# Chapter 5
# Synthetic Biology

**Abstract** Synthetic biology started with an emphasis in experimental molecular biology through the demonstration that characterized DNA sequences which can be taken out of their native context and re-implemented in novel ways. The scope of synthetic biology research has rapidly increased with the improvement and development of tools for direct DNA synthesis and assembly of DNA molecules. These tools now make it possible to engineer biological systems precisely and accurately to reflect specific DNA-level designs. Application of synthetic biology techniques to biofuels research expands the scope of biological engineering that can be achieved where it is now possible to conceive, design, and implement large-scale changes to a cellular system.

While systems biology has provided a strong biological knowledge base for information and analysis, synthetic biology mainly focuses on tools and methods to manipulate or modify a biological system. A general goal of synthetic biology, which builds on advances in molecular and systems biology, is to expand the uses and applications of biology in the same way that chemical synthesis expanded the uses and applications of chemistry. Currently, synthetic biology has a main focus on nucleic acid methodologies (DNA, RNA) with one aim to provide standardized methods for genetic engineering.

Just as systems biology was enabled by technological developments, synthetic biology was also enabled by technology advances. Specifically, improved methods for DNA synthesis and molecular tools for assembling DNA are foundational to synthetic biology. Generally speaking, the improved ability to synthesize and construct DNA has led to the ability to more carefully interrogate genotype-phenotype relationships and also enabled the generation of novel genetically encoded biological function. Synthetic biology includes the design and construction of new biological entities, such as enzyme and even whole cells in order to create novel combinations of processes. The complexity of biological systems provides multiple types of machinery and a variety of options to include in those parts when engineering a system.

In its current form, synthetic biology operates primarily with nucleic acids. This means that design and implementation are done at a genetic base-by-base level. The most common type of sequence that is used is a gene where an average

S. M. Clay and S. S. Fong, *Developing Biofuel Bioprocesses Using Systems and Synthetic Biology*, SpringerBriefs in Systems Biology,
DOI: 10.1007/978-1-4614-5580-6_5, © The Author(s) 2013

length is about 1,000 DNA bases. To facilitate communication at this level, synthetic biology has adopted several terms to describe different levels of organization (see Box 5.1). The term DNA "part" is used to refer to a standalone DNA sequence that has a discrete function. Parts can vary in length, but are typically of the order of tens to thousands of DNA bases in length. Representative parts could be a gene, a promoter, or a terminator.

If several parts are used in concert to achieve a more complicated function that collection of parts is termed a "device." The early demonstrations of the genetic toggle switch (Gardner et al. 2000) or repressilator (Elowitz and Leibler 2000) can be considered devices as well as the bacterial photography device (Levskaya et al. 2005). At a similar level of organization, the term genetic circuit is often used. Genetic circuits also typically incorporate a collection of DNA parts, but the difference in terminology is born from some of the early parallels to electrical engineering concepts that helped to lay the foundation for design. Some of the classic genetic circuits that have been constructed to date are biological equivalents to logic gates used in electrical circuitry (Wang et al. 2011; Zhan et al. 2010).

The host organism for implementing synthetic biology constructs is termed the "chassis." Currently, the most commonly used chasses in synthetic biology are *Escherichia coli* and *Saccharomyces cerevisiae* due to the large amount of information available for these organisms and the relative ease of working and genetically manipulating these organisms. In a generic sense, a chassis can be any system that contains all of the components necessary to functionally express a genetic construct, so it may be possible to engineer a biological chassis that is specialized for a given application. In the future, there may be a cellulolytic chassis that can be used as the starting point for biofuel applications that is specifically designed to efficiently breakdown lignocellulosic biomass and be streamlined for target fuel production.

**Box 5.1: Synthetic Biology Terminology**

Part: a single, relatively short DNA sequence with discrete, defined function
Device: a collection of multiple DNA sequences that integrates individual functions to achieve a novel coordinated function
Genetic circuit: a collection of multiple DNA sequences designed to operate as a functional circuit (design parallels to electrical engineering)
Chassis: host organism for implementing genetic constructs

## 5.1 Experimental Synthetic Biology

As with systems biology, the field of synthetic biology is not clearly defined (some would consider synthetic biology a natural progression of molecular biology), but there are some commonalities demonstrated by pioneering synthetic biology research. The earliest synthetic biology experiments (repressilator and toggle switch)

utilized different genetic components found in various systems to conceptually and experimentally implement novel, controlled functions into a biological system. These early projects demonstrated some of the hallmarks of synthetic biology: novel design and utilization of genetic tools for experimental implementation. In parallel to a growing number of developed genetic circuits to demonstrate novel function, much of the work in experimental synthetic biology has been focused on developing methods for genetic engineering.

### *5.1.1 Core Experimental Methods*

As the methods for DNA synthesis became more standardized, synthetic biologists took the opportunity to build a database of parts that could be combined in endless variations to build organisms with new functions. These DNA building blocks are called BioBrick parts and they are categorized into their different functions. There are BioBrick primers, ribosome binding sites, protein domains, protein coding sequences, terminators, and plasmid backbones. The database includes other various parts and combinations of existing parts as well. The ease of use with BioBrick parts comes from the systematic use of restriction enzymes specific to BioBrick parts, which makes assembling the DNA much like putting a puzzle together.

Different methods have been developed for assembling DNA fragments and BioBrick users can choose between 3A, Scarless, and Gibson assemblies. The name 3A refers to the three antibiotics used for selection with antibiotic resistance and it has the highest success rate with BioBrick parts. While there is no PCR or gel purification needed for this assembly, there is a scar left behind from the restriction and ligation process.

Additional methodologies have been developed to implement scarless assembly of DNA fragments. As implied by the name, these methods do not leave a scar from linkers. The absence of scars is very useful in assembling proteins and also allows the user to assemble parts that may not be compatible otherwise. Polymerase cycling assembly (PCA) runs similar to a PCR and uses oligonucleotides that all have flanking regions that combine to leave single-stranded gaps that a DNA polymerase then fills in. The DNA strands can be up to 50 base pairs and should overlap about 20 base pairs. Similar to PCA is another method called Isothermal Assembly (Gibson Assembly) where there is an overlap of about 20–40 base pairs and multiple strands of DNA can be joined in one reaction. Unlike PCA, Gibson is an isothermal assembly that occurs at 50 °C and runs for up to an hour making it one of the quickest assembly methods. In this method however the oligonucleotides contain the complete sequence; therefore, there is no need to fill in missing sections of the DNA although DNA polymerase is included in the reaction in case there are any gaps. This method becomes specifically useful when combining blunt-ended fragments. A T5 exonuclease is used to eliminate 20–40 base pairs from each end, leaving a single-stranded sticky end for ligation.

In order to make a recombinant gene more efficient it is important to make sure that the codons being used are easily used by the host. Codons are the three base pair sequences that code for a single amino acid and there are multiple codons that code for the same amino acid. Depending on what organism the recombinant genes are coming from or going to, the preferred codons will be slightly different. Codon optimization is the changing of a base pair in a codon in order to gain optimal production of the amino acid in the host organism. Optimizing the codons is most important when the recombinant DNA comes from a source that is genetically distant from the host organism such as plant DNA into bacteria. When optimized, it helps to improve improved translation rates, protein yields, and enzymatic activities.

Metabolic evolution provides a route for optimization and an option when determining the strongest strains. Allowing the organisms to compete for a food source allows nature to take over and the best strain can thrive and adapt and then evolve to be most suited for the given environment. Small mutations in the cell may happen naturally over time or they can be influenced by duplicating a gene using an enzyme with a high error rate. Allowing the cell to adapt makes the strain more stable and long lasting. Once the strain has evolved into a more robust state, the new genes can be used for the redesign of other systems.

### *5.1.2 Progress for Biofuels*

While synthetic biology is a relatively young field, the global interest in biofuel research has led to the application of synthetic biology to several successful biofuel studies. One of the general approaches that are used is to use optimized heterologous expression of targeted genes to introduce novel biofuel production capabilities into an amenable host strain. Examples of this include the expression of different alcohol dehydrogenase genes from *Saccharomyces cerevisiae* and *Lactococcus lactis* in *Escherichia coli* to generate a strain of *E. coli* that produces isobutanol (Atsumi et al. 2010). Another demonstration was the engineering of the cyanobacterium *Synecoccocus elongatus* to produce isobutyraldehyde (Atsumi et al. 2009).

These demonstrations exhibit the ability to effectively use synthetic biology techniques to identify and express genes to modify specific pathways within an organism. This approach largely leaves the majority of an organism's biochemical network unaltered and intact. With the generation of an entire synthetic genome (Gibson et al. 2010) it may become possible to change the scale of synthetic biology engineering to include whole-cell design, not just pathway-specific design.

## 5.2 Computational Synthetic Biology

To complement and facilitate experimental synthetic biology research, computational methods are being developed. Due to the difference in system size and goals, the methods developed for synthetic biology are different from systems

biology computational methods. Generally, the systems that are considered for synthetic biology are smaller in scale, but more detailed molecular-level dynamics are important.

## *5.2.1 Core Computational Methods*

### 5.2.1.1 Genomic Information

One of the shifts associated with synthetic biology is the ability to explicitly control the base-by-base sequence of a genetic construct. With this level of control, it is possible to directly interrogate the effect of specific genetic changes to function. This is at the core of establishing genotype-phenotype relationships.

At one level, there is the need to compile and interpret sequence information. At a course-grain level, this is achieved during genome sequencing by genome annotation. There are several methods and pipelines that have been used to achieve genome annotation using computational means (though some input from experts is always beneficial). These include the Integrated Microbial Genomes (IMG) pipeline (http://img.jgi.doe.gov), the SEED system (www.theseed.org/wiki/Home_of_the_SEED), and a pipeline that is being developed through the National Institutes of Health (www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

At a more detailed level, there are also programs such as GenoCAD (Wilson et al. 2011) that begins by developing a "grammar" for genetic parts. This approach considers genetic sequences as a language with specific rules that dictate the structure and function of different genetic parts. This grammar can be applied to not only studying DNA sequences for functional sequences, but can also be used as a basis for designing constructs to achieve new functional units.

### 5.2.1.2 Design Tools

A variety of tools are being established to help to design genetic circuits. Given the ability to experimentally construct any desired DNA sequence exactly, the design process has become truly open-ended. Any gene from any organism can be utilized in combination with any other gene. In a broader sense, even novel (previously undocumented) gene function can be proposed and tested.

Some of the design approaches that focus on utilizing existing biological information attempt to mine database information to propose a collection of genes (from any organism) to create a pathway that would achieve the desired goal. The "From Metabolite to Metabolite" tool demonstrates one iteration of this approach (http://fmm.mbc.nctu.edu.tw/). Using this tool, a user only needs to input a starting metabolite and a desired end metabolite. The algorithm then uses information from online databases such as KEGG, UniProt, and GeneBank to identify metabolic, protein, and sequence information, respectively. The output is a list of

proposed pathways that could achieve the desired biochemical conversion from one metabolite to another.

Other tools have been developed to approach the design problem from a more generic approach. In these approaches, different methods represent chemicals/metabolites in a standardized form so that individual biochemical transformations can be considered in a stepwise fashion. Each proposed biochemical transformation can then be correlated to enzymes that would have the closest reaction mechanism (often as dictated by the enzyme commission number—EC#). This approach is conceptually similar to the old word game of changing one word to another by changing only one letter at a time while maintaining a valid word at each intermediate step (see Box 5.2). The different methods that have been implemented for this type of approach largely different on the method by which chemicals are represented (atomic mapping onto graph coordinates or linearized representation of atoms).

**Box 5.2: Illustration of Stepwise Transformation**

C A T

Step 1: Conversion of "A" to "O"

C O T

Step 2: Conversion of "C" to "D"

D O T

Step 3: Conversion of "T" to "G"

D O G

#### 5.2.1.3 Dynamic Simulation

With tools to study the basic information content of different DNA sequences and to propose different collections of genes to achieve a desired outcome, the final step that computational methods have addressed is the ability to simulate the function of the designed genetic circuit dynamically. A variety of different methods can be implemented to dynamically simulate small gene circuits including differential equation modeling, stochastic simulations, and agent-based modeling.

Differential equation models are the staple of dynamic simulations and can be implemented for small systems. Tools such as TinkerCell and SynBioSS can be used to develop computational models for small synthetic systems.

Stochastic simulations provide a simulation method that is different from differential equation models in which they are not deterministic and therefore account for some of the variability and noise that are inherent in biological systems. While this has the advantage of being a better representation of biological processes, there is often a tradeoff in terms of the size of the system that can be simulated and the computational time required to run simulations.

Agent-based models can be considered a subset of stochastic simulations, but with one major distinction. Agent-based models are formulated with discrete agents representing the physical entities within the system and thus it is possible to account for density and spatial effects. As with other stochastic simulations, there are limitations to the size of the system that can be studied largely due to computational resource limitations.

### *5.2.2 Progress for Biofuels*

Currently, the number of studies linking computational modeling, synthetic biology, and biofuels is relatively limited. The majority of the work in this area has thus far focused on computational tools that can help with the design process, specifically in terms of helping identify non-native pathways and chemical targets that can be implemented using controlled heterologous gene expression.

As mentioned previously, one Web-based tool that can be used to help with pathway design independent of organism is the "From Metabolite to Metabolite" algorithm. Another recently developed algorithm used *Escherichia coli*, *Saccharomyces cerevisiae*, and *Cornyebacterium glutamicum* as host organisms and searched for non-native metabolites that could potentially be produced by heterologous expression (Chatsurachai et al. 2012). This algorithmic search was then coupled with flux balance analysis to determine feasibility.

Another common approach for studying the diversity of metabolites that can be produced by an organism via biochemical means is the use of graph theory or graph-based algorithms (Brunk et al. 2012). There exist different implementations of this approach to studying biochemical conversions. A recent implementation was used to specifically study the potential of different organisms to produce 1-butanol as a biofuel target (Ranganathan and Maranas 2010).

## References

Atsumi S, Higashide W, Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nat Biotechnol 27(12):1177–1180. doi:10.1038/nbt.1586

Atsumi S, Wu TY, Eckl EM, Hawkins SD, Buelter T, Liao JC (2010) Engineering the isobutanol biosynthetic pathway in Escherichia coli by comparison of three aldehyde reductase/alcohol dehydrogenase genes. Appl Microbiol Biotechnol 85(3):651–657. doi:10.1007/s00253-009-2085-6

Brunk E, Neri M, Tavernelli I, Hatzimanikatis V, Rothlisberger U (2012) Integrating computational methods to retrofit enzymes to synthetic pathways. Biotechnol Bioeng 109(2):572–582. doi:10.1002/bit.23334

Chatsurachai S, Furusawa C, Shimizu H (2012) An in silico platform for the design of heterologous pathways in nonnative metabolite production. BMC Bioinformatics 13(1):93. doi:10.1186/1471-2105-13-93

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403(6767):335–338. doi:10.1038/35002125

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403(6767):339–342. doi:10.1038/35002131

Gibson DG et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329(5987):52–56. doi:10.1126/science.1190719

Levskaya A, Chevalier AA, Tabor JJ, Simpson ZB, Lavery LA, Levy M, Davidson EA, Scouras A, Ellington AD, Marcotte EM, Voigt CA (2005) Synthetic biology: engineering Escherichia coli to see light. Nature 438(7067):441–442. doi:10.1038/nature04405

Ranganathan S, Maranas CD (2010) Microbial 1-butanol production: Identification of non-native production routes and in silico engineering interventions. Biotechnol J 5(7):716–725. doi:10.1002/biot.201000171

Wang B, Kitney RI, Joly N, Buck M (2011) Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat Commun 2:508. doi:10.1038/ncomms1516

Wilson ML, Hertzberg R, Adam L, Peccoud J (2011) A step-by-step introduction to rule-based design of synthetic genetic constructs using GenoCAD. Methods Enzymol 498:173–188. http://www.ncbi.nlm.nih.gov/pubmed/21601678. Accessed 13 July 2012

Zhan J, Ding B, Ma R, Ma X, Su X, Zhao Y, Liu Z, Wu J, Liu H (2010) Develop reusable and combinable designs for transcriptional logic gates. Mol Syst Biol 6:388