombineering to understand the synergy between the expression of genes that are relevant for *E. coli* tolerance to different selective pressures. Specifically, over 108 variants with combinatorial mutations were generated by recursive multiplex recombineering that regulate expression of genes identified by TRMR.
48. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW: **Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy.** *Nat Genet* 1996, **14**:450-456.
49. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM et al.: **Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis.** *Science* 1999, **285**:901-906.
50. Engel P, Mladenov R, Wulforth H, Jager G, Spiess AC: **Point by point analysis: how ionic liquid affects the enzymatic hydrolysis of native and modified cellulose.** *Green Chem* 2010, **12**:1959-1966.
51. Peralta-Yahya P, Carter BT, Lin HN, Tao HY, Comish VW: **High-throughput selection for cellulase catalysts using chemical complementation.** *J Am Chem Soc* 2008, **130**:17446-17452.
52. Wahlstrom R, Rovio S, Suurnakki A: **Partial enzymatic hydrolysis of microcrystalline cellulose in ionic liquids by *Trichoderma reesei* endoglucanases.** *RSC Adv* 2012, **2**:4472-4480.
53. Pinkert A, Marsh KN, Pang SS, Staiger MP: **Ionic liquids and their interaction with cellulose.** *Chem Rev* 2009, **109**:6712-6728.
54. El Seoud OA, Koschella A, Fidale LC, Dorn S, Heinze T: **Applications of ionic liquids in carbohydrate chemistry: a window of opportunities.** *Biomacromolecules* 2007, **8**:2629-2647.
55. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ et al.: **Precise manipulation of chromosomes in vivo enables genome-wide codon replacement.** *Science* 2011, **333**:348-353.