ACS Sustainable Chemistry & Engineering

Research Article

pubs.acs.org/journal/ascecg

Elucidating Xylose Metabolism of *Scheffersomyces stipitis* for Lignocellulosic Ethanol Production

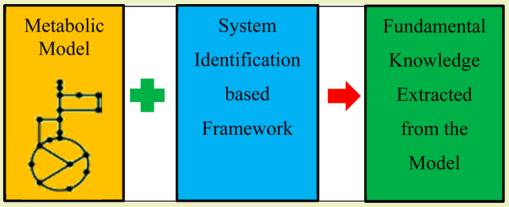
Meng Liang,[†] Andrew Damiani,[†] Q. Peter He,[‡] and Jin Wang*[,†]

[†]Department of Chemical Engineering, Auburn University, Auburn, Alabama 36849, United States

[‡]Department of Chemical Engineering, Tuskegee University Tuskegee, Alabama 38088, United States

Supporting Information

ABSTRACT: The conversion of pentose to ethanol is one of the major barriers of industrializing the lignocellulosic ethanol processes. As one of the most promising native strains for pentose fermentation, *Scheffersomyces stipitis* (formerly known as *Pichia stipitis*) has been widely studied for its xylose fermentation. In spite of the abundant experimental evidence regarding ethanol and byproducts production under various aeration conditions, the mathematical descriptions of the processes are rare. In this work, a constraint-based metabolic network model for the central carbon metabolism of *S. stipitis* was reconstructed by integrating genomic (*S. stipitis* v2.0, KEGG), biochemical (ChEBI, PubChem), and physiological information available for this microorganism and other related yeast. The model consists of the stoichiometry of metabolic reactions, biosynthetic requirements for growth, and other constraints. Flux balance analysis is applied to characterize the phenotypic behavior of *S. stipitis* grown on xylose. The model predictions are in good agreement with published experimental results. To understand the effect of redox balance on xylose fermentation, we propose a system identification-based metabolic analysis framework to extract biological knowledge embedded in a series of designed in silico experiments. In the proposed framework, we first design in silico experiments to perturb the metabolic network in order to investigate the interested properties and then perform system identification, whereby applying principal component analysis (PCA) to the data generated by the designed in silico experiments. By combining the in silico perturbation experiments with system identification tools, biologically meaningful information contained in the complex network structure can be decomposed and translated into easily interpretable information that is useful for biologist. The PCA analysis identifies the phenotypic changes caused by oxygen supply and reveals key metabolic reactions related to redox homeostasis in different phenotypes. In addition, the influence of the cofactor preference of key enzyme (xylose reductase) in xylose metabolism is investigated using the proposed approach, and the results provide important insights on cofactor engineering of xylose metabolism.



KEYWORDS: *Scheffersomyces stipitis, Xylose metabolism, Redox balance, Principal component analysis, Cofactor engineering, Systems biology*

INTRODUCTION

Lignocellulosic ethanol, considered as a second-generation biofuel,[1] has received increasing attention in recent years due to the increasing demand of fuel and issues caused by first-generation biofuels (i.e., fuels from food crops).[2] Efficient utilization of xylose is one of the major obstacles for commercial production of lignocellulosic ethanol.[3] *Scheffersomyces stipitis* (formerly known as *Pichia stipitis*)[4] has a set of physiological traits that make it potentially a valuable candidate for lignocellulosic ethanol production.[5] It is considered as one of the few yeasts that can metabolize xylose in high efficiency with few byproducts.[6] It has been shown that oxygen availability plays a critical role in xylose metabolism of *S. stipitis* due to redox balance.[5] In particular, the NAD(P)H-dependent XR and NAD^+-dependent XDH of *S. stipitis* create a cofactor imbalance resulting in xylitol formation under oxygen limited conditions.[7−10] Therefore, a better understanding of the cellular redox balance would provide valuable insight on how to improve ethanol production of *S. stipitis*. However, in spite of the abundant experimental evidence, little is known about the cellular details on how redox balance affects xylose metabolism.

Constraint-based stoichiometric metabolic network modeling has been proven to be an effective way to study the metabolism of many microorganisms. Among the many approaches, flux balance analysis (FBA) has been adopted in many studies and can be used to study many aspects of biochemical networks.[11] However, one of the difficulties in the existing metabolic network modeling approach is how to identify key reactions among hundreds or thousands of reactions when cellular metabolism is influenced by a given factor. To address this challenge, we propose a new approach to analyze xylose metabolism by integrating FBA and principal component analysis (PCA), and we apply this framework to study the

Special Issue: Sustainable Chemical Product and Process Engineering

Received: July 31, 2013
Revised: November 5, 2013
Published: November 8, 2013

   dx.doi.org/10.1021/sc400265g | ACS Sustainable Chem. Eng. 2014, 2, 38−48

central carbon metabolic network of *S. stipitis*. The proposed approach can reveal the key metabolism details of how different oxygen utilization rates would cause metabolism shifts. The influences of cofactor specificity of a critical enzyme in the xylose metabolism, xylose reductase (XR), was also studied. The analysis results provide valuable insights on possible cofactor engineering of *S. stipitis*.

METHODS

Construction of Metabolic Model. The central carbon metabolic network model of *S. stipitis* is constructed following the published protocol.[12] The network model is reconstructed based on the genomic and biochemical information of the organism available in its genome project,[5] KEGG database,[13] and available biochemical information. An overview of the reconstructed metabolic network model is shown in Figure 1, with detailed reactions provided in the Supporting Information. The metabolites involved in the model have all been verified and balanced with both element and charge based on biochemical information from ChEBI[14] and the PubChem compound database.[15] The model captures cell growth on glucose and xylose. Included in the model are 117 reactions with 66 as reversible reactions and 51 as irreversible reactions (including transport reactions). Fifteen compounds allowed to exchange with the external environment are glucose, xylose, NH_4^+, urea, O_2, CO_2, SO_4^{2-}, H^+, HO_4P^{2-} (Pi^{2-}), H_2O, ethanol, acetate, glycerol, xylitol, and biomass. The nongrowth associated maintenance requirement is tested within the range of [0.5, 3.5] mmol/gDCW/h according to different literatures,[16−19] and the value of 3.5 mmol ATP/gDCW/h is adopted (DCW stands for dry cell weight). The model consists of the reactions from intermediary metabolism including glycolysis, pentose phosphate pathway, tricarboxylic acid (TCA) cycle, glyoxylate and dicarboxylate metabolism, oxidative phosphorylation, nitrogen metabolism, and nicotinate and nicotinamide metabolism. The model also includes reactions of cell mass formation and synthesis of various precursors and common byproducts such as ethanol, glycerol, xylitol, and acetate. Some linear reactions in the model are lumped together for simplicity. Transport reactions, including passive diffusion, facilitated diffusion, and active transport, are also incorporated. Cell mass reaction in the model is assembled from the macromolecular components of the cells (i.e., proteins, nucleic acids, lipids, and carbohydrates).[12,20] The contribution of each component to cell mass and the appropriate coefficients for every building block are estimated from the genome data of *S. stipitis* and information of *S. cerevisiae*.[5,21,22] The cell mass term presented here reconnect the cell composition to 13 precursors along with energy (ATP), redox cofactors (NAD^+/NADH and $NADP^+$/NADPH), and other nutrients (Pi^{2-}, NH_4^+ and SO_4^{2-}). The 13 precursors are glutamate, acetyl-CoA, glycerol, oxaloacetate, phosphoenolpyruvate, glucose-6-phosphate, glyceraldehyde-3-phosphate, glutamine, ribose-5-phosphate, pyruvate, erythose-4-phosphate, and 3-phospho-D-glycerate.

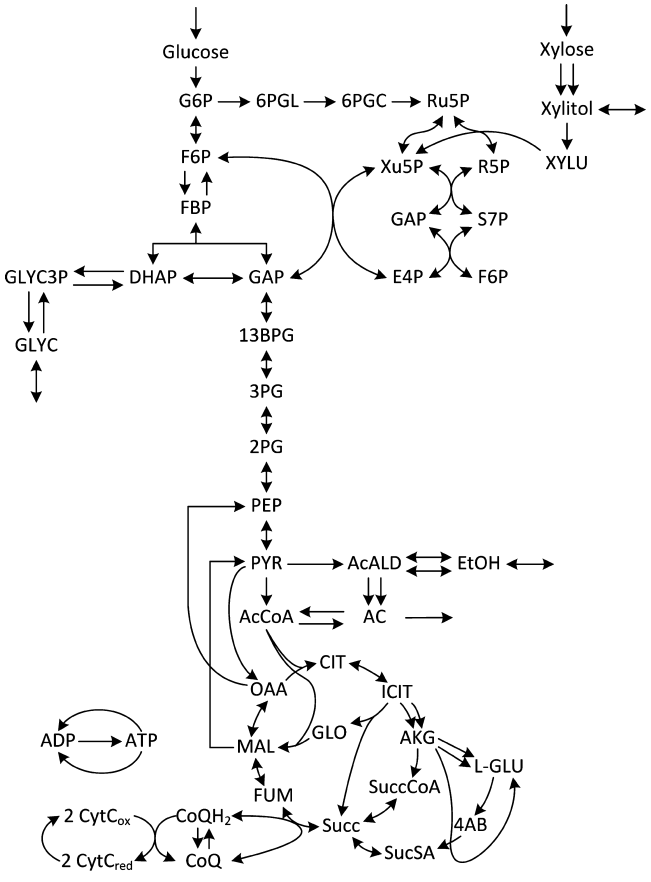


Figure 1. Overview of the metabolic network model. The double arrows in the same direction for a reaction indicate that the enzyme that catalyzes the reaction has affinity to different cofactors (NADH/NAD^+ and NADPH/$NADP^+$). This also applies to the other figures in this work.

Flux Balance Analysis (FBA). Flux balance analysis (FBA) is a commonly applied method to study metabolic networks, especially genome-scale metabolic networks.[23] In this work, FBA is applied to study the central carbon metabolism of *S. stipitis* using a publicly available constraints-based reconstruction and analysis (COBRA) toolbox for Matlab version 2.05.[23] The upper limits of uptake rate of xylose and oxygen under various conditions are defined in FBA to predict external secretion rates and internal net fluxes. Other exchange fluxes are constrained accordingly. Maximizing cellular growth rate is used as the objective function for all FBA simulations. The initial concentrations of different compounds that are allowed to exchange with the environment are not needed for FBA calculation. The carbon source (glucose or xylose) and oxygen uptake rates are fixed, while the other exchange compounds (NH_4^+, SO_4^{2-}, H^+, HO_4P^{2-}(Pi^{2-}), H_2O) are allowed to enter and exit the network freely. The products (CO_2, ethanol, acetate, glycerol, xylitol) are expelled from the network liberally, with the biomass being computed. The simulation results are analyzed further to reveal the intracellular mechanism of xylose metabolism.

Principal Component Analysis (PCA). Principal component analysis (PCA) is a commonly used multivariate analysis method for dimension reduction, which extracts the directions corresponding to the largest variations among different variables in a high dimensional data set.[24] It has been applied in the metabolomics studies to analyze metabolites profiles at given conditions.[25] In this work, we propose a new way of applying PCA to extract the underlying biological knowledge embedded in the data obtained through designed in silico experiments.

Proposed Method: FBA−-PCA. In microbial metabolism, hundreds and even thousands of reactions are involved. Existing approaches, such as elementary mode analysis (EMA) and flux balance analysis (FBA), can provide detailed flux distributions under different conditions and therefore provide descriptions of different phenotypes. However, by simply comparing different flux distributions, it is very difficult to extract the underlying biological knowledge, such as what key reactions or key correlations of different pathways govern the cellular metabolism of a given phenotype. In this work, we propose a system identification based metabolic flux analysis framework to extract such knowledge by integrating PCA with FBA. Specifically, in the proposed framework, we first design in silico experiments to perturb the metabolic network in order to investigate the interested properties, and then we perform system identification by applying PCA to the high dimensional data generated through the designed in silico experiments. By combining the in silico perturbation experiments with system identification tools, biologically meaningful information contained in the complex network structure can be extracted from a sufficient amount of in silico experimental data in a

dx.doi.org/10.1021/sc400265g | ACS Sustainable Chem. Eng. 2014, 2, 38−48

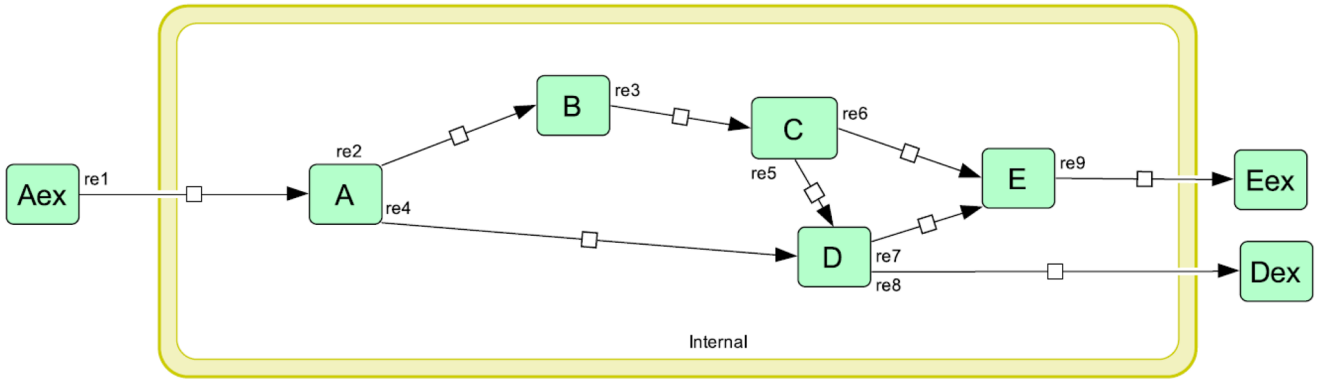


Figure 2. Reaction network scheme of the illustrative example.

form that is easily interpretable by biologists. It is worth noting that because the metabolic network is linear, if only one degree of perturbation is introduced within a series of in silico experiments, then one principal component (PC) is sufficient to capture 100% of the variation, provided that there is no saturation (i.e., flux reaching its upper/lower limit) nor network structure changes (i.e., activation/deactivation of reactions). In this case, the correlations among different reactions are fully captured by the loading of the PC. Therefore, by examining the loading, we can easily identify how the introduced perturbation propagates through the whole network and what reactions are affected most by the introduced perturbation. Here, we use an illustrative example to explain how the proposed method works.

Illustrative Example. In this illustrative example, we consider a simple reaction network as shown in Figure 2. The network consists of five metabolites and nine reactions, which are listed in Table 1. The resulted stoichiometric matrix is given in Table 2. Among all reactions, three are exchange reactions and six are internal reactions. The constraints we consider are: $0 \leq$ re1, ..., re9 ≤ 10. Below, we consider two case studies: one with maximizing production of metabolite D as the objective function of FBA and the other with maximizing production of metabolite E as the objective function. For both cases, we investigate how the flux distribution would be affected if we increase the pickup rate of substrate A. In particular, we would like to identify what reactions are affected most significantly if pickup rate of A increases.

Table 1. Internal and Exchange Reactions of the Illustrative Example

internal reactions		exchange reactions
re2: A → B	re5: C → D	re1: A_{ex} → A
re3: B → 0.5C	re6: C → 2E	re8: D → D_{ex}
re4: A → 2D	re7: 0.5 D → E	re9: E → E_{ex}

Table 2. Stoichiometric Matrix

	re1	re2	re3	re4	re5	re6	re7	re8	re9
A	1	−1	0	0	0	0	0	0	0
B	0	1	−1	−1	0	0	0	0	0
C	0	0	0.5	0	−1	−1	0	0	0
D	0	0	0	2	1	0	−0.5	−1	0
E	0	0	0	0	0	2	1	0	−1

Case Study I: FBA Objective Function = Maximizing Flux of re8 (Production of D). In this case study, we first conduct a series of 100 in silico experiments by varying the flux of re1 (pick up rate of A) from 2 to 4 mmol/gDCW/h with a step size of 0.02. This set of experiments results in a 9 × 100 data matrix, where each column represents the nine reaction fluxes for a given substrate pick up rate. We then perform PCA on the data matrix, which confirms that one principal component (PC) captures 100% of the variance contained in the data matrix. The scaled loading of the PC is plotted in Figure 3. The loadings show that

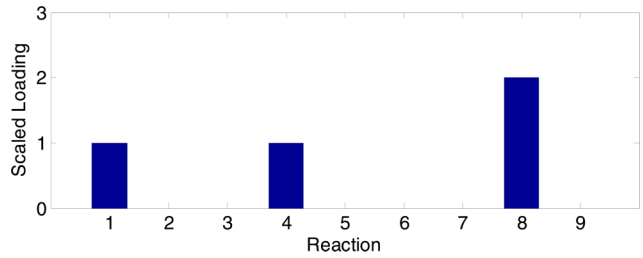


Figure 3. Scaled PCA loading for case study I.

with increased substrate pickup rate (which is scaled to be 1 as the basis) only re4 and re8 are affected with a scale of 1 and 4, which indicates that flux of re4 increases with the same amount as that of re1, while flux of re8 increase 4 times the amount of increase in flux of re1. It is worth noting that a negative loading in this case would indicate a decreased flux. Figure 4 visualizes the analysis result by highlighting the reactions that are affected by increasing flux of re1.

Case Study II: FBA Objective Function = Maximizing Flux of re9 (Production of E). In this case study, similar steps as in case study I were carried out, with the only difference in the objective function of FBA. In this case study, the objective function is to maximize the production of E. The PC loading and network visualization are plotted in Figures 5 and 6.

Both case studies show that even though the "hypothetical cell" has an alternative route to produce D and E, i.e., the one with intermediate metabolite C, it does not choose the alternative route because the route does not maximize the objective function. This is due to the difference in stoichiometric coefficients (A → 0.5C → 0.5D while A → 2D for the chosen route). If the alternative route were chosen, less product would be produced.

This illustrative example shows that the proposed method can systematically identify the reactions that would be affected by the introduced perturbation (e.g., increased substrate pickup rate in this case) without going through the detailed examination of the network stoichiometry. Such examination is nontrivial even for relatively small network models, such as central carbon metabolic networks, and quickly becomes infeasible when the size of the network increases. But with the proposed method, we can easily examine how a perturbation would affect the whole network and identify the key reactions that are affected the most by the perturbation.

dx.doi.org/10.1021/sc400265g | ACS Sustainable Chem. Eng. 2014, 2, 38−48

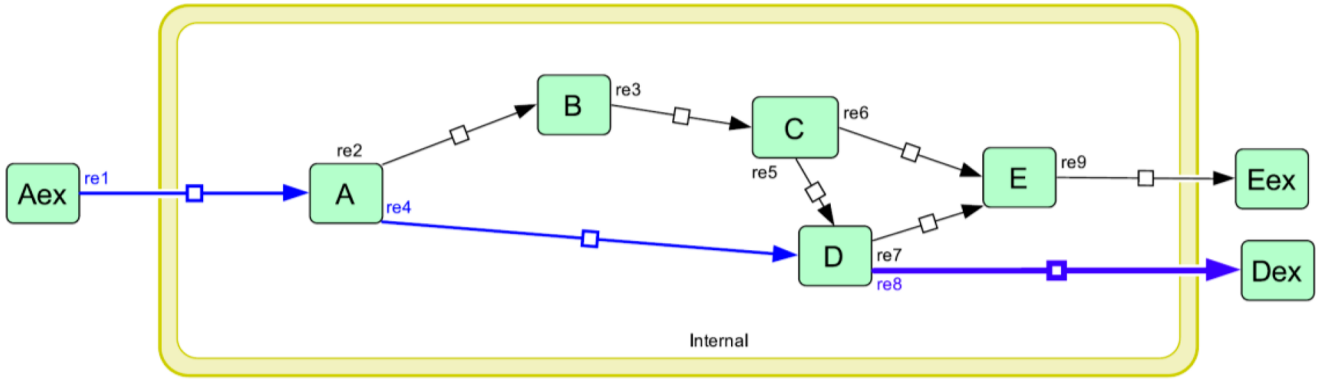


Figure 4. Visualization of the analysis results for case study I. The reactions that are affected by increasing flux of re1 are highlighted in blue. The line thickness is proportional to its loading.

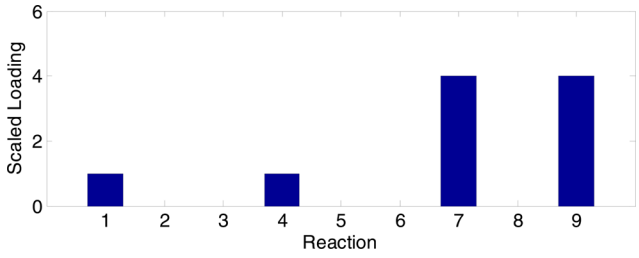


Figure 5. Scaled PCA loading for case study II.

VALIDATION AND ANALYSIS OF *S. stipitis* METABOLIC NETWORK

In this section, we apply the framework outlined and illustrated in the previous section to validate and analyze the reconstructed *S. stipitis* metabolic network model.

Validation of the Model. After the metabolic network model is reconstructed, we first validate the model by comparing the computed flux distribution with published experimental data. Specifically, five sets of published experimental results[26] are used to validate the proposed metabolic network model. Figures 7 −10 compare the model predicted flux distribution with batch experiments under four different growth regimes: glucose and xylose aerobic, and glucose and xylose microaerobic.[26] In these simulations, the carbon and oxygen uptake rates were constrained to be the same as the experimental values. Figures 7−10 show that the model predictions in general agree with the experimental results very well, with the only exceptions of biomass under glucose microaerobic and CO_2 under xylose microaerobic condition, which show bigger discrepancies. In addition, a set of reported chemostat experimental results[27] are used to validate the model, which is shown in Figure 11. In Grootjen et al.,[27] a series of chemostat growth experiments on glucose were conducted, and yields of biomass and ethanol with varying oxygen uptake rates were reported. Figure 11 compares the experimental and model predicted biomass and ethanol under oxygen-limited conditions. In these simulations, the glucose and oxygen uptake rates were constrained to be the same as the experimental

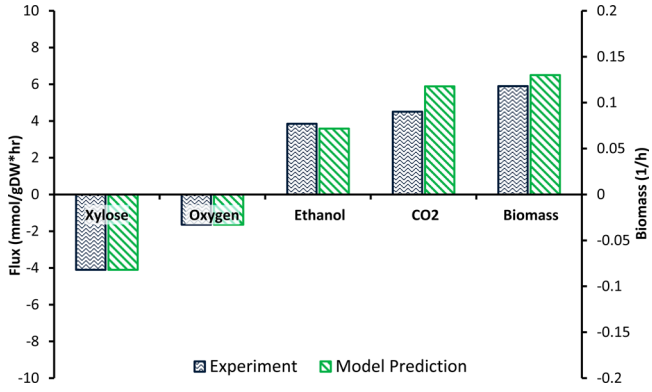


Figure 7. Comparison of model predicted flux distribution with batch experiments under xylose microaerobic condition.

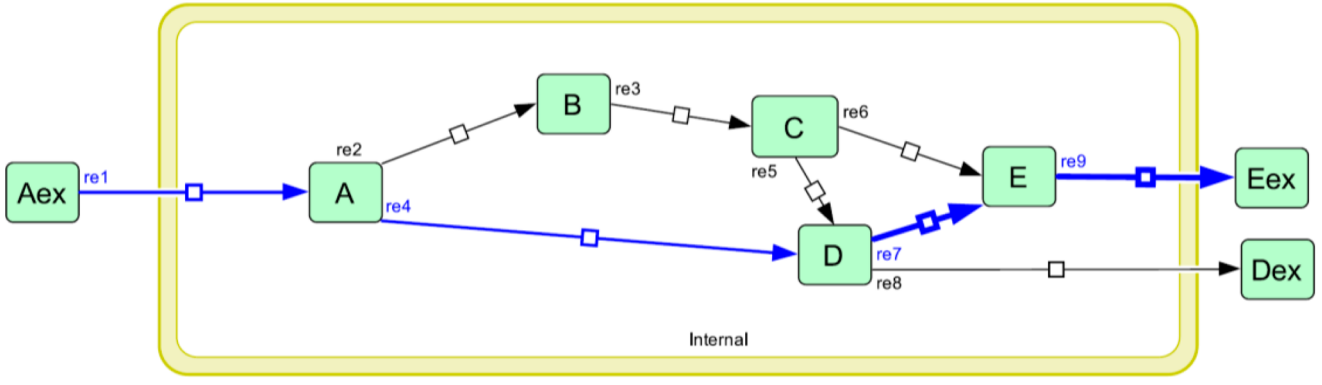


Figure 6. Visualization of the analysis results for case study II. The reactions that are affected by increasing flux of re1 are highlighted in blue. The line thickness is proportional to its loading.

dx.doi.org/10.1021/sc400265g | *ACS Sustainable Chem. Eng.* 2014, 2, 38−48

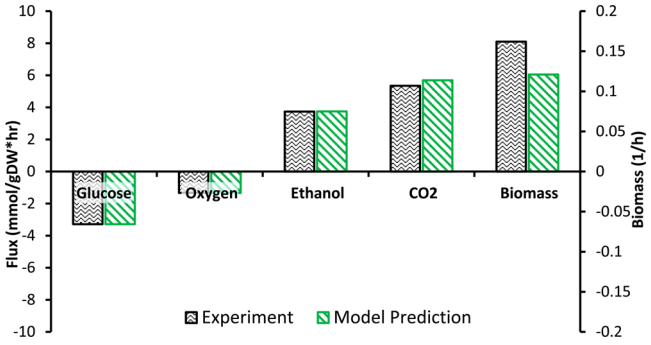


Figure 8. Comparison of model predicted flux distribution with batch experiments under glucose microaerobic condition.

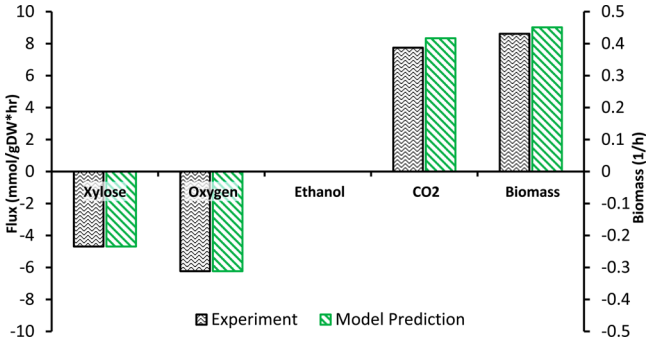


Figure 9. Comparison of model predicted flux distribution with batch experiments under xylose aerobic condition.

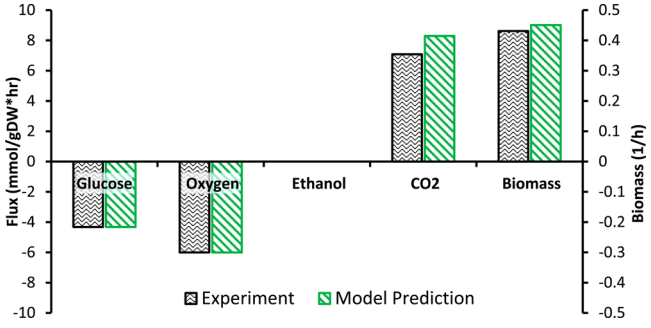


Figure 10. Comparison of model predicted flux distribution with batch experiments under glucose aerobic condition.

values. As shown in Figure 11, the model predictions agree well with the reported experimental data.

Analysis of the Model. With the model validated, the topological properties and intracellular flux distribution of the central carbon metabolic model were studied using FBA. In all simulations carried out in this work, the model is constrained for cells to grow on minimal defined medium.[28] The topological study shows that very few metabolites are highly connected while most metabolites participate only in a few reactions, which indicates that the metabolic network is scale-free. The average reaction participation, i.e., the number of metabolites per reaction, is approximately four, which indicates that the most common reaction mode in the model is four reactants and products combined. The topological analysis results are in agreement with the results from other metabolic network models, e.g., *E. coli*[11] and *S. cerevisiae*.[22] The reactions that are essential to cell growth are evaluated as whether its removal is fatal to the model. The results indicate that the

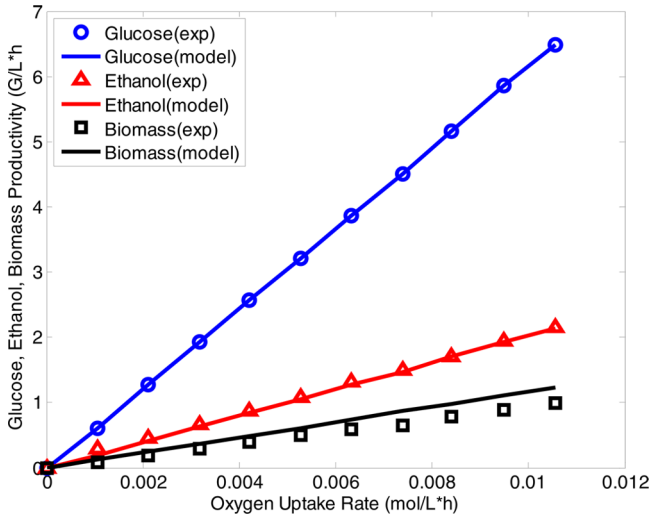


Figure 11. Model predicted flux distribution of *S. stipitis* during glucose metabolism compared with experimental chemostat data. Only substrate uptake rates are constrained, i.e., Glucose(exp) = Glucose-(model).

identified key reactions depend on the carbon source (glucose or xylose) and oxygen condition (aerobic or oxygen-limited). With aerobic glucose culture, it is shown that a total of 14 reactions are essential to cell growth. When switched to oxygen-limited condition (glucose), the number changed to 18. For xylose culture, the number is 16 and 26 for aerobic and oxygen-limited conditions, respectively, among which 10 reactions are essential under all conditions. They are reactions in glycolysis, pentose phosphate pathway, and urea metabolism. The larger difference of essential reaction number in xylose metabolism under different aeration conditions indicates the higher sensitivity of xylose metabolism to oxygen condition change compared with glucose metabolism, which agrees with experimental findings.[29]

■ ANALYSIS OF XYLOSE METABOLISM IN *S. stipitis*

It is well-known that oxygen plays an important role in cell growth, redox balance, functioning of the mitochondria and generation of energy for xylose transport in *S. stipitis*.[29] However, how oxygen influences the intracellular flux distribution and redox balance and which reactions would be the most important for redox balance are not well understood. In this section we design a series of in silico experiments to perturb the central carbon metabolic network of *S. stipitis* and apply PCA to analyze the in silico experimental results. The goal is to identify the key reactions or pathways that are affected by the introduced perturbation.

Phenotype Identification. *Designed in Silico Experiments.* In this case study, in order to study how different oxygen availabilities affect cellular metabolism, we performed FBA to determine the intracellular fluxes by varying the oxygen pickup rate (i.e., oxygen utilization rate) from 0 to 20 mmol/gDCW/h with a step size of 0.01. The xylose uptake rate has an upper limit of 10 mmol/gDCW/h. The specified xylose and oxygen uptake rates were chosen based on reported experimental results.[26] This set of experiments resulted in a 117 × 2001 matrix, where each column represents the 117 intracellular fluxes under a certain OUR. In the model, the flux distribution ratio through NADPH-dependent and NADH-dependent xylose reductase (XR) was set to be 1.0 (the setting

dx.doi.org/10.1021/sc400265g | *ACS Sustainable Chem. Eng.* 2014, 2, 38−48

and influence of such ratio is discussed later). Phenotype phase plane (PhPP) analysis[30,31] was also carried out under the same conditions for comparison.

Phenotypes Identified. PCA was applied to analyze the in silico experimental results. As shown in Figure 12(a) where scores corresponding to the first two PCs are plotted, a total of six phenotypes of metabolism are identified. One phenotype is distinguished from another phenotype when the correlation among fluxes has changed, which is shown on the PCA score plot as two different straight lines, each represents a distinctive correlation among fluxes. The distinction between phenotype 2 and phenotype 3 is not very clear in the figure because of the scale. The distinction is clear if the figure is zoomed in. The result from PhPP is given in Figure 12(b), where the same six phenotypes are identified. Figure 12(c) plots the model predicted cell growth rate and ethanol production rate under different aeration conditions, which reveals some difference among different phenotypes. The main characteristics of the different phenotypes are summarized in Table 3.

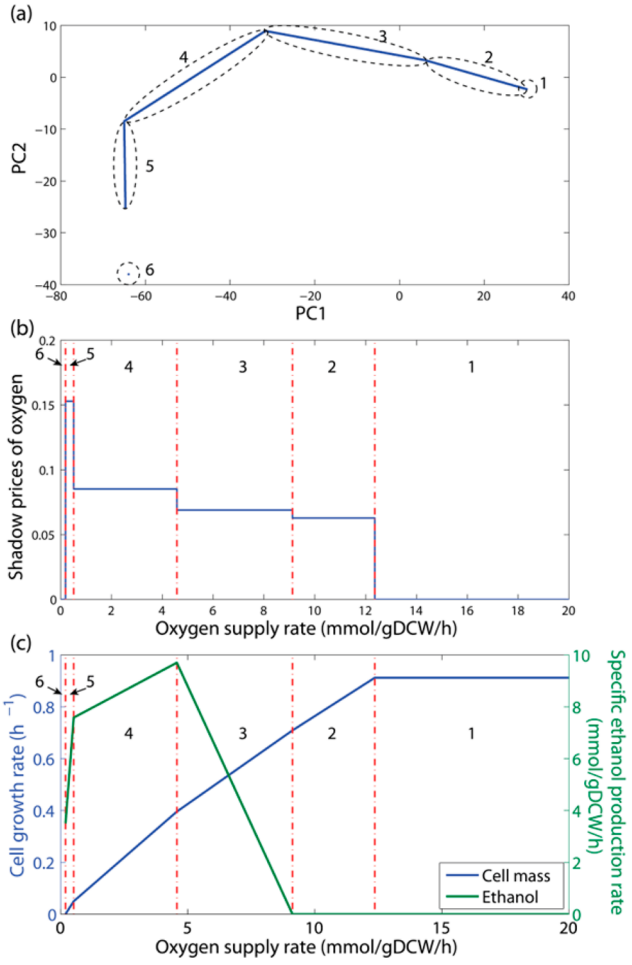


Figure 12. Phenotypes identified with PCA when OUR changes within [0, 20] mmol/gDCW/h. (a) phenotypes identified by PCA. (b) Phenotypes identified by PhPP. (c) Model predicted cell growth rates and specific ethanol production rates. The numbers 1−6 correspond to the identified phenotypes.

Although PhPP and the proposed approach identify the same six phenotypes, they are completely different in revealing the cellular details that underlie the specific phenotype. For PhPP, it can easily identify whether oxygen or carbon source is a limiting factor by examining the shadow price,[30] which determines the effect of each metabolite (b_i^v) on the objective function (Z_i) and can be expressed mathematically as $\gamma_i = -((dZ_i)/(db_i^v))$, but it is very difficult to identify what contribute to the change in the shadow price, as it only examines the objective function as a whole and does not provide the detail on how different reactions are affected by changing each metabolite. On the other hand, for the proposed approach, the limiting factor can be identified by checking whether the corresponding fluxes hit their upper limits. More importantly, one significant advantage of the proposed FBA−PCA based approach is that it can reveal the cellular details, particularly the key reactions that differentiate different phenotypes, by examining the PCA loading matrix.

Table 3. Summary of Characteristics of Identified Phenotypes

phenotype	growth limitation	metabolic product(s)	main metabolic characteristics
1	xylose	cell mass	aerobic growth
2	xylose, oxygen	cell mass, acetic acid	increasing acetic acid production
3	xylose, oxygen	cell mass, ethanol, acetic acid	ethanol production and declined acetic acid production
4	xylose, oxygen	cell mass, ethanol, xylitol	declined ethanol production and increasing xylitol production
5	oxygen	cell mass	declined ethanol and xylitol production
6	–	–	cannot maintain metabolism (no growth)

Here, we use an example to demonstrate the effectiveness of the proposed FBA−PCA method. The reactions that are affected the most by changing OUR in both phenotype 2 and 3 are plotted in Figure 13, where the metabolic fluxes are colored according to their loadings.

From Figure 13, several key differences can be observed. First, the importance of TCA cycle for cell growth in phenotype 3 has decreased compared to phenotype 2. Further examination shows that this is caused by the turning off of 2-oxoglutarate dehydrogenase due to decreased oxygen supply in phenotype 3, which further leads to an incomplete (or branched) TCA cycle as shown in Figure 14 (also shown as a gray TCA cycle in Figure 13(b) but as a green TCA cycle in Figure 13(a)). This branched TCA cycle has been previously reported in *S. cerevisiae*.[32,33] This prediction is further supported by Jeffries et al., where expressed sequence tags (EST) from oxygen-limited growth of *S. stipitis* on xylose showed that KGD2 (the TCA cycle reaction being bypassed) was downregulated. Second, fermentation pathway, i.e., ethanol production, has been activated by the branched TCA cycle to resolve the redox balance of NADH/NAD^+, which is indicated by gray in phenotype 2 and red in phenotype 3. Third, due to the decreased cell growth, the requirement of NADPH has been reduced and caused the down-regulation of fluxes through pentose phosphate pathway (PPP) as shown in Figure 13 by the color of PPP changing from light green in phenotype 2 to light yellow in phenotype 3.

Effect of OUR on Redox Balance in Phenotype 5. In this subsection, we apply the proposed FBA−PCA method to study the effect of OUR on redox balance in phenotype 5. Specially, we study the OUR range of [0.2, 0.5] mmol/gDCW/h, as Figure 12 shows that ethanol production is the most sensitive to OUR in this range. We first conducted a series of in

dx.doi.org/10.1021/sc400265g | ACS Sustainable Chem. Eng. 2014, 2, 38−48

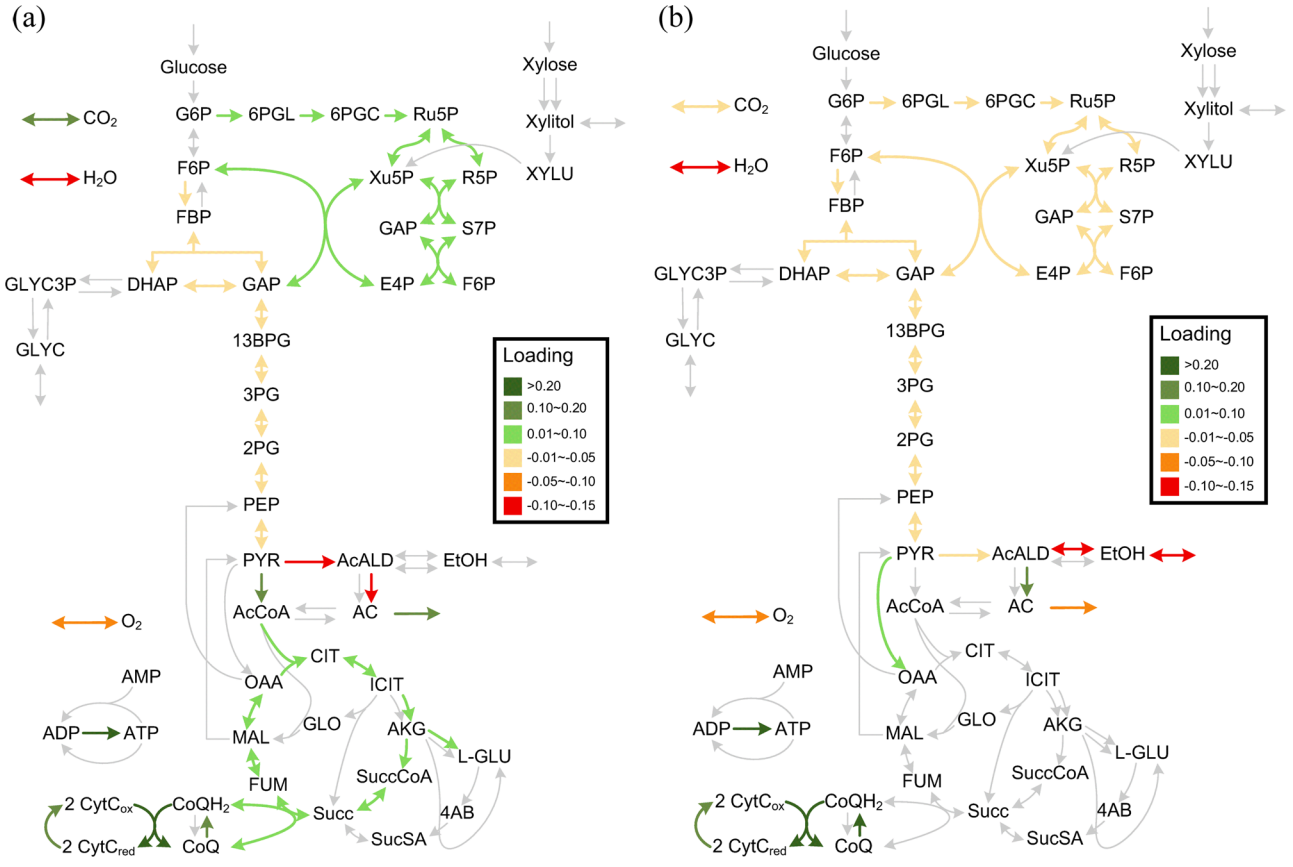


Figure 13. Metabolic maps for phenotype 2 (a) and phenotype 3 (b) identified in Figure 12.

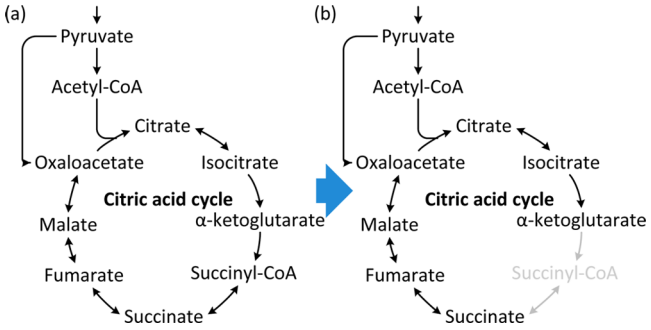


Figure 14. TCA cycle change occurred in phenotype 3: (a) complete TCA cycle, (b) branched TCA.

silico experiment where FBA was performed to compute the flux distribution by varying OUR from 0.2 to 0.5 mmol/gDCW/h, with step size 0.01. Then PCA was applied to analyze the resulted data matrix. Again, one PC captures 99.9% of all variance. All reactions that involve cofactor consumption and regeneration are listed in Table 4. The loadings corresponding to the involved reactions are plotted in Figure 15. The loadings are scaled by the rate of change in OUR. The fluxes of key reactions that are affected the most by the increase of OUR are tabulated in Table 5 for two conditions with OUR of 0.2 and 0.5 mmol/gDCW/h. The seven key reactions identified in Table 5 cover 99% of the total redox shift. The metabolic map with identified key reactions for phenotype 5 is shown in Figure 16. Both Figure 15 and Table 5 show that the proposed approach can reveal key information about metabolism shift and therefore help interpret the predictions from metabolic network model and provide insights into microorganism metabolism.

Effect of Xylose Reductase (XR) Cofactor Specificity on Redox Balance. It has been reported that *S. stipitis* can ferment xylose to ethanol with little xylitol production.[5] This is due to the dual affinity of XR to both NADH and NADPH, as illustrated in Figure 17. In the first step where xylose is reduced to xylitol, XR prefers NADPH over NADH.[34−38] Together with the difference in cellular cofactor concentrations,[39] more NADPH than NADH is consumed in the first step. In the second step where xylitol is converted to xylulose, XDH is reported to use NAD+ only. Therefore, there exists a cofactor imbalance when xylose is converted to xylulose. To reduce such cofactor imbalance, researchers have tried to apply protein engineering to alter the cofactor preferences of XR toward NADH in order to improve ethanol production and/or to reduce byproducts production.[40−45] In the previous subsection, it has been shown that a perturbation will propagate through the whole metabolic network and significantly affect multiple reactions involving cofactor consumption and regenerations. Therefore, systematically studying the effect of XR cofactor specificity on overall cellular cofactor balance would help understand the biological details of xylose fermentation in *S. stipitis* and engineered *S. cerevisiae* as well as provide rational design strategy for cofactor engineering.

We performed a series of in silico experiments with gradually increased XR flux ratio to study the effect of XR cofactor specificity on cellular redox balance. Here, XR flux ratio is defined as the ratio of the flux through the reaction that utilizes NADPH to the flux through the reaction that utilizes NADH when converting xylose to xylitol. On the basis of the reported

dx.doi.org/10.1021/sc400265g | ACS Sustainable Chem. Eng. 2014, 2, 38−48

Table 4. All Reactions That Involve Cofactor Consumption and Regeneration

R1	xyl[c] + nadh[c] + h[c] → xylt[c] + nad[c]
R2	dhap[c] + nadh[c] + h[c] → glyc3p[c] + nad[c]
R3	acald[c] + nadh[c] + h[c] ↔ etoh[c] + nad[c]
R4	akg[c] + gln[c] + h[c] + nadh[c] → 2 glu[c] + nad[c]
R5	akg[c] + nh4[c] + nadh[c] + h[c] → glu[c] + nad[c] + h2o[c]
R6	nadh[c] + q[c] + 5 h[c] → qh2[c] + nad[c] + 4 h[e]
R7	xylt[c] + nad[c] ↔ xylu[c] + nadh[c] + h[c]
R8	atp[c] + nad[c] → adp[c] + nadp[c] + h[c]
R9	mal[c] + nad[c] → pyr[c] + nadh[c] + co2[c]
R10	acald[c] + h2o[c] + nad[c] → ac[c] + nadh[c] + 2 h[c]
R11	gap[c] + nad[c] + pi[c] ↔ 13bpg[c] + nadh[c] + h[c]
R12	mal[c] + nad[c] ↔ oaa[c] + nadh[c] + h[c]
R13	akg[c] + coa[c] + nad[c] → succoa[c] + co2[c] + nadh[c]
R14	icit[c] + nad[c] → akg[c] + co2[c] + nadh[c]
R15	xyl[c] + nadph[c] + h[c] → xylt[c] + nadp[c]
R16	acald[c] + nadph[c] + h[c] ↔ etoh[c] + nadp[c]
R17	akg[c] + nh4[c] + nadph[c] + h[c] → glu[c] + nadp[c] + h2o[c]
R18	acald[c] + h2o[c] + nadp[c] → ac[c] + nadph[c] + 2 h[c]
R19	g6p[c] + nadp[c] → 6pgl[c] + nadph[c] + h[c]
R20	6pgc[c] + nadp[c] → nadph[c] + co2[c] + ru5p[c]
R21	sucsal[c] + nadp[c] + h2o[c] → succ[c] + nadph[c] + 2 h[c]
R22	icit[c] + nadp[c] → akg[c] + co2[c] + nadph[c]

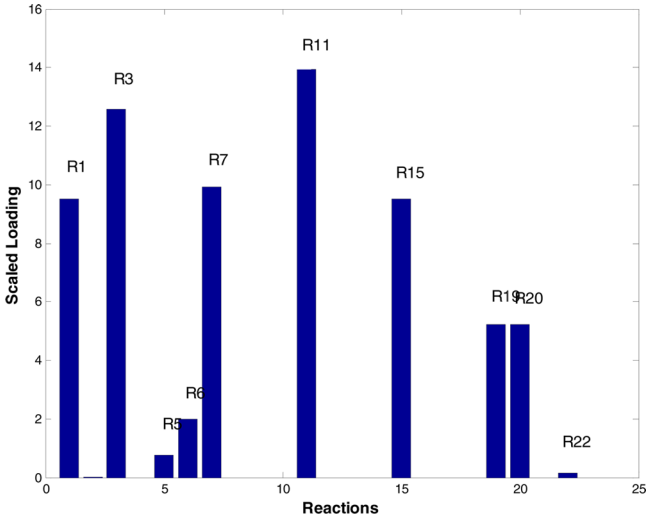


Figure 15. Loadings of the reactions involved in cofactor consumption and regeneration.

Table 5. Shift of Cofactor Consumption and Regeneration in Phenotype 5

cofactor	reaction	OUR = 0.200	OUR = 0.500	total shift
NADH consumption	R1	2.00	4.85	6.63
	R3	3.61	7.39	
NADH regeneration	R7	2.38	5.36	7.16
	R11	3.63	7.81	
NADPH consumption	R15	2.00	4.85	2.85
NADPH regeneration	R19	1.00	2.57	3.14
	R20	1.00	2.57	

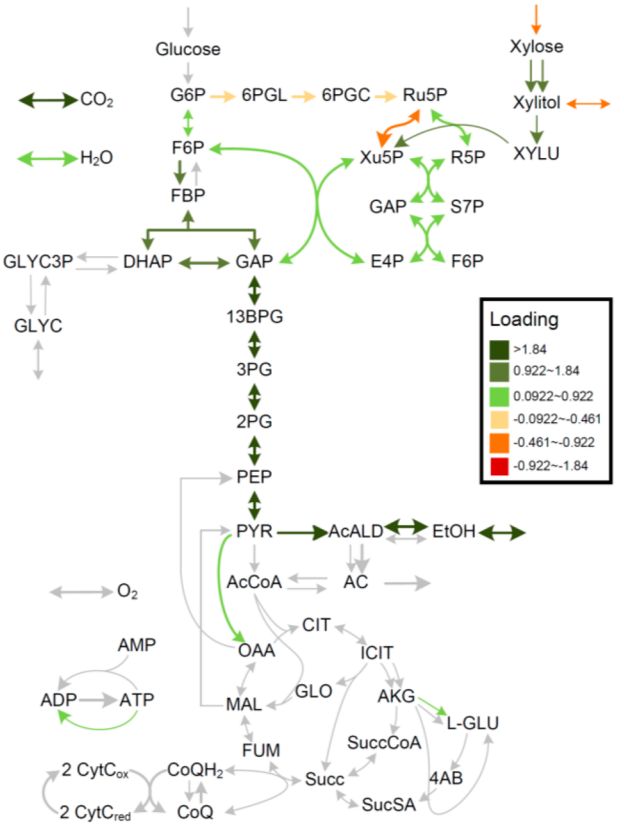


Figure 16. Metabolic map for phenotype 5 with key reactions (with dark green arrows) identified by FBA–PCA.

results and general knowledge of the XR affinity to different cofactors and the in vivo concentrations of NADH/NAD$^+$ and NADPH/NADP$^+$ pools,[46] as well as the reported results on cofactor affinity of XR mutants, in these experiments the XR flux ratio was varied within [0, 2] to study its effect on redox balance and ethanol production.

First, we performed simulations to study the general influence of XR activity ratio to model predictions under various oxygenation conditions through FBA. In these experiments, the xylose uptake rate was again constrained to be 10 mmol/gDCW/h, and OUR was varied between 0 and 14

dx.doi.org/10.1021/sc400265g | ACS Sustainable Chem. Eng. 2014, 2, 38−48

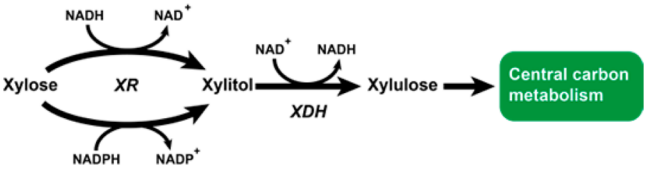


Figure 17. Illustration of xylose metabolism in *S. stipitis*.

mmol/gDCW/h with a step size of 0.1. The XR flux ratio compared in these experiments varied between [0, 2] with a step-size of 0.2, with an additional XR flux ratio of 10 as an extreme case. The resulted cell growth, ethanol production, and xylitol production are shown in Figure 18. It shows that the increase of NADH affinity of XR can improve the ethanol production and reduce xylitol production.[47]

In order to elucidate the cellular details that underlie the predicted cell growth and ethanol production, we carried out a second set of in silico experiments, where we fixed both xylose and oxygen uptake rates to 10 mmol/gDCW/h and 0.4 mmol/

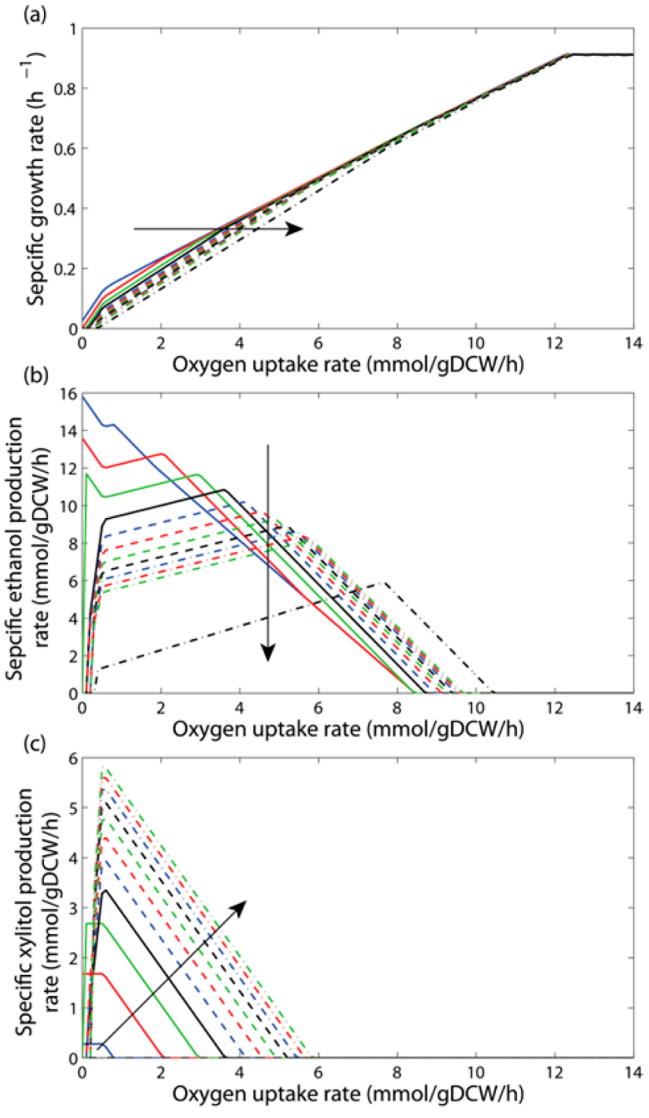


Figure 18. Influences of XR flux ratio on specific cell growth rate (a), specific ethanol production rate (b), and specific xylitol production rate (c). The arrows in the plots indicate the increase of XR flux ratio.

gDCW/h respectively. Within this set of in silico experiments, the XR flux ratio was again changed between [0, 2] with a step size of 0.01. PCA was applied to analyze the resulted flux distribution matrix to identify the key changes among different reactions when the XR flux ratio is changed. The loadings corresponding to the reactions involving cofactor consumption and regeneration are plotted in Figure 19. The loadings are

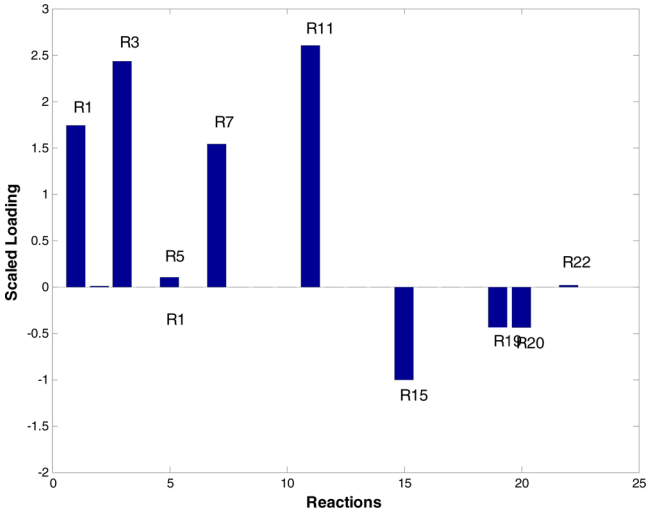


Figure 19. Loadings of the reactions involved in cofactor consumption and regeneration.

scaled by the rate of change in XR flux ratio. The fluxes of key reactions that are affected most by increase of XR flux ratio are tabulated in Table 6 for two conditions with XR flux ratio of 0.5

Table 6. Shift of Cofactor Consumption and Regeneration

cofactor	reaction	XR = 0.5	XR = 2.0	total shift
NADH consumption	R1	6.67	2.64	−9.66
	R3	10.1	4.47	
NADH regeneration	R7	6.92	3.32	−9.63
	R11	10.6	4.57	
NADPH consumption	R15	3.33	5.28	1.95
NADPH regeneration	R19	1.83	2.68	1.7
	R20	1.83	2.68	

and 2.0. The seven key reactions identified in Table 6 cover 98% of the total redox shift. The metabolic map with identified key reactions is shown in Figure 20.

By comparing Figures 16 and 20 as well as Tables 5 and 6, it is interesting to notice that the same seven reactions that involve cofactor consumption and regeneration are affected the most by changing OUR or changing XR flux ratio. However, how they are affected by different perturbations are different. It could be because the in silico experiments were conducted under the same phenotype, i.e., phenotype 5. We are currently running further experiments to test whether different reactions would be affected the most for different phenotypes.

CONCLUSIONS

In this work, the central carbon metabolic network of *S. stipitis* is reconstructed. The model is validated against experimental results reported in the literature. Even though the reconstructed metabolic network model does not capture all the cellular details, for example, it only considers one compartment (the

dx.doi.org/10.1021/sc400265g | ACS Sustainable Chem. Eng. 2014, 2, 38−48

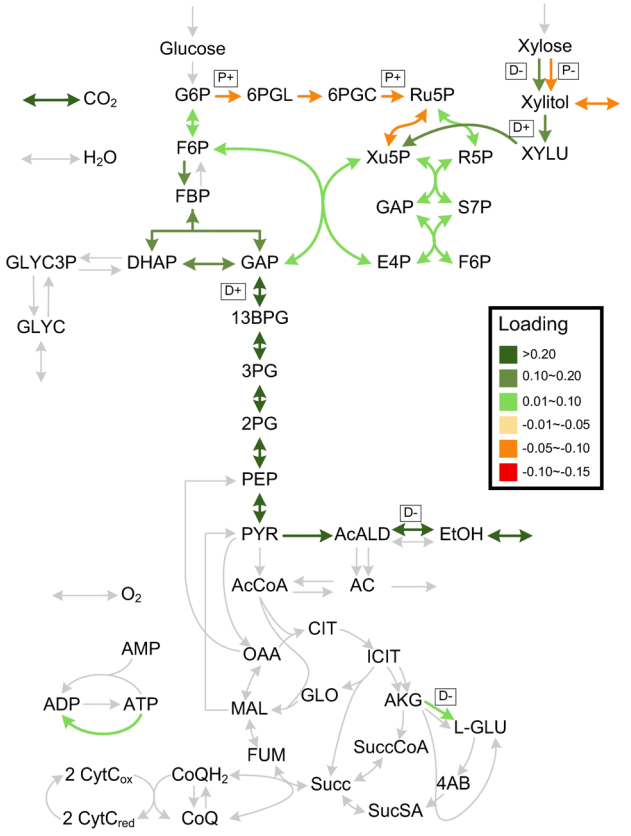


Figure 20. Metabolic map with key reactions (with dark green arrows) identified by FBA–PCA.

cytosol) and does not contain any gene regulatory mechanism, it still provides a comprehensive picture of the central carbon metabolism of *S. stipitis*. Such a model enables us to elucidate the xylose metabolism using a systems approach.

To investigate how the cellular redox balance is affected by change in the OUR and XR cofactor specificity, we developed a system identification-based metabolic flux analysis framework to extract the underlying biological knowledge embedded in the network structure. By applying the proposed framework, we were able to identify the key reactions that dominate the cellular redox balance. It is interesting to find out that under an oxygen-limited condition it is the same set of the key reactions that dominate the redox shift caused by a change in OUR or a change in the XR cofactor specificity, although they are affected in different ways by different factors. The in silico experiments and PCA analysis results show that xylose reductase plays a key role in xylose fermentation to ethanol. In particular, its cofactor specificity, if adjusted toward favoring NADH, could improve ethanol yield. Finally, the set of key reactions (total of seven reactions) should be considered together when designing a mutant to improve ethanol yield through shifting cellular redox balance.

■ ASSOCIATED CONTENT

Supporting Information

All reactions and metabolites of the central carbon metabolic network model of *S. stipitis* constructed in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: wang@auburn.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The financial support from USDA-AFRI 2010-65504-20358, Sun Grant 8500014160, NSF 1248388, 1264861, and NSF IGERT programs are gratefully appreciated.

■ ABBREVIATIONS

FBA, flux balance analysis; PCA, principal component analysis; XR, xylose reductase; DCW, dry cell weight; TCA, tricarboxylic acid; COBRA, constraints-based reconstruction and analysis; EMA, elementary mode analysis; PhPP, phenotype phase plane analysis; OUR, oxygen utilization rate; PPP, pentose phosphate pathway

■ REFERENCES

(1) Naik, S. N.; Goud, V. V.; Rout, P. K.; Dalai, A. K. Production of first and second generation biofuels: A comprehensive review. *Renewable Sustainable Energy Rev.* **2010**, *14* (2), 578–597.

(2) Eisentraut, A. Sustainable Production of Second-Generation Biofuels: Potential and perspectives in major economies and developing countries; Information Paper; International Enery Agency (IEA): Paris, 2010.

(3) Margeot, A.; Hahn-Hagerdal, B.; Edlund, M.; Slade, R.; Monot, F. New improvements for lignocellulosic ethanol. *Curr. Opin. Biotechnol.* **2009**, *20* (3), 372–380.

(4) Kurtzman, C. P.; Suzuki, M. Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera Babjeviella, Meyerozyma, Millerozyma, Priceomyces, and Scheffersomyces. *Mycoscience* **2010**, *51* (1), 2–14.

(5) Jeffries, T. W.; Van Vleet, J. R. H. *Pichia stipitis* genomics, transcriptomics, and gene clusters. *FEMS Yeast Res.* **2009**, *9* (6), 793–807.

(6) Jeffries, T. W.; Jin, Y. S. Ethanol and thermotolerance in the bioconversion of xylose by yeasts. *Adv. Appl. Microbiol.* **2000**, *47*, 221–268.

(7) Slininger, P.; Branstrator, L.; Bothast, R.; Okos, M.; Ladisch, M. Growth, death, and oxygen uptake kinetics of *Pichia stipitis* on xylose. *Biotechnol. Bioeng.* **1991**, *37*, 973–980.

(8) Slininger, P.; Bothast, R.; Ladisch, M.; Okos, M. Optimum pH and temperature conditions for xylose fermentation by Pichia stipitis. *Biotechnol. Bioeng.* **1990**, *35*, 727–731.

(9) Slininger, P.; Branstrator, J.; Lomont, J.; Dien, B.; Okos, M.; Ladisch, M.; Bothast, R. Stoichiometry and kinetics of xylose fermentation by *Pichia stipitisa*. *Ann. N.Y. Acad. Sci.* **1990**, *589*, 25–40.

(10) Slininger, P.; Bothast, R.; Okos, M.; Ladisch, M. Comparative evaluation of ethanol production by xylose-fermenting yeasts presented high xylose concentrations. *Biotechnol. Lett.* **1985**, *7*, 431–436.

(11) Orth, J. D.; Thiele, I.; Palsson, B. O. What is flux balance analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248.

(12) Thiele, I.; Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **2010**, *5* (1), 93–121.

(13) Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2011**, DOI: 10.1093/nar/gkr988.

(14) Degtyarenko, K.; Hastings, J.; de Matos, P.; Ennis, M. *ChEBI: An Open Bioinformatics and Cheminformatics Resource* **2009**, DOI: 10.1002/0471250953.bi1409s26.

(15) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: Integrated Platform of Small Molecules and Biological

Activities. In *Annual Reports in Computational Chemistry*; American Chemical Society: Washington, DC, 2008; Vol. *4*.
(16) Balagurunathan, B.; Jonnalagadda, S.; Tan, L.; Srinivasan, R. Reconstruction and analysis of a genome-scale metabolic model for *Scheffersomyces stipitis*. *Microb. Cell Fact.* **2012**, *11*, 27.
(17) Caspeta, L.; Shoaie, S.; Agren, R.; Nookaew, I.; Nielsen, J. Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in-silico evaluation of their potentials. *BMC Syst. Biol.* **2012**, *6* (1), 24−24.
(18) Guebel, D. V.; Cordenons, A.; Nudel, B. C.; Giulietti, A. M. Influence of oxygen transfer rate and media composition on fermentation of d-xylose by *Pichia stipitis* NRRL Y-7124. *J. Ind. Microbiol.* **1991**, *7* (4), 287−291.
(19) Rizzi, M.; Klein, C.; Schulze, N. A.; Bui-Thanh, A.; Dellweg, H. Mathematical model for the semiaerobic fermentation of xylose by the yeast *Pichia stipitis*. In *Proceedings of 4th European Congress of Biotechnology*, **1987**.
(20) Senger, R. S. Biofuel production improvement with genome-scale models: The role of cell composition. *Biotechnol. J.* **2010**, *5* (7), 671−685.
(21) Vanrolleghem, P. A.; de Jong-Gubbels, P.; van Gulik, W. M.; Pronk, J. T.; van Dijken, J. P.; Heijnen, S. Validation of a metabolic network for *Saccharomyces cerevisiae* using mixed substrate studies. *Biotechnol. Prog.* **1996**, *12* (4), 434−448.
(22) Duarte, N. C.; Herrgard, M. J.; Palsson, B. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **2004**, *14* (7), 1298−1309.
(23) Schellenberger, J.; Que, R.; Fleming, R. M. T.; 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. Ø. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6* (9), 1290−1307.
(24) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer, 2002; pp 487−487.
(25) Griffin, J. L. Metabolic profiles to define the genome: Can we hear the phenotypes? *Philos. Trans. R. Soc., B* **2004**, *359* (1446), 857−871.
(26) Li, P. Y. In Silico Metabolic Network Reconstruction of *Scheffersomyces stipitis*. Ph.D. Thesis, University of Toronto, 2012.
(27) Grootjen, D.; Van der Lans, R.; Luyben, K. Effects of the aeration rate on the fermentation of glucose and xylose by *Pichia stipitis* CBS 5773. *Enzyme Microb. Technol.* **1990**, *12*, 20−23.
(28) Jeffries, T. W.; Grigoriev, I. V.; Grimwood, J.; Laplaza, J. M.; Aerts, A.; Salamov, A.; Schmutz, J.; Lindquist, E.; Dehal, P.; Shapiro, H. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol.* **2007**, *25*, 319−326.
(29) Skoog, K.; Hahn-Hagerdal, B. Effect of oxygenation on xylose fermentation by *Pichia stipitis*. *Appl. Environ. Microbiol.* **1990**, *56* (11), 3389−3394.
(30) Edwards, J. S.; Ramakrishna, R.; Palsson, B. O. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* **2002**, *77* (1), 27−36.
(31) Bell, S.; Palsson, B. Phenotype phase plane analysis using interior point methods. *Comput. Chem. Eng.* **2005**, *29* (3), 481−486.
(32) Vargas, F. a.; Pizarro, F.; Perez-Correa, J. R.; Agosin, E. Expanding a dynamic flux balance model of yeast fermentation to genome-scale. *BMC Syst. Biol.* **2011**, *5* (1), 75−75.
(33) Nissen, T. L.; Schulze, U.; Nielsen, J.; Villadsen, J. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* **1997**, *143* (1), 203−218.
(34) Hou, X. Anaerobic xylose fermentation by *Spathaspora passalidarum*. *Appl. Microbiol. Biotechnol.* **2012**, *94* (1), 205−214.
(35) Verduyn, C.; Van Kleef, R.; Frank, J.; Schreuder, H.; Van Dijken, J.; Scheffers, W. Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast *Pichia stipitis*. *Biochem. J.* **1985**, *226* (3), 669−677.
(36) Yablochkova, E. N.; Bolotnikova, O. I.; Mikhailova, N. P.; Nemova, N. N.; Ginak, A. I. The activity of xylose reductase and xylitol dehydrogenase in yeasts. *Microbiology* **2003**, *72* (4), 414−417.
(37) Yablochkova, E. N.; Bolotnikova, O. I.; Mikhailova, N. P.; Nemova, N. N.; Ginak, A. I. The activity of key enzymes in xylose-assimilating yeasts at different rates of oxygen transfer to the fermentation medium. *Microbiology* **2004**, *73* (2), 129−133.
(38) Slininger, P. J.; Thompson, S. R.; Weber, S.; Liu, Z. L.; Moon, J. Repression of xylose-specific enzymes by ethanol in *Scheffersomyces (Pichia) stipitis* and utility of repitching xylose-grown populations to eliminate diauxic lag. *Biotechnol. Bioeng.* **2011**, *108* (8), 1801−1815.
(39) Watanabe, S.; Saleh, A. A.; Pack, S. P.; Annaluru, N.; Kodaki, T.; Makino, K. Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein-engineered NADH-preferring xylose reductase from Pichia stipitis. *Microbiology* **2007**, *153*, 3044−3054.
(40) Matsushika, A.; Inoue, H.; Watanabe, S.; Kodaki, T.; Makino, K.; Sawayama, S. Efficient bioethanol production by a recombinant flocculent *Saccharomyces cerevisiae* strain with a genome-integrated NADP+-dependent xylitol dehydrogenase gene. *Appl. Environ. Microbiol.* **2009**, *75* (11), 3818−3822.
(41) Chu, B. C. H.; Lee, H. Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation. *Biotechnol. Adv.* **2007**, *25* (5), 425−441.
(42) Watanabe, S.; Kodaki, T.; Makino, K. Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *J. Biol. Chem.* **2005**, *280* (11), 10340−10349.
(43) Krahulec, S.; Klimacek, M.; Nidetzky, B. Analysis and prediction of the physiological effects of altered coenzyme specificity in xylose reductase and xylitol dehydrogenase during xylose fermentation by *Saccharomyces cerevisiae*. *J. Biotechnol.* **2012**, *158* (4), 192−202.
(44) Bengtsson, O.; Hahn-Hagerdal, B.; Gorwa-Grauslund, M. F. Xylose reductase from *Pichia stipitis* with altered coenzyme preference improves ethanolic xylose fermentation by recombinant *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* **2009**, *2* (1), 9−9.
(45) Liang, L.; Zhang, J.; Lin, Z. Altering coenzyme specificity of *Pichia stipitis* xylose reductase by the semi-rational approach CASTing. *Microb. Cell Fact.* **2007**, *6*, 36.
(46) Bergdahl, B.; Heer, D.; Sauer, U.; Hahn-Hagerdal, B.; van Niel, E. W. Dynamic metabolomics differentiates between carbon and energy starvation in recombinant *Saccharomyces cerevisiae* fermenting xylose. *Biotechnol. Biofuels* **2012**, *5* (1), 34−34.
(47) Jeppsson, M.; Bengtsson, O.; Franke, K.; Lee, H.; Hahn-Hägerdal, B.; Gorwa-Grauslund, M. F. The expression of a *Pichia stipitis* xylose reductase mutant with higher KM for NADPH increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **2006**, *93*, 665−673.