Genetic/Bio Design Automation for (Re-)Engineering Biological Systems

Soha Hassoun

Department of Computer Science

Tufts University

Medford, MA 02155

Soha.Hassoun@tufts.edu

Abstract—Constructing biological circuits in a bottom-up modular fashion using design methodologies similar to those used in electronics has gained tremendous attention in the past decade. The end goal, however, is engineering biological systems and not only individual components in the context of pursuing applications useful in improving human health or enhancing the environment. This article reviews the basics of biological system design rooted in Metabolic Engineering and Systems Biology and outlines current system-level modeling, analysis, optimization, and synthesis with emphasis on some current bottlenecks in establishing more rigorous design tools and methodologies for engineering biological systems.

I. INTRODUCTION

Continued interest in engineering biological cellular systems is driven by the desire to create products or engineer novel behaviors with current applications for producing pharmaceuticals, repairing defective genes, destroying cancer cells, generating biofuels, and assisting climate change through carbon capture. The excitement, however, is about what is yet to be achieved [1], [2]. Innovations including high-throughput data technologies such as DNA sequencing [3], the construction of synthetic gene regulatory circuits using principled engineering methods [4], [5] and directed evolution approaches [6] will drive biological engineering in a manner that resembles how, during the past five decades, IC manufacturing technology enabled current electronic system design. Systematic and predictable engineering approaches, and in particular Computer-Aided Design tools and methodologies, are poised to play critical roles.

The view that biological systems are engineerable has emerged from multiple disciplines [7]. Metabolic Engineering aims to introduce, modify and optimize biochemical reactions within a living cell to produce novel compounds of industrial and medical interest. Systems Biology seeks to understand in quantitative predictable ways the regulation of complex cellular pathways and intercellular communication with the purpose of understanding living systems. Emphasizing the need for standards, abstraction hierarchy, and composability in engineering biology, Synthetic Biology calls for the assembly of *de novo* genetic circuitry (gene regulatory networks) from a set of genetic parts. Conceptually, gene regulatory networks provide control over the data flow (flux) in a biological system. Additional efforts come from the disciplines of molecular biology (dealing with individual macromolecules of living organisms, genetic engineering focusing on gene manipulation, cloning, and recombinant DNA technology) and computational biology (applying computer science, statistics and applied mathematics to biological problems in the form of data analysis bioinformatics, molecular modeling, and protein structure prediction and simulation). Advances have occurred within each discipline and together shape our understanding of biology and therefore our ability to engineer it. The analogy is in the synergy among various design efforts targeting devices, circuits, architecture, hardware systems and software.

Biological systems and electronics are different. Biological cells grow changing size and functional characteristics, evolve through mutations, die, reproduce, exhibit robustness to environmental variables such as nutrients, temperature, and PH balance, and adapt due to redundant components. While the notion and merits of designing a minimal synthetic biological system have been debated, biological systems naturally exist as microorganisms and within plants, animals and humans. Design efforts have typically focused on "re-engineering" biological systems by integrating novel genetic material, either naturally existing or synthetic, within the cell and then tweaking the system to produce a desired behavior or product. Biological systems themselves are thus design platforms (chassis) and re-engineering biological systems is a form of platform-based design. One spectacular success is designing and implementing a nonnative synthesis pathway in yeast to produce artemisinic acid, a direct precursor of the antimalarial compound artemisinin [8]. In addition to identifying and inserting an efficient biosynthetic pathway, significant production of the target compounds required further modifications to yeast's metabolic network to remove bottlenecks in precursor supply and alleviate the accumulation of harmful intermediates. More recently, a similar approach was adopted by Pfeifer and co-workers to synthesize the complete antibiotic erythromycin A (rather than one of its precursors) in *E. coli* [3]. "In-human" medicinal applications of synthetic biology to analyze or control cancer cells will require integrating within mammalian cells [9]. Considerable effort is required to create predictive models for each unique biological platform (cellular organism) and to validate our understanding of both biological structure and function. Currently, it is more expedient to use *in vivo* experimental methods guided by coarse computational

guidance instead of relying on *in silico* design approaches.

This article provides an engineering perspective on current computational methods to model, analyze, optimize, and synthesize the (re)-engineering of biological systems. Validation/verification of engineered designs is not specifically addressed because it is currently conducted experimentally. Often, the experimental work drives model construction and refinement and hypothesis generation. The reviewed techniques are complementary to bio/genetic design automation (G/BDA) tools that target gene regulatory networks [10], [11], [12].

II. BACKGROUND

A cellular process is represented using a biochemical network consisting of a set of reactions and compounds (Figure 1(a)). The reactions describe the chemicalprocesses in livingorganisms. A reaction's stoichiometry, invariant to the cell's operating conditions, specifies the relative number of atoms consumed or produced due to a chemical reaction. A biochemical network with m compounds and n reactions is represented using a $m \times n$ stoichiometric matrix N (Figure 1(b)). Each column describes a reaction. A column entry represents the stoichiometric coefficient of a compound participating in the relevant reaction. A column entry is zero if the compound does not participate in the reaction, positive if the compound is produced and negative if consumed. Reactions in a network can be classified as internal or exchange reactions linking a biochemical network to its external environment, as defined by the user, providing either uptake or production of external metabolites. Each row summarizes how a compound participates in various reactions. When utilizing the N matrix during analysis, typically only rows corresponding to internal compounds are included. The matrix can be viewed as a graph (Figure 1(c)). A pathway refers to a sequence of reactions, just like in graph theory, however, it is often implied that the pathway is atomically balanced. Reactions maybe be reversible, and are sometimes split into forward and reverse reactions during analysis. A network, once in steady state (consuming and producing internal compounds at a steady rate), exhibits a large number of possible functional states dependent on operating conditions. *Flux*, the reaction rate or the turnover rate of the molecules associated with a reaction or pathway at steady state, characterizes cellular functional states. Often, however, there are more compounds than reactions, and only flux ranges can be mathematically determined.

Kinetic models capture dynamic behaviors such as how fast a reaction occurs as a function of the relevant concentration. Unlike purely chemical reactions, a biochemical reaction experiences saturation when catalyzed by an enzyme. An example reaction rate as a function of substrate concentration is shown in Figure 1(d), along with the simplest equation, in Michaelis-Menten form, used to describe enzyme kinetics of a single-substrate single-enzyme catalyzed reaction. V_{max} is the maximum rate (velocity) for a particular reaction. V_{max}

changes if more enzyme is added. The Michaelis-Menten constant, K'_m, is determined experimentally and represents the substrate concentration at which the reaction rate is half of V_{max}. A biochemical system can thus be described by coupled ordinary differential equations (ODEs). Enzyme kinetics is however complex and several proposed mathematical models have been proposed as approximations including S-systems and convenience kinetics [13]. Parameters are however often unknown or characterized for specific conditions, and expression forms are approximations.

III. COMPUTER-AIDED TOOLS

A. Modeling

Functional specification using specialized languages such as Verilog or SystemC have become common practice when designing electronics. Within Metabolic Engineering, however, a model refers to a mathematical representation of an existing biological system. Creating the model is referred to as a "reconstruction". The process is iterative, and, until recently, manual. For example, the construction of the biochemical reaction network for *Escherichia coli* spanned over 15 years [14]. Manually constructed models often serve a particular purpose and are therefore partial and only partially validated.

The recent availability of genome, reaction, and organism-specific databases have allowed for the automatic reconstruction of genome-scale models represented as stoichiometric matrices. A protocol encompassing 94 steps details the process including obtaining a draft construction from databases, collecting experimental data, refining the reconstruction by adding details to ensure that the network is mass- and charge-balanced and that missing reactions steps are properly flagged, to test the ability of the model to grow, and to compare against known properties [15]. The tool, Model SEED, expedites this process by automating most of the steps but manual curation is needed to match the quality of those that are manually constructed [16]. The quality of the model, whether obtained manually or automatically, is as comprehensive as the availability of the reconstruction and experimental data.

Kinetic models, while preferable as they provide time-varying analysis, are difficult to reconstruct. Some reaction-catalyzing enzymes have a simple structure but many are complex. Enzymes bind to reactants to form an enzyme-substrate complex and lower the activation energy, which results in accelerating the reaction rate. The interaction between reactant and enzyme varies and the exact molecular aspects of enzyme-substrate interactions are not fully understood. Reaction mechanisms are partially or completely unknown and often experimental conditions under which kinetic parameters have been determined are not available through the databases [17]. Importantly, models must be validated in an *in vivo* context. Using approximate mathematical models of the enzyme kinetics and measured data, several methods have been proposed to estimate the kinetic parameters. The problem is generally cast as a non-linear optimization. Given a set of ODEs, the objective function is to minimize the sum-squared difference between the calculated and measured dependent
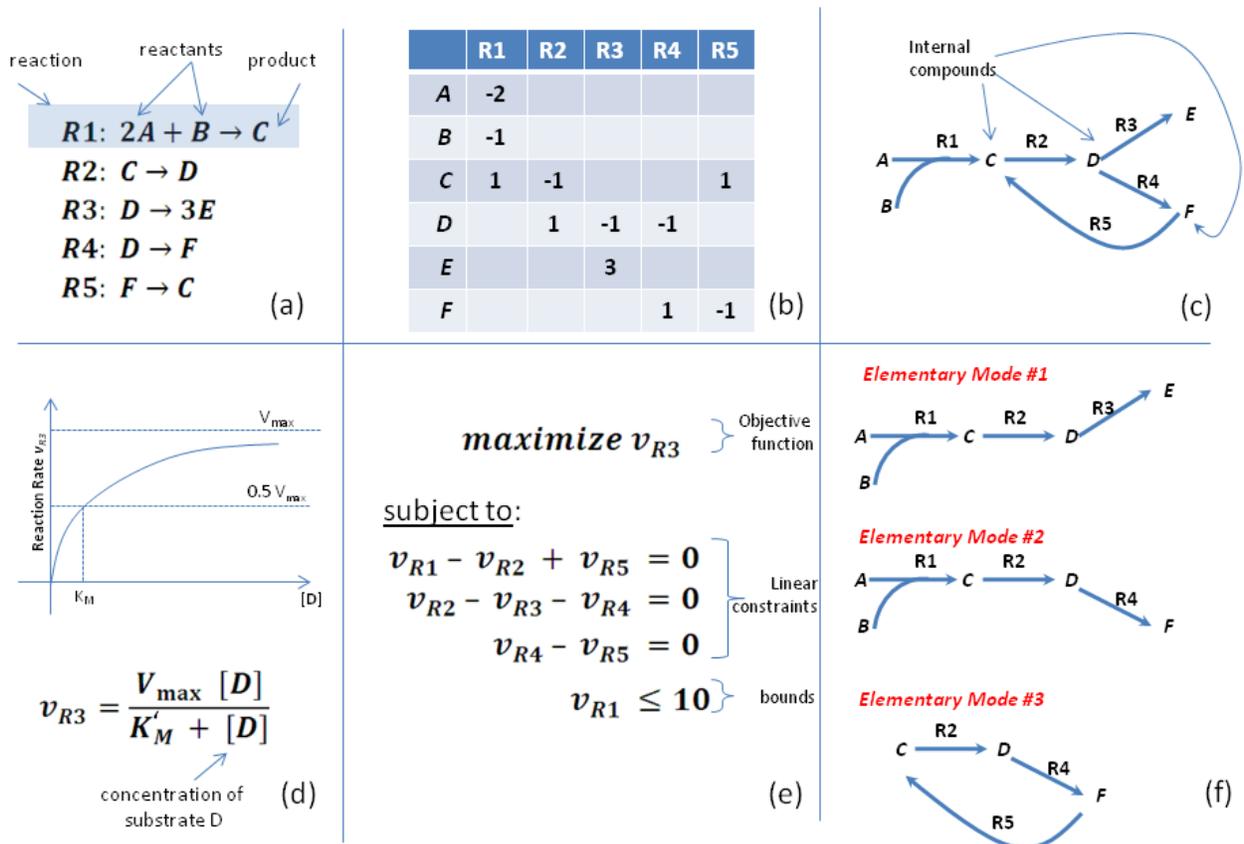
Fig. 1. System analysis fundamentals. (a) The systems are modeled as a set of biochemical reactions. (b) Reaction stoichiometry is captured using a stoichiometric matrix. Zero entries are removed from the matrix for simplicity. (c) The network of reactions can be represented using a graph. (d) Example equation and graph for reaction rate as a function of substrate concentration. (e) Flux Balance Analysis example to maximize the flux of reaction R3. (f) Elementary modes for the network in Figure (c).

variables based on a set of parameter choices. The inputs to the problem are the measured initial values of the independent variables (e.g. metabolite concentrations). As with other non-linear optimization problems, guaranteeing a globally optimal solution is difficult. Using gradient-based local methods that repeatedly solve the problem with different initial conditions fail to arrive at satisfactory solutions [10]. It is generally agreed that global methods, while computationally expensive, are more appropriate[18]. Despite computational advances in parameter estimation, the size and complexity of biochemical networks reconstructed from genome databases has greatly increased over the years rendering the estimation of kinetic or regulatory parameters, or fitting against in vitro experimental data, either impractical or outright infeasible. Often, steady-state analysis is the only means to analyze a biochemical system.

B. Analysis

An interesting feature of biochemical networks is that they exhibit a large number of possible functional states, resulting in a great variety of phenotypes. At best, metabolic engineers and system biologists today can utilize known constraints,

such as conservation of mass, energy and momentum, to limit possible functional states. Flux Balance Analysis (FBA) is a technique to analyze flux distributions at steady state, when the net production and consumption rates are equal. Flux for a reaction i, is typically denoted by v_i. Equivalent to Kirchhoff's current law, mass conservation at steady state declares that the rate of consumption and production of internal compounds must be equal for each metabolite (Figure 1(e)). Specifying mass balance constraints for all internal compounds generates a set of linear equations. An objective function can be defined to correspond to maximizing the flux through a reaction leading to a desired target metabolite. For example, as shown in Figure 1(e), specifying the uptake rate of R_1 to be 10 and maximizing v_{R3}, results in v_{R3} equal to 10. There are, however, several flux distributions that maximize v_R3 as the set of linear equations describing the system is undetermined. One possible distribution is $\{10, 15, 10, 5, 5\}$, corresponding to the flux in reactions 1 through 5. Another is $\{10, 10, 10, 0, 0\}$. Constrained-based analysis has been used to analyze flux variability, flux coupling, and to identify optimal gene (reaction) knockout strategies. In recent years, FBA has been enriched with various extensions, including energy balance

constraints [19], pathway thermodynamics [20], [21], and multi-objective optimization [22]. See [23] for a review.

Pathways play an important part in the analysis of biochemical networks. Elementary Flux Mode (EFM) analysis is a pathway analysis technique that decomposes a biochemical network into an independent set of stoichiometrically balanced pathways called elementary flux modes (EFMs). When applied to the example in Figure 1(c), the resulting three elementary modes are as illustrated in Figure 1(g). A feasible flux distribution (e.g. {10, 15, 10, 5, 5}) can be expressed as a linear combination of the flux in the EFMs, (e.g. using weights 10 and 5 for elementary modes 1 and 2, respectively). EFM analysis exhaustively enumerates all stoichiometrically balanced pathways and cycles. Once all EFM are identified, they can be analyzed individually or within EFM families and used to make engineering decisions. For example, enhancing enzyme activities along a particular pathways can increase the target's yield. Eliminating competing pathways through gene knockouts (i.e. eliminate the reactions from the network) can further enhance yield.

EFM analysis is computationally intractable [24]. Even a modest size model can have millions of EFMs, and computation time and/or memory become bottlenecks. Dividing the network into subsystems (modularization) is one approach [25]. However, the modularization may eliminate some EFMs that span several subsystems in the larger network [26]. Abstracting the network, e.g. by eliminating redundant pathways, is yet another approach; however, the abstraction choices are critical and influence the pathways that are detected [26]. Parallel implementations [27], [28], [29] and algorithmic improvements are possible [30]. Recent analysis using a parallelized algorithm on Blue Gene/P of IBM enumerated approximately 49 million EFMs for the *S. cerevisiae* metabolic network consisting of 63 reactions and 83 compounds [27]. These results highlight the practical challenge in applying EFM analysis to comprehensive models of metabolism. Improvement in runtime and memory requirements are needed.

From an analysis and redesign perspective, not all elementary modes are of interest. Avoiding exhaustive enumeration provides an excellent and familiar alternative, similar to shortest and longest path delay analysis in timing analysis. Identifying a stoichiometrically balanced pathway was conceptually addressed within the context of pathway synthesis [32]. The Dominant Edge algorithm identifies a pathway containing the best thermodynamic bottleneck reaction, from a source metabolite to a destination metabolite using Gibbs free energy change as edge weights[31]. Results for several tests cases indicated that thermodynamically feasible paths are either identical, a proper subset, or overlapping with EFMs. The Dominant-Edge algorithm can be utilized with flux values as edge weights to identify a path that contains the flux-limiting reaction, or to find the pathway with the least flux variability. A method for decomposing a given flux distribution into a set of constituent elementary modes without initially generating a full set of elementary modes was proposed [33]. Pathway analysis methods circumventing explicit enumeration will play an important computational role in the future.

C. Optimization: Knock outs, Up/Down Regulation

Constrained optimization has played a fundamental role in developing computational design frameworks to systematically re-engineer cell behavior, where the goal is to identify reactions in the system whose fluxes should be changed, e.g. through gene up-regulation, down-regulation, or knockout, to achieve a flux distribution in the network that corresponds to a desired cellular objective. OptKnock identifies gene knockouts leading to targeted overproductions of chemicals in *E. coli* using a bi-level program that can be formulated as a mixed-integer linear program [21]. OptReg extends this framework to also consider not only knockouts, i.e. eliminations of reaction fluxes, but also up- and down-regulations of reaction fluxes [22]. Very recently, OptForce was introduced to distinguish between flux changes that are forced via genetic modifications and flux changes that are allowed by the natural variability of the system. The optimization goal of OptForce is to identify minimal sets of engineering modifications that must be imposed to overproduce a target metabolite above a desired threshold [34]. The comparison results provide information on the set of reactions whose fluxes must change to overproduce the target. While FBA-based techniques have significantly advanced strain optimization and were able to predict successful pathway modifications that were achieved experimentally by others, the techniques have limitations. In particular, multiple solutions may maximize the production of a particular target. Exploring the solution space and arriving at practical implementation remains difficult. Moreover, bi-level optimization is computationally inefficient [35]. OptGene finds gene deletions utilizing genetic algorithms to find multiple plausible solutions for the flux, and to solve a non-linear objective function, MOMA, Minimization of Metabolic Adjustments [36]. Optimizing for MOMA minimizes the Euclidian distance between the new fluxes and the pre-engineered ones. MOMA has been shown to support the hypothesis that knockout metabolic fluxes undergo a minimal redistribution with respect to the flux generation of the pre-engineered (wild type) system [37]. Genetic Design through Local Search (GDLS) is based on evolutionary algorithms and simulated annealing and was shown to have significant computational speed over Mixed Integer Linear Programming (MILP) approaches [35].

The precise controlling pathway flux through up/down regulation of enzyme activity is still not well understood. Metabolic Control Analysis (MCA) provides a quantitative framework to describe changes in substrate flux in response to changes in system parameters such as enzyme activities [38]. A flux control coefficient represents the enzyme's potential to affect a metabolite flux if that enzyme concentration is modified. Enzyme elasticity coefficients represent how a change in metabolite concentration changes the reaction rate. MCA provided several key insights. The flux control coefficient of a particular enzyme is a function of elasticity terms of other enzymes. Flux values are thus not dependent on the kinetic

properties of a single enzyme, but are a system property dependent on the kinetic properties of other enzymes. MCA is in essence a small-signal model, a common analysis technique to approximate the behavior of nonlinear devices at a particular DC bias point. While MCA provides a conceptual and quantitative framework, control coefficients computed through MCA are valid for specific steady states, and are not applicable with large changes in enzyme amounts, thus limiting the practical use of MCA in cell re-engineering. The need to characterize gene expression/flux relationships will become more pressing due to advances in synthetic biology and the need to tightly integrate gene circuitry within hosts [45].

D. De novo *Pathway Synthesis*

Pathway synthesis is the process of identifying a series of reactions to form a pathway to produce a desirable metabolite, such as a drug precursor, in a host organism. In some cases, the choice for a synthesis pathway is obvious. For example, there is only one known pathway for biosynthesis of 1,3-propanediol (a building block for synthetic polymers such as laminates and adhesives) from glycerol. This pathway consists of two reactions, each catalyzed by a singular enzyme. More generally, the number of alternative pathways for a given target may be too large for computational and experimental exploration, especially if the goal is to exploit the diversity of metabolic enzymes across many different organisms. A database such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), which currently lists over 8000 reactions, must be searched to produce the final product molecule from one or more reactant metabolites in the host organism. The search process needs to take into account not only the main reactants, but also cofactors.

Due to the combinatorial nature of the problem, an exhaustive search for candidate pathways is impractical. Over the past several years, a number of heuristic approaches have been developed for particular applications (e.g. predicting novel pathways for degradation of xenobiotics or biosynthesis of native and nonnative compounds). PathMiner constructs a synthesis pathway with the objective of minimizing the transformation cost in the biochemical state space where compounds define the states and transformations between compounds define state-transitions [39]. ReTrace is a graph-theoretical method that constructs plausible synthesis routes by searching for shortest paths between reactant-product pairs based on atom mapping [40]. PathPred is another method to construct plausible reaction pathways based on the chemical structure transformation patterns of small molecules [41]. PathPred specifically exploits the KEGG RPAIR database, which contains biochemical structure transformation patterns for substrate-product pairs (reactant pairs) of known enzymatic reactions. OptStrain uses mixed integer programming to identify stoichiometrically balanced pathways by adding or deleting reactions to selected host metabolic reaction networks [42]. A key advantage of this approach is to couple the selection of reactions with the ranking of the synthesis pathways in terms of theoretical yields. Success of the optimization

however critically depends on thoroughly pre-processing the database, which remains a non-trivial task. Another method for constructing synthesis pathways utilizes a graph-based probabilistic-search approach and ranking the pathways using FBA [43]. This approach is promising as when compared to an exhaustive search enumerating all possible reaction routes consisting of 10 reaction steps, the search returned nearly identical distributions of maximal yields, while requiring far less computing time.

Once candidate pathways are identified, deciding the best pathway within the selected host system should be based on a performance metric such as maximal predicted yield. This evaluation must assess whether the introduction of the synthesis pathway will negatively impact the host organism's capacity for balanced growth. There currently is a lack of data and consensus on the best synthesis pathway scoring methods. The number of pathway steps does not necessarily correlate with yield or the implementation practicality. Thermodynamic feasibility, which tries to compute the change in the Gibbs free energy of the reaction along the pathways, is another possible ranking metric. Metabolic burden is another such metric but poorly defined. Tighter integration between synthesis and evaluation, or host pre-characterizing, could computationally improve finding the optimal pathway.

IV. CONCLUSION

This review of computational tools to engineer biological systems is cast along traditional engineering tasks including modeling, analysis, synthesis and optimization. While some tasks bear conceptual similarities to those in Electronic Design Automation, modeling is fundamentally different. It aims to create mathematical models that explain existing complex design platforms. In contrast, models in electronics accurately capture hardware and software behaviors. Prototyping is occasionally needed to speed-up the design cycle. In biology, incomplete models and techniques that don't account for noise and variability are insufficient in enabling complete *in silico* design cycles. Screening and experimental analysis will continue to play a critical role in designing biology until modeling capabilities outpace experimental ones. Until then, a one-to-one correspondence in design methodologies for electronics and for biological systems is unlikely. Addressing grand challenges in health, food, energy and the environment however will require engineering sophisticated biological systems. Computational tools that analyze, optimize and explore large design spaces will play a complimentary role to experimental methods. Creating better predictive computational tools and developing future integrative tools that address the impact of gene regulatory networks on hosts and that co-optimize control and data flow will become a necessity.

REFERENCES

[1] W. C. Ruder, T. Lu, and J. J. Collins, "Synthetic biology moving into the clinic," *Science*, vol. 333, no. 6047, pp. 1248–1252, 2011.

[2] Y.-S. Jang, J. M. Park, S. Choi, Y. J. Choi, D. Y. Seung, J. H. Cho, and S. Y. Lee, "Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches." *Biotechnology advances*, 2011.

[3] Y. Wang, J. Wu, K. Skalina, and B. Pfeifer, "Complete biosynthesis of erythromycin a and designed analogs using E. coli as a heterologous host," *Chemistry and Biology*, vol. 17, no. 11, pp. 1232–1240, Nov 2010.

[4] R. P. Shetty, D. Endy, and T. F. Knight, "Engineering biobrick vectors from biobrick parts." *Journal of Biological Engineering*, vol. 2, 2008.

[5] J. C. Anderson, J. E. Dueber, M. Leguia, G. C. Wu, J. A. Goler, A. P. Arkin, and J. D. Keasling, "Bglbricks: A flexible standard for biological part assembly." *Journal of Biological Engineering*, vol. 4, no. 1, 2010.

[6] H. H. Wang, F. J. Isaacs, P. A. Carr, Z. Z. Sun, G. Xu, C. R. Forest, and G. M. Church, "Programming cells by multiplex genome engineering and accelerated evolution," *Nature*, vol. 460, no. 7257, pp. 894–898, Jul 2009.

[7] A. P. Arkin and D. A. Fletcher, "Fast, cheap and somewhat in control," *Genome Biol*, vol. 7, no. 8, p. 114, 2006.

[8] J. R. Anthony, L. C. Anthony, F. Nowroozi, G. Kwon, J. D. Newman, and J. D. Keasling, "Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene," *Metabolic Engineering*, vol. 11, no. 1, pp. 13–9, Jan 2009.

[9] Y. Y. Chen and C. D. Smolke, "From dna to targeted therapeutics: bringing synthetic biology to the clinic," *Science Translational Medicine*, vol. 3, no. 106, p. 106ps42, Oct 2011.

[10] H. Koeppl, D. Densmore, G. Setti, and M. E. di Bernardo, *Design and Analysis of Biomolecular Circuits*. Springer, 2011.

[11] M. W. Lux, B. W. Bramlett, D. A. Ball, and J. Peccoud, "Genetic design automation: Engineering fantasy or scientific renewal?" *Trends in Biotechnology*, pp. 1–7, 2011.

[12] C. Myers, *Engineering Genetic Circuits*. Chapman & Hall/CRC, 2009.

[13] W. Liebermeister and E. Klipp, "Bringing metabolic networks to life: convenience rate law and thermodynamic constraints." *Theoretical biology & medical modelling*, vol. 3, p. 41, 2006.

[14] A. M. Feist and B. Ø. Palsson, "The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli." *Nature biotechnology*, vol. 26, pp. 659–67, 2008.

[15] I. Thiele and B. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction," *Nature Protocols*, vol. 5, pp. 93–121, 2010.

[16] C. S. Henry, M. DeJongh, A. a Best, P. M. Frybarger, B. Linsay, and R. L. Stevens, "High-throughput generation, optimization and analysis of genome-scale metabolic models." *Nature biotechnology*, vol. 28, pp. 977–82, 2010.

[17] R. S. Costa, D. Machado, I. Rocha, and E. C. Ferreira, "Critical perspective on the consequences of the limited availability of kinetic data in metabolic dynamic modelling." *IET systems biology*, vol. 5, no. 3, pp. 157–163, 2011.

[18] 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. 2467–74, 2003.

[19] D. A. Beard, S. D. Liang, and H. Qian, "Energy balance for analysis of complex metabolic networks," *Biophysical Journal*, vol. 83, no. 1, pp. 79–86, Jul 2002.

[20] M. D. Jankowski, C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Group contribution method for thermodynamic analysis of complex metabolic networks," *Biophysical Journal*, vol. 95, no. 3, pp. 1487–99, Aug 2008.

[21] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Thermodynamics-based metabolic flux analysis," *Biophysical Journal*, vol. 92, no. 5, pp. 1792–805, Oct 2007.

[22] A. A. Alonso and J. R. B. José-Oscar H. Sendín, "Multi-objective optimization of biological networks for prediction of intracellular fluxes," *International Workshop on Practical Applications of Computational Biology and Bioinformatics*, vol. 49, pp. 197–205, 2009.

[23] N. Price, J. Reed, and B. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints," *Nature Reviews Microbiology*, vol. 2, no. 11, pp. 886– 97, 2004.

[24] M. Yeung, I. Thiele, and B. O. Palsson, "Estimation of the number of extreme pathways for metabolic networks," *BMC Bioinformatics*, vol. 8, no. 1, p. 363, 2007.

[25] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar, "Exploring the pathway structure of metabolism: decomposition into subnetworks and application to mycoplasma pneumoniae," *Bioinformatics*, vol. 18, no. 2, pp. 351–61, Feb 2002.

[26] C. Kaleta, L. F. de Figueiredo, and S. Schuster, "Can the whole be less than the sum of its parts? pathway analysis in genome-scale metabolic networks using elementary flux patterns," *Genome Res*, vol. 19, no. 10, pp. 1872–83, Oct 2009.

[27] D. Jevremović, C. Trinh, F. Srienc, C. Sosa, and D. Boley, "Parallelization of nullspace algorithm for the computation of metabolic pathways," *Parallel computing*, vol. 37, pp. 261–278, 2011.

[28] S. Klamt, J. Gagneur, and A. von Kamp, "Algorithmic approaches for computing elementary modes in large biochemical reaction networks," *Syst Biol (Stevenage)*, vol. 152, no. 4, pp. 249–55, Dec 2005.

[29] N. Samatova, A. Geist, G. Ostrouchov, and M. A, "Parallel out-of-core algorithm for genome-scale enumeration of metabolic systemic pathways," *International Parallel and Distributed Processing Symposium*, 2002.

[30] R. Urbanczik and C. Wagner, "An improved algorithm for stoichiometric network analysis: theory and applications," *Bioinformatics*, vol. 21, no. 7, pp. 1203–10, Apr 1 2005.

[31] E. Ullah, K. Lee, and S. Hassoun, "A weighted graph algorithm for identifying dominant-edge metabolic pathways," in *International Conference on Computer-Aided Design*, Nov, 2009, pp. 144 – 150.

[32] M. L. Mavrovouniotis, G. Stephanopoulos, and G. Stephanopoulos, "Computer-aided synthesis of biochemical pathways," *Biotechnol Bioeng*, vol. 36, no. 11, pp. 1119–32, Dec 1990.

[33] K. Ip, C. Colijn, and D. S. Lun, "Analysis of complex metabolic behavior through pathway decomposition," *BMC Systems Biology*, vol. 5, no. 1, 2011.

[34] S. Ranganathan, P. F. Suthers, and C. D. Maranas, "OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions," *PLoS Computational Biology*, vol. 6, no. 4, Apr 2010.

[35] D. S. Lun, G. Rockwell, N. J. Guido, M. Baym, J. A. Kelner, B. Berger, J. E. Galagan, and G. M. Church, "Large-scale identification of genetic design strategies using local search," *Molecular Systems Biology*, vol. 5, no. 296, 2009.

[36] 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," *Curr Opin Biotechnol*, vol. 19, no. 6, pp. 556–63, Dec 2008.

[37] D. Segre, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proc Natl Acad Sci U S A*, vol. 99, no. 23, pp. 15 112–7, Nov 2002.

[38] D. Fell, *Understanding the Control of Metabolism*. Portland Pr, 1997.

[39] D. C. McShan, S. Rao, and I. Shah, "Pathminer: predicting metabolic pathways by heuristic search," *Bioinformatics*, vol. 19, pp. 1692–1698, Sep 2003.

[40] E. Pitkanen, P. Jouhten, and J. Rousu, "Inferring branching pathways in genome-scale metabolic networks," *BMC Syst Biol*, vol. 3, p. 103, 2009.

[41] Y. Moriya, D. Shigemizu, M. Hattori, T. Tokimatsu, M. Kotera, S. Goto, and M. Kanehisa, "Pathpred: an enzyme-catalyzed metabolic pathway prediction server," *Nucleic Acids Research*, vol. 38, pp. 138–143, Jul 2010.

[42] 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.

[43] M. Yousofshahi, K. Lee, and S. Hassoun, "Probabilistic pathway construction," *Metabolic Engineering*, vol. 13, no. 4, pp. 435–44, Jul 2011.

[44] A. M. Weeks and M. C. Chang, "Constructing de novo biosynthetic pathways for chemical synthesis inside living cells," *Biochemistry*, vol. 50, no. 24, pp. 5404–18, Jun 2011.

[45] P. M. Boyle and P. A. Silver, "Parts plus pipes: Synthetic biology approaches to metabolic engineering," *Metabolic Engineering*, 2011.