Chapter 7
Building Engineered Strains

Abstract A critical step in engineering design is the ability to accurately construct something to directly translate a conceptual design to practice. Due to the relatively short history of molecular biology and the limitations in knowledge and technology, biological engineering has been hindered by a lack of ability to easily implement conceptual designs. Recent developments have attempted to standardize different aspects of biology to facilitate and expedite biological engineering. Standardization of genetic parts will not only ease the methodological hurdles for biological engineering, but it will also enable biological engineering to focus more exclusively on conceptual design of function rather than being constrained by practical limitations.

A classic hallmark of humans is the ability to conceive, build, and use tools to achieve desired goals. Advances in construction methods relevant for most disciplines allow building or construction to occur quickly and with minimal thought. For example, humans have been working with iron for more than 3000 years and mass-produced steel for more than 160 years. This established history of metallurgy allows for devices to be constructed out of metal quickly and efficiently. In terms of the overall process of biological process development, this chapter will focus on the specific aspect of building/construction as it is one of the major hurdles for biological engineering.

The ability to quickly and efficiently construct desired devices has not been possible for biological systems, especially if considering a functional cell to be the device to be constructed. Contrary to the long history and established methodologies in other disciplines, scientists have only been working in detail with the fundamental material of biological systems (DNA) for a little over 50 years. While great advances have been made in a short period of time, there are still fundamental pieces of information that are still being discovered, such as the debated possibility of growing and incorporating arsenic instead of phosphorus into biological building blocks [including DNA (Wolfe-Simon et al. 2011)].

The shorter history of working with DNA in biological systems leads to very practical limitations in terms of what can be achieved in building biological systems. In most engineering disciplines other than biological engineering

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_7, © The Author(s) 2013

(mechanical, electrical, chemical, computer, civil), the historical foundations of those fields have provided sufficient depth of knowledge and tools to prospectively design and physically implement a product (Table 7.1). In the more established engineering disciplines, the different aspects of knowledge, design, and construction are typically segregated and discrete steps. The implication of this is that each of the steps can be addressed independently without major concerns for the other downstream steps. For example, an internal combustion engine can be designed based upon principles of combustion and mechanics. Design efforts produce a blueprint/schematic of the proposed engine based upon theory and desired functionality. The design step can be considered almost exclusively based upon the fundamental design principles and desired function without much regard to the actual construction of the engine. The techniques and capability to construct the engine are assumed as methodologies for metal working and machining are available.

The ability to segregate knowledge, design, and construction into discrete, independent steps has not always been possible for biological systems and biological engineering. Often all three of these components are interwoven due to different limitations. The design process often is limited by the amount of information/knowledge available and sometimes, biological design cannot occur due to a lack of information. For example, an organism can be engineered to produce specific polyketides through genetic engineering, but this ability is contingent upon having an understanding of polyketide synthases. In other areas such as terpenoid production, the design process is not as straightforward as the level of knowledge for terpene synthases is not as well established as for polyketide synthases. One current example of this is the desire to heterologously express and produce the terpenoid paclitaxel (Taxol) for cancer treatment, but the complete biosynthetic pathway has not yet been elucidated. Thus, an engineered strain for bioproduction of paclitaxel is not yet possible.

Of specific concern for this section, there have also been limitations in construction that influence the design step. Currently, the single most commonly used approach for implementing biological designs is the use of genetic engineering methodologies. Thus, there is really no consideration on what the materials of construction will be (DNA), but the ability to exactly build something to match a *de novo* design specification and to have it compatible with other components by a set of standards has been problematic. Practically speaking what this has led to in biological engineering is a design process where designs are constrained by what information is known and also how possible it is implement a proposed design. Until recently, this has caused the engineering process for biological systems to be a process of continual feedback and iteration rather than a linear building/construction process.

In terms of building engineered strains, recent developments (mainly associated with technology and methodology improvements) have sought to address the constraints and limitations associated with the implementation or construction of genetic constructs. The developments have taken different approaches to address the two main problems: (1) the ability to construct something that exactly matches the blueprint/design and (2) standardization to facilitate compatibility.

Table 7.1 Examples of aspects of design in various engineering disciplines

Engineering Discipline	Goal	Theory	Parts	Product
Electrical	Electronic device	Kirchoff's law Ohm's law	Transistors, resistors, capacitors, inductors, diodes	Circuit board
Chemical	Chemical process	Conservation of mass and energy, reaction kinetics	Catalysts, chemicals	Chemical reactor
Biological	Biological process	Central dogma of molecular biology	DNA	Cell

7.1 Standardization of DNA

Standardization has many different facets but the two that we will focus on here are the standardization of requirements and the standardization of part interoperability. When mentioning the standardization of requirements, what is meant is a set of guidelines or rules that are accepted in the field and must be abided by. For standardization of part interoperability, the intention is to facilitate the actual construction and implementation during the actual building process.

Standardization of requirements occurs in all engineering disciplines and impacts various facets of that field. Every field has some form of standards that related to safety and ethics. In addition, there are also technical specifications/standards that are established. For biological systems, these standards in the areas of ethics and technical specifications are often unspoken and can vary from lab to lab.

In the area of safety and ethics, some of the guidelines are well established whereas others are almost left to the individual to decide. Safety standards are relatively uniform, especially for work that is conducted in an academic setting or sponsored by federal funding. For example, the National Institutes of Health have established, published guidelines regarding research using recombinant DNA (http://oba.od.nih.gov/rdna/nih_guidelines_oba.html). One of the challenges is that methodologies and capabilities are continually changing and thus, policy guidelines must also evolve. This is demonstrated by the discussion on amending the recombinant DNA guidelines to account for synthetic nucleic acids (http://oba.od.nih.gov/rdna_rac/rac_pub_con.html).

Ethical considerations are less well-defined than safety in terms of policies and guidelines. The ethical implications of genetic engineering and designed organisms are more grounded in personal world views rather than technical detail. Some of the more widely debated related subject areas are genetically modified foods and whole organism cloning. Specific to the topic of biofuel production

and synthetic or recombinant DNA are questions regarding the scope of genetic interventions. Are we ethically comfortable with modifying a single gene in an organism? An entire pathway? The entire organism? As mentioned previously, the technical capability exists to synthetically create the DNA for an organism's entire genome. If an entire genome is synthetically created and no known organism's genome is used as template, would the new genome and related organism be a synthetically created new life form? How much modification to an existing organism would need to be made to designate a new species? The answers to these questions and how the scientific community should proceed are pressing issues that need to be addressed.

The other major area of standardization that is being addressed is the standardization of technical specifications. This is largely a practical consideration and follows from the concept of interchangeable parts. A nut and bolt combination works well when there are standardized threads (depth and pitch) that match between the nut and bolt. Furthermore, if depth and pitch of the thread is standardized, replacement nuts can be easily found for a given bolt if a nut should be lost. This type of technical standardization is common in most fields where something is built or constructed.

Biological research has long been one where standardization is not common. Polymerase chain reaction (PCR) is one of the most useful and prevalent methods used to amplify DNA, but every individual DNA sequence to be amplified requires the design and construction of unique DNA primer sequences to be used in the PCR reaction. Furthermore, depending upon the characteristics of the DNA sequence to be amplified, changes may need to be made to the actual experimental protocol of the method. The challenges associated with the uniqueness of biology are pervasive and can be seen from designing individual probes for gene expression microarrays to having mass and fragmentation patterns for mass spectrometry applications.

In terms of genetic engineering, one of the most comprehensive attempts to establish standardized technical specifications for DNA is the BioBrick formalism (Shetty et al. 2008). The BioBrick concept establishes a standard format for all DNA sequences where a sequence of interest is flanked upstream and downstream by standard DNA sequences. The added DNA sequences are cut sites that are recognized by specific restriction enzymes. In this format, the desired DNA sequence is flanked upstream by the restriction enzyme recognition cut sites for EcoRI and XbaI and downstream by the restriction enzyme recognition cut sites for SpeI and PstI (Fig. 7.1). Using this formalism, the desired DNA sequence (DNA part) can be of any length with any sequence (as long as the sequence does not contain cut sites for EcoRI, XbaI, SpeI, or PstI). Different DNA sequences that have this format can be manipulated by the same protocol by using the four restriction enzymes EcoRI, XbaI, SpeI, and PstI. Currently, thousands of DNA sequences of varying length and function exist in a centralized DNA repository (www.partsregistry.org) and all of these parts can be worked with using standardized protocols (contrast this with having to design and synthesize primers for each sequence individually).

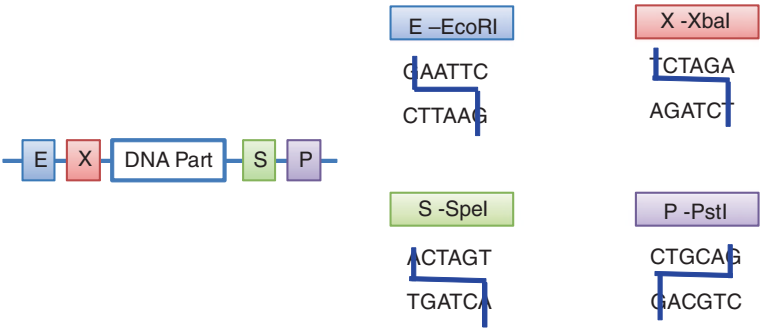

Fig. 7.1 Graphical depiction of the BioBrick format for standardizing DNA parts. A target sequence (DNA part) is flanked upstream by restriction enzyme cut sites EcoRI (E) and XbaI (X) and downstream by SpeI (S) and PstI (P)

When using the BioBrick format, DNA sequences can be isolated and assembled with other DNA sequences quickly by cutting the sequence using restriction enzymes and ligating the desired sequences together using DNA ligase. This can be done for as many DNA sequences as desired and in any order that is desired. Furthermore, there exist a number of DNA plasmids that have been constructed with replication origins and different antibiotic resistance genes as markers. Thus, the same methodologies used to assemble DNA sequences together can be used to generate a self-replicating plasmid housing DNA sequences of interest.

7.2 Interoperability of DNA Constructs

The challenges associated with building or constructing biological systems continues beyond developing methods to standardize genetic parts. One of the characteristics of biological systems is that the components in an organism are all highly connected (metabolic network, regulatory network, protein interaction network). Thus, even after being able to construct a desired DNA sequence properly, there is a challenge in having the DNA sequence expressed and functional within the network context of existing components. This can be viewed as a challenge of interoperability or compatibility.

In terms of controlling expression of the designed DNA sequence, various tools are available to help to control or to dictate the expression level of an introduced or modified DNA sequence. Fundamentally, expression is controlled at the transcription and translation steps by the pairings of promoter and DNA polymerase for transcription and ribosome binding site (RBS) and ribosome for translation. The level of transcription can be altered by modifying the binding strength between a promoter and DNA polymerase. To date, the most effective means of achieving this has been accomplished empirically by developing and

characterizing promoter libraries that contain promoters with small variations in sequence. The level of translation can be similarly altered by modifying the interaction between the RBS and the ribosome. In this case, the interaction between the RBS and ribosome follows nucleic acid pairing, so predictions on the strength of this interaction can be made to help to guide design of this interaction (an example of this is the ribosome binding site calculator (Salis et al. 2009)). In addition to these primary methods, other methodologies such as codon optimization and designing RNA hairpins can influence the expression levels of a target DNA sequence.

Even when a DNA sequence is designed, constructed, and care is taken to try to control its expression, the DNA sequence may not function as desired within the context of a living organism. In these instances, the ability to efficiently build and modify a biological system is severely limited by the state of our knowledge. It is often difficult to predict all of the downstream consequences of a genetic modification and in some instances there may be no means of predicting how an introduced gene/protein (for example) will interact with other existing genes/proteins. In these instances, the most common approach to address interoperability issues is an empirical approach where molecular evolution is used. Variants of the desired gene/protein are generated and screened in context for the desired function. Currently, the use of molecular evolution is often necessary for implementation of even small genetic constructs that contain only a couple genes.

References

Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27(10):946–950. doi:10.1038/nbt.1568

Shetty RP, Endy D, Knight TF Jr (2008) Engineering BioBrick vectors from BioBrick parts. J Biol Eng 2:5. doi:10.1186/1754-1611-2-5

Wolfe-Simon F, Switzer Blum J, Kulp TR, Gordon GW, Hoeft SE, Pett-Ridge J, Stolz JF, Webb SM, Weber PK, Davies PC, Anbar AD, Oremland RS (2011) A bacterium that can grow by using arsenic instead of phosphorus. Science 332(6034):1163–1166. doi:10.1126/science.1197258