Biochemical Engineering Journal 127 (2017) 229–241

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

Regular article

Analyses of extracellular protein production in *Bacillus subtilis* – I: Genome-scale metabolic model reconstruction based on updated gene-enzyme-reaction data

Pınar Kocabaş[a], Pınar Çalık[b], Güzide Çalık[a], Tunçer H. Özdamar[a,*]

[a] *Chemical Engineering Department, Biochemical Reaction Engineering Laboratory, Ankara University, Tandoğan, 06100 Ankara, Turkey*
[b] *Chemical Engineering Department, Industrial Biotechnology and Metabolic Engineering Laboratory, Middle East Technical University, 06800, Ankara, Turkey*

ARTICLE INFO

Article history:
Received 20 December 2016
Received in revised form 7 July 2017
Accepted 11 July 2017
Available online 23 July 2017

Keywords:
Bacillus subtilis
Gene-enzyme-reaction data
Genome scale model
Flux balance analysis
Serine alkaline protease
Metabolic engineering sites

ABSTRACT

Bacillus subtilis genome-scale model (GEM) reconstruction was stimulated by the recent sequencing and consequent re-annotations. The updated gene-enzyme-reaction data were collected from databases to reconstruct *B. subtilis* reaction network BsRN-2016 containing 1144 genes linked to 1955 reactions and 1103 metabolites. Thermodynamic analysis was conducted to identify reversibility and directionality of the reactions. By elimination of unconnected-reactions from BsRN-2016, reconstruction process of the first third-generation GEM *i*Bsu1144 employing 1083 reactions linked to 719 genes was completed. The stoichiometric flux-balance based model was solved using time-profiles of serine alkaline protease fermentation-data at three different oxygen-transfer conditions creating perturbations on the intracellular reaction-network. Testing *i*Bsu1144 dignity with three different objective functions indicated superior robustness of the GEM. Comparison of *i*Bsu1144 results with a second-generation GEM was demonstrated. Insights obtained from flux distributions were used to determine metabolic engineering sites. Asparagine, isoleucine, threonine, and aspartic acid were determined as the primary rate-limiting amino acids to be considered as corresponding metabolic engineering sites in SAP synthesis in *B. subtilis*. Flux variability analysis carried out for the optimum condition reveals that 288 reactions are active and linked to 317 genes, so called substantial genes. The blocked 735 reactions linked to 533 genes, formed a platform for guided gene deletions in *B. subtilis* to generate simplified host strains.

1. Introduction

Bacillus subtilis and its relatives are prodigious producers of industrial enzymes, such as proteases and amylases. Thus, *B. subtilis* has been the model of Firmiculates for decades, since the absence of an outer membrane combined with an efficient *Sec*-dependent secretion pathway [1] means that proteins can be secreted directly into the culture medium at high concentrations [2]. In recent years, considerable effort has been aimed at developing *B. subtilis* as host for the production of heterologous proteins. Meanwhile, the studies on: (i) promoters and suitable ribosome-binding sites [3–5], (ii) signal peptides [6,7], (iii) secretion pathway [5,8–10], (iv) proteolytic degradation of proteins [11], and (v) genome reductions for host generation [12,13] underpin the crucial aim to increase recombinant protein (r-protein) production in *B. subtilis* and stimulate more research directed to metabolic and bioreactor engineering. Moreover, being sequenced and updated totally, *B. subtilis* [14–16] has also been shown to be an excellent model organism for systems biological analyses [2].

Intracellular reaction fluxes can be calculated by solving flux-balance based stoichiometric models based on intracellular reaction networks together with elaborate fermentation data [17]. Therefore, based on reliable models calculation of reaction fluxes is an important tool in metabolic engineering, allowing detailed

Abbreviations: BsRN-2016, *Bacillus subtilis* reaction network generated in this work containing 1955 reactions and 1103 metabolites; *i*Bsu1144, third-generation reconstructed genome-scale model based on BsRN-2016 containing 1083 reactions and 743 metabolites; GEM, Genome scale model; HOT, high- oxygen transfer condition; *i*Bsu1103, first-generation GEM reconstructed by Henry et al. (2009); *i*Bsu1103V2, second-generation GEM reconstructed by Tanaka et al. (2009); *i*Bsu1103V2$_{SAP}$, *i*Bsu1103V2 including SAP synthesis reaction; KO, gene knockout; LOT, low- oxygen transfer condition; MOT, medium- oxygen transfer condition; PPP, pentose phosphate pathway; PSR, protein synthesis reaction; SAP, serine alkaline protease; TCA, tricarboxylic acid.

* Corresponding author.
E-mail address: thozdamar@ankara.edu.tr (T.H. Özdamar).

http://dx.doi.org/10.1016/j.bej.2017.07.005
1369-703X/

Nomenclature

A	Stoichiometric matrix of the metabolic network
c(t)	Metabolite accumulation vector
$\mathbf{c_1(t)}$	Extracellular metabolite accumulation vector
$\mathbf{c_2(t)}$	Intracellular metabolite accumulation vector
C_C	Citrate concentration ($kg\,m^{-3}$)
$G_{f_i}{}^o$	Standard molar Gibbs free energy of formation of compounds
$\Delta G_R\,o$	Standard Gibbs free energy change of a reaction
K	Thermodynamic equilibrium constant
N	Agitation rate (min^{-1})
Q_o	Volumetric air feed rate ($m^3\,min^{-1}$)
r(t)	Vector of reaction fluxes
R#	*i*Bsu1144 reaction number
*R#	BsRN-2016 reaction number
$R_{LF\text{-}i}$	Amino acid rate limiting function ($R_{LF\text{-}i} = \alpha_{ai\text{-}PSR}/\sum r_{ai}$, g_{DW} h/mmole)
t	Cultivation time (h)
V	Volume of the bioreaction medium (m^3)
Z	Objective function
T	Temperature (°C)
$\sum r_{ai}$	Total synthesis flux of amino acid- *i* (mmole/g_{DW}/h)

Greek letters

μ	Specific growth rate (h^{-1})
α_i	Coefficient of component-*i* in the stoichiometric equation of the corresponding reaction
$\alpha_{ai\text{-}PSR}$	Stoichiometric coefficient of amino acid-*i* in the protein synthesis reaction not detected extracellularly

quantification of *in vivo* fluxes which provide valuable information on cellular physiology that can be applied for engineering metabolic and regulatory pathways, dignity of which is dependent on the quantity and reliability of intracellular reactions.

Genome sequencing and subsequently annotation studies enable construction of genome scale reaction networks with high number of reactions interrelated with gene- and enzyme- data. *B. subtilis* genome was sequenced first by Kunst et al. [14] with 4100 protein-coding genes, and the annotated genome sequence stimulated the studies for *B. subtilis* network constructions [18–20], which were converted into *in silico* genome scale models (GEMs). Consequently, the first three GEMs reported by Oh et al. [18], Goelzer et al. [19], and Henry et al. [20] which can be regarded as the first-generation *B. subtilis* GEMs were based on the genome annotation of Kunst et al. [14]; thereafter, the second-generation GEMs of Tanaka et al. [21] and Hao et al. [22] were reconstructed after the genome annotation of Barbe et al. [15] (Table 1a). In order to evaluate the *B. subtilis* GEMs, type of data used in the construction [18] and reconstructions [19–22] and in validation processes of the first- and second- generation *B. subtilis* GEMs are summarized in Table 1b. Furthermore, the strategies applied for optimum flux distributions in model constructions and their validation are summarized in Table 1c. Also, the accuracy of the GEMs are summarized in Table 1d.

Genome-scale reaction networks need to involve all accurate reactions based on correct gene-enzyme-reaction annotations. Therefore, resequencing of a genome stimulates reconstruction of genome scale model(s). In this context, the third re-sequencing of *B. subtilis* genome in 2013 by Belda et al. [16] and the subsequent re-annotation studies renewed the gene-enzyme-reaction data and created an updated data platform for the field of biochemical reaction engineering to reconstruct a renewed and expanded genome scale *B. subtilis* reaction network. The flow of updated gene-enzyme-reaction data enabled the authors to reconstruct the updated third-generation *B. subtilis* reaction network, BsRN-2016. Through thermodynamic analyses and elimination of unconnected-reactions, BsRN-2016 was converted into the first third-generation GEM, so-called *i*Bsu1144. The dignity and robustness of *i*Bsu1144 was tested in predicting intracellular reaction fluxes in a protein fermentation process, *i.e.*, serine alkaline protease (SAP), using elaborate experimental data used in the first flux analysis work for a protein fermentation [29,30]. The model presents a step towards more complete description of protein production in *B. subtilis* which is required to understand and optimize fermentation processes. Insights obtained from the intracellular flux distributions at well-defined low-, medium-, and high-oxygen transfer conditions throughout the bioprocess during batch-fermentations were used to determine metabolic engineering sites for the protein synthesis and for host cell constructions for protein productions. The reconstructed third generation GEM *i*Bsu1144 was compared with a second- generation GEM [21], and in the light of information gathered from the GEM *i*Bsu1144, future perspectives were discussed.

2. Method

2.1. Mass flux balance-based analysis

Stoichiometric analysis of metabolic flux distributions [31] with flux balance (FB) methodology aims calculation of flow of metabolites through the mathematically represented intracellular reaction network without kinetic parameters. As it does not use kinetic parameters, it cannot predict metabolite concentrations, and it is only suitable for the calculation of fluxes at a mathematically well defined state. When fermentation data are used as constraints, an allowable solution space is obtained. Through optimization of a defined objective function, FBA can identify a single optimal flux distribution on the edge of the allowable solution space [32,33].

Mass-balance based stoichiometric equations are constructed for each metabolite, assuming the cell as a semi-batch microbioreactor, where algebraic sum of all the conversion reactions of each metabolite-*i* in the defined reactions plus the transport of metabolite-*i* are balanced with the accumulation of metabolite-*i*, as reported elsewhere [17,29]. The scalar flux balance equations can be represented in the form of a linear vector differential equation, as follows:

$$\mathbf{A} \times \mathbf{r(t)} = \mathbf{c(t)} \tag{1}$$

where, **A** is the $m \times n$ stoichiometric coefficients matrix of the metabolic reaction network, m is the number of the metabolites and n is the number of reactions [17,29], **r(t)** is the vector of fluxes and **c(t)** is the metabolite accumulation vector. The elements of **c(t)** are divided into two sub-vectors:

$$\mathbf{c(t)} = \mathbf{c_1}(t) + \mathbf{c_2}(t) \tag{2}$$

where, $\mathbf{c_1(t)}$ and $\mathbf{c_2(t)}$ correspond to extracellular and intracellular metabolite accumulation vectors, respectively. Using pseudo-steady state (PSS) approximation for the intracellular metabolites, the intracellular metabolite flux vector $\mathbf{c_2(t)}$ is set to zero. Intracellular fluxes can be calculated by minimizing or maximizing the objective function Z (Eq. (3)), specified for a selected component-*i* [29]:

$$Z = \sum \alpha_i r_i \tag{3}$$

where, Z is a linear combination of fluxes (r_i) multiplied by corresponding stoichiometric coefficient of component-*i* (α_i) for every reaction wherein component-*i* is consumed or synthesised.