ELSEVIER

Directed evolution: new parts and optimized function

Michael J Dougherty[1] and Frances H Arnold[1,2]

Constructing novel biological systems that function in a robust and predictable manner requires better methods for discovering new functional molecules and for optimizing their assembly in novel biological contexts. By enabling functional diversification and optimization in the absence of detailed mechanistic understanding, directed evolution is a powerful complement to 'rational' engineering approaches. Aided by clever selection schemes, directed evolution has generated new parts for genetic circuits, cell–cell communication systems, and non-natural metabolic pathways in bacteria.

Addresses
[1] Division of Chemistry & Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[2] Division of Biochemistry & Molecular Biophysics, California Institute of Technology, Pasadena, CA 91125, USA

Corresponding author: Arnold, Frances H (frances@cheme.caltech.edu)

Introduction

The emerging field of synthetic biology promises useful fuels and chemicals, sensors, pharmaceuticals, therapies, and materials as well as increased understanding of natural biological systems. To rationalize the time-consuming, *ad hoc* construction of new biological systems [1], some researchers advocate efforts to 'standardize' biological parts in such a way that their behavior in novel assemblies or environments becomes more predictable [2]. The notorious complexity and context-dependency of the behavior of biological parts and their assemblies, however, make such standardization extremely challenging. It is unlikely, in fact, that biological parts can ever be fully standardized, and engineering methods that enable rapid optimization of synthetic biological systems will be very useful. Evolution is Nature's optimization algorithm; it is also the mechanism by which all the natural biological parts were generated. Here we will discuss how evolution in the laboratory — directed evolution — supports synthetic biology efforts, both in the generation of new parts and in the optimization of their assemblies.

Genome sequencing efforts have provided an enormous number of sequences that encode new biological parts. Unfortunately, a sequence only provides a hint as to its detailed biological properties, and the vast majority of these parts remain uncharacterized, much less standardized. Furthermore, these parts are all from the large but nonetheless limited set of biologically relevant ones. Many synthetic biology applications will require parts that are not biologically relevant, parts that solve human problems and not necessarily problems for the organisms that make them. An alternative to culling parts from natural sources is to construct them in the laboratory. In many cases it may be more efficient to start from a well-characterized part, for example, an enzyme or a transcription factor, and engineer it to exhibit new properties than to try to find the exact desired sequence in Nature. By tuning one property, for example ligand-binding specificity, without dramatically changing others (e.g. ability to be expressed in a particular host organism), this approach can in principle generate families of parts that differ along one desired dimension but whose behaviors along others are at least similar and are therefore more likely to be predictable. Ellis *et al.* demonstrated this very nicely by generating a library of promoters and using them to construct feed-forward networks with different, predictable input–output characteristics [3]. The diverse families of biological molecules that exist in Nature were all generated by evolution. Not surprisingly, directed evolution — iterations of mutagenesis and artificial selection or screening — can be an efficient way to generate new biological parts in the laboratory [4,5].

The next step is assembly into circuits or pathways. Unfortunately, novel arrangements or contexts can significantly affect how these molecules perform, and these effects are very difficult to predict. In addition to its role in creating new functional molecules, evolution also plays the role of Chief Editor in nature, turning disjoint paragraphs into great literature. In synthetic biology, however, natural evolution is tremendously destructive, relentlessly chipping away at painstakingly engineered networks and allowing the organism to escape control. Nonetheless, evolution can be very helpful when tamed. Thus directed evolution can be used to discover mutations that fine-tune circuits and pathways and optimize their performance, all without requiring a detailed understanding of the mechanisms by which those improvements are achieved.

This review covers recent applications of directed evolution in synthetic biology. The discussion is limited to bacterial systems, where directed evolution experiments

are relatively easy to perform. Applications have included modifying protein transcription factors, constructing synthetic genetic regulatory and cell–cell communication circuits, and engineering enzymes for non-natural metabolic pathways. Some future challenges and opportunities for directed evolution in synthetic biology are also discussed.

Engineering protein transcription factors

A significant fraction of synthetic biology research has focused on the construction of synthetic genetic circuits based on transcriptional regulators. The goal is to program cellular behavior in both time and space, which extends genetic engineering to a much richer set of behaviors and may provide a better understanding of how much more complex natural circuits work. The construction of simple biological devices in bacteria, such as toggle switches, pulse generators, and oscillators, as well as 'minimal circuits' that aim to reconstruct specific biological functions, has relied largely on a very limited number of well-studied transcription factors: LacI, TetR, AraC, LuxR, and λCI [6]. The construction of more complex circuits will require a larger set of parts to choose from. Because 'wiring' in biological circuits occurs through chemical binding specificity, these transcription factors will need to respond to diverse signaling molecules and activate or repress transcription at diverse DNA sequences. While there are a few reports of novel functions that have been engineered rationally into transcription factors [7], designing a protein that can bind a chemical signal, transmit that signal into a conformational change that alters DNA binding and regulates transcription is extremely challenging. A very high degree of modularity has significantly simplified the problem for eukaryotic signaling pathways [8,9]; many bacterial transcription factors, however, do not exhibit the same degree of modularity and appear far more difficult to modify.

The *E. coli* transcription factor AraC is widely used in combination with the P_{BAD} promoter for arabinose-inducible gene expression in bacteria. Unfortunately, crosstalk in the form of inhibition of the AraC–P_{BAD} system by IPTG prevents its use in combination with the popular LacI–P_{lac} system over the full range of inducer concentrations. Directed evolution improved the compatibility of these systems by increasing the AraC system's sensitivity to arabinose 10-fold, which also reduced its sensitivity to inhibition by IPTG and therefore crosstalk between the two [10]. The binding specificity of AraC has also been altered by random mutagenesis of specific residues of the binding pocket followed by FACS-based dual screening to identify variants that respond to D-arabinose and not the native effector L-arabinose [11•].

Acyl-homoserine lactone (AHL)-based quorum sensing (QS) systems from bacteria have been an important source of transcription factors for synthetic cell–cell communication circuits. These systems consist of a signal synthase, which produces a diffusible signal, and a receptor protein, which binds to the signal and activates transcription at a specific promoter. The well-studied QS transcription factor LuxR, from *Vibrio fischeri*, has been a target for directed evolution. Random mutagenesis and dual selection were used to switch the specificity of LuxR so that it responds to decanoyl-homoserine lactone rather than the native 3-oxo-hexanoyl-homoserine lactone signal [12]. This work demonstrated the utility of dual selection for identifying variants of transcription factors that could respond to new signals and no longer to the native ones. LuxR sensitivity to its native signal was also increased using directed evolution [13]. In addition, these authors demonstrated how artificial positive feedback loops can increase the effective sensitivity of LuxR-based circuits. The signal synthase component of QS systems has also been engineered: the substrate specificity of the RhlI synthase was modified to increase the production of the non-native signal hexanoyl-homoserine lactone [14].

These successes notwithstanding, efforts to direct the evolution of protein transcription factors have all been limited to conferring the ability to respond to new signaling molecules that are closely related to the native signals. Large changes in a signal that elicits a regulatory response have not been achieved, at least yet. DNA-binding specificity appears to be similarly difficult to modify in these proteins. No code for DNA binding has been deduced. One recent study reported some success in using comparative genomic data to guide changes in DNA-binding specificity, but unanticipated activation between noncognate protein–operator pairs was also frequently found [15]. In our experience, significant changes in ligand-binding and DNA-binding functions are difficult to obtain, at least in the LuxR family of transcriptional activators. There may be inherent limitations to transcription factor engineering, particularly when ligand binding, DNA binding, and interaction with RNA polymerase and other transcription factors are coupled through protein conformational changes. Mutations that alter one property often have negative effects on another, and there may simply be fewer single mutations that lead to measurable improvements in overall function when subfunctions are coupled.

Protein engineers may be able to glean useful information about choosing evolvable scaffolds by looking at natural evolutionary history. A recent analysis of genome sequence data, for example, showed uneven evolutionary expansion of different families of transcription factors along the human lineage: the number of zinc finger proteins increased at several stages, while the number of helix–loop–helix proteins has not significantly changed since originating in metazoans [16]. Evolution and functional diversification is clearly much easier in modular

systems. In any case, much more work will have to be done to develop synthetic systems that can respond to diverse ligands and regulate gene expression at diverse DNA sites. In addition to their utility for constructing complex regulatory and cell–cell communication circuits, such proteins will be useful as biosensors and in the regulation of novel metabolic pathways.

Construction of synthetic genetic circuits

The construction of synthetic gene circuits is hampered by our inability to accurately predict both the behavior of components placed in a new biological context and overall system behavior, which can be highly nonlinear and dependent on stochastic fluctuations, or 'molecular noise'. While natural evolution readily takes care of such design problems, the explosive growth in mutational possibilities with increasing sequence length and our limited ability to screen or select for desired functions make this a challenge for directed evolution. Combining mathematical modeling and directed evolution can greatly reduce this combinatorial complexity [17]. Systematic perturbation of system components and characterization of system behavior can allow the construction of a good predictive model with a relatively small number of perturbations. Model analysis then allows identification of the component(s) whose modification is(are) most likely to lead to the desired function, significantly reducing the search space for directed evolution (Figure 1). Alternatively, having a library of tuned components already at hand allows rapid fine-tuning of circuit function, particularly with the aid of a validated mathematical model of the system [3].

In an early example of the directed evolution approach, random mutagenesis of one protein component rescued a circuit that was nonfunctional because of mismatches in the activities or concentrations of chemicals that 'wire' one component to the next [18]. A similar approach was used recently to generate a genetic AND gate using LuxR and LacI [14]. Ribosome binding sites (RBS) are good targets for tuning mismatches in component activities. In a nice example, a modular AND gate was constructed using two inputs to drive expression of an amber suppressor tRNA and a T7 RNA polymerase gene modified to contain two amber codons [19]. The initial construct did not function as an AND gate, but saturation mutagenesis of the RBS and screening the resulting clones identified functional circuits. Targeting the RBS greatly reduces the experiment complexity — this RBS library contained only ~128 variants.

Directed evolution is likely to be extremely useful for identifying functional circuits and tuning their performance. Fluorescent protein gene expression reporters allow reasonably simple functional screening [11•], and selections based on antibiotic resistance are also generally accessible [12].

Figure 1
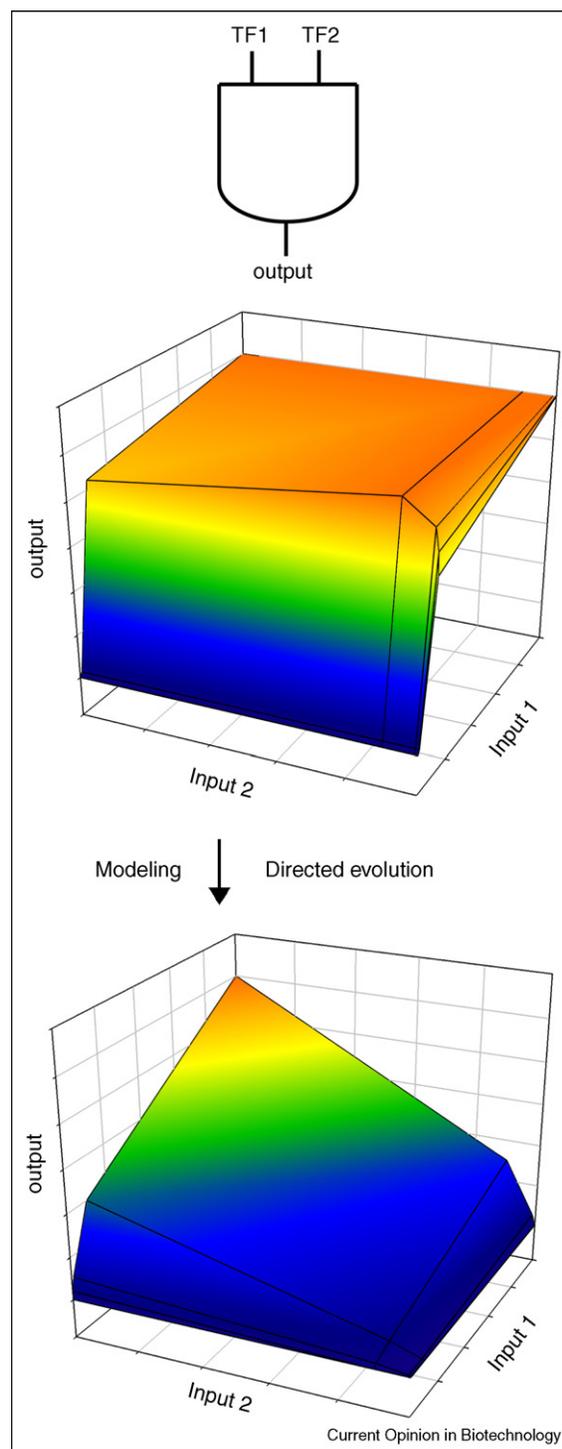
Synthetic gene circuits frequently do not function as designed because of mismatches in component properties. Mathematical modeling combined with directed evolution can rapidly tune circuit performance. For example, the sensitivities of two transcription factors to individual inputs can be tuned to generate a functional AND gate from a nonfunctional circuit.

4 **Systems and synthetic biology**

Figure 2
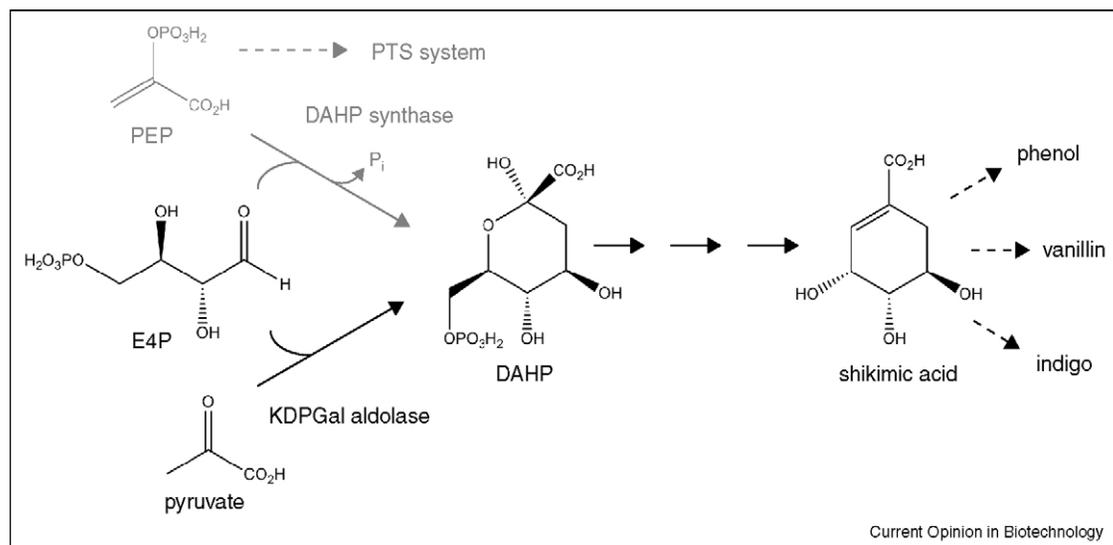
Directed evolution can be used to generate novel metabolic routes to useful chemicals. A phosphoenolpyruvate (PEP)-independent route (shown in black) to the key intermediate, 3-deoxy-D-*arabino*-heptulosonic acid 7-phosphate (DAHP), was engineered by using directed evolution to enhance a weak DAHP-synthesizing activity present in 2-keto-3-deoxy-6-phosphogalactonate (KDPGal) aldolase [23•]. The original route using DAHP synthase is shown in gray.

Engineering non-native metabolic pathways

An exciting challenge for synthetic biology is establishing and optimizing non-native, and sometimes non-natural, metabolic pathways in various host microorganisms for the production of fuels, bulk chemicals, specialty chemicals, and pharmaceuticals. New molecular biology tools and fast, inexpensive DNA synthesis methods represent significant improvements to the genetic engineering toolbox and have allowed synthetic biologists to rapidly generate, characterize, and optimize new enzymes for these applications [20,21•]. Biosynthesis offers significant advantages in terms of reaction selectivity, waste reduction and the ability to use inexpensive feedstocks such as plant sugars, but these applications require significant optimization in terms of enzyme function, timing of gene expression, and balancing of metabolic fluxes in order to be cost-competitive with chemical processes [22].

Directed evolution has been very useful for improving enzymes in engineered metabolic pathways. It has also been used to engineer new activities and establish whole new routes to chemicals with the potential for higher yields and/or fluxes. The microbial shikimate pathway is interesting for the production of a variety of chemicals traditionally produced from petroleum-derived benzene [23•]. Competition for phosphoenolpyruvate (PEP) between the phosphotransferase (PTS) system of *E. coli* and 3-deoxy-D-*arabino*-heptulosonic acid 7-phosphate (DAHP) synthase, a key enzyme in the shikimate pathway, limits the theoretical yield of shikimic acid that can be derived from glucose. A creative strategy for alleviat-

ing this limitation involves replacing DAHP synthase with an enzyme that would produce DAHP by condensing D-erythrose 4-phosphate and pyruvate rather than PEP (Figure 2). The low activity of *E. coli* 2-keto-3-deoxy-6-phosphogalactonate (KDPGal) aldolase for this key reaction was enhanced by directed evolution, where the enzyme's ability to support growth of a DAHP-deficient strain in the absence of aromatic amino acids was used to identify improved variants [23•]. Although this experiment clearly established this new pathway, the observed 60-fold increase in k_{cat}/k_M was not sufficient to improve shikimic acid production relative to an optimized strain that used DAHP synthase.

An exciting development in the field of biofuels is a route to the production of higher alcohols in *E. coli* using the 2-keto acid intermediates from branched-chain amino acid biosynthesis [24•]. Initial work demonstrated the feasibility of making isobutanol, 1-butanol, 2-methyl-1-butanol, 3-methyl-1-butanol, and 2-phenylethanol from glucose; more recent work has focused on optimizing the pathways for the production of specific alcohols. In one example, directed evolution of citramalate synthase (CimA) from *Methanococcus jannaschii*, which in combination with two *E. coli* enzymes allows the intermediate 2-ketobutyrate to be produced directly from pyruvate, improved 1-propanol and 1-butanol production 10–20-fold. The selection took advantage of the isoleucine auxotrophy of an *E. coli ikvA tdcB* mutant to identify CimA variants that improved flux through the alternate 2-ketobutyrate pathway [24•]. In contrast, a structure-

guided approach was used to engineer improved variants of 2-ketoisovalerate decarboxylase and 2-isopropylmalate synthase, which increased the production of five to eight carbon alcohols [25•]. This pair of papers nicely illustrates that there is no one-size-fits-all approach to engineering non-natural metabolic pathways; both directed evolution and rational design will be useful.

Natural products produced by nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) are a rich source of important pharmaceuticals. The chemical diversity generated by such enzyme complexes can be expanded beyond that of known natural pathways by engineering the complexes. One approach has been to swap genes between NRPSs to generate chimeras that produce new variations of the natural product. Chimeras produced in this way, however, frequently have little or no activity. But recent work has demonstrated that directed evolution can quickly improve chimera activity, even with modest library sizes and limited screening [26•]. Subjecting two chimeric NRPSs to random mutagenesis by error-prone PCR achieved ∼10-fold improvements in activity with 3 or fewer rounds of screening (10^3–10^4 clones). The availability of a suitable high-throughput assay (or selection) with which to screen mutant libraries is critical for directed evolution. Lee and Khosla engineered a strain of *E. coli* that produced small amounts of the macrolide 6-deoxyerythromycin D and then used a colony-based bioassay to identify mutant bacteria, which produced more of the antibiotic [27].

Directed evolution has been widely used to improve activities and discover novel activities in a wide range of enzymes. Any of these is potentially a part for a new non-natural metabolic pathway, limited only by the creativity of the synthetic biologist. For example, many useful natural products are glycosylated, and the identity and position of sugar moieties often have large effects on a given compounds pharmacological properties. The field has lacked general tools for diversifying the glycosylation of natural products, but Williams *et al.* described the engineering of a promiscuous glycosyltransferase that accepts a wide variety of sugar substrates [28•]. Three rounds of mutagenesis and a high-throughput screen using a fluorescent surrogate substrate were used to identify a variant of the oleandomycin glycosyltransferase capable of transferring a variety of sugar substrates to diverse relevant acceptor molecules, including macrolides, flavonoids, and coumarins. This group later used information from the screen to identify targets for saturation mutagenesis, allowing further improvements in activity on a therapeutically relevant substrate, novobiocic acid [29].

Conclusions

Directed evolution is a powerful optimization and engineering strategy for synthetic biology, but it is important to understand its limitations. Success requires that there exist an incremental pathway of beneficial mutations to the desired function and a good screening or selection strategy to identify it. Mathematical modeling can identify appropriate targets for optimization, dramatically reducing the search space for optimization. Many enzymes have proven to be remarkably evolvable, and this plasticity of function has been exploited to establish whole new biosynthetic pathways to fuels and chemicals. The more modest successes reported with engineered transcription factors may reflect inherent limitations to their evolvability, or it may simply reflect the much lower level of effort devoted to engineering them. The ease of coupling the function of transcription factors to widely available fluorescent protein reporters or to cell growth, however, offers significant advantages for their directed evolution. Biological parts evolved in bacteria can often be incorporated directly into pathways and circuits constructed in higher organisms. While directed evolution in yeast is straightforward and indeed offers some advantages [30], extension of evolutionary methods to other bacteria and higher organisms is limited mainly by the genetic tools available, particularly the ability to make large libraries of variants and by the longer time-scales for cell growth. Novel *in vivo* evolution methods may overcome some of these hurdles in the future [31].

References and recommended reading

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

• of special interest

1. Marguet P, Balagadde F, Tan C, You L: **Biology by design: reduction and synthesis of cellular components and behaviour**. *J R Soc Interface* 2007, **4**:607-623.

2. Canton B, Labno A, Endy D: **Refinement and standardization of synthetic biological parts and devices**. *Nat Biotechnol* 2008, **26**:787-793.

3. Ellis T, Wang X, Collins JJ: **Diversity-based, model-guided construction of synthetic gene networks with predicted functions**. *Nat Biotechnol* 2009, **27**:465-471.

4. Jackel C, Kast P, Hilvert D: **Protein design by directed evolution**. *Annu Rev Biophys* 2008, **37**:153-173.

5. Tracewell CA, Arnold FH: **Directed enzyme evolution: climbing fitness peaks one amino acid at a time**. *Curr Opin Chem Biol* 2009, **13**:3-9.

6. Voigt CA: **Genetic parts to program bacteria**. *Curr Opin Biotechnol* 2006, **17**:548-557.

7. Wegner SV, Boyaci H, Chen H, Jensen MP, He C: **Engineering a uranyl-specific binding protein from NikR**. *Angew Chem Int Ed Engl* 2009, **48**:2339-2341.

8. Dent CL, Lau G, Drake EA, Yoon A, Case CC, Gregory PD: **Regulation of endogenous gene expression using small molecule-controlled engineered zinc-finger protein transcription factors**. *Gene Ther* 2007, **14**:1362-1369.

6   Systems and synthetic biology

9. Bashor CJ, Helman NC, Yan S, Lim WA: **Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics**. *Science* 2008, **319**:1539-1543.

10. Lee SK, Chou HH, Pfleger BF, Newman JD, Yoshikuni Y, Keasling JD: **Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters**. *Appl Environ Microbiol* 2007, **73**:5711-5715.

11. Tang SY, Fazelinia H, Cirino PC: **AraC regulatory protein**
•   **mutants with altered effector specificity**. *J Am Chem Soc* 2008, **130**:5267-5271.
The authors use saturation mutagenesis of the binding pocket of AraC and dual-screening via FACS to obtain a switch in ligand specificity from L-arabinose to D-arabinose.

12. Collins CH, Leadbetter JR, Arnold FH: **Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR**. *Nat Biotechnol* 2006, **24**:708-712.

13. Sayut DJ, Niu Y, Sun L: **Construction and engineering of positive feedback loops**. *ACS Chem Biol* 2006, **1**:692-696.

14. Kambam PK, Eriksen DT, Lajoie J, Sayut DJ, Sun L: **Altering the substrate specificity of RhlI by directed evolution**. *Chembiochem* 2009, **10**:553-558.

15. Desai TA, Rodionov DA, Gelfand MS, Alm EJ, Rao CV: **Engineering transcription factors with novel DNA-binding specificity using comparative genomics**. *Nucleic Acids Res* 2009.

16. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM: **A census of human transcription factors: function, expression and evolution**. *Nat Rev Genet* 2009, **10**:252-263.

17. Haseltine EL, Arnold FH: **Synthetic gene circuits: design with directed evolution**. *Annu Rev Biophys Biomol Struct* 2007, **36**:1-19.

18. Yokobayashi Y, Weiss R, Arnold FH: **Directed evolution of a genetic circuit**. *Proc Natl Acad Sci U S A* 2002, **99**:16587-16591.

19. Anderson JC, Voigt CA, Arkin AP: **Environmental signal integration by a modular AND gate**. *Mol Syst Biol* 2007, **3**:133.

20. Heinzelman P, Snow CD, Wu I, Nguyen C, Villalobos A, Govindarajan S, Minshull J, Arnold FH: **A family of thermostable fungal cellulases created by structure-guided recombination**. *Proc Natl Acad Sci U S A* 2009, **106**:5610-5615.

21. Bayer TS, Widmaier DM, Temme K, Mirsky EA, Santi DV, Voigt CA:
•   **Synthesis of methyl halides from biomass using engineered microbes**. *J Am Chem Soc* 2009, **131**:6508-6515.
Great example of creative pathway engineering, using total gene synthesis to make and test many natural versions of a key enzyme.

22. Keasling JD: **Synthetic biology for synthetic chemistry**. *ACS Chem Biol* 2008, **3**:64-76.

23. Ran N, Frost JW: **Directed evolution of 2-keto-3-deoxy-6-**
•   **phosphogalactonate aldolase to replace 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase**. *J Am Chem Soc* 2007, **129**:6130-6139.
A laboratory-evolved enzyme establishes a novel shikimate pathway that bypasses a yield-limiting step. This is an excellent demonstration of how protein engineering can enable creative metabolic pathway engineering.

24. Atsumi S, Liao JC: **Directed evolution of *Methanococcus***
•   ***jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli***. *Appl Environ Microbiol* 2008, **74**:7802-7808.
Directed evolution of citramalate synthase (CimA) in an engineered pathway to 1-propanol and 1-butanol improved production 10–20-fold. A clever selection for improved variants used complementation of isoleucine auxotrophy, where isoleucine synthesis required the newly established pathway to 2-ketobutyrate.

25. Zhang K, Sawaya MR, Eisenberg DS, Liao JC: **Expanding**
•   **metabolism for biosynthesis of nonnatural alcohols**. *Proc Natl Acad Sci U S A* 2008, **105**:20653-20658.
Structure-guided protein engineering of two enzymes in a non-natural pathway increased the production of five to eight-carbon alcohols.

26. Fischbach MA, Lai JR, Roche ED, Walsh CT, Liu DR: **Directed**
•   **evolution can rapidly improve the activity of chimeric assembly-line enzymes**. *Proc Natl Acad Sci U S A* 2007, **104**:11951-11956.
This paper demonstrates that chimeric NRPSs with poor activity can rapidly be improved ∼10-fold with relatively small libraries and limited screening.

27. Lee HY, Khosla C: **Bioassay-guided evolution of glycosylated macrolide antibiotics in *Escherichia coli***. *PLoS Biol* 2007, **5**:e45.

28. Williams GJ, Zhang C, Thorson JS: **Expanding the promiscuity of**
•   **a natural-product glycosyltransferase by directed evolution**. *Nat Chem Biol* 2007, **3**:657-662.
These authors use a fluorescent surrogate substrate to screen for natural product glycosyltransferases with broadened substrate profiles. This work provides a starting point for the generation of natural products with diverse and novel glycosylation patterns.

29. Williams GJ, Goff RD, Zhang C, Thorson JS: **Optimizing glycosyltransferase specificity via ''hot spot'' saturation mutagenesis presents a catalyst for novobiocin glycorandomization**. *Chem Biol* 2008, **15**:393-401.

30. Roodveldt C, Aharoni A, Tawfik DS: **Directed evolution of proteins for heterologous expression and stability**. *Curr Opin Struct Biol* 2005, **15**:50-56.

31. Wang L, Jackson WC, Steinbach PA, Tsien RY: **Evolution of new nonantibody proteins via iterative somatic hypermutation**. *Proc Natl Acad Sci U S A* 2004, **101**:16745-16749.