Experimental and *In Silico* Fermentation of Glucose and Xylose with *Scheffersomyces stipitis*

by

Meng Liang

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 3, 2014

Keywords: *Scheffersomyces stipitis*, flux balance analysis, redox balance, constraint-based metabolic network model, bioethanol

Approved by

Jin Wang, Chair, B. Redd Associate Professor of Chemical Engineering
Qinghua Peter He, Associate Professor of Chemical Engineering, Tuskegee University
Yoon Y. Lee, Uthlaut Family Professor of Chemical Engineering
Mario R. Eden, Joe T. & Billie Carole McMillan Professor of Chemical Engineering

Abstract

Fossil fuel reserves are running out, global warming is becoming a reality, waste recycling is becoming ever more costly and problematic, and unrelenting population growth will require more and more energy and consumer products. There is now an alternative to the 100% oil economy; it is a renewable resource based on biomass. Production and development of these new products are based on biorefinery concept. The substitution of oil products by bio-based products will develop a new bio-economy and industrial processes respecting the sustainable development concept. The carbohydrate fraction of biomass feedstock (i.e. cellulose and hemicellulose in lignocellulosic biomass) is expected to play the biggest role as a renewable carbon source for biochemical products.

Scheffersomyces stipitis, a novel yeast for lignocellulosic bioconversion, accepts various substrates and shows good overall performance in hydrolysate. As one of the best xylose-fermenting yeast, it has worked long as the gene provider and now it has the potential to be host for further genetic modification. With the genome sequenced, it is very necessary now to study *S. stipitis* in a systematic way.

In this study, the fermentation of glucose and xylose with *S. stipitis* has been studied both experimentally and computationally. First, the fermentation of glucose and xylose were studied via experiment. To solve the washout caused by low growth rate with limited oxygen supply, a "pseudo-continuous" fermentation was used. The system proved its efficiency and also provided a better approach for improving ethanol tolerance, which was evaluated by the significant improvements of five different definitions on ethanol tolerance. Following the experimental results, a constraint-based core carbon metabolic network model has been constructed based on literatures, databases, and genome data. Flux balance analysis (FBA)

was used to investigate the properties of the model under various conditions. To evaluate the performance of the constructed model, bioethanol production was chosen as the study system. The model was verified qualitatively and quantitatively with experimental observations and reported literature data. Different phenotypes in glucose or xylose metabolism with *S. stipitis* have been identified via phenotype analysis and thus studied via flux distribution. To further extract the underlying biological knowledge under the phenotype shifts, we proposed a new system identification based framework, FBA-PCA, and showed its power on analyzing metabolic network model through the identification of key reactions when oxygen supply rate or the ratio through NADPH- and NADH-linked reactions catalyzed by xylose reductase changes. The methodologies proposed in this dissertation can be applied to other biological system and therefore can broaden the application of the metabolic network models.

Acknowledgments

First of all, I would like to thank Dr. Jin Wang for her guidance and support in this research and the preparation of this dissertation. I also acknowledge Dr. Qinghua He for his valuable discussions. I would like to express my gratitude towards my committee members, Dr. Yoon Y. Lee, and Dr. Mario R. Eden for their time and assistance in helping me complete this work.

I would like to thank my group members, Hector Galicia, Min Hea Kim, Zi Xiu Wang, Andrew Damiani and Kyle Stone for useful discussions and enjoyable moments. Thanks are also due to Li Kang, Suan Shi, and other fellow graduate students who have made my time spent at Auburn University both educational and enjoyable.

Special thanks to my parents, Qingtian Liang and Yunxia Si, and my elder brother Lin Liang, for their lasting love and support throughout my life. I would like to thank my wife Xiaowei Zhou, whose unconditional love and constant motivation have made this research possible. Especially, I want to thank my little lovely daughters, Liying and Katherine. They have brought me joy of being a father and taken light to my life.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Sustainable Development from Biomass

The world's primary source of energy for the transport sector and production of chemicals is oil. World demand is approximately 84 million barrels a day in 2009 and is projected to increase to about 99 million barrels a day by 2035, with transport accounting for some 60% of such a rising demand (Eisentraut, 2010). Concerning chemicals, their dependence on fossil resources is even stronger. The majority of chemical products are produced from oil refinery and almost 4% of oil is worldwide used for chemical and plastic production. Besides, the strong dependence on fossil fuels comes from the intensive use and consumption of petroleum derivatives which, combined with diminishing petroleum resources, causes environmental and political concerns (Cherubini, 2010).

In order to simultaneously reduce the dependence on oil and mitigate climate change in transport and chemical sectors, alternative production chains are necessary. It is increasingly recognized that there is not a single solution to these problems and that combined actions are needed, including changes in behavior, changes in vehicle technologies, expansion of public transport and introduction of innovative fuels and technologies (Cherubini, 2010). Recently, society began to recognize the opportunities offered by a future sustainable economy based on renewable sources and has been starting to finance R&D activities for its implementation. In 2009, the cost of consumption subsidies to fossil fuels in the world is $312 billion, while the funding support given to renewable energy in 2009 is $57 billion (Eisentraut, 2010). It is increasingly acknowledged globally that plant-based raw materials (i.e. biomass) have the

potential to replace a large fraction of fossil resources as feedstocks for industrial productions, addressing both the energy and non-energy (i.e. chemicals and materials) sectors. At national, regional and global levels there are three main drivers for using biomass in biorefinery for production of bioenergy, biofuels and biochemicals. These are climate change, energy security and rural development. The political motivation to support renewable sources of energy and chemicals arises from each individual driver or combinations. Policies designed to target one driver can be detrimental to another. For example, policies aimed at ensuring energy security may result in increased GHG emissions where local coal reserves are preferentially exploited at the expense of imported oil or gas. In addition, electricity and heat can be provided by a variety of renewable alternatives (wind, sun, water, biomass and so on), while biomass is very likely to be the only viable alternative to fossil resources for production of transportation fuels and chemicals, since it is the only C-rich material source available on the Earth, besides fossils. As a consequence, the sustainable biomass production is a crucial issue, especially concerning a possible fertile land competition with food and feed industries (Langeveld, Sanders, and Meeusen, 2010).

In the following parts, the possibilities to use biomass feedstocks as raw materials in biorefinery are reviewed briefly. First, the current status in biofuel production is provided (Naik et al., 2010) and then the emerging biorefinery concept is described. The latter is done through an overview of the most promising biomass feedstocks, technological processes and final products. The current oil refinery industry is taken as benchmark.

1.1.1 State of the art in biofuel production

Currently, transportation fuels based on biomass (i.e. biofuels) are identified as 1st and 2nd generation biofuels. First generation biofuels usually refer to the biofuels produced from raw materials in competition with food and feed industries. Because of this competition, these biofuels give rise to ethical, political and environmental concerns. In order to overcome these issues, production of second generation biofuels (i.e. from raw materials based on waste,

residues or non-food crop biomass) gained an increasing worldwide interest in the last decade as a possible greener alternative to fossil fuels and conventional biofuels.

1.1.1.1 First generation biofuels

First generation biofuels are produced from sugar, starch, vegetable oil or animal fats using conventional technologies. The basic feedstocks are often seeds and grains such as wheat, corn and rapeseed. The most common first generation biofuels are bioethanol, biodiesel and starch-derived biogas (Deublein and Steinhauser, 2008), but also straight vegetable oils, biomethanol and bioethers may be included in this category (Cherubini, 2010). Among the varieties of first-generation biofuels, bioethanol has received extended attention and will be discussed separately.

The main advantages of first generation biofuels are due to the high sugar or oil content of the raw materials and their easy conversion into biofuel. Many biofuel production chains have been analyzed by means of Life Cycle Assessment (LCA) in order to point out their environmental performances. With the exception of a few studies, most LCAs have found a net reduction in global warming emissions and fossil energy consumption when the most common transportation biofuels (bioethanol and biodiesel) are used to replace conventional diesel and gasoline.

In addition, 1st generation biofuels are in competition with food and feed industries for the use of biomass and agricultural land, giving rise to ethical implications: as prices for fossil fuels increase, a larger proportion of cereals or agricultural land will be dedicated to biofuel production instead of using it to produce food. In conclusion, first generation biofuels currently produced from sugars, starches and vegetable oils cause several concerns: these productions compete with food for their feedstock and fertile land, their potential availability is limited by soil fertility and per hectare yields and the effective savings of CO_2 emissions and fossil energy consumption are limited by the high energy input required for

crop cultivation and conversion (Lange, 2007; Marris, 2006). These limitations are expected to be partially overcome by developing the so-called 2nd generation biofuels.

1.1.1.2 Second generation biofuels

Second generation biofuels are produced from a variety of nonfood crops. These include the utilization of lignocellulosic materials, such as residues from agriculture, forestry and industry and dedicated lignocellulosic crops. In the scientific literature, the term "2nd generation" shows wide variation in usage and can variably refer to feedstocks (e.g. lignocellulosic material), conversion routes (e.g. thermochemical, flash pyrolysis, enzymatic, etc.) and end products (e.g. gas or synthetic liquid biofuels) (Clark, 2007; Naik et al., 2010; Osmont et al., 2010).

Contrarily to first generation biofuels, where the utilized fraction (grains and seeds), represents only a small portion of the above-ground biomass, second generation biofuels can rely on the whole plant for bioenergy production and thus eliminate the disadvantages of first generation biofuels mentioned above.

Thanks to technology development, environmental performances of 2nd generation biofuels could benefit of the use of high quantities of lignocellulosic residues and waste which are already available: they can constitute the main raw material sources, which can be also supplemented with non-food crops such as perennial grasses, and short-rotation forestry. Most processes and technologies for 2nd generation biofuels from biomass residues are still at a pre-commercial stage, but could enter the market within 10–15 years if corresponding investments (R&D, infrastructure) are achieved (Cherubini, 2010; Naik et al., 2010; Osmont et al., 2010).

On the one side the raw material situation is optimum (widespread, relatively cheap and easily available); on the other side, their use could allow the co-production of valuable

biofuels, chemical compounds as well as electricity and heat, leading to better energy, environmental and economic performances through the development of biorefinery concepts (Kamm et al., 2006).

1.1.2 From oil refinery to biorefinery

The structure of biorefinery raw materials is totally different from that on which the current oil refinery is based. In fact, the crude oil is a mixture of many different organic hydrocarbon compounds. The first step of oil refinery is to remove water and impurities, then distill the crude oil into its various fractions as gasoline, diesel fuel, kerosene, lubricating oils and asphalts. Then, these fractions can be chemically changed further into various industrial chemicals and final products.

Unlike petroleum, biomass composition is not homogeneous, because the biomass feedstock might be made of grains, wood, grass, biological waste and so on, and the elemental composition is a mixture of C, H and O (plus other minor components such as N, S and other mineral compounds). Chemical and elemental compositions of some lignocellulosic biomass feedstocks have been reported (EERE, 2006). If compared to petroleum, biomass generally has too little hydrogen, too much oxygen, and a lower fraction of carbon. The compositional variety in biomass feedstocks is both an advantage and a disadvantage. An advantage is that biorefineries can make more classes of products that can petroleum refineries and can rely on a wider range of raw materials. A disadvantage is that a relatively larger range of processing technologies is needed, and most of them are still at a pre-commercial stage (Dale and Kim, 2006). The detailed composition of biomass will be discussed in next section. In order to be used for production of biofuels and chemicals, biomass needs to be depolymerized and deoxygenated. Deoxygenation is required because the presence of O in biofuels reduces the heat content of molecules and usually gives them high polarity, which hinders blending with

existing fossil fuels (Lange, 2007). Chemical applications may require much less deoxygenation, since the presence of O often provides valuable physical and chemical properties to the product.

Unlike petroleum, biomass experiences seasonal changes, since harvesting is not possible throughout the entire year. A switch from crude oil to biomass may require a change in the capacity of chemical industries, with a requirement to generate the materials and chemicals in a seasonal time-frame. Alternatively, biomass may have to be stabilized prior to long-term storage in order to ensure continuous, year-round, operation of the biorefinery (Clark, Deswarte, and Farmer, 2009).

Biorefinery represents a change from the traditional oil refinery based on large exploitation of natural resources and large waste production towards integrated systems in which all resources are used (Kamm et al., 2006). Various processing technologies are used in biorefinery processes, such as biological, chemical, thermal, thermo-chemical and physical processes, etc. (Kamm and Kamm, 2004).

Today's chemical industry processes crude oil into a limited number of base fractions. Using numerous cracking and refining catalysts and using distillation as the dominant separation process, crude oil is refined into fractions such as naphtha, gasoline, kerosene, gas oil and residues. A biorefinery industry aiming at producing bulk chemicals from biomass will be based on a different selection of simple platforms than those currently used in the petrochemical industry. Given the chemical complexity of biomass, there is some choice of which platform chemicals to produce since, within limits, different processing strategies of the same material can lead to various breakdown products. Although, in principle, all oil refinery platform chemicals can be also derived from biomass, but with lower yields and higher costs (Haveren, Scott, and Sanders, 2008), the future biorefineries are expected to be based on a limited number of platforms, from which all the other commodity and bulk chemicals can be derived. In particular, the carbohydrate fraction of biomass feedstock (i.e. cellulose and hemicellulose in lignocellulosic biomass) is expected to play the biggest role

as a renewable carbon source for biochemical products. In fact, biomass polysaccharides can be effectively hydrolyzed to monosaccharides (e.g., glucose, fructose and xylose) which can then be converted, via fermentations or chemical synthesis, to an array of bio Platform Molecules (bPM building block chemicals with potential use in the production of numerous value-added chemicals), analogous to the petro-platform molecules of the current oil refinery (Werpy et al., 2004).

1.2 Composition of Biomass

Biomass is defined as consisting of all plant and plant-derived materials including livestock manures (Mansfield et al., 2006). Through photosynthesis, plants use light energy from the sun to convert water and carbon dioxide to sugars that can be stored. Some plants, like sugar cane and sugar beets, store the energy as simple sugars. Other plants, like corn, potatoes and root crops, store the energy as more complex sugars, called starches. Currently, industrial ethanol production is carried out by using starchy materials such as corn, wheat starch and potatoes. However, bioethanol from starchy materials has put the effort into direct competition with the food industry. Lignocellulosic biomass is the non-starch, fibrous part of plant materials. It is an attractive resource because it is renewable, abundant and low cost (Perlack et al., 2005). In recent years, more and more attention is being focused on the use of lignocellulosic biomass for the production of bioethanol via fermentation. Lignocellulosic biomass that can be used as feedstocks to produce bioethanol includes: 1) agricultural residues (leftover material from crops, such as corn stover and wheat straw); 2) forestry wastes (chips and sawdust from lumber mills, dead trees, and tree branches); 3) municipal solid wastes (household garbage and paper products); 4) food processing and other industrial wastes (black liquor, a paper manufacturing by-product); and 5) energy crops (fast-growing trees and grasses, such as switchgrass, poplar and willow) (Mansfield et al., 2006).

For this discussion, the word "biomass" will refer to "lignocellulosic biomass". The primary components of biomass are carbohydrate polymers (cellulose, hemicellulose) and

phenolic polymers (lignin). Low concentration of various other compounds, such as proteins, acids, salts, and minerals, are also present. The general composition of biomass is shown in Figure 1.1 (Lee et al., 2007).

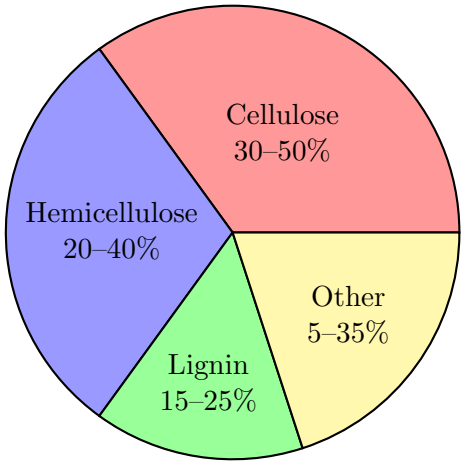

Figure 1.1: General composition of biomass

Cellulose is the most common form of carbon in biomass, accounting for 30%-50% by weight. It is a glucose polymer linked by β–1, 4 glycosidic bonds. The basic building block of this linear polymer is cellubiose, a glucose-glucose dimer. Hydrolysis of cellulose results in individual glucose monomers, which can be fermented to ethanol directly. Hemicellulose is a short, highly branched polymer containing five-carbon sugars (usually xylose and arabinose) and six-carbon sugars (glucose, galactose and mannose). It is at levels of between 20% and 40% by weight depending on the biomass types. Hemicellulose is more easily hydrolyzed than cellulose because of its branched, amorphous nature. When hydrolyzed, the hemicellulose from hardwoods releases products high in xylose (a five-carbon sugar). Lignin which provides structural integrity in plants is the largest non-carbohydrate fraction of lignocellulose. It makes up 15% to 25% by weight of biomass. Unlike cellulose and hemicellulose, lignin cannot be utilized in the fermentation process. However, it contains a lot of energy and can be burned to produce steam and electricity for the biomass-to-bioethanol process. The composition of cellulose, hemicellulose and lignin varies with the sources of biomass. Table 1.1 shows the composition of several selected agricultural residues, forestry wastes and energy

crops. An extensive list of chemical compositions of 82 varieties of biomass plus algae have been published (Vassilev et al., 2010).

Table 1.1: Cellulose, hemicellulose, and lignin content in various sources of biomass

Feedstock	**Cellulose**	**Hemicellulose**	**Lignin**	**Reference**
Corn stover	36.4	22.6	16.6	Mansfield et al. (2006)
Corn cob	42.0	39.0	14.0	Kuhad and Singh (1993)
Rice straw	32.0	24.0	13.0	Kuhad and Singh (1993)
Wheat straw	30.0	24.0	18.0	Kuhad and Singh (1993)
Rice hulls	36.0	15.0	21.0	Kuhad and Singh (1993)
Saw dust	55.0	14.0	21.0	Olsson and Hahn-Hägerdal (1996)
Willow	37.0	23.0	21.0	Olsson and Hahn-Hägerdal (1996)
Switchgrass	31.0	24.4	17.6	Mansfield et al. (2006)
Poplar	49.9	20.4	18.1	Mansfield et al. (2006)

1.3 Bioethanol

1.3.1 Development of bioethanol

The principle fuel used as a gasoline substitute for road transport vehicles is bio-ethanol. Bioethanol has a number of advantages over fossil fuels. Firstly, it comes from a renewable resource. Secondly, it is biodegradable, low in toxicity and causes little environmental pollution. Bioethanol is a high octane fuel and can be added into gasoline as an octane enhancer. In the United States, ethanol is blended with gasoline at a 10:90 ethanol-to-gasoline ratio to boost the fuel's octane rating, which allows it to burn more cleanly, reducing urban smog (Service, 2007). Thirdly, the use of bioethanol can reduce the greenhouse gas emissions. Relative to fossil fuels, greenhouse gas emissions are reduced about 18% by the use of corn-based ethanol, but it can be up to 88% if using cellulosic ethanol (Service, 2007).

A closed carbon dioxide cycle can be formed by using bioethanol as fuels. After combustion of bioethanol, the released carbon dioxide is recycled back into crops because crops use carbon dioxide to synthesize cellulose during photosynthesis (Chandel et al., 2007). In addition, blending bioethanol with gasoline will help extend the life of the diminishing fossil oil supplies and ensure greater fuel security, avoiding heavy reliance on oil producing nations that have not always been very stable. Another advantage of encouraging bioethanol use is that the rural economy would receive a boost from growing the necessary crops and creating new employment opportunities (Mansfield et al., 2006). In addition, using agricultural and industrial residues to produce bioethanol can solve the waste disposal problem and provide environmental benefits. Currently bioethanol is recovered from biomass feedstocks such as sugarcane, sugar beet and starch crops (mainly corn and wheat). In 2006, total world production reached 51.3 billion liters. USA is currently the largest producer of bioethanol with a production of 19.8 billion liters per year, with corn as primary feedstock. Sugarcane is used as primary feedstock in Brazil, currently the world's second largest producer (17.8 billion liters per year). The European Union produces 3.44 billion liters of bioethanol, mainly from sugar beet and starch crops (Cherubini, 2010). However, the increased demand for bioethanol will result in serious problems, such as supply scarcity and dramatic increases in the cost of the food. Moreover, even converting all the starch to bioethanol, it can only reduce 10% of the gasoline demand (Service, 2007). Therefore, lignocellulosic bioethanol is thought to be the answer for solving these problems.

1.3.2 Conversion from biomass to ethanol

Basically, the overall process for converting lignocellulose to bioethanol is comprised of four major unit operations: pretreatment, hydrolysis, fermentation and product separation/distillation. Figure 1.2 shows the basic features of this process.

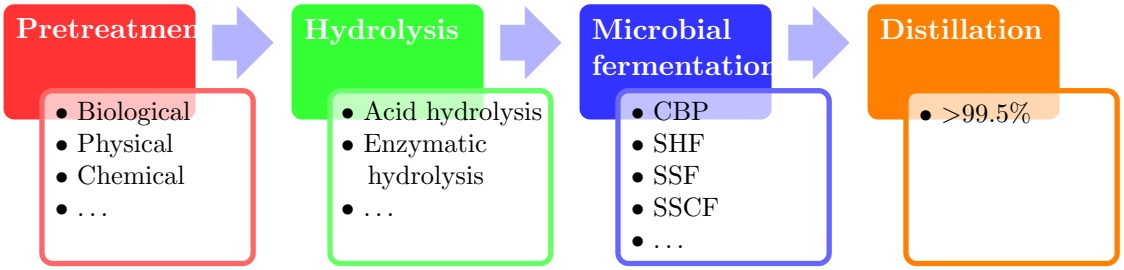

Figure 1.2: Stages of producing bioethanol

1.3.2.1 Pretreatment

Pretreatment is an important first step in the conversion process of biomass to bioethanol. This step reduces the biomass size and opens up the plant structure since native lignocellulosic biomass is extremely recalcitrant to hydrolysis, i.e., to make the lignocellulosic biomass amenable to hydrolysis. There are several pretreatment methods such as mechanical combination, steam explosion, ammonia fiber explosion, acid or alkaline pretreatment and biological treatment (Chandel et al., 2007). Each of these is suitable for different types of biomass. Currently, pretreatment is still one of the most expensive processing steps with the cost as high as 30 cents per gallon produced (Mosier et al., 2005). Therefore, lowering the cost of the pretreatment process is necessary in order to achieve the production of bioethanol from lignocellulosic biomass on commercial scale.

1.3.2.2 Hydrolysis

After pretreatment, the cellulose and hemicellulose portions need to be broken down further by enzymes or acids into monomeric sugars for the fermentation into ethanol. There are three principle methods of extracting sugars from biomass: dilute acid hydrolysis, concentrated acid hydrolysis and enzymatic hydrolysis.

The dilute acid hydrolysis process is one of the oldest and simplest methods of extracting fermentable sugars from biomass. This process is carried out in two stages. Different concentration sulfuric acid and temperature were applied in the two stages to optimize the

process. Dilute acid hydrolysis has some limitations. If higher temperatures or longer residence time are applied, the monomeric sugars derived from hemicellulose will degrade to form some fermentation inhibitors, such as furan compounds and weak carboxylic acids (Olsson and Hahn-Hägerdal, 1996). In order to remove these fermentation inhibitors, several chemical and biological methods could be used, such as ion exchange, charcoal adsorption and biological detoxification (Chandel et al., 2007). However, this will increase operating cost.

The concentrated acid hydrolysis process can provide complete and rapid conversion of cellulose to glucose and hemicellulose to xylose with little degradation. Approximately, 90% of both cellulose and hemicellulose can be depolymerized into their monomeric sugars with concentrated hydrolysis, so this process has the advantage of high sugar recovery efficiency (Chandel et al., 2007).

The enzymatic hydrolysis of cellulose into glucose is a slow and complex process because of the physical nature of the substrate. Cellulose in its native form has a highly crystalline structure. In addition, the cellulose is embedded in a matrix of lignin and hemicellulose, where the number of active enzyme binding sites available is limited. The factors that affect the enzymatic hydrolysis of lignocellulosic biomass include cellulose property, substrates, and reaction conditions (temperature, pH, etc.). At the same time, the process is very expensive compared with acid hydrolysis due to the high enzyme cost. Although the cost of cellulolytic enzyme has come down to 20 to 30 cents per gallon of ethanol produced, this conversion process cannot be competitive with the process of ethanol production from starch in corn kernels at a cost of 3 to 4 cents per gallon of ethanol (Stephanopoulos, 2007).

1.3.2.3 Fermentation

After hydrolysis, the primary fermentable sugars in hydrolysate are pentose and hexose, such as glucose and xylose. Different microorganisms are used to ferment these sugars to produce bioethanol, such as *Saccharomyces cerevisiae*, *Scheffersomyces stipitis*, *Kluyveromyces*

marxianus, *Candida shehatate*, *Zymomonas mobilis* and *Escherichia coli*. Currently, the fermentation of a mixture of hexose and pentose is inefficient because no wild organism has been found that can convert all sugars into ethanol at a high yield (Ragauskas et al., 2006). There are several strategies for fermentation process: Consolidated BioProcessing (CBP) (Lynd et al., 2005), Separate Hydrolysis and Fermentation (SHF) (Toon et al., 1997), Simultaneous Saccharification and Fermentation (SSF) (Eken-Saraçolu and Arslan, 2000; Tomás-Pejó et al., 2008), and Simultaneous Saccharification and Co-Fermentation (SSCF) (Chandel et al., 2007; Ohgren et al., 2007).

1.3.3 Current issues and strategies

Lignocellulosic bioethanol is proposed as having such benefits as: reduction of greenhouse gas emissions, reduction of fossil fuel use, increased national energy security, increased rural development, a sustainable fuel supply for the future. Although significant advances have been made at bench scale toward the bioethanol generation from lignocellulose, there are still technical and economic barriers, which make the bioethanol program unsuccessful on a commercial scale. Currently, the challenges include: 1) low bulk density feedstock; 2) high viscosity substrate; 3) optimization of hydrolysis and fermentation; 4) fermentability of substrate; 5) xylose fermentation; 6) cost challenges.

1.4 *Scheffersomyces stipitis*

1.4.1 Introduction

Scheffersomyces stipitis (*S. stipits*, formerly known as *Pichia stipitis*) (Kurtzman and Suzuki, 2010) has a set of physiological traits that make it very useful for the bioconversion of lignocellulose. In addition to its extensively studied capacity for xylose fermentation, it is also able to ferment, glucose, mannose, galactose and cellobiose along with mannan and xylan oligomers. This makes it a potent organism for hydrolysate or SSF (Jeffries and Van Vleet, 2009). After glucose, xylose is the second most abundant hemicellulosic component in

agricultural residues and fast-growing hardwood species, and cellobiose is the primary sugar formed in enzymatic hydrolysis.

Currently researchers in numerous laboratories have borrowed genes from *S. stipitis* and other fermentative microorganisms to modify *S. cerevisiae* for xylose, xylan or cellulose metabolism. While partly successful, efficient xylose utilization has been impaired by *S. cerevisiae*'s generally low rate of xylose consumption and its inappropriate regulatory responses. It lacks sufficient levels of the assimilatory genes, sugar transporters and mechanisms for balancing cofactor levels under oxygen-limiting conditions. Besides providing genes to other microorganism, the fermentation performance of *S. stipitis* has been compared with other strains widely used in biochemical industry. *S. stipitis* showed to be one of the best overall performance (Rumbold et al., 2009, 2010).

No matter *S. stipitis* works as a gene provider for other microorganisms or as a host to accept genes from other microorganisms, a detailed understanding of physiology, biochemistry and genetics of *S. stipitis* is required. This is possible only when the major pathways and mechanisms are known. Biochemical and genetic characterization of xylose fermentation by *P. stipitis* Pignal (1967) (*Yamadazyma stipitis*) has been underway for at least 15 years since the development of systems for its genetic transformation (Laplaza et al., 2006; Lu et al., 1998; Yang et al., 1994) and mating (Melake, Passoth, and Klinner, 1996). Relatively few researchers, however, have attempted its rational modification despite the fact that native strains produce more ethanol from xylose than any other studied yeast – including genetically modified *S. cerevisiae*.

S. stipitis has the highest native capacity for xylose fermentation of any known microorganism (Dijken et al., 1986; Preez, Driessel, and Prior, 1989). This yeast was originally isolated from insect larvae and is closely related to several yeast endosymbionts of passalid beetles (Nardi et al., 2006) that inhabit and degrade white-rotted hardwood (Suh et al., 2003). It is a predominantly haploid, homothallic, hemiascomycetous yeast (Gupthar, 1994;

Kurtzman, 1990; Melake, Passoth, and Klinner, 1996) that forms buds along with pseudomycelia during vegetative growth, and two hat-shaped ascospores from each ascus. Fed batch cultures of *S. stipitis* produce up to 47 g/L of ethanol from xylose at 30 °C (Preez, Driessel, and Prior, 1989) with ethanol yields of 0.35–0.44 g/g xylose (Hahn-Hägerdal and Pamment, 2004), and they are capable of fermenting sugars from hemicellulosic acid hydrolysates with a yield equivalent to about 80% of the maximum theoretical conversion efficiency (Nigam, 2001a,b).

The genome of *S. stipitis* codes for cellulases, mannases, xylanase and other degradative enzymes that enable survival and growth in a wood-inhabiting, insect-gut environment (Nardi et al., 2006). *S. stipitis* has the capacity to ferment xylose, xylan (Lee et al., 1986; Özcan, Kötter, and Ciciary, 1991) and cellobiose, and to use all of the major sugars found in wood, including arabinose and rhamnose (Koivistoinen et al., 2008). For these reasons, *S. stipitis* has been a common source of genes for engineering xylose metabolism in *S. cerevisiae* (Jeffries and Jin, 2004).

S. stipitis also has a number of other bioconversion related traits: it modifies low-molecular-weight lignin moieties (Jeffries and Van Vleet, 2009), reduces acyclic enones to the corresponding alcohols (Conceição, Moran, and Rodrigues, 2003), forms various esters and aroma components (Fuganti et al., 1993) and can be engineered to produce lactic acid (Ilmén et al., 2007) or xylitol (Kim et al., 2001; Rodrigues et al., 2008) in high yield. Strains of *S. stipitis* have also been selected for resistance to furfural and hydroxy-methyl furfural (Liu, Slininger, and Gorsich, 2005).

Metabolic engineering and adaptive evolution of *S. cerevisiae* for xylose fermentation has been successful to varying degrees (Harhangi et al., 2003; Karhumaa, Hahn-Hägerdal, and Gorwa-Grauslund, 2005; Sonderegger et al., 2004). Engineering it with the basic assimilatory machinery of *XYL*1, *XYL*2, *XYL*3 (or *XKS*1), *TAL*1, *TKL*1, *RPE*1 and *RPI*1 enables ethanol production. Expressing xylose isomerase (Maris et al., 2007; Wiedemann

and Boles, 2008) or xylose reductases and xylitol dehydrogenases with altered cofactor specificities (Matsushika and Sawayama, 2008; Petschacher and Nidetzky, 2008) reduces cofactor imbalances, and increases the ethanol yield. It is not yet clear as to which of these engineering approaches will prove to be more successful in *S. cerevisiae* (Karhumaa et al., 2007).

Overexpression of *S. stipitis* or other fungal sugar transporters can also improve the performance of engineered *S. cerevisiae* on xylose (Hector et al., 2008; Katahira et al., 2008; Leandro, Spencer-Martins, and Gonçalves, 2008; Saloheimo et al., 2007; Weierstall, Hollenberg, and Boles, 1999), but additional regulatory engineering is necessary because *S. cerevisiae* does not possess mechanisms to coordinate ethanol production in response to xylose (Jin, Laplaza, and Jeffries, 2004). Therefore, even though the genetic tools, detailed biochemical knowledge and physiological properties of *S. cerevisiae* hold great promise for engineering the fermentation of xylose, xylan, cellulose, arabinose, rhamnose (Koivistoinen et al., 2008) and other sugars, much remains to be learned from *S. stipitis* and other yeasts that use these substrates natively. Conversely, the mechanisms *S. cerevisiae* use to ferment xylose can be adapted to improve the performance of *S. stipitis*.

S. stipitis shunts most of its metabolic flux into ethanol, and produces very little xylitol, but its fermentation rate on xylose is low relative to that of *S. cerevisiae* on glucose. Glucose and xylose are not equivalent fermentations for many reasons, but increasing the capacity of *S. stipitis* for rapid xylose fermentation could greatly improve its usefulness in commercial applications. Unlike *S. cerevisiae*, which regulates fermentation by sensing the presence of glucose, *S. stipitis* induces fermentative activity in response to oxygen limitation (Klinner et al., 2005; Passoth, Zimmermann, and Klinner, 1996; Passoth et al., 2003). This, however, does not constitute the only fermentative regulatory mechanism. Global expression array analysis has shown specific response patterns for xylose, cellobiose, arabinose, rhamnose and other lignocellulosic substrates. It is not fully known whether these are attributable to carbon catabolite derepression or specific induction. Our expression array results show evidence for both.

1.4.2 Physiological features of *S. stipitis*

Most of the research with *S. stipitis* has focused on its capacity to ferment xylose. Even so, relatively little has been established concerning the rate-limiting steps in ethanol production from this sugar. An early work (Bicho et al., 1988) showed that xylose reductase (*Xyl*1) and xylitol dehydrogenase (*Xyl*2) are repressed by glucose and induced during growth on xylose. Xylose is generally not consumed in the presence of glucose; hence, under glucose repression, these activities, along with xylose transport, are rate limiting. In a respiration-limited, cyc1 mutant of *S. stipitis*, however, xylose is used coincidently with glucose as compared with the CYC1 parental strain (Shi et al., 1999), suggesting that reducing ATP production can bring about a partial derepression of xylose assimilation.

Increasing the expression of *XYL*1 for xylose reductase (Takuma et al., 1991) increased the enzymatic activity almost twofold, but had no beneficial effect on ethanol production (Dahn et al., 1996). To date, the overexpression of *XYL2* in *S. stipitis* has not been examined (Kötter et al., 1990); however, deletion of *XYL2* blocks xylose utilization at the level of xylitol and prevents its growth on this carbon source (Kim et al., 2001; Laplaza et al., 2006; Shi et al., 2000). D-Xylulokinase activity (*Xks*1) (Ho, Chen, and Brainard, 1998) limits the rate of xylose assimilation by *S. cerevisiae* (Karhumaa, Hahn-Hägerdal, and Gorwa-Grauslund, 2005; Richard, Toivari, and Penttilä, 2000). D-Xylokinase (*Xyl*3) does not, however, appear to be rate limiting in *S. stipitis* once it is induced on xylose. *Xyl*3 from *S. stipitis* exhibits about three times the specific activity of *Xks1* from *S. cerevisiae*, and cells can still metabolize xylose via a bypass pathway even in a xyl3D background (Jin et al., 2002). This indicates that a second pentose kinase pathway is active in *S. stipitis*.

Deletion of the *S. stipitis* *ADH*1 and *ADH*2 genes (Cho and Jeffries, 1998; Passoth et al., 1998) decreases ethanol production dramatically, while increasing xylitol production. Adh activities for *S. stipitis* increase under oxygen-limiting conditions (Cho and Jeffries, 1999; Passoth et al., 2003) in much the same manner as that observed with *Candida shehatae* (Alexander, Chapman, and Jeffries, 1987). Pyruvate decarboxylase activities are also

induced under oxygen-limiting conditions along with increasing fermentative activity (Lu, Davis, and Jeffries, 1998; Passoth et al., 1998). Taken together, these findings suggest that the final steps of the fermentative pathway direct the flow of the reductant from xylitol to ethanol.

The respiratory capacity of *S. stipitis* is notably greater than that of *S. cerevisiae*. Particularly, *S. stipitis* possesses an alternative, nonphosphorylating terminal oxidase (Shi et al., 2002) in addition to a fully functional NADH dehydrogenase complex (respiratory complex I), both of which are lacking in *S. cerevisiae*. While these enable much higher growth yields and the capacity to grow at very low oxygen levels, they also reduce the intracellular NADH supply for fermentation and result in higher cell yields than with *S. cerevisiae*. Deleting *S. stipitis* cytochrome c (*CYC*1) reduces the cell yield and growth rate, while shunting more substrate into ethanol (Shi et al., 1999). Deleting the alternative oxidase reduces the capacity of *S. stipitis* to scavenge oxygen at low levels.

S. stipitis possesses β-xylosidase (Basaran and Ozcan, 2008; Manzanares, Ramón, and Querol, 1999) and native Family 11 xylanase activities. The latter has been cloned and characterized from *S. stipitis* NRRL Y-11543 (Basaran et al., 2001). The published xylanases sequence does not match with any identified ORF in the sequenced genome of *S. stipitis* CBS 6054 (= NRRL Y-11545, ATCC 58785), but the sequenced genome does include Family 10 endo-1,4-b-xylanase, and endoglucanase activities that might also act on xylan. *S. stipitis*'s native xylanase activity has been supplemented through heterologous expression (Görgens et al., 2005; Passoth and Hahn-Hägerdal, 2000).

1.5 Systems biology

In the twentieth century, engineering sciences have inspired numerous successful applications in the fields of manufacturing, electronics, communications, transportation, computer and networks, and so on. Compared to the engineering systems, biological systems are more complex and their mechanisms are less known. Historically, biological questions have been

approached by a reductionist paradigm that is completely different from methodologies being applied to engineering systems. This reductionist way of thinking was based on the assumption that by unraveling the function of all the different components the information gained could be used to piece together the puzzle of complex cellular networks (Hofmeyr and Westerhoff, 2001). The research paradigm has dominated mainstream biology with enormous progresses in accumulating biological information at genetic and protein levels. However, this is a slow and exhaustive process that fails to adequately approach the true complexities of living phenomena and is of limited relevance to biological systems as a whole.

The fast-growing applications of genomics and high-throughput technologies (Wheeler et al., 2006) have led to recognition of the limitations of the reductionist/atomistic view of the world. It is realized that a new systems biology paradigm is needed for the next level of understanding of the functions of the genes and proteins, and the regulation of intracellular networks that cannot be obtained by studying the individual constituents on a part-by-part basis. It is also realized that there is great similarity between biology and engineering at the system level, despite their obviously different physical implementation, and that important research challenges in biology may have parallels with those uncomplicated engineering systems (Fu et al., 2009). This similarity forms a basis for the introduction of synthetic biology or the engineering applications within biological systems, which is beyond the scope of this dissertation and will not be discussed here.

Systems biology attempts to investigate the behavior and relations of all the elements in a particular biological system while it is functioning (Ideker, Galitski, and Hood, 2001; Palsson, 2000). It aims at system-level understanding of biological processes and biochemical networks as a whole. This "system-oriented" new biology is shifting our focus from examining particular molecular details to studying the information flows at all biological levels: Genomic DNA, mRNA, proteins, informational pathways, and regulatory networks. Systems biology approaches seek to study the complexity of life to help in understanding how the cellular networks work together. To this end, the approach emphasizes the investigation of

biological phenomena by considering system structures, system dynamics, control methods, and design methods (Kitano, 2002; Wolkenhauer, 2001). It requires a broad interdisciplinary integration of molecular and cell biology, biochemistry, informatics, mathematics, computing, and engineering. It does not apply to genome-scale studies that are focused solely on discovery. Rather, it is a framework for using genome-scale experiments to perform predictive, hypothesis driven science (Figure 1.3) (Chuang, Hofree, and Ideker, 2010).

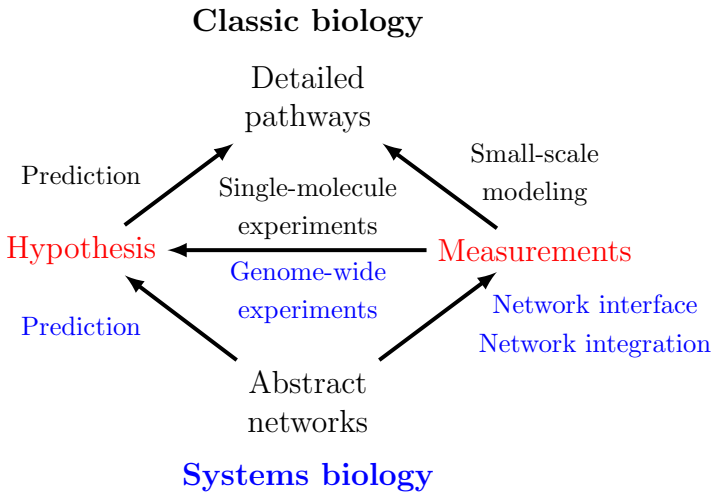

Figure 1.3: Overview of the experimental process in classic biology (top) versus systems biology (bottom). (Reproduced from Chuang, Hofree, and Ideker (2010))

There are two main approaches to computational analysis of biological data. The causal approach makes concrete deterministic or stochastic models (differential equations, stochastic differential equations, Boolean networks, et cetera) of biological processes. The probabilistic view is associated with probabilistic inference approaches, using pattern recognition or learning algorithms (such as neural networks and graphical models) for analysis of data from large-scale experimental methods. These two approaches rest on a large part of applied mathematics (including numerical integration, optimization, interpolation, and control theory) and computer science (search theory, coding theory, and database design). This breadth necessitates collaborations between people with diverse backgrounds, but an inadequate understanding of the limitations and applicability of techniques and concepts from different fields hinders such collaborations. The background information required makes biological

modeling a difficult task, but the real challenge remains that of making computational models effective and efficient representations of biological systems (Szallasi, Stelling, and Periwal, 2006).

1.6 Modeling and analysis of metabolic network

Metabolic network modeling can be classified into dynamic and structural metabolic network modeling. An accurate dynamic metabolic network model can help elucidate complex dynamic cellular phenotypes under different environmental and genetic perturbations. A general framework has been proposed to develop a dynamic model for a genome-scale metabolic network (Jamshidi and Palsson, 2008). Its development is hampered by the unavailability of kinetic parameters and is limited to small-size metabolic networks (typically ¡100 reactions). Several approaches have been developed to generate kinetic parameters under uncertainty to enable dynamic metabolic network modeling, for instance, cybernetic kinetic modeling and ensemble metabolic network modeling through various assumptions (Machado et al., 2012; Song and Ramkrishna, 2012; Tran, Rizk, and Liao, 2008; Wang, Birol, and Hatzimanikatis, 2004; Young et al., 2008).

Due to lack of kinetic parameters, structural metabolic network modeling has been widely applied for analyzing cellular metabolism under steady-state. Depending on what assumptions are made and whether experimental data are required, different techniques have been developed to analyze the invariant of metabolic networks such as metabolic flux analysis (MFA), flux balance analysis (FBA), and metabolic path-wayanalysis (MPA) including elementary mode (EMA) and extreme pathway analyses (EPA) (Lewis2012, Stephanopoulos1998, Trinh2009, Trinh2012).

1.6.1 Flux balance analysis

Flux Balance Analysis (FBA) is a causal approach for analyzing the flow of metabolites through a metabolic network (Orth, Thiele, and Palsson, 2010), in particular the genome-scale metabolic network reconstructions that have been built in the past decade (Duarte et al., 2007; Feist and Palsson, 2008; Feist et al., 2007; Oberhardt, Palsson, and Papin, 2009). FBA calculates the flow of metabolites through this metabolic network, thereby making it possible to predict the growth rate of an organism or the rate of production of a biotechnologically important metabolite. With metabolic models for more than 35 organisms already available (see on-line list) and high-throughput technologies enabling the construction of many more each year, FBA is an important tool for harnessing the knowledge encoded in these models.

FBA is based on the fundamental physicochemical constraints on metabolic networks. It does not require kinetic parameters and can be computed very quickly even for large networks. This makes it well suited to studies that characterize many different perturbations such as different substrates or genetic manipulations.

FBA has limitations, however. Because it does not use kinetic parameters, it cannot predict metabolite concentrations. It is also only suitable for determining fluxes at steady state. Except in some modified forms, FBA does not account for regulatory effects such as activation of enzymes by protein kinases or regulation of gene expression. Therefore, its predictions may not always be accurate.

Flux Balance Analysis usually assumes time-invariant extracellular conditions and generates steady-state predictions consistent with continuous fermentation (Palsson, 2006). However, large-scale production of metabolic products is often achieved in batch and fed-batch culture. Therefore, in order to combine the extracellular kinetics with FBA, Dr. M.A. Henson has dedicated into the study of Dynamic Flux Balance Analysis (DFBA) (Hjersted and Henson, 2006; Hjersted, Henson, and Mahadevan, 2007).

1.6.1.1 General procedure of flux balance analysis

The general procedure of formulation of an FBA problem is shown by Figure 1.4 (Orth, Thiele, and Palsson, 2010). From the procedure we can tell that the procedure can be roughly separated into two parts: model preparation, mathematical description and solution. The modeling part includes the genome-scale metabolic reconstruction, finding the proper constraints and objective function that are biologically meaningful. The mathematical description and solution part includes describing the reconstructed network with stoichiometric matrix, mathematically describing the constraints and objective functions, solve the optimization problem with LP.

1.6.1.2 Model preparation

A comprehensive guide to creating, preparing and analyzing a metabolic model using FBA, in addition to other techniques, has been published (Thiele and Palsson, 2010). It is an iterative procedure (Figure 1.5). The key parts of model preparation are: creating a metabolic network without holes, adding constraints to the model and finally adding an objective function (often called the Biomass function), usually to simulate the growth of the organism being modeled.

The network Metabolic networks can vary in scope from those describing the metabolism in a single pathway, up to the cell, tissue or organism. The only requirement of a metabolic network that forms the basis of an FBA-ready network is that it contains no gaps. This typically means that extensive manual curation is required, making the preparation of a metabolic network for flux balance analysis a process that can take months or years. Software packages exist to speed up the creation of new FBA-ready metabolic networks. Generally models are created in SBML format so that further analysis or visualization can take place in other software although this not a requirement. The details of different steps to reconstruct the network have been discussed (Thiele and Palsson, 2010).

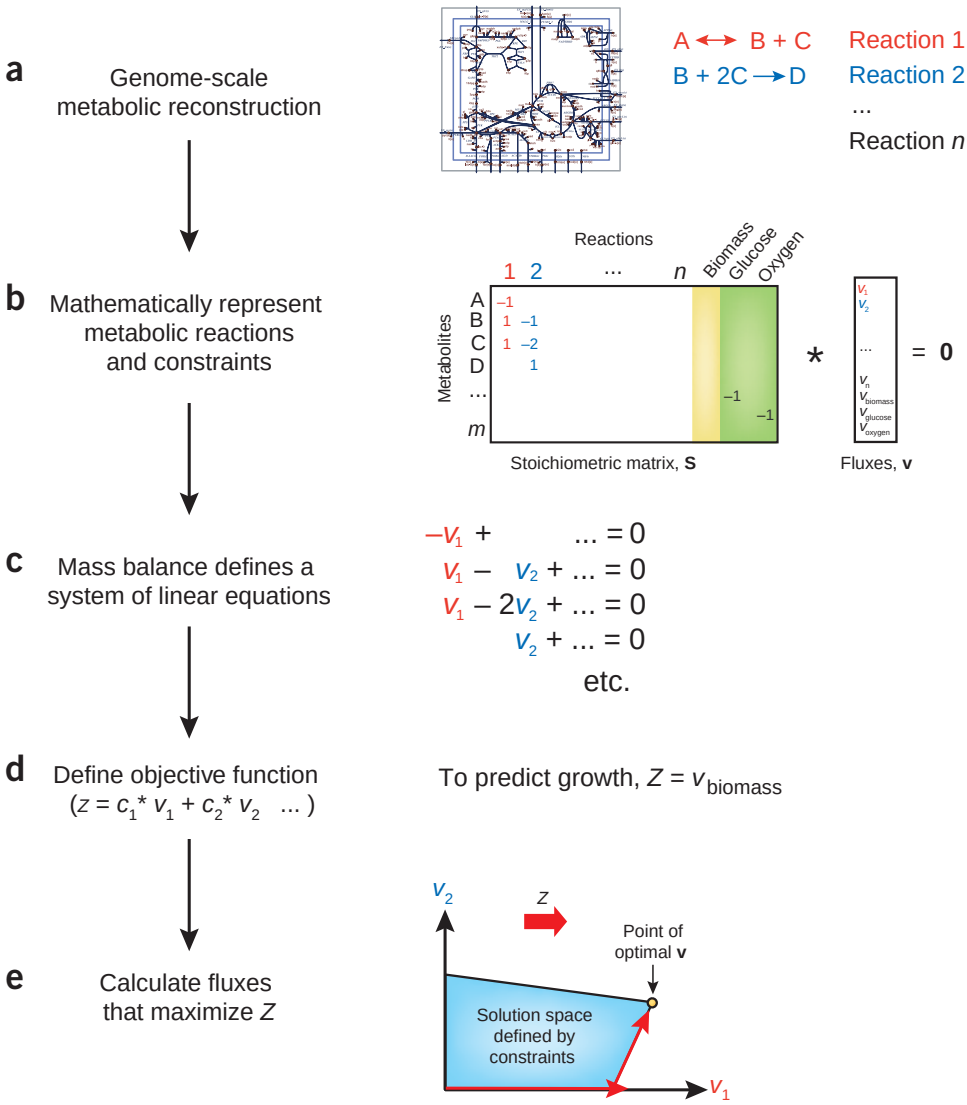

Figure 1.4: Formulation of an FBA problem. Adopted from Orth, Thiele, and Palsson (2010).

Constraints A key part of FBA is the ability to add constraints to the flux rates of reactions within networks, forcing them to stay within a range of selected values. This lets the model more accurately simulate real metabolism and can be thought of biologically in two subsets: constraints that limit nutrient uptake and excretion and those that limit the flux through reactions within the organism. FBA-ready metabolic models that have had constraints added can be analyzed using software such as the COBRA toolbox (Becker et al., 2007; Schellenberger et al., 2011). Different kind of constraints can be used to limit possible

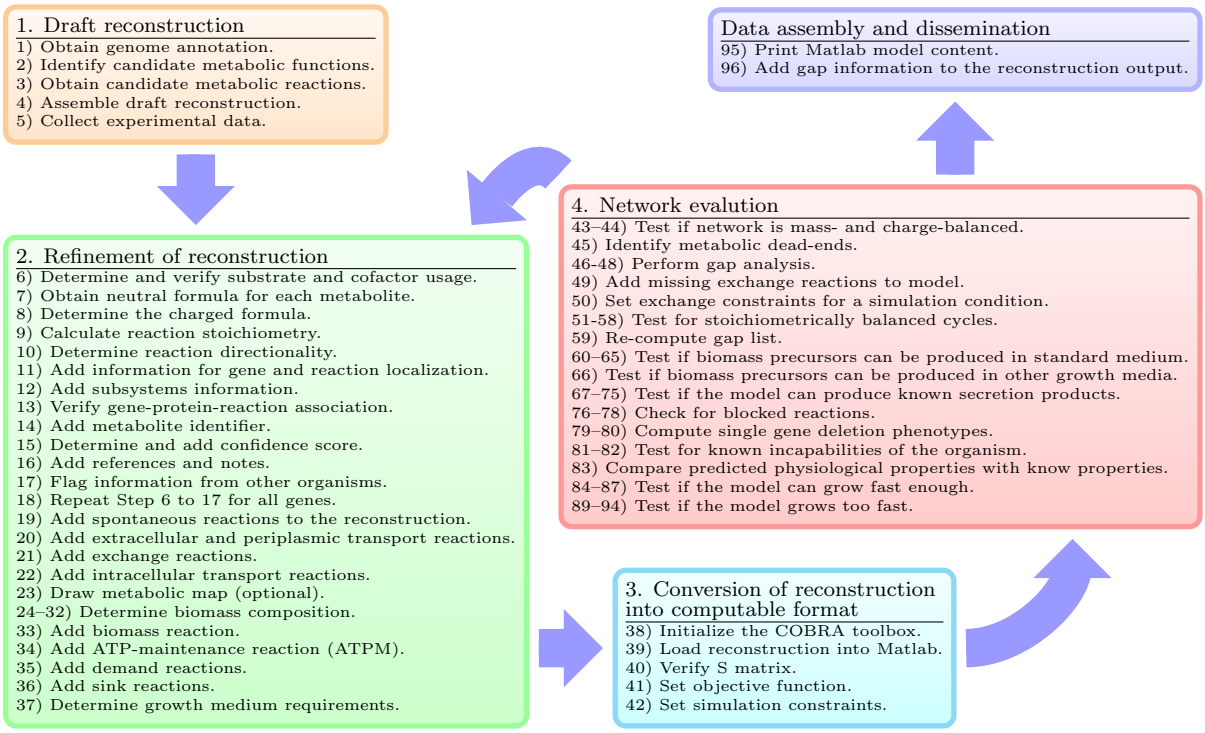

Figure 1.5: Overview of the procedure to iteratively reconstruct metabolic network. Adopted from Thiele and Palsson (2010).

solutions. An overview of constraints used by flux balance analysis is given by Hjersted, Henson, and Mahadevan (2007).

- **Physico-chemical constraints** are the so-called 'hard' constraints on cell functions. These constraints will not change due to environmental conditions. Examples of these hard constraints are mass, energy and momentum which is conserved in the cell, which means that during the experiment the total quantity of these variables, will be equal to these initial values. The conservation of mass is described by the mass balance equations and will form the most important constraints for the FBA model.

- **Topobiological constraints.** The crowding of molecules in cells leads to topobiological or three-dimensional constraints. An example is the DNA tightly packed within the nucleus because DNA stretched will be a 1000 times larger than the size of the cell. At the same time DNA has to be accessed for transcription in high quantities and

fast. Another example is the ratio between the available amount of tRNA molecules and ribosomes, which are respectively the building blocks and factories for proteins.

- **Environmental conditions** in cells are time and conditions dependent. Nutrient concentrations, pH value and temperature are examples of environmental constraints.

Organisms, and all other metabolic systems, require some input of nutrients. Typically the rate of uptake of nutrients is dictated by their availability, their concentration and diffusion constants (higher concentrations of quickly-diffusing metabolites are absorbed more quickly) and the method of absorption (such as active transport or facilitated diffusion versus simple diffusion).

If the rate of absorption (and/or excretion) of certain nutrients can be experimentally measured then this information can be added as a constraint on the flux rate at the edges of a metabolic model. This ensures that nutrients that are not present or not absorbed by the organism do not enter its metabolism (the flux rate is constrained to zero) and also means that known nutrient uptake rates are adhered to by the simulation. This provides a secondary method of making sure that the simulated metabolism has experimentally verified properties rather than just mathematically acceptable ones. In mathematical terms, the application of constraints can be considered to reduce the solution space of the FBA model.

In addition to those applied at the edges of a metabolic network, constraints can be applied to reactions deep within the network. These constraints are normally usually simple; they may constrain the direction of a reaction due to energy considerations or constrain the maximum speed of a reaction due to the finite speed of all reactions in nature.

Objective function In FBA there are a large number of mathematically acceptable solutions to the steady-state problem ($S\vec{v} = 0$) but the ones that are biologically interesting are those that produce the desired metabolites in the correct proportion. The set of metabolites, in the correct proportions, that an FBA model tries to create is called the objective function. When modeling an organism the objective function is generally the biomass of the organism

and simulates growth and reproduction. If the biomass function is defined sensibly, or exactly measured experimentally, it can play an important role in making the results of FBA biologically applicable: by ensuring that the correct proportion of metabolites are produced by metabolism and by predicting exact rates of Biomass production for example.

When modeling smaller networks the objective function can be changed accordingly. An example of this would be in the study of the carbohydrate metabolism pathways where the objective function would probably be defined as a certain proportion of ATP and NADH and thus simulate the production of high energy metabolites by this pathway.

1.6.1.3 Mathematical description

A biological network can be thought of as a set of nodes (compounds) connected by directional edges (reactions) and therefore represented as a matrix. The properties of this matrix are well known and thus a biological problem becomes amenable to computational analysis. A real biological system is extremely complex which in turn leads to problems measuring enough parameters to define the system and in some cases requiring a huge amount of computing time to perform simulations. Flux balance analysis simplifies the representation of the biological system, requiring fewer parameters (such as enzyme kinetic rates, compound concentrations and diffusion constants) and greatly reduces the computer time required for simulations.

Homeostasis Much of the power of flux balance analysis comes from applying the principle of homeostasis to the problem. Since the internal concentrations of metabolites within a biological system remain more or less the same over time we can apply the homeostatic condition that,

$$\frac{\mathrm{d}[C]_i}{\mathrm{d}t} = 0 \tag{1.1}$$

and thus simplify the problem to one of simply balancing the fluxes within the system, hence the name flux balance analysis.

The stoichiometric matrix The representation of the equations above can be generalized to any similar biological network and represented in a more powerful manner by using matrices. In this stoichiometry matrix (S) of size $m \times n$, every row represents one unique compound (for a system with m compounds) and every column represents one reaction (n reactions). The entries in each column are the stoichiometric coefficients of the metabolites participating in a reaction. There is a negative coefficient for every metabolite consumed and a positive coefficient for every metabolite that is produced. A stoichiometric coefficient of zero is used for every metabolite that does not participate in a particular reaction. S is a sparse matrix because most biochemical reactions involve only a few different metabolites. The flux through all of the reactions in a network is represented by the vector $\vec{v}$, which has a length of n.

The systems of mass balance equations at steady state ($dC_i/dt = 0$) is given:

$$S\vec{v} = 0 \tag{1.2}$$

This general operation is called taking the Null Space of the stoichiometric matrix S and the technique is valid for all stoichiometric matrices. Since a typical stoichiometric matrix contains many more metabolites than reactions ($m > n$) and the majority of reactions are linearly independent there is no unique solution to this system of equations.

1.6.1.4 Application to the biology of the system

The analysis of the null space of matrices is common within linear algebra and many software packages such as Matlab can help with this process. Nevertheless, knowing the null space of S only tells us all the possible collections of flux vectors (or linear combinations thereof) that balance fluxes within the biological network. Flux balance analysis has two further aims, to accurately represent the biology limits of the system and to return the flux distribution closest to that naturally occurring within the target system/organism.

The stoichiometric matrix is almost always underdetermined meaning that the solution space to $S\vec{v} = 0$ is very large. The size of the solution space can be reduced, and made more reflective of the biology of the problem through the application of certain constraints on the solutions.

Constrains are represented in two ways, as equations that balance reaction inputs and outputs and as inequalities that impose bounds on the system. The matrix of stoichiometries imposes flux (that is, mass) balance constraints on the system, ensuring that the total amount of any compound being produced must be equal to the total amount being consumed at steady state. Every reaction can also be given upper and lower bounds, which define the maximum and minimum allowable fluxes of the reactions. These balances and bounds define the space of allowable flux distributions of a system — i.e. the rates at which every metabolite — is consumed or produced by each reaction.

Certain flux rates can be measured experimentally and the fluxes within a metabolic model can be constrained to ensure these known flux rates are accurately reproduced in the simulation. Flux rates are most easily measured for nutrient uptake at the edge of the network but measurements of internal fluxes are possible, generally using radioactively labeled or NMR visible metabolites.

Even after the application of constraints there are usually a large number of possible solutions to the flux balance problem. If an optimization goal is defined, linear programming can be used to find a single optimal solution. The most common biological optimization goal for a whole organism metabolic network would be to choose the flux vector $\vec{v}$ that maximizes the flux through a biomass function composed of the constituent metabolites of the organism placed into the stoichiometric matrix and denoted $v_{biomass}$ or simply v_b

$$\max_{\vec{v}} v_b \quad s.t.\ S\vec{v} = 0 \tag{1.3}$$

In the more general case any reaction be defined and added defined as a biomass function with either the condition that it be maximized or minimized if a single "optimal" solution is desired. Alternatively, and in the most general case, a vector $\vec{c}$ can be defined which defines the weighted set of reactions that the linear programming model should aim to maximize or minimize,

$$\max_{\vec{v}} \vec{v} \cdot \vec{c} \quad s.t.\ S\vec{v} = 0 \tag{1.4}$$

In the case of there being only a single separate biomass function/reaction within the stoichiometric matrix $\vec{c}$ would simplify to all zeroes with a value of 1 (or any non-zero value) in the position corresponding to that biomass function. Where there were multiple separate objective functions $\vec{c}$ would simplify to all zeroes with weighted values in the positions corresponding to all objective functions.

1.6.1.5 Dynamic flux balance analysis

FBA assumes time-invariant extracellular conditions and generates steady-state predictions consistent with continuous culture. However, large-scale production of metabolic products is often achieved in batch and fed-batch culture. Dynamic flux balance models are obtained by combining stoichiometric equations for intracellular metabolism with dynamic mass balances on key extracellular substrates and products assuming fast intracellular dynamics. The intracellular and extracellular descriptions are coupled through the cellular growth rate and substrate uptake kinetics. Dynamic flux balance analysis (DFBA) offers the possibility of formulating substrate uptake kinetics to account for known regulatory processes. DFBA has been primarily used to generate dynamic predictions of substrate, biomass and product concentrations for wild type growth in batch culture. The utility of yeast dynamic flux balance models for optimization of fed-batch operating strategies (Hjersted and Henson, 2006) and identification of ethanol overproduction mutants (Hjersted, Henson, and Mahadevan, 2007) have been shown.

Chapter 2
Experimental Fermentation of Glucose and Xylose

Although many experimental report of *S. stipitis* have been published, published results on continuous fermentation of *S. stipitis* with glucose and xylose (Fiaux et al., 2003; Grootjen, Lans, and Luyben, 1990; Skoog and Hahn-Hägerdal, 1990; Skoog, Jeppsson, and Hahn-Hägerdal, 1992) as carbon source are very limited. One of the reasons is that it is difficult to maintain continuous fermentation by *S. stipitis* with xylose under oxygen limited condition due to low growth rate. Two important aspects of *S. sitpitis* as discussed in Section 1.4 have been experimentally studied in this work: (i) influence of oxygen availability; (ii) ethanol tolerance. In this chapter, the fermentation of glucose and xylose were carried out in a modified commercial bioreactor under "pseudo-continuous" mode to study the general performance of *S. stipitis* CBS 6054 and the influence of oxygen availability. The investigation on the impact of "pseudo-continuous" fermentation to the ethanol tolerance of *S. stipitis* will be discussed later in Chapter 3.

2.1 Materials and methods

2.1.1 Microorganism and media

S. stipitis CBS 6054 was obtained from Dr. Thomas W. Jeffries at the University of Wisconsin-Madison. The strain was maintained at 4 °C on YPX agar plates containing 10 g yeast extract, 20 g peptone, 20 g xylose and 15 g agar per liter deionized (DI) water.

The culture media are modified based on the minimal medium (Jeffries et al., 2007). The pre-culture medium contained (per liter DI water): 20 g D-glucose or D-xylose, 1.7 g yeast nitrogen base without amino acids and ammonium sulfate (YNB w/o AA & AS) (BD, Fanklin Lakes, NJ) and 2.27 g urea.

The culture medium for batch fermentation and initial phase of pseudo-continuous fermentation contained (per liter DI water): 20 g D-glucose or D-xylose, 1.7 g YNB w/o AA & AS and 2.27 g urea.

The feed medium for pseudo-continuous fermentation contained (per liter DI water): 50 g D-glucose or D-xylose, 1.7 g YNB w/o AA & AS and 2.27 g urea.

2.1.2 Pseudo-continuous fermentation

The concept of "pseudo-continuous" fermentation is not a new concept. It has been applied in animal cell culture, which is termed as "perfusion" (Butler, 2005). The main reason that we introduce "pseudo-continuous" fermentation here is that it can completely eliminates the potential washout, which allows us to operate and study the fermentation under various conditions, such as different dilution rates and oxygen supply rates. Note that for traditional continuous fermentation of xylose using *S. stipitis*, one major difficulty is the frequent washout due to the slow or even negative growth rate under micro-aerobic condition, which has been reported (Rizzi et al., 1989; Shi and Jeffries, 1998) and shown in our experiments. Besides, there are also two other advantages associated with the pseudo-continuous fermentation. First, the pseudo-continuous fermentation can help improve the fermentation rate, as very high cell density can be easily obtained with cell retention, which will increase the ethanol throughput. Second, this system can be used for ethanol adaptation. Because the adaptation is provided by the ethanol produced by the cells, it is assured that the environment pressure always exists. Therefore, whether or not the improved ethanol tolerance is achieved through mutation, there is no risk of losing the obtained capability

during the fermentation process.It can also be used for similar situations for strain adaptation or evolution.

The pseudo-continuous fermentation system has been built upon a modified Bioflo 110 fermenter with the in-house developed cell retention module, which is shown in Figure 2.1. During the experiments, the temperature was controlled at 30 °C and pH was maintained at 5.0 by automatic addition of 3.0 M KOH. Agitation speed was set to be 300 rpm.

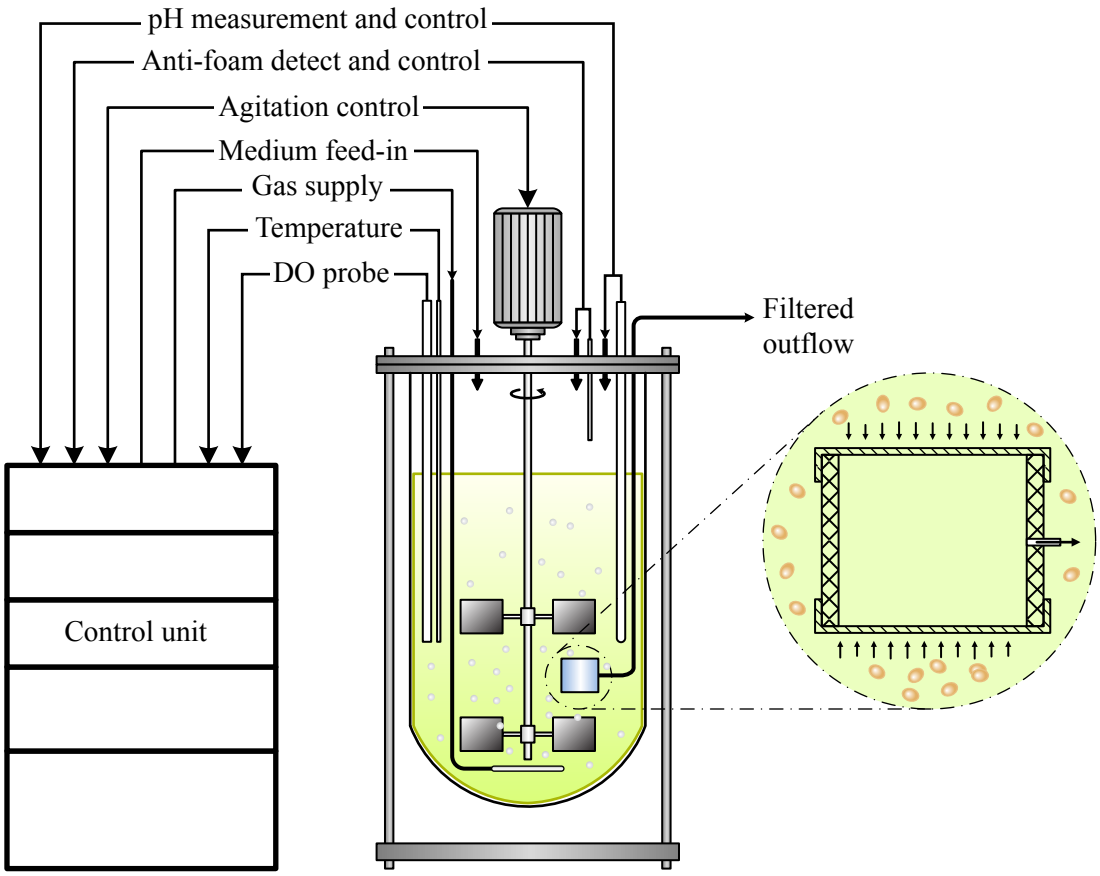

Figure 2.1: Reactor setup for "pseudo-continuous" fermentation.

Based on our experience, the biggest challenge in pseudo-continuous operation is maintaining the stability and effectiveness of cell filtration module. With the continuous withdrawal of fermentation broth, there will be a buildup of cells on the filtration surface if no action is taken to prevent it, which will decrease and eventually completely block the flow rate. To address this difficulty, we have developed a cell retention module in house, which

has two membrane surfaces (as shown in Figure 2.1). The cell retention module is installed to make the two surfaces located right against the propeller blades. In this way, the filtration surface can be constantly cleaned through shear stress resulted from agitation, which enabled us to achieve sustained pseudo-continuous fermentation for more than two months.

To evaluate the efficiency of the cell retention module, the cell retention ratio has been defined as

$$R = 1 - \frac{OD_{600(\text{efflux})}}{OD_{600(\text{bioreator})}} \tag{2.1}$$

where R is the cell retention ratio, $OD_{600(\text{efflux})}$ and $OD_{600(\text{bioreactor})}$ are the optical density measurements at 600 nm for efflux and broth in the bioreactor.

As normal continuous fermentation, the pseudo-continuous process can be separated into three phases after inoculation: phase I (cell growth), phase II (batch fermentation) and phase III (pseudo-continuous fermentation).

Phase I (cell growth) This phase is for *S. stipitis* to grow from initial inoculum. The pre-cultured *S. stipitis* cells were used to inoculate the bioreactor. The cells were cultivated till OD_{600} reached certain value with aeration rate of air at 1 vvm (volume volume per minute).

Phase II (batch fermentation) Phase II is for *S. stipitis* to transit from growth state to fermentation state. The temperature and agitation rate were kept the same as in Phase I, but the aeration rate of air was reduced. To maintain the total gas flow rate (therefore its impact on agitation) the same as in Phase I, nitrogen gas was introduced to make up the reduced air flow rate. The phase continued till the sugar (i.e., glucose or xylose) concentration fell below 1 g/L.

Phase III (pseudo-continuous fermentation) Within Phase III, the system transits from batch fermentation to pseudo-continuous fermentation. The feed medium defined in Section 2.1.1 was fed into the bioreactor and the broth was withdrawn from the reactor

through the cell retention module so that the cells were retained in the reactor. Both feed-in and withdrawal rates were kept at 0.2 mL/min to maintain a dilution rate of 0.008 h^{-1}. Other conditions except the aeration rate were the same as previous phases. In Phase III, various aeration rates were used to test the influence of oxygenation to the ethanol fermentation.

2.1.3 Chemical analytical procedures

The concentrations of glucose/xylose, xylitol, ethanol, glycerol and acetic acid were measured using an Agilent 1200 series high performance liquid chromatography (HPLC) with UV/Vis and IR detectors. They were analyzed on an Aminex HPX-87H column (Bio-Rad, Hercules, CA) at 45 °C with 0.05 M H_2SO_4 solution as the mobile phase at a flow rate of 0.6 mL/min. The time for each run was 25 min.

2.2 Fermentation of glucose with *S. stipitis*

To study the performance of *S. stipitis* with glucose as carbon source under various oxygenation conditions, a pseudo-continuous fermentation has been carried out.

For glucose fermentation, two stages have been applied in Phase III (pseudo-continuous fermentation). The only difference between different stages is the oxygenation condition, which has been shown in Table 2.1. In stage 1 (S1), oxygen-limited condition has been applied. Stage 2 is the strict anaerobic condition. The general performance of the fermentation is shown in Figure 2.2.

Table 2.1: Aeration conditions for glucose pseudo-continuous fermentation

Phases/Stages		**Air Supply Rate (mL/min)**	**Nitrogen Supply Rate (mL/min)**	**Duration (h)**
PI		1500	0	22.5
PII		30	1470	47.25
PIII	S1	30	1470	115.25
	S2	0	14	465

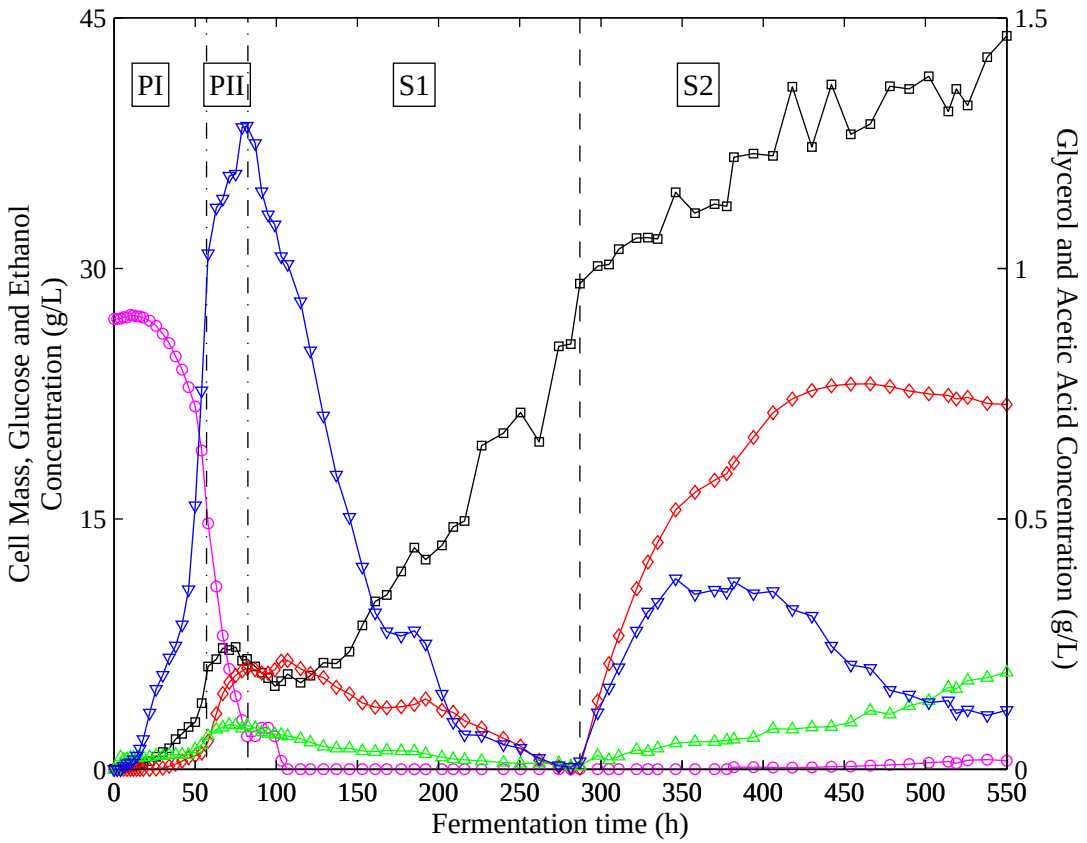

Figure 2.2: Performance of *S. stipitis* with glucose as carbon source under various oxygenation conditions. (open square, □) cell mass concentration, (open circle, ○) glucose concentration, (open diamond, ◇) ethanol concentration, (open triangle, △) glycerol concentration, (open inverted triangle, ▽) acetic acid concentration.

In PI when aeration was high, ethanol was produced when the measured DO dropped to zero (data not shown) and began to produce in a higher rate when the aeration condition switched to oxygen-limited condition. A relatively high concentration of acetic acid was accumulated in these two phases (up to 1.35 g/L). A small amount of glycerol was also produced. After the fermentation switched to oxygen-limited condition under "pseudo continuous" fermentation (S1), the concentrations of the products gradually dropped to zero due to the continuous flow-out and low production rates under carbon starvation (the glucose concentration dropped to zero). Under anaerobic condition (S2), the cells began to produce ethanol, acetic acid and glycerol at higher rates compared with S1.

2.3 Fermentation of xylose with *S. stipitis*

Xylose fermentation with *S. stipitis* has been reported to be oxygen-sensitive (Jeffries and Van Vleet, 2009). Therefore, in the pseudo-continuous fermentation of xylose, several oxygenation conditions have been tested in order to study its influence to ethanol production. The oxygenation conditions have been summarized in Table 2.2.

Table 2.2: Aeration conditions for xylose pseudo-continuous fermentation

Phases/Stages		**Air Supply Rate (mL/min)**	**Nitrogen Supply Rate (mL/min)**	**Duration (h)**
PI		1000	0	26
PII		20	70	6
PIII	S1	20	70	228
	S2	4.7	0	335
	S3	13.9	0	216
	S4	19.74	0	313
	S5	11.57	0	346

The general performance of the process is shown in Figure 2.3. With xylose as the carbon source, *S. stipitis* showed different behaviors for ethanol production with that shown in glucose metabolism and a much higher sensitivity to oxygen supply rate.

During the whole fermentation process, no glycerol production has been observed. Although the cells experienced xylose starvation at the first 500 hours, it didn't consume ethanol as the carbon source with oxygen supplied, which is proven by the increase of ethanol concentration in S1 and has also been reported. The oxygen supply has big influence on xylose metabolism. It influences the xylose uptake, cell growth, ethanol production, xylitol production as well as acetic acid production.

For different stages in pseudo-continuous fermentation, the oxygenation condition and various yields for cell mass, ethanol, xylitol as well as acetic acid have been summarized

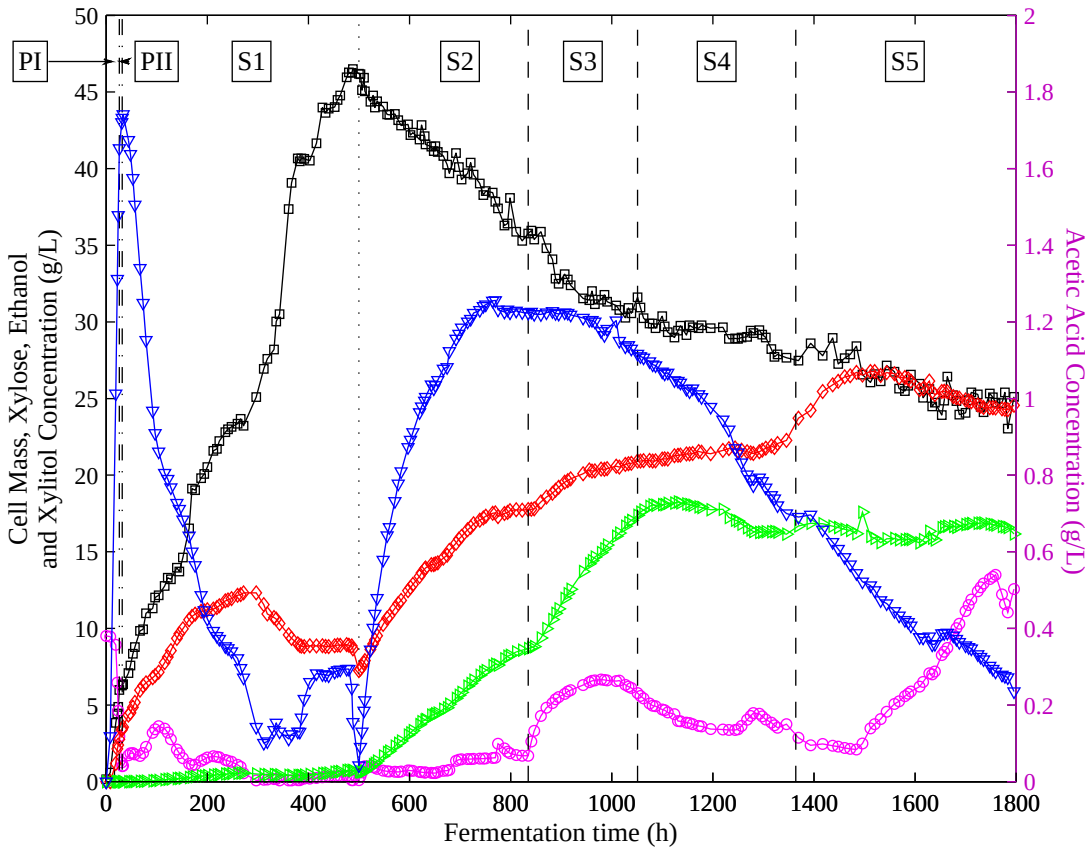

Figure 2.3: Performance of *S. stipitis* with glucose as carbon source under various oxygenation conditions. (open square, □) cell mass concentration, (open circle, ○) xylose concentration, (open diamond, ◇) ethanol concentration, (open right-pointing triangle, ▷) xylitol concentration, (open inverted triangle, ▽) acetic acid concentration.

in Table 2.3. From the yields, it showed even more clearer how the oxygenation influenced xylose metabolism. Another factor we needs to consider here is the adaptation of the strain to the ethanol with the fermentation going on for a long time, which will be discussed later in Chapter 3.

2.4 Conclusion

S. stipitis shows different phenotypes for glucose and xylose fermentation with minimal medium. They have different by-products patterns, sensitivities to oxygen conditions and optimal ethanol production conditions. For glucose, glycerol and acetic acid are the main

Table 2.3: Yields of various products under different oxygenation conditions for pseudo-continuous fermentation of xylose

Stages	**Specific growth rate**	**Ethanol yield**	**Xylitol yield**	**Acetic acid yield**
S1	0.0012	0.32	0.018	0.0030
S2	-0.00011	0.39	0.21	0.031
S3	-0.0013	0.33	0.30	0.017
S4	-0.0012	0.32	0.22	0.0075
S5	-0.00048	0.41	0.31	0.0059

detected by-products; while for xylose, xylitol and acetic acid are the main detected by-products. Compared to glucose fermentation with *S. stipitis*, xylose metabolism is very sensitive to oxygenation condition. The best ethanol production rate has been reached with anaerobic condition for glucose while for xylose metabolism oxygen is a much for cell growth and ethanol production with the minimal medium.

Chapter 3
Impact of pseudo-continuous fermentation on the ethanol tolerance

3.1 Abstract

In this work we conducted the pseudo-continuous fermentation, i.e., continuous fermentation with cell retention, using *Scheffersomyces stipitis*, and studied its effect on ethanol tolerance of the strain. During the fermentation experiments, *S. stipitis* was adapted to a mild concentration of ethanol (20-26 g/L) for two weeks. Two substrates (glucose and xylose) were used in different fermentation experiments. After fermentation, various experiments were performed to evaluate the ethanol tolerance of adapted cells and unadapted cells. Compared to the unadapted cells, the viability of adapted cells increased by 8 folds with glucose as the carbon source and 6 folds with xylose as the carbon source following exposure to 60 g/L ethanol for 2 h. Meanwhile, the ethanol limit concentration of cell growth increased 28% (glucose) and 32% (xylose) respectively. Improved ethanol tolerance of the adapted cells was also revealed in the effects of ethanol on plasma membrane permeability, extracellular alkalization and acidification. The mathematical modeling of cell leakage, extracellular alkalization and acidification revealed that cells cultured on glucose show better ethanol tolerance than cells cultured on xylose but the differences become smaller for adapted cells. The results show that pseudo-continuous fermentation can effectively improve cell's ethanol tolerance due to the environmental pressure during the fermentation process.

3.2 Introduction

In view of rising concerns over energy sustainability and global warning, the need for biofuels is expected to increase sharply in the coming years (Eisentraut, 2010; US Congress, 2007). At present, the U.S. biofuels market is dominated by corn-derived ethanol. However, substantial future growth of fuel ethanol will depend upon developments of cellulosic ethanol processes due to the following reasons: (i) further growth of corn ethanol is restrained by the availability of agricultural land, water resources, and the food vs. fuel trade-off which is already causing concern (Cherubini, 2010; Hoekman, 2009); (ii) cellulosic ethanol reduces both energy input and greenhouse gas emission by over 85% compared to corn ethanol (Chandel et al., 2007; Farrell et al., 2006); and (iii) cellulosic biomass is the most abundant and inexpensive renewable feedstock for ethanol production (Mansfield et al., 2006; Perlack et al., 2005).

Complete substrate utilization is one of the prerequisites for rendering lignocellulosic ethanol processes economically competitive. This means that both hexoses and pentoses in cellulose and hemicellulose must be converted to ethanol, and that microorganisms must be obtained to efficiently perform this conversion under industrial conditions. While ethanolic fermentation of hexoses derived from cellulosic biomass, i.e., glucose, mannose and galactose, using baker's yeast *Saccharomyces cerevisiae* is well established on large scale, the conversion of the pentoses (e.g., xylose and arabinose) to ethanol is still one of the major barriers of industrializing lignocellulosic ethanol processes.

To address the difficulties of pentose fermentation, vast majority of existing research has been focusing on genetically modifying a single strain so that it can ferment both hexose and pentose efficiently into ethanol. The genetic modifications usually include inserting pentose fermenting pathways and various sugar transporters, as well as making necessary adjustments to balance the cell's redox potential (Jeffries and Jin, 2004; Jeffries and Van Vleet, 2009). By introducing multiple changes to the host genome, numerous recombinant strains have been successfully developed in laboratories to ferment both glucose and xylose

simultaneously. Among them, the representative ones are the recombinant *S. cerevisiae* (424A) (Sedlak and Ho, 2004), engineered *Zymomonas mobilis* (CP4) (Rogers et al., 1982; Zhang et al., 1995), and engineered *Escherichia coli* (KO-11) (Ohta et al., 1991). However, despite the promising results in ethanol production from these recombinant strains, few of them have been realized in industrial applications due to issues related genetically stabilities, diauxic growth or other reasons (Hahn-Hägerdal et al., 2007).

In parallel to the genetic recombinant strategy, enhancing the innate capabilities of strains that can ferment both hexose and pentose into ethanol has been explored as well (Jeffries, 2008). Among all the native xylose fermenting strains, *Scheffersomyces stipitis* (*S. stipitis*, formerly known as *Pichia stipitis*) (Kurtzman and Suzuki, 2010) has been shown to be one of the most promising organisms for direct high-yield fermentation of xylose without significant by-product formation (Ferrari et al., 1992; Jeffries and Van Vleet, 2009; Jeffries and Shi, 1999). In addition, it naturally ferments both xylose and arabinose (the two major pentose contained in lignocellulosic material) into ethanol. In fact, *S. stipitis* has been a source of genes for engineering xylose metabolism in *S. cerevisiae*. However, several limitations prevent *S. stipitis* from being used to ferment mixed sugars at an industrial scale (McMillan, 1993). Among them, an important one is the low tolerance of *S. stipitis* to ethanol and other inhibitors such as acetic acid.

How ethanol affects yeast cells is very complicated. The reported effects include: influence on the cell membrane (changing the composition, structure, and function of the plasma and mitochondrial inner membrane) (D'Amore et al., 1990; Jeffries and Jin, 2000), influences on proteins (enzymes, transporters, signal protein, etc.) (Casey and Ingledew, 1986; Ma and Liu, 2010), inhibition of cell division (Mikami, Haseba, and Ohno, 1997), decrease of cell viability (D'Amore et al., 1990; Ma and Liu, 2010), and reduced metabolic activity (Liu and Qureshi, 2009; Ma and Liu, 2010), etc. The possible targets of ethanol in yeast cells are illustrated in Figure 3.1. The mechanisms of the tolerance of cells to ethanol are reported to be composed of many factors: lipid composition of cell membrane (Casey and Ingledew,

1986; D'Amore et al., 1990; Ding et al., 2009; Jeffries and Jin, 2000; Ma and Liu, 2010), amino acid compositions of membrane and protein (Ding et al., 2009; Ma and Liu, 2010), H^+-ATPase activity (Ding et al., 2009; Jeffries and Jin, 2000; Ma and Liu, 2010), factors that stabilize or repair denatured proteins (Ding et al., 2009; Jeffries and Jin, 2000), temperature (Casey and Ingledew, 1986; Jeffries and Jin, 2000; Preez, Bosch, and Prior, 1987; Van Uden, 1983), and different nutrients and ions (Birch and Walker, 2000; Ding et al., 2009; Furukawa et al., 2004). Correspondingly, different approaches have been proposed to improve cell's ethanol tolerance. Among them, adaptation has been shown to be effective. For example, Watanabe et al. (2011) reported that 20 cycles' batch adaptation can notably improve the ethanol tolerance of *S. stipitis*. In addition, *S. stipitis* can also be adapted to tolerate higher concentration of acetic acid (Mohandas, Whelan, and Panchal, 1995), hardwood hydrolysate (Nigam, 2001b) and rice straw hydrolysate (Huang et al., 2009) through batch fermentation process. However, there are a few limitations associated with the traditional batch adaptation approach. First, the commonly used batch adaptation method may take a long time to reach the desired improvement. Second, it has been suggested that endogenous ethanol (i.e., ethanol generated within the cells) is more toxic to the cells (D'Amore et al., 1990). Therefore, exogenous ethanol (i.e., ethanol added externally) used in batch adaptation may be less effective for adaptation. Finally, if the improved ethanol tolerance through adaptation is not caused by mutation, the adapted strain may lose its acquired capability when the environmental pressure is removed. At the same time, ethanol tolerance of yeast strains can be enhanced by cell immobilization (Ciesarová et al., 1998; Desimone et al., 2002; Jirku, 1999; Krisch and Szajáni, 1997), and it is argued that the improved ethanol tolerance is due to the enhanced hydration layer stability resulted from attaching cells to the carrier. However, because the cells maintain their enhanced ethanol tolerance after being removed from the carrier (Zhou, Martin, and Pamment, 2008), the ethanol tolerance enhancement associated with cell immobilization may be due to adaptation caused by endogenous ethanol. Due to the mass transfer resistance introduced by the carrier, the ethanol concentration in

the small vicinity of the immobilized cells is higher than the bulk concentration. Compared to the free-floating cells, the higher ethanol concentration around the carrier exerts environmental pressure to the immobilized cells. Consequently, during the course of fermentation process, the immobilized cells gradually adapt to the higher ethanol concentration and show improved ethanol tolerance.

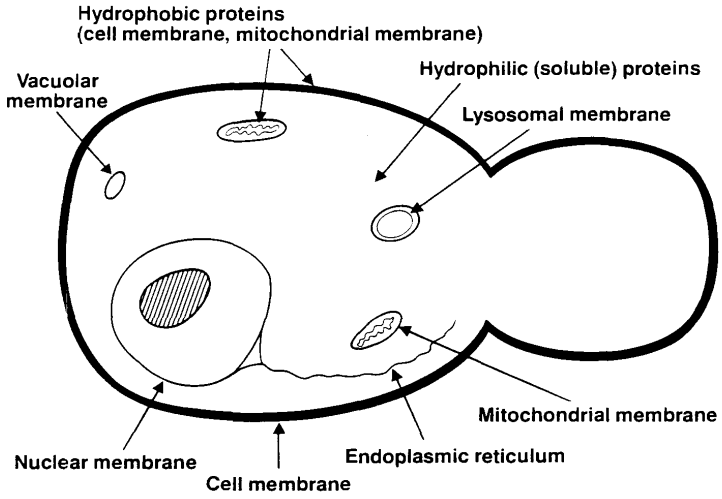

Figure 3.1: Possible targets of ethanol in yeast cells (D'Amore et al., 1990)

In this work, we propose to ferment different sugars (glucose and xylose) using *S. stipitis* under a pseudo-continuous operation, characterized by continuous nutrient feeding and continuous cell-free broth withdrawal. We hypothesize that pseudo-continuous fermentation provides an ideal environment for the cells to adapt to higher ethanol tolerance while they produce ethanol. In this work, we will verify this hypothesis by performing fermentation experiments using a modified New Brunswick Bioflo 110 fermentor. By installing an in-house developed cell retention module in the Bioflo 110 fermentor, we were able to perform pseudo-continuous operation. With adapted cells obtained through pseudo-continuous operation, we further examine some of their properties to validate our claim.

3.3 Materials and methods

3.3.1 Microorganism, media, culture condition and chemical analysis procedure

Microorganism, media, culture condition and chemical analysis procedure have been described in details in Chapter 2.

3.3.2 Ethanol tolerance evaluation

As mentioned in Introduction, the cellular response of yeasts to ethanol stress is very complex ((Ma and Liu, 2010) and references cited therein). Therefore, many approaches have been proposed to evaluate ethanol tolerance of different strains. In this work, we adopt the following criteria to evaluate the ethanol tolerance of adapted and unadapted strains: (i) cell viability after ethanol shock; (ii) growth limitation concentration of ethanol; (iii) ethanol induced membrane leakage; (iv) extracellular alkalization and acidification. Cells in different physiological states may show different levels of ethanol tolerance (Slininger, Gorsich, and Liu, 2009). Here the unadapted strain means *S. stipitis* CBS6054 cells cultured directly from minimal medium. To eliminate the potential impact of different physiological states on cells' ethanol tolerance, before conducting the following evaluation experiments, both adapted and unadapted cells were cultured in the pre-culture medium for 24 hours and the cells were harvested in mid-exponential growth phase.

3.3.2.1 Cell viability after ethanol shock

The rate of viability loss in the presence of ethanol has been used as a means to assess ethanol tolerance in a number of different yeast strains as well as to measure the influences of nutritional and environmental conditions on ethanol tolerance. Losses in cell viability found in high-gravity brewery fermentations have been attributed to the killing effects of ethanol and it has been suggested that resistance to killing by ethanol may be related to ethanol

tolerance (Casey and Ingledew, 1986). Compared with the other definitions of ethanol tolerance, this might show the capability of cells to survive in high ethanol concentration and therefore be more suitable to describe the performance in continuous fermentation.

In this method, the ethanol tolerance of cells was evaluated by comparing their survival rates after two hours exposure to fermentation media containing 15, 30, 45, and 60 g/L ethanol following a modified procedure of Zhou, Martin, and Pamment (2008). In this work, the unadapted cells means *S. stipitis* CBS 6054 cells cultured directly from minimal medium. To eliminate the potential impact of different physiological states on cells' ethanol tolerance, before conducting this and all other ethanol tolerance evaluation experiments, both adapted and unadapted cells were cultured in the pre-culture medium containing glucose or xylose for 24 hours and the cells were harvested in mid-exponential growth phase. Cells were collected by centrifugation, washed twice with fresh fermentation medium, re-suspended and diluted with fresh fermentation medium so that the OD_600 of the resulted suspension is about 7.0. 0.4 mL of the suspension was then added to 3.6 mL of each of the control and ethanol-containing treatment solutions. After two hours incubation at 30 °C, the cells were harvested and washed, re-suspended in fresh fermentation medium. The total viable numbers of cells were determined by plate counts (30 °C, 28 hours for cells grown on glucose and 36 hours for cells grown on xylose due to slower growth on xylose). The ethanol tolerance was calculated as the number of viable cells that survived exposure to ethanol for two hours as a percentage of those that survived exposure to the ethanol-free solution for the same period. Three replicates were conducted for each experiment and then they were pooled to obtain average ethanol tolerance measures.

3.3.2.2 Graphical determination of the ethanol limitation to growth

This is one of the most widely used methods to define ethanol tolerance. It is usually defined as the concentration of ethanol which will completely suppress batch growth. This

could be measured by adding different concentration of ethanol into the medium and measuring the growth rate. In this work the theoretical ethanol concentration to inhibit cell growth (E_m), i.e. the ethanol concentration above which cells do not grow, is graphically determined following nonlinear kinetics of ethanol inhibition to cell growth proposed by Luong (1985):

$$\mu_i/\mu_0 = 1 - (E/E_m)^\alpha \tag{3.1}$$

where μ_i is the maximum specific growth rate in the presence of ethanol, μ_0 the maximum specific growth rate at zero initial ethanol concentration, E the initial ethanol concentration, and α a dimensionless constant. By rearranging Eqn. 3.1 we obtain

$$\ln(1 - \mu_i/\mu_0) = \alpha \ln E - \alpha \ln E_m \tag{3.2}$$

so that a plot of $\ln(1 - \mu_i/\mu_0)$ vs. $\ln E$ has a slope equal to α and an intercept equal to $-\alpha \ln E_m$.

The specific growth rates were measured based on batch culture experiments. The adapted and unadapted cells were inoculated into flasks as described in the previous section respectively. The optical density of the culture (OD_600) was monitored after inoculation. The specific growth rate was calculated by linear regression of $\ln \mathrm{OD}_{600}$ versus time to find the slope during logarithmic growth phase based on the following equation

$$\mu = \frac{1}{X} \cdot \frac{\mathrm{d}X}{\mathrm{d}t} = \frac{1}{\mathrm{OD}_{600}} \cdot \frac{\mathrm{dOD}_{600}}{\mathrm{d}t} = \frac{\mathrm{d}\ln(\mathrm{OD}_{600})}{\mathrm{d}t} \tag{3.3}$$

where X is the cell density (g/L).

3.3.2.3 Ethanol induced leakage of 260-nm-light-absorbing compounds

Following the procedure described by Salgueiro, Sá-Correia, and Novais (1988), the adapted and unadapted cells were washed twice with sterile phosphate buffer (50 mM, pH

5.0), and suspended in sterile buffer plus different ethanol concentrations (0, 5, 10, 15, 20 and 25 g/L) contained in conical flasks closed with a rubber bung. OD_{600} of these cellular suspensions was adjusted to around 3. The suspensions were shaken slowly (40 rpm) at 30 °C. Samples (1 mL) taken every half an hour were centrifuged for 5 min and the supernatants were immediately examined for 260-nm-light-absorbing compounds in a UV/Vis spectrophotometer at 260 nm.

To quantitatively compare the plasma membrane permeability of intracellular metabolites induced by ethanol for both adapted and unadapted strains, we developed a mathematical model to describe the dynamics of ethanol induced leakage of 260-nm-light-absorbing compounds as shown in Eqn. 3.4. The detailed model derivation is given in .

$$\mathrm{OD}_{260} = K_L \left[1 - \exp\left(\frac{d_L - t}{\tau_L} \right) \right] \tag{3.4}$$

where t denotes time, K_L, d_L, and τ_L are the model parameters. K_L represents the ultimate effect of a given ethanol concentration on the leakage of 260-nm compounds, which is the maximum or steady state OD_{600} observed for a given ethanol concentration when t is large. d_L represents the onset of the effect after putting the cells in contact with ethanol. τ_L indicates how fast the leakage will reach its maximum.

3.3.2.4 Extracellular alkalization and acidification

The extracellular alkalization and acidification experiments were carried out based on the procedures published by Meyrial et al. (1997).

In the extracellular alkalization experiments, cells were collected by centrifuge, washed twice and re-suspended in a solution of NaCl (0.9%) to a solution with $OD_{600} = 16$. The re-suspended cells were incubated at 30 °C for 2.5 h on a rotary shaker (50 rpm), to deplete the cells of energy and, thus, deactivate the plasma membrane ATPase. In parallel, 125 mL flasks containing increasing ethanol concentrations (0, 15, 30, 45, 60, 90, and 120 g/L) in a total volume of 54 mL, adjusted by addition of a solution of NaCl (0.9%), were equipped

with a magnetic stirrer and incubated at 30 °C. The initial pH of the assay mixture was adjusted to 4.00 with 0.1 M HCl before the reaction was started by the addition of the cell suspension (6 mL). The extracellular pH was monitored with a digital pH meter accurate to two decimal places for 10 min or till the pH value reached steady state. The recorded pH values are then converted to proton concentration ($[H^+]$) in M. Following a similar modeling approach as in ethanol induced leakage, we derive a first-order dynamic model to describe the evolution of $[H^+]$ with time, as shown below.

$$[H^+] = [H^+]_0 + K_p \left[1 - \exp\left(\frac{d_p - t}{\tau_p}\right)\right] \tag{3.5}$$

where the definitions of model parameters K_p, d_p, and τ_p are similar to those in Eqn. 3.4. $[H^+]_0$ is a constant representing initial $[H^+]$. The time series $[H^+]$ values are fitted to Eqn. 3.5 to estimate the maximum net proton influx, which is calculated by K_p/τ_p in μmol/L/min then convered to nmol/mg/min by dividing the dry cell weight concentration of 9.9824 g/L for the broth. All experiments were independently performed three times.

The extracellular acidification experiments were carried out with the starved cells prepared as described above. Compared with the alkalization procedure, the solution in acidification was prepared with the addition of glucose/xylose (10 g/L). The initial pH of the solution was adjusted to 4.50 with 0.1 M HCl before the reaction was started by adding 6 mL suspension of starved cells ($OD_{600} = 16$). The external pH was measured with a digital pH meter accurate to two decimal places for every 1 minute till 30 minutes or till the pH value reached steady state. All experiments were independently carried out three times.

3.4 Results and discussion

3.4.1 General results of the continuous fermentation with cell retention and adaptation

Two fermentation experiments were carried out in the modified bioreactor with cell retention module, using glucose and xylose as the carbon source respectively. The overall processes of the two fermentation experiments are shown in Figure 2.2 and Figure 2.3. The cell retention ratio was above 99% throughout the experiments.

The detailed general performance of the pseudo-continuous fermentation of glucose and xylose with *S. stipitis* has been discussed in Chapter 2. For the experiments, if we consider the cells entered adaptation when the ethanol concentration was higher than 20 g/L, the cells have been adapted to mild endogenous ethanol environment for more than two weeks before sampled for further evaluation. The adaptation time is summarized in Table 3.1.

Table 3.1: Adaptation process summary

Carbon Source	**Adaptation Ethanol Concentration (g/L)**	**Adaptatiion Duration**
Glucose	20 – 23	420 h
Xylose	20 – 26	425 h

There are several advantages associated with the pseudo-continuous fermentation. First, the pseudo-continuous fermentation can help improve the fermentation rate, as very high cell density can be easily obtained with cell retention, which will increase the ethanol throughput. Second, it also completely eliminates the potential washout, which allows us to operate and study the fermentation under various conditions, such as different dilution rates and oxygen supply rates. Note that for traditional continuous fermentation of xylose using *S. stipitis*, one major difficulty is the frequent washout due to the slow cell growth, or even negative growth rate as shown in our experiment, under micro-aerobic condition (Rizzi et al., 1989; Shi and Jeffries, 1998). In other words, if the cell growth rate is lower than the rate of cells

being carried out from the bioreactor, the cell mass within the bioreactor would eventually drop to zero, or being washed out. Finally, because the adaptation is provided by the ethanol produced by the cells, it is assured that the environment pressure always exists. Therefore, whether or not the improved ethanol tolerance is achieved through mutation, there is no risk of losing the obtained capability during the fermentation process.

3.4.2 Cell viability under ethanol shock

Ethanol tolerance tests were performed on both unadapted and adapted cells with glucose and xylose as carbon sources. Cell survival rates after two hours exposure to fermentation medium containing 15, 30, 45, and 60 g/L ethanol are summarized and compared in Figure 3.2. It should be pointed out that both adapted and unadapted cells were harvested in the mid-exponential growth phase in order to eliminate possible effect of different physiological states on cells' ethanol tolerance. Figure 3.2 shows that the ethanol tolerance of *S. stipitis* CBS 6054 cells was noticeably improved by adaptation during the pseudo-continuous fermentation. The cell viability difference between the unadapted and adapted cells increases as ethanol concentration increases. In the case of 60 g/L ethanol shock, the survival rates increased by 8 folds for the adapted cells cultured on glucose, and 6 folds on xylose. In terms of the effect of the carbon source on adaptation, for all ethanol concentrations, cells with glucose as carbon source, either unadapted or adapted, showed a slightly higher ethanol tolerance than the cells with xylose as carbon source.

3.4.3 Ethanol limitation to growth

The ethanol limitations (E_m) to cell growth for different cells have been determined by Eqn. 3.2. The results are shown in Figure 3.3.

Compared with the cells grown on glucose, the cells grown on xylose always showed a lower ethanol limit concentration for both adapted and unadapted cells. For the unadapted cells, the cells grown on glucose has an ethanol limitation concentration of 57.29 g/L while

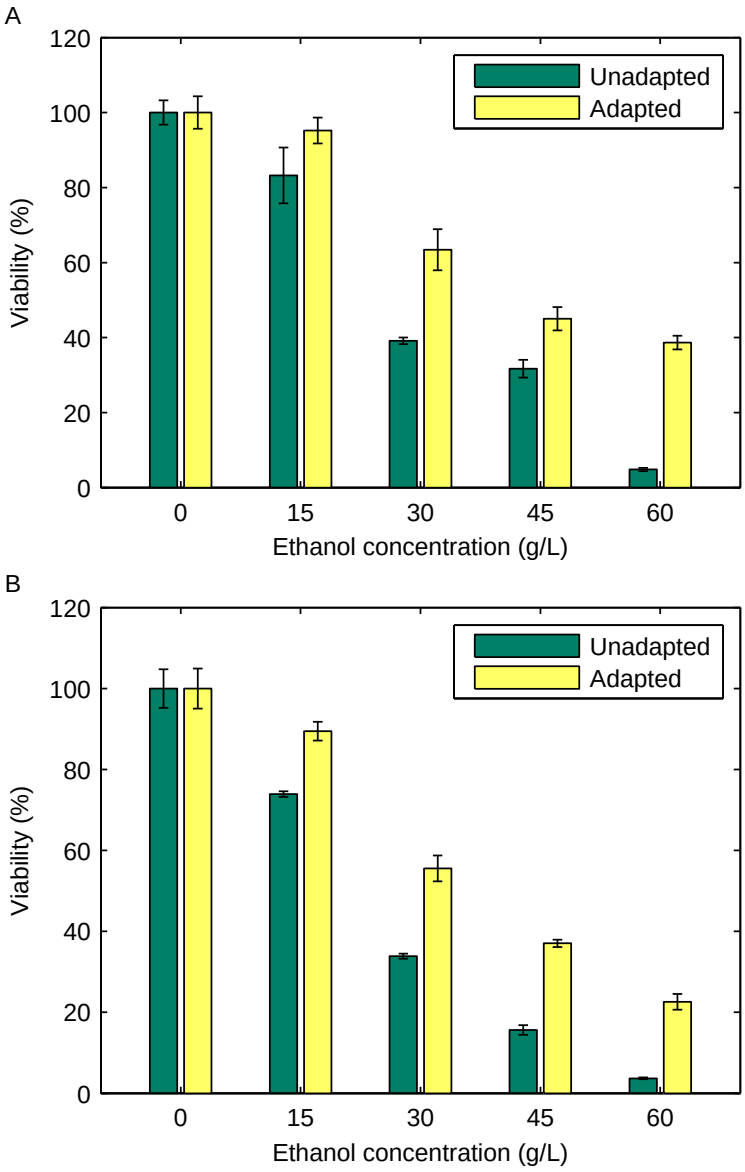

Figure 3.2: Viability of adapted and unadapted cells under ethanol shock. A: glucose as the carbon source; B: xylose as the carbon source. The error bar represents one standard deviation ($\pm\sigma$) estimated based on 3 independent measurement.

the cells grown on xylose is 50.24 g/L, which is about 7 g/L lower. After the adaptation, no matter what the carbon source is, the ethanol limit concentration increased. The ethanol limitation concentration of adapted cells grown on glucose increased from 57.29 g/L to 73.40 g/L, which corresponded to a 28.12% increase. The ethanol limitation concentration of adapted cells grown on xylose increased from 50.24 g/L to 66.38 g/L, corresponded to a

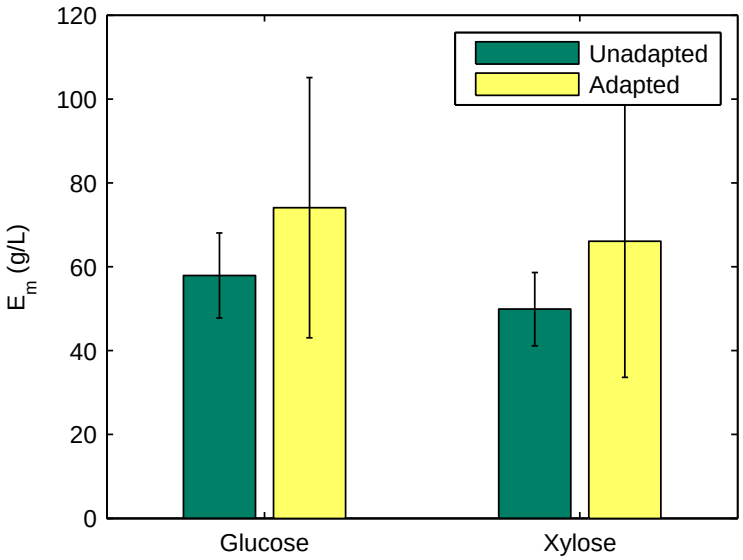

Figure 3.3: Limited ethanol concentration for growth of *S. stipitis*. The results are calculated from Eqn. 3.2: (red) unadapted cells, (cyan) adapted cells. The results are grouped based on the carbon source.

32.13% increase although it is still about 7 g/L lower than the adapted cells grown on glucose. For both cases, the increased upper ethanol limitation concentration further confirms the increased ethanol tolerance of adapted cells.

3.4.4 Ethanol induced leakage of 260-nm-light-absorbing compounds

Ethanol could induce the leakage of 260-nm-light-absorbing compounds in cells, such as purine, pyrimidine bases and nucleotides (LEE and LEWIS, 1968). This can be used to character the resistance of the plasma membrane to membrane permeabilization by ethanol, which is reported related to the ethanol tolerance of the cells (Jirku, 1999; Salgueiro, Sá-Correia, and Novais, 1988).

The adapted cells were tested for their plasma membrane permeability of 260-nm-light-absorbing compounds and compared with the unadapted cells grown on glucose and xylose respectively. The results are shown in Figure 3.4, where the adapted cells show a much slower 260-nm-light-absorbing compounds leakage compared to the unadapted cells no matter which carbon source was used.

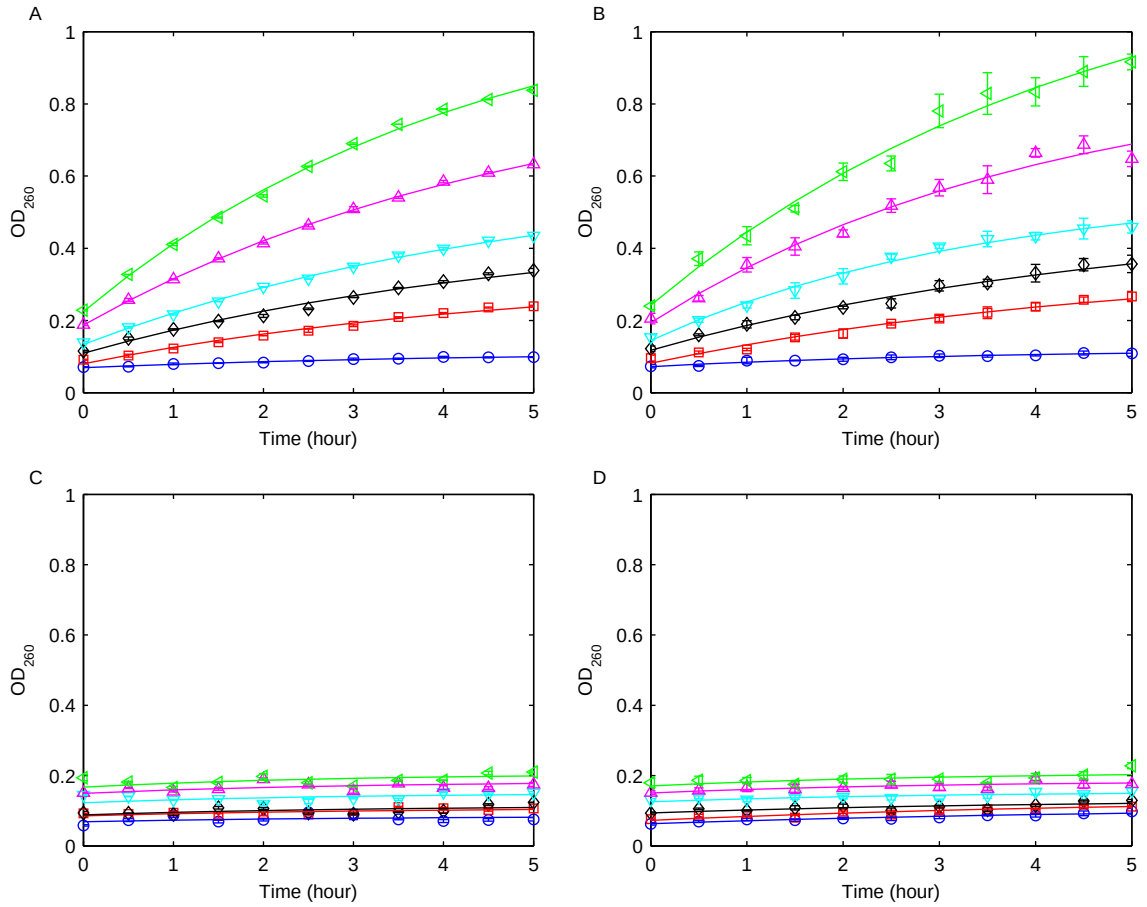

Figure 3.4: Comparison of 260-nm-light-absorbing leakage among cells under different ethanol concentration (g/L). 0 (open circle), 5(open square), 10 (open diamond), 15 (open inverted triangle), 20 (open triangle), 25 (open left-pointing triangle). A: unadapted cells grown on glucose; B: unadapted cells grown on xylose; C: adapted cells grown on glucose; D: adapted cells grown on xylose. Solid lines are model fittings.

The nonlinear least squares fittings of the measurements OD_{260} to Eqn. 3.4 show well agreement with the experimental data as shown in Figure 3.4. The maximum leakage rate (K_L/τ_L), in absorbance units per hour or AU/h, estimated based on the fittings are shown in Figure 3.5. The maximum leakage rate is exponentially correlated with ethanol concentration for all four cases studied and the dash and solid lines are the nonlinear least squares fittings of exponential functions. The effect of adaptation on cells' ethanol tolerance is clearly shown in Figure 3.5: although the maximum leakage rate increases exponentially with ethanol concentration for both adapted and unadapted cells, the increases in the adapted cells are

dramatically slower than that of the unadapted ones. Specifically, the maximum leakage rate induced by 25 g/L ethanol decreased by 73.71% and 74.72% respectively after adaptation for glucose and xylose grown cells. In addition, Figure 3.5 shows the differences on ethanol tolerance between glucose and xylose grown cells become much smaller after adaptation.

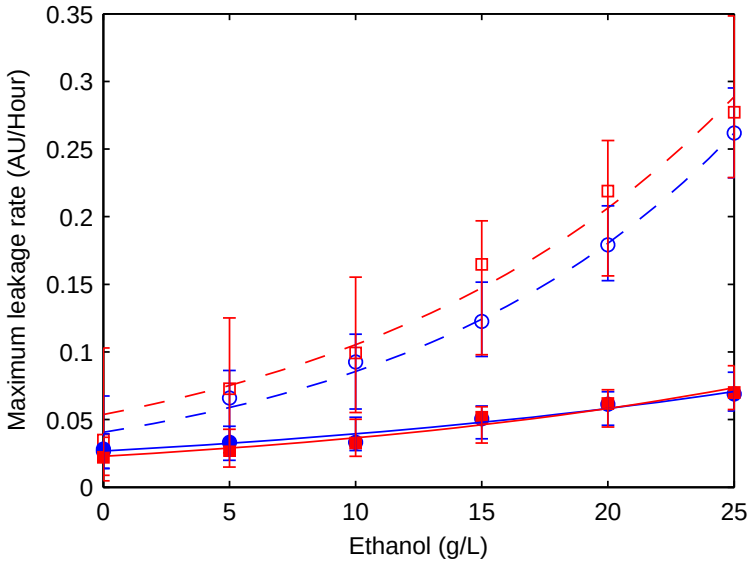

Figure 3.5: Comparison of maximum leakage rate under different ethanol concentration for unadapted and adapted cells. (open circle) unadapted cells grown on glucose, (open square) unadapted cells grown on xylose, (filled circle) adapted cells grown on glucose, (filled square) adapted cells grown on xylose.

3.4.5 Extracellular alkalization

In this experiment, the de-energized cell suspension of *S. stipitis* is added to the solution with different concentration of ethanol and the extracellular pH is measured for 10 minutes. The recorded pH values are then converted to proton concentration ($[H^+]$) in μM. Under the experimental condition (i.e. de-energized cells, no sugar available), passive proton influx would be the dominant mechanism for proton transport and extracellular alkalization is expected. The evolution of $[H^+]$ over time for unadapted and adapted cells with different ethanol concentrations shows extracellular alkalization. In addition, we observe that after adaptation the cells show much slower extracellular alkalization, which indicates that the

adapted cells show a higher resistance to ethanol. The reduced alkalization was observed for adapted cells grown on both glucose and xylose, although with slight difference in the changing magnitude.

The maximum net proton influxes are plotted in Figure 3.6. It is shown that the maximum net proton influx increases with ethanol concentration for both unadapted and adapted cells. However, the adapted cells show much slower increase, as indicated by the much smaller slopes compared to those of the unadapted cells. Specifically, with the exposure to 120 g/L ethanol, the maximum net proton influxes of cells grown on glucose and xylose decreased to 79.67% and 80.44% respectively after adaptation. In addition, it is worth noting that within each group (adapted or unadapted), the xylose grown cells always have higher maximum net proton influx than the glucose grown cells. This is consistent with the results reported by Meyrial et al. (1997).

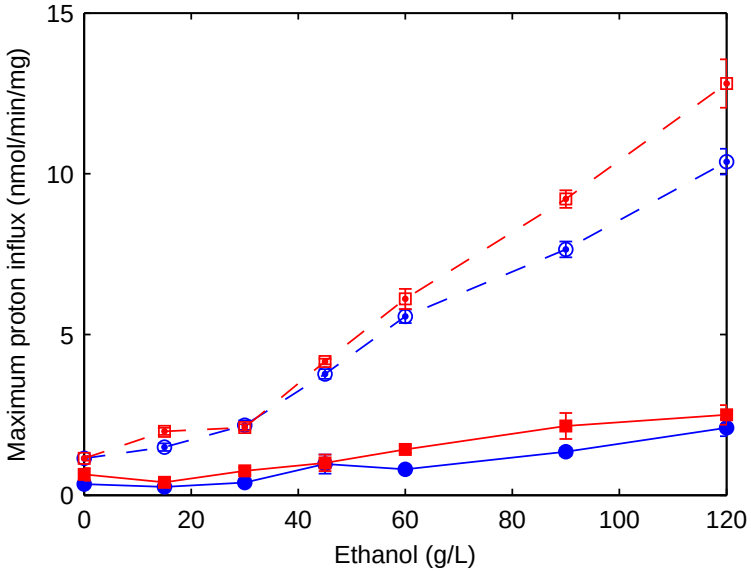

Figure 3.6: Effect of ethanol on maximum net proton influx for unadapted and adapted cells. (open circle) unadapted cells grown on glucose, (open square) unadapted cells grown on xylose, (filled circle) adapted cells grown on glucose, (filled square) adapted cells grown on xylose.

3.4.6 Extracellular acidification

In this experiment, de-energized cells are added to the solution with different concentration of ethanol plus 10 g/L of sugar (glucose or xylose) and the extracellular pH is measured over 40 minutes. Due to the presence of sugar, any of the three mechanisms of proton transport could play a significant role. In our experiments, extracellular acidification is observed, which indicates that the active proton efflux through proton pump outpaces the proton influx via sugar symport and passive influx. Such extracellular acidification was also observed and reported by Jiménez and Uden (1985). In our experiments it also shows that the acidification rate decreases with the increase of ethanol concentration, probably due to the ethanol inhibited ATPase activity and ethanol induced plasma membrane leakage. Also, compared to the unadapted cells, the adapted cells showed a faster acidification rate, which can be observed at higher ethanol concentration. This indicates that the adapted cells might have a higher ATPase activity and/or a lower permeability to proton in the presence of ethanol. Again, xylose grown cells showed a higher ethanol influence compared to glucose grown cells in both adapted and unadapted groups.

The $[H^+]$ changes during extracellular acidification are fitted to the same first-order dynamic model in extracellular alkalization, i.e. Eqn. 3.5. Based on the first-order dynamic models, the maximum net proton efflux rates were estimated and plotted in Figure 3.7. From Figure 3.7, we observe that adapted cells showed higher maximum net proton efflux compared to unadapted cells in each medium group. Specifically with 120 g/L of ethanol in presents, the maximum net proton efflux increased by 2.67 folds and 3.9 folds respectively for cells grown on glucose and xylose after adaptation. In addition, the glucose grown cells showed a higher maximum net proton efflux rate compared to the xylose grown cells in the same adapted or unadapted group. With increasing ethanol concentration, the difference in the maximum net proton efflux rate of glucose grown and xylose grown cells become smaller. This indicates that at high ethanol concentrations, due to strong effect of ethanol to ATPase

activity and plasma membrane permeability, the influence of carbon source to the apparent proton extrusion rates diminishes.

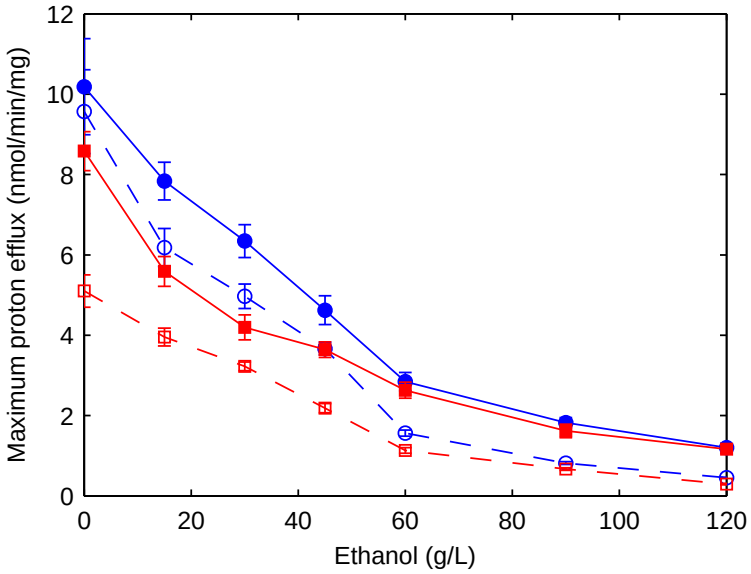

Figure 3.7: Effect of ethanol on proton extrusion rate in cells: (open circle) unadapted cells grown on glucose, (open square) unadapted cells grown on xylose, (filled circle) adapted cells grown on glucose, (filled square) adapted cells grown on xylose.

3.4.7 Remarks

It is worth noting that pseudo-continuous fermentation was proposed as an effective way to conduct high throughput fermentation for slow growing (or non-growing) cells, not intended for strain development. Therefore, for all experiments, we did not isolate any single colony of the adapted strain. Instead, we collected some fermentation broth from the bioreactor, and cultured all cells together for comparison experiments. In addition, the adapted samples were collected after about 400 hours of mild adaptation (i.e., with ethanol concentration higher than 20 g/L), which is usually too short for any ethanol-tolerant mutant to be isolated. To confirm this claim, we also cultured the adapted cells under non-stressed condition. We observed that after 3 rounds of batch culture, the adapted cells showed almost no difference from the unadapted cells in terms of cell viability after 2 h ethanol shock. Therefore, we believe that the improved ethanol tolerance is mainly due to cell's physiological

response to the environmental pressure presented during pseudo-continuous fermentation, instead of genetic changes. It is also worth noting that the final effect to ethanol tolerance is a combination of different factors such as fermentation conditions, fermentation time and ethanol concentration, as well as carbon sources. When comparing the effects of carbon sources on ethanol tolerance, because other conditions (i.e., aeration condition, fermentation duration, ethanol concentration, etc.) were not the same, and in fact some of them cannot be controlled to be the same (e.g., aeration conditions in order to produce ethanol), all the results and discussions on carbon source effect are qualitative, or semi-quantitative at best.

3.5 Conclusion

Continuous fermentation with cell retention offers an effective means for cell adaptation. In our experiment it shows that the adaptation by continuous fermentation with cell retention has been proved to be an effective method of improving ethanol tolerance of *S. stipitis*, which was one of the biggest obstacle for the cells to be applied in industry. Ethanol tolerance of *S. stipitis* has been significantly enhanced. We used different ways to characterize the features of the adapted cells. The viability under 2 h 60 g/L ethanol shock increased by 8 folds on glucose and 6 folds on xylose respectively. At the same time, the 260-nm-light-absorbing compounds leakage as well as the passive proton influx and active proton extrusion showed the plasma membrane compositions and plasma ATPase activities changed significantly after the adaptation, which means that during the adaptation the cells might change the metabolism a lot. The ethanol limitation concentration calculated from the specific growth rate increased 28% (on glucose) and 32% (on xylose) respectively. Future work to measure the change of plasma compositions and intracellular metabolites may well lead to additional gains in tolerance stability.

Chapter 4
Reconstruction and validation of central carbon metabolic network model

S. stipitis, as previously mentioned at Section 1.4, is an important strain for xylose metabolism. This chapter reports the reconstruction and validation of a central carbon network model of *S. stipitis* metabolism. We begin with the draft reconstruction of the metabolic network model, followed by the refinement of the model. In the end the model is validated qualitatively and quantitatively.

4.1 Metabolic network model reconstruction

Follow the construction procedure by Orth, Thiele, and Palsson (2010) and Thiele and Palsson (2010) as discussed in 1.6.1.1, the first stage of reconstruction of metabolic network is to draft the model based on biochemistry textbooks (Berg, Tymoczko, and Stryer, 2006; Voet and Voet, 2010; Voet, Voet, and Pratt, 2008) and network databases of genome and biochemistry.

4.1.1 Draft construction of the reaction list

The very first step of reconstruction is to selection proper reactions for the model. This has been done by information from biochemistry textbooks, literatures, and online databases. The textbooks provide the basic structure of the central carbon metabolism. The essential pathways includes:

1. glycolysis/gluconeogenesis (Voet and Voet, 2010; Voet, Voet, and Pratt, 2008);

2. pentose phosphate pathway and xylose metabolism (Hahn-Hägerdal et al., 1994a,b; Jeffries and Van Vleet, 2009; Jeffries et al., 2007; Jeppsson, Alexander, and Hahn-Hagerdal, 1995; Verduyn et al., 1985; Voet and Voet, 2010; Voet, Voet, and Pratt, 2008);

3. pentose and glucuronate interconversions (Jeffries and Van Vleet, 2009; Jeffries et al., 2007; Voet and Voet, 2010; Voet, Voet, and Pratt, 2008);

4. glycerolipid metabolism (Costenoble et al., 2007; Jeffries and Van Vleet, 2009);

5. pyruvate metabolism (Agbogbo and Wenger, 2006; Jeffries and Van Vleet, 2009; Nigam, 2002; Parekh, Yu, and Wayman, 1986);

6. TCA cycle (Jeffries and Van Vleet, 2009; Voet and Voet, 2010);

7. oxidative phosphorylation (Jeffries and Van Vleet, 2009; Klinner et al., 2005; Preez, 1994; Shi et al., 1999; Shi et al., 2002);

8. glutamate metabolism;

9. glyoxylate and dicarboxylate metabolism;

10. nicotinate and nicotinamide metabolism (Jeffries and Van Vleet, 2009; Voet and Voet, 2010);

11. transport (Boles and Hollenberg, 1997; Jeffries and Van Vleet, 2009; Kilian and Uden, 1988; Ligthelm, Prior, and Preez, 1988a; Voet and Voet, 2010).

Particularly, the xylose metabolism and oxidative phosphorylation are described in details here.

Xylose metabolism The first three steps of xylose metabolism with *S. stipitis* has been shown in Figure 4.1. As mentioned in Chapter 1, xylose reductase of *S. stipitis* has dual cofactor preferences, i.e. it can utilize NADH and NADPH at the same time. For xylitol, it is

considered to be NAD^+-dependent only. However, reports on its preference to $NADP^+$ also exist. The details will be discussed in Section 4.2.4. Therefore, in the model reconstruction, the four reactions have been all been included.

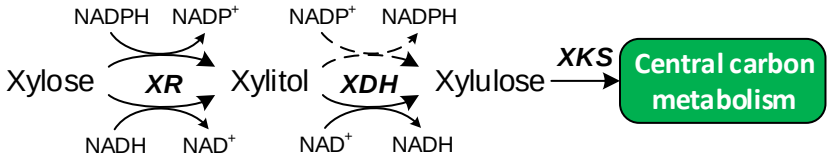

Figure 4.1: Illustration of xylose metabolism in *S. stipitis*. XR: xyluse reductase; XDH: xylitol dehydrogenase; XKS: xylulose kinase. The dash line indicates that there's debate with the existence of the reaction.

Oxidative phosphorylation In the electron transport chain (ETC) of *S. stipitis*, alternative compounds in addition to the standard respiratory systems exist (Shi et al., 2002 and the references therein). Upstream of the ETC, rotenone-insensitive NAD(P)H dehydrogenases (non-proton-translocating) are present in addition to the rotenone-sensitive NADH dehydrogenase Complex I (proton-translocating). Downstream, an alternative terminal oxidase that is sensitive to salicylhydroxamic acid (SHAM) is present in addition to the standard cytochrome c oxidase (Cox). The overview of the system is shown in Figure 4.2. The corresponding reactions have been constructed in the model.

4.1.2 Charge- and element-balancing the model

After selecting reactions of the central carbon metabolism, online biochemistry, genome databases are used to assist and facilitate the reconstruction. Some free and commercial data sources are already available Thiele and Palsson, 2010. In this work, the most involved databases are: *S. stipitis* genome database (Pichia stipitis v2.0) (Jeffries et al., 2007), KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2010, 2012), ChEBI (Degtyarenko et al., 2008, 2009; Matos et al., 2010) and PubChem Compound(Bolton et al., 2008). The KEGG and *S. stipitis* genome database are used to verify the existence of the reactions from gene

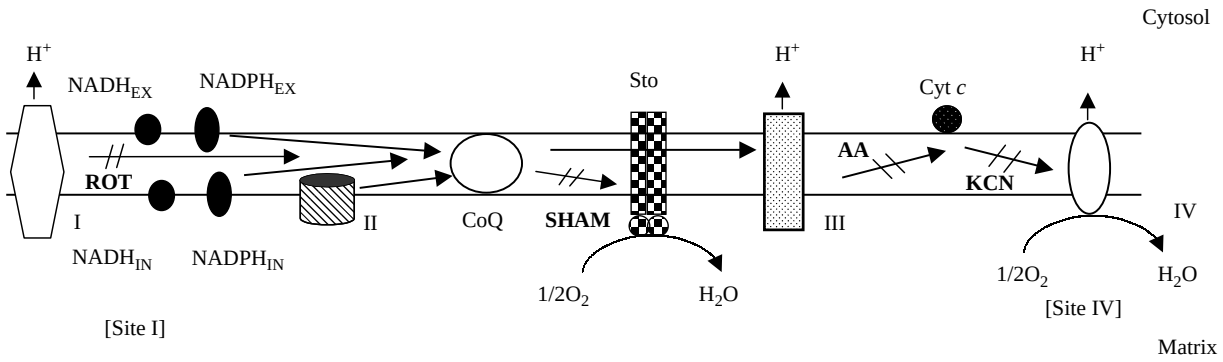

Figure 4.2: A diagram of alternative and standard redox components present in the electron transport chain (ETC) of *S. stipitis*. The action site of SHAM, antimycin A (AA), rotenone (ROT) and KCN are marked. I, rotenone-sensitive NADH dehydrogenase complex (Complex I); $NADH_{IN}$, rotenone-insensitive NADH dehydrogenase (internal); $NADH_{EX}$, rotenone-insensitive NADH dehydrogenase (external); $NADPH_{IN}$, rotenone-insensitive NADPH dehydrogenase (internal); $NADPH_{EX}$, rotenone-insensitive NADPH dehydrogenase (external); II, succinate dehydrogenase complex (Complex II); CoQ, ubiquinone complex; Sto, SHAM-sensitive terminal oxidase; III, cytochrome bc1 complex; Cyt c, cytochrome c; IV, cytochrome c oxidase (Cox). Site I, electron entry site; Site IV, electron quenching site. Adopted from Shi et al. (2002).

presentation, location and cofactor specifications by gene annotation in the genome database; while ChEBI and PubChem Compound are used to provide detailed information for the compounds, such as, formula, charge, etc.

In databases, the metabolites are generally listed with their uncharged formula. However, many metabolites are protonated or deprotonated in medium and cells. Therefore, the charge and the corresponding formula of the compounds are very important to the prediction of the model. The protonation state, and thus, the charged formula, depends on the pH of interest. The intracellular pH is very important in maintaining normal physiological activities of the cell and therefore should be controlled within a very narrow range (Madshus, 1988). Once the charged formula is obtained for each metabolite, the reaction stoichiometry can be determined by counting different elements on the left- and right-hand side of the reaction. Addition of protons and water may be required in this step, as many databases and biochemical textbooks omit these molecules from the reactions. The change of proton in the reactions would have influence to the prediction of the model due to the power of proton

in cellular physiology. It is intimately involved in the capture of energy from oxidation, substrates and ion transports, etc., and is therefore necessary to balance every element and charge on both sides of the reaction.

To calculate the charge of the compounds, the intracellular pH must be determined. Just as mentioned above, the intracellular pH should be maintained at a relatively constant value to make normal physiological functions (Halperin, Goldstein, and Kamel, 2010; Madshus, 1988). The measured extracellular pH is usually between 7.4 (Halperin, Goldstein, and Kamel, 2010; Llopis et al., 1998; Madshus, 1988; Roos and Boron, 1981). With HCO_3^- existing intracellularly, the intracellular pH would be a little lower. Therefore, 7.2 is usually used to construct the model (Thiele and Palsson, 2010). Although organelles intracellular might have a different pH as cytosol, the mitochondrion usually has a pH range between 7.4 and 8.0 Porcelli et al., 2005. This would not have a big influence to the charge status of the different compounds. So pH 7.2 is a reasonable value used in the construction of the model. With intracellular pH determined, the charged formulas of the compounds are calculated based on pKa values. In this work, the pKa values for the compounds are from CRC Handbook of Chemistry and Physics by Lide (2009). In case there's no pKa value available, the computation of the charges is based on the functional groups or structures of the compounds (Thiele and Palsson, 2010) using ChemAxon (*Calculator Plugins*).

Once the charged formula is obtained for each metabolite, the reaction stoichiometry can be determined by counting different elements on the left- and right-hand side of the reaction.

The balanced reactions should fulfill the following equation:

$$M_E \times S = 0 \tag{4.1}$$

where S is the stoichiometric vector of the reactions, M_E is the element matrix, which is in the format of

$$M_E = \begin{array}{c} \\ C \\ H \\ O \\ N \\ P \end{array} \begin{array}{c} \begin{array}{cccccc} c_1 & c_2 & c_3 & \dots & c_n \end{array} \\ \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \end{array}$$

in which the row vector is the corresponding element number of the chemical compounds in the reaction, the column vector is the composition of chemical compounds in the reaction.

4.1.3 Compartmentalization of the reactions

Compartmentalization is very important in microorganisms. However, the detailed distribution of metabolic functions among compartments is often unknown. It is very important to exam carefully before determining which compartment to include in the model and which to exclude from the model. But by systematically constraining some individual fluxes in a de-compartmentalized version of the model, the compartments can be removed without introducing significant errors (Klitgord and Segrè, 2010). At the same time, for *S. cerevisiae*, due to the existence of a symmetrical electron transport chain unlike mammalian mitochondria, the localization of reducing cofactors plays a minor role in *S. cerevisiae* and can be ignored for a small scale model (Jin and Jeffries, 2004).

Currently, no obvious information on compartmentalization of *S. stipitis* has been found. Meanwhile, the transport reactions of the redox cofactors are very important to correct prediction of the phenotypes of the *in silico* strain. Without proper information on the constraints applied on these reactions, it is hard to predict proper phenotypes, e.g. the xylitol production under microaerobic conditions without further constraints, which are not shown in the results of the published genome-scale models for *S. stipitis* (Balagurunathan et al., 2012; Caspeta et al., 2012; Liu et al., 2012). Therefore, in this work, only one intracellular compartment, cytosol, is included in the model.

4.1.4 Determination of the objective function

To identify optimal solutions in the vast solution space, FBA objective functions are defined to solve the system of linear equations that represent the mass balance constraints. While different objectives were proposed for various biological systems (Burgard and Maranas, 2003; Ebenhöh and Heinrich, 2001; Helnrlch et al., 1997; Holzhütter, 2004; Kacser and Beeby, 1984; Knorr, Jain, and Srivastava, 2007; Price, Schellenberger, and Palsson, 2004), by far the most common assumption is that microbial cells maximize their growth, especially for continuous culture (Schuetz, Kuepfer, and Sauer, 2007). Therefore, biomass reaction is widely used in the construction of metabolic network model and is selected as the objective function in this work. The biomass reaction accounts for all known biomass constituents and their fractional contributions to the overall cellular biomass. A detailed biomass composition of the target organism needs to be determined experimentally for cells growing in log phase (Izard and Limberger, 2003).

Due to the lack of physiological data for *S. stipitis*, the biomass reaction is constructed based on the available information on genome and comparison with the model of *S. cerevisiae* (Duarte, Herrgå rd, and Palsson, 2004; Vanrolleghem et al., 1996). The biomass reaction of *S. stipitis* (SLININGER et al., 1990) has been abandoned because it is too simple to reflect the requirement of cell growth to different nutrients and pre-cursors.

First, the synthesis pathways of DNA, RNA and amino acids of *S. stipitis* have been searched based on gene information from KEGG (Kanehisa et al., 2012) and *S. stipitis* genome database (Jeffries et al., 2007). The pathways have been lumped by Matlab to simplify the reactions. Then the compositions of DNA, RNA, and amino acids are extracted from the data from *S. stipitis* genome database vis Biopython (Cock et al., 2009).

4.2 Model refinement

After the draft procedure, the model needs to be tuned to remove incorrect information and make its prediction more reliable. In the published protocol (Thiele and Palsson, 2010)

the refine procedure is clearly defined and described. In this part the drafted model have been tuned in several steps to improve the predictions under various conditions. First, some "reaction cycles" have been evaluated and most confidential reactions have been kept. This is due to the limited availability of physiological information fro *S. stipitis* and therefore some reactions reported in KEGG and literatures for other strains might not exist. Second, more constraints have been applied to exchange reactions to make the prediction of the model fit better with real experimental results. Thereafter the influence of maintenance energy to the predicted phenotypes have been studied and the proper constraint has been applied for the further investigations.

4.2.1 Tune-up of exchange reaction constraints

Due to lack of transcription and regulation data, the constraints of reactions are kind of arbitrary at the beginning. The prediction of the model might be vary far away from the real phenotypes without proper constraints. Among all kinds of constraints, the ones applied on exchange reactions have very important impact to the simulation results due to their affects on the uptake of nutrients or ions from extracellular or the excretion of intracellular products to fulfill objective function or maintain the redox balance inside the cells. The drafted exchange reactions might not exist or should be blocked in the real metabolism. Therefore, the adjustment of constraints on exchange reactions during simulation is crucial to the accuracy of predication.

In the initial draft model, the exchange reaction of succinate is unbounded, i.e., the cell is supposed to be able to uptake or excrete succinate without limitation. During the simulation, the unlimited lower and upper boundaries make the reaction a sink for oxygen and a pool for ATP generation. Therefore, the simulation results predict flux of succinate for both glucose and xylose fermentation, which has not been reported in literatures and is also not shown in our experiments. Thus, the excretion of succinate has been blocked.

In the same way, the exchange of acetaldehyde has also been blocked. Otherwise, the main by-product in anaerobic fermentation of glucose will be acetaldehyde, which hasn't been observed in real experiments. After blocking the exchange reaction of acetaldehyde, the main by-product changed to acetic acid, which has been detected in real experiments.

4.2.2 Futile cycles identification

Futile cycle here is defined the cycle formed by reactions that are only used to import or export proton or to generate energy but are isolated from other parts of the reaction network or maybe not be isolated but are not biologically meaningful. One significant characteristics of the futile cycle is that the fluxes of the reactions are usually pushed to very high value (close to the upper bound). This kind of pathway exists commonly in the published metabolic network models. For example, in genome-scale model of *S. cerevisiae* (Duarte, Herrgå rd, and Palsson, 2004), many futile cycles exist in the solution as shown in Table 4.1.

Table 4.1: Futile cycles shown in genome-scale model of *S. cerevisiae*

Glucose uptake rate	**Oxygen uptake rate**	**Growth rate**	**Active reactions**	**Reaction in futile cycles**
5	Unconstrained	0.4779	316	47
5	0	0.0853	302	38

The futile cycles are usually added into the model due to incomplete annotation to the genome and undetermined thermo properties of the reactions. The large flux values are biologically meaningless and may cause unexpected results in model analysis. Therefore, it should be identified and modified to make predictions of the model more reliable.

By setting proper boundary conditions and constraints, *in silico* culture of *S. stipitis* under different aeration conditions and on glucose or xylose have been carried out. The futile reactions have been identified in the old model and listed in Table 4.2. With few physiological information on the reaction pairs found, the model was modified based on the comparison with *S. cerevisiae* genome-scale model (Duarte, Herrgå rd, and Palsson,

2004). The reactions involved in the futile cycles were removed or changed the reversibility to irreversible to prevent the occurrences of futile cycles. The corresponding actions are also listed in Table 4.2.

Table 4.2: List of futile cycles and the actions applied

Reaction pair	**Formula**	**Action**
ADHx	acald[c] + nadh[c] + h[c] ⇔ etoh[c] + nad[c]	
ADHy	acald[c] + nadph[c] + h[c] ⇔ etoh[c] + nadp[c]	Removed
GA3PDx+PGK	ga3p[c] + nad[c] + pi[c] ⇔ 13dpg[c] + nadh[c] + h[c] 13dpg[c] + adp[c] ⇔ 3pg[c] + atp[c]	
GA3PDHy	ga3p[c] + nadp[c] + h2o[c] → 3pg[c] + nadph[c] + 2 h[c]	Removed
ICITDHx	icit[c] + nad[c] ⇔ akg[c] + co2[c] + nadh[c]	To irreversible
ICITDHy	icit[c] + nadp[c] ⇔ akg[c] + co2[c] + nadph[c]	To irreversible

4.2.3 Influence of non-growth-associated maintenance energy

The ATP generated in metabolism is used in two ways intracellularly. First, the ATP is required for biomass synthesis (i.e. precursor biosynthesis and polymerization) (growth-associated maintenance, GAM), and then is used for the maintenance of cell metabolism not related to cell growth (non-growth-associated maintenance, NGAM), e.g. for keeping intracellular pH stable. It is usually expressed in the following equation:

$$r_{ATP} = Y_{XATP} \cdot \mu + m_{ATP} \tag{4.2}$$

where r_{ATP} specifies the total amount of ATP being utilized, Y_{XATP} corresponds to GAM, μ is the specific growth rate, and m_{ATP} is NGAM. The GAM is shown in biomass reaction while the NAGAM is defined in the reaction named ATPM:

ATPM: h2o[c] + atp[c] → h[c] + adp[c] + pi[c]

In this work, the GAM has been calculated as 81.50 mmol/gDCW. The NGAM has been reported to be between 0.5-3.5 mmmol/gDCW/h (Balagurunathan et al., 2012; Caspeta et al., 2012; Guebel et al., 1991; Liu et al., 2012; Rizzi et al., 1987). Due to the nature of FBA, the optimization process is always trying to minimize the NGAM to favor biomass production. Therefore, a proper lower boundary for reaction ATPM is necessary. However, the NGAM will change with the physiological status of the cells. In order to set a proper value for NGAM, we study the influence of NGAM to the phenotype prediction of the model under various oxygenation condition, especially for xylose metabolism considering its high sensitivity to oxygen.

The *in silico* experiment is carried out with sugar uptake rate (glucose/xylose) of 10 mmol/gDCW/h, oxygen uptake rate of [0, 3] mmol/gDCW/h, and NGAM of [0, 4] mmol/gDCW/h. Due to the significant influence of oxygen to ethanol production, especially for xylose metabolism, the investigation must be done within a range of OUR instead of a fixed value. The simulation results of glucose/xylose metabolism are shown in Figure 4.3.

From the results, it showed that for cell growth and ethanol production, NGAM doesn't have a significant influence, which is expected. Because NGAM is small compared with GAM, the change of NGAM within the investigation range is almost negligible. While the main physiological function of ethanol production is to produce energy, it have, therefore, smaller influence with higher OUR when energy is most produced from oxidative phosphorylation. Meanwhile at lower OUR, the portion of NGAM in total energy produced is higher and the influence is larger. However, for xylitol and acetic acid production, NGAM plays an important role, which is mostly due to the production of hydrogen ion in this reaction and thus the influence to intracellular redox state. Regarding glucose metabolism, it shows similar trend of influence (data not shown). It is hard to determine which data is more confidential comparing with other sources. In this dissertation, NGAM of 3.5 mmol/gDCW/h from Guebel et al. (1991) has been used.

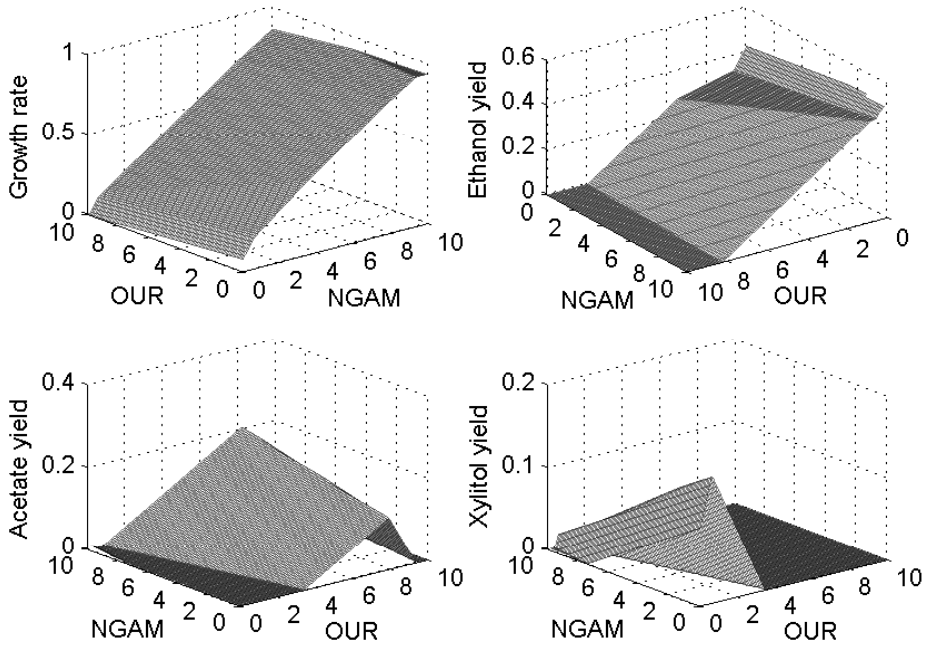

Figure 4.3: Influence of non-growth-associated maintenance energy (NGAM) and oxygen uptake rate (OUR) to *in silico* xylose fermentation

4.2.4 Flux coupling constraints on xylose reductase and xylitol dehydrogenase

The experimental determination of metabolic flux and pathway usage through the use of isotope tracers has significantly contributed to the overall understanding of regulated metabolism. One approach to characterize metabolism is through the use of metabolic flux ratio analysis (Fischer, Zamboni, and Sauer, 2004; Sauer et al., 1999). This method is used to determine the degree of converging pathway usage to produce a metabolite pool when multiple synthesis routes exist, i.e. it describes the dependencies between reactions. Early results revealed the robustness of central carbon metabolism of *E. coli* (Fischer, Zamboni, and Sauer, 2004; Sauer et al., 1999) as many calculated flux ratios were found impervious to genetic perturbations. Thus, with the availability of physiological data, the flux ratio will be an additional constraint on the metabolic network model.

The general form of flux ratio constraint between reaction 1 ($r1$) and reaction 2 ($r2$) is described in Equation 4.3:

$$R_{12} = \frac{v_{r1}}{v_{r2}} \tag{4.3}$$

where R_{12} is the flux ratio, v_{r1} and v_{r2} are the fluxes through $r1$ and $r2$ respectively. It is very clear that this constraint is nonlinear. The early algorithms developed (Bühler et al., 2008; Sauer et al., 1997) used nonlinear programming methods and was effective given small metabolic networks of primary metabolism. Unfortunately, these aspects have limited the applicability to large genome-scale metabolic networks, which often must rely on linear programming.

McAnulty et al. (2012) linearized the flux ratio constraints and used them to study the metabolism of *Clostridium acetobutylicum* via the gnome-scale model. The linearization of the flux ratio constraint described in Equation 4.3 is done by converting it into the linear form in Equantion 4.4:

$$v_{r1} - R_{12} \times v_{r2} = 0 \tag{4.4}$$

which can be easily solved by linear programming. However, in the work of McAnulty et al. (2012), a fix flux ratio constraint was applied. The dependencies of reactions are more than fixed ratios. Flux coupling analysis (FCA) has been developed to analyze the dependicies of reactions in stoichiometric metabolic network model. With the relationship between evolution of metabolic genes and flux coupling addressed (Notebaart et al., 2008, 2009; Pál, Papp, and Lercher, 2005a,b; Seshasayee et al., 2009), the studies on genomes can also reveal more information on dependencies of reactions, i.e. flux coupling.

There are totally five flux coupling relationships have been defined (Burgard et al., 2004; Marashi and Bockmayr, 2011): directional coupling, partial coupling, full coupling, mutually exclusive, and uncoupled (or sometimes coupled). The definition is listed as below and illustrated in Figure 4.4.

Directional coupling ($v_1 \rightarrow v_2$), if a non-zero flux for v_1 implies a non-zero flux for v_2 but not necessarily the reverse.

Partial coupling ($v_1 \leftrightarrow v_2$), if a non-zero flux for v_1 implies a non-zero, though variable, flux ratio for v_2 and vice versa.

Full coupling ($v_1 \Leftrightarrow v_2$), if a non-zero flux for v_1 implies not only a non-zero but also a fixed flux for v_2 and vice versa.

Mutually exclusive ($v_1 \bigoplus v_2$), if a non-zero flux for v_1 implies a zero flux for v_2 and vice versa.

Uncoupled ($v_1 \nleftrightarrow v_2$), if the reaction pair do not fall into any of the above catagories.

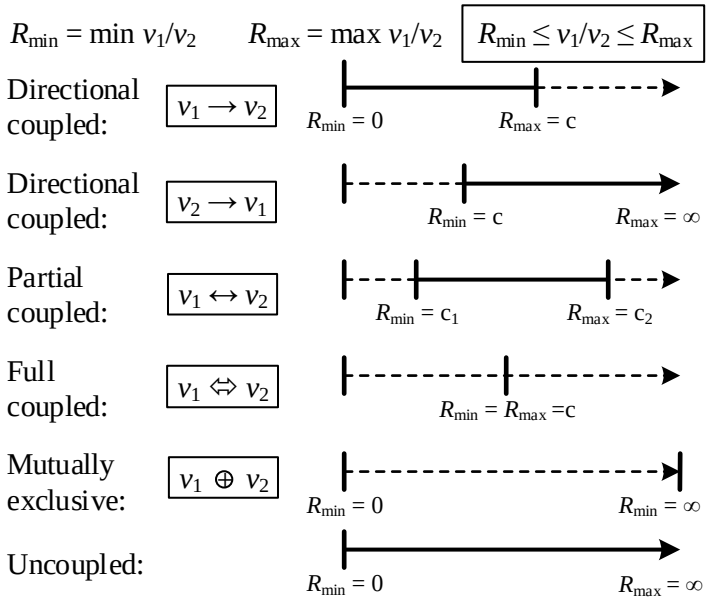

Figure 4.4: Illustration of possible flux coupling between two reactions. Various types of coupling are related to the flux ratio limits R_{min} and R_{max} as shown. Adopted and Modified from Burgard et al. (2004)

Marashi and Bockmayr (2011) proved that FCA is sensitive to the missing reactions. For the small scale network, FCA can reveal limited and even wrong information. Therefore, it is meaningless to apply FCA to the developed model.

Inspired by flux coupling analysis (FCA), the flux ratio constraint is further extend to be flux coupling constraint (FCC). Totally four types of constraints have been developed: directional coupling constraint, partial coupling constraint, full coupling constraint, and mutually exclusive constraint. It is not necessary to apply any constraint if the reaction pair is uncoupled.

Directional coupling $(v_1 \rightarrow v_2)$ indicates that $v_1/v_2 \leq c$. Therefore, after the linearization it is changed to:

$$c \times v_2 - v_1 \geq 0 \tag{4.5}$$

Partial coupling $(v_1 \leftrightarrow v_2)$ indicates that $c_1 \leq v_1/v_2 \leq c_2$, which can be converted to:

$$\begin{aligned} v_1 - c_1 \times v_2 \geq 0 \\ c_2 \times v_2 - v_1 \geq 0 \end{aligned} \tag{4.6}$$

Full coupling $(v_1 \Leftrightarrow v_2)$ indicates $v_1/v_2 = c$, therefore the linearized form is the same as Equation 4.4.

Mutually exclusive $(v_1 \oplus v_2)$ cannot be implemented by linear programming but mixed integeral linear programming (MILP). First, two binary variables, i_1 and i_2, are introduced for the two fluxes v_1 and v_2 respectively, which follow the constraint:

$$i_1 + i_2 = 1 \tag{4.7}$$

Second a big number B is combined together with the binary variable defines following constraints:

$$\begin{aligned} B \times i_1 - v_1 \geq 0 \\ B \times i_1 - v_1 \leq B \\ B \times i_2 - v_2 \geq 0 \\ B \times i_2 - v_2 \leq B \end{aligned} \tag{4.8}$$

Changing the problem to MILP problem leads to a NP-hard computation complexity instead of a P-hard. Thus the computation loads will increase significantly.

Usually most flux coupling constraints are implicated by the network structure. However, for the branched nodes, at least partial coupling or full coupling constraints will be very useful to further constrain the solution space of the metabolic flux distribution.

For the central carbon metabolism of *S. stipitis*, the xylose reductase and xylitol dehydrogenase involved in the first two steps of xylose metabolism are dual cofacotr preferred as discussed in the background and shown in Figure 4.1. It is also reported to be the reason for low xylitol production in xylose metabolism of *S. stipitis*, comparing to the mono cofactor preference of NADPH in other strains. Therefore, the redox imbalance caused by the NAD^+-dependence of xylitol dehydrogenase (XDH) could be reduced (Agbogbo and Coward-Kelly, 2008). Many results on experimental study of XR preference to NADPH and NADH have been reported (Hou, 2012; Slininger et al., 2011; Verduyn et al., 1985; Yablochkova et al., 2003, 2004). It would be a very useful constraint to increase the accuracy of *in silico* study for xylose metabolism with *S. stipitis*.

From the network structure, the reactions catalyzed by XR and XDH are intrinsically partial coupling, i.e. $v_{XRx} \leftrightarrow v_{XRy}$ and $v_{XDHx} \leftrightarrow v_{XDHy}$, where subscriptions of x and y correspond to NAD(H)- and NADP(H)-linked reactions. Due to the difficulty in measurement, no data of real flux values of R_{min} and R_{max} through the reactions is available. Thus, the dependencies of the reactions have to be estimated from *in vitro* measurements of specific enzyme activities or kinetic parameters of the enzymes. The reported ratios of specific enzyme activities are summarized in Table 4.3.

From Table 4.3, specific enzyme activitiy ratio of between XRy and XRx is within the range [0.5 2.3] while the range for XDH is [0, 0.01). Considering that the *in vitro* measurements were carried out with saturated substrate concentration in dilute solution compared with the intracellular system as well as the *in vivo* concentrations of the cofactors, the *in vivo* flux ratios of XR and XDH are within [0.3 3] and [0 0.02] respectively.

Table 4.3: Summary of the specific enzymes activities to NAD(H) and NAP(H) of xylose reductase and xylitol dehydrogenase

Enzyme	**Specific activitiy (U/mg)**		**Ratio**	**Citation**
	NADP(H)	**NAD(H)**		
XR	0.23 ± 0.06	0.10 ± 0.02	2.3	Bengtsson, Hahn-Hägerdal, and Gorwa-Grauslund (2009)
	0.73	0.65	1.1	Hou (2012)
	0.25 ± 0.05	0.13 ± 0.04	1.9	Eliasson et al. (2000)
	47.8	30.7	1.6	Rizzi et al. (1988)
	1.61	1.25	1.3	Skoog and Hahn-Hägerdal (1990)
	1.39	0.86	1.6	Skoog and Hahn-Hägerdal (1990)
	0.38	0.29	1.3	Verduyn et al. (1985)
	0.08 ± 0.005	0.03 ± 0.004	2.7	Watanabe et al. (2007b)
	10.6	5.96	1.8	Yablochkova et al. (2004)
	4.84	10.37	0.5	Yablochkova et al. (2004)
	8.24	4.53	1.8	Yablochkova et al. (2004)
XDH	ND	0.31	-	Hou (2012)
	0.009 ± 0.001	1.272 ± 0.152	0.007	Matsushika et al. (2008)
	ND	0.17 ± 0.01	-	Watanabe et al. (2007b)
	0.56	17.50	0.03	Yablochkova et al. (2004)
	0.27	8.37	0.03	Yablochkova et al. (2004)
	0.42	9.92	0.04	Yablochkova et al. (2004)
	0.001 ± 0.0001	1.11 ± 0.09	0.001	Watanabe, Kodaki, and Makino (2005)

Note: due to the different definition of the enzyme activity, the absolute values may differ significantly among the researchers.

With the partial coupling constraints applied, the optimal flux distributions calculated via linear programming always pick up the R_{min} of XR and R_{max} of XDH (data not shown), i.e. XR always tries to utilize NADH as much as possible while XDH always tries to generate NADPH in its highest capacity allowed. Therefore, $v_{XRx} \leftrightarrow v_{XRy}$ and $v_{XDHx} \leftrightarrow v_{XDHy}$ have been changed to $v_{XRx} \Leftrightarrow v_{XRy}$ and $v_{XDHx} \Leftrightarrow v_{XDHy}$ in the model. In other words, the values

of R_{XR} and R_{XDH} are fixed in the model simulation. In the following studies, R_{XR} is 1 and R_{XDH} is 0 without explicit statement.

4.2.5 Statistics of the model

The model has been constructed from textbook and on-line databases. It captures the central metabolism of glucose and xylose. Included in the model are 118 reactions with 66 as reversible and 52 as irreversible (including transport reactions) and 65 metabolites (ignorance of compartmentalization). Details of the reactions and metabolites are listed in Appendix A and B respectively. Fifteen compounds allowed to exchange with 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. Full coupling constraint have been applied on XRx and XRy, XDHx and XDHy as 1 and 0 respectively.

4.3 Validation of the prediction capacity of the model

Metabolic network model must be validated after the reconstruction to prove the reliability of the predictions. Various validation, qualitative and quantitative ones, have been applied to different models. The most straightforward validation is to compare the model prediction with the experimental results. However, with little intracellular physiological information, it is very hard to be confident that the applied constraints are true in the real metabolism. Therefore, for genome-scale models, the most common validation methods are (Balagurunathan et al., 2012; Caspeta et al., 2012; Duarte, Herrgå rd, and Palsson, 2004; Liu et al., 2012; Mo, Palsson, and Herrgå rd, 2009; Orth et al., 2011): 1) validating the utilization of substrates, i.e. whether a substrate can be metabolized *in silico* when it can be fermented in wet experiments; 2) studying the influence of gene deletions, i.e. whether the model shows similar phenotypes *in silico* and in web experiments when certain gene has been deleted; 3) comparing the *in silico* predictions and the results of experiments, which

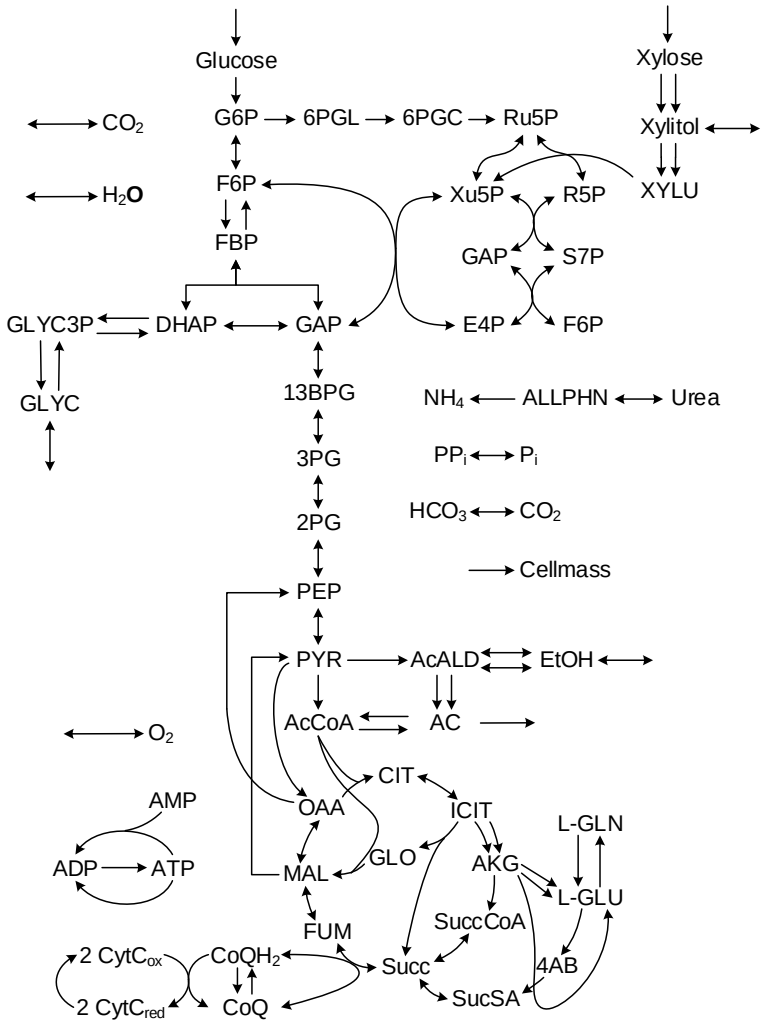

Figure 4.5: Overview of the metabolic network model. The double arrows in the same direction for a reaction indicate that the enzyme catalyzes the reaction has affinity to different cofactors (NADH/NAD^+ and NADPH/$NADP^+$). This also applies to the other figures in this work.

is usually done roughly with limited experimental results due to the wide varieties exists in wet experiments.

Just as mentioned at the beginning of the chapter, the objective of our work is to study the glucose and xylose metabolism of *S. stipitis*. Thus we construct a small-scale central carbon metabolic network model with only glucose and xylose as the substrates. To validate the model, we first qualitatively compare the general predicted and real phenotypes as well

as the flux ratio between glycolysis and pentose phosphate pathway (PPP). Then we focus on confirm the performance of the model with wet experimental data.

4.3.1 Qualitative validation with general prediction

To validate the model qualitatively, we first compared the model prediction with what we observed in the experiments, e.g. the main products under various oxygenation. The results are shown in Table 4.4.

Table 4.4: The comparison of general performance of the model and the experiments

Carbon source	**Aeration**	**Products**	
		Experiments	**Predictions**
Glucose	Aerobic	Cellmass, CO2	Cellmass, CO2
	Micro-aerobic	Cellmass, ethanol, glycerol, acetic acid, CO2	Cellmass, ethanol, glycerol or acetic acid, CO2
	Anaerobic	Cellmass, ethanol, glycerol, acetic acid, CO2	Cellmass, ethanol, glycerol, CO2
Xylose	Aerobic	Cellmass, CO2	Cellmass, CO2
	Micro-aerobic	Cellmass, ethanol, xylitol, acetic acid, CO2	Cellmass, ethanol, xylitol or acetic acid, CO2
	Anaerobic	No growth	Infeasible solution

There is good agreement between the simulated results and the experimental observations, which indicates that the simulated intracellular fluxes might agree with the ones in live cells. A major discrepancy between the model predictions and experimental data is the production of acetic acid. From experimental data, acetic acid is produced with limited rate and can be secreted simultaneously with glycerol or xylitol. However, in simulations, acetic acid is produced in a high rate not observed in experiments. Meanwhile, the production of glycerol or xylitol and acetic acid are mutually exclusive.

The differences between glucose and xylose metabolism, especially the specific growth rate and ethanol yield, are shown in Figure 4.6. It shows that *S. stipitis* grows faster with

glucose than with xylose while ethanol yield with xylose is higher vice versa. These agree well with the experimental observations.

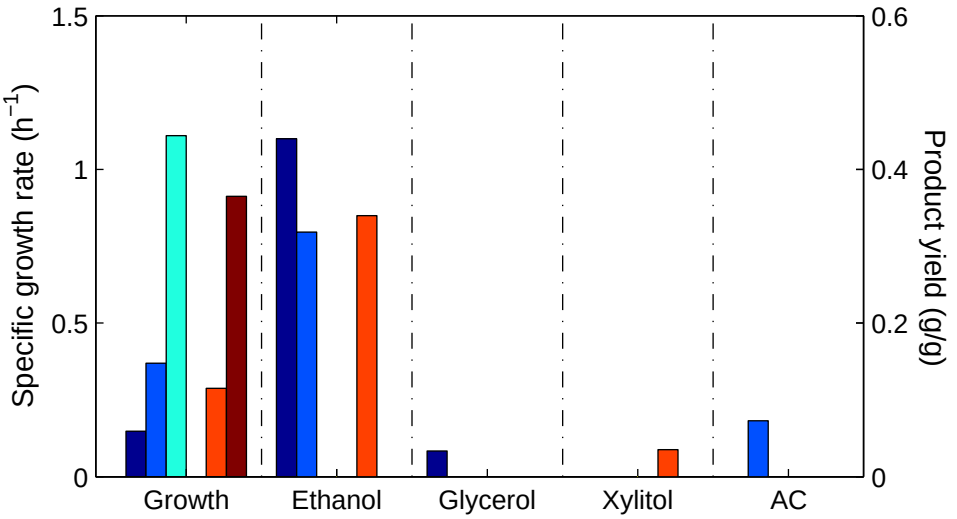

Figure 4.6: Growth and products formation with glucose or xylose under various oxygen conditions

To study the intracellular metabolism, metabolic flux profiling or metabolic flux analysis (MFA) is a very useful tool. However, comparing with *S. cerevisiae*, *S. stipitis* has very limited information available. To my best knowledge, few papers has been published (Feng and Zhao, 2013; Fiaux et al., 2003). The information might provide another valuable validation method to the model. Therefore, the ratio of carbon flux through PPP has been studied and compared with the reported data (shown in Table 4.5). Because there's no detailed OUR available for oxygen-limited condition from the original papers, the OUR has been set to be 0.4 mmol/gDCW/h in the simulations.

From the result, the value of the ratio of carbon flux through PPP for glucose metabolism under aerobic condition predicted by the model falls into the range of experimental data. This agreement adds more confidence to the structure of the model.

Table 4.5: Ratio of carbon flux through PPP of simulated and experimental results

Carbon source	**Aeration condition**	**Ratio of carbon flux through PPP**	
		Simulated result	**Experimental result**
Glucose	Aerobic	61.66%	57±9% [a]
	Oxygen-limited	15.24%	15.56±4.67% [b]
Xylose	Aerobic	49.66%	N/A
	Oxygen-limited	18.26%	16.36±4.91% [b]

[a]: Fiaux et al. (2003)

[b]: Feng and Zhao (2013)

4.3.2 Quantitative validation with experimental data

FBA supposes steady state of the cell metabolism but only a few studies on continuous culturing of *S. stipitis* to investigate its physiology have been done (Fiaux et al., 2003; Li, 2012; Skoog and Hahn-Hägerdal, 1990; Skoog, Jeppsson, and Hahn-Hägerdal, 1992). Particularly, the oxygen transfer rate is a crucial factor for xylose metabolism that, at defined levels, can maximize the productivity and yield of ethanol. The role of oxygen in xylose fermentation of *S. stipitis*can be explained by the fact that cells have to maintain redox balance, xylose transportation, cell growth or keep mitochondrial function (Skoog, Jeppsson, and Hahn-Hägerdal, 1992). Results from *in silico* predictions of specific growth rate (μ), ethanol yields, and CO_2 yield were compared with experimental data (Figure 4.7). The simulation conditions are listed in Table 4.6.

Figure 4.7 shows that the model predicts the correlation of oxygen transfer rate with metabolism, which passes from fermentative to respiratory. These results are in agreement with experimental data (Li, 2012). Furthermore, *in silico* simulations predicted the inability of *S. stipitis* to grow in anaerobic conditions with the minimal medium.

From Figure 4.7, the predictions of the cell growth and ethanol yield agree very well with the experimental data. The computed CO_2 yield is higher than experimental data,

Table 4.6: Model setup for validation with experimental data from Li (2012)

Condition	Sugar uptake rate	Oxygen uptake rate	XR ratio	XDH ratio
Glucose aerobic (GAO)	4.25 (Glucose)	5.98	-	-
Glucose microaerobic (GMA)	3.32 (Glucose)	1.40	-	-
Xylose aerobic (XAO)	4.82 (Xylose)	6.40	0.5	0
Xylose microaerobic (XMA)	4.07 (Xylose)	1.67	0.5	0

especially for xylose metabolism under microaerobic condition. The discrepancy should be caused by the utilization of CO_2 in the anabolism which is lumped in the model and some alternative pathways may be absent. The experimental data for acetic acid production is not available in Li (2012). The model predicts the secretion of acetic acid in most of the simulation setup when oxygen is limited. The calculated results show that, however, the production of acetic acid cannot exist simultaneously with xylitol and glycerol production in xylose and glucose metabolism respectively. One should note that in the work of Skoog and Hahn-Hägerdal (1990) no xylitol production was observed. However, in the *in silico* evaluations (model setup and results do not list here), other published works (Slininger et al., 2011) and our experiments, it showed that xylitol is produced in the xylose metabolism under oxygen-limited condition.

4.4 Conclusion

In this chapter, following the published procedure (Thiele and Palsson, 2010), a central carbon metabolic network model for *S. stipits* has been reconstructed, refined and validated. The compartmentalization of the reactions has been discussed and only one compartment, cytosol, is involved in the model. The objective function has been chosen to be cellmass reaction, i.e. the cell growth, which is formed based on the genome and biochemical information from *S. stipitis* as well as *S. cerevisiae* whenever the information specific to *S. stipitis* is not available. The model was further improved by tuning up its reactions and

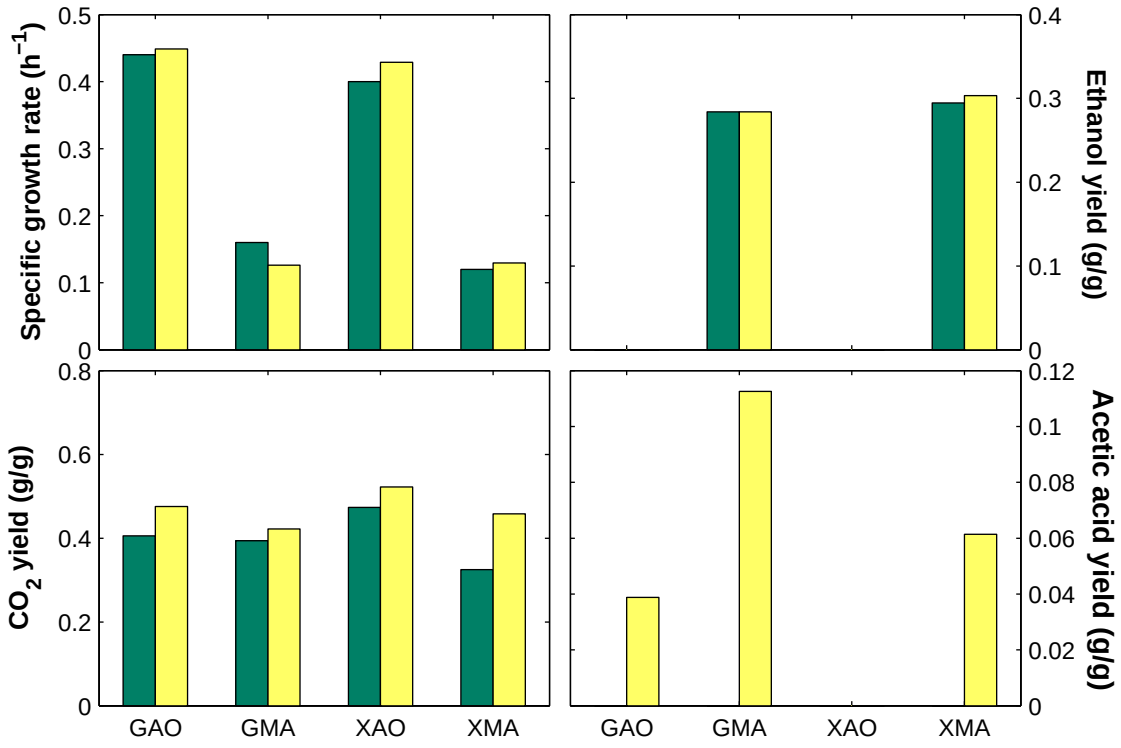

Figure 4.7: Comparison of cell growth and product yields between computed and experimental results. GAO: glucose aerobic; GMA; glucose microaerobic; XAO: xylose aerobic; XMA: xylose microaerobic. The dark green bars are experimental data; simulated results are presented by light yellow bars.

constraints. The futile cycles in the model has been identified and revised. The boundaries of the exchange reactions have also been evaluated to make the simulated results more close to the experimental phenotypes. Non-growth-associated maintenance energy has been found to be important to the prediction of by-products and a suitable value has been chosen based on literature data and simulation results. Due to the dual cofactor specificities of xylose reductase and xylitol dehydrogenase, the *in vivo* fluxes through different reactions are hard to determine. To further constrain the model, flux coupling constraints, an extension of flux ratio constraint, have been applied to the reactions catalyzed by the two enzymes. The ratios were determined based on literature data and model simulation. The finalized model

totally has 118 reactions (66 reversible and 52 irreversible) with transport reactions and 58 metabolites.

After the reconstruction of the model, simulation results were compared with reported experimental data qualitatively and quantitatively to validate the prediction capacity of the model. The general performance, intracellular flux ratio of pentose phosphate pathway all agree well with reported data. Furthermore, the model was evaluated quantitatively with published experimental data and showed a good agreement. The validation process gives us confidence on the quality and prediction capacity of the model. Thus in next chapter, we will analyze and compare the glucose and xylose metabolism of *S. stipitis* with the model.

Chapter 5
Analysis of the reconstructed model

With the metabolic network model reconstructed and validated, various approaches have been applied to study glucose and xylose metabolism of *S. stipitis*. In this chapter, the topological properties have been checked first and compared with the other metabolic network models. Then we investigated the influence of oxygen to glucose and xylose metabolism, which have been reported to be important to efficient ethanol production. Different phenotypes have been identified and the difference of internal fluxes have been studied.

5.1 Methods

5.1.1 Flux balance analysis (FBA)

Flux balance analysis (FBA) is a widely used approach for studying biochemical networks (Orth, Thiele, and Palsson, 2010). FBA calculates the flow of metabolites through this metabolic network, thereby making it possible to predict the growth rate of an organism or the rate of production of a biotechnologically important metabolite. With metabolic models for 35 organisms already available (see on-line list) and high-throughput technologies enabling the construction of many more each year, FBA is an important tool for harnessing the knowledge encoded in these models.

5.1.2 Robustness analysis

Biological systems exist and have evolved in the face of internal and external perturbations. These perturbations have resulted in a particular biological system organization

manifested in multiple layered and inter-related components (Yamada and Bork, 2009). The unprecedented progress in molecular biology propels the understanding of living systems through integration of these components, and complements the reductionist approach that has prevailed in various biological disciplines. Moreover, systems biology has revived the interest in gleaning the determinants which contribute to the robustness of biological systems, i.e., their inherent property to maintain normal performance in presence of perturbations of changing environments and internal modification (Kitano, 2002; Koonin and Wolf, 2010; Shinar and Feinberg, 2010; Visser et al., 2003). This has resulted in studies of robustness at the different levels of organization, from gene regulation to population level (Shinar and Feinberg, 2010; Visser et al., 2003).

Due to the critical role of metabolism, it is becoming generally accepted that robustness is one of its salient properties. Determinants of robustness is usually stemmed from graph-theoretic and stoichiometry-based formalisms: robustness from network structure and from constraint-based approaches (Larhlimi et al., 2011). Here we adopt the latter approach to study the robustness of the reconstructed model due to the model size and combinations of the reactions. Robustness, defined here with respect to metabolic networks, is a measure of the change in the maximal flux of the objective function when the optimal flux through any particular metabolic reaction is changed (Edwards and Palsson, 2000). This definition reveals how sensitive the objective is to a particular reaction.

FBA and a variety of it, such as flux variability analysis (FVA) (Edwards and Palsson, 2000) and phenotype phase plane analysis (PhPP) (Edwards, Covert, and Palsson, 2002), have been developed to study the robustness based on optimization. In this work, FBA have been used to study the robustness of particular reaction, i.e. the flux through one reaction is varied and the optimal objective value is calculated as a function of this flux via FBA.

5.1.3 Phenotype analysis

All steady-state metabolic flux distributions are mathematically confined to the flux cone defined for the given metabolic map, where each solution in the flux cone corresponds to a particular internal flux distribution or a particular metabolite phenotype (Varma and Palsson, 1994). Under specified growth conditions, the optimal phenotype in the cone can be determined using linear programming (LP). If the constraints vary, the shape of the cone changes, and the optimal flux vector may qualitatively change. Phenotype analysis is to consider all possible variations in constraints (Edwards, Ramakrishna, and Palsson, 2002). The investigated constraining variables can be one or two. If two varibles are studied simultaneously, it is called phenotype phase plane analysis (PhPP) (BELL and PALSSON, 2005; Edwards, Ramakrishna, and Palsson, 2002).

The phenotype is constructed by calculating the shadow prices throughout the solution space of the particular problem. The shadow price defines the intrinsic value of each metabolite toward the objective function. Changes in shadow prices are used to interpret the metabolic behavior.

Mathematically, the shadow price is the dual solution of the linear programming problem (Bazaraa, Jarvis, and Sherali, 2009). It is defined as (Edwards, Ramakrishna, and Palsson, 2002; Varma and Palsson, 1993):

$$\gamma_i = -\frac{dZ}{db_i^v} \tag{5.1}$$

The shadow price defines the sensitivity of the objective function (Z) to changes in the availability of each metabolite (b_i^v defines the violation of a mass balance constraint and is equivalent to an uptake reaction). The shadow price can be either negative, zero, or positive, depending on the value of the metabolite. The direction and magnitude of the shadow price vector in each phenotype is different and thus related to the state of the metabolic system.

5.2 Topological properties of the model

5.2.1 Degrees of metabolites

The degree of a node in a network (sometimes referred to incorrectly as the connectivity) is the number of connections or edges the node has to other nodes. The metabolites are connected to each other by reactions in the metabolic network. Therefore, the degree of metabolites can be calculated by counting the reactions that they present. This is done by converting the stoichiometric matrix S into a binary matrix and thus counting the non-zero items in each row. The result is shown in Figure 5.1.

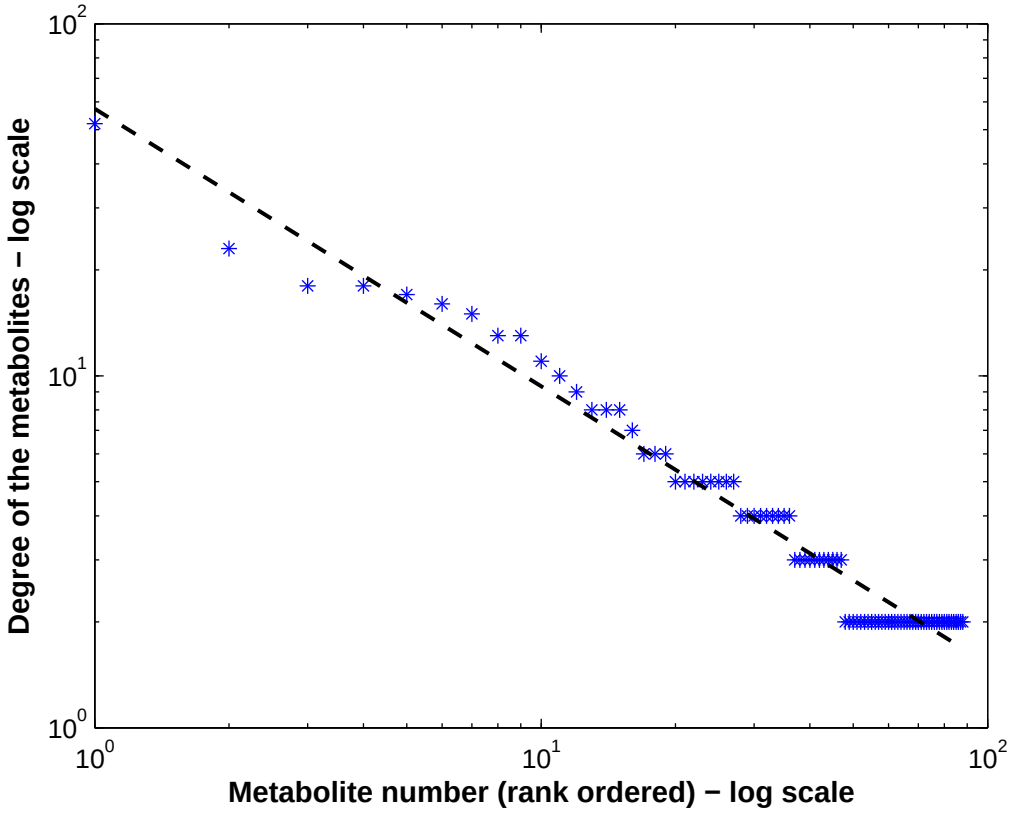

Figure 5.1: Degree distribution of the metabolites in the *S. stipitis* model. The dash line shows the exponential relationship between degree of metabolite and reaction number.

There are total 17 different degrees. The lowest degree is 2 while the highest value is 52. As Figure 5.1 shows, there are very few metabolites that are high-degree, while

most metabolites participate only in a few reactions. The few high-degree metabolites are "global" players, similar to hubs in protein-protein interaction networks, while the low-degree metabolites are "local" players, many of which only occur in linear pathways (Barabási and Oltvai, 2004; Palsson, 2006). The approximate linear appearance of the curve of degree of the metabolites corresponds to a "power law" degree distribution of metabolite, which indicates that the network is scale-free (Jeong et al., 2000). This result agrees with the structure of the genome-scale models (Duarte, Herrgå rd, and Palsson, 2004; Orth et al., 2011) although this one is small. The metabolite with highest degree is hydrogen ion, which has a value of 52.

5.2.2 Correlation between reaction essentiality and degree of metabolite

Previous research shows that network topology, especially the degree distribution of nodes provides significant information about the presumed robustness of microorganisms to perturbation (Barabási and Oltvai, 2004; Hartwell et al., 1999). Earlier works (Mahadevan and Palsson, 2005; Samal et al., 2006) reveals that low degree metabolites explain essential reactions in reaction networks of *E. coli*, *S. cerevisiae* and *Staphylococcus aureus*. Therefore, it is very interesting to investigate whether the structure of the model implies the same conclusion.

The reaction essentiality is determined as whether it would cause the cell stop growing once it has been removed. The results depends on the constraint setup. Specifically, in our model, it depends on the carbon source (glucose or xylose) and oxygen supply condition (aerobic or oxygen-limited). With aerobic glucose culture, it shows that a total of 14 reactions are essential to cell growth. When the aeration condition changed to oxygen-limited (glucose), the number became 18. For xylose culture, the number is 16 and 26 for aerobic and oxygen-limited conditions respectively. 10 reactions are essential under all conditions. They are distributed in glycolysis, pentose phosphate pathway, sulfite 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 supply change compared with glucose metabolism, which agrees with experimental findings (Skoog and Hahn-Hägerdal, 1990).

The correlation between reaction essentiality and degree of metabolite is evaluated as the fraction of essential reactions in all the reactions associated with the metabolite. The result for the reconstructed *S. stipitis* model is shown in Figure 5.2. Essentiality is evaluated for glucose (A and C) and xylose metabolism (B and D) respectively.

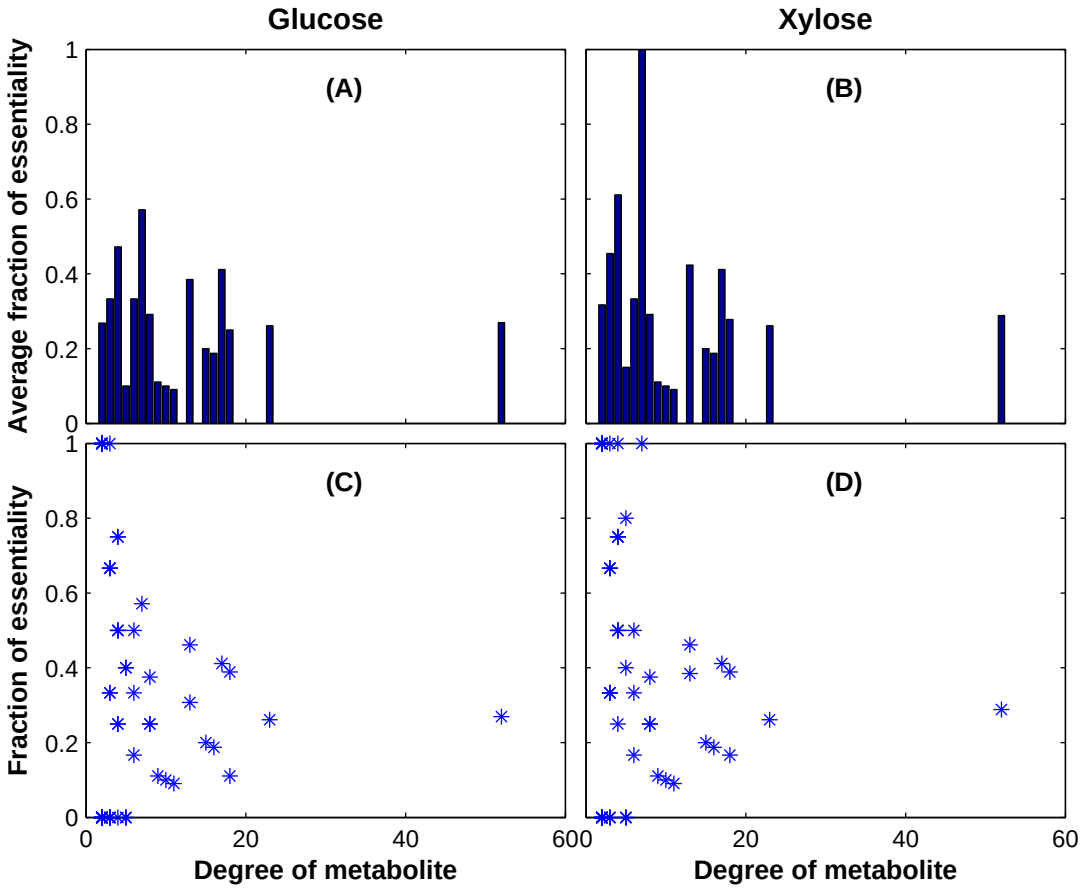

Figure 5.2: Correlation between reaction essentiality and degree of metabolite

The results (Figure 5.2) shows that the small-scale model generally follows the trend that the reactions linked with low degree metabolites have more possibility to be essential no matter what carbon source is utilized. However, due to size of the model, many reactions

have been lumped into the cellmass reaction. Therefore, it is not as clear as the results in genome-scale models. In fact, for the *S. stipitis* core model, most fractions of essentiality lie between 0.1 and 0.5 for the majority of the metabolites, regardless of their degree. There are a few compounds that have a degree of 2 and have a fraction of 1 (e.g., glc[e], 2pg[c]). These metabolites often occur in a linear pathway, at the end of which an essential biomass precursor is produced.

The results show differences between glucose metabolism and xylose metabolism. 8 out of 17 degrees have different average fractions of essentiality. Further investigation shows that the difference is caused by the unique reactions for glucose or xylose metabolism, e.g. the first three steps of xylose metabolism, and also some reactions in non-oxidative PPP, which is not essential to glucose metabolism. The capacity of the model that can grow without these reactions is most likely due to the highly lumped cellmass reaction.

5.2.3 Reaction participation

From the structure of the model, the reaction participation can be calculated, which is defined as the number of metabolites per reaction. For the current model, the average reaction participation is 3.8559, which is very close to the value in *E. coli* metabolic network model (Orth, Fleming, and Palsson, 2010). The number indicates that the most common reaction mode in the model is totally 4 reactants and products and agree with the published results for genome-scale model (Papin, Price, and Palsson, 2002).

5.3 Influence of oxygenation to glucose and xylose metabolism

It is well known that the metabolism of *S. stipitis* is sensitive to oxygenation, especially with xylose (Jeffries and Van Vleet, 2009; Jeffries et al., 2007; Skoog and Hahn-Hägerdal, 1990). Oxygen has also been reported to be important for efficient ethanol production (Grootjen, Lans, and Luyben, 1990; Ligthelm, Prior, and Preez, 1988b; Skoog, Jeppsson, and Hahn-Hägerdal, 1992). Meanwhile, by comparing the different performance of glucose

and xylose metabolism, the impact of the first two steps of xylose metabolism to the internal redox balance can be clearer.

5.3.1 *In silico* experiment design

In order to study the impact of aeration condition, or the robustness of oxygen uptake rate (OUR) in glucose and xylose metabolism, a series of *in silico* simulations were constructed. The intracellular fluxes were calculated with FBA through varying the oxygen uptake rate from 0 to 20 mmol/gDCW/h with a step size of 0.01. The flux of incoming glucose or xylose has an upper limit of 10 mmol/gDCW/h, a realistic value. The flux of uptaking carbon source is set to have an upper limit instead of a fixed value because this will also provide information on how oxygenation will affect sugar input. Specifically for xylose uptake, since it is a symport process, i.e. with proton import simultaneously. Other constraints are set the same values as described above.

The set of experiments resulted in 118×2001 matrix for each carbon source, where each column represents the 118 intracellular fluxes under a certain OUR. The generated results were then analyzed with phenotype analysis to extract biological sensible information.

5.3.2 Cell growth and product formations in glucose and xylose metabolism

With the varying oxygenation, the cell growth and product formations are shown in Figure 5.3 and 5.4 for glucose and xylose metabolism respectively. Based on phenotype analysis, phenotypes have been identified along the aeration conditions and are also marked in Figure 5.3 and Figure 5.4. The detailed descriptions of the characteristics of different phenotypes have been summarized in Table 5.1. These results generally agree well with experiment observations.

From the results in Table 5.1, *S. stipitis* shows to be more sensitive to OUR, which is indicated by more phenotypes under oxygen limited condition.

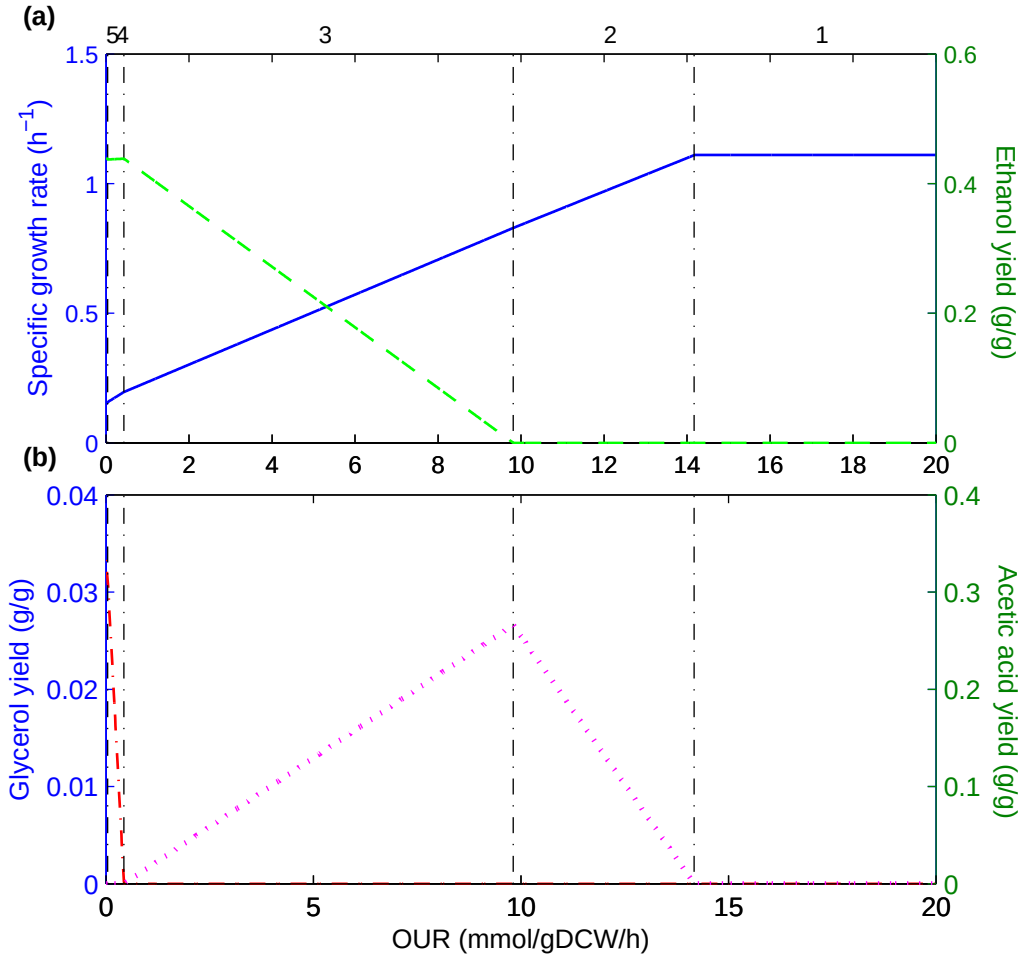

Figure 5.3: Cell growth and product formations in glucose metabolism. Solid line: cell growth; dash line: ethanol yield; dash-dot line (non-verticle): glycerol yield; dot line: acetic acid yield. The vertical dash-dot lines define the boundaries of different phenotypes. The phenotype number is shown at the top of the figure.

Besides the influence to cell growth and product formation, oxygen also impacts the xylose uptake rate but not glucose uptake rate. In glucose metabolism, from anaerobic to aerobic growth, the glucose uptake rate always hit the upper boundary, specifically 10 mmol/gDCW/h. For xylose metabolism, the xylose uptake rate begins to decrease with the decreasing OUR at phenotype 5, which is shown in Figure 5.5. Therefore, in phenotype 5, oxygen is the only limiting factor for cell growth under current setup. This phenomenon is because the uptake process of xylose is symport, i.e. a proton is imported into the cell with

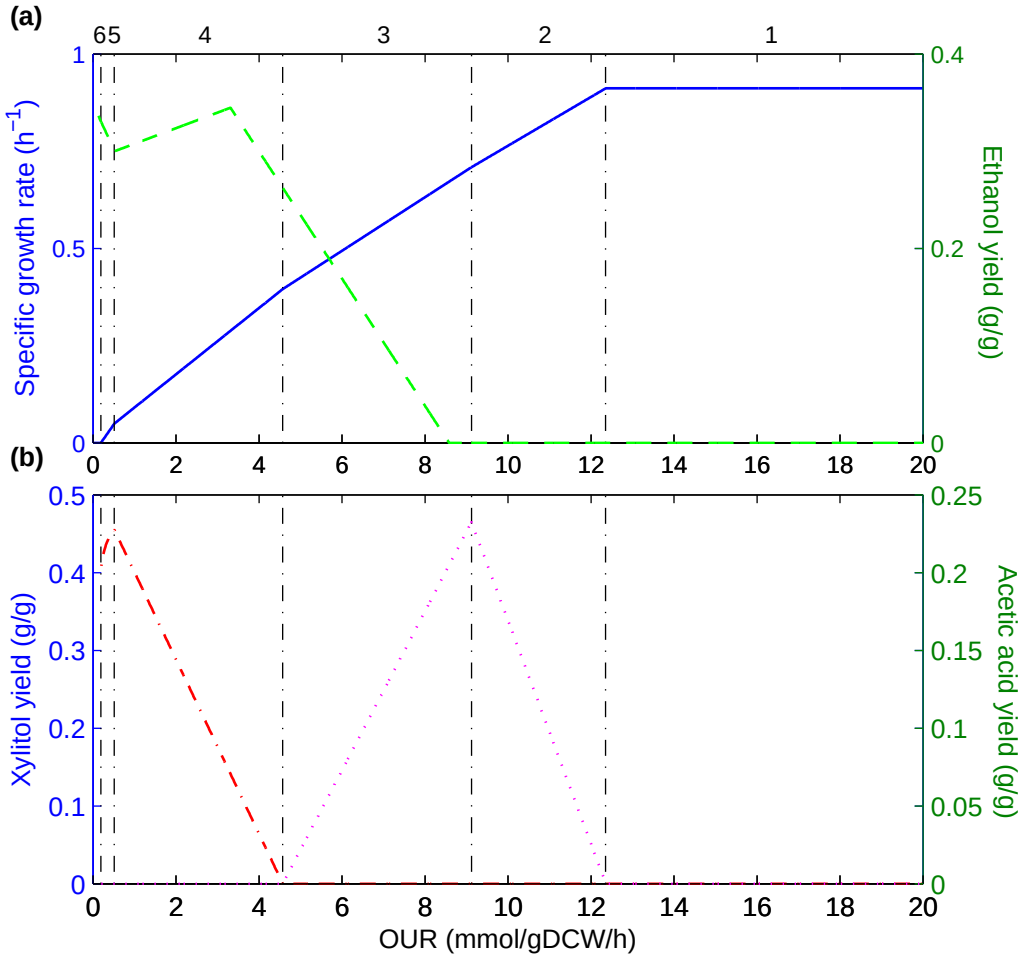

Figure 5.4: Cell growth and product formations in xylose metabolism. Solid line: cell growth; dash line: ethanol yield; dash-dot line (non-verticle): xylitol yield; dot line: acetic acid yield. The vertical dash-dot lines define the boundaries of different phenotypes. The phenotype number is shown at the top of the figure.

one xylose molecule. With the decrease of aeration, the power of oxidative cannot support this perturbation and xylose uptake rate decreases consequently.

5.3.3 Interpreting changes of metabolism among phenotypes

The phenotypes show different physiological characteristics, which indicate that the optimal flux distribution calculated in each phenotype are different. In this part, the flux

Table 5.1: Summary of the characteristics of identified phenotypes

Carbon source	Pheno	Growth limitation	Metabolite product(s)	Main metabolic characteristics
Glucose	1	Glc	X	Aerobic growth
	2	Glc, O_2	X, ac	Increasing acetic acid production
	3	Glc, O_2	X, etoh, ac	Ethanol production and declined acetic acid production
	4	Glc, O_2	X, etoh, glyc	Stable ethanol production and increasing glycerol production
	5	Glc	X, etoh, glyc	Maximal ethanol and glycerol production
Xylose	1	Xyl	X	Aerobic growth
	2	Xyl, O_2	X, ac	Increasing acetic acid production
	3	Xyl, O_2	X, etoh, ac	Ethanol production and declined acetic acid production
	4	Xyl, O_2	X, etoh, xylt	Declined ethanol production and increasing xylitol production
	5	O_2	X, etoh, xylt	Declined ethanol and xylitol production
	6	-	-	Cannot maintain metabolism (no growth)

Glc: glucose, Xyl: xylose, X: cell mass, etoh: ethanol, ac: acetic acid, glyc: glycerol, xylt: xylitol.

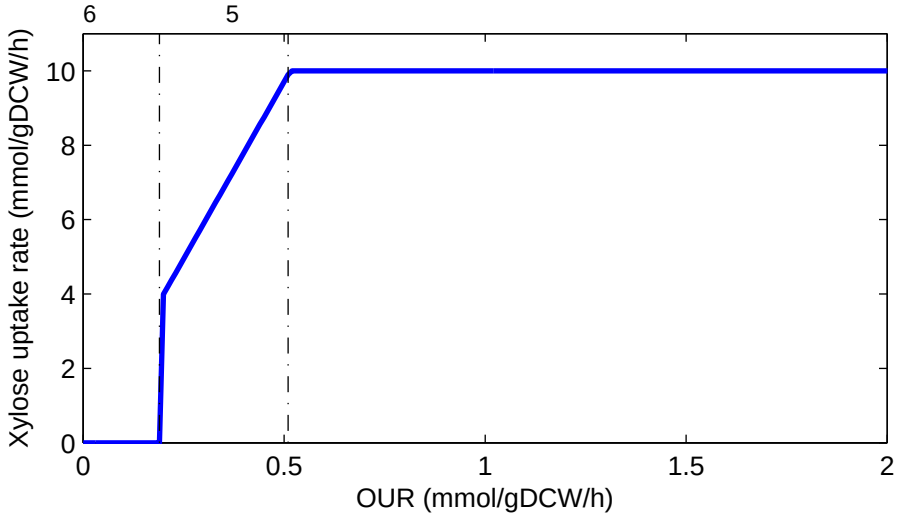

Figure 5.5: Oxygen influence to the xylose uptake rate. The number above the figure is the phenotype number.

distributions of different phenotypes were compared to study the change of intracellular metabolism.

5.3.3.1 The metabolism changes with glucose as carbon source

Phenotype 1 to phenotype 2 From Table 5.1, in this change of phenotype acetic acid is produced. The comparison of flux distribution also confirmed this: the only change is the activation of acetic acid from acetyle coenzyme A, which is a intermediate product of pyruvate metabolism. While with the decrease of oxygenation, the electron transport chain (ETC) cannot utilize all the reductive power generated by TCA cycle. Therefore, acetic acid is produced.

Phenotype 2 to phenotype 3 In this stage, the metabolism of the cell switches from respiratory to fermentation. Part of the TCA cycle has been shut down and an alternative route through glyoxylate cycle is activated. This kind of incomplete (or branched) TCA cycle as shown in Figure 5.6 has been reported in *S. cerevisiae* (Nissen et al., 1997; Vargas et al., 2011). This prediction is further supported by Jeffries et al. (2007), where expressed sequence tags (EST) from oxygen-limited growth of *S. stipitis* on xylose showed that KGD2 (the TCA cycle reaction being passed) was down-regulated. Due to partly block of the TCA cycle, the regeneration of NADH decreases in phenotype 3. The ethanol production pathway is activated to utilize the extra carbon flux decreased through TCA and also to consume cofactor and thus provide another way to redox balance.

Phenotype 3 to phenotype 4 With the further decrease of oxygen supply, the oxidative power of the cell declined as well. So does the cell growth. One consequence of this change is to change the redox balance. The fluxes through ETC decrease even more and most of the carbon is consumed by ethanol production. Therefore, the model adapts to this change by shifting ethanol production way from NADPH-preferred to NADH-preferred reaction. However, this kind of change lacks validation from experiments.

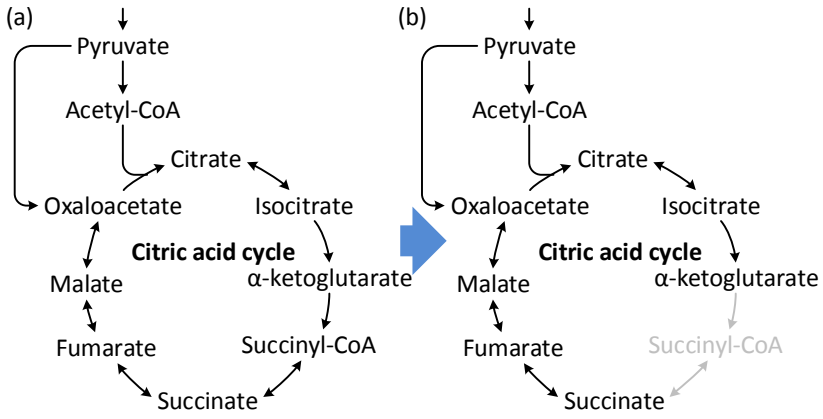

Figure 5.6: TCA cycle change occurred in phenotype 3: (a) complete TCA cycle; (b) branched TCA.

Phenotype 4 to phenotype 5 During this stage, the comparison of flux distribution doesn't reveal any further information besides the decrease fluxes through ETC. Phenotype 5 actually is just a point: the anaerobic fermentation.

5.3.3.2 Differences between phenotypes in xylose metabolism

Phenotype 1 to phenotype 2 The physiological characteristics are the same as in glucose metabolism. However, the activation of acetic acid production is totally different. In glucose metabolism, the acetic acid is produced from acetyle coenzyme A, which is an intermediate metabolite between pyruvate and TCA cycle. While in xylose metabolism, acetic acid is produced via activating production of acetaldehyde and then the generation of acetic acid. Compared with the acetyle coenzyme A pathway, this pathway generates more NADPH, which reflects the influence of oxygenation to redox balance in xylose metabolism.

Phenotype 2 to phenotype 3 The same as the shift from phenotype 2 to phenotype 3, the branched TCA cycle again appears when the xylose metabolism shifts from phenotype 2 to phenotype 3. Except this, the shift of metabolism didn't introduce any other perturbation to the intracellular system.

Phenotype 3 to phenotype 4 During this phenotype change, the decrease of oxidative power leads to decreasing buildup of building blocks and energy. Therefore, the capacity of central carbon metabolism declines. Extra generated xylitol excretes outwards. The regeneration of NADPH through acetic acid production is no longer needed and turned off.

Phenotype 4 to phenotype 5 By simply comparison of flux distribution, it is hard to extract information for this phenotype variation. After carefully examining intracellular fluxes, the shift is mainly due to the decrease of xylose uptake rate, i.e. the redox balance cannot be solved even with xylitol excretion. From Figure 5.4 (a), it shows that the ethanol yield even increases although the specific production rate drops fast with the decreasing oxygen supply.

Phenotype 5 to phenotype 6 At the end of phenotype 5, the system cannot generate enough energy and maintain the redox balance. Therefore, there's no feasible solution for the whole system, i.e. no cell growth predicted.

5.3.3.3 Discussion

From the comparison of the results of phenotype analysis on glucose and xylose metabolism, redox balance is more complicated in xylose metabolism than in glucose metabolism, which is confirmed by more reactions related to cofactor identified. This further causes the sensitivity of xylose metabolism to oxygen.

Phenotype analysis definitely shows its power in the study on influences of oxygen. It can easily identify the phenotypes when one flux varies and is a useful tool for the robustness analysis. However, the shortcoming of phenotype analysis is also shown clear here. Very little information can be extracted from the phenotype identified. It is hard to find out how the system responds to the perturbation introduced. For example, which reactions are most responsible for the cofactor balance change? In this small scale model, it is possible but still very hard to answer the question by examining the intracellular fluxes carefully.

But what if for a genome-scale model? It is very difficult or even impossible to extract biological information from hundreds of fluxes. Therefore, in next chapter we proposed a system identification based framework for this purpose.

5.4 Conclusion

In this chapter, flux balance analysis, robustness analysis and phenotype analysis have been applied to the metabolic network model constructed in the previous chapter. The topoligical properties of the model was first studied and the results confirmed that the constructed small-scale network model shows similar properties as in genome-scale model. The model also shows some variations which is caused by the combination of reactions in the construction process. The oxygen influences to glucose and xylose metabolism were also studied thereafter. Robustness analysis on glucose and xylose metabolism shows that xylose metabolism is more sensitive to oxygen supply. The best aeration condition identified by model simulation is anaerobic while the one for xylose metabolism is more complicated and requires careful control on oxygen supply rate. Different phenotypes have been identified and some biological sensible information has been revealed by comparison of flux distribution. While it shows the power of phenotype analysis, the disadvantage of this method also becomes obvious. A new approach to analyze the results from FBA simulations and thus to extract biological sensible information is needed, which is proposed and applied to elucidate xylose metabolism in next chapter.

Chapter 6
Study of redox balance in xylose metabolism

6.1 Introduction

The catabolism of sugars by microorganisms is accomplished by a variety of metabolic pathways. Yeasts, as a group, are more homogeneous with respect to sugar catabolism than are bacteria. All yeasts described so far are able to grow on glucose. Invariably, the major portion of this sugar is catabolized via the Embden-Meyerhof pathway; respiration proceeds only with oxygen as the terminal electron acceptor, and if fermentation occurs, ethanol is the major end product. Despite these similarities, however, many differences may be observed between different yeasts, especially with respect to the ability to utilize various sugars and the regulation of respiration and fermentation.

In the metabolism of sugars by yeasts the nicotinamide adenine dinucleotides NAD(H) and NADP(H) play separate and distinct roles. NADH may be regarded as a predominantly catabolic reducing equivalent, whereas NADPH is mainly involved in anabolic processes (Dijken et al., 1986). This is not always the case, however. In fact, the distinction between assimilatory and dissimilatory reactions in heterotrophic organisms is to some extent artificial. For example, glycolysis plays an essential role in sugar dissimilation, but also generates building blocks for biosynthesis. Furthermore, although most biosynthetic reactions use NADPH as a reductant, some NADH-linked reductions occur in the conversion of central metabolites (pyruvate, oxaloacetate, acetyl CoA) to cellular monomers, for example in amino acid biosynthesis (Bakker et al., 2001).

Under conditions of oxygen depletion, NADH generated in glycolysis can be re-oxidized in the conversion of pyruvate to ethanol and CO_2. In the presence of oxygen many yeasts do not form ethanol and NADH, generated during catabolism, is re-oxidized with oxygen. Since catabolic and anabolic pathways share the initial reactions of sugar metabolism, NADH is also formed during the assimilation of sugars to cell material. The formation of NADH during assimilation is even higher than is anticipated on the basis of a comparison of the reduction levels of sugar and biomass. This is due to the fact that the NADH produced during the formation of intermediates of glycolysis and TCA cycle is not the principal reductant for the conversion of these intermediates to the building blocks of cell polymers.

The specific requirement for NADPH in the assimilation of sugars to cell material, in combination with the absence of transhydrogenase activity, necessitates the conversion of part of the sugar exclusively for the purpose of generating reducing power in the form of NADPH. This is accomplished in the oxidative steps of the hexose monophosphate pathway. Thus, in the overall process of aerobic growth and biomass formation, two separate flows of reducing equivalents can be distinguished: production of NADH for the purpose of ATP formation and production of NADPH for reductive processes in the cell's anabolism, mainly in the synthesis of amino acids and fatty acids. A similar scheme holds for anaerobic growth. In this case, however, NADH plays no direct role in ATP formation and is re-oxidized in the final reaction of the alcoholic fermentation.

Apart from ethanol, yeasts may excrete a variety of metabolic products. These include polyalcohols (glycerol, erythritol, arabinitol, xylitol, ribitol), monocarboxylic acids (mainly acetic and pyruvic acid), dicarboxylic and tricarboxylic acids (succinic, citric, and isocitric acid). Generally, the metabolic basis for the formation of these products and the fate of reducing power during their synthesis is poorly understood. Excretion of metabolites can occur under either aerobic or anaerobic conditions, and is dependent on the particular species and on environmental conditions.

In this chapter, the redox balance in *S. stipitis* will be investigated with the reconstructed model. The redox balance influence to the distribution of products and cell growth will be investigated.

6.2 Methods

6.2.1 Flux Balance Analysis (FBA)

Flux balance analysis was performed to study the central carbon metabolism of *S. stipitis* using a publicly available COBRA toolbox for MATLAB version 2.05 (Schellenberger et al., 2011). 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 simulation results are analyzed further to reveal the intracellular mechanism of xylose metabolism.

6.2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is commonly used multivariate analysis method of dimension reduction, which extracts the directions corresponding to the largest variations among different variables in a high dimensional data set (Jolliffe, 2002). It has been applied in the metabolomics studies to analyze metabolites profiles at given conditions (Griffin, 2004). 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.

6.2.3 Proposed method: FBA-PCA

In microbial metabolism, hundreds and even thousands 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 to 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. To fill the gap, we propose a system identification based metabolic flux analysis framework to extract such knowledge by integrating PCA with FBA in this work. Specifically, in the proposed framework, we first design *in silico* experiments to perturb the metabolic network in order to investigate the interested properties, 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 sufficient amount of *in silico* experimental data in the 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. In Appendix D, an illustrative example explains how the proposed method works.

6.3 Elucidating influence of oxygen in xylose metabolism with FBA-PCA

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* (Skoog and Hahn-Hägerdal, 1990). 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.

6.3.1 Designed *in silico* experiments

In order to study how different oxygen availability affects cellular metabolism, we performed FBA to calculate the intracellular fluxes by varying the oxygen uptake rate from 0 to 20 mmol/gDCW/h with a step of 0.001,i.e. totally 20001 runs of experiments. This set of experiments resulted in a 118 $\times$ 2001 matrix, where each column represents the 118 intracellular fluxes under a certain OTR and a xylose uptake rate with upper limit of 10 mmol/gDCW/h. In the model, the flux distribution ratio through NADPH-dependent and NADH-dependent xylose reductase (XR) was set to 1.0 while the xylitol dehydrogenase was supposed to be NAD^+-dependent only. Phenotypical Phase Plane Analysis (PhPP) (BELL and PALSSON, 2005; Edwards, Ramakrishna, and Palsson, 2002) was also carried out under the same conditions for comparison.

6.3.2 Phenotype identification

PCA is applied to analyze the *in silico* experimental results. As shown in Figure 6.1 (a) where scores corresponding to the first two PCs are plotted, totally 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 represent 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 results from PhPP is given in Figure 6.1 (b), where the same six phenotypes are identified. Figure 6.1 (c) plots the 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 6.1.

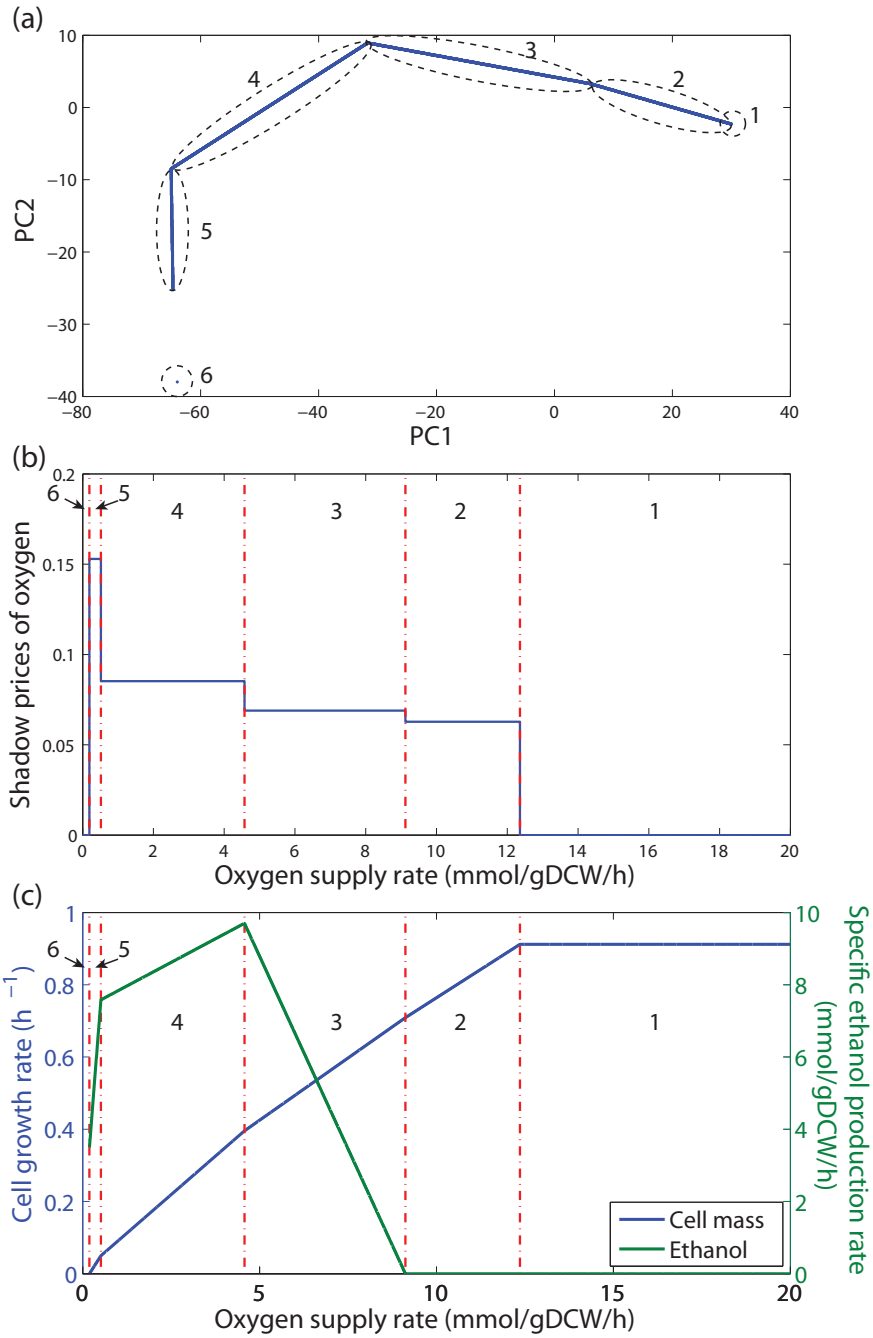

Figure 6.1: Phenotypes identified with PCA when OUR changes within [0, 20] mmol/gDCW/h. (a) phenotypes identified by PCA; (b) phenotypes identified by PhPP; (c) model predicted cell growth rates and specific ethanol production rates. The numbers 1-6 correspond to the identified phenotypes.

Although PhPP and the proposed approach identify the same 6 phenotypes, they are completely different in revealing the cellular details that underlie the specific phenotype. For

Table 6.1: Summary of the 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)

PhPP, it can easily identify whether oxygen or carbon source is a limiting factor by examining the shadow price (Edwards, Covert, and Palsson, 2002), but it is very difficult 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 approach is that it can reveal the cellular details, particularly the key reactions that differentiate different phenotypes, by examing the loading matrix.

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 6.2, where the metabolic fluxes are colored according to their loadings.

From Figure 6.2, 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 turning off of 2-oxoglutarate dehydrogenase due to decreased oxygen supply in phenotype 3, which further leads to an incomplete (or branched)

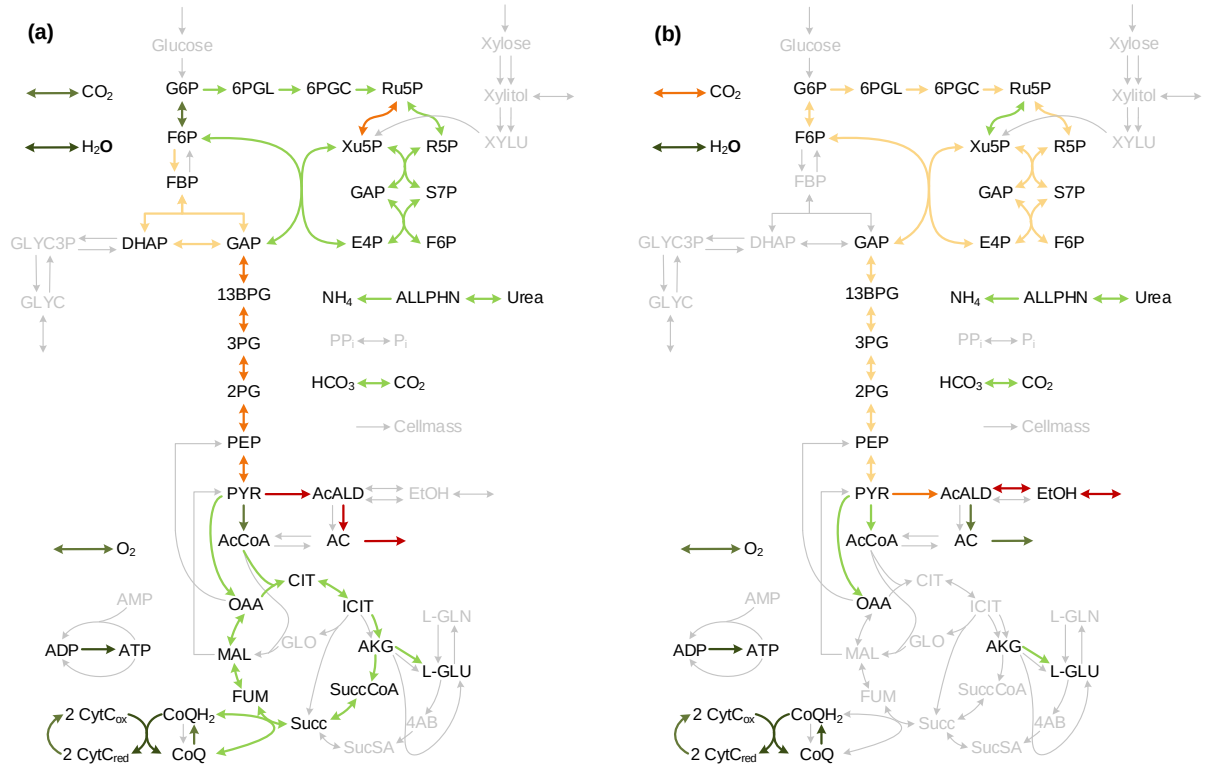

Figure 6.2: Metabolic maps for phenotype2 (a) and phenotype 3 (b) identified in Figure 6.1.

TCA cycle as shown in Figure 6.3 (also shown as gray TCA cycle in Figure 6.2 (b) but green TCA cycle in Figure 6.2 (a)). This branched TCA cycle has been previously reported in *S. cerevisiae* (Nissen et al., 1997; Vargas et al., 2011). This prediction is further supported by Jeffries et al. (2007), where expressed sequence tags (EST) from oxygen-limited growth of *S. stipitis* on xylose showed that KGD2 (the TCA cycle reaction being passed) was down-regulated. Second, fermentation pathway, i.e. ethanol production, has been activated by the branched TCA cycle to resolve the redox balance of $NADH/NAD^+$ which are indicated by gray in phenotype 2 and red in phenotype 3. Third, due to the decrease of cell growth, the requirement of NADPH has been reduced and caused the down-regulation of fluxes through pentose phosphate pathway as shown in Figure 6.2 by the color of PPP changing from light green in phenotype 2 to light yellow in phenotype 3.

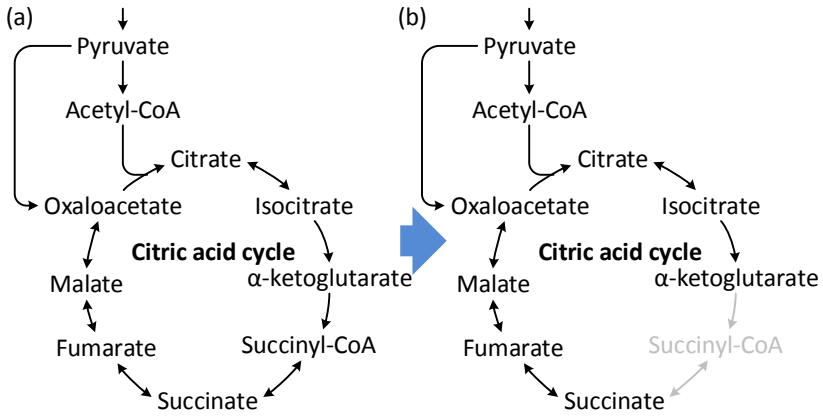

Figure 6.3: TCA cycle change occurred in phenotype 3: (a) complete TCA cycle; (b) branched TCA.

6.3.3 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 6.1 shows that ethanol production is the most sensitive to OUR in this range. We first conducted a series of *in 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 6.2. The loadings corresponding to the involved reactions are plotted in Figure 6.4. 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 6.3 for two conditions with OUR of 0.2 and 0.5 mmol/gDCW/h. The seven key reactions identified in Table 6.3 cover 99% of the total redox shift. The metabolic map with identified key reactions for phenotype 5 is shown in Figure 6.5. Both Figure 6.5 and Table 6.3 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.

Table 6.2: All reactions that involve cofactor consumption and regeneration

No.	Abbreviation	Formula
1	Cellmass_PsScC	4.4318 glu[c] + 3.1644 accoa[c] + 73.3883 h2o[c] + ...
2	AKGDH	nad[c] + akg[c] + coa[c] → nadh[c] + co2[c] + succoa[c]
3	ICITDHxm	nad[c] + icit[c] → akg[c] + nadh[c] + co2[c]
4	ICITDHym	nadp[c] + icit[c] → nadph[c] + akg[c] + co2[c]
5	MDH	nad[c] + mal[c] ⇔ oaa[c] + h[c] + nadh[c]
6	GDHx	nh4[c] + akg[c] + h[c] + nadh[c] → glu[c] + h2o[c] + nad[c]
7	GDHy	nadph[c] + nh4[c] + akg[c] + h[c] → glu[c] + h2o[c] + nadp[c]
8	GOGATx	gln[c] + akg[c] + h[c] + nadh[c] → 2 glu[c] + nad[c]
9	SSADHy	h2o[c] + nadp[c] + sucsal[c] → nadph[c] + 2 h[c] + succ[c]
10	GLYC3PDHx	h[c] + nadh[c] + dhap[c] → nad[c] + glyc3p[c]
11	ADHx	h[c] + nadh[c] + acald[c] ⇔ nad[c] + etoh[c]
12	ADHy	nadph[c] + h[c] + acald[c] ⇔ nadp[c] + etoh[c]
13	GAPDHx	nad[c] + pi[c] + gap[c] ⇔ h[c] + nadh[c] + 13bpg[c]
14	PDHm	nad[c] + pyr[c] + coa[c] → accoa[c] + nadh[c] + co2[c]
15	CPLXI	5 h[c] + nadh[c] + q[c] → nad[c] + qh2[c] + 4 h[e]
16	G6PDH	g6p[c] + nadp[c] → nadph[c] + h[c] + 6pgl[c]
17	GND	nadp[c] + 6pgc[c] → nadph[c] + ru5p[c] + co2[c]
18	ALDx	h2o[c] + nad[c] + acald[c] → 2 h[c] + nadh[c] + ac[c]
19	ALDy	h2o[c] + nadp[c] + acald[c] → nadph[c] + 2 h[c] + ac[c]
20	MAExm	nad[c] + mal[c] → pyr[c] + nadh[c] + co2[c]
21	XDHx	nad[c] + xylt[c] ⇔ h[c] + nadh[c] + xylu[c]
22	XRx	h[c] + nadh[c] + xyl[c] → nad[c] + xylt[c]
23	XRy	nadph[c] + h[c] + xyl[c] → nadp[c] + xylt[c]
24	XDHy	nadp[c] + xylt[c] ⇔ nadph[c] + h[c] + xylu[c]

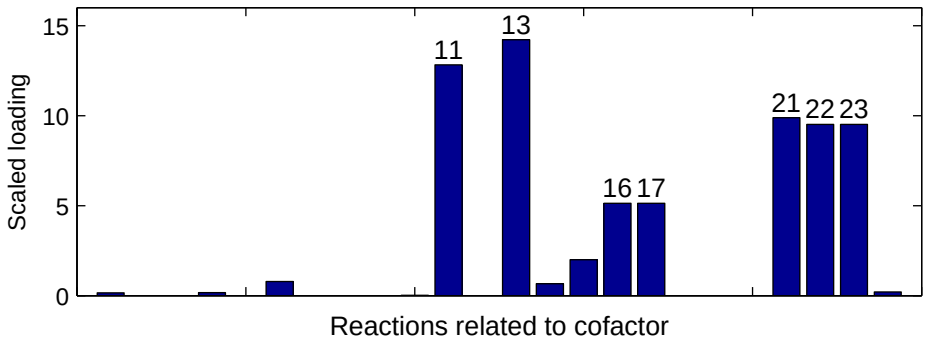

Figure 6.4: The loadings of the reactions involved in cofactor consumption and regeneration with varying OUR

Table 6.3: Shift of cofactor consumption and regeneration in phenotype 5

Cofactor	**Reaction**	**OUR = 0.200**	**OUR = 0.500**	**Total shift**
NADH consumption	R11	2.00	4.85	6.63
	R22	3.61	7.39	
NADH regeneration	R13	2.38	5.36	7.16
	R21	3.63	7.81	
NADPH consumption	R23	2.00	4.85	2.85
NADPH regeneration	R16	1	2.57	3.14
	R17	1	2.57	

6.4 Influence of cofactor specificity of xylose reductase

As discussed before in Section 4.2.4, the flux ratio of XR is very important to xylose metabolism with *S. stipitis*. Many results on experimental study of XR preference to NADPH and NADH have been reported (Hou, 2012; Slininger et al., 2011; Verduyn et al., 1985; Yablochkova et al., 2003, 2004). However, the reported results are not consistent with each other (Slininger et al., 2011; Yablochkova et al., 2004). In additional, all reported enzyme activities are measured *in vitro* with saturated substrate concentration in dilute solution. These all deviate from the *in vivo* condition within the cell. Meanwhile, researchers have tried to apply protein engineering to alter the cofactor preferences of XR to improve

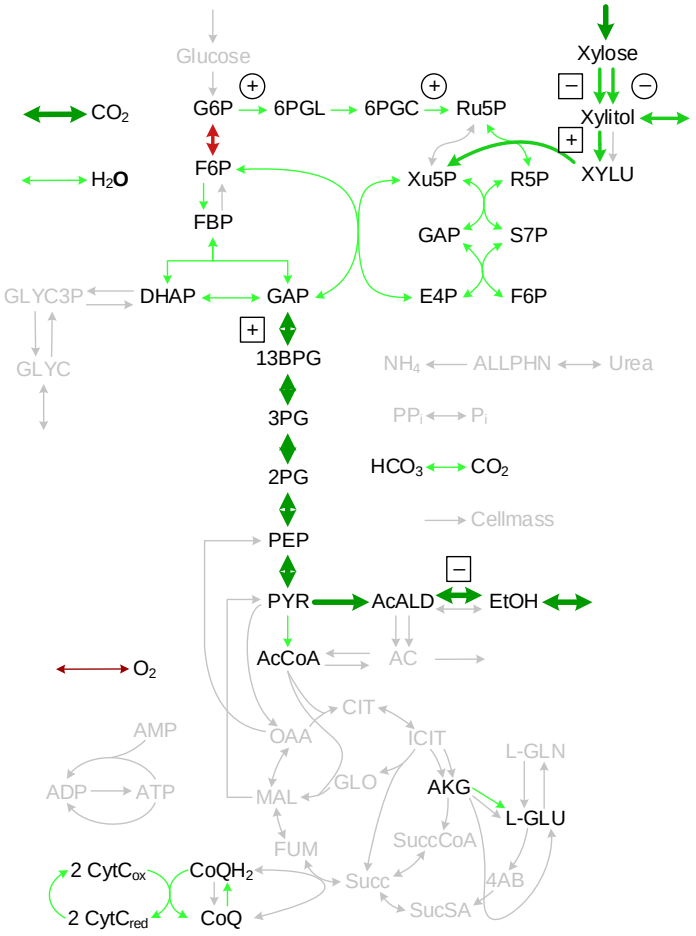

Figure 6.5: Metabolic map for phenotype 5 with key reactions (with thick green arrows) identified by FBA-PCA. The green color indicates positive loading; red means negative loading. Open square implies the reaction related to NAD(H) while circle for reactions related to NADP(H). "+" means regeneration and "−" represents consumption.

ethanol production and/or to reduce by-product productions (Bengtsson, Hahn-Hägerdal, and Gorwa-Grauslund, 2009; Chu and Lee, 2007; Krahulec et al., 2012; Liang, Zhang, and Lin, 2007; Matsushika et al., 2009; Watanabe, Kodaki, and Makino, 2005; Watanabe et al., 2007a). Therefore, studying the influence of cofactor specificity of XR by altering the flux ratio will 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.

Here we define XR activity ratio (R_{XR}) 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. Based on the reported results and general knowledge of the *in vivo* concentrations of NADH/NAD^+ and NADPH/$NADP^+$ pools (Bergdahl et al., 2012), in this section we first vary R_{XR} within [0, 2] and study its effect on redox balance and ethanol production.

First we performed simulations to study the general influence of R_{XR} to model predictions under various oxygenation conditions through FBA. In these experiments, xylose uptake rate is constrained to be 10 mmol/gDCW/h, oxygen supply rate is changed between 0 to 14 mmol/gDCW/h with a step of 0.1, while R_{XR} is varied between [0, 2] with a step of 0.2 plus 10 as an extreme case. The resulted cell growth, ethanol production and xylitol production are shown in Figure 6.6. It shows that the increase of NADH affinity of XR can improve the ethanol production and reduce xylitol production. The results show different patterns in ethanol production rate caused by different ratios through the reactions (shown in Figure 6.6 (b) as different combination of increase and decrease), which can be used for experimental validation and thus provide insights on flux ratio through different cofactor-linked reactions and intracellular cofactor pool size.

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/gDCW/h respectively. The activity ratio of NADPH- and NADH-linked XR is changed incrementally within [0, 2] with a step of 0.001, which results in 2001 in silico experiments. PCA was applied to identify the key changes among different reactions when the ratio is changed. The loadings corresponding to the reactions involving cofactor consumption and regeneration are plotted in Figure 6.7. The loadings are 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 6.4 for two conditions with XR flux of 0.5 and 2.0. The seven key reactions identified in Table 6.4 cover 98% of the total redox shift. The map with identified key reactions is shown in Figure 6.8.

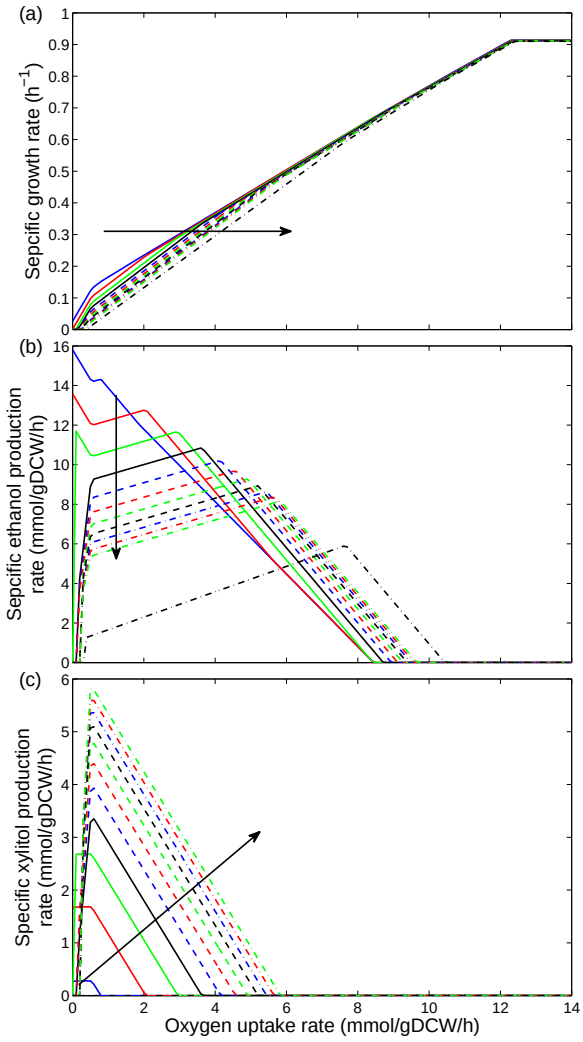

Figure 6.6: Influences of XR flux ratio on specific cell growth rate (a), specific ethanol production rate (b) and specific xylitol production rate (c) with varying aeration. The arrows in the plots indicate the increase of XR flux ratio.

6.5 Conclusion

In this chapter, to investigate how cellular redox balance is affected by change in 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

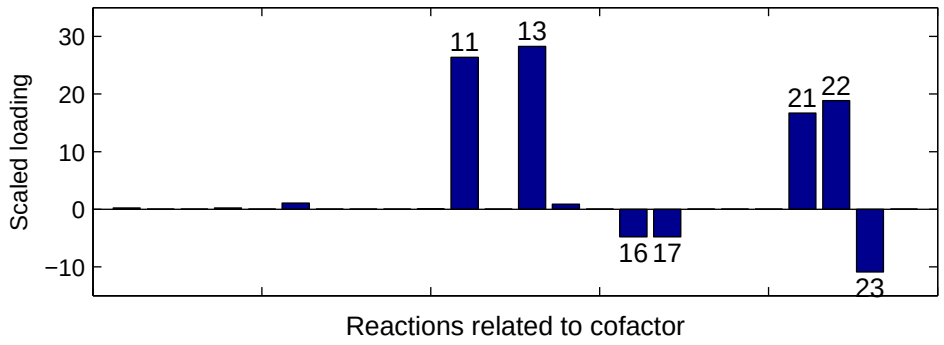

Figure 6.7: The loadings of the reactions involved in cofactor consumption and regeneration with varying XR ratio. Numbers in the figure correspond to the reaction numbers in Table 6.2.

Table 6.4: Shift of cofactor consumption and regeneration

Cofactor	**Reaction**	$R_{XR} = 0.5$	$R_{XR} = 2.0$	**Total shift**
NADH consumption	R11	6.67	2.64	-9.66
	R22	10.1	4.47	
NADH regeneration	R13	6.92	3.32	-9.63
	R21	10.6	4.57	
NADPH consumption	R23	3.33	5.28	1.95
NADPH regeneration	R16	1.83	2.68	1.7
	R17	1.83	2.68	

that dominant the cellular redox balance. It is interesting to find out that under oxygen-limited condition, it is the same set of the key reactions that dominate the redox shift caused by change of OUR or change of 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 (totally 7 reactions) should be considered together when designing mutant to improve ethanol yield through shifting cellular redox balance.

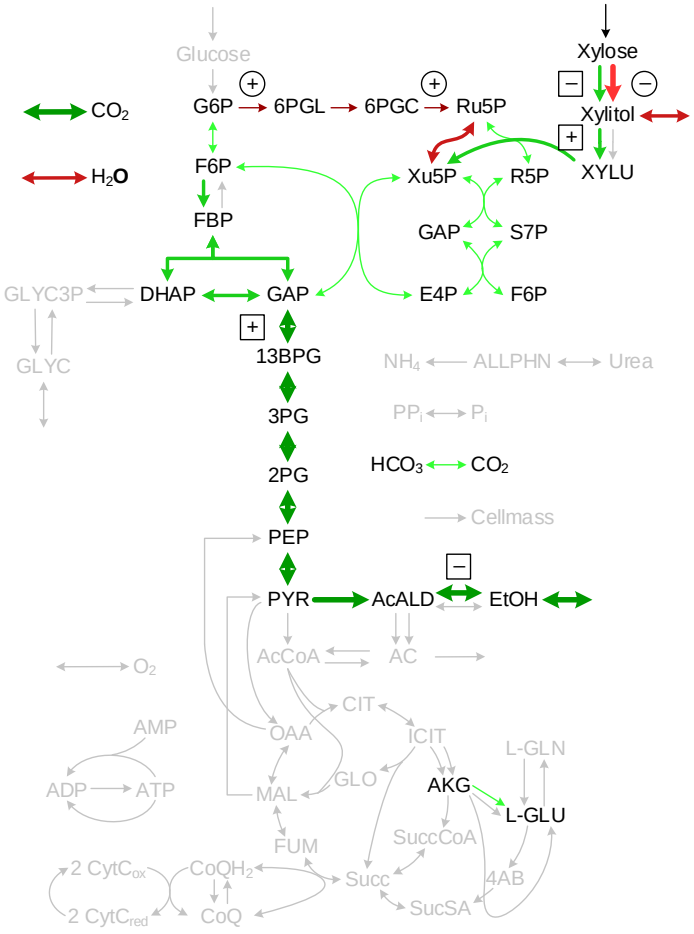

Figure 6.8: Metabolic map with key reactions (with thick arrows) identified by FBA-PCA. The green color indicates positive loading; red means negative loading. Open square implies the reaction related to NAD(H) while circle for reactions related to NADP(H). "+" means regeneration and "−" represents consumption.

Chapter 7
Conclusion and Outlook

7.1 Conclusion

Lignocellulosic ethanol production represents an attractive alternative source for long-term renewable energy supply due to the rising concerns over energy sustainability, global warming and feed stock availability. However, many barriers exist for industrializing lignocellulosic ethanol processes and one of them is the effective conversion of xylose, the second abundant mono-saccharide and representative pentose in the hydrolysate of lignocellulosic biomass.

As the most promising native strain for xylose fermentation, *Scheffersomyces stipitis* shows good overall performance on lignocellulosic hydrolysate. Understanding its metabolism, especially the central carbon metabolism, is very important to improve the strain, or to provide hints for metabolic adjustment of other strains. In this work, the glucose and xylose metabolism of *S. stipitis* have been studied via construction of a central carbon metabolic network as well as wet experiments.

First, the glucose and xylose metabolism were studied via experiments. Obstacles exist for industrial adoptation of *S. stiptis*, e.g. low ethanol tolerance, sensitive to oxygen supply and low growth rate under micro-aeration condition. To solve the problem, in this work, continue culture with cell retention module have been carried out with both glucose and xylose. The system provides a high cell retention ratio and can make cell evolving for long time and in harsh environment without considering the possible wash out. Meanwhile, the continuous application of environmental selection pressure makes the strain evolve in the

desired direction. *S. stipitis* has been cultured in the system for more than 3 months, which improved the ethanol tolerance significantly. Various approaches have been evaluated for the ethanol tolerance improvement. In the experiment, it also confirmed the high sensitivity of xylose metabolism in *S. stipitis* to oxygen. To produce ethanol in high yield, an accurate control over oxygen supply is a necessary.

To study the influence of environmental perturbations to the metabolism of *S. stipitis* systematically, a stoichiometric central carbon metabolic network is reconstructed. During the construction, futile cycles in the model have been identified. The impact of particular constraints, i.e. non-growth-associated maintenance energy and flux coupling constraints, has been studied and proper values have been adopted from literature data. The model is validated against experimental results reported in literature. Even though the reconstructed metabolic network model does not capture all the cellular details, for example, it only considers one compartment (the cytosol) and does not contain any gene regulatory mechanism; it still provides a comprehensive picture of the central carbon metabolism of *S. stipitis*. Such model enables us to elucidate the xylose metabolism using a systems approach.

The model properties have also been investigated after the reconstruction. The topological properties generally agree with that of the genome-scale model. Some variants exist due to the model size. Flux balance analysis, robustness analysis and phenotype analysis have been combined together to study the oxygen influence to the glucose and xylose metabolism. The changes of intracellular metabolism with various aeration conditions have been studied by phenotype identification and comparison of flux distribution. However, this approach shows its limitation in our research. Therefore, a system identification based metabolic flux analysis framework, FBA-PCA, has been proposed and applied to study the influence of OUR to xylose metabolism. The application of FBA-PCA shows clearly that it can extract the underlying biological knowledge embedded in the network structure. By applying the proposed framework, we were able to identify the key reactions that dominant the cellular redox balance. Meanwhile, the proposed framework is utilized to analyze the impact of change

of XR cofactor specificity, which has been considered to be an important aspect for xylose metabolism. It is interesting to find out that under oxygen-limited condition, it is the same set of the key reactions that dominate the redox shift caused by change of OUR or change of 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. The set of key reactions (totally 7 reactions related to cofactors) should be considered together when designing mutant to improve ethanol yield through shifting cellular redox balance.

7.2 Outlook

As mentioned above, the constructed model has some intrinsic disadvantages, e.g. the lack of mitochondrial compartment and regulation information integration. Therefore, to continue study on this model, several research directions show interesting future.

Improvement of the model Although several genome-scale model of *S. stipitis* have been published, they have more or less variances from the experimental phenotypes and literature data. The model needs more development so that more biological knowledge could be extracted *in silico* and work together with experimental results to promote the understanding to the strain.

Dynamic model development Approaches based on flux balance analysis or optimization-based methods suppose that the intracellular metabolism is in steady state. However, this is usually not the case with the environmental perturbations. Combining the kinetics information from wet experiments, metabolic network model may be applied in wider range. Also, the study of dynamic properties systematically will also promote the development of human knowledge.

Application of FBA-PCA on other system FBA-PCA has been proven to be effective to extract underlying biological knowledge in our research. With many genome-scale models constructed, it can be applied to other biological systems to study the metabolism details.

Bibliography

Agbogbo, Frank K and Coward-Kelly, Guillermo (2008). Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast, Pichia stipitis. *Biotechnology letters*, 30 (9), pp. 1515–1524. DOI: 10.1007/s10529-008-9728-z.

Agbogbo, Frank K and Wenger, Kevin S (2006). Effect of pretreatment chemicals on xylose fermentation by Pichia stipitis. *Biotechnology letters*, 28 (24), pp. 2065–2069. DOI: 10.1007/s10529-006-9192-6.

Alexander, M A, Chapman, T W, and Jeffries, T W (1987). Continuous ethanol production from D-xylose by Candida shehatae. *Biotechnology and bioengineering*, 30 (5), pp. 685–691. DOI: 10.1002/bit.260300515.

BELL, S and PALSSON, B (2005). Phenotype phase plane analysis using interior point methods. *Computers and Chemical Engineering*, 29 (3), pp. 481–486. DOI: 10.1016/j.compchemeng.2004.08.019.

Bakker, B M et al. (2001). Stoichiometry and compartmentation of NADH metabolism in Saccharomyces cerevisiae. *FEMS microbiology reviews*, 25 (1), pp. 15–37. DOI: 10.1111/j.1574-6976.2001.tb00570.x.

Balagurunathan, Balaji et al. (2012). Reconstruction and analysis of a genome-scale metabolic model for Scheffersomyces stipitis. *Microbial cell factories*, 11 (1), p. 27. DOI: 10.1186/1475-2859-11-27.

Barabási, Albert-László and Oltvai, Zoltán N (2004). Network biology: understanding the cell's functional organization. *Nature reviews. Genetics*, 5 (2), pp. 101–113. DOI: 10.1038/nrg1272.

Basaran, P et al. (2001). Cloning and heterologous expression of xylanase from Pichia stipitis in Escherichia coli. *Journal of applied microbiology*, 90 (2), pp. 248–255. DOI: 10.1046/j.1365-2672.2001.01237.x.

Basaran, Pervin and Ozcan, Meltem (2008). Characterization of beta-xylosidase enzyme from a Pichia stipitis mutant. *Bioresource technology*, 99 (1), pp. 38–43. DOI: 10.1016/j.biortech.2006.11.056.

Bazaraa, Mokhtar S., Jarvis, John J., and Sherali, Hanif D. (2009). *Linear Programming and Network Flows*. 4th ed. Wiley, p. 768. ISBN: 978-0470462720.

Becker, Scott A et al. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2 (3), pp. 727–738. DOI: 10.1038/nprot.2007.99.

Bengtsson, Oskar, Hahn-Hägerdal, Bärbel, and Gorwa-Grauslund, Marie F (2009). Xylose reductase from Pichia stipitis with altered coenzyme preference improves ethanolic xylose fermentation by recombinant Saccharomyces cerevisiae. *Biotechnology for biofuels*, 2 (1), p. 9. DOI: 10.1186/1754-6834-2-9.

Berg, Jeremy M, Tymoczko, John L, and Stryer, Lubert (2006). *Biochemistry*. 6th ed. W. H. Freeman, p. 1026. ISBN: 0716787245.

Bergdahl, Basti et al. (2012). Dynamic metabolomics differentiates between carbon and energy starvation in recombinant Saccharomyces cerevisiae fermenting xylose. *Biotechnology for biofuels*, 5 (1), p. 34. DOI: 10.1186/1754-6834-5-34.

Bicho, Paul A et al. (1988). Induction of Xylose Reductase and Xylitol Dehydrogenase Activities in Pachysolen tannophilus and Pichia stipitis on Mixed Sugars. *Applied and environmental microbiology*, 54 (1), pp. 50–54.

Birch, Rosslyn M and Walker, Graeme M (2000). Influence of magnesium ions on heat shock and ethanol stress responses of Saccharomyces cerevisiae. *Enzyme and Microbial Technology*, 26 (9-10), pp. 678–687. DOI: 10.1016/S0141-0229(00)00159-9.

Boles, Eckhard and Hollenberg, Cornelis P (1997). The molecular genetics of hexose transport in yeasts. *FEMS microbiology reviews*, 21 (1), pp. 85–111. DOI: 10.1111/j.1574-6976.1997.tb00346.x.

Bolton, Evan E et al. (2008). "PubChem : Integrated Platform of Small Molecules and Biological Activities". *Annual Reports in Computational Chemistry*. Vol. 4. Washington DC: American Chemical Society. Chap. 12.

Bühler, Bruno et al. (2008). NADH availability limits asymmetric biocatalytic epoxidation in a growing recombinant Escherichia coli strain. *Applied and environmental microbiology*, 74 (5), pp. 1436–46. DOI: 10.1128/AEM.02234-07.

Burgard, Anthony P and Maranas, Costas D (2003). Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and bioengineering*, 82 (6), pp. 670–677. DOI: 10.1002/bit.10617.

Burgard, Anthony P et al. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome research*, 14 (2), pp. 301–312. DOI: 10.1101/gr.1926504.

Butler, Michael (2005). Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. *Applied microbiology and biotechnology*, 68 (3), pp. 283–291. DOI: 10.1007/s00253-005-1980-8.

Casey, G P and Ingledew, W M (1986). Ethanol tolerance in yeasts. *Critical reviews in microbiology*, 13 (3), pp. 219–280. DOI: 10.3109/10408418609108739.

Caspeta, Luis et al. (2012). Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in-silico evaluation of their potentials. *BMC Systems Biology*, 6 (1), p. 24. DOI: 10.1186/1752-0509-6-24.

Chandel, Anuj Kumar et al. (2007). Economics and environmental impact of bioethanol production technologies : an appraisal. *Molecular Biology*, 2 (February), pp. 14–32.

ChemAxon. *Calculator Plugins*.

Cherubini, Francesco (2010). The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Conversion and Management*, 51 (7), pp. 1412–1421. DOI: 10.1016/j.enconman.2010.01.015.

Cho, Jae-Yong and Jeffries, Thomas W (1999). Transcriptional Control of ADH Genes in the Xylose-Fermenting Yeast Pichia stipitis. *Appl. Environ. Microbiol.* 65 (6), pp. 2363–2368.

Cho, Jae-yong and Jeffries, Thomas W (1998). Pichia stipitis Genes for Alcohol Dehydrogenase with Fermentative and Respiratory Functions. *Appl. Environ. Microbiol.* 64 (4), pp. 1350–1358. DOI: 0099-2240/98/04.00+0.

Chu, Byron C H and Lee, Hung (2007). Genetic improvement of Saccharomyces cerevisiae for xylose fermentation. *Biotechnology advances*, 25 (5), pp. 425–441. DOI: 10.1016/j.biotechadv.2007.04.001.

Chuang, Han-Yu, Hofree, Matan, and Ideker, Trey (2010). A decade of systems biology. *Annual review of cell and developmental biology*, 26, pp. 721–744. DOI: 10.1146/annurev-cellbio-100109-104122.

Ciesarová, Z et al. (1998). Comparison of ethanol tolerance of free and immobilized Saccharomyces uvarum yeasts. *Folia Microbiologica*, 43 (1), pp. 55–58. DOI: 10.1007/BF02815543.

Clark, James H (2007). Green chemistry for the second generation biorefinerysustainable chemical manufacturing based on biomass. *Journal of Chemical Technology & Biotechnology*, 82 (7), pp. 603–609. DOI: 10.1002/jctb.1710.

Clark, James H, Deswarte, Fabien E I, and Farmer, Thomas J (2009). The integration of green chemistry into future biorefineries. *Biofuels, Bioproducts and Biorefining*, 3 (1), pp. 72–90. DOI: 10.1002/bbb.119.

Cock, Peter J A et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics (Oxford, England)*, 25 (11), pp. 1422–3. DOI: 10.1093/bioinformatics/btp163.

Conceição, Gelson J.Andrade, Moran, Paulo J S, and Rodrigues, J.Augusto R (2003). Highly efficient extractive biocatalysis in the asymmetric reduction of an acyclic enone by the yeast Pichia stipitis. *Tetrahedron: Asymmetry*, 14 (1), pp. 43–45. DOI: 10.1016/S0957-4166(02)00798-X.

Costenoble, Roeland et al. (2007). 13C-Labeled metabolic flux analysis of a fed-batch culture of elutriated Saccharomyces cerevisiae. *FEMS yeast research*, 7 (4), pp. 511–526. DOI: 10.1111/j.1567-1364.2006.00199.x.

D'Amore, T et al. (1990). A study of ethanol tolerance in yeast. *Critical reviews in biotechnology*, 9 (4), pp. 287–304. DOI: 10.3109/07388558909036740.

Dahn, Kristine M et al. (1996). Increased xylose reductase activity in the xylose-fermenting yeast Pichia stipitis by overexpression of XYL1. *Applied Biochemistry and Biotechnology*, 57-58 (1), pp. 267–276. DOI: 10.1007/BF02941707.

Dale, Bruce E and Kim, Seungdo (2006). "Biomass Refining Global Impact The Biobased Economy of the 21st Century". *Biorefineries - Industrial Processes and Products: Status Quo and Future Directions*. Ed. by B Kamm, P R Gruber, and M Kamm. 1st ed. Vol. 1. Weinheim, Germany: Wiley-VCH. Chap. 2, pp. 41–66. ISBN: 978-3527310272. DOI: 10.1002/9783527619849.ch2.

Degtyarenko, Kirill et al. (2008). ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic acids research*, 36 (Database issue), pp. D344–50. DOI: 10.1093/nar/gkm791.

Degtyarenko, Kirill et al. (2009). "ChEBI: an open bioinformatics and cheminformatics resource." *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* Vol. Chapter 14. Chap. 14, Unit 14.9. DOI: 10.1002/0471250953.bi1409s26.

Desimone, Martín F et al. (2002). Ethanol tolerance in free and sol-gel immobilised Saccharomyces cerevisiae. *Biotechnology Letters*, 24 (19), pp. 1557–1559. DOI: 10.1023/A:1020375321009.

Deublein, Dieter and Steinhauser, Angelika (2008). *Biogas from Waste and Renewable Resources: An Introduction*. 1st ed. Wiley-VCH, p. 472. ISBN: 3527318410.

Dijken, J P van et al. (1986). Alcoholic fermentation by 'non-fermentative' yeasts. *Yeast (Chichester, England)*, 2 (2), pp. 123–127. DOI: 10.1002/yea.320020208.

Ding, Junmei et al. (2009). Tolerance and stress response to ethanol in the yeast Saccharomyces cerevisiae. *Applied microbiology and biotechnology*, 85 (2), pp. 253–263. DOI: 10.1007/s00253-009-2223-1.

Duarte, Natalie C, Herrgå rd, Markus J, and Palsson, Bernhard Ø (2004). Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome research*, 14 (7), pp. 1298–1309. DOI: 10.1101/gr.2250904.

Duarte, Natalie C et al. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (6), pp. 1777–1782. DOI: 10.1073/pnas.0610772104.

EERE (2006). *Biomass Feedstock Composition and Property Database*.

Ebenhöh, O and Heinrich, R (2001). Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. *Bulletin of mathematical biology*, 63 (1), pp. 21–55. DOI: 10.1006/bulm.2000.0197.

Edwards, J S and Palsson, B O (2000). Robustness analysis of the Escherichia coli metabolic network. *Biotechnology progress*, 16 (6), pp. 927–939. DOI: 10.1021/bp0000712.

Edwards, Jeremy S, Covert, Markus, and Palsson, Bernhard (2002). Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology*, 4 (3), pp. 133–140. DOI: 10.1046/j.1462-2920.2002.00282.x.

Edwards, Jeremy S, Ramakrishna, Ramprasad, and Palsson, Bernhard O (2002). Characterizing the Metabolic Phenotype : A Phenotype Phase Plane Analysis. *Biotechnology and bioengineering*, 77 (302), pp. 27–36. DOI: 10.1002/bit.10047.

Eisentraut, Anselm (2010). *Sustainable Production of Second-Generation Biofuels: Potential and perspectives in major economies and developing countries*. Tech. rep. Paris: IEA, p. 221.

Eken-Saraçolu, Nurdan and Arslan, Yeim (2000). Comparison of different pretreatments in ethanol fermentation using corn cob hemicellulosic hydrolysate with Pichia stipitis and Candida shehatae. *Biotechnology Letters*, 22 (10), pp. 855–858. DOI: 10.1023/A:1005663313597.

Eliasson, A. et al. (2000). Anaerobic xylose fermentation by recombinant Saccharomyces cerevisiae carrying XYL1, XYL2, and XKS1 in mineral medium chemostat cultures. *Applied and environmental microbiology*, 66 (8), pp. 3381–6. DOI: 10.1128/AEM.66.8.3381-3386.2000.

Farrell, Alexander E et al. (2006). Ethanol can contribute to energy and environmental goals. *Science (New York, N.Y.)* 311 (5760), pp. 506–508. DOI: 10.1126/science.1121416.

Feist, Adam M and Palsson, Bernhard Ø (2008). The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nature biotechnology*, 26 (6), pp. 659–667. DOI: 10.1038/nbt1401.

Feist, Adam M et al. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology*, 3 (121), p. 121. DOI: 10.1038/msb4100155.

Feng, Xueyang and Zhao, Huimin (2013). Investigating glucose and xylose metabolism in Saccharomyces cerevisiae and Scheffersomyces stipitis via 13 C metabolic flux analysis. *AIChE Journal*, 59 (9), pp. 3195–3202. DOI: 10.1002/aic.14182.

Ferrari, M D et al. (1992). Ethanol production from eucalyptus wood hemicellulose hydrolysate by Pichia stipitis. *Biotechnology and bioengineering*, 40 (7), pp. 753–759. DOI: 10.1002/bit.260400702.

Fiaux, Jocelyne et al. (2003). Metabolic-flux profiling of the yeasts Saccharomyces cerevisiae and Pichia stipitis. *Eukaryotic cell*, 2 (1), pp. 170–180. DOI: 10.1128/EC.2.1.170-180.2003.

Fischer, Eliane, Zamboni, Nicola, and Sauer, Uwe (2004). High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. *Analytical biochemistry*, 325 (2), pp. 308–316. DOI: 10.1016/j.ab.2003.10.036.

Fu, Nan et al. (2009). A novel co-culture process with Zymomonas mobilis and Pichia stipitis for efficient ethanol production on glucose/xylose mixtures. *Enzyme and Microbial Technology*, 45 (3), pp. 210–217. DOI: 10.1016/j.enzmictec.2009.04.006.

Fuganti, Claudio et al. (1993). On the microbial biogeneration of (R) γ-jasmolactone. *Bioorganic & Medicinal Chemistry Letters*, 3 (12), pp. 2777–2780. DOI: 10.1016/S0960-894X(01)80762-9.

Furukawa, Keiji et al. (2004). Effect of cellular inositol content on ethanol tolerance of Saccharomyces cerevisiae in sake brewing. *Journal of bioscience and bioengineering*, 98 (2), pp. 107–113. DOI: 10.1016/S1389-1723(04)70250-9.

Görgens, Johann F et al. (2005). Amino acid supplementation, controlled oxygen limitation and sequential double induction improves heterologous xylanase production by Pichia stipitis. *FEMS yeast research*, 5 (6-7), pp. 677–683. DOI: 10.1016/j.femsyr.2004.12.003.

Griffin, Julian L (2004). Metabolic profiles to define the genome: can we hear the phenotypes? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 359 (1446), pp. 857–871. DOI: 10.1098/rstb.2003.1411.

Grootjen, D R J, Lans, R.G.J.M. van der, and Luyben, K.Ch.A.M. (1990). Effects of the aeration rate on the fermentation of glucose and xylose by Pichia stipitis CBS 5773. *Enzyme and Microbial Technology*, 12 (1), pp. 20–23. DOI: 10.1016/0141-0229(90)90174-0.

Guebel, D V et al. (1991). Influence of oxygen transfer rate and media composition on fermentation of d-xylose by Pichia stipitis NRRL Y-7124. *Journal of Industrial Microbiology*, 7 (4), pp. 287–291. DOI: 10.1007/BF01577657.

Gupthar, Abindra S (1994). Theoretical and practical aspects of ploidy estimation in Pichia stipitis. *Mycological Research*, 98 (7), pp. 716–718. DOI: 10.1016/S0953-7562(09)81042-3.

Hahn-Hägerdal, B et al. (1994a). An interlaboratory comparison of the performance of ethanol-producing micro-organisms in a xylose-rich acid hydroysate. *Applied Microbiology and Biotechnology*, 41 (1), pp. 62–72. DOI: 10.1007/BF00166083.

Hahn-Hägerdal, B et al. (1994b). Biochemistry and physiology of xylose fermentation by yeasts. *Enzyme and Microbial Technology*, 16 (11), pp. 933–943. DOI: 10.1016/0141-0229(94)90002-7.

Hahn-Hägerdal, Bärbel and Pamment, Neville (2004). Microbial pentose metabolism. *Applied biochemistry and biotechnology*, 113-116 (2001), pp. 1207–1209.

Hahn-Hägerdal, Bärbel et al. (2007). Towards industrial pentose-fermenting yeast strains. *Applied microbiology and biotechnology*, 74 (5), pp. 937–953. DOI: 10.1007/s00253-006-0827-2.

Halperin, Mitchell L, Goldstein, Marc B, and Kamel, Kamel S (2010). *Fluid, Electrolyte and Acid-Base Physiology: A Problem-Based Approach*. 4th ed. Saunders, p. 616. ISBN: 978-1416024422.

Harhangi, Harry R et al. (2003). Xylose metabolism in the anaerobic fungus Piromyces sp. strain E2 follows the bacterial pathway. *Archives of microbiology*, 180 (2), pp. 134–141. DOI: 10.1007/s00203-003-0565-0.

Hartwell, L H et al. (1999). From molecular to modular cell biology. *Nature*, 402 (6761 Suppl), pp. C47–52. DOI: 10.1038/35011540.

Haveren, Jacco Van, Scott, Elinor L, and Sanders, Johan (2008). Bulk chemicals from biomass. *Biofuels, Bioproducts and Biorefining*, 2 (1), pp. 41–57. DOI: 10.1002/bbb.43.

Hector, Ronald E et al. (2008). Expression of a heterologous xylose transporter in a Saccharomyces cerevisiae strain engineered to utilize xylose improves aerobic xylose consumption. *Applied microbiology and biotechnology*, 80 (4), pp. 675–684. DOI: 10.1007/s00253-008-1583-2.

Helnrlch, Reinhart et al. (1997). Theoretical Approaches to the Evolutionary Optimization of Glycolysis. Thermodynamic and Kinetic Constraints. *European Journal of Biochemistry*, 243 (1-2), pp. 191–201. DOI: 10.1111/j.1432-1033.1997.0191a.x.

Hjersted, Jared L and Henson, Michael A (2006). Optimization of fed-batch Saccharomyces cerevisiae fermentation using dynamic flux balance models. *Biotechnology Progress*, 22 (5), pp. 1239–1248. DOI: 10.1021/bp060059v.

Hjersted, Jared L, Henson, Michael A, and Mahadevan, Radhakrishnan (2007). Genome-scale analysis of Saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture. *Biotechnology and bioengineering*, 97 (5), pp. 1190–1204. DOI: 10.1002/bit.21332.

Ho, N W, Chen, Zhengdao, and Brainard, Adam P (1998). Genetically engineered Saccharomyces yeast capable of effective cofermentation of glucose and xylose. *Applied and environmental microbiology*, 64 (5), pp. 1852–9.

Hoekman, S Kent (2009). Biofuels in the U.S. Challenges and Opportunities. *Renewable Energy*, 34 (1), pp. 14–22. DOI: 10.1016/j.renene.2008.04.030.

Hofmeyr, J H and Westerhoff, H V (2001). Building the cellular puzzle: control in multi-level reaction networks. *Journal of theoretical biology*, 208 (3), pp. 261–285. DOI: 10.1006/jtbi.2000.2216.

Holzhütter, Hermann-Georg (2004). The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European journal of biochemistry / FEBS*, 271 (14), pp. 2905–2922. DOI: 10.1111/j.1432-1033.2004.04213.x.

Hou, X (2012). Anaerobic xylose fermentation by Spathaspora passalidarum. *Applied microbiology and biotechnology*, 94 (1), pp. 205–214. DOI: 10.1007/s00253-011-3694-4.

Huang, Chiung-Fang et al. (2009). Enhanced ethanol production by fermentation of rice straw hydrolysate without detoxification using a newly adapted strain of Pichia stipitis. *Bioresource technology*, 100 (17), pp. 3914–3920. DOI: 10.1016/j.biortech.2009.02.064.

Ideker, Trey, Galitski, Timothy, and Hood, Leroy (2001). A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, 2, pp. 343–372. DOI: 10.1146/annurev.genom.2.1.343.

Ilmén, Marja et al. (2007). Efficient production of L-lactic acid from xylose by Pichia stipitis. *Applied and environmental microbiology*, 73 (1), pp. 117–123. DOI: 10.1128/AEM.01311-06.

Izard, Jacques and Limberger, Ronald J (2003). Rapid screening method for quantitation of bacterial cell lipids from whole cells. *Journal of microbiological methods*, 55 (2), pp. 411–418. DOI: 10.1016/S0167-7012(03)00193-3.

Jamshidi, Neema and Palsson, Bernhard Ø (2008). Formulating genome-scale kinetic models in the post-genome era. *Molecular systems biology*, 4, p. 171. DOI: 10.1038/msb.2008.8.

Jeffries, T W and Jin, Y S (2000). Ethanol and thermotolerance in the bioconversion of xylose by yeasts. *Advances in applied microbiology*, 47, pp. 221–268. DOI: 10.1016/S0065-2164(00)47006-1.

Jeffries, T W and Jin, Y.-S. (2004). Metabolic engineering for improved fermentation of pentoses by yeasts. *Applied microbiology and biotechnology*, 63 (5), pp. 495–509. DOI: 10.1007/s00253-003-1450-0.

Jeffries, Thomas W (2008). "Engineering the Pichia stipitis Genome for Fermentation of Hemicellulose Hydrolysates". *Bioenergy*. Ed. by Judy D Wall, Caroline S Harwood, and Arnold Demain. Washington, DC: ASM Press. Chap. 3, pp. 37–48. ISBN: 978-1-55581-478-6.

Jeffries, Thomas W and Van Vleet, Jennifer R Headman (2009). Pichia stipitis genomics, transcriptomics, and gene clusters. *FEMS yeast research*, 9 (6), pp. 793–807. DOI: 10.1111/j.1567-1364.2009.00525.x.

Jeffries, Thomas W et al. (2007). Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis. *Nature biotechnology*, 25 (3), pp. 319–326. DOI: 10.1038/nbt1290.

Jeffries, Thomas and Shi, Nian-Qing (1999). Genetic Engineering for Improved Xylose Fermentation by Yeasts. *Adv. Biochem. Eng. Biotechnol.* Advances in 65, pp. 117–161.

Jeong, H et al. (2000). The large-scale organization of metabolic networks. *Nature*, 407 (6804), pp. 651–654. DOI: 10.1038/35036627.

Jeppsson, H, Alexander, N J, and Hahn-Hagerdal, B (1995). Existence of Cyanide-Insensitive Respiration in the Yeast Pichia stipitis and Its Possible Influence on Product Formation during Xylose Utilization. *Applied and environmental microbiology*, 61 (7), pp. 2596–2600.

Jiménez, J and Uden, N van (1985). Use of extracellular acidification for the rapid testing of ethanol tolerance in yeasts. *Biotechnology and bioengineering*, 27 (11), pp. 1596–1598. DOI: 10.1002/bit.260271113.

Jin, Yong-Su and Jeffries, Thomas W (2004). Stoichiometric network constraints on xylose metabolism by recombinant Saccharomyces cerevisiae. *Metabolic engineering*, 6 (3), pp. 229–238. DOI: 10.1016/j.ymben.2003.11.006.

Jin, Yong-Su, Laplaza, Jose M, and Jeffries, Thomas W (2004). Saccharomyces cerevisiae engineered for xylose metabolism exhibits a respiratory response. *Applied and environmental microbiology*, 70 (11), pp. 6816–6825. DOI: 10.1128/AEM.70.11.6816-6825.2004.

Jin, Yong-Su et al. (2002). Molecular cloning of XYL3 (D-xylulokinase) from Pichia stipitis and characterization of its physiological function. *Applied and environmental microbiology*, 68 (3), pp. 1232–1239. DOI: 10.1128/AEM.68.3.1232-1239.2002.

Jirku, V (1999). Whole cell immobilization as a means of enhancing ethanol tolerance. *Journal of Industrial Microbiology and Biotechnology*, 22 (3), pp. 147–151. DOI: 10.1038/sj.jim.2900620.

Jolliffe, I T (2002). *Principal Component Analysis*. 2nd ed. Springer Series in Statistics. Springer, p. 487. ISBN: 9780387954424.

Kacser, Henrik and Beeby, Richard (1984). Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. *Journal of molecular evolution*, 20 (1), pp. 38–51. DOI: 10.1007/BF02101984.

Kamm, B and Kamm, M (2004). Principles of biorefineries. *Applied microbiology and biotechnology*, 64 (2), pp. 137–145. DOI: 10.1007/s00253-003-1537-7.

Kamm, Birgit et al. (2006). "Biorefinery Systems An Overview". *Biorefineries - Industrial Processes and Products: Status Quo and Future Directions*. Ed. by Birgit Kamm, Patrick R Gruber, and Michael Kamm. 1st ed. Vol. 1. Wiley-VCH Verlag GmbH. Chap. 1, pp. 1–40. ISBN: 9783527619849. DOI: 10.1002/9783527619849.ch1.

Kanehisa, M and Goto, S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28 (1), pp. 27–30.

Kanehisa, Minoru et al. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research*, 38 (Database issue), pp. D355–60. DOI: 10.1093/nar/gkp896.

Kanehisa, Minoru et al. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, 40 (Database issue), pp. D109–14. DOI: 10.1093/nar/gkr988.

Karhumaa, Kaisa, Hahn-Hägerdal, Bärbel, and Gorwa-Grauslund, Marie-F. (2005). Investigation of limiting metabolic steps in the utilization of xylose by recombinant Saccharomyces cerevisiae using metabolic engineering. *Yeast (Chichester, England)*, 22 (5), pp. 359–368. DOI: 10.1002/yea.1216.

Karhumaa, Kaisa et al. (2007). High activity of xylose reductase and xylitol dehydrogenase improves xylose fermentation by recombinant Saccharomyces cerevisiae. *Applied microbiology and biotechnology*, 73 (5), pp. 1039–46. DOI: 10.1007/s00253-006-0575-3.

Katahira, Satoshi et al. (2008). Improvement of ethanol productivity during xylose and glucose co-fermentation by xylose-assimilating S. cerevisiae via expression of glucose transporter Sut1. *Enzyme and Microbial Technology*, 43 (2), pp. 115–119. DOI: 10.1016/j.enzmictec.2008.03.001.

Kilian, S G and Uden, N van (1988). Transport of xylose and glucose in the xylose-fermenting yeast Pichia stipitis. *Applied Microbiology and Biotechnology*, 27 (5), pp. 545–548. DOI: 10.1007/BF00451629.

Kim, Min-Soo et al. (2001). High-yield production of xylitol from xylose by a xylitol dehydrogenase defective mutant of Pichia stipitis. *Journal of Microbiology and Biotechnology*, 11 (4), pp. 564–569.

Kitano, Hiroaki (2002). Systems biology: a brief overview. *Science (New York, N.Y.)* 295 (5560), pp. 1662–1664. DOI: 10.1126/science.1069492.

Klinner, U et al. (2005). Aerobic induction of respiro-fermentative growth by decreasing oxygen tensions in the respiratory yeast Pichia stipitis. *Applied microbiology and biotechnology*, 67 (2), pp. 247–253. DOI: 10.1007/s00253-004-1746-8.

Klitgord, Niels and Segrè, Daniel (2010). The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. *Genome informatics. International Conference on Genome Informatics*, 22, pp. 41–55. DOI: 10.1142/9781848165786_0005.

Knorr, Andrea L, Jain, Rishi, and Srivastava, Ranjan (2007). Bayesian-based selection of metabolic objective functions. *Bioinformatics (Oxford, England)*, 23 (3), pp. 351–357. DOI: 10.1093/bioinformatics/btl619.

Koivistoinen, Outi M et al. (2008). Identification in the yeast Pichia stipitis of the first L-rhamnose-1-dehydrogenase gene. *The FEBS journal*, 275 (10), pp. 2482–2488. DOI: 10.1111/j.1742-4658.2008.06392.x.

Koonin, Eugene V and Wolf, Yuri I (2010). Constraints and plasticity in genome and molecular-phenome evolution. *Nature reviews. Genetics*, 11 (7), pp. 487–98. DOI: 10.1038/nrg2810.

Kötter, P et al. (1990). Isolation and characterization of the Pichia stipitis xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing Saccharomyces cerevisiae transformant. *Current genetics*, 18 (6), pp. 493–500. DOI: 10.1007/BF00327019.

Krahulec, Stefan et al. (2012). Comparison of Scheffersomyces stipitis strains CBS 5773 and CBS 6054 with regard to their xylose metabolism: implications for xylose fermentation. *MicrobiologyOpen*, 1 (1), pp. 64–70. DOI: 10.1002/mbo3.5.

Krisch, Judit and Szajáni, Béla (1997). Ethanol and acetic acid tolerance in free and immobilized cells of Saccharomyces cerevisiae and Acetobacter aceti. *Biotechnology Letters*, 19 (6), pp. 525–528. DOI: 10.1023/A:1018329118396.

Kuhad, Ramesh Chander and Singh, Ajay (1993). Lignocellulose Biotechnology: Current and Future Prospects. *Critical Reviews in Biotechnology*, 13 (2), pp. 151–172. DOI: 10.3109/07388559309040630.

Kurtzman, C P (1990). Candida shehatae–genetic diversity and phylogenetic relationships with other xylose-fermenting yeasts. *Antonie van Leeuwenhoek*, 57 (4), pp. 215–222.

Kurtzman, Cletus P and Suzuki, Motofumi (2010). Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera Babjeviella, Meyerozyma, Millerozyma, Priceomyces, and Scheffersomyces. *Mycoscience*, 51 (1), pp. 2–14. DOI: 10.1007/s10267-009-0011-5.

LEE, T C and LEWIS, M J (1968). Identifying Nucleotidic Materials Released by Fermenting Brewer's Yeast. *Journal of Food Science*, 33 (2), pp. 119–123. DOI: 10.1111/j.1365-2621.1968.tb01333.x.

Lange, Jean-paul (2007). Lignocellulose conversion: an introduction to chemistry, process and economics. *Biofuels, Bioproducts and Biorefining*, 1 (1), pp. 39–48. DOI: 10.1002/bbb.7.

Langeveld, Hans, Sanders, Johan, and Meeusen, Marieke, eds. (2010). *The Biobased Economy: Biofuels, Materials and Chemicals in the Post-oil Era*. Routledge, p. 416. ISBN: 9781844077700.

Laplaza, Jose M et al. (2006). Sh ble and Cre adapted for functional genomics and metabolic engineering of Pichia stipitis. *Enzyme and Microbial Technology*, 38 (6), pp. 741–747.

Larhlimi, Abdelhalim et al. (2011). Robustness of metabolic networks: a review of existing definitions. *Bio Systems*, 106 (1), pp. 1–8. DOI: 10.1016/j.biosystems.2011.06.002.

Leandro, Maria José, Spencer-Martins, Isabel, and Gonçalves, Paula (2008). The expression in Saccharomyces cerevisiae of a glucose/xylose symporter from Candida intermedia is affected by the presence of a glucose/xylose facilitator. *Microbiology (Reading, England)*, 154 (Pt 6), pp. 1646–1655. DOI: 10.1099/mic.0.2007/015511-0.

Lee, DoKyoung et al. (2007). *Composition of Herbaceous Biomass Feedstocks*. Tech. rep. June. Sun Grant Initiative, p. 16.

Lee, H et al. (1986). Utilization of Xylan by Yeasts and Its Conversion to Ethanol by Pichia stipitis Strains. *Applied and environmental microbiology*, 52 (2), pp. 320–324.

Li, Peter Yan (2012). "In Silico Metabolic Network Reconstruction of Scheffersomyces Stipitis". Masters of Applied Science. University of Toronto, p. 69.

Liang, Ling, Zhang, Jingqing, and Lin, Zhanglin (2007). Altering coenzyme specificity of Pichia stipitis xylose reductase by the semi-rational approach CASTing. *Microbial cell factories*, 6 (1), p. 36. DOI: 10.1186/1475-2859-6-36.

Lide, David R, ed. (2009). *CRC Handbook of Chemistry and Physics*. 90th ed. CRC Press, p. 2804. ISBN: 978-1420090840.

Ligthelm, Magdalena E, Prior, Bernard A, and Preez, James C du (1988a). The induction of D-xylose catabolizing enzymes in Pachysolen tannophilus and the relationship to anaerobic D-xylose fermentation. *Biotechnology Letters*, 10 (3), pp. 207–212. DOI: 10.1007/BF01134831.

— (1988b). The oxygen requirements of yeasts for the fermentation of d-xylose and d-glucose to ethanol. *Applied Microbiology and Biotechnology*, 28 (1), pp. 63–68. DOI: 10.1007/BF00250500.

Liu, Siqing and Qureshi, Nasib (2009). How microbes tolerate ethanol and butanol. *New biotechnology*, 26 (3-4), pp. 117–121. DOI: 10.1016/j.nbt.2009.06.984.

Liu, Ting et al. (2012). A constraint-based model of Scheffersomyces stipitis for improved ethanol production. *Biotechnology for biofuels*, 5 (1), p. 72. DOI: 10.1186/1754-6834-5-72.

Liu, Z Lewis, Slininger, Patricia J, and Gorsich, Steve W (2005). Enhanced Biotransformation of Furfural and Hydroxymethylfurfural by Newly Developed Ethanologenic Yeast Strains. *Applied Biochemistry and Biotechnology*, 121 (1-3), pp. 451–460. DOI: 10.1385/ABAB:121:1-3:0451.

Llopis, J et al. (1998). Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 95 (12), pp. 6803–6808.

Lu, P et al. (1998). Cloning and disruption of the beta-isopropylmalate dehydrogenase gene (LEU2) of Pichia stipitis with URA3 and recovery of the double auxotroph. *Applied microbiology and biotechnology*, 49 (2), pp. 141–146.

Lu, Ping, Davis, Brian P, and Jeffries, Thomas W (1998). Cloning and Characterization of Two Pyruvate Decarboxylase Genes from Pichia stipitis CBS 6054. *Appl. Environ. Microbiol.* 64 (1), pp. 94–97.

Luong, J H (1985). Kinetics of ethanol inhibition in alcohol fermentation. *Biotechnology and bioengineering*, 27 (3), pp. 280–285. DOI: 10.1002/bit.260270311.

Lynd, Lee R et al. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Current opinion in biotechnology*, 16 (5), pp. 577–583. DOI: 10.1016/j.copbio.2005.08.009.

Ma, Menggen and Liu, Z Lewis (2010). Mechanisms of ethanol tolerance in Saccharomyces cerevisiae. *Applied microbiology and biotechnology*, 87 (3), pp. 829–845. DOI: 10.1007/s00253-010-2594-3.

Machado, Daniel et al. (2012). Exploring the gap between dynamic and constraint-based models of metabolism. *Metabolic engineering*, 14 (2), pp. 112–9. DOI: 10.1016/j.ymben.2012.01.003.

Madshus, I H (1988). Regulation of intracellular pH in eukaryotic cells. *The Biochemical journal*, 250 (1), pp. 1–8.

Mahadevan, R and Palsson, B O (2005). Properties of metabolic networks: structure versus function. *Biophysical journal*, 88 (1), pp. L07–9. DOI: 10.1529/biophysj.104.055723.

Mansfield, Betty Kay et al. (2006). *Breaking the Biological barriers to Cellulosic Ethanol: A Joint Research Agenda*. Tech. rep. Germantown, MD, USA: US DOE.

Manzanares, P, Ramón, D, and Querol, A (1999). Screening of non-Saccharomyces wine yeasts for the production of beta-D-xylosidase activity. *International journal of food microbiology*, 46 (2), pp. 105–112. DOI: 10.1016/S0168-1605(98)00186-X.

Marashi, Sayed-Amir and Bockmayr, Alexander (2011). Flux coupling analysis of metabolic networks is sensitive to missing reactions. *Bio Systems*, 103 (1), pp. 57–66. DOI: 10.1016/j.biosystems.2010.09.011.

Maris, Antonius J A van et al. (2007). Development of efficient xylose fermentation in Saccharomyces cerevisiae: xylose isomerase as a key component. *Advances in biochemical engineering/biotechnology*, 108 (April), pp. 179–204. DOI: 10.1007/10_2007_057.

Marris, Emma (2006). Sugar cane and ethanol: drink the best and drive the rest. *Nature*, 444 (7120), pp. 670–672. DOI: 10.1038/444670a.

Matos, Paula de et al. (2010). Chemical Entities of Biological Interest: an update. *Nucleic acids research*, 38 (Database issue), pp. D249–54. DOI: 10.1093/nar/gkp886.

Matsushika, Akinori and Sawayama, Shigeki (2008). Efficient bioethanol production from xylose by recombinant saccharomyces cerevisiae requires high activity of xylose reductase and moderate xylulokinase activity. *Journal of bioscience and bioengineering*, 106 (3), pp. 306–309. DOI: 10.1263/jbb.106.306.

Matsushika, Akinori et al. (2008). Bioethanol production from xylose by recombinant Saccharomyces cerevisiae expressing xylose reductase, NADP(+)-dependent xylitol dehydrogenase, and xylulokinase. *Journal of bioscience and bioengineering*, 105 (3), pp. 296–299. DOI: 10.1263/jbb.105.296.

Matsushika, Akinori et al. (2009). Efficient bioethanol production by a recombinant flocculent Saccharomyces cerevisiae strain with a genome-integrated NADP+-dependent xylitol dehydrogenase gene. *Applied and environmental microbiology*, 75 (11), pp. 3818–3822. DOI: 10.1128/AEM.02636-08.

McAnulty, Michael J et al. (2012). Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. *BMC systems biology*, 6 (1), p. 42. DOI: 10.1186/1752-0509-6-42.

McMillan, J D (1993). *Xylose Fermentation to Ethanol : A Review*. Tech. rep. January. U.S. Department of Energy, p. 53.

Melake, T, Passoth, V, and Klinner, U (1996). Characterization of the Genetic System of the Xylose-Fermenting Yeast Pichia stipitis. *Current microbiology*, 33 (4), pp. 237–242. DOI: 10.1007/s002849900106.

Meyrial, Valérie et al. (1997). Relationship between effect of ethanol on proton flux across plasma membrane and ethanol tolerance, in Pichia stipitis. *Anaerobe*, 3 (6), pp. 423–429. DOI: 10.1006/anae.1997.0124.

Mikami, K, Haseba, T, and Ohno, Y (1997). Ethanol induces transient arrest of cell division (G2 + M block) followed by G0/G1 block: dose effects of short- and longer-term ethanol exposure on cell cycle and cell functions. *Alcohol and alcoholism (Oxford, Oxfordshire)*, 32 (2), pp. 145–152.

Mo, Monica L, Palsson, Bernhard O, and Herrgå rd, Markus J (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC systems biology*, 3 (1), p. 37. DOI: 10.1186/1752-0509-3-37.

Mohandas, Devaki V, Whelan, Douglas R, and Panchal, Chandra J (1995). Development of xylose-fermenting yeasts for ethanol production at high acetic acid concentrations. *Applied Biochemistry and Biotechnology*, 51-52 (1), pp. 307–318. DOI: 10.1007/BF02933434.

Mosier, Nathan et al. (2005). Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource technology*, 96 (6), pp. 673–686. DOI: 10.1016/j.biortech.2004.06.025.

Naik, S N et al. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 14 (2), pp. 578–597. DOI: 10.1016/j.rser.2009.10.003.

Nardi, James B et al. (2006). Communities of microbes that inhabit the changing hindgut landscape of a subsocial beetle. *Arthropod structure & development*, 35 (1), pp. 57–68. DOI: 10.1016/j.asd.2005.06.003.

Nigam, J N (2001a). Development of xylose-fermenting yeast Pichia stipitis for ethanol production through adaptation on hardwood hemicellulose acid prehydrolysate. *Journal of applied microbiology*, 90 (2), pp. 208–215. DOI: 10.1046/j.1365-2672.2001.01234.x.

— (2001b). Ethanol production from hardwood spent sulfite liquor using an adapted strain of Pichia stipitis. *Journal of Industrial Microbiology and Biotechnology*, 26 (3), pp. 145–150. DOI: 10.1038/sj/jim/7000098.

— (2002). Bioconversion of water-hyacinth (Eichhornia crassipes) hemicellulose acid hydrolysate to motor fuel ethanol by xylosefermenting yeast. *Journal of Biotechnology*, 97 (2), pp. 107–116. DOI: 10.1016/S0168-1656(02)00013-5.

Nissen, T L et al. (1997). Flux Distributions in Anaerobic, Glucose-Limited Continuous Cultures of Saccharomyces Cerevisiae. *Microbiology*, 143 (1), pp. 203–218. DOI: 10.1099/00221287-143-1-203.

Notebaart, Richard A et al. (2008). Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS computational biology*, 4 (1), e26. DOI: 10.1371/journal.pcbi.0040026.

Notebaart, Richard A et al. (2009). Asymmetric relationships between proteins shape genome evolution. *Genome biology*, 10 (2), R19. DOI: 10.1186/gb-2009-10-2-r19.

Oberhardt, Matthew A, Palsson, Bernhard Ø, and Papin, Jason A (2009). Applications of genome-scale metabolic reconstructions. *Molecular systems biology*, 5 (320), p. 320. DOI: 10.1038/msb.2009.77.

Ohgren, Karin et al. (2007). Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. *Bioresource technology*, 98 (13), pp. 2503–2510. DOI: 10.1016/j.biortech.2006.09.003.

Ohta, Kazuyoshi et al. (1991). Genetic improvement of Escherichia coli for ethanol production: chromosomal integration of Zymomonas mobilis genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Applied and environmental microbiology*, 57 (4), pp. 893–900.

Olsson, Lisbeth and Hahn-Hägerdal, Bärbel (1996). Fermentation of lignocellulosic hydrolysates for ethanol production. *Enzyme and Microbial Technology*, 18 (5), pp. 312–331. DOI: 10.1016/0141-0229(95)00157-3.

Orth, Jeffrey D, Fleming, R M T, and Palsson, Bernhard Ø (2010). "Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide". *EcoSalEscherichia coli and Salmonella: Cellular and Molecular Biology*. Ed. by A Böck et al. Washington, DC: ASM Press. Chap. 10.2.1, p. 60.

Orth, Jeffrey D, Thiele, Ines, and Palsson, Bernhard Ø (2010). What is flux balance analysis? *Nature biotechnology*, 28 (3), pp. 245–248. DOI: 10.1038/nbt.1614.

Orth, Jeffrey D et al. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Molecular systems biology*, 7, p. 535. DOI: 10.1038/msb.2011.65.

Osmont, a. et al. (2010). Second generation biofuels: Thermochemistry of glucose and fructose. *Combustion and Flame*, 157 (6), pp. 1230–1234. DOI: 10.1016/j.combustflame.2009.12.002.

Özcan, Sabire, Kötter, Peter, and Ciciary, Michael (1991). Xylan-hydrolysing enzymes of the yeast Pichia stipitis. *Applied Microbiology and Biotechnology*, 36 (2), pp. 190–195. DOI: 10.1007/BF00164418.

Pál, Csaba, Papp, Balázs, and Lercher, Martin J (2005a). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature genetics*, 37 (12), pp. 1372–5. DOI: 10.1038/ng1686.

— (2005b). Horizontal gene transfer depends on gene content of the host. *Bioinformatics (Oxford, England)*, 21 Suppl 2, pp. ii222–3. DOI: 10.1093/bioinformatics/bti1136.

Palsson, Bernhard Ø (2006). *Systems Biology: Properties of Reconstructed Networks*. 1st ed. Cambridge University Press, p. 336. ISBN: 978-0521859035.

Palsson, Bernhard (2000). The challenges of in silico biology. *Nature Biotechnology*, 18 (11), pp. 1147–1150. DOI: 10.1038/81125.

Papin, Jason A, Price, Nathan D, and Palsson, Bernhard Ø (2002). Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome research*, 12 (12), pp. 1889–1900. DOI: 10.1101/gr.327702.

Parekh, SaradR. R, Yu, Shiyuan, and Wayman, Morris (1986). Adaptation of Candida shehatae and Pichia stipitis to wood hydrolysates for increased ethanol production. *Applied Microbiology and Biotechnology*, 25 (3), pp. 300–304. DOI: 10.1007/BF00253667.

Passoth, Volkmar and Hahn-Hägerdal, B (2000). Production of a heterologous endo-1,4-beta-xylanase in the yeast Pichia stipitis with an O(2)-regulated promoter. *Enzyme and microbial technology*, 26 (9-10), pp. 781–784. DOI: 10.1016/S0141-0229(00)00171-X.

Passoth, Volkmar, Zimmermann, Martin, and Klinner, Ulrich (1996). Peculiarities of the regulation of fermentation and respiration in the crabtree-negative, xylose-fermenting yeast Pichia stipitis. *Applied biochemistry and biotechnology*, 57-58 (1), pp. 201–212. DOI: 10.1007/BF02941701.

Passoth, Volkmar et al. (1998). Molecular cloning of alcohol dehydrogenase genes of the yeast ¡I¿Pichia stipitis¡/I¿ and identification of the fermentative ADH. *Yeast*, 14 (14), pp. 1311–1325.

Passoth, Volkmar et al. (2003). Analysis of the hypoxia-induced ADH2 promoter of the respiratory yeast Pichia stipitis reveals a new mechanism for sensing of oxygen limitation in yeast. *Yeast (Chichester, England)*, 20 (1), pp. 39–51. DOI: 10.1002/yea.933.

Perlack, Robert D et al. (2005). *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasability of a Billion-Ton Annual Supply*. Tech. rep. Oak Ridge, TN: DOE and Oak Ridge National Laboratory, p. 78. DOI: 10.2172/885984.

Petschacher, Barbara and Nidetzky, Bernd (2008). Altering the coenzyme preference of xylose reductase to favor utilization of NADH enhances ethanol yield from xylose in a metabolically engineered strain of Saccharomyces cerevisiae. *Microbial cell factories*, 7, p. 9. DOI: 10.1186/1475-2859-7-9.

Porcelli, Anna Maria et al. (2005). pH difference across the outer mitochondrial membrane measured with a green fluorescent protein mutant. *Biochemical and biophysical research communications*, 326 (4), pp. 799–804. DOI: 10.1016/j.bbrc.2004.11.105.

Preez, J C du (1994). Process parameters and environmental factors affecting d-xylose fermentation by yeasts. *Enzyme and Microbial Technology*, 16 (11), pp. 944–956. DOI: 10.1016/0141-0229(94)90003-5.

Preez, J C du, Driessel, B van, and Prior, B A (1989). Ethanol tolerance of Pichia stipitis and Candida shehatae strains in fed-batch cultures at controlled low dissolved oxygen levels. *Applied Microbiology and Biotechnology*, 30 (1), pp. 53–58. DOI: 10.1007/BF00255996.

Preez, J.C. du, Bosch, M., and Prior, B.A. (1987). Temperature profiles of growth and ethanol tolerance of the xylose-fermenting yeasts Candida shehatae and Pichia stipitis. *Applied Microbiology and Biotechnology*, 25 (6), pp. 547–550. DOI: 10.1007/BF00252010.

Price, Nathan D, Schellenberger, Jan, and Palsson, Bernhard O (2004). Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophysical journal*, 87 (4), pp. 2172–2186. DOI: 10.1529/biophysj.104.043000.

Ragauskas, Arthur J et al. (2006). The path forward for biofuels and biomaterials. *Science (New York, N.Y.)* 311 (5760), pp. 484–489. DOI: 10.1126/science.1114736.

Richard, P, Toivari, M H, and Penttilä, M (2000). The role of xylulokinase in Saccharomyces cerevisiae xylulose catabolism. *FEMS microbiology letters*, 190 (1), pp. 39–43. DOI: 10.1111/j.1574-6968.2000.tb09259.x.

Rizzi, M et al. (1987). "Mathematical model for the semiaerobic fermentation of xylose by the yeast Pichia stipitis." *Proceedings of 4th European Congress of Biotechnology*. Ed. by O M Neijssel, R R van der Meer, and K Ch. Luyben. Amsterdam: Elsevier.

Rizzi, Manfred et al. (1988). Xylose fermentation by yeasts 4.Purification and kinetic studies of xylose reductase from Pichia stipitis. *Applied Microbiology and Biotechnology*, 29 (2-3), pp. 148–154. DOI: 10.1007/BF00939299.

Rizzi, Manfred et al. (1989). Xylose fermentation by yeasts. 5. Use of ATP balances for modeling oxygen-limited growth and fermentation of yeast Pichia stipitis with xylose as carbon source. *Biotechnology and bioengineering*, 34 (4), pp. 509–514. DOI: 10.1002/bit.260340411.

Rodrigues, Rita et al. (2008). Fermentation Kinetics for Xylitol Production by a Pichia stipitis D-Xylulokinase Mutant Previously Grown in Spent Sulfite Liquor. *Applied Biochemistry and Biotechnology*, 148 (1), pp. 199–209. DOI: 10.1007/s12010-007-8080-4.

Rogers, P et al. (1982). "Ethanol production by Zymomonas mobilis". *Microbial Reactions, Advances in Biochemical Engineering/Biotechnology*. Ed. by A Fiechter. Vol. 23. Advances in Biochemical Engineering/Biotechnology. Springer Berlin / Heidelberg, pp. 37–84. ISBN: 978-3-540-11698-1. DOI: 10.1007/3540116982_2.

Roos, Albert and Boron, Walter F (1981). Intracellular pH. *Physiological reviews*, 61 (2), pp. 296–434.

Rumbold, Karl et al. (2009). Microbial production host selection for converting second-generation feedstocks into bioproducts. *Microbial cell factories*, 8, p. 64. DOI: 10.1186/1475-2859-8-64.

Rumbold, Karl et al. (2010). Microbial renewable feedstock utilization: a substrate-oriented approach. *Bioengineered bugs*, 1 (5), pp. 359–366. DOI: 10.4161/bbug.1.5.12389.

SLININGER, P J et al. (1990). Stoichiometry and Kinetics of Xylose Fermentation by Pichia stipitis. *Annals of the New York Academy of Sciences*, 589 (1 Biochemical E), pp. 25–40. DOI: 10.1111/j.1749-6632.1990.tb24232.x.

Salgueiro, Sancha P, Sá-Correia, I, and Novais, Julio M (1988). Ethanol-Induced Leakage in Saccharomyces cerevisiae: Kinetics and Relationship to Yeast Ethanol Tolerance and Alcohol Fermentation Productivity. *Applied and environmental microbiology*, 54 (4), pp. 903–909.

Saloheimo, Anu et al. (2007). Xylose transport studies with xylose-utilizing Saccharomyces cerevisiae strains expressing heterologous and homologous permeases. *Applied microbiology and biotechnology*, 74 (5), pp. 1041–1052. DOI: 10.1007/s00253-006-0747-1.

Samal, Areejit et al. (2006). Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC bioinformatics*, 7, p. 118. DOI: 10.1186/1471-2105-7-118.

Sauer, U et al. (1997). Metabolic fluxes in riboflavin-producing Bacillus subtilis. *Nature biotechnology*, 15 (5), pp. 448–52. DOI: 10.1038/nbt0597-448.

Sauer, U et al. (1999). Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. *Journal of bacteriology*, 181 (21), pp. 6679–88.

Schellenberger, Jan et al. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, 6 (9), pp. 1290–1307. DOI: 10.1038/nprot.2011.308.

Schuetz, Robert, Kuepfer, Lars, and Sauer, Uwe (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular systems biology*, 3, p. 119. DOI: 10.1038/msb4100162.

Sedlak, Miroslav and Ho, Nancy W Y (2004). Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant Saccharomyces yeast. *Yeast (Chichester, England)*, 21 (8), pp. 671–684. DOI: 10.1002/yea.1060.

Service, Robert F (2007). Cellulosic ethanol. Biofuel researchers prepare to reap a new harvest. *Science (New York, N.Y.)* 315 (5818), pp. 1488–1491. DOI: 10.1126/science.315.5818.1488.

Seshasayee, Aswin S N et al. (2009). Principles of transcriptional regulation and evolution of the metabolic system in E. coli. *Genome research*, 19 (1), pp. 79–91. DOI: 10.1101/gr.079715.108.

Shi, N Q and Jeffries, T W (1998). Anaerobic growth and improved fermentation of Pichia stipitis bearing a URA1 gene from Saccharomyces cerevisiae. *Applied microbiology and biotechnology*, 50 (3), pp. 339–345. DOI: 10.1007/s002530051301.

Shi, N Q et al. (1999). Disruption of the cytochrome c gene in xylose-utilizing yeast Pichia stipitis leads to higher ethanol production. *Yeast (Chichester, England)*, 15 (11), pp. 1021–1030. DOI: 10.1002/(SICI)1097-0061(199908)15:11<1021::AID-YEA429>3.0.CO;2-V.

Shi, N Q et al. (2000). Characterization and complementation of a Pichia stipitis mutant unable to grow on D-xylose or L-arabinose. *Applied biochemistry and biotechnology*, 84-86, pp. 201–216. DOI: 10.1385/ABAB:84-86:1-9:201.

Shi, Nian-Qing et al. (2002). SHAM-sensitive alternative respiration in the xylose-metabolizing yeast Pichia stipitis. *Yeast (Chichester, England)*, 19 (14), pp. 1203–1220. DOI: 10.1002/yea.915.

Shinar, Guy and Feinberg, Martin (2010). Structural sources of robustness in biochemical reaction networks. *Science (New York, N.Y.)* 327 (5971), pp. 1389–91. DOI: 10.1126/science.1183372.

Skoog, Kerstin and Hahn-Hägerdal, Bärbel (1990). Effect of Oxygenation on Xylose Fermentation by Pichia stipitis. *Applied and environmental microbiology*, 56 (11), pp. 3389–3394.

Skoog, Kerstin, Jeppsson, Helena, and Hahn-Hägerdal, Bärbel (1992). The effect of oxygenation on glucose fermentation with pichia stipitis. *Applied Biochemistry and Biotechnology*, 34-35 (1), pp. 369–375. DOI: 10.1007/BF02920560.

Slininger, Patricia J, Gorsich, Steven W, and Liu, Zonglin L (2009). Culture nutrition and physiology impact the inhibitor tolerance of the yeast Pichia stipitis NRRL Y-7124. *Biotechnology and Bioengineering*, 102 (3), pp. 778–790. DOI: 10.1002/bit.22110.

Slininger, Patricia J et al. (2011). Repression of xylose-specific enzymes by ethanol in Scheffersomyces (Pichia) stipitis and utility of repitching xylose-grown populations to eliminate diauxic lag. *Biotechnology and bioengineering*, 108 (8), pp. 1801–1815. DOI: 10.1002/bit.23119.

Sonderegger, Marco et al. (2004). Molecular basis for anaerobic growth of Saccharomyces cerevisiae on xylose, investigated by global gene expression and metabolic flux analysis. *Applied and environmental microbiology*, 70 (4), pp. 2307–2317. DOI: 10.1128/AEM.70.4.2307-2317.2004.

Song, Hyun-Seob and Ramkrishna, Doraiswami (2012). Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metabolic engineering*, 14 (2), pp. 69–80. DOI: 10.1016/j.ymben.2012.02.003.

Stephanopoulos, Gregory (2007). Challenges in Engineering Microbes for Biofuels Production. *Science (New York, N.Y.)* 315 (5813), pp. 801–804. DOI: 10.1126/science.1139612.

Suh, Sung-Oui et al. (2003). Wood ingestion by passalid beetles in the presence of xylose-fermenting gut yeasts. *Mol. Ecol.* 12 (11), pp. 3137–3145. DOI: 10.1046/j.1365-294X.2003.01973.x.

Szallasi, Zoltan, Stelling, Jörg, and Periwal, Vipul (2006). *System modeling in cellular biology: from concepts to nuts and bolts*. 1st ed. MIT Press, p. 448. ISBN: 0262195488.

Takuma, Shinya et al. (1991). Isolation of xylose reductase gene of Pichia stipitis and its expression in Saccharomyces cerevisiae. *Applied Biochemistry and Biotechnology*, 28-29 (1), pp. 327–340. DOI: 10.1007/BF02922612.

Thiele, Ines and Palsson, Bernhard Ø (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5 (1), pp. 93–121. DOI: 10.1038/nprot.2009.203.

Tomás-Pejó, Elia et al. (2008). Comparison of SHF and SSF processes from steam-exploded wheat straw for ethanol production by xylose-fermenting and robust glucose-fermenting Saccharomyces cerevisiae strains. *Biotechnology and bioengineering*, 100 (6), pp. 1122–1131. DOI: 10.1002/bit.21849.

Toon, S T et al. (1997). Enhanced cofermentation of glucose and xylose by recombinant Saccharomyces yeast strains in batch and continuous operating modes. *Applied biochemistry and biotechnology*, 63-65, pp. 243–255. DOI: 10.1007/BF02920428.

Tran, Linh M, Rizk, Matthew L, and Liao, James C (2008). Ensemble modeling of metabolic networks. *Biophysical journal*, 95 (12), pp. 5606–17. DOI: 10.1529/biophysj.108.135442.

US Congress (2007). *Energy Independence and Security Act of 2007*.

Van Uden, N (1983). Effects of Ethanol on the Temperature Relations of Viability and Growth in Yeast. *Critical Reviews in Biotechnology*, 1 (3), pp. 263–272. DOI: 10.3109/07388558309077982.

Vanrolleghem, Peter A et al. (1996). Validation of a metabolic network for Saccharomyces cerevisiae using mixed substrate studies. *Biotechnology progress*, 12 (4), pp. 434–448. DOI: 10.1021/bp960022i.

Vargas, Felipe a et al. (2011). Expanding a dynamic flux balance model of yeast fermentation to genome-scale. *BMC systems biology*, 5 (1), p. 75. DOI: 10.1186/1752-0509-5-75.

Varma, Amit and Palsson, Bernhard O (1993). Metabolic capabilities of Escherichia coli: I. synthesis of biosynthetic precursors and cofactors. *Journal of theoretical biology*, 165 (4), pp. 477–502. DOI: 10.1006/jtbi.1993.1202.

— (1994). Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology*, 12 (10), pp. 994–998. DOI: 10.1038/nbt1094-994.

Vassilev, Stanislav V et al. (2010). An overview of the chemical composition of biomass. *Fuel*, 89 (5), pp. 913–933. DOI: 10.1016/j.fuel.2009.10.022.

Verduyn, C et al. (1985). Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast Pichia stipitis. *The Biochemical journal*, 226 (3), pp. 669–677.

Visser, J. Arjan G. M. de et al. (2003). PERSPECTIVE: EVOLUTION AND DETECTION OF GENETIC ROBUSTNESS. *Evolution*, 57 (9), pp. 1959–1972. DOI: 10.1111/j.0014-3820.2003.tb00377.x.

Voet, Donald and Voet, Judith J (2010). *Biochemistry*. 4th Editio. Wiley, p. 1520. ISBN: 978-0470570951.

Voet, Donald, Voet, Judith J, and Pratt, Charlotte W (2008). *Fundamentals of Biochemistry: Life at the Molecular Level*. 3rd Editio. Wiley, p. 1240. ISBN: 978-0470129302.

Wang, Liqing, Birol, Inanç, and Hatzimanikatis, Vassily (2004). Metabolic control analysis under uncertainty: framework development and case studies. *Biophysical journal*, 87 (6), pp. 3750–63. DOI: 10.1529/biophysj.104.048090.

Watanabe, Seiya, Kodaki, Tsutomu, and Makino, Keisuke (2005). Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *The Journal of biological chemistry*, 280 (11), pp. 10340–10349. DOI: 10.1074/jbc.M409443200.

Watanabe, Seiya et al. (2007a). Ethanol production from xylose by recombinant Saccharomyces cerevisiae expressing protein engineered NADP+-dependent xylitol dehydrogenase. *Journal of biotechnology*, 130 (3), pp. 316–9. DOI: 10.1016/j.jbiotec.2007.04.019.

Watanabe, Seiya et al. (2007b). The positive effect of the decreased NADPH-preferring activity of xylose reductase from Pichia stipitis on ethanol production using xylose-fermenting

recombinant Saccharomyces cerevisiae. *Bioscience, biotechnology, and biochemistry*, 71 (5), pp. 1365–1369. DOI: 10.1271/bbb.70104.

Watanabe, Takashi et al. (2011). A UV-induced mutant of Pichia stipitis with increased ethanol production from xylose and selection of a spontaneous mutant with increased ethanol tolerance. *Bioresource technology*, 102 (2), pp. 1844–1848. DOI: 10.1016/j.biortech.2010.09.087.

Weierstall, T, Hollenberg, C P, and Boles, E (1999). Cloning and characterization of three genes (SUT1-3) encoding glucose transporters of the yeast Pichia stipitis. *Molecular microbiology*, 31 (3), pp. 871–883. DOI: 10.1046/j.1365-2958.1999.01224.x.

Werpy, T et al. (2004). *Top Value Added Chemicals from Biomass Volume I Results of Screening for Potential Candidates from Sugars and Synthesis Gas*. Tech. rep. US DOE, p. 76. DOI: 10.2172/926125.

Wheeler, David L et al. (2006). Database resources of the National Center for Biotechnology Information. *Nucleic acids research*, 34 (Database issue), pp. D173–80. DOI: 10.1093/nar/gkj158.

Wiedemann, Beate and Boles, Eckhard (2008). Codon-optimized bacterial genes improve L-Arabinose fermentation in recombinant Saccharomyces cerevisiae. *Applied and environmental microbiology*, 74 (7), pp. 2043–2050. DOI: 10.1128/AEM.02395-07.

Wolkenhauer, O (2001). Systems biology: the reincarnation of systems theory applied in biology? *Briefings in bioinformatics*, 2 (3), pp. 258–270. DOI: 10.1093/bib/2.3.258.

Yablochkova, E N et al. (2003). The activity of xylose reductase and xylitol dehydrogenase in yeasts. *Microbiology*, 72 (4), pp. 414–417. DOI: 10.1023/A:1025032404238.

Yablochkova, E N et al. (2004). The Activity of Key Enzymes in Xylose-Assimilating Yeasts at Different Rates of Oxygen Transfer to the Fermentation Medium. *Microbiology*, 73 (2), pp. 129–133. DOI: 10.1023/B:MICI.0000023977.73456.eb.

Yamada, Takuji and Bork, Peer (2009). Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nature reviews. Molecular cell biology*, 10 (11), pp. 791–803. DOI: 10.1038/nrm2787.

Yang, Vina W et al. (1994). High-efficiency transformation of Pichia stipitis based on its URA3 gene and a homologous autonomous replication sequence, ARS2. *Applied and environmental microbiology*, 60 (12), pp. 4245–4254.

Young, Jamey D et al. (2008). Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. *Biotechnology and bioengineering*, 100 (3), pp. 542–559. DOI: 10.1002/bit.21780.

Zhang, M et al. (1995). Metabolic Engineering of a Pentose Metabolism Pathway in Ethanologenic Zymomonas mobilis. *Science (New York, N.Y.)* 267 (5195), pp. 240–243. DOI: 10.1126/science.267.5195.240.

Zhou, Bin, Martin, Gregory J O, and Pamment, Neville B (2008). Increased phenotypic stability and ethanol tolerance of recombinant Escherichia coli KO11 when immobilized in continuous fluidized bed culture. *Biotechnology and bioengineering*, 100 (4), pp. 627–633. DOI: 10.1002/bit.21800.

Appendices

Appendix A
List of reactions in the model

In this appendix, the reactions used in the model have been listed with the information of abbreviation, name/description, formula, lower and upper boundaries (LB/UB), subsystem and confidence scores (CS).

The detailed information is shown in Table A.2 (next page). The confidence score is defined as in Table A.1.

Table A.1: Definition of Confidence Score

Confidence Score	**Evidence type**	**Examples**
4	Biochemical data	Direct evidence for gene production function and biochemical reaction: protein purification, biochemical assays, experimentally solved protein structures and comparative gene-expression studies
3	Genetic data	Direct and indirect evidence for gene function: knockout characterization, knock-in characterization and overexpression
2	Physiological data	Indirect evidence for biochemical reactions based on physiological data: secretion products or defined medium components serve as evidence for transport and metabolic reactions
2	Sequence data	Evidence for gene function: genome annotation and SEED annotation
1	Modeling data	No evidence is available, but reaction is required for modeling. The included function is a hypothesis and needs experimental verfication. The reaction mechanism may be different from the included reaction(s)
0	Not evaluated	

Table A.2: List of reactions in the model

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
Cellmass	Cell mass equation from *S. cerevisiae* core model and *P. stipitis* 2.0	4.431762 glu[c] + 3.164425 accoa[c] + 73.388348 h2o[c] + 81.501005 atp[c] + 2.463950 nad[c] + 7.118377 nadph[c] + 0.087923 nh4[c] + 0.088298 glyc[c] + 0.976698 oaa[c] + 0.545806 pep[c] + 2.303566 g6p[c] + 0.621820 gln[c] + 0.313086 ru5p[c] + 1.472231 pyr[c] + 0.679970 3pg[c] + 0.272903 e4p[c] + 0.059757 so4[c] → 4.000793 akg[c] + 3.164425 coa[c] + 76.106880 h[c] + 2.463950 nadh[c] + 7.118377 nadp[c] + 81.501005 adp[c] + 1.185253 co2[c] + 85.371381 pi[c] + 0.028927 gap[c]	0/Inf	Cellmass Formation	1
ACONT	Aconitase	cit[c] ⇔ icit[c]	-Inf/Inf	Citric Acid Cycle	2
AKGDH	2-oxoglutarate dehydrogenase complex	akg[c] + coa[c] + nad[c] → succoa[c] + co2[c] + nadh[c]	0/Inf	Citric Acid Cycle	3
CISYm	Citrate synthase	accoa[c] + oaa[c] + h2o[c] → coa[c] + cit[c] + h[c]	0/Inf	Citric Acid Cycle	3
FUM	Fumarase	fum[c] + h2o[c] ⇔ mal[c]	-Inf/Inf	Citric Acid Cycle	2
ICITDHxm	Isocitrate dehydrogenase (NAD), mitochondria	icit[c] + nad[c] → akg[c] + co2[c] + nadh[c]	0/Inf	Citric Acid Cycle	3

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
ICITDHym	Isocitrate dehydrogenase (NADP), mitochondria	icit[c] + nadp[c] → akg[c] + co2[c] + nadph[c]	0/Inf	Citric Acid Cycle	3
MDH	Malate dehydrogenase	mal[c] + nad[c] ⇔ oaa[c] + nadh[c] + h[c]	-Inf/Inf	Citric Acid Cycle	2
SDHm	Succinate dehydrogenase (ubiquinone), mitochondrial (Complex II)	succ[c] + q[c] ⇔ fum[c] + qh2[c]	-Inf/Inf	Citric Acid Cycle	2
SUCCOAL	Succinyl-CoA ligase (ADP-forming)	succoa[c] + pi[c] + adp[c] ⇔ succ[c] + coa[c] + atp[c]	-Inf/Inf	Citric Acid Cycle	2
EX_ac	Acetate exchange	ac[e] ⇔	0/Inf	Exchange	2
EX_acald	Acetaldehyde exchange	acald[e] ⇔	0/Inf	Exchange	2
EX_akg	2-Oxoglutarate exchange	akg[e] ⇔	0/Inf	Exchange	2
EX_cit	Citrate exchange	cit[e] ⇔	0/Inf	Exchange	2
EX_co2	CO2 exchange	co2[e] ⇔	-Inf/Inf	Exchange	2
EX_etoh	Ethanol exchange	etoh[e] ⇔	0/Inf	Exchange	2
EX_fum	Fumarate exchange	fum[e] ⇔	0/Inf	Exchange	2
EX_glc	Glucose exchange	glc[e] ⇔	-10/Inf	Exchange	2
EX_gln	L-glutamine exchange	gln[e] ⇔	0/Inf	Exchange	2
EX_glu	L-glutamate exchange	glu[e] ⇔	0/Inf	Exchange	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
EX_glyc	Glycerol exchange	glyc[e] ⇔	0/Inf	Exchange	2
EX_h	H+ exchange	h[e] ⇔	-Inf/Inf	Exchange	2
EX_h2o	H2o exchange	h2o[e] ⇔	-Inf/Inf	Exchange	2
EX_mal	malate exchange	mal[e] ⇔	0/Inf	Exchange	2
EX_nh4	Ammonium exchange	nh4[e] ⇔	0/Inf	Exchange	2
EX_o2	O2 exchange	o2[e] ⇔	-Inf/Inf	Exchange	2
EX_pi	Phosphate exchange	pi[e] ⇔	-Inf/Inf	Exchange	2
EX_pyr	Pyruvate exchange	pyr[e] ⇔	0/Inf	Exchange	2
EX_so4	Sulfurate exchange	so4[e] ⇔	-Inf/Inf	Exchange	1
EX_succ	Succinate exchange	succ[e] ⇔	0/Inf	Exchange	2
EX_ure	Urea exchange	ure[e] ⇔	-Inf/Inf	Exchange	2
EX_xyl	Xylose exchange	xyl[e] ⇔	0/Inf	Exchange	2
EX_xylt	Xylitol exchange	xylt[e] ⇔	0/Inf	Exchange	2
ABTA	4-aminobutyrate transaminase	4abut[c] + akg[c] → sucsal[c] + glu[c]	0/Inf	Glutamate metabolism	2
GDC	Glutamate decarboxylase	glu[c] + h[c] → 4abut[c] + co2[c]	0/Inf	Glutamate metabolism	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	Name/Description	Formula	LB/UB	Subsystem	CS
GDHx	Glutamate dehydrogenase (NAD)	akg[c] + nh4[c] + nadh[c] + h[c] → glu[c] + nad[c] + h2o[c]	0/Inf	Glutamate metabolism	1
GDHy	Glutamate dehydrogenase (NADP)	akg[c] + nh4[c] + nadph[c] + h[c] → glu[c] + nadp[c] + h2o[c]	0/Inf	Glutamate metabolism	4
GLNS	Glutamine synthetase	glu[c] + atp[c] + nh4[c] → gln[c] + adp[c] + h[c] + pi[c]	0/Inf	Glutamate metabolism	2
GOGATx	Glutamate synthase (NAD)	akg[c] + gln[c] + h[c] + nadh[c] → 2 glu[c] + nad[c]	0/Inf	Glutamate metabolism	1
SSADHy	Succinate-semialdehyde dehydrogenase (NADP)	sucsal[c] + nadp[c] + h2o[c] → succ[c] + nadph[c] + 2 h[c]	0/Inf	Glutamate metabolism	2
GLYC3PDHx	Glycerol-3-phosphate dehydrogenase (NAD)	dhap[c] + nadh[c] + h[c] → glyc3p[c] + nad[c]	0/Inf	Glycerolipid metabolism	3
GLYC3PDHzm	Glycerol-3-phosphate dehydrogenase (FAD), mitochondrial	glyc3p[c] + fad[c] → dhap[c] + fadh2[c]	0/Inf	Glycerolipid metabolism	3
GLYK	Glycerol kinase	glyc[c] + atp[c] → glyc3p[c] + adp[c] + h[c]	0/Inf	Glycerolipid metabolism	3
GPP	Glycerol-3-phosphatase	glyc3p[c] + h2o[c] → glyc[c] + pi[c]	0/Inf	Glycerolipid metabolism	3
ADHx	Alcohol dehydrogenase	acald[c] + nadh[c] + h[c] ⇔ etoh[c] + nad[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2
ADHy	Alcohol dehydrogenase	acald[c] + nadph[c] + h[c] ⇔ etoh[c] + nadp[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	1

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
ENO	Enolase	2pg[c] ⇔ pep[c] + h2o[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2
FBPA	Fructose-bisphosphate aldolase	fbp[c] ⇔ dhap[c] + gap[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2
FBPP	Fructose bisphosphate phosphatase	fbp[c] + h2o[c] → f6p[c] + pi[c]	0/Inf	Glycolysis/ Gluconeogenesis	2
GAPDHx	Glyceraldehyde-3-phosphate dehydrogenase	gap[c] + nad[c] + pi[c] ⇔ 13bpg[c] + nadh[c] + h[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2
GLK	Glucokinase	glc[c] + atp[c] → g6p[c] + adp[c] + h[c]	0/Inf	Glycolysis/ Gluconeogenesis	2
PCK	Phosphoenolpyruvate carboxykinase	oaa[c] + atp[c] → pep[c] + co2[c] + adp[c]	0/Inf	Glycolysis/ Gluconeogenesis	2
PDHm	Pyruvate dehydrogenase	pyr[c] + coa[c] + nad[c] → accoa[c] + co2[c] + nadh[c]	0/Inf	Glycolysis/ Gluconeogenesis	2
PFK	6-phosphofructokinase	f6p[c] + atp[c] → fbp[c] + adp[c] + h[c]	0/Inf	Glycolysis/ Gluconeogenesis	2
PGI	Glucose-6-phosphate isomerase	g6p[c] ⇔ f6p[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2
PGK	Phosphoglycerate kinase	13bpg[c] + adp[c] ⇔ 3pg[c] + atp[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2
PGM	Phosphoglycerate mutase	3pg[c] ⇔ 2pg[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
PYK	Pyruvate kinase	pep[c] + adp[c] + h[c] → pyr[c] + atp[c]	0/Inf	Glycolysis/ Gluconeogenesis	2
TPI	Triose-phosphate isomerase	dhap[c] ⇔ gap[c]	-Inf/Inf	Glycolysis/ Gluconeogenesis	2
ICITL	Isocitrate lyase	icit[c] → glx[c] + succ[c]	0/Inf	Glyoxylate and dicarboxylate metabolism	2
MLS	Malate synthase	glx[c] + accoa[c] + h2o[c] → mal[c] + coa[c] + h[c]	0/Inf	Glyoxylate and dicarboxylate metabolism	2
NADK	NAD kinase	atp[c] + nad[c] → adp[c] + nadp[c] + h[c]	0/Inf	Nicotinate and nicotinamide metabolism	2
ALPHNH	Allophanate hydrolase	allphn[c] + 3 h[c] + h2o[c] → 2 co2[c] + 2 nh4[c]	0/Inf	Nitrogen metabolism	2
URC	Urea carboxylase	ure[c] + atp[c] + hco3[c] ⇔ adp[c] + allphn[c] + h[c] + pi[c]	-Inf/Inf	Nitrogen metabolism	2
DPPH	Diphosphate phosphohydrolase	ppi[c] + h2o[c] ⇔ 2 pi[c] + h[c]	-Inf/Inf	Other	2
HCO3E	HCO3 equilibration reaction	co2[c] + h2o[c] ⇔ h[c] + hco3[c]	-Inf/Inf	Other	4
ADK	Adenylate kinase	amp[c] + atp[c] ⇔ 2 adp[c]	-Inf/Inf	Oxidative phosphorylation	2
ATPM	ATP maintenance requirement	atp[c] + h2o[c] → adp[c] + h[c] + pi[c]	3.5/Inf	Oxidative phosphorylation	2
ATPSF	F-type ATPase	adp[c] + 3 h[e] + pi[c] → atp[c] + 2 h[c] + h2o[c]	0/Inf	Oxidative phosphorylation	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	Name/Description	Formula	LB/UB	Subsystem	CS
CPLXI	Complex I (NADH dehydrogenase)	nadh[c] + q[c] + 5 h[c] → qh2[c] + nad[c] + 4 h[e]	0/Inf	Oxidative phosphorylation	2
CPLXIII	Complex III (Cytochrome bc1 complex)	qh2[c] + 2 cytco[c] + 2 h[c] → q[c] + 2 cytcr[c] + 4 h[e]	0/Inf	Oxidative phosphorylation	2
CPLXIV	Complex IV (Cytochrome c oxidase)	4 cytcr[c] + 8 h[c] + o2[c] → 4 cytco[c] + 2 h2o[c] + 4 h[e]	0/Inf	Oxidative phosphorylation	2
STO	Alternative oxidase (SHAM-sensitive terminal oxidase)	2 qh2[c] + o2[c] → 2 q[c] + 2 h2o[c]	0/Inf	Oxidative phosphorylation	2
SUCDHm	Succinate dehydrogenase (ubiquinone), mitochondrial	fadh2[c] + q[c] ⇔ fad[c] + qh2[c]	-Inf/Inf	Oxidative phosphorylation	2
G6PDH	Glucose 6-phosphate dehydrogenase	g6p[c] + nadp[c] → 6pgl[c] + nadph[c] + h[c]	0/Inf	Pentose phosphate pathway	2
GND	Phosphogluconate dehydrogenase	6pgc[c] + nadp[c] → nadph[c] + co2[c] + ru5p[c]	0/Inf	Pentose phosphate pathway	2
PGL	6-phosphogluconolactonase	6pgl[c] + h2o[c] → 6pgc[c] + h[c]	0/Inf	Pentose phosphate pathway	2
RPE	Ribulose 5-phosphate 3-epimerase	ru5p[c] ⇔ xu5p[c]	-Inf/Inf	Pentose phosphate pathway	2
RPI	Ribose-5-phosphate isomerase	ru5p[c] ⇔ r5p[c]	-Inf/Inf	Pentose phosphate pathway	2
TAL	Transaldolase	gap[c] + s7p[c] ⇔ e4p[c] + f6p[c]	-Inf/Inf	Pentose phosphate pathway	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	Name/Description	Formula	LB/UB	Subsystem	CS
TKT1	Transketolase	r5p[c] + xu5p[c] ⇔ gap[c] + s7p[c]	-Inf/Inf	Pentose phosphate pathway	2
TKT2	Transketolase	xu5p[c] + e4p[c] ⇔ f6p[c] + gap[c]	-Inf/Inf	Pentose phosphate pathway	2
ACOAHim	Acetyl-CoA hydrolase, mitochondrial (irrevsible)	accoa[c] + h2o[c] → coa[c] + ac[c] + h[c]	0/Inf	Pyruvate Metabolism	2
ACOAS	Acetyl-CoA synthetase	ac[c] + atp[c] + coa[c] → accoa[c] + amp[c] + ppi[c]	0/Inf	Pyruvate Metabolism	2
ALDx	Aldehyde dehydrogenase (acetaldehyde, NAD)	acald[c] + h2o[c] + nad[c] → ac[c] + nadh[c] + 2 h[c]	0/Inf	Pyruvate Metabolism	2
ALDy	Aldehyde dehydrogenase (acetaldehyde, NADP)	acald[c] + h2o[c] + nadp[c] → ac[c] + nadph[c] + 2 h[c]	0/Inf	Pyruvate Metabolism	2
MAExm	Malic enzyme (NAD), mitochondria	mal[c] + nad[c] → pyr[c] + nadh[c] + co2[c]	0/Inf	Pyruvate Metabolism	2
PDC	Pyruvate decarboxylase	pyr[c] + h[c] → acald[c] + co2[c]	0/Inf	Pyruvate Metabolism	2
PYC	Pyruvate carboxylase	pyr[c] + atp[c] + hco3[c] → oaa[c] + adp[c] + pi[c] + h[c]	0/Inf	Pyruvate Metabolism	2
ACALDt	Acetaldehyde reversible transport	acald[c] ⇔ acald[e]	-Inf/Inf	Transport	2
ACts	Acetate reversible transport via proton symport	ac[c] + h[c] ⇔ ac[e] + h[e]	-Inf/Inf	Transport	2
AKGt	2-oxoglutarate reversible transport via symport	akg[c] + h[c] ⇔ akg[e] + h[e]	-Inf/Inf	Transport	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
ATPS	ATPase, cytosolic	atp[c] + h2o[c] → adp[c] + h[e] + pi[c]	0/Inf	Transport	2
CITts	Citrate reversible transport via symport	cit[c] + h[c] ⇔ cit[e] + h[e]	-Inf/Inf	Transport	2
CO2t	CO2 transport via diffusion	co2[c] ⇔ co2[e]	-Inf/Inf	Transport	2
ETOHt	Ethanol transport via diffusion	etoh[c] ⇔ etoh[e]	-Inf/Inf	Transport	2
FUMts	Fumarate reversible transport via symport	fum[c] + h[c] ⇔ fum[e] + h[e]	-Inf/Inf	Transport	2
GLCt	Glucose transport (uniport)	glc[e] → glc[c]	0/Inf	Transport	2
GLNts	L-glutamine reversible transport via proton symport	gln[c] + h[c] ⇔ gln[e] + h[e]	-Inf/Inf	Transport	2
GLUts	L-glutamate reversible transport via proton symport	glu[c] + h[c] ⇔ glu[e] + h[e]	-Inf/Inf	Transport	2
GLYCtc	Glycerol transport via channel	glyc[c] ⇔ glyc[e]	-Inf/Inf	Transport	2
H2Ot	H2O transport via diffusion	h2o[e] ⇔ h2o[c]	-Inf/Inf	Transport	2
MALts	Malate reversible transport via symport	mal[c] + h[c] ⇔ mal[e] + h[e]	-Inf/Inf	Transport	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
NH4t	Ammonia reservible transport	nh4[e] ⇔ nh4[c]	-Inf/Inf	Transport	2
O2t	Oxygen transport (diffusion)	o2[e] ⇔ o2[c]	-Inf/Inf	Transport	2
PIts	Phosphate reversible transport via symport	h[e] + pi[e] ⇔ h[c] + pi[c]	-Inf/Inf	Transport	2
PYRt	Pyruvate exchange, diffusion	pyr[c] → pyr[e]	0/Inf	Transport	2
PYRts	Pyruvate transport via proton symport	pyr[e] + h[e] → pyr[c] + h[c]	0/Inf	Transport	2
SO4t	Sulfurate Transport	so4[e] → so4[c]	0/Inf	Transport	2
SUCCts	Succinate transport via proton symport	succ[c] + h[c] ⇔ succ[e] + h[e]	-Inf/Inf	Transport	2
URts	Urea transport via proton symport	ure[e] + 2 h[e] ⇔ ure[c] + 2 h[c]	-Inf/Inf	Transport	2
XYLts	Xylose transport in via proton symporter	xyl[e] + h[e] → xyl[c] + h[c]	0/Inf	Transport	2
XYLTt	Xylitol transport via passive diffusion	xylt[c] ⇔ xylt[e]	-Inf/Inf	Transport	2
XDHx	Xylitol dehydrogenase (NAD)	xylt[c] + nad[c] ⇔ xylu[c] + nadh[c] + h[c]	-Inf/Inf	Xylose metabolism	2
XDHy	Xylitol dehydrogenase (NADP)	xylt[c] + nadp[c] ⇔ xylu[c] + nadph[c] + h[c]	-Inf/Inf	Xylose metabolism	2

Continue on next page...

Table A.2 – *Continued from previous page*

Abbreviation	**Name/Description**	**Formula**	**LB/UB**	**Subsystem**	**CS**
XKS	Xylulokinase	xylu[c] + atp[c] → xu5p[c] + adp[c] + h[c]	0/Inf	Xylose metabolism	2
XRx	NADH-dependent D-xylose reductase	xyl[c] + nadh[c] + h[c] → xylt[c] + nad[c]	0/Inf	Xylose metabolism	2
XRy	NADPH-dependent D-xylose reductase	xyl[c] + nadph[c] + h[c] → xylt[c] + nadp[c]	0/Inf	Xylose metabolism	2

Appendix B
List of metabolites in the model

In this appendix, the reactions used in the model have been listed with the information of abbreviation, name/description, formula, and charge. The formula listed in Table B.1 is the charged formula.

Table B.1: List of metabolites in the model

Abbreviation	Full Name	Formula	Charge
13bpg	1,3-Bisphospho-D-glycerate	C3H4O10P2	-4
2pg	2-Phospho-D-glycerate	C3H4O7P	-3
3pg	3-Phospho-D-glycerate	C3H4O7P	-3
4abut	4-Aminobutanoate	C4H9NO2	0
6pgc	6-Phospho-D-gluconate	C6H10O10P	-3
6pgl	D-Glucono-1,5-lactone 6-phosphate	C6H9O9P	-2
ac	Acetate	C2H3O2	-1
acald	Acetaldehyde	C2H4O	0
accoa	Acetyl coenzyme A	C23H34N7O17P3S	-4
adp	ADP	C10H12N5O10P2	-3
akg	2-Oxoglutarate (alpha-Ketoglutaric acid)	C5H4O5	-2
allphn	Allophanate (urea-1-carboxylate)	C2H3N2O3	-1
amp	AMP	C10H12N5O7P	-2
atp	ATP	C10H12N5O13P3	-4
cit	Citrate	C6H5O7	-3

Continued on next page...

Table B.1– *Continued from previous page*

Abbreviation	Full Name	Formula	Charge
co2	Carbon dioxide	CO2	0
coa	Coenzyme A	C21H32N7O16P3S	-4
cytco	Ferricytochrome c	C42H52FeN8O6S2	3
cytcr	Ferrocytochrome c (Reduced cytochrome c)	C42H52FeN8O6S2	2
dhap	Dihydroxyacetone phosphate	C3H5O6P	-2
e4p	D-Erythrose 4-phosphate	C4H7O7P	-2
etoh	Ethanol	C2H6O	0
f6p	Fructose 6-phosphate	C6H11O9P	-2
fad	Flavin adenine dinucleotide	C27H31N9O15P2	-2
fadh2	FADH2	C27H33N9O15P2	-2
fbp	Fructose 1,6-bisphosphate	C6H10O12P2	-4
fum	Fumarate	C4H2O4	-2
g6p	Glucose 6-phosphate	C6H11O9P	-2
gap	D-glyceraldehyde 3-phosphate	C3H5O6P	-2
glc	D-Glucose	C6H12O6	0
gln	L-Glutamine	C5H10N2O3	0
glu	L-Glutamate	C5H8NO4	-1
glx	Glyoxylate	C2HO3	-1
glyc	Glycerol	C3H8O3	0
glyc3p	Glycerol 3-phosphate	C3H7O6P	-2
h	Hydron	H	1
h2o	Water	H2O	0
hco3	Bicarbonate	CHO3	-1
icit	Isocitrate	C6H5O7	-3
mal	Malate	C4H4O5	-2
nad	Nicotinamide adenine dinucleotide	C21H26N7O14P2	-1

Continued on next page...

Table B.1– *Continued from previous page*

Abbreviation	Full Name	Formula	Charge
nadh	Nicotinamide adenine dinucleotide	C21H27N7O14P2	-2
nadp	Nicotinamide adenine dinucleotide phosphate	C21H25N7O17P3	-3
nadph	Reduced nicotinamide adenine dinucleotide phosphate	C21H26N7O17P3	-4
nh4	Ammoniam	H4N	1
o2	Oxygen	O2	0
oaa	Oxaloacetate	C4H2O5	-2
pep	Phosphoenolpyruvate	C3H2O6P	-3
pi	Orthophosphate	HO4P	-2
ppi	Diphosphate	HO7P2	-3
pyr	Pyruvate	C3H3O3	-1
q	Ubiquinone-6 (Coenzyme Q)	C39H58O4	0
qh2	Ubiquinol-6	C39H60O4	0
r5p	Ribose 5-phosphate	C5H9O8P	-2
ru5p	D-Ribulose 5-phosphate	C5H9O8P	-2
s7p	Sedoheptulose 7-phosphate	C7H13O10P	-2
so4	Sulfate	O4S	-2
succ	Succinate	C4H4O4	-2
succoa	Succinyl CoA	C25H35N7O19P3S	-5
sucsal	Succinate semialdehyde (conjugate acid of 4-oxobutanoate)	C4H5O3	-1
ure	Urea	CH4N2O	0
xu5p	D-Xylulose 5-phosphate	C5H9O8P	-2
xyl	Xylose	C5H10O5	0
xylt	Xylitol	C5H12O5	0
xylu	D-xylulose	C5H10O5	0

Appendix C
Modeling of ethanol induced leakage

To model the time response of ethanol induced leakage, we use the following simplified pathway to describe the leakage:

$$A \rightarrow B \tag{C.1}$$

where A and B denote the 260nm-light-absorbing compounds that are within the cell and are in the external environment respectively. Assuming that when external ethanol concentration is zero, leakage can be described by a first order kinetics, i.e.,

$$r_A = -\frac{\mathrm{d}C_A}{\mathrm{d}t} = k_A C_A \tag{C.2}$$

We can derive the first-order dynamics of C_B, the concentrations of 260nm-absorbing compounds in the extracellular environment that are measured experimentally, as a function of time. Note that for the batch experiment,

$$V_c C_A + V_e C_B = N_T \tag{C.3}$$

where N_T is a constant, which accounts for the total amount of 260nm-absobing compounds in the system; V_c is the volume of the cells; V_e is the volume of the external environment. By rearranging the above equation, we have

$$\begin{aligned} C_A &= \frac{N_T}{V_c} - \frac{V_e}{V_c} C_B \\ &= C_0 - \lambda C_B \end{aligned} \tag{C.4}$$

where C_0 denotes the initial concentration of 260nm-absorbing compounds in the cell, and λ is the ratio of V_e to V_c.

By taking derivation of the above equation, we have

$$\frac{\mathrm{d}C_A}{\mathrm{d}t} = -\lambda \frac{\mathrm{d}C_B}{\mathrm{d}t} \tag{C.5}$$

By plugging Eqn. C.4 and Eqn. C.5 into Eqn. C.2, we have

$$\begin{aligned} \lambda \frac{\mathrm{d}C_B}{\mathrm{d}t} &= k_A \left(C_0 - \lambda C_B\right) \\ \frac{\mathrm{d}C_B}{\mathrm{d}t} &= -k_A C_B + \frac{k_A C_0}{\lambda} \end{aligned} \tag{C.6}$$

with the initial condition of $C_B(0) = 0$. By solving the above ODE, we have

$$C_B(t) = \frac{C_0}{\lambda}\left[1 - \exp(-k_A t)\right] \tag{C.7}$$

To model the time delay (d) introduced by experiment operation, we modify the above equation to

$$C_B(t) = \frac{C_0}{\lambda}\left\{1 - \exp\left[-k_A(t+d)\right]\right\} \tag{C.8}$$

which describe the evolution of C_B as a function of time.

Appendix D
Illustrative example for FBA-PCA

In this appendix, an illustrative example shows how the proposed FBA-PCA method works.

D.1 Model setup

A simple network is constructed as shown in Figure D.1. The network consists of 5 metabolites and 9 reactions, which are listed in Table D.1. Among all reactions, 3 are and 6 are internal reactions. The corresponding stoichiometric matrix S is shown in Equation D.1, in which rows correspond to the metabolites while columns represents the reactions. The constraints we consider are: $0 \leq re1, \cdots, re9 \leq \text{Inf}$.

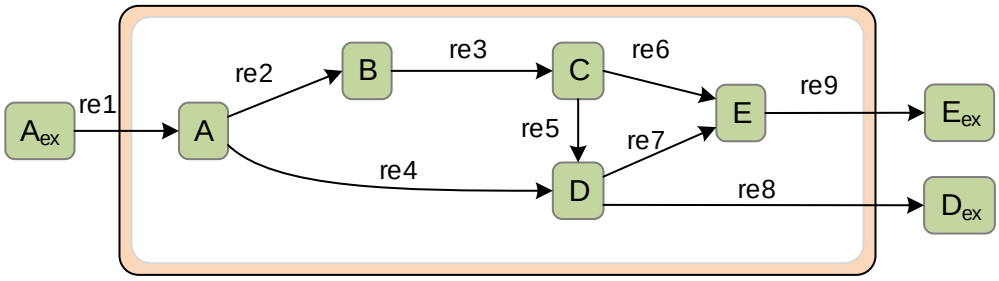

Figure D.1: Reaction network scheme of the illustrative example

Table D.1: Internal and exchange reactions of the illustrative example

Internal reactions		**Exchange reactions**
re2: A $\rightarrow$ B	re5: C $\rightarrow$ D	re1: $A_ex \rightarrow$ A
re3: B $\rightarrow$ 0.5 C	re6: C $\rightarrow$ 2 E	re8: D $\rightarrow D_{ex}$
re4: A $\rightarrow$ 2 D	re7: 0.5 D $\rightarrow$ E	re9: E $\rightarrow E_{ex}$

$$S = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & -0.5 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & -1 \end{bmatrix} \tag{D.1}$$

D.2 Case studies

Two case studies have been used here. The first one is to maximize production of metabolite D as the objective function of FBA, the second one picks maximal production of metabolite E as the object 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.

D.2.1 Case Study I

Objective function: maximal 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/hr with a step size of 0.02. This set of experiments results in a 9 $\times$ 101 data matrix, with each column represents the 9 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 D.2. 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 D.3 visualizes the analysis result by highlighting the fluxes that are affected by increasing flux of re1.

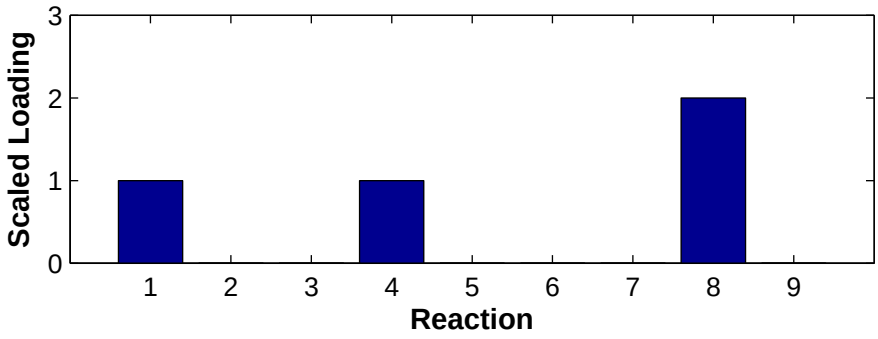

Figure D.2: Scaled PCA loading for case study I

D.2.2 Case Study II

Objective function: maximal 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

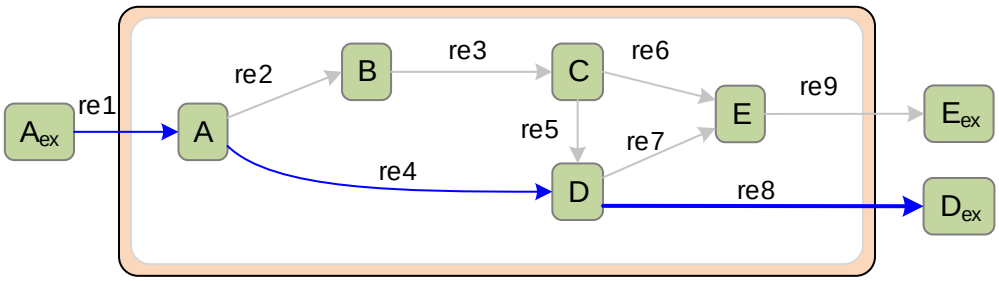

Figure D.3: Visualization of the analysis results for Case I. The reactions that are affected by increasing flux of re1 are highlighted in blue. The line thickness is proportional to its loading.

maximize the production of E. The PC loading and network visualization are plotted in Figure D.4 and Figure D.5.

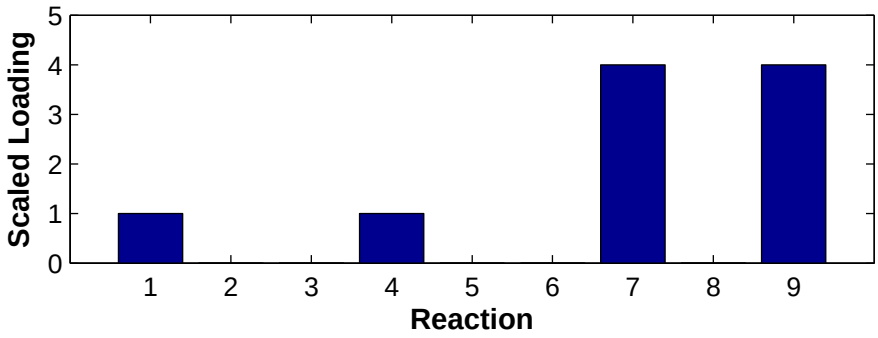

Figure D.4: Scaled PCA loading for case study II

D.3 Discussion

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 $\rightarrow$ 0.5 C $\rightarrow$ 0.5 D while A $\rightarrow$ 2 D 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

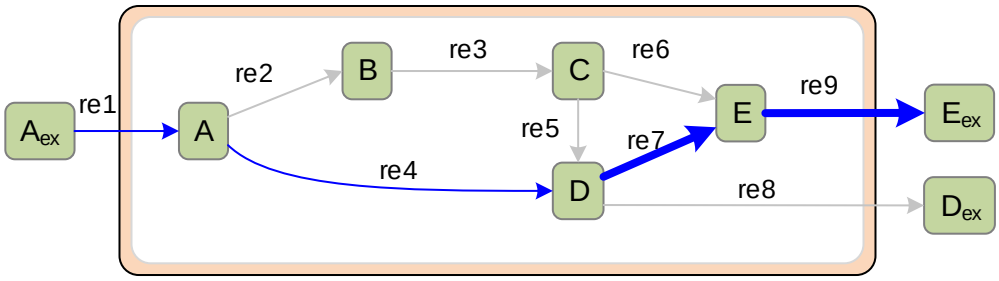

Figure D.5: Visualization of the analysis results for Case II. The reactions that are affected by increasing flux of re1 are highlighted in blue. The line thickness is proportional to its loading.

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.