

journal homepage: www.elsevier.com/locate/csbj

Mini Review

Computational approaches to metabolic engineering utilizing systems biology and synthetic biology

Stephen S. Fong*

Department of Chemical and Life Science Engineering, Virginia Commonwealth University, 601 W. Main St., Richmond, VA 23284, United States

ARTICLE INFO

Available online 27 August 2014

Keywords:

Metabolic engineering
Genome-scale modeling
Synthetic biology
Computational design
Biotechnology

ABSTRACT

Metabolic engineering modifies cellular function to address various biochemical applications. Underlying metabolic engineering efforts are a host of tools and knowledge that are integrated to enable successful outcomes. Concurrent development of computational and experimental tools has enabled different approaches to metabolic engineering. One approach is to leverage knowledge and computational tools to prospectively predict designs to achieve the desired outcome. An alternative approach is to utilize combinatorial experimental tools to empirically explore the range of cellular function and to screen for desired traits. This mini-review focuses on computational systems biology and synthetic biology tools that can be used in combination for prospective in silico strain design. © 2014 Fong. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	28
2. Main Text	29
2.1. Genome-scale Models and Metabolic Engineering	29
2.2. Synthetic Biology	31
2.3. Transcription	31
2.4. Translation	32
3. Summary and Outlook	32
References	33

1. Introduction

One of the central challenges to biology is understanding biological information flow such as how genotypes manifest into functional phenotypes. Historically, biological experiments have been difficult to conduct leading to a scarcity of data. However, technological improvements in experimental high-throughput measurements have shifted biology to being a data-rich field, driving a need for analytical tools to facilitate the analysis and interpretation of biological data.

Metabolic engineering applies biological information to genetically modify cellular function, usually toward production of a targeted chemical or protein product. Metabolic engineering research requires knowledge of integrated cellular function and molecular detail to be successful. Broadly, there are two types of approaches that lead to successful metabolic engineering results: directed designs built upon knowledge or

combinatorial screening that leverages high-throughput experimental techniques. Within the past fifteen years, computational tools have been developed to leverage biological data in the analysis and design of microbial strains for metabolic engineering and have facilitated prospective metabolic engineering design. These tools began with genome-scale metabolic models that aid in the analysis and prediction of whole cell function and have expanded to include tools for predicting the function of specific DNA sequences.

Here, a brief overview is presented on the development and progression of computational tools that can be applied to metabolic engineering. Individual fields of systems biology, synthetic biology, computational biology, or metabolic engineering are expansive enough for multiple reviews. The specific focus of this mini-review is to focus on a subset of tools from systems biology and synthetic biology that can be used in combination to enable prospective in silico strain design. Key developments associated with genome-scale metabolic models and algorithms that can be used to computationally propose microbial strain design will be discussed. Specific developments in synthetic biology associated

* Tel.: +1 804 827 7038; fax: +1 804 828 3846.
E-mail address: ssfong@vcu.edu.

with transcriptional and translational control will also be presented and placed within the context of genome-scale modeling and metabolic engineering.

2. Main Text

Systems biology emphasizes data-intensive, integrative analyses that account for extended network function. With the introduction of whole genome sequencing and genomics technologies, one of the first objectives was to develop methods utilizing genomic information to understand and predict phenotypic function. The constraint-based modeling approach [1] was implemented to generate genome-scale metabolic models of some of the first organisms with genome sequences [2–4], demonstrating the conceptual value of this computational approach. The initial genome-scale models were constructed based upon genomic data (sequence information) and biochemical data (reaction stoichiometry) in conjunction with linear programming to apply mass balancing principles to a whole-cell system. The conceptual framework provided a context for analyzing attributes of a cellular system [5] and was shown to be able to predict cellular growth phenotypes [6,7].

Since the initial development of genome-scale models, a wide variety of improvements have been made to address different needs. These range from understanding the underlying structure of networks by using elementary modes [8] or extreme pathways [9], model-building approaches [10,11], and progressively more cellular detail including thermodynamics [12], transcriptional regulation [13,14], and signaling pathways [15]. All of these have contributed to improve the predictive capability and accuracy of genome-scale metabolic models and can be used to study a variety of aspects of cellular systems. Here, we are specifically going to focus on the existing tools and challenges associated with genome-scale metabolic models, particularly as they apply to metabolic engineering applications. An overview of computational tools presented here is shown in Table 1, which only represents a select subset of the numerous available tools and algorithms that have been developed. A more comprehensive (and continually updated) list of tools associated with constraint-based models is curated online (cobramethods.wikidot.com).

2.1. Genome-scale Models and Metabolic Engineering

Using the natural ability of genome-scale metabolic models to simulate the behavior of cellular metabolism one can predict cellular designs for maximizing chemical production. Metabolic engineering goals of identifying and modifying pathway fluxes to optimize the production of a desired chemical product align well with the pathway-level predictions that are generated from a genome-scale model. The foundation for this work was demonstrated when it was shown that the constraint-based modeling approach could reasonably predict the cellular growth phenotypes resulting from genetic modifications (gene deletions) [16]

in *Escherichia coli*. This quickly led to a demonstration of using a genome-scale model of *E. coli* to predict strain designs for the over-production of lactic acid [17], which set the stage for genome-scale models as powerful computational tools for strain design.

The first iterations of combining computer-aided strain design with experimental implementation relied on strain designs that incorporated gene deletions. This approach was computationally achievable through the removal of pathways associated with genes (following gene–protein–reaction relationships) and could be achieved experimentally with established methods for targeted gene deletions using homologous recombination [18]. Initial results were promising from the standpoint that the designs improved overall production of the desired chemical, but there was still a quantitative mismatch between the computationally calculated theoretical yield and the experimental yield.

The discrepancy between computationally predicted function and actual function led to the development of an algorithm to predict targets for iterative improvement of the experimental strain [19]. By utilizing transcriptomic data of the experimental strain, algorithmic analysis predicted areas of metabolism with the largest difference between the theoretical and experimental function. This analysis predicted specific genes to be targeted for synthetic regulation of gene expression (increased or decreased expression). The problem remained in connecting the computational prediction with tools for direct experimental implementation. Recently, several developments have occurred in parallel both computationally and experimentally.

For constraint-based genome-scale metabolic models, new methodologies and analyses continue to be developed that improve the accuracy of these models to predict cellular phenotypes. One major consideration for genome-scale metabolic models is that the mathematical representation for a biological system is underdetermined and thus, the same cellular phenotype can be reproduced from different underlying flux states/pathway usage. This problem complicates metabolic engineering design. For example, a normal growth phenotype may have numerous proposed flux states that vary in specific pathway use, but produce the same cellular growth rate/product yield. However, once genetic modification of the network is implemented, all of the possible flux states may no longer be functionally equivalent. When considering growth phenotypes at the level of cellular growth, it may not be necessary to explicitly identify the exact flux state of the cell. Knowledge of the starting in vivo flux state is important for pathway-specific metabolic engineering design.

The problem of identifying in vivo flux states within the context of genome-scale metabolic models has been approached using a combination of high-throughput experimental data and computational algorithms. The initial formulation of this approach used transcriptomic or proteomic data with a human metabolic model to identify tissue-specific metabolic differences [20]. In this approach, the experimental data was translated to a binary present/absent scoring for each individual transcript/protein. The scored experimental data was then algorithmically

Table 1
Overview of computational tools discussed.

Tool	Description	Reference
OMNI	Reconciles discrepancies between in silico and in vivo phenotypes using transcriptomics	[19]
MILP	Refined flux state predictions based upon high-throughput experimental data	[20]
E matrix	Prediction of gene and protein expression levels	[25]
DFBA	Dynamic flux balance analysis	[27]
OptCom	Multi-level optimization for modeling microbial consortia	[30]
d-OptCom	Dynamic variant of OptCom	[31]
OptKnock	Bi-level optimization for strain design using gene deletions	[32]
EMLiO	Strain design incorporating increased/decreased gene expression	[37]
CosMos	Flux-based strain design	[38]
CASOP	Strain design using elementary modes	[39]
FBrAtio	Strain design based upon flux ratios at critical nodes	[40]
k-OptForce	Strain design incorporating substrate-level inhibition	[41]
DySScO	Strain design incorporating process kinetics	[42]
PWM	Prediction of DNA sequence variation on promoter strength	[48]
RBS calculator	Prediction of protein translation initiation rates	[50]

integrated with the metabolic model framework using mixed integer linear programming (MILP) to calculate a flux state that is parsimonious with the experimental data. The underlying principle seeks to maximize the agreement between experimental data and computational fluxes where experimentally present entities can carry flux and experimentally absent entities should have zero flux.

The MILP approach to experimental data integration and prediction of in vivo flux states has started to be adopted for metabolic engineering purposes. One example of this is the development and analysis of a metabolic model for the cellulolytic anaerobe, *Clostridium thermocellum*. *C. thermocellum* naturally produces a cellulosome that efficiently breaks down cellulosic biomass in carbohydrate monomers that *C. thermocellum* can ferment into ethanol and hydrogen as metabolic end-products. Thus, *C. thermocellum* has been a focal organism for biofuel development. After developing a *C. thermocellum* model that explicitly produces a cellulosome [21], RNA sequencing (RNAseq) data was obtained to study the transcriptome. MILP integration of the RNAseq data with the metabolic model helped improve the accuracy of computational predictions and helped identify genetic targets to improve biofuel production [22].

Another major improvement in genome-scale metabolic modeling was the explicit inclusion of charge-balancing for biochemical reactions. While the initial formulation of constraint-based models was based upon material balances, reactions were not necessarily charge-balanced with-out careful curation of individual reactions. This was first implemented with the model for *E. coli* [23] and has become standard practice for constraint-based model reconstruction. This detailed aspect of the models becomes integral to metabolic engineering especially in cases involving fermentation where restricted energetics lead to higher fluxes and distributions of metabolic end-products need to maintain redox balance for the cell. This was shown to be particularly important for organisms with growth phase shifts such as the commercial organism *Clostridium acetobutylicum* that is central to the acetone–butanol–ethanol process. *C. acetobutylicum* goes through distinct acidogenic and solventogenic growth phases with widely different metabolic end-products produced in each phase. Critical to understanding and modeling the shift and difference in end-products produced was the development and analysis of proton fluxes for *C. acetobutylicum* [24].

One of the most recent major conceptual advances in the constraint-based modeling methodology was the formulation of the Expression matrix, or E matrix, in *E. coli* [25,26]. Constraint-based models to this point have been developed using biochemical reaction stoichiometry to form a stoichiometric matrix (S) that provides and establishes a gene–protein–reaction (GPR) relationship for metabolic functions. The stoichiometric matrix provides the basis for all simulations utilizing flux balance analysis (FBA), but considers the metabolic network as an abstraction from the cell where basically any gene product can be produced at any amount with no cellular cost. The E matrix represents a major advancement in detail and prediction as it explicitly accounts for all mechanisms required for transcription, translation, and modification of each gene product. Along with adding a new level of mechanistic biological detail, the development of the E matrix enables the prediction of two important aspects of cellular function. One is that the E matrix allows a priori prediction of the transcriptome and proteome directly based upon a genetic sequence. The second major benefit is that cellular costs can be associated with producing each gene product, thus the cellular cost–benefit aspect of expressing pathways (e.g. heterologous pathways) can now be considered.

Based upon the stoichiometric matrix (or even including the expression matrix), standard simulations using genome-scale metabolic models utilize flux balance analysis (FBA) as a staple for making predictions of pathway fluxes and phenotypes. This analysis assumes a pseudo-steady state hypothesis and does not explicitly account for any process kinetics. A methodological improvement to FBA was the development of dynamic flux balance analysis (DFBA) [27]. DFBA was initially developed considering two different approaches to explicitly integrate

kinetics into FBA simulations. The first approach was to reformulate the problem to account for metabolite concentrations and the dynamic optimization problems were solved using non-linear programming over the entire designated time period of interest. The second approach discretized the time period of interest into smaller time steps that were solved individually using linear programming with a final integration of time steps. Both approaches extended the analytical capabilities of the *E. coli* model when used to study diauxic growth on glucose [27].

The concept of DFBA has been expanded to also explicitly include Michaelis–Menten kinetics for processes where reasonable rate parameters could be found and used as a basis for modeling microbial consortia. Designer microbial consortia are gaining interest as a potential means of efficiently utilizing some of the unique capabilities of different microbes. In the cellulosic biorefinery for example, since many of the highly efficient cellulolytic microbes are strict anaerobes and have poor tools for genetic manipulation the process consists of two steps: saccharification by a cellulolytic microbe and fermentation of freed sugars to a metabolic product by a genetically tractable microbe. Thus, a single-step consolidated bioprocess (CBP) could conceptually be designed from a microbial consortium with an anaerobic cellulolytic microorganism for the degradation of lignocellulosic material, a supporting facultative anaerobe that can be used to consume oxygen to produce an anaerobic environment and cross-feed off of secreted metabolic end-products to produce a desired chemical product.

Microbial consortia are often difficult to propagate and stabilize in a laboratory setting making experimental approaches to consortium design difficult. Ideally, genome-scale metabolic models can be used to study and predict microbial consortium behavior. While the mass balance conceptual framework of metabolic inputs and outputs is foundational to computational study of microbial consortia, analytical frameworks needed to be developed to interface individual organism models. To date, there have been two noted formulations of constraint-based modeling approaches applied to microbial consortia. The first was an iteration of DFBA that modeled a co-culture of *C. acetobutylicum* and *Clostridium cellulolyticum* [28]. This formulation was used as a basis for studying some of the interactions between the two *Clostridia* to elucidate some of the synergies that arise from co-culture [29]. Independently, a separate multi-level optimization computational framework known as OptCom was developed as a general and scalable means of studying microbial communities and the interactions within those communities [30]. The OptCom framework considers the community by splitting optimization between individual organism fitness considerations and the overall community fitness as a secondary optimization. The OptCom framework was later expanded to explicitly incorporate dynamic behavior in the d-OptCom framework [31]. The d-OptCom framework was initially used to study auxotrophic strains of *E. coli* and a microbial community composed of *Geobacter sulfurreducens*, *Rhodospirillum rubrum*, and *Shewanella oneidensis* that reduce uranium [31].

The final analytical component when using genome-scale metabolic models as a tool for microbial strain design is an algorithm that can facilitate the design process. A growing number of algorithms have been developed that expand the predictive capabilities of genome-scale models to simulate different strain design parameters. One of the first strain design algorithms to be developed for use with genome-scale metabolic models was OptKnock [32] that formulated a bi-level optimization where gene deletions were considered to increase the production of a desired chemical while maintaining cellular growth. Underlying the algorithm development was the premise that secretion of a target chemical could be stoichiometrically coupled to growth such that the faster a cell grew the faster the chemical would be produced. *E. coli* strains designed by OptKnock to secrete lactic acid experimentally demonstrated that it is possible to couple growth with chemical secretion. Subsequent to OptKnock, several variant algorithms have been developed including OptStrain [33], OptORF [34], OptFlux [35], and OptForce [36].

In addition to the family of algorithms related to OptKnock, there have been numerous variant algorithms developed to address computational or algorithmic limitations. EMILiO [37] was developed as a computationally efficient algorithm for searching through a broader range of possible genetic manipulations. CosMos [38] developed designs for a variety of manipulations, but considered the problem based upon fluxes (and flux modifications). Elementary modes were used as the design basis for identifying knockout or overexpression targets in the CASOP [39] method and relative flux ratios were used as a design criterion for FBrAtio [40]. The recent algorithms such as k-OptForce [41] and DySScO [42] have incorporated process kinetics into the design consideration.

Using any of these approaches allows the prediction of how genetic modification will impact an organism production of a target chemical. These predictions even provide sufficient detail to identify on a pathway-by-pathway basis the exact flux that is required to achieve the optimal production rate. The challenge has remained in the translation of the computationally predicted strain design (and corresponding optimal flux state) to the in vivo system. In other words, if you can identify exactly what pathways you want to express and exactly what flux you want through each of those pathways, how do you get your cell to exactly match the computational prediction?

2.2. Synthetic Biology

Experimentally, the rapid development of molecular biology and genetic engineering tools typified by synthetic biology has significantly increased the ability to explicitly design and implement controllable genetic constructs. These improvements were enabled by the technological advances in DNA synthesis and rapid, high-fidelity assembly of DNA fragments [43–45]. Once genetic circuits could be constructed in an efficient and inexpensive manner with base-by-base specificity, it was possible to interrogate sequence-to-function relationships. While central to the investigation of all biological functions, for metabolic engineering applications this represented a broad advance where finer control over the expression of genes in a given pathway could be designed and implemented. Designed expression control for individual genes typically occurs either at the transcription or translation levels. Developments in synthetic biology have occurred in parallel with

systems biology developments (Fig. 1) and enable multi-scale approaches to metabolic engineering.

2.3. Transcription

The first level of functional control for specific genes occurs during transcription. The transcriptional process involves a binding event between a DNA sequence (promoter) and RNA polymerase to initiate polymerization of an mRNA. When naturally occurring promoter sequences are analyzed, there are conserved sequence motifs corresponding to regions that physically bind to the sigma subunit of the RNA polymerase. Sequence variation in the promoter affects transcriptional strength; however, it has been difficult to predict the binding interaction between a promoter and RNA polymerase as it involves binding between DNA and a folded protein.

Some of the early work in this area was empirically based (out of necessity). Naturally occurring variations in DNA sequences associated with transcription were identified and characterized in terms of transcriptional effect (strength). Identification of these DNA sequences was the basis for constructing and characterizing libraries of transcription-associated DNA sequences. In the early stages of the development of synthetic biology and the Registry of Standard Biological Parts (parts.igem.org), transcriptional DNA parts were created and cataloged allowing for more detailed characterization and use of genetic variants for transcriptional control [46].

The decreased cost of direct DNA synthesis further expedited the investigation of sequence variation to function for transcriptional control. Base-by-base changes in promoter (or UP element) sequences could be accurately synthesized and rapidly tested in synthetic DNA constructs. Controlled sequence-to-function relationships could then be extrapolated using mathematical correlation methods such as a position weight matrix (PWM).

PWMs are a mathematical representation of a pattern and were initially applied to biological systems as a quantitative means of investigating conserved DNA sequences [47]. Recently, PWMs have been used to quantitatively describe the sequence-to-function relationship for promoters in *E. coli* [48]. By dividing promoter sequences into 6 motifs (−35, spacer, −10, disc start, initial transcribed region), 60 different promoter sequences were characterized for promoter strength and used to evaluate the influence of specific genetic changes on function [48].

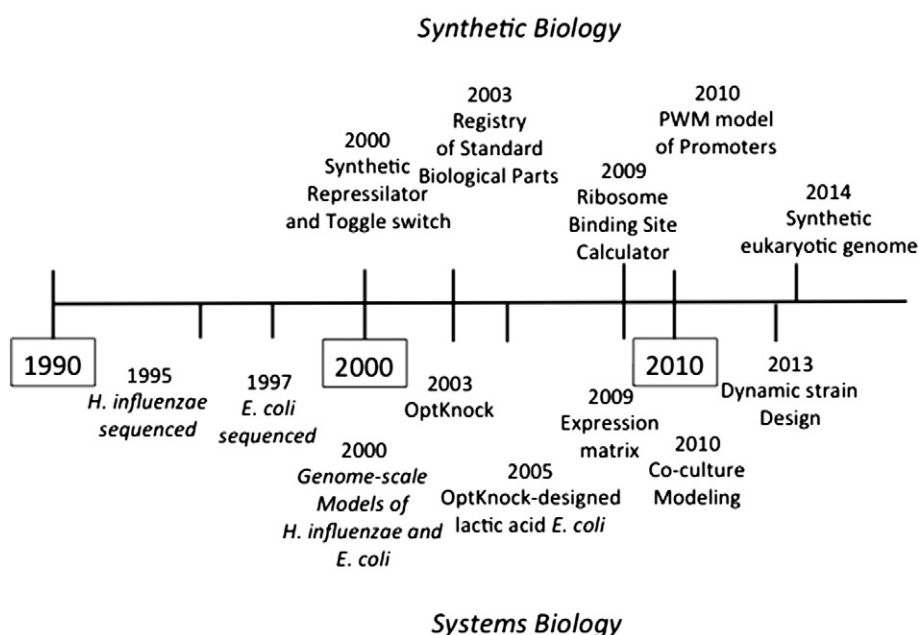


Fig. 1. Timeline showing several major developments in synthetic biology and systems biology of significance to metabolic engineering applications.

This approach was also applied to a class of transcriptional modulating sequences called UP elements [49], which occur upstream of core promoter sequences. In these cases, PWMs were used to mathematically model the relationship between sequence and function. Thus, with the recent modeling developments, transcriptional control has progressed from empirical understanding to quantitatively modeled predictions that can direct synthetic promoter design to produce a desired level of transcriptional strength.

2.4. Translation

Following transcription, mRNA is translated into proteins. Translation initiates when a ribosome interacts with a ribosome binding site (RBS) and facilitates the subsequent tRNA binding to mRNA codons to produce polypeptides by the addition of amino acids. Translation involves three steps: Initiation, elongation, and termination. Of these three steps, translation initiation is the rate-limiting step and within each cell different rates of translation initiation are found due to variation in the DNA sequence of the RBS. Thus, the functional amount of protein produced in terms of translation is tied to the thermodynamics of RBS–ribosome interaction.

Recently, a computational approach has been developed to predict translation initiation rates for all start codons in a given DNA sequence based upon a thermodynamic calculation of Gibbs free energy [50,51]. This calculation specifically considers the interaction of the 30S ribosomal subunit with a specific mRNA sequence. The total change in Gibbs free energy (ΔG_{Tot}) is overall calculated as the difference in Gibbs free energy between the 30S subunit complex bound to the mRNA sequence and the 30S subunit and mRNA sequence with no interaction. This “Ribosome Binding Site Calculator” [50] can be used not only to predict translation initiation rates for existing sequences, but also to design de novo RBS sequences for synthetically controlling translated protein levels.

Synthetic biology has now developed a complement of experimental and computational tools to design and control individual gene expression levels at both the transcriptional and translational levels [52].

These tools enable a finer level of design control for biological systems and can be implemented for metabolic engineering applications where the production of a desired product is typically linked to carefully balancing multiple pathway fluxes. Progress in synthetic biology is providing a suite of molecular-level tools that is a natural complement to network/systems-level analyses.

3. Summary and Outlook

As a field, metabolic engineering leverages the vast diversity of biochemical capabilities found in biological systems to produce a desired product. Successful metabolic engineering projects typically balance systems-level strain design and molecular-level detail to modify cell function. As such, metabolic engineering projects often incorporate the most recent biological knowledge and tools as these projects span all levels of biological organization and function within a cell. Due to the complexity and multi-scale nature of cellular systems computational modeling methods are instrumental for analyzing and predicting function, Fig. 2.

For system-level design, genome-scale metabolic models have developed into a robust computational tool for predicting consequences of network modifications on cellular function. Building upon the stoichiometric foundation underlying genome-scale metabolic models, numerous improvements have been made in terms of model contents and analytical tools. It is now possible to computationally simulate the major cellular components starting with a genome through mechanistic expression and function of proteins. These simulations can be used for prospective computational design of microbial strains with modified function.

Depending upon the specific strain design, experimental implementation can involve the combinations of gene deletions, gene additions (heterologous expression), gene knockdown, or gene over-expression. At a molecular-level, it becomes important to understand how best to achieve the desired gene construct and how to experimentally implement the construct to achieve the desired expression level. Work in synthetic biology has started to address both of these aspects through the

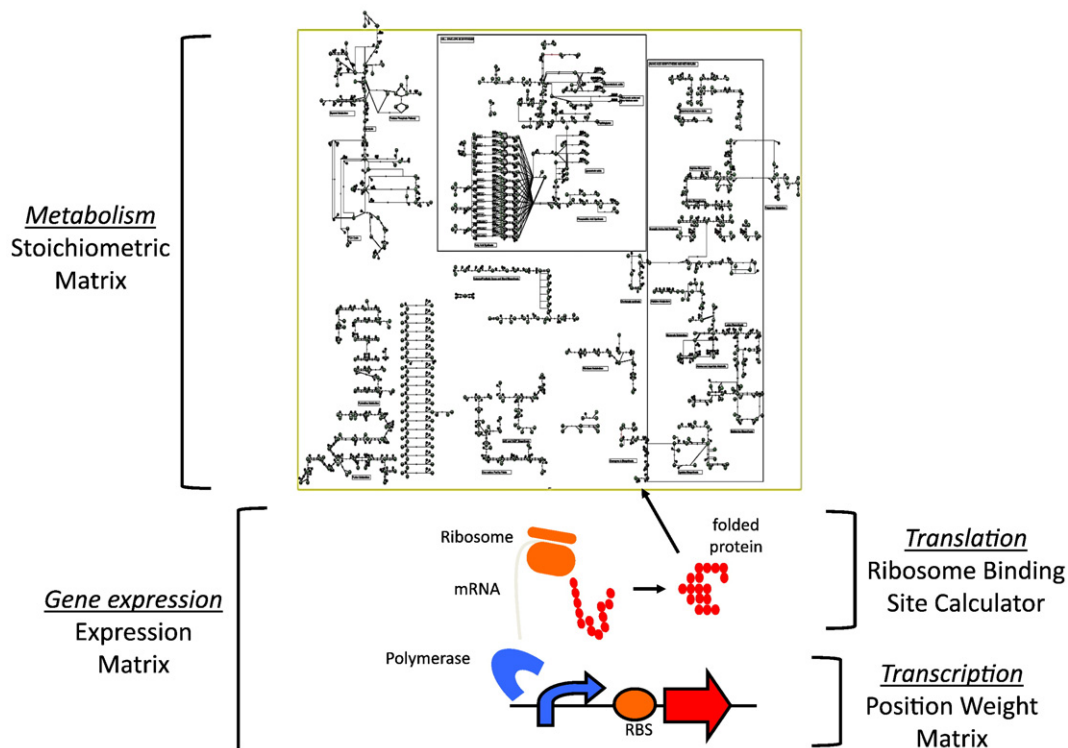


Fig. 2. Schematic depiction of multi-scale biological processes and modeling approaches used to analyze each process.

development of molecular tools for the design and synthesis of DNA constructs and specific tools for transcription and translation. These molecular-level tools allow DNA sequences to be designed to provide desired levels of transcription and translation to achieve desired levels of protein function. Furthermore, the designed sequences can then be directly implemented experimentally using DNA synthesis and assembly methods.

Together, it is now possible to approach metabolic engineering projects using multi-scale computational modeling tools to direct all major facets of experimental research. Metabolic network analysis and design can be achieved using genome-scale models and optimization algorithms. Specific desired pathway fluxes predicted during network analysis can then be individually considered using molecular tools. Control of transcription and translation of each individual gene can then be designed and genetically encoded into a synthetic DNA construct for implementation in an engineered microbial strain.

Progressively, the complexity of strain designs that can be implemented will also increase opening avenues for targeting difficult products or processes. Synthetic feedback loops or embedded biosensors can be used as built-in control mechanisms for monitoring or triggering cellular processes. Controlling the kinetics or throughput of upstream and downstream pathways can minimize the effects of chemically toxic intermediates. It may even be possible to completely redesign [53] or refactor [54,55] major components of metabolism to more efficiently utilize resource pools to minimize material drains going to biomass. Undoubtedly, there will be a growing number of studies that successfully incorporate and utilize systems biology and synthetic biology such as the development of sequence-expression-activity maps (SEAMAPs) [56] that help identify optimal sequences and expression levels.

References

- Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2007;2:727–38. <http://dx.doi.org/10.1038/nprot.2007.99>.
- Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, et al. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol* 2002;184:4582–93. <http://dx.doi.org/10.1128/JB.184.16.4582-4593.2002>.
- SCHILLING CH, PALSSON BØ. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a Genome-scale pathway analysis. *J Theor Biol* 2000;203:249–83. <http://dx.doi.org/10.1006/jtbi.2000.1088>.
- Edwards JS, Palsson BØ. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci* 2000;97:5528–33. <http://dx.doi.org/10.1073/pnas.97.10.5528>.
- Price ND, Reed JL, Palsson BØ. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* 2004;2:886–97. <http://dx.doi.org/10.1038/nrmicro1023>.
- Edwards JS, Ibarra RU, Palsson BØ. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 2001;19:125–30. <http://dx.doi.org/10.1038/84379>.
- Fong SS, Marciniak JY, Palsson BØ. Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale in silico metabolic model. *J Bacteriol* 2003;185:6400–8. <http://dx.doi.org/10.1128/JB.185.21.6400-6408.2003>.
- Trinh CT, Wlaschin A, Srien F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol* 2009;81:813–26. <http://dx.doi.org/10.1007/s00253-008-1770-1>.
- Price ND, Papin JA, Palsson BØ. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. *Genome Res* 2002;12:760–9. <http://dx.doi.org/10.1101/gr.218002>.
- Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, et al. Automated genome annotation and metabolic model reconstruction in the SEED and model SEED. In: Alper HS, editor. *Systems Metabolic Engineering. Methods in Molecular Biology*. Humana Press; 2013. p. 17–45 [Available: http://link.springer.com/protocol/10.1007/978-1-62703-299-5_2. Accessed 29 May 2014].
- Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 2010;28:977–82. <http://dx.doi.org/10.1038/nbt.1672>.
- Beard DA, Babson E, Curtis E, Qian H. Thermodynamic constraints for biochemical networks. *J Theor Biol* 2004;228:327–33. <http://dx.doi.org/10.1016/j.jtbi.2004.01.008>.
- Covert MW, Palsson BØ. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem* 2002;277:28058–64.
- COVERT MW, SCHILLING CH, PALSSON B. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* 2001;213:73–88. <http://dx.doi.org/10.1006/jtbi.2001.2405>.
- Papin JA, Palsson BØ. The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis. *Biophys J* 2004;87:37–46. <http://dx.doi.org/10.1529/biophysj.103.029884>.
- Fong SS, Palsson BØ. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet* 2004;36:1056–8. <http://dx.doi.org/10.1038/ng1432>.
- Fong SS, Burgard AP, Herrig CD, Knight EM, Blattner FR, et al. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005;91:643–8. <http://dx.doi.org/10.1002/bit.20542>.
- Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci* 2000;97:6640–5. <http://dx.doi.org/10.1073/pnas.120163297>.
- Herrgård MJ, Fong SS, Palsson BØ. Identification of genome-scale metabolic network models using experimentally measured flux profiles. *PLoS Comput Biol* 2006;2:e72. <http://dx.doi.org/10.1371/journal.pcbi.0020072>.
- Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 2008;26:1003–10. <http://dx.doi.org/10.1038/nbt.1487>.
- Roberts SB, Gowen CM, Brooks JP, Fong SS. Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Syst Biol* 2010;4:31. <http://dx.doi.org/10.1186/1752-0509-4-31>.
- Gowen CM, Fong SS. Genome-scale metabolic model integrated with RNAseq data to identify metabolic states of *Clostridium thermocellum*. *Biotechnol J* 2010;5:759–67. <http://dx.doi.org/10.1002/biot.201000084>.
- Reed JL, Vo TD, Schilling CH, Palsson BØ. An expanded genome-scale model of *Escherichia coli* K-12 (ijR904 GSM/GPR). *Genome Biol* 2003;4:R54. <http://dx.doi.org/10.1186/gb-2003-4-9-r54>.
- Senger RS, Papoutsakis ET. Genome-scale model for *Clostridium acetobutylicum*: part II. Development of specific proton flux states and numerically determined sub-systems. *Biotechnol Bioeng* 2008;101:1053–71. <http://dx.doi.org/10.1002/bit.22009>.
- Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, et al. Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PLoS One* 2012;7:e45635.
- Thiele I, Jamshidi N, Fleming RMT, Palsson BØ. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol* 2009;5:e1000312. <http://dx.doi.org/10.1371/journal.pcbi.1000312>.
- Mahadevan R, Edwards JS, Doyle III FJ. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 2002;83:1331–40. [http://dx.doi.org/10.1016/S0006-3495\(02\)73903-9](http://dx.doi.org/10.1016/S0006-3495(02)73903-9).
- Salimi F, Zhuang K, Mahadevan R. Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. *Biotechnol J* 2010;5:726–38. <http://dx.doi.org/10.1002/biot.201000159>.
- Mahadevan R, Henson MA. Genome-based modeling and design of metabolic interactions in microbial communities. *Comput Struct Biotechnol J* 2012;3:e201210008. <http://dx.doi.org/10.5936/csbj.201210008>.
- Zomorodi AR, Maranas CD. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol* 2012;8:e1002363. <http://dx.doi.org/10.1371/journal.pcbi.1002363>.
- Zomorodi AR, Islam MM, Maranas CD. d-OptCom: dynamic multi-level and multi-objective metabolic modeling of microbial communities. *ACS Synth Biol* 2014;3:247–57. <http://dx.doi.org/10.1021/sb4001307>.
- Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 2003;84:647–57. <http://dx.doi.org/10.1002/bit.10803>.
- Pharkya P, Burgard AP, Maranas CD. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res* 2004;14:2367–76. <http://dx.doi.org/10.1101/gr.2872004>.
- Kim J, Reed JL. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol* 2010;4:53. <http://dx.doi.org/10.1186/1752-0509-4-53>.
- Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, et al. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst Biol* 2010;4:45. <http://dx.doi.org/10.1186/1752-0509-4-45>.
- Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 2010;6:e1000744. <http://dx.doi.org/10.1371/journal.pcbi.1000744>.
- Yang L, Cluett WR, Mahadevan R. EMILIO: a fast algorithm for genome-scale strain design. *Metab Eng* 2011;13:272–81. <http://dx.doi.org/10.1016/jymben.2011.03.002>.
- Cotten C, Reed JL. Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. *Biotechnol J* 2013;8:595–604. <http://dx.doi.org/10.1002/biot.201200316>.
- Hádıcke O, Klamt S. CASOP: a computational approach for strain optimization aiming at high productivity. *J Biotechnol* 2010;147:88–101. <http://dx.doi.org/10.1016/j.jbiotec.2010.03.006>.
- McAnulty MJ, Yen JJ, Freedman BG, Senger RS. Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. *BMC Syst Biol* 2012;6:42. <http://dx.doi.org/10.1186/1752-0509-6-42>.
- Chowdhury A, Zomorodi AR, Maranas CD. k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput Biol* 2014;10:e1003487. <http://dx.doi.org/10.1371/journal.pcbi.1003487>.
- Zhuang K, Yang L, Cluett WR, Mahadevan R. Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. *DySSCo strategy for strain design. BMC Biotechnol* 2013;13:8.

- [43] Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison 3rd CA, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 2009;6:343–5. <http://dx.doi.org/10.1038/nmeth.1318>.
- [44] Sleight SC, Sauro HM. BioBrick™ assembly using the in-fusion PCR cloning kit. *Methods Mol Biol* 2013;1073:19–30. http://dx.doi.org/10.1007/978-1-62703-625-2_3.
- [45] Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* 2014;344:55–8. <http://dx.doi.org/10.1126/science.1249252>.
- [46] Kelly JR, Rubin AJ, Davis JH, Ajo-Franklin CM, Cumbers J, et al. Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 2009;3:4. <http://dx.doi.org/10.1186/1754-1611-3-4>.
- [47] Stormo GD, Schneider TD, Gold L, Ehrenfeucht A. Use of the "Perceptron" algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res* 1982;10:2997–3011.
- [48] Rhodius VA, Mutalik VK. Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigmaE. *Proc Natl Acad Sci U S A* 2010;107:2854–9. <http://dx.doi.org/10.1073/pnas.0915066107>.
- [49] Rhodius VA, Mutalik VK, Gross CA. Predicting the strength of UP-elements and full-length *E. coli* oE promoters. *Nucleic Acids Res* 2012;40:2907–24. <http://dx.doi.org/10.1093/nar/gkr1190>.
- [50] Salis HM. Chapter two — the ribosome binding site calculator. In: Voigt Christopher, editor. *Methods in Enzymology. Synthetic Biology, Part B Computer Aided Design and DNA Assembly*. Academic Press; 2011. p. 19–42 [Available: <http://www.sciencedirect.com/science/article/pii/B9780123851208000024>. Accessed 29 May 2014].
- [51] Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 2009;27:946–50. <http://dx.doi.org/10.1038/nbt.1568>.
- [52] Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* 2013;10:354–60. <http://dx.doi.org/10.1038/nmeth.2404>.
- [53] Bogorad IW, Lin T-S, Liao JC. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* 2013;502:693–7. <http://dx.doi.org/10.1038/nature12575>.
- [54] Temme K, Zhao D, Voigt CA. Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proc Natl Acad Sci U S A* 2012;109:7085–90. <http://dx.doi.org/10.1073/pnas.1120788109>.
- [55] Chan LY, Kosuri S, Endy D. Refactoring bacteriophage T7. *Mol Syst Biol* 2005;1. <http://dx.doi.org/10.1038/msb4100025> [2005.0018].
- [56] Farasat I, Kushwaha M, Collens J, Easterbrook M, Guido M, et al. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol Syst Biol* 2014;10:731.