BioSystems 117 (2014) 10–14

Contents lists available at ScienceDirect

BioSystems

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

Genome-scale reconstruction of a metabolic network for *Gluconobacter oxydans* 621H

Xinsen Wu [1], Xiaoyang Wang [1], Wenyu Lu [*]

Department of Biological Engineering and Key Laboratory of Systems Bioengineering of the Ministry of Education, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 21 May 2012
Received in revised form
10 November 2013
Accepted 2 January 2014

Keywords:
Gluconobacter oxydans
Genome-scale metabolic network
Flux balance analysis
Flux variability analysis
Gene essentiality analysis

ABSTRACT

Gluconobacter oxydans is a Gram-negative bacterium with a number of biotechnological applications. Although the genome of *G. oxydans* has been reported in 2005, the systematical cellular metabolism in this high-value bacterium, however, remains unclear. In this study, a genome-scale metabolic network of *G. oxydans* 621H, iXW433, was reconstructed and validated on the basis of the known genome annotations and biochemical information. This reconstructed model included 433 genes, 859 reactions, and 985 metabolites. To test the capability of the model, gene and reaction essentiality analysis, flux variability analysis, and robustness analysis simulations were performed. The metabolic states predicted by the model were highly consistent with the experimental data of *G. oxydans*. According to the result, 92 genes and 137 reactions were identified to be essential, 194 reactions were found to be variable by flux variability analysis, and 2 possible genetically modified targets were determined. The model would be valuable for further research on *G. oxydans* and thereby expanding its application.

1. Introduction

Gluconobacter oxydans is a Gram-negative bacterium belonging to the family of acetic acid bacteria (Acetobacteraceae), which are characterized by their ability to rapidly and incompletely oxidize a wide range of carbohydrates and alcohols (Deppenmeier et al., 2002). Because of the ability of incompletely oxidation, *G. oxydans* is widely used in biotechnological industry such as the production of L-sorbose from D-sorbitol (vitamin C synthesis) (Hu et al., 2011), 1,3-dihydroxyacetone (DHA) from glycerol (Bremus et al., 2006), 5-keto-D-gluconic acid from 6-(2-hydroxyethyl)amino-6-deoxy-α-L-sorbofuranose (Merfort et al., 2006), and 6-amino-L-sorbose from 1-amino-D-sorbitol (De Muynck et al., 2007). The most two interesting industrial applications are as follows: 5-keto-D-gluconic acid has been successfully used in the production of L-(+)-tartaric acid, an antioxidant in food industry (Lichtenthaler, 2009); Miglitol, a new α-glucosidase inhibitor, has been successfully synthesized in industry with D-glucose as the raw material and 6-(2-hydroxyethyl)amino-6-deoxy-α-L-sorbofuranose as the intermediate (Keliang and Dongzhi, 2006). This organism contains many membrane-bound dehydrogenases that are crucial to its ability to incompletely oxidize biotechnologically important substrates (Katrlík et al., 2007). It has been shown that membrane-bound dehydrogenases channel electrons into the respiratory chain, while the energy-transducing efficiency of the respiratory chain in *G. oxydans* is very low because of the absence of a proton-translocating NADH: ubiquinone oxidoreductase (complex I) and a cytochrome c oxidase (complex IV) (Prust et al., 2005). The poor coupling between respiratory chain and ATP production reduce the inhibition of the membrane-bound redox reactions, which results in a rapid oxidation of the substrates. The unique metabolism makes it an ideal organism for industrial application.

Due to the industrial value of *G. oxydans*, there were a lot of research about this strain, and the genome of *G. oxydans* has been reported in 2005 (Prust et al., 2005). Since the publication of the genome of *G. oxydans*, the biological knowledge about this strain has significantly increased and many valuable achievements have resulted from this knowledge (Schleyer et al., 2008). On the other hand, systematic analyses of its metabolic and biotechnological capacities yet to be performed. Thus it is very meaningful to develop a systematic model for *G. oxydans*.

Genome-scale metabolic network reconstruction has become an important tool for studying the systems biology of metabolism (Feist and Palsson, 2008). The genome-scale metabolic models are becoming available for some organisms (Feist et al., 2009) since the model of *Haemophilus influenza* Rd was reconstructed (Edwards and Palsson, 1999). Up to now, more than 50 genome-scale metabolic reconstructions have been published, such as *E. coli* (Feist et al., 2007), *Saccharomyces cerevisiae* (Forster et al., 2003), maize (Saha et al., 2011), and human (Duarte et al., 2007). These metabolic networks have been used toward contextualization of

* Corresponding author. Tel.: +86 22 27892132; fax: +86 22 27400973.
E-mail addresses: wenyulu@tju.edu.cn, luwenyu1973@hotmail.com (W. Lu).
[1] These authors contributed equally to this work.

0303-2647/$ – see front matter © 2014 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biosystems.2014.01.001

high-throughput data, guidance of metabolic engineering, directing hypothesis-driven discovery, interrogation of multi-species relationships, and network property discovery (Oberhardt et al., 2009). Genome-scale metabolic networks organize numerous available data of different species, enabling easier using of these data. With appropriate algorithm, genome-scale metabolic networks can be used to simulate metabolic state of an organism in a given environmental condition. For example, flux balance analysis (FBA) can simulate the maximal growth of a cell, and find the key factors on growth, so the simulation results can be a useful guide for metabolic engineering (Orth et al., 2010). Some useful algorithms have been development in recent years, such as steady-state flux balance analysis (SR-FBA) (Shlomi et al., 2007) and regulatory on/off minimization (ROOM) (Shlomi et al., 2005), making genome-scale metabolic networks more useful. By studying the network topology structure of metabolic networks of 80 organisms, Ma and Zeng (2003) obtained the structure character of these metabolic networks, such as path length and connection degree distribution. Besides, 20 hub metabolites, which were defined as the metabolites holding the biggest connection degree, were also found in their metabolic networks. These hub metabolites participated in numerous reactions and have great influence on the flux distribution.

In this paper, a genome-scale network reconstruction of *G. oxydans* was built based on genomic, biochemical and physiological information. Being the first comprehensive metabolic model of *G. oxydans*, it was named iXW433 following the naming convention proposed by the published literature (Reed et al., 2003). FBA was performed to test the predict potential of the model with the focus on the capability of dihydroxyacetone synthesis with glycerol as the carbon source, and the result was validated against literature data.

2. Methods

The complete genome sequence of *G. oxydans* 621H has been reported in 2005 by Prust et al. (2005), and genome annotation is available online at numerous databases. *G. oxydans* 621H genome size is 2.922 Mb with 2664 candidate protein-encoding open reading frames (ORFs). Genome annotation is the basic material for reconstruction of metabolic network. The *in silico* simulation of the metabolic network was carried out using MATLAB7.10 (The Mathworks, Inc.) and COBRA toolbox (Becker et al., 2007).

2.1. Metabolic reconstruction process

A three-stage process was designed to reconstruct the genome-scale metabolic network of *G. oxydans* 621H based on a published protocol (Thiele and Palsson, 2010). A draft reconstruction was firstly created. All the necessary information, containing genome annotation and the biochemistry information of the enzymes, can be searched in KEGG (Kanehisa et al., 2006). 992 reactions were included in the model as the candidate metabolic reactions, which were collected by identifying enzymes based on the information in biochemical database, such as KEGG (Kanehisa et al., 2006) and BRENDA (Scheer et al., 2011). The following factors were included in the draft: (1) the basic reaction information (ID, name, equation) and (2) the enzymes and genes relevant to particular reaction. However, many reactions were redundant in the draft while some necessary ones were absent. A good example was the ignoring of some reactions which synthesize or modify the macromolecules (protein, nucleic acid, *etc.*). Besides, the overall and sub-step reactions of the same multi-step reaction might exist at the same time, which might result in futile cycles.

So a lot of work was done to refine the draft model in the second stage. Actually, many of the above mentioned mistakes could be corrected through manual refinement. The refinement mainly focused on these parts. (1) The reactions which synthesize or modify the macromolecules were replaced by the biomass formulation reactions. (2) Only one form of the multi-step reaction was remained in the model. For example, the overall reaction R00742 was kept while its sub-step reactions R04385 and R04386 were removed. (3) Redundant compound numbers were removed to ensure a metabolite had only one compound number. This could be achieved by removing the generic reactions in the metabolic network. Finally, cofactors and reactions were determined and verified. Based on the refined metabolic network, metabolic map was drawn and topological analysis was carried out using the method proposed by Ma and Zeng (2003).

In the last stage, biomass formation reactions, transport reactions and exchange reactions were added to the metabolic network to get a complete model. Biomass formation (*i.e.*, protein, RNA, DNA, lipids, peptide glycan and carbohydrate) was predominantly based on the chemicals of the cell. Transport reactions were added based on the following rules: The ones containing the metabolites which were verified to be taken up from the medium were added; the ones containing the metabolites which can diffuse through the membranes must be added. Exchange reactions, which represented the system boundaries, were added for all extracellular metabolites (Thiele and Palsson, 2010).

The reconstructed metabolic network was automatically converted into a mathematical model that could be analyzed by MATLAB and COBRA Toolbox (Becker et al., 2007). Some functions in COBRA Toolbox were used to debug the metabolic network. And based on the result of debugging, gaps, futile cycles and loops in the metabolic network could be found.

2.2. Biomass formulation

The elementary cell composition of *G. oxydans* was found to be $C_5H_{8.9}NO_{1.9}$ (Olijve and Kok, 1979). The lipids composition of phospholipids and fatty acids has been reported (Heefner and Claus, 1976, 1978), but it remains unknown for other components composition. Hence, the determination of biomass composition was based on the composition of *E. coli* (Stephanopoulos et al., 1998). Besides, the amino acid and deoxynucleotide composition were estimated using genome information from Comprehensive Microbial Resource (CMR) database. RNA was actually composed of mRNA, rRNA and tRNA, and the percentages of each kind were supposed to be 5%, 75% and 20%, respectively (Borodina et al., 2005). The nucleotide composition of mRNA, rRNA and tRNA were estimated using genome information from CMR database.

2.3. Constraints-based flux analysis

Constraints-based flux analysis is a general way for optimization-based simulation techniques, and various algorithms are available for it (Kim et al., 2008). Metabolic reactions are represented as a stoichiometric matrix $\boldsymbol{S}$ ($m \times n$), where m is the number of metabolites and n is the number of reactions. The flux through all of the reactions in a network is represented by the vector $\mathbf{v}$. FBA is based on an assumption of pseudo-steady state in which the net sum of all the production and consumption rates is set to be zero for each internal metabolite. Mass balance and lower/upper bounds of flux through each reaction act as the constraints, which are represented by the formulas $\boldsymbol{S}\cdot\boldsymbol{v}=\mathbf{0}$ and $\boldsymbol{v}_{\mathbf{min}} \leqq \boldsymbol{v} \leqq \boldsymbol{v}_{\mathbf{max}}$. The formula $\boldsymbol{S}\cdot\boldsymbol{v}=\mathbf{0}$ defines a system of linear equations, so FBA can maximize or minimize a selected objective function by solving the linear programming equations. FBA has been widely used in physiological studies, gap-filling efforts and genome-scale synthetic biology, and numerous algorithms have been developed based on it, such as robustness analysis (Edwards