Metabolic Engineering Communications 5 (2017) 54–59

Contents lists available at ScienceDirect

Metabolic Engineering Communications

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

UP Finder: A COBRA toolbox extension for identifying gene overexpression strategies for targeted overproduction

Xi Wang, Liang Yu, Shulin Chen*

Department of Biological Systems Engineering, Washington State University, Pullman, WA 99164, USA

ARTICLE INFO

Keywords:
COBRA toolbox
Gene overexpression strategy
Metabolic engineering
Targeted overproduction
Rational pathway design

ABSTRACT

Overexpression of key genes is a basic strategy for overproducing target products via metabolic engineering. Traditionally, identifying those key genes/pathways largely relies on the knowledge of biochemistry and bioinformatics. In this study, a modeling tool named UP Finder was developed to facilitate the rapid identification of gene overexpression strategies. It was based on the COBRA toolbox under MATLAB environment. All the key gene/pathway targets are identified in one click after simply loading a Systems Biology Markup Language model and specifying a metabolite as the targeted product. The outputs are also quantitatively ranked to show the preference for determining overexpression strategies in pathway design. Analysis examples for overproducing lycopene precursor in *Escherichia coli* and fatty acyl-ACP in the cyanobacterium *Synechocystis* sp. PCC 6803 by the UP Finder showed high degree of agreement with the reported key genes in the literatures.

1. Introduction

Engineering microorganisms to overproduce interested products is an important practice in metabolic engineering. In the successful examples, overexpressing key genes of metabolic pathways is a widely used strategy for achieving overproduction (Ajikumar et al., 2010; Alper et al., 2005). The purpose is to up-regulate the flux for substrate synthesis or to intensify the shunt at key metabolic nodes toward an improved flux to targeted metabolites. Since the overproduction of natively synthesized metabolites is usually achieved by genetically manipulating metabolic pathways, identifying the key pathways and gene targets is a key step to determine gene overexpression strategies for consequential manipulations. Traditionally, completion of such tasks was largely relying on the experience of metabolic pathways and enzymatic kinetics. However, with the increasing practices of metabolic engineering in overproducing fuels, chemicals and natural products (Stephanopoulos, 2012), empirical predictions have been hardly satisfying the analysis of sophisticated pathways, such as the multiple-repeated pathways in fatty acid synthesis and the rarely explored secondary metabolite biosynthesis. Therefore, it is critical to establish a standard procedure for identifying gene overexpression strategies.

The rapid advances of constraint-based models provide the possibilities for quantitative evaluation of cellular metabolism (Bordbar et al., 2015; Kauffman et al., 2003), allowing to develop the standard method for rational pathway design. According to the annotated genome information, the reconstructed constraint-based models could represent the current knowledge of full metabolic reactions and their associated genes for an organism. With those constraint-based models, algorithms such as flux balance analysis (FBA) were developed to perform the *in silico* analysis of metabolic fluxes (Orth et al., 2010). Relying on the principle of mathematical optimization and mass balance, metabolic fluxes can be simulated within determined constraints. Such efforts have advanced the development of modeling approaches such as OptKnock (Burgard et al., 2003) that facilitates the procedures for identifying gene targets and pathway design.

Unlike gene knockout based simulation, *in silico* identifying gene overexpression targets has more uncertainties to be experimentally verified because of the difficulties for exactly manipulating fluxes to certain values. To overcome this challenge, methodologies have been developed for simulating gene overexpression, such as OptForce (Ranganathan et al., 2010) and FSEOF (Choi et al., 2010), as well as their derivatives (Chowdhury et al., 2014; Park et al., 2012). By using enforced flux and flux variability analysis, gene targets with desired up-regulation were successfully simulated and experimentally verified. However, those overexpressed gene targets were mostly identified to coordinate with additional manipulations (e.g. knockouts or down-regulation), whereby overexpressing some targets such as targets in glycolysis may not always independently contribute to an overproduction. Therefore, it is important to know the contribution of each candidate targets toward the theoretical maximum yield to fulfill the growing needs on customized pathway design. In addition, most current modeling methods still require specific programming skills that restricts

* Corresponding author.
E-mail address: chens@wsu.edu (S. Chen).

http://dx.doi.org/10.1016/j.meteno.2017.08.001
Received 7 March 2016; Received in revised form 20 April 2017; Accepted 14 August 2017
Available online 16 August 2017

the access for biologists and broad users. It is highly desirable to develop the software platform that can bridge the technical gap between computational modeling and bench works. In this paper, we present a software package, UP Finder that facilitates the identification of gene overexpression strategies for the metabolic engineering of targeted overproduction. It highlighted the quantitative evaluation for each overexpression candidate on yield contribution. The graphical user interface of the UP Finder also provided easier access for broad users. Two typical examples in metabolic engineering that lycopene precursor and fatty acyl-ACP overproduction were used to evaluate feasibilities of the UP Finder for analyzing biosynthesis pathways of natural products and biofuels. The identified gene targets by the UP Finder showed high degree of agreement with the reported key genes in the literatures.

2. Materials and methods

2.1. Models and FBA

The metabolic reconstructed model of *Escherichia coli* iJO1366 (Orth et al., 2011) was used for analyzing gene overexpression strategies in lycopene precursor overproduction. And the reconstructed model of *Synechocystis* sp. PCC 6803 iJN678 (Nogales et al., 2012) was used for the analysis of fatty acyl-ACP overproduction.

FBA was used for all model analysis. For wild-type model, the defaulted biomass formulation was used as the objective function for maximizing cell growth. For theoretical maximum yield model, the targeted product was used as the objective function for maximizing the production of targeted product, such as farnesyl pyrophosphate and fatty acyl-ACP discussed in Results.

All computation was performed on Mac OS × 10.6.8, 1.86 GHz Inter Core 2 Duo Processor, 2 GB 1067 MHz DDR3 Memory. COBRA toolbox v2.0.5 was added to the path of MATLAB_R2012b, including SBML Toolbox_4.1.0 bundled in the package. libSBML_5.7.0 was installed to access the Systems Biology Makeup Language. Gurobi_5.1.0 was used as the LP solver.

2.2. Definition of parameters

The parameter $flux_{wt}$ represents wild-type flux that is the flux solution of the wild-type model, and $flux_{opt}$ represents the optimum flux that is the flux solution of the theoretical maximum yield model. The up-regulation ratio (*Ratio*) is defined as the ratio of $flux_{opt}$ to $flux_{wt}$ of a reaction ($Ratio = flux_{opt} / flux_{wt}$). And the *Yield* is simulated product yield of the targeted product by using $flux_{opt}$ of a reaction as the constraint, in which maximizing cell growth is the objective function.

2.3. Development of the UP Finder

UP Finder is an interfacial modeling tool based on the COBRA toolbox in MATLAB, which is developed by the MATLAB Graphical User Interface Development Environment (GUIDE). It is used to identify all the key gene targets for overexpression that directly related to the overproduction of a metabolite in a microorganism. The working procedure of the UP Finder is composed of following steps (Fig. 1):

(1) Identification of up-regulated fluxes. The main concept is to compare the flux distributions between the wild-type and overproducing metabolic networks by calculating theoretical maximum yield of a targeted product. Thus, up-regulated fluxes and their associated pathways (termed as up-regulated pathways in this study) can be identified through this comparison.
(2) Re-verification of identified pathways. Since not all the identified pathways from Step (1) are directly related to the overproduction, a re-verification is necessary to filter the low-relevant targets. For these identified pathways, their fluxes under overproducing networks were considered as the optimum fluxes ($flux_{opt}$) to achieve theoretical maximum yield of the product. The simulated product yields (*Yield*) constrained by each $flux_{opt}$ for the wild-type network were used to evaluate the best contribution of each up-regulated pathway toward overproduction. Pathways with $Yield > 0$ are considered as the key targets that directly lead to the overproduction.
(3) Rank of the output. The output of the UP Finder is the abbreviated reaction names of the selected key pathways in Step (2). A termed parameter, *Ratio,* which is the ratio of each $flux_{opt}$ over their associated wild-type fluxes ($flux_{wt}$) was used for ranking the output from high to low. Because *Ratio* reflects the up-regulated level for each reaction, the one with the highest *Ratio* value suggests the highest preference when considering gene overexpression in engineering of the targeted overproduction.

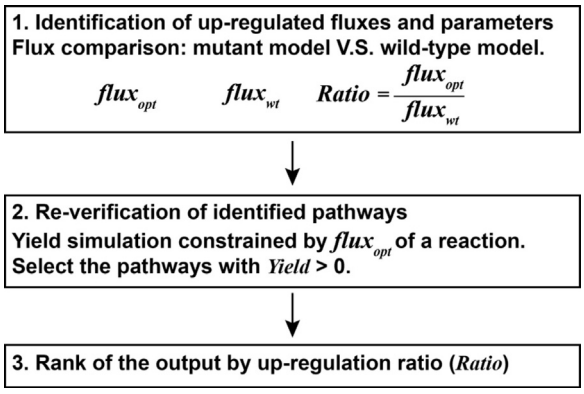

Fig. 1. The working procedure of the UP Finder.

Mutant model, the overproducing metabolic network, the model with flux distribution under the theoretical maximum yield conditions (flux distribution for reaching theoretical maximum yield of a metabolite); $flux_{wt}$, wild-type flux, which is the flux distribution of the wild-type conditions; $flux_{opt}$, optimum flux, which is the flux distribution of the theoretical maximum yield conditions; *Ratio*, up-regulation ratio, which is the ratio of the optimum flux to the wild-type flux of a reaction; *Yield*, simulated yield of the targeted product by using the optimum flux of a reaction as the constraint.

2.4. Implementation

Through the interface of the UP Finder, after loading a Systems Biology Markup Language (SBML) model in the *Organism* item, all metabolites included in this model will be shown in the *Targeted product* item. Users can simply specify one metabolite as the target for overproduction. By choosing UPA in the *Method* option and running the program, a list of reaction names presented with their associated genes, reaction formulas, *Yield* and *Ratio* values will be returned as the output. All computation is based on the COBRA toolbox and MATLAB, and all optimization uses FBA for the solutions (Orth et al., 2010; Schellenberger et al., 2011). Initializing the COBRA toolbox is necessary in MATLAB before loading SBML models. The default uptake and growth constraints of the reconstructed model are used for the analysis. Users can also adjust the uptake and growth conditions to simulate metabolisms with special requirements. In addition, the UP Finder integrates FBA optimization in the *Method* option, which allows the basic function for computing growth rates under different conditions (Fig. 2). The UP Finder is freely available from GitHub (https://github.com/MEpathway/UP-Finder.git).

The interface contains 5 major functional units, including the *Organism*, *Condition/Growth*, *Targeted product*, *Method* and *Output*. *Condition* is the exchange reactions of SBML models with their constraints, and *Growth* indicates the specific biomass objective function