Journal of Biotechnology 161 (2012) 42–48

Contents lists available at SciVerse ScienceDirect

Journal of Biotechnology

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

Reconstruction and analysis of a genome-scale metabolic model of the vitamin C producing industrial strain *Ketogulonicigenium vulgare* WSH-001

Wei Zou[a,c], Liming Liu[a,b,c,*], Jing Zhang[d], Haoru Yang[a,c], Maoda Zhou[a,c], Qiang Hua[b], Jian Chen[a,c]

[a] State Key Laboratory of Food Science and Technology, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu 214122, China
[b] State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China
[c] The Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, 1800 Lihu Road, Wuxi, Jiangsu 214122, China
[d] College of Bioengineering, Sichuan University of Science and Engineering, 180 Huixing Road, Zigong, Sichuan 643000, China

ARTICLE INFO

Article history:
Received 27 December 2011
Received in revised form 12 May 2012
Accepted 18 May 2012
Available online 19 June 2012

Keywords:
Genome-scale metabolic model
Ketogulonicigenium vulgare
2-keto-L-Gulonic acid
Sulfate metabolism
Vitamin C

ABSTRACT

Ketogulonicigenium vulgare WSH001 is an industrial strain commonly used in the vitamin C producing industry. In order to acquire a comprehensive understanding of its physiological characteristics, a genome-scale metabolic model of *K. vulgare* WSH001, *i*WZ663, including 830 reactions, 649 metabolites, and 663 genes, was reconstructed by genome annotation and literature mining. This model was capable of predicting quantitatively the growth of *K. vulgare* under L-sorbose fermentation conditions and the results agreed well with experimental data. Furthermore, phenotypic features, such as the defect in sulfate metabolism hampering the syntheses of L-cysteine, L-methionine, coenzyme A (CoA), and glutathione, were investigated and provided an explanation for the poor growth of *K. vulgare* in monoculture. The model presented here provides a validated platform that can be used to understand and manipulate the phenotype of *K. vulgare* to further improve 2-KLG production efficiency.

1. Background

As the safest and most effective nutrient, L-ascorbic acid or vitamin C is widely used in the pharmaceutical, food, beverage, and feedstock industries (Bremus et al., 2006; Hancock and Viola, 2002), and the demand is increasing. In general, vitamin C is produced from a glucose source by two main processes. The first synthetic route is the Reichstein process, which was developed in the 1930s (Reichstein and Grüssner, 1934). The second is the modern two-step fermentation process, originally developed in China in the 1960s (Yin et al., 1980). In the two-step fermentation process, *Ketogulonicigenium vulgare* (previously regarded as *Gluconobacter oxydans*) (Urbance et al., 2001) and *Bacillus megaterium* are used to synthesize the L-ascorbic acid precursor 2-keto-L-gulonic acid (2-KLG) from L-sorbose (Yin et al., 1990). A high 2-KLG concentration (75.8 g/L) and yield (94.8%) can be achieved by providing L-sorbose to the co-culture of *K. vulgare* and *B. megaterium* during 72 h of cultivation, with the 2-KLG yield significantly enhanced by the use of two bacterial strains, rather than just *K. vulgare* alone (Xu et al., 2004). The reasons why the presence of *B. megaterium* increases the efficiency of 2-KLG synthesis *via K. vulgare*, whereas sole culture of *K. vulgare* yields little 2-KLG production are unknown. Furthermore, *K. vulgare* and *B. megaterium* have different nutritional requirements and different optimal growth conditions. In addition to the uncertain optimal ratio of *K. vulgare* and *B. megaterium*, the fermentation process is characterized by low stability and complex fermentation conditions that are difficult to control (Zhang et al., 2010b), which means there has been difficulty in increasing the yield and productivity of 2-KLG. Therefore, a key issue for the two-step fermentation process is a comprehensive understanding of the interaction mechanism between *K. vulgare* and *B. megaterium*.

In order to address this issue, much effort has been devoted to investigating the physiological relationship between *K. vulgare* and *B. megaterium*. Biochemical and molecular methods have demonstrated that endogenous and exogenous metabolites, such as some proteins or amino acids from *B. megaterium*, promote *K. vulgare* growth and 2-KLG production. Based on these results, in our previous study we used lysozyme to lyse *B. megaterium* to release the intracellular components and, as a result, the growth rate of *K. vulgare* and 2-KLG productivity increased 27.4% and 28.2%, respectively (Zhang et al., 2010a). However, those results still do not provide a clear and comprehensive answer to the question: what does *B. megaterium* provide and what is the limiting factor for *K. vulgare* growth? In order to address this issue, recently, Zhou et al. (2011) employed metabolomic strategies to determine that *B. megaterium* secretes some metabolites, such as erythrose,

* Corresponding author at: State Key Laboratory of Food Science and Technology, Jiangnan University, 1800 Lihu Road, Wuxi 214122, China. Tel.: +86 0510 85197875; fax: +86 0510 85197875.
E-mail address: mingll@jiangnan.edu.cn (L. Liu).

0168-1656/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jbiotec.2012.05.015

erythritol, guanine, and inositol, that assist *K. vulgare* growth. Furthermore, Ma et al. (2011) used quantitative systems biological analysis to demonstrate that the *B. megaterium* sporulation process releases adenine, guanine, xanthine, and hypoxanthine to help *K. vulgare* to resist reactive oxygen species (ROS), enhance energy production, and promote growth.

In fact, the essential requirement for understanding the *K. vulgare* and *B. megaterium* physiological interaction is to provide insight into what is insufficient in *K. vulgare*. As early as 2004, Leduc et al. (2004) suggested that the purine nucleotides and deoxythymidylate biosynthesis pathways were probably insufficient in *K. vulgare*. Moreover, based on the annotation of genome sequences of *K. vulgare*, we reconstructed the synthesis pathways of the amino acids and found that *K. vulgare* is deficient in one or more key enzymes in the *de novo* synthesis pathways of certain amino acids. The addition of those amino acids increased 2-KLG productivity by 11.8–20.4% (Liu et al., 2011a). Currently, the genome sequences of three *K. vulgare* strains, namely WB0104, Y25, and WSH001, have been completed and published (Liu et al., 2011b; Xiong et al., 2011; Yang et al., 2006). It is believed that the genome sequences may provide a global view of the physiological characteristics of *K. vulgare*.

Based on the integration of genome sequence information and detailed biochemical information, the reconstruction of a genome-scale metabolic model (GSMM) is an important platform for gaining comprehensive insight into microbial physiology (Kim et al., in press; Liu et al., 2010).

In this study, on the basis of genome annotation and biochemical and physiological data on *K. vulgare*, a genome-scale metabolic model for *K. vulgare* WSH001, a 2-KLG-producing bacterium used in the vitamin C industry, is presented. Furthermore, the metabolic features of *K. vulgare* were comprehensively elucidated with the help of algorithm methods. The model presented here provides a useful framework for understanding and exploring the metabolic capacity and physiology of *K. vulgare*.

2. Materials and methods

2.1. Metabolic network reconstruction

The main reconstruction procedure follows the published protocol (Thiele and Palsson, 2010). The genome information of *K. vulgare* WSH001 was obtained from our previous sequencing study (GenBank accession no. CP002018, CP002019 and CP002020) (Liu et al., 2011b). To start the reconstruction, the rapid annotation using subsystem technology (RAST) server (Aziz et al., 2008) and the KEGG automatic annotation server (KAAS) (Moriya et al., 2007) were used to re-annotate the genome of *K. vulgare* WSH001. Then, RAST annotation was put through the Model SEED pipeline (Henry et al., 2010) to obtain an initial network reconstruction. This initial network was integrated with information extracted from KAAS annotation and the KEGG database (Kanehisa et al., 2010), which led to a draft network. Refinement of the draft metabolic network was completed with the help of several public databases, including KEGG (Kanehisa et al., 2010), MetaCyc (Caspi et al., 2010), UniProt (Jain et al., 2009), BRENDA (Scheer et al., 2011), and TCDB (Saier et al., 2009). The gene–protein-reaction (GPR) association of the ABC transporter was mainly identified by KEGG Modules (Kanehisa et al., 2010). Reactions that were manually extracted from the published literature and experimental data were added. New assigned open reading frames (ORFs) obtained from RAST annotation in the reconstruction were identified by the known physiological features of *K. vulgare* or a local sequence similarity search (BLASTp) of the UniProtKB/Swiss-Prot database (Jain et al., 2009). Metabolic gaps that existed in the construction process were filled by reviewing published papers and our own unpublished experimental observations and adding the necessary reactions taken from the KEGG or MetaCyc database (for detail, see Additional File 1).

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2012.05.015.

2.2. Biomass composition

The biomass composition of *K. vulgare* was assumed to consist of eight components: proteins, DNA, RNA, lipids, peptidoglycans, lipopolysaccharides (LPS), glycogen, and soluble pool (Feist et al., 2007). Since no detailed information on the biomass composition of *K. vulgare* has been published, the protein, RNA, DNA, and lipid contents were treated as the same as that of *Escherichia coli* (Feist et al., 2007). For the detailed composition of each macromolecule, information from the genome sequence and the published literature on *K. vulgare* was applied (Liu et al., 2011b; Urbance et al., 2001). The detailed description of the biomass composition can be found in Additional File 2.

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2012.05.015.

2.3. Flux balance analysis

Flux balance analysis (FBA) is a widely used mathematical approach for analyzing flux distributions in a metabolic network (Orth et al., 2010). The reconstructed metabolic network formed a stoichiometric matrix $\mathbf{S}_{m\times n}$, where m represents the number of metabolites and n the number of reactions or fluxes to be computed. In the pseudo-steady-state, metabolite concentrations and fluxes are assumed to be constant. The model can be represented as: $\mathbf{S} \times \mathrm{v} = 0$, where vector v signifies the reaction flux. As the number of reactions (n) typically exceeds the number of metabolites (m), the system is under-determined. In order to get a feasible solution, linear programming (LP), subject to mass balance and flux constraints, was used by introducing an optimization problem. Setting the biomass equation as the objective function (Z), this LP problem was formulated as follows: maximize $Z = \mathrm{c}^{\mathrm{T}} \times \mathrm{v}$, subject to: $\mathbf{S} \times \mathrm{v} = 0$ and $\mathrm{v}_{i,\min} \le \mathrm{v}_i \le \mathrm{v}_{i,\max}$ (for $i = 1,\ldots,n$). Here, c is a row vector indicating the influence of individual fluxes on the objective function, $\mathrm{v}_{i,\min}$ and $\mathrm{v}_{i,\max}$ are the lower and upper bounds of the flux v_i. FBA was performed using COBRAToolbox-2.0 in MATLAB, with GLPK as the linear programming solver (Schellenberger et al., 2011). Constraint conditions used in the simulation process can be seen in Additional File 3.

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2012.05.015.

2.4. Single gene and reaction deletion

Single reaction deletion analysis was carried out by setting the flux of each reaction to zero, then running FBA to simulate the optimal growth rate. Reactions that led to poor growth, lower than 10^{-3} of the optimal value, were considered essential, whereas reactions that led to a decrease in the growth rate of *K. vulgare* that was nevertheless larger than 10^{-3} of the optimal value were considered partial essential. For gene deletion, the fluxes of all reactions associated with a particular gene were set to zero and FBA was used to compute the optimal growth rate. The criteria for the identification of essential and partial essential genes were the same as for reactions. Gene and reaction essentiality analysis was performed using COBRAToolbox-2.0 in MATLAB, with GLPK as the linear programming solver (Schellenberger et al., 2011).

Table 1
General features of the model *i*WZ663.

Features	Number
Genome features	
Genome size (bp)	3,288,404
Total open reading frames (ORFs)[a]	3096
Metabolic model	
Total reactions	830
Biochemical reactions	576
Transport reactions	137
Exchange reactions	117
Metabolites	649
ORFs associated in model	663
ORF coverage[b] (%)	21.4

[a] The total number of ORFs involving the 41 new ORFs annotated by RAST.
[b] The number of ORFs incorporated in model *i*WZ663 divided by the total number of ORFs.

3. Results and discussion

3.1. Metabolic network reconstruction

The procedure of metabolic network reconstruction was elaborated in Section 2, but there were two special treatments that need to be elucidated. First, the KEGG modules, which include KEGG ORTHOLOGY (KO) identifiers and KAAS annotation, were used to build the gene–protein-reaction associations for the ABC transporters (see Additional File 3). Second, one notable characteristic in the *K. vulgare* WSH001 genome is that there are 97 ORFs responsible for the peptide (not including dipeptide) ABC transporters. However, this model did not include those proteins, due to difficulties in identifying the substrates and metabolites. Information about the peptide ABC transporters is listed in Additional File 4.

Supplementary material related to this article found, in the online version, at http://dx.doi.org/10.1016/j.jbiotec.2012.05.015.

Based on the facts in the literature about the physiological properties of *K. vulgare*, many new biochemical reactions that had no gene assignments were added during the gap-filling process. Those reactions were supplementary to the genome annotation and make the model more reliable (Additional File 4). For example, in the literature *K. vulgare* has been shown to utilize 17 different substrates as carbon sources (Urbance et al., 2001), whereas only 10 substrates could be metabolized according to the draft model. The utilization of other carbon sources could be realized by adding certain reactions that are responsible for assimilating these substrates through KEGG database annotation (Kanehisa et al., 2010). Another example is that the model could not synthesize LPS, a representative structural component of the gram-negative strain. The results demonstrated that some genes encoding the key enzymes for lipopolysaccharide synthesis, such as ADP-L-glycero-D-manno-heptos-6-epimerase (EC 5.1.3.20), could not be assigned by genome annotation.

Following the above protocols, a genome-scale metabolic model for *K. vulgare* WSH001, named *i*WZ663 (Reed et al., 2003), was reconstructed. Model *i*WZ663 consists of 663 genes, 649 metabolites, and 830 reactions. The total of 663 genes in this model represents about 21.4% of *K. vulgare*'s ORFs (Table 1). Among the annotated genes, 41 genes obtained from the RAST server (Aziz et al., 2008) were completely new when compared with the published genome annotation (Liu et al., 2011b). Some of them are essential for the biosynthesis of biomass components, such as peg.1743, annotated as UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanine ligase (EC 6.3.2.10), which is the key enzyme in the peptidoglycan biosynthesis pathway. These ORFs were identified by the known physiological characteristics of *K. vulgare* or a homology search (BLASTp) with UniProtKB/Swiss-Prot (Jain et al., 2009). The new ORFs identified can be found in Additional File 4. Meanwhile, 209 reactions were involved with no gene assignment for biochemical evidence or network connectivity. The 830 reactions belong to 14 metabolic subsystems (Fig. 1). The full list of reactions and metabolites is available in Additional File 3.

3.2. Simulations of the growth

The genome-scale metabolic model involves nearly all the metabolic activities in a strain, so the basic physiology of the specific strain can be inferred from it. Here, FBA with certain constraints was used to investigate the metabolic properties of the model *i*WZ663. In this study, using the consumption rate of L-sorbose (19.7 mmol/gDW/h) and the amino acids (values were all set to 0.1 mmol/gDW/h) as constraint conditions (see Additional File 3 for details), the *in silico* cell growth rate of *K. vulgare* was simulated as $\mu = 0.695\ h^{-1}$. Compared with the experimental result ($\mu = 0.507\ h^{-1}$) (Zhang et al., 2010a), the simulated result was higher by 37.1%. The sensitivity of the predicted growth rate to the L-sorbose consumption rate was investigated (Additional File 1). When the uptake rate of L-sorbose was lower than 6.6 mmol/gDW/h, cell growth was L-sorbose-limited. However, when the L-sorbose uptake rate was higher than that value, the growth rate did not change, indicating the existence of other constraints that could inhibit the cell growth.

Previous studies showed that the compositions of corn steep liquor powder (CSLP) from different batches could affect the growth of *K. vulgare* and the production of 2-KLG (Zhang et al., 2011). Amino acids, which account for 30% of CSLP by dry weight, were shown to be the key factors. In order to study the influence of the uptake rates of amino acids on cell growth, we set the amino acids uptake rates to different levels (0.01–0.25 mmol/gDW/h) and FBA was performed to predict the optimum cell growth rates. A linear positive correlation between the uptake rates of the amino acids and the growth rate was found, which proved that amino acids are important limiting factors of the growth of *K. vulgare* (Additional File 1). The detail discussion of amino acid metabolism in *K. vulgare* can be found in the Section 3.4.2. As the growth rate of *K. vulgare* in our batch culture ranged from 0.08 to $0.9\ h^{-1}$ (Zhang et al., 2010a, 2011, 2010c), the appropriate uptake rates of amino acids that are used for *in silico* analysis should be set to 0.01–0.13 mmol/gDW/h.

In addition, oxygen is also an important factor for the growth of *K. vulgare* in L-sorbose-CSLP medium. Robustness of the oxygen consumption rate indicated the growth predicted by the model is oxygen-limited when the oxygen consumption rate is below 15.2 mmol/gDW/h (Additional File 1). However, when the oxygen consumption rate exceeds a certain value (91 mmol/gDW/h), growth of *K. vulgare* will be inhibited. Further, the growth rate is unchanged when the oxygen consumption rate ranges between 15.2 and 91 mmol/gDW/h. When the oxygen consumption rate is set to 9.9 mmol/gDW/h, the predicted growth rate is $0.492\ h^{-1}$, lower by 3% than the experimental result. Thus, with proper constraints, the model *i*WZ663 could be of great help in predicting the growth of *K. vulgare*.

However, as the production of 2-KLG from L-sorbose by *K. vulgare* has a non-growth associated character (Takagi et al., 2009), similar to secondary metabolite production such as penicillin, the model could not exactly predict the 2-KLG production rate when taking the biomass equation as the objective function. To simulate the production of 2-KLG, the exchange reaction of 2-KLG was set as the objective function. The uptake rate of L-sorbose was set to 6.6 mmol/gDW/h, and the growth rate was set to $0.0365\ h^{-1}$ (Takagi et al., 2009). When the oxygen uptake rate was

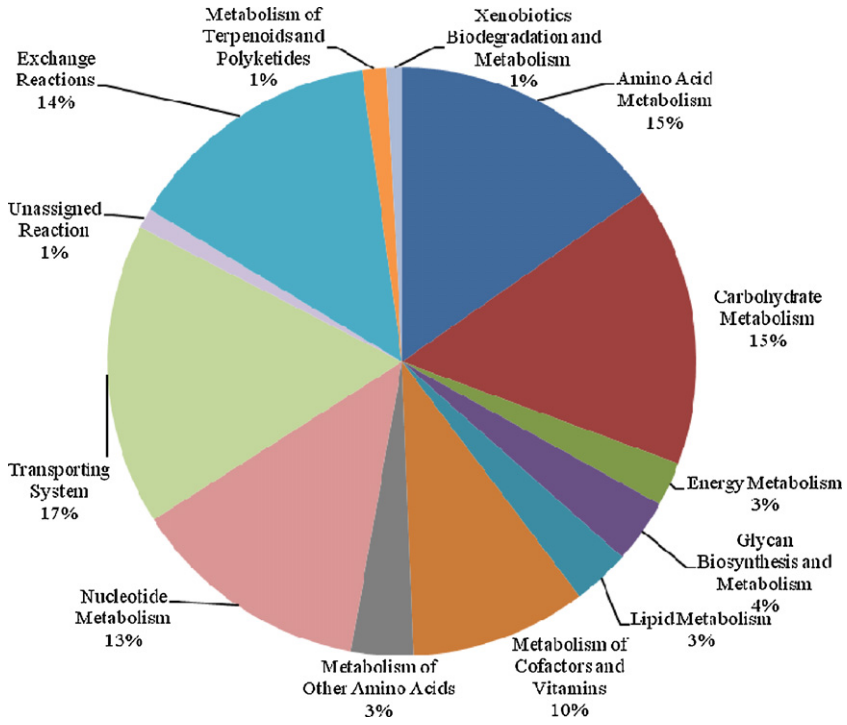

Fig. 1. Distributions of metabolic subsystems in *i*WZ663.

constrained to 3.4 mmol/gDW/h, the rate of production of 2-KLG was 5.9 mmol/gDW/h, about 1.7% lower than the experimental data (6 mmol/gDW/h) (Takagi et al., 2009).

3.3. Essential genes and reactions analysis

When cultured on the above L-sorbose fermentation medium, 116 genes (17.5% of the total genes in the model) in *K. vulgare* were identified as essential genes for growth by the FBA method (Additional Files 1 and 4). Moreover, it was interesting to note that all those essential genes are located on the chromosome, although the genes of two plasmids account for about 15% of the genome. Distributions of these essential genes are shown in Additional File 1 (Fig. S6), and nearly half of the essential genes are involved in nucleotide metabolism (27.6%) and amino acid metabolism (17.4%), suggesting the growth of *K. vulgare* could be greatly affected by these two metabolic subsystems (Leduc et al., 2004; Liu et al., 2011a). On the contrary, single deletions of 15 genes (2.3% of the total genes in the model) reduced the growth rate of *K. vulgare*. However, for the metabolic route from L-sorbose to 2-KLG, no essential gene was identified because there were multiple copies or isoenzyme genes for the enzymes in that metabolic route. Five copies of L-sorbose/L-sorbosone dehydrogenase, which catalyzes the conversion of L-sorbose to L-sorbosone, and two kinds of isoenzymes (L-sorbose/L-sorbosone dehydrogenase and L-sorbosone dehydrogenase), which catalyze the conversion of L-sorbosone to 2-KLG, were found in *K. vulgare* (see Section 3.4.1 for details). As shown in Additional File 1 (Fig. S5), 153 essential reactions (18.4% of the total reactions) and 22 partial essential reactions (2.7% of the total reactions) in model *i*WZ663 were predicted by the FBA method. Three of the essential reactions were amino acid exchange reactions, indicating that L-asparagine, L-cysteine, and L-methionine are necessary for *K. vulgare* growth. The important role of these 3 amino acids will be discussed in the next section.

3.4. Metabolic characteristics of *K. vulgare*

For a long time, research work regarding the metabolic characteristics of *K. vulgare* was focused on elucidation of the metabolic pathway of 2-KLG production from L-sorbose. Less attention was paid to an understanding of its physiological characteristics. Here, with the help of FBA and other constraint-based approaches, *i*WZ663 was used to investigate the metabolic characteristics of *K. vulgare* by simulating the optimal growth and global metabolic flux distributions.

3.4.1. Metabolism of L-sorbose

As shown in Fig. 2, the metabolic fate of L-sorbose in *K. vulgare* was analyzed by FBA in model *i*WZ663. The results demonstrated that L-sorbose is directly metabolized to 2-KLG or vitamin C, and a part of the carbon flux is channeled into the central carbon metabolic pathways (Fig. 2, Additional File 3). In the metabolic pathway from L-sorbose to 2-KLG, *K. vulgare* has five copies of L-sorbose/L-sorbosone dehydrogenase (KVU_0203, KVU_1366, KVU_2142, KVU_2159, KVU_PA0245) (Asakura and Hoshino, 1999), which convert L-sorbose to L-sorbosone and further to 2-KLG, and two copies of L-sorbosone dehydrogenase (KVU_0095, KVU_PB0115) (Miyazaki et al., 2006), which transform L-sorbosone to both 2-KLG and vitamin C. Previous researchers reported that vitamin C could be further degraded to other substrates (Sugisawa et al., 2005). However, model *i*WZ663 did not include such biochemical reactions, due to the fact that the degradation products remain unknown (Sugisawa et al., 1995). Moreover, 2-KLG can be converted into L-idonate by gluconate 2-dehydrogenase (KVU_PB_0008). L-Idonate is then metabolized into 6-phospho-D-gluconate through two enzymes: L-idonate 5-dehydrogenase (KVU_1353) and gluconate 5-dehydrogenase (KVU_1351). Finally, 6-phospho-D-gluconate enters the Entner-Doudoroff (ED) pathway and the pentose phosphate pathway (PPP) for energy and biomass production.

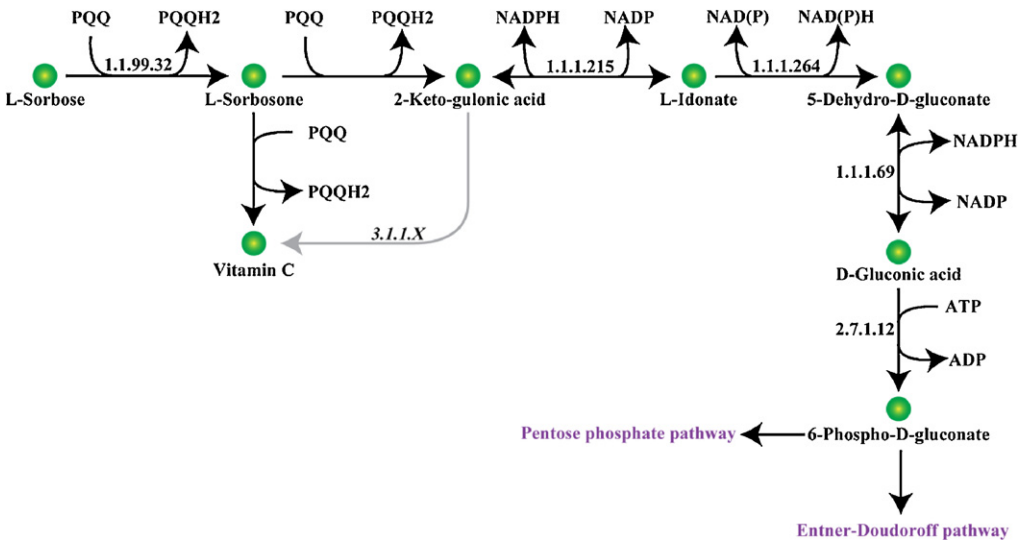

Fig. 2. L-sorbose metabolic pathway of *K. vulgare*. The pathways from L-sorbose to 2-keto-L-gulonic acid or vitamin C have been validated in the literature; other biochemical conversions in the map are based on the genome annotation of *K. vulgare* WSH001 (shown as black arrows and EC numbers). The lactonase, also known as carboxylic ester hydrolase (3.1.1.X), can be found in the *K. vulgare* genome, but this conversion has not been identified through biochemical experiments (shown as a gray arrow).

In order to increase the concentration and yield of 2-KLG, one metabolic engineering strategy is to remove or inhibit gluconate 2-dehydrogenase to block further metabolism of 2-KLG. When that gene was deleted *in silico*, L-sorbose was completely converted into 2-KLG according to the FBA simulation. Furthermore, three genes coding EC 3.1.1.17 (KVU_1414, KVU_2383) and EC 3.1.1.24 (KVU_PB0206) were annotated to lactonase (3.1.1.X), which could catalyze the conversion of 2-KLG to vitamin C in other species such as *E. coli* and *Zymomonas mobilis* (Asakura et al., 2000). Thus, exploration of these genes may afford the potential to realize the large-scale conversion of 2-KLG to vitamin C in the cell, replacing the current chemical step in the modern two-step fermentation (Yin et al., 1990).

3.4.2. Amino acid metabolism

Previous studies demonstrated that the kinds and concentrations of amino acids significantly affect the cell growth and 2-KLG production in 2-KLG fermentations (Li et al., 1996; Liu et al., 2011a; Zhang et al., 2011). *K. vulgare* has a strong transporting system for uptaking extracellular amino acids, and 48 genes coding amino acid transporters are involved in model *i*WZ663 (Additional File 3). Meanwhile, in the genome of *K. vulgare*, 103 genes are responsible for transporting exogenous peptides and about 58 genes for coding aminopeptidases or peptidases that can hydrolyze the peptides into amino acids (Additional File 4). The high levels of amino acids around *K. vulgare* in a solid rich medium observed by Zhou et al. (2011) could be attributed to those peptidases.

The number of reactions in which each amino acid participates in *i*WZ663 is shown in Table 2. It was found that L-glutamic acid and glycine were the most connected; L-glutamine, L-aspartic acid, L-alanine, L-serine, L-methionine, L-proline, and L-cysteine were highly connected; whereas the other 11 amino acids were poorly connected in *i*WZ663. Further, through the single-gene deletion method, L-asparagine, L-cysteine, and L-methionine were identified as essential amino acids for *K. vulgare* growth. Meanwhile, under an assumed environment in which amino acids are the sole carbon source and nitrogen source, FBA predicted that L-asparagine, L-aspartic acid, L-glutamic acid, L-glutamine, glycine, L-alanine, L-proline, L-serine, and L-threonine could be completely assimilated. These 8 kinds of amino acids could be converted into intermediate products of the tricarboxylic acid cycle (TCA cycle) and used for generating energy or other biomass components (Table 2, Additional File 3). Previous studies revealed that purines, such as guanine and adenine, from the media or the lysis of *B. megaterium*, play a vital role in *K. vulgare* growth and 2-KLG production (Leduc et al., 2004; Ma et al., 2011). Glycine, L-serine, and L-threonine, which serve as donors of one-carbon units in purine biosynthesis, exhibit the same promotion role (Zhang et al., 2011). Furthermore, L-aspartic acid and L-glutamine are also involved in purine biosynthesis.

3.4.3. Influence of sulfate metabolism on growth

The genome-scale metabolic model can be a helpful tool for understanding microbial physiology (Liu et al., 2010). As described above, single-reaction deletion analysis revealed that L-asparagine, L-cysteine, and L-methionine are essential components of *K. vulgare*. As L-cysteine and L-methionine are both sulfur-containing amino acids, the metabolic pathway of assimilatory sulfate

Table 2
Metabolic features of the amino acids in *i*WZ663.

Amino acid[a]	Reaction number[b]	TCA cycle[c]	Purine biosynthesis[d]
L-Isoleucine	3	–	–
L-Valine	3	–	–
L-Arginine	3	–	–
L-Lysine	3	–	–
L-Asparagine	4	+	–
L-Tryptophan	4	–	–
L-Histidine	4	–	–
L-Threonine	5	+	+
L-Leucine	5	–	–
L-Tyrosine	5	–	–
L-Phenylalanine	5	–	–
L-Cysteine	8	–	–
L-Proline	8	+	–
L-Methionine	10	–	–
L-Serine	11	+	+
L-Alanine	13	+	–
L-Aspartic acid	13	+	+
L-Glutamine	15	+	–
Glycine	24	+	+
L-Glutamic acid	40	+	+

[a] Essential amino acids for *K. vulgare* growth appear in bold and italics.
[b] Reaction number represents the number of reactions involving each amino acid.
[c] TCA cycle lists the amino acids that can (+) or cannot (–) be metabolized into intermediates of the TCA cycle.
[d] Purine biosynthesis indicates amino acids that can (+) or cannot (–) participate in purine biosynthesis.

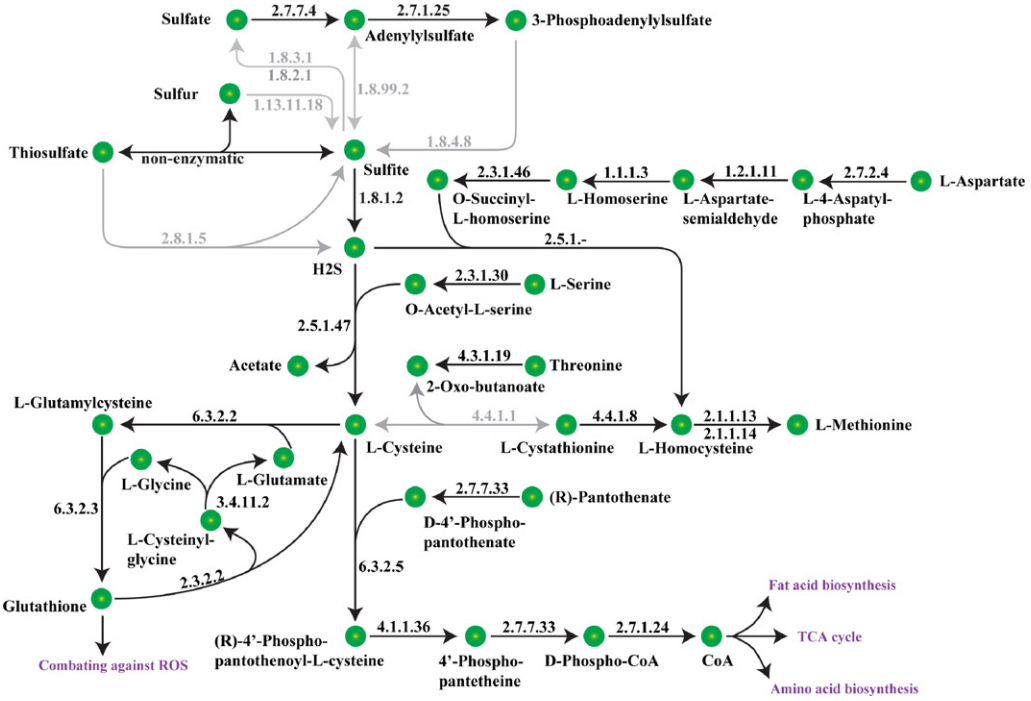

Fig. 3. Sulfate metabolism and its correlated pathways in *K. vulgare*. The black arrows and EC numbers indicate the existing reactions and enzymes in *K. vulgare*, whereas the gray arrows and EC numbers represent those that are absent. *K. vulgare* cannot reduce sulfate to sulfite and further to hydrogen sulfide, which is used for the biosynthesis of L-cysteine and L-methionine as a thiol group donor. The inability to synthesize L-cysteine hampers the formation of CoA and glutathione, which could influence amino acid and fatty acid biosynthesis, the TCA cycle, and defenses against ROS in *K. vulgare*.

reduction was carefully analyzed in this study. With sulfite as substrate, model *i*WZ663 exhibited the capability of synthesizing L-cysteine and L-methionine. It was shown that *K. vulgare* cannot reduce sulfate to sulfite and further to hydrogen sulfide, which can be assimilated into L-cysteine biosynthetic pathways. To investigate this issue, a metabolic network that involves sulfur metabolism and its correlated pathways was reconstructed and is shown in Fig. 3. As described in Fig. 3, sulfate cannot be converted into sulfite and further used to synthesize L-cysteine and L-methionine. L-Cysteine is an important precursor for CoA and glutathione. CoA can further form acetyl-CoA, which plays crucial roles in the TCA cycle and fatty acid biosynthesis. This analysis result was supported by the experimental data, when separate additions of 0.8 g/L L-cysteine and 1 g/L glutathione into the *K. vulgare* culture medium increased the growth of *K. vulgare* about 17.7% and 16.1% (unpublished results).

Metabolome profiling revealed that a high concentration of pyruvic acid accumulates in pure monocultures of *K. vulgare* (Zhou et al., 2011). This phenomenon can be partially explained by the fact that pyruvic acid cannot channel into the TCA cycle because of the shortage of acetyl-CoA. In the cytoplasm, no L-cysteine or L-methionine was detected (Zhou et al., 2011); this result is in accordance with the model predictions. Glutathione is known to play important roles in the detoxification of ROS-modified compounds in *K. vulgare* (Ma et al., 2011). In the two-step fermentation, the lysis of *B. megaterium* in the sporulation process, which may release the precursors of glutathione, could help *K. vulgare* in combating ROS (Ma et al., 2011).

4. Conclusions

In this study, a genome-scale metabolic model, *i*WZ663, for the industrial strain *K. vulgare* WSH001 that produces 2-KLG was reconstructed, based on genome annotation and literature mining. The model includes 663 genes, 649 metabolites, and 830 biochemical reactions, and the *in silico* growth rate showed high consistency with the batch culture. Using this model, the essential characteristics of *K. vulgare* that prevent its growth in pure culture were elucidated: defects in key biochemical reactions that involve amino acids, sulfate, cofactors, and vitamins in biosynthesis pathways.

Acknowledgements

The author would like to thank Dr. Jens Nielsen, Professor at Chalmers University of Technology, for technological help on the model reconstruction. This work was supported by grants from the Key Program of National Natural Science Foundation of China (20836003), Program for New Century Excellent Talents in University (NCET-10-0456), the Priority Academic Program Development of Jiangsu Higher Education Institutions, Program for Advanced Talents within Six Industries of Jiangsu Province (2011-NY033), the Open Funding Project of the State Key Laboratory of Bioreactor Engineering, the 111 Project (111-2-06) and the Fundamental Research Funds for the Central Universities (JUDCF10017).

References

Asakura, A., Hoshino, T., 1999. Isolation and characterization of a new Quinoprotein dehydrogenase, L-sorbose/L-sorbosone dehydrogenase. Bioscience, Biotechnology, and Biochemistry 63, 46–53.

Asakura, A., Hoshino, T., Kiyasu, T., Shinjoh, M., 2000. Manufacture of L-ascorbic acid and D-erythorbic acid. United States Patent 6146860.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

Bremus, C., Herrmann, U., Bringer-Meyer, S., Sahm, H., 2006. The use of microorganisms in L-ascorbic acid production. Journal of Biotechnology 124, 196–205.

Caspi, R., Altman, T., Dale, J.M., Dreher, K., Fulcher, C.A., Gilham, F., Kaipa, P., Karthikeyan, A.S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Paley, S., Popescu, L., Pujar, A., Shearer, A.G., Zhang, P., Karp, P.D., 2010. The

MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research 38, D473–D479.
Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.O., 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology 3, 121.
Hancock, R.D., Viola, R., 2002. Biotechnological approaches for L-ascorbic acid production. Trends in Biotechnology 20, 299–305.
Henry, C.S., Dejongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation optimization and analysis of genome-scale metabolic models. Nature Biotechnology 28, 977–982.
Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B.E., Martin, M.J., McGarvey, P., Gasteiger, E., 2009. Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics 10, 136.
Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M., 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research 38, D355–D360.
Kim, T.Y., Sohn, S.B., Kim, Y.B., Kim, W.J., Lee, S.Y. Recent advances in reconstruction and applications of genome-scale metabolic models. Current Opinion in Biotechnology, http://dx.doi.org/10.1016/j.copbio.2011.10.007, in press.
Leduc, S., de Troostembergh, J.C., Lebeault, J.M., 2004. Folate requirements of the 2-keto-L-gulonic acid-producing strain *Ketogulonigenium vulgare* LMP P-20356 in L-sorbose/CSL medium. Applied Microbiology and Biotechnology 65, 163–167.
Li, Q., Diao, J., Xiang, B., Cao, Z., 1996. Studies on metabolism of nitrogen source in fermentation of 2-keto-L-gulonic acid. Acta Microbiologica Sinica 36, 19–24.
Liu, L., Chen, K., Zhang, J., Liu, J., Chen, J., 2011a. Gelatin enhances 2-keto-L-gulonic acid production based on *Ketogulonigenium vulgare* genome annotation. Journal of Biotechnology 156, 182–187.
Liu, L., Li, Y., Zhang, J., Zhou, Z., Liu, J., Li, X., Zhou, J., Du, G., Wang, L., Chen, J., 2011b. Complete genome sequence of the industrial strain *Ketogulonicigenium vulgare* WSH-001. Journal of Bacteriology 193, 6108–6109.
Liu, L.M., Agren, R., Bordel, S., Nielsen, J., 2010. Use of genome-scale metabolic models for understanding microbial physiology. FEBS Letters 584, 2556–2564.
Ma, Q., Zhou, J., Zhang, W., Meng, X., Sun, J., Yuan, Y.J., 2011. Integrated proteomic and metabolomic analysis of an artificial microbial community for two-step production of vitamin C. PLoS One 6, e26108.
Miyazaki, T., Sugisawa, T., Hoshino, T., 2006. Pyrroloquinoline quinone-dependent dehydrogenases from *Ketogulonicigenium vulgare* catalyze the direct conversion of L-sorbosone to L-ascorbic acid. Applied and Environmental Microbiology 72, 1487–1495.
Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research 35, W182–W185.
Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? Nature Biotechnology 28, 245–248.
Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O., 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biology 4, R54.
Reichstein, T., Grüssner, A., 1934. Eine ergiebige Synthese der L-Ascorbinsäure (C-Vitamin). Helvetica Chimica Acta 17, 311–328.
Saier, M.H., Yen, M.R., Noto, K., Tamang, D.G., Elkan, C., 2009. The Transporter Classification Database: recent advances. Nucleic Acids Research 37, D274–D278.
Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Sohngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA the enzyme information system in 2011. Nucleic Acids Research 39, D670–D676.
Schellenberger, J., Que, R., Fleming, R.M., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.O., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6, 1290–1307.
Sugisawa, T., Miyazaki, T., Hoshino, T., 2005. Microbial production of L-ascorbic acid from D-sorbitol L-sorbose, L-gulose, and L-sorbosone by *Ketogulonicigenium vulgare* DSM 4025. Bioscience, Biotechnology, and Biochemistry 69, 659–662.
Sugisawa, T., Ojima, S., Matzinger, K.P., Hoshino, T., 1995. Isolation and characterization of a new vitamin C producing enzyme (L-gulono-(-lactone dehydrogenase) of bacterial origin. Bioscience, Biotechnology, and Biochemistry 59, 190–196.
Takagi, Y., Sugisawa, T., Hoshino, T., 2009. Continuous 2-keto-L-gulonic acid fermentation from L-sorbose by *Ketogulonigenium vulgare* DSM 4025. Applied Microbiology and Biotechnology 82, 1049–1056.
Thiele, I., Palsson, B.O., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5, 93–121.
Urbance, J.W., Bratina, B.J., Stoddard, S.F., Schmidt, T.M., 2001. Taxonomic characterization of *Ketogulonigenium vulgare* gen. nov., sp. nov. and *Ketogulonigenium robustum* sp. nov. which oxidize L-sorbose to 2-keto-L-gulonic acid. International Journal of Systematic and Evolutionary Microbiology 51, 1059–1070.
Xiong, X.H., Han, S., Wang, J.H., Jiang, Z.H., Chen, W., Jia, N., Wei, H.L., Cheng, H., Yang, Y.X., Zhu, B., You, S., He, J.Y., Hou, W., Chen, M.X., Yu, C.J., Jiao, Y.H., Zhang, W.C., 2011. Complete genome sequence of the bacterium *Ketogulonicigenium vulgare* Y25. Journal of Bacteriology 193, 315–316.
Xu, A., Yao, J., Yu, L., Lv, S., Wang, J., Yan, B., Yu, Z., 2004. Mutation of *Gluconobacter oxydans* and *Bacillus megaterium* in a two-step process of L-ascorbic acid manufacture by ion beam. Journal of Applied Microbiology 96, 1317–1323.
Yang, F., Jia, Q., Xiong, Z.H., Zhang, X.B., Wu, H.T., Zhao, Y., Yang, J., Zhu, J.P., Dong, J., Xue, Y., Sun, L.L., Shen, Y., Jin, Q., 2006. Complete genome analysis of *Ketogulonigenium* sp WB0104. Chinese Science Bulletin 51, 941–945.
Yin, G.L., Tao, Z.X., Yan, Z.Z., Ning, W.Z., Wang, C.H., Wang, S.D., 1990. Fermentation process. United State Patent 4935359.
Yin, G.L., Tao, Z.X., Yu, L.H., Wang, D.S., Dan, J.L., Yan, Z.Z., Ning, W.Z., Wang, C.H., Wang, S.D., Jiang, H.F., Zhang, X.D., Feng, X.Y., Zhao, Q., Wei, W.Q., 1980. Studies on the production of vitamin C precursor—2-keto-L-gulonic acid from L-sorbose by fermentation I. isolation screening and identification of 2-keto-L-gulonic acid producing bacteria. Acta Microbiologica Sinica 20, 246–251.
Zhang, J., Liu, J., Shi, Z.P., Liu, L.M., Chen, J., 2010a. Manipulation of *B. megaterium* growth for efficient 2-KLG production by *K. vulgare*. Process Biochemistry 45, 602–606.
Zhang, J., Zhou, J., Liu, J., Chen, K., Liu, L., Chen, J., 2011. Development of chemically defined media supporting high cell density growth of *Ketogulonicigenium vulgare* and *Bacillus megaterium*. Bioresource Technology 102, 4807–4814.
Zhang, J., Zhou, J., Liu, L., Liu, J., Chen, K., Du, G., Chen, J., 2010b. Enhancement of 2-keto-L-gulonic acid production using three-stage pH control strategy. Chinese Journal of Biotechnology 26, 1263–1268.
Zhang, J., Zhou, J.W., Liu, L.M., Liu, J., Chen, K.J., Du, G.C., Chen, J., 2010c. Enhancement of 2-keto-L-gulonic acid production using three-stage pH control strategy. Chinese Journal of Biotechnology 26, 1263–1268.
Zhou, J., Ma, Q., Yi, H., Wang, L., Song, H., Yuan, Y.-J., 2011. Metabolome profiling reveals metabolic cooperation between *Bacillus megaterium* and *Ketogulonicigenium vulgare* during induced swarm motility. Applied and Environmental Microbiology 77, 7023–7030.