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

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

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 overproduction 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]
EMILiO	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]