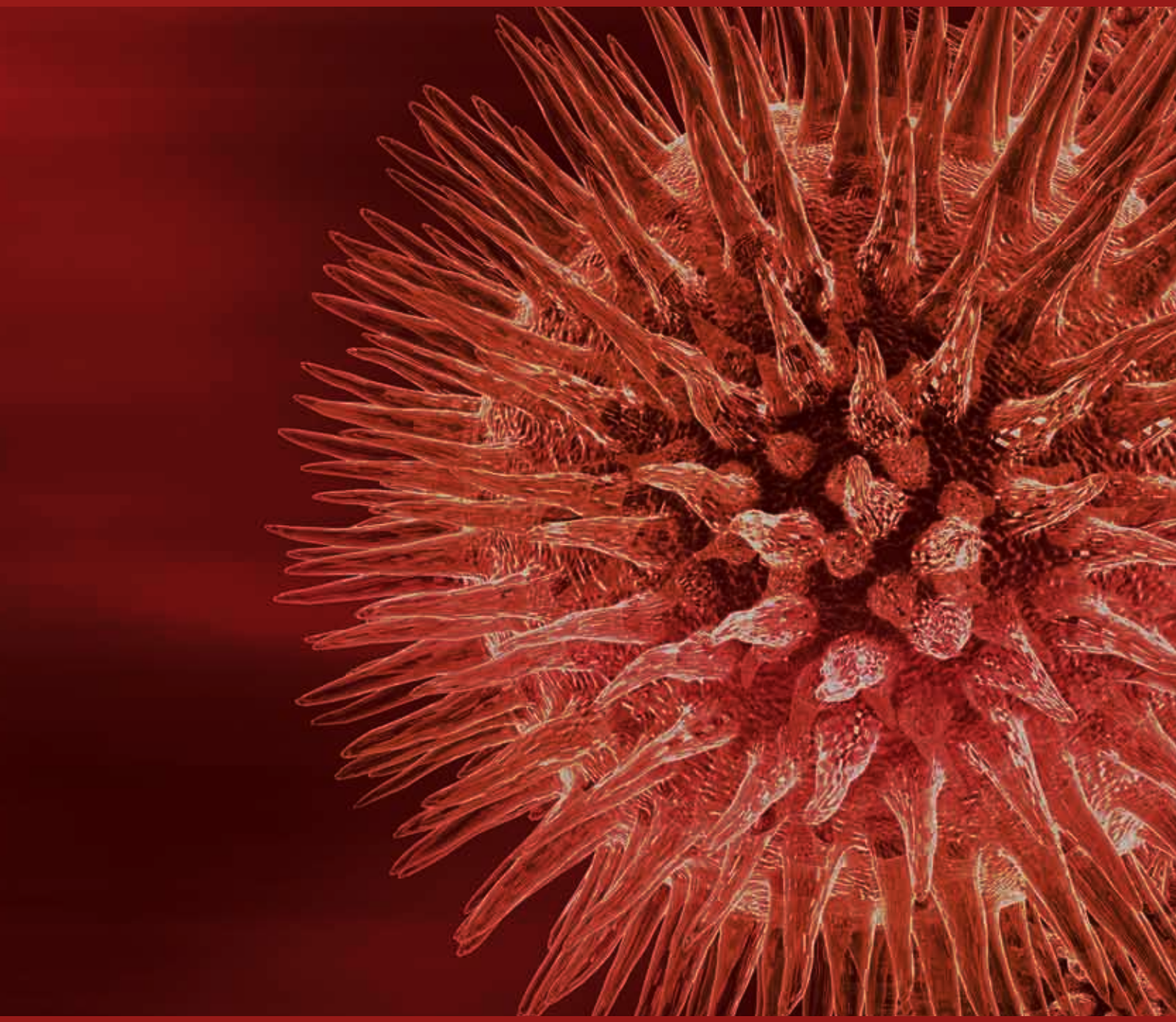
Journal of Biomedicine and Biotechnology

Applications of Synthetic Biology in Microbial Biotechnology

Guest Editors: Hal Alper, Patrick Cirino, Elke Nevoigt, and Ganesh Sriram



Applications of Synthetic Biology in Microbial Biotechnology

Journal of Biomedicine and Biotechnology

Applications of Synthetic Biology in Microbial Biotechnology

Guest Editors: Hal Alper, Patrick Cirino, Elke Nevoigt, and Ganesh Sriram

Editorial Board

The editorial board of the journal is organized into sections that correspond to the subject areas covered by the journal.

Agricultural Biotechnology

Guihua Bai, USA
Christopher P. Chanway, Canada
Ravindra N. Chibbar, Canada
Ian Godwin, Australia
Hari B. Krishnan, USA
Carol A. Mallory-Smith, USA
Dennis P. Murr, Canada
Rodomiro Ortiz, Mexico
Badal C. Saha, USA
Mariam B. Sticklen, USA
Chiu-Chung Young, Taiwan

Animal Biotechnology

Ernest S. Chang, USA
Hans H. Cheng, USA
Bhanu P. Chowdhary, USA
Noelle E. Cockett, USA
Peter Dovc, Slovenia
Scott C. Fahrenkrug, USA
Dorian J. Garrick, USA
Thomas A. Hoagland, USA
Tosso Leeb, Switzerland
James D. Murray, USA
Anita M. Oberbauer, USA
Jorge A. Piedrahita, USA
Daniel Pomp, USA
Kent M. Reed, USA
Lawrence Reynolds, USA
Sheila M. Schmutz, Canada
Lawrence B. Schook, USA
Mari A. Smits, The Netherlands
Leon Spicer, USA
John P. Verstegen, USA
Matthew B. Wheeler, USA
Kenneth L. White, USA

Biochemistry

Robert Blumenthal, USA
David R. Brown, UK
Saulius Butenas, USA
Vittorio Calabrese, Italy
Francis J. Castellino, USA
Roberta Chiaraluce, Italy
David M. Clarke, Canada
Francesca Cutruzzolà, Italy
Paul W. Doetsch, USA
Hicham Fenniri, Canada
Nick V. Grishin, USA
J. Guy Guillemette, Canada
Paul W. Huber, USA
Chen-Hsiung Hung, Taiwan
Michael Kalafatis, USA
Bruce E. Kemp, Australia
Phillip E. Klebba, USA
Wen-Hwa Lee, USA
Richard D. Ludescher, USA
George Makhatadze, USA
Leonid Medved, USA
Susan A. Rotenberg, USA
Jason Shearer, USA
Mark Smith, USA
Andrei Surguchov, USA
John B. Vincent, USA
Yujun George Zheng, USA

Bioinformatics

Tatsuya Akutsu, Japan
Miguel A. Andrade, Germany
Mark Borodovsky, USA
Rita Casadio, Italy
Artem Cherkasov, Canada
David Wolfe Corne, UK
Sorin Draghici, USA
Stavros J. Hamodrakas, Greece
Paul Harrison, Canada
George Karypis, USA
J. A. Leunissen, The Netherlands
Guohui Lin, Canada
Satoru Miyano, Japan
Zoran Obradovic, USA
Florencio Pazos, Spain
Zhirong Sun, China
Ying Xu, USA
Alexander Zelikovsky, USA
Albert Zomaya, Australia

Biophysics

Miguel Castanho, Portugal
P. Bryant Chase, USA
Kuo-Chen Chou, USA
Rizwan Khan, India
Ali A. Khraibi, Saudi Arabia
Rumiana Koynova, USA
Serdar Kuyucak, Australia
Jianjie Ma, USA
S. B. Petersen, Denmark
Peter Schuck, USA
Claudio M. Soares, Portugal

Cell Biology

Ricardo Benavente, Germany
Omar Benzakour, France
Sanford I. Bernstein, USA
Phillip I. Bird, Australia
Eric Bouhassira, USA
Mohamed Boutjdir, USA
Chung-Liang Chien, Taiwan
Richard Gomer, USA
Paul J. Higgins, USA
Pavel Hozak, Czech Republic
Xudong Huang, USA
Anton M. Jetten, USA
Seamus J. Martin, Ireland
Manuela Martins-Green, USA
Shoichiro Ono, USA
George Perry, USA
Mauro Piacentini, Italy
George E. Plopper, USA
Lawrence Rothblum, USA
Ulrich Scheer, Germany
Michael Sheetz, USA
James L. Sherley, USA
G. S. Stein, USA
Richard Tucker, USA
Thomas van Groen, USA
Andre Van Wijnen, USA
Steve Winder, UK
Chuanyue Wu, USA
Bin-Xian Zhang, USA

Genetics

Adewale Adeyinka, USA
Claude Bagnis, France
James Birchler, USA
Susan Blanton, USA
Barry J. Byrne, USA
Ranajit Chakraborty, USA
Domenico Coviello, Italy
Sarah H. Elsea, USA
Celina Janion, Poland
J. Spencer Johnston, USA
M. Ilyas Kamboh, USA
Feige Kaplan, Canada
Manfred Kayser, The Netherlands
Brynn Levy, USA
Xiao Jiang Li, USA
Thomas Liehr, Germany
James M. Mason, USA
Mohammed Rachidi, France
Raj S. Ramesar, South Africa
Elliot D. Rosen, USA
Dharambir K. Sanghera, USA
Michael Schmid, Germany
Markus Schuelke, Germany
Wolfgang A. Schulz, Germany
Jorge Sequeiros, Portugal
Mouldy Sioud, Norway
Meena Upadhyaya, UK
Rongjia Zhou, China

Genomics

Vladimir Bajic, Saudi Arabia
Margit Burmeister, USA
Settara Chandrasekharappa, USA
Yataro Daigo, Japan
J. Spencer Johnston, USA
Vladimir Larionov, USA
Thomas Lufkin, Singapore
Joakim Lundeberg, Sweden
John L. McGregor, France
John V. Moran, USA
Yasushi Okazaki, Japan
Gopi K. Podila, USA
Paul B. Samollow, USA
Momiao Xiong, USA

Immunology

Hassan Alizadeh, USA
Peter Bretscher, Canada
Robert E. Cone, USA
Terry L. Delovitch, Canada
Anthony L. DeVico, USA
Nick Di Girolamo, Australia
Dorothy Mark Estes, USA
Soldano Ferrone, USA
Jeffrey A. Frelinger, USA
John Gordon, UK
John Robert Gordon, Canada
James D. Gorham, USA
Silvia Gregori, Italy
Thomas Griffith, USA
Young S. Hahn, USA
Dorothy E. Lewis, USA
Bradley W. McIntyre, USA
R. Lee Mosley, USA
Marija Mostarica-Stojković, Serbia
Hans Konrad Muller, Australia
Ali Ouaissi, France
Kanury V. S. Rao, India
Yair Reisner, Israel
Harry W. Schroeder, USA
Wilhelm Schwaeble, UK
Nilabh Shastri, USA
Yufang Shi, China
Piet Stinissen, Belgium
Hannes Stockinger, Austria
J. W. Tervaert, The Netherlands
Graham R. Wallace, UK

Microbial Biotechnology

Jozef Anné, Belgium
Yoav Bashan, Mexico
Marco Bazzicalupo, Italy
Nico Boon, Belgium
Luca Simone Cocolin, Italy
Peter Coloe, Australia
Daniele Daffonchio, Italy
Han de Winde, The Netherlands
Yanhe Ma, China
Bernd Rehm, New Zealand
Angela Sessitsch, Austria
Effie Tsakalidou, Greece
Juergen Wiegel, USA

Microbiology

David Beighton, UK
Steven R. Blanke, USA
Stanley Brul, The Netherlands
Isaac K. O. Cann, USA
Peter Dimroth, Switzerland
Stephen K. Farrand, USA
Alain Filloux, UK
Gad Frankel, UK
Roy Gross, Germany
Hans-Peter Klenk, Germany
Tanya Parish, UK
Gopi K. Podila, USA
Frederick D. Quinn, USA
Didier Raoult, France
Isabel Sá-Correia, Portugal
Pamela L. C. Small, USA
Lori Snyder, UK
Michael Thomm, Germany
Henny van der Mei, The Netherlands
Schwan William, USA

Molecular Biology

Rudi Beyaert, Belgium
Michael Bustin, USA
Douglas Cyr, USA
Kostas Iatrou, Greece
Lokesh Joshi, Ireland
David W. Litchfield, Canada
Noel F. Lowndes, Ireland
Wuyuan Lu, USA
Patrick Matthias, Switzerland
John L McGregor, France
Sherry Mowbray, Sweden
Elena Orlova, UK
Yeon-Kyun Shin, USA
William S. Trimble, Canada
Lisa Wiesmuller, Germany
Masamitsu Yamaguchi, Japan

Oncology

Colin Cooper, UK
F. M. J. Debruyne, The Netherlands
Nathan Ames Ellis, USA
Dominic Fan, USA
Gary E. Gallick, USA
Daila S. Gridley, USA
Xin-yuan Guan, Hong Kong
Anne Hamburger, USA
Manoor Prakash Hande, Singapore
Beric Henderson, Australia
Steve B. Jiang, USA
Daehee Kang, South Korea
Abdul R. Khokhar, USA
Rakesh Kumar, USA
Macus Tien Kuo, USA
Eric W Lam, UK
Sue-Hwa Lin, USA
Kapil Mehta, USA
Orhan Nalcioglu, USA
Vincent C. O. Njar, USA
Peter J. Oefner, Germany
Allal Ouhtit, USA
Frank Pajonk, USA
Waldemar Priebe, USA
Fernando Carlos Schmitt, Portugal
Sonshin Takao, Japan
Ana M. Tari, USA
Henk G. Van Der Poel, The Netherlands
Haodong Xu, USA
David J. Yang, USA

Pharmacology

Abdel A. Abdel-Rahman, USA
Mostafa Z. Badr, USA
Stelvio M. Bandiera, Canada
Ronald E. Baynes, USA
R. Keith Campbell, USA
Hak-Kim Chan, Australia
Michael D. Coleman, UK
Jacques Descotes, France
Dobromir Dobrev, Germany
Ayman El-Kadi, Canada
Jeffrey Hughes, USA
Kazim Husain, USA
Farhad Kamali, UK
Michael Kassiou, Australia
Joseph J. McArdle, USA
Mark McKeage, New Zealand
Daniel T. Monaghan, USA
T. Narahashi, USA
Kennerly S. Patrick, USA
Vickram Ramkumar, USA
Michael J. Spinella, USA
Quadiri Timour, France
Todd W. Vanderah, USA
Val J. Watts, USA
David J. Waxman, USA

Plant Biotechnology

Prem L. Bhalla, Australia
Jose Botella, Australia
Elvira Gonzalez De Mejia, USA
H. M. Häggman, Finland
Liwen Jiang, Hong Kong
Pulugurtha B. Kirti, India
Yong Pyo Lim, South Korea
Gopi K. Podila, USA
Ralf Reski, Germany
Sudhir Kumar Sopory, India
Neal Stewart, USA

Toxicology

Michael Aschner, USA
Michael L. Cunningham, USA
Laurence D. Fechter, USA
Hartmut Jaeschke, USA
Youmin James Kang, USA
M. Firoze Khan, USA
Pascal Kintz, France
Ronald Tjeerdema, USA
Kenneth Turteltaub, USA
Brad Upham, USA

Virology

Contents

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 918391, 2 pages
doi:10.1155/2010/918391

Editorial

Applications of Synthetic Biology in Microbial Biotechnology

Hal Alper,[1] Patrick Cirino,[2] Elke Nevoigt,[3] and Ganesh Sriram[4]

[1] *Department of Chemical Engineering, The University of Texas at Austin, 1 University Station, C0400, Austin, TX 78712, USA*
[2] *Department of Chemical Engineering, The Pennsylvania State University, 226A Fenske Laboratory, University Park, PA 16802-4400, USA*
[3] *Molecular Biotechnology, Jacobs University gGmbH, School of Engineering and Science, Research II - Room 113, Campus Ring 1, 28759 Bremen, Germany*
[4] *Department of Chemical and Biomolecular Engineering, University of Maryland, 1208D, Chemical and Nuclear Engineering Building 090, College Park, MD 20742, USA*

Correspondence should be addressed to Hal Alper, halper@che.utexas.edu

Received 12 December 2010; Accepted 12 December 2010

Recent advances in synthetic biology have greatly expanded the capacity to improve pathway performance and cellular phenotype. The design, use, and importation of constructed genetic control elements, optimized genes, and functional genetic circuits can be used to modulate the function of metabolic pathways. The modularity and hierarchy of parts, devices, systems, and chassis provide a framework for studying and controlling cellular functions. Moreover, synthetic biology provides an interface where protein engineering, directed evolution, DNA synthesis, and *in silico* design can direct the field of metabolic engineering. Many of these advances are influencing how scientists and engineers view and perturb cellular systems. Perhaps the paramount application of these tools is in the arena of microbial biotechnology where the overarching goal is to rewire cellular systems for biochemical production. This task is quite complex given the complexity and interconnectedness of metabolism and regulation. Synthetic biology approaches provide a way to both intervene in as well as bypass cellular complexities.

This special issue contains six review articles that highlight two major aspects of synthetic biology: (1) general design principles and modeling of synthetic biology and (2) applications of synthetic biology to biochemical and biofuels production. The articles are ordered based on these two areas. In the first class, review articles describe the basic paradigm of synthetic biology and include the design and modeling of systems.

In the first article, E. Young and H. Alper discuss the areas of synthetic biology within a process engineering paradigm, through a three-tiered model that uses the central dogma of biology as a pivot. The first tier comprises the units of the central dogma and the possible flows between them (DNA, transcription, RNA, translation, and protein), the second tier includes intrinsic interactions between these units by means of regulatory mechanisms, and the third tier comprises external or environmental influences on the units. This conceptualization permits engineering approaches in synthetic biology to be classified by which components they influence. The authors provide several examples to illustrate how synthetic biology approaches can be combined toward realizing the goal of integrative synthetic biology, which the authors exemplify with two examples from the literature.

In the second article, Fritz et al. discuss a "biology by design" paradigm, which involves the conceptualization, design, and construction of a synthetic biological system, followed by the expectation that the system will perform as desired. Surmounting the challenges en route to this ambitious objective will require answering several "top-down" and "bottom-up" design questions. Bottom-up questions relate to the parts and modules of the synthetic biological system as well as their characterization and assembly in isolation, whereas top-down questions focus on desirable and undesirable interactions between the synthetic system and its biological context. This general framework can be applied to problems of increasing biological sophistication, ranging from the engineering of novel polymers, metabolic

networks, and cellular devices. The authors illustrate the applications of these ideas to three areas: biochemical transformations, cellular devices, and therapies as well as engineering the chemistry of life.

The third paper highlights how mathematical models can predict the dynamics of a network under different conditions. This paper by Y. Zheng and G. Sriram provides detailed insights into mathematical modeling concepts and methodologies as relevant to synthetic biology. Any model is formulated on the basis of certain assumptions regarding the system under investigation as discussed by the authors. The paper introduces the two broad types of modeling frameworks: deterministic and stochastic modeling. The importance of parameter estimation and optimization in modeling is emphasized. In addition, mathematical techniques used to analyze a model such as sensitivity analysis and bifurcation analysis are presented. The authors discuss the role of modeling in phenotype analysis and conclude with three in-depth case studies about mathematical modeling in the framework of synthetic biology.

The second class of articles highlights examples and applications of synthetic biology for improved biofuels and biochemicals production. M. R. Connor and S. Atsumi review progress in improving microbial production of renewable liquid fuels and the unique role of synthetic biology in advancing these efforts. The fuel challenges facing society are first highlighted. They then provide an overview of research for the production of fermentative alcohols, nonfermentative higher alcohols, isoprenoids, and fatty acids. The importance of photosynthetic organisms in biofuels research is also discussed, and the reliance on synthetic biology tools to improve productivities and titers is emphasized.

Next, L. R. Jarboe et al. review the combined use of synthetic biology tools with traditional metabolic engineering strategies for designing and improving microbial production of chemicals from renewable resources. They begin with an overview of technologies and tools available for biocatalyst redesign. They next describe strain engineering examples in which existing pathways were modified for the overproduction of various chemicals (e.g., succinate and D-lactate). These examples are contrasted with strain re-design via the introduction of foreign or nonnatural genes and pathways for production of compounds such as L-alanine and xylitol. The latter scenario includes the use of enzymes which have been engineered to have unnatural activities.

Finally, G. H. McArthur IV and S. S. Fong focus in their paper on the impact of synthetic biology on engineering microbial metabolism. They discuss recent progress and current challenges in applying synthetic biology to metabolic engineering. In particular, the authors separately focus on each of the four workflow steps, design, modeling, synthesis, and analysis, and distinguish between parts (individual functional units) and pathways (part-based systems) for each single level.

Collectively, these papers describe the current state of synthetic biology in the context of microbial biotechnology. Current and ongoing advances in this field will certainly propel our capacity to engineer improved microorganisms in the future. In this regard, these articles all provide prospective into the future of this field.

Hal Alper
Patrick Cirino
Elke Nevoigt
Ganesh Sriram

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 130781, 12 pages
doi:10.1155/2010/130781

Review Article

Synthetic Biology: Tools to Design, Build, and Optimize Cellular Processes

Eric Young and Hal Alper

Department of Chemical Engineering, The University of Texas at Austin, 1 University Station, C0400, Austin, TX 78712, USA

Correspondence should be addressed to Hal Alper, halper@che.utexas.edu

Received 5 August 2009; Accepted 28 October 2009

Academic Editor: Ganesh Sriram

The general central dogma frames the emergent properties of life, which make biology both necessary and difficult to engineer. In a process engineering paradigm, each biological process stream and process unit is heavily influenced by regulatory interactions and interactions with the surrounding environment. Synthetic biology is developing the tools and methods that will increase control over these interactions, eventually resulting in an integrative synthetic biology that will allow ground-up cellular optimization. In this review, we attempt to contextualize the areas of synthetic biology into three tiers: (1) the process units and associated streams of the central dogma, (2) the intrinsic regulatory mechanisms, and (3) the extrinsic physical and chemical environment. Efforts at each of these three tiers attempt to control cellular systems and take advantage of emerging tools and approaches. Ultimately, it will be possible to integrate these approaches and realize the vision of integrative synthetic biology when cells are completely rewired for biotechnological goals. This review will highlight progress towards this goal as well as areas requiring further research.

1. Introduction

The central dogma of biology is simply and elegantly stated; however it is less straightforward to engineer, control, and rewire for biotechnological purposes. This difficulty stems from our limited understanding of the multiscale, and often stochastic, operation, and regulation of biological systems [1–3]. Nevertheless, rapid progress in uncovering the basic framework and information flow within the central dogma has helped fuel the current biotechnological revolution. Yet, elucidating the specific components and control mechanisms inherent in this process has lagged significantly [4–6]. This limitation prevents the creation of custom-built cellular factories using modeling and *de novo* design. However, this limitation is only temporary. Recent advances in high-throughput biology are quickly uncovering the identity and details of these components and control schemes [7–10]. While not yet complete, this global, systems biology approach repeatedly depicts the central dogma as a multistep process subject to exquisite regulatory mechanisms established to maintain cellular homeostasis and to respond to environmental stimuli. Once our understanding is advanced, it will be possible to synthetically create desired functions at all levels of the central dogma.

The integrative complexity of the central dogma (and biological systems in general) has analogies and parallels to chemical or electrical systems. The rationale for drawing these analogies is twofold: (1) it helps to contextualize the various parts of a cellular process and (2) it facilitates the possible transfer of knowledge between the analogous systems. In this regard, understanding the central dogma processes, the process controls, and the environmental influences within a cell is as vital as understanding analogous components within a traditional chemical factory. Uncovering and studying these components will ultimately lead to a factory-like cellular blueprint—a detailed catalogue of parts, interactions, and functions. Moreover, compiling such a blueprint for all species will expand the number of parts we are able to access, characterize, and employ when trying to design cells and circuits from scratch. Thus, this understanding will enhance our ability to predict, control, and design cellular systems—major tenets in the emerging field of synthetic biology.

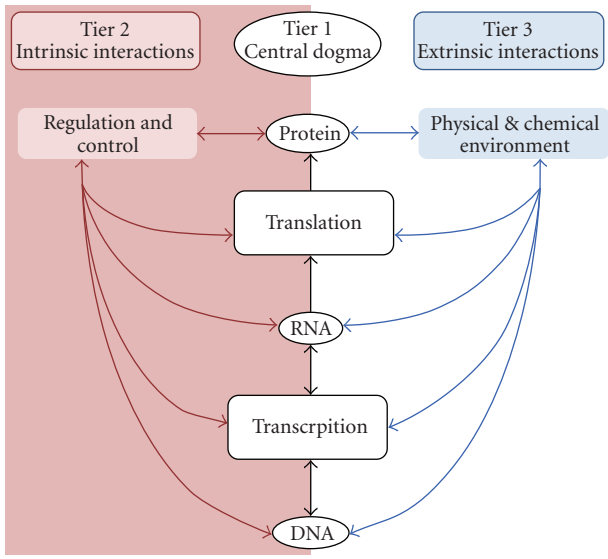


Figure 1: *The central dogma with regulatory and environmental influences acting upon the process.* The two central dogma process units (transcription and translation) and the three process streams (DNA, RNA, and protein) are depicted. These units and streams are all subject to control by both internal, regulatory and external, environmental conditions. These influences alter the central dogma process and regulatory mechanisms. Large bold arrows are used to indicate that proteins are the major workhorses of the cell, participating both in regulatory mechanisms and responding to the environment. Whether the system of interest is a signaling cascade or metabolic pathway, proteins are essential components and must become well understood and modifiable to bring about ground-up cellular optimization. Synthetic biology is developing tools to modify and control each unit and stream in this process.

Due to its youth, the field of synthetic biology has yet to have a concrete, comprehensive definition. Yet, in its broadest sense, *synthetic biology aims to harness the emergent properties of the central dogma* for biotechnological and human use. This description of the field is comprehensive since even synthetically designed biological circuits actually interface with existing central dogma machinery in the cell. In this regard, tools for synthetic biology harness the complexity of the central dogma process in a predictable, designed fashion.

Within the context of engineering the central dogma, the seemingly wide variety of themes and aims in the synthetic biology research field become more unified. Considering the central dogma as a simple process diagram (Figure 1), it can be seen that the varied areas of synthetic biology research all influence the central dogma albeit at different access points in the process. As Figure 1 illustrates, this system has three tiers, specifically: (1) the central dogma *process units* (transcription and translation) and associated streams (DNA, RNA, and protein), (2) the *intrinsic* regulatory mechanisms in the cells, and (3) the *extrinsic* physical and chemical environment of the cells. These three tiers are depicted separately, but in reality are thoroughly enmeshed with one another as a result of evolved biological complexity. Yet, this very complexity provides a multitude of access points, or nodes, for synthetic biologists to engineer.

Synthetic biology research is at the forefront of engineering the three tiers of biological systems. For example, the newly developed ability to design and chemically synthesize genetic sequences [11–13] provides a greater ability to manipulate DNA, the "feed stream" molecule for the first tier. Contributions from systems biology have broadened our ability to understand and engineer biological networks [14–18], providing impetus for modifying tier two intrinsic control systems and tier three extrinsic signaling interactions. Other frontiers in synthetic biology have greatly expanded our capacity to construct and improve pathways and global cellular phenotype [19–24], which engineers the third tier interaction between proteins and the chemical environment. In the same vein, protein engineering provides the synthetic biologist a great deal of flexibility for introducing and optimizing new function at any node [25–31], since proteins are such universal components throughout the central dogma process. All of these areas of synthetic biology are building toward a single goal: integrative control of the central dogma for biotechnological and human use.

From this viewpoint, developing powerful new tools that manipulate biology at each of the three tiers will empower scientists and engineers with the ability to rewire and program cellular systems for both medical and biotechnological applications. Combining these tools to work in concert would define the field of *integrative synthetic biology*. This culminating point of synthetic biology development will usher in the age of ground-up cellular design and optimization. However, much of current synthetic biology research is focused on tool development, a required foundation for integrative synthetic biology. As a result, it is not yet clear how to best integrate these approaches. Therefore, the purpose of this review is to provide an overview of synthetic biology research, focusing on microbial hosts, and to highlight areas where more work must be done before realizing the potential of ground-up synthetic cellular engineering.

2. The First Tier—Process Optimization of the Central Dogma

The first tier of synthetic biology focuses on altering the general process flow—specifically modifications to the function and behavior of the process units (transcription [32] and translation) and the associated process streams (DNA, RNA [33], and protein). These manipulations are made possible through detailed knowledge of the central dogma process. While this capacity has existed for several decades [34], novel capabilities and genetic tools afforded by synthetic biology may help overcome some of the limitations and time-consuming bottlenecks inherent in established techniques. In this regard, synthetic biology aims to develop foundational technologies such as large-scale, economical *de novo* DNA synthesis [35] that would increase the efficiency of traditional recombinant DNA technology and genetic engineering. Collectively, synthetic engineering of the central dogma aims to optimize and expand the capabilities of native cellular machinery. The methods and technologies

developed from this research will contribute to a more powerful and efficient toolbox for the microbial engineer. In this section, we will review synthetic biology technologies and applications for influencing components within the first tier.

2.1. Engineering DNA. DNA manipulation began very early in the biotechnological revolution with recombinant DNA methods [36–38] and DNA sequencing technology [39–41]. Mutagenesis techniques and the establishment of standardized molecular biology methods [34] expanded these tools and empowered metabolic engineers with more powerful approaches to improve metabolic phenotypes [42–46]. Despite being straightforward and robust, these approaches are inherently limited by template-based DNA synthesis and restriction enzyme cloning. However, inexpensive, large scale synthetic (*de novo*) DNA manufacturing technology has the potential to revolutionize this process once again. Unlike traditional methods, *de novo* synthesis removes the need to engineer cellular systems using preexisting DNA as a template. In this regard, this technology brings about a new power to synthetically design genes, control elements, and circuits that do not exist in nature—thus creating novel function from the basic building blocks of nucleic acids and amino acids.

Already, there are multiple companies with expertise in synthesizing DNA (Blue Heron, DNA2.0, GENEART, IDT, etc.), from small fragments to whole genes and genetic elements. Moreover, improvements and new technologies are continually being published [12, 35, 47–50] which expand the potential applications and drive down prices. As a result, synthesis capabilities have moved beyond the scale of single genes and into the scale of chemically synthesized genomes [11, 13]. Moreover, efforts are being made to introduce this synthetic DNA into a generic host [51] in an effort to completely reprogram a cell. The combination of these powerful new DNA synthesis techniques coupled with low-cost DNA sequencing has the potential to confer a great deal of freedom to researchers. With these advances, DNA design and cloning is no longer limited by existing fragments of template DNA and available restriction sites in plasmids. In essence, this technology serves as the basis for other synthetic biology tools, since DNA is the vehicle of almost every biological perturbation, regardless of the tier of interest.

However, our ability to create DNA *de novo* is not equally matched by a capacity to predict the ideal DNA sequence *a priori* for a given application. Attempts have been made to catalogue DNA elements [52, 53] and predict the function of synthetic networks using models [9, 15, 17, 18, 54]. Nevertheless, our knowledge base for constructing predictive models of global cellular behavior is limited as is our ability to design large operons and circuits *de novo*. Future work on characterizing these elements as well as their dynamics and interaction will allow for synthetically created custom-designed genetic circuits.

Simply synthesizing and importing designed DNA is not enough to ensure desired function. Specifically, for these elements to operate efficiently, synthetic DNA operons must act independently and not be negatively influenced by other cellular processes. One solution to mitigate this problem embodies another area of synthetic DNA engineering research: the quest for a minimal cell [55–57]. A minimal cell only contains the essential genetic information required to maintain viability under controlled conditions. In following with the industrial process analogy, this would correspond to a factory containing only the equipment necessary for a given process application. It is clear that this minimization makes sense in a process plant as superfluous equipment would be a waste of precious resources such as money and space. However, cells contain many more parts than are necessary for a given biotechnological application. Thus, taking a cell "off the shelf" can result in limited efficiency. The search for a minimal cell provides a noninterfering "chassis" suitable for manipulation by the biological engineer. Recent advances in cataloging essential genes continue to move the minimal cell closer to reality [58, 59]. However, it is currently unclear whether the genetic definition of a minimal cell will be generic or process specific. Thus, there may be a suite of minimal cells required; each one suited for different classes of bioproducts.

Another area of synthetic DNA engineering aims to expand the basic genetic code by adding synthetic base pairs [60–63]. Incorporating synthetic codons provides a means of utilizing nonnatural amino acids (see Section 2.3) and introducing nonnative DNA-protein binding pairs. Already, alternative genetic codes have led to new applications for engineered biology [61]. One of the potential difficulties of incorporating synthetic base pairs into DNA is that the three-dimensional structure of the molecule may change and key binding proteins and polymerases may not be able to recognize the new genetic language. However, initial results are promising [63, 64] and suggest that drastic changes to innate cellular architecture are not required. Thus, alternative base pairs provide a newfound flexibility in genetic code and DNA manipulation technology. Furthermore, this approach is an excellent application for *de novo* DNA synthesis: the coupling of synthetic base pairs with DNA synthesis technology could create a powerful tool for designing synthetic circuits. Regardless of the application, the capacity to engineer DNA using synthetic biology tools provides new access points to the cell unachievable by previous technology.

2.2. Engineering Transcription. Since the central dogma is so highly integrated, DNA-level perturbations can cause significant alterations in downstream process units (Figure 1). As a result, microbial engineers must be able to synthetically optimize each of the process units. The first process unit in the central dogma is transcription. A large number of proteins, small molecules, and even small RNAs can participate in this process step [5]. Nevertheless, the ultimate goal of this process is RNA transcription—converting DNA into an mRNA message. As a result, synthetic control of this process step influences the rate and capacity of mRNA synthesis.

Not surprisingly, the key step of RNA polymerase II binding to a promoter sequence has been targeted by synthetic

tools, such as promoter engineering, for the purposes of controlling gene expression levels [65–68]. By creating a library of promoter sequence mutants, a graduated expression profile can be developed. This resulting range of expression affords a more detailed investigation of expression levels beyond traditional wild type—knockout—strong overexpression studies. Furthermore, well-characterized promoters enable more precise gene delivery [52]. A similar requirement for controlled expression is critical for genetic circuits where protein expression must be balanced to maintain a desired steady state. Often these circuits use inducible promoters, and a similar approach can be used to augment the expression capacity of inducible promoters. Thus, well-documented genetic elements will be extremely useful in creating synthetic cells and circuits. However, transcription is a two-body problem requiring both proteins and DNA. Most previous work focused on the DNA aspect of the problem; however, proteins involved in transcription can also be engineered to synthetically control a cell [32]. Moreover, altering the DNA sequence focuses the change to one particular genetic locus, whereas changing the involved proteins has a profound, global impact.

It is often necessary to alter the transcriptional profile of many genes simultaneously to obtain a desired complex phenotype. This level of synthetic control in the cell is essential for rewiring cells into biofactories. In this regard, another synthetic biology tool termed global transcription machinery engineering (gTME) [30, 69, 70] aims to alter the proteins responsible for the process step of transcription in an effort to exert a pleiotropic downstream effect. The gTME approach operates by creating a mutant library of proteins responsible for transcription (such as sigma factors and TATA binding proteins) and subjecting the library to a high-throughput phenotype screen. This technique is useful for a phenotype that is typically under the control of a multitude of genes. This approach of synthetically rewiring cells at the transcriptional level provides a means of creating large changes within the transcriptome and provides a novel approach to modulating the process step of transcription.

The rationale behind gTME was recently applied to the RNA polymerase II protein itself [32]. By creating a mutant library of the polymerase α subunit and applying selective pressures, the authors demonstrated increases in the tolerance of *E. coli* to 1-butanol. More studies such as these are required not only to optimize the transcription process unit for synthetic biology applications, but also to gain more fundamental knowledge of the process unit. With enough information, rational design of synthetic transcriptional machinery may be possible. However, large-library based selection techniques are currently required to identify promising mutants, limiting the capacity to design transcription machinery *de novo*.

Also of note for future synthetic biology tool development is reverse transcription, illustrated by the double-pointed arrows in Figure 1. Reverse transcription as a method of gene delivery for disease treatment [71, 72] merits further exploration by synthetic biologists, although this type of work is generally outside the focus of microbial synthetic biologists. Yet, a great deal of work still remains prior to gaining full, synthetic control of the transcription process. Specifically, more studies focusing on the complex interactions between participatory molecules must be performed. These studies will also implicate future molecular targets for strategies similar to promoter engineering and gTME.

2.3. Engineering Translation. The second major process unit in the central dogma, translation, has also been the subject of recent synthetic biology research. Similar to transcription, translation encompasses many different classes of molecules that can serve as good targets for optimization and rewiring cells. However, less is known about the most essential molecules in this process.

One of the most successful examples of synthetically engineering translation machinery involves the incorporation of unnatural amino acids into proteins [73]. In this work, mutant aminoacyl tRNA synthetases incorporate amino acids with diverse R-group chemistries into proteins. This approach holds a great deal of promise for the synthetic biologist as a means of creating wholly new biological functions and chemistries [29, 74–76]. This direction towards designer proteins is akin to nontemplate based DNA synthesis. However, as with *de novo* DNA synthesis, it is not always clear which amino acid(s) should be changed to an unnatural analog to confer a desired protein function of interest. In this regard, more work is required to develop a computational linkage between sequence and function. Nevertheless, this approach gives significant leverage to synthetic biologists to create custom proteins with desired functions.

Another example of engineering translation is gene codon optimization [33]. Codon optimization is a method to bias the redundant codons for each amino acid toward the codons most commonly found in the host organism. This approach is greatly expedited by sequencing and synthesis technologies that can produce the required alternately coded genes. Codon optimization has been shown to be successful in many cases [77–82] and has led to improved translation rates, protein yields, and enzymatic activities. When combined with pathway engineering, codon-optimized pathways are typically more efficient than their unengineered counterpart. This approach is especially important when attempting to produce natural products found in systems distantly related to the host organisms (such as importing plant genes into *E. coli* [83]). Finally, computational techniques are continually being developed to perform the task of codon optimization and assembly design which will improve our ability to control this process step [84]. However, recent evidence has shown that codon optimization may be more effective due to changes in mRNA secondary structure, as opposed to making more abundant tRNA available for translation [85]. Engineering mRNA secondary structure is a second-tier, regulatory method of controlling the central dogma, and is discussed in that section. With these emerging questions, it remains to be seen if the combination of *de novo* DNA synthesis with codon optimization algorithms will greatly expedite this process and remove some translation-level limitations in cellular systems.

2.4. Engineering the First Tier—Summary. The techniques and approaches described above focus on a synthetic approach aimed to redesign the information and process flow in the central dogma. As Figure 1 suggests, manipulation of the central dogma process at any one of the nodes often results in changes at the protein level. As the major catalytic, structural, and signaling components of cells, synthetically modifying proteins is one of the primary goals of engineering biology. Some of these manipulations, such as promoter engineering and codon optimization, are designed to alter protein level, while others, such as directed evolution [25, 86, 87] and unnatural amino acids, are intended to synthetically alter protein function directly. In either case, the change must be made at the first tier in order to create the downstream effect.

The tools of synthetic biology in this arena have improved the rate and precision of changes that can be made to this first tier. Moreover, they open the capacity to design novel elements and process units that serve a higher biotechnological goal. Yet, more work is required to enable full *de novo* design of these custom-made elements. In addition, the complete rewiring of cells will likely require multiple modifications in the first tier, thus these approaches must be used in combination to obtain the best results for bioprocessing applications.

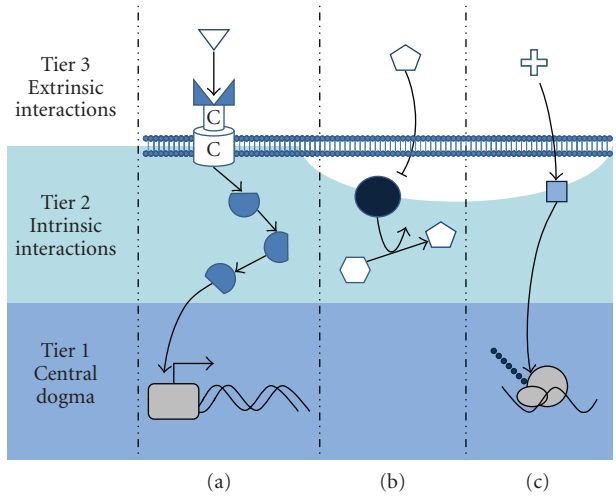


Figure 2: *Interaction of components illustrating synthetic biology at the three tiers.* (a) Canonical signalling pathway depicting a tier three ligand-receptor binding event that induces a tier two phosphorylation cascade which subsequently alters a promoter at the first tier. (b) Canonical allosteric protein inhibition by a molecule downstream in the pathway. The inhibition binding event is a tier three interaction which results in a tier two alteration in the pathway "set point." (c) Hypothetical signalling pathway in which a signal molecule alters an initiation factor at tier three which then changes the tier two protein translation set point.

3. The Second Tier—Engineering Process Controls

A myriad of control systems have evolved to regulate the highly complex process steps of the central dogma. As a result of tight integration between process and control, it is sometimes difficult to clearly delineate between the biological components of control mechanisms and the central dogma process flow. For this review, we propose that microbial control mechanisms are largely those components that interface with the central dogma process steps, but do not function in a catalytic manner with respect to the process step. Using the process control analogy, these components establish, alter, and regulate the biological "set points." Figure 2 depicts hypothetical control mechanisms within the cell to more clearly delineate between the various tiers of cellular processes. Generally, these control elements form a functional link between the central dogma process (the first tier) and the extrinsic environment (the third tier) and relay control messages into changes in the process units and streams. Actual control mechanisms within the cell are much more complex than this simple depiction, as studies on the RpoS subunit of *E. coli* RNA polymerase II have shown [88]. To this end, the systems biology approach of surveying global protein-protein interactions is pivotal to understanding these complex mechanisms.

Even without full knowledge of cellular regulatory machinery, synthetic biologists have developed tools that establish desirable synthetic set points for the central dogma process units. These tools are required for the function of genetic circuits and control of metabolic pathways. In this section, we will review enabling technologies and applications for synthetic engineering of process control systems within the second tier.

3.1. Transcriptional Control. As stated previously, the control elements involved in many signaling and regulatory networks are being uncovered using a systems biology approach. In particular, these studies are uncovering the important roles that transcription factors play in the process unit of transcription [7, 9, 59, 89]. Typically, these interaction networks are reconstructed using high throughput data obtained from two-hybrid, coimmunoprecipitation, or bioinformatic mining protocols. These studies have produced a wealth of data for analysis, although the data is often collected outside of biologically relevant conditions [90]. Recent work is attempting to improve upon these techniques by collecting protein interaction data in native systems with natural protein expression levels [90]. Even with preliminary data and targets, attempts to engineer transcription factor networks have shown promise [31]. If this knowledge gap is closed, imported synthetic circuits and genes could act independently and not be negatively influenced by other cellular processes. Moreover, components of transcription represent a target and provide an access point for synthetic biologists to effect change in biological systems. Therefore, transcriptional control networks should remain an active area of research for synthetic biologists in order to close this knowledge gap and open the possibility for ground-up cellular optimization.

3.2. Translational Control. In the past few years, a great interest has arisen in the synthetic engineering of RNA [91, 92]. From a process engineering perspective, RNA serves both as the "feed stream" to the translation process unit and as a central component of the translational machinery itself. Therefore, depending on the application and desired output, engineering of RNA to alter translation can be classified as a first-tier or a second-tier approach to synthetic biology. In terms of translational control, it has been demonstrated that modifying mRNA structure can modulate protein levels [93, 94]. As an example, by optimizing intergenic regions in an operon [95], hairpins that sequestered ribosomal start codons were introduced and afforded synthetic control of a bicistronic message. Also, similar work has been performed to identify and engineer synthetic ribosome entry (IRES) sites for polycistronic transcripts [96]. Finally, Breaker and colleagues have focused a great deal of effort to the understanding of native RNA response to small molecules [97, 98]. These findings along with research on riboswitches could begin to link translational control with signaling networks or environmental signals (for more on this topic, see Tier 3).

Despite these successes, the capacity to engineer translational control mechanisms is quite limited due to the fact that many initiation factors (IF) are yet to be fully characterized and explored. Yet, initial work is beginning to unveil the complex molecular interactions that occur at this level [99–101]. However, studies have shown that mRNA levels do not always correlate with protein levels [10, 102]. This disjoint does not reconcile with the simple, reasonable hypothesis embodied by the central dogma that increasing transcript levels should increase protein levels. However, it does provide circumstantial evidence for unknown control mechanisms influencing translation. More studies, perhaps borrowing from the systems biology approach to studying transcription factors [103, 104], are required to uncover these translation-level control mechanisms. If obtained, the full detail of translational control systems will provide a novel means to synthetically control the translation process unit. Therefore, the study of translational control mechanisms should be more fully explored by synthetic biologists.

3.3. Protein Regulation. Regulation occurs at all levels within the cell and is not limited to simply the process units of transcription and translation. In this regard, a great deal of regulation takes place at the protein level, especially upon activity and degradation rate. Also, modifying events such as glycosylation, phosphorylation, and acetylation allow protein function to be modulated and localized within the cell [105–107]. These events regulate cellular activity through activation, inhibition, and signaling. Systems biology is delving into these complex protein-protein interaction networks [108–115]. Supplied with this increasing amount of data, synthetic biologists are constructing synthetic regulatory networks [116, 117] that take advantage of these control processes.

As stated previously, protein level within a cell is partially regulated by degradation rates. Synthetic control of these processes is important to (1) aid our understanding of how to extend the half-lives of desirable proteins and (2) further our control of regulatory and signaling systems. To this end, a synthetic protein degradation network has been constructed in *Saccharomyces cerevisiae* [118]. Eukaryotic systems have elegant processes such as ubiquitination to control degradation, while prokaryotic systems were thought not to possess such capability. However, this has been found to not be the case. Many degradation mechanisms such as the AAA+ protease family in *E. coli* [119], prokaryotic ubiquitin-like protein (Pup) [120], and others have reversed this thinking. As a result, understanding and engineering protein degradation in prokaryotes can prove to be a fertile area for synthetic biology research in the future.

Protein regulation has been studied since the very beginning of the biotechnological revolution, and is presently becoming influenced by the synthetic biology paradigm. The work previously discussed exemplifies this and opens the door to ground-up cellular optimization.

3.4. Engineering the Second Tier—Summary. The regulatory mechanisms that act upon the central dogma process are vital to optimizing cellular function, from improving product titer to switching gene expression profiles. Synthetic biology has demonstrated that these mechanisms can be effectively engineered at all levels of the central dogma process. By altering regulatory and control systems, it will be possible to ensure that a cell will respond in a desired manner. This is an important trait when considering the behavior of synthetic circuits as well as the function of metabolically engineered organisms. However, more work needs to be done to establish the foundation necessary for ground-up cellular optimization and designer control elements that can leverage these critical regulatory networks.

4. The Third Tier—Engineering Environmental Signal Reception

The central dogma and the internal control mechanisms present in the cell have evolved to integrate and respond to a wide array of environmental signals. A short list of stimuli includes temperature, metabolite concentration, light, toxins, ions, and molecular signals from other cells. As we begin to probe cellular systems further, a greater appreciation is being given to cell communication and signaling pathways. New advances in small molecule detection are uncovering key small molecule elicitors. Figure 2, in addition to showing interactions at the second tier, also depicts cellular response to a hypothetical environmental signal at the third tier. This third tier is focused on extracellular molecules and sensing proteins. Once an extracellular molecule interacts with a component of either the internal control mechanisms or of the central dogma, the information has been translated into either the first or second tier. Synthetic engineering of environmental response and communication pathways provides a unique opportunity to exert control in a cellular system. In this section, we will review novel synthetic approaches for signal integration and communication.

Engineering at this third tier is producing some of the newest work at the frontier of synthetic biology. This area encompasses the work of genetic circuit construction as well as engineering bacterial quorum sensing and microbial consortia [121]. Advancing the toolbox for the microbial engineer at this level is critical for external control of microbial populations. Furthermore, this work provides a newfound method for process control of fermentations. Constructing synthetic sensors that can detect a particular signaling molecule can serve as a responsive switch in a simple genetic circuit as well as a biosensor used to detect toxins in the environment. Furthermore, these systems can be combined to achieve higher logic functions such as AND gates [122]. Engineering at this final tier is the last link in the whole-cell chain of events: environmental signal affects control mechanism, which then elicits change in central dogma, which affects protein levels, and then produces the desired result. As a result, synthetic control of molecular input capability is vital for the complex functions that are currently being designed into microbes.

4.1. Signal Receptive Genetic Circuits. Modeling biological networks has borrowed from electrical circuits theory [123], which has given rise to a great deal of analogies between the two fields. The concept of the genetic circuit has become one of the key contributions of synthetic biology [1, 124, 125]. By definition, genetic circuit design requires an input signal, which represents engineering at the third tier. Due to the many environmental properties a microbe is able to sense, there are a plethora of inputs available to engineer. These inputs can include metabolites [126], proteins [127], temperature [128], and light [129]. Using these inputs, circuits such as switches [127], oscillators [126], bandpass filters [130], and feedback loops [131] may be constructed. Also, it has been shown that genetic circuits can be improved using the directed evolution algorithm typically applied to single proteins [26].

One of the most common positive feedback loops used in genetic circuit engineering is bacterial quorum sensing [132–136]. This natural system of cell-cell communication has been used to initiate cell death for population control, cancer cell invasion [137], artificial predator-prey relationships [138], and cell motility [139]. In these studies, a synthetic quorum sensing response has been introduced into nonnative cells to drive a desired phenotype.

Beyond feedback loops, synthetic biologists could incorporate environmental signals through receptor-ligand interactions [140]. Not surprisingly, protein engineering techniques are important to manipulate these receptor-mediated interactions. Methods such as directed evolution [26, 87, 141] are proven, effective methods to accomplishing this end. However, there is also an ongoing push towards rational design of proteins [142, 143]. As this develops, synthetic biologists will have better avenues towards *a priori* protein design. Already, there is a great deal of published work on engineering binding pockets of proteins [140] and modeling protein folding changes based on binding events [144]. While these studies had a direct metabolic and medical [145] application, this approach could contribute to the construction of artificial regulatory networks for proteins that lack native regulation.

In addition, RNA engineering has shown that riboswitches represent a powerful and promising way to incorporate an environmental signal. These small fragments of RNA are a noncoding portion of an mRNA transcript that binds to small molecules, allosterically affecting protein activation levels in the cell [97, 146]. Moreover, the ability of these fragments to work in an independent fashion allows for a portable, modular assembly of responsive elements [98]. Riboswitches present an exciting new approach to environmental signal recognition because the sequence-structure-function relationship is more predictable than with that of a normal length protein [130]. Success has also been demonstrated designing a riboswitch as a biosensor [147], and constructing artificial switches with natural aptamer domains [148]. There is a great deal more to uncover about the function and applications of riboswitches, and ongoing work in this field will continue in that effort.

4.2. Engineering the Third Tier—Summary. Environmental signal input is essential for programming cellular function. Using the paradigm of genetic circuits, much work has been accomplished to this end. In this regard, the powerful capability of bacterial quorum sensing appears to be a very effective means of synthetically controlling cells. Also, engineering activation and inhibition sites into proteins would allow greater control over biological processes, but work in this area lags due to the unsolved protein sequence-structure-function problem. A third way to engineer environmental inputs is through riboswitches, an approach that utilizes RNA as the receptor molecule. All of these areas continue to increase the degree to which the central dogma is rationally controlled for specific uses. This capability, coupled with the powerful techniques at the other two tiers, will lead to ground-up engineering of biology.

5. Integrative Synthetic Biology

As the development of synthetic biology tools continues to mature, one can envision studies moving towards ground-up cellular optimization on a level never before seen. Armed with powerful DNA writing technology, along with the knowledge and ability to manipulate cellular processes, the possibilities for the microbial engineer may become almost limitless. This will be the stage where all of the disparate synthetic biology tools and approaches will be able to be fused with one another, creating the field of *integrative synthetic biology*.

To date, very few examples of this kind exist since some of the tools and understanding of the parts are lacking. However, recent work from the Keasling laboratory and colleagues provides an initial picture of the type of studies that will be done in this new field. In order to produce the medically valuable product artemisinin, many different approaches have been used, including synthetic pathway construction [83], codon optimization [83], environmental

signal detection [149], and mRNA secondary structure engineering [95]. Each of these strategies and tools was used in tandem to improve the biotechnological goal, and represent the earliest examples of the power of integrative synthetic biology.

Another example illustrating the reality of integrative synthetic biology is work on orthogonal central dogma machinery. It has already been discussed that the process units of transcription and translation are under a great deal of control, and the mechanisms concerning this control are not yet fully understood. This may overly complicate genetic manipulations, since heterologous gene expression in a host organism is subject to this native central dogma machinery, and the regulatory mechanisms may be unknown or inhibitory. One approach synthetic biologists could use to overcome this difficulty is by designing orthogonal, or noninteracting, transcription and translation machinery [150]. The concept of an orthogonal biological system is an excellent way to avoid undesired host interference, and this has recently been accomplished by combining the specific T7 bacterial promoter system with a previously developed orthogonal ribosome-rRNA pair [151]. This, in effect, creates two AND gates that must be satisfied for heterologous gene expression. Using these types of constructs for engineering control systems and integration of environmental signals may provide the fundamental knowledge needed to understand the more complex and interrelated systems of natural organisms, leading us one step closer to ground-up cellular optimization.

The integrative synthetic biology approach should bring the ability to perform wholesale cellular remodeling into the scope of a single research project. However, as has been iterated at many points in this review, much more basic knowledge and foundational research must be made in order to realize this scenario. In particular, our ability to predict and design components lags behind our ability to engineer them. However, the transformative work that is continually being done in synthetic biology inspires confidence that these techniques may soon be at the fingertips of the microbial engineer.

References

[1] D. Sprinzak and M. B. Elowitz, "Reconstruction of genetic circuits," *Nature*, vol. 438, no. 7067, pp. 443–448, 2005.

[2] J.-W. Veening, W. K. Smits, and O. P. Kuipers, "Bistability, epigenetics, and bet-hedging in bacteria," *Annual Review of Microbiology*, vol. 62, pp. 193–210, 2008.

[3] N. Mitarai, K. Sneppen, and S. Pedersen, "Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization," *Journal of Molecular Biology*, vol. 382, no. 1, pp. 236–245, 2008.

[4] V. Ramakrishnan, "Ribosome structure and the mechanism of translation," *Cell*, vol. 108, no. 4, pp. 557–572, 2002.

[5] R. J. Sims III, S. S. Mandal, and D. Reinberg, "Recent highlights of RNA-polymerase-II-mediated transcription," *Current Opinion in Cell Biology*, vol. 16, no. 3, pp. 263–271, 2004.

[6] P. B. Moore, "The ribosome returned," *Journal of Biology*, vol. 8, no. 1, article 8, 2009.

[7] K. Tan, T. Shlomi, H. Feizi, T. Ideker, and R. Sharan, "Transcriptional regulation of protein complexes within and across species," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 4, pp. 1283–1288, 2007.

[8] A. Roguev, S. Bandyopadhyay, M. Zofall, et al., "Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast," *Science*, vol. 322, no. 5900, pp. 405–410, 2008.

[9] M. M. Babu, B. Lang, and L. Aravind, "Methods to reconstruct and compare transcriptional regulatory networks," *Methods in Molecular Biology*, vol. 541, pp. 163–180, 2009.

[10] J. F. Moxley, M. C. Jewett, M. R. Antoniewicz, et al., "Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 16, pp. 6477–6482, 2009.

[11] H. O. Smith, C. A. Hutchison III, C. Pfannkoch, and J. C. Venter, "Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 26, pp. 15440–15445, 2003.

[12] J. Tian, H. Gang, N. Sheng, et al., "Accurate multiplex gene synthesis from programmable DNA microchips," *Nature*, vol. 432, no. 7020, pp. 1050–1054, 2004.

[13] D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, et al., "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome," *Science*, vol. 319, no. 5867, pp. 1215–1220, 2008.

[14] T. Ideker, V. Thorsson, J. A. Ranish, et al., "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network," *Science*, vol. 292, no. 5518, pp. 929–934, 2001.

[15] B.-K. Cho, P. Charusanti, M. J. Herrgard, and B. Palsson, "Microbial regulatory and metabolic networks," *Current Opinion in Biotechnology*, vol. 18, no. 4, pp. 360–364, 2007.

[16] D. Greber and M. Fussenegger, "Mammalian synthetic biology: engineering of sophisticated gene networks," *Journal of Biotechnology*, vol. 130, no. 4, pp. 329–345, 2007.

[17] S. Noack, A. Wahl, M. Haunschild, E. Qeli, B. Freisleben, and W. Wiechert, "Visualizing regulatory interdependencies and parameter sensitivities in biochemical network models," *Mathematics and Computers in Simulation*, vol. 79, no. 4, pp. 991–998, 2008.

[18] T. Ellis, X. Wang, and J. J. Collins, "Diversity-based, model-guided construction of synthetic gene networks with predicted functions," *Nature Biotechnology*, vol. 27, no. 5, pp. 465–471, 2009.

[19] A. Meyer, R. Pellaux, and S. Panke, "Bioengineering novel in vitro metabolic pathways using synthetic biology," *Current Opinion in Microbiology*, vol. 10, no. 3, pp. 246–253, 2007.

[20] K. E. Tyo, H. S. Alper, and G. N. Stephanopoulos, "Expanding the metabolic engineering toolbox: more options to engineer cells," *Trends in Biotechnology*, vol. 25, no. 3, pp. 132–137, 2007.

[21] J. D. Keasling, "Synthetic biology for synthetic chemistry," *ACS Chemical Biology*, vol. 3, no. 1, pp. 64–76, 2008.

[22] S. K. Lee, H. Chou, T. S. Ham, T. S. Lee, and J. D. Keasling, "Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels," *Current Opinion in Biotechnology*, vol. 19, no. 6, pp. 556–563, 2008.

[23] K. L. J. Prather and C. H. Martin, "De novo biosynthetic pathways: rational design of microbial chemical factories," *Current Opinion in Biotechnology*, vol. 19, no. 5, pp. 468–474, 2008.

[24] Z. Shao, H. Zhao, and H. M. Zhao, "DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways," *Nucleic Acids Research*, vol. 37, no. 2, 2009.

[25] H. Zhao, Y. Li, and F. H. Arnold, "Strategy for the directed evolution of a peptide ligase," *Annals of the New York Academy of Sciences*, vol. 799, pp. 1–5, 1996.

[26] Y. Yokobayashi, R. Weiss, and F. H. Arnold, "Directed evolution of a genetic circuit," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 26, pp. 16587–16591, 2002.

[27] A. J. Link, M. L. Mock, and D. A. Tirrell, "Non-canonical amino acids in protein engineering," *Current Opinion in Biotechnology*, vol. 14, no. 6, pp. 603–609, 2003.

[28] H. Bayley and L. Jayasinghe, "Functional engineered channels and pores," *Molecular Membrane Biology*, vol. 21, no. 4, pp. 209–220, 2004.

[29] J. Xie and P. G. Schultz, "A chemical toolkit for proteins—an expanded genetic code," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 10, pp. 775–782, 2006.

[30] H. Alper and G. Stephanopoulos, "Global transcription machinery engineering: a new approach for improving cellular phenotype," *Metabolic Engineering*, vol. 9, no. 3, pp. 258–267, 2007.

[31] T. A. Desai, D. A. Rodionov, M. S. Gelfand, E. J. Alm, and C. V. Rao, "Engineering transcription factors with novel DNA-binding specificity using comparative genomics," *Nucleic Acids Research*, vol. 37, no. 8, pp. 2493–2503, 2009.

[32] D. Klein-Marcuschamer, C. N. S. Santos, H. Yu, and G. Stephanopoulos, "Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes," *Applied and Environmental Microbiology*, vol. 75, no. 9, pp. 2705–2711, 2009.

[33] B. P. Cormack, G. Bertram, M. Egerton, N. A. R. Gow, S. Falkow, and A. J. P. Brown, "Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in *Candida albicans*," *Microbiology*, vol. 143, no. 2, pp. 303–311, 1997.

[34] J. Sambrook, *Molecular Cloning: A Laboratory Manual*, edited by D. W. Russell, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 2000.

[35] J. Tian, K. Ma, and I. Saaem, "Advancing high-throughput gene synthesis technology," *Molecular BioSystems*, vol. 5, no. 7, pp. 714–722, 2009.

[36] D. V. Goeddel, D. G. Kleid, F. Bolivar, et al., "Expression in *Escherichia coli* of chemically synthesized genes for human insulin," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 1, pp. 106–110, 1979.

[37] S. Nagata, H. Taira, A. Hall, et al., "Synthesis in *Escherichia coli* of a polypeptide with human leukocyte interferon activity," *Nature*, vol. 284, no. 5754, pp. 316–320, 1980.

[38] S. Slater, T. Gallaher, and D. Dennis, "Production of poly-(3-hydroxybutyrate-co-3-hydroxyvalerate) in a recombinant *Escherichia coli* strain," *Applied and Environmental Microbiology*, vol. 58, no. 4, pp. 1089–1094, 1992.

[39] J. Messing, R. Crea, and P. H. Seeburg, "A system for shotgun DNA sequencing," *Nucleic Acids Research*, vol. 9, no. 2, pp. 309–321, 1981.

[40] G. M. Church and S. Kieffer-Higgins, "Multiplex DNA sequencing," *Science*, vol. 240, no. 4849, pp. 185–188, 1988.

[41] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 12, pp. 5463–5467, 1977.

[42] J. E. Bailey, "Toward a science of metabolic engineering," *Science*, vol. 252, no. 5013, pp. 1668–1675, 1991.

[43] S.-W. Kim and J. D. Keasling, "Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production," *Biotechnology & Bioengineering*, vol. 72, no. 4, pp. 408–415, 2001.

[44] W. Wiechert, "^{13}C metabolic flux analysis," *Metabolic Engineering*, vol. 3, no. 3, pp. 195–206, 2001.

[45] H. Alper, Y.-S. Jin, J. F. Moxley, and G. Stephanopoulos, "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*," *Metabolic Engineering*, vol. 7, no. 3, pp. 155–164, 2005.

[46] S. Atsumi, A. F. Cann, M. R. Connor, et al., "Metabolic engineering of *Escherichia coli* for 1-butanol production," *Metabolic Engineering*, vol. 10, no. 6, pp. 305–311, 2008.

[47] G. Linshiz, T. B. Yehezkel, S. Kaplan, et al., "Recursive construction of perfect DNA molecules from imperfect oligonucleotides," *Molecular Systems Biology*, vol. 4, 2008.

[48] D. Bang and G. M. Church, "Gene synthesis by circular assembly amplification," *Nature Methods*, vol. 5, no. 1, pp. 37–39, 2008.

[49] H. Ye, M. C. Huang, M.-H. Li, and J. Y. Ying, "Experimental analysis of gene assembly with TopDown one-step real-time gene synthesis," *Nucleic Acids Research*, vol. 37, no. 7, p. e51, 2009.

[50] S. Blair, K. Richmond, M. Rodesch, M. Bassetti, and F. Cerrina, "A scalable method for multiplex LED-controlled synthesis of DNA in capillaries," *Nucleic Acids Research*, vol. 34, no. 16, 2006.

[51] C. Lartigue, J. I. Glass, N. Alperovich, et al., "Genome transplantation in bacteria: changing one species to another," *Science*, vol. 317, no. 5838, pp. 632–638, 2007.

[52] J. R. Kelly, A. J. Rubin, J. H. Davis, et al., "Measuring the activity of BioBrick promoters using an in vivo reference standard," *Journal of Biological Engineering*, vol. 3, article 4, 2009.

[53] R. P. Shetty, D. Endy, and T. F. Knight Jr., "Engineering BioBrick vectors from BioBrick parts," *Journal of Biological Engineering*, vol. 2, article 5, 2008.

[54] L. M. Tran, M. L. Rizk, and J. C. Liao, "Ensemble modeling of metabolic networks," *Biophysical Journal*, vol. 95, no. 12, pp. 5606–5617, 2008.

[55] J. Carrera, G. Rodrigo, and A. Jaramillo, "Model-based redesign of global transcription regulation," *Nucleic Acids Research*, vol. 37, no. 5, 2009.

[56] A. C. Forster and G. M. Church, "Towards synthesis of a minimal cell," *Molecular Systems Biology*, vol. 2, article 45, 2006.

[57] A. Moya, R. Gil, A. Latorre, J. Pereto, M. P. Garcillan-Barcia, and F. de La Cruz, "Toward minimal bacterial cells: evolution vs. design," *FEMS Microbiology Reviews*, vol. 33, no. 1, pp. 225–235, 2009.

[58] J. I. Glass, N. Assad-Garcia, N. Alperovich, et al., "Essential genes of a minimal bacterium," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 2, pp. 425–430, 2006.

[59] R. Zhang and Y. Lin, "DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes," *Nucleic Acids Research*, vol. 37, supplement 1, pp. D455–D458, 2009.

[60] C. Switzer, S. E. Moroney, and S. A. Benner, "Enzymatic incorporation of a new base pair into DNA and RNA," *Journal of the American Chemical Society*, vol. 111, no. 21, pp. 8322–8323, 1989.

[61] S. A. Benner and A. M. Sismour, "Synthetic biology," *Nature Reviews Genetics*, vol. 6, no. 7, pp. 533–543, 2005.

[62] Z. Y. Yang, D. Hutter, P. Sheng, A. M. Sismour, and S. A. Benner, "Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern," *Nucleic Acids Research*, vol. 34, no. 21, pp. 6095–6101, 2006.

[63] S. A. Havemann, S. Hoshika, D. Hutter, and S. A. Benner, "Incorporation of multiple sequential pseudothymidines by DNA polymerases and their impact on DNA duplex structure," *Nucleosides, Nucleotides & Nucleic Acids*, vol. 27, no. 3, pp. 261–278, 2008.

[64] Z. Y. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, and S. A. Benner, "Enzymatic incorporation of a third nucleobase pair," *Nucleic Acids Research*, vol. 35, no. 13, pp. 4238–4249, 2007.

[65] P. R. Jensen and K. Hammer, "The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters," *Applied and Environmental Microbiology*, vol. 64, no. 1, pp. 82–87, 1998.

[66] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos, "Tuning genetic control through promoter engineering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 36, pp. 12678–12683, 2005.

[67] M. Venter, "Synthetic promoters: genetic control through cis engineering," *Trends in Plant Science*, vol. 12, no. 3, pp. 118–124, 2007.

[68] F. S. Hartner, C. Ruth, D. Langenegger, et al., "Promoter library designed for fine-tuned gene expression in *Pichia pastoris*," *Nucleic Acids Research*, vol. 36, no. 12, 2008.

[69] H. Yu, K. Tyo, H. Alper, D. Klein-Marcuschamer, and G. Stephanopoulos, "A high-throughput screen for hyaluronic acid accumulation in recombinant *Escherichia coli* transformed by libraries of engineered sigma factors," *Biotechnology & Bioengineering*, vol. 101, no. 4, pp. 788–796, 2008.

[70] H. Liu, M. Yan, C. Lai, L. Xu, and P. Ouyang, "gTME for improved Xylose fermentation of *Saccharomyces cerevisiae*," to appear in *Applied Biochemistry and Biotechnology*.

[71] D. V. Schaffer, J. T. Koerber, and K.-I. Lim, "Molecular engineering of viral gene delivery vehicles," *Annual Review of Biomedical Engineering*, vol. 10, pp. 169–194, 2008.

[72] R. Alemany, C. Balague, and D. T. Curiel, "Replicative adenoviruses for cancer therapy," *Nature Biotechnology*, vol. 18, no. 7, pp. 723–727, 2000.

[73] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, and P. G. Schultz, "A general method for site-specific incorporation of unnatural amino acids into proteins," *Science*, vol. 244, no. 4901, pp. 182–188, 1989.

[74] K. C. Schultz, L. Supekova, Y. Ryu, J. Xie, R. Perera, and P. G. Schultz, "A genetically encoded infrared probe," *Journal of the American Chemical Society*, vol. 128, no. 43, pp. 13984–13985, 2006.

[75] C. C. Liu, A. V. Mack, M.-L. Tsao, et al., "Protein evolution with an expanded genetic code," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 46, pp. 17688–17693, 2008.

[76] H. S. Lee, G. Spraggon, P. G. Schultz, and F. Wang, "Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe," *Journal of the American Chemical Society*, vol. 131, no. 7, pp. 2481–2483, 2009.

[77] N. A. Burgess-Brown, S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, and O. Gileadi, "Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study," *Protein Expression and Purification*, vol. 59, no. 1, pp. 94–102, 2008.

[78] H. Foster, P. S. Sharp, T. Athanasopoulos, et al., "Codon and mRNA sequence optimization of microdystrophin transgenes improves expression and physiological outcome in dystrophic mdx mice following AAV2/8 gene transfer," *Molecular Therapy*, vol. 16, no. 11, pp. 1825–1832, 2008.

[79] S. Sasaguri, J.-I. Maruyama, S. Moriya, T. Kudo, K. Kitamoto, and M. Arioka, "Codon optimization prevents premature polyadenylation of heterologously-expressed cellulases from termite-gut symbionts in *Aspergllus oryzae*," *Journal of General and Applied Microbiology*, vol. 54, no. 6, pp. 343–351, 2008.

[80] M. Tokuoka, M. Tanaka, K. Ono, S. Takagi, T. Shintani, and K. Gomi, "Codon optimization increases steady-state mRNA levels in *Aspergillus oryzae* heterologous gene expression," *Applied and Environmental Microbiology*, vol. 74, no. 21, pp. 6538–6546, 2008.

[81] H. Wang, Q. Wang, F. Zhang, Y. Huang, Y. Ji, and Y. Hou, "Protein expression and purification of human Zbtb7A in *Pichia pastoris* via gene codon optimization and synthesis," *Protein Expression and Purification*, vol. 60, no. 2, pp. 97–102, 2008.

[82] Z.-L. Wu, J. Qiao, Z.-G. Zhang, F. P. Guengerich, Y. Liu, and X.-Q. Pei, "Enhanced bacterial expression of several mammalian cytochrome P450s by codon optimization and chaperone coexpression," *Biotechnology Letters*, vol. 31, no. 10, pp. 1589–1593, 2009.

[83] V. J. J. Martin, D. J. Piteral, S. T. Withers, J. D. Newman, and J. D. Keasling, "Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids," *Nature Biotechnology*, vol. 21, no. 7, pp. 796–802, 2003.

[84] K. S. Sandhu, S. Pandey, S. Maiti, and B. Pillai, "GASCO: genetic algorithm simulation for codon optimization," *In Silico Biology*, vol. 8, no. 2, pp. 187–192, 2008.

[85] G. Kudla, A. W. Murray, D. Tollervey, and J. B. Plotkin, "Coding-sequence determinants of expression in *Escherichia coli*," *Science*, vol. 324, no. 5924, pp. 255–258, 2009.

[86] J. C. Moore and F. H. Arnold, "Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents," *Nature Biotechnology*, vol. 14, no. 4, pp. 458–467, 1996.

[87] L. You and F. H. Arnold, "Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide," *Protein Engineering*, vol. 9, no. 1, pp. 77–83, 1996.

[88] R. Hengge-Aronis, "Signal transduction and regulatory mechanisms involved in control of the σ^s (RpoS) subunit of RNA polymerase," *Microbiology and Molecular Biology Reviews*, vol. 66, no. 3, pp. 373–395, 2002.

[89] E. Balleza, L. N. López-Bojorquez, A. Martínez-Antonio, et al., "Regulation by transcription factors in bacteria: beyond description," *FEMS Microbiology Reviews*, vol. 33, no. 1, pp. 133–151, 2009.

[90] J. Muller and N. Johnsson, "Split-ubiquitin and the split-protein sensors: chessman for the endgame," *ChemBioChem*, vol. 9, no. 13, pp. 2029–2038, 2008.

[91] F. J. Isaacs, D. J. Dwyer, and J. J. Collins, "RNA synthetic biology," *Nature Biotechnology*, vol. 24, no. 5, pp. 545–554, 2006.

[92] M. N. Win, J. C. Liang, and C. D. Smolke, "Frameworks for programming biological function through RNA parts and devices," *Chemistry & Biology*, vol. 16, no. 3, pp. 298–310, 2009.

[93] C. D. Smolke, T. A. Carrier, and J. D. Keasling, "Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures," *Applied and Environmental Microbiology*, vol. 66, no. 12, pp. 5399–5405, 2000.

[94] C. D. Smolke, V. J. J. Martin, and J. D. Keasling, "Controlling the metabolic flux through the carotenoid pathway using directed mRNA processing and stabilization," *Metabolic Engineering*, vol. 3, no. 4, pp. 313–321, 2001.

[95] B. F. Pfleger, D. J. Pitera, C. D. Smolke, and J. D. Keasling, "Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes," *Nature Biotechnology*, vol. 24, no. 8, pp. 1027–1032, 2006.

[96] A. Venkatesan and A. Dasgupta, "Novel fluorescence-based screen to identify small synthetic internal ribosome entry site elements," *Molecular and Cellular Biology*, vol. 21, no. 8, pp. 2826–2837, 2001.

[97] M. Mandal and R. R. Breaker, "Gene regulation by riboswitches," *Nature Reviews Molecular Cell Biology*, vol. 5, no. 6, pp. 451–463, 2004.

[98] J. N. Kim, K. F. Blount, J. Lim, K. H. Link, and R. Breaker, "Design and antimicrobial action of purine analogues that bind Guanine riboswitches," *ACS Chemical Biology*, vol. 4, no. 11, pp. 915–927, 2009.

[99] A.-C. Gingras, B. Raught, and N. Sonenberg, "eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation," *Annual Review of Biochemistry*, vol. 68, pp. 913–963, 1999.

[100] A. Pacheco, S. L. de Quinto, J. Ramajo, N. Fernandez, and E. Martinez-Salas, "A novel role for Gemin5 in mRNA translation," *Nucleic Acids Research*, vol. 37, no. 2, pp. 582–590, 2009.

[101] P. Saini, D. E. Eyler, R. Green, and T. E. Dever, "Hypusine-containing protein eIF5A promotes translation elongation," *Nature*, vol. 459, no. 7243, pp. 118–121, 2009.

[102] T. J. Griffin, S. P. Gygi, T. Ideker, et al., "Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*," *Molecular & Cellular Proteomics*, vol. 1, no. 4, pp. 323–333, 2002.

[103] A. Subramanian, P. Tamayo, V. K. Mootha, et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.

[104] L. M. Tran, M. P. Brynildsen, K. C. Kao, J. K. Suen, and J. C. Liao, "gNCA: a framework for determining transcription factor activity based on transcriptome: identifiability and numerical implementation," *Metabolic Engineering*, vol. 7, no. 2, pp. 128–141, 2005.

[105] C. L. Brooks and W. Gu, "Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation," *Current Opinion in Cell Biology*, vol. 15, no. 2, pp. 164–171, 2003.

[106] J. Seo and K.-J. Lee, "Post-translational modifications and their biological functions: proteomic analysis and systematic approaches," *Journal of Biochemistry and Molecular Biology*, vol. 37, no. 1, pp. 35–44, 2004.

[107] M. Mann and O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature Biotechnology*, vol. 21, no. 3, pp. 255–261, 2003.

[108] N. J. Krogan, G. Cagney, H. Yu, et al., "Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*," *Nature*, vol. 440, no. 7084, pp. 637–643, 2006.

[109] P. Shannon, A. Markiel, O. Ozier, et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.

[110] I. Xenarios, Ł. Salwínski, X. J. Duan, P. Higney, S.-M. Kim, and D. Eisenberg, "DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions," *Nucleic Acids Research*, vol. 30, no. 1, pp. 303–305, 2002.

[111] Y. Ho, A. Gruhler, A. Heilbut, et al., "Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry," *Nature*, vol. 415, no. 6868, pp. 180–183, 2002.

[112] M. Mann, R. C. Hendrickson, and A. Pandey, "Analysis of proteins and proteomes by mass spectrometry," *Annual Review of Biochemistry*, vol. 70, pp. 437–473, 2001.

[113] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, "A comprehensive two-hybrid analysis to explore the yeast protein interactome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 8, pp. 4569–4574, 2001.

[114] T. Ito, K. Tashiro, S. Muta, et al., "Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 3, pp. 1143–1147, 2000.

[115] S. Fields and R. Sternglanz, "The two-hybrid system: an assay for protein-protein interactions," *Trends in Genetics*, vol. 10, no. 8, pp. 286–292, 1994.

[116] P. M. Pryciak, "Designing new cellular signaling pathways," *Chemistry & Biology*, vol. 16, no. 3, pp. 249–254, 2009.

[117] R. P. Bhattacharyya, A. Remenyi, B. J. Yeh, and W. A. Lim, "Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits," *Annual Review of Biochemistry*, vol. 75, pp. 655–680, 2006.

[118] C. Grilly, J. Stricker, W. L. Pang, M. R. Bennett, and J. Hasty, "A synthetic gene network for tuning protein degradation in *Saccharomyces cerevisiae*," *Molecular Systems Biology*, vol. 3, article 127, 2007.

[119] S. Gottesman, E. Roche, Y. N. Zhou, and R. T. Sauer, "The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system," *Genes & Development*, vol. 12, no. 9, pp. 1338–1347, 1998.

[120] K. H. Darwin, "Prokaryotic ubiquitin-like protein (Pup), proteasomes and pathogenesis," *Nature Reviews Microbiology*, vol. 7, no. 7, pp. 485–491, 2009.

[121] K. Brenner, L. You, and F. H. Arnold, "Engineering microbial consortia: a new frontier in synthetic biology," *Trends in Biotechnology*, vol. 26, no. 9, pp. 483–489, 2008.

[122] J. C. Anderson, C. A. Voigt, and A. P. Arkin, "Environmental signal integration by a modular and gate," *Molecular Systems Biology*, vol. 3, article 133, 2007.

[123] H. H. McAdams and L. Shapiro, "Circuit simulation of genetic networks," *Science*, vol. 269, no. 5224, pp. 650–656, 1995.

[124] D. J. Sayut, P. K. R. Kambam, and L. Sun, "Engineering and applications of genetic circuits," *Molecular BioSystems*, vol. 3, no. 12, pp. 835–840, 2007.

[125] K. Michalodimitrakis and M. Isalan, "Engineering prokaryotic gene circuits," *FEMS Microbiology Reviews*, vol. 33, no. 1, pp. 27–37, 2009.

[126] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S. G. Lee, and J. C. Liao, "A synthetic gene-metabolic oscillator," *Nature*, vol. 435, no. 7038, pp. 118–122, 2005.

[127] W. Weber, C. Lienhart, D. M. El Baba, and M. Fussenegger, "A biotin-triggered genetic switch in mammalian cells and mice," *Metabolic Engineering*, vol. 11, no. 2, pp. 117–124, 2009.

[128] T. Waldminghaus, J. Kortmann, S. Gesing, and F. Narberhaus, "Generation of synthetic RNA-based thermosensors," *Biological Chemistry*, vol. 389, no. 10, pp. 1319–1326, 2008.

[129] A. Levskaya, A. A. Chevalier, J. J. Tabor, et al., "Synthetic biology: engineering *Escherichia coli* to see light," *Nature*, vol. 438, no. 7067, pp. 441–442, 2005.

[130] T. Sohka, R. A. Heins, R. M. Phelan, J. M. Greisler, C. A. Townsend, and M. Ostermeier, "An externally tunable bacterial band-pass filter," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 25, pp. 10135–10140, 2009.

[131] T. Kobayashi, L. Chen, and K. Aihara, "Modeling genetic switches with positive feedback loops," *Journal of Theoretical Biology*, vol. 221, no. 3, pp. 379–399, 2003.

[132] M. B. Miller and B. L. Bassler, "Quorum sensing in bacteria," *Annual Review of Microbiology*, vol. 55, pp. 165–199, 2001.

[133] W. C. Fuqua, S. C. Winans, and E. P. Greenberg, "Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators," *Journal of Bacteriology*, vol. 176, no. 2, pp. 269–275, 1994.

[134] E. L. Haseltine and F. H. Arnold, "Implications of rewiring bacterial quorum sensing," *Applied and Environmental Microbiology*, vol. 74, no. 2, pp. 437–445, 2008.

[135] S. Hooshangi and W. E. Bentley, "From unicellular properties to multicellular behavior: bacteria quorum sensing circuitry and applications," *Current Opinion in Biotechnology*, vol. 19, no. 6, pp. 550–555, 2008.

[136] L. You, R. S. Cox III, R. Weiss, and F. H. Arnold, "Programmed population control by cell-cell communication and regulated killing," *Nature*, vol. 428, no. 6985, pp. 868–871, 2004.

[137] J. C. Anderson, E. J. Clarke, A. P. Arkin, and C. A. Voigt, "Environmentally controlled invasion of cancer cells by engineered bacteria," *Journal of Molecular Biology*, vol. 355, no. 4, pp. 619–627, 2006.

[138] F. K. Balagadde, H. Song, J. Ozaki, et al., "A synthetic *Escherichia coli* predator-prey ecosystem," *Molecular Systems Biology*, vol. 4, article 187, 2008.

[139] L. E. Weiss, J. P. Badalamenti, L. J. Weaver, et al., "Engineering motility as a phenotypic response to LuxI/R-dependent quorum sensing in *Escherichia coli*," *Biotechnology & Bioengineering*, vol. 100, no. 6, pp. 1251–1255, 2008.

[140] J. T. Koh, "Engineering selectivity and discrimination into ligand-receptor interfaces," *Chemistry & Biology*, vol. 9, no. 1, pp. 17–23, 2002.

[141] Z. Shao and F. H. Arnold, "Engineering new functions and altering existing functions," *Current Opinion in Structural Biology*, vol. 6, no. 4, pp. 513–518, 1996.

[142] L. Jiang, E. A. Althoff, F. R. Clemente, et al., "De novo computational design of retro-aldol enzymes," *Science*, vol. 319, no. 5868, pp. 1387–1391, 2008.

[143] D. Rothlisberger, O. Khersonsky, A. M. Wollacott, et al., "Kemp elimination catalysts by computational enzyme design," *Nature*, vol. 453, no. 7192, pp. 190–195, 2008.

[144] B. K. Ho and D. A. Agard, "Probing the flexibility of large conformational changes in protein structures through local perturbations," *PLoS Computational Biology*, vol. 5, article 4, 2009.

[145] N. V. Wase and P. C. Wright, "Systems biology of cyanobacterial secondary metabolite production and its role in drug discovery," *Expert Opinion on Drug Discovery*, vol. 3, no. 8, pp. 903–929, 2008.

[146] A. Roth and R. R. Breaker, "The structural and functional diversity of metabolite-binding riboswitches," *Annual Review of Biochemistry*, vol. 78, pp. 305–334, 2009.

[147] N. Muranaka, V. Sharma, Y. Nomura, and Y. Yokobayashi, "Efficient design strategy for whole-cell and cell-free biosensors based on engineered riboswitches," *Analytical Letters*, vol. 42, no. 1, pp. 108–122, 2009.

[148] M. Wieland, A. Benz, B. Klauser, and J. S. Hartig, "Artificial ribozyme switches containing natural riboswiteh aptamer domains," *Angewandte Chemie*, vol. 48, no. 15, pp. 2715–2718, 2009.

[149] B. F. Pfleger, D. J. Pitera, J. D. Newman, V. J. J. Martin, and J. D. Keasling, "Microbial sensors for small molecules: development of a mevalonate biosensor," *Metabolic Engineering*, vol. 9, no. 1, pp. 30–38, 2007.

[150] O. Rackham and J. W. Chin, *Synthesizing Cellular Networks from Evolved Ribosome-mRNA Pairs*, Portland Press, London, UK, 2006.

[151] W. An and J. W. Chin, "Synthesis of orthogonal transcription-translation networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 21, pp. 8477–8482, 2009.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 232016, 11 pages
doi:10.1155/2010/232016

Review Article

Biology by Design: From Top to Bottom and Back

Brian R. Fritz,[1,2] Laura E. Timmerman,[1,2] Nichole M. Daringer,[1] Joshua N. Leonard,[1,3] and Michael C. Jewett[1,2,3]

[1] *Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*
[2] *Chemistry of Life Processes Institute, Northwestern University, 2170 Campus Drive, Evanston, IL 60208, USA*
[3] *Member, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*

Correspondence should be addressed to Joshua N. Leonard, j-leonard@northwestern.edu and Michael C. Jewett, m-jewett@northwestern.edu

Received 5 July 2010; Accepted 22 September 2010

Academic Editor: Ganesh Sriram

Synthetic biology is a nascent technical discipline that seeks to enable the design and construction of novel biological systems to meet pressing societal needs. However, engineering biology still requires much trial and error because we lack effective approaches for connecting basic "parts" into higher-order networks that behave as predicted. Developing strategies for improving the performance and sophistication of our designs is informed by two overarching perspectives: "bottom-up" and "top-down" considerations. Using this framework, we describe a conceptual model for developing novel biological systems that function and interact with existing biological components in a predictable fashion. We discuss this model in the context of three topical areas: biochemical transformations, cellular devices and therapeutics, and approaches that expand the chemistry of life. Ten years after the construction of synthetic biology's first devices, the drive to look beyond what does exist to what can exist is ushering in an era of biology by design.

1. Introduction

Our understanding of physical laws and knowledge of material properties allow us to engineer bridges that do not collapse and car engines that convert energy into mechanical motion. Engineering biology, however, is different. Even the simplest bacterium comprises a system whose complexity is humbling [1, 2]. In recognition of such challenges, the central goal of synthetic biology is to transform biology into a system that can be engineered just as we engineer bridges and mechanical systems today (reviewed in [3–8]). This approach promises to provide an improved understanding of the living world and will enable us, in the coming years, to harness the diverse repertoire of biology for compelling applications. These include next-generation biofuels, renewable "green" chemicals and industrial feedstocks, programmable and personalizable biological therapies, materials with novel properties, cheap and deployable diagnostics and therapeutics to promote global health, and technologies enabling environmental stewardship and remediation. The guiding question is this: how can we make biology engineerable? The new paradigm of biology by design can be summarized as follows:

(i) *conceive* a desired biological function,

(ii) *design* an engineered biological system to perform this function,

(iii) *build* the system,

(iv) the system *performs* as predicted.

Achieving this ambitious goal will require an improved understanding of the mechanisms by which biological parts function and by which they interact with one another and their environment.

The first wave of synthetic biology has focused on the development of simple biological modules, involving anywhere between two and ten well-characterized biological "parts" (e.g., genes) that are connected to perform defined functions (reviewed in [7]). In principle, these "parts" should still operate in a predictable manner when transplanted into different biological contexts. This conceptual framework

connects synthetic biology with other engineering disciplines. Whether the parts are electrical, chemical, physical, or biological, engineers depend on the predictable behavior of connected components to design and construct complex systems. For example, the automotive industry must develop engines that work within the context of an entire automobile. Construction of an engine is possible due to an understanding of how the individual components (e.g., pistons, crankshafts, and spark plugs) interact with one another, how these individual functions can be assembled to achieve a desired task (e.g., conversion of energy released by combusting fuel into mechanical motion), and how this assembled system interacts with the broader context of the vehicle (e.g., drawing fuel from the fuel lines and powering the transmission to transmit energy to the wheels). When synthetic biologists design simple biological systems from individual components in order to perform novel functions, these systems often do not behave as expected, particularly when transported from one biological context to another. In effect, a synthetic biologist today must endeavor to design an engine from imperfectly characterized parts and without knowing how the engine will eventually connect to the rest of the vehicle.

To overcome the limitations of our incomplete knowledge of biology, synthetic biologists approach the construction of user-defined functions through an iterative design cycle that incorporates both bottom-up and top-down design perspectives (Figure 1).

> *Bottom-up considerations* include the following: What is the desired function of the synthetic system? What parts could be used to construct this system (e.g., promoters, ribosome binding sites, etc.)? Do the parts exist and do they require additional characterization? How should the parts be configured? What is the predicted behavior of the synthetic module? Conceptually, this design perspective focuses on the characteristics of individual modules (which may be parts or subnetworks of parts) and their assembly into novel configurations in isolation from the endogenous cellular context into which they will eventually be placed.
>
> *Top-down considerations* include the following: How might the engineered module be decoupled from or insulated from aspects of the broader biological context (e.g., metabolic state, cell cycle progression, and epigenetic modifications) that may complicate the function of the engineered network? What interactions are necessary to connect a synthetic controller to the endogenous system it will control? How can the synthetic system take advantage of the existing biological infrastructure? The top-down perspective focuses on potential interactions—both desirable and undesirable—between the engineered subsystem and its biological context, and on development of strategies for harnessing or compensating for these influences.

Top-down and bottom-up perspectives guide our development of improved strategies for engineering biology

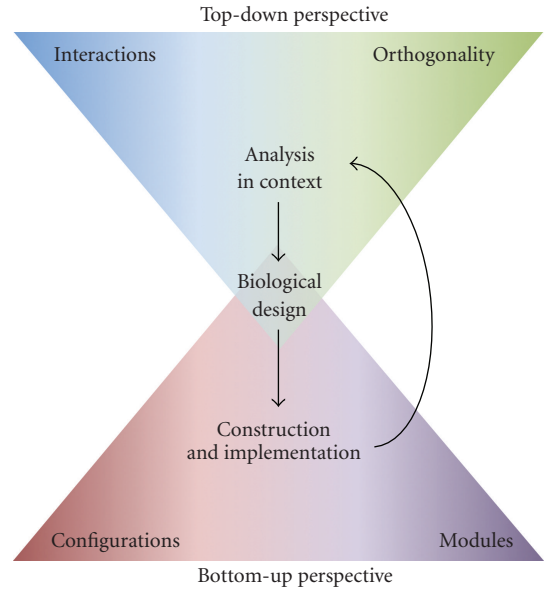


Figure 1: General conceptual framework for incorporating top-down and bottom-up perspectives in the synthetic biology design process. Due to our incomplete knowledge of biology, the design of biological systems through synthetic biology is currently an iterative process that incorporates both top-down and bottom-up design considerations. First, a design objective is identified. Next, a suitable synthetic biological system is designed given the known properties of well-characterized components (bottom-up). The synthetic system is then constructed and inserted into a larger biological context with which the synthetic system may interact (top-down), and performance of the combined system is assessed. If the system fails to meet performance requirements, this new information can be used to refine the design and repeat the cycle. Our ever-improving understanding of biology should reduce the number of iterations necessary to achieve a specific design objective.

(Figure 1), even in the face of incomplete information. Here, we discuss these perspectives in the context of three rapidly progressing areas which exemplify the opportunities and challenges encountered at different scales of biological organization: biochemical transformations, cellular devices and therapeutics, and approaches that expand the chemistry of life (Figure 2). Each of these topics is leading to new applications and advancing frontiers in synthetic biology.

2. Applications and New Frontiers

2.1. Biochemical Transformations. In the nearly four decades since the molecular biology revolution, it has become clear that the biological world provides a rich and diverse repertoire that may help to address some of humanity's most significant challenges. For example, harnessing biology for production of small molecules has led to applications in biofuels, chemical feedstocks, and therapeutics [3, 9]. However, to achieve these advances, substantial investments of resources, time, and labor are currently necessary. For example, it has been estimated that engineering microbes for

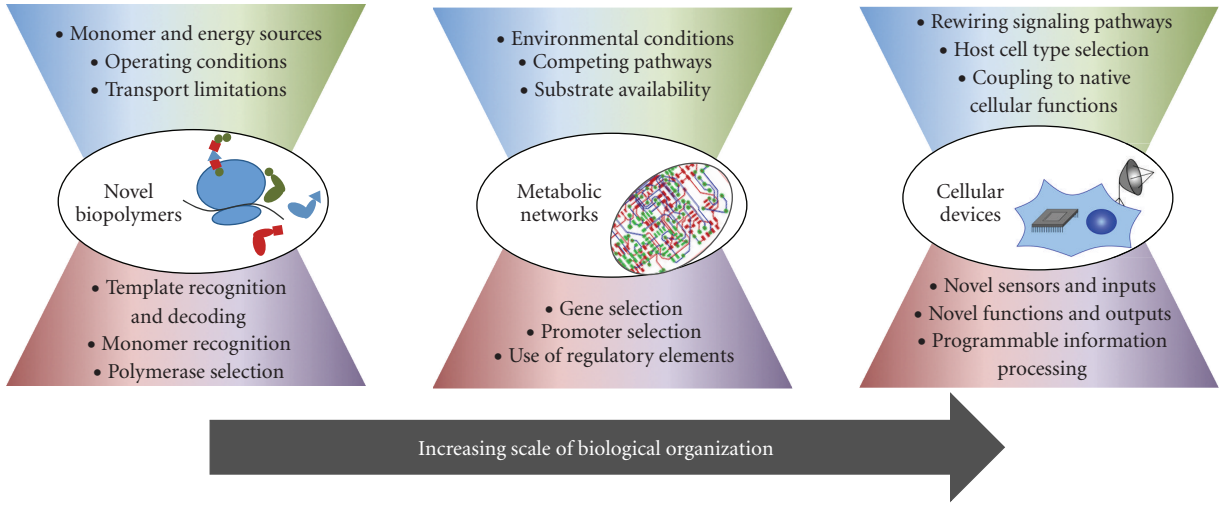


Figure 2: Application of the general framework to specific design objectives. At each scale of biological organization, designing synthetic biological systems invokes unique instances of the top-down and bottom-up considerations described in Figure 1.

expression of the antimalarial drug artemisinin required 150-person years of work [10].

Biochemical transformations comprise a family of applications for which synthetic biology may complement and expand efforts in metabolic and genetic engineering, by providing a robust technological framework for addressing common challenges. These challenges include targeting substrate flux towards a specific product, avoiding the loss of intermediates to competing pathways, reducing the accumulation of toxic intermediates, and preventing saturation of pathway enzymes [11, 12]. Synthetic biology tools are used to remove rate-limiting steps and increase titers of target biochemicals through an iterative design cycle of analysis, design, and implementation (Figure 1). Each cycle invokes both bottom-up and top-down considerations (Figure 2). Bottom-up considerations may include gene selection, promoter selection, and use of regulatory elements. Top-down considerations may include selection of the host organism, potential interactions of the synthetic metabolic network with existing cellular components, and competition with native pathways that impact substrate or intermediate availability. Host selection must often include making a choice between transferring heterologous pathways to a well-characterized organism or improving the production capacity (or other properties) of organisms for which robust genetic tools are not available.

Despite the numerous design requirements and constraints, many biochemical transformation projects in biotechnology have been realized. Here, we highlight four types: (i) incorporating nonnative genes to extend natural metabolism, (ii) incorporating whole pathways to add function, (iii) creating new pathways that have never existed before in nature, and (iv) organizing pathways to enhance activity (e.g., through compartmentalization or scaffolds). It should be noted that significant engineering of host metabolism is often a necessity to ensure precursor availability and inhibition of competitive pathways [12]. Moreover, in many cases the scope of genetic modifications requires multiple approaches to achieve a desired user-defined objective [13–15].

To extend natural metabolism, network optimization and selection of nonnative enzymes that catalyze the formation of a desired biochemical is required. In one example of this approach, Liao and colleagues converted 2-keto acid metabolites into branched alcohols for use as biofuels [13]. This was accomplished in *Escherichia coli* by first engineering the amino acid biosynthetic pathways that supply precursors required for 2-keto acid metabolite production. Then, incorporation of 2-keto acid decarboxylases and alcohol dehydrogenases from a variety of organisms yielded the desired alcohols. In later work, Atsumi et al. improved production titers for isobutanol through the use of alcohol dehydrogenases from different organisms, *Saccharomyces cerevisiae* and *Lactococcus lactis*, rather than the native alcohol dehydrogenase of *E. coli* [16]. In another example, an *E. coli* strain was engineered to convert simple sugar to biodiesel through modified fatty-acid biosynthesis pathways [15]. Here, native host metabolism was reprogrammed by modulating more than 8 enzymes (such as thioesterase, acyl-ligase, and wax ester synthase) to overproduce fatty acids, deregulate fatty acids, and produce fatty acid ethyl and methyl esters (biodiesel). Then, incorporation of nonnative hemicellulase genes enabled direct conversion of biomass to biodiesel in a single *E. coli* strain. Other poignant examples of heterologous gene transfer have come from meeting global health needs, namely, the treatment of malaria. In a landmark study, production of the antimalarial drug artemisinin required incorporation of amorphadiene synthase and cytochrome P450 monoxygenase from the plant *Artemisia annua* [14]. These genes were incorporated into a strain of *S. cerevisiae* modified to produce large quantities of the amorphadiene precursor, farnesyl pyrophosphate, by a complete overhaul of the mevalonate pathway.

Beyond heterologous gene transfer, protein engineering is being explored to create nonnatural enzymes that can be combined with biosynthetic pathways to create novel biochemical products. New enzymes can be created through either mutagenesis of existing enzymes or rational design of new enzymes. For example, Liao and colleagues expanded their work on the biosynthesis of alcohols by utilizing promiscuous enzyme activity [17]. By altering substrate specificity of an enzyme through site-directed mutagenesis, they were able to synthesis enzymes to make nonnatural alcohols with possible value as biofuels. A similar approach was used to develop substrate-specific cytochrome P450 enzymes [18]. These enzymes were created through random and site-directed mutagenesis to selectively deprotect different monosaccharide substrates for polysaccharide synthesis. In addition, protein engineering has advanced to a point where design of new enzymes from scratch is possible. An enzyme to perform a Diels-Alder reaction was created through computational design and experimental fine-tuning to achieve substrate specificity and stereoselectivity [19]. The designed enzyme was then altered to demonstrate the ability to perform similar reactions with different substrates. Such work demonstrates a powerful new approach to achieving molecular transformations that cannot be efficiently performed by either nature or conventional chemistry.

As the number of nonnative and unnatural genes transferred into a host increases to include entire pathway branches, a critical consideration is seamless integration with the organism's regular metabolic functions. One issue with metabolic network transfer is that the global energy resources of the cell must support both synthetic modules and the host cell. Despite this challenge, as well as other hurdles associated with constructing, testing, and balancing large networks of molecular pathways (where design complexity can be intimidating), some examples have shown success. Two notable applications include production of high butanol titers through the transfer of the *Clostridium acetobutylicum* butanol pathway to a variety of bacterial hosts [20] and production of valencene in yeast by transferring an isoprenoid pathway consisting of 7 *E. coli* genes [21].

In addition to manipulating and extending native pathways, it is also possible to engineer novel biochemical pathways that have never been observed in biology. For example, Moon et al. designed a synthetic pathway using genes from three sources: *myo*-inositol-1-phosphate synthase from *S. cerevisiae*, *myo*-inositol oxygenase from mice, and uronate dehydrogenase from *Pseudomonas syringae* [22]. The novel combined pathway was inserted into *E. coli* to convert *myo*-inositol into glucaric acid, a value-added chemical used as a dietary supplement, therapeutic, and potential feedstock for polymers. Similarly, Hawkins and Smolke synthesized benzylisoquinoline alkaloids in *S. cerevisiae* through the combination of genes from three different plant species as well human P450 enzyme [23]. These examples illustrate that complex molecules of therapeutic interest may be produced by combining existing enzymes into novel assemblies to create novel biosynthetic pathways.

Protein scaffolds that hold tagged pathway enzymes in close proximity have also been used to improve biochemical transformations [24]. In this model, scaffolds have been designed for biosynthesis pathways including mevalonate and glucaric acid pathways, and spatial colocalization of pathway enzymes has lead to increased productivity [24]. The use of scaffolds increases the local concentration of pathway intermediates near the desired downstream enzymes, which improves overall pathway kinetics, avoids the accumulation of hazardous intermediates, and minimizes consumption of intermediates by competing pathways [25, 26]. Such synthetic biology tools provide novel mechanisms for manipulating metabolic pathways and complement traditional pathway optimization.

Characterizing transcriptional and translational architectures and developing technologies for manipulating these processes provide other useful approaches for engineering microbial biotransformations. For instance, new genetic elements have been created to carry out circuit-like regulation that mimic toggle switches, AND gates, and oscillators [27–29]. Inserting such regulatory elements into cellular devices could lead to high-resolution control and novel functional behavior. Posttranscriptional regulation is also possible. Translation efficiency, for example, can be adjusted by modifying ribosome binding site affinity [30].

To fine-tune imperfect designs, genetic screening and evolutionary tools are also being developed. Sommer et al. screened a metagenomic library of arbitrary environmental DNA to identify three novel genes that convey tolerance to the biomass chemicals syringaldehyde and 2-furoic acid [31]. Similarly, Bayer et al. used a metagenomic screen of 89 suspected methyl halide transferase enzymes in *E. coli* to increase the yield of methyl halides from biomass to serve as precursors for chemicals and fuels [32]. Another screening technique called SCALEs (scalar analysis of library enrichments) uses libraries consisting of genomic DNA inserts of specific sizes to identify regions of a genome responsible for a particular phenotype [33]. This method has been used to identify gene regions conveying tolerance to inhibitors of metabolic pathways involving aspartic acid [34]. Such screens demonstrate approaches that can be used to discover genes responsible for other desirable or undesirable phenotypes. New robust evolutionary techniques such as multiplex automated genome engineering (MAGE) and global transcription machinery engineering (gTME) have also been developed to enable rapid fine-tuning of metabolic networks. MAGE generates genomic diversity within cells and across populations by using parallel, site-specific modification of DNA. Using MAGE, the lycopene synthesis pathway in *E. coli* was optimized by genetically modifying several targeted chromosome locations associated with genes of the pathway [35]. The gTME approach uses random mutagenesis of transcriptional machinery to alter the transcriptional program of an organism and has been used to develop *E. coli* and *S. cerevisiae* strains with desirable phenotypes for biofuel production, such as high ethanol tolerance [36, 37]. Evolutionary approaches provide a complementary strategy to design-based engineering that may reveal unexpected routes toward higher system performance and provide a better understanding of biological systems that may be incorporated into subsequent design strategies.

The next few years will bring increased attempts to construct and program large-scale user-defined pathways, and even whole organisms. Recently, researchers at the J. Craig Venter Institute (JCVI) assembled, modified, and implanted a complete *Mycoplasma mycoides* donor genome into a related *Mycoplasma* recipient cell [38]. This technological milestone marks the dawn of synthetic genomics. While this achievement is an important step towards designer organisms, applications of this approach remain challenging. For example, if one knew how to design a cell that efficiently converts sunlight, water, and CO_2 into fuel, then JCVI's technology could enable the synthesis and transplantation of that chemically synthesized genome into a recipient cell. However, we do not yet know how to design such a genome from scratch, and simply screening astronomical numbers of synthetic genome variants is impractical or even impossible. In addition, the costs of genome construction currently prohibit most researchers from applying this approach. In the long run, however, there is no doubt that JCVI's technology will join a suite of other whole-genome engineering techniques to accelerate the development of microbial factories for producing fuels, pharmaceuticals, green biochemicals, and novel materials.

2.2. Cellular Devices and Therapies. For applications in medicine, synthetic biology may also be used to create new cellular functions, such that the engineered cell itself—not just a product that it produces—serves a therapeutic role. In this context, a cell can be viewed as a device that receives inputs, processes this information, and produces outputs all based upon its genetic programming. Therefore, novel cellular functions may be constructed through strategies that create new connections between existing and novel mechanisms for input (e.g., receptors), processing (e.g., intracellular signaling cascades and regulatory interactions), and output (e.g., the production of bioactive molecules or the induction of specialized cellular effector functions). Every such strategy must take into account both bottom-up considerations (including the choice of biological parts and their configuration) and top-down considerations (including the choice of cell type to be engineered and the interactions—both desirable and undesirable—of the engineered components with native cellular functions) (Figure 2). Each of these design choices is guided by the type of cellular function required for a given application.

For some strategies, a cell may be reprogrammed to change the way it relates existing inputs to existing outputs by rewiring intracellular signal processing. In one example of this approach, signaling through the ErbB2 receptor tyrosine kinase, which mediates mitogenic or transformative signaling in tumor cells, was redirected into a proapoptotic pathway [39]. This was accomplished by engineering novel intracellular signaling proteins, in which SH2 or PTB phosphotyrosine-recognition domains from Grb2 or ShcA, respectively, (which interact with ErbB2), were fused to the death effector domain of Fadd. Such intracellular signaling proteins can be engineered to modulate multiple functional characteristics, including autoregulation, ligand specificity, and signaling dynamics using modular functional domains [40, 41] and various scaffold motifs and configurations [42]. An alternative approach is one often used in basic research, whereby the extracellular ligand-binding domain of one receptor is fused to the intracellular signaling domain of another receptor to generate a chimeric receptor. Strategies such as these could be used to generate cellular therapies for situations in which the natural input/output relationship is dysfunctional, such as tumor-mediated conversion of tumor-reactive T cells to an immunosuppressive phenotype. An important consideration when using this approach is that the engineered signaling pathway does not replace native signaling pathways, but rather engineered and native pathways are coupled by the shared use of native receptors or downstream components.

In other applications, cells may be engineered to respond to a stimulus to which the cell does not naturally respond. An early example of this approach is the use of chimeric antigen receptors (CARs) for cancer immunotherapy. In this model, ligand-binding domains from antibodies specific for tumor antigens are fused to the intracellular portion of T cell receptors in order to generate T cells capable of killing tumor cells that express the targeted antigen. Early implementations of this approach are in clinical trials, and while most benefits have been modest, some responses are quite promising [43]. In the most recent version of this strategy, third generation CARs are constructed by fusing the ligand-binding domains of tumor-specific antibodies to intracellular domains of CD28, CD137, and the TCR-ζ chain, in sequence [44]. In this construct, the TCR-ζ chain confers TCR signaling upon binding of the antibody domain to its target, while the CD28 and CD137 domains enhance TCR signaling and T cell survival *in vivo*. In a related approach, Xu et al. constructed tumor-responsive dendritic cells that recognize the tumor antigen, erbB2, and signal through the Toll-like Receptor 4 (TLR4) pathway to induce an inflammatory immune response against the tumor [45]. This was accomplished by fusing the ligand-binding domain of an anti-erbB2 antibody to various intracellular mediators of TLR4 signaling [45]. While each of these examples used receptors or antibodies that naturally recognize a particular ligand, it is also possible to re-engineer receptors, such as GPCRs, so that they respond to novel small molecule ligands [46]. Such an approach may be applied to generate cells whose functions are regulated by an inducible trigger, such as a small molecule drug. As an example of this strategy, T cells were engineered to express interleukin 2 (IL-2, a cytokine necessary for T cell survival) under the control of a synthetic ribozyme switch that responds to the inducer theophylline. When these cells were transferred to recipient mice, theophylline treatment induced IL-2 production and engineered T cell survival *in vivo* [47]. Similar ribozyme switches may be designed to respond to a variety of small molecule cues. Other RNA-based approaches can be used to program cells to perform more complicated functions, including the logical evaluation of multiple inputs, such that activation of a target gene (or genes) is conditioned upon the presence or absence of multiple extracellular cues [48].

Like mammalian cells, bacteria may also be programmed for therapeutic applications in which the bacteria itself is the therapeutic. For example, bacteria naturally home to tumor environments, so this propensity may be harnessed by engineering bacteria to selectively invade and destroy cancer cells *in vivo* [49]. In this model, bacteria were engineered to express the invasin protein in response to the hypoxic conditions that typify tumor microenivronments. Invasin expression enables these bacteria to invade mammalian cells expressing β1-integrins, which is a common marker on some types of cancer. This approach can be further enhanced to include the hypoxia-inducible release of cytotoxic agents to destroy the tumor cells after invasion [49]. Similarly, synthetic biology could be used to integrate more complex functions and regulatory features into other bacteria-based therapeutics. Important considerations when manipulating bacteria for therapeutic applications include both interactions with host biology and preventing the engineered bacteria from spreading beyond the patient being treated. For example, *Lactococcus lactis* was engineered to treat inflammatory bowel diseases by programming these microbes to secrete the immunosuppressive cytokine Interleukin-10 (IL-10) and by deleting their thymidylate synthase gene [50]. This strategy restricted bacterial growth to cultures supplemented with thymidine, which prevents these cells from replicating efficiently in the gut. This treatment has already shown promise in clinical trials [51]. Because the engineered bacteria die quickly within the gut, however, frequent treatments would be required to treat chronic disease. Synthetic biology could be used to overcome such limitations. For example, the immunosuppressive bacteria described above could be modified such that they require survival signals from the gut and quickly die outside of this environment, or such that they secrete IL-10 only when they detect disease flareups. Such modifications could allow for safer and more effective treatments while still restricting these bacteria to their targeted milieu. Synthetic biology provides the framework required to design and implement such sophisticated biotherapy strategies.

Engineering communication between cells also may be useful for therapeutic applications. Already, simple synthetic communication systems have been constructed in bacteria, yeast, and even mammalian cells. In a representative example, a yeast "sender" population was engineered to secrete the plant hormone, isopentenyladenine, and a "receiver" yeast population was made responsive to this signal by expressing the receptor protein, AtCRE1 [52]. Bidirectional communication can also be used to coordinate multicellular functions. For example, two populations of *E. coli* were engineered to communicate bidirectionally using *P. aeruginosa* quorum sensing components, such that each cell type secretes a unique "sender" signal and expresses its own reporter gene only when it detects the signal secreted by the other cell type [53]. Similar information exchange between sender and receiver cells can also be used to program spatial organization and pattern formation [54]. Other applications of intercellular communication are "polling" strategies, which can be used to coordinate and synchronize dynamic cellular functions and potentially improve performance by reducing sensitivity to variations in environmental conditions [55]. These approaches can also be extended to interkingdom communication, including bacterial, yeast, and mammalian cells [56]. When applied in mammalian cells, such approaches could eventually be used to program communication and spatial organization for the construction of synthetic organs or organ-like devices. These strategies may also provide useful tools for programming the differentiation of stem cells in a manner that does not require recapitulating the complex extracellular cues that guide natural differentiation programs. Interkingdom communication platforms could enable engineered symbiotic microbes, which might patrol the colon or skin for signs of disease, to communicate this information to their human hosts. To date, existing intercellular communication systems rely upon simple small molecule-based signaling intermediates, and sophisticated multicellular communication strategies will require the development of novel orthogonal, high-information-content signaling platforms.

Finally, synthetic biology could lead to new therapeutics through *in vitro* applications such as screening and diagnostics. Current drug screens are typically based upon measurements such as treatment-induced changes in the replication and viability of diseased cells relative to changes in healthy control cells. This limits the type of information about biological activity and safety that can be assessed at the screening stage. Screens constructed through synthetic biology can be used to evaluate more complicated biological effects. For example, engineered cells can be used to identify drugs that target specific pathogen functions. Such an approach was used to identify substances that block the *M. tuberculosis* protein, EthR, which otherwise confers resistance to the prodrug ethionamide [57]. This approach may be especially useful for pathogens that are dangerous or difficult to culture *in vitro*. Similar approaches could be used to construct screens that evaluate the impact of test molecules on human cellular functions that do not naturally produce an easily-assessed phenotype like apoptosis. Synthetic biology may also be useful for the construction of biological diagnostics. For example, bacterial bioreporters have been constructed to detect various compounds and toxins [58]. In most cases, this detection relies upon natural sensing mechanisms and involves expressing a reporter gene in response to the activation of an existing bacterial promoter. Synthetic biology could be employed to engineer novel sensing mechanisms and build bacterial bioreporters that perform diagnostic assessments such as detecting biochemical markers within human blood samples. These technologies could be designed to function and produce readouts without requiring external equipment, so that they could be readily deployed for mobile operations or in resource-poor areas. As we continue to develop improved technologies for engineering both human cells and the microorganisms with which they interact, synthetic biology promises to transform medicine just as it has already begun to transform biotechnology.

2.3. Expanding the Chemistry of Life: Novel Biopolymer Synthesis. To lessen undesired and unpredictable interactions

that often hamper design goals, modules of nonnatural components that operate independently of the cell's natural machinery are envisioned. Such orthogonal modules offer tremendous flexibility for optimization without disrupting the cell's operating system and, as a bonus, could exploit the renewability and evolvability of biology to synthesize nonbiological materials. In nature, complex and diverse chemical tasks are carried out across a variety of length scales using biopolymers of very simple composition. The expansion of the central dogma of biology to include diverse, nonnatural analogs may lead to a set of general solutions featuring orthogonal information coding, orthogonal cellular machinery, orthogonal scaffolds, orthogonal compartments, orthogonal interactions, or orthogonal communication pathways, which could form the basis for entirely artificial living systems. As an example, even minor modifications to natural biopolymers can endow them with altered interactions and enhanced functionalities. Peptide nucleic acids (PNAs), which have backbones similar to peptides but nucleobase functional groups (as opposed to the twenty natural amino acids) [59], hybridize to short stretches of DNA, and exhibit increased resistance to both proteases and endonucleases.

To expand the chemistry of life, orthogonal modules may be constructed through evolved catalytic ensembles that enable synthesis, evolution, and organization of nonnatural or hybrid polymers. In this case, bottom-up design considerations must include template recognition, substrate recognition, catalytic activity, error correction, product folding, and transport properties (Figure 2). Top-down considerations may require that synthesis should occur under physiological conditions (temperature, buffer conditions, etc.) and that monomer building blocks must be available to the host or able to be synthesized by the host. Implicit mass transfer limitations arising from cellular structure must also be addressed.

Increasing the capacity of nucleic acids to store, recall, and propagate specific information can augment genomic information [60, 61]. Early experiments with modified nucleic acids (termed xenonucleic acids, or XNAs [62]) were motivated by the desires to understand how DNA and RNA developed in the prebiotic world [63] and to create antisense molecules that could be easily taken up by cells [64]. More recently, efforts have shifted towards modulating function by exploring diversity in the composition of bases [65], sugars [61], and phosphates [61]. Backbone modifications of natural nucleic acids generally decrease recognition by endonucleases and provide an alternative strategy for modulating XNA binding affinity to complementary strands. The most studied backbone modifications of natural nucleic acids include PNAs [59], locked nucleic acids (LNAs) [66], threose nucleic acids (TNAs) [67], phosphorothioates [68], and Morpholinos [69]. Beyond antisense applications, these XNAs have the potential to vastly expand biopolymer information storage.

In order for novel biopolymers to become a viable tool for synthetic biologists, sequence-defined synthesis of the polymers is required. Beyond low-efficiency solid phase synthesis techniques, enzymatic template-directed synthesis of nonnatural polymers can be performed. In one approach, naturally occurring DNA and RNA polymerases have been identified that accept nonnatural substrates, thus allowing the templated synthesis of TNA from DNA [70]. In addition, these polymerases can be evolved *in vitro* to improve processivity and decrease error rates with nonnatural monomer units [71]. Improved understanding of structure-function relationships in polymerases has further resulted in the rational modification of a natural transcriptase to a polymerase that accepts nonstandard nucleotides [72]. This illustrates how directed evolution can be incorporated into the design cycle. Advances in enzyme engineering and our understanding of nature's design rules will speed progress toward engineered polymerases and enable us to probe the feasibility of enzymatic XNA synthesis and application.

In another approach, the cell's protein polymerase, the ribosome, could be engineered to selectively accept and polymerize synthetic monomers with high template fidelity. Driven by the promise of new therapeutics, evolvable abiological polymers, and biophysical probes, there has long been interest in harnessing natural translation machinery to synthesize polymers of nonnatural building blocks [73]. In concept, ribosome engineering may provide a more general, modular solution to novel biopolymer synthesis due to the use of transfer RNAs (tRNAs) that serve as adapters between the genetic code and the biopolymer product. Incorporating nonnatural monomer units on the ribosome depends on (i) reassignment of codons to nonnatural monomer units and (ii) acceptance of these units in the catalytic center. Remarkably, dozens of nonnatural amino acids have already been site-specifically incorporated into polypeptides by the ribosome. This was accomplished by engineering synthetic routes toward aminoacylation of tRNAs [74, 75] and orthogonal synthetase-tRNA pairs [76] and showing that natural ribosomes could accept nonnatural substrates. Even nonpeptide backbones, such as N-substituted amino acids [77] and hydroxy acids [78] have been incorporated by the ribosome, although with varying degrees of success. When using orthogonal synthetase-tRNA pairs, one potential issue is nonspecific incorporation that results from competing natural aminoacyl-tRNAs. To obtain greater control of homogeneous sequence composition, genetic code reprogramming [79, 80] or synthetic genetic systems (based on XNAs or modified bases) can be applied.

Although a variety of nonnatural monomer units have been incorporated by native ribosomes, it is clear that these substrates are inherently suboptimal. The tuning of polymerase machines to their natural substrates is exquisite [81]. Therefore, a major design challenge lies not only in the bottom-up aspect of incorporating nonnatural building blocks, but also in reversing the evolutionary optimization of finely tuned machinery and aligning it onto an entirely new path. To address this limitation, the ribosome and its machinery can be modified to enable increased promiscuity and efficiency at polymerizing nonnatural monomer units. In one case, EF-Tu was evolved to more efficiently incorporate amino acids containing bulky aromatic groups [82]. In examples that modify the ribosome directly, ribosomal RNA has been mutated or chemically modified to enable more efficient nonnatural substrate incorporation [83, 84].

Although the ribosome has been shown to be tolerant of modifications, there are relatively few examples because constraints imposed by living cells (e.g., dominant lethality) have hindered efforts to engineer ribosomes.

Specialized ribosomes overcome the dominant lethality constraint and provide an exemplary illustration of orthogonal performance. Originally pioneered by de Boer and colleagues [85], anti-Shine-Dalgarno sequences can be engineered into 16S ribosomal RNA to enable a group of ribosomes (orthogonal from the host) to translate specific mRNA populations. This has shown advantages for performing logic operations [85, 86], synthesizing proteins with nonnatural amino acids, such as p-benzoyl-(L)-phenylalanine [87], and efficiently incorporating quadruplet codons [88]. Orthogonal ribosomes will provide useful tools in programming synthetic function and expanding the genetic code to make nonnatural products for biotechnology.

Cell-free systems are being used to remove the most stringent top-down design constraints confronted when attempting to expand the chemistry of life [89–92]. Cellular overhead is reduced (e.g., genes of unknown functions, metabolic redundancies, and unintended interactions), transport barriers are removed, cellular viability is not required, and resources are more efficiently directed towards user-defined objectives. An additional advantage is direct access to the system of interest, similar to opening the hood of a car and accessing the engine. Cell-free protein synthesis systems [93–95] are one of the most prominent examples of cell-free biology and can be employed to create large amounts of nonnatural peptide [96]. Complex systems of biochemical transformations can also be built [97]. Looking forward, both *in vitro* and *in vivo* efforts towards novel catalytic ensembles promise to meld the complexity and diversity of chemistry with the accuracy and reproducibility of biology. The main rewards will be increased understanding of biology and new classes of evolvable molecules, which may find use as therapeutics or novel materials.

3. Moving Forward

Given the complexity that characterizes living systems developed through eons of evolution, the challenge to transform biology into an engineerable system is formidable. Considering biological design from both top-down and bottom-up perspectives provides a useful framework for proceeding towards this goal. At each scale of biological organization, this approach guides both the design and analysis of novel systems, as well as our investigations into the fundamental mechanisms that impact the performance of engineered biological systems. Synthetic biology is already changing the way we think about producing useful molecules, engineering our own biology, and expanding the chemistry of life. Making progress on these and other fronts will require both interdisciplinary approaches and new pedagogical strategies for training a new cadre of whole-brain thinkers. If successful, these efforts may revolutionize our ability to meet pressing societal needs through a new paradigm of biology by design.

Acknowledgments

The authors gratefully acknowledge funding from the National Institutes of Health (Grant Number R00GM081450), the National Academies Keck *Futures Initiative* (Grant Numbers NAFKI-SB5 and NAKFI-SB6 to M. Jewett and J. Leonard, resp.), and the National Science Foundation (Grant Numbers MCB-0943393 and PHY-0943390 to M. Jewett and J. Leonard, resp.). B. R. Fritz, L. E. Timmerman, and N. M. Daringer contributed equally.

References

[1] M. Costanzo, A. Baryshnikova, J. Bellay et al., "The genetic landscape of a cell," *Science*, vol. 327, no. 5964, pp. 425–431, 2010.

[2] M. Güell, V. van Noort, E. Yus et al., "Transcriptome complexity in a genome-reduced bacterium," *Science*, vol. 326, no. 5957, pp. 1268–1271, 2009.

[3] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age," *Nature Reviews Genetics*, vol. 11, no. 5, pp. 367–379, 2010.

[4] S. A. Benner and A. M. Sismour, "Synthetic biology," *Nature Reviews Genetics*, vol. 6, no. 7, pp. 533–543, 2005.

[5] D. Baker, G. Church, J. Collins et al., "Engineering life: building a FAB for biology," *Scientific American*, vol. 294, no. 6, pp. 44–51, 2006.

[6] D. Endy, "Foundations for engineering biology," *Nature*, vol. 438, no. 7067, pp. 449–453, 2005.

[7] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 6, pp. 410–422, 2009.

[8] F. J. Isaacs, D. J. Dwyer, and J. J. Collins, "RNA synthetic biology," *Nature Biotechnology*, vol. 24, no. 5, pp. 545–554, 2006.

[9] J. M. Carothers, J. A. Goler, and J. D. Keasling, "Chemical synthesis using synthetic biology," *Current Opinion in Biotechnology*, vol. 20, no. 4, pp. 498–503, 2009.

[10] R. Kwok, "Five hard truths for synthetic biology," *Nature*, vol. 463, no. 7279, pp. 288–290, 2010.

[11] J. E. Bailey, "Toward a science of metabolic engineering," *Science*, vol. 252, no. 5013, pp. 1668–1675, 1991.

[12] J. Nielsen, "Metabolic engineering," *Applied Microbiology and Biotechnology*, vol. 55, no. 3, pp. 263–283, 2001.

[13] S. Atsumi, T. Hanai, and J. C. Liao, "Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels," *Nature*, vol. 451, no. 7174, pp. 86–89, 2008.

[14] D.-K. Ro, E. M. Paradise, M. Quellet et al., "Production of the antimalarial drug precursor artemisinic acid in engineered yeast," *Nature*, vol. 440, no. 7086, pp. 940–943, 2006.

[15] E. J. Steen, Y. Kang, G. Bokinsky et al., "Microbial production of fatty-acid-derived fuels and chemicals from plant biomass," *Nature*, vol. 463, no. 7280, pp. 559–562, 2010.

[16] S. Atsumi, T.-Y. Wu, E.-M. Eckl, S. D. Hawkins, T. Buelter, and J. C. Liao, "Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes," *Applied Microbiology and Biotechnology*, vol. 85, no. 3, pp. 651–657, 2010.

[17] K. Zhang, M. R. Sawaya, D. S. Eisenberg, and J. C. Liao, "Expanding metabolism for biosynthesis of nonnatural alcohols," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 52, pp. 20653–20658, 2008.

[18] J. C. Lewis, S. Bastian, C. S. Bennett et al., "Chemoenzymatic elaboration of monosaccharides using engineered cytochrome $P450_{BM3}$ demethylases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 39, pp. 16550–16555, 2009.

[19] J. B. Siegel, A. Zanghellini, H. M. Lovick et al., "Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction," *Science*, vol. 329, no. 5989, pp. 309–313, 2010.

[20] D. R. Nielsen, E. Leonard, S.-H. Yoon, H.-C. Tseng, C. Yuan, and K. L. J. Prather, "Engineering alternative butanol production platforms in heterologous bacteria," *Metabolic Engineering*, vol. 11, no. 4-5, pp. 262–273, 2009.

[21] J. Maury, M. A. Asadollahi, K. Møller et al., "Reconstruction of a bacterial isoprenoid biosynthetic pathway in *Saccharomyces cerevisiae*," *FEBS Letters*, vol. 582, no. 29, pp. 4032–4038, 2008.

[22] T. S. Moon, S.-H. Yoon, A. M. Lanza, J. D. Roy-Mayhew, and K. L. J. Prather, "Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*," *Applied and Environmental Microbiology*, vol. 75, no. 3, pp. 589–595, 2009.

[23] K. M. Hawkins and C. D. Smolke, "Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*," *Nature Chemical Biology*, vol. 4, no. 9, pp. 564–573, 2008.

[24] J. E. Dueber, G. C. Wu, G. R. Malmirchegini et al., "Synthetic protein scaffolds provide modular control over metabolic flux," *Nature Biotechnology*, vol. 27, no. 8, pp. 753–759, 2009.

[25] E. W. Miles, S. Rhee, and D. R. Davies, "The molecular basis of substrate channeling," *Journal of Biological Chemistry*, vol. 274, no. 18, pp. 12193–12196, 1999.

[26] H. O. Spivey and J. Ovádi, "Substrate channeling," *Methods*, vol. 19, no. 2, pp. 306–321, 1999.

[27] J. C. Anderson, C. A. Voigt, and A. P. Arkin, "Environmental signal integration by a modular and gate," *Molecular Systems Biology*, vol. 3, article 133, 2007.

[28] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[29] M. Tigges, T. T. Marquez-Lago, J. Stelling, and M. Fussenegger, "A tunable synthetic mammalian oscillator," *Nature*, vol. 457, no. 7227, pp. 309–312, 2009.

[30] H. M. Salis, E. A. Mirsky, and C. A. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology*, vol. 27, no. 10, pp. 946–950, 2009.

[31] M. O. A. Sommer, G. M. Church, and G. Dantas, "A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion," *Molecular Systems Biology*, vol. 6, article 360, 2010.

[32] T. S. Bayer, D. M. Widmaier, K. Temme, E. A. Mirsky, D. V. Santi, and C. A. Voigt, "Synthesis of methyl halides from biomass using engineered microbes," *Journal of the American Chemical Society*, vol. 131, no. 18, pp. 6508–6515, 2009.

[33] M. D. Lynch, T. Warnecke, and R. T. Gill, "SCALEs: multiscale analysis of library enrichment," *Nature Methods*, vol. 4, no. 1, pp. 87–93, 2007.

[34] J. Bonomo, M. D. Lynch, T. Warnecke, J. V. Price, and R. T. Gill, "Genome-scale analysis of anti-metabolite directed strain engineering," *Metabolic Engineering*, vol. 10, no. 2, pp. 109–120, 2008.

[35] H. H. Wang, F. J. Isaacs, P. A. Carr et al., "Programming cells by multiplex genome engineering and accelerated evolution," *Nature*, vol. 460, no. 7257, pp. 894–898, 2009.

[36] H. Alper, J. Moxley, E. Nevoigt, G. R. Fink, and G. Stephanopoulos, "Engineering yeast transcription machinery for improved ethanol tolerance and production," *Science*, vol. 314, no. 5805, pp. 1565–1568, 2006.

[37] H. Alper and G. Stephanopoulos, "Global transcription machinery engineering: a new approach for improving cellular phenotype," *Metabolic Engineering*, vol. 9, no. 3, pp. 258–267, 2007.

[38] D. G. Gibson, J. I. Glass, C. Lartigue et al., "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science*, vol. 329, no. 5987, pp. 52–56, 2010.

[39] P. L. Howard, M. C. Chia, S. Del Rizzo, F.-F. Liu, and T. Pawson, "Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 20, pp. 11267–11272, 2003.

[40] J. E. Dueber, B. J. Yeh, K. Chak, and W. A. Lim, "Reprogramming control of an allosteric signaling switch through modular recombination," *Science*, vol. 301, no. 5641, pp. 1904–1908, 2003.

[41] J. E. Dueber, B. J. Yeh, R. P. Bhattacharyya, and W. A. Lim, "Rewiring cell signaling: the logic and plasticity of eukaryotic protein circuitry," *Current Opinion in Structural Biology*, vol. 14, no. 6, pp. 690–699, 2004.

[42] C. J. Bashor, N. C. Helman, S. Yan, and W. A. Lim, "Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics," *Science*, vol. 319, no. 5869, pp. 1539–1543, 2008.

[43] V. Marcu-Malina, S. van Dorp, and J. Kuball, "Re-targeting T-cells against cancer by gene-transfer of tumor-reactive receptors," *Expert Opinion on Biological Therapy*, vol. 9, no. 5, pp. 579–591, 2009.

[44] M. C. Milone, J. D. Fish, C. Carpenito et al., "Chimeric receptors containing CD137 signal transduction domains mediate enhanced survival of T cells and increased antileukemic efficacy *in vivo*," *Molecular Therapy*, vol. 17, no. 8, pp. 1453–1464, 2009.

[45] Y. Xu, P. K. Darcy, and M. H. Kershaw, "Tumor-specific dendritic cells generated by genetic redirection of Toll-like receptor signaling against the tumor-associated antigen, erbB2," *Cancer Gene Therapy*, vol. 14, no. 9, pp. 773–780, 2007.

[46] B. R. Conklin, E. C. Hsiao, S. Claeysen et al., "Engineering GPCR signaling pathways with RASSLs," *Nature Methods*, vol. 5, no. 8, pp. 673–678, 2008.

[47] Y. Y. Chen, M. C. Jensen, and C. D. Smolke, "Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 19, pp. 8531–8536, 2010.

[48] K. Rinaudo, L. Bleris, R. Maddamsetti, S. Subramanian, R. Weiss, and Y. Benenson, "A universal RNAi-based logic evaluator that operates in mammalian cells," *Nature Biotechnology*, vol. 25, no. 7, pp. 795–801, 2007.

[49] J. C. Anderson, E. J. Clarke, A. P. Arkin, and C. A. Voigt, "Environmentally controlled invasion of cancer cells by engineered bacteria," *Journal of Molecular Biology*, vol. 355, no. 4, pp. 619–627, 2006.

[50] L. Steidler, S. Neirynck, N. Huyghebaert et al., "Biological containment of genetically modified *Lactococcus lactis* for intestinal delivery of human interleukin 10," *Nature Biotechnology*, vol. 21, no. 7, pp. 785–789, 2003.

[51] H. Braat, P. Rottiers, D. W. Hommes et al., "A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease," *Clinical Gastroenterology and Hepatology*, vol. 4, no. 6, pp. 754–759, 2006.

[52] M.-T. Chen and R. Weiss, "Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*," *Nature Biotechnology*, vol. 23, no. 12, pp. 1551–1555, 2005.

[53] K. Brenner, D. K. Karig, R. Weiss, and F. H. Arnold, "Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 44, pp. 17300–17304, 2007.

[54] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, no. 7037, pp. 1130–1134, 2005.

[55] T. Danino, O. Mondragón-Palomino, L. Tsimring, and J. Hasty, "A synchronized quorum of genetic clocks," *Nature*, vol. 463, no. 7279, pp. 326–330, 2010.

[56] W. Weber, M. Daoud-El Baba, and M. Fussenegger, "Synthetic ecosystems based on airborne inter- and intrakingdom communication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 25, pp. 10435–10440, 2007.

[57] W. Weber, R. Schoenmakers, B. Keller et al., "A synthetic mammalian gene circuit reveals antituberculosis compounds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 29, pp. 9994–9998, 2008.

[58] J. R. van der Meer and S. Belkin, "Where microbiology meets microengineering: design and applications of reporter bacteria," *Nature Reviews Microbiology*, vol. 8, no. 7, pp. 511–522, 2010.

[59] P. E. Nielsen, "PNA technology," *Molecular Biotechnology*, vol. 26, no. 3, pp. 233–248, 2004.

[60] Y. Brudno and D. R. Liu, "Recent progress toward the templated synthesis and directed evolution of sequence-defined synthetic polymers," *Chemistry & Biology*, vol. 16, no. 3, pp. 265–276, 2009.

[61] D. Loakes and P. Holliger, "Polymerase engineering: towards the encoded synthesis of unnatural biopolymers," *Chemical Communications*, no. 31, pp. 4619–4631, 2009.

[62] P. Herdewijn and P. Marlière, "Toward safe genetically modified organisms through the chemical diversification of nucleic acids," *Chemistry & Biodiversity*, vol. 6, no. 6, pp. 791–808, 2009.

[63] S. A. Benner, "Understanding nucleic acids using synthetic chemistry," *Accounts of Chemical Research*, vol. 37, no. 10, pp. 784–797, 2004.

[64] S. M. Freier and K.-H. Altmann, "The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes," *Nucleic Acids Research*, vol. 25, no. 22, pp. 4429–4443, 1997.

[65] C. R. Geyer, T. R. Battersby, and S. A. Benner, "Nucleobase pairing in expanded Watson-Crick-like genetic information systems," *Structure*, vol. 11, no. 12, pp. 1485–1498, 2003.

[66] H. Kaur, B. R. Babu, and S. Maiti, "Perspectives on chemistry and therapeutic applications of locked nucleic acid (LNA)," *Chemical Reviews*, vol. 107, no. 11, pp. 4672–4697, 2007.

[67] J. K. Ichida, K. Zou, A. Horhota, B. Yu, L. W. McLaughlin, and J. W. Szostak, "An in vitro selection system for TNA," *Journal of the American Chemical Society*, vol. 127, no. 9, pp. 2802–2803, 2005.

[68] T. Wu and L. E. Orgel, "Disulfide-linked oligonucleotide phosphorothioates: novel analogues of nucleic acids," *Journal of Molecular Evolution*, vol. 32, no. 4, pp. 274–277, 1991.

[69] J. Summerton and D. Weller, "Morpholino antisense oligomers: design, preparation, and properties," *Antisense and Nucleic Acid Drug Development*, vol. 7, no. 3, pp. 187–195, 1997.

[70] J. C. Chaput and J. W. Szostak, "TNA synthesis by DNA polymerases," *Journal of the American Chemical Society*, vol. 125, no. 31, pp. 9274–9275, 2003.

[71] M. Fa, A. Radeghieri, A. A. Henry, and F. E. Romesberg, "Expanding the substrate repertoire of a DNA polymerase by directed evolution," *Journal of the American Chemical Society*, vol. 126, no. 6, pp. 1748–1754, 2004.

[72] A. M. Sismour, S. Lutz, J.-H. Park et al., "PCR amplification of DNA containing non-standard base pairs by variants of reverse transcriptase from human immunodeficiency virus-1," *Nucleic Acids Research*, vol. 32, no. 2, pp. 728–735, 2004.

[73] L. Wang and P. G. Schultz, "Expanding the genetic code," *Angewandte Chemie International Edition*, vol. 44, no. 1, pp. 34–66, 2005.

[74] S. M. Hecht, B. L. Alford, Y. Kuroda, and S. Kitano, ""Chemical aminoacylation" of tRNA's," *Journal of Biological Chemistry*, vol. 253, no. 13, pp. 4517–4520, 1978.

[75] H. Murakami, A. Ohta, H. Ashigai, and H. Suga, "A highly flexible tRNA acylation method for non-natural polypeptide synthesis," *Nature Methods*, vol. 3, no. 5, pp. 357–359, 2006.

[76] L. Wang, A. Brock, B. Herberich, and P. G. Schultz, "Expanding the genetic code of *Escherichia coli*," *Science*, vol. 292, no. 5516, pp. 498–500, 2001.

[77] T. Kawakami, H. Murakami, and H. Suga, "Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides," *Chemistry & Biology*, vol. 15, no. 1, pp. 32–42, 2008.

[78] S. Fahnestock and A. Rich, "Ribosome-catalyzed polyester formation," *Science*, vol. 173, no. 994, pp. 340–343, 1971.

[79] A. Ohta, H. Murakami, E. Higashimura, and H. Suga, "Synthesis of polyester by means of genetic code reprogramming," *Chemistry & Biology*, vol. 14, no. 12, pp. 1315–1322, 2007.

[80] A. Ohta, Y. Yamagishi, and H. Suga, "Synthesis of biopolymers using genetic code reprogramming," *Current Opinion in Chemical Biology*, vol. 12, no. 2, pp. 159–167, 2008.

[81] S. Ledoux and O. C. Uhlenbeck, "Different aa-tRNAs are selected uniformly on the ribosome," *Molecular Cell*, vol. 31, no. 1, pp. 114–123, 2008.

[82] Y. Doi, T. Ohtsuki, Y. Shimizu, T. Ueda, and M. Sisido, "Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system," *Journal of the American Chemical Society*, vol. 129, no. 46, pp. 14458–14462, 2007.

[83] L. M. Dedkova, N. E. Fahmi, S. Y. Golovine, and S. M. Hecht, "Enhanced D-amino acid incorporation into protein by modified ribosomes," *Journal of the American Chemical Society*, vol. 125, no. 22, pp. 6616–6617, 2003.

[84] A. Chirkova, M. Erlacher, R. Micura, and N. Polacek, "Chemically engineered ribosomes: a new frontier in synthetic biology," *Current Organic Chemistry*, vol. 14, no. 2, pp. 148–161, 2010.

[85] A. Hui and H. A. de Boer, "Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 14, pp. 4762–4766, 1987.

[86] O. Rackham and J. W. Chin, "Cellular logic with orthogonal ribosomes," *Journal of the American Chemical Society*, vol. 127, no. 50, pp. 17584–17585, 2005.
[87] K. Wang, H. Neumann, S. Y. Peak-Chew, and J. W. Chin, "Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion," *Nature Biotechnology*, vol. 25, no. 7, pp. 770–777, 2007.
[88] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, and J. W. Chin, "Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome," *Nature*, vol. 464, no. 7287, pp. 441–444, 2010.
[89] A. C. Forster and G. M. Church, "Towards synthesis of a minimal cell," *Molecular Systems Biology*, vol. 2, article 45, 2006.
[90] A. C. Forster and G. M. Church, "Synthetic biology projects in vitro," *Genome Research*, vol. 17, no. 1, pp. 1–6, 2007.
[91] C. Hold and S. Panke, "Towards the engineering of in vitro systems," *Journal of the Royal Society Interface*, vol. 6, no. Suppl 4, pp. S507–S521, 2009.
[92] M. C. Jewett and A. C. Forster, "Update on designing and building minimal cells," *Current Opinion in Biotechnology*, vol. 21, no. 5, pp. 697–703, 2010.
[93] M. C. Jewett, K. A. Calhoun, A. Voloshin, J. J. Wuu, and J. R. Swartz, "An integrated cell-free metabolic platform for protein production and synthetic biology," *Molecular Systems Biology*, vol. 4, article 220, 2008.
[94] M. C. Jewett and J. R. Swartz, "Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis," *Biotechnology and Bioengineering*, vol. 86, no. 1, pp. 19–26, 2004.
[95] O. P. T. Barrett and J. W. Chin, "Evolved orthogonal ribosome purification for in vitro characterization," *Nucleic Acids Research*, vol. 38, no. 8, pp. 2682–2691, 2010.
[96] A. R. Goerke and J. R. Swartz, "High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids," *Biotechnology and Bioengineering*, vol. 102, no. 2, pp. 400–416, 2009.
[97] M. Bujara, M. Schümperli, S. Billerbeck, M. Heinemann, and S. Panke, "Exploiting cell-free systems: implementation and debugging of a system of biotransformations," *Biotechnology and Bioengineering*, vol. 106, no. 3, pp. 376–389, 2010.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 541609, 16 pages
doi:10.1155/2010/541609

Review Article

Mathematical Modeling: Bridging the Gap between Concept and Realization in Synthetic Biology

Yuting Zheng and Ganesh Sriram

Department of Chemical and Biomolecular Engineering, University of Maryland, 1208D, Chemical and Nuclear Engineering Building 090, College Park, MD 20742, USA

Correspondence should be addressed to Ganesh Sriram, gsriram@umd.edu

Received 21 December 2009; Accepted 7 March 2010

Academic Editor: Hal Alper

Mathematical modeling plays an important and often indispensable role in synthetic biology because it serves as a crucial link between the concept and realization of a biological circuit. We review mathematical modeling concepts and methodologies as relevant to synthetic biology, including assumptions that underlie a model, types of modeling frameworks (deterministic and stochastic), and the importance of parameter estimation and optimization in modeling. Additionally we expound mathematical techniques used to analyze a model such as sensitivity analysis and bifurcation analysis, which enable the identification of the conditions that cause a synthetic circuit to behave in a desired manner. We also discuss the role of modeling in phenotype analysis such as metabolic and transcription network analysis and point out some available modeling standards and software. Following this, we present three case studies—a metabolic oscillator, a synthetic counter, and a bottom-up gene regulatory network—which have incorporated mathematical modeling as a central component of synthetic circuit design.

1. Introduction

Synthetic biology aims to design novel biological circuits for desired applications, implemented through the assembly of biological parts including natural components of cells and artificial molecules that emulate biological behavior [1, 2]. Because of its parts-to-whole approach, synthetic biology has a significant engineering component. Engineering endeavors typically involve the three classical engineering strategies: standardization (ensuring that components of a system are compatible and exchangeable), decoupling (dissecting a system into less complicated subsystems), and abstraction (streamlining a problem to focus only on the pertinent facets) [3–5]. It may appear that it should be possible to apply these strategies toward constructing a synthetic biological circuit in a manner similar to constructing an electric or electronic circuit. The attainment of this ideal goal is, however, impeded by the overwhelming complexity of biological systems with their myriad biomolecules and interconnections as well as sparse databases of gene function [3]. Consequently it is challenging to convert design concepts to predicted results.

This stumbling block in synthetic biology can be alleviated by the use of computer-aided mathematical modeling. Modeling is a powerful and often indispensable link between design and realization in engineering. It can predict the dynamics of a network under several different conditions and combinations thereof. Due to this, a user can search large parameter spaces *in silico* to identify the small regions of parameter space that produce the desired behavior or the most effective design or, alternatively, avoid parameter values that result in undesired responses. Modeling also provides the capability of using knowledge about the constituent parts of a system to predict the behavior of a system as a whole. Therefore, mathematical modeling serves as a bridge connecting a conceptual design idea to its biological realization (Figure 1).

In this review we present mathematical modeling concepts as relevant to synthetic biology and illustrate their application through the discussion of three case studies [6–8]. While the role of modeling in synthetic biology has been expertly reviewed before (e.g., [4, 9, 10]), this review aims to build upon the previous reviews by collecting

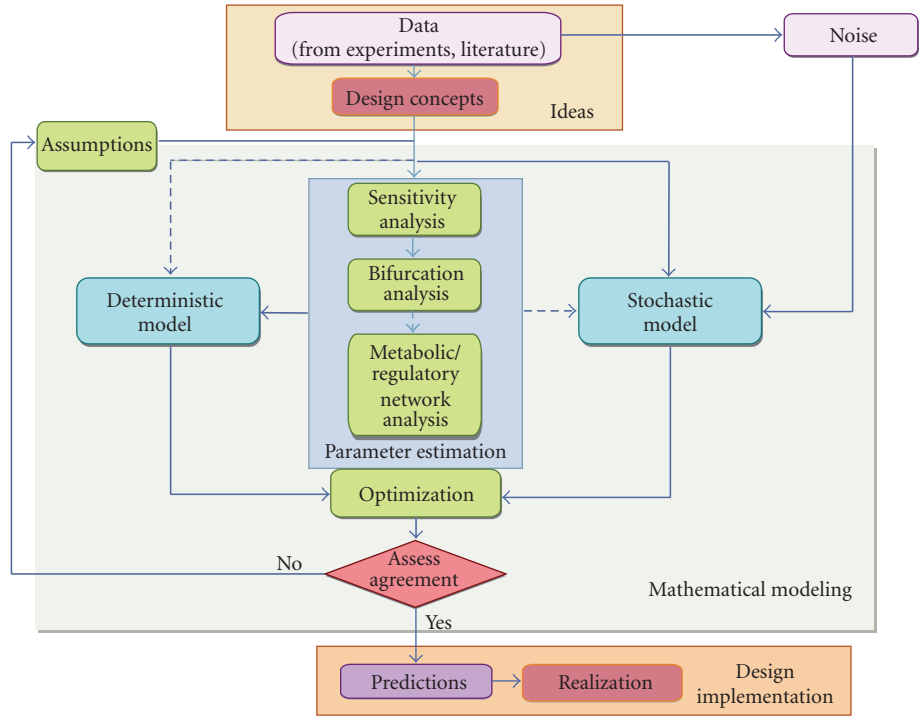


Figure 1: The role of mathematical modeling in synthetic biology. Computer-aided mathematical modeling bridges a design concept to realization in synthetic biology. Solid lines depict typical steps that have to be performed while developing a model; dashed lines depict unusual scenarios or conditions under which the steps shown by the corresponding solid lines are trivial or can be bypassed. A concept or ideas for designing a circuit for a particular function may be inspired by data from experiments or the literature. A mathematical model is then formulated on the basis of certain assumptions. The framework of a model could be deterministic or stochastic. The development of a model generally begins with the estimation of parameters that govern the model; this is a process that involves sensitivity analysis, bifurcation analysis, and, under certain circumstances, metabolic and transcription (regulatory) network analysis. The dashed line from design concepts to deterministic model indicates that, in some cases, parameter estimation is trivial or can be bypassed for this type of model. A stochastic model is developed by employing statistical functions to mimic system dynamics and considering fluctuations in the data. The dashed line from parameter estimation to stochastic model indicates that in some cases, parameter estimation may offer information in choosing statistical functions when constructing a stochastic model. Optimization is required for both models and is complete when the model exhibits an agreement (goodness of fit) with experimental data. A good agreement enables reliable prediction of system behavior and further biological realization, whereas unsatisfactory agreement requires the revision of the initial assumptions and the beginning of the next modeling cycle. See text and Figure 2 for explanations of terms.

several modeling methodologies into a single article and exemplifying them with in-depth case studies.

Figure 1 summarizes the important role played by mathematical modeling in synthetic biology and the key steps in the modeling process. Briefly, a model is formulated on the basis of certain assumptions about the system. Two broad types of modeling frameworks are available: deterministic modeling and stochastic modeling. Depending on the model type and the system, the estimation of parameters in the model could be a crucial step in obtaining a satisfactory model, and optimization may play a major role in this process. The predictions of a completed model are used, in conjunction with sensitivity analysis, bifurcation analysis, or, in certain cases, metabolic and regulatory network analyses to obtain insights toward an effective design. Below, we present a detailed discussion of the steps involved in modeling.

2. Assumptions Underlying a Model

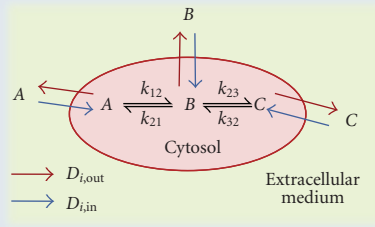
Biological systems are difficult to model and simulate despite a wealth of data on the structure and function of biomolecules and on cellular mechanisms. This is because biological systems exhibit complexity on several scales. Firstly metabolites, metabolic fluxes, proteins, RNA, and genes network in a highly complex manner; furthermore, their interconnections could constitute feedback or feed-forward loops that respond at various time scales [11]. Secondly, living systems can be highly sensitive to time-variant environmental conditions such as light, humidity, and supply of nutrients. These and other unknown causes of uncertainty result in "biological errors", which are distinct and usually greater in magnitude than instrumentation or measurement errors. Therefore, it is difficult to exactly predict the output of a biological system as compared to

Mathematical modeling definitions and techniques

Agreement. See goodness of fit.
Assumption. A presupposition that forms the basis of a model and trims down the complexity of the model. Examples are spatial or cell population homogeneity, lack of intracellular compartmentation or the existence of equilibrium, steady state or quasi-steady state.
Bifurcation analysis. An analysis that identifies the boundaries between the regions in parameter space that result in drastically different system dynamics, e.g. stable versus oscillatory. Two popular methods of bifurcation analysis are saddle node bifurcation and Hopf bifurcation.
Deterministic model. A model that endeavors to mimic a real system with analytical equations (usually ODEs or PDEs) that include numerical parameters. The output of such a model is predictable and reproducible (see Example, right).
Goodness of fit. A criterion used to judge the conformity between experimental data points and simulated points, using which parameter values can be accurately estimated. A chi-squared statistic is commonly used as a goodness of fit criterion. Agreement assessment involves statistical methods such as t tests and confidence interval estimation.
Metabolic flux analysis. Quantification of carbon flow through a network of metabolic pathways using stoichiometry, carbon atom rearrangements and isotopomer data from nuclear magnetic resonance and mass spectrometry.
Noise. Unpredictable fluctuation in the variables that describe the state of the system. Deterministic models usually do not consider noise, where as it is a significant component of stochastic models.
Optimization. A method to locate the minimum or maximum of an objective function. Local optimization searches for the optimum within a local neighborhood where as global optimization searches for the best optimum in the constrained parameter space.
Sensitivity analysis. The study of how the variations of certain parameters in a constrained domain will affect the response of a model; it enables the identification of parameters that are crucial to the model.
Stochastic model. A model that endeavors to mimic a real system with equations and parameters that vary stochastically. Such a model incorporates fluctuations inherent in real systems.

Transcription network analysis. An analysis of the relationships between genes and their regulators (usually transcription factors). Such an analysis aims to identify regulatory motifs underlying observed phenotypes.

Example of a deterministic model
Consider the development of a deterministic model for the following simple network.



Assumptions:
·Homogeneous
·Not compartmented
·1st order reactions

Model equations

$$\frac{dA}{dt} = 0 \cdot A + k_{12}B + 0 \cdot C + D_1$$
$$\frac{dB}{dt} = k_{21}A + 0 \cdot B + k_{23}C + D_2$$
$$\frac{dC}{dt} = 0 \cdot A + k_{32}B + 0 \cdot C + D_3$$

Parameters
$i, j = 1, 2, 3$
k_{ij}: reaction rate constants
D_i: the rate of transport of each metabolite in/out of the cell
$D_i = X_{i,\text{in}} - X_{i,\text{out}}$

These equations can be written in matrix form:

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} 0 & k_{12} & 0 \\ k_{21} & 0 & k_{23} \\ 0 & k_{23} & 0 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} \Longrightarrow \frac{dX}{dt} = F(N, t; \theta)$$

Where $X$and N are vectors comprised of concentrations of the species A, B and C (and are identical in this case), dX/dt is the rate of change of X, θ is a vector of model parameters.

With D_i measurable, this model can be solved for the time course trajectories of the metabolites A, B and C. The steady state(s) of the model can be analyzed by setting $dX/dt = 0$.

Figure 2: Terms used in mathematical modeling and an example of a simple (deterministic) mathematical model.

a mechanical or an electrical system and one has to often reconcile with an approximate reproduction.

However, a biological system can often be simplified to a level that permits a user to obtain insights toward synthetic circuit construction [11]. For example, Ma'ayan et al. [12] demonstrated how simplifying the dynamics of single components could lead to valuable information on a system's function. Simplification of a model requires the making of various assumptions. A commonly used assumption is homogeneity, both within the cell and within a cell population. Spatially homogeneous time-variant systems can be modeled by ordinary differential equations (ODEs) (equation (1), see Figure 2, e.g.). However, time-variant systems that feature compartmentation [13], spatial segregation, or intracellular gradients [14] may require the use of partial differential equations (PDEs). Although solving PDEs (and thereby non-homogenous models) is computationally much more intensive than solving ODEs, it can pay off well. For example, effects such as the spatial segregation of two enzymes that may generate intracellular gradients or the effect of protein diffusivity on enzyme activity can be simulated by using nonhomogenous models [14]. Closely related to spatial homogeneity is the assumption of cell population homogeneity, which is very frequently employed in models of biological systems. However, the modeling of heterogeneous populations in chemical reactors [15] has found application in the modeling of heterogeneous cell populations, and stochastic models frequently employ it. Besides the homogeneity assumption, most models involving enzyme kinetics or transcriptional regulations also assume

equilibrium [8], steady state [7], or quasi-steady state [6]. Such an assumption can remove time-dependence from the model and converts ODEs to simpler algebraic equations. The task of formulating the assumptions underlying a model is a fine balancing act between trimming down the complexity of the system while retaining the features of the system that are crucial to making reliable predictions for the application at hand. If a model based on certain assumptions does not agree with experimentally observed behavior, then the assumptions have to be revised [9]. This makes mathematical modeling an iterative process (Figure 1).

3. Types of Model Frameworks

3.1. Deterministic Mathematical Models. Mathematical models of biological systems can be categorized into two major types: deterministic and stochastic [4]. A deterministic model emulates a real system with analytical equations (usually ODEs or PDEs) that include numerical parameters. These equations are usually mass balances on cellular species (Figure 2), and the state of the system that is predicted by such a model is reproducible [4]. In contrast, a stochastic model endeavors to represent a real system with randomly interacting particles or species. The rate of each reaction between the species follows a probabilistic equation [9]; additionally, the time between the reactions can also vary. Stochastic models usually incorporate the fluctuations and noise inherent in real biological systems and examine the effect of noise on system dynamics [4].

Figure 2 explains the construction of a deterministic model for a simple network. Deterministic models usually employ differential equations used to describe interactions or reactions between biomolecules:

$$\frac{dX}{dt} = F(N, t; \theta), \tag{1}$$

where X and N are vectors comprised of species concentrations (and may be identical), dX/dt is the rate of change of X, θ is a vector of model parameters (see the following section on model parameters), and $F(N, t; \theta)$ is a nonlinear vector function that relates rates of change to concentrations [4]. Dynamic simulations of a system modeled by equation (1) are quite straightforward and will reveal the time-dependent characteristics of the system by generating time series trajectories of the species concentrations. Furthermore, the simulations help to analyze the behavior of a network when feedforward or feedback regulations are integrated into it. Such analysis has shown that the dynamic properties of feedforward loops depend on their specific architecture [16, 17], that positive or double negative feedbacks often introduce ultrasensitive or bistable switches [18, 19], and that negative feedbacks may reduce instability of the system [20]. Furthermore, multiple steady states or oscillations may occur as the result of positive and negative feedbacks [21, 22].

To analyze steady states of a time-dependent biological system, the time derivatives in (1) are set to zero [4]:

$$F(N, t; \theta) = 0, \tag{2}$$

which represents the steady state(s) of the system, is usually combined with bifurcation analysis to obtain the range of parameters in which the system will exhibit certain desired behaviors such as oscillations [23] (see Section 5 below).

Numerous deterministic models have been developed for biological systems, including several for synthetic circuits. Case studies I and II discussed in this article [6, 7] employ deterministic models.

3.2. Stochastic Mathematical Models. In deterministic models, every interaction and every parameter value is certain. Therefore, such models predict identical system dynamics for the same set of parameter values and initial conditions. However, real systems are characterized by unexpected and irreproducible fluctuations. To capture these fluctuations and their consequences on the behavior of the system, an alternative type of model, the stochastic model, is used. Such models mimic a system as a collection of interacting particles, with the reaction rates being governed by probabilistic rate laws [9]. An example of such a rate law is the chemical Master equation [10]. Stochastic simulation algorithms (SSAs) [10] such as Gillespie's algorithm are then used to simulate the state of the system.

One approach in stochastic modeling is to assume that a system is comprised of randomly interacting biomolecules, wherein the reactions between the molecules are modeled as Poisson processes with a probabilistically determined rate parameter [9]. Another approach is to perceive a time-variant system as a discrete time stochastic process. This approach uses a random variable or a vector X_n to indicate the discrete state of the system amongst several (finite or infinite) possible states [24]. The fewer the system states, the easier it is to construct a stochastic model. Guido et al. [8] have employed the latter approach with six system states to develop a stochastic model for a bottom-up gene regulatory network. In this type of stochastic model, the probability p_i at time n of each system state S_i is computed based on certain assumptions [24]:

$$p_i(n) = P(X_n = i). \tag{3}$$

System responses or outputs such as the rate of synthesis of green fluorescent protein (GFP) are then described in terms of the state probabilities:

$$\gamma = \sum_{i=1}^{n} g_i p_i, \tag{4}$$

where γ is the net output resulting from a combination of n states and g_i is the synthesis rate contributed by each state S_i. The probability of a system state $p_i(n)$ could be estimated by taking into account noises from synthesis and degradation of mRNA, GFP, or transcription factors and physical properties of the cell or system [8]. Finally, equation-free stochastic models have also been developed [25].

Several stochastic models have been developed for synthetic biological circuits and related simple biological systems [8, 25–29]. In case study III of this article, we discuss a stochastic model for a bottom-up gene regulatory network reported in Guido et al. [8].

4. Parameters in a Model

Any model contains several variables that do not represent the system state, but whose values govern the dynamics of the equations in the model. Such variables include reaction rate constants, equilibrium constants, diffusivities, and other physical properties. These are termed "parameters" of the model, as opposed to "state variables" such as species concentrations that represent the state of the system. To make useful predictions from a model, the parameters in the model have to be accurately estimated.

Mechanistic models, which are based on physical and chemical laws, include parameters that carry physical, chemical, or biological meaning. However, there could be many instances where not much information is available about a system, and constructing a "black box" model is the only option available. The parameters of such a model do not carry physical or biological meaning, but their estimation is nevertheless indispensable to the success of the model. Occasionally, information about a system could be so meager that even a black box model cannot be constructed. In such cases, a reverse engineering approach is employed to translate observable information to not only parameters but also model equations. This approach involves searching through (discrete) topological space instead of (continuous) numerical parameter space. Sometimes, combining the topological and numerical parameters for a system and simultaneously searching for both types of parameters has many advantages in understanding systems that are sparsely known [30]. Sometimes, parameter estimation can be largely bypassed. Recently, Tran et al. [31] described an approach called ensemble modeling, suited to modeling metabolic networks. This technique bypasses the requirement of obtaining accurate parameter values by reducing an initial set of models. The final set of models so obtained is capable of describing phenotypes of enzyme perturbations. Such approaches offer versatile ways to attack parameter problems in modeling.

4.1. Parameter Estimation. Parameter estimation is known as the "inverse problem" or "model calibration" and is both a key step and a limiting step of model construction [32, 33]. Parameter estimation typically leads to a first working model. If this initial model exhibits significant departures from experimental data (a frequent occurrence), further experiments may need to be performed to refine the parameter values. This process is repeated iteratively until a satisfactory model is constructed [34, 35]. Model parameters can occasionally be found from the literature or estimated manually, although this is usually feasible for small parameter spaces and simple systems. In general, parameter estimation from experimental observations requires sophisticated techniques such as described below. Amongst these, (global) optimization, a computationally intensive but powerful and robust method, is widely used.

4.2. Optimization. Parameter estimation is generally an optimization problem that involves locating the optimum (minimum or maximum) of an objective function that represents how well the model simulations agree with experimental data. This can be expressed as

$$\min_{\theta} \Phi(\theta). \tag{5}$$

The function $\Phi(\theta)$, which represents the goodness of fit between experiment and simulation, is a scalar function of the parameter vector θ. Its optimum is determined by iteratively adjusting the values of the components of θ and sometimes revising model assumptions [4]. The function $\Phi(\theta)$ is most frequently a weighted sum-of-squares error (a chi-squared statistic) between the experimental data points and the corresponding simulated points [36]. Other objective functions such as the Bayesian estimator and the maximum likelihood estimator also work well [33]. The process of adjusting and refining parameter values to reach the optimum of this objective function can be performed manually for linear or piecewise linear models [37]. However, many biological processes are not only nonlinear but also random, thus necessitating nonlinear models, stochastic models, or both [37–39]. In such cases, several general methods of parameter estimation are available, including some that are particularly suited to nonlinear dynamic systems typical in biology [40–43]. Even while using sophisticated algorithms, parameter estimation can involve unexpected complications such as the inability of a given optimization algorithm to effectively search a parameter space. In such cases, an exhaustive searching of parameter space can sometimes be accomplished well by stochastic Monte Carlo algorithms [44]. However, the computation involved in such exhaustive searching may often become prohibitively expensive when the number of parameters runs into hundreds or thousands. Additionally, unobserved or unknown interactions that were not accounted for in the formulation of the model can result in unsuccessful parameter searches and will require the model assumptions to be revisited.

The calculation of the first and second partial derivatives of the objective function is sometimes useful in optimization. Gradient search optimization algorithms depend on the partial derivatives of the objective function (or the partial derivative matrix, the Jacobian) for their success [4]. The Jacobian J is defined as

$$J = \frac{\partial \Phi}{\partial \theta}. \tag{6}$$

Occasionally it is also useful to analyze how the combination of two parameters may affect the system dynamics; this is accomplished through the Hessian matrix H:

$$H = \frac{\partial^2 \Phi}{\partial \theta \partial \theta^T}, \tag{7}$$

which is a matrix containing the second partial derivatives of the objective function with respect to pairs of parameters.

In gradient search methods it is not always possible to reach the global optimum of the objective function, especially for nonlinear objective functions that may have several local optima far away from the global optimum.

Therefore, it may become necessary to sacrifice the speed of the gradient search methods for the exhaustive searching abilities of probabilistic methods [45]. Such methods avoid the inferior solutions often found by gradient search methods [46]. Examples of probabilistic optimization methods are simulated annealing, genetic, and evolutionary algorithms [47, 48] or scatter searches [49]. Conversely, a local minimum may sometimes be sufficient to generate parameter values that are practical. In some particularly difficult combinatorial optimization problems the local minimum found near the global minimum may turn out to be a better choice [50]. On some rare occasions, perhaps many of the optima could be useful for the researcher.

5. Model Analysis Techniques

After a satisfactory model is constructed, the analysis of the model and its predictions provides crucial input toward designing synthetic circuits that exhibit a given behavior. Here, we discuss two analysis techniques: sensitivity analysis and bifurcation analysis.

5.1. Sensitivity Analysis. Sensitivity analysis, which analyzes how sensitive a system is with respect to changes in parameter values [51], is useful in quantifying the significance of a parameter or parameters on system performance [4]. Local sensitivity analysis analyzes the effects of small perturbations whereas global sensitivity analysis is used to analyze the effects of perturbing parameter values over the entire parameter ranges. Caution should be exercised while extrapolating the results of sensitivity analyses to an operating point far away in parameter space, as this may not accurately predict system behavior at the operating point.

For local sensitivity analysis a sensitivity s is defined as

$$s = \frac{\partial G(N)}{\partial \theta}, \tag{8}$$

where N is a vector comprised of species concentrations, $G(N)$ represents a system state or output, and θ is a vector of parameters. The magnitudes of the elements of the vector s are proportional to the effect of the corresponding parameters [4]. In global sensitivity analysis, the parameter space (constrained by physical limitations, mass fraction summations, etc.) is explored by methodologies such as random sampling-high dimensional model representation [52], multiparametric sensitivity analysis, or Monte Carlo simulation [53].

5.2. Bifurcation Analysis. Bifurcation analysis is crucial to understanding and analyzing steady states, oscillations, and other dynamic features of a system and has found use in numerous modeling studies [6, 54, 55]. In many nonlinear models, the parameter space can be divided into regions that lead to a stable system, an unstable system, or a periodic (oscillatory) system. Identifying the boundaries of these regions will enable the design of a synthetic circuit that exhibits desired behavior. Two popular methods of bifurcation analysis are saddle node bifurcation analysis and Hopf bifurcation analysis. Saddle node analysis endeavors to investigate the threshold where a system functions as a biological switch and thus separates the region of the parameter space that confers monostability [56]. Hopf bifurcation analysis is predominantly used to analyze oscillators and can characterize the critical parameter values that enable a system to transition from a stable steady-state solution to a periodic solution [57, 58]. In Hopf bifurcation analysis the eigenvalues of the Jacobian matrix, equation (6), are used to obtain the threshold parameter values at which the system's behavior changes drastically. A phase diagram is then constructed by identifying the points where the real parts of a pair of complex conjugate eigenvalues crosses zero while all other eigenvalues have negative real parts (see [6], e.g.; this paper is discussed in case study I below). While there are numerous examples of bifurcation analysis of deterministic models, one work [27] has comprehensively treated the applications of bifurcation analysis to a stochastic model of an oscillatory system.

6. Modeling as a Part of Phenotype Analysis

Synthetic biology can also benefit from metabolic flux and transcription network analyses, which combine high-throughput experimental observations (such as metabolome, isotopomer, and gene expression profiles) with mathematical modeling to quantitatively describe the phenotype of a biological system. This type of analysis could be particularly useful for highly complex systems. For example, Noirel et al. [59] presented a probabilistic metabolic model that was useful in analyzing the systemic metabolic effects of inserting synthetic circuits into a cell. Of specific relevance to synthetic biology is an elegant work [60] that used optimization to identify how regulation could be superposed on a metabolic network to optimize the network.

6.1. Metabolic Pathway Analysis. Isotope-assisted [61–63] metabolic flux analysis is a powerful tool to evaluate carbon flow in metabolic networks and could be relevant in synthetic biology. In this method, material balance models for cellular species are used together with measurements of extracellular metabolites or isotopomers (from nuclear magnetic resonance or mass spectrometry) to obtain metabolic flux maps of a system. While the mathematical modeling approaches discussed above are generally still valid in flux analysis, these models can be enormously complex due to the vast numbers of reactions, myriad carbon atom rearrangements, reaction reversibilities, and variations of intracellular fluxes in real time. This complexity is significantly reduced through the use of compartmental matrix techniques such as cumulative isotopomers (cumomers) [64], bond isomers (bondomers) [65, 66], or elementary metabolite units [63]. Despite this reduction in complexity, metabolic flux analysis of compartmented [61] or instationary systems [62, 67] requires tremendous amounts of computation.

Another set of powerful techniques for modeling metabolic networks includes flux balance analysis (FBA) [68–70] or genome-scale metabolic modeling [71–73] and

Table 1: Modeling techniques used in the case studies.

	Case study	Model types		Modeling techniques			
		Deterministic	Stochastic	Parameter estimation	Optimization	Sensitivity analysis	Bifurcation analysis
Case study I	Fung et al. [6]; designed a synthetic oscillator in *E. coli.*	×		×	×	×	×
Case study II	Friedland et al. [7]; developed synthetic counters in *E. coli* that can count up to two or three induction events.	×		×	×	×	
Case study III	Guido et al. [8]; constructed regulatory networks by assembling simpler building blocks.	×	×	×	×	×	

ensemble modeling [31]. In FBA, a metabolic network is modeled with linear stoichiometric equations, constrained by factors such as extracellular flux measurements and reaction irreversibilities. This model is usually solved by linear optimization and results in a map of steady-state flux values. Through such a reconstruction of a metabolic network, FBA can offer important insights towards selecting gene deletion targets.

Another valuable technique, metabolic control analysis [74–76], aims to elucidate the interdependence of various parts of a metabolic network. The outcomes of this technique are metrics such as flux control coefficients [77], which represent the amount of control exerted by one system component (such as a metabolite) on another system component (such as an enzyme). This method has much to offer toward the important problem of linking genome and phenotype [78].

6.2. Transcription Network Analysis. Determining how genes are controlled by regulatory motifs is an important problem in biology. Because synthetic circuits are composed of well-characterized components, they can be used to investigate and quantify transcription networks. Such an investigation would employ a combinatorial technique to construct a circuit comprised of numerous genes and a smaller number of regulatory motifs [79]. The high-dimensional output of such a network is gene expression data and is the end product of the low-dimensional regulatory signals (transcription factor activities) and the strengths of the connectivities between the transcriptional motifs and genes [80]. These transcription factor activities and connectivities are quantified by analyzing the measured gene expression data, using one or more of several available methods [81]. These methods include principal component analysis [82], singular value decomposition [83, 84], independent component analysis [85], network component analysis [80], or state-space models [86]. Network component analysis is a powerful method that uses a priori knowledge about connectivities between transcription factors and genes together with gene expression data to quantitatively infer transcription factor activities and the strengths of the transcription factor-gene connectivities. The a priori information is obtained from databases or experimental techniques such as ChIP-chip analysis [87].

7. Modeling Standards and Software

Several standards and software are now available to simplify the process of building mathematical models and thereby bridge the gap between model description and prediction of the system's behavior. System biology markup language (SBML) [88] and synthetic biology open language (SBOL) (http://dspace.mit.edu/handle/1721.1/49523) are two examples of standards. Both are computer-readable formats for representing models and facilitate the sharing of models between researchers and between different software platforms. Several modeling software are available to practitioners of synthetic biology. These software, which are usually written in popular computer languages such as C++, feature a user interface, relatively simple ways to input information and graphical output of the modeling outcomes, thus relieving users of the burden of setting up and solving mathematical equations. A nonexhaustive listing of these software includes Athena (http://www.codeplex.com/athena), BioJade (http://web.mit.edu/jagoler/www/biojade), Gepasi (http://www.gepasi.org), SynBioSS (http://synbioss.sourceforge.net/), which reads and writes in SBML, and TinkerCell (http://www.tinkercell.com/Home) which enables users to incorporate new features through custom programs written in C or Python.

8. Case Studies

Below we present three case studies that illustrate several of the previously discussed modeling methodologies. The case studies feature two deterministic models [6, 7] and one stochastic model [8]. Table 1 summarizes the principal modeling techniques used in the case studies.

8.1. Case Study I: Deterministic Model of a Synthetic Oscillator. Fung et al. [6] reported the mathematical model-aided design of a gene-metabolic oscillatory circuit called

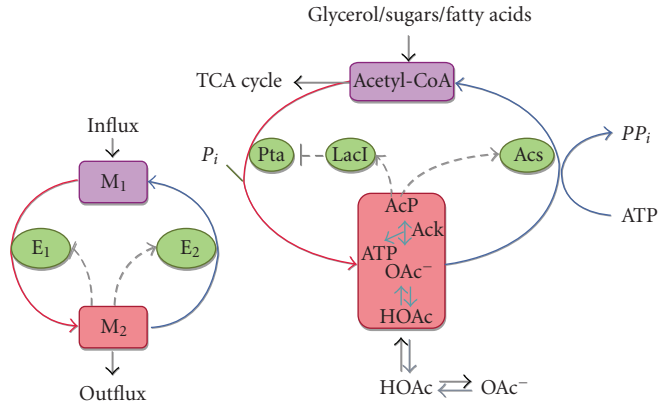


FIGURE 3: Conceptual design and biological implementation of the oscillatory circuit metabolator in Fung et al. [6]. (a) Conceptual design. The metabolator consists of two interconverting metabolite pools M_1 and M_2; their interconversions are catalyzed by the enzymes E_1 and E_2. Dashed lines indicate positive (arrow) and negative (blunt bar) regulation by M_2 at the transcriptional or translational level; the accumulation of M_2 represses E_1 and induces E_2. The circuit functions as follows. Influx into the circuit (from upstream processes) increases the concentration of M_1, which is converted to M_2 by E_1. Initially the concentration of M_1 is high and M_2 is low. However, M_2 gradually accumulates causing E_1 to be repressed and E_2 to be induced, eventually causing a net conversion of M_2 to M_1, which then starts a new cycle. (b) Biological implementation. The design of the metabolator was implemented using the acetate pathway in *E. coli*. The M_1 pool is acetyl-CoA; the M_2 pool consists of AcP, OAc^-, and HOAc. Pta and Acs correspond to enzymes E_1 and E_2. Pta converts Acetyl-CoA to AcP, and AcP is further converted to OAc^- by Ack. The protonated form of OAc^- (HOAc) is permeable across the cell membrane. AcP is used as a signaling molecule and functions as follows. When AcP builds up, it will activate promoter *glnAp2* through phosphorylation. The promoter *glnAp2* controls the expression of Acs and lac repressor (LacI), and LacI in turn represses the expression of Pta. Ack: acetate kinase; AcP: acetyl phosphate; Acs: acetyl-CoA synthetase; HOAc: protonated form of acetate; LacI: lac repressor; OAc^-: acetate; Pta: phosphate acetyltransferase (adapted from Fung et al. [6]).

the metabolator. Toward designing an oscillatory circuit the authors conceived a network with two interconverting metabolite pools wherein one metabolite differently regulates the enzymes that interconvert the two pools. Such a network is theoretically capable of oscillation (see Figure 3(a)). The circuit was implemented in *Escherichia coli* using the acetate pathway (see Figure 3(b)). Under certain circumstances, the sizes of the pools M_1 (Acetyl-CoA) and M_2 (lumped pool of Ack, AcP, OAc^-, and HOAc) can oscillate. The readout of this network was a GFP, located downstream of the network such that the oscillations of the GFP readout reflect any oscillations occurring in the network. A mathematical model was developed to analyze the behavior of the metabolator and determine the conditions under which oscillations would occur. The model was deterministic and employed ODEs of the following form:

$$\frac{dX}{dt} = V_{\text{in}} - V_{\text{out}}, \tag{9}$$

$$\frac{dX}{dt} = \text{synthesis rate} - \text{degradation rate}, \tag{10}$$

where X represents any pool and V_{in}, V_{out} indicate its influx and outflux, respectively. Equation (9) was used to describe the metabolite pools M_1 and M_2 while equation (10) and Michaelis-Menten rate laws were used to describe the kinetics of the enzymes driving the M_1-M_2 interconversions. Using parameter values or ranges typical for this system, the authors implemented the model with a fourth-order Runge-Kutta algorithm. Parameter sensitivity analysis showed that increasing glycolytic flux increases the oscillatory capability of the system (Figure 4(a)). This prediction was tested by experimentally varying the glycolytic flux through the feeding of different carbon sources (glucose, fructose, mannose, and glycerol), each of which resulted in a different value of this flux. An explanation for this observation is that there existed a threshold value of the glycolytic flux beyond which the system would oscillate. Conversely, high external acetate was predicted to suppress oscillation (Figure 4(c)), which was also verified by experiments.

Hopf bifurcation was then used to characterize the dynamics of the model and determine the transition point at which the steady state would turn to a periodic state. Figure 4(b) depicts the phase diagrams constructed by mapping the locus of Hopf bifurcation. The oscillations approach a limit cycle and were stable, as determined through Floquet analysis. As previously inferred, oscillations occur above a threshold glycolytic flux value and are not sustained at a high acetate concentration. Furthermore, by comparing the modeling simulations with experimental data, the authors found that the inherent noise in gene expression was an important determinant of the amplitude of oscillation.

This work represents a universal approach to construct a synthetic biological circuit with interesting dynamics and beautifully demonstrates the key role played by mathematical modeling in realizing a design concept. Modeling offered valuable insights on the analysis of experiment data and made nontrivial predictions of the system dynamics. The

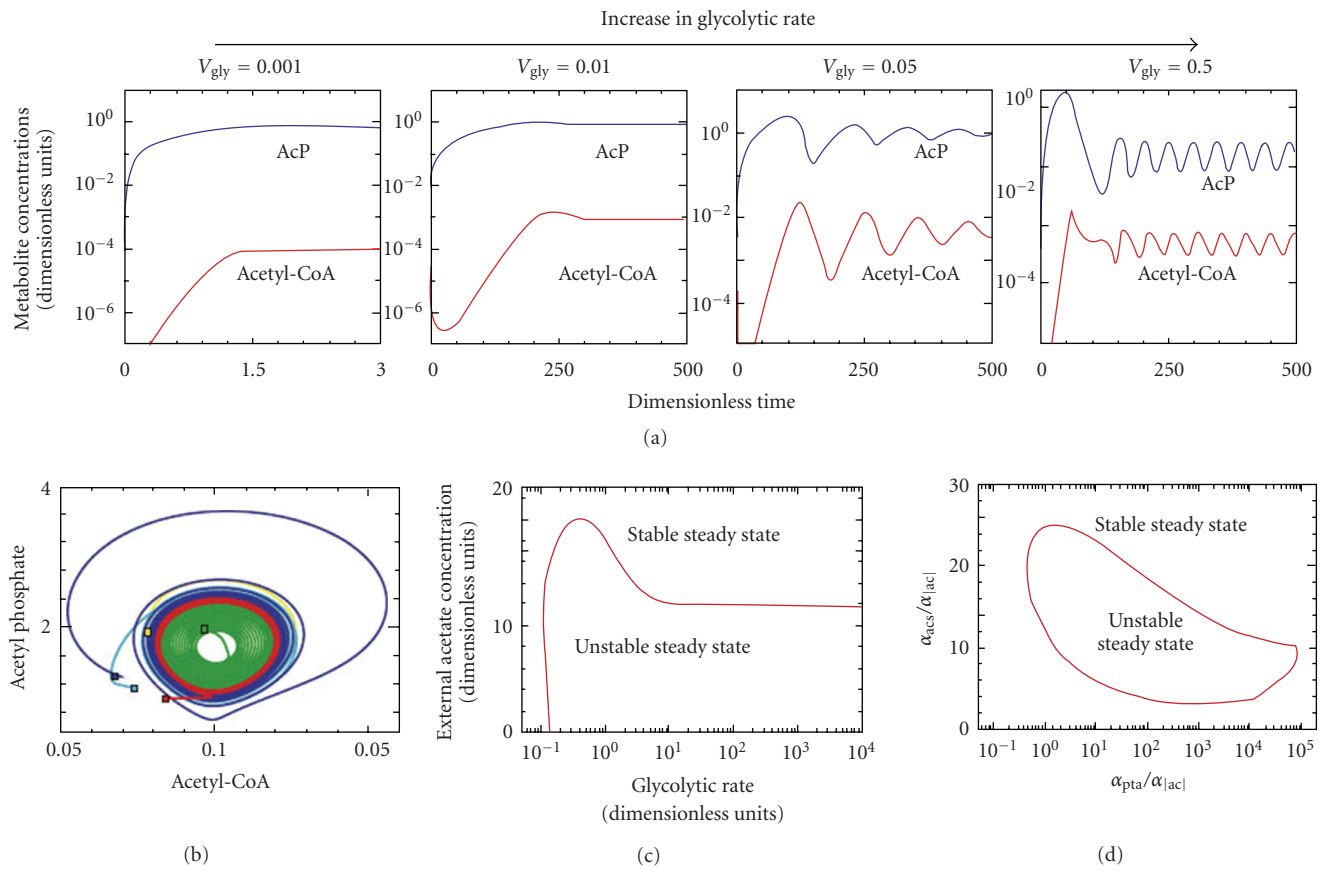


FIGURE 4: Sensitivity and bifurcation analyses of the metabolator model in Fung et al. [6]. (a) Sensitivity analysis: increase in glycolytic rate increases the oscillatory capability of the metabolator. The glycolytic rate V_{gly} is equal to 0.001, 0.01, 0.05, and 0.5 in the top four panels from left to right. (b) Phase plots obtained by perturbing the steady-state solution at $V_{gly} = 1$ show that the oscillatory dynamics is limit cycle oscillation irrespective of the initial condition. The initial state of the oscillator is depicted with squares. (c) Hopf bifurcation analysis was used to construct a phase diagram of glycolytic rate versus external acetate concentration. The flux-sensitive nature of the oscillations is evident here; low glycolytic fluxes lead to a stable steady state with oscillations setting in beyond a threshold value of the glycolytic flux. (d) Another phase diagram suggests that at $V_{gly} = 10$, specific combinations of three protein levels are required to sustain oscillation. The variable α_i represents rate of synthesis of protein i (i: LacI, Pta or Acs) (from Fung et al. [6], with permission).

use of bifurcation analysis was particularly useful as it facilitated the determination of the points at which the system transitions between stable and periodic states. We expect that as oscillator design develops [21], modeling will become ever more relevant in the design process.

8.2. Case Study II: Deterministic Model of a Synthetic Counter. Another example of deterministic modeling is that of Friedland et al. [7], who developed synthetic counters in *E. coli* that can count up to two or three induction events. The first of these, a riboregulated transcriptional cascade (RTC) two-counter, has two nodes and is able to count up to two arabinose pulses by expressing a different protein in response to each pulse. This was extended to the RTC three-counter (Figure 5) which has three nodes and can count up to three pulses as follows. The constitutive promoter $p_{Ltet0\text{-}1}$ drives transcription of T7 RNA polymerase (T7 RNAP), whose gene product binds the T7 promoter, which in turn drives the transcription of T3 RNAP. Similarly the protein of T3 RNAP binds the T3 promoter and ultimately drives the transcription of GFP. All genes are further downregulated and upregulated by cis and trans elements of riboregulators. A cis repressor (cr) interacts with the downstream ribosome binding site (RBS) in such a manner as to prevent translation. The arabinose promoter p_{BAD} drives a transactivating, noncoding RNA (taRNA) that binds to cr in trans, thus relieving the inhibition of translation. Due to this design, protein synthesis at each node requires both transcription and translation to independently happen. Thus this cascade is able to count three arabinose pulses by expressing a different protein in response to each pulse.

The authors constructed and analyzed a mathematical model for the two- and three-counters. The model used equations of the form of (10) to describe the dynamics of the species in the circuit. The degradation terms in these equations were assumed to be simple exponential decays with a different rate constant for each species whereas the synthesis rates were rate laws that reflected how each species was synthesized. The Hill function was used to describe arabinose induction and the dynamics of GFP. Arabinose

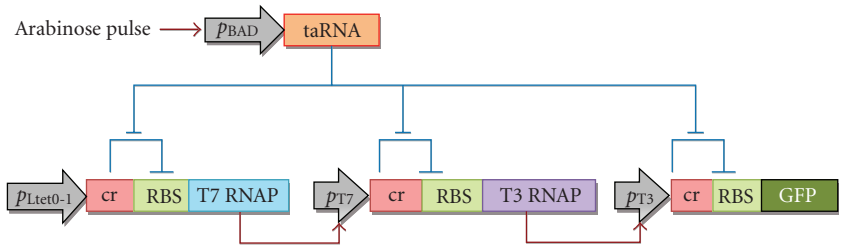


FIGURE 5: Design concepts for the RTC three-counter in Friedland et al. [7]. To count three induction events, the RTC three-counter employs a transcriptional cascade that has three nodes. The constitutive promoter $p_{Ltet0-1}$ drives transcription of T7 RNA polymerase (T7 RNAP), whose gene product binds the T7 promoter, which in turn controls the transcription of T3 RNAP. Similarly the protein of T3 RNAP binds the T3 promoter and ultimately controls the transcription of GFP. All genes are further downregulated and upregulated by cis and trans elements of riboregulators. A cis repressor (cr) interacts with the downstream ribosome binding site (RBS) in such a manner as to prevent translation. The arabinose promoter p_{BAD} drives a transactivating, noncoding RNA (taRNA) that binds to cr in trans, thus relieving this inhibition of translation. Due to this design, protein synthesis at each node requires both transcription and translation to independently happen. Thus this cascade is able to count three arabinose pulses by expressing a different protein in response to each pulse (adapted from Friedland et al. [7]).

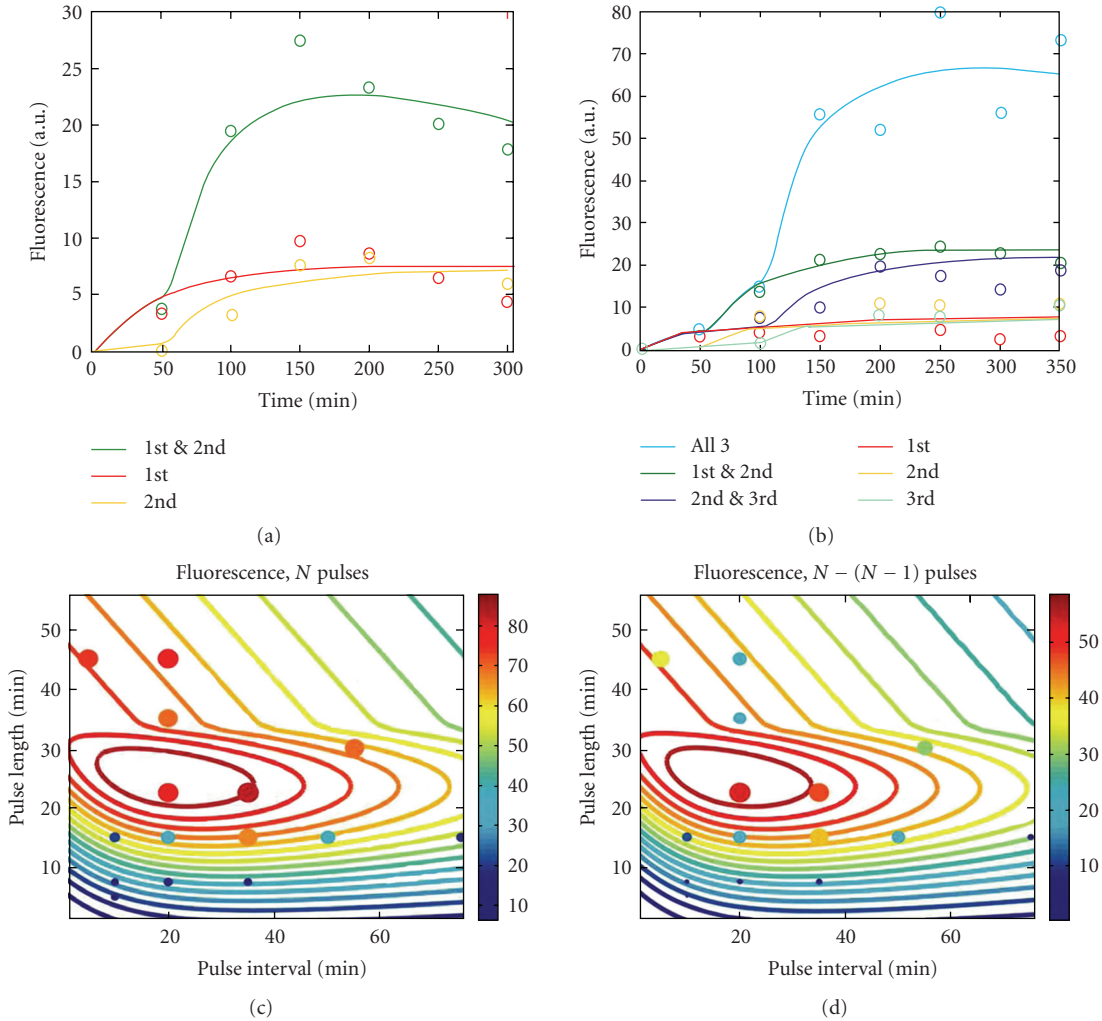


FIGURE 6: Modeling predictions and analysis of the RTC three-counter in Friedland et al. [7]. (a) A model of (a) the RTC two-counter and (b) the RTC three-counter with fitted parameters was simulated (solid line) and agrees well with experimental data (normalized fluorescence, solid dots). (c) The output of the RTC three-counter (N pulses) is simulated for a range of pulse lengths and intervals. The predictions (colored contour lines) match experimental results (solid circles), whose levels are indicated by both color and size. (d) Similar to (c), except that values shown here are the differences in the output of the three-counter after three (N) pulses and two ($N-1$) pulses (from Friedland et al. [7], with permission).

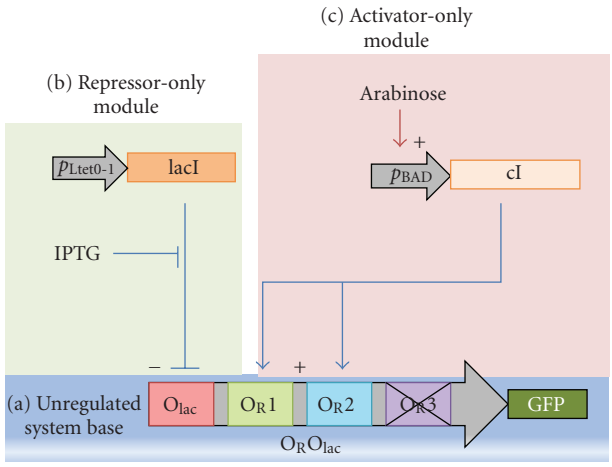


FIGURE 7: Design concept for the repressor-activator system in Guido et al. [8]. The repressor-activator system was assembled by combining (a) an unregulated system base with (b) a repressor-only module and (c) an activator-only module. The unregulated system consists of an O_RO_{lac} promoter driving GFP expression. The rectangular boxes in the promoter are the binding sites for the lacI and cI proteins. The cross at O_R3 indicates a point mutation that inhibits cI binding at this site. In (b), the repressor-only module, the promoter $p_{Ltet0-1}$ controls the transcription of lacI, which binds to the O_{lac} site to repress the O_RO_{lac} promoter. The extent of this effect can be tuned by using the lacI inhibitor IPTG, which reduces the binding of lacI protein to the O_{lac} site. In (c), the activator-only module, the p_{BAD} promoter controls the transcription of cI, which can bind to either the O_R1 or the O_R2 sites, sequentially or cooperatively, to activate the O_RO_{lac} promoter. The extent of this effect can be tuned through arabinose, which activates the p_{BAD} promoter (adapted from Guido et al. [8]).

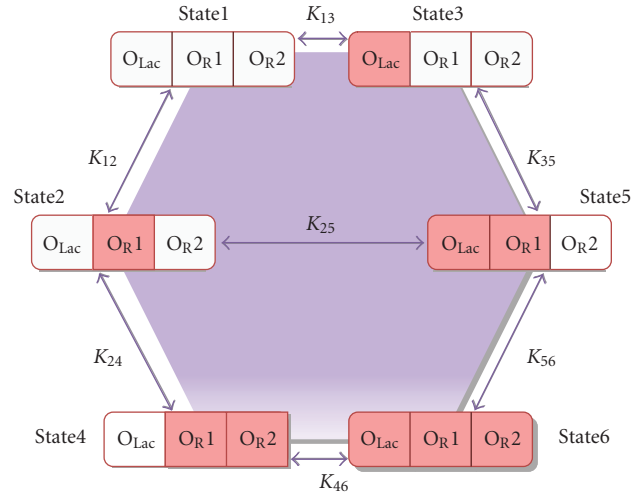


FIGURE 8: The six possible binding states in the repressor-activator model in Guido et al. [8]. The rectangular boxes represent the three available operator sites of engineered O_RO_{lac} promoter. Unbound sites are depicted with unfilled (white) boxes; filled (pink) boxes depict sites bound by the appropriate protein (lacI to O_{lac}, cI to O_R1 and O_R2). Of the eight (2^3) possible states, two are not feasible. This is because cI binds sequentially to the O_R1 and O_R2 sites so that O_R2 cannot bind unless O_R1 is also bound. The model parameters K_{12} through K_{56} are the rate constants for the transitions between each possible binding state (adapted from Guido et al. [8]).

pulse dynamics were modeled with two differential equations as follows. Arabinose consumption from the medium was modeled as a zero order rate law:

$$\frac{d[ara]}{dt} = -k_c, \tag{11}$$

whereas the decay of intracellular arabinose in cells suspended in arabinose-free media was modeled as a first order rate law:

$$\frac{d[ara]}{dt} = -k_d \cdot [ara]. \tag{12}$$

The arabinose pulses were mimicked by alternately using equation (11) and equation (12). The authors used optimization (implemented by a MATLAB routine *lsqcurvefit*) to evaluate (fit) the parameters in the model so as to agree with experimental data (Figures 6(a) and 6(b)). The model with fitted parameters was used to examine the effects of arabinose pulse frequency and length on the performance of the RTC three-counter (Figures 6(c) and 6(d)). The simulations indicated the ranges of pulse intervals (10 to 40 minutes) and pulse length (20 to 30 minutes) that would result in maximal system response (GFP fluorescence); these predictions were later confirmed by experiments (Figures 6(c) and 6(d)). The simulations indicated that the system response increased significantly when two or three pulses were delivered. Also indicated by the simulations was a small amount of leakage that occurred in response to partial pulses; this was also verified experimentally (Figures 6(a) and 6(b)). The authors further used this approach to construct a DNA invertase cascade (DIC) counter and discussed extending this counter with the use of other unique polymerases or recombinases to record N sequenced events.

The RTC two- and three-counters constitute an elegant example showing how synthetic circuit elements can be combined to recognize sequential events. Here too, the mathematical model was important in investigating the system dynamics and identifying the pulse length and interval that yielded the most effective response. Parameter estimation (and thereby optimization) provided crucial insights toward improving counter performance. As these counters are expanded to become capable of counting larger numbers of events (and thereby increasing in complexity), the role played by mathematical modeling in design will become increasingly important.

8.3. Case Study III: Stochastic Model of a Bottom-Up Gene Regulatory Circuit. Guido et al. [8] constructed regulatory networks by assembling simpler modules, such that the behavior of the network was predictable from that of the components. The authors engineered the O_RO_{lac} promoter such that it caused both repression and activation of gene expression in *E. coli*. For this, they combined an unregulated promoter, a repressor-only and an activator-only system (Figure 7). The unregulated system (Figure 7(a)) consists

Steps in the model of Guido et al. [8]

Step 1. Calculating the state probability of each promoter

Using p_1 as the reference, the probabilities for other five states could be described as:

$$p_2 = k_{12}p_1$$
$$p_3 = k_{13}p_1$$
$$p_4 = k_{12}k_{24}p_1$$
$$p_5 = k_{13}k_{35}p_1$$
$$p_6 = k_{12}k_{24}k_{46}p_1$$

Probability constraint:

$$\sum_{i=1}^{6} p_i = 1$$

Thermodynamic equilibrium constraints:

$$\frac{k_{12}k_{25}}{k_{13}k_{35}} = 1, \frac{k_{24}k_{46}}{k_{25}k_{56}} = 1$$

From the first constraint, p_1, the probability of the state of the unregulated promoter, is obtained as:

$$p_1 = \frac{1}{1 + k_{12} + k_{13} + k_{12}k_{24} + k_{13}k_{35} + k_{12}k_{24}k_{46}}$$

Step 2. Estimating parameters

Factors controlling equilibrium constants:

k_{13}, k_{25} and k_{46} depend on the concentration of IPTG

k_{12}, k_{24}, k_{35} and k_{56} depend on the concentration of arabinose

For the repressor-only system:

$k_{12} = k_{24} = k_{35} = k_{56} = 0$

For activator-only system:

$k_{13} = k_{25} = k_{46} = 0$

Step 3. The model response (average mRNA synthesis rate)

The average synthesis rate is given by:

$$\overline{y}_m = p_1 + g_2p_2 + g_3p_3 + g_4p_4 + g_5p_5 + g_6p_6$$

where g_i is a nonlinear function that represents the mRNA synthesis rate in the ith promoter state.

Step 4. Extend the model with stochastic effects

Incorporate cell growth, synthesis and degradation rates of mRNA and GFP, multimerization of cI.

Figure 9: Steps in the repressor-activator mathematical model in Guido et al. [8].

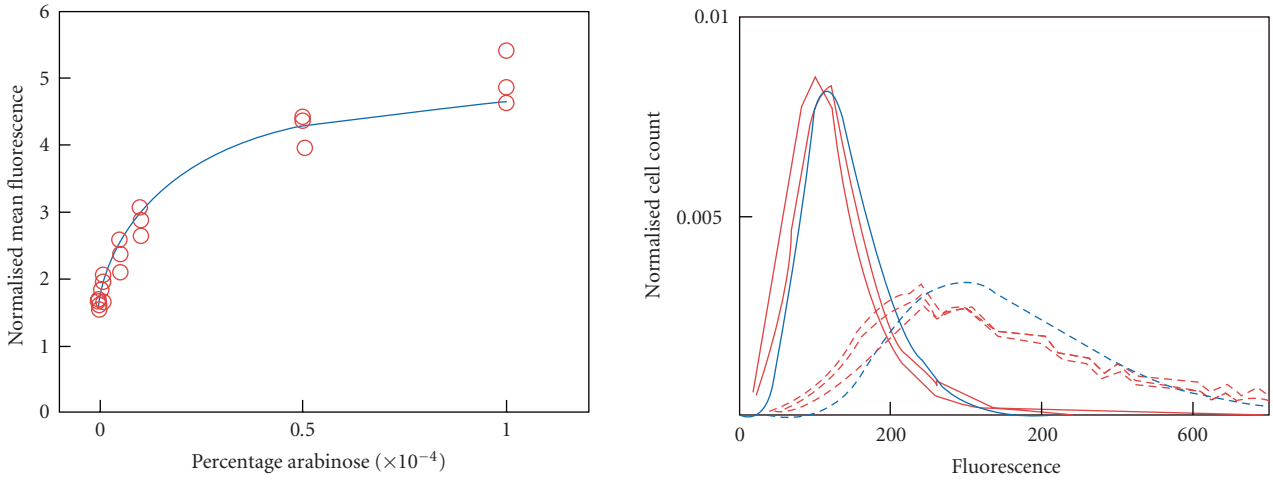


Figure 10: Model predictions and analysis of the repressor-activator system in Guido et al. [8]. (a) Simulations with baseline deterministic model: dependence of the model response (GFP fluorescence) of the repressor-activator system on arabinose concentration, with IPTG concentration maintained constant. The model simulation is shown with a blue line and experimental data points are shown with red circles. (b) Simulations with the extended stochastic model: distribution of GFP fluorescence within the cell population for levels of arabinose ranging from no arabinose (solid lines toward left) to the maximum possible level of arabinose (dashed lines toward right); IPTG concentration is maintained constant throughout. Stochastic model simulations are shown with blue lines and experimental data is shown with red lines (from Guido et al. [8], with permission).

of an O_RO_{lac} promoter driving GFP expression. In the repressor-only module (Figure 7(b)), the promoter $p_{Ltet0\text{-}1}$ controls the transcription of lacI, which binds to the O_{lac} site to repress the O_RO_{lac} promoter. This binding can be tuned by with the lacI inhibitor isopropyl-b-D-thiogalactopyranoside (IPTG), which reduces the binding of lacI protein to the O_{lac} site. In the activator-only module (Figure 7(c)), the p_{BAD} promoter controls the transcription of cI, which can bind sequentially or cooperatively to either the O_R1 or the O_R2 sites of the O_RO_{lac} promoter to activate it. This effect is

tunable with arabinose, which activates the p_{BAD} promoter. The repressor-activator system is constructed by combining all three of these modules.

The authors first developed a baseline deterministic model for the repressor-activator system and then extended it to include stochastic effects. A quasi-equilibrium state was assumed for the promoter $O_R O_{lac}$ such that there were six possible chemical states of the promoter (Figure 8). The interconversions between the states were modeled with the equation:

$$K_{ij} = K_{ij}^{eq}[P], \tag{13}$$

where K_{ij} is the transition rate from promoter state S_i to S_j, K_{ij}^{eq} is the equilibrium constant for the interconversion, and $[P]$, depending on the reaction, is the concentration of either *lacI* or *cI*. The probabilities of each state were functions of the transition rates K_{ij} and the average GFP synthesis rate was in turn a function of the state probabilities (see discussion around equation (4) and Figure 9 for details). The authors extended the model with stochastic effects so that it was able to simulate cell growth and the distribution of GFP within the cell population. For example cell growth was modeled by treating the cell volume as a random variable with an exponential distribution. The model assumed that at every instance of cell division, the volume halves and the cellular components (mRNA and GFP) are distributed amongst the daughter cells in a binomial distribution. The simulations of the deterministic and stochastic parts of the model are shown in Figure 10. The parameters in this model were determined by a gradient search local optimization algorithm and all stochastic modeling was implemented through the Gillespie Monte Carlo method realized by BioNetS software. The authors experimentally measured all mRNA synthesis rates and GFP fluorescence relative to the unregulated system and chose the parameter values that best fitted the fluorescence results shown in Figure 10. The model slightly underestimates the experimentally observed variability of fluorescence within the cell population which indicates that effects not considered in the model might be occurring.

To further verify the predictive power of the model and test its stochastic aspects, the authors expanded the circuit to include a positive feedback. This was accomplished adding the *cI* gene in front of GFP so that the repressor-activator system transcribes both *cI* and GFP simultaneously. This amounts to positive feedback because the product of the *cI* gene activates gene expression in the system. To effectively model this feedback loop, the model was extended to include synthesis, degradation, and multimerization of *cI*. This model with positive feedback exhibited very good agreement with experimental data. This agreement validated the bottom-up approach employed by the authors to study regulatory systems. Furthermore, the model was used to make the counterintuitive prediction that cessation of cell growth and division increases noise in protein expression levels, which was also verified experimentally.

Although other researchers similarly built larger regulatory systems from simple ones (e.g., [89]), the work by Guido et al. emphasized the important role played by mathematical modeling in the design and in making counterintuitive predictions about system dynamics. We expect that in the future, mathematical modeling along with transcription network analyses will become indispensable in the design of more complex regulatory networks.

9. Conclusion and Outlook

Mathematical modeling is an often indispensable tool in synthetic biology. The mathematical techniques of parameter estimation as well as sensitivity and bifurcation analyses can be crucial to the development of a model intended to mimic a complex system. Modeling also plays an important role in phenotypic analyses such as metabolic flux analysis or transcription network analysis.

A mathematical model is akin to a road map that provides a visualization of a geographical area. Although the map may not describe every detail of the landscape, it contains adequate information to enable users to plan a journey; a mathematical model is similar in scope [4]. A seminal review on mathematical modeling [90] stated that the purpose of models is to discern which parts and connections of a system are significant, to unravel new strategies, or sometimes to correct conventional wisdom. Examples of these uses abound in synthetic biology, where models have been employed to identify which regions in parameter space cause a system to behave in a desired manner (e.g., [6] and case study I) or what parameter values result in the most effective design (e.g., [7] and case study II). Furthermore, models have also been employed to understand the global dynamics of a system from known behaviors of its component units and have made counterintuitive predictions that were later verified experimentally (e.g., [8] and case study III). We anticipate that as experimental advances in synthetic biology produce increasingly complex circuits, mathematical modeling will play an ever more important function as a bridge between concept and realization.

Acknowledgments

This work was funded by the Department of Chemical and Biomolecular Engineering, University of Maryland (faculty startup grant to GS), the A. James Clark School of Engineering, University of Maryland (Minta Martin award to GS), and Maryland Industrial Partnerships (MIPS) (award no. 4426). Y. Zheng was cofunded by a Jan and Anneke Sengers fellowship.

References

[1] F. J. Isaacs, D. J. Dwyer, and J. J. Collins, "RNA synthetic biology," *Nature Biotechnology*, vol. 24, no. 5, pp. 545–554, 2006.

[2] M. Heinemann and S. Panke, "Synthetic biology—putting engineering into biology," *Bioinformatics*, vol. 22, no. 22, pp. 2790–2799, 2006.

[3] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging

discipline," *Molecular Systems Biology*, vol. 2, Article ID 2006.0028, 2006.

[4] E. L. Haseltine and F. H. Arnold, "Synthetic gene circuits: design with directed evolution," *Annual Review of Biophysics and Biomolecular Structure*, vol. 36, pp. 1–19, 2007.

[5] D. Endy, "Foundations for engineering biology," *Nature*, vol. 438, no. 7067, pp. 449–453, 2005.

[6] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S.-G. Lee, and J. C. Liao, "A synthetic gene-metabolic oscillator," *Nature*, vol. 435, no. 7038, pp. 118–122, 2005.

[7] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins, "Synthetic gene networks that count," *Science*, vol. 324, no. 5931, pp. 1199–1202, 2009.

[8] N. J. Guido, X. Wang, D. Adalsteinsson, et al., "A bottom-up approach to gene regulation," *Nature*, vol. 439, no. 7078, pp. 856–860, 2006.

[9] D. Chandran, W. B. Copeland, S. C. Sleight, and H. M. Sauro, "Mathematical modeling and synthetic biology," *Drug Discovery Today: Disease Models*, vol. 5, no. 4, pp. 299–309, 2008.

[10] Y. N. Kaznessis, "Models for synthetic biology," *BMC Systems Biology*, vol. 1, article 47, 2007.

[11] O. Brandman, J. E. Ferrell Jr., R. Li, and T. Meyer, "Systems biology: interlinked fast and slow positive feedback loops drive reliable cell decisions," *Science*, vol. 310, no. 5747, pp. 496–498, 2005.

[12] A. Ma'ayan, S. L. Jenkins, S. Neves, et al., "Formation of regulatory patterns during signal propagation in a mammalian cellular network," *Science*, vol. 309, no. 5737, pp. 1078–1083, 2005.

[13] J. E. Lunn, "Compartmentation in plant metabolism," *Journal of Experimental Botany*, vol. 58, no. 1, pp. 35–47, 2007.

[14] B. N. Kholodenko, "Cell-signalling dynamics in time and space," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 3, pp. 165–176, 2006.

[15] M. J. Hounslow, R. L. Ryall, and V. R. Marshall, "A discretized population balance for nucleation, growth, and aggregation," *AIChE Journal*, vol. 34, no. 11, pp. 1821–1832, 1988.

[16] A. Ma'ayan, "Insights into the organization of biochemical regulatory networks using graph theory analyses," *Journal of Biological Chemistry*, vol. 284, no. 9, pp. 5451–5455, 2009.

[17] S. Mangan and U. Alon, "Structure and function of the feed-forward loop network motif," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 11980–11985, 2003.

[18] J. E. Ferrell Jr., "Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability," *Current Opinion in Cell Biology*, vol. 14, no. 2, pp. 140–148, 2002.

[19] U. S. Bhalla and R. Iyengar, "Emergent properties of networks of biological signaling pathways," *Science*, vol. 283, no. 5400, pp. 381–387, 1999.

[20] A. Becskei and L. Serrano, "Engineering stability in gene networks by autoregulation," *Nature*, vol. 405, no. 6786, pp. 590–593, 2000.

[21] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator," *Nature*, vol. 456, no. 7221, pp. 516–519, 2008.

[22] M. B. Elowitz and S. Leibier, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.

[23] J. J. Tyson, K. Chen, and B. Novak, "Network dynamics and cell physiology," *Nature Reviews Molecular Cell Biology*, vol. 2, no. 12, pp. 908–916, 2001.

[24] L. Allen, *An Introduction to Stochastic Processess with Applications to Biology*, Pearson Education, Upper Saddle River, NJ, USA, 2003.

[25] H. Salis and Y. N. Kaznessis, "An equation-free probabilistic steady-state approximation: dynamic application to the stochastic simulation of biochemical reaction networks," *Journal of Chemical Physics*, vol. 123, no. 21, Article ID 214106, 16 pages, 2005.

[26] H. Salis and Y. Kaznessis, "Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions," *Journal of Chemical Physics*, vol. 122, no. 5, Article ID 054103, 13 pages, 2005.

[27] L. M. Tuttle, H. Salis, J. Tomshine, and Y. N. Kaznessis, "Model-driven designs of an oscillating gene network," *Biophysical Journal*, vol. 89, no. 6, pp. 3873–3883, 2005.

[28] B. Zhou, D. Beckwith, L. R. Jarboe, and J. C. Liao, "Markov chain modeling of pyelonephritis-associated pili expression in uropathogenic *Escherichia coli*," *Biophysical Journal*, vol. 88, no. 4, pp. 2541–2553, 2005.

[29] L. R. Jarboe, D. Beckwith, and J. C. Liao, "Stochastic modeling of the phase-variable pap operon regulation in uropathogenic *Escherichia coli*," *Biotechnology and Bioengineering*, vol. 88, no. 2, pp. 189–203, 2004.

[30] M. Sugimoto, S. Kikuchi, and M. Tomita, "Reverse engineering of biochemical equations from time-course data by means of genetic programming," *BioSystems*, vol. 80, no. 2, pp. 155–164, 2005.

[31] L. M. Tran, M. L. Rizk, and J. C. Liao, "Ensemble modeling of metabolic networks," *Biophysical Journal*, vol. 95, no. 12, pp. 5606–5617, 2008.

[32] E. O. Voit, *Computational Analysis of Biochemical Systems*, Cambridge University Press, Cambridge, UK, 2000.

[33] M. Rodriguez-Fernandez, P. Mendes, and J. R. Banga, "A hybrid approach for efficient and robust parameter estimation in biochemical pathways," *BioSystems*, vol. 83, no. 2-3, pp. 248–265, 2006.

[34] E. Walter and L. Pronzato, *Identification of Parameter Models from Experimental Data*, Springer, Berlin, Germany, 1997.

[35] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ, USA, 1999.

[36] W. E. Stewart, M. Caracotsios, and J. P. Sørensen, "Parameter estimation from multiresponse data," *AIChE Journal*, vol. 38, no. 5, pp. 641–650, 1992.

[37] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann, "Qualitative simulation of the carbon starvation response in *Escherichia coli*," *BioSystems*, vol. 84, no. 2, pp. 124–152, 2006.

[38] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, "Regulation of noise in the expression of a single gene," *Nature Genetics*, vol. 31, no. 1, pp. 69–73, 2002.

[39] J. Paulsson, "Models of stochastic gene expression," *Physics of Life Reviews*, vol. 2, no. 2, pp. 157–175, 2005.

[40] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.

[41] I. Swameye, T. G. Müller, J. Timmer, O. Sandra, and U. Klingmüller, "Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 3, pp. 1028–1033, 2003.

[42] J. Timmer, T. G. Müller, I. Swameye, O. Sandra, and U. Klingmüller, "Modeling the nonlinear dynamics of cellular

signal transduction," *International Journal of Bifurcation & Chaos in Applied Sciences & Engineering*, vol. 14, no. 6, pp. 2069–2079, 2004.

[43] S. Audoly, L. D'Angiò, M. P. Saccomani, and C. Cobelli, "Global identifiability of linear compartmental models—a computer algebra algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 36–47, 1998.

[44] E. Cinquemani, A. Milias-Argeitis, S. Summers, and J. Lygeros, "Stochastic dynamics of genetic networks: modelling and parameter identification," *Bioinformatics*, vol. 24, no. 23, pp. 2748–2754, 2008.

[45] P. Mendes and D. B. Kell, "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation," *Bioinformatics*, vol. 14, no. 10, pp. 869–883, 1998.

[46] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: a comparison of global optimization methods," *Genome Research*, vol. 13, no. 11, pp. 2467–2474, 2003.

[47] G. R. Wood, D. L. J. Alexander, and D. W. Bulger, "Approximation of the distribution of convergence times for stochastic global optimisation," *Journal of Global Optimization*, vol. 22, no. 1–4, pp. 271–284, 2002.

[48] E. Onbaşoğlu and L. Özdamar, "Parallel simulated annealing algorithms in global optimization," *Journal of Global Optimization*, vol. 19, no. 1, pp. 27–50, 2001.

[49] E. Balsa-Canto, M. Rodriguez-Fernandez, and J. R. Banga, "Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation," *Journal of Food Engineering*, vol. 82, no. 2, pp. 178–188, 2007.

[50] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, Calif, USA, 1979.

[51] K.-H. Cho, S.-Y. Shin, W. Kolch, and O. Wolkenhauer, "Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo method: a case study for the TNFα-mediated NF-κB signal transduction pathway," *Simulation*, vol. 79, no. 12, pp. 726–739, 2003.

[52] X.-J. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz, "Optimizing genetic circuits by global sensitivity analysis," *Biophysical Journal*, vol. 87, no. 4, pp. 2195–2202, 2004.

[53] J. D. Keasling, "Synthetic biology for synthetic chemistry," *ACS Chemical Biology*, vol. 3, no. 1, pp. 64–76, 2008.

[54] C. Tan, H. Song, J. Niemi, and L. You, "A synthetic biology challenge: making cells compute," *Molecular BioSystems*, vol. 3, no. 5, pp. 343–353, 2007.

[55] V. Chickarmane, B. N. Kholodenko, and H. M. Sauro, "Oscillatory dynamics arising from competitive inhibition and multisite phosphorylation," *Journal of Theoretical Biology*, vol. 244, no. 1, pp. 68–76, 2007.

[56] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[57] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, "Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 30, pp. 10955–10960, 2004.

[58] J. Hasty, M. Dolnik, V. Rottschäfer, and J. J. Collins, "Synthetic gene network for entraining and amplifying cellular oscillations," *Physical Review Letters*, vol. 88, no. 14, Article ID 148101, 4 pages, 2002.

[59] J. Noirel, S. Y. Ow, G. Sanguinetti, and P. C. Wright, "Systems biology meets synthetic biology: a case study of the metabolic effects of synthetic rewiring," *Molecular BioSystems*, vol. 5, no. 10, pp. 1214–1223, 2009.

[60] V. Hatzimanikatis, C. A. Floudas, and J. E. Bailey, "Analysis and design of metabolic reaction networks via mixed-integer linear optimization," *AIChE Journal*, vol. 42, no. 5, pp. 1277–1292, 1996.

[61] G. Sriram, D. B. Fulton, V. V. Iyer, et al., "Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing," *Plant Physiology*, vol. 136, no. 2, pp. 3043–3057, 2004.

[62] K. Nöh and W. Wiechert, "Experimental design principles for isotopically instationary ^{13}C labeling experiments," *Biotechnology and Bioengineering*, vol. 94, no. 2, pp. 234–251, 2006.

[63] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, "Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions," *Metabolic Engineering*, vol. 9, no. 1, pp. 68–86, 2007.

[64] W. Wiechert, M. Möllney, N. Isermann, M. Wurzel, and A. A. De Graaf, "Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems," *Biotechnology and Bioengineering*, vol. 66, no. 2, pp. 69–85, 1999.

[65] G. Sriram and J. V. Shanks, "Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping," *Metabolic Engineering*, vol. 6, no. 2, pp. 116–132, 2004.

[66] W. A. van Winden, J. J. Heijnen, and P. J. T. Verheijen, "Cumulative bondomers: a new concept in flux analysis from 2D [^{13}C,^{1}H] COSY NMR data," *Biotechnology and Bioengineering*, vol. 80, no. 7, pp. 731–745, 2002.

[67] J. D. Young, J. L. Walther, M. R. Antoniewicz, H. Yoo, and G. Stephanopoulos, "An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis," *Biotechnology and Bioengineering*, vol. 99, no. 3, pp. 686–699, 2008.

[68] J. M. Lee, E. P. Gianchandani, and J. A. Papin, "Flux balance analysis in the era of metabolomics," *Briefings in Bioinformatics*, vol. 7, no. 2, pp. 140–150, 2006.

[69] J. S. Edwards and B. O. Palsson, "Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions," *BMC Bioinformatics*, vol. 1, article 1, 2000.

[70] K. J. Kauffman, P. Prakash, and J. S. Edwards, "Advances in flux balance analysis," *Current Opinion in Biotechnology*, vol. 14, no. 5, pp. 491–496, 2003.

[71] O. Resendis-Antonio, J. L. Reed, S. Encarnación, J. Collado-Vides, and B. ∅. Palsson, "Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli," *PLoS Computational Biology*, vol. 3, no. 10, article e192, 2007.

[72] M. AbuOun, P. F. Suthers, G. I. Jones, et al., "Genome scale reconstruction of a salmonella metabolic model: comparison of similarity and differences with a commensal *Escherichia coli* strain," *Journal of Biological Chemistry*, vol. 284, no. 43, pp. 29480–29488, 2009.

[73] A. M. Feist, C. S. Henry, J. L. Reed, et al., "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information," *Molecular Systems Biology*, vol. 3, article 121, 2007.

[74] P. Geigenberger, M. Stitt, and A. R. Fernie, "Metabolic control analysis and regulation of the conversion of sucrose to starch

in growing potato tubers," *Plant, Cell & Environment*, vol. 27, no. 6, pp. 655–673, 2004.

[75] U. S. Ramli, J. J. Salas, P. A. Quant, and J. L. Harwood, "Metabolic control analysis reveals an important role for diacylglycerol acyltransferase in olive but not in oil palm lipid accumulation," *FEBS Journal*, vol. 272, no. 22, pp. 5764–5770, 2005.

[76] B. M. Bakker, H. V. Westerhoff, F. R. Opperdoes, and P. A. M. Michels, "Metabolic control analysis of glycolysis in trypanosomes as an approach to improve selectivity and effectiveness of drugs," *Molecular and Biochemical Parasitology*, vol. 106, no. 1, pp. 1–10, 2000.

[77] D. A. Fell, "Metabolic control analysis: a survey of its theoretical and experimental development," *Biochemical Journal*, vol. 286, no. 2, pp. 313–330, 1992.

[78] D. Fell, "Metabolic control analysis," in *Systems Biology*, L. Alberghina and H. Westerhoff, Eds., pp. 69–80, Springer, Berlin, Germany, 2005.

[79] C. C. Guet, M. B. Elowitz, W. Hsing, and S. Leibler, "Combinatorial synthesis of genetic networks," *Science*, vol. 296, no. 5572, pp. 1466–1470, 2002.

[80] J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury, "Network component analysis: reconstruction of regulatory signals in biological systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 26, pp. 15522–15527, 2003.

[81] Z. Li and C. Chan, "Extracting novel information from gene expression data," *Trends in Biotechnology*, vol. 22, no. 8, pp. 381–383, 2004.

[82] S. Raychaudhuri, J. M. Stuart, and R. B. Altman, "Principal components analysis to summarize microarray experiments: application to sporulation time series," in *Pacific Symposium on Biocomputing*, pp. 455–466, Island of Oahu, Hawaii, USA, 2000.

[83] M. K. S. Yeung, J. Tegnér, and J. J. Collins, "Reverse engineering gene networks using singular value decomposition and robust regression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 9, pp. 6163–6168, 2002.

[84] N. S. Holter, A. Maritan, M. Cieplak, N. V. Fedoroff, and J. R. Banavar, "Dynamic modeling of gene expression data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 4, pp. 1693–1698, 2001.

[85] W. Liebermeister, "Linear modes of gene expression determined by independent coponent analysis," *Bioinformatics*, vol. 18, no. 1, pp. 51–60, 2002.

[86] Z. Li, S. M. Shaw, M. J. Yedwabnick, and C. Chan, "Using a state-space model with hidden variables to infer transcription factor activities," *Bioinformatics*, vol. 22, no. 6, pp. 747–754, 2006.

[87] C. T. Harbison, D. B. Gordon, T. I. Lee, et al., "Transcriptional regulatory code of a eukaryotic genome," *Nature*, vol. 431, no. 7004, pp. 99–104, 2004.

[88] M. Hucka, A. Finney, H. M. Sauro, et al., "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models," *Bioinformatics*, vol. 19, no. 4, pp. 524–531, 2003.

[89] S. Bornholdt, "Less is more in modeling large genetic networks," *Science*, vol. 310, no. 5747, pp. 449–451, 2005.

[90] J. E. Bailey, "Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities," *Biotechnology Progress*, vol. 14, no. 1, pp. 8–20, 1998.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 541698, 9 pages
doi:10.1155/2010/541698

Review Article

Synthetic Biology Guides Biofuel Production

Michael R. Connor and Shota Atsumi

Department of Chemistry, University of California, Davis, CA 95616, USA

Correspondence should be addressed to Shota Atsumi, atsumi@chem.ucdavis.edu

Received 16 May 2010; Accepted 5 July 2010

Academic Editor: Patrick Cirino

The advancement of microbial processes for the production of renewable liquid fuels has increased with concerns about the current fuel economy. The development of advanced biofuels in particular has risen to address some of the shortcomings of ethanol. These advanced fuels have chemical properties similar to petroleum-based liquid fuels, thus removing the need for engine modification or infrastructure redesign. While the productivity and titers of each of these processes remains to be improved, progress in synthetic biology has provided tools to guide the engineering of these processes through present and future challenges.

1. Introduction

The desire for the discovery of renewable liquid fuels, as well as commodity chemicals, has escalated due to the environmental impact, supply security, and decreasing total reserve of petroleum-based fuels and chemicals [1]. Petroleum consumption reached 37.1 quadrillion BTU in the United States in 2008, of which a large majority (71%) was used as a liquid fuel in the transportation sector [2]. This has lead to an increased focus to find sustainable replacements or supplements to petroleum derived diesel fuel, jet fuel, and motor gasoline [3–5]. The largest effort thus far has been the production of ethanol, which is often used as a supplement to gasoline but is also available in high percentage blends such as E85. Ethanol production via fermentation reached 9.2 billion gallons in the United States in 2008, an increase of over 40% from 2007 [2]. According to the newest Renewable Fuels Standard (RFS2) set aside in 2010, the mandate for renewable fuels production is 36 billion gallons by 2022. These renewable fuels are classified into 4 categories: cellulosic biofuels, which must be derived from renewable lignocellulosics and achieve a 60% lifecycle greenhouse gas (GHG) emission reduction over their petroleum-derived counterparts; biomass-based diesel (50% GHG emission reduction); advanced biofuels, which consist of any renewable fuel other than corn ethanol that reduces GHG emissions by 50%; total renewable fuel, of which any fuel that achieves a 20% reduction of GHG emissions is counted [6]. Of the biomass-derived diesel fuels, biodiesel has gained momentum as a supplement or replacement to traditional petrodiesel, with production reaching just under 700 million gallons in 2008 [2]. Biodiesel can be made by several methods, but is most commonly synthesized by the transesterification of oils and fat triglycerides with methanol to make fatty acid methyl esters (FAME).

Although each of these fuel alternatives provides initial platforms for biofuel development, their increased commercialization to replace petroleum is not without its limitations. Ethanol is incompatible with the current fuel infrastructure, and the supply of raw materials for biodiesel production from plant oils and waste animal fats may become a concern. These opportunities for refinement have lead researchers to look for alternative fuels and production processes to replace petroleum derived fuels, including fermentative alcohols [7–12], nonfermentative higher chain alcohols [13], isoprenoid [14] and lipid fuels [15–18], and fuels synthesized directly from CO_2 via photosynthesis [19, 20]. These microbial-based processes are critical first steps in designing processes to provide renewable drop-in liquid fuels.

Aiding in the design and continued development of these processes, among a host of others, has been synthetic biology. Synthetic biology aims to design, synthesize, and characterize new biological elements, or redesign natural systems, that can be lumped together in a "toolbox." These elements

can include promoters [21–23], regulatory proteins and RNAs [24–28], and scaffolds [29, 30]. With this "toolbox," synthetic biologists assemble these individually characterized parts into hierarchal structures to perform new, novel, or nonnative tasks [31], such as synthetic oscillators [32] and toggle switches [33]. This "toolbox" also allows for the investigation of several designs to achieve the same function, often with varying levels of success, as in the case of heterologous 1-butanol production [7, 9, 10]. This differs from traditional engineering approaches in that the design focal point is on the core components, which can be fine-tuned to meet strict guidelines for specific tasks [34]. Much of the work accomplished in biofuel research until now has relied on the identification of target pathways and the design of synthetic expression systems for enzymes responsible for fuel production. As these technologies progress and mature, the design, implementation, and optimization of new functions, as well as the upgrading and rewiring of existing components, will be essential for the successful discovery and production of new biofuels, as many challenges still limit their productivity. This review will investigate recent progress made in the microbial production of biofuels to supplant petrodiesel and motor gasoline and will discuss how existing and newly developed synthetic biology tools may aid in the advancement of these processes.

2. Current Biofuels Research

2.1. Traditional Fermentative Processes. Ethanol, isopropanol, and 1-butanol are the only naturally produced alcohol biofuels. Isopropanol can be used directly as a fuel supplement to gasoline or as a feedstock for the transesterification of fats into biodiesel [35]. Both isopropanol and 1-butanol are produced in a mixed product fermentation in various strains of *Clostridium* [36], with maximum production levels reaching 2 g/L and 20 g/L, respectively [37, 38]. With a renewed interest in alternative fuels, the production of isopropanol and 1-butanol has been recently investigated in genetically tractable heterologous organisms. These organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*, facilitate the design and optimization of new biofuels processes by combining an increasing synthetic biology toolbox with a well-studied metabolism. Isopropanol production in *E. coli* has surpassed that of *Clostridium* by assembling the pathway for acetone production and a secondary alcohol dehydrogenase [8, 12]. The production of 1-butanol, however, has proven to be more difficult. Initial efforts were able to produce ~0.5 g/L using *E. coli* as a host [7]. Construction of a new strain harboring a single construct resulted in an increase in production to 1.2 g/L [9]. In addition to *E. coli*, 1-butanol production has been investigated in *Pseudomonas putida*, *Bacillus subtilis*, and *S. cerevisiae* [10, 11], although production in *E. coli* has thus far shown the most promise. Each of these processes, however, is far from industrial feasibility, as yields (~0.05 g/g) and productivities (~0.01 g/L/h) must increase significantly to match the same figures for corn ethanol (~0.5 g/g and 2 g/L/h). The advancement of these processes is thought to be limited by the low activity of pathway enzymes due to poor expression, solubility, or oxygen sensitivity, as well as the metabolic imbalance introduced by these heterologous pathways. While productivity in each of these platforms is low in comparison with Clostridial fermentation, the ability to engineer and manipulate these user-friendly hosts will facilitate the development of these processes.

2.2. Nonfermentative Higher Alcohols. The production of biofuels from native organisms can present unique challenges to synthetic biologists as oftentimes the availability of genetic tools and physiological knowledge of the hosts is limited. Additionally, the engineering of heterologous hosts for biofuel production may decrease the overall fitness of the cell and require delicate pathway balancing that is oftentimes difficult [7]. It is therefore advantageous to use native pathways to generate immediate precursors for biofuel production. This was accomplished in *E. coli* by using the hosts' amino acid biosynthesis pathways to generate 2-keto acid precursors, which can be converted to alcohols through a single heterologous reaction (Figure 1). These alcohols can serve as direct replacements to gasoline, or can be polymerized to form a variety of potential fuel molecules [39]. Expression of keto acid decarboxylase (KDC) from *Lactococcus lactis* enabled *E. coli* to convert these 2-keto acids to aldehydes, which can be reduced to alcohols using alcohol dehydrogenase (ADH) [13]. A total of 6 of these higher chain alcohols were detected after expression of KDC. The production of isobutanol, in particular, was able to surpass 20 g/L after host optimization and amplification of genes responsible for the synthesis of 2-ketoisovalerate, the precursor to both isobutanol and valine [13]. Subsequent efforts have also been made to engineer the production of 1-propanol and 1-butanol [40], 2-methyl-1-butanol [41], and 3-methyl-1-butanol [42].

An alternative and more direct route to 2-ketobutyrate, the citramalate pathway from *Methanococcus jannaschii*, was utilized to produce 1-butanol and 1-propanol independently of threonine. By leveraging a growth selection based on a requirement for 2-keto acids, the directed evolution of citramalate synthase was able to enhance the activity of this heterologous pathway and increase the production of 1-propanol and 1-butanol by 9-fold and 22-fold, respectively [43]. These pathways were also expanded to produce longer chain alcohols from nonstandard 2-keto acids. These longer chain alcohols can be used as commodity chemicals and also possess advantageous fuel properties similar to other high chain alcohols. The leucine biosynthesis pathway was engineered to catalyze the elongation of larger substrates by mutation of LeuA [44]. Similarly, KDC was also designed to fit larger substrates, after which *E. coli* was able to produce several longer chain alcohols from C5–C8 from these nonstandard 2-keto acids [44].

One distinct advantage in the production of these alcohols is the ability to apply existing amino acid production technology. Amino acids are produced microbially from several microorganisms, yet *Corynebacterium glutamicum* has been the most successful host for the production of many amino acids, including valine [45]. In addition to industrial success in amino acid production, *C. glutamicum*

shows an increased tolerance to isobutanol relative to *E. coli*, making the production of isobutanol from *C. glutamicum* promising. Optimization of the host by gene deletion and overexpression of isobutanol synthesis genes resulted in the production of 4.9 g/L of isobutanol from glucose [46]. Similarly, an amino acid strain development technique was adapted for the production of 3-methyl-1-butanol (3 MB), which shares a common precursor with leucine. Whole cell mutagenesis and selection with a leucine analogue resulted in a strain able to produce 2.8 g/L 3 MB, greater than a 5-fold improvement from wild-type (WT) [47]. Addition of an in situ extraction technique to remove 3 MB from the aqueous culture media resulted in the production of 9.5 g/L of 3 MB [47].

2.3. Isoprenoids. Isoprenoids represent a diverse group of hydrocarbons synthesized from the C5 isomers isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These precursors can be produced from acetyl coenzyme A (CoA) via the mevalonate pathway or from glyceraldehyde-3-phosphate and pyruvate through the methylerythritol pathway (MEP) (Figure 1). Isoprenoids are found naturally as hormones, photosynthetic pigments, and a variety of other specialized secondary metabolites, and have been previously investigated as nutraceuticals [48] and pharmaceuticals [49]. Because isoprenoids possess vast structural diversity, including saturated, unsaturated, branched, or cyclic alkenes or alkanes, their potential as fuel candidates, such as isopentenol (C5) for motor gasoline, or farnesene (C15) for diesel fuel, is promising [3].

Recently, the production of isopentenol (3-methyl-3-buten-1-ol or 3-methyl-2-buten-1-ol), a C5 unsaturated alcohol, has been investigated in *E. coli*. Isopentenol can be produced by the dephosphorylation of the isoprenoid building blocks IPP and DMAPP (Figure 1). By screening a *Bacillus subtilis* genomic library for relief of prenyl diphosphate (IPP and DMAPP) accumulation, two isopentenol biosynthetic genes, *nudF* and *yhfR*, were identified [14]. By overexpression of *nudF* in a previously optimized strain [50], production of isopentenol reached 110 mg/L [14]. Although the production of isoprenoid fuels is still limited, synthetic biology can provide the framework for improving isopentenol production as well as the development of processes for other potential isoprenoid fuels.

2.4. Lipids. The fatty acid biosynthesis pathways are of great importance to the production of renewable fuels. Fatty acids are commonly used now in the synthesis of biodiesel, and the production of long chain alkanes, alkenes, aldehydes, and alcohols offer promise in the development of alternative diesel and jet fuels. The fatty acid elongation cycle begins with the carboxylation of acetyl-CoA to malonyl-CoA. After transacylation of acetyl-CoA and malonyl-CoA to acyl carrier protein (ACP), acetyl-ACP and malonyl-ACP are condensed into acetoacetyl-ACP. After a reduction, dehydration, and second reduction reaction, a saturated fatty acyl-ACP (butyryl-ACP) is formed. In each subsequent elongation cycle, malonyl-ACP is condensed with the saturated fatty acyl-ACP to add 2 carbons to the growing hydrocarbon. After elongation, these hydrocarbons can be released as free fatty acids by a thioesterase. Additionally, fatty acyl-CoA can be reduced to a fatty aldehyde, which itself can be reduced to a fatty alcohol, or decarbonylated to an alkane or alkene [51] (Figure 1).

The microbial production of biodiesel has been approached from two angles First, by producing short chain alcohols and performing the transesterification *in vivo* with exogenously added fatty acids, and second, by producing free fatty acids that can be harvested for transesterification *in vitro*. The *in vivo* production of biodiesel using endogeneously produced ethanol was recently demonstrated in *E. coli*. Expression of the ethanol production pathway from *Zymomonas mobilis*, along with a broad substrate range acyltransferase (AtfA) from *Acinetobacter baylyi*, lead to the production of 1.3 g/L of fatty acid ethyl esters (FAEE) after addition of exogenous oleic acid [18].

Research on the production of fatty acids have centered around the discovery of oligeanous algae, yeast, and even bacteria, in which the lipid content, mainly composed of triacylglycerols (TAG), can reach 60%–80% of the total biomass produced [51, 52]. The identification of the genetic elements involved in fatty acid synthesis and the implementation and development of synthetic biology tools to facilitate strain development will become critical for process refinement [51]. Recently, efforts have also been made to produce fatty acids from user-friendly organisms such as *E. coli*. The overexpression of an endogenous and exogenous thioesterase along with acetyl-CoA carboxylase in a Δ*fadD* strain resulted in the production of 2.5 g/L of fatty acids from glycerol [16]. By redesigning the expression of these genes the production of fatty acids was increased to 4.5 g/L at a 6% yield with a specific productivity of 0.04 g/h/g dry cell weight [17].

Although each of these approaches has been successful in producing precursors for biodiesel synthesis, the supply of raw materials (lipid) or downstream processing (transesterification) can be cost-intensive [18]. It would be advantageous, therefore, to create a consolidated process to reduce costs. Consolidated bioprocessing (CBP), which involves the simultaneous production of saccharolytic enzymes with the hydrolysis of pretreated biomass and the fermentation of hexose and pentose sugars [53], significantly reduces processing costs by converting abundant, inexpensive biomass into useful fuels or chemicals in a single step. Unfortunately, no organisms possess both the ability to digest lignocellulosic biomass and ferment sugars to fuels at high yields. The solution to this challenge is being approached in two directions: the introduction of high yield fuel production pathways into cellulolytic organisms, and the engineering of substrate (lignocellulosic) utilizing pathways into organisms with superior product formation. Much of the work until this point has focused around the production of ethanol [53, 54], until recently when this strategy was applied by engineering *E. coli* for the production of biodiesel (FAEE), fatty alcohols, and wax esters [15]. *E. coli* was chosen as a host due to its high fatty acid synthesis rate (0.2 g/L/h/g dry cell mass [51]) and straightforwardness in genetic manipulation. As in previous

studies, the ethanol production pathway from *Z. mobilis* (*pdc*, *adhB*) was overexpressed to produce ethanol for FAEE production. By combining this pathway with a cytosolic version of an endogenous thioesterase (‘*tesA*) and an ester synthase from *A. baylyi* (*atfA*), a fatty acid oxidation deficient strain of *E. coli* (Δ*fadE*) was able to produce 37 mg/L of FAEE directly from glucose [15]. To increase the production of FAEE, two CoA ligases, *fadD* from *E. coli* and *FAA2* from *S. cerevisiae*, were overexpressed along with another copy of *atfA* to bring production of FAEE up to 674 mg/L [15]. In order to mitigate the cost of processing of raw cellulosic biomass into refined sugars, this process was engineered to use xylan, a pentose polysaccharide component of hemicellulose. Expression of the endoxylanase *xyn10B* from *Clostridium stercorarium* and the xylanase *xsa* from *Bacteroides ovatus* as chimeras with OsmY allowed *E. coli* to grow on xylan as a sole carbon source [15]. Assembly of the xylan degradation pathway with the previously described FAEE production strain resulted in a strain able to produce 12 mg/L of FAEE from xylan [15]. Future work may focus on the development of secreted cellulases to increase the substrate utilization capacity of *E. coli*, in addition to optimizing the fatty acid pathway by prospecting for enzymes or expression systems with increased activity or stability. This work demonstrates the first consolidated process for the production of fatty acid based fuels and chemicals from complex polysaccharides, and while the process yields and productivity remained to be improved to merit commercialization, this work gives engineers and synthetic biologists the foundation to advance this process.

2.5. Direct Incorporation of CO_2. The role of photosynthesis in any biofuel production process is critical. Many current technologies, such as biomass derived biofuels and algal lipids, have received attention as viable fuel replacement technologies [55, 56], yet rely on intermediate stages to incorporate CO_2 or recover biomass to process precursors into useable fuels, which can increase costs. Photosynthetic organisms such as cyanobacteria, algae, and plants use light energy to generate reducing power to directly incorporate CO_2 into organic metabolites. The use of these organisms to directly produce fuels can limit production costs and CO_2 emissions during intermediate processing, and may also help reduce net CO_2 emissions by scrubbing CO_2 enriched flue gases from traditional power plants and producing useful fuels or chemicals, although their potential is not limited to this scenario. Initial efforts have focused on the production of ethanol in *Rhodobacter* [57] and *Synechococcus* [58].

The production of advanced biofuels such as isobutyraldehyde, isobutanol, and isoprene has recently been investigated in cyanobacteria. This was first demonstrated by transferring the 2-keto acid pathways to higher chain alcohols into the cyanobacterium *Synechococcus elongatus* (Figure 2). Isobutyraldehyde was chosen as an initial target since its boiling point (63°C) allows it to be easily stripped from the culture medium, thus avoiding any toxicity effects. Chromosomal integration of 2-ketoisovalerate biosynthesis genes (*alsS*, *ilvCD*) and 2-ketoisovalerate decarboxylase (*kivd*) resulted in the production of 723 mg/L of isobutyraldehyde from dissolved CO_2 ($NaHCO_3$) [19]. To improve the low activity of ribulose-1,5-bisphophate carboxylase/oxygenase (RubisCO), the *rbcLS* genes from a similar cyanobacterium were integrated downstream of the endogenous *rbcLS* genes. Production of isobutyraldehyde in this strain was elevated to 1.1 g/L, with a productivity of 6.2 mg/L/h [19], an encouraging figure considering microalgal biodiesel has been estimated to be near 4 mg/L/h [19]. The production of isobutanol was also investigated by expressing the $NADP^+$ dependent alcohol dehydrogenase YqhD from *E. coli*. The production of isobutanol reached 450 mg/L, and although encouraging, is currently thought to be limited by end product toxicity [19].

A similar study was also recently conducted in which the isoprenoid biosynthesis pathways were exploited for biofuel production in the cyanobacterium *Synechocystis* sp. PCC6803 (Figure 2). The volatile hydrocarbon isoprene, most notably produced in plants [59], is a potential feedstock for biofuel or chemical production. The *ispS* gene from kudzu vine (*Pueraria montana*), encoding for an isoprene synthase, which catalyzes the conversion of DMAPP to isoprene, was codon optimized and cloned into *Synechocystis* under the control of a light dependent promoter. Expression of *ispS* under high intensity light resulted in a small accumulation of isoprene (50 μg/day/g dry cell mass) [20] relative to healthy oak leaves (~1,650 μg/day/g dry cell mass) [60], demonstrating that while successful, this process remains to be improved.

A significant obstacle to biofuel production in photosynthetic organisms is the design of scale up processes. Photosynthetic organisms require more complex reactor designs and their growth and productivity in suboptimal conditions (temperature, salt concentration, etc.) is not well understood. However, two reactor designs in use today, the raceway pond and the photobioreactor, have shown promise in their ability to accumulate photosynthetic biomass using sunlight [61]. The raceway pond is inexpensive and simple to construct, but it is subject to contamination and is less photosynthetically efficient than a photobioreactor. Photobioreactors, having several designs [62], have increased capital costs compared to raceway ponds, but have superior productivities due to their increased biomass concentration, and therefore, greater photosynthetic efficiency. Hybrid systems comprised of both raceway ponds and photobioreactors have also been investigated to maximize the advantages of each design [62].

3. Synthetic Biology for Biofuels

Synthetic biology is an increasingly expanding discipline focusing on the design and construction of artificial systems to achieve a desired goal. These systems are derived from the assembly of standardized components in a hierarchal manner to create a population of programmed cells carrying out a desired function [31]. For biofuels, this is of particular interest as the production of these chemicals requires efficient integration of foreign genes and pathways into central metabolism. Delicate optimization and fine-tuning of these processes to maximize productivity and

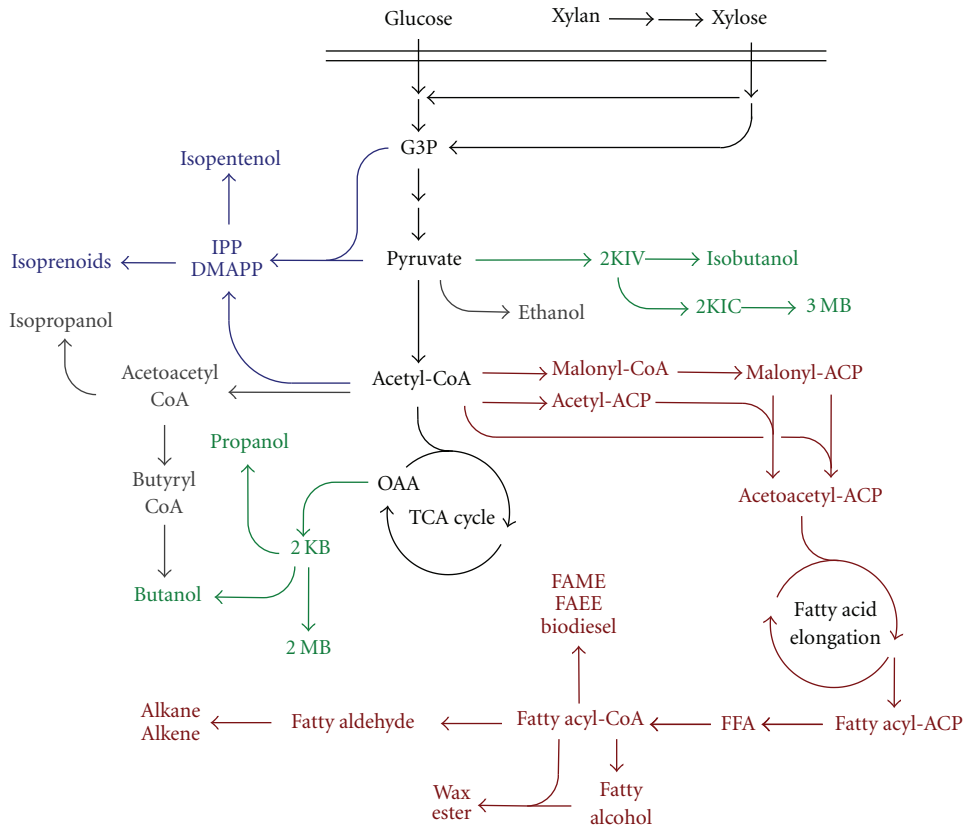


Figure 1: Metabolic schematic for biofuel production. Biofuel production pathways for traditional fermentative processes (grey), nonfermentative higher chain alcohols (green), isoprenoid fuels (blue), and fatty acid fuels (red) from central metabolism. Abbreviations: 2 KB (2-ketobutyrate), 2 KIV (2-ketoisovalerate), 2KIC (2-ketoisocaproate), 2 MB (2-methyl-1-butanol), 3 MB (3-methyl-1-butanol), ACP (acyl carrier protein), CoA (coenzyme A), DMAPP (dimethylallyl diphosphate), FAEE (fatty acid ethyl ester), FAME (fatty acid methyl ester), FFA (free fatty acid), G3P (glyceraldehyde-3-phosphate), IPP (isopentyl diphosphate), and OAA (oxaloacetate).

yield is of equal concern as the viability of any biofuel processes is extremely sensitive to production costs [3], such as raw material supply, total production, and downstream processing. Synthetic biology can provide tools and design principles to guide the development of such processes (Figure 3).

A goal of synthetic biology is to create a library of biological parts that can be used independently or as part of a larger assembly for a higher function. These biological parts can have simple or complex behaviors relayed through a variety of outputs such as gene expression. A simple but powerful example is the design of synthetically regulated promoters [21–23] to accurately control gene expression. More complex examples include the modulation of the expression of multiple genes through tunable intergenic regions (TIGR) [27], the activation and silencing of gene expression by riboregulators and ribozymes [24, 26, 28], or successive increases in gene expression through chromosomal amplification [63], which will become particularly important in the design of stable strains for industrial biofuel production. Additionally, thermodynamic models have been developed to rationally design ribosome binding sites to achieve robust expression levels varying by as much as 100,000 fold [64]. Preliminary designs for these biofuel gene expression systems will need to evolve to regulate and fine-tune the gene expression of these pathways, as was previously discovered for isoprenoids, a pathway discussed earlier for biofuel production [65]. Synthetic biology has also achieved the adaptation of posttranslation systems to regulate enzyme activity such as allosteric protein gates [25] and synthetic scaffolds [30].

The balance between the metabolic capacity of biofuel production pathways and host fitness will also play a key role in the productivity and yield of the process. This balancing of biofuel pathways, which aims to maximize the flow of carbon toward product formation without drastically altering the metabolic load or intracellular cofactors (NAD(P)H, ATP, ACP, etc.) can be achieved through the use of these synthetic biology tools. This was again demonstrated in the mevalonate pathway to isoprenoids by employing a synthetic scaffold to increase the metabolic capacity while limiting protein expression [29]. In addition to the balancing of biofuel production pathways, the expression of multienzyme complexes, which often require delicate balancing of catalytic subunits, is of great importance. The most direct example is the production of complexed cellulases, which will be crucial to increase the substrate utilization capacity of biomass-based biofuel processes.

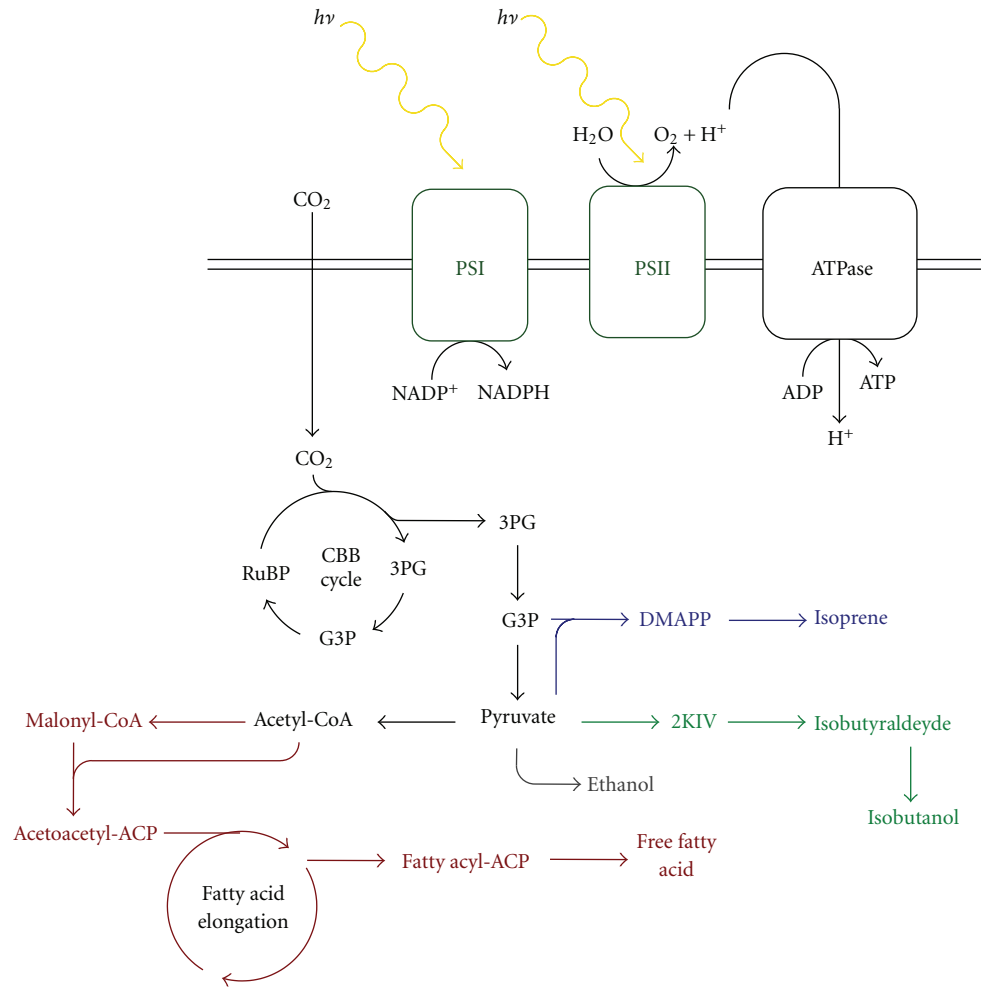


Figure 2: Metabolic schematic for direct photosynthetic biofuel production. Biofuel production pathways for traditional fermentative processes (grey), nonfermentative higher chain alcohols (green), isoprenoid fuels (blue), and fatty acid fuels (red) from CO_2 through photosynthesis. Abbreviations: 2 KIV (2-ketoisovalerate), 3PG (3-phosphoglycerate), ACP (acyl carrier protein), CoA (coenzyme A), DMAPP (dimethylallyl diphosphate), G3P (glyceraldehyde-3-phosphate), PSI (photosystem I), PSII (photosystem II), and RuBP (ribulose-1,5-bisphosphate).

The regulation of flux through divergent or branched pathways will also be critical, as the balance of carbon and electronic cofactors such as NADH must be considered to achieve an efficient process. These divergent pathways are especially common in the nonfermentative higher chain alcohol and isoprenoid pathways. To direct carbon flow in the desired manner, scaffolds can be engineered to connect the preferred branches of the pathway together. Regulatory RNA molecules can also be employed to minimize the expression of competing but essential pathways. Product yields may also be dramatically affected by the availability of NADH or NADPH. One possible solution is to engineer pathways with alternative cofactor specificity to increase their availability [66] or interconversion [67].

Synthetic biology will also be key in rewiring existing regulation. The design of new regulatory pathways from synthetic genetic elements will be important in sensing the extracellular or intracellular environment and producing a programmed cellular response. For biofuel production from lignocellulosic biomass, this is of great importance as the efficient uptake of a mixture of hexose and pentose sugars simultaneously is desirable. Synthetic biology can provide tools to construct new circuits devoid of unwanted regulation from the bottom up [68] to sense the extracellular environment and produce the necessary response to digest the sugars. Ultimately, the bottom-up construction of biological circuits may extend beyond individual pathways toward all of metabolism, as the synthesis of entirely synthetic, replicable, and functional genomes has recently been accomplished [69]. In the future, this will allow synthetic biologists to build heterologous pathways into organisms devoid of unwanted pathways or properties, therefore, increasing the selectivity and yield of the process.

The development and optimization of many aspects of biofuel production technology can benefit from the work already accomplished through synthetic biology. As the

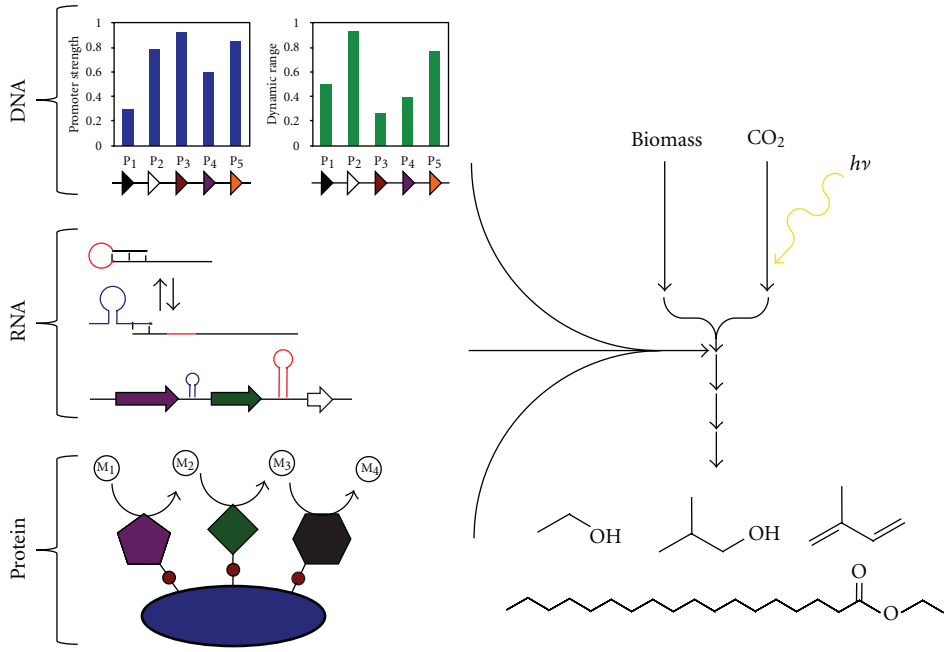


FIGURE 3: Synthetic biology for biofuels. Synthetic biology provides tools at the DNA, RNA, and protein levels that can guide the development of biofuel production processes.

number of available tools increases, the standardization of these parts will become increasingly important. One of the current challenges facing synthetic biology is the reproducibility of the developed tools in different systems, as the variability in cellular regulation and physiology can vary greatly from host to host. Proposals for standardization have been made [70], although as the complexity of these systems increase so will their variability. However, current technology developed by synthetic biology has already allowed for the successful design of many heterologously expressed biofuel production and substrate utilization pathways extending beyond the most user-friendly organisms. The advancement of synthetic biology toward new diagnostic tools and high throughput screening systems will aid in the further development of these biofuel processes for pathway optimization and enzyme discovery or improvement. The success that synthetic biology has already afforded biofuel production technology lends confidence to future synergistic developments and breakthroughs.

4. Conclusions

The desire for renewable liquid fuel replacements to petroleum has steadily increased with concerns about the current fuel economy's stability and environmental impact. The development of new biofuel production processes has sought to mitigate some of these issues. These fuels, whether designed for motor gasoline, diesel fuel, or jet fuel, will face challenges in strain development and productivity in a cost-sensitive market. The integration of synthetic biology with the development of these processes will be significant in bringing the biofuels industry from its infancy to a commercially viable alternative to petroleum.

References

[1] G. Stephanopoulos, "Challenges in engineering microbes for biofuels production," *Science*, vol. 315, no. 5813, pp. 801–804, 2007.

[2] K. E. Serferlein, *Annual Energy Review 2008*, E.I. Administration, 2009.

[3] S. K. Lee, H. Chou, T. S. Ham, T. S. Lee, and J. D. Keasling, "Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels," *Current Opinion in Biotechnology*, vol. 19, no. 6, pp. 556–563, 2008.

[4] P. P. Peralta-Yahya and J. D. Keasling, "Advanced biofuel production in microbes," *Biotechnology Journal*, vol. 5, no. 2, pp. 147–162, 2010.

[5] S. Atsumi and J. C. Liao, "Metabolic engineering for advanced biofuels production from Escherichia coli," *Current Opinion in Biotechnology*, vol. 19, no. 5, pp. 414–419, 2008.

[6] *EPA Finalizes Regulations for the National Renewable Fuel Standard Program for 2010 and Beyond*, O.o.T.a.A. Quality, 2010.

[7] S. Atsumi, A. F. Cann, M. R. Connor et al., "Metabolic engineering of Escherichia coli for 1-butanol production," *Metabolic Engineering*, vol. 10, no. 6, pp. 305–311, 2008.

[8] T. Hanai, S. Atsumi, and J. C. Liao, "Engineered synthetic pathway for isopropanol production in Escherichia coli," *Applied and Environmental Microbiology*, vol. 73, no. 24, pp. 7814–7818, 2007.

[9] M. Inui, M. Suda, S. Kimura et al., "Expression of Clostridium acetobutylicum butanol synthetic genes in Escherichia coli,"

Applied Microbiology and Biotechnology, vol. 77, no. 6, pp. 1305–1316, 2008.

[10] D. R. Nielsen, E. Leonard, S.-H. Yoon, H.-C. Tseng, C. Yuan, and K. L. J. Prather, "Engineering alternative butanol production platforms in heterologous bacteria," *Metabolic Engineering*, vol. 11, no. 4-5, pp. 262–273, 2009.

[11] E. J. Steen, R. Chan, N. Prasad et al., "Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol," *Microbial Cell Factories*, vol. 7, article 36, 2008.

[12] T. Jojima, M. Inui, and H. Yukawa, "Production of isopropanol by metabolically engineered Escherichia coli," *Applied Microbiology and Biotechnology*, vol. 77, no. 6, pp. 1219–1224, 2008.

[13] S. Atsumi, T. Hanai, and J. C. Liao, "Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels," *Nature*, vol. 451, no. 7174, pp. 86–89, 2008.

[14] S. T. Withers, S. S. Gottlieb, B. Lieu, J. D. Newman, and J. D. Keasling, "Identification of isopentenol biosynthetic genes from Bacillus subtilis by a screening method based on isoprenoid precursor toxicity," *Applied and Environmental Microbiology*, vol. 73, no. 19, pp. 6277–6283, 2007.

[15] E. J. Steen, Y. Kang, G. Bokinsky et al., "Microbial production of fatty-acid-derived fuels and chemicals from plant biomass," *Nature*, vol. 463, no. 7280, pp. 559–562, 2010.

[16] X. Lu, H. Vora, and C. Khosla, "Overproduction of free fatty acids in E. coli: implications for biodiesel production," *Metabolic Engineering*, vol. 10, no. 6, pp. 333–339, 2008.

[17] T. Liu, H. Vora, and C. Khosla, "Quantitative analysis and engineering of fatty acid biosynthesis in E. coli," *Metabolic Engineering*, vol. 12, no. 4, pp. 378–386, 2010.

[18] R. Kalscheuer, T. Stölting, and A. Steinbüchel, "Microdiesel: Escherichia coli engineered for fuel production," *Microbiology*, vol. 152, part 9, pp. 2529–2536, 2006.

[19] S. Atsumi, W. Higashide, and J. C. Liao, "Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde," *Nature Biotechnology*, vol. 27, no. 12, pp. 1177–1180, 2009.

[20] P. Lindberg, S. Park, and A. Melis, "Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism," *Metabolic Engineering*, vol. 12, no. 1, pp. 70–79, 2010.

[21] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos, "Tuning genetic control through promoter engineering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 36, pp. 12678–12683, 2005.

[22] U. Baron, M. Gossen, and H. Bujard, "Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential," *Nucleic Acids Research*, vol. 25, no. 14, pp. 2723–2729, 1997.

[23] R. Lutz and H. Bujard, "Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements," *Nucleic Acids Research*, vol. 25, no. 6, pp. 1203–1210, 1997.

[24] T. S. Bayer and C. D. Smolke, "Programmable ligand-controlled riboregulators of eukaryotic gene expression," *Nature Biotechnology*, vol. 23, no. 3, pp. 337–343, 2005.

[25] J. E. Dueber, B. J. Yeh, K. Chak, and W. A. Lim, "Reprogramming control of an allosteric signaling switch through modular recombination," *Science*, vol. 301, no. 5641, pp. 1904–1908, 2003.

[26] F. J. Isaacs, D. J. Dwyer, C. Ding, D. D. Pervouchine, C. R. Cantor, and J. J. Collins, "Engineered riboregulators enable post-transcriptional control of gene expression," *Nature Biotechnology*, vol. 22, no. 7, pp. 841–847, 2004.

[27] B. F. Pfleger, D. J. Pitera, C. D. Smolke, and J. D. Keasling, "Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes," *Nature Biotechnology*, vol. 24, no. 8, pp. 1027–1032, 2006.

[28] M. N. Win and C. D. Smolke, "A modular and extensible RNA-based gene-regulatory platform for engineering cellular function," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 36, pp. 14283–14288, 2007.

[29] J. E. Dueber, G. C. Wu, G. R. Malmirchegini et al., "Synthetic protein scaffolds provide modular control over metabolic flux," *Nature Biotechnology*, vol. 27, no. 8, pp. 753–759, 2009.

[30] S.-H. Park, A. Zarrinpar, and W. A. Lim, "Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms," *Science*, vol. 299, no. 5609, pp. 1061–1064, 2003.

[31] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline," *Molecular Systems Biology*, vol. 2, Article ID 2006.0028, 2006.

[32] M. B. Elowitz and S. Leibier, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.

[33] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[34] J. D. Keasling, "Synthetic biology for synthetic chemistry," *ACS Chemical Biology*, vol. 3, no. 1, pp. 64–76, 2008.

[35] I. Lee, L. A. Johnson, and E. G. Hammond, "Use of branched-chain esters to reduce the crystallization temperature of biodiesel," *Journal of the American Oil Chemists' Society*, vol. 72, no. 10, pp. 1155–1160, 1995.

[36] D. T. Jones and D. R. Woods, "Acetone-butanol fermentation revisited," *Microbiological Reviews*, vol. 50, no. 4, pp. 484–524, 1986.

[37] C.-K. Chen and H. P. Blaschek, "Acetate enhances solvent production and prevents degeneration in Clostridium beijerinckii BA101," *Applied Microbiology and Biotechnology*, vol. 52, no. 2, pp. 170–173, 1999.

[38] J.-S. Chen and S. F. Hiu, "Acetone-butanol-isopropanol production by Clostridium beijerinckii (synonym, Clostridium butylicum)," *Biotechnology Letters*, vol. 8, no. 5, pp. 371–376, 1986.

[39] M. R. Connor and J. C. Liao, "Microbial production of advanced transportation fuels in non-natural hosts," *Current Opinion in Biotechnology*, vol. 20, no. 3, pp. 307–315, 2009.

[40] C. R. Shen and J. C. Liao, "Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways," *Metabolic Engineering*, vol. 10, no. 6, pp. 312–320, 2008.

[41] A. F. Cann and J. C. Liao, "Production of 2-methyl-1-butanol in engineered Escherichia coli," *Applied Microbiology and Biotechnology*, vol. 81, no. 1, pp. 89–98, 2008.

[42] M. R. Connor and J. C. Liao, "Engineering of an Escherichia coli strain for the production of 3-methyl-1-butanol," *Applied and Environmental Microbiology*, vol. 74, no. 18, pp. 5769–5775, 2008.

[43] S. Atsumi and J. C. Liao, "Directed evolution of Methanococcus jannaschii citramalate synthase for biosynthesis of 1-propanol and 1-butanol by Escherichia coli," *Applied and Environmental Microbiology*, vol. 74, no. 24, pp. 7802–7808, 2008.

[44] K. Zhang, M. R. Sawaya, D. S. Eisenberg, and J. C. Liao, "Expanding metabolism for biosynthesis of nonnatural alcohols," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 52, pp. 20653–20658, 2008.

[45] W. Leuchtenberger, K. Huthmacher, and K. Drauz, "Biotechnological production of amino acids and derivatives: current status and prospects," *Applied Microbiology and Biotechnology*, vol. 69, no. 1, pp. 1–8, 2005.

[46] K. M. Smith, K.-M. Cho, and J. C. Liao, "Engineering Corynebacterium glutamicum for isobutanol production," *Applied Microbiology and Biotechnology*, vol. 87, no. 3, pp. 1045–1055, 2010.

[47] M. R. Connor, A. F. Cann, and J. C. Liao, "3-methyl-1-butanol production in Escherichia coli: random mutagenesis and two-phase fermentation," *Applied Microbiology and Biotechnology*, vol. 86, no. 4, pp. 1155–1164, 2010.

[48] D.-K. Ro, E. M. Paradise, M. Quellet et al., "Production of the antimalarial drug precursor artemisinic acid in engineered yeast," *Nature*, vol. 440, no. 7086, pp. 940–943, 2006.

[49] W. R. Farmer and J. C. Liao, "Improving lycopene production in Escherichia coli by engineering metabolic control," *Nature Biotechnology*, vol. 18, no. 5, pp. 533–537, 2000.

[50] V. J. J. Martin, D. J. Piteral, S. T. Withers, J. D. Newman, and J. D. Keasling, "Engineering a mevalonate pathway in Escherichia coli for production of terpenoids," *Nature Biotechnology*, vol. 21, no. 7, pp. 796–802, 2003.

[51] M. A. Rude and A. Schirmer, "New microbial fuels: a biotech perspective," *Current Opinion in Microbiology*, vol. 12, no. 3, pp. 274–281, 2009.

[52] Q. Li, W. Du, and D. Liu, "Perspectives of microbial oils for biodiesel production," *Applied Microbiology and Biotechnology*, vol. 80, no. 5, pp. 749–756, 2008.

[53] L. R. Lynd, W. H. Van Zyl, J. E. McBride, and M. Laser, "Consolidated bioprocessing of cellulosic biomass: an update," *Current Opinion in Biotechnology*, vol. 16, no. 5, pp. 577–583, 2005.

[54] W. H. Van Zyl, L. R. Lynd, R. Den Haan, and J. E. McBride, "Consolidated bioprocessing for bioethanol production using Saccharomyces cerevisiae," *Advances in Biochemical Engineering/Biotechnology*, vol. 108, pp. 205–235, 2007.

[55] Y. Chisti, "Biodiesel from microalgae beats bioethanol," *Trends in Biotechnology*, vol. 26, no. 3, pp. 126–131, 2008.

[56] A. J. Ragauskas, C. K. Williams, B. H. Davison et al., "The path forward for biofuels and biomaterials," *Science*, vol. 311, no. 5760, pp. 484–489, 2006.

[57] T. M. Wahlund, T. Conway, and F. R. Tabita, "Bioconversion of CO_2 to ethanol and other compounds," *American Chemical Society Division of Fuel Chemistry*, vol. 41, pp. 1403–1405, 1996.

[58] M.-D. Deng and J. R. Coleman, "Ethanol synthesis by genetic engineering in cyanobacteria," *Applied and Environmental Microbiology*, vol. 65, no. 2, pp. 523–528, 1999.

[59] T. D. Sharkey and S. Yeh, "Isoprene emission from plants," *Annual Review of Plant Biology*, vol. 52, pp. 407–436, 2001.

[60] L. J. Anderson, P. C. Harley, R. K. Monson, and R. B. Jackson, "Reduction of isoprene emissions from live oak (Quercus fusiformis) with oak wilt," *Tree Physiology*, vol. 20, no. 17, pp. 1199–1203, 2000.

[61] Y. Chisti, "Biodiesel from microalgae," *Biotechnology Advances*, vol. 25, no. 3, pp. 294–306, 2007.

[62] P. M. Schenk, et al., "Second generation biofuels: high efficiency microalgae for biodiesel production," *Bioenergy Research*, vol. 1, pp. 20–43, 2008.

[63] K. E. J. Tyo, P. K. Ajikumar, and G. Stephanopoulos, "Stabilized gene duplication enables long-term selection-free heterologous pathway expression," *Nature Biotechnology*, vol. 27, no. 8, pp. 760–765, 2009.

[64] H. M. Salis, E. A. Mirsky, and C. A. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology*, vol. 27, no. 10, pp. 946–950, 2009.

[65] D. J. Pitera, C. J. Paddon, J. D. Newman, and J. D. Keasling, "Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli," *Metabolic Engineering*, vol. 9, no. 2, pp. 193–207, 2007.

[66] I. Martínez, J. Zhu, H. Lin, G. N. Bennett, and K.-Y. San, "Replacing Escherichia coli NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from Clostridium acetobutylicum facilitates NADPH dependent pathways," *Metabolic Engineering*, vol. 10, no. 6, pp. 352–359, 2008.

[67] A. M. Sánchez, J. Andrews, I. Hussein, G. N. Bennett, and K.-Y. San, "Effect of overexpression of a soluble pyridine nucleotide transhydrogenase (UdhA) on the production of poly(3-hydroxybutyrate) in Escherichia coli," *Biotechnology Progress*, vol. 22, no. 2, pp. 420–425, 2006.

[68] N. J. Guido, X. Wang, D. Adalsteinsson et al., "A bottom-up approach to gene regulation," *Nature*, vol. 439, no. 7078, pp. 856–860, 2006.

[69] D. G. Gibson, J. I. Glass, C. Lartigue et al., "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science*, vol. 329, no. 5987, pp. 52–56, 2010.

[70] B. Canton, A. Labno, and D. Endy, "Refinement and standardization of synthetic biological parts and devices," *Nature Biotechnology*, vol. 26, no. 7, pp. 787–793, 2008.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 761042, 18 pages
doi:10.1155/2010/761042

Review Article

Metabolic Engineering for Production of Biorenewable Fuels and Chemicals: Contributions of Synthetic Biology

Laura R. Jarboe,[1] Xueli Zhang,[2] Xuan Wang,[2] Jonathan C. Moore,[2] K. T. Shanmugam,[2] and Lonnie O. Ingram[2]

[1] *Department of Chemical and Biological Engineering, Iowa State University, 3051 Sweeney Hall, Ames, IA 50011, USA*
[2] *Department of Microbiology and Cell Science, University of Florida, P.O. Box 110700, Gainesville, FL 32611, USA*

Correspondence should be addressed to Lonnie O. Ingram, ingram@ufl.edu

Received 14 August 2009; Revised 18 December 2009; Accepted 13 January 2010

Academic Editor: Patrick Cirino

Production of fuels and chemicals through microbial fermentation of plant material is a desirable alternative to petrochemical-based production. Fermentative production of biorenewable fuels and chemicals requires the engineering of biocatalysts that can quickly and efficiently convert sugars to target products at a cost that is competitive with existing petrochemical-based processes. It is also important that biocatalysts be robust to extreme fermentation conditions, biomass-derived inhibitors, and their target products. Traditional metabolic engineering has made great advances in this area, but synthetic biology has contributed and will continue to contribute to this field, particularly with next-generation biofuels. This work reviews the use of metabolic engineering and synthetic biology in biocatalyst engineering for biorenewable fuels and chemicals production, such as ethanol, butanol, acetate, lactate, succinate, alanine, and xylitol. We also examine the existing challenges in this area and discuss strategies for improving biocatalyst tolerance to chemical inhibitors.

1. Introduction

Human society has always depended on biomass-derived carbon and energy for nutrition and survival. In recent history, we have also become dependent on petroleum-derived carbon and energy for commodity chemicals and fuels. However, the nonrenewable nature of petroleum stands in stark contrast to the renewable carbon and energy present in biomass, where biomass is essentially a temporary storage unit for atmospheric carbon and sunlight-derived energy. Thus there is increasing demand to develop and implement strategies for production of commodity chemicals and fuels from biomass instead of petroleum. Specifically, in this work we are interested in the microbial fermentation of biomass-derived sugars to commodity fuels and chemicals.

In order for a fermentation process to compete with existing petroleum-based processes, the target chemical must be produced at a high yield, titer and productivity. Sometimes there are additional constraints on the fermentation process, such as the presence of potent inhibitors in biomass hydrolysate or the need to operate at an extreme pH or temperature [1]. These goals can be difficult to attain with naturally-occurring microbes. Therefore, microorganisms with these desired traits often must be developed, either by modification of existing microbes or by the *de novo* design of new microbes. While significant progress has been made towards *de novo* design [2, 3], this work focuses on the modification of existing microbes.

Humanity has long relied on microbial biocatalysts for production of fermented food and beverages and eukaryotic biocatalysts for food and textiles. We have slowly modified these biocatalysts by selecting for desirable traits without understanding the underlying biological mechanisms. But upon elucidation of the biological code and the development of recombinant DNA technology, we now have the tools to do more than just select for observable traits—we are now able to rationally modify and design metabolic pathways, proteins, and even whole organisms.

Much of this rational modification has been in the form of Metabolic Engineering. Metabolic Engineering was

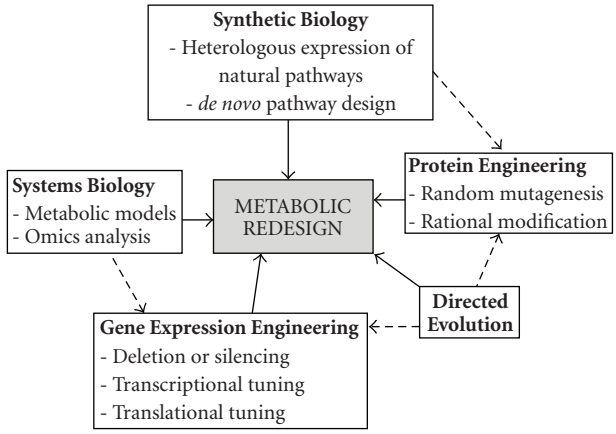


Figure 1: Overview of tools for metabolic redesign.

defined in 1991 [4, 5] and here we use the definition of "*the directed improvement of production, formation, or cellular properties through the modification of specific biochemical reactions or the introduction of new ones with the use of recombinant DNA technology* " [6]. While Metabolic Engineering has enabled extraordinary advances in the production of commodity chemicals and fuels from biomass, some of which are discussed in this work, we have now reached the point where biological functions that do not exist in nature are desired. Synthetic biology aims to develop and provide these nonnatural biological functions.

For many years, the term Synthetic Biology was used to describe concepts that would be classified today as Metabolic Engineering [7]. However in the last 10 years, terms such as "unnatural organic molecules" [7], "unnatural chemical systems [8], "novel behaviors" [9], "artificial, biology-inspired systems" [10], and "functions that do not exist in nature" [11] have been used to describe Synthetic Biology. For the purpose of this review, we will apply the Synthetic Biology definition of "*the design and construction of new biological components, such as enzymes, genetic circuits, and cells, or the redesign of existing biological systems*" [12].

Synthetic biology has application to many fields, including cell-free synthesis [13], tissue and plant engineering [14] and drug discovery [15], but here we are interested in the modification of microbes for the biorenewable production of commodity chemicals and fuels. Other recent reviews have also dealt with this topic [16–18].

Synthetic biology for the production of a target compound can be expressed as a sequence of the following events, each of which will be discussed in more detail and demonstrated below. (1) Design the metabolic pathways and phenotypic properties of the desired system. What are the desired substrates and products? What are the expected environmental stressors? (2) Choose an appropriate host organism (chassis) based on the following criteria. Which organisms display at least some of the desired properties? How well characterized and annotated are these organisms? Are there molecular biology tools for modification of this chassis? (3) Formulate an implementation approach. What modifications are necessary to achieve the pathways and properties identified in step (1)? Do metabolic pathways need to be added, removed, or tuned? Does the desired pathway or phenotype exist in nature, or does it need to be designed *de novo*? (4) Optimize the redesigned system and assess the system properties relative to the ideal. Can the chassis be improved further?

Even a simple biocatalyst, such as the laboratory workhorse *Escherichia coli*, is a complex system of an estimated 4603 genes, 2077 reactions, and 1039 unique metabolites [19, 20], and while the steps outlined above are relatively straightforward, it is still difficult to quickly and reliably engineer a biocatalyst to perform desired behaviors [21]. Systems biology, the standardization of biological systems, and metabolic evolution are all vital to the compensation for this disconnect between the expected and actual biocatalyst behaviors. Through a combination of these powerful techniques, biocatalysts have been redesigned for the production of an astounding array of commodity fuels and chemicals, both natural and unnatural (Figure 1 and Table 1). Here we discuss successful examples involving the production of commodity fuels and chemicals, with a focus on D- and L-lactate, L-alanine, succinate, ethanol, and butanol.

2. Methods and Tools for Biocatalyst Redesign

2.1. Chassis. A robust and stable chassis enables efficient and economical production of fuels and chemicals at an industrial level. Since we are specifically interested in biocatalysts that can utilize biomass, a desirable chassis has the following characteristics: (1) growth in mineral salts medium with inexpensive carbon sources, (2) utilization of hexose and pentose sugars, so that all the sugar components in lignocellulosic biomass can be converted to the desired product, (3) high metabolic rate, essential for high rate of productivity, (4) simple fermentation process to reduce the manipulation cost and minimize failure risks in large-scale production, (5) robust organism (high temperature and low pH where possible) to reduce the requirement for external cellulase during cellulose degradation, as well as to reduce the required amount of base addition, (6) ease of genetic manipulation and genetic stability, (7) resistance to inhibitors produced during the biomass pretreatment process, and (8) tolerance to high substrate and product concentrations in order to obtain high titers of target compound.

Enteric bacteria, especially *E. coli*, have many of the above mentioned physiological characteristics and are, thus, an excellent chassis for synthetic biology. Most of the examples discussed here use *E. coli*, but other important microbial model systems have been redesigned, including *Clostridium acetobutylicum* [28], *Corynebacterium glutamicum* [29], *Saccharomyces cerevisiae* [30], and *Aspergillus niger* [31]. *E. coli* has been used as a model organism since the beginning of genetic engineering [32]. While K-12 strain MG1655 (ATCC# 47076) is one of the most commonly used *E. coli* strains [33], there are other lineages, such as B (ATCC# 11303), C (ATCC# 8739), and W (ATCC# 9637), that are also generally regarded as safe since they are unable to

Table 1: Summary of engineered *E. coli* biocatalysts for production of renewable fuels and chemicals in our laboratory.

Product	Fermentation condition[(1)]	Titer (g/L)	Yield(g/g)	Productivity (g/L/h)	Reference
Redesign through modification of existing pathways					
D-lactate	Anaerobic, batch	118	0.98	2.88	[22]
Acetate	Aerobic, fed-batch	53	0.50	1.38	[23]
Succinate	Anaerobic, batch	83	0.98	0.90	[24]
Redesign through introduction of foreign pathways					
Ethanol	Anaerobic, batch	43	0.48	2.00	[25]
L-lactate	Anaerobic, batch	116	0.98	2.29	[22]
Xylitol	Aerobic, fed-batch	38	1.40	0.81	[26]
L-alanine	Anaerobic, batch	114	0.95	2.38	[27]

[(1)] All fermentations were done in mineral salts medium with glucose, except for the ethanol fermentations which used xylose.

colonize the human gut [34]. Although K-12 is the most characterized and widely used strain, *E. coli* W (ATCC# 9637) and C (ATCC# 8739) have proven to be better chassis for synthesizing fuels and chemicals. For example, K-12-derived strains were unable to completely ferment 10% (w/v) glucose in either complex or mineral salts medium [1, 35], while derivatives of strains W or C can completely ferment more than 10% (w/v) of glucose with higher cell growth and sugar utilization rates than K-12. Additionally, *E. coli* W strains have the native ability to ferment sucrose [1, 36].

Foreign genes may be unstable in host cells due to recombination facilitated by mobile DNA elements, and thus the mobile DNA elements in *E. coli* K-12 strain have been deleted [37]. This minimal genome construction strategy is an excellent approach to improve this chassis for the production of fuels and chemicals.

2.2. Systems Biology Tools

2.2.1. Genome-Scale Models and In Silico Simulation. Given the rational basis of metabolic engineering and synthetic biology, models and simulations are critical predictive and tools. Genome sequencing and automatic annotation tools have enabled construction of genome-scale metabolic models of nearly 20 microorganisms [38]. These constraint-based models and *in silico* simulations can be used to predict metabolic flux redistribution after genetic manipulation, or to predict other cellular functions, such as substrate preference, outcomes of adaptive evolution and shifts in expression profiles [39]. They can also aid in pathway design to obtain desired phenotypes [40–42]. For example, the *E. coli* *i*JE660a GSM model was used to successfully simulate single- and multiple-gene knockouts to improve lycopene production [42]. The computational framework, Optknock, was developed to identify gene deletion targets for system optimization [41], and simulation results for gene deletions for succinate, lactate, and 1,3-propanediol production were in agreement with experimental data. Another simulation program, OptStrain, was developed to guide metabolic pathway modification for target compound production, through both the addition of heterologous metabolic reactions and deletion of native reactions [40]. However, most of the current models only have stoichiometric information, while kinetic and regulatory effects are not included [38, 39]. Integration of kinetic and regulatory information will improve the accuracy and predictive power of these models.

2.2.2. High-Throughput Omics Analysis. High-throughput omics analysis, such as transcriptome, proteome, metabolome, and fluxome [43–45], aids in characterization of cellular function on multiple levels, and therefore provide a "debugging" capability for system optimization [12, 45].

Genetic manipulations can disturb the metabolic balance or impair cell growth due to depletion of important precursors [46, 47], accumulation of toxic intermediates [48], or redox imbalance [1]. For example, high NADH levels in *E. coli* reengineered for ethanol production inhibited citrate synthase activity, thereby limiting cell growth by lowering production of the critical metabolite 2-ketoglutarate [49]. Metabolome and fluxome analysis can quickly identify the limiting metabolites or altered metabolic flux distribution, providing the basis for problem solving [45, 50]. For example, metabolite measurements of *Aspergillus terreus* were implemented in the rational metabolic redesign for increased production of lovastatin [45, 50]. Changes of mRNA and protein profiles can be identified by transcriptome and proteome analysis, providing gene targets for further engineering [46, 47]. The work of Choi et al. demonstrate this concept: transcriptome analysis of *E. coli* producing the human insulin-like growth factor I fusion protein aided in selection for targets for gene deletion. The resulting redesigned strain showed a greater than 2-fold increase in product titer and volumetric productivity [46, 47]. Additionally, comparative genome sequence analysis facilitates identification of mutated genes or regulators during evolution, and these mutations can be used to redesign the systems for better synthetic capability. For example, in an effort described as "genome-based strain reconstruction", evolved strains of *Corneybacterium glutamicum* selected for L-lysine production were compared to the parental strain, and mutations were found that were proposed as beneficial to L-lysine production. Three of these mutations were introduced into the parent strain and enabled production of up to 3.0 g/L/hr L-lysine [51].

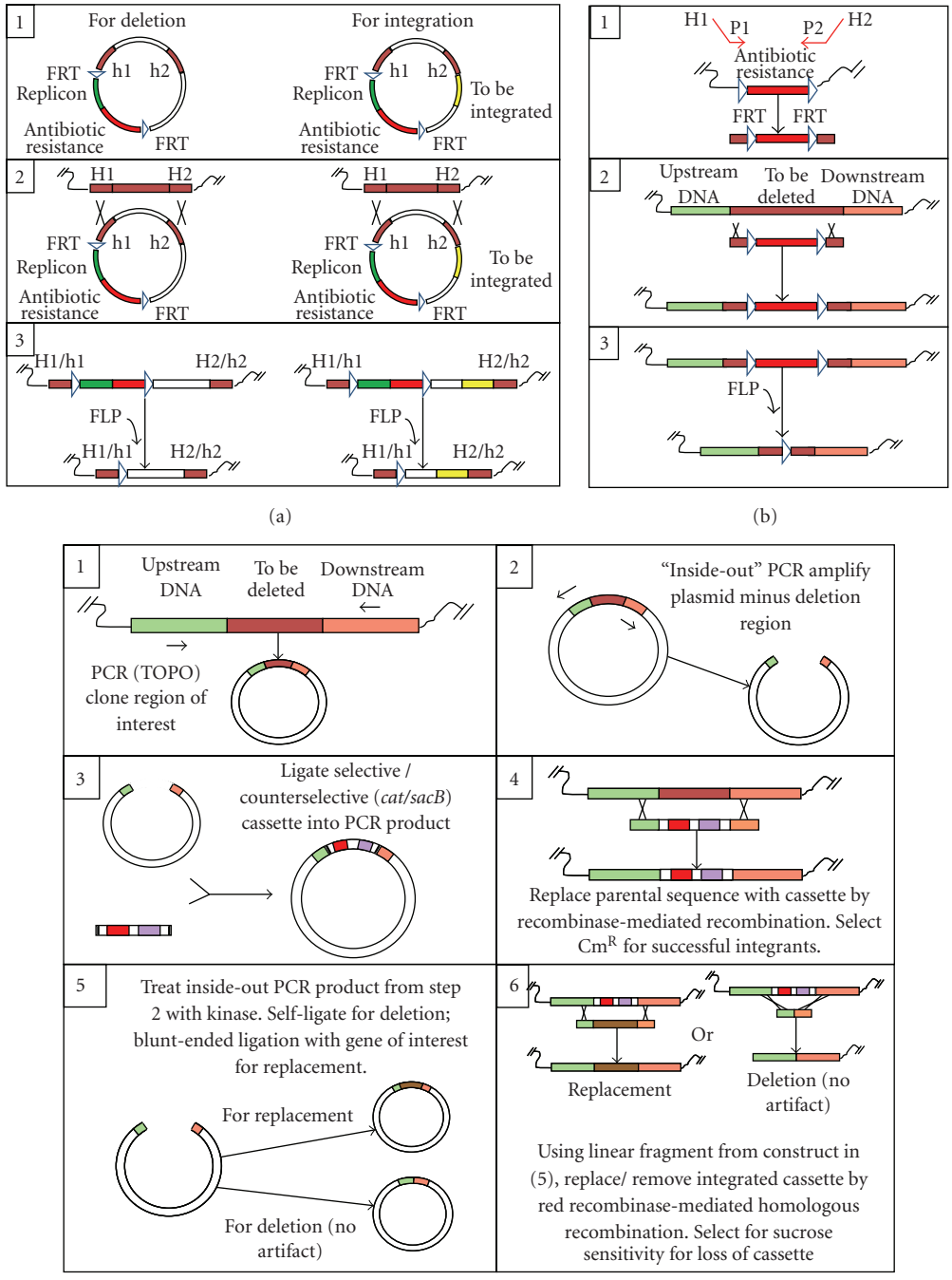


Figure 2: Comparison of three-gene deletion methods in *E. coli*. These methods can also be used in other enteric bacteria. The first and third methods can also be used for gene integration into the chromosome and promoter replacement for tuning gene expression. 2(a) *plasmid-based method*. Step 1 is construction of the deletion plasmid containing DNA fragments homologous to the target gene (h1 and h2), a selectable marker, and either a temperature sensitive or conditional replicon. Step 2 is double-crossover recombination; the plasmid cannot replicate in the host strain, and antibiotic-resistant colonies are selected. In step 3, the FRT, replicon, and antibiotic resistance marker are removed by FLP. 2(b) *Linear DNA-based method*. Step 1 is construction of the linear DNA fragment by PCR (H1-P1 and H2-P2 as primers). H1 and H2 refer to short DNA fragments homologous to target gene. Step 2 is replacement of the target gene with the antibiotic resistance gene through crossover recombination with the help of Red recombinase. Step 3 is removal of FRT and antibiotic marker by FLP. 2(c) *Two-stage recombination-based method developed in our lab*. Steps 1, 2, 3, and 5 describe construction of the plasmids and linear DNA fragments for the two-stage recombinations. Step 4 describes the first recombination step, in which the *cat, sacB* cassette is inserted into the target gene. Step 6 is the second recombination step, in which the *cat, sacB* cassette is removed by selection on sucrose.

2.3. Genetic Manipulation Tools

2.3.1. Gene Deletion. Gene deletion can redistribute carbon flux toward the target product by deleting genes critical to competing metabolic pathways and, thus, is widely used in metabolic redesign strategies. Homologous recombination is the most frequently used strategy for gene-deletion (Figure 2). Historically, plasmids containing a selectable marker flanked by DNA fragments homologous to the target gene and either temperature sensitive or conditional replicons were needed for efficient gene deletion in bacteria [52] (Figure 2(a)). In contrast, genes can be directly disrupted in yeast by linear PCR fragments with short flanking DNA fragments homologous to chromosomal DNA. Linear DNA is not as easy to transform into *E. coli* because of the intracellular exonuclease system and low recombination efficiency. Gene deletion systems based on bacteriophage λ Red recombinase facilitate chromosomal gene deletion using a linear PCR fragment [53]. In this method, the chromosomal gene is replaced by the selectable marker flanked by two FRT (FLP recognition target) fragments (Figure 2(b)) and then the marker can be removed by the FLP recombinase [54]. However, this method leaves a 68bp-FRT scar on the chromosome after each excision [52], reducing further gene deletion efficiency. Repeated use of this FRT/FLP system for specific gene deletions has the potential to generate large unintended chromosomal deletions.

To facilitate sequential gene deletions, our lab has developed a two-stage recombination strategy (Figure 2(c)), using the sensitivity of *E. coli* to sucrose when *Bacillus subtilis* levansucrase (*sacB*) is expressed [24, 27, 55]. Gene deletions created by this method do not leave foreign DNA, antibiotic resistance markers, or scar sequences at the site of deletion. In the first recombination, part of the target gene is replaced by a DNA cassette containing a chloramphenicol resistance gene (*cat*) and levansucrase gene (*sacB*). In the second recombination, the *cat, sacB* cassette is removed by selection for resistance to sucrose. Cells containing the *sacB* gene accumulate levan during incubation with sucrose and are killed [55]. Surviving recombinants are highly enriched for loss of the *cat, sacB* cassette [24, 27].

2.3.2. Gene Expression Tuning. Like gene deletions, plasmid-based expression systems are ubiquitous to metabolic redesign. However, plasmid-based systems have several disadvantages. (1) Plasmid maintenance is a metabolic burden on the host cell, especially for high-copy number plasmids [56]. Note that high copy numbers are not essential, considering that most central metabolic enzymes are encoded by a single gene; (2) plasmid-based expression is dependent on plasmid stability, with only few natural unit-copy plasmids having the desired stability [12]; (3) only low-copy number plasmids have replication that is timed with the cell cycle, and thus maintaining a consistent copy number in all cells is challenging [12]; (4) metabolic redesign can require construction of a complex heterologous pathway, and thus several genes, encoded in large pieces of DNA, need to be incorporated. Most commercial plasmids have difficulties carrying large DNA fragments.

Chromosomal integration of the target genes followed by fine-tuning their expression could eliminate these plasmid-associated problems. The abovementioned two-step recombination strategy for gene deletion can also be used for gene integration or promoter replacement (Figure 2).

Gene expression in prokaryotes is mainly controlled at the transcriptional level, and therefore the promoter is the most tunable element. While inducible promoters, such as *lac* and *ara*, have been traditionally used to modulate gene expression, large-scale inducer use is cost prohibitive for production of fuels and bulk chemicals. However, several strategies have been developed to construct constitutive promoter libraries for fine-tuning gene expression. Some methods rely on the use of natural promoters. For example, *Zymomonas mobilis* genomic DNA was used to construct a promoter library for screening optimal expression of *Erwinia chrysanthemi* endoglucanase genes (*celY* and *celZ*) in *Klebsiella oxytoca* P2 in order to improve ethanol production from cellulose [57]. Other methods rely on random modification of existing promoters, such as the randomization of the spacer sequences between the consensus sequences [58], or mutagenesis of a constitutive promoter [59]. This promoter modification method was used to assess the impact of phosphoenolpyruvate carboxylase levels on cell yield and deoxy-xylulose-P synthase levels on lycopene production, and the optimal expression levels of these genes were identified for maximal desired phenotype [59]. These synthetic promoter libraries could also be integrated into the chromosome directly, which could facilitate expression modulation of chromosomal genes [60, 61].

The fine-tuning methods described above rely on the selection of the best natural promoter or random alteration of existing promoters. One of the goals of synthetic biology is construction of standard parts, and posttranscriptional processes, such as transcriptional termination, mRNA degradation, and translation initiation, have been engineered with this goal in mind. Examples include construction of a synthetic library of 5' secondary structures to successfully manipulate mRNA stability [62], and modulation of the ribosome binding site (RBS) as well as Shine-Dalgarno (SD) and AU-rich sequences to tune gene expression at the translation initiation process [60, 63]. Riboregulators were also developed to tune gene expression via RNA-RNA interactions [64]. A final method of fine-tuning gene expression is codon optimization, which can improve translation of foreign genes [65]. These optimized gene sequences often do not exist in nature and must be generated using DNA synthesis techniques.

In many cases, more than one gene needs to be introduced into the chassis and expression of these genes needs to be coordinated to attain desired biocatalyst performance. One such method is modulation of the expression of each individual gene via its own promoter. However, it is difficult to predict the appropriate expression level of each gene. Another option is to combine multiple genes into a synthetic operon with a single promoter, and fine-tune expression of each gene through posttranscriptional processes [12] with tunable control elements (such as mRNA secondary structure, RNase cleavage sites, ribosome binding sites, and

sequestering sequences) at intergenic regions. Libraries of tunable intergenic regions (TIGRs) were generated and screened to tune expression of several genes in an operon [48]. This method was used to coordinate expression of three genes in an operon that encodes a heterologous mevalonate biosynthetic pathway, improving mevalonate production by 7-fold [48]. Another method to control expression of more than one gene is to engineer global transcription machinery by random mutagenesis of transcription factors [66, 67]. This method was shown to efficiently improve tolerance to toxic compounds and production of metabolites, and to alter phenotypes [66, 67].

2.3.3. Protein Engineering. Natural proteins may not meet the required criteria for specific and efficient system performance, and thus alteration for a specific application may be needed. Directed evolution of proteins offers a way to rapidly optimize enzymes, even in the absence of structural or mechanistic information [68]. For directed evolution, a protein library is usually generated by random mutagenesis [68], recombination of a target gene [69], or a family of related genes [70] and then the library is analyzed by high-throughput screening. This method has been used to successfully increase enzyme activity [71, 72], increase protein solubility and expression, invert enantioselectivity, and increase stability and activity in unusual environments [68]. For example, a mutation library of the gene-encoding geranylgeranyl diphosphate synthase of *Archaeoglobus fulgidus* was generated to screen for mutants with higher activity, enabling lycopene production in *E. coli*. Screening of more than 2,000 variants identified eight with increased activity; one of which increased lycopene production by 100% [71].

Of particular relevance to the field of synthetic biology is the creation of novel enzymatic activity through protein engineering [73, 74]. For example, the unnatural isomerization of α-alanine to β-alanine was attained by evolving a lysine 2,3-aminomutase to expand its substrate specificity to include α-alanine [73].

Rational design is another powerful tool to increase protein properties, especially with the aid of computational analysis [75, 76]. Based on knowledge of protein structure and function, one can predict which amino acid(s) to change in order to obtain the desired function. In the redesign of *Lactobacillus brevis* for the production of secondary alcohols, it was desired to change the cofactor preference of the R-specific alcohol dehydrogenase from NADPH to NADH. A structure-based computational model was used to identify potentially beneficial amino acid substitutions and one of these changes increased NADH-dependent activity four-fold [77].

While these examples demonstrate the power of rational enzyme (re-)design, this approach requires detailed information about the protein structure and mechanism, while random mutagenesis does not. Recent advances have combined directed evolution and rational design in a so-called "semi-rational" approach to successfully improve enzyme activity when only limited information is available [78, 79]. When the mutagenesis is limited to specific residues, as chosen from existing structural or functional knowledge, these "smart" libraries are more likely to yield positive results [79]. For example, the catalytic activity of pyranose-2-oxidase was improved by mutagenesis of the known active site [80].

While the 20 natural amino acids supply enzymes with a wide range of possible activity, this range can be expanded even further by the use of unnatural amino acids (UAAs). There are more than 40 UAAs available at this time and they have been used to probe protein function, photocage critical residues, and alter metalloprotein properties [81, 82]. While this technology is still in the developmental stage, at least one study has shown an improvement in enzyme activity following insertion of UAAs. Site 124 of *E. coli*'s nitroreductase was replaced with a variety of natural and unnatural amino acids and certain UAA variants had a greater than 2-fold increase in activity over the best natural amino acid variant [83]. This biomimetic approach has been expanded to other metabolites, such as carbohydrates [84] and lipids [85].

2.4. Evolution. As described above, a robust biocatalyst with high yield, titer, and productivity is critical for a fermentation process to compete with petrochemical-based production. Current models and simulation tools provide a framework given the constraints of known protein functions. But the many reactions and enzymes that remain uncharacterized cannot be included in this theoretical analysis. Therefore rational design methods often result in a biocatalyst that performs poorly relative to the model. Metabolic evolution provides a complementary approach to improve biocatalyst productivity and robustness, dependent upon the design of an appropriate selection pressure. Where feasible, synthesis of the target compound can be coupled to the production of ATP, redox balance, or key metabolites that are essential for growth, and selection for improvements in growth during metabolic evolution (serial transfers) can be used to coselect for higher rates or titers of target compounds (Figure 3). Both redox balance and net ATP production in such a synthetic system are requisites for successful evolution.

We have used this metabolic evolution strategy to optimize biocatalysts redesigned for production of several fermentation products [1], including ethanol, D-lactate, L-lactate, L-alanine (Figure 3), and succinate, as described in more detail below. A frequently-used design scheme is to couple synthesis of the target product to growth by inactivating competing NADH-consuming pathways. Thus, the only way for cells to regenerate NAD^+ for glycolysis is to produce the target compound. Increased cell growth, supported by higher ATP production rate during glycolysis, is coupled with higher NADH oxidization rate, and thus tightly coupled with synthesis of target product. This evolution strategy has been shown to increase productivity by up to two orders of magnitude.

Computational frameworks based on genome-scale metabolic models have been used to construct biocatalysts that couple biomass formation with chemical production [40, 41], and therefore provide a basis for selective pressure for high productivity. For example, Optknock identified gene deletion targets for the construction of lactate-producing

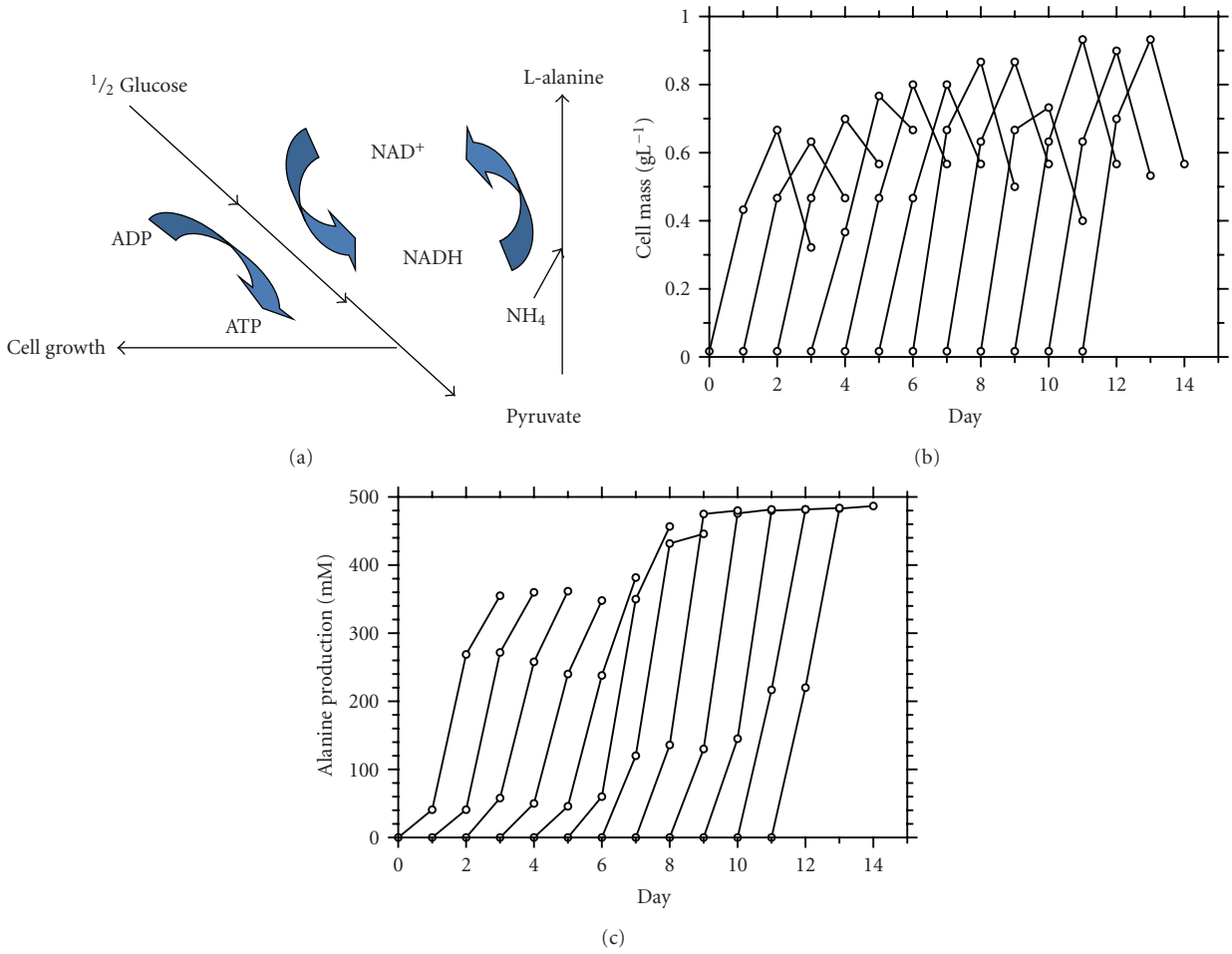


Figure 3: Metabolic evolution for improving L-alanine production in *E. coli* [27]. 3(a) Redesigned metabolic pathway for L-alanine production: ATP production and cell growth is coupled to NADH oxidation and L-alanine production. 3(b) Directed evolution improves cell growth. Parental strain XZ112 reaches a maximum cell mass of 0.7 gL^{-1} after 48 hours of fermentation; evolved strain XZ113 attains 0.7 gL^{-1} after 24 hours and a maximum of 0.9 gL^{-1} after 48 hours; 3(c) metabolic evolution to improve cell growth also improves alanine production. Parental strain XZ112 produces 355 mM alanine after 72 hours of fermentation; evolved strain XZ113 produces 484 mM in 48 hours.

E. coli, and then directed evolution improved production capability [86]. Although rational design of metabolic pathways based on current metabolic models is a common method for maximizing yield of the target compound, this method is not always the best strategy, due to our limited understanding of the complicated metabolic network and dynamic kinetics of each reaction. Metabolic evolution provides an excellent alternative method for strain improvement, through which reactions that are not currently predictable would be selected to improve biocatalyst performance [87]. As our knowledge of biocatalyst behavior and metabolism improves, predictive models will become even more powerful.

3. Redesign through Modification of Existing Pathways

In this section, we highlight projects that have redesigned a chassis to produce target compounds at high yield and titer without the introduction of foreign pathways. In the next section, we describe biocatalyst redesigns which used foreign or nonnatural pathways.

3.1. Succinate. Succinate, a four-carbon dicarboxylic acid, is currently used as a specialty chemical in food, agricultural, and pharmaceutical industries [88] but can also serve as a starting point for the synthesis of commodity chemicals used in plastics and solvents, with a potential global market of $15 billion [89]. Succinate is primarily produced from petroleum and there is considerable interest in the fermentative production of succinate from sugars [89].

Several rumen bacteria can produce succinate from sugars with a high yield and productivity [90–92], but require complex nutrients. Alternatively, native strains of *E. coli* ferment glucose effectively in simple mineral salts medium but produce succinate only as a minor product [93]. Therefore *E. coli* strain C (ATCC 8739) was redesigned for

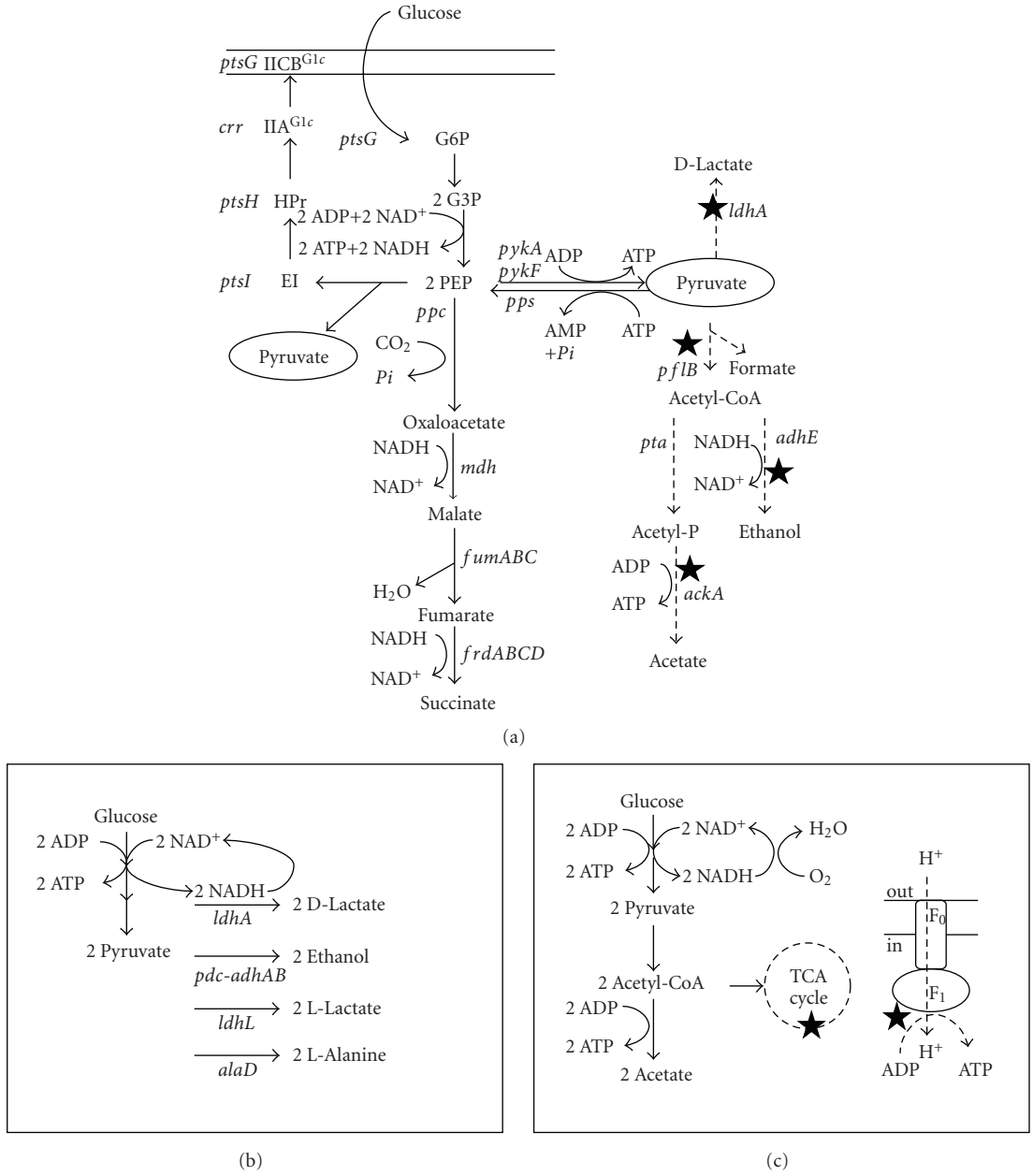


Figure 4: Continued.

succinate production at high yield, titer, and productivity [94].

The initial redesign strategy focused on inactivation of competitive pathways, specifically deletion of lactate dehydrogenase (*ldhA*), alcohol/aldehyde dehydrogenase (*adhE*), and acetate kinase (*ackA*). However, the resulting strain grew poorly in mineral salts medium under anaerobic condition and accumulated only trace amounts of succinate. Because NADH oxidization is coupled to succinate synthesis in this strain, metabolic evolution was used to improve both the cell growth and succinate production. After inactivation of pyruvate formate-lyase and methylglyoxal synthase to eliminate formate and lactate production, the final strain, KJ073, produced near 670 mM succinate (80 g/L) in mineral salts medium with a high yield (1.2 mol/mol glucose) and high productivity (0.82 g/L/h) [94]. Inactivation of threonine decarboxylase (*tdcD*), 2-ketobutyrate formate-lyase (*tdcE*), and aspartate aminotransferase (*aspC*) further increased succinate yield (1.5 mol/mol glucose), titer (700 mM), and productivity (0.9 g/L/h) [24].

Despite its power in improving biocatalyst performance, metabolic evolution has the undesirable property of being a black box; evolved strains show the desired biocatalyst properties, but the metabolic evolution process does not improve our understanding of the biocatalyst. Therefore, reverse engineering of evolved strains can help us identify

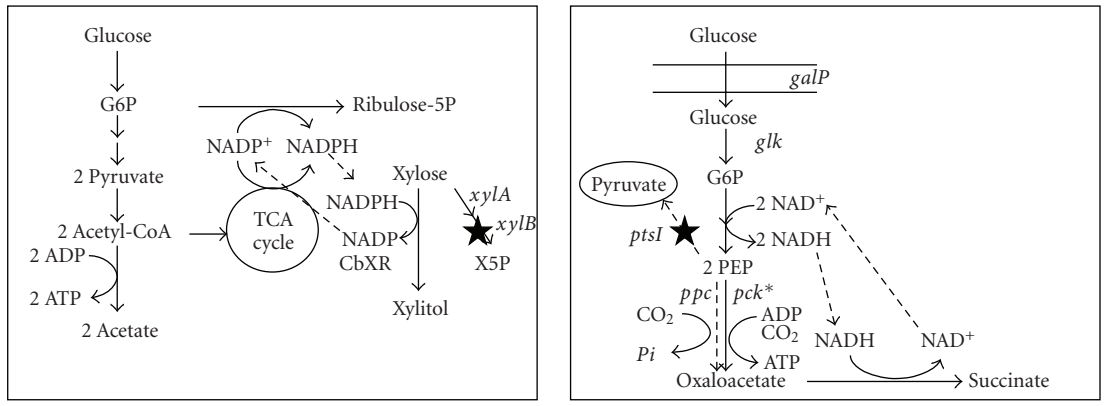


Figure 4: Synthetic pathways of *E. coli* for production of fuels and chemicals in our lab: 4(a) Native metabolic pathways of glucose fermentation in *E. coli*; 4(b) synthetic pathways for production of D-lactate, ethanol, L-lactate and L-alanine; 4(c) synthetic pathways for production of pyruvate and acetate; 4(d) synthetic pathway for production of xylitol, 4(e) synthetic pathway for production of succinate. ★ indicate gene deletion. Genes and enzymes: *ackA*, acetate kinase; *adhAB*, alcohol dehydrogenase (*Z. mobilis*); *adhE*, alcohol/aldehyde dehydrogenase; *alaD*, L-alanine dehydrogenase (*G. stearothermophilus*); *crr*, glucose-specific phosphotransferase enzyme IIA component; *frd*, fumarate reductase; *fum*, fumarase; *galP*, galactose-proton symporter (glucose permease); *glk*, glucokinase; *ldhA*, D-lactate dehydrogenase; *ldhL*, L-lactate dehydrogenase (*P. acidilactici*); *mdh*, malate dehydrogenase; *pdc*, pyruvate decarboxylase; *pflB*, pyruvate formate-lyase; *ppc*, phosphoenolpyruvate carboxylase; *pta*, phosphate acetyltransferase; *ptsG*, PTS system glucose-specific EIICB component; *ptsH*, phosphocarrier protein HPr; *ptsI*, phosphoenolpyruvate-protein phosphotransferase (Phosphotransferase system, enzyme I); *pyk*, pyruvate kinase; *xrd*, xylose reductase (*C. boidinii*); *xylB*, xylulokinase. Metabolites: G6P, glucose-6-phosphate; G3P, glycerol-3-phosphate; PEP, phosphoenol pyruvate; X5P, D-xylulose-5-phosphate.

key mutations that can then be rationally applied to other biocatalysts. Reverse engineering of the succinate-producing strain revealed two significant changes in cellular metabolism that increased energy efficiency [87]. The first change is that PEP carboxykinase (*pck*), which normally functions in gluconeogenesis during the oxidative metabolism of organic acids [90, 95, 96], became the major carboxylation pathway for succinate production. High-level expression of PCK dominated CO_2 fixation and increased ATP yield (1 ATP per oxaloacetate produced). The second change is that the native phosphoenolpyruvate- (PEP-) dependent phosphotransferase system for glucose uptake was inactivated and replaced by an alternative glucose uptake pathway: GalP permease (*galP*) and glucokinase (*glk*). These changes increased the pool of PEP available for maintaining redox balance, as well as increasing energy efficiency by eliminating the need to produce additional PEP from pyruvate, a reaction that requires two ATP equivalents [97].

While rational design based on current metabolic understanding is a key component of metabolic engineering and synthetic biology, our limited understanding of the complicated metabolic network and dynamic kinetics of each reaction can lead to failure of predictive models. In this example, metabolic evolution was demonstrated as an excellent alternative method for strain improvement, through which currently unpredictable reactions would be selected to expand cellular metabolic capability [87]. By understanding the mutations that enabled desirable performance of the succinate-producing strain, we have more options available for the redesign of future systems. To demonstrate this, *E. coli* was again redesigned based on the findings from the evolved strain [98]. This time, the design strategy shifted from inactivating competitive fermentation pathways to recruiting energy conserving pathways for efficient succinate production (Figure 4(e)). After increasing *pck* gene expression and inactivating the native glucose PTS system, the native *E. coli* metabolic system was converted to an efficient succinate synthetic system, equivalent to the native pathway of succinate-producing rumen bacteria [98].

3.2. D-Lactate. D-lactate is widely used as a specialty chemical in the food and pharmaceutical industry. It can also be combined with L-lactate for the production of polylactic acid (PLA), an increasingly popular biorenewable and biodegradable plastic [99, 100] whose commercial success obviously depends on the production cost. Although glucose is the current substrate for fermentative production of lactate, it is desirable to produce this commodity chemical from lignocellulosic feedstock, which contains a mixture of sugars. Some lactic acid bacteria have the desirable native ability to produce large amount of D-lactate under low pH condition, where the low pH reduces the process cost [101, 102]. However, these lactic acid bacteria require complex nutrients, and many of them lack the ability to ferment pentose sugars. The lactic acid bacteria that do ferment pentose sugars unfortunately produce a mixture of lactate and acetate and, thus, are not a good chassis for commercial production of D-lactate. While *E. coli* can ferment many sugars effectively in a simple mineral salts medium, inherent D-lactate productivity is low and other undesirable metabolites are also produced [103]. Therefore, the *E. coli* metabolic system was redesigned to attain

the desired properties of high yield and productivity of D-lactate.

E. coli strain W3110 was used as the chassis for D-lactate production with a redesign strategy that focused on inactivation of competitive fermentation pathways [104]. After deleting the genes encoding fumarate reductase (*frd-ABCD*), alcohol/aldehyde dehydrogenase (*adhE*), pyruvate formate lyase (*pflB*), and acetate kinase (*ackA*), the resulting strain, SZ63, can only oxidize NADH via D-lactate synthesis (Figure 4(b)). Although this strain could completely utilize 5% (w/v) glucose in a mineral salts medium with a yield near theoretical maximum (96%), the volumetric D-lactate productivity of 0.42 g/L/h was relatively low compared with lactic acid bacteria [35]. In addition, this strain can neither utilize sucrose nor completely utilize 10% (w/v) sugar [35]. Therefore, an *E. coli* W derivative strain was chosen as chassis for more robust D-lactate production [35, 105]. After redesigning central metabolism so that D-lactate production was the sole means of oxidizing NADH, metabolic evolution was used to further improve cell growth and D-lactate productivity. The resulting strain, SZ194, efficiently consumed 12% (w/v) glucose in mineral salts medium and produced 110 g/L D-lactate [105] with a volumetric productivity of 2.14 g/L/h, a 5-fold increase over the W3110 derivative. The biocatalyst was further optimized by deleting methylglyoxal synthase gene (*mgsA*) to eliminate L-lactate production, and by metabolic evolution to increase yield and productivity. The final D-lactate producing strain, TG114, could convert 12% (w/v) glucose to 118 g/L D-lactate with an excellent yield (98%) and productivity (2.88 g/L/h) [22].

3.3. Acetate. Acetate is a commodity chemical with 2001 worldwide production estimated at 6.8 million metric tons [23]. Biological production of acetate accounts for only 10% of world production, mainly in the form of vinegar, with the remainder of production through petrochemical routes [106–108]. Biological production of commodity chemicals has historically focused on anaerobic production of reduced products, since substrate loss as cell mass and CO_2 is minimal and product yields are high. Contrastingly, acetate is an oxidized chemical, and traditional biological production involves a complex two-stage process: fermentation of sugars to ethanol by *Saccharomyces*, followed by aerobic oxidation of ethanol to acetate by *Acetobacter* [106–108]. To enable microbial production of redox-neutral or oxidized products at high yield, the biocatalyst metabolism needs to be redesigned to combine attributes of both fermentative and oxidative metabolisms.

Redesign of *E. coli* W3110 metabolism for acetate production focused on three major pathways: fermentative metabolism, oxidative metabolism, and energy supply (Figure 4(c)) [23]. The competitive fermentation pathways (*pflB, ldhA, frd, adhE*) were inactivated to prevent the consumption of common precursor pyruvate, and the oxidative tricarboxylic acid (TCA) cycle was interrupted to reduce the carbon loss as CO_2. Finally, oxidative phosphorylation was disrupted (*atpFH*) to reduce ATP production while maintaining the ability to oxidize NADH by the electron transport system, thus increasing the glycolytic flux for more ATP production through substrate-level phosphorylation. Although rationally designed, the resulting strain, TC32, had an undesirable auxotrophic requirement for succinate during growth in glucose-minimal medium. Evolution was used to eliminate this auxotrophy and the final strain, TC36, produced 878 mM acetate (53 g/L) in mineral salts medium with 75% of the maximal theoretical yield. Although this is a lower titer than acetate produced from ethanol oxidation by *Acetobacter*, TC36 has a two-fold higher production rate, requires only mineral salts medium, and can metabolize a wide range of carbon sources in a simple one-step process [23].

3.4. Others. Butanol is an excellent alternative transportation fuel with several advantages compared to ethanol, including higher-energy content, lower volatility, less hydroscopicity, and less corrosivity [109]. Redesign of *E. coli* for butanol production is discussed below. *C. acetobutylicum* ATCC 824 naturally produces butanol and was redesigned to increase butanol production and decrease coproduct accumulation. Metabolic engineering-type modifications, such as overexpression of the acetone formation pathway to increase formation of butanol precursor butyryl-CoA, inactivation of the transcriptional repressor SolR, and overexpression of alcohol/aldehyde dehydrogenase all increased butanol production [110–112]. In an excellent example of synthetic biology-type applications, expression of the butyrate kinase gene was fine-tuned by a rationally designed antisense RNA to increase butanol production [113].

1,2-propanediol (1,2-PD) is a major commodity chemical currently derived from propylene. *E. coli* naturally produces low amounts of 1,2-PD, and therefore its metabolism was redesigned to produce 1,2-PD at high yield and titer from glucose this was achieved by inactivation of competing pathways (lactate dehydrogenase and glyoxalase I), and overexpression of essential genes of 1,2-PD synthetic pathway (methylglyoxal synthase, glycerol dehydrogenase, and 1,2-PD oxidoreductase) [114]. Evolution was also used in combination with rational design for increased 1,2-PD production [115].

L-valine, an essential hydrophobic and branched-chain amino acid, is used in cosmetics, pharmaceuticals, and animal feed additives [116]. *E. coli* was redesigned for L-valine production at high yield and titer from glucose through a combination of traditional metabolic engineering and synthetic biology. Traditional metabolic engineering was used to inactivate competing pathways and overexpress acetohydroxy acid synthase I (*ilvBN*), part of the valine biosynthesis pathway. Unfortunately, the *E. coli* chassis has regulatory elements that tightly control L-valine biosynthesis, making production of valine at high yield and titer difficult. Feedback inhibition was eliminated by rational site-directed mutagenesis of acetohydroxy acid synthase III. In an excellent demonstration of the gene expression tuning techniques discussed above, transcriptional attenuation of valine biosynthesis genes *ilvGMEDA* was eliminated by replacing the attenuator leader region with the constitutive *tac* promoter. Transcriptome analysis and *in silico* simulation guided selection of additional target genes for amplification

and deletion, and the final biocatalyst produced 0.378 g L-valine per g glucose, giving a titer of 7.55 g/L valine from 20%(w/v) glucose [116]. A similar strategy was also used for L-threonine production [117].

4. Redesign through Introduction of Foreign or Nonnatural Pathways

4.1. Foreign Pathways

4.1.1. Ethanol. Ethanol is a renewable transportation fuel. Replacement of gasoline with ethanol would significantly reduce US import oil dependency, increase the national security, and reduce environmental pollution [118]. However, only 9 billion gallons of ethanol were produced in 2008, and all were from corn-based production. Lignocellulose is generally regarded as an excellent source of sugars for conversion into fuel ethanol. It is, thus, desirable to design or obtain biocatalysts that can utilize all the sugar components in lignocellulose and convert them to ethanol with high yield and productivity in mineral salts medium. Native *S. cerevisiae* and *Z. mobilis* strains can efficiently convert glucose to ethanol, but cannot utilize pentose sugars. In contrast, *E. coli* strains can utilize all the sugar components of lignocelluloses but ethanol is only a minor fermentation product, with mixed acids accumulating as the major fermentation product [103]. While recent advances have been made engineering the native *E. coli* metabolic pathways for ethanol production [119], the most successful example used a foreign metabolic pathway to enable ethanol production from *E. coli* strain W (ATCC# 9637) [1].

Redesign for ethanol production was decoupled to three parts: construction of a metabolic pathway for production of ethanol as the major fermentation product, elimination of competitive NADH oxidization pathways, and disruption of side-product formation. The *Z. mobilis* homoethanol pathway (pyruvate decarboxylase and alcohol dehydrogenase) was introduced as a foreign pathway, enabling redox-balanced production of ethanol at high yield [120] (Figure 4(b)). Then fumarate reductase (*frd*) was disrupted to increase ethanol yield. The resulting strain, KO11, produced ethanol at a yield of 95% in a complex medium [121]. This strain was developed at the dawn of metabolic engineering and has been used to produce ethanol from a variety of lignocellulosic materials, as reviewed in [1].

Although the ethanol production rate of KO11 was as high as yeast, the ethanol tolerance and performance in minimal medium did not meet the desired standards. Therefore strain SZ110, a derivative of KO11 modified for lactate production in mineral salts media [35], was redesigned for ethanol production [122]. As with the design of KO11, redesign of SZ110 was decoupled to construction of an ethanol synthetic pathway, elimination of competitive NADH oxidization pathways, and blockage of side-product formation. However, this redesign strategy also included the acceleration of mixed sugar co-utilization. The lactate producing pathway was disrupted and the *Z. mobilis* homoethanol pathway was integrated into the chromosome by random insertion to select for optimal expression. The *Pseudomonas putida* short-chain esterase (*estZ*) [123] was introduced to decrease ethyl acetate levels in the fermentation broth and decrease the downstream purification cost. In addition, methylglyoxal synthase (*mgsA*) was inactivated, resulting in co-metabolism of glucose and xylose, and accelerated the metabolism of a 5-sugar mixture (mannose, glucose, arabinose, xylose, and galactose) to ethanol [25]. After using evolution to increase cell growth and production, the final strain, LY168, could concurrently metabolize a complex combination of the five principal sugars present in lignocellulosic biomass with a high yield and productivity in mineral salts medium [25].

4.1.2. L-Lactate. As described above, L-lactate is the major component of the biodegradable plastic PLA. Although many lactic acid bacteria produce L-lactate with high yield and productivity [124], they usually require complex nutrients. *E. coli* does not have a native pathway for L-lactate production, and therefore introduction of a foreign pathway was necessary.

The strategy for redesigning *E. coli* W3110 for L-lactate production was to eliminate competitive NADH oxidization pathways and then construct the desired L-lactate synthetic pathway (Figure 4(b)) [125]. The L-lactate production pathway, L-lactate dehydrogenase (*ldhL*) from *Pediococcus acidilactici,* was used and its coding region and terminator were integrated into the *E. coli* chromosome at the *ldhA* site, so that *ldhL* could be expressed under the native *ldhA* promoter. In addition, since the *ldhL* gene contains a weak ribosomal-binding region, this region was rationally replaced with *ldhA*'s RBS [125]. Following a period of metabolic evolution, the resulting strain, SZ85, synthesized 45 g/L L-lactate in a mineral salts medium with yield near theoretical maximum (94%). However, this strain was a K-12 derivative and displayed the same problems seen with the K12-based D-lactate-producing strain described above, meaning that it was unable to completely ferment high sugar concentrations and had a low productivity (0.65 g/L/h). Therefore, the same design strategy was implemented in an *E. coli* W (ATCC# 9637) derivative. After further deleting *mgsA* gene to improve chiral purity and using metabolic evolution to improve cell growth and productivity, the final L-lactate-producing strain, TG108, could convert 12% glucose to 116 g/L L-lactate with an excellent yield (98%) and productivity (2.29 g/L/h) [22].

4.1.3. Xylitol. The pentahydroxy sugar alcohol xylitol is commonly used to replace sucrose in food and as a natural, non-nutritive sweetener that inhibits dental caries [126]. Xylitol can also be used as a building block for synthesizing new polymers [127]. Current xylitol commercial production involves hydrogenation of hemicellulose-derived xylose with an active metal catalyst [127]. Biological-based processes have also recently been developed, but although high xylitol titer was achieved by some yeast, the process requires complex medium with numerous expensive vitamin supplements [128]. While *E. coli* does not have the native capability to

synthesize xylitol, a redesign strategy for strain W3110 was proposed involving a foreign metabolic pathway [26]. In the proposed redesign, glucose would support cell growth and provide reducing equivalents, while xylose would be used as substrate for xylitol synthesis (Figure 4(d)). The design strategy consisted of three major components: enabling co-utilization of glucose and xylose, separation of xylose metabolism from central metabolism, and construction of a xylitol production pathway (Figure 4(d)). In order to enable co-utilization of glucose and xylose, glucose-mediated repression of xylose metabolism was eliminated by replacing the native *crp* gene with a cAMP-independent mutant (CRP*). Xylose metabolism was separated from central metabolism by deleting the xylulokinase (*xylB*) gene, preventing the loss of xylose carbon to central metabolism. Finally, xylose reductase and xylitol dehydrogenase from several microorganisms were tested for xylitol synthetic capability, and the NADPH-dependent xylose reductase from *C. boidinii* (CbXR) was found to support optimal xylitol production. The final strain, PC09 (CbXR), could produce 250 mM (38 g/L) xylitol in mineral salts medium. The yield was 1.7 mol xylitol per mol glucose consumed, which was improved to 4.7 mol/mol by using resting cells. It was proposed that xylitol production could be further improved by increasing supply of reducing equivalents [129].

4.1.4. L-Alanine. L-alanine can be used with other L-amino acids as a pre- and postoperative nutrition therapy in pharmaceutical and veterinary applications [130]. It is also used as a food additive because of its sweet taste. The annual worldwide production of L-alanine is around 500 tons [131], and this market is currently limited by production costs. The current commercial production process converts aspartate to alanine via aspartate decarboxylase, where aspartate is produced from fumarate by aspartate ammonia-lyase catalysis [27]. An efficient fermentative process with a renewable feedstock such as glucose offers the potential to reduce L-alanine cost and facilitate a broad expansion of the alanine market into other products.

SZ194, a derivative of *E. coli* W (ATCC# 9637) that was previously engineered for D-lactate production, was used as the chassis for L-alanine production [27] (Figure 4(b)). Alanine production in the native strain uses glutamate- and NADPH-dependent glutamate-pyruvate aminotransferase. It is preferable to produce L-alanine directly from pyruvate and ammonia using an NADH-dependent enzyme, and therefore L-alanine dehydrogenase (*alaD*) of *Geobacillus stearothermophilus* was employed. The native ribosome binding site, coding region, and terminator of *alaD* gene were integrated into the *E. coli* chromosome at the *ldhA* site, so that expression of *alaD* could be controlled by the native promoter of *ldhA*, a promoter that has worked well for production of D- and L-lactate, as described above. Further redesign focused on elimination of trace amounts of lactate and increasing the L-alanine chiral purity by deleting *mgsA* and the major alanine racemase gene (*dadX*). Metabolic evolution increased the final titer and productivity by 15- and 30-fold, respectively (Figure 3). The latest L-alanine producing strain, XZ132, converted 12% glucose to 114 g/L L-alanine with a 95% yield and the excellent volumetric productivity of 2.38 g/L/h [27].

4.1.5. Combining Multiple Foreign Pathways in a Single Chassis. Although the work described above relied on the introduction of a single foreign pathway, there are other excellent examples that employ pathways from more than one organism in a single host.

E. coli was redesigned for 1,3-propanediol production using *S. cerevisiae* pathway to convert glucose to glycerol and a *K. pneumonia* pathway to convert glycerol to 1,3-propanediol [132]. *E. coli* was also redesigned for isopropanol production by combining acetyl CoA acetyltransferase (*thl*) and acetoacetate decarboxylase (*adc*) from *C. acetobutylicum* with the second alcohol dehydrogenase (*adh*) from *C. beijerinckii* and *E. coli*'s own acetoacetyl-CoA transferase (*atoAD*) [133]. Artemisinic acid, a precursor of antimalarial drug artemisin, was produced by *E. coli* following the combination of a mevalonate pathway from *S. cerevisiae* and *E. coli*, amorphadiene synthase, and a novel cytochrome P450 monooxygenase (CYP71AV1) from *Artemisia annua* [12, 134].

S. cerevisiae was redesigned for flavanone production by combining *Arabidopsis thaliana* cinnamate 4-hydroxylase (C4H), *Petroselinum crispum* 4-coumaroyl: CoA-ligase (4CL), and *Petunia* chalcone synthase (CHS), *Petunia* chalcone isomerase (CHI) [135]. A similar synthetic system producing hydroxylated flavonols was also constructed in *E. coli* with additional amplification of *C. roseus* P450 flavonoid 3′, 5′-hydroxylase (F3′5′H) fused with P450 reductase, *Malus domestica* flavanone 3β-hydroxylase (FHT), and *Arabidopsis thaliana* flavonol synthase (FLS) [136]. The flavonoid production was significantly increased through further redesigning of the central metabolic system of *E. coli* to increase precursor (*Malonyl-CoA*) supply [137].

4.2. Modification of Natural Pathways for Production of Unnatural Compounds. One of the goals of synthetic biology is to design or construct new genetic circuits. In the examples given thus far, existing biological parts have been reassembled to engineer a biocatalyst that efficiently produces a product that already exists in nature. However, metabolic pathways can also be constructed to produce unnatural compounds.

As discussed above, directed evolution of proteins can modify their activity such that new substrates are recognized or new products are formed [138]. For example, novel carotenoid compounds were generated by evolution of two key carotenoid synthetic enzymes, phytoene desaturase, and lycopene cyclase [139]. Additionally, combinatorial biosynthesis, which combines genes from different organisms into a heterologous host, can also generate new products [140]. For example, four previously unknown carotenoids were produced by combinatorial biosynthesis in *E. coli* [141].

4.3. De Novo Pathway Design. In order to broaden the available biosynthesis space, it is essential to go beyond the natural

pathways and design pathways *de novo* [142]. Although this exciting design strategy still has many challenges, several successful examples have been reported.

For example, a synthetic pathway for 3-hydroxypropionic acid (3-HP) production was designed involving the unnatural isomerization of α-alanine to β-alanine, as mentioned above. In this example the researchers used directed evolution to expand the substrate specificity of lysine 2,3-aminomutase to include α-alanine [73]. The resulting β-alanine can then be converted to 3-HP through existing metabolic pathways.

Unnatural pathways for higher alcohol production in *E. coli* were designed by combining the native amino acid synthetic pathways with a 2-keto acid decarboxylase from *Lactococcus lactis* and alcohol dehydrogenase from *S. cerevisiae* [143]. The 2-keto acid intermediates in amino acid biosynthesis pathways were redirected from amino acid production to alcohol production, enabling production of 3-methyl-1-pentanol. This pathway was then expanded for production of unnatural alcohols by rational redesign of two enzymes, with the resulting biocatalysts having the ability to synthesize various unnatural alcohols ranging in length from five to eight carbons [144].

4.4. Engineering Tolerance to Inhibitory Compounds. As our repertoire of biologically-produced compounds increases, tolerance to high product titers becomes more important. Biofuels, such as ethanol and butanol, can inhibit biocatalyst growth, and therefore the tolerance of the biocatalyst needs to be improved [145–147]. As described above, our goal is to use lignocellulosic biomass as a substrate for production of commodity fuels and chemicals. Unfortunately, the processes used to convert biomass to soluble sugars also produce a mixture of minor products, such as furfural and acetic acid, that inhibit biocatalyst metabolism [148]. Although most of these inhibitors could be removed by detoxification [149], this additional process would increase operational cost. It is, thus, desirable to obtain microorganisms that are tolerant to these inhibitors and can directly ferment hemicellulose hydrolysate.

One approach to increasing tolerance is to understand the mechanism of inhibition. Transcriptome analysis has been used to probe the response to ethanol [145, 150], furfural [151], and butanol [147]. Another approach is to use directed evolution, as highlighted by the following example. Ethanologenic *E. coli* strain LY180 (a derivative of LY168 with restored lactose utilization and integration of an endoglucanase, and cellobiose utilization) was used as the chassis to select for furfural resistance through evolution [148]. The evolved strain, EMFR9, had significantly increased furfural resistance. Reverse engineering efforts, including transcriptome analysis, attributed furfural resistance to the silencing expression of several oxidoreductases. These oxidoreductases use NADPH for furfural reduction, depleting the available pools for biosynthesis. Thus furfural-mediated growth inhibition can be attributed to NADPH depletion [148], an insight that can be applied to other biocatalyst design projects.

5. Perspectives

Although many biocatalysts have been successfully redesigned for production of industrially important fuels and chemicals through traditional metabolic engineering, we are just beginning to see the potential of synthetic biology in this area. One of the foremost goals in our lab is the improvement of biocatalysts for biomass utilization. To attain this goal, tolerance to hydrolysate-derived inhibitors needs to be improved. For all applications, tolerance to high substrate and product titers is also important. This goal of redesigning a biocatalyst's phenotype, that is, tolerance, is not as clear as redesigning metabolism and a rational redesign strategy is particularly difficult when the mechanism of inhibition is not known.

As the understanding of our biocatalysts improves, particularly through reverse engineering of evolved strains, genome-scale models can be improved. Inclusion of kinetic and regulatory effects will also improve the accuracy and predictive power of these models. Note that some models have recently been developed that bypass the need for kinetic data, though [152]. Since enzymes are the major functional part performing the metabolic synthesis, improved protein engineering tools and new protein catalytic capability will aid in advancement of this field. It is important to generate high-quality protein mutagenesis libraries (relatively small libraries with a high diversity of enzymes) to facilitate efficient screening efforts [138]. Direct screening from metagenomic libraries of environmental samples can aid in isolation of enzymes with new functions, which cannot be obtained by the traditional strain isolation methods [153]. Enzymes can even be synthesized from scratch by a rational design strategy with computational aid [154]. Finally, new tools for better *de novo* design of synthetic pathways need to be developed. Several databases, such as BNICE (Biochemical Network Integrated Computational Explorer) [155] and ReBiT (Retro-Biosynthesis Tool) [142], have already been established to facilitate identification of enzymes to construct a complete synthetic pathway for producing target compounds. It is important to establish guidelines, such as redox balance, energy production, and thermodynamic feasibility, to screen among these enormous pathways for the optimal routes.

By including synthetic biology tools in metabolic engineering projects, and vice versa, these two fields can significantly advance the replacement of petroleum-derived commodity products with those produced from biorenewable carbon and energy.

References

[1] L. R. Jarboe, T. B. Grabar, L. P. Yomano, K. T. Shanmugan, and L. O. Ingram, "Development of ethanologenic bacteria," *Advances in Biochemical Engineering/Biotechnology*, vol. 108, pp. 237–261, 2007.

[2] D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, et al., "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome," *Science*, vol. 319, no. 5867, pp. 1215–1220, 2008.

[3] C. Laitigue, S. Vashee, M. A. Algire, et al., "Creating bacterial strains from genomes that have been cloned and engineered in yeast," *Science*, vol. 325, no. 5948, pp. 1693–1696, 2009.
[4] J. E. Bailey, "Toward a science of metabolic engineering," *Science*, vol. 252, no. 5013, pp. 1668–1675, 1991.
[5] G. Stephanopoulos and J. J. Vallino, "Network rigidity and metabolic engineering in metabolite overproduction," *Science*, vol. 252, no. 5013, pp. 1675–1681, 1991.
[6] G. N. Stephanopoulos, A. A. Aristidou, and J. Nielsen, *Metabolic Engineering: Principles and Methodologies*, Academic Press, San Diego, Calif, USA, 1998.
[7] P. Ball, "Synthetic biology: starting from scratch," *Nature*, vol. 431, no. 7009, pp. 624–626, 2004.
[8] S. A. Benner and A. M. Sismour, "Synthetic biology," *Nature Reviews Genetics*, vol. 6, no. 7, pp. 533–543, 2005.
[9] R. McDaniel and R. Weiss, "Advances in synthetic biology: on the path from prototypes to applications," *Current Opinion in Biotechnology*, vol. 16, no. 4, pp. 476–483, 2005.
[10] J. Pleiss, "The promise of synthetic biology," *Applied Microbiology and Biotechnology*, vol. 73, no. 4, pp. 735–739, 2006.
[11] L. Serrano, "Synthetic biology: promises and challenges," *Molecular Systems Biology*, vol. 3, article 158, 2007.
[12] J. D. Keasling, "Synthetic biology for synthetic chemistry," *ACS Chemical Biology*, vol. 3, no. 1, pp. 64–76, 2008.
[13] A. C. Forster and G. M. Church, "Synthetic biology projects in vitro," *Genome Research*, vol. 17, no. 1, pp. 1–6, 2007.
[14] D. Greber and M. Fussenegger, "Mammalian synthetic biology: engineering of sophisticated gene networks," *Journal of Biotechnology*, vol. 130, no. 4, pp. 329–345, 2007.
[15] W. Weber and M. Fussenegger, "The impact of synthetic biology on drug discovery," *Drug Discovery Today*, vol. 14, no. 19-20, pp. 956–963, 2009.
[16] C. E. French, "Synthetic biology and biomass conversion: a match made in heaven?" *Journal of the Royal Society Interface*, vol. 6, supplement 4, pp. S547–S558, 2009.
[17] S. K. Lee, H. Chou, T. S. Ham, T. S. Lee, and J. D. Keasling, "Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels," *Current Opinion in Biotechnology*, vol. 19, no. 6, pp. 556–563, 2008.
[18] S. Picataggio, "Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals," *Current Opinion in Biotechnology*, vol. 20, no. 3, pp. 325–329, 2009.
[19] A. M. Feist, C. S. Henry, J. L. Reed, et al., "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information," *Molecular Systems Biology*, vol. 3, article 121, 2007.
[20] I. M. Keseler, C. Bonavides-Martínez, J. Collado-Vides, et al., "EcoCyc: a comprehensive view of *Escherichia coli* biology," *Nucleic Acids Research*, vol. 37, supplement 1, pp. D464–D470, 2009.
[21] D. Endy, "Foundations for engineering biology," *Nature*, vol. 438, no. 7067, pp. 449–453, 2005.
[22] T. B. Grabar, S. Zhou, K. T. Shanmugam, L. P. Yomano, and L. O. Ingram, "Methylglyoxal bypass identified as source of chiral contamination in L(+) and D(-)-lactate fermentations by recombinant *Escherichia coli*," *Biotechnology Letters*, vol. 28, no. 19, pp. 1527–1535, 2006.
[23] T. B. Causey, S. Zhou, K. T. Shanmugam, and L. O. Ingram, "Engineering the metabolism of *Escherichia coli* W3110 for the conversion of sugar to redox-neutral and oxidized products: homoacetate production," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 3, pp. 825–832, 2003.
[24] K. Jantama, X. Zhang, J. C. Moore, K. T. Shanmugam, S. A. Svoronos, and L. O. Ingram, "Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C," *Biotechnology and Bioengineering*, vol. 101, no. 5, pp. 881–893, 2008.
[25] L. P. Yomano, S. W. York, K. T. Shanmugam, and L. O. Ingram, "Deletion of methylglyoxal synthase gene (mgsA) increased sugar co-metabolism in ethanol-producing *Escherichia coli*," *Biotechnology Letters*, vol. 31, no. 9, pp. 1389–1398, 2009.
[26] P. C. Cirino, J. W. Chin, and L. O. Ingram, "Engineering *Escherichia coli* for xylitol production from glucose-xylose mixtures," *Biotechnology and Bioengineering*, vol. 95, no. 6, pp. 1167–1176, 2006.
[27] X. Zhang, K. Jantama, J. C. Moore, K. T. Shanmugam, and L. O. Ingram, "Production of L-alanine by metabolically engineered *Escherichia coli*," *Applied Microbiology and Biotechnology*, vol. 77, no. 2, pp. 355–366, 2007.
[28] R. S. Senger and E. T. Papoutsakis, "Genome-scale model for Clostridium acetobutylicum—part I: metabolic network resolution and analysis," *Biotechnology and Bioengineering*, vol. 101, no. 5, pp. 1036–1052, 2008.
[29] K. R. Kjeldsen and J. Nielsen, "In silico genome-scale reconstruction and validation of the corynebacterium glutamicum metabolic network," *Biotechnology and Bioengineering*, vol. 102, no. 2, pp. 583–597, 2009.
[30] J. Forster, I. Famili, P. Fu, et al., "Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network," *Genome Research*, vol. 13, no. 2, pp. 244–253, 2003.
[31] M. R. Andersen, M. L. Nielsen, and J. Nielsen, "Metabolic model integration of the bibliome, genome, metabolome and reactome of Aspergillus niger," *Molecular Systems Biology*, vol. 4, article 178, 2008.
[32] F. R. Blattner, et al., "The complete genome sequence of *Escherichia coli* K-12," *Science*, vol. 277, no. 5331, pp. 1453–1474, 1997.
[33] R. U. Ibarra, J. S. Edwards, and B. O. Palsson, "*Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth," *Nature*, vol. 420, no. 6912, pp. 186–189, 2002.
[34] A. P. Bauer, S. M. Dieckmann, W. Ludwig, and K.-H. Schleifer, "Rapid identification of *Escherichia coli* safety and laboratory strain lineages based on Multiplex-PCR," *FEMS Microbiology Letters*, vol. 269, no. 1, pp. 36–40, 2007.
[35] S. Zhou, L. P. Yomano, K. T. Shanmugam, and L. O. Ingram, "Fermentation of 10% (*w*/*v*) sugar to D: (−)-lactate by engineered *Escherichia coli B*," *Biotechnology Letters*, vol. 27, no. 23-24, pp. 1891–1896, 2005.
[36] M. Moniruzzaman, X. Lai, S. W. York, and L. O. Ingram, "Isolation and molecular characterization of high-performance cellobiose- fermenting spontaneous mutants of ethanologenic *Escherichia coli* KO11 containing the Klebsiella oxytoca casAB operon," *Applied and Environmental Microbiology*, vol. 63, no. 12, pp. 4633–4637, 1997.
[37] G. Posfai, G. Plunkett III, T. Fehér, et al., "Emergent properties of reduced-genome *Escherichia coli*," *Science*, vol. 312, no. 5776, pp. 1044–1046, 2006.
[38] M. Durot, P.-Y. Bourguignon, and V. Schachter, "Genome-scale models of bacterial metabolism: reconstruction and applications," *FEMS Microbiology Reviews*, vol. 33, no. 1, pp. 164–190, 2009.

[39] N. D. Price, J. A. Papin, C. H. Schilling, and B. O. Palsson, "Genome-scale microbial in silico models: the constraints-based approach," *Trends in Biotechnology*, vol. 21, no. 4, pp. 162–169, 2003.

[40] P. Pharkya, A. P. Burgard, and C. D. Maranas, "OptStrain: a computational framework for redesign of microbial production systems," *Genome Research*, vol. 14, no. 11, pp. 2367–2376, 2004.

[41] A. P. Burgard, P. Pharkya, and C. D. Maranas, "OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnology and Bioengineering*, vol. 84, no. 6, pp. 647–657, 2003.

[42] H. Alper, Y.-S. Jin, J. F. Moxley, and G. Stephanopoulos, "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*," *Metabolic Engineering*, vol. 7, no. 3, pp. 155–164, 2005.

[43] C. Bro and J. Nielsen, "Impact of 'ome' analyses on inverse metabolic engineering," *Metabolic Engineering*, vol. 6, no. 3, pp. 204–211, 2004.

[44] S. Y. Lee, D. Y. Lee, and T. Y. Kim, "Systems biotechnology for strain improvement," *Trends in Biotechnology*, vol. 23, no. 7, pp. 349–358, 2005.

[45] T. Hermann, "Using functional genomics to improve productivity in the manufacture of industrial biochemicals," *Current Opinion in Biotechnology*, vol. 15, no. 5, pp. 444–448, 2004.

[46] J. H. Choi, S. J. Lee, and S. Y. Lee, "Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling," *Applied and Environmental Microbiology*, vol. 69, no. 8, pp. 4737–4742, 2003.

[47] M.-J. Han, K. J. Jeong, J.-S. Yoo, and S. Y. Lee, "Engineering *Escherichia coli* for increased productivity of serine-rich proteins based on proteome profiling," *Applied and Environmental Microbiology*, vol. 69, no. 10, pp. 5772–5781, 2003.

[48] B. F. Pfleger, D. J. Pitera, C. D. Smolke, and J. D. Keasling, "Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes," *Nature Biotechnology*, vol. 24, no. 8, pp. 1027–1032, 2006.

[49] S. A. Underwood, M. L. Buszko, K. T. Shanmugam, and L. O. Ingram, "Flux through citrate synthase limits the growth of ethanologenic *Escherichia coli* KO11 during xylose fermentation," *Applied and Environmental Microbiology*, vol. 68, no. 3, pp. 1071–1081, 2002.

[50] M. Askenazi, E. M. Driggers, D. A. Holtzman, et al., "Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains," *Nature Biotechnology*, vol. 21, no. 2, pp. 150–156, 2003.

[51] J. Ohnishi, S. Mitsuhashi, M. Hayashi, et al., "A novel methodology employing Corynebacterium glutamicum genome information to generate a new L-lysine-producing mutant," *Applied Microbiology and Biotechnology*, vol. 58, no. 2, pp. 217–223, 2002.

[52] F. Martinez-Morales, A. C. Borges, A. Martinez, K. T. Shanmugam, and L. O. Ingram, "Chromosomal integration of heterologous DNA in *Escherichia coli* with precise removal of markers and replicons used during construction," *Journal of Bacteriology*, vol. 181, no. 22, pp. 7143–7148, 1999.

[53] K. A. Datsenko and B. L. Wanner, "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 12, pp. 6640–6645, 2000.

[54] G. Pósfai, M. D. Koob, H. A. Kirkpatrick, and F. R. Blattner, "Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the *Escherichia coli* O157:H7 genome," *Journal of Bacteriology*, vol. 179, no. 13, pp. 4426–4428, 1997.

[55] P. Gay, D. Le Coq, M. Steinmetz, et al., "Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria," *Journal of Bacteriology*, vol. 164, no. 2, pp. 918–921, 1985.

[56] G. de la Cueva-Mendez and B. Pimentel, "Gene and cell survival: lessons from prokaryotic plasmid R1," *The EMBO Reports*, vol. 8, no. 5, pp. 458–464, 2007.

[57] S. Zhou, F. C. Davis, and L. O. Ingram, "Gene integration and expression and extracellular secretion of Erwinia chrysanthemi endoglucanase CelY (celY) and CelZ (celZ) in ethanologenic Klebsiella oxytoca P2," *Applied and Environmental Microbiology*, vol. 67, no. 1, pp. 6–14, 2001.

[58] P. R. Jensen and K. Hammer, "The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters," *Applied and Environmental Microbiology*, vol. 64, no. 1, pp. 82–87, 1998.

[59] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos, "Tuning genetic control through promoter engineering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 36, pp. 12678–12683, 2005.

[60] I. Meynial-Salles, M. A. Cervin, and P. Soucaille, "New tool for metabolic pathway engineering in *Escherichia coli*: one-step method to modulate expression of chromosomal genes," *Applied and Environmental Microbiology*, vol. 71, no. 4, pp. 2140–2144, 2005.

[61] C. Solem and P. R. Jensen, "Modulation of gene expression made easy," *Applied and Environmental Microbiology*, vol. 68, no. 5, pp. 2397–2403, 2002.

[62] T. A. Carrier and J. D. Keasling, "Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*," *Biotechnology Progress*, vol. 15, no. 1, pp. 58–64, 1999.

[63] Y. S. Park, S. W. Seo, S. Hwang, et al., "Design of 5'-untranslated region variants for tunable expression in *Escherichia coli*," *Biochemical and Biophysical Research Communications*, vol. 356, no. 1, pp. 136–141, 2007.

[64] F. J. Isaacs, D. J. Dwyer, C. Ding, D. D. Pervouchine, C. R. Cantor, and J. J. Collins, "Engineered riboregulators enable post-transcriptional control of gene expression," *Nature Biotechnology*, vol. 22, no. 7, pp. 841–847, 2004.

[65] S. Jana and J. K. Deb, "Strategies for efficient production of heterologous proteins in *Escherichia coli*," *Applied Microbiology and Biotechnology*, vol. 67, no. 3, pp. 289–298, 2005.

[66] H. Alper, J. Moxley, E. Nevoigt, G. R. Fink, and G. Stephanopoulos, "Engineering yeast transcription machinery for improved ethanol tolerance and production," *Science*, vol. 314, no. 5805, pp. 1565–1568, 2006.

[67] H. Alper and G. Stephanopoulos, "Global transcription machinery engineering: a new approach for improving cellular phenotype," *Metabolic Engineering*, vol. 9, no. 3, pp. 258–267, 2007.

[68] I. P. Petrounia and F. H. Arnold, "Designed evolution of enzymatic properties," *Current Opinion in Biotechnology*, vol. 11, no. 4, pp. 325–330, 2000.

[69] H. Zhao, L. Giver, Z. Shao, J. A. Affholter, and F. H. Arnold, "Molecular evolution by staggered extension process (StEP)

in vitro recombination," *Nature Biotechnology*, vol. 16, no. 3, pp. 258–261, 1998.

[70] A. Crameri, S.-A. Raillard, E. Bermudez, and W. P. C. Stemmer, "DNA shuffling of a family of genes from diverse species accelerates directed evolution," *Nature*, vol. 391, no. 6664, pp. 288–291, 1998.

[71] C. Wang, M.-K. Oh, and J. C. Liao, "Directed evolution of metabolically engineered *Escherichia coli* for carotenoid production," *Biotechnology Progress*, vol. 16, no. 6, pp. 922–926, 2000.

[72] P. C. Cirino and F. H. Arnold, "Protein engineering of oxygenases for biocatalysis," *Current Opinion in Chemical Biology*, vol. 6, no. 2, pp. 130–135, 2002.

[73] H. H. Liao, "Alanine 2,3-aminomutase," US patent 7309597, 2007.

[74] M. M. Altamirano, J. M. Blackburn, C. Aguayo, and A. R. Fersht, "Directed evolution of new catalytic activity using the α/β-barrel scaffold," *Nature*, vol. 403, no. 6770, pp. 617–622, 2000.

[75] M. Leisola and O. Turunen, "Protein engineering: opportunities and challenges," *Applied Microbiology and Biotechnology*, vol. 75, no. 6, pp. 1225–1232, 2007.

[76] A. Korkegian, M. E. Black, D. Baker, and B. L. Stoddard, "Computational thermostabilization of an enzyme," *Science*, vol. 308, no. 5723, pp. 857–860, 2005.

[77] R. Machielsen, L. L. Looger, J. Raedts, et al., "Cofactor engineering of Lactobacillus brevis alcohol dehydrogenase by computational design," *Engineering in Life Sciences*, vol. 9, no. 1, pp. 38–44, 2009.

[78] J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan, and F. H. Arnold, "Evolving strategies for enzyme engineering," *Current Opinion in Structural Biology*, vol. 15, no. 4, pp. 447–452, 2005.

[79] R. A. Chica, N. Doucet, and J. N. Pelletier, "Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design," *Current Opinion in Biotechnology*, vol. 16, no. 4, pp. 378–384, 2005.

[80] O. Spadiut, I. Pisanelli, T. Maischberger, et al., "Engineering of pyranose 2-oxidase: improvement for biofuel cell and food applications through semi-rational protein design," *Journal of Biotechnology*, vol. 139, no. 3, pp. 250–257, 2009.

[81] Q. Wang, A. R. Parrish, and L. Wang, "Expanding the genetic code for biological studies," *Chemistry and Biology*, vol. 16, no. 3, pp. 323–336, 2009.

[82] Y. Lu, "Design and engineering of metalloproteins containing unnatural amino acids or non-native metal-containing cofactors," *Current Opinion in Chemical Biology*, vol. 9, no. 2, pp. 118–126, 2005.

[83] J. C. Jackson, S. P. Duffy, K. R. Hess, and R. A. Mehl, "Improving nature's enzyme active site with genetically encoded unnatural amino acids," *Journal of the American Chemical Society*, vol. 128, no. 34, pp. 11124–11127, 2006.

[84] D. B. Walker, G. Joshi, and A. P. Davis, "Progress in biomimetic carbohydrate recognition," *Cellular and Molecular Life Sciences*, vol. 66, no. 19, pp. 3177–3191, 2009.

[85] M. P. Cashion and T. E. Long, "Biomimetic design and performance of polymerizable lipids," *Accounts of Chemical Research*, vol. 42, no. 8, pp. 1016–1025, 2009.

[86] S. S. Fong, A. P. Burgard, C. D. Herring, et al., "In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid," *Biotechnology and Bioengineering*, vol. 91, no. 5, pp. 643–648, 2005.

[87] X. Zhang, K. Jantama, J. C. Moore, L. R. Jarboe, K. T. Shanmugam, and L. O. Ingram, "Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 48, pp. 20180–20185, 2010.

[88] J. G. Zeikus, M. K. Jain, and P. Elankovan, "Biotechnology of succinic acid production and markets for derived industrial products," *Applied Microbiology and Biotechnology*, vol. 51, no. 5, pp. 545–552, 1999.

[89] J. B. McKinlay, C. Vieille, and J. G. Zeikus, "Prospects for a bio-based succinate industry," *Applied Microbiology and Biotechnology*, vol. 76, no. 4, pp. 727–740, 2007.

[90] N. S. Samuelov, R. Lamed, S. Lowe, and I. G. Zeikus, "Influence of CO_2-HCO_3- levels and ph on growth, succinate production, and enzyme activities of *Anaerobiospirillum succiniciproducens*," *Applied and Environmental Microbiology*, vol. 57, no. 10, pp. 3013–3019, 1991.

[91] M. J. Vander Werf, M. V. Guettler, M. K. Jain, and J. G. Zeikus, "Environmental and physiological factors affecting the succinate product ratio during carbohydrate fermentation by Actinobacillus sp. 130Z," *Archives of Microbiology*, vol. 167, no. 6, pp. 332–342, 1997.

[92] P. Lee, S. Lee, S. Hong, and H. Chang, "Isolation and characterization of a new succinic acid-producing bacterium, Mannheimia succiniciproducens MBEL55E, from bovine rumen," *Applied Microbiology and Biotechnology*, vol. 58, no. 5, pp. 663–668, 2002.

[93] D. G. Fraenkel, "Glycolysis," in Escherichia coli *and Salmonella: Cellular and Molecular Biology*, F. C. Neidhardt, Ed., ASM Press, Washington, DC, USA, 1996.

[94] K. Jantama, M. J. Haupt, S. A. Svoronos, et al., "Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate," *Biotechnology and Bioengineering*, vol. 99, no. 5, pp. 1140–1153, 2008.

[95] M.-K. Oh, L. Rohlin, K. C. Kao, and J. C. Liao, "Global expression profiling of acetate-grown *Escherichia coli*," *Journal of Biological Chemistry*, vol. 277, no. 15, pp. 13175–13183, 2002.

[96] K. C. Kao, L. M. Tran, and J. C. Liao, "A global regulatory role of gluconeogenic genes in *Escherichia coli* revealed by transcriptome network analysis," *Journal of Biological Chemistry*, vol. 280, no. 43, pp. 36079–36087, 2005.

[97] R. Patnaik, W. D. Roof, R. F. Young, and J. C. Liao, "Stimulation of glucose catabolism in *Escherichia coli* by a potential futile cycle," *Journal of Bacteriology*, vol. 174, no. 23, pp. 7527–7532, 1992.

[98] X. Zhang, K. Jantama, K. T. Shanmugam, and L. O. Ingram, "Reengineering *Escherichia coli* for succinate production in mineral salts medium," *Applied and Environmental Microbiology*, vol. 75, no. 24, pp. 7807–7813, 2009.

[99] S. S. Ray and M. Bousmina, "Biodegradable polymers and their layered silicate nano composites: in greening the 21st century materials world," *Progress in Materials Science*, vol. 50, no. 8, pp. 962–1079, 2005.

[100] A. K. Agrawal and R. Bhalla, "Advances in the production of poly(lactic acid) fibers. A review," *Journal of Macromolecular Science-Polymer Reviews C*, vol. 43, no. 4, pp. 479–503, 2003.

[101] S. Benthin and J. Villadsen, "Production of optically pure D-lactate by lactobacillus bulgaricus and purification by crystallisation and liquid/liquid extraction," *Applied Microbiology and Biotechnology*, vol. 42, no. 6, pp. 826–829, 1995.

[102] A. Demirci and A. L. Pometto, "Enhanced production of D(-)-lactic acid by mutants of lactobacillus-delbrueckii

ATCC-9649," *Journal of Industrial Microbiology*, vol. 11, no. 1, pp. 23–28, 1992.

[103] D. G. Fraenkel, "Glycolysis," in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, F. C. Neidhardt, Ed., vol. 1, chapter 14, ASM Press, Washington DC, USA, 2nd edition, 1996.

[104] S. Zhou, T. B. Causey, A. Hasona, K. T. Shanmugam, and L. O. Ingram, "Production of optically pure D-lactic acid in mineral salts medium by metabolically engineered *Escherichia coli* W3110," *Applied and Environmental Microbiology*, vol. 69, no. 1, pp. 399–407, 2003.

[105] S. Zhou, K. T. Shanmugam, L. P. Yomano, T. B. Grabar, and L. O. Ingram, "Fermentation of 12% (*w*/*v*) glucose to 1.2 M lactate by *Escherichia coli* strain SZ194 using mineral salts medium," *Biotechnology Letters*, vol. 28, no. 9, pp. 663–670, 2006.

[106] M. Cheryan, S. Parekh, M. Shah, and K. Witjitra, "Production of acetic acid by Clostridium thermoaceticum," *Advances in Applied Microbiology*, vol. 43, pp. 1–33, 1997.

[107] C. Berraud, "Production of highly concentrated vinegar in fed-batch culture," *Biotechnology Letters*, vol. 22, no. 6, pp. 451–454, 2000.

[108] S. N. Freer, "Acetic acid production by Dekkera/Brettanomyces yeasts," *World Journal of Microbiology and Biotechnology*, vol. 18, no. 3, pp. 271–275, 2002.

[109] S. Y. Lee, J. H. Park, S. H. Jang, L. K. Nielsen, J. Kim, and K. S. Jung, "Fermentative butanol production by clostridia," *Biotechnology and Bioengineering*, vol. 101, no. 2, pp. 209–228, 2008.

[110] L. D. Mermelstein, E. T. Papoutsakis, D. J. Petersen, and G. N. Bennett, "Metabolic engineering of Clostridium acetobutylicum ATCC 824 for increased solvent production by enhancement of acetone formation enzyme activities using a synthetic acetone operon," *Biotechnology and Bioengineering*, vol. 42, no. 9, pp. 1053–1060, 1993.

[111] L. Harris, L. Blank, R. P. Desai, N. E. Welker, and E. T. Papoutsakis, "Fermentation characterization and flux analysis of recombinant strains of Clostridium acetobutylicum with an inactivated soIR gene," *Journal of Industrial Microbiology and Biotechnology*, vol. 27, no. 5, pp. 322–328, 2001.

[112] R. V. Nair, E. M. Green, D. E. Watson, G. N. Bennett, and E. T. Papoutsakis, "Regulation of the sol locus genes for butanol and acetone formation in Clostridium acetobutylicum ATCC 824 by a putative transcriptional repressor," *Journal of Bacteriology*, vol. 181, no. 1, pp. 319–330, 1999.

[113] R. P. Desai and E. T. Papoutsakis, "Antisense RNA strategies for metabolic engineering of Clostridium acetobutylicum," *Applied and Environmental Microbiology*, vol. 65, no. 3, pp. 936–945, 1999.

[114] N. E. Altaras and D. C. Cameron, "Enhanced production of (R)-1,2-propanediol by metabolically engineered *Escherichia coli*," *Biotechnology Progress*, vol. 16, no. 6, pp. 940–946, 2000.

[115] P. Soucaille, I. Meynial-Salles, F. Voelker, and R. Figge, "Microorganisms and methods for prodcution of 1,2-propanediol and acetol," WO, 2008, 2008/116853.

[116] J. H. Park, K. H. Lee, T. Y. Kim, and S. Y. Lee, "Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 19, pp. 7797–7802, 2007.

[117] K. H. Lee, J. H. Park, T. Y. Kim, H. U. Kim, and S. Y. Lee, "Systems metabolic engineering of *Escherichia coli* for L-threonine production," *Molecular Systems Biology*, vol. 3, article 149, 2007.

[118] L. O. Ingram, P. F. Gomez, X. Lai, et al., "Metabolic engineering of bacteria for ethanol production," *Biotechnology and Bioengineering*, vol. 58, no. 2-3, pp. 204–214, 1998.

[119] Y. Kim, L. O. Ingram, and K. T. Shanmugam, "Construction of an *Escherichia coli* K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes," *Applied and Environmental Microbiology*, vol. 73, no. 6, pp. 1766–1771, 2007.

[120] L. O. Ingram, T. Conway, D. P. Clark, G. W. Sewell, and J. F. Preston, "Genetic engineering of ethanol production in Escherichia coli," *Applied and Environmental Microbiology*, vol. 53, no. 10, pp. 2420–2425, 1987.

[121] K. Ohta, D. S. Beall, J. P. Mejia, K. T. Shanmugam, and L. O. Ingram, "Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of Zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase II," *Applied and Environmental Microbiology*, vol. 57, no. 4, pp. 893–900, 1991.

[122] L. P. Yomano, S. W. York, S. Zhou, K. T. Shanmugam, and L. O. Ingram, "Re-engineering *Escherichia coli* for ethanol production," *Biotechnology Letters*, vol. 30, no. 12, pp. 2097–2103, 2008.

[123] A. Hasona, S. W. York, L. P. Yomano, L. O. Ingram, and K. T. Shanmugam, "Decreasing the level of ethyl acetate in ethanolic fermentation broths of *Escherichia coli* KO11 by expression of Pseudomonas putida estZ esterase," *Applied and Environmental Microbiology*, vol. 68, no. 6, pp. 2651–2659, 2002.

[124] K. Hofvendahl and B. Hahn-Hagerdal, "Factors affecting the fermentative lactic acid production from renewable resources," *Enzyme and Microbial Technology*, vol. 26, no. 2–4, pp. 87–107, 2000.

[125] S. D. Zhou, K. T. Shanmugam, and L. O. Ingram, "Functional replacement of the *Escherichia coli* D-(-)-lactate dehydrogenase gene (ldhA) with the L-(+)-lactate dehydrogenase gene (ldhL) from Pediococcus acidilactici," *Applied and Environmental Microbiology*, vol. 69, no. 4, pp. 2237–2244, 2003.

[126] J. C. Parajó, H. Domínguez, and J. M. Domínguez, "Biotechnological production of xylitol—part 1: interest of xylitol and fundamentals of its biosynthesis," *Bioresource Technology*, vol. 65, no. 3, pp. 191–201, 1998.

[127] T. Werpy and G. Petersen, Eds., *Top Value Added Chemicals from Biomass: Volume I—Results of Screening for Potential Candidates from Sugars and Synthesis Gas*, Pacific Northwest National Laboratory and National Renewable Energy Laboratory, 2004.

[128] T. B. Kim and D. K. Oh, "Xylitol production by Candida tropicalis in a chemically defined medium," *Biotechnology Letters*, vol. 25, no. 24, pp. 2085–2088, 2003.

[129] J. W. Chin, R. Khankal, C. A. Monroe, C. D. Maranas, and P. C. Cirino, "Analysis of NADPH supply during xylitol production by engineered *Escherichia coli*," *Biotechnology and Bioengineering*, vol. 102, no. 1, pp. 209–220, 2009.

[130] P. Hols, M. Kleerebezem, A. N. Schanck, et al., "Conversion of Lactococcus lactis from homolactic to homoalanine fermentation through metabolic engineering," *Nature Biotechnology*, vol. 17, no. 6, pp. 588–592, 1999.

[131] M. Ikeda, "Amino acid production processes," *Advances in Biochemical Engineering/Biotechnology*, vol. 79, pp. 1–35, 2003.

[132] C. E. Nakamura and G. M. Whited, "Metabolic engineering for the microbial production of 1,3-propanediol," *Current Opinion in Biotechnology*, vol. 14, no. 5, pp. 454–459, 2003.

[133] T. Hanai, S. Atsumi, and J. C. Liao, "Engineered synthetic pathway for isopropanol production in *Escherichia coli*," *Applied and Environmental Microbiology*, vol. 73, no. 24, pp. 7814–7818, 2007.

[134] D.-K. Ro, E. M. Paradise, M. Quellet, et al., "Production of the antimalarial drug precursor artemisinic acid in engineered yeast," *Nature*, vol. 440, no. 7086, pp. 940–943, 2006.

[135] Y. Yan, A. Kohli, and M. A. G. Koffas, "Biosynthesis of natural flavanones in Saccharomyces cerevisiae," *Applied and Environmental Microbiology*, vol. 71, no. 9, pp. 5610–5613, 2005.

[136] E. Leonard, Y. Yan, and M. A. G. Koffas, "Functional expression of a P450 flavonoid hydroxylase for the biosynthesis of plant-specific hydroxylated flavonols in *Escherichia coli*," *Metabolic Engineering*, vol. 8, no. 2, pp. 172–181, 2006.

[137] E. Leonard, K. H. Lim, P.-N. Saw, and M. A. G. Koffas, "Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*," *Applied and Environmental Microbiology*, vol. 73, no. 12, pp. 3877–3886, 2007.

[138] A. L. de Boer and C. Schmidt-Dannert, "Recent efforts in engineering microbial cells to produce new chemical compounds," *Current Opinion in Chemical Biology*, vol. 7, no. 2, pp. 273–278, 2003.

[139] C. Schmidt-Dannert, D. Umeno, and F. H. Arnold, "Molecular breeding of carotenoid biosynthetic pathways," *Nature Biotechnology*, vol. 18, no. 7, pp. 750–753, 2000.

[140] G. Sandmann, "Combinatorial biosynthesis of carotenoids in a heterologous host: a powerful approach for the biosynthesis of novel structures," *ChemBioChem*, vol. 3, no. 7, pp. 629–635, 2002.

[141] M. Albrecht, S. Takaichi, S. Steiger, Z.-Y. Wang, and G. Sandmann, "Novel hydroxycarotenoids with improved antioxidative properties produced by gene combination in *Escherichia coli*," *Nature Biotechnology*, vol. 18, no. 8, pp. 843–846, 2000.

[142] K. L. J. Prather and C. H. Martin, "De novo biosynthetic pathways: rational design of microbial chemical factories," *Current Opinion in Biotechnology*, vol. 19, no. 5, pp. 468–474, 2008.

[143] S. Atsumi, T. Hanai, and J. C. Liao, "Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels," *Nature*, vol. 451, no. 7174, pp. 86–89, 2008.

[144] K. Zhang, M. R. Sawaya, D. S. Eisenberg, and J. C. Liao, "Expanding metabolism for biosynthesis of nonnatural alcohols," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 52, pp. 20653–20658, 2008.

[145] R. Gonzalez, H. Tao, J. E. Purvis, S. W. York, K. T. Shanmugam, and L. O. Ingram, "Gene array-based identification of changes that contribute to ethanol tolerance in ethanologenic *Escherichia coli*: comparison of KO11 (parent) to LY01 (resistant mutant)," *Biotechnology Progress*, vol. 19, no. 2, pp. 612–623, 2003.

[146] L. P. Yomano, S. W. York, and L. O. Ingram, "Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* KO11 for fuel ethanol production," *Journal of Industrial Microbiology and Biotechnology*, vol. 20, no. 2, pp. 132–138, 1998.

[147] M. P. Brynildsen and J. C. Liao, "An integrated network approach identifies the isobutanol response network of *Escherichia coli*," *Molecular Systems Biology*, vol. 5, article 277, 2009.

[148] E. N. Miller, et al., "Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic *Escherichia coli*," *Applied and Environmental Microbiology*, vol. 75, no. 13, pp. 4315–4323, 2009.

[149] A. Martinez, M. E. Rodriguez, M. L. Wells, S. W. York, J. F. Preston, and L. O. Ingram, "Detoxification of dilute acid hydrolysates of lignocellulose with lime," *Biotechnology Progress*, vol. 17, no. 2, pp. 287–293, 2001.

[150] R. Gonzalez, H. Tao, K. T. Shanmugam, S.W. York, and L. O. Ingram, "Transcriptome analysis of ethanologenic *Escherichia coli* strains: tolerance to ethanol," in *Proceedings of the 225th ACS National Meeting*, p. U200, New Orleans, La, USA, March 2003.

[151] E. N. Miller, L. R. Jarboe, P. C. Turner, et al., "Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180," *Applied and Environmental Microbiology*, vol. 75, no. 19, pp. 6132–6141, 2009.

[152] L. M. Tran, M. L. Rizk, and J. C. Liao, "Ensemble modeling of metabolic networks," *Biophysical Journal*, vol. 95, no. 12, pp. 5606–5617, 2008.

[153] P. D. Schloss and J. Handelsman, "Biotechnological prospects from metagenomics," *Current Opinion in Biotechnology*, vol. 14, no. 3, pp. 303–310, 2003.

[154] L. Jiang, E. A. Althoff, F. R. Clemente, et al., "De novo computational design of retro-aldol enzymes," *Science*, vol. 319, no. 5868, pp. 1387–1391, 2008.

[155] V. Hatzimanikatis, C. Li, J. A. Ionita, C. S. Henry, M. D. Jankowski, and L. Broadbelt, "Exploring the diversity of complex metabolic networks," *Bioinformatics*, vol. 21, no. 8, pp. 1603–1609, 2005.

Hindawi Publishing Corporation
Journal of Biomedicine and Biotechnology
Volume 2010, Article ID 459760, 10 pages
doi:10.1155/2010/459760

Review Article

Toward Engineering Synthetic Microbial Metabolism

George H. McArthur IV and Stephen S. Fong

Department of Chemical and Life Science Engineering, Virginia Commonwealth University, 601 W. Main Street, Richmond, VA 23284-3028, USA

Correspondence should be addressed to Stephen S. Fong, ssfong@vcu.edu

Received 15 August 2009; Accepted 9 October 2009

Academic Editor: Hal Alper

The generation of well-characterized parts and the formulation of biological design principles in synthetic biology are laying the foundation for more complex and advanced microbial metabolic engineering. Improvements in *de novo* DNA synthesis and codon-optimization alone are already contributing to the manufacturing of pathway enzymes with improved or novel function. Further development of analytical and computer-aided design tools should accelerate the forward engineering of precisely regulated synthetic pathways by providing a standard framework for the predictable design of biological systems from well-characterized parts. In this review we discuss the current state of synthetic biology within a four-stage framework (design, modeling, synthesis, analysis) and highlight areas requiring further advancement to facilitate true engineering of synthetic microbial metabolism.

1. Introduction

As with its sister field of systems biology, synthetic biology is perhaps best described not by *what* you do, but *how* you do it. From this perspective, one way of summarizing synthetic biology is by its intended goal of making biological systems explicitly tractable through careful modularization of biology [1]. In this vein, the core emphases in synthetic biology are synthesis, abstraction, and standardization of biological components [2]. Synthesis refers to the generation of modular, reusable biological parts (normally as DNA). Abstraction refers to the implementation of synthesized parts by function rather than composition (e.g., "this is a promoter" rather than "this is ttgacagctagctcagtcctaggtataatgctagc") and can facilitate the design of more complex biological devices and systems. One of the hallmarks of this approach is the push for standardization where all synthesized parts and devices are characterized, reproducible, and interchangeable.

Although nascent, the approach employed in synthetic biology is already making significant contributions in areas within biological engineering and basic biological research by providing a rigorous framework for building biological systems from the ground-up [3]. Progress in synthetic biology is providing the concepts and tools to develop biological engineering as an application of biology, just as chemical engineering is an application of chemistry. Although the scope of applied synthetic biology is wide, including areas that are "beyond the bioreactor" such as bioremediation, agriculture, and human health [4], it is apparent that many of the applications of synthetic biology that will materialize in the near future are in metabolic engineering [5]. In this context, the synthetic biology approach aims to generate scalable, reusable genetic parts for controlling genetically encoded unit operations (e.g., enzymes, transporters), for the predictable design and construction of metabolic pathways [6].

Since its inception, made possible by the advent of recombinant DNA technology and the polymerase chain reaction (PCR), the goal of metabolic engineering has been to optimize cellular metabolism for a particular process application—often through extensive genetic modification. While research in metabolic engineering has continued to advance with a growing number of genetic, analytical, and computational tools [7], the application of synthetic biology to metabolic engineering has the potential to create a paradigm shift [8]. Rather than starting with the full complement of components in a wild-type organism and piecewise modifying and streamlining its function, metabolic

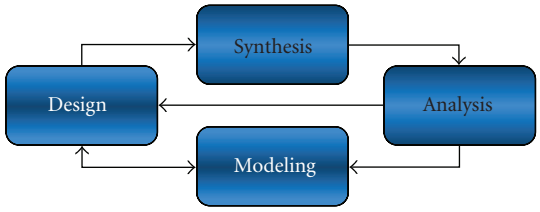


Figure 1: *Synthetic biology four-step engineering workflow.* The design and modeling steps (white text) are computational in nature, whereas the synthesis and analysis steps (black text) are experimentally-based. The workflow presents the relationships between the individual steps as well as a logical progression using a synthetic biology framework to generate a product that meets determined specifications.

engineering can be attempted from a bottom-up, parts-based approach to design by carefully and rationally specifying the inclusion of each necessary component. In this way, the metabolic engineering process becomes uniform for any goal and eliminates laborious cycles of trial and error.

In this review, we discuss synthetic biology as it applies specifically to metabolic engineering applications. For the purposes of this discussion, we are presenting the typical workflow of synthetic biology within a four step framework (Figure 1). As depicted in the flow diagram, the typical stages involved are: design, modeling, synthesis, and analysis. The workflow is often highly cyclical, although progress is being made on moving from iterative biological design to more predictable, directed design. In the following sections, we discuss recent progress and current challenges in applying synthetic biology to metabolic engineering for each of the four workflow steps (design, modeling, synthesis, and analysis) in general and specifically at the level of genetic parts and metabolic pathways. While synthetic biology applies to all levels of biological organization, we are delineating parts (individual functional units such as promoters, ribosomal binding sites, protein coding sequences, and terminators) from pathways (part-based systems) as there are often different considerations or methodologies employed at these levels.

2. Design

Imagine an architect preparing blueprints for someone's dream house. In the design process, the architect will largely focus on including all of the desired functional and aesthetic components but there will also need to be some consideration given to standard safety and logistical requirements. A draft blueprint can be generated to best satisfy the functional and logistical requirements and after the design is approved, the materials needed for construction can be purchased. In this scenario, the design and construction (synthesis) phases are decoupled due to the availability of standard, readily available building materials. Currently, the design and synthesis of biological systems are not decoupled. For example, the construction of metabolic pathways from genetic parts first requires a collection of well-characterized parts, which do not yet exist. This shortcoming is beginning to be addressed by the synthetic biology community through the development and compilation of standard biological building blocks such as those found in the Registry of Standard Biological Parts. Although virtually nonexistent now, the future availability of modular, functionally interoperable genetic parts that are carefully characterized will allow biological design to focus more exclusively on functional, aesthetic, and logistical concerns independent of construction concerns.

2.1. Parts. In designing individual biological parts, the base-by-base content of that part (promoter, ribosome-binding site (RBS), protein coding region, terminator, etc.) is explicitly dictated. Rules and guidelines for designing genetic parts at this level are just beginning to be established [9]. In particular, a major consideration in designing protein-coding parts is codon optimization, encoding the same amino acid sequence with an alternative, preferred nucleotide sequence. Although a particular sequence, when expressed, may be theoretically functional, its expression may be far from optimal or even completely suppressed due to codon usage bias in the heterologous host [10]. Although the design principles for expressing exogenous genes have not been perfected, existing codon optimization strategies are used routinely to enhance pathway performance. Codon optimization of coding sequences can be achieved using freely available algorithms such as Gene Designer, stand-alone software made by DNA2.0. In addition to codon optimization, other considerations such as compliance with standard assembly requirements and part-specific objectives including activity or specificity modifications are often considered. For example, the BioBrick methodology requires that BioBrick parts exclude four standard restriction enzyme sites, which are reserved for use in assembly [11].

Extensive collections of parts can be generated by using a naturally occurring part as a template and rationally modifying it to create a library of that particular genetic part. Significant progress in this area has been recently demonstrated for promoters and RBSs [12, 13]. The work by Ellis et al. documents the generation and characterization of two promoter libraries that can be used to tune network behavior *a priori* by fitting mathematical promoter models with measured parameters [12]. This model-guided design approach limits system variability, increasing predictability and decreasing the time spent on combinatorial system construction, testing and debugging. Noisy or leaky promoters can complicate system design by adding variability. In these cases, finer control over expression can be established by weakening the binding strength of the downstream RBS [14] or by using two promoter inputs to drive transcription of an output via a modular AND gate [15].

In addition to increasing the number of parts within *existing* families (e.g., constitutive promoters), there is an opportunity to develop entirely *new* parts families such as those used to build RNA devices that can regulate expression posttranscriptionally [16]. Win and Smolke developed RNA

devices that incorporate RNA aptamers for sensing small molecules, transmitter sequences that transmit information by way of a conformational change, and actuators such as hammerhead ribozymes that may render the mRNA untranslatable depending on the state of the sensor and transmitter [17]. The benefit of these devices is that they are composed of modular parts and are completely scalable, since RNA aptamers can theoretically be designed to sense any small molecule ligand [18].

2.2. Pathways. Rational pathway design can be thought of as mixing and matching well-known, modular parts and modulating gene expression through various control mechanisms. One of the primary motivations behind the attempt to achieve fine control of gene expression in the context of metabolic engineering is simply to balance and optimize flux through a desired pathway [19]. Thus, design at the pathway-level is not only concerned with including the necessary parts, but also with controlling the expressed functionality of those parts (e.g., a specific gene may be used in different pathways, the level of required expression may differ based upon the context).

Parts-based synthetic metabolic pathways will require tunable control, just as their natural counterparts which often employ feedback and feedforward motifs to achieve complex regulation [20]. The temporal and spatial dynamics necessary for life (e.g., diffusion, evolution, growth/death) make it difficult to achieve this goal. For example, biological signals must diffuse throughout the cell until it is received by the signal's receptor, unlike an electrical signal that is insulated and directed along a path. Therefore, information transfer in biological systems is intrinsically noisy, an issue that has been reviewed and continues to be explored [21, 22]. This is the case for gene regulation at the transcription level in which there is a DNA-protein interaction between regulatory proteins such as the LacI repressor and its operator site, lacO. Although a handful of operator-regulator pairs have been used successfully to construct simple genetic circuits [23, 24], more complex networks will likely require a large repetoire of distinct regulatory elements to eliminate molecular cross-talk. Developing regulatory parts such as promoters, operator-regulator pairs, ribosome binding sites (RBSs), and riboswitches is a proven way to tackle this problem [25]. Table 1 lists a sampling of regulatory parts for control of various biological processes involved in gene expression.

Another approach, based on well-studied natural systems [26], was recently demonstrated by Dueber et al. to complement these methods. This elegant approach uses protein scaffolds that bear modular interaction domains to physically link pathway enzymes that have been tagged with the appropriate corresponding peptide ligands [27]. Not only does enzyme colocalization limit the loss of intermediates to competing pathways, it also enables the direct control of metabolic flux by adjusting the number of interaction domains on the scaffold thereby adjusting the enzyme complex composition.

Using a synthetic biology approach, the design of DNA sequences encoding metabolic pathways (e.g., operons) should be relatively straightforward. However, the prerequisite collection of modular genetic parts has not yet been developed, but progress has been made [28]. Synthetic scaffolds and well-characterized families of regulatory parts have emerged as powerful tools for engineering metabolism by providing rational methodologies for coordinating control of multigene expression as well as decoupling pathway design from construction. This approach begins to solve the "impedance problem" by providing component variants with covering a range of input/output transfer functions, allowing pathway tenability [29]. Pathway design should not overlook the fact that exogenous pathways interact with native cellular components and have their own specific energy requirements. Therefore, modifying endogenous gene expression (e.g., gene knockouts, knockdowns, overexpression) may be necessary in addition to balancing cofactor fluxes (e.g., ATP, NADH) and installing membrane transporters [30]. Figure 2 illustrates how well-characterized parts can be composited to form functional devices or modules that can then be used to build complex systems, including metabolic pathways.

3. Synthesis

Just as cars are routinely built from modular components in assembly lines, rapid, modularized construction of microbial chemical factories is on the horizon. Enabling this manufacturing process is DNA synthesis technology, the chemical fabrication of gene- and even genome-length DNA molecules. For example, putative genes identified computationally in a metagenomics study need not be isolated and cloned; the sequence information can be outsourced to synthesis companies as Bayer et al. recently reported [31]. However, protocols exist to synthesize DNA in laboratories using common equipment and techniques [32]. This capability opens many opportunities for metabolic engineers. First of all, the convenience of this approach over traditional cloning allows for the systematic generation of genetic part variants such as promoter libraries (see Section 2). Secondly, it provides a practical way for eliminating restriction sites or undesirable RNA secondary structures and to codon optimize genes for the expression in heterologous hosts. Thirdly and perhaps most importantly, the generation of biological building blocks will allow for the decoupling of construction from design.

3.1. Parts. Generating genetic parts is as easy as emailing the part's nucleotide sequence to a DNA synthesis company such as DNA2.0, GENEART, or Genscript. Depending on the length and complexity of the part, it could take as little as a week to synthesize and ship. Although synthesis technology is available for individual laboratories, the time and money required for equipment upkeep may not make investment worthwhile, especially with commercial synthesis prices steadily dropping. Still, the cost of DNA synthesis is relatively high and can be a prohibitive factor for some projects. It is anticipated that DNA synthesis technology

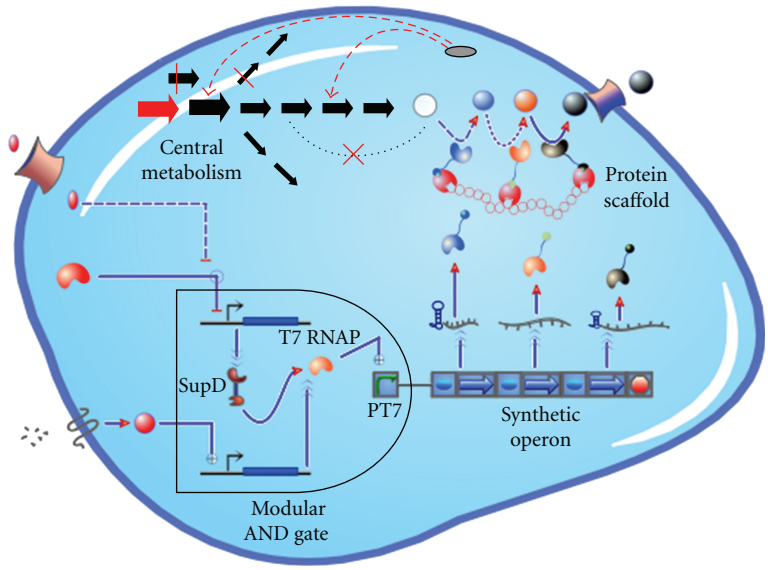


Figure 2: *Cartoon of an engineered cell.* Four distinct strategies have been combined to illustrate a synthetic biology approach to metabolic engineering. The engineered pathway is encoded by a *synthetic operon* designed from standard parts/components including tunable intergenic regions (TIGRs) that provide differential control over mRNA stability [41]. An upstream device/module, a transcriptional *modular AND gate* acts as a master regulator by requiring two separate inputs to turn on expression of the pathway [15]. The enzymes have been tagged with a peptide ligand that binds them to a synthetic *protein scaffold*, forming a complex of colocalized pathway enzymes [27]. The *central metabolism* of the host/chassis has been modified to interface with the engineered system/network, providing a sufficient flux of necessary precursors and cofactors as well as increased transporters to facilitate product secretion [7]. This image was generated in TinkerCell.

will follow in the footsteps of DNA sequencing technology where research and improvements have decreased costs by orders of magnitude. Czar et al. provide an excellent review of synthesis technologies and strategies, including those that enable in-house synthesis [33].

3.2. Pathways. The ability to make large changes to DNA molecules has resulted in standardized methods for assembling basic genetic parts into larger composite devices, which facilitate part-sharing and faster system-level construction. For example, the Registry of Standard Biological Parts supports the use of the BioBrick methodology to assemble parts into devices and devices into systems. The restriction sites flank BioBrick parts (two serve as the prefix and the other two as the suffix) allowing for restriction-ligation all the while maintaining the same restriction sites. A standardized assembly method, 3A BioBrick assembly, enables a three-way ligation of two insert parts into a plasmid backbone and eliminates the need to gel-purify digested parts and plasmid backbones [11]. This approach can be automated using programmable liquid-handling robots, providing a scalable assembly process that can be used to quickly build systems from existing genetic components. In this respect, high-throughput pathway assembly is more easily implemented in individual laboratories than *de novo* DNA part synthesis, but there remains the option to outsource this aspect of synthesis through commercial avenues as well.

Other approaches based on type II restriction enzymes, such as Golden Gate Shuffling, provide ways to assemble many more components together in one step [45]. A similar one-step assembly approach, circular polymerase extension cloning (CPEC), avoids the need for restriction-ligation or single-stranded homologous recombination altogether. This scalable, sequence-independent cloning technology works by extending overlapping sequences between the part and the plasmid backbone (e.g., BioBrick prefix and suffix), just as polymerase extension occurs in PCR, resulting in a complete circular plasmid [46]. Not only is this useful for cloning single genes, but also for assembling parts into a larger sequences encoding entire metabolic pathways and for generating combinatorial part libraries.

Multiplex automated genome engineering (MAGE) was recently introduced as another scalable, combinatorial method for producing large-scale genomic diversity [47]. This approach makes chromosomal modification easier by simultaneously mutating target sites across the chromosome. In this method, a pool of degenerate oligonucleotides that target specific genomic sites is cyclically introduced into various cells within the population, which generates diversity across the population. MAGE was used to modify twenty four genetic parts in the DXP pathway, including ribosomal binding sites and enzyme coding sequences for tuning expression levels and knocking out gene function. This strategy resulted in a fivefold increase in lycopene production in only three days, demonstrating a practical alternative to encoding pathways on plasmids that often require many iterations of testing and optimization.

Plasmid-based expression and chromosomal integration are the two common vehicles for implementing synthetic metabolic pathways and each method has its advantages, described in an excellent review on gene-expression tools

Table 1: Compilation of biological controllers.

Biological process	Genetic part or mechanism
Replication	Replication origin (copy number) [34]
Transcription	
Initation	RNA polymerase [35], transcription factors [36], specificity factors [37], promoters [38], enhancers, operator-regulator pairs [39], methylation
Elongation	RNA polymerase [35], insulators, terminators [40], attenuators
Processing	Splice sites, polyA sites, TIGRs [41]
Posttranscription	
Degradation	TIGRs [41], stabilizers, RNase sites [42]
Translation	
Initiation	Ribosome-RBS pairs [43], Riboregulators, Riboswitches, Ribozymes, Antisense RNA, MicroRNA, Small interfering RNA. For a review on RNA parts see [44].
Elongation	Codon bias, rare codons, stop codon-supressors pairs [15]
Processing	Chaperones
Posttranslational	
Degradation	Proteases
Localization	Target sequences, anchor sequences, interaction domain-peptide ligand pairs [27]

[48]. Recently, a new long-term expression method called chemically inducible chromosomal evolution (CIChE) was presented as an alternative to existing methods [49]. CIChE avoids complications associated with plasmid replication and segregation, and can be used to integrate multiple copies of genes (up to 40 copies demonstrated) into the genome. These techniques, coupled with in vivo strategies for assembling genetic parts into systems encoding metabolic pathways [50–52] and large-scale (several hundred kilobases) in vitro assembly methods [53], will provide technical platforms for the rapid synthesis of parts and subsequent pathways. On a larger scale, bacterial genome transplantation and the chemical synthesis of a natural bacterial genome have been demonstrated [54, 55].

4. Analysis

Cells are not unlike computers in that they process input information to realize output functions. A cell's genome can therefore be thought of as an operating system (OS) and, similarly, engineered operons encoding metabolic pathways as individual programs that are built from well-characterized, modular parts (the equivalent of a function in computer science terms). A twist on this analogy is that there already exists a large pool of parts and pathways in natural and synthetic biological systems and we are collectively learning their biological function through biochemical and bioinformatic analyses. Therefore, before metabolic pathways can be rationally designed and constructed, the genetic parts from which the pathways are built must be carefully characterized.

4.1. Parts. With the capability to synthesize biological parts, it is possible to use molecular-level controls to study and analyze each part [56]. One of the major stumbling blocks for the development of well-characertized parts is figuring out exactly what parameters should be measured and how. Although made from five basic parts, the well-characterized BioBrick signaling device BBa_F2620 developed by Canton et al. provided the first example of what can be measured and a possible standard documentation format—a datasheet (see Canton et al. [57] for a detailed example of a datasheet). The device's performance characteristics, including input compatibility, reliability, and transcriptional output demand, are neatly summarized. Datasheets may facilitate rapid widespread adoption of synthetic biological parts, devices, and systems if they include relevant information that is available to the community [58].

Recent work toward characterizing promoters demonstrates the utility of measuring the relative strengths of promoters in a library as they are built [12]. A separate effort by Kelly et al. has proposed a promoter reference standard, a fluoresense-dependent measurement kit and a measurement unit used to report promoter activity as an attempt to normalize absolute measurements taken by various instruments under various conditions [59]. The conceptual framework developed for promoter measurement could easily be applied to other genetic parts such as RBSs and transcriptional terminators. However, these tools will only be beneficial if they are supported and adopted by the community.

4.2. Pathways. Although well-characterized parts will drastically improve metabolic engineering efforts, there will always be a need for debugging tools and troubleshooting methods. Systems biology provides a framework for analyzing metabolic pathway performance and identifying possible solutions. Integrating high-throughput global measurements (i.e., transcriptomics, proteomics, metabolomics, and fluxomics) and computational tools (see Section 5) offers unprecedented understanding of how engineered pathways affect cell physiology and vice versa across the entire cell [60]. This kind of analysis can be used to identify bottlenecks in the

pathway as well as detrimental side effects that may not have been obvious at the initial design stage [61].

For example, Kizer et al. used transciptional and metabolite profiling to determine the biochemical interactions of an engineered pathway and the endogenous metabolic network [62]. In this work, they were not only able to identify the bottleneck that resulted in toxic accumulation of a pathway intermediate, but also the mechanism by which the accumulated metabolite inhibits cellular growth—and a way to overcome the problem. Using a system biology analytical approach, metabolic engineers are able to intelligently modulate synthetic pathway behavior if the measured performance does not meet specifications. Furthermore, the synthetic biology framework should enable the rational forward-engineering of pathways in a more plug-and-play fashion and significantly decrease the time spent on debugging the genetic programs used to encode metabolic function.

5. Modeling

Highly complex electrical circuits are able to be designed with ease because of the existence of computer-aided design (CAD) software. This technology is dependent on the high-fidelity virtual representation of physical circuit componentry, which necessitates an accurate mathematical model of each component (e.g., transistors, capacitors) based on a deep understanding of the physics involved. Once the fundamental physics of genetic part behavior is understood, then parts can be designed to be insulated from each other. Subsequently, abstract design based on standard biological models would be possible, allowing CAD of intricate metabolic pathways and regulatory networks. While this is the ideal, there are a variety of modeling approaches being applied to biology that have various strengths and weaknesses leaving a consolidated, standard modeling approach something to strive for in the future.

5.1. Parts. At the part-level, it is possible to develop highly detailed, quantitative, dynamic models when considering a single or small number of parts. The software tool ProMoT was developed for modeling parts with ordinary differential equations (ODEs) and enabling interfacing by quantifying the exchange of biological signals (i.e., RNA polymerases, ribosomes, transcription factors and environmental stimuli) between parts [63]. In a similar spirit, SBML- and CellML-based models of genetic parts and well-characterized modules have been proposed for establishing a registry for standard biological models [64, 65]. Standard, quantitative models of genetic parts would allow biological engineers to seamlessly integrate the design process with modeling, enabling simulation of engineered systems prior to actual construction (see design/model loop in Figure 1).

5.2. Pathways. As the scale of the system to be modeled increases from a single pathway to all pathways within an organism, more attention must be given to selecting a modeling approach as the quality of simulation results can vary simply due to the practical limitations imposed by the increasing number of variables. It is anticipated that as synthetic biology and modeling methods progress, the line between design and modeling will be blurred. This will condense the workflow into the common design-build-test engineering cycle. A variety of CAD tools are emerging to make this happen [66, 67]. Other computational tools are being developed to make novel pathway design a high-throughput process. From Metabolite to Metabolite (FMM) is a web-based tool that constructs a variety of possible enzymatic pathways from an input metabolite to an output metabolite, both of which are provided by the user [68]. This saves a large amount of time spent searching through the databases for relevant enzymes used in engineered pathways. Although not automated, the Retro-Biosynthesis Tool (ReBiT) provides access to a searchable database of enzyme-catalyzed reactions categorized by the relevant functional groups of the reactants and products involved in the reaction [69]. ReBiT returns a list of enzymes capable of consuming or generating any user-defined functional group, which is particularly helpful for designing novel metabolic pathways (i.e., nonnatural).

Although these tools will facilitate the design of engineered systems, they will also need to take into account connectivity with the existing cellular network. At this point, even the most well-understood model microorganisms like *E. coli* and *S. cerevisiae* are far from being completely modeled [70, 71]. Therefore, interfacing engineered pathways with natural networks remains a difficult task [72]. However, existing metabolic models, when combined with powerful CAD tools to develop novel biological systems, will streamline the metabolic engineering process. Current genome-scale metabolic models are developed by deriving reaction list from genome annotation, refined using literature information and experimental data, and probed using constraint-based algorithms such as flux balance analysis (FBA) [73]. This modeling approach should enable decision-making when it comes to choosing an appropriate microbe to host a particular synthetic pathway and provide a means of predicting the in-context consequences of heterologous expression.

6. Perspectives

Genetically programming microbes to perform desired tasks has great potential to transform modern biotechnology. Significant progress has already been made in implementing novel biological functionality in microorganisms, including complex sensing/actuation algorithms and genetic circuits that count [74–77]. These advances have implications for metabolic engineering. For example, genetic circuits have been engineered to control the transcription of an operon that encodes enzymes and machinery necessary to manufacture and secrete recombinant spider silk monomers in *Salmonella* [78]. As progress is made in synthetic biology to develop well-characterized parts and increasingly complex devices, the design, modeling, synthesis, and analysis of synthetic metabolic pathways will become easier.

In metabolic engineering, there is an implicit view of cells as chemical factories. As such, it can be useful to borrow conceptual frameworks from chemical engineering. Chemical plants are built from components that generally fall into one of three classes: unit operations (e.g., mixers, reactors, separators), connectors (pipes, tanks), and process controllers (measurement instruments, valves). These component classes exist within cells, but need to be further refined so that they can be used at the level of sophistication found in a chemical plant. In addition to advancing metabolic control through regulation of gene expression (e.g., engineered promoters, tunable intergenic regions [41]) and enzyme colocalization (modular protein scaffolds, or perhaps synthetic organelles), the enzymes themselves (i.e., specificity, activity, and allostery) can be modulated through directed evolution [79, 80] or perhaps designed/redesigned computationally. In fact, a recent study showed that swapping catalytic domains from three fungal cellulases can produce chimeras with superior properties [81]. Furthermore, engineered enzymes with novel catalytic properties would enable nonnatural metabolic pathway implementation. However, rationally engineering enzymes remain difficult and must be conducted on a case-by-case basis.

Therefore, it would be advantageous for the metabolic engineering community to work within the synthetic biology framework and focus on generating and populating families of genetic parts that are scalable, measured, and modeled. The simultaneous construction and characterization of parts will lead to model-based design of pathways in the future, just as complex electrical circuits are built *in silico* from well-modeled components. High-throughput screening technologies have become powerful platforms for quantifying component performance of combinatorially-generated libraries [41]. Likewise, high-throughput measurements provide insight about evolved systems that have adapted to meet performance criteria [82]. These approaches could be very useful in identifying the genetic basis of the biological solution that evolution has provided, a strategy commonly referred to as inverse metabolic engineering [83]. The interplay of rational, combinatorial, and evolutionary approaches is crucial to the continued elucidation of biological design principles, which allow further forward-engineering of biological systems [84].

In order to achieve the forward-engineering of systems-level metabolism, the design, modeling, synthesis, and analysis stages must incorporate shared standards in order to efficiently work together. In particular, well-characterized, off-the-shelf genetic components *and* computational models of these components must be available for use within a standard biological CAD tool. As models of components and modules develop, metabolic engineers will be able to move away from using the equivalent of biological machine code (DNA bases: ATCGs) for design to using compiled biological programs (genes and genetic circuits) that will enable faster pathway design by reusing pieces of previously optimized pathways. Similarly, characterization methods and standard measurement units must be agreed on so that parts are characterized as they are made in a way that is useful to the community.

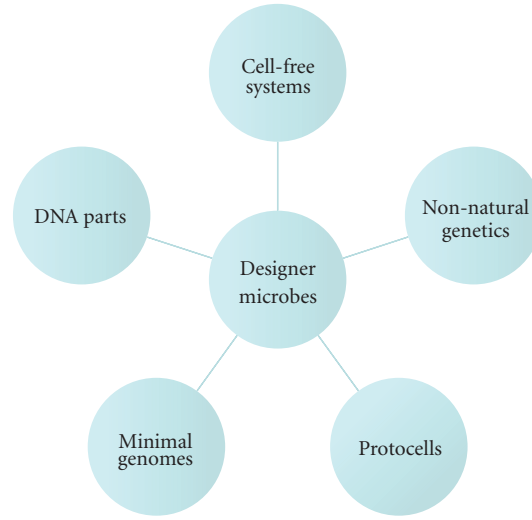


Figure 3: *Convergence of the different realms of synthetic biology.* In addition to efforts to refine and characterize genetic parts based on DNA, research on expanding the genetic code and implementing biological processes in cell-free systems will likely integrate with metabolic engineering projects in the future. Genome engineering and synthetic genomics will enable the manufacture of customizable minimal genomes. Alternatively, protocell engineering may provide a non-biological chassis to house synthetic chemistry.

As already envisioned by a number of people [56, 61], integration of synthetic biology and systems biology will likely occur as both fields grow, eventually leading to tools that can allow for true cellular-scale network and pathway engineering. In many respects the fields of synthetic biology and systems biology are perfect complements to each other in regard to outlook and approach as we move toward whole-cell engineering. While some overlap already exists between the two fields, synthetic biology currently has significant strengths in the steps of biological design and synthesis whereas systems biology has strengths in analysis and modeling (Figure 1). To properly integrate the detailed, dynamic, small-network components of synthetic biology with the large-scale analyses and algorithms of systems biology, it will be necessary to ensure that work in each field is compatible with work in the other field. For example, large-scale systems biology models will need to be able to incorporate dynamic behavior demonstrated by constructed synthetic circuits, and synthetic circuits should be constructed and implemented to have minimal cross-talk with existing cellular components. One can imagine a time when design can happen at a large-scale level using systems biology algorithms [85–87] followed by molecular-level design using synthetic biology tools, *de novo* synthesis of components and pathways, analysis of results using high-throughput systems biology methods, and a hybrid large-scale, dynamic model of the entire process.

As the field of synthetic biology matures, various branches of synthetic biology that were not discussed here

will soon be relevant to implementing novel metabolism in microbes (Figure 3). For example, a streamlined chassis based on a minimal, refactored genome would simplify host/system interactions and potentially minimize the effect of the metabolic burden the exogenous pathway would place on the cell. Likewise, the concept of designer microbes with engineered genomes could be realized by the developing area of synthetic genomics, which is dependent on the continued improvement in DNA synthesis technology. In the near future, host strains that have been engineered to confer complex phenotypes (e.g., solvent tolerance) using techniques such global transcription machinery engineering (gTME) may provide excellent chassis for synthetic metabolic systems [37, 88]. Even so, recent successes in synthetic biology-enabled metabolic pathway engineering in common lab strains are very encouraging [31, 89–93]. Furthermore, work done toward engineering microbial communities may provide an approach to industrial biosynthesis that is superior to monocultures due to increased diversity and flexibility [94]. The future of this field is extremely exciting and will likely include rapid prototyping, testing, and debugging of synthetic metabolism for the industrial-scale microbial production of a wide variety of natural and novel chemicals and materials.

Acknowledgment

The authors thank Julius B. Lucks for thoughtful discussions and input.

References

[1] J. W. Chin, "Modular approaches to expanding the functions of living matter," *Nature Chemical Biology*, vol. 2, no. 6, pp. 304–311, 2006.

[2] D. Endy, "Foundations for engineering biology," *Nature*, vol. 438, no. 7067, pp. 449–453, 2005.

[3] S. A. Benner and A. M. Sismour, "Synthetic biology," *Nature Reviews Genetics*, vol. 6, no. 7, pp. 533–543, 2005.

[4] T. S. Ham, S. K. Lee, J. D. Keasling, and A. P. Arkin, "Design and construction of a double inversion recombination switch for heritable sequential genetic memory," *PLoS ONE*, vol. 3, no. 7, article e2815, 2008.

[5] J. D. Keasling, "Synthetic biology for synthetic chemistry," *ACS Chemical Biology*, vol. 3, no. 1, pp. 64–76, 2008.

[6] J. B. Lucks, L. Qi, W. R. Whitaker, and A. P. Arkin, "Toward scalable parts families for predictable design of biological circuits," *Current Opinion in Microbiology*, vol. 11, no. 6, pp. 567–573, 2008.

[7] T. Y. Kim, S. B. Sohn, H. U. Kim, and S. Y. Lee, "Strategies for systems-level metabolic engineering," *Biotechnology Journal*, vol. 3, no. 5, pp. 612–623, 2008.

[8] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology: from modules to systems," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 6, pp. 410–422, 2009.

[9] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging discipline," *Molecular Systems Biology*, vol. 2, Article ID 2006.0028, 14 pages, 2006.

[10] M. Welch, A. Villalobos, C. Gustafsson, and J. Minshull, "You're one in a googol: optimizing genes for protein expression," *Journal of the Royal Society Interface*, vol. 6, supplement 4, pp. S467–S476, 2009.

[11] R. P. Shetty, D. Endy, and T. F. Knight Jr., "Engineering BioBrick vectors from BioBrick parts," *Journal of Biological Engineering*, vol. 2, article 5, 2008.

[12] T. Ellis, X. Wang, and J. J. Collins, "Diversity-based, model-guided construction of synthetic gene networks with predicted functions," *Nature Biotechnology*, vol. 27, no. 5, pp. 465–471, 2009.

[13] H. M. Salis, E. A. Mirsky, and C. A. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology*, vol. 27, no. 10, pp. 946–950, 2009.

[14] T. S. Ham, S. K. Lee, J. D. Keasling, and A. P. Arkin, "A tightly regulated inducible expression system utilizing the fim inversion recombination switch," *Biotechnology and Bioengineering*, vol. 94, no. 1, pp. 1–4, 2006.

[15] J. C. Anderson, C. A. Voigt, and A. P. Arkin, "Environmental signal integration by a modular and gate," *Molecular Systems Biology*, vol. 3, article 133, 2007.

[16] F. J. Isaacs, D. J. Dwyer, C. Ding, D. D. Pervouchine, C. R. Cantor, and J. J. Collins, "Engineered riboregulators enable post-transcriptional control of gene expression," *Nature Biotechnology*, vol. 22, no. 7, pp. 841–847, 2004.

[17] M. N. Win and C. D. Smolke, "A modular and extensible RNA-based gene-regulatory platform for engineering cellular function," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 36, pp. 14283–14288, 2007.

[18] M. N. Win, J. C. Liang, and C. D. Smolke, "Frameworks for programming biological function through RNA parts and devices," *Chemistry and Biology*, vol. 16, no. 3, pp. 298–310, 2009.

[19] D. J. Pitera, C. J. Paddon, J. D. Newman, and J. D. Keasling, "Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*," *Metabolic Engineering*, vol. 9, no. 2, pp. 193–207, 2007.

[20] M. E. Wall, W. S. Hlavacek, and M. A. Savageau, "Design of gene circuits: lessons from bacteria," *Nature Reviews Genetics*, vol. 5, no. 1, pp. 34–42, 2004.

[21] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: from theories to phenotypes," *Nature Reviews Genetics*, vol. 6, no. 6, pp. 451–464, 2005.

[22] A. Raj and A. van Oudenaarden, "Nature, nurture, or chance: stochastic gene expression and its consequences," *Cell*, vol. 135, no. 2, pp. 216–226, 2008.

[23] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[24] M. B. Elowitz and S. Leibier, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.

[25] T. E. Landrain, J. Carrera, B. Kirov, G. Rodrigo, and A. Jaramillo, "Modular model-based design for heterologous bioproduction in bacteria," *Current Opinion in Biotechnology*, vol. 20, no. 3, pp. 272–279, 2009.

[26] R. J. Conrado, J. D. Varner, and M. P. DeLisa, "Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy," *Current Opinion in Biotechnology*, vol. 19, no. 5, pp. 492–499, 2008.

[27] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, et al., "Synthetic protein scaffolds provide modular control over metabolic flux," *Nature Biotechnology*, vol. 27, no. 8, pp. 753–759, 2009.

[28] C. A. Voigt, "Genetic parts to program bacteria," *Current Opinion in Biotechnology*, vol. 17, no. 5, pp. 548–557, 2006.

[29] M. R. Bennett and J. Hasty, "Overpowering the component problem," *Nature Biotechnology*, vol. 27, no. 5, pp. 450–451, 2009.

[30] J. H. Park, S. Y. Lee, T. Y. Kim, and H. U. Kim, "Application of systems biology for bioprocess development," *Trends in Biotechnology*, vol. 26, no. 8, pp. 404–412, 2008.

[31] T. S. Bayer, D. M. Widmaier, K. Temme, E. A. Mirsky, D. V. Santi, and C. A. Voigt, "Synthesis of methyl halides from biomass using engineered microbes," *Journal of the American Chemical Society*, vol. 131, no. 18, pp. 6508–6515, 2009.

[32] S. J. Reisinger, K. G. Patel, and D. V. Santi, "Total synthesis of multi-kilobase DNA sequences from oligonucleotides," *Nature Protocols*, vol. 1, no. 6, pp. 2596–2603, 2007.

[33] M. J. Czar, J. C. Anderson, J. S. Bader, and J. Peccoud, "Gene synthesis demystified," *Trends in Biotechnology*, vol. 27, no. 2, pp. 63–72, 2009.

[34] K. L. Jones, S.-W. Kim, and J. D Keasling, "Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria," *Metabolic Engineering*, vol. 2, no. 4, pp. 328–338, 2000.

[35] D. Klein-Marcuschamer, C. N. S. Santos, H. Yu, and G. Stephanopoulos, "Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes," *Applied and Environmental Microbiology*, vol. 75, no. 9, pp. 2705–2711, 2009.

[36] T. Sera, "Zinc-finger-based artificial transcription factors and their applications," *Advanced Drug Delivery Reviews*, vol. 61, no. 7-8, pp. 513–526, 2009.

[37] H. Alper and G. Stephanopoulos, "Global transcription machinery engineering: a new approach for improving cellular phenotype," *Metabolic Engineering*, vol. 9, no. 3, pp. 258–267, 2007.

[38] H. Alper, C. Fischer, E. Nevoigt, and G. Stephanopoulos, "Tuning genetic control through promoter engineering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 36, pp. 12678–12683, 2005.

[39] M. Krueger, O. Scholz, S. Wisshak, and W. Hillen, "Engineered Tet repressors with recognition specificity for the tetO-4C5G operator variant," *Gene*, vol. 404, no. 1-2, pp. 93–100, 2007.

[40] L. Du, R. Gao, and A. C. Forster, "Engineering multigene expression in vitro and in vivo with small terminators for T7 RNA polymerase," *Biotechnology and Bioengineering*, vol. 104, no. 6, pp. 1189–1196, 2009.

[41] B. F. Pfleger, D. J. Pitera, C. D. Smolke, and J. D. Keasling, "Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes," *Nature Biotechnology*, vol. 24, no. 8, pp. 1027–1032, 2006.

[42] C. D. Smolke, T. A. Carrier, and J. D. Keasling, "Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures," *Applied and Environmental Microbiology*, vol. 66, no. 12, pp. 5399–5405, 2000.

[43] O. Rackham and J. W. Chin, "Cellular logic with orthogonal ribosomes," *Journal of the American Chemical Society*, vol. 127, no. 50, pp. 17584–17585, 2005.

[44] F. J. Isaacs, D. J. Dwyer, and J. J. Collins, "RNA synthetic biology," *Nature Biotechnology*, vol. 24, no. 5, pp. 545–554, 2006.

[45] C. Engler, R. Gruetzner, R. Kandzia, and S. Marillonnet, "Golden gate shuffling: a one-pot DNA shuffling method based on type ils restriction enzymes," *PLoS ONE*, vol. 4, no. 5, article e5553, 2009.

[46] J. Quan and J. Tian, "Circular polymerase extension cloning of complex gene libraries and pathways," *PLoS ONE*, vol. 4, no. 7, article e6441, 2009.

[47] H. H. Wang, F. J. Isaacs, P. A. Carr, et al., "Programming cells by multiplex genome engineering and accelerated evolution," *Nature*, vol. 460, no. 7257, pp. 894–898, 2009.

[48] J. D. Keasling, "Gene-expression tools for the metabolic engineering of bacteria," *Trends in Biotechnology*, vol. 17, no. 11, pp. 452–460, 1999.

[49] K. E. J. Tyo, P. K. Ajikumar, and G. Stephanopoulos, "Stabilized gene duplication enables long-term selection-free heterologous pathway expression," *Nature Biotechnology*, vol. 27, no. 8, pp. 760–765, 2009.

[50] Z. Shao, H. Zhao, and H. Zhao, "DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways," *Nucleic Acids Research*, vol. 37, no. 2, article e16, 2009.

[51] R. M. Q. Shanks, N. C. Caiazza, S. M. Hinsa, C. M. Toutain, and G. A. O'Toole, "Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria," *Applied and Environmental Microbiology*, vol. 72, no. 7, pp. 5027–5036, 2006.

[52] D. G. Gibson, G. A. Benders, K. C. Axelrod, et al., "One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 51, pp. 20404–20409, 2008.

[53] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature Methods*, vol. 6, no. 5, pp. 343–345, 2009.

[54] C. Lartigue, J. I. Glass, N. Alperovich, et al., "Genome transplantation in bacteria: changing one species to another," *Science*, vol. 317, no. 5838, pp. 632–638, 2007.

[55] D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, et al., "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome," *Science*, vol. 319, no. 5867, pp. 1215–1220, 2008.

[56] C. L. Barrett, T. Y. Kim, H. U. Kim, B. Ø. Palsson, and S. Y. Lee, "Systems biology as a foundation for genome-scale synthetic biology," *Current Opinion in Biotechnology*, vol. 17, no. 5, pp. 488–492, 2006.

[57] B. Canton, A. Labno, and D. Endy, "Refinement and standardization of synthetic biological parts and devices," *Nature Biotechnology*, vol. 26, no. 7, pp. 787–793, 2008.

[58] A. Arkin, "Setting the standard in synthetic biology," *Nature Biotechnology*, vol. 26, no. 7, pp. 771–774, 2008.

[59] J. R. Kelly, A. J. Rubin, J. H. Davis, et al., "Measuring the activity of BioBrick promoters using an in vivo reference standard," *Journal of Biological Engineering*, vol. 3, article 4, 2009.

[60] S. J. Park, S. Y. Lee, J. Cho, et al., "Global physiological understanding and metabolic engineering of microorganisms based on omics studies," *Applied Microbiology and Biotechnology*, vol. 68, no. 5, pp. 567–579, 2005.

[61] A. Mukhopadhyay, A. M. Redding, B. J. Rutherford, and J. D. Keasling, "Importance of systems biology in engineering microbes for biofuel production," *Current Opinion in Biotechnology*, vol. 19, no. 3, pp. 228–234, 2008.

[62] L. Kizer, D. J. Pitera, B. F. Pfleger, and J. D. Keasling, "Application of functional genomics to pathway optimization for increased isoprenoid production," *Applied and Environmental Microbiology*, vol. 74, no. 10, pp. 3229–3241, 2008.

[63] M. A. Marchisio and J. Stelling, "Computational design of synthetic gene circuits with composable parts," *Bioinformatics*, vol. 24, no. 17, pp. 1903–1910, 2008.

[64] L. Endler, N. Rodriguez, N. Juty, et al., "Designing and encoding models for synthetic biology," *Journal of the Royal Society Interface*, vol. 6, supplement 4, pp. S405–S417, 2009.

[65] V. Rouilly, B. Canton, P. Nielsen, and R. Kitney, "Registry of BioBricks models using CellML," *BMC Systems Biology*, vol. 1, article 79, 2007.

[66] M. J. Czar, Y. Cai, and J. Peccoud, "Writing DNA with GenoCADTM," *Nucleic Acids Research*, vol. 37, supplement 2, pp. W40–W47, 2009.

[67] J. A. Goler, B. W. Bramlett, and J. Peccoud, "Genetic design: rising above the sequence," *Trends in Biotechnology*, vol. 26, no. 10, pp. 538–544, 2008.

[68] C.-H. Chou, W.-C. Chang, C.-M. Chiu, C.-C. Huang, and H.-D. Huang, "FMM: a web server for metabolic pathway reconstruction and comparative analysis," *Nucleic Acids Research*, vol. 37, supplement 2, pp. W129–W134, 2009.

[69] K. L. J. Prather and C. H. Martin, "De novo biosynthetic pathways: rational design of microbial chemical factories," *Current Opinion in Biotechnology*, vol. 19, no. 5, pp. 468–474, 2008.

[70] D. W. Selinger, M. A. Wright, and G. M. Church, "On the complete determination of biological systems," *Trends in Biotechnology*, vol. 21, no. 6, pp. 251–254, 2003.

[71] D. J. Dwyer, M. A. Kohanski, and J. J. Collins, "Networking opportunities for bacteria," *Cell*, vol. 135, no. 7, pp. 1153–1156, 2008.

[72] H. Kobayashi, M. Kærn, M. Araki, et al., "Programmable cells: interfacing natural and engineered gene networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 22, pp. 8414–8419, 2004.

[73] J. M. Park, T. Y. Kim, and S. Y. Lee, "Constraints-based genome-scale metabolic simulation for systems metabolic engineering," *Biotechnology Advances*, vol. 27, no. 6, pp. 979–988, 2009.

[74] J. C. Anderson, E. J. Clarke, A. P. Arkin, and C. A. Voigt, "Environmentally controlled invasion of cancer cells by engineered bacteria," *Journal of Molecular Biology*, vol. 355, no. 4, pp. 619–627, 2006.

[75] J. J. Tabor, H. M. Salis, Z. B. Simpson, et al., "A synthetic genetic edge detection program," *Cell*, vol. 137, no. 7, pp. 1272–1281, 2009.

[76] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins, "Synthetic gene networks that count," *Science*, vol. 324, no. 5931, pp. 1199–1202, 2009.

[77] A. Levskaya, O. D. Weiner, W. A. Lim, and C. A. Voigt, "Spatiotemporal control of cell signalling using a light-switchable protein interaction," *Nature*, vol. 461, no. 7266, pp. 997–1001, 2009.

[78] D. M. Widmaier, D. Tullman-Ercek, E. A. Mirsky, et al., "Engineering the Salmonella type III secretion system to export spider silk monomers," *Molecular Systems Biology*, vol. 5, article 309, 2009.

[79] Y. Yoshikuni, J. A. Dietrich, F. F. Nowroozi, P. C. Babbitt, and J. D. Keasling, "Redesigning enzymes based on adaptive evolution for optimal function in synthetic metabolic pathways," *Chemistry and Biology*, vol. 15, no. 6, pp. 607–618, 2008.

[80] J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan, and F. H. Arnold, "Evolving strategies for enzyme engineering," *Current Opinion in Structural Biology*, vol. 15, no. 4, pp. 447–452, 2005.

[81] P. Heinzelman, C. D. Snow, I. Wu, et al., "A family of thermostable fungal cellulases created by structure-guided recombination," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 14, pp. 5610–5615, 2009.

[82] S. S. Fong, A. P. Burgard, C. D. Herring, et al., "In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid," *Biotechnology and Bioengineering*, vol. 91, no. 5, pp. 643–648, 2005.

[83] C. N. S. Santos and G. Stephanopoulos, "Combinatorial engineering of microbes for optimizing cellular phenotype," *Current Opinion in Chemical Biology*, vol. 12, no. 2, pp. 168–176, 2008.

[84] L. Rohlin, M.-K. Oh, and J. C. Liao, "Microbial pathway engineering for industrial processes: evolution, combinatorial biosynthesis and rational design," *Current Opinion in Microbiology*, vol. 4, no. 3, pp. 330–335, 2001.

[85] A. P. Burgard, P. Pharkya, and C. D. Maranas, "Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnology and Bioengineering*, vol. 84, no. 6, pp. 647–657, 2003.

[86] P. Pharkya, A. P. Burgard, and C. D. Maranas, "OptStrain: a computational framework for redesign of microbial production systems," *Genome Research*, vol. 14, no. 11, pp. 2367–2376, 2004.

[87] D. S. Lun, G. Rockwell, N. J. Guido, et al., "Large-scale identification of genetic design strategies using local search," *Molecular Systems Biology*, vol. 5, article 296, 2009.

[88] J. R. Warner, R. Patnaik, and R. T. Gill, "Genomics enabled approaches in strain engineering," *Current Opinion in Microbiology*, vol. 12, no. 3, pp. 223–230, 2009.

[89] K. M. Hawkins and C. D. Smolke, "Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae," *Nature Chemical Biology*, vol. 4, no. 9, pp. 564–573, 2008.

[90] T. S. Moon, S.-H. Yoon, A. M. Lanza, J. D. Roy-Mayhew, and K. L. J. Prather, "Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*," *Applied and Environmental Microbiology*, vol. 75, no. 3, pp. 589–595, 2009.

[91] K. Zhang, M. R. Sawaya, D. S. Eisenberg, and J. C. Liao, "Expanding metabolism for biosynthesis of nonnatural alcohols," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 52, pp. 20653–20658, 2008.

[92] V. J. J. Martin, D. J. Piteral, S. T. Withers, J. D. Newman, and J. D. Keasling, "Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids," *Nature Biotechnology*, vol. 21, no. 7, pp. 796–802, 2003.

[93] D.-K. Ro, E. M. Paradise, M. Quellet, et al., "Production of the antimalarial drug precursor artemisinic acid in engineered yeast," *Nature*, vol. 440, no. 7086, pp. 940–943, 2006.

[94] K. Brenner, L. You, and F. H. Arnold, "Engineering microbial consortia: a new frontier in synthetic biology," *Trends in Biotechnology*, vol. 26, no. 9, pp. 483–489, 2008.