

Synthetic biology: putting synthesis into biology

Jing Liang,^{1†} Yunzi Luo^{1†} and Huimin Zhao^{1,2,3,4,5*}

The ability to manipulate living organisms is at the heart of a range of emerging technologies that serve to address important and current problems in environment, energy, and health. However, with all its complexity and interconnectivity, biology has for many years been recalcitrant to engineering manipulations. The recent advances in synthesis, analysis, and modeling methods have finally provided the tools necessary to manipulate living systems in meaningful ways and have led to the coining of a field named synthetic biology. The scope of synthetic biology is as complicated as life itself—encompassing many branches of science and across many scales of application. New DNA synthesis and assembly techniques have made routine customization of very large DNA molecules. This in turn has allowed the incorporation of multiple genes and pathways. By coupling these with techniques that allow for the modeling and design of protein functions, scientists have now gained the tools to create completely novel biological machineries. Even the ultimate biological machinery—a self-replicating organism—is being pursued at this moment. The aim of this article is to dissect and organize these various components of synthetic biology into a coherent picture. © 2010 John Wiley & Sons, Inc.

WIREs Syst Biol Med 2011 3 7–20 DOI: 10.1002/wsbm.104

INTRODUCTION

The field of synthetic biology lies at the interface of many different biological research areas, such as functional genomics, protein engineering, chemical biology, metabolic engineering, systems biology, and bioinformatics. Not surprisingly, synthetic biology means different things to different people, even to leading practitioners in the field.¹ To avoid possible confusion for the reader, here we would define it as ‘deliberate design of improved or novel biological systems that draws on principles elucidated by biologists, chemists, physicists, and

engineers’. It is true that scientists have been attempting to design biological systems for decades. However, synthetic biology has become a field of its own only recently, mostly driven by the advances in systems biology and the development of new powerful tools for DNA synthesis and sequencing. Synthetic biology has broad applications in medical, chemical, food, and agricultural industries. In addition to practical applications, synthetic biology also aims to increase our understanding of basic life sciences.

There are numerous reviews on synthetic biology.^{2–11} In this article, we will focus on recent advances in the synthetic biology field, mostly from the past 2 years. We will first introduce a number of new synthetic biology methods developed for synthesis, analysis, and modeling of biological systems. We will then illustrate how these new tools have been used to address various applications in synthetic biology at different levels, including the molecular, pathway or network, cell, and multi-cell levels (Figure 1). Finally, we will offer our perspectives on the challenges and possible future development in synthetic biology.

†These authors contributed equally to this article.

*Correspondence to: zhao5@illinois.edu

¹Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

²Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA

³Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL, USA

⁴Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

⁵Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

DOI: 10.1002/wsbm.104

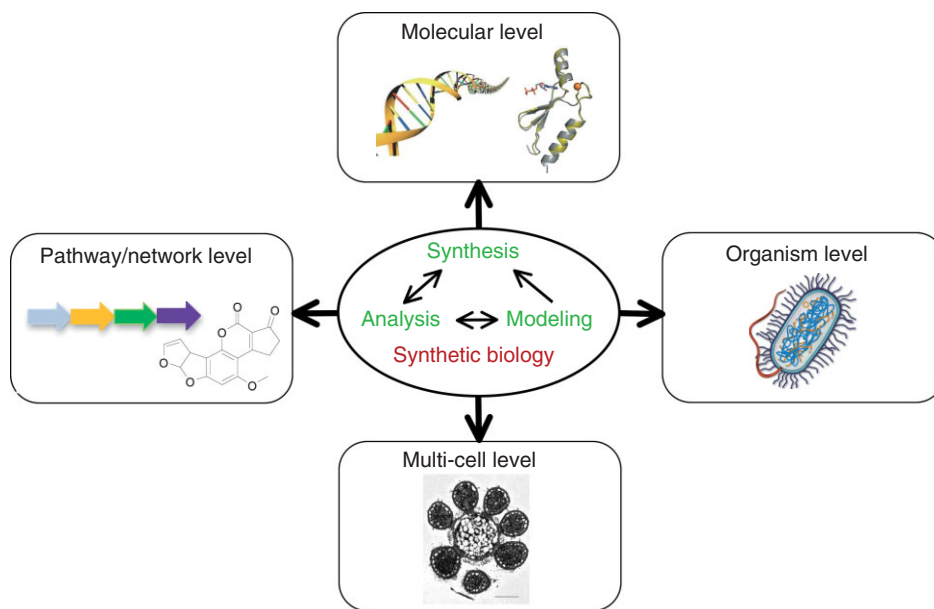


FIGURE 1 | Overview of synthetic biology. This scheme describes the relationship among the various methods used in synthetic biology and different applications of synthetic biology on different levels.

METHODS IN SYNTHETIC BIOLOGY

Synthesis Tools

Tools for synthesizing and modifying a broad range of biological entities such as DNA, proteins, pathways, organelles, viruses, and genomes in an efficient and cost-effective manner are the bedrock of synthetic biology. In recent years, a number of new and powerful tools were developed for low-cost synthesis of ever-increasing sizes of DNA and efficient modification of proteins, pathways, and genomes.

DNA Synthesis

Unlike traditional recombinant DNA technologies such as DNA cloning, chemical synthesis of DNA enables scientists to rationally design any new DNA sequence. Commercial DNA synthesis services capable of synthesizing tens of kb of DNA are now readily available.¹² Almost any gene can be mail-ordered simply by sending its sequence to a DNA synthesis company, such as Blue Heron, Geneart, DNA 2.0, and GenScript. The convenience of this approach allows rapid generation of genes, elimination of restriction sites or undesirable RNA secondary structures, and optimization of codon for gene expression. DNA synthesis tools are used in most synthetic biology applications, and they are essential in the creation of artificial genomes and customization of biosynthetic pathways.

The need for accessing larger DNA molecules has greatly stimulated the development of new DNA

synthesis technologies. Between 1970 and 2008, the size of DNA synthesized has increased from 75 to 582,970 bp.¹³ The current record was set by the J. Craig Venter Institute (JCVI) in which the 583 kb *Mycoplasma genitalium* genome was constructed from 101 pieces of DNA fragments, each 5–6 kb in length.¹⁴ Initially, synthesis was achieved in two steps: the first step involves multiple *in vitro* enzymatic recombination to create six large pieces of DNA, while the second step takes advantage of the *in vivo* yeast recombination mechanism to produce the final synthetic genome (Figure 2). Later, the entire genome assembly process was simplified by directly assembling 25 overlapping DNA fragments in a single step in yeast.¹⁵ A similar method called DNA assembler was developed by Shao and coworkers to assemble a biochemical pathway.¹⁶ As proof of concept, a biochemical pathway consisting of eight genes (a total of ~19 kb) was assembled in a single step in yeast, resulting in a recombinant strain that could utilize D-xylose to produce a nutraceutical zeaxanthin. Because it is a bottom-up approach, this method offers unmatched versatility and flexibility in pathway engineering.

Protein Engineering

Although *de novo* protein synthesis remains a challenge, many tools have been developed to modify naturally occurring proteins to create variants with desired properties. Broadly speaking, there are two protein engineering approaches—rational design and

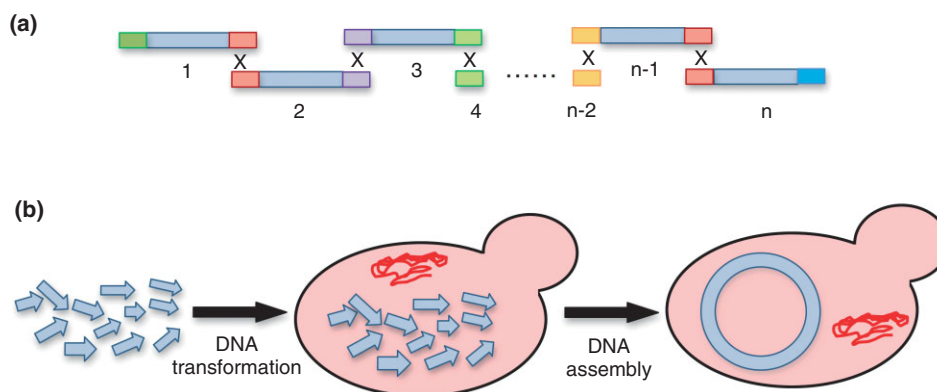


FIGURE 2 | Synthesis of large DNA molecules in yeast. (a) Yeast homologous recombination mechanism. DNA fragments sharing an overlap region at 3'- and 5'-ends with the neighboring DNA fragments can be assembled into a single larger DNA molecule. (b) Construction of a synthetic *M. genitalium* genome. Twenty-five different overlapping DNA segments (blue arrows, 17–35 kb each) composing the genome were co-transformed into yeast followed by assembly of the entire genome in a single step.

directed evolution.¹⁷ Rational design is often difficult because of incomplete knowledge of protein structure, function, and dynamics, whereas directed evolution is more suited for improving preexisting protein functions.

Directed evolution mimics the Darwinian evolution process in the test tube, involving repeated cycles of generating genetic diversity by random mutagenesis or gene recombination, and screening or selecting functionally improved or novel protein variants. Directed evolution has been successfully used to tailor a wide variety of protein properties such as activity, stability, and selectivity.^{18–22} However, directed evolution has its own limitations too. For example, the use of microorganisms in directed evolution to link the protein activity with its genetic code imposes an upper limit on the protein variant library size that can be sieved through. This limit is generally lower than 10^{10} , which is the potential library size that would be generated if a typical protein (300 amino acids) were to be diversified with three randomly introduced point mutations. Thus, much effort has been devoted to combine the advantages of both directed evolution and rational design in the past few years.

One notable approach is the Iterative Saturation Mutagenesis (ISM) method, which comprises iterative cycles of saturation mutagenesis at rationally chosen sites that are individually randomized and screened.^{23,24} There are also other alternative combinatorial methods for multi-site saturation mutagenesis. One is the Incorporating Synthetic Oligonucleotides *via* Gene Reassembly (ISOR),²⁵ which uses mutagenic primers to partially saturate multiple target positions in a protein sequence. Another is the One-pot Simple methodology for CAssette Randomization

and Recombination (OSCARR),²⁶ which allows advantageously randomizing DNA fragments with high GC content that are barely mutated by conventional error-prone PCR due to polymerase biases. Other methods include Overlap-Primer-Walk Polymerase Chain Reaction (OPW-PCR)²⁷ and rapid Site Directed Domain Scanning Mutagenesis (SDDSM).²⁸

Computational protein design has also been combined with directed evolution to create proteins with novel functions. A fine example of such strategy is the creation of a Kemp elimination enzyme through computational design followed by directed evolution optimization.²⁹ Briefly, Röthlisberger and coworkers first decided on the mechanism of catalysis, and then, with the aid of computational tools, placed catalytic residues at their exact positions to maximize transition state stabilization. The resulting rationally designed enzymes showed poor activity, but it was significantly improved by directed evolution.

Pathway Engineering

Although *de novo* synthesis of a biochemical pathway is possible using the above-mentioned DNA synthesis tools, most efforts focus on construction and optimization of an existing biochemical pathway either in a native or in a heterologous host. Pathway engineering can be designed rationally by mixing and matching well-known modular parts and modulating gene expression through various control mechanisms.

However, the design at the pathway level is concerned not only with including the necessary biological parts such as promoters, genes, and proteins, but also with optimizing the expressed functionality of those parts. Failure to balance the flux in the synthetic pathway will result in a bottleneck and the accumulation of intermediates.³⁰

One way to balance the flux is through transcription optimization of the various genes in the pathway. By fusing a library of promoters to the various enzymes in the isoprenoid production pathway, Pitera and coworkers have managed to engineer a flux-balanced pathway with improved yield and reduced metabolic burden.³⁰ These well-characterized families of transcription regulators have emerged as powerful tools in metabolic engineering as they allow rational coordination and control of multigene expression, thereby decoupling pathway design from construction.^{31,32} Pathway engineering tools are most directly relevant in biochemical production at the pathway level.

Flux balancing can also be achieved through protein co-localization. For example, Dueber and coworkers have recently developed a method where protein scaffolds with modular interaction domains are used to physically link pathway enzymes together.³³ The resulting enzyme co-localization limits the loss of intermediates to competing pathways. At the same time, it enables the direct control of metabolic flux by adjusting the number of interaction domains on the scaffold, thereby adjusting the enzyme complex composition. In addition to transcription and localization control, regulatory parts such as ribosome binding sites (RBSs), riboswitches, and operator–regulator pairs can also be redesigned and engineered to help solve problems involved in synthetic metabolic pathways.³⁴

Genome Engineering

The increasing capability to generate and manipulate large DNA molecules has opened a new era for genome engineering. Depending on purpose, editing of genomes can be carried out by a wide variety of approaches, such as constructing a genome *de novo*,¹⁴ eliminating unstable DNA elements,³⁵ combining one genome with another genome,³⁶ reorganizing genome components,³⁷ and whole genome shuffling.³⁸ Most recently, a new powerful approach called Multiplex Automated Genome Engineering (MAGE) was developed for large-scale programming and evolution of cells.³⁹ In this method, a pool of degenerate oligonucleotides is cyclically introduced into various cells within the population, to generate sequence diversity across the chromosome by allelic replacement. The entire process can be automated to facilitate rapid and continuous generation of a diverse set of genetic changes, such as mismatches, insertions, and deletions. Because it can simultaneously target many locations on the chromosome for modification in a single cell or across a population of cells, MAGE can be used to produce combinatorial genomic

diversity on a large scale. As its name implies, genome engineering is most valuable in whole-cell level applications.

Analysis Tools

Biological engineering is often a data-driven iterative process. Traditionally, biological engineers have relied on the standard analytical techniques of molecular biology, microbiology, and genetics to provide the necessary data. However, recent advancements in the ‘omic’ profiling have changed the landscape and provided the modern synthetic biologists with powerful tools to enquire the systematic effects of an engineering effort. Here we will highlight a few new technologies for analyzing biological systems at the genomic, proteomic, and metabolomic levels. These ‘omic’ tools are especially valuable in pathway, whole-cell, and multi-cell applications.

Genomics Tools

Microarray is a well-established method to make mRNA level comparison among a large number of genes. By analyzing a collection of genes simultaneously, scientists have the capability to investigate the systematic response of a genetic or environmental change. Microarray techniques have been widely employed for examination of simultaneous changes in the expression of large numbers of genes in response to experimental manipulation or environmental variations.^{40,41} Our ability to make genotype–phenotype correlations has been improved by the developments in genomics technologies including high-throughput sequencing and DNA microarrays.

While knowing the relative abundance of a transcript is a useful piece of information, it says nothing about the mechanism that controls the relative abundance. It is gratifying to know that the need for a genomic level characterization of transcription control has been addressed by a method called Chromatin Immunoprecipitation (ChIP).⁴² ChIP is a technique in which a protein of interest is selectively immunoprecipitated from a chromatin preparation to determine the DNA sequences associated with it. ChIP has been widely used to map the localization of posttranslationally modified histones, histone variants, transcription factors, or chromatin modifying enzymes on the genome or on a given locus.⁴³ Recently, a new method called Chromatin Immunoprecipitation followed by sequencing (ChIP–seq) was developed for genome-wide profiling of DNA-binding proteins, histone modifications or nucleosomes, and genome alignment.^{44,45} The advantages of ChIP–seq lie in

the higher resolution, lower noise, and greater coverage than its array-based predecessor ChIP–chip method. Of note, since ChIP–seq experiments typically generate large amount of data, powerful computational tools are needed to uncover biological mechanisms.⁴⁶

Proteomics Tools

The field of proteomics is relatively new compared to that of genomics. While it is relatively easy to characterize a genome and its mRNA expression profile, DNA and mRNA are not the actual molecules that perform the function. Protein level expression can be decoupled from mRNA level expression due to translational control, posttranslational modification, and localization. This is the motivation behind the development of the proteomics tools.

Proteomics can be regarded as the identification and quantification of all the expressed gene products of a cell type, tissue, or organism.⁴⁷ The central analytical technique for proteomics research is mass spectrometry.⁴⁸ One of the most important but also most challenging technical tasks in proteomics is the quantification of differences between two or more physiological states of a biological system.⁴⁹ Mass spectrometry-based quantification methods always employ differential stable isotope labeling to create a specific mass tag that can be recognized by a mass spectrometer and at the same time provide the basis for quantification.^{50,51} In contrast, label-free quantification approaches aim to correlate the mass spectrometric signal of intact proteolytic peptides or the number of peptide sequencing events with the relative or absolute protein quantity directly.

An alternative approach ideally suited for the global analysis of protein function is activity-based proteomics. Activity-based protein profiling has emerged as a powerful chemoproteomic strategy to decipher the physiological functions of enzymes through the use of chemical probes that target large groups of enzymes that share active-site features.⁵² The key components of this method are the chemical probes, which are activity-based and contain two main elements: (1) a reactive group to label mechanistically related enzymes in an active site-directed manner, and (2) a reporter tag to visualize, enrich, and identify probe labeled enzymes. This strategy has been successfully applied to many enzyme classes.⁵³

Metabolomics Tools

Metabolomics can be seen as the complementary tool to proteomics. While proteomics can give information on what and how much of proteins are present, metabolomics can investigate what those proteins are

doing to a chemical molecule. Using a typical chemical pathway diagram for analogy, proteomics fill in the enzyme names, while metabolomics will fill in the chemical names.

Metabolomics refers to the analysis of the metabolome, i.e., the metabolic profile of a biological system.⁵⁴ Metabolomic tools have been used to debug synthetic metabolism for industrial-scale microbial production of a variety of natural and novel chemicals. One recent example is the use of transcriptional and metabolite profiling to determine the biochemical interactions of an engineered pathway and the endogenous metabolic network.⁵⁵

Modeling Tools

Due to the combinatorial nature of biological systems, the solution space of any potential biological problem can be immense. As mentioned earlier in protein engineering, just three random mutations in a 300 amino acid protein can already result in 10^{10} possible combinations. Part of the motivation behind mathematical modeling is to gain the ability to explore those combinations *in silico*. Because many of the synthetic biological systems are relatively small and largely independent of evolutionary contexts, they can be represented with mathematical models. A mathematical model can be used to describe synthetic constructs and explain how biological phenotypic complexity emerges as a result of well-defined biomolecular interactions.⁵⁶ Mathematical modeling can dramatically increase the speed of the design process as well as reduce the cost of development. Cooling and coworkers described the development of an online repository of Standard Virtual Biological Parts (SVBPs)—mathematical model components, which can be any biological building blocks.⁵⁷ These SVBPs can be downloaded, extended, and recombined to aid the *in silico* design of synthetic biological systems.

The scale of biological systems to be modeled can be extended from genes or proteins to pathways. From Metabolite to Metabolite (FMM) is a web-based tool for construction of a variety of possible enzymatic pathways from an input metabolite to an output metabolite, which is useful in pathway engineering.⁵⁸ Zhang and coworkers developed a new method for genome-wide association studies, which employs mixed models to improve the ability to detect phenotype–genotype associations in the presence of population stratification and multiple levels of relatedness.⁵⁹ Once the software for analyzing the models is available, the use of mixed models with increasingly large data sets can be developed.

In addition, computational modeling can be carried out on a genome scale. Lehner and coworkers described methods that can be used to map genetic interactions in yeast and *Caenorhabditis elegans*, and investigate how these networks can provide information to study human disease.⁶⁰ The mechanistic interpretation of genetic interaction networks can be used to understand gene functions and methods that have been developed to predict genetic interactions on a genome-wide scale can also be studied using modeling. In addition, genome-scale metabolic models were established for a wide variety of organisms to connect genome-derived biochemical information and metabolic phenotypes.^{61,62} Such genome-scale models can provide a solid interpretative framework for experimental data related to metabolic states, and enable simple *in silico* experiments with whole-cell metabolism.

APPLICATIONS IN SYNTHETIC BIOLOGY

In the past few years, synthetic biology has permeated across many scales of application—from proteins, the basic functional unit of life, to life itself. The focus of this section is the utility of the application rather than the experimental details. We will start by looking at synthetic biology at the molecular level, where basic functions are conferred to polymers (proteins and nucleic acids). Then, we will look at the pathway and network level, where basic functional units are integrated into pathways and networks that perform higher level functions. Moving higher, we will reach the cellular level, where an intricate biochemical network gives rise to a self-replicating entity—life. Lastly, we will end with synthetic biology at the multi-cellular level, which may also be called synthetic ecology.

Molecular Level

Synthetic biology at the molecular level seeks to alter the fundamental properties of biological macromolecules such as nucleic acids and proteins in the hope that some of these properties will give rise to molecules Nature has not produced.

Non-Natural Nucleic Acids

DNA is the informational storage molecule at the foundation of the central dogma and Nature employs a triplet code of A, T, C, G to represent all the complexity of life. There are three fundamental design parameters: (1) the choice of nucleotide, (2) the number of nucleotides, and (3) the triplet code.

Engineering efforts in the area of non-natural nucleic acids have focused on the first two areas. When engineering non-natural nucleic acids, three design criteria need to be met, including the need to have some type of base pairing, compatibility with the natural or engineered DNA polymerase, and compatibility with RNA polymerase.⁶³

One of the objectives of using non-natural nucleic acids is to improve the stability of polynucleotides. The phosphodiester backbone of DNA is the attack site for enzymatic cleavage. By replacing the phosphate group of the nucleotides, the most common degradation site is thus eliminated. Examples include phosphorothioate,⁶⁴ boranophosphate,⁶⁵ and phosphonate.⁶⁶ Another objective of using non-natural nucleic acids is to expand the number of base pairs. Engineered base pairs such as isoC:isoG⁶⁷ and Z:P⁶⁸ are compatible with natural DNA polymerase and almost orthogonal to the ATCG system of base pairing.

Non-Natural Amino Acids

In nature, a large diversity of protein functions has been achieved through the incorporation of only 20 amino acids. It is thought that by expanding the repertoire of the natural translation machinery, novel protein properties, and functions may be achieved.⁶⁹ However, in order to maintain orthogonality to the host system, there are not many codons that can be used to incorporate non-natural amino acids. This is an area where the non-natural nucleic acids and non-natural amino acids act synergistically to create a whole meaningfully expanded genetic code.⁷⁰ Protein engineering is an especially valuable synthesis tool in non-natural amino acids research.

In vivo site-specific incorporation⁷¹ of non-natural amino acids requires a few components. First, there needs to be an amino-acyl transferase that charges a tRNA with an unnatural amino acid and, at the same time, act orthogonally to the host system. Then, the tRNA would have to recognize a unique codon that normally does not code for any amino acid. The selected codon is most often the UAG (amber) codon, while the tRNA and amino-acyl transferase pair is taken from a heterologous source to maintain orthogonality. All of the components are engineered to work together by protein engineering, mostly via the directed evolution approach.^{69,72}

Over the years, a large number of amino acid analogs have been developed. Many of them have interesting properties such as fluorescence and photo-regulation.^{73,74} In order to maximize the potential of non-natural proteins, we need to be able to specify more than one non-natural amino acid. To this

end, a few strategies have been proposed and are currently being pursued. One of them is to expand the genetic code by introducing non-natural base pairs as mentioned earlier. Another is to use a quadruplet code for non-natural amino acids.⁷⁵ In order to create a ribosome that reads such a code, protein engineering is once again employed.

Pathway/Network Level

Synthetic biology at the pathway and network level has two main focuses, including exploring the gene circuitry design, and performing more complex transformations that cannot be accomplished in a single step. Many tools are applied for engineering at this level. For example, pathway engineering methods are used for the optimization of biosynthetic pathways, the genomic and metabolic engineering tools in combination with the ‘omic’ tools are used in the optimization of strains, and modeling can also be used to predict network behaviors.

Gene Circuits

Gene circuits draw their analogies from electronic circuits. While these biological circuits are unlikely to match the computational power of their silicon counterparts, they nevertheless offer a channel to control biological behavior in a predictable fashion. Through the design of synthetic biological circuits, we can also gain insights into the workings of natural circuits that have been refined by evolution.

The output of a gene circuit can be controlled at any of the three steps—before transcription, before translation, and after translation. Control at different points will result in different dynamic properties, e.g., controlling activity by posttranslational modifications will have a much faster response time than controlling

transcription. Due to its ease of manipulation, transcriptional regulation has been the most common mode of control, and it will be the focus for this part of the article.

Transcriptional gene circuits are typically constructed by chemically interconnecting multiple promoters, repressors, and activators. In earlier experiments, basic circuit components such as logic gates and oscillators have been constructed.^{76,77} The current effort is in the development of higher function circuits such as digital-to-analog converters, analog-to-digital converters, timers, adaptive learning networks, and decision-making circuits.⁷⁸ The mechanistic description of these advanced networks is not within the scope of this review. Here, we will use an analog-to-digital converter circuit as an example to illustrate the utility of these next generation gene circuits (Figure 3).

The motivation behind an analog-to-digital converter is noise control, and a particularly relevant application is in whole-cell sensors. When we use cells to detect molecules, we may want a ‘yes or no’ answer instead of an ‘ x relative to y ’ answer. For example, to tell if a neurological toxin is above a dangerous threshold, we would prefer a device that turns red when above threshold and colorless otherwise, as oppose to one that changes from colorless to pink to red as the concentration increases, which will require the user to differentiate between light red (below threshold) and red (above threshold). In a digital system, a variable has only two states and is therefore more tolerant to noise than in an analog system. However, the natural world is inherently analog. To interface the two systems, we need an analog-to-digital converter. Such a converter can be constructed using an array of toggle switches with increasing response threshold.⁷⁸

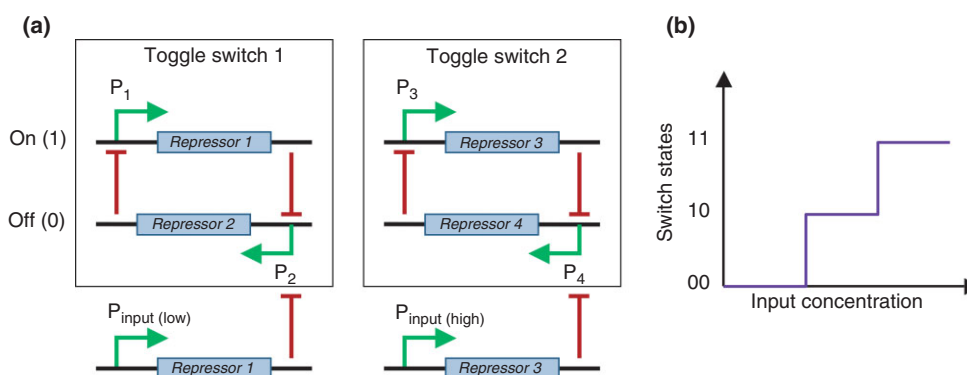


FIGURE 3 | Analog-to-digital converter. (a) In this example system, the two toggle switches are initially ‘off’, i.e., repressor 2 and repressor 4 are expressed. Separately in a sensor array, repressor 1 has an input dependent promoter with a low input threshold. When input concentration rises above its threshold, repressor 1 is expressed, switching the state of toggle switch 1 to ‘on’. Using another input dependent promoter that has a higher threshold, switch 2 is toggled ‘on’ at a higher input concentration. (b) The corresponding digital response of the example gene switches.

Other than direct applications, constructing gene circuits also allow scientists to understand the properties of nature's circuits. In a recent example, Cağatay and coworkers asked why nature picked a certain circuit design when there are others that are functionally equivalent.⁷⁹ By constructing the alternative circuit and comparing that to the wild type, they found that the wild type circuit is noisier, i.e., has more fluctuation, than the alternative. Further experiments then revealed that while the alternative circuit is more precise and is able to perform its function better in a specific environment, the noisy wild type circuit is able to perform its function over a wider range of environments. Thus in a changing environment similar to that of nature, a noisy and adaptable circuit is selected over a precise circuit.

Biochemical Production

Pathway and metabolic engineering for biochemical production is another major research area at the pathway and network level. This is distinct from the over-production of therapeutic proteins. In the latter case, a protein is the final product, and machineries already exist in cells to produce proteins from a DNA template. In biochemical production however, for example artemisinin (an anti-malarial drug), cells do not typically have the machineries for its production, and multiple enzymes need to be heterologously expressed to form a complete pathway. The unique challenges in such engineering efforts include balancing of promoter strengths,

enzyme expression and coordination of expression, and interface with native metabolism.²

The motivation behind using cells for biochemical production is different depending on the class of biochemicals produced. For specialty chemicals, the structure is sometimes so complex that chemical synthesis becomes prohibitively expensive, difficult, or both. In this case, the motivation of using the biological production route is to circumvent difficult chemistry. Well known for their specificity, enzymes can perform transformations that may take many more steps to accomplish through chemical synthesis. For bulk chemicals, where chemical synthesis is viable, the motivation of using the biological production route is cost and sustainability. When properly interfaced with native metabolism, chemicals can be produced from cheap and renewable substrates such as glucose and plant biomass.

In the artemisinin production example, genes encoding mevalonate biosynthesis, farnesyl pyrophosphate biosynthesis, as well as artemisinin biosynthesis have been gathered from *Saccharomyces cerevisiae*, *Escherichia coli*, and *Artemisia annua*.² The genes are assembled into two operons and transformed into *E. coli* for the production of artemisinin (Figure 4). The pathway has also been subjected to various optimizations, such as the codon optimization of amorphaadiene synthase (ADS), protein engineering, and expression balancing of individual genes. Through this series of engineering efforts, the production of amorphaadiene was improved by more than

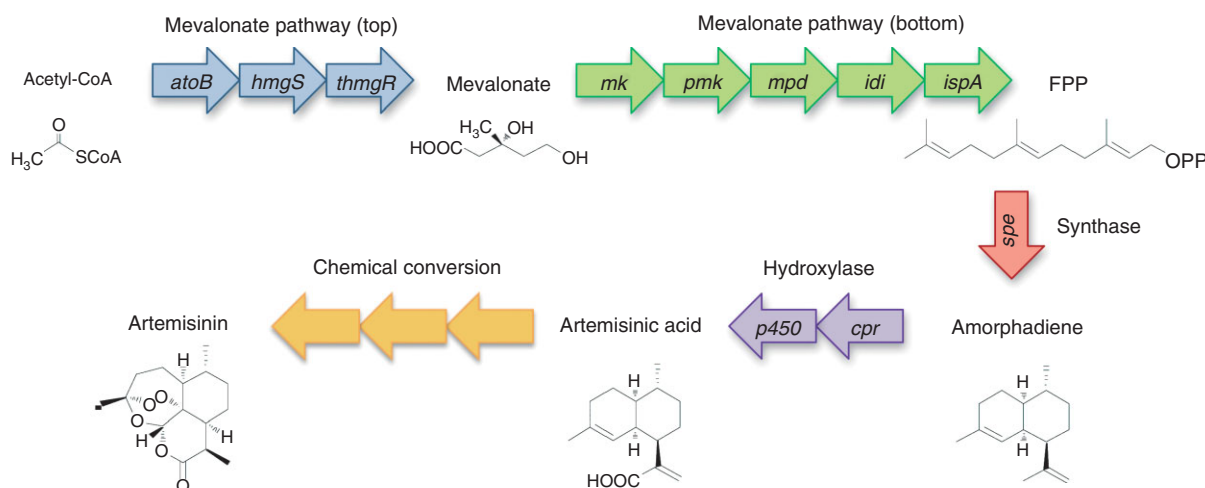


FIGURE 4 | Artemisinin biosynthetic pathway. The mevalonate-based FPP biosynthetic pathway has been assembled from *S. cerevisiae* (HMG-CoA synthase, *hmgS*; N-terminally truncated HMG-CoA reductase, *thmgR*; mevalonate kinase, *mk*; phosphomevalonate kinase, *pmk*; and mevalonate diphosphate decarboxylase, *mpd*) and *E. coli* (acetoacetyl-CoA synthase, *atoB*; IPP isomerase, *idi*; and FPP synthase, *ispA*), and expressed in *E. coli*. The pathway is assembled in two operons, one is responsible for converting acetyl-CoA to mevalonate, and the other is responsible for converting mevalonate to FPP. Further introduction of amorphaadiene synthase (*ads*), oxidase (*p450*), and redox partner (*cpr*) allows the strain to convert FPP to artemisinic acid which can be chemically converted to artemisinin.

10 million-fold to 25 g/L in a fed-batch reactor.⁸⁰ Artemisinin can be chemically converted from either amorphaadiene or artemisinic acid.

Ethanol is a good example of a bulk chemical that is currently being produced in a biochemical process. Yeast, a natural producer of ethanol, is the logical biological catalyst. However, yeast has its limitations—it is unable to break down cellulose and hemicellulose, major components of biomass, and it is unable to utilize pentoses such as D-xylose and L-arabinose, which come from the breakdown of hemicellulose. Thus, engineering yeast to break down cellulose and hemicellulose and utilize all the sugars from their breakdown is currently a major research effort in the area of bioenergy.^{81,82} In order to endow yeast with the above functions, multiple enzymes and pathways need to be inserted and optimized. To break down cellulose to glucose, at least three classes of enzymes need to be introduced—namely cellobiohydrolase, endoglucanase, and beta-glucosidase. To break down hemicellulose to its component sugars (including D-xylose and L-arabinose), at least four classes of enzymes need to be introduced—namely xylanase, xylosidase, glucuronidase, and arabinosidase. All these enzymes can be either secreted or surface-anchored, and they need to be expressed in the right proportion for the best synergy. Furthermore, to increase the utilization efficiency of the pentose sugars, pentose transporters may be introduced too.

Whole-Cell/Organism Level

Over the course of evolution, cells have developed systems that are robust and efficient for their niche environment and for their survival. This means that they are unlikely to selflessly submit themselves to human manipulation and human benefits. Synthetic biology at the whole-cell level aims to one day develop organisms from the ground up that are designed for the sole purpose of serving people—be it as a microbial factory, as an environmental mediator, or to fight diseases.

To accomplish this higher goal, scientists first need to understand what is life. To address the problem, scientists have taken two main approaches—top down and bottom up. In the top down approach, an organism is stripped to its bare minimum genome in an effort to understand what is required to create a self-replicating organism. In the bottom up approach, genes are synthesized and inserted into a ‘shell’ to create a living organism. Synthetic biology at the whole-cell level and above often employs the full array of tools currently available.

Blattner and coworkers developed a series of reduced-genome *E. coli* strains with up to 15% of its genome deleted.³⁵ Overall, over 700 non-essential genes, mobile DNA elements, and cryptic virulence genes were deleted. Surprisingly, the resulting strain has comparable growth and protein production characteristics to the wild type strain. Furthermore, the genome reduction had led to beneficial properties such as high electroporation efficiency and high recombinant gene stability.

The genome reduced *E. coli* has over 3600 genes, and is still too complex to understand completely. Scientists at the J. Craig Venter Institute have taken the approach further on *M. genitalium* by reducing its already tiny genome of 482 genes to 382 genes. The number of essential genes is greater than expected, and there are about 100 of these with unknown functions.⁸³ In addition, as mentioned above, they have also taken a bottom up approach with the *M. genitalium* genome. The whole 583 kb genome of wild type *M. genitalium* has been chemically synthesized and then assembled in yeast.¹⁴ Attempts to reboot the aforementioned *M. genitalium* genome have been, for a while, challenging.⁸⁴ However, they have recently succeeded in rebooting a chemically synthesized *M. mycoides* genome transplanted into *M. capricolum*. The rebooted cell is capable of self-replication and exhibits the expected phenotype of *M. mycoides*.⁸⁵

While the chemical synthesis of a self-replicating organism is still underway, the chemical synthesis of a host-dependent organism has already been achieved. DNA and RNA viruses, with their small genomes (8–30 kb) and relatively simple function, are more amenable to chemical synthesis. So far, the whole genome synthesis of poliovirus, 1918 ‘Spanish’ influenza virus, HIVcpz, coronavirus, and ΦX174 have been achieved and proven infectious.⁸⁶

One immediate use of chemically synthesized viruses is in the area of vaccine production. It is now possible to make large-scale changes to a viral genome so as to attenuate its virulence. For example, codon-deoptimization has led to a strain of virus that is phenotypically identical to its wild type cousin but is unable to replicate in the normal host because of multiple rare codon usage.⁸⁶

Szostak and coworkers have led a separate effort to create life in laboratory. The main motivation behind their work is to understand how a self-replicating organism can be formed from prebiotic chemistry—i.e., without the use of any enzymes, from simple organic molecules such as sugars, amino acids, nucleic acids, and fatty acids. Solving the mystery of life that has been created billions of

years ago is a daunting task, and their work offers a plausible physical and chemical origin of life on earth. The non-enzymatic assembly and replication of lipid vesicles has already been demonstrated, while the non-enzymatic polymerization of nucleic acid is currently being attempted.⁸⁷

Multi-Cell Level

In higher order organisms, cells have differentiated and are grouped into organs to carry out complex functions that no individual cell can perform. Synthetic biology at the multi-cellular level is based on the same ideas: communication and division of labor. This approach of engineering is also called the consortium approach.

Communication can be achieved either by the exchange of metabolites or dedicated signal molecules. For example, Balagaddé and coworkers constructed a predator and prey bacterial ecosystem where the predator causes the prey to kill itself but at the same time requires signals from the prey to survive (Figure 5).⁸⁸ The result is an expected predator and prey fluctuation in population. Communication has also been demonstrated between different domains of life, e.g., bacterial and mammalian cells.⁸⁹

Division of labor can make cellular engineering easier by reducing the metabolic load of individual

populations.⁹⁰ Consortia are well utilized in nature, one example being the cellulolytic consortia found in the guts of herbivores, where many different strains of microbes act synergistically and sometimes symbiotically to degrade cellulose.⁹¹ The same synergy has been explored in a synthetic consortium of *E. coli*, demonstrating the co-utilization of sugars with two strains each capable of using just one.⁹²

CONCLUSION

Thirty-six years after the coining of the term ‘synthetic biology’ by Waclaw Szybalski, the field has finally become a blooming and vibrant area of research. The ability to design a biological system that behaves predictably and functions superior to the natural counterpart is the dream of synthetic biologists. Groundwork is now being laid to make the dream into reality. Nevertheless, challenges across all scales abound.

According to one school of thought, to be able to come up with design rules, we first need to understand what we are designing—and we do not. We are not yet able to make a protein perform any physically possible function that we can conceive. Neither can we introduce any pathway into any organism and make it function. These abilities are the epitome of complete understanding of biological systems, and we are not there yet. We do, however, already have the ability to make any protein, form any pathway, and synthesize a modest-sized genome, which are by themselves huge accomplishments.

While some try to understand biological design from first principles, others try to understand it by trial and error. The idea behind this is that even though a process is not completely understood, by trying it out many times across many conditions, a correlation that can predict its behavior should at least be obtained. Both efforts are being fervently pursued by their respective believers.

As mentioned in this article, meaningful applications exist at every scale. However, even more can be achieved when applications are combined across the scales. For example, enzymes can first be improved for better solubility, stability, and functions; they can then be integrated into a pathway that produces a desirable product. To improve the titer and purity, genome scale engineering can be employed to optimize the production strain. And lastly, multiple strains can be developed this way to complement each other in a consortium to perform a full series of chemical transformation that no single strain can independently perform.

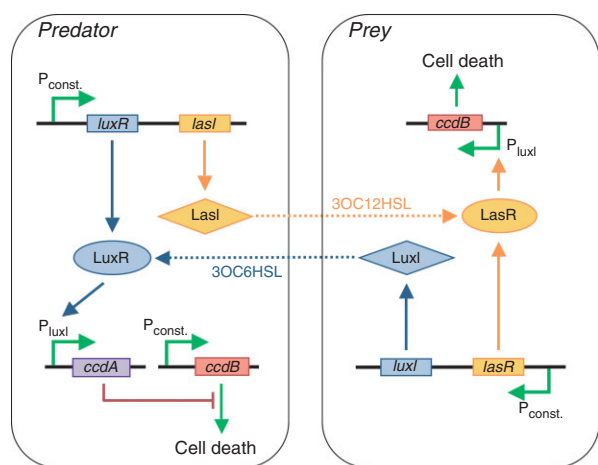


FIGURE 5 | A synthetic predator–prey ecosystem. The system consists of two engineered bacterial populations that control each other’s survival and death. CcdB is a cytotoxic protein that kills the cell and CcdA is its antidote. In the predator, CcdB is constitutively expressed, and will die off without an external signal. The antidote CcdA is expressed only when the predator receives signaling molecules (3OC6HSL) from the prey. At high enough prey concentration, the predator will survive. However, the prey also receives signaling molecules (3OC12HSL) from the predator. In the prey, CcdB is expressed in response to the predator signal. Therefore, in high enough predator concentration, the prey will die.

ACKNOWLEDGEMENTS

We thank the National Institutes of Health (GM077596), the National Academies Keck Futures Initiative on Synthetic Biology, the Biotechnology Research and Development Consortium (BRDC) (Project 2-4-121), the British Petroleum Energy Biosciences Institute, and the National Science Foundation for financial support in our synthetic biology projects. J. Liang and Y. Luo also acknowledge fellowship support from the Singapore A*STAR and National Research Foundation of Korea (NRFK) (220-2009-1-D00033), respectively.

REFERENCES

1. What's in a name?. *Nat Biotechnol* 2009, 27:1071–1073.
2. Keasling JD. Synthetic biology for synthetic chemistry. *ACS Chem Biol* 2008, 3:64–76.
3. Young E, Alper H. Synthetic biology: tools to design, build, and optimize cellular processes. *J Biomed Biotechnol* 2010, 12.
4. Mukherji S, van Oudenaarden A. Synthetic biology: understanding biological design from synthetic circuits. *Nat Rev Genet* 2009, 10:859–871.
5. Yeh BJ, Lim WA. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat Chem Biol* 2007, 3:521–525.
6. Tigges M, Fussenegger M. Recent advances in mammalian synthetic biology—design of synthetic transgene control networks. *Curr Opin Biotechnol* 2009, 20:449–460.
7. Purnick PEM, Weiss R. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* 2009, 10:410–422.
8. Benner SA, Sismour AM. Synthetic biology. *Nat Rev Genet* 2005, 6:533–543.
9. Marchisio MA, Stelling J. Computational design tools for synthetic biology. *Curr Opin Biotechnol* 2009, 20:479–485.
10. Khalil AS, Collins JJ. Synthetic biology: applications come of age. *Nat Rev Genet* 2010, 11:367–379.
11. Neumann H, Neumann-Staubitz P. Synthetic biology approaches in drug discovery and pharmaceutical biotechnology. *Appl Microbiol Biotechnol* 2010, DOI: 10.1007/s00253-010-2578-3
12. May M. Engineering a new business. *Nat Biotechnol* 2009, 27:1112–1120.
13. Mueller S, Coleman JR, Wimmer E. Putting synthesis into biology: a viral view of genetic engineering through *de novo* gene and genome synthesis. *Chem Biol* 2009, 16:337–347.
14. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA, Smith HO. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 2008, 319:1215–1220.
15. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *P Natl Acad Sci USA* 2008, 105:20404–20409.
16. Shao Z, Zhao H. DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* 2009, 37:e16.
17. Dougherty MJ, Arnold FH. Directed evolution: new parts and optimized function. *Curr Opin Biotechnol* 2009, 20:486–491.
18. Schmidt-Dannert C. Directed evolution of single proteins, metabolic pathways, and viruses. *Biochemistry-US* 2001, 40:13125–13136.
19. Rubin-Pitel SB, Zhao HM. Recent advances in biocatalysis by directed enzyme evolution. *Comb Chem High Throughput Screen* 2006, 9:247–257.
20. Shivange AV, Marienhagen J, Mundhada H, Schenk A, Schwaneberg U. Advances in generating functional diversity for directed protein evolution. *Curr Opin Chem Biol* 2009, 13:19–25.
21. Arnold FH. Design by directed evolution. *Acc Chem Res* 1998, 31:125–131.
22. Zhao HM, Chockalingam K, Chen ZL. Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr Opin Biotechnol* 2002, 13:104–110.
23. Chockalingam K, Chen ZL, Katzenellenbogen JA, Zhao HM. Directed evolution of specific receptor-ligand pairs for use in the creation of gene switches. *Proc Natl Acad Sci USA* 2005, 102:5691–5696.
24. Reetz MT, Carballeira JD. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc* 2007, 2:891–903.
25. Herman A, Tawfik DS. Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng Des Sel* 2007, 20:219–226.
26. Hidalgo A, Schliessmann A, Molina R, Hermoso J, Bornscheuer UT. A one-pot, simple methodology

- for cassette randomisation and recombination for focused directed evolution. *Protein Eng Des Sel* 2008, 21:567–576.
27. Kumar R, Rajagopal K. Single-step overlap-primer-walk polymerase chain reaction for multiple mutagenesis without overlap extension. *Anal Biochem* 2008, 377:105–107.
28. Deng QW, Luo WS, Donnenberg MS. Rapid site site-directed domain scanning mutagenesis of enteropathogenic *Escherichia coli* espD. *Biol Proced Online* 2007, 18–26.
29. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D. Kemp elimination catalysts by computational enzyme design. *Nature* 2008, 453:190–195.
30. Pitera DJ, Paddon CJ, Newman JD, Keasling JD. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab Eng* 2007, 9:193–207.
31. Bennett MR, Hasty J. Overpowering the component problem. *Nat Biotechnol* 2009, 27:450–451.
32. Ellis T, Wang X, Collins JJ. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* 2009, 27:465–471.
33. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KLJ, Keasling JD. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* 2009, 27:753–759.
34. Landrain TE, Carrera J, Kirov B, Rodrigo G, Jaramillo A. Modular model-based design for heterologous bioproduction in bacteria. *Curr Opin Biotechnol* 2009, 20:272–279.
35. Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR. Emergent properties of reduced-genome *Escherichia coli*. *Science* 2006, 312:1044–1046.
36. Itaya M, Tsuge K, Koizumi M, Fujita K. Combining two genomes in one cell: Stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc Natl Acad Sci USA* 2005, 102:15971–15976.
37. Chan LY, Kosuri S, Endy D. Refactoring bacteriophage T7. *Mol Syst Biol* 2005, 1.
38. Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WPC, del Cardayre SB. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* 2002, 415:644–646.
39. 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.
40. Forêt S, Kassahn KS, Grasso LC, Hayward DC, Iguchi A, Ball EE, Miller DJ. Genomic and microarray approaches to coral reef conservation biology. *Coral Reefs* 2007, 26:475–486.
41. Weniger M, Engelmann JC, Schultz J. Genome expression pathway analysis tool—analysis and visualization of microarray gene expression data under genomic, proteomic and metabolic context. *BMC Bioinformatics* 2007, 8:179.
42. Grandori C, Mac J, Siebelt F, Ayer DE, Eisenman RN. Myc-Max heterodimers activate a DEAD box gene and interact with multiple E box-related sites *in vivo*. *EMBO J* 1996, 15:4344–4357.
43. Collas P. The current state of chromatin immunoprecipitation. *Mol Biotechnol* 2010, 45:87–100.
44. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. *Nat Biotechnol* 2009, 27:455–457.
45. Farnham PJ. Insights from genomic profiling of transcription factors. *Nat Rev Genet* 2009, 10:605–616.
46. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009, 10:669–680.
47. Griffiths WJ, Wang Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chem Soc Rev* 2009, 38:1882–1896.
48. Domon B. Mass spectrometry and protein analysis. *Science* 2006, 312:212–217.
49. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 2007, 389:1017–1031.
50. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999, 17:994–999.
51. Mann M. Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* 2006, 7:952–958.
52. Simon GM, Cravatt BF. Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. *J Biol Chem* 2010, 285:11051–11055.
53. Cravatt BF, Wright AT, Kozarich JW. Activity-based protein profiling: From enzyme chemistry. *Annu Rev Biochem* 2008, 77:383–414.
54. Goldsmith P, Fenton H, Morris-Stiff G, Ahmad N, Fisher J, Prasad KR. Metabonomics: a useful tool for the future surgeon. *J Surg Res* 2008, 160:122–132.
55. Kizer L, Pitera DJ, Pflieger BF, Keasling JD. Application of functional genomics to pathway optimization for increased isoprenoid production. *Appl Environ Microbiol* 2008, 74:3229–3241.
56. Kaznessis YN. Computational methods in synthetic biology. *Biotechnol J* 2009, 4:1392–1405.
57. Cooling MT, Rouilly V, Misirli G, Lawson J, Yu T, Hallinan J, Wipat A. Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics* 2010, 26:925–931.

58. Chou CH, Chang WC, Chiu CM, Huang CC, Huang HD. FMM: a web server for metabolic pathway reconstruction and comparative analysis. *Nucleic Acids Res* 2009, 37:W129–W134.
59. Zhang Z, Buckler ES, Casstevens TM, Bradbury PJ. Software engineering the mixed model for genome-wide association studies on large samples. *Brief Bioinform* 2009, 10:664–675.
60. Lehner B. Modelling genotype–phenotype relationships and human disease with genetic interaction networks. *J Exp Biol* 2007, 210:1559–1566.
61. Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev* 2009, 33:164–190.
62. Mo ML, Palsson BO. Understanding human metabolic physiology: a genome-to-systems approach. *Trends Biotechnol* 2009, 27:37–44.
63. Appella DH. Non-natural nucleic acids for synthetic biology. *Curr Opin Chem Biol* 2009, 13:687–696.
64. Wang L, Chen S, Xu T, Taghizadeh K, Wishnok JS, Zhou X, You D, Deng Z, Dedon PC. Phosphorothioation of DNA in bacteria by *dnd* genes. *Nat Chem Biol* 2007, 3:709–710.
65. Shaw BR, Dobrikov M, Wang X, Wan J, He K, Lin JL, Li P, Rait V, Sergueeva ZA, Sergueev D. Reading, writing, and modulating genetic information with boranophosphate mimics of nucleotides, DNA, and RNA. *Ann NY Acad Sci* 2003, 1002:12–29.
66. Renders M, Lievrouw R, Krecmerová M, Holý A, Herdewijn P. Enzymatic polymerization of phosphonate nucleosides. *Chembiochem* 2008, 9:2883–2888.
67. Benner SA. Understanding nucleic acids using synthetic chemistry. *Acc Chem Res* 2004, 37:784–797.
68. Yang Z, Hutter D, Sheng P, Sismour AM, Benner SA. Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res* 2006, 34:6095–6101.
69. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of *Escherichia coli*. *Science* 2001, 292:498–500.
70. Hirao I, Ohtsuki T, Fujiwara T, Mitsui T, Yokogawa T, Okuni T, Nakayama H, Takio K, Yabuki T, Kigawa T, Kodama K, Yokogawa T, Nishikawa K, Yokoyama S. An unnatural base pair for incorporating amino acid analogs into proteins. *Nat Biotechnol* 2002, 20:177–182.
71. Ibba M, Hennecke H. Towards engineering proteins by site-directed incorporation *in vivo* of non-natural amino acids. *Biotechnology (NY)* 1994, 12:678–682.
72. Liu W, Brock A, Chen S, Chen S, Schultz PG. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* 2007, 4:239–244.
73. Wang Q, Parrish AR, Wang L. Expanding the genetic code for biological studies. *Chem Biol* 2009, 16:323–336.
74. Rowe L, Ensor M, Mehl R, Daunert S. Modulating the bioluminescence emission of photoproteins by *in vivo* site-directed incorporation of non-natural amino acids. *ACS Chem Biol* 2010, DOI: 10.1021/cb9002909
75. Taira H, Fukushima M, Hohsaka T, Sisido M. Four-base codon-mediated incorporation of non-natural amino acids into proteins in a eukaryotic cell-free translation system. *J Biosci Bioeng* 2005, 99:473–476.
76. Leloup JC, Goldbeter A. A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J Biol Rhythms* 1998, 13:70–87.
77. Hasty J, McMillen D, Collins JJ. Engineered gene circuits. *Nature* 2002, 420:224–230.
78. Lu TK, Khalil AS, Collins JJ. Next-generation synthetic gene networks. *Nat Biotechnol* 2009, 27:1139–1150.
79. Cadatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Süel GM. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* 2009, 139:512–522.
80. Tsuruta H, Paddon CJ, Eng D, Lenihan JR, Horning T, Anthony LC, Regentin R, Keasling JD, Renninger NS, Newman JD. High-level production of amorpho-4,11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. *PLoS ONE* 2009, 4:e4489.
81. van Zyl WH, Lynd LR, Den Haan R, McBride JE. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Adv Biochem Eng Biotechnol* 2007, 108:205–235.
82. Carere CR, Sparling R, Cicek N, Levin DB. Third generation biofuels via direct cellulose fermentation. *Int J Mol Sci* 2008, 9:1342–1360.
83. Glass JL, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA 3rd, Smith HO, Venter JC. Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* 2006, 103:425–430.
84. Marshall A. The sorcerer of synthetic genomes. *Nat Biotechnol* 2009, 27:1121–1124.
85. Gibson DG, Glass JL, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, DOI: 2010:10.1126/science.1190719
86. Wimmer E, Mueller S, Tumpey TM, Taubenberger JK. Synthetic viruses: a new opportunity to understand and prevent viral disease. *Nat Biotechnol* 2009, 27:1163–1172.
87. Budin I, Szostak JW. Expanding roles for diverse physical phenomena during the origin of life. *Annu Rev Biophys* 2010, 39:245–263.
88. Balagaddé FK, Song H, Ozaki J, Collins CH, Barnet M, Arnold FH, Quake SR, You L. A synthetic *Escherichia*

- coli* predator-prey ecosystem. *Mol Syst Biol* 2008, 4:187.
89. Brenner K, You L, Arnold FH. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol* 2008, 26:483–489.
 90. Alper H, Stephanopoulos G. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat Rev Microbiol* 2009, 7:715–723.
 91. Morrison M, Pope PB, Denman SE, McSweeney CS. Plant biomass degradation by gut microbiomes: more of the same or something new? *Curr Opin Biotechnol* 2009, 20:358–363.
 92. Eiteman MA, Lee SA, Altman E. A co-fermentation strategy to consume sugar mixtures effectively. *J Biol Eng* 2008, 2:3.