Engineering Life: A Review of Synthetic Biology

Martin M. Hanczyc*
University of Trento
  Department of Cellular,
    Computational, and Integrative
    Biology (CIBIO)
University of New Mexico
  Chemical and Biological Engineering
  Martin.hanczyc@unitn.it

Abstract Synthetic biology is a field of scientific research that applies engineering principles to living organisms and living systems. It is a field that is increasing in scope with respect to organisms engineered, practical outcomes, and systems integration. There is a commercial dimension as well, where living organisms are engineered as green technologies that could offer alternatives to industrial standards in the pharmaceutical and petroleum-based chemical industries. This review attempts to provide an introduction to this field as well as a consideration of important contributions that exemplify how synthetic biology may be commensurate or even disproportionate with the complexity of living systems. The engineerability of living systems remains a difficult task, yet advancements are reported at an ever-increasing pace.

1 Introduction

This article is intended as a perspective on the field of synthetic biology. It attempts to address the question: Can living organisms and living systems be engineered effectively towards desired outcomes? I come from the disciplines of evolution and organismal biology. When I was a student, certain technical advances in the field of recombinant DNA—or genetic engineering, as it was known at the time—began to pop up in the scientific literature and in university lectures. There was a staunch belief that, considering the complex molecular biology, intricate ecology, and long evolutionary history of living cells, our blunt instruments, however fascinating, could not direct or redirect living systems effectively in a way enabling us to call the approach engineering. It was all down to trial and error. Sometimes one got lucky and the manipulated cell behaved more or less as desired. The field has come a long way since then, with more and more sophisticated techniques being added to the toolbox. Reports of success in synthetic biology reach beyond the hopes of tackling genetic disease, and so we now must also consider futuristic supra-Darwinism [4].

2 Genetic Manipulation—Plasmids

There are two main targets for genetic manipulation in cells: the genome and extrachromosomal plasmids (see Figure 1). Traditionally plasmids as transferable elements were the target of choice. They were easy to manipulate in the laboratory, using restriction enzymes to mix and match different genes. And once a new genetic manipulation was made, they were then relatively easy to transfer into

---

 * Corresponding author.

System design: Modeling and design of genetic changes or circuits through rational design, combinatorics, or directed evolution. Increasingly, deterministic or stochastic modeling is used.

Implementation and genetic modification: Plasmids are constructed and used in cells and also in cell-free systems, or genomes are modified in cells by, for example, CRISPR or MAGE.

System integration and refinement: Often initial tests reveal that system parameters need to be modified, for example, to match input and output parameters. Feedback control can be implemented to dynamically change parameters. System noise quantified and perhaps modulated.

Outcomes: Control of gene expression by external or internal signals. Production or optimization of gene product(s).

Standardization: Biological parts defined; modularity tested; standardization of genes and outputs.

Figure 1. Summary of concepts.

cells via electroporation or other methods. For an interesting historical account of the terminology for plasmids by Joshua Lederberg see [36]. Plasmids are found naturally and are a known mechanism by which genetic information and functionality can be exchanged between organisms, for example in the development of metabolic pathways to degrade man-made toxins in the environment [34, 46].

The plasmid called pSC101 was the first cloning vector [16]. It hosted a single tetracycline resistance gene and a single restriction site for the restriction enzyme EcoRI. This single restriction site allowed for the insertion of genetic element(s) into the plasmid. The pSC101 plasmid became the first patented commercial DNA cloning vector. The "SC" stands for one of the inventors of this construct, Stanley Cohen. It was used in the first demonstration of the addition and expression of the genes from one species into another species [12].

We have developed further the use of plasmids to design and construct genetic elements and introduce them into organisms. But, as is often the case in biology, control is difficult. Due to different origins of replication, different plasmids could be maintained inside the host cells at different copy numbers. Roughly, the more copies of a gene that are present, the more the gene is expressed in the cell. This is an interesting and natural system of genetic control that is often exploited in synthetic biology.

However, the number of plasmids in a cell can drastically change over time in response to environmental changes. For example, when the genetically manipulated plasmid contains an essential antibiotic resistance gene, necessary for the survival of the organism, the copy number of the plasmid may increase 10-fold (for example, from 30 copies to 300 copies) under strong selection pressure [47].

In addition, there are costs associated with maintaining multiple copies of plasmids in cells even if these plasmids are necessary for survival. As the copy number of plasmids increases, the relative fitness of a cell (i.e., its growth rate) decreases (Figure 2) [47, 52]. This decrease in fitness is seen when for example the plasmid carries an essential antibiotic resistance gene and the antibiotic is removed from the system. The burden of carrying, copying, and maintaining extra genetic information is quantifiable.

There is another common problem with using plasmids as the carrier of genetic manipulations. They are often transient and lost from the growing cell line over time [36]. For example, Moser et al. [38] constructed logic gates with the genetic elements hosted on plasmids. It was found that the system behaved as desired, providing expected outputs based on controlled inputs. However, the system began to break down within a short time. They discovered that the breakdown was due to loss of plasmids, one of the essential plasmids being completely lost within 48 hours. This demon-
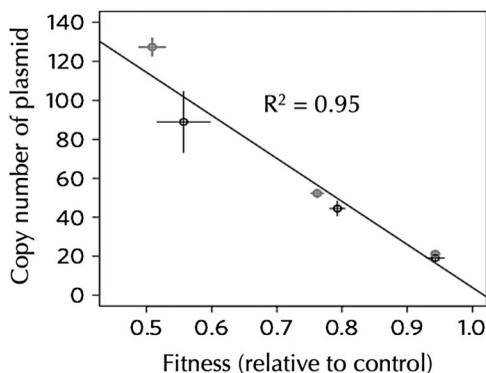
Figure 2. Estimation of how the fitness of a population of cells (estimated as the growth rate relative to a wild type control line without plasmids) varies with the copy number of the same plasmid. Figure adapted from [47].

strates the inherent instability of using plasmids as the vehicle for genetic manipulations in synthetic biology. However, due to their ease of use, they are often still employed in a number of applications. On the other hand, as tools for genomic manipulation are growing and their ease of use is increasing (e.g., CRISPR), more and more genetic manipulations in synthetic biology are on the level of the genome.

3   Genetic Manipulations—Genome

There have been several methods for genome modification, starting with Muller's bombardment of *Drosophila melanogaster* with X rays to produce more than 100 mutants in a short time [39]. There was no control over the position or type of spontaneous mutations produced by such methods. Then early successes with highly specific genetic manipulations using restriction enzymes allowed for the insertion for the first time of new genetic information into the genome of a virus [28].

Today genome editing technology has the potential to revolutionize synthetic biology and society, approaching the idea of supra-Darwinian evolution [4]. The clustered regularly interspaced short palindromic repeats (CRISPR) gene editing system is derived from a natural defense mechanism against phage and plasmid transfer in bacteria. It uses CRISPR with associated protein 9 (Cas9). With CRISPR one can edit and change the sequence of the genome in a controlled way resulting in an increase in modulation of gene expression, from a complete knockout to specific and detailed changes in the translated protein sequence and structure [30]. One can also manipulate several genome targets simultaneously in a single experiment, allowing for the elucidation of higher-order interactions and expression of complex phenotypes [17]. From an evolutionary perspective, the inclusion of the CRISPR system in the germ line can be used to support guided gene drives where wild-type target genes are selectively modified each generation, leading to highly biased inheritance in favor of the genetic modifications [5, 18, 23, 24, 27]. Recent announcements report that the genomes of human babies have been edited using this technique to specifically alter the chances of the humans to contract HIV [61]. There are many reviews about this technique [1, 56, 59], and I recommend a particular piece that strikes at the core of the excitement for CRISPR; it includes the phrases "Any idiot can do it" and "the tool has limitless possibilities" [15]. As technical advances in genetic manipulations are being presented, there is increasing pressure on society and lawmakers to devise and impose limits on what types of genetic manipulations will be allowed.

4   An Early Example of Synthetic Biology

To provide a perspective on the state of the art of genetic manipulation, we can look at a set of studies published nearly 20 years ago. The goal of this work was to use genetic manipulation in

plants so that they would produce more α-tocopherol, or vitamin E. The problem at hand was that plant oils, the main dietary source of tocopherols, typically contain α-tocopherol as a minor component with high levels of its biosynthetic precursor, γ-tocopherol. A genomics-based approach was used to find and clone the final enzyme in α-tocopherol synthesis, γ-tocopherol methyltransferase (γ-TMT), and manipulate an overexpression of γ-tocopherol methyltransferase in *Arabidopsis* seeds to shift the oil composition in favor of α-tocopherol [50].

The metabolic pathway of α-tocopherol anabolism was known at that time [51]. However, only one enzyme in the entire pathway was known for *Arabidopsis*. All of the other enzymes were not yet identified, including the target enzyme, γ-TMT. So the investigators took a comparative approach to compare what was known of the pathway in *Arabidopsis* versus the same pathway in cyanobacteria (*Synechocystis* PCC6803). Both photosynthetic organisms produce tocopherol, including α-tocopherol, and therefore γ-TMT enzyme must be present in both organisms. In addition, *Synechocystis* PCC6803 produces α-tocopherol as more than 95% of its total tocopherol, which suggests that γ-TMT is present and very active at a desired level. Can this high level of activity be reproduced in *Arabidopsis*?

To find the elusive gene in *Arabidopsis*, they started a search in its distant cousin, *Synechocystis*. Because related biosynthetic genes in bacteria are often grouped into operons, they hypothesized that such an operon might contain other genes involved in tocopherol biosynthesis, including γ-TMT. Since only one gene in the tocopherol biosynthesis pathway in *Arabidopsis* was known (the enzyme HPPDase), they searched for this known gene in the *Synechocystis* genome. They identified an open reading frame (ORF) that shared 35% amino acid sequence identity with the *Arabidopsis* HPPDase. The *Synechocystis* HPPDase gene was located within a predicted 10-gene operon. The operon included one candidate for a *Synechocystis* γ-TMT, ORF SLR0089. This candidate looked promising, as the ORF was predicted to produce a protein that shared a high degree of amino acid sequence similarity with known plant sterol methyltransferases. ORF SLR0089 also contained several structural features of a γ-TMT, including domains binding S-adenosylmethionine (SAM). They therefore hypothesized that the SLR0089 ORF encoded the *Synechocystis* γ-TMT.

The next step was to test the putative gene γ-TMT from *Synechocystis* for the production of α-tocopherol. A null mutant for the SLR0089 gene was created by replacing the wild-type gene with a nonfunctional SLR0089 gene through homologous recombination. This resulted in altering both the amount and the type of tocopherol produced: The mutant was devoid of α-tocopherol and accumulated the biosynthetic precursor γ-tocopherol, and produced less total tocopherol overall. This phenotype was consistent for a genetic variant that lacked γ-TMT activity. This null mutant of the SLR0089 gene then provided evidence that the γ-TMT gene was found, but in *Synechocystis*.

The same gene and activity had yet to be identified in the target organism, *Arabidopsis*. To identify a γ-TMT gene from higher plants, they used the *Synechocystis* γ-TMT protein sequence as a query in a search of the *Arabidopsis* expressed sequence tag (EST) database. One cDNA clone, *Arabidopsis* EST 165H5T7, showed 66% amino acid sequence similarity with *Synechocystis* γ-TMT. The 165H5T7 protein contained structural features common to γ-TMT, including SAM-binding domains. The recombinant 165H5T7 protein from *Arabidopsis* was expressed in *E. coli* and, when tested with various potential substrates, was able to produce α-tocopherol, the desired γ-TMT activity.

Now that the target γ-TMT gene in *Arabidopsis* was identified, the final step was to genetically manipulate *Arabidopsis* to produce more α-tocopherol. The target gene was inserted into a plasmid that would not only allow production of the gene in plants but would overproduce the gene product due to a strong promoter sequence. The plasmid also included a gene for antibiotic resistance to allow for selection of plants that received this new genetic material. The plasmid was then introduced into the *Arabidopsis* plants by transformation. Primary transformants were selected by antibiotic resistance and grown to maturity. Seeds were analyzed for changes in tocopherol content and composition by high-performance liquid chromatography (HPLC). Seed-specific overexpression of the *Arabidopsis* γ-TMT gene increased seed α-tocopherol levels 80-fold. Seeds of the lines overexpressing the largest amounts of γ-TMT contained 95% of their total tocopherol pool as α-tocopherol.

This long story is presented as an early example of synthetic biology in a time that predates the "-omics" and "big data" era of systems biology. At that time the genomic sequences of the species were not known, and only a few genes here and there were characterized. Such painstaking detective work to find genes and manipulate them is largely a challenge of the past. But this example does present the state of the art 20 years ago. This study was largely successful and demonstrated the potential of synthetic biology and that biological species are able to be engineered. Also, what is clarified by this example is that knowledge of the system to be manipulated is essential for success. Current research in synthetic biology has less to do with gene hunting and more to do with design, integration, and control of systems.

5   System Control and System Noise

An important aspect of synthetic biology is the demonstration of control. For example, it can be desirable that a specific gene or pathway be regulated by the addition or subtraction of an external chemical trigger. Commonly this is done by the addition of small diffusible inducers such as IPTG (isopropyl β-D-1-thiogalactopyranoside) or ATc (anhydrotetracycline). In such systems the gene of interest is situated behind a region of DNA that can be regulated by these molecules. For example, when IPTG enters a cell, it binds to the lac repressor protein, which is bound as a tetramer to the DNA of interest. The binding of IPTG to the lac repressor removes the repressor from the DNA, allowing for active transcription of the gene or genes. Of note are studies that creatively use other types of control. For example, blue light was used to induce the remote-controlled expression of insulin genes to enhance blood-glucose homeostasis in mice [57].

If one potential goal of synthetic biology is to engineer the regulated expression of a target gene, what amount of control versus noise is inherent in a typical biological system? How precise are the typical control mechanisms? A single-copy chromosomal gene with an inducible promoter was incorporated into the chromosome of *Bacillus subtilis* [43]. The gene was a commonly used reporter, the green fluorescent protein gene (GFP). They chose to integrate GFP into the chromosome itself, rather than in the form of plasmids, as variation in plasmid copy number can act as an additional and unwanted source of noise. Transcriptional efficiency was regulated by using an IPTG-inducible promoter, P_{spac}, upstream of GFP. The amount of GFP transcription was controlled by varying the concentration of the IPTG inducer in the growth medium.

In addition, the authors wanted to study the regulation of translation. Translational efficiency was regulated by constructing a series of *B. subtilis* strains that contained point mutations in the ribosome binding site (RBS) and initiation codon of GFP. By manipulating both the transcription and translation levels, the relative contributions of these processes to biochemical noise could be studied; see the summary of the data in Figure 3. They found that the phenotypic noise strength shows a strong positive correlation with translational efficiency (slope 21.8), in contrast to the weak positive correlation observed for transcriptional efficiency (slope 6.5). This is a demonstration that phenotypic variation can be controlled by genetic parameters, and low translation rates will lead to reduced fluctuations in protein expression. Such results also suggest that in nature, several inefficiently translated regulatory genes could have been naturally selected for their low-noise characteristics, even though efficient translation is energetically favorable.

Chizzolini et al. [14] explored the noise of a synthetic biological gene expression system in vitro. Both *E. coli* cell extract and the *E. coli* PURE system were tested for gene expression of a genetic construct with products including different fluorescent proteins (to quantitate protein production) and Spinach aptamer (to quantitate RNA transcript production). They found several sources of variability: the RBS sequence, whether the expressed genes are added in single expression vectors or cascaded in one vector, the GC content of the coding sequences, and the type of RNA polymerase used. Their results are consistent with the conclusions of Ozbudak et al. [43] that translation contributes more noise than transcription, perhaps due to the need for more molecular components to
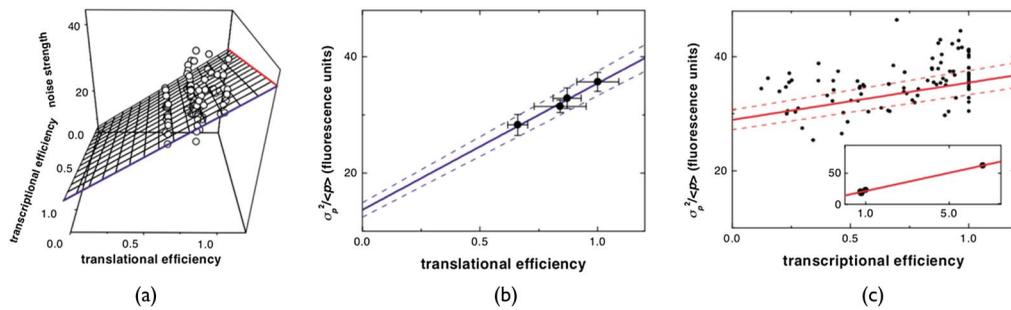
Figure 3. The amount of noise in fluorescent protein output versus translation and transcriptional efficiencies. Output noise was estimated from a population of 10^4–10^5 cells assayed by flow cytometry and represents the cell-to-cell variation in fluorescent signal as standard deviation. Translational and transcriptional levels were normalized to a wild-type strain. Figure adapted from [43].

carry out the task. They point out that the accurate production of one gene product cannot be used to predict the production of another gene product using the same genetic construct and context. Further work on the role of mRNA dynamics could help to understand the source of noise and allow models to be more predictable.

The regulation of a target gene versus noise production has also been explored by Collins and colleagues [40]. They developed a combinatorial promoter design strategy to characterize how the position and number of $tetO_2$ operator sites within the GAL1 promoter affect the target gene expression levels and expression noise in *Saccharomyces cerevisiae*. The promoters were designed with one, two, or three operators at varying proximity to the transcription start site. The inducer ATc was titrated into the system to control gene transcription levels. They found that the multiple operator elements did not behave independently, but rather a certain amount of cooperativity was seen when the ATc was added, with the largest amount of transcriptional variation found when the operators were farthest away from the start codon and when multiple operators were used. Such systematic effects could be modeled and both expression levels and noise levels predicted. These studies exemplify how much control versus biochemical noise could be expected from a biological cell engineered through synthetic biology.

6   Optimization of a Specific Product

The addition of a gene from one organism to another can have substantial societal and economic impact. For example, the introduction of the human insulin gene into *E. coli* or yeast produces useful insulin at a large scale for human use, and this application of recombinant DNA technology has almost completely replaced the use of pigs and cattle as sources of insulin [31]. Therefore, substantial effort has been made to optimize the production of value-added products through synthetic biology.

In the following example, Stephanopoulos and colleagues [2] described how complicated it can be to optimize the production of a target product through genetic manipulation. The goal of the project was to overproduce a Taxol precursor in *E. coli*. To do this they partitioned the metabolic pathway into two modules: an upstream module and a downstream module, each containing several steps in the biosynthetic pathway. They then systematically varied each module to maximize Taxol precursor production. To do so, they constructed 16 different strains of bacteria where both the promoters and the copy numbers of the plasmids were varied. This variation was used as a genetic basis to control the amount of gene products produced in critical parts of the metabolic pathway.

A simplistic assumption would be that to increase the output of a downstream product, each critical metabolic step along the pathway should be upregulated, which would result in higher metabolic

flux through that pathway and more product. This would be consistent with thinking of a biological metabolic pathway as an assembly line. Increasing the rate of production of each individual step (or, more specifically, the rate-limiting steps) should give more product at the end in an additive fashion. However, biological systems are not so simple, and highly reductionist ideas about the modularity or additivity of genetic elements might not be correct. While systematically controlling the expression of the upstream and downstream modules, they found that the highest production rate of the Taxol precursor occurred when the upstream module was expressed at a low level and the downstream module was expressed at a moderate level. High expression of both modules led to very little product (Figure 4). This emphasizes that the context of genetic constructs plays a dominant role in the outcome.

The authors identified a reason for these results, which exemplify the complexities of living systems. They discovered a correlation of Taxol precursor production and the production of indole, a toxic compound. Although the direct biochemical mechanism between the target metabolic pathway and indole production was not known, it would explain why the upstream and downstream metabolic modules needed to be balanced to increase production of the target molecule while limiting the production of toxin.

This study provides an example of the complications of engineering an organism with synthetic biology. In some aspects, genes and gene products can be considered as modular independent units, but often the wider context and embeddedness becomes an important consideration if not a limitation of the system.

7 Parts, Modularity, Standardization

As a discipline, synthetic biology aims to fully characterize the genetic elements it needs to build new genetic circuits and modify the behavior of organisms. The field terms these elements as *parts*, which acknowledges the intent for each genetic element to be used in a modular fashion. For example, promoters, repressors, and reporters can be used again and again in disparate studies with the hope that each part will perform consistently. This has led to the creation of a parts registry [62] and a database where increasing numbers of biological parts are added [22]. Several coordinated efforts have been aimed at defining and refining each part for not only their activity but quality [6, 41].

This collective effort to produce modular and standardized genetic parts to aid in synthetic biology applications is bolstered in large part by community building [21]. Apart from international conferences, the field of synthetic biology is fed by the recruitment and activity of young researchers through the highly active International Genetically Engineered Machine Competition (iGEM),
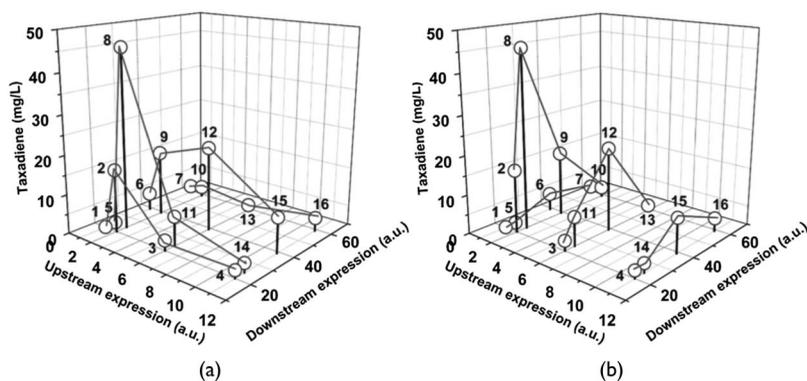

Figure 4. Optimization of a target metabolite (taxadiene) while varying both the upstream and the downstream metabolic paths independently. The highest production of taxadiene was in strain 8, which has low expression of the upstream module and moderate expression of the downstream module. Figure from [2].

which has been held annually since 2004 [63]. As a requirement for this student competition, each group contributes new parts and refinements of parts to the BioBricks database. Such collection and standardization of parts can be used to build not only devices such as logic gates [37], but also larger and larger systems [44], using the principles of engineering.

8   Model-Driven Design in Synthetic Biology

Technical innovation and increase of knowledge may ultimately lead to the design and modification of living systems with *predicted* functionality. Many published examples of synthetic biology demonstrate that several iterative steps of refinement are needed to demonstrate desired function or control of the system. Post hoc refinement shows that our abilities to perform genetic engineering are good but often need improvement. A demonstration of our ability to engineer a system from rational *de novo* design with predicted outcomes would be a valuable step in developing the practicality of synthetic biology as biotechnology.

By using model-driven design principles, important strides towards modularity and standardization have been taken. As an example, a thermodynamic model based on free-energy values was constructed to demonstrate a forward engineering approach to precisely control protein expression levels. By modeling and then testing 119 predictions of the ribosome binding site (RBS) sequence, the authors showed that the translation initiation rates can be controlled over at least a 100,000-fold range [45].

Ellis et al. [20] took a diversity-based, model-guided approach to construct gene networks and tested them for their actual performance versus predicted performance dictated by the model. They started by constructing a library of 20 promoters and tested the expression of a fluorescent reporter protein for each promoter when it was fully repressed and fully expressed to determine the possible dynamic range of each construct. The idea was to quantitate the functionality of each promoter beyond the relative descriptions of "weak" and "strong."

The dynamic range for each promoter in the library was used to parameterize a mathematical model based on deterministic ordinary differential equations, which was useful in predicting population average behaviors under steady state conditions. This model was then applied to a specific construct, a negative feedforward loop where the expression of the reporter gene was modulated by two independent inputs, ATc and IPTG (Figure 5). The output of the mathematical model when both inputs were varied was compared with the output from lab tests and showed very good agreement. They further demonstrated the utility of their model-driven method by predicting and then demonstrating the timing of yeast sedimentation using their promoters and simple genetic control networks. A similar strategy was followed using designed RNA-based devices to quantitatively control gene expression with good correlation between predicted and measured expression levels ($r = 0.94$) [7]. Other approaches to modeling in synthetic biology have been applied, such as using Bayesian statistics to explore high-dimensional parameter spaces [3]. It is hoped that such modeling approaches will begin to make the design and implementation of modular genetic control networks reliable and predictable as a step towards engineering biological systems.
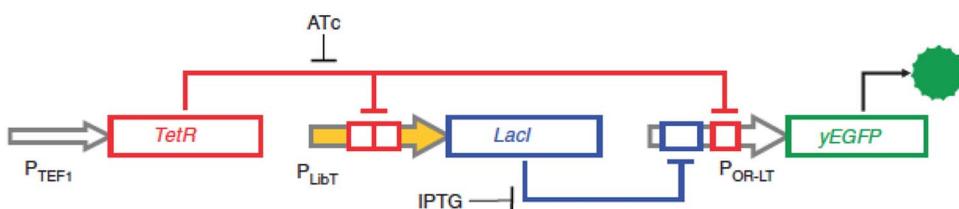

Figure 5. A negative feedforward loop where the expression of the reporter gene (yEGFP in green) was modulated by two independent inputs, ATc and IPTG. (Image adapted from [20].)

9   Combinatorial and Directed Evolution Strategies

Although many current approaches strive to create functional synthetic biology devices through standardization and first-principles design, often the understanding of the biological system or component is not complete enough to allow for successful design implementation. As an alternative method of exploration and construction, combinatorial approaches and directed evolution have also been demonstrated as useful tools towards system creation. For example, promoter libraries can be generated using synthetic degenerate oligonucleotides, resulting in incremental promoter activities when assayed. Such an artificial promoter library then can be used to select gene expression levels over at least an order of magnitude [29]. Combinatorial promoter design was used by Collins and colleagues [40] to measure and predict the amount of noise in a gene expression system, as described above.

In addition to the modulation of single-component functionality, whole genetic circuit integration can be performed using combinatorial and directed evolution approaches [25]. For example, library construction and genetic circuit construction that resulted in nonfunctional mismatched components were shown to evolve into functional devices. In this case the evolved devices outperformed the best-guess rational design constructs [58]. This demonstrates the utility of using irrational design approaches to match the often unknown parameters of multicomponent synthetic genetic devices in their proper context.

10   Cell-Free Approaches

Synthetic biology often employs typical laboratory model organisms such as *E. coli* to host genetic elements and desired functionality. However, some approaches dispense with living cells altogether, demonstrating the advantages of using a cell-free approach. There are two main types of cell-free protein synthesis systems. The first is made by extracting the living protoplasm from living cells. The four commercially available extract types are from *E. coli*, rabbit reticulocytes, wheat germ, and insect cells. The choice of extract depends on the source of the genes one would like to express cell-free. The other main type of cell-free protein synthesis is called the PURE system [48, 49]. This system is derived from the *E. coli* translation system; each component is cloned, purified, and reconstituted to make a functional protein synthesis system in a test tube.

There are several advantages to using cell-free systems for synthetic biology [53]. Cell viability and homeostasis restraints are removed, and conditions or products that would normally kill a living cell are no longer a problem. The struggle to meld the cell's inherent objectives such as production of biomass with the engineer's objectives such as the production of a heterologous metabolite is obviated. Finally, transport barriers are removed, allowing for the introduction or extraction of material, systematic sampling, and direct feeding of metabolites without the need for transmembrane transport. This could allow for prolonged metabolic activity, making this system useful for practical applications such as drug production.

Cell-free protein synthesis as a technological application has been in use for decades. Notably, the first coding assignment of a codon to an amino acid (UUU to phenylalanine) was discovered using cell extract [42]. Cell-free systems have been applied to the production and optimization of valuable chemical targets such as 2,3-butanediol [33] and *n*-butanol [32]. In addition, the cell-free protein synthesis system itself is available for manipulation, refinement, and optimization [8–10, 19].

11   MAGE

DNA replication proceeds differently for each strand of the double helix, with the leading strand being copied more or less continuously and the lagging strand in fragments, necessitating an RNA primer and resulting in an intermediate stage in which Okazaki fragments are produced. George Church and colleagues devised a way to reprogram the genetic information in the cell by supplying

exogenous ssDNA fragments to outcompete the endogenous Okazaki fragments [54]. By introducing synthetic DNA oligonucleotides that are not completely complementary to the original genomic DNA, multiple DNA mutations can be introduced. This technique is termed MAGE (multiplex automated genomic engineering). The possible genetic changes include nucleotide sequence changes, deletions and insertions with the target sequence modified by 1–60 nucleotides. Wang et al. [54] demonstrated this ability to genetically reprogram a cell using an engineered strain in *E. coli* that lacks mismatch repair and expresses a protein that directs ssDNA to the lagging strand and promotes strand annealing. The bacterium to be modified is subjected to multiple iterations of ssDNA introduction through electroporation, recovery, and growth. It is expected that only a proportion of the cells will incorporate the exogenous ssDNA information during a round of DNA replication. Therefore, the process is repeated until the majority of the surviving population contains the desired mutations. The number of cycles has been experimentally determined; some larger genetic changes require more cycles. Since the whole process is iterative, a dedicated automated platform was designed to perform these experiments [55].

To demonstrate an application of the MAGE process, Wang et al. optimized the metabolic flux through a biosynthesis pathway to overproduce the isoprenoid lycopene in an *E. coli* strain [54]. Specifically, for each of the 20 genes, 90-mer oligos containing degenerate ribosome binding site (RBS) sequences flanked by homologous regions on each side were used, with a total pool complexity of 4.7×10^5. Additionally, four genes from secondary pathways were targeted for inactivation by oligos that introduced two nonsense mutations in the open reading frame, to limit the flux through competitive pathways and improve the flux through the desired pathway. In effect, they changed 24 genes simultaneously to maximize lycopene production. As many as 15 billion genetic variants (4.3×10^8 bp variations per cycle for 35 MAGE cycles) were generated and screened. For lycopene production in the *E. coli* strain, fivefold improvement in yield was shown after three days, demonstrating the fast and efficient tuning of genetic diversity in *E. coli*.

Using the same approach, the Church group attempted to recode the *E. coli* genome by selecting 42 highly expressed essential genes and removing all instances of 13 rare codons [35]. In addition, they attempted to show the tolerance of a living organism for large-scale sequence alterations by shuffling all possible codons to synonymous alternatives. They showed that the removal of 13 codons is feasible. Most of the 42 genes were able to be genomically recoded by this method, with a certain cost in growth rate or fitness. 16 of the genes failed to be recoded, demonstrating the limitations to such large-scale alterations.

12   Feedback Control

As mentioned above, a certain number of genetic changes or additions may be tolerated by an organism before the fitness or growth rate decreases due to the metabolic burden. Avoidance of growth deficiencies is of obvious importance when considering the economic objectives in using synthetic biology. Integrated feedback systems that can dynamically sense the concentrations of critical metabolic intermediates could help to balance the metabolism of the organism with the engineered metabolic pathways. Zhang et al. [60] designed such a dynamic sensor-regulator system and used it to produce fatty-acid-based products in *E. coli* for biodiesel production. Dynamically sensing the metabolites in this case is important because ethanol can be produced by this system, which is toxic for the cell, reduces the cell growth rate, and consumes carbon courses needed for producing the desired fatty-acid-based products, among other concerns. They developed a biosensor for a key intermediate, fatty-acyl CoA, and in addition developed two synthetic promoters to increase the dynamic range of gene expression. This integrated feedback system substantially improved the stability of the engineered strains and increased their titers threefold. They argue that many such sensor-regulator systems can be designed for nearly any biosynthetic pathway as long as a sensor either exists or can easily be obtained. In addition, a quorum-sensing circuit in *E. coli* was used to

dynamically regulate the expression of heterologous genes [26]. Using this type of feedback control, the system produced a 5.5-fold increase in a target compound, myo-inositol. Usually synthetic biology uses variation in promoters to fine-tune the regulation of protein expression. Chasin et al. instead used programmed protein degradation from a modular degron library in negative feedback circuits for timed pulses of protein expression [13]. The metabolic burden itself has been used in a feedback loop to control gene expression [11]. These systems demonstrate that dynamic control of an organism's metabolism can balance the needs of the organism to proliferate with the needs of the engineer to produce desired products.

13   Conclusions

Genes can be transferred successfully from one organism to another, and the level of gene expression and activity can be changed using different promoters and/or copy numbers of the genes or balanced with rates of degradation. The gene editing system CRISPR makes genetic changes easy [15] (even with commercial kits being offered to the home market (e.g., [64]).

Challenges remain in several areas. One of the most overlooked aspects of living systems is the amount of inherent noise. In efforts to program organisms with precisely designed genetic constructs, it is often the case that the construct does not perform as intended, due to the variability produced by the host organism. One challenge on this front is to understand how to modulate or control the inherent noise to an acceptable level (if possible) or to understand how to use the inherent noise to the benefit of the system. Related to that challenge is how to produce sustainable genetically modified systems using standard genetic elements such as plasmids. Genetic circuits can function in a host cell for minutes to days before the system becomes nonfunctional, typically due to the loss of plasmids over time. The development of sophisticated feedback mechanisms may allow for sustained function of genetic constructs over time. Again related to the above challenges is the challenge of understanding the toll on the fitness of the engineered organism. How does the metabolic burden manifest itself, what are the underlying mechanisms, and are the effects additive?

Standardization and modularity continue to play a primary role in the refinement of synthetic biology research and practices. With the development of tools and techniques in synthetic biology, deeper questions in biology may become experimentally tractable: for example, the nature of noise, attractor states, embodiment, mutation regimes, biological optimization, and information transfer, to name a few. So far, most of these considerations are appropriately considered for short time scales, for example, from hours to days of functionality. The long-term visions of genetic manipulations are not clearly defined, although they are being discussed. Although synthetic biology provides more control over genetic systems, the limits to the amount of engineering possible in biological systems are yet to be elucidated.

References

1. Ahmad, H. I., Ahmad, M. J., Asif, A. R., Adnan, M., Iqbal, M. K., Mehmood, K., Muhammad, S. A., Bhuiyan, A. A., Elokil, A., Du, X., & Zhao, C. (2018). A review of CRISPR-based genome editing: Survival, evolution and challenges. *Current Issues in Molecular Biology*, *28*, 47–68.

2. Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., & Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science*, *330*(6000), 70–74.

3. Barnes, C. P., Silk, D., Sheng, X., & Stumpf, M. P. (2011). Bayesian design of synthetic biological systems. *Proceedings of the National Academy of Sciences of the USA*, *108*(37), 15190–15195.

4. Benner, S. A. (2010). Defining life. *Astrobiology*, *10*(10), 1021–1030.

5. Callaway, E. (2018). Controversial CRISPR 'gene drives' tested in mammals for the first time. *Nature*, *559*(7713), 164–165.

6. Cambray, G., Guimaraes, J. C., Mutalik, V. K., Lam, C., Mai, Q. A., Thimmaiah, T., Carothers, J. M., Arkin, A. P., & Endy, D. (2013). Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Research*, *41*(9), 5139–5148.

7. Carothers, J. M., Goler, J. A., Juminaga, D., & Keasling, J. D. (2011). Model-driven engineering of RNA devices to quantitatively program gene expression. *Science*, *334*(6063), 1716–1719.

8. Caschera, F., Bedau, M. A., Buchanan, A., Cawse, J., de Lucrezia, D., Gazzola, G., Hanczyc, M. M., & Packard, N. H. (2011). Coping with complexity: Machine learning optimization of cell-free protein synthesis. *Biotechnology and Bioengineering*, *108*(9), 2218–2228.

9. Caschera, F., & Noireaux, V. (2014). Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription–translation system. *Biochimie*, *99*, 162–168.

10. Caschera, F., & Noireaux, V. (2015). A cost-effective polyphosphate-based metabolism fuels an all *E. coli* cell-free expression system. *Metabolic Engineering*, *27*, 29–37.

11. Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., Awan, A. R., Gilbert, C., & Ellis, T. (2018). Burden-driven feedback control of gene expression. *Nature Methods*, *15*(5), 387–393.

12. Chang, A. C., & Cohen, S. N. (1974). Genome construction between bacterial species in vitro: Replication and expression of *Staphylococcus* plasmid genes in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA*, *71*(4), 1030–1034.

13. Chassin, H., Müller, M., Tigges, M., Scheller, L., Lang, M., & Fussenegger, M. (2019). A modular degron library for synthetic circuits in mammalian cells. *Nature Communications*, *10*(1), 1–11.

14. Chizzolini, F., Forlin, M., Yeh Martín, N., Berloffa, G., Cecchi, D., & Mansy, S. S. (2017). Cell-free translation is more variable than transcription. *ACS Synthetic Biology*, *6*(4), 638–647.

15. Cohen, J. (2016). 'Any idiot can do it.' CRISPR could put mutant mice in everyone's reach. *Science Magazine*, 3 Nov. https://doi.org/10.1126/science.aal0334

16. Cohen, S. N., Chang, A. C., Boyer, H. W., & Helling, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the USA*, *70*(11), 3240–3244.

17. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823.

18. DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M., & Church, G. M. (2015). RNA-guided gene drives can efficiently and reversibly bias inheritance in wild yeast. bioRxiv, 013896.

19. Dudley, Q. M., Karim, A. S., & Jewett, M. C. (2015). Cell-free metabolic engineering: Biomanufacturing beyond the cell. *Biotechnology Journal*, *10*(1), 69–82.

20. Ellis, T., Wang, X., & Collins, J. J. (2009). Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnology*, *27*(5), 465–471.

21. Endy, D. (2005). Foundations for engineering biology. *Nature*, *438*(7067), 449–453.

22. Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M., & Gennari, J. H. (2011). Standard biological parts knowledgebase. *PLoS One*, *6*(2), e17005.

23. Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science*, *348*(6233), 442–444.

24. Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*. *Proceedings of the National Academy of Sciences of the USA*, *112*(49), E6736–E6743.

25. Guet, C. C., Elowitz, M. B., Hsing, W., & Leibler, S. (2002). Combinatorial synthesis of genetic networks. *Science*, *296*(5572), 1466–1470.

26. Gupta, A., Reizman, I. M. B., Reisch, C. R., & Prather, K. L. (2017). Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nature Biotechnology*, *35*(3), 273.

27. Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., et al. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nature Biotechnology*, *34*(1), 78.

28. Jackson, D. A., Symons, R. H., & Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of simian virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the USA*, *69*(10), 2904–2909.

29. Jensen, P. R., & Hammer, K. (1998). Artificial promoters for metabolic optimization. *Biotechnology and Bioengineering*, *58*(2–3), 191–195.

30. Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 structures and mechanisms. *Annual Review of Biophysics*, *46*, 505–529.

31. Johnson, I. S. (1983). Human insulin from recombinant DNA technology. *Science*, *219*(4585), 632–637.

32. Karim, A. S., & Jewett, M. C. (2016). A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metabolic Engineering*, *36*, 116–126.

33. Kay, J. E., & Jewett, M. C. (2015). Lysate of engineered *Escherichia coli* supports high-level conversion of glucose to 2,3-butanediol. *Metabolic Engineering*, *32*, 133–142.

34. Kellogg, S. T., Chatterjee, D. K., & Chakrabarty, A. M. (1981). Plasmid-assisted molecular breeding: New technique for enhanced biodegradation of persistent toxic chemicals. *Science*, *214*(4525), 1133–1135.

35. Lajoie, M. J., Kosuri, S., Mosberg, J. A., Gregg, C. J., Zhang, D., & Church, G. M. (2013). Probing the limits of genetic recoding in essential genes. *Science*, *342*(6156), 361–363.

36. Lederberg, J. (1952). Cell genetics and hereditary symbiosis. *Physiological Reviews*, *32*(4), 403–430.

37. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., & Voigt, C. A. (2012). Genetic programs constructed from layered logic gates in single cells. *Nature*, *491*(7423), 249–253.

38. Moser, F., Broers, N. J., Hartmans, S., Tamsir, A., Kerkman, R., Roubos, J. A., et al. (2012). Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synthetic Biology*, *1*(11), 555–564.

39. Muller, H. J. (1927). Artificial transmutation of the gene. *Science*, *66*(1699), 84–87.

40. Murphy, K. F., Balázsi, G., & Collins, J. J. (2007). Combinatorial promoter design for engineering noisy gene expression. *Proceedings of the National Academy of Sciences of the USA*, *104*(31), 12726–12731.

41. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Mai, Q. A., Christoffersen, M. J., Martin, L., et al. (2013). Quantitative estimation of activity and quality for collections of functional genetic elements. *Nature Methods*, *10*(4), 347.

42. Nirenberg, M. W., Matthaei, J. H., Jones, O. W., Martin, R. G., & Barondes, S. H. (1963). Approximation of genetic code via cell-free protein synthesis directed by template RNA. *Federation Proceedings*, *22*(1, Jan.), pt 1, 55–61.

43. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nature Genetics*, *31*(1), 69–73.

44. Purnick, P. E., & Weiss, R. (2009). The second wave of synthetic biology: From modules to systems. *Nature Reviews Molecular Cell Biology*, *10*(6), 410–422.

45. Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, *27*(10), 946.

46. Sander, P., Wittich, R. M., Fortnagel, P., Wilkes, H., & Francke, W. (1991). Degradation of 1,2,4-trichloro- and 1,2,4,5-tetrachlorobenzene by *Pseudomonas* strains. *Applied and Environmental Microbiology*, *57*(5), 1430–1440.

47. San Millan, A., Escudero, J. A., Gifford, D. R., Mazel, D., & MacLean, R. C. (2017). Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. *Nature Ecology & Evolution*, *1*(1), 1–8.

48. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature Biotechnology*, *19*(8), 751–755.

49. Shimizu, Y., Kanamori, T., & Ueda, T. (2005). Protein synthesis by pure translation systems. *Methods*, *36*(3), 299–304.

50. Shintani, D., & DellaPenna, D. (1998). Elevating the vitamin E content of plants through metabolic engineering. *Science*, *282*(5396), 2098–2100.

51. Threlfall, D. R., & Whistance, G. R. (1971). *Aspects of terpenoid chemistry and biochemistry*. London and New York: Academic Press.

52. Vogwill, T., & MacLean, R. C. (2015). The genetic basis of the fitness costs of antimicrobial resistance: A meta-analysis approach. *Evolutionary Applications*, *8*(3), 284–295.

53. Wang, H. H., Huang, P. Y., Xu, G., Haas, W., Marblestone, A., Li, J., et al. (2012). Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multienzyme catalysis. *ACS Synthetic Biology*, *1*(2), 43–52.

54. Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, *460*(7257), 894–898.

55. Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., & Church, G. M. (2012). Genome-scale promoter engineering by coselection MAGE. *Nature Methods*, *9*(6), 591.

56. Wilkinson, R. A., Martin, C., Nemudryi, A. A., & Wiedenheft, B. (2019). CRISPR RNA-guided autonomous delivery of Cas9. *Nature Structural & Molecular Biology*, *26*(1), 14–24.

57. Ye, H., Daoud-El Baba, M., Peng, R. W., & Fussenegger, M. (2011). A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science*, *332*(6037), 1565–1568.

58. Yokobayashi, Y., Weiss, R., & Arnold, F. H. (2002). Directed evolution of a genetic circuit. *Proceedings of the National Academy of Sciences of the USA*, *99*(26), 16587–16591.

59. Zhan, T., Rindtorff, N., Betge, J., Ebert, M. P., & Boutros, M. (2019). CRISPR/Cas9 for cancer research and therapy. In *Seminars in cancer biology* (*Vol. 55*, pp. 106–119). London: Academic Press.

60. Zhang, F., Carothers, J. M., & Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*, *30*(4), 354.

61. https://www.bbc.com/news/world-asia-china-50944461

62. https://biobricks.org/

63. http://igem.org/

64. http://www.the-odin.com/diy-crispr-kit/