Engineering Bacteria for Drug Production

Jay Keasling
Berkeley Center for Synthetic Biology
University of California, Berkeley &
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

The need for new drugs and better ways to synthesize existing drugs

It has been estimated that 5,000-10,000 compounds must be introduced into the drug discovery pipeline for every successful candidate. On average, it takes $802 million and 10-15 years to develop a successful drug. Given the very low success rates and incredible costs in developing a drug, it is important for any drug company to introduce as many drug candidates into their pipeline as possible.

An important source for drug leads has been natural products. As high as 60% of the successful drugs are of natural origin (Cragg, Newman et al. 1997). Some of the most potent natural products have found use as anti-cancer, anti-bacterial, and anti-fungal drugs. However, most of these natural products have evolved for purposes other than the treatment of human disease. Thus, even though these natural products function as human therapeutics, their pharmacological properties may not be optimal. Furthermore, many of these drugs are produced in miniscule amounts in their native hosts, making the drugs expensive to harvest.

Organic chemistry methodologies are widely used to synthesize many pharmaceuticals (whether of natural origin or not) and functionalize many of the pharmaceutically-relevant natural products in use today. With the appropriate protection and deprotection steps, chiral centers and functionalities can be introduced into molecules with precision. The advent of combinatorial chemical synthesis enables one to construct entire families of molecules substituted at several positions with several different substituents, allowing drug companies to

fill their drug discovery pipelines with variations of promising leads. While advances in organic synthesis methodologies have allowed the creation of complicated molecules, they are still far from approaching the ease, specificity, and "green-ness" of enzymes. Indeed, many organic synthesis routes now incorporate one or more enzymes for transformations that are particularly difficult using non-enzymatic routes. Furthermore, enzymes are now being used for *in vitro*, combinatorial functionalization of complex molecules. The next logical step in the synthesis of chemotherapeutics is the use of enzymes for combinatorial synthesis inside the cell. This would allow one to produce drug candidates from inexpensive starting materials and avoid purification of the enzymes, which may be necessary for *in vitro* synthesis.

Biological engineering for synthesis of drugs

The richness and versatility of biological systems make them ideally suited to solve some of the world's most significant challenges, such as converting cheap, renewable resources into energy-rich molecules; producing high-quality, inexpensive drugs to fight disease; detecting and destroying chemical or biological agents; and remediating polluted sites. Over the years, significant strides have been made in engineering microorganisms to solve many of these problems. For example, microorganisms have been engineered to produce ethanol, bulk chemicals, and valuable drugs from inexpensive starting materials; to detect and degrade nerve agents as well as less toxic organic pollutants; and to accumulate metals and reduce radionuclides. However, these biological engineering challenges have long development times, in large part due to a lack of useful tools that would enable engineers to easily and predictably reprogram existing systems, let alone build new enzymes, signal transduction pathways, genetic circuits, and, eventually, whole cells. The ready availability of these tools would drastically alter the biotechnology industry, leading to less expensive pharmaceuticals, renewable energy, and

biological solutions to problems that do not currently have sufficient monetary returns to justify the high cost of today's biological research.

Most of the biological engineering tools currently available to scientists and engineers have not changed significantly since the dawn of genetic engineering in the 1970s: biologists largely use natural, gene expression control systems (promoters with their cognate repressors/activators). The ability to place a single heterologous gene under the control of one of these native promoters and produce large quantities of a protein of interest is the basis for the modern biotechnology industry. While these redesigned biological control systems have been generally effective for their intended purpose (controlling rather roughly the expression of a single gene or a few genes), not surprisingly they are often inadequate for more complicated engineering tasks: e.g., control of very large, heterologous, metabolic pathways or signal transduction systems. Further, these borrowed "biological parts" retain many of the features that were beneficial in their native form but which make them difficult to use for purposes other than that for which they evolved. Well-characterized standard biological parts and larger devices made from such parts would make biological engineering more predictable and allow construction and integration of larger systems than is currently possible.

In almost every other field of engineering, standards have been developed to allow one to easily assemble components from various manufacturers to build a large integrated system. Biologists and engineers have not yet defined the standards for the various parts that might allow them to build larger biological devices. The design and construction of new devices (genetic control systems, for example) would benefit greatly from a set of standards that would govern how the various parts (regulatory proteins, promoter, ribosome binding site, for example) should

interact and be assembled. Setting a standard will, in turn, encourage manufacturing firms to develop parts as well.

Biological engineering is hindered by the fact that many of the most effective biological parts (promoters, genes, plasmids, etc.) have been patented and are available only to those companies that can afford the royalty payments, which increase the cost of drug development and hamper the development of new biological solutions to problems where the eventual monetary payoff is not significant (basically anything other than drug development). Open-source biological parts, devices, and eventually whole cells will decrease the cost of engineering biological systems, make biological engineering more predictable with less guesswork, and encourage the development of novel biological solutions to some of our most challenging problems. The development of open-source biological technology will lead to better awareness of, and minimization of possible future biological risks, in the same way that open source software tends to promote a constructive and responsive community of users and developers.

Synthetic biology

Synthetic biology is the design and construction of new biological entities such as enzymes, genetic circuits, and cells or the redesign of existing biological systems. Synthetic biology builds on the advances in molecular, cell, and systems biology and seeks to transform biology in the same way that synthesis transformed chemistry and integrated circuit design transformed computing. The element that distinguishes synthetic biology from traditional molecular and cellular biology is the focus on the design and construction of core components (parts of enzymes, genetic circuits, metabolic pathways, etc.) that can be modeled, understood, and tuned to meet specific performance criteria, and the assembly of these smaller parts and devices into larger integrated systems that solve specific problems. Just as engineers now design

integrated circuits based on the known physical properties of materials and then fabricate functioning circuits and entire processors (with relatively high reliability), synthetic biologists will soon design and build engineered biological systems. Unlike many other areas of engineering, biology is incredibly non-linear and less predictable, and there is less knowledge of the parts and how they interact. Hence, the overwhelming physical details of natural biology (gene sequences, protein properties, biological systems) must be organized and recast via a set of design rules that hide information and manage complexity, thereby enabling the engineering of many-component integrated biological systems. It is only when this is accomplished that designs of significant scale will be possible.

Synthetic biology arose from four different intellectual agendas. The first is the scientific idea that one practical test of understanding is an ability to reconstitute a functional system from its basic parts. Using synthetic biology, scientists are testing models of how biology works by building systems based on models and measuring differences between expectation and observation. Second, the idea arose that, to some, biology is an extension of chemistry and thus synthetic biology is an extension of synthetic chemistry. Attempts to manipulate living systems at the molecular level will likely lead to a better understanding, and new types, of biological components and systems. Third is the concept that natural living systems have evolved to continue to exist, rather than being optimized for human understanding and intention. By thoughtfully redesigning natural living systems it is possible to simultaneously test our current understanding, and may become possible to implement engineered systems that are easier to study and interact with. Fourth, the idea emerged that biology can be used as a technology, and that biotechnology can be broadly redefined to include the engineering of integrated biological

systems for the purposes of processing information, producing energy, manufacturing chemicals, and fabricating materials.

While the emergence of the discipline of synthetic biology is motivated by these agendas, progress towards synthetic biology has only been made practical by the more recent advent of two foundational technologies, DNA sequencing and synthesis. Sequencing has increased our understanding of the components and organization of natural biological systems and synthesis has provided the ability to begin to test the designs of new, synthetic biological parts (McDaniel, Kao et al. 1997; Becskei and Serrano 2000; Datsenko and Wanner 2000; Gardner, Cantor et al. 2000; Gardner and Collins 2000; De Luca and Laflamme 2001; Geerlings, Redondo et al. 2001; Cane, Kudo et al. 2002; Gerasimenko, Sheludko et al. 2002; Godfrin-Estevenon, Pasta et al. 2002; Guet, Elowitz et al. 2002; Allert, Rizk et al. 2004; Basu, Mehreja et al. 2004; Dwyer and Hellinga 2004; Kobayashi, Kaern et al. 2004) and systems (Irmler, Schroder et al. 2000; Judd, Laub et al. 2000; Bignell and Thomas 2001; Le Borgne, Palmeros et al. 2001; Martin, Yoshikuni et al. 2001; Hughes and Shanks 2002; Martin, Smolke et al. 2002; Martin, Pitera et al. 2003; Blake and Isaacs 2004; Iijima, Gang et al. 2004; Kumar, Khosla et al. 2004; Okamoto, Tanaka et al. 2004). While these examples each individually demonstrate the incredible potential of synthetic biology, they also illustrate that many foundational scientific and engineering challenges must be solved in order to make the engineering of biology routine.

References

Allert, M., S. S. Rizk, et al. (2004). "Computational design of receptors for an organophosphate surrogate of the nerve agent soman." Proc Natl Acad Sci U S A **101**(21): 7907-12.

Basu, S., R. Mehreja, et al. (2004). "Spatiotemporal control of gene expression with pulse-generating networks." Proc Natl Acad Sci U S A **101**: 6355-60.

Becskei, A. and L. Serrano (2000). "Engineering stability in gene networks by autoregulation." Nature **405**(6786): 590-3.

Bignell, C. and C. M. Thomas (2001). "The bacterial ParA-ParB partitioning proteins." J Biotechnol **91**(1): 1-34.

Blake, W. J. and F. J. Isaacs (2004). "Synthetic biology evolves." Trends Biotechnol **22**: 321-324.
Cane, D. E., F. Kudo, et al. (2002). "Precursor-directed biosynthesis: biochemical basis of the remarkable selectivity of the erythromycin polyketide synthase twoard unsaturated triketides." Chem. Biol. **9**: 131-142.
Cragg, G. M., D. J. Newman, et al. (1997). "Natural products in drug discovery and development." J. Nat. Prod. **60**: 52-60.
Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-5.
De Luca, V. and P. Laflamme (2001). "The expanding universe of alkaloid biosynthesis." Curr Opin Plant Biol **4**(3): 225-33.
Dwyer, M. A. and H. W. Hellinga (2004). "Periplasmic binding proteins: a versatile superfamily for protein engineering." Curr Opin Struct Biol **14**(4): 495-504.
Gardner, T. S., C. R. Cantor, et al. (2000). "Construction of a genetic toggle switch in Escherichia coli." Nature **403**(6767): 339-42.
Gardner, T. S. and J. J. Collins (2000). "Neutralizing noise in gene networks." Nature **405**(6786): 520-1.
Geerlings, A., F. J. Redondo, et al. (2001). "Biotransformation of tryptamine and secologanin into plant terpenoid indole alkaloids by transgenic yeast." Appl Microbiol Biotechnol **56**(3-4): 420-4.
Gerasimenko, I., Y. Sheludko, et al. (2002). "Heterologous expression of a Rauvolfia cDNA encoding strictosidine glucosidase, a biosynthetic key to over 2000 monoterpenoid indole alkaloids." Eur J Biochem **269**(8): 2204-13.
Godfrin-Estevenon, A. M., F. Pasta, et al. (2002). "The parAB gene products of Pseudomonas putida exhibit partition activity in both P. putida and Escherichia coli." Mol Microbiol **43**(1): 39-49.
Guet, C. C., M. B. Elowitz, et al. (2002). "Combinatorial synthesis of genetic networks." Science **296**(5572): 1466-70.
Hughes, E. H. and J. V. Shanks (2002). "Metabolic engineering of plants for alkaloid production." Metab Eng **4**(1): 41-8.
Iijima, Y., D. R. Gang, et al. (2004). "Characterization of geraniol synthase from the peltate glands of sweet basil." Plant Physiol **134**(1): 370-9.
Irmler, S., G. Schroder, et al. (2000). "Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase." Plant J **24**(6): 797-804.
Judd, E. M., M. T. Laub, et al. (2000). "Toggles and oscillators: new genetic circuit designs." Bioessays **22**(6): 507-9.
Kobayashi, H., M. Kaern, et al. (2004). "Programmable cells: interfacing natural and engineered gene networks." Proc Natl Acad Sci U S A **101**(22): 8414-9.
Kumar, P., C. Khosla, et al. (2004). "Manipulation and analysis of polyketide synthases." Methods Enzymol **388**: 269-93.
Le Borgne, S., B. Palmeros, et al. (2001). "Improvement of the pBRINT-Ts plasmid family to obtain marker-free chromosomal insertion of cloned DNA in E. coli." Biotechniques **30**(2): 252-4, 256.
Martin, V. J. J., D. J. Pitera, et al. (2003). "Engineering the mevalonate pathway in *Escherichia coli* for production of terpenoids." Nat. Biotechnol. **21**: 796-802.

Martin, V. J. J., C. D. Smolke, et al. (2002). "Redesigning cells for production of complex organic molecules." ASM News **68**: 336-343.
Martin, V. J. J., Y. Yoshikuni, et al. (2001). "The in vivo synthesis of plant sesquiterpenes in Escherichia coli." Biotechnol. Bioeng. **75.**: 497-503.
McDaniel, R., C. M. Kao, et al. (1997). "Engineered intermodular and intramodular polyketide synthase fusions." Chem Biol **4**(9): 667-74.
Okamoto, A., K. Tanaka, et al. (2004). "DNA logic gates." J Am Chem Soc **126**(30): 9458-63.