Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Next-generation genome-scale models for metabolic engineering

Zachary A King[1], Colton J Lloyd[1], Adam M Feist[1,2] and Bernhard O Palsson[1,2,3]

Constraint-based reconstruction and analysis (COBRA) methods have become widely used tools for metabolic engineering in both academic and industrial laboratories. By employing a genome-scale *in silico* representation of the metabolic network of a host organism, COBRA methods can be used to predict optimal genetic modifications that improve the rate and yield of chemical production. A new generation of COBRA models and methods is now being developed — encompassing many biological processes and simulation strategies — and next-generation models enable new types of predictions. Here, three key examples of applying COBRA methods to strain optimization are presented and discussed. Then, an outlook is provided on the next generation of COBRA models and the new types of predictions they will enable for systems metabolic engineering.

Addresses
[1] Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[2] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark
[3] Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Corresponding author: Palsson, Bernhard O (palsson@ucsd.edu)

Current Opinion in Biotechnology 2015, **35**:23–29

This review comes from a themed issue on **Chemical biotechnology**

Edited by **Uwe T Bornscheuer** and **Alex Toftgaard Nielsen**

http://dx.doi.org/10.1016/j.copbio.2014.12.016

Introduction

The demand for raw material inputs to agriculture, industry, and energy are growing steadily, and concerns about environmental sustainability are becoming more acute; thus, alternatives to traditional, fossil-fuel based chemical production are becoming economically viable [1]. Cell factories, which use microorganisms to produce materials from renewable biomass, are an attractive alternative, and an increasing number of platform chemicals are being produced at industrial scale using engineered microorganisms [2]. To meet the demand for robust, high-yield production strains, an initial metabolic engineering strategy was developed, which used random mutagenesis and screening to identify strains with improved production performance (i.e., titer, productivity, and yield). However, the permutations possible in genomic sequences are so numerous that a mutagenesis and screening approach can only explore a small subset of possible strains, so the highest performance strains might never be identified. On the other hand, if *targeted* strain improvements can be predicted, then strains can be engineered which would not be found using untargeted mutagenesis and screening, and this can be accomplished with the tools of systems biology.

The goal of systems biology is to encode detailed information about an organism into a computational framework, with the goal of predicting cellular behavior from cellular genotype [3]. Thus, systems biology tools can provide targeted predictions of production strain modifications that improve their performance. The first major advancement toward this goal came with the development of the first genome-scale stoichiometric models of cellular metabolism. From these models emerged constraint-based reconstruction and analysis (COBRA) methods, with flux balance analysis (FBA) as a staple computational tool [4]. These advancements in turn led to the first studies in the field of systems metabolic engineering, where production strains are systematically built and improved through a combination of systems biology, synthetic biology, and evolutionary engineering [5–8].

An important first step in designing a production strain for a non-native metabolite is to identify and build a synthetic pathway [9]. COBRA methods have been employed successfully for pathway prediction and optimization [9,10]. After a pathway has been designed, strain optimization is performed to increase the yield and productivity of the strain. This commentary presents the most successful types of predictions made by COBRA tools for that process — optimizing microbial production strains — and then introduces a new generation of COBRA methods and prediction types which will advance systems metabolic engineering in the coming years.

Types of COBRA predictions used in strain optimization

A great number of COBRA methods have been developed [4], and the methods that have led to experimental

improvements in production strains can be categorized according to the types of predictions made. Three types of predictions have generally been employed: firstly, predicting maximum theoretical yield for native and non-native pathways, secondly, predicting gene deletions and their effects on a cell's ability to generate biomass and the target product, and thirdly, predicting the up and down expression of native genes necessary to improve product yield. Recent examples of strain engineering in each category are highlighted here.

Theoretical yield

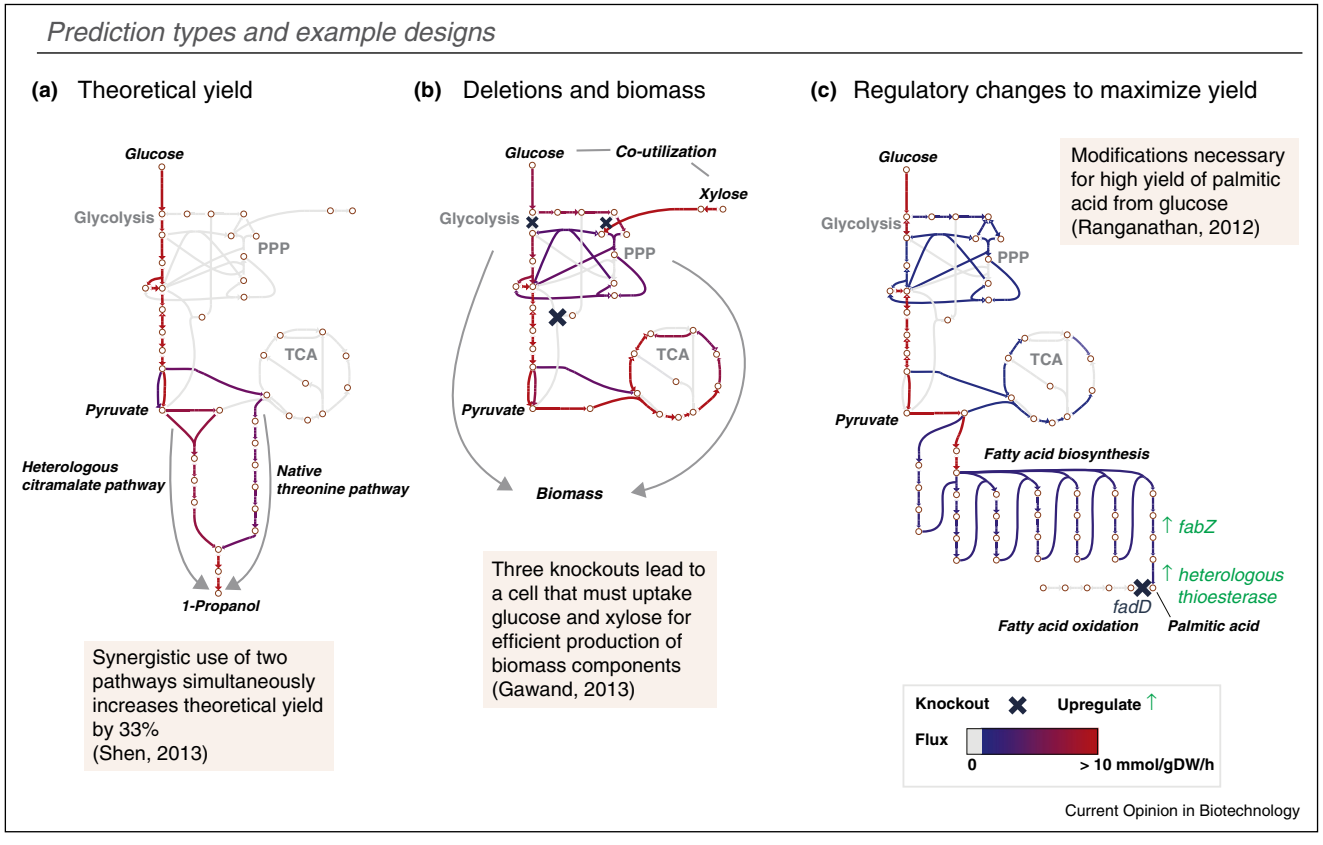
The most common prediction of COBRA methods in systems metabolic engineering has been the calculation of maximum *theoretical yield*, the percentage of substrate carbon that can be converted to a target molecule, given the limitations of carbon and redox balance in the stoichiometric network (Figure 1a). Yield is a critical consideration when determining the economic viability of chemical production, so these analyses have direct consequences for cell factory design.

Shen and Liao [11•] recently demonstrated the importance of theoretical yield calculations for optimizing production strains. In order to design a strain of *Escherichia coli* that produces 1-propanol, the authors used a simple mass-balanced and redox-balanced stoichiometric model and FBA to compare the theoretical yields of three routes to 1-propanol: the native threonine pathway, the non-native citramalate pathway, and a synergistic employment of both pathways (Figure 1a). The calculations revealed that the two pathways together have a theoretical yield of 1.33 mol 1-propanol per mol glucose, 33% higher than either individual pathway. Indeed, the authors constructed production strains for all three 1-propanol routes, and the synergistic employment of both pathways had the highest observed yield, ~30% greater yield than the citramalate pathway and ~55% greater yield than the threonine pathway.

Gene deletions and biomass requirements

Another class of COBRA predictions uses the biomass objective function — a representation of all the metabolite

Figure 1



Three types of COBRA predictions have been successfully implemented for systems metabolic engineering. **(a)** Theoretical yield. As an example, theoretical yield was maximized through synergistic use of two production pathways [11]. **(b)** Gene deletions and biomass. As an example, the SIMUP algorithm identifies three gene knockouts that force co-utilization of glucose and xylose to achieve maximum growth [19•]. **(c)** Regulatory changes to increase yield. As an example, gene up-regulations and deletions were used to increase the production of fatty acids [23•]. Figures were generated using Escher [49].