ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Genome-Scale Modeling and In Silico Analysis of Ethanologenic Bacteria *Zymomonas mobilis*

Hanifah Widiastuti,[1] Jae Young Kim,[2] Suresh Selvarasu,[3] Iftekhar A. Karimi,[1] Hyungtae Kim,[2,4] Jeong-Sun Seo,[2,4] Dong-Yup Lee[1,3]

[1]Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576, Singapore; telephone: +65-6516-6907; fax: +65-6779-1936; e-mail: cheld@nus.edu.sg

[2]Macrogen Inc., World Meridian Venture Center, 60-24, Gasan-dong, Seoul 153-781, Korea

[3]Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01, Centros, Singapore 138668, Singapore

[4]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Yeongeon-dong 28, Jongno-gu, Seoul 110-799, Korea

Received 24 February 2010; revision received 23 August 2010; accepted 24 September 2010

Published online 21 October 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bit.22965

ABSTRACT: Bioethanol has been recognized as a potential alternative energy source. Among various ethanol-producing microbes, *Zymomonas mobilis* has acquired special attention due to its higher ethanol yield and tolerance. However, cellular metabolism in *Z. mobilis* remains unclear, hindering its practical application for bioethanol production. To elucidate such physiological characteristics, we reconstructed and validated a genome-scale metabolic network (*i*ZM363) of *Z. mobilis* ATCC31821 (ZM4) based on its annotated genome and biochemical information. The phenotypic behaviors and metabolic states predicted by our genome-scale model were highly consistent with the experimental observations of *Z. mobilis* ZM4 strain growing on glucose as well as NMR-measured intracellular fluxes of an engineered strain utilizing glucose, fructose, and xylose. Subsequent comparative analysis with *Escherichia coli* and *Saccharomyces cerevisiae* as well as gene essentiality and flux coupling analyses have also confirmed the functional role of *pdc* and *adh* genes in the ethanologenic activity of *Z. mobilis*, thus leading to better understanding of this natural ethanol producer. In future, the current model could be employed to identify potential cell engineering targets, thereby enhancing the productivity of ethanol in *Z. mobilis*.

Biotechnol. Bioeng. 2011;108: 655–665.

KEYWORDS: *Zymomonas mobilis*; genome-scale metabolic network; constraints-based flux analysis; bioethanol; industrial systems biotechnology

Correspondence to: Dong-Yup Lee

Contract grant sponsor: National University of Singapore (Academic Research Fund)
Contract grant number: (R-279-000-258-112)
Contract grant sponsor: Biomedical Research Council of A*STAR (Agency for Science, Technology and Research) Singapore
Additional supporting information may be found in the online version of this article.

Introduction

In recent years, continuous growth in global population and industries has increased energy consumption, resulting in a potential energy crisis. This situation is compounded by the increase in fuel prices and sharp decline of fossil fuels. Thus, there is a global stimulus to look for alternative renewable energy sources (Fukuda et al., 2009). In this regard, biofuels can be considered as major alternatives for conventional fossil fuels. Recently, US Energy Information Administration has reported that biomass-derived sustainable biofuels contributed to 53% of total renewable energy production (Energy, 2009); and amongst the known biofuel options, bioethanol has been widely used due to its highly desirable fuel characteristics such as high octane number, low cetane number, and high heat of vaporization (Balat et al., 2008). It is expected that the use of bioethanol can significantly reduce the dependency on the fossil fuels as well as their negative environmental effects (Balat et al., 2008; Sánchez and Cardona, 2008). In general, bioethanol can be produced from a variety of potential microbes including *Saccharomyces cerevisiae, Escherichia coli, Klebsiella oxytoca*, and *Zymomonas mobilis* by fermenting biological resources such as energy-rich crops or lignocellulosic biomass (Rubin, 2008). Of these, *S. cerevisiae* still remains the most widely employed microbe for ethanologenic fermentation (Bai et al., 2008). However, *Z. mobilis* has acquired special interest. It has numerous advantages over *Saccharomyces sp.* in terms of higher sugar uptake rate (q_s of about 10 g glucose/gDCW/h), 5–10% more ethanol per fermented glucose, productivity of over 4 g ethanol/L volume/h in batch culture, tolerance to toxic ethanol concentration of

up to 16% (vol/vol) in the medium (Dien et al., 2003), and lower downstream processing cost for purifying ethanol.

Despite the aforementioned valuable characteristics, most wild *Z. mobilis* strains can ferment only selected carbon sources such as glucose, fructose, and sucrose (Claassen et al., 1999), thus limiting their industrial use with other cheaper cellulosic materials, for example, xylose and arabinose (Mohagheghi et al., 2004). Although previous studies have successfully engineered *Z. mobilis* strains capable of fermenting various carbon sources for ethanol production (Gao et al., 2002; Lawford and Rousseau, 2002; Mohagheghi et al., 2002), the cellular metabolism pertaining to the various physiological states during carbon source utilization is poorly understood. Thus, it is highly required to develop a systematic approach for improving *Z. mobilis* strain possessing better fermentation capability. To this end, genome-scale modeling combined with metabolic engineering is desirable for characterizing the cellular physiology as well as designing engineered strains to improve ethanol production (Durot et al., 2009). This will provide the initial stepping-stone toward the ultimate aim of developing a competitive lignocellulosic ethanol plant.

With advancements in the genomic area, in silico models of various organisms are being reconstructed at the genome-scale, thereby enhancing our understanding of their cellular physiology (Durot et al., 2009; Francke et al., 2005; Lee et al., 2005; Reed et al., 2006). Such genome-scale in silico models have been built for various microbes, for example, *E. coli* (Feist et al., 2007), *S. cerevisiae* (Mo et al., 2009), and *Pichia pastoris* (Chung et al., 2010), and even for mammalian cells (Duarte et al., 2007; Selvarasu et al., 2009). There has been an initial attempt to model the metabolic network of *Z. mobilis* covering the amino acid biosynthetic pathway with emphasis on the central carbon metabolism (Tsantili et al., 2007). In addition, a dynamic model of *Z. mobilis* was also developed for analyzing the inhibitory effect of ethanol on glucose and sucrose fermentation (Lee and Huang, 2000). However, these simplified models have clearly overlooked the details of other major metabolic pathways, thus possibly leading to inaccurate predictions. Therefore, we reconstructed a genome-scale model of *Z. mobilis* metabolic network based on the annotated genome data of the wild type strain ZM4 (ATCC31821) and biochemical information from available literature. The model was then validated using batch culture data of *Z. mobilis* grown on various carbon sources. Subsequently, the model was functionally characterized by gene essentiality analysis to identify essential genes for ethanol production. It was followed by flux coupling analysis for confirming the functional roles in the ethanologenic activity of *Z. mobilis*. The knowledge gained from the analysis could be applied to identify potential cell engineering targets for enhancing carbon source utilization in *Z. mobilis* and improve the production of other desirable metabolites.

Materials and Methods

Metabolic Network Reconstruction

Figure 1 outlines the overall procedure for reconstructing a genome-scale network of *Z. mobilis* metabolism. The reconstruction was initiated based on the open reading frame (ORF) information from two sources of the annotated

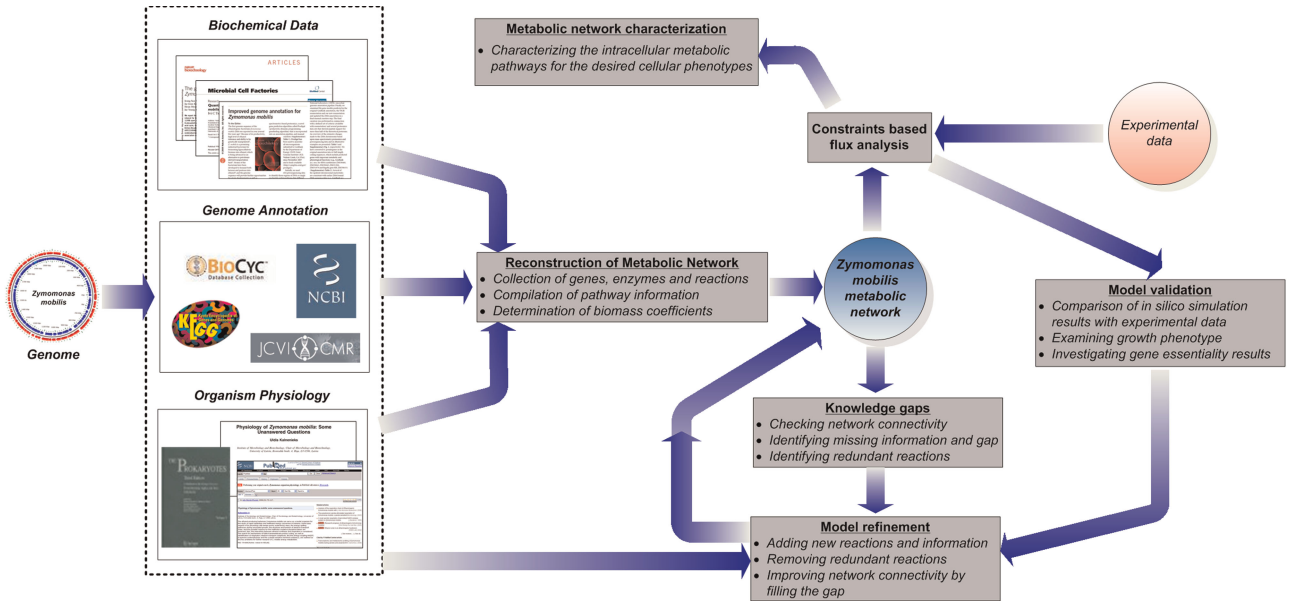


Figure 1. Schematic representation of the workflow involved in the reconstruction of *Z. mobilis* genome-scale model. [Color figure can be seen in the online version of this article, available at wileyonlinelibrary.com]

genome of *Z. mobilis* ATCC31821 (ZM4) (Seo et al., 2005; Yang et al., 2009a). The biochemical reactions corresponding to the genetic products were mainly compiled from KEGG (Kanehisa et al., 2006) and BioCyc (Caspi et al., 2006), resulting in a draft model comprising enzymatic reactions and metabolites. For each reaction, we identified and assigned gene–protein–reaction (GPR) association, enzyme and reaction localization, coenzyme and substrate specificity, and reaction directionality accordingly (Notebaart et al., 2006). Additionally, some spontaneous as well as non-gene-associated reactions were included, if their existence was evidenced by any physiological or experimental data (presence of substrate or product) from the literature and databases. The network connectivity was then checked by maximizing the production of individual metabolite precursors to biomass synthesis or the metabolites that lead to known by-products such as ethanol, acetate, sorbitol, acetoin, lactate, etc. from various carbon sources. The identified missing links were then filled using transport reactions to allow material exchange between the cell and its surrounding environment. In addition, the reactions were carefully examined for charge and mass balancing. The detailed list of genes, protein, reactions, and metabolites can be obtained from Supplementary 1, and also available as Systems Biology Markup Language (SBML) file (level 2, version 1, http://sbml.org/).

Biomass Composition and Maintenance Energy

An equation describing the conversion of various cellular components into cell biomass is provided in addition to a list of metabolic reactions in the reconstructed genome-scale model. This empirical reaction represents the conversion of the metabolic precursors into the building blocks of cell biomass such as carbohydrates, proteins, nucleic acids (DNA and RNA), and fatty acids. (Feist et al., 2007). In this study, the biomass composition was obtained from a number of references pertaining to the various cellular components of *Z. mobilis*. The energetic coefficients for both cell growth and maintenance were also included in the model based on the experimental evidence (Beyeler et al., 1984) while the method for estimating growth-associated (GAM) and non-growth-associated maintenance (NGAM) energies was adopted from Feist et al. (2006). The detailed calculation of biomass composition is provided in Supplementary 2.

Fermentation of *Zymomonas mobilis* ZM4 Strain

The reconstructed genome-scale model of *Z. mobilis* was validated using experimental data from batch fermentation. The batch cultures were prepared in two separate experiments under buffered and non-buffered conditions. The fermentation was carried out in RM medium, composed of 20.0–100.0 g L^{-1} glucose, 10.0 g L^{-1} yeast extract, 1.0 g L^{-1} $(NH_4)_2SO_4$, 1.0 g L^{-1} $MgSO_4 \cdot 7H_2O$, and 2.0 g L^{-1} KH_2PO_4. The initial culture pH was adjusted to 5.5 with 1 N NaOH. Glucose and phosphate were sterilized separately. Yeast extract and other minerals were then added into the medium aseptically. The main fermentations (100 g/L glucose) were performed from two sequential seed-cultures (20–50 g/L glucose) using mid-logarithmic cells under non-aerated conditions at 30 °C with mild-agitation. In the case of the buffered condition, the culture pH was simultaneously maintained at 5.0 with 2 N NaOH solution.

Analytical Methods

During fermentations, cell growth was monitored by measuring the optical density at 660 nm. Each absorbance value ($OD_{660\,nm}$) was then converted to the cell mass following the correlation coefficient previously reported by Young et al. (2005). Glucose, ethanol, and organic acid concentrations were measured by Shimadzu LC-20A (Shimadzu, Nakagyo-ku, Japan) high performance liquid chromatography (HPLC) system equipped with a UV detector (SPD-20A) and a refractive index detector (RID-10A). The culture supernatants taken from each culture were filtered with a 0.45 μm filter unit, then analyzed via two organic acid columns, Rspak KC-811 (300 × 8 mm, Shodex, Japan) and Aminex HPX-87H (300 × 7.8 mm, Bio-RAD Labs., Richmond, CA) using 5 mM sulfuric acid as the mobile phase at 0.6 mL min^{-1} (60 °C).

Constraints-Based Flux Analysis

After constructing the list of reactions involved in the *Z. mobilis* metabolism and balancing those reactions stoichiometrically, the cellular objective of biomass synthesis (specific growth rate) was maximized to determine the metabolic flux distributions attaining the optimal phenotype during the exponential growth phase. Assuming pseudo-steady state, the net sum of all production and consumption rates was set to zero for each internal metabolite. While carbon sources such as glucose, fructose, and xylose were constrained as limiting nutrients using their measured uptake rates, other metabolites (phosphate, biphosphate, sulfate, carbon dioxide, and water), and by-products (ethanol, acetate, lactate, succinate, acetoin, acetaldehyde, sorbitol, and xylitol) were unconstrained. Anaerobic cell growth condition was specified by fixing oxygen uptake rate at zero while the aerobic case was specified by keeping it unconstrained. In this study, the linear programming (LP) optimization problem was solved by using a stand-alone flux analysis program, MetaFluxNet (Lee et al., 2003). Model validation was carried out by comparing the in silico simulation results with the experimentally measured ethanol and biomass production rates.

Gene Essentiality and Flux Coupling Analyses

Gene essentiality analysis was performed by imposing additional constraints in the current linear flux model. The essentiality of a given gene was determined by evaluating the maximal growth rate, while constraining the corresponding reaction to be zero (in silico knock-out). The gene was identified as lethal, if the removal of the corresponding reaction leads to zero growth. Based on the identified essential reactions, we also conducted flux-coupling analysis to explore the dependency of their fluxes, following the flux-coupling finder method (Burgard et al., 2004). This involves maximizing and minimizing the ratio of any two fluxes of interest to obtain the degree of their dependency. More specifically, the flux of one reaction is fixed at a unit value, and the other flux is maximized and minimized. The resulting flux-coupling relationship can be fully coupled, directionally coupled, partially coupled, or uncoupled. Since gene essentiality and flux coupling may depend on medium nutrient condition, we assumed anaerobic growth on glucose by setting glucose and oxygen uptake rates at 10 mmol/gDCW/h and zero, respectively. For both analyses, we used CPLEX 12.1.0 solver in the General Algebraic Modeling System (GAMS) 23.2 platform (McCarl, 2009).

Results and Discussion

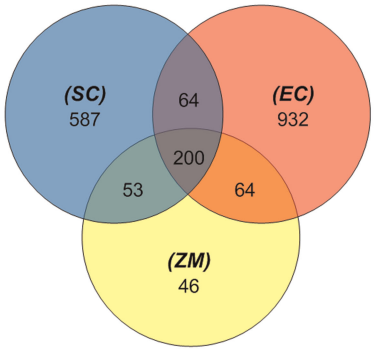
Genome-Scale Reconstruction of the *Z. mobilis* Metabolic Network

The reconstructed genome-scale metabolic network (*i*ZM363) of *Z. mobilis* ZM4 (ATCC31821) contained 704 metabolites and 747 reactions, classified into 48 specific pathways or subsystems based on their functional roles with the complete distribution presented in Supplementary 1. The model also accounts for 363 ORFs based on various reaction databases. These 363 ORFs correspond to 571 (out of 747) reactions. One hundred forty-four reactions describe metabolite transports via the cell membrane, while the remaining 33 reactions are included in the model to account for some physiological considerations with experimental evidence. For example, lysine and methionine synthesis pathways were not completed due to the gaps within the initial metabolic network constructed mainly based on the genome annotation (Seo et al., 2005). However, it has been reported that *Z. mobilis* can synthesize all amino acids (Goodman et al., 1982). Therefore, the gap was filled by adding genes *yfdZ* and *metB* for synthesizing lysine and methionine, respectively, based on the information obtained from the BioCyc database. Note that these two genes are putatively identified and updated in the database using the Bayesian method (Green and Karp, 2004) for gap filling.

We further fine-tuned the initial model by using a reannotated genomic sequence of *Z. mobilis* ZM4 (Yang et al., 2009a). This new set of gene annotation covers 634 genes (coding 463 enzymes) compared to 641 genes (coding 466 enzymes) in the previous genome sequence (Seo et al., 2005). Amongst the enzymes, 368 enzymes are common between new and old sequences, while 98 and 95 enzymes are unique for old and new ones, respectively. Several unique enzymes in the new sequence support the non-gene-associated reactions that have been included in the model during the gap-filling process. One of these reactions is the pyruvate conversion to acetaldehyde in the carbohydrate metabolism. This is catalyzed by the pyruvate decarboxylase enzyme (E.C. 4.1.1.1) encoded by *pdc* gene (ZMO1360). This *pdc*-catalyzed reaction is included to update the model after confirming its existence from Tittmann et al. (1998) and the BioCyc database. The presence of this gene was indeed confirmed by the new annotation results.

On the contrary, the new sequence excludes the phosphoglycerate phosphomutase (E.C. 5.4.2.1) which is the enzyme catalyzing one of the backbone glycolyctic reactions, namely the conversion from 3-phospho-D-glycerate to 2-phospho-D-glycerate. A similar gene (ZMO1240, *gpm*) coding this enzyme was reannotated for bisphosphoglycerate mutase (E.C. 5.4.2.4). The other important enzyme that is not included in the new sequence is D-glyceraldehyde-3-phosphate-lyase (E.C. 4.1.2.14) which catalyzes an important Entner–Doudoroff reaction, that is, the conversion of 2-dehydro-3-deoxy-6-phospho-D-gluconate to glyceraldehydes-3-phosphate and pyruvate. Akin to *gpm*, the gene ZMO0997 (*eda*) for D-glyceraldehyde-3-phosphate-lyase is available in the new sequence, but the gene is only considered to code 4-hydroxy-2-oxoglutarate aldolase catalyzing the conversion reaction from hydroxy-2-oxoglutarate to pyruvate and glyoxylate. Despite the existence of such conflict cases between previous annotation and reannotation results, we left those reactions in the model to increase the coverage. However, it clearly indicates a need for future corrections with experimental confirmation. These examples also show that mere genomic data are not sufficient to build a complete genome-scale model and it is necessary to refer to biochemical and experimental data for validation in an iterative way (Park et al., 2009).

Figure 2 presents overall network features of *Z. mobilis* genome and *i*ZM363 model, and its comparison with other ethanol producers, *E. coli* and *S. cerevisiae*. Major metabolic pathways for *Z. mobilis*, including the glycolytic pathway, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, and amino acids biosynthesis, are illustrated in Figure 3. As a unique feature of *Z. mobilis* metabolism, the PPP and TCA cycle were incomplete because the genes for phosphogluconate dehydrogenase (*pgd*) and transaldolase (*tal*) in PPP as well as 2-oxoglutarate dehydrogenase complex (s*ucABCD*) and malate dehydrogenase (*mdh*) in TCA were not found. However, the deficiency of the genes provides some operational advantages in *Z. mobilis*: less carbon flux is diverted to the PPP and TCA cycle, thus leading to a highly efficient and homofermentative ethanol production at lower biomass yield. This insight into the *Z. mobilis* metabolism could offer an idea of how this microbe has evolved toward efficient ethanol synthesis. In the reconstructed model, we identified 193 dead end



Features	*Z. mobilis*	*E. coli*	*S. cerevisiae*
Total genes (ORFs)	1998	4289	4841
Genes involved in genome-scale model	363	1260	904
ORF coverage (%)	18	29	19
Reactions in genome scale model	747	2077	1577
Metabolites in genome scale model	704	1387	1228
Number of essential reactions for cell growth (%)	196 (26%)	352 (17%)	200 (13%)
Ethanol yield (Yp/s) [anaerobic/aerobic]	0.49/0.21	0.22/0.00	0.41/0.07
Specific growth rate (μ) [anaerobic/aerobic]	0.08/0.08	0.17/0.92	0.18/0.97

Figure 2. Comparison among *S. cerevisiae*, *E. coli*, and *Z. mobilis*. Venn diagram shows the number of genes in each organism. (ZM, *Z. mobilis*; EC, *E. coli*; SC, *S. cerevisiae*). [Color figure can be seen in the online version of this article, available at wileyonlinelibrary.com]

metabolites, which were either only produced or consumed, and 71 isolated reactions that were not connected with main network. These provide room for further model improvement by wet-lab and computational approaches in the future (Pitkänen et al., 2008).

Model Validation

We validated the model using two data sources: one is our own batch culture data of wild type *Z. mobilis* strain grown on glucose while the other is NMR-measured internal flux data of an engineered *Z. mobilis* strain grown on glucose, fructose and xylose.

Batch Culture of Wild Type Strain Grown on Glucose

The predictive capability of *i*ZM363 was evaluated by employing constraints-based flux analysis based on batch culture data of *Z. mobilis* grown on glucose medium with the addition of yeast extract. It is believed that during the exponential growth phase, wild-type organisms typically have evolved to maximize growth rate (Schuster et al., 2008). Hence, the physiological behavior in the growth phase can be simulated by maximizing the cell biomass, while constraining the experimentally measured glucose uptake rate (q_s, mmol/gDCW/h). It should be noted that the concentrations of nutrients in the yeast extract were negligible when compared to glucose. Thus, glucose was assumed as the sole carbon source for this validation. At the outset, the simulation was carried out without incorporating the maintenance energy requirement (Simulation 1). The initial model prediction showed higher growth rates and lower ethanol production rates than experimental observations (Fig. 4A). This significant discrepancy between the observed and simulated growth rates might be attributed to various modeling artifacts such as incomplete biomass formulation, maintenance energy requirement, and some missing regulatory mechanisms involved in glucose utilization, which were not accounted for in the model (Reed et al., 2006). Among them, operational condition such as maintenance energy requirements including GAM and NGAM can be considered as one major factor affecting the simulated results. Previous studies have reported that the NGAM values for different strains of *Z. mobilis* varied between 2.88 and 4.97 mmol ATP/gDCW/h (Beyeler et al., 1984; Oliveira et al., 1992), showing decreasing tendency with reducing glucose concentration (Beyeler et al., 1984). Similarly, the GAM values may vary in different fermentation conditions (Lazdunski and Belaich, 1972). For instance, the GAM value reported by Oliveira et al. (1992) was 40.928 mmol ATP/gDCW/h for *Z. mobilis* Z-1-81. In order to obtain the appropriate GAM and NGAM values to best represent current physiological and fermentation conditions, glucose uptake rate (q_s, mmol/gDCW/h) was plotted against specific growth rate (which can be seen as the dilution rate in batch culture) for the experimental data with and without buffer addition (pH 5). The profile is linear for the buffer condition (Fig. 5). This plot can be typically generated from a steady-state continuous culture (chemostat condition). Apparently, the maintenance energy in *Z. mobilis* might be highly affected by the buffer addition, which could compensate the energy requirement for maintenance. This observation can be explained by the fact that one function of maintenance energy is to preserve the concentration gradients affecting the electrical potential gradients (Δ), which are highly dependent on pH in *Z. mobilis*. Thus, it is reasonable to take 82.638 mmol ATP/gDCW/h and 3.8964 mmol ATP/gDCW/h for the GAM and NGAM values, respectively, estimated from Figure 5. Surprisingly, the model predictions with those values (Simulation 2) resulted in much better prediction for both ethanol and biomass productions (Fig. 4B). It should be highlighted that unlike other genome-scale models, which can give good growth predictions by fixing most of uptake/secretion measurements, *i*ZM363 could predict both ethanol secretion and cell growth rates well without constraining other metabolite products, indicating the current model is capable of elucidating physiological characteristics of this natural ethanol producer.

NMR-Measured Internal Fluxes of an Engineered Strain on Various Carbon Sources

*i*ZM363 was further validated against published data on internal flux measurements in a xylose-fermenting recombinant *Z. mobilis* strain growing on glucose, fructose, and xylose (De Graaf et al., 1999). Internal fluxes analyzed by ^{13}C- and ^{31}P-NMR spectroscopy under different

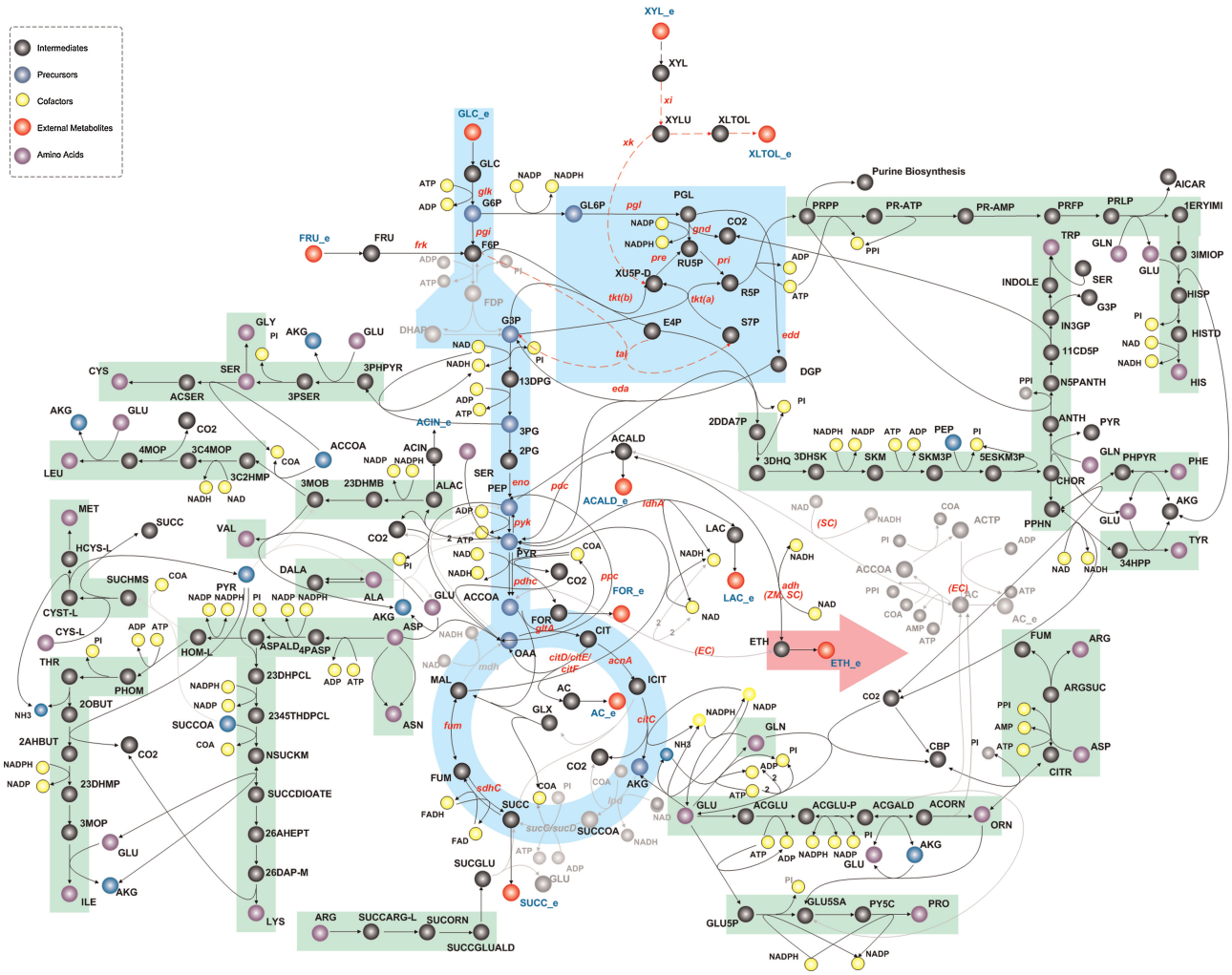


Figure 3. Central metabolic pathways of *Z. mobilis*, *E. coli*, and *S. cerevisiae*. Black lines represent active reactions in *Z. mobilis* while gray lines indicate *S. cerevisiae* and *E. coli* reactions that are not found in *Z. mobilis*. Dashed red lines are the inserted reactions in engineered strain for xylose utilization. Unique reactions are labeled by capital letters in bracket (ZM, *Z. mobilis*; EC, *E. coli*; SC, *S. cerevisiae*).

carbon-source conditions and simulation results are given in Table I. In order to mimic the recombinant xylose-fermenting strain (Zhang et al., 1995), *i*ZM363 was modified by inserting three heterelogous genes encoding xylose isomerase (*xylA*, EC. 5.3.1.5), xylulokinase (*xylB*, EC. 2.7.1.17), and translaldolase (*talB*, EC. 2.2.1.2) from *E. coli*. The growth simulation on each carbon source was initially carried out by using a GAM value of 82.638 mmol ATP/gDCW/h and a NGAM value of 3.8964 mmol ATP/gDCW/h (Simulation A). Note that both values were obtained from our batch-culture experimental data for the wild-type strain grown on glucose. Although the resulting internal fluxes through the glycolysis (ED pathway) were consistent with the NMR-measurements, the predicted cell growth was much higher than observed biomass yield (Table I). In order to resolve this discrepancy between the predicted and experimental biomass yield, the NGAM value was adjusted to fit the biomass yield (Simulation B). Remarkably, we observed much better prediction on the internal flux distribution with slightly improved ethanol production, when the value reached 40, 30, and 23 mmol ATP/gDCW/h for glucose, fructose, and xylose, respectively. Clearly, the adjusted NGAM values are beyond the range (2.88–4.97 mmol ATP/gDCW/h) reported for the wild-type *Z. mobilis* strains (Beyeler et al., 1984; Oliveira et al., 1992). Thus, more experimental data for this engineered strain are required to estimate maintenance energy during cell growth and confirm our predictions.

In the simulation with xylose as the carbon source, the reactions coded by the transketolase gene (tklB/tkt(a) and tkt(b), EC. 2.2.1.1), which had zero flux during growth on glucose and fructose, were activated. During the growth on xylose, the fluxes along the whole PPP had higher values compared to the glucose and fructose cases (Table I). Based on the model predictions and experimental observations, ethanol yield on xylose (0.38 g ethanol/g xylose) was lower

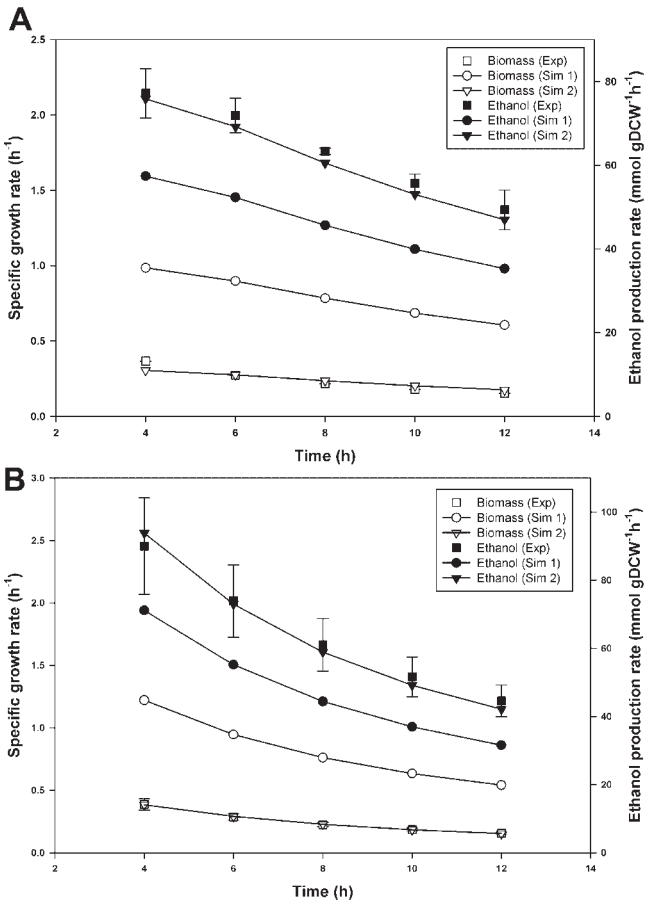


Figure 4. Experimental and simulated results for different time points during the exponential growth phase of the culture by using different values for GAM and NGAM. (**A**) Without buffer and (**B**) With buffer (pH was set at 5). (Exp, Experiment; Sim, Simulation).

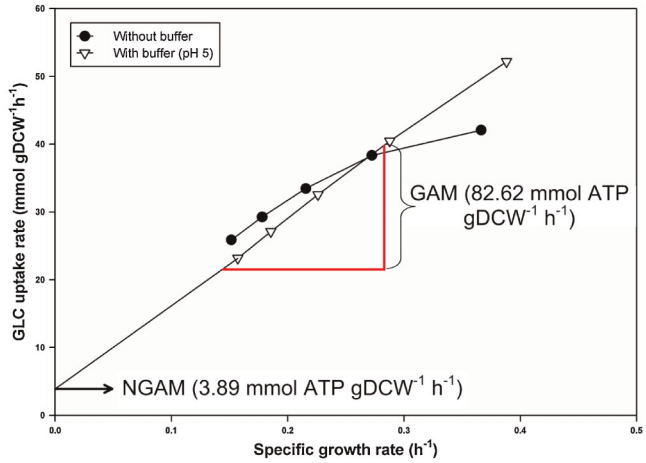


Figure 5. Determination of GAM and NGAM energy requirements in *Z. mobilis* from batch culture experiments with buffer. [Color figure can be seen in the online version of this article, available at wileyonlinelibrary.com]

than that on glucose (0.49 g ethanol/g glucose), while there is no difference between xylose and glucose utilizations in the biomass yield. This result showed a good agreement with the experimental observation. However, the deviations in the internal flux distributions between NMR-measurements and simulation results are relatively higher than the deviations in glucose and fructose utilizations. Note that xylose is transported via the glucose facilitated diffusion system (protein) that is mediated by the *glf* gene (ZMO0366) at the identical rate as glucose (Weisser et al., 1996). However, it was reported that in general *Z. mobilis* growth on xylose is slower than its growth on glucose (Kim et al., 2000). Hence, there must be a metabolic bottleneck in xylose fermentation that hinders xylose transportation possibly due to high intracellular xylose concentration. This regulatory mechanism was not entirely captured by the current model. Therefore, additional regulatory information is required to further improve the model's predictive capability.

Comparative Analysis With Other Ethanol Producers

We compared *Z. mobilis* (*i*ZM4363) with other genome-scale models of ethanol producing microbes, *E. coli* (*i*AF1260) (Feist et al., 2007) and *S. cerevisiae* (*i*MM904) (Mo et al., 2009) to investigate the similarity and conserved reactions among them and the uniqueness of each model. *i*AF1260 and *i*MM904 have 2077 and 1577 metabolic reactions with 954 and 761 gene-associated reactions, respectively. We found that all three models share 200 common genes in their metabolisms (Fig. 2). From the phylogenetic point of view, *Z. mobilis* is slightly closer to *E. coli* since 64 genes mainly from the amino acids synthetic pathways are shared with *E. coli* compared to 53 genes with *S. cerevisiae*. However, *Z. mobilis* shares highly relevant genes for various byproducts formation, particularly ethanol fermentation, with *S. cerevisiae*. In both *Z. mobilis* and *S. cerevisiae*, ethanol can be produced by fermenting carbon source (e.g., glucose) through pyruvate into ethanol via the pyruvate decarboxylase (*pdc*) and alcohol dehydrognease (*adh*) enzymes, whereas *E. coli* produces ethanol from acetyl-CoA (accoa) mediated by the pyruvate dehydrogenase (*pdhc*) enzyme complexes instead of the *pdc* enzyme. Thus, we observed the topological and functional similarity between *Z. mobilis* and *S. cerevisiae* from their anabolic properties mainly controlled by the *pdc* gene, indicating that *pdc* might play an important role in the ethanologenic activity for stimulating anaerobic ethanol production (Fig. 3).

To further explore how metabolism can be utilized for ethanol production under aerobic and anaerobic conditions, in silico simulations were conducted for three ethanol producers for a hypothetical glucose uptake rate of 10 mmol/gDCW/h as the sole carbon source. Overall, ethanol production was double in *Z. mobilis* than those in *E. coli* and *S. cerevisiae*, but the cell growth was lower (Fig. 2). Unlike *E. coli* and *S. cerevisiae*, where aerobic cell

Table I. Comparison of intracellular flux distribution by using simulation and NMR-fermentation data (De Graaf et al., 1999).

	Flux value (%) for growth on glucose			Flux value (%) for growth on fructose			Flux value (%) for growth on xylose		
Gene name	NMR	Simulation A	Simulation B	NMR	Simulation A	Simulation B	NMR	Simulation A	Simulation B
glk	100.0	100.0	100.0	—	—	—	—	—	—
frk	—	—	—	100.0	100.0	100.0	—	—	—
pgi	0.2	0.2	0.2	99.9	96.8	97.3	71.6	60.1	63.6
pgl	99.8	99.7	99.9	99.9	96.7	97.2	71.6	60.0	63.5
edd	99.4	99.1	99.6	99.6	96.1	99.9	58.6	60.0	50.4
eda	99.4	99.1	99.6	99.6	96.1	99.9	58.6	60.0	50.4
eno	98.8	96.8	98.7	97.1	95.0	98.7	93.1	92.0	93.5
pyk	97.4	91.6	96.6	95.5	89.3	96.1	92.4	89.1	92.3
pdc	192.6	184.0	191.5	190.3	183.6	189.4	149.8	145.5	136.2
adh	190.2	183.5	191.3	188.8	184.2	189.2	145.4	148.9	148.2
pre	0.1	0.1	0.1	0.1	0.1	0.1	23.1	34.7	31.2
pri	0.3	0.7	0.3	0.3	0.7	0.4	36.1	34.6	44.3
tkt(a)	0.0	0.0	0.0	0.0	0.0	0.0	35.8	30.1	31.8
tkt(b)	0.0	0.0	0.0	0.0	0.0	0.0	35.8	30.0	31.8
gnd	0.3	0.6	0.3	0.3	0.6	0.3	13	13.0	13.1
ppc	0.8	2.3	1.0	0.9	3.1	1.2	0.7	2.0	0.9
pdhc	1.0	1.0	1.1	1.2	1.0	1.0	0.9	2.9	1.2
xi	—	—	—	—	—	—	94.8	94.8	94.8
xk	—	—	—	—	—	—	94.8	94.8	94.8
tal	—	—	—	—	—	—	35.8	30.1	31.8
Biomass yield	0.017	0.043	0.017	0.022	0.043	0.022	0.017	0.035	0.017
Ethanol yield	0.49	0.47	0.49	0.48	0.47	0.48	0.37	0.38	0.38

Note: Simulation A was carried out with GAM of 82.638 mmol ATP/gDCW/h and NGAM of 3.8964 mmol ATP/gDCW/h while Simulation B used GAM of 82.638 mmol ATP/gDCW/h and NGAM of 40, 30 and 23 mmol ATP/gDCW/h for glucose, fructose, and xylose, respectively.

growths were about three- and two-fold than the anaerobic growths, respectively, there is no significant difference in *Z. mobilis* growth rates under two conditions. A recent study on transcriptomic and metabolomic pattern analysis confirmed similar specific growth rates for the anaerobic and aerobic cultures of *Z. mobilis* (Yang et al., 2009b). We also observed similar flux distribution patterns under both conditions, except the reactions for the acetate production. We believe that incomplete TCA and PPP in *Z. mobilis* can be considered as the main reason for these observations and the fact that major fluxes run in the glycolysis pathway and only small parts of them are directed to TCA and PPP. The observed rigidity in the *Z. mobilis* metabolism may also explain the lower cell growth compared to other ethanol producers (Fig. 2). On the other hand, there is a significant change in the metabolic fluxes for the anaerobic versus aerobic growths in *E. coli*. For the anaerobic growth, most fluxes passed through various pyruvate-producing metabolites such as acetate, glycolate, formate, ethanol, and succinate, while fluxes in PPP and TCA reached almost zero, resulting in much lower cell growth compared to the aerobic growth.

From the above comparative analysis, we can suggest various strategies to achieve improved ethanol production from *E. coli* and *S. cerevisiae*. An in silico analysis of *S. cerevisiae* (*i*MM904) under anaerobic growth condition for a glucose uptake rate of 10 mmol/gDCW/h showed that a portion (around 2.7 mmol/gDCW/h) of the glycolytic fluxes was used for biomass synthesis leaving 14.4 mmol/gDCW/h for the pyruvate branch (Fig. 6). From pyruvate, less than 10% (~1 mmol/gDCW/h) of the pyruvate-consuming flux entered TCA and was converted mainly into acetaldehyde, which eventually led to ethanol (12.1 mmol/gDCW/h) and acetate (0.2 mmol/gDCW/h). In comparison, both *E. coli* and *Z. mobilis* had a flux of approximately 19 mmol/gDCW/h from glycolysis to pyruvate. In *E. coli*, pyruvate was further converted to formate and accoa (17 mmol/gDCW/h each) by the pyruvate–formate lyase enzyme (*pfl* gene). While formate was secreted as a by-product, accoa was then converted to ethanol or diverted to acetate production through acetyl-phosphate (actp). Both acetate and ethanol were fermentation by-products with an equal flux of 8.3 mmol/gDCW/h. Uniquely, in *Z. mobilis*, a major flux (19.4 mmol/gDCW/h) from pyruvate was converted into acetaldehyde, which then mostly contributed to ethanol production. Therefore, from the observed flux distributions for three microbes, it could be suggested that the flux drain to biomass synthesis in *S. cerevisiae* should be reduced to enhance fluxes to pyruvate, which leads to the increased ethanol production. For *E. coli*, the improved ethanol production may be achievable by removing or down-regulating the flux from pyruvate to accoa (*pfl*, coA (coenzyme A) + pyruvate → accoa + formate) which would inhibit formate production and increase ethanol production. Similar strategies have been suggested by identifying possible candidates for genetic manipulation via in silico simulations (Pharkya and Maranas, 2006).

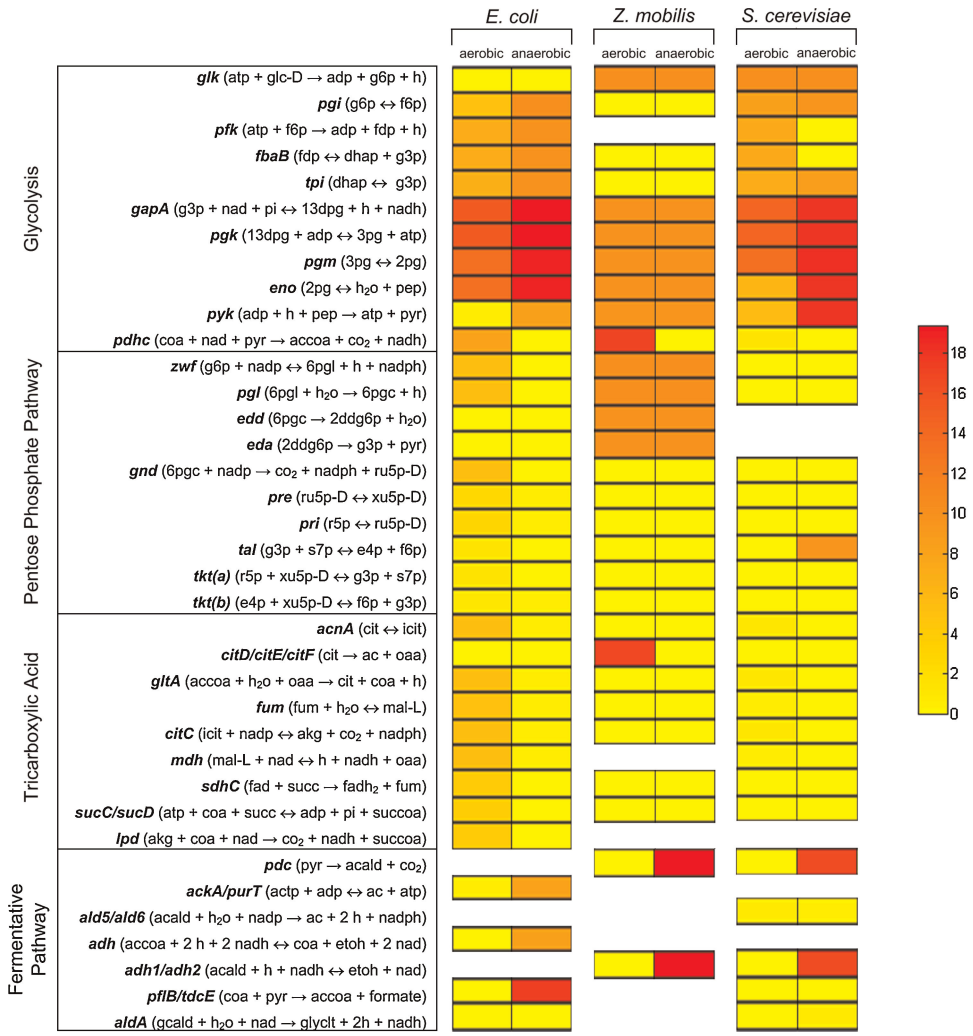
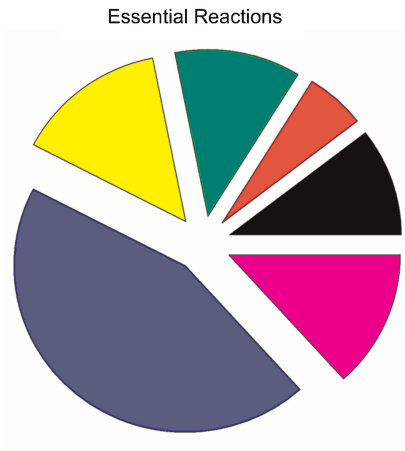


Figure 6. Flux distributions in *Z. mobilis*, *E. coli*, and *S. cerevisiae* under aerobic and anaerobic growth conditions (Unit: mmol/gDCW/h). [Color figure can be seen in the online version of this article, available at wileyonlinelibrary.com]

Gene Essentiality and Flux Coupling Analysis

We conducted gene essentiality analysis to explore the effect of gene deletion on *Z. mobilis* growth and concomitant ethanol production by simulating growth on glucose under the anaerobic condition, constraining each flux to be zero consecutively (see Materials and Method Section). The resultant 195 essential reactions for cell growth are almost 30% of total reactions. An overall distribution of these reactions along various pathways is presented in Figure 7. Interestingly, a number of lethal reactions (93 out of 195) were found in the amino acids metabolism (43% of the total amino acid reactions). Thus, it is apparent that amino acid synthesis is the rigid part of *Z. mobilis* metabolism, while other metabolic pathways are relatively flexible. Overall, it is likely that *Z. mobilis* metabolism has very rigid highway backbone for the biosynthetic precursors and fermentative products without many branches from a network perspective. Hence, there exist fewer alternative pathways to attain similar cellular functions, giving rise to such a higher percentage of essential reactions (26%) compared to other microbes, *E. coli* (17%) and *S. cerevisiae* (13%) as can be seen in Figure 2. It should also be noted that gene essentiality results are quite sensitive to model completeness. The growing research on this bacterium is expected to provide additional information on *Z. mobilis* metabolism, which can be used to fill the relatively higher remaining gaps (192 dead ends) in the current model and as such improve its connectivity in the future.

Essentiality analysis also showed that the removal of 102 lethal reactions resulted in zero ethanol production, indicating that redox potential can be balanced for cell growth only by utilizing those reactions and consequently secreting ethanol. Of them, *adh* (E.C. 1.1.1.1) and *pdc* (E.C. 4.1.1.1) were recognized as the most significant genes for high ethanologenic activity in *Z. mobilis* ethanol fermentation and have been used to construct *pet* operon (Gunasekaran and Raj, 2008), which was introduced to



Subsystem	Total no. of reactions	Essential reactions	% Essential
Carbohydrate metabolism	105	22	21
Energy metabolism	48	12	25
Lipid metabolism	86	25	29
Cofactors and vitamins metabolism	77	30	39
Amino acids metabolism	214	93	43
Nucleotides metabolism	81	28	35

Figure 7. Distribution of essential reactions in main subsystems of *Z. mobilis* metabolism. [Color figure can be seen in the online version of this article, available at wileyonlinelibrary.com]

engineer several microorganisms for ethanol production, for example, *E. coli KO11* (Yomano et al., 2008). In order to identify other key genes for ethanol production, we investigated the correlations among the essential reactions by using the flux-coupling analysis (Burgard et al., 2004). Amongst the 102 essential reactions for both cell growth and ethanol production, 35 reactions are partially coupled and 67 are either fully or directionally coupled (Data not shown). Not surprisingly, two consecutive reactions coded by *pdc* and *adh* for the ethanol production are fully coupled, together with ethanol transportation (secretion) and a reaction coded by *citDEF* leading to the precursors production (malate and oxaloacetate) for amino acid biosynthesis. Thus, this flux coupling result confirmed the pivotal role of *pdc* and *adh* for ethanol production and cell growth via two functional links with ethanol secretion and amino acids biosynthesis, respectively.

Concluding Remarks

In the current work, we reconstructed a genome-scale model of *Z. mobilis* ZM4 (ATCC31821) followed by model validation using batch fermentation data for the wild type and engineered *Z. mobilis* grown on various carbon sources. The predicted phenotypic behaviors and metabolic states were in good agreement with the experimental observation, thus allowing us to elucidate and systematically investigate the physiological characteristics and metabolic capability of this natural ethanol producer. From comparative analysis with *E. coli* and *S. cerevisiae*, we found that the incomplete PPP and TCA cycle in *Z. mobilis* provide some operational advantages leading to highly efficient and homofermentative ethanol production. Subsequently, we performed gene essentiality and flux coupling analysis to identify key genes for the ethanol production. It was found that *pdc* and *adh* genes are highly coupled with the ethanol secretion and anaplerotic reaction for amino acid synthesis, indicating their essential roles in the ethanologenic activity of *Z. mobilis*.

The work was supported by the Academic Research Fund (R-279-000-258-112) from the National University of Singapore and the Biomedical Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

References

Bai FW, Anderson WA, Moo-Young M. 2008. Ethanol fermentation technologies from sugar and starch feedstocks. Biotechnol Adv 26(1):89–105.

Balat M, Balat H, Öz C. 2008. Progress in bioethanol processing. Prog Energy Combust Sci 34(5):551–573.

Beyeler W, Rogers PL, Fiechter A. 1984. A simple technique for the direct determination of maintenance energy coefficient: An example with *Zymomonas mobilis*. Appl Microbiol Biotechnol 19(4):277–280.

Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res 14(2):301–312.

Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, Krummenacker M, Paley S, Pick J, Rhee SY, Tissier C, Zhang P, Karp PD. 2006. MetaCyc: A multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res 34: (Database issue) D511–D516.

Chung B, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, Lee H, Lee D-Y. 2010. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Factories 9(1):50.

Claassen PAM, van Lier JB, Lopez Contreras AM, van Niel EWJ, Sijtsma L, Stams AJM, de Vries SS, Weusthuis RA. 1999. Utilisation of biomass for the supply of energy carriers. Appl Microbiol Biotechnol 52(6):741–755.

De Graaf AA, Striegel K, Wittig RM, Laufer B, Schmitz G, Wiechert W, Sprenger GA, Sahm H. 1999. Metabolic state of *Zymomonas mobilis* in glucose-, fructose-, and xylose-fed continuous cultures as analysed by 13C- and 31P-NMR spectroscopy. Arch Microbiol 171(6):371–385.

Dien BS, Cotta MA, Jeffries TW. 2003. Bacteria engineered for fuel ethanol production: Current status. Appl Microbiol Biotechnol 63(3):258–266.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BÅ. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc. Natl Acad Sci 104(6):1777–1782.

Durot M, Bourguignon P-Y, Schachter V. 2009. Genome-scale models of bacterial metabolism: Reconstruction and applications. FEMS Microbiol Rev 33(1):164–190.

Energy USDo. 2009. Renewable energy consumption and electricity preliminary statistics 2008. Washington, DC: Energy Information Administration.

Feist AM, Scholten JCM, Palsson BO, Brockman FJ, Ideker T. 2006. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. Mol Syst Biol 2.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that

accounts for 1260 ORFs and thermodynamic information. Mol Syst Bioly 3.
Francke C, Siezen RJ, Teusink B. 2005. Reconstructing the metabolic network of a bacterium from its genome. Trends Microbiol 13(11): 550–558.
Fukuda H, Kondo A, Tamalampudi S. 2009. Bioenergy: Sustainable fuels from biomass by yeast and fungal whole-cell biocatalysts. Biochem Eng J 44(1):2–12.
Gao Q, Zhang M, McMillan J, Kompala D. 2002. Characterization of heterologous and native enzyme activity profiles in metabolically engineered *Zymomonas mobilis* strains during batch fermentation of glucose and xylose mixtures. Appl Biochem Biotechnol 98(1):341–355.
Goodman AE, Rogers PL, Skotnicki ML. 1982. Minimal medium for isolation of auxotrophic *Zymomonas* mutants. Appl Environ Microbiol 44(2):496–498.
Green M, Karp P. 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. BMC Bioinf 5(1):76.
Gunasekaran P, Raj KC. 2008. Ethanol fermentation technology—*Zymomonas mobilis*. Curr Sci 77:56–68.
Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: New developments in KEGG. Nucleic Acids Res 34: (Database issue) D354–357.
Kim IS, Barrow KD, Rogers PL. 2000. Kinetic and nuclear magnetic resonance studies of xylose metabolism by recombinant *Zymomonas mobilis* ZM4(pZB5). Appl Environ Microbiol 66(1):186–193.
Lawford H, Rousseau J. 2002. Performance testing of *Zymomonas mobilis* metabolically engineered for cofermentation of glucose, xylose, and arabinose. Appl Biochem Biotechnol 98–100(1):429–448.
Lazdunski A, Belaich JP. 1972. Uncoupling in bacterial growth: ATP pool variation in *Zymomonas mobilis* cells in relation to different uncoupling conditions of growth. J Gen Microbiol 70(2):187–197.
Lee W-C, Huang C-T. 2000. Modeling of ethanol fermentation using *Zymomonas mobilis* ATCC 10988 grown on the media containing glucose and fructose. Biochem Eng J 4(3):217–227.
Lee D-Y, Yun H, Park S, Lee SY. 2003. MetaFluxNet: The management of metabolic reaction information and quantitative metabolic flux analysis. Bioinformatics 19(16):2144–2146.
Lee SY, Lee D-Y, Kim TY. 2005. Systems biotechnology for strain improvement. Trends Biotechnol 23(7):349–358.
McCarl BA. 2009. McCarl's GAMS User Guide. GAMS Development Corporation.
Mo M, Palsson B, Herrgard M. 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol 3(1):37.
Mohagheghi A, Evans K, Chou Y-C, Zhang M. 2002. Cofermentation of glucose, xylose, and arabinose by genomic DNA-integrated xylose/arabinose fermenting strain of *Zymomonas mobilis* AX101. Appl Biochem Biotechnol 98–100(1–9):885–898.
Mohagheghi A, Dowe N, Schell D, Chou Y-C, Eddy C, Zhang M. 2004. Performance of a newly developed integrant of *Zymomonas mobilis* for ethanol production on corn stover hydrolysate. Biotechnol Lett 26(4):321–325.
Notebaart R, van Enckevort F, Francke C, Siezen R, Teusink B. 2006. Accelerating the reconstruction of genome-scale metabolic networks. BMC Bioinf 7(1):296.
Oliveira EG, Morais JO, Pereira N. 1992. Determination of the energy maintenance coefficient of *Zymomonas mobilis*. Biotechnol Lett 14(11): 1081–1084.
Park JM, Kim TY, Lee SY. 2009. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. Biotechnol Adv 27(6): 979–988.
Pharkya P, Maranas CD. 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8(1):1–13.
Pitkänen E, Rantanen A, Rousu J, Ukkonen E. 2008. A computational method for reconstructing gapless metabolic networks. Bioinf Res Dev 13:288–302.
Reed JL, Famili I, Thiele I, Palsson BO. 2006. Towards multidimensional genome annotation. Nat Rev Genet 7(2):130–141.
Rubin EM. 2008. Genomics of cellulosic biofuels. Nature 454(7206):841–845.
Sánchez ÓJ, Cardona CA. 2008. Trends in biotechnological production of fuel ethanol from different feedstocks. Bioresour Technol 99(13):5270–5295.
Schuster S, Pfeiffer T, Fell DA. 2008. Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol 252(3):497–504.
Selvarasu S, Wong VVT, Karimi IA, Lee D-Y. 2009. Elucidation of metabolism in hybridoma cells grown in fed-batch culture by genome-scale modeling. Biotechnol Bioeng 102(5):1494–1504.
Seo J-S, Chong H, Park HS, Yoon K-O, Jung C, Kim JJ, Hong JH, Kim H, Kim J-H, Kil J-I, et al. 2005. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat Biotechnol 23(1): 63–68.
Tittmann K, Mesch K, Pohl M, Hubner G. 1998. Activation of thiamine diphosphate in pyruvate decarboxylase from *Zymomonas mobilis*. FEBS Lett 441(3):404–406.
Tsantili I, Karim MN, Klapa M. 2007. Quantifying the metabolic capabilities of engineered *Zymomonas mobilis* using linear programming analysis. Microb Cell Factories 6(1):8.
Weisser P, Kramer R, Sprenger G. 1996. Expression of the *Escherichia coli pmi* gene, encoding phosphomannose—Isomerase in *Zymomonas mobilis*, leads to utilization of mannose as a novel growth substrate, which can be used as a selective marker. Appl Environ Microbiol 62(11):4155–4161.
Yang S, Pappas KM, Hauser LJ, Land ML, Chen G-L, Hurst GB, Pan C, Kouvelis VN, Typas MA, Pelletier DA, et al. 2009a. Improved genome annotation for *Zymomonas mobilis*. Nat Biotechnol 27(10):893–894.
Yang S, Tschaplinski T, Engle N, Carroll S, Martin S, Davison B, Palumbo A, Rodriguez M, Brown S. 2009b. Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. BMC Genomics 10(1):34.
Yomano L, York S, Zhou S, Shanmugam K, Ingram L. 2008. Re-engineering *Escherichia coli* for ethanol production. Biotech Lett 30(12):2097–2103.
Young JJ, Svenson CJ, Rogers PL. 2005. Over-expression of xylulokinase in a xylose-metabolising recombinant strain of *Zymomonas mobilis*. FEMS Microbiol Lett 244(1):85–92.
Zhang M, Eddy C, Deanda K, Finkelstein M, Picataggio S. 1995. Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. Science 267(5195):240–243.