Synthetic biology

A foundation for multi-scale molecular biology

Adam G. Bower, Maria K. McClintock and Stephen S. Fong*

Department of Chemical and Life Science Engineering; Virginia Commonwealth University; Richmond, VA USA

Key words: synthetic biology, multi-scale, DNA synthesis, biological standards, metabolic engineering

Abbreviations: SBGN, systems biology graphical notation; SBML, systems biology markup language; SBOL, synthetic biology open language; SDK, software development kit

Submitted: 04/30/10

Revised: 05/17/10

Accepted: 05/18/10

Previously published online:
www.landesbioscience.com/journals/biobugs/article/12391

DOI:10.4161/bbug.1.5.12391

*Correspondence to: Stephen S. Fong;
Email: ssfong@vcu.edu

Commentary to: McArthur GH, 4th, Fong SS. Toward engineering synthetic microbial metabolism. *J Biomed Biotechnol* 2010;2010:459760; PMID: 20037734; DOI: 10.1155/2010/459760.

The field of synthetic biology has made rapid progress in a number of areas including method development, novel applications and community building. In seeking to make biology “engineerable,” synthetic biology is increasing the accessibility of biological research to researchers of all experience levels and backgrounds. One of the underlying strengths of synthetic biology is that it may establish the framework for a rigorous bottom-up approach to studying biology starting at the DNA level. Building upon the existing framework established largely by the Registry of Standard Biological Parts, careful consideration of future goals may lead to integrated multi-scale approaches to biology. Here we describe some of the current challenges that need to be addressed or considered in detail to continue the development of synthetic biology. Specifically, discussion on the areas of elucidating biological principles, computational methods and experimental construction methodologies are presented.

The recent review by McArthur and Fong¹ discussed the field of synthetic biology with a focus on its application to metabolic engineering. Specifically, recent developments and advances were presented within a general framework consisting of design, modeling, synthesis and analysis components. One of the larger goals of synthetic biology is to help make biological systems “engineerable,” and thus metabolic engineering is a natural application of synthetic biology. The original review was narrowly focused on metabolic engineering and details of appropriate methodologies. Here

we would like to add some broader context to the state of synthetic biology and highlight challenges and prospects for future work (Fig. 1).

New fields and technologies often initially benefit from built-up excitement and promotion concurrent with their announcement. For example, Apple’s iPhone and subsequent iPad were faced with much anticipation, excitement and high expectations when they were initially released. Many technologies or new developments receive high levels of attention during their initial release. The significant question is how best to maintain interest and excitement beyond the initial release to make the technology sustainable and integral to people’s lives (see inset Box 1). The field of synthetic biology is in a state analogous to the early stages of a new technology release where there is excitement, vast potential and a number of early adopters. One of the underlying challenges for synthetic biology is to develop a plan to maintain excitement and scientific momentum, continue its growth, and expand its utility for biological research. A critical component of this is to develop a vision for the field to include goals that will help guide synthetic biology research and provide realistic expectations from the broader community.

Since its inception, synthetic biology has undergone a rapid initial phase of growth due in part to its potential and in part to its broad accessibility. Within the overarching goal of making biology “engineerable,” a number of studies have simultaneously laid the foundation of synthetic biology and provided concrete demonstrations of the fields potential. From the early

Figure 1. Overview schematic depicting the relationship between core components of synthetic biology (standardization of parts, methods and analyses) and applications. Previous discussion¹ has focused on the application of synthetic biology to metabolic engineering.

Box 1. Expanding the iPhone analogy

For the Apple iPhone, a software development kit (SDK) was developed and released. The SDK established a level of standardization and functionality across all applications that are developed. In synthetic biology, the rough equivalent is the Registry of Standard Biological Parts whose contents follow set guidelines and are the foundation for building different biological functions or applications. Naturally, the complexity and multi-scale aspects of biology necessitate the development of additional standardized platforms that are compatible and add to the existing Registry.

construction of biological oscillators,²⁻⁶ genetic switches^{3,7-12} and logic gates,¹³⁻¹⁶ to recent developments of counting cells,¹⁷ bacterial photography¹⁸ and biosensors.¹⁹ Synthetic biology has a rapidly growing number of examples that show the potential of biological applications if we can engineer biology through implementing well-defined compatible biological parts.

Coupled with the excitement of the early successes in synthetic biology a growing number of people are officially or unofficially becoming synthetic biologists. As synthetic biology progresses and continues to develop standardized, well-characterized biological parts and methodologies, a vast range of biological research will be enabled. Researchers of all experience levels, high school through post-doctoral researchers, and from diverse backgrounds (biology to engineering) find

that they can make significant contributions to synthetic biology. There is a subtle shift occurring where biological research will be limited by the speed and creativity in generating ideas or hypotheses and not necessarily by methodologies. While not fully there yet, synthetic biology has expedited this shift, so the real need right now is identify the long-range goals of synthetic biology and to develop an NIH-style roadmap to move the field from where it is to where we want it to be. We propose that synthetic biology may provide the framework for establishing a true bottom-up approach to study biology in a multi-scale manner through foresight and consensus-building of standards.

The Registry of Standard Biological Parts (www.biobricks.org) is currently the de facto starting point for the synthetic biology community. It calls upon the community as a whole to contribute research in the collective effort to develop, characterize and deposit modular biological parts (BioBricks). By having scientists around the world contribute their biological “parts” as BioBricks into a centralized system using a specified, standardized format, the Registry of Standard Biological Parts is a storehouse of biological building blocks. This, combined with the developed 3A assembly method for combining parts,²⁰ is a powerful tool for bottom-up construction of biological systems. These developments in synthetic biology are necessary,

but not sufficient stepping-stones to make biology a truly engineerable process.

If the Registry of Standard Biological Parts is our starting point, where would we like to be? We suggest that the end goal is to study and understand the genotype-phenotype relationship with sufficient multi-scale detail that we can: (1) elucidate biological design principles (beginning at the DNA level), (2) computationally design and model systems based upon fundamental principles, and (3) control or construct biological subsystems or systems. Naturally, research progress reflects the tension between where we would like to be and what is technically possible.

Elucidating Biological Principles

Research aimed at elucidating biological design principles has been and continues to progress largely facilitated by the ability to synthesize DNA directly. If biological function is encoded and dictated by DNA, then explicitly controlling and studying the consequences of base-by-base, scientist-dictated sequences can potentially lead to elucidation of biological design principles. Ideally, the downstream consequences of changes in DNA sequence would be measured at all relevant levels leading to multi-scale characterization.

Ideally, multi-scale characterization would entail standardized analytical methods that can be used to directly measure function at each biological level. In the past, there have been attempts to standardize data reporting²¹ and to create measurement standards.²² Now might be the time to develop and implement a detailed data standard that will allow for the understanding of genetic changes at multiple levels such that every characterized biological part carries with it a wealth of technical and provenance information. If we implement a directed change in a DNA sequence, does the DNA folding change? As the DNA is transcribed to RNA, is there a change in transcription rate or level? Is the rate of translation altered affecting protein levels or is the protein structurally altered and/or altered in function? How are any associated biochemical reactions affected? Are there changes in metabolite pools that influence

additional metabolic fluxes? How do flux alterations affect cellular function? How does the altered cellular function affect organism or community behavior?

Take for example, a standard promoter that has conserved -10 and -35 regions that are important for the binding affinity for RNA polymerase. While the biological significance of promoters for initiating transcription is well-established, ongoing research is being conducted to understand the details on the kinetics of the process and specifically how genetic variation leads to different transcriptional outcomes. Detailed studies of promoters and transcription are necessary for both constitutive and inducible promoters and the methods required to study and characterize them may have subtle differences. In the characterization of promoters, a natural direct measurement would involve quantitative measurement of mRNA transcripts (real-time PCR, microarrays, RNA sequencing), however, would it also be useful to characterize the effect on pooled resources (polymerase, dNTPs, ATP, etc.) or on cellular function? The challenge is to define a standard set of information that is required to sufficiently characterize a part, but without wasting resources by employing every possible analytical technique. Efforts to carefully construct and categorize biological parts are underway²² and largely are coordinated by the Registry of Standard Biological Parts, however, community consensus on the minimal characterization standards still need to be developed. One of the first concrete steps in the area is the proposed Synthetic Biology Open Language (SBOL).

The continued development of a centralized, well-characterized repository of biological parts and information will be a critical resource for studying broader biological principles. Developing an understanding of biology without sufficient well-defined component information is like trying to use a language without knowing all the parts of speech. In the case of biological systems, the DNA is like the hieroglyphs on the Rosetta stone. Before the meaning of the hieroglyphs had been discovered, the hieroglyphs were undecipherable symbols. The functions and systems that are well defined can help provide the means to making DNA decipherable.

On the Rosetta stone, by “cracking the code”, the rules and structure that govern hieroglyphs were discovered, opening the door to examine other ancient artifacts/texts. Once we learn the proper punctuation, structure and grammar we will be able to not only read the prose and poetry, but to write it as well in a true *de novo* fashion.

Biological Modeling

A second concept in the development of synthetic biology is the development and use of computational modeling to accurately describe a system. Sound modeling approaches allow for both the testing of existing knowledge and the prediction of function. For example, a model can allow the researcher to see what effect the removal of gene would have on the function of a pathway, thus simultaneously assessing the predicted function of the gene and determining the functional outcome of a specified change.

Currently, diverse modeling methods are developed and employed to study biology for different system sizes or time-scales. This is necessary as a single method of modeling an entire biological system from DNA to phenotype with spatial and temporal accuracy is not yet feasible. Building upon biological modeling methods focused on DNA-level analyses (such as GenoCAD^{23,24}), dynamic analysis of genetic circuits (such as SynBioSS,^{25,26} TinkerCell²⁷ and differential equations²⁸) and genome-scale networks (constraint-based models and flux balance analysis²⁹⁻³²), multi-scale modeling may eventually be possible through the combination of existing techniques. The main challenge is to generate standard input-output interfaces for the different modeling methodologies to ensure compatibility. This effort has also begun with the development and evolution of the Systems Biology Markup Language (SBML)³³⁻³⁷ and the associated Systems Biology Graphical Notation (SBGN).³⁸

The development of a unified, multi-scale computational modeling pipeline could revolutionize aspects of biological research ranging from fundamental function to biological applications such as personalized medicine and metabolic

engineering. Beyond the technical challenges, specific areas that need to be addressed to fully realize an era of computer-aided or computer-driven biological research are to compile and interface the disparate modeling methods and to increase the general usability for inexperienced modelers. One can envision a time in the not-too-distant future where DNA sequence information can be analyzed and parsed into functional predictions, fed into molecular-level dynamic simulations, and analyzed within the context of a network model.

Constructing Synthetic Systems

If biological principles can be abstracted, modeled and utilized for functional predictions, the last component would be to have established methods for experimentally constructing and implementing the desired design. Along with the development of the Registry of Standard Biological Parts, methods for combining discrete BioBrick parts were established. Thus, this area has some standard and well-utilized methodologies and a number of ongoing research efforts to potentially address additional needs in synthetic construction methods. Length scales delineate the different methodologies in this area where different methodologies exist for base-by-base DNA synthesis, combination of short DNA fragments, and construction of large multi-gene pathways. Advancements in these areas are critical enabling technologies to successful experimental implementation of synthetic biology.

Currently, the greatest need is for advancements to facilitate the construction of larger, multi-gene DNA constructs. Base-by-base DNA synthesis is driven directly through instrumentation and technology improvements and there are established idempotent methods for combining BioBrick parts based upon restriction enzyme digestion and ligation protocols. The overall utility of building DNA-based biological devices (collections of BioBrick parts) will continue to increase as the number of individual BioBricks available increases through community contributions and efforts such as the recently formed BIOFAB (International Open Facility Advancing Biotechnology).

The field is rapidly moving towards the generation of large, multi-gene DNA constructs to enable more complex function. These efforts would be greatly expedited by the development of easily scalable methodologies. Direct base-by-base synthesis of large constructs is possible but cost-prohibitive. Construction based upon BioBrick assembly is also possible, but is not a naturally scalable method. Currently variants of PCR-based methodologies appear to be the most feasible approaches, but still remain costly and difficult to scale.

Summary

There have been a number of rapid developments in synthetic biology that have maintained a high-level of interest and research effort. As a young field, work has been conducted concurrently in developing methodologies, demonstrating applications and building community standards. As the field continues to mature, balanced efforts in each of these areas must proceed and with some thoughtfulness and community consensus-building, synthetic biology can lay the foundation for true multi-scale approaches to biology. The application of such an approach will have far-reaching impact on biological research.

Acknowledgements

The authors would like to thank George McArthur for discussions related to synthetic biology.

References

- McArthur GH, Fong SS. Toward engineering synthetic microbial metabolism. *J Biomed Biotechnol* 2010; 459760.
- Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature* 2000; 403:335-8.
- Atkinson MR, Savageau MA, Myers JT, Ninfa AJ. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* 2003; 113:597-607.
- Fung E, Wong WW, Suen JK, Bulter T, Lee SG, Liao JC. A synthetic gene-metabolic oscillator. *Nature* 2005; 435:118-22.
- Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J. A fast, robust and tunable synthetic gene oscillator. *Nature* 2008; 456:516-9.
- Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M. A tunable synthetic mammalian oscillator. *Nature* 2009; 457:309-12.
- Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 2000; 403:339-42.
- Bayer TS, Smolke CD. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* 2005; 23:337-43.
- Deans TL, Cantor CR, Collins JJ. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* 2007; 130:363-72.
- Dueber JE, Yeh BJ, Chak K, Lim WA. Reprogramming control of an allosteric signaling switch through modular recombination. *Science* 2003; 301:1904-8.
- Ham TS, Lee SK, Keasling JD, Arkin AP. Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLoS ONE* 2008; 3:2815.
- Ham TS, Lee SK, Keasling JD, Arkin AP. A tightly regulated inducible expression system utilizing the *fim* inversion recombination switch. *Biotechnol Bioeng* 2006; 94:1-4.
- Anderson JC, Voigt CA, Arkin AP. Environmental signal integration by a modular AND gate. *Mol Syst Biol* 2007; 3:133.
- Guet CC, Elowitz MB, Hsing W, Leibler S. Combinatorial synthesis of genetic networks. *Science* 2002; 296:1466-70.
- Rackham O, Chin JW. Cellular logic with orthogonal ribosomes. *J Am Chem Soc* 2005; 127:17584-5.
- Rinaudo K, Bleris L, Maddamsetti R, Subramanian S, Weiss R, Benenson Y. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol* 2007; 25:795-801.
- Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. Synthetic gene networks that count. *Science* 2009; 324:1199-202.
- Levskaya A, Chevalier AA, Tabor JJ, Simpson ZB, Lavery LA, Levy M, et al. Synthetic biology: engineering *Escherichia coli* to see light. *Nature* 2005; 438:441-2.
- Khalil AS, Collins JJ. Synthetic biology: applications come of age. *Nat Rev Genet* 11:367-79.
- Shetty RP, Endy D, Knight TF Jr. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2008; 2:5.
- Arkin A. Setting the standard in synthetic biology. *Nat Biotechnol* 2008; 26:771-4.
- Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, Czar MJ, et al. Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 2009; 3:4.
- Cai Y, Wilson ML, Peccoud J. GenoCAD for iGEM: a grammatical approach to the design of standard-compliant constructs. *Nucleic Acids Res* 38:2637-44.
- Czar MJ, Cai Y, Peccoud J. Writing DNA with GenoCAD. *Nucleic Acids Res* 2009; 37:40-7.
- Kaznessis YN. Computational methods in synthetic biology. *Biotechnol J* 2009; 4:1392-405.
- Hill AD, Tomshine JR, Weeding EM, Sotiropoulos V, Kaznessis YN. SynBioSS: the synthetic biology modeling suite. *Bioinformatics* 2008; 24:2551-3.
- Chandran D, Bergmann FT, Sauro HM. TinkerCell: modular CAD tool for synthetic biology. *J Biol Eng* 2009; 3:19.
- Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* 2005; 436:588-92.
- Schellenberger J, Park JO, Conrad TM, Palsson BO. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11:213.
- Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol* 28:245-8.
- Hyduke DR, Palsson BO. Towards genome-scale signalling-network reconstructions. *Nat Rev Genet* 11:297-307.
- Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5:93-121.
- Hucka M, Finney A, Bornstein BJ, Keating SM, Shapiro BE, Matthews J, et al. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Syst Biol (Stevenage)* 2004; 1:41-53.
- Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003; 19:524-31.
- Keating SM, Bornstein BJ, Finney A, Hucka M. SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics* 2006; 22:1275-7.
- Finney A, Hucka M. Systems biology markup language: Level 2 and beyond. *Biochem Soc Trans* 2003; 31:1472-3.
- Gillespie CS, Wilkinson DJ, Proctor CJ, Shanley DP, Boys RJ, Kirkwood TB. Tools for the SBML Community. *Bioinformatics* 2006; 22:628-9.
- Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The systems biology graphical notation. *Nat Biotechnol* 2009; 27:735-41.