STRAIN OPTIMIZATION THROUGH THEORETICAL AND EXPERIMENTAL TOOLS

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Pornkamol Unrean

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Friedrich Srienc, Adviser

August 2010

Acknowledgements

I would like to express my appreciation to my supervisor, Prof. Dr. Friedrich Srienc, for his supervision, assistance, valuable guidance throughout my research work at University of Minnesota. I want to express my sincere appreciation to the past and present members of the Srienc Laboratory including Dr. Zhigang Zhang, Dr. Dipen Sangurdekar, Dr. Cong Trinh, Christopher McChalicher, Dr. Greg Sitton, Dr. John Gorke, Dr. Pedro Pena, Janelle Cockrell, Christopher Flynn, Daniel Rouse, Tyler Price and John Barrett for their practical help and helpful suggestions. I want to thank Dr. Erik Holtzapple and Dr. Claudia Schimidt Dannert for supplying the pACMNOx and pACEBI plasmids that were used for this work. Dr. Trinh also constructed *E. coli* mutants, TCS083 and CT1101 used in this dissertation. I would like thank my committee Dr. W. S. Hu, Dr. Y. Kaznessis, Dr. L. Wackett for agreeing to serve on my committee and for reviewing my dissertation. In addition, I must thank funding sources including the National Institutes of Health (NIH), Initiative for Renewable Energy and the Environment (IREE) and Mascoma Corporation for supporting my works. Finally, I am also very grateful to my parents and my sister for their support and understanding.

Dedication

This dissertation is dedicated to my parents

Ratana Unrean and Phiasan Unrean

Abstract

In this dissertation, metabolic network analysis based on elementary mode analysis (EMA), metabolic control analysis (MCA) and thermodynamic analysis of pathways are applied to quantitatively analyze cell metabolism and to engineer a cell for improved performance. By applying EMA, we acquire knowledge of all the pathways of a cell's metabolism. The pathway information permits the systematic implementation of cell manipulation to develop a strain with a desired phenotype. The rational strain improvement is achieved by limiting the cell's functionality to only efficient pathways through gene knockout mutations. This way, the functional space of the designed strain is minimized to a set of pathways that only support the efficient production of the desired product. The EMA approach has been implemented for enhanced synthesis of carotenoid in *E. coli* and ethanol in *T. saccharolyticum*.

Metabolic control analysis and thermodynamic analysis of pathways are employed to examine changes in metabolic pathways within a cell during metabolic evolution. The evolution approach is utilized to select for a mutant of the designed strain that shows a further improvement in product synthesis or strain robustness. The approach is demonstrated in *E. coli* for enhanced carotenoid production, improved ethanol production in the presence of inhibitors, and in *T. saccharolyticum* for increased ethanol productivity. MCA is used as a guiding tool to identify a controlling step in the pathways, while thermodynamic analysis is used to determine changes in the distribution of pathway flux during evolution.

The fermentation process is optimized to enhance production efficiency of the products. A controlled fed-batch fermentation process is designed and conducted to produce a high titer of carotenoid. The process of immobilized mixed cells of two substrate-selective strains of *E. coli* that allow for an efficient conversion of mixed sugars into ethanol at a high yield and a high productivity is also designed and implemented.

Table of Contents

List of Tables

List of Figures

Chapter 1

Thesis overview

1.1 Motivation

The metabolism of a cell is composed of a complex network of biochemical reactions that involve interactions between the number of metabolites and enzymes. An organism utilizes this reaction network to transport, synthesize, or degrade metabolites for growth and energy synthesis. The cellular metabolism has been selected by evolution for growth and survival but has not been optimized for production of high-valued bio-based products. Hence, the organism often does not produce these products at the best possible yield, titer, and productivity. This dissertation describes the implementation of combined strategies for developing an efficient and/or robust strain that can operate at its optimal cellular performance. The strategies involve metabolic engineering based on metabolic pathway analysis and metabolic evolution. These strategies are applied to determine the cellular capability and the genetic modifications needed to enhance cell performance. Fermentation technology for optimization of product synthesis is also presented. The research described in this dissertation entails the study of both native metabolism and recombinant metabolism for the production of primary and secondary metabolites.

1.2 Inverse metabolic engineering

The classical approach for strain improvement has been achieved through a non-directed method of random mutagenesis followed by screening. However, this method involves a large number of mutants that need to be screened and the resulting phenotype is often unpredictable. With this random approach, it is unclear what type of cell

manipulation is responsible for the observed cell response. Therefore, a more systematic, rational approach for strain development is preferred.

Metabolic engineering has evolved and expanded for more than 20 years. Metabolic engineering is defined as a way to enhance cellular capacity using recombinant DNA technology and genetic engineering (Bailey 1991). The scope of the metabolic engineering of a cell includes both manipulation of enzymatic reactions via gene overexpression and gene deletion or recombination of heterologous genes into a cell. Metabolic engineering plays a critical role in strain development. The application of metabolic engineering is varied from an investigation of cellular operation of individual pathways to a complex, interconnected network of a single cell for optimizing product synthesis or for creating a novel compound.

Cell metabolism is comprised of a network of enzyme-catalyzed reactions that are involved in different metabolic pathways and in membrane transport. Functional genes encode these enzymes, creating the direct link between cellular phenotype and genotype. The cell's phenotype can, therefore, be predicted from the manipulation of these enzyme-catalyzed reactions. Inverse metabolic engineering applies this link to facilitate the prediction of cell capacity and the systematic designed strategy for cell manipulation to acquire the selected cell phenotype (Bailey, et al. 1996). Thus, unlike the classical approach, inverse metabolic engineering offers a rational approach to strain design.

This dissertation implements the concept of inverse metabolic engineering for rational strain development. Specifically, the metabolism of a cell is modeled and analyzed to obtain the knowledge of all pathway operations, the distribution of pathways, and the

control of cellular metabolic networks. A number of different techniques are applied to the analysis of cell metabolism and to metabolic engineering for strain improvement including (1) elementary mode analysis (EMA), (2) metabolic control analysis (MCA), and (3) thermodynamic analysis. Details of these approaches are discussed in Chapter 3. Experimental tests are used to validate these methods.

1.3 Experimental organisms

The work outlined in this dissertation focuses on two organisms, *Escherichia coli* and *Thermoanaerobacterium saccharolyticum*.

E. coli is selected as a recombinant host for the production of primary end-products such as ethanol and secondary end-products like carotenoid. The genotype and physiology of *E. coli* are well-characterized, which makes *E. coli* an ideal organism to be used in this research (Neidhardt 1996; Blattner, et al. 1997). The availability of a wide array of host backgrounds, vectors, and regulated strong promoters also makes the sophisticated genetic manipulations of *E. coli* much quicker and easier than manipulation of other organisms.

T. saccharolyticum is used as a native host for the production of ethanol. The organism is useful as a biocatalyst for conversion of biomass into chemicals and fuels because of its ability to ferment a wide variety of biomass-derived sugars including poly-, di-, and mono-saccharides. The organism is also attractive due to its tolerance to low pH and high temperature.

The following sections review topics relevant to the research presented in later chapters.

1.3.1 Basic *E. coli* physiology

E. coli is a gram negative, rod-shaped bacterium. Typically, it is 0.5 μm in diameter and 2.0 μm long. The organism has a genome size of about 4.7 millions bases that are encoded in 4,290 annotated open reading frames (Blattner, et al. 1997). The bacterium is facultatively anaerobic, using both a respiratory and a fermentative type metabolism (Neidhardt 1996). *E. coli* can grow at temperatures ranging from 10 to 45°C and at a pH between 6 and 8. Its optimal culture condition is at 37°C and pH 7 with a doubling time of about 30 minutes in rich medium and about 60 minutes in a defined medium (Neidhardt, et al. 1990). *E. coli* is easy to grow on various types of substrates such as hexose sugars or pentose sugars. A typical composition of an *E. coli* cell is shown in Table 1.1.

Table 1.1 Mass compositions of an *E. coli* cell under aerobic growth at a specific growth rate of 1.0 hr^{-1} at 30°C (adapted from Neidhardt 1996)

Components	% Mass Dry Weight
Protein	55
RNA	20.5
DNA	3.1
Lipid	9.1
Lipopolysaccharides	3.4
Peptidoglycan	2.5
Glycogen	2.5
Polyamines	0.4
Metabolites, Cofactors, and Ions	3.5

1.3.2 *E. coli* metabolism

The metabolism of *E. coli* can be grouped into two classes based on the type of electron acceptors. In Aerobic metabolism, oxygen serves as an external electron acceptor. Anaerobic metabolism occurs in the absence of oxygen where other compounds are used as an external electron acceptor. The metabolic map of *E. coli* used in this dissertation is given in Fig. 1.1. The detailed description of the network and reaction designation can be found in the Appendix.

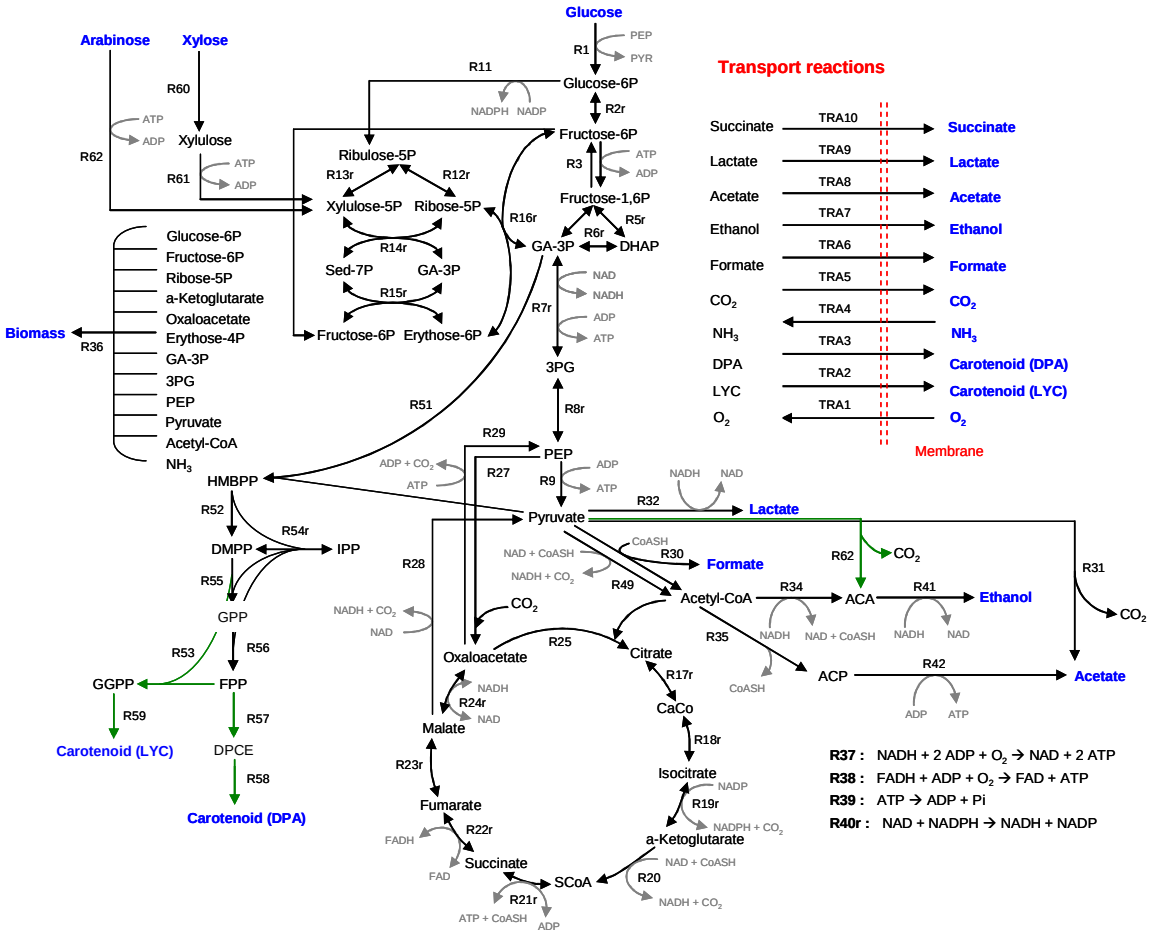


Figure 1.1 Metabolic map of *E. coli*

1.3.2.1 Aerobic metabolism of *E. coli*

E. coli operates under aerobic metabolism when there is oxygen present. Aerobic metabolism consists of four major processes: glycolysis, the pentose phosphate pathway (PPP), the tricarboxylic acid cycle (TCA cycle), and oxidative phosphorylation. Glycolysis is the process in which an organic substrate such as sugar is converted to an intermediated metabolite, pyruvate. PPP is an anabolic pathway that converts a 6 carbons metabolite into 5 carbons metabolite and reducing equivalents. PPP is carried out to generate precursors and cofactors used for biosynthesis within the cell. The pyruvate generated by glycolysis is decarboxylated to acetyl CoA before entering the TCA cycle. The cycle undergoes a series of reactions to strip electrons from acetyl CoA and transfer them to electron acceptor molecules such as NAD^+ (Nicotinamide Adenine Dinucleotide), $NADP^+$ (Nicotinamide Adenine Dinucleotide Phosphate) or FAD^+ (Flavin Adenine Dinucleotide). Through the TCA cycle, the substrate is completely oxidized to a carbon dioxide end-product. These reduced cofactors enter the oxidative phosphorylation process where the electrons are passed down through a series of intermediate electron carriers and finally combined with oxygen to form a water end product. At this step, a proton gradient across the cell membrane is generated that is used for the synthesis of energy in form of ATP (Neidhardt 1996; Lodish, et al. 2000).

1.3.2.2 Anaerobic metabolism of *E. coli*

Anaerobic metabolism is incapable of utilizing oxygen as the external electron acceptor. Therefore, some other small compounds such as nitrate or sulfate are used as external electron acceptors in anaerobic respiration. Another option is to utilize metabolic

intermediates as electron acceptors. *E. coli* metabolism can utilize fermentative products such as lactic acid, acetic acid, ethanol, succinic acid, and formic acid as final electron receptors for regenerating redox components and creating energy required for cell growth. These fermentation products are synthesized through a mixed acid fermentation pathway (Demain, et al. 1999; Lodish, et al. 2000). The fermentative pathway transforms pyruvate precursor to acetyl CoA, producing formate as a byproduct. The acetyl CoA precursor is then converted to fermentation products such as acetic acid or ethanol. During anaerobic growth, the TCA cycle is interrupted and split into two braches. One branch of the cycle is used for the synthesis of precursors required for biosynthesis, while another branch results in the production of succinic acid. The energy produced in anaerobic process via substrate phosphorylation is less efficient compared to the energy yielding process through oxidative phosphorylation in aerobic metabolism (Voet D and Voet J 1995).

1.3.3 Physiological background of *T. saccharolyticum*

Thermoanaerobacterium saccharolyticum is a gram positive, thermophilic, anaerobic bacterium. The organism has a rod shape with approximated size of 1.0 by 3.0 pm (Lee, et al. 1993). The bacterium belongs to the phylum Firmicutes, class Clostridia, order Thermoanaerobacteriales, and family Noctuoidea. *T. saccharolyticum* contains the genome size of about 3 millions bases with an average GC-content of 35% (Liu, et al. 1996). It was first isolated from Frying Pan Springs in Yellowstone National Park. *T. saccharolyticum* grows chemoorganotrophically by utilizing xylan, the predominant form of hemicellulosic polysaccharide in plant biomass and starch but not cellulose (Liu, et al.

1996; Mai and Wiegel 2000). It can also utilize a number of disaccharides and monosaccharides. The organism grows under a wide range of pH from 3.8-6.8 and in high-temperature environments with temperatures ranging from 50 to 65°C. It has the optimum pH for growth between 5.7 and 6.6 and optimal culture temperature between 55 and 60°C.

1.3.4 *T. saccharolyticum* metabolism

T. saccharolyticum ferments a wide variety of carbohydrates including glucose, fructose, mannose, galactose, maltose, cellobiose, sucrose, lactose, trehalose, xylose, starch, rhamnose, raffinose, and xylan. The fermentation products include acetate, ethanol, lactate, CO_2, and H_2. Degradation of hemicellulose by *T. saccharolyticum* is achieved by the concerted action of several enzymes including endoxylanases, exoxylanases, glucuronidase, and acetyl xylan esterases (Shao, et al. 1995; Liu, et al. 1996; Walter-Lorenz and Wiegel 1997). The hydrolysis process converts xylan to xylose oligomers and ultimately to xylobiose and xylose.

The pathway of *T. saccharolyticum* for the production of fermentative end products is similar to the mixed acid fermentation pathway of *E. coli*. However, unlike *E. coli* where pyruvate is decarboxylated to acetyl CoA via pyruvate dehydrogenase, *T. saccharolyticum* uses pyruvate-ferredoxin oxidoreductase to make acetyl CoA, carbon dioxide, and reduced ferredoxin from pyruvate. In addition, the ethanol synthesis pathway of *T. saccharolyticum* can use either NAD or NADP as electron carriers, which make the pathway more flexible comparing to the NAD-linked ethanol synthesis pathway of *E. coli*. The *T. saccharolyticum* metabolic network model used in this dissertation is

shown in Fig. 1.2. The detailed description of the network and reaction designation is listed in the Appendix.

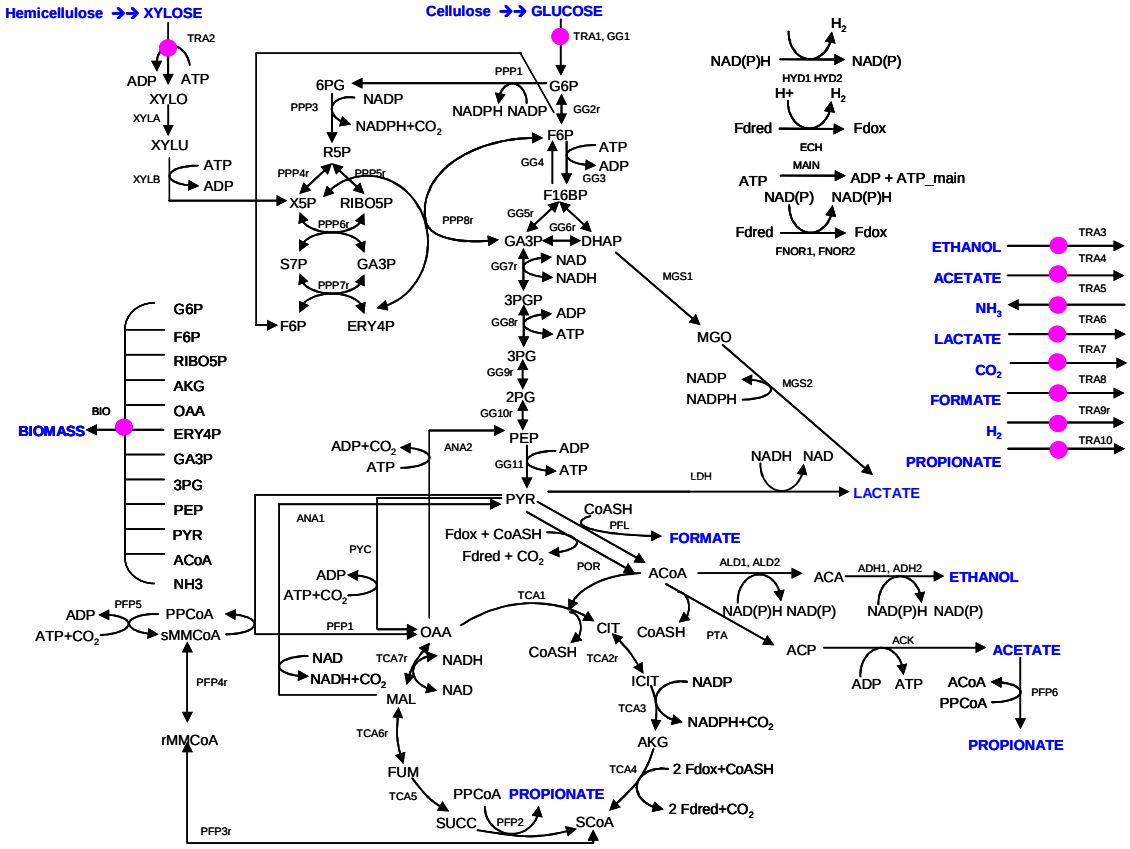


Figure 1.2 Metabolic map of *T. saccharolyticum*

1.3.5 Cofactors and redox balance

Maintaining a favorable redox balance is critical for cell survival. Two major nicotinic cofactors that involve in the redox state of *E. coli* cell are NAD^+/NADH and $NADP^+$/NADPH. The cofactor NAD^+/NADH is produced primarily from the catabolism of energy sources, whereas the cofactor $NADP^+$/NADPH is commonly engaged in anabolic reactions. *T. saccharolyticum* also uses ferredoxin as another source of electron carrier besides NAD^+/NADH and $NADP^+$/NADPH (Shaw, et al. 2008). All these

cofactors serve as electron carriers during enzymatic reactions. They also provide a cell with reducing power that can later be retrieved. These reducing equivalents generated throughout the cell metabolism must be recycled back to their oxidized forms in order for the metabolism to proceed (Clark DP 1989). The balance of redox potential is achieved using an external electron acceptor such as oxygen during aerobic growth or other metabolic intermediates during mixed acid fermentation for recycling the cofactors. The redox ratio also plays a critical role in regulatory mechanisms to control the electron flow from aerobic respiration to anaerobic fermentation (Iuchi, et al. 1996). An increase in the $NADH/NAD^+$ ratio can lead to a shift from the TCA cycle and subsequent respiration to the production of mixed acid fermentation products (de Graef, et al. 1999; Alexeeva, et al. 2000). Therefore, the redox ratio is an important consideration in engineering cell metabolism to favor a certain end-product.

1.3.6 Application of *E. coli* in biotechnology

E. coli is considered to be the first generation organism used in many biotechnological applications and industrial processes for expressing target genes or producing biological products. The fast growth rate of *E. coli* and its capacity for utilizing a variety of substrates allow the use of this organism for commercial production. *E. coli* can produce a variety of products like succinic acid, lactic acid, acetic acid, and ethanol. Using recombinant DNA technology one can engineer *E. coli* to produce several heterologous products that are not native to the organism. The first successful application in the use of engineered *E. coli* was for the production of human insulin (Thomas, et al. 1987). Since then, engineered *E. coli* have been used to synthesize products ranging from

simple biochemicals such as 1,3-propandiol, 3-hydroxybutyric acid, and butanol to products with structural complexity like carotenoid or biodegradable plastic (Tong, et al. 1991; Fidler and Dennis 1992; Lindsay, et al. 1995; Chang, et al. 1999; El-Mansi 2004; Carlson, et al. 2005; Martinez, et al. 2006; Sanchez, et al. 2006; Yoon, et al. 2006; Liu, et al. 2007; Atsumi, et al. 2008; Lee, et al. 2008).

1.3.7 Application of *T. saccharolyticum* in biotechnology

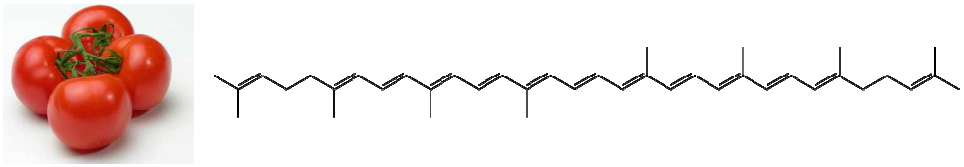
T. saccharolyticum is potentially of interest as a biological catalyst in consolidated bioprocessing (CBP). In consolidated bioprocessing, the production of saccharolytic enzymes, the hydrolysis of a renewable biomass to sugars, and the fermentation of those sugars to commodity products such as ethanol are performed in a single step by the organism. Using this hemicellulose-fermenting thermophile in combination with cellulose-fermenting thermophile such as *Clostridium thermocellum* would provide a substantial time and cost advantage for biomass conversion to biofuels and biochemicals (Lynd, et al. 2002; Lynd, et al. 2005).

1.4 Carotenoid

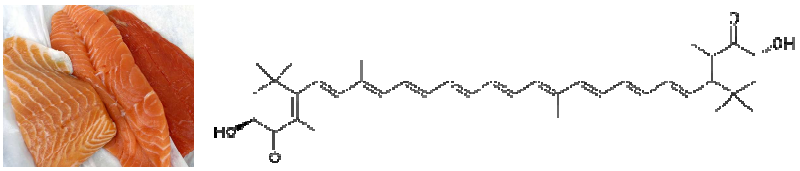
Carotenoid is a natural color pigment consisting of long, unsaturated hydrocarbon chains of 30 to 50 carbons that are synthesized in certain plants and microorganisms (Britton 1995). The product can function as a light-harvesting pigment in the photosynthetic process, a UV-protective compound, a regulator of membrane fluidity, and an antioxidant to protect cells against harmful oxygen radicals. It has been found that carotenoid also plays an important role in reducing the effects of aging, as well as in the prevention of cancer and certain chronic degenerative diseases such as heart disease and

immune deficiency (Ong and Tee 1992; Conn, et al. 1992; Nishino H. 1998; Vershinin A. 1999; Fazekas, et al. 2003). Carotenoid is currently used in the food and pharmaceutical industries as a nutrient supplement, a pharmaceutical, and as a food colorant. To date, more than 600 natural carotenoids have been characterized (Fraser and Sandmann 1999). Some common carotenoid products found in nature are shown in Fig. 1.3.

(1) Lycopene, a red pigment in tomato



(2) Astaxanthin, a pink color in salmon



(3) β-carotene, an orange color in carrot

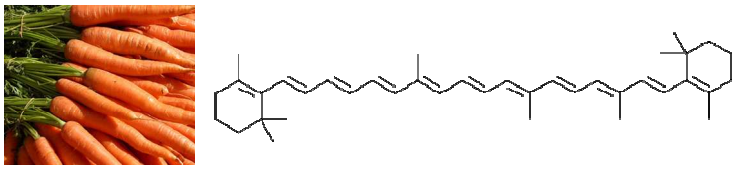


Figure 1.3 Examples of carotenoids

1.5 Carotenoid synthesis pathway

The synthesis of carotenoids can be divided into three major steps. The first step is creating the subunits of carotenoid, isopentenyl diphosphate (IPP), and dimethylallyl diphosphate (DMPP). The second step involves the synthesis of the carotenoid backbone through a serial condensation of the subunits. The last step includes the production of carotenoid end-products by means of modification steps of the backbone.

1.5.1 Biosynthesis of carotenoid subunits

Carotenoid subunits IPP and DMPP can be synthesized using one of two biosynthetic routes, either the non-mevalonate or the mevalonate pathway. The non-mevalonate pathway uses precursors of glycerol-3-phosphate and pyruvate (Rohmer, et al. 1993; Lange, et al. 2000; Lichtenthaler 2000). The pathway begins with a condensation of pyruvate and glycerol-3-phosphate, yielding 1-deoxyxylulose-5-phosphate (DPP) which is later rearranged to form 2-C-methyl-D-erythritol-4-phosphate (MEPP). The MEPP is undertaken by a series of phosphorylations to produce 1-hydroxy-2-methyl-2-butenyl-4-diphosphate (HMBPP). HMBPP is then transformed to IPP and DMPP.

The mevalonate pathway utilizes acetyl CoA as a precursor for the synthesis of IPP and DMPP (Kuzuyama 2002; Lichtenthaler, et al. 1997; Zhong and Yue 2005). The pathway starts with a conversion of acetyl CoA to acetoacetyl CoA which is later condensed to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). HMG-CoA is then reduced and serially phosphorylated to form mevalonate-5-pyrophosphate (MPP). MPP is eventually converted to IPP and DMPP. Biosynthesis of carotenoid subunits through the non-mevalonate and mevalonate pathways is summarized in Fig. 1.4.

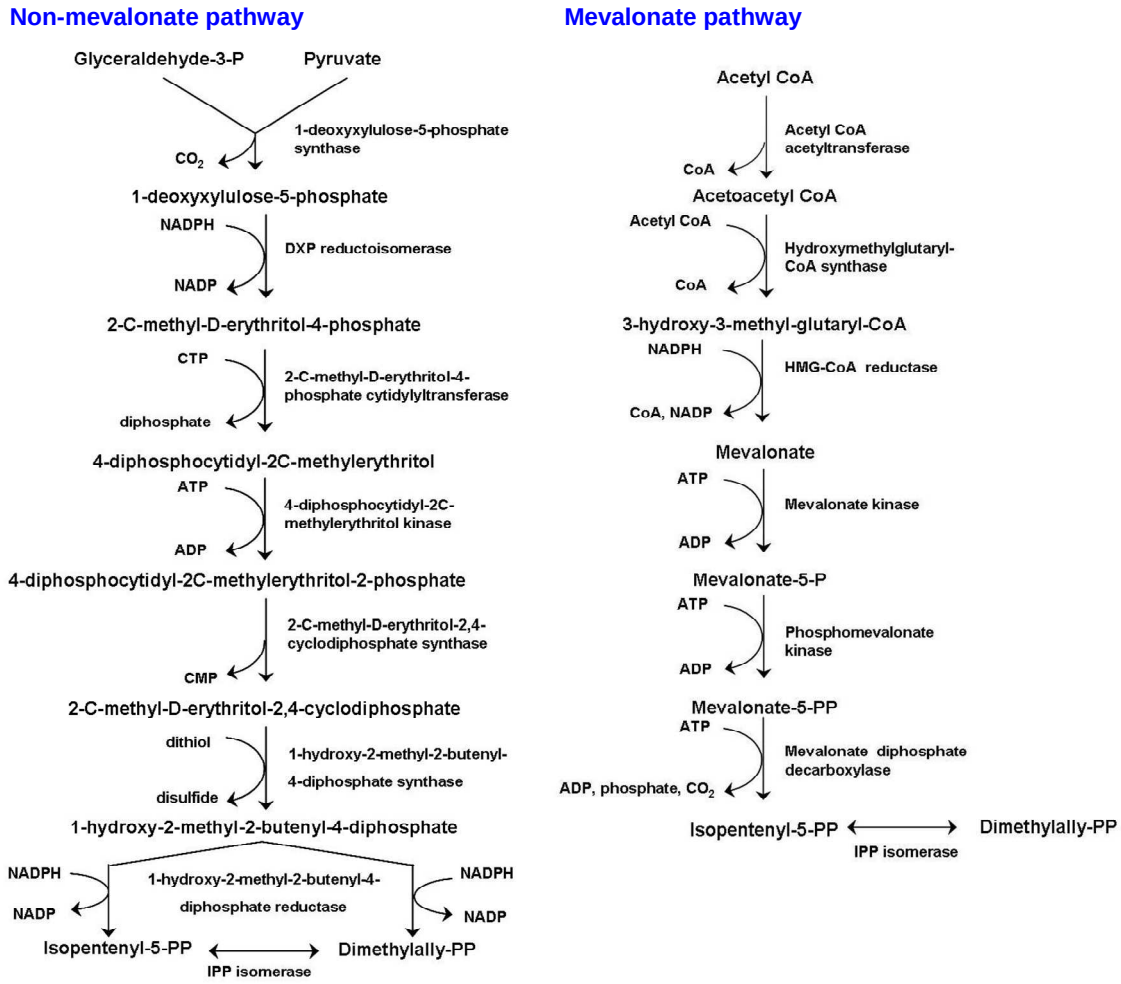


Figure 1.4 Biosynthesis pathways of carotenoid subunits IPP and DMPP

1.5.2 Biosynthesis of the carotenoid backbone

Carotenoid subunits IPP and DMPP from either the mevalonate or the non-mevalonate pathway go through a series of condensations to form the carotenoid backbone. The backbone can range from 30 carbons to 40 carbons long, depending upon the number of iterative steps of condensation it accomplishes (Britton 1998). In general, biosynthesis of 30 carbon carotenoid is a result of three consecutive condensation steps of carotenoid subunits. First, two molecules of IPP and DMPP are combined to form a 10

carbon chain of geranyl diphosphate (GPP). Afterward, GPP is combined with another IPP to form a 15 carbon chain of farnesyl diphosphate (FPP). The two molecules of FPP are then recombined with each other to form a 30 carbon carotenoid backbone, Diapophytoene. Biosynthesis of a 40 carbon carotenoid is similar to the synthesis of the 30 carbon carotenoid except that one more step of condensation is carried out to add another IPP subunit to FPP forming a 20 carbon chain of geranyl geranyl diphosphate (GGPP). The two molecules of GGPP are combined to form a 40 carbon carotenoid backbone, Phytoene. Fig 1.5 shows the biosynthesis of 30 carbon and 40 carbon carotenoid backbones.

In this dissertation, two carotenoid products are under investigation: a 30 carbon carotenoid Diapolycopen-diacid (DPA) and a 40 carbon carotenoid Lycopene (LYC). The biosynthesis pathway of DPA and LYC are introduced into *E. coli* through the plasmid pACMNOx (Mijts, et al. 2005) and the plasmid pACEBI (Lee, et al. 2002) respectively. Details of DPA and LYC biosynthesis pathways were described in Mijts, et al. 2005 and Fraser, et al. 1992 respectively.

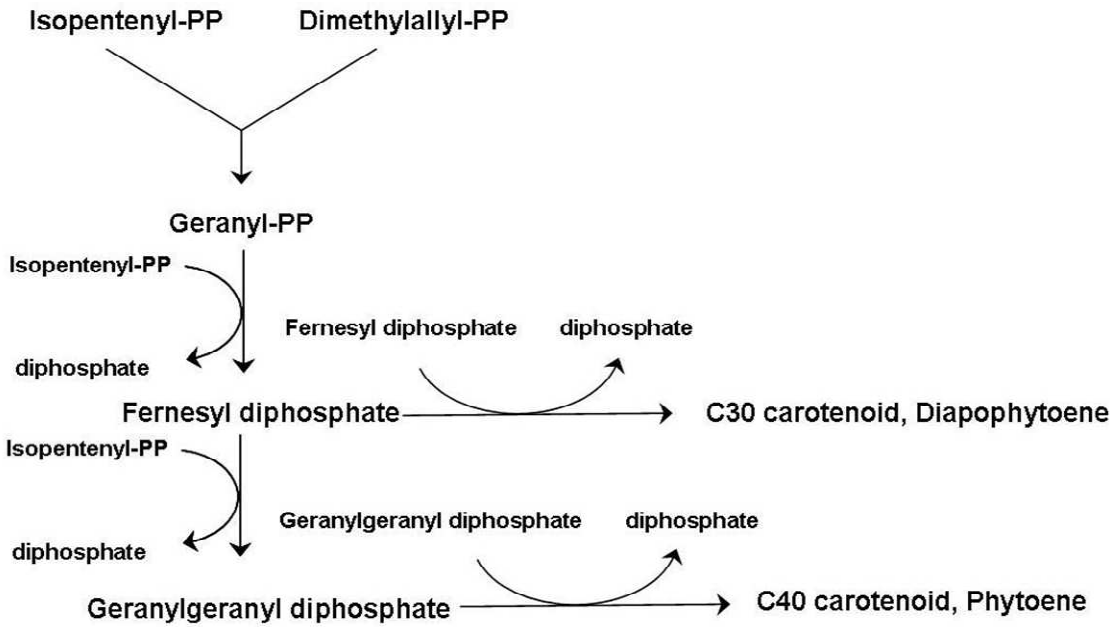


Figure 1.5 Biosynthesis pathway of 30 carbon and 40 carbon carotenoids from carotenoid subunits IPP and DMPP

1.6 Recombinant ethanol producing pathway

Currently, ethanol is the most common biofuel used worldwide. The product has emerged as an important renewable and sustainable energy source; it has the potential to replace gasoline, reducing our reliance on fossil resources. Ethanol can be produced inexpensively from a simple sugar fermentation using microorganisms. A recombinant ethanol biosynthesis pathway that is introduced into *E. coli* through the plasmid pLOI297 (ATCC68239, Alterthum and Ingram 1989) is studied in this dissertation. The pathway is composed of a conversion of pyruvate to acetaldehyde via pyruvate decarboxylase before convertion to ethanol via alcohol dehydrogenase. This pathway is selected for the production of ethanol in *E. coli* because it is more efficient than the native pathway due to its high affinity and its independence of cofactor CoA. The pathway is also commonly

found in other native ethanol producers such as *Zymomonas mobilis* and *Saccharomyces cerevisiae.*

1.7 Thesis outline

One of the challenges in metabolic engineering is to produce a desirable phenotype through the reconstruction of a cell's metabolic pathways. This dissertation discusses approaches that can be used to guide metabolic engineering of a cell for improved performance.

Chapter 1 presents the motivation of the dissertation including key problems and solutions. This chapter also reviews the concept of inverse metabolic engineering applied to the development of strains, the background of experimental organisms, and the biosynthesis pathways of the products of interest. **Chapter 2** describes materials and methods used in the experiment. This chapter explains a technique to construct genetically modified strains using recombinant DNA technology, a technique to improve strain performance through metabolic evolution, and a technique to improve production processes via cell immobilization. This section also includes procedures used for strain characterization. The methods of fermentation in batch, fed-batch, and continuous mode under aerobic and anaerobic conditions are also presented. **Chapter 3** introduces elementary mode analysis (EMA) and provides the basic concept of the computational algorithm to determine metabolic flux. A detailed discussion emphasizes the application of EMA as a useful metabolic pathway analysis tool for characterizing cellular metabolism and for developing an efficient strain. This chapter also includes other forms of metabolic pathway analysis including metabolic control analysis and thermodynamic

analysis. **Chapter 4** discusses metabolic engineering, guided by EMA, to improve the production of carotenoid diapolycopendioic acid (DPA) in recombinant *E. coli*. This chapter demonstrates the application of inverse metabolic engineering through EMA for rational strain design. The model prediction of strain performances is also validated with the achieved experimental results. The production of DPA in an optimized fed-batch fermentation is also included. **Chapter 5** covers the development of metabolically engineered *E. coli* for the efficient production of lycopene using the combined approaches of inverse metabolic engineering and metabolic evolution. **Chapter 6** presents the development of a robust mutant of *E. coli* TCS083 for lignocellulose-based ethanol production through evolution engineering in a chemostat. **Chapter 7** demonstrates the efficient production of ethanol from mixed hexoses and pentoses using immobilized mixed cells of two substrate-specific strains of *E. coli*. **Chapter 8** extends the application of EMA to metabolic engineering of *T. saccharolyticum* for improved ethanol synthesis. Metabolic evolution is further applied to enhance the kinetics of ethanol production. **Chapter 9** presents the application of statistical thermodynamics for describing cell metabolism. Thermodynamic analysis of pathways permits the prediction of change in pathway fluxes during metabolic evolution. **Chapter 10** applies metabolic control analysis for identifying rate-controlling enzymes targeted by metabolic pathway evolution. **Chapter 11** summarizes the dissertation's work for strain development and suggests new avenues for future research.

Chapter 2

Materials and methods

2.1 *E. coli* strains and plasmids

The K12 derived *E. coli* strain, MG1655 (Blattner, et al. 1997) and its mutant derivatives were used as hosts for the synthesis of products of interest. Individual strainss containing genetic knockouts were obtained from the *E. coli* knockout library, Keio collection (Baba, et al. 2006). Table 2.1 summarizes all *E. coli* strains used in the experiments. A summary of plasmids used for creating the recombinant strains is shown in Table 2.2.

Table 2.1 List of *E. coli* strains used in this dissertation

Strain name	Parental strain	Modifications	References
Wild-type	MG1655	None	Bachmann 1996
JW1375	BW25113	ΔldhA::kan+	Baba et al 2006
JW4115	BW25113	ΔfrdA::kan+	Baba et al 2006
JW0855	BW25113	ΔpoxB::kan+	Baba et al 2006
JW2294	BW2229	Δpta::kan+	Baba et al 2006
JW1228	BW25113	ΔadhE::kan+	Baba et al 2006
JW1666	BW25113	ΔpykF::kan+	Baba et al 2006
JW1841	BW25113	Δzwf::kan+	Baba et al 2006
JW2447	BW25113	ΔmaeB::kan+	Baba et al 2006
JW3366	BW25113	ΔpckA::kan+	Baba et al 2006
JW2449	BW25113	ΔtktB::kan+	Baba et al 2006
TCS083	MG1655	ΔzwfΔndhΔsfcAΔmaeBΔldhAΔfrdAΔpoxB Δpta::kan-	Trinh et al 2008

CT1101	MG1655	ΔzwfΔndhΔsfcAΔmaeBΔldhAΔfrdAΔpoxBΔpta ΔglkΔptsGΔmanX::kan-	Trinh et al 2008
CRT028	MG1655	ΔldhAΔfrdAΔpoxBΔptaΔadhEΔpykFΔzwf ΔmaeB::kan-	This study
LYC018	MG1655	ΔpoxBΔpckAΔfrdAΔldhAΔtktBΔadhEΔpykF Δpta::kan-	This study
LYC018E1	MG1655	LYC018, metabolic evolution in NBS medium with 1% glucose	This study
AFF01	MG1655	TCS083, metabolic evolution in LB medium with 1.5% acetic acid and 0.2% furfural	This study

Table 2.2 List of plasmids expressed in *E. coli*

Plasmids	Characteristics	References
pftA	Ampicilin^R, containing flippase recombinase, Temperature sensitive	Posfai et al 1997
pACMNOx	Chloramphenicol^R, containing crtM, crtN, crtOx from *Staphylococcus aureus*	Mijts et al 2005
pACEBI	Chloramphenicol^R, containing crtE, crtB, crtI from *Erwinia euridova*	Lee and Schimidt-Dannert 2002
pLOI297	Ampicilin^R and tetracycline^R containing pdc and adhB from *Zymomonas mobilis*	ATCC 68239

2.2. *T. saccharolyticum* strains and plasmids

T. saccharolyticum strain, JW/SL-YS485 (Lee, et al. 1993) and its mutant derivatives were used for ethanol production from glucose and xylose. Strains containing chromosomal gene knockouts were constructed by homologous recombination with non-replicative knockout plasmid. List of *T. saccharolyticum* strains and the plasmids used in this dissertation are given in Table 2.3 and 2.4 respectively.

Table 2.3 List of *T. saccharolyticum* strains used in this dissertation

Strain name	Modifications	References
Wild-type	None	Lee et al 1993
M355	ΔldhAΔpta-ack::kan-	Shaw et al 2008
AS001	ΔldhA::kan+	This study
AS002	Δzwf::kan+	This study
AS003	Δsbm::kan+	This study
AS004	Δmgs::kan+	This study
AS031	M355 with additional deletion of Δmgs::kan-	This study
AS411	M355 with additional deletion of Δmgs and Δzwf::kan-	This study
AS421	M355 with additional deletion of Δmgs and Δsbm::kan+	This study
AS511	M355 with additional deletion of Δmgs, Δzwf and Δsbm::kan+	This study
AS411E0	AS411, metabolic evolution in MTC medium with 2% glucose for 75 generations	This study
AS411E1	AS411, metabolic evolution in MTC medium with 2% glucose for 159 generations	This study
AS411E2	AS411, metabolic evolution in MTC medium with 2% glucose for 251 generations	This study
AS411E3	AS411, metabolic evolution in MTC medium with 2% glucose for 280 generations	This study
AS511E1	AS411E1 with additional deletion of Δsbm::kan+	This study

Table 2.4 List of plasmids used for construction of genetic knockouts in *T. saccharolyticum*

Plasmids	Characteristics	References
pMU433	*S. cerevisiae-bacteria* shuttle vector containing ura3, kanamycin^R from *Streptococcus faecalis*	Shaw et al 2008
pMU-zwf	pMU433 with zwf up- and down- flanking regions	This study
pMU-mgs	pMU433 with mgs up- and down- flanking regions	This study
pMU-sbm	pMU433 with sbm up- and down- flanking regions	This study

2.3 Storage of strains and plasmids

Cultures of *E. coli* were stored at -80^{o}C in 15% glycerol in cryogenic tubes (Ausubel, et al. 1995). *T. saccharolyticum* cell cultures were stored at -80^{o}C in 5% DMSO in a wheaton vial. The plasmids were either isolated and stored at -20^{o}C or stored as transgenic cell stocks at -80^{o}C.

2.4 Growth media and buffers

2.4.1 Complex medium for *E. coli*

Complex media were used in preparing cells for plasmid transformation, strain construction, or cell immobilization. The three media used were Luria-Bertani (LB), SOC, or Z-broth. LB contained 5 g/l NaCl, 5 g/l yeast extract, and 10 g/l tryptone. SOC contained 5 g/l yeast extract, 20 g/l tryptone, 0.59 g/l NaCl, 0.19 g/l KCl, 0.95 g/l $MgCl_2$, 2.4 g/l $MgSO_4$, and 3.6 g/l glucose. Z-broth was LB medium supplemented with 5 mM $CaCl_2$.

The media used during the immobilization experiment were either an LB or a semi-defined medium composed of 5 g/l yeast extract, 0.5 g/l NaCl, 1 mM $CaCl_2$, 1 mM $MgSO_4$, 1 mg/l Thiamine-HCl, 1 mM Betaine-HCl, and 1 g/l NH_4Cl. All media were autoclaved or filter-sterilized before use. Antibiotics added to the culture for plasmid maintenance were 100 μg/ml for kanamycin, 100 μg/ml for ampicillin, 10 μg/ml for tetracycline, and 50 μg/ml for chloramphenicol. The carbon source was prepared separately and added to the medium if needed. For solid medium, 20 g/l bacto agar was added.

2.4.2 Minimal media for *E. coli*

Two types of minimal media used were M9 and NBS. The M9 medium contained 2 g/l NH_4Cl, 1 g/l NaCl, 25.6 g/l $Na_2HPO_4 \cdot 7H_2O$, 6.0 g/l KH_2PO_4, 0.01 g/l $CaCl_2$, and 0.24 g/l $MgSO_4$. The NBS medium (Martinez, et al. 2007) contained 3.5 g/l $(NH_4)_2HPO4$, 5.0 g/l K_2HPO_4, 3.5 g/l KH_2PO_4, 0.1 mM $CaCl_2$, 1mM $MgSO_4$, and 1 mM Betain-HCl. The media were supplemented with trace metals and minerals, 5.5 mg/l $CaCl_2 \cdot 2H_2O$, 1.0 mg/l $MnCl_2 \cdot 4H_2O$, 1.7 mg/l $ZnCl_2$, 0.43 mg/l $CuCl_2$, 0.60 mg/l $CoCl_2 \cdot 6H_2O$, 0.60 mg/l $NaMoO_4 \cdot 2H_2O$, 8.42 mg/l $FeSO_4 \cdot 7H_2O$, 10 mg/l $Fe(NH_4)$-citrate, 0.38 mg/l $Na_2B_4O_7 \cdot 10H_2O$, and 1 mg/l of thiamine. All media were supplemented with a carbon source and an antibiotic as indicated, adjusted to pH as specified, and filter-sterilized prior to use.

2.4.3 Media for *T. saccharolyticum*

Three culture media used for *T. saccharolyticum* were M122C, M122-defined, and MTC (Hogsett 1995; Zhang and Lynd 2003, Zhang and Lynd 2005). M122C and M122-defined media were used for routine culture maintenance and genetic manipulation, while MTC was used for all fermentation studies. M122C contained 1.3 g/l $(NH_4)_2SO_4$, 1.43 g/l KH_2PO_4, 1.8 g/l $K_2HPO_4 \cdot 3H_2O$, 2.6 g/l $MgCl_2 \cdot 6H_2O$, 0.13 g/l $CaCl_2 \cdot 2H_2O$, 6 g/l glycerol-2-phosphate disodium, 4.5 g/l yeast extract, 1.1 mg/l $FeSO_4 \cdot 7H_2O$, 0.5 g/l cystiene-HCl, 1 mg/l resasurin, and a 5 g/l carbon source. The M122-defined medium was the same as the M122C except the yeast extract was replaced by RPMI vitamin solution (Sigma# R7131) and the MEM amino acid solution (Sigma# M5550). MTC contained 10 g/l yeast extract, 5 g/l tryptone, 2 g/l $C_6H_5O_7K_3 \cdot H_2O$, 1.25 g/l $C_6H_8O_7 \cdot H_2O$, 1 g/l Na_2SO_4, 1 g/l KH_2PO_4,

2.5 g/l $NaHCO_3$, 5 g/l CH_4N_2O, 1 g/l $MgCl_2 \cdot 6H_2O$, 0.2 g/l $CaCl_2 \cdot 2H_2O$, 0.1 g/l $FeCl_2 \cdot 4H_2O$, 1 g/l cystiene-HCl, 1 mg/l resasurin, and a 20g/l carbon source. The pH was adjusted to 6.0 using sulfuric acid. All media were flushed with nitrogen gas for 15 minutes before autoclaving at 121°C for 1 hour. Cysteine-HCl and the carbon source were filter-sterilized separately and added to the media prior to use. Resasurin is used as a color indicator for determining anaerobic condition. The compound turns pink (the oxidized state) in the presence of oxygen, while it becomes colorless (the reduced state) in the absence of oxygen. For a solid medium, 12 g/l bacto agar was added. Selective agents including 200 μg/ml kanamycin or 0.2 mM chloroacetate were added to the culture as appropriate.

2.4.4 Buffers

The citrate buffer used during the knockout construction of the *E. coli* contained 23.5 mM citric acid and 76.5 mM sodium citrate. The TE buffer (pH 7.4) containing 10 mM Tris.HCl and 1 mM EDTA was used for DNA storage. The TBE buffer (pH 8.0) used in agarose electrophoresis contained 10.8 g/l Tris base, 27.5 g/l boric acid, and 2 mM EDTA.

2.5 Culture conditions

2.5.1 Growth in shake flasks

Shake-flask aerobic cultures were performed in 250 ml baffled shake flasks containing 50 ml of the culture medium. For micro-aerobic cultures, cells were grown in capped shake flasks containing 100 ml of the culture medium. All experiments were carried out in a Labline incubator shaker (Model 3528, Melrose Park, IC, USA) at 30°C

or 37°C at a shaking rate of 200 rpm. Culture samples were periodically collected for determining the concentration of cells, the residual substrate, and the product.

2.5.2. Growth in anaerobic shake tubes

Glass tubes (18x150 mm) (Bellco Glass, Inc., Vineland, N. J.) were used for anaerobic shake tube experiments of *T. saccharolyticum*. The tubes contained 20 ml of the culture medium and were sealed with a blue butyl rubber stopper and an aluminum crimp seal. Cultures were inoculated into the tubes through the stopper using disposable syringes and sterile 23 gauge needles. Cell inoculum was 5% vol/vol. The shake tube experiments were carried out in a Labline Orbit Shaker Bath (Melrose Park, IC, USA) at 55°C without shaking.

2.5.3 Growth in bioreactors

2.5.3.1 Batch and repeated batch bioreactors used for *E. coli* culture

The batch bioreactor experiments were conducted in 2l or 10l Braun bioreactors (Biostat MD, B. Braun Biotech International, Melsungen, Germany) with a 50% working volume for the suspension cell cultures. For the immobilized-cell, fluidized-bed reactors, the culture was initiated by inoculating with 50% (v/v) gel beads containing immobilized cells into the bioreactor with a total working volume of approximately 0.7l.

Under aerobic conditions, the reactor was operated at a 1 vvm aeration rate. Under anaerobic conditions, nitrogen was sparged into the bioreactor through a 0.2 μm filter at a volumetric flow rate of 100 ml/min for 6 hours prior to inoculation or continuously flushed into the reactor at a flow rate of 20 ml/min. Agitation speed was constant throughout the fermentation. Temperature was controlled at either 30^{o}C or 37^{o}C. The pH

was controlled using 28% (v/v) NH_4OH or 3M NaOH and 40% (v/v) H_3PO_4. The sterilized antifoam solution was added manually as required. Repeated batch fermentation was implemented similar to batch fermentation. After each batch run, the used medium was drained. The reactor was washed with 0.1% $CaCl_2$ solution before re-suspending in the fresh medium to start a new batch run. Samples (3ml) were periodically removed for further analysis.

2.5.3.2 Batch bioreactor used for *T. saccharolyticum* culture

Batch fermentation was operated in a 2l Braun bioreactor (Biostat MD, B. Braun Biotech International, Melsungen, Germany) with 1l working volume of MTC medium. The temperature was kept at 55°C and the pH was controlled at 6.0 using N_2-flushed 6M NaOH. The agitation speed was set at 100 rpm. The fermentor was made anaerobic by sparging with nitrogen for 6 hours or until the culture medium became colorless prior to cell inoculation. The colorless of culture medium indicates the absence of oxygen in the medium. The cell inoculum was an overnight-grown culture in the same medium. A 5% vol/vol inoculation was used. Culture samples (3ml) were periodically collected for further analysis.

2.5.3.3 Fed-batch bioreactors

Fed-batch fermentation was conducted in a 5l Braun bioreactor (Biostat MD, B. Braun Biotech International, Melsungen, Germany). The fermentation was started as a batch culture with an initial volume of 1.5l. After consumption of the initial substrate, feedstock solution was fed into the reactor at a feeding rate as estimated in Eq(2.3). The feedstock composition used in fed-batch culture for DPA production in *E. coli* was 10 g/l

yeast extract, 50 g/l glucose, 25.6 g/l $Na_2HPO_4.7H_2O$, 6 g/l KH_2PO_4, 1 g/l NaCl, 2 g/l NH_4Cl, 0.24 g/l $MgSO_4$, 50 mg/l chloramphenicol, 0.5 mg/l $MnCl_2.4H_2O$, 0.85 mg/l $ZnCl_2$, 0.22 mg/l $CuCl_2$, 0.3 mg/l $CoCl_2.6H_2O$, 0.3 mg/l $NaMoO_4.2H_2O$, 4.21 mg/l $FeSO_4.7H_2O$, 5 mg/l Fe(NH_4)citrate, 0.19 mg/l $Na_2B_4O_7.10H_2O$, 1.0 mg/l Thiamine HCl, and 1.3 mg/l $CaCl_2.2H_2O$. Aeration rate was set at 4 l/min. Temperature was set at 30°C. The pH was partially controlled with 28% NH_4OH. Dissolved oxygen (DO%) was controlled through a feedback loop control of agitation speed. During cultivation, the mass of feeding medium, temperature, agitation rate, aeration rate, and the pH in the reactor were monitored continuously. Samples were periodically removed for further analysis.

2.5.3.4 Theory of fed-batch processes

The operating conditions in the fed-batch fermentation were designed based on a mathematical model that describes a balance of limiting substrate S and cell growth. The model was used to determine an optimal feed rate of substrate for maintaining a constant growth rate and product synthesis.

Under non-limiting growth conditions, the cell mass balance is

$$\frac{d\ X \cdot V}{dt} = \mu \cdot X \cdot V \qquad \textbf{(2.1)}$$

The substrate balance for a well-mixed bioreactor is given by

$$\frac{d\ S \cdot V}{dt} = F \cdot S_F - (\frac{\mu}{Y_{X/S}} + \frac{\mathrm{Q}_P}{Y_{P/S}} + m) \cdot X \cdot V \qquad \textbf{(2.2)}$$

During a fed-batch process, no accumulation of the dissolved substrate concentration is assumed ($d\,S \cdot V/dt = 0$). Combination of Eq (2.1) and Eq (2.2) yields an optimal, balanced substrate feeding rate

$$M_S = F \cdot S_F = (\frac{\mu}{Y_{X/S}} + \frac{Q_P}{Y_{P/S}} + m) \cdot X_0 \cdot V_0 \cdot \exp[\mu \cdot (t - t_0)] \qquad \textbf{(2.3)}$$

where M_S is the mass-feeding rate of the substrate glucose (g/h); F is the feeding flow rate (l/h); S is the substrate concentration in the culture (g/l); S_F is the substrate concentration in the feed (g/l); X is the cell concentration (g/l); X_0 is the initial cell concentration before feeding (g/l); V is the culture volume (l); V_0 is the initial culture volume before feeding; μ is the specific growth rate (hr^{-1}); $Y_{X/S}$ is cell yield on substrate (g-cell/g-substrate); Q_P is specific production rate of product, DPA (DPA/cell-hr); $Y_{P/S}$ is product yield on glucose (g-DPA/g-glucose); m is specific maintenance coefficient (g-glucose/g-cell-h); t_0 is time of feeding start; t is processing time.

2.5.3.5 Continuous bioreactor

The continuous bioreactor was operated in a 2l Braun bioreactor. The culture was initially conducted as a batch for 24-48 hrs before switching to continuous mode. In the continuous culture, the sterile feed medium containing carbon source, as specified, was pumped into the reactor at a dilution rate under investigation. Effluent liquid was continuously withdrawn by an outlet pump operated at a higher flow rate than the inlet pump to keep the culture volume constant. The pH was partially maintained by the addition of 3.0 M NaOH. Temperature was set at 37°C. Samples of fermentation broth were taken daily for further analysis. For the immobilized-cell fluidized-bed continuous

culture, an outlet tubing with a 50 mesh stainless steel screen at the end was used to retain the immobilized cells in the reactor. The tube was set at a level to maintain the liquid culture volume constant at 0.7 l.

2.5.3.6 Theory of continuous culture using immobilized mixed cells

A kinetic model describing continuous fermentation of a mixture of sugars by immobilized cells was developed. In mixed-sugar fermentation, immobilized cells of multiple strains ($X_{IM,j}$) is considered. Balance of each sugar substrate (C_{Si}) in the continuous culture is

$$\frac{dC_{Si}}{dt} = D_L(C_{Si,F}-C_{Si})-\sum_j X_{IM,j}q_{Si,j} \qquad \textbf{(2.4)}$$

where C_{Si} (mmole/l) is the concentration of substrate i in the culture; D_L (hr^{-1}) is dilution rate based on liquid volume of the culture; $C_{Si,F}$ (mmole/l) is concentration of substrate i in the feed medium; $X_{IM,j}$ (g-immobilized cells/l) is immobilized cell concentration of strain j; $q_{Si,j}$ represents specific uptake rate of substrate i by the immobilized cells of strain j (mmole/immobilized cells-hr).

For a completed conversion of each sugar at the same retention time, the optimal dilution rate $D_{L,opt}$ by immobilized cells becomes

$$D_{L,opt} = \frac{\sum_j X_{IM,j}q_{Si,j}}{C_{Si,F}} \qquad \textbf{(2.5)}$$

For fermentation of a two-sugar mixture (C_{S1} and C_{S2}) by two immobilized cells ($X_{IM,1}$ and $X_{IM,2}$), an optimized cell ratio for a completed utilization of all sugars at steady state is

$$X_{IM,1} \Big/ X_{IM,2} = \frac{C_{S1,F} q_{S2,IM2} - C_{S2,F} q_{S1,IM2}}{C_{S2,F} q_{S1,IM1} - C_{S1,F} q_{S2,IM1}} \qquad \textbf{(2.6)}$$

where $C_{S1,F}$ and $C_{S2,F}$ are the feed concentration of substrate S_1 and S_2 respectively; $q_{Si,IMj}$ represents specific uptake rate of substrate S_i by the immobilized cells IM_j.

2.6 DNA manipulation

2.6.1 Construction of plasmids

All plasmids were constructed using *S. cerevisiae* in vivo recombination as previously described (Shank, et al. 2006). Briefly, a *S. cerevisiae-E. coli* shuttle vector to be modified is digested with restriction enzymes. Inserts to be joined with the cut vector were amplified using the designed primers. The primers also contained 30 to 40-bp tails of homologous sequence to the cut vector on each side of the restriction enzyme cut site. The PCR amplicon(s) and the linearized vector were co-transformed into *S. cerevisiae* strain D603. The cell recombined the amplicon(s) and the cut vector through homologous recombination using its native recombination enzyme, creating a stable circularized plasmid. Plasmids were transferred into *E. coli* and screened for the desired construct.

2.6.2 Isolation of chromosomal DNA and plasmid DNA from *E. coli*

A technique for crude DNA extraction was used to isolate chromosomal DNA. An overnight grown culture of *E. coli* was lysed using 1 mg/ml lysozyme and incubated at 37°C for 1 hour or overnight at room temperature. The enzyme was inactivated by

placing the solution in an 80°C water bath for 10 minutes. The crude DNA was stored at -20°C for later use. QIAprep Spin Miniprep kit (QIAGen Worldwide, Germantown, USA) was used for plasmid isolation from a cell culture containing the target plasmid.

2.6.3 Isolation of plasmid DNA from *S. cerevisiae*

The QIAprep Spin Miniprep kit was used for isolating plasmid DNA from yeast cells. Briefly, 10 ml of an overnight-grown culture in selective medium was harvested and resuspended in a suspension buffer. The cell suspension was mixed with approximately 100 µl of nitric acid-washed glass beads and vortexed at maximum speed for 10 minutes. Lysis buffer was added to the mixture which was incubated then at room temperature for 5 minutes. The lysate was transferred to a QIAprep Spin Column. Plasmids were eluted from the column with 50 µl water. For subsequent transformation into *E. coli,* 20 µl of eluate was mixed with 100 µl of competent cells which was electroporated at 2.5 kV, 25µF and 400Ω for 10 microseconds.

2.6.4 Isolation of chromosomal DNA from *T. saccharolyticum*

Approximately 5 ml exponentially growing culture of *T. saccharolyticum* was harvested by centrifugation and the cell pellet was mixed with 3 mg/ml lysozyme and incubated at 37°C for 2 hours or overnight at room temperature. The mixture was then incubated at 80°C for 10 minutes to inactivate the enzyme. The crude DNA solution was used for further analysis. To further purify DNA from cell lysate, phenol-chloroform extraction was used. In brief, an equal volume of phenol and chloroform was mixed with 200 µl of the crude DNA solution. The aqueous layer was collected and washed by liquid-liquid extraction with chloroform to remove the trace of phenol. DNA was then

precipitated with ethanol by mixing the DNA solution with an ethanol/acetate solution at a ratio of 1/10 volume of 3M sodium acetate and 3 volumes of ethanol. The mixture was left in a -80°C freezer for 20 minutes for DNA precipitation. The pellet DNA was washed once with 70% ethanol and resuspended in water.

2.7 Construction of gene knockouts of *E. coli*

Strains containing genetic knockouts were obtained from the Keio knockout library (AGAC facility, University of Minnesota) which was originally generated by Mori, et al. 2000. The knockouts were created via homologous recombination to replace a gene on the chromosome with the kanamycin resistance gene (Datsenko and Wanner 2000). P1 transduction technique was then used to transfer the knockout gene from the Keio library strain into the MG1655 derived recipient strain. Two main steps were involved in the P1 transduction, preparation of P1 lysate and infection of recipient cells with the lysate.

2.7.1 Preparation of P1 lysate

The knockout strains from the library were grown overnight in Z-broth. Approximately 2 ml of cell culture sample was mixed with 100 µl of bacteriophage P1 stock solution and incubated at 37°C for 20 minutes. The mixture was then resuspended in Z-broth to a final volume of 10 ml and incubated at 37°C for 12 hours or until a clear lysate was observed. Then 100 µl of chloroform was added into the culture and vortexed for 30 seconds. The cell debris was removed by centrifugation at 5000 rpm for 10 minutes. The clear lysate was collected and stored at 4°C for later use.

2.7.2 Transferring the knockout gene

The gene transfer process was conducted by infecting a recipient strain with the P1 lysate of the Keio knockout strain. Approximately 100µl of the lysate was mixed with 2 ml of an overnight grown recipient cell grown in Z-broth. The mixture was incubated at 37°C for 30 minutes. Then the cells were washed twice with ice-cold citrate buffer to stop infection and to remove any remaining phages. The cells were allowed to recover in 10ml of LB broth supplemented with 25 mM sodium citrate for 30 minutes at 37°C and were washed twice with the ice-cold citrate buffer before plated on an LB agar plate supplied with kanamycin for screening. The procedure was repeated for the construction of a multiple-gene knockout. The knockout was confirmed via PCR and gel electrophoresis with primers designed to bind to the region outside and inside of the knockout gene (Table 2.5).

Table 2.5 Primer sequences designed for testing specific deleted genes in *E. coli*

Deleted genes	Outside primers[1]	Inside primers[1]
ldhA	5'-CGGCTTTATATTTACCCAGC-3'	5'-TACCCAACGAACCAATTTTC-3'
	5'-CGCAACAAACGCGGCTAC-3'	5'-GCTGGAAGAGCTGAAAAAGC-3'
frdA	5'-ACGCTTCAACCTTCATACCG-3'	5'-TTACGTGCCATTGCGGAGTG-3'
	5'-GATGCCGTTTCGCTCATAGT-3'	5'-TCACGATACAGTAGCGGGTG-3'
poxB	5'-ATGGATATCGTCGGGTTTGA-3'	5'-CCACCAGCTTTCATCTCCAT-3'
	5'-AAGCAATAACGTTCCGGTTG-3'	5'-TATTCCCTCCAGCGAAATTG-3'
pta	5'-GTTCGCCTGCTTCGTTAGTC-3'	5'-TCAGATCCGGGAAGATGAAC-3'
	5'-CTGCCGCTATGTTGAAGACA-3'	5'-TGTGCTGATGGAAGAGATCG-3'
adhE	5'-AAAGCGATGCTGAAAGGTGT-3'	5'-TGAATGCAGTCTGCTTGGTC-3'
	5'-AGAAAGCGTCAGGCAGTGTT-3'	5'-AAAACGTTGGTCCGAAACAC-3'

pykF	5'-GCGCCAATTGACTCTTGAAT-3'	5'-CACCGTACTGGTTGACGATG-3'
	5'-CCTGCCAGCAGAGTAGAACC-3'	5'-CACAACGCCTTTGCTCAGTA-3'
zwf	5'-CGCGTAACAATTGTGG-3'	5'-TCTACCCATTTCCAGGCTTC-3'
	5'-CTGGATTTTTTCCAGC-3'	5'-TGGGACACCCTGAGTGCACG-3'
maeB	5'-CTGTTTGATGCCGTCTAACTCGTTC-3'	5'-GAGCTGTCCGGCATACGGTC-3'
	5'-CTTTATCCATGAGTCGCCGCCTGTG-3'	5'-CGCTGGCAGGCAAACCGGTG-3'
pckA	5'-AGGAATGCGATTCCACTCAC-3'	5'-AACTAATCTGGGTGCCGTTG-3'
	5'-ATGGCTGGATCAAAGTCAGC-3'	5'-GGCAGAGTGAAGGTTTCTGC-3'
tktB	5'-ACTGAAGGAGCTGCAGGAAA-3'	5'-ATCTGACCGGTTACGACCTG-3'
	5'-GTGTCGCGTAAAAAGCCACT-3'	5'-CCAGATTCTGCCTTGAGAGG-3'

[1] Inside primers mean the primers specified to the deleted part of the gene, while outside primers means the primers specified to the undeleted portion of the gene.

2.7.3 Removal of antibiotic resistance marker

The kanamycin resistance marker in the knockout strain was removed using a flippase recombinase encoded on a helper plasmid pftA (Posfai, et al. 1997; Martinez-Morales, et al. 1999). The flippase recombinase excised the resistance gene between homologous sequence flanking sites by a single crossover into the sites. The knockout cells containing the pftA plasmid were incubated overnight at 30°C in LB media supplemented with ampicillin and heat-inactivated chlorotetracycline to promote the recombination. The strain was then screened for the loss of resistance marker by comparison plating. The plasmid was cured by incubating the cell overnight at 37°C.

2.8 Construction of genetic knockouts of *T. saccharolyticum*

2.8.1 Construction of knockout plasmids

The plasmid used for genetic knockout construction of *T. saccharolyticum* contained upstream and downstream homology regions of the targeted gene with a kanamycin marker inserted between them. The knockout plasmid was made using the method of in vivo recombination *S. cerevisiae*.

The upstream and downstream fragments of the target gene were amplified from genomic DNA of *T. saccharolyticum* using the designed primer pairs (Table 2.6). pMU433 was used as a vector which contained the URA3 blaster cassette (*ura3*) and the kanamycin resistance cassette (kan^R) as selection markers (Mai, et al. 1997; Mai and Wiegel 2000). The vector was digested at the BamH1 and BspE1 sites. The amplicons and the cut vector were recombined in the yeast cell through homologous recombination creating the knockout plasmid.

The plasmid was transformed into the *T. saccharolyticum* cell and integrated into its genome by homologous recombination of the two upstream and downstream flanking regions of the target gene. The knockout cells with the gene being replaced by the resistance marker were selected on kanamycin.

Table 2.6 Primers for construction of knockout plasmids used in *T. saccharolyticum*

Deleted genes	Up- primers[1]	Down- primers[1]
zwf	5'-TTCGACTGAGCCTTTCGTTTTA TTTGATGCCTGGTTTAGGTGAGC TTGCTATGTCG-3'	5'-GGGGTCCCGAGCGCCTACGA GGAATTTGTATCGACGATGAAAA TATATGATGTTTCAAT-3'
	5'-TTGTAGAATACAATCCACTTCA CAATGGGCACGCTTTCTATGAGT CAAGTCTCCTGTGCC-3'	5'-CCGTCAGTAGCTGAACAGGA GGGACAGCTGATAGACAATTGCC TTATGATTTCTTTAGCT-3'
mgs	5'-TCGACTGAGCCTTTCGTTTTAT TTGATGCCTGGTACAATCAAATT GCCAAGCTCTC-3'	5'-AGGGGTCCCGAGCGCCTACG AGGAATTTGTATCGTAAAAAATCA TAAACCTCCTTTC-3'
	5'-TTGTAGAATACAATCCACTTC ACAATGGGCACGTCAATATCTCT TTGTAGGCAATAGC-3'	5'-CCGTCAGTAGCTGAACAGGA GGGACAGCTGATAGAACTATCGA AAACAGCCGCTC-3'
sbm	5'-TTTCGACTGAGCCTTTCGTTTT ATTTGATGCCTGGGCAGCTAATA TAGGTTCCAATAATG-3'	5'-AGGGGTCCCGAGCGCCTACG AGGAATTTGTATCGTAATTTAAGC TTGATGAAAAGAGGC-3'
	5'-TTGTAGAATACAATCCACTTCA CAATGGGCACGCTTTAACTTTGT CCAAACAATCT-3'	5'-CCGTCAGTAGCTGAACAGGA GGGACAGCTGATAGATTACCATT TGAAGTCCTCCTTGA-3'

[1] The primers were designed based on up- and downstream sequences of the targeted gene knockout of *T. saccharolyticum*

2.8.2 Removal of the kanamycin resistance marker

The kanamycin resistance marker in the knockout cells was removed by transforming cells with a PCR product containing the two homologous flanking regions linked together without selection markers. The PCR product amplified using primers given in Table 2.7 was integrated into the genome at the flanking regions resulting in an excision of selective markers. The strain was screened for the loss of kanamycin by using a chloroacetate sensitivity counterselection system. Markerless gene deletion was confirmed by PCR and gel electrophoresis through inside and outside primers (Table 2.8). The process was repeated for construction of multiple-gene knockouts.

Table 2.7 Primers used for removal of the kanamycin resistance marker in *T. saccharolyticum*

Deleted genes	Up- primers[1]	Down- primers[1]
zwf	5'-TCGACTGAGCCTTTCGTTTTATTTGATGCCTGGTTTAGGTGAGCTTGCTATGTCG-3'	5'-CAGCACATCCTTCGCCAGCTGGCGTAATAGCGAACGATGAAAATATATGATGTTTCAAT-3'
	5'-CTATTACGCCAGCTGGCGAAGGATGTGCTGCTTTCTATGAGTCAAGTCTCCT-3'	5'- TGTAGAATACAATCCACTTCACAATGGGCACGCAATTGCCTTATGATTTCTTTAGCT-3'
mgs	5'- TCGACTGAGCCTTTCGTTTTATTTGATGCCTGGTACAATCAAATTGCCAAGCTCTC-3'	5'-CAGCACATCCTTCGCCAGCTGGCGTAATAGCGAAGTAAAAAATCATAAACCTCCTTTC-3'
	5'-CTTCGCTATTACGCCAGCTGGCGAAGGATGTGCTGTCAATATCTCTTTGTAGGCAATAGC-3'	5'- TTGTAGAATACAATCCACTTCACAATGGGCACGACTATCGAAAACAGCCGCTC-3'
sbm	5'-TTTCGACTGAGCCTTTCGTTTTATTTGATGCCTGGGCAGCTAATATAGGTTCCAATAATG-3'	5'-CAGCACATCCTTCGCCAGCTGGCGTAATAGCGAAGTAATTTAAGCTTGATGAAAAGAGGC-3'
	5'-CTTCGCTATTACGCCAGCTGGCGAAGGATGTGCTGCTTTAACTTTGTCCAAACAATCT-3'	5'- GTAGAATACAATCCACTTCACAATGGGCACGTTACCATTTGAAGTCCTCCTTGA-3'

[1] The primers were designed based on up- and downstream sequences of the targeted gene knockout of *T. saccharolyticum*

Table 2.8 Primer sequences designed for testing specific deleted genes in *T. saccharolyticum*

Deleted genes	Outside primers[1]	Inside primers[1]
zwf	5'-AGTTGACGCTGCCACAATAG-3'	5'-TTTGTGTCGTATCGGTTGGA-3'
	5'-TCCAGCATTGTGTTTTCACC-3'	5'-CGCACCGGTATCACCTTATT-3'
mgs	5'-TCAAATTGCCAAGCTCTCAA-3'	5'-ATTACAGGGATTCCGCTTCC-3'
	5'-TGTGGATCTATCGGCTTTCC-3'	5'-CTGCATCTCTCCATTCCAAA-3'
sbm	5'-GGTGGATGCTATGCAAAGTG-3'	5'-ATTCAGAGCTGCGAGAAGGA -3'
	5'-AGCCGCCTTTCTTTAAGACC-3'	5'-TCTTCTGCTTCCATGTCGTG -3'

[1] Inside means the primers specified to deleted part of the gene, while outside means the primers specified to up and down region of the gene.

2.9 Transformation

2.9.1 Transformation of *E. coli*

Electroporation was used to transform *E. coli* cells with plasmid (Gene Pulser II, BioRad Laboratories, Hercules, USA). Competent *E. coli* cell was prepared by washing overnight grown cells twice with sterile, deionized, ice-cold water. 100 µl of competent cells was then mixed with approximately 5 µl of the transforming plasmid. The mixture was transferred to a sterile electroporation cuvette and electroporated for 10 microseconds at 2.5 kV, 25µF and 400Ω. The electroporated cells were recovered in SOC medium and incubated for 1 hour at 37°C before plating on LB agar plates containing the selection antibiotic.

2.9.2 Transformation of *S. cerevisiae*

The 5 ml overnight-grown culture of *S. cerevisiae* in YPD medium was transformed using the lazy bones transformation protocol outlined by Soni, et al. 1993. Briefly, the cell pellet was collected and resuspended in 50 µl of the culture medium. 10 µl of sonicated salmon sperm DNA (10 mg/ml stock) and the plasmid DNA to be transformed was added to the cell suspension and vortexed briefly. 0.5 ml of yeast transformation solution (40% PEG 3350, 0.1 M LiAc, 10 mM TrisCl pH 7.5, 1 mM EDTA) was added along with 6 µl of DMSO. The suspension was mixed by inverting the vials, incubated at room temperature for 15 minutes, then heat shocked at 42°C for 15 minutes, and placed on ice for 2 minutes. The cell pellet was collected, washed once in sterile water, resuspended in 200 µl of sterile water, and spread on agar plates containing the appropriate selective agent. The plates were incubated at 30°C until colonies appeared.

2.9.3 Transformation of *T. saccharolyticum*

To introduce foreign DNA into *T. saccharolyticum,* 10 µl of frozen cells of *T. saccharolyticum* from a freezer culture was resuspended in 10 ml of M122C. Approximately 1 µg of plasmid DNA or PCR product was mixed with 1 ml of the culture. The mixture was allowed to grow at 50°C for 12 hours or until the culture reached mid-exponential phase. The culture was then plated using a pour plate technique by mixing the culture with 25 ml of liquid agar. The mixture was then poured immediately into sterile culture dish. The agar plates were incubated at 50°C for 2-4 days or until colonies appeared in an anaerobic jar containing carbon dioxide as the gas atmosphere. An anaerobic atmosphere was generated using an AnaeroGen sachet (Oxoid, Cambridge, UK). The strain of *T. saccharolyticum* carrying the plasmid of interest was isolated.

2.10 Polymerase chain reaction

Polymerase chain reaction (PCR) is a technique used to amplify a DNA fragment. The PCR was done using a puReTaq Ready-To-Go PCR beads (Amersham Biosciences, catalog # 27-9557-01). The reaction mixture contained 20 µl of sterilized deionized water, 1.5 µl of each primer (40 µM), 3 µl of DNA template, and 1 PCR bead. The standard PCR process consists of three major steps: denaturing, annealing and elongating. Denaturing the dual strand of DNA at 94°C unwound the DNA template to form a single stranded DNA. A complimentary primer then bound onto each strand of DNA in the annealing step. The annealing temperature was varied depending on the melting temperature of the primer. The last step involved elongation at 72°C at an approximate

extension rate of 1 kb/min. At this step, an enzyme polymerase attached on the primer-DNA pair and started copying the template. The process was repeated for 30 cycles in the MJ Mini Thermocycler (Biorad, catalog # PTC-1148C).

2.11 Gel electrophoresis

Agarose gel electrophoresis is a method used to separate DNA fragments based on size. The technique was used to isolate linear DNA fragments for DNA transformation, for sequencing, and for verifying the gene knockouts. 0.9% (w/v) of agarose concentration was used for the separation of 0.5-10 kb of DNA. The dissolved agarose gel was mixed with ethidium bromide at a final concentration of 0.5 μg/ml and poured into the casting tray. The solidified gel was placed in the Hoeffer Scientific Instruments model HE33 electrophoresis chamber (San Francisco, CA) filled with TBE electrophoresis buffer. DNA samples were mixed 10X BlueJuice Gel Loading Buffer (Invitrogen, catalog # 10816015) before loading into the gel. After loading the samples into the gel, the gel was run at 10 V/cm for 45-60 minutes. An ultraviolet transilluminator was used to visualize DNA bands on the gel.

2.12 Cell immobilization

Cells were grown in 2l Erlenmeyer shake flasks with a working volume of 1.5 l LB supplemented with 20g/l sugar under aerobic condition at 30°C and 200 rpm. An overnight-grown cell culture was used to prepare the immobilized cells. The cell pellet was collected through centrifugation at 5000 rpm, 4°C for 15 minutes, washed once with 1g/l ice-cold NaCl solution and concentrated. The sterilized agar solution containing 35g/l sodium alginate and 1g/l NaCl was thoroughly mixed with the concentrated cells at

concentrations as specified and extruded through 14 gauge tubing as drops into 40 g/l $CaCl_2$ solution at room temperature. The immobilized cell concentration was determined as the total cell dry weight per liter of alginate agar solution. The cell dry weight was estimated based on a determined correlation of 1 OD_{600} = 0.29 g-cdw/l. The bead size can be varied by varying the size of tubing used for generating the bead. Gel beads were suspended in $CaCl_2$ solution at 4°C for 1 hour for gelation and then washed twice with ice-cold sterile distilled water before use. Immobilized mixed cells were prepared by mixing cells at a proportion as specified before immobilization.

2.13 Analysis procedures

2.13.1 Cell dry weight determination

Cell samples were collected from the culture and washed once with ice-cold deionized water. The sample was then transferred to a pre-weighed tube and dried at 100˚C for 24 hours. The cell dry weight was determined by subtracting the sample mass from the mass of the empty tube.

2.13.2 Cell concentration measurements

The cell concentration was monitored via optical density at a wavelength of 600nm in a 1 cm Acryl-Cuvettes (No. 67.740, Sarstedt, Newton, NC) using a Hewlett Packard 8453 Diode Array spectrophotometer (Palo Alto, CA). The cell sample was diluted so that the optical density reading fell within the range of 0.05-0.80. Immobilized cell concentration was determined by dissolving gel beads in 4% (w/v) EDTA solution for 24 hours (King & Zall 1985). The resulting cell suspension was centrifuged, washed and resuspended in

sterile water before measuring the absorption. A standard curve was used for correlating optical density to cell concentration.

2.13.3 Metabolite quantification

Metabolite concentrations including sugars and other secreted fermentation products were determined using a HPLC system (Shimadz10A, Shimadzu, Columbia, MD) equipped with an autosampler (SIL-10AF), a cation exchange column (Biorad Labs, Hercules, CA), and two detectors in series including a UV–vis detector (SPD-10A) and a refractive index detector (RID-10A). The HPX-87H column was run in an isocratic mode at 65°C (or 50°C) and 0.5 ml/min (or 0.6 ml/min) using a mobile phase of 5 mM H_2SO_4. The HPX-87P column was operated at 80°C with deionized water as mobile phase at a flow rate of 0.6 ml/min. The standard curve correlating peak area to concentration of metabolites was used to determine the quantity of metabolites in the sample.

2.13.4 Carotenoid assay

Intracellular carotenoid was extracted from 5 ml cell culture. The centrifuged cell pellet was first washed with a 1 ml 0.9% ice-cold NaCl solution. Diapolycopendial (DPL) and lycopene (LYC) was extracted by resuspending the washed cell pellet in acetone with vigorous vortexing for 5 minutes and centrifugation at 16,000 rpm for 5 minutes to separate the cell debris. Diapolycopendioic acid (DPA) was extracted from the cell pellet using 10% (g/g) KOH with incubation at 30°C for 24 hours. The KOH extract solution was later purified by liquid-liquid extraction with chloroform. The carotenoid contents in the extract solutions were quantified by measuring the absorbance at 476 nm for LYC, at 506 nm for DPL (reported λmax, 507nm, Tao, et al. 2005) and at 490 nm for DPA

(reported λmax, 490nm, Tao, et al. 2005). The concentration of the carotenoid was determined using the calibration curve or extinction coefficients obtained from CaroteNature (Lupsingen, Switzerland) and the appropriate dilution factor.

2.14 Preparation of cell-free extracts

A cell culture sample in the mid-exponential phase (200 ml) was collected by centrifugation at 5000 rpm for 20 minutes at 4°C. The cells were washed twice in 50mM Tris–HCl pH 8.0 supplemented with 5mM dithiothreitol, resuspended in 10 ml of the same buffer, and frozen at -80°C. Cell disruption was conducted by a passing three times through a French pressure cell at 14,000 psi. Cell debris and unbroken cells were removed by centrifugation for 5 minutes at 5000 rpm at 4°C. Protein concentration was determined using the Bradford method (Bio-Rad Laboratories, Hercules, CA) with bovine serum albumin as a standard.

2.15 Enzyme assays

Enzymatic assays were conducted at 55°C at the absorbance of 340nm using a Hewlett Packard 8453 Diode Array spectrophotometer. The absorbance was measured every 10 seconds for approximately 5-10 minutes. The enzyme activities for phosphoglucose isomerase (PGI) and hexokinase (GLK) were determined by monitoring the level of NADP reduction in a coupled reaction with glucose-6-phosphate dehydrogenase (Hansen, et al. 2001). The phosphofructokinase (PFK) enzyme activity was measured by monitoring the oxidation of NADH via pyruvate kinase and lactate dehydrogenase (Schroder, et al. 1994). Test reactions of the enzyme assays are:

GLK: glucose + ATP → glucose-6P + ADP

PGI: fructose-6P → glucose-6P

PFK: fructose-6P + ATP → fructose-1,6P + ADP.

The assay solution for PGI enzyme activity contained 100 mM Tris-HCl (pH 7); 0.5 mM NADP; 1 U glucose-6-phosphate dehydrogenase; cell extracts (approximately 0.1 mg protein); and varied amount of fructose-6-phosphate (0-15 mM). The assay solution for GLK enzyme activity contained 60 mM Tris-HCl (pH 9); 20 mM $MgCl_2$; 4 mM ATP; 0.9 mM NADP; 3 U glucose-6-phosphate dehydrogenase; cell extracts; and 15 mM glucose. The assay solution for PFK enzyme activity contained 100 mM Tris-HCl (pH 8); 5 mM dithioerythritol; 10 mM $MgCl_2$; 15 mM fructose-6-phosphate; 2 mM ATP; 5 mM phosphoenol pyruvate; 0.3 mM NADH; 8 U pyruvate kinase; 30 U lactate dehydrogenase; and cell extracts.

2.16 Metabolic evolution

Two methods were applied for metabolic evolution with growth based selection. The first method was carried out through a serial dilution by sequentially sub-culturing cells in fresh medium broth containing the substrate of interest. *E. coli* was evolved in an aerobic shake flask, while *T. saccharolyticum* was evolved in an anaerobic shake tube. The evolution experiment was accomplished through serial dilution of the culture at a dilution ratio of 1:10 to 1:100 at a 8-24 hour interval. At each serial dilution step, cell growth was monitored via the change of optical density (OD_{600}) over time until the OD_{600} reached 0.8-1.0. A mutant with an improved growth rate was isolated from the final culture by plating on solid medium and stored at -80°C.

The second method was done in a chemostat culture with a working volume of 100 ml and a dilution rate as specified. The cell culture was initially treated with 100 μg/ml of a mutagen, nitrosoguanidine (NTG) before inoculating in the chemostat. The media used contained substrates and inhibitors of interest. The chemostat culture was continued for 30 days before a single clone was isolated from the culture after plating on solid medium and preserved as a frozen stock.

2.17 Estimation of half maximal inhibitory concentration

The concentration of inhibitor that caused 50% inhibition of *E. coli* growth (IC_{50}) was determined from batch shake-flask growth experiments containing a culture medium supplemented with various concentrations of the inhibitor of interest. The culture was initiated by adding sufficient cells to provide an approximated OD_{600} of 0.1. The specific growth rate in each culture was determined from the slope of the linear regression of semi-log time plots of the optical density OD_{600} during the exponential growth phase. The tolerance curve was constructed by plotting the relative specific growth rate against the concentration of the inhibitor. The relative specific growth rate is the ratio of a specific growth rate in presence of an inhibitor to that in absence of an inhibitor. The tolerance curve was then used to estimate the value of IC_{50} for each inhibitor.

2.18 Metabolic network analysis

The metabolic network of *E. coli* and *T. saccharolyticum* was reconstructed using the EcoCyc (http://ecocyc.org) and the MetaCyc Database (http://metacyc.org). Details of the metabolic map used in this study are in Fig. 1.1 for *E. coli* and in Fig. 1.2 for *T. saccharolyticum*. Reaction and gene notations are listed in the Appendix. Some linear

reactions in the model were lumped together for simplification. The biomass term in the network was determined by reconnecting the experimental measurement of cell composition to the precursors and energy of the central metabolism (Carlson and Srienc 2004). The cell composition of *E. coli* was obtained from Neidhardt 1987, while the cell composition of *T. saccharolyticum* was from Tang, et al. 2009. Both models have been experimentally validated (Trinh, et al. 2008; Unrean, et al. 2009; Unrean and Srienc. 2010 in preparation). The carotenoid synthesis pathway and the foreign ethanol producing pathway were also included in the *E. coli* model to account for the production of carotenoid and ethanol respectively. The metabolic network model was evaluated using METATOOL software (von Kamp and Schuster 2006). Elementary mode results were analyzed using Excel (Microsoft Corp.) for mode sorting and filtering.

Chapter 3

Metabolic network analysis tools for characterizing and engineering cell metabolism

3.1 Introduction

A cell functions through a highly branched and highly coupled reaction network. The impact of engineering a particular pathway on cellular metabolism is complicated due to the interconnections of metabolic pathways. Understanding how the network is organized and controlled and how the metabolic pathways are distributed is crucial for characterizing and engineering cellular metabolism.

Constructing an *in silico* model of a cellular metabolic network has been attractive since the model can provide the metabolic state of a cell under various genotypic and environmental conditions. Several groups have developed approaches for analyzing and processing metabolic networks (Edwards, et al. 1999; Ouzounis, et al. 2000; Burgard, et al. 2001; Edwards, et al. 2001). This chapter explains metabolic network analysis and methods for analyzing cell metabolism including (1) metabolic flux analysis with the focus on elementary mode analysis, (2) metabolic control analysis, and (3) thermodynamic analysis of metabolic pathways. The application of these techniques allows for a rational approach to strain development that will be illustrated in later chapters.

3.2 Theory of metabolic network analysis

A cellular metabolic network is composed of interconnected chemical reactions that connect a vast number of intermediary metabolites and cofactors with the stoichiometric

relationships. A framework of metabolic network analysis reflects the fundamental mass and energy balances around the reaction network. This includes the conservation of metabolites and cofactors, the balanced compartmentalization, and the thermodynamic constraints for reversibility/irreversibility of reactions. Information about metabolic fluxes through various pathways within a cell can be gathered based on these balance constraints.

With the assumption of a well-mixed metabolite concentration inside a cell, the accumulation rate of a metabolite being consumed or produced by cells can be predicted from the balance of constant fluxes around that metabolite (Roels 1983). Based on a simple mass balance of a metabolite M within a reaction network, the rate of accumulation is equivalent to the sum of chemical conversion rates (synthesis, degradation, and utilization) and the transportation rate (Fig. 3.1).

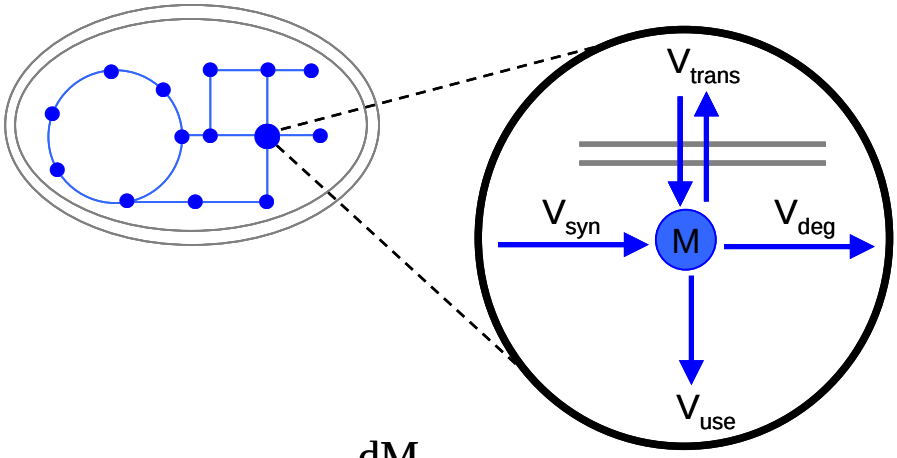


Figure 3.1 Flux balance of a cellular metabolic network. The accumulation rate of a metabolite M (dM/dt **) is balanced by the sum of conversion rates (rate of synthesis** $V_{synthesis}$ **, degradation** $V_{degradation}$ **and utilization** V_{use} **) and transportation rate** $V_{transport}$

The mass balance of metabolite M within a metabolic network can be written in mathematical form:

$$\frac{dC_M}{dt} = S \cdot J \pm \Phi - \mu \cdot C_M \qquad \textbf{(3.1)}$$

The concentration of metabolite M inside the cell is represented by C_M (mole/l). The accumulation rate dC_M / dt (mole/l-hr) represents the net flux of metabolite M. The multiplication of S and J represents the chemical rate of synthesis, degradation and utilization of metabolite M. S is the stoichiometric coefficient of reaction in the network where metabolite M is involved. The sign of the coefficient is positive for synthesis and negative for degradation or consumption. J (mole/l-hr) is the metabolic flux passing through these reactions. The rate of transportation including import or export flux of metabolite M is defined as Φ (mole/l-hr). The intracellular concentration of metabolite M can also be affected by a change of cell volume during cell growth. The effect is called dilution effect which is represented by the product between specific cell growth rate μ (1/hr) and concentration of metabolite M.

Usually, the change in the intracellular concentration of a metabolite caused by enzymatic reaction fluxes is much more significant than the effect due to change in cellular volume (Clarke 1988; Stephanopoulos, et al. 1998). Therefore, the dilution term ($\mu \cdot C_M$) can be neglected. The cellular metabolism often involves fast enzymatic reactions and high turnover of a substance, which makes the time constant characterizing metabolic transients very fast compared to that of cell growth and process dynamic (Edwards and Palsson 2000; Klamt and Stelling 2002). Therefore, the concentration of

metabolites can be assumed to be at a quasi steady state. The mass balance of metabolite M is reduced to

$$S \cdot J = \Phi \tag{3.2}$$

For internal metabolites, there is no transport flux passing across the cell system boundary ($\Phi = 0$). Thus, the rate of synthesis, degradation, and consumption of the internal metabolite is balanced by a linear combination of all the reactions that involve the metabolite fluxes (Roels 1983):

$$S \cdot J = 0 \tag{3.3}$$

For a cellular reaction network containing *m* metabolites and *q* reactions, the material balance equation of the overall metabolism can be expressed in a matrix vector:

$$\underline{\underline{S}} \cdot \underline{J} = \underline{0} \tag{3.4}$$

where $\underline{\underline{S}}$ is an *m* by *q* stoichiometric matrix of metabolite *m* in reaction r_q and $\underline{J}$ is a column vector comprised of net metabolite fluxes of reaction r_q. Due to thermodynamic constraints, some reactions are considered irreversible and their fluxes are constrained in the positive direction. That is, the inequality constraint for $\underline{J} \in$ irreversible reaction is

$$J \geq 0 \tag{3.5}$$

Since stoichiometry of these reactions is fixed by the reaction mechanism, metabolic fluxes $\underline{J}$ of the reaction network are the only unknown in the equation. Therefore, by solving the balance equation Eq (3.4) together with inequaility constraint Eq (3.5), the rate of metabolites being produced or consumed within the cell can be determined as shown in Fig. 3.2.

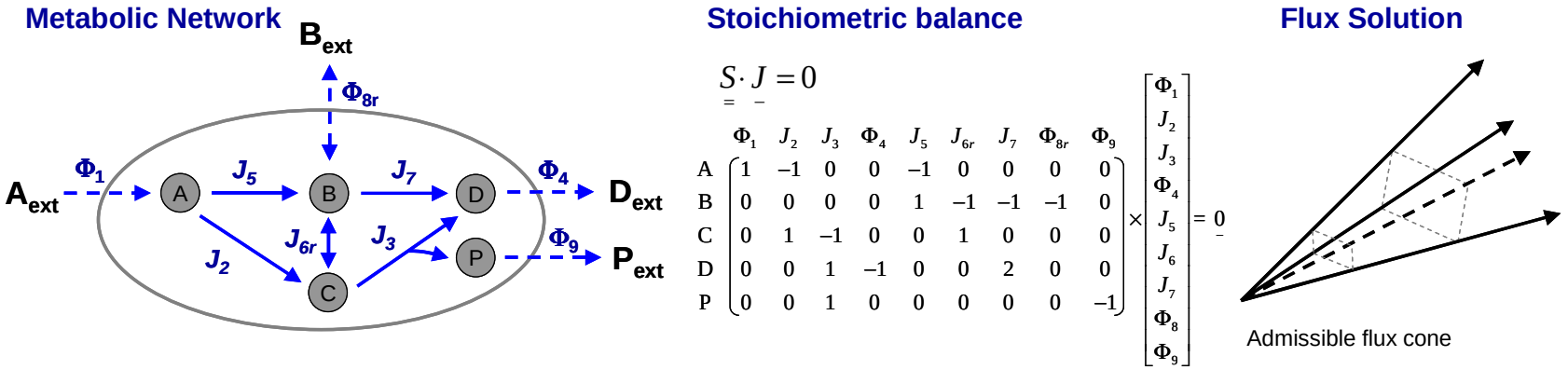


Figure 3.2 Analysis of a metabolic network. An example network is composed of 5 internal metabolites (M) and 4 external metabolites (M_{ext}) that are linked through 9 reactions, 2 reversible and 7 irreversible ones. S denotes a stoichiometry matrix where each element is the stoichiometry coefficient of an internal metabolites (row) in a reaction (column). The positive and negative coefficients are for the produced and the consumed metabolites respectively. The flux solution can be viewed as a convex flux cone where any point within the cone represents the admissible flux solution that satisfies the stoichiometric balance constraint.

3.3 Methods of analyzing a metabolic network

Several approaches exist to investigate Eq (3.4) which include metabolic flux analysis, flux balance analysis, extreme pathway analysis, and elementary mode analysis. Each of these approaches introduces some mathematically fundamental and biologically important aspects of metabolic modeling. Their differences are briefly discussed.

3.3.1 Metabolic flux analysis

Metabolic flux analysis (MFA) applies linear algebra to solve the metabolite balance equation expressed by Eq (3.6):

$$S_u \cdot J_u = S_m \cdot J_m \qquad \textbf{(3.6)}$$

It should be noted that Eq (3.6) is Eq (3.4) with the flux vector partitioned into the unknown flux vector (J_u) and the measured flux vector (J_m). Using the known stoichiometry and the measured flux vector, Eq (3.6) can be iteratively solved for the unknown flux (Wiechert 2001). The limitation of the MFA approach is its reliance of the measured flux vector. Accuracy of the determination of the intracellular flux with MFA is dependent on that of the measured flux.

3.3.2 Flux balance analysis

Flux balance analysis (FBA) utilizes a pathway of balanced reactions expressed by Eq (3.4) using the convex space analysis. The approach requires imposing an objective function subjected to constraints such as substrate uptake rate or cell growth rate in order to determine the metabolic flux vector (Price, et al. 2004). Based on the optimization strategy, FBA identifies only one optimal flux vector under a given set of constraints. As

a result, the FBA approach will depend on the validity of the formulated objective function.

3.3.3 Extreme pathway analysis

Extreme pathway analysis (EPA) is based on concepts from the convex space analysis. Unlike flux balance analysis, the EPA approach does not require imposing any objective functions. In EPA, all reversible reactions are decomposed into a set of two irreversible reactions in opposing directions for convex space analysis. Thus, solving the balance equation Eq (3.4) using EPA results in a set of spanning vectors with only positive or zero magnitudes that possibly exist in the network (Schilling, et al. 2000; Gagneur and Klamt 2004; Bazaraa, et al. 2005). These vectors are referred to as extreme pathways. Each vector represents a non-divisible, unique, linearly independent pathway and the sum of these vectors defines all positions within the convex space. For application to biological networks, the extreme pathways can be viewed as a genetically independent set of fluxes that operate within cellular metabolism. The overall cell metabolism can be also defined as a weighted average of these extreme pathways.

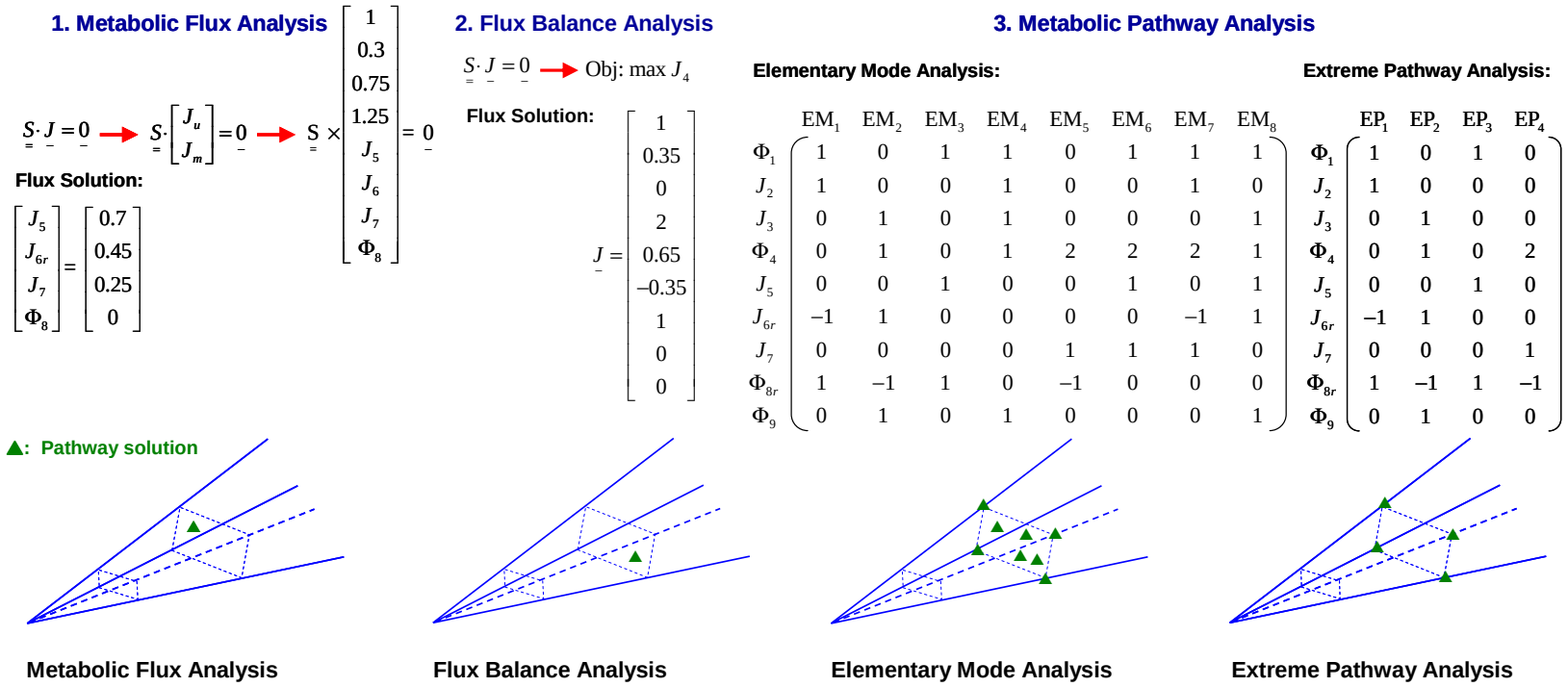
3.3.4 Elementary mode analysis

Elementary mode analysis (EMA) applies the same concepts as extreme pathway analysis for solving the balance equation. Different from EPA, EMA does not confine a flux to the positive domain (Pfeiffer, et al. 1999; Schuster, et al. 2000; Schuster, et al. 2002). That is, the reversible reaction is allowed to operate in reverse direction. Elementary modes represent a unique, non-decomposible, minimal set of reactions within the metabolic network that operate under the steady-state stoichiometrically balanced

reaction constraint. These elementary modes are determined by blocking a reaction (i.e., flux) and determining if the balance equation is satisfied. If not, another reaction is inhibited and the results are recalculated. This processes repeats until the balanced reaction constraint is fulfilled. Similar to EPA, the overall metabolism of a cell is a result of a weighted combination of all elementary modes in the network

$$J_{TOT} = p_1 \cdot E_1 + p_2 \cdot E_2 .... + p_n \cdot E_n \qquad \textbf{(3.7)}$$

where J_{TOT} is flux vector describing the overall network. E_i and p_i are flux vector of elementary mode and the weighted fraction of that mode respectively. It should be noted that EMA examines all possible unique and indivisible flux solutions available to the entire network, while EPA identifies only the flux solutions of the convex cone vertices (Fig. 3.3). That is, extreme pathways are a subset of elementary modes. Thus, EMA is the only approach that introduces a systematic way of extracting all biologically meaningful, balanced pathways from a metabolic network. Fig. 3.3 compares the methods for metabolic network analysis. The EMA approach can be applied to guide the development of a useful metabolic engineered strain.



	EM_1	EM_2	EM_3	EM_4	EM_5	EM_6	EM_7	EM_8
Φ_1	1	0	1	1	0	1	1	1
J_2	1	0	0	1	0	0	1	0
J_3	0	1	0	1	0	0	0	1
Φ_4	0	1	0	1	2	2	2	1
J_5	0	0	1	0	0	1	0	1
J_{6r}	−1	1	0	0	0	0	−1	1
J_7	0	0	0	0	1	1	1	0
Φ_{8r}	1	−1	1	0	−1	0	0	0
Φ_9	0	1	0	1	0	0	0	1

	EP_1	EP_2	EP_3	EP_4
Φ_1	1	0	1	0
J_2	1	0	0	0
J_3	0	1	0	0
Φ_4	0	1	0	2
J_5	0	0	1	0
J_{6r}	−1	1	0	0
J_7	0	0	0	1
Φ_{8r}	1	−1	1	−1
Φ_9	0	1	0	0

Figure 3.3 Methods of analyzing a metabolic network. (1) Metabolic flux analysis (MFA) where the flux vectors are partitioned into the unknown flux (J_u) and the measured flux (J_m). The calculation of the vector J_u is based on the known stoichiometry and the measured flux J_m. MFA identifies only one existing pathway solution within the cone based on available measured fluxes. (2) Flux balance analysis (FBA) where the metabolic flux vector J is determined based on the convex space analysis with the objective function, $\max J_4$. FBA identifies only one existing pathway solution within the cone based on the defined objective function. (3) Metabolic pathway analysis where the flux vector is identified using the convex space analysis without imposing of any objective functions. Elementary mode analysis determines all pathway solutions that lie anywhere of the cone (on the edge, on the face and inside the cone) whereas extreme pathway analysis identifies a set of pathway solutions that lie on the edge of the cone.

3.4 Example of elementary mode analysis

An example of a reaction network (Fig. 3.2) is used to help explain the application of EMA. The circle indicates the boundary of the system. The metabolites can be broken down into two types: external metabolites (outside the boundary) and internal metabolites (inside the boundary). Internal metabolites (M) are balanced, while external metabolites (M_{ext}) serve as sources or sinks for the network. Each reaction flux is limited to its designated direction. Flux Φ_i and J_i represent the transport flux and internal flux respectively. The example is comprised of 4 extracellular metabolites, 5 intracellular metabolites, and 9 reactions. The balance of each internal metabolite in a matrix form based on Eq (3.4) is given in Fig. 3.2. Negative and positive signs of reaction stoichiometry mean the consumption and production of metabolite respectively.

The software package used to calculate elementary mode is METATOOL (Schuster, et al. 1994; Pfeiffer, et al. 1999; Klamt and Schuster 2002). All reactions and their stoichiometric coefficients are written in an input text file. The reactions and metabolites are designated as reversible/irreversible and internal/external respectively. The program reads an input file and returns a result of elementary mode in a matrix form representing relative fluxes of reactions in each mode. In the example reaction network, a total of 8 elementary modes are identified (Fig. 3.4).

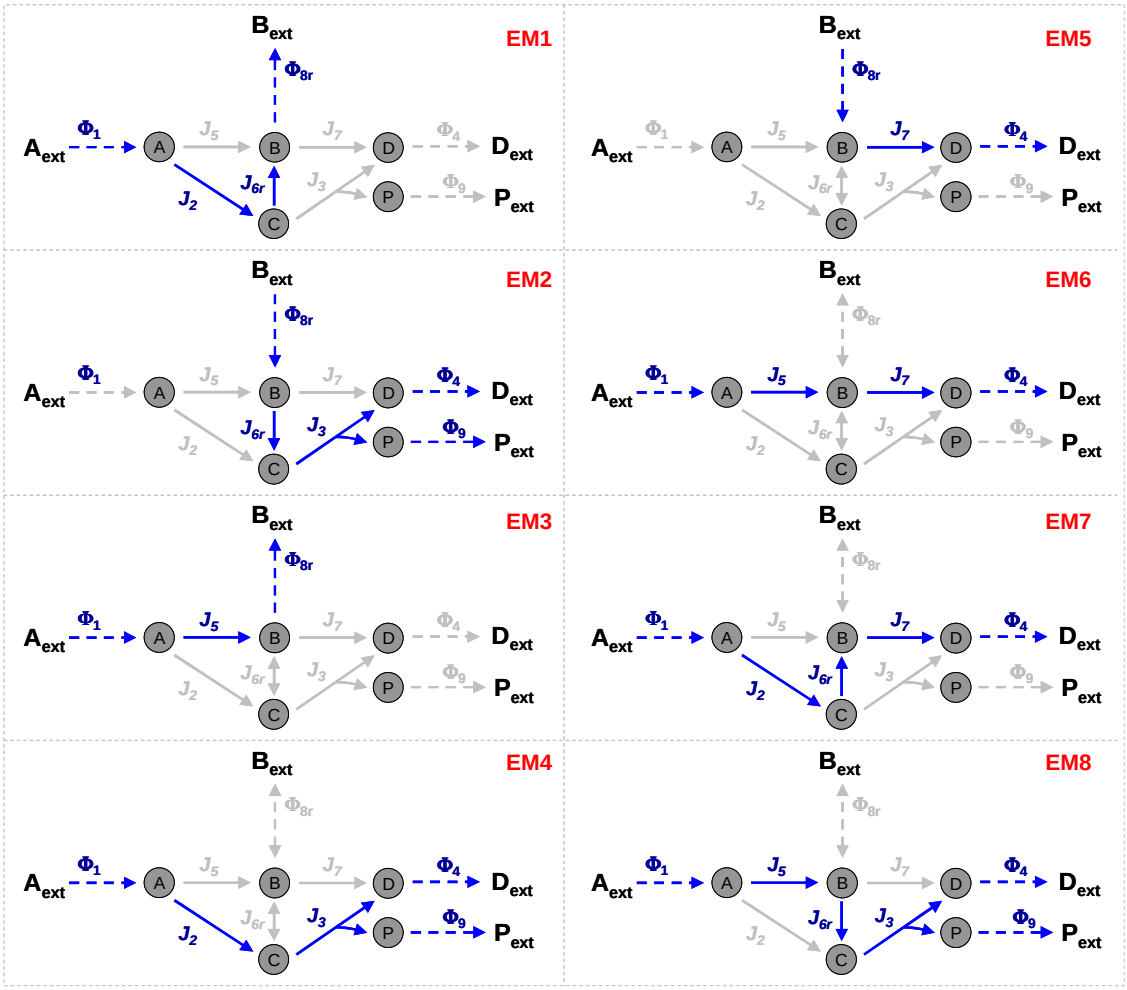


Figure 3.4 Representation of elementary modes based on elementary mode analysis of the example metabolic network

3.5 Application of elementary mode analysis

EMA decomposes a metabolic network of a cell into a set of unique, non-divisible pathways inherent to the stoichiometry and direction of reactions under steady state balance operation (Pfeiffer, et al. 1999; Schuster, et al. 2000; Schuster, et al. 2002). Thus, all physiological states corresponding to all pathway possibilities within a cell can be identified based on EMA without requiring the kinetic parameters of cellular metabolism.

This pathway information along with the direct correspondence between metabolic reaction network and genetic network allows the prediction of a cellular phenotype resulting from genotype perturbation. EMA has proven to be valuable in metabolic engineering where it is used as a guiding tool for modifying cellular metabolism towards the desired direction such as to improve the production of desired metabolites.

The EMA approach has been applied for analysis of the metabolic networks of *E. coli* (Liao, et al. 1996), *Methylobacterium exorequens* (Van Dien and Lidstrom 2002), *Saccharomyces cerevisiae* (Carlson, et al. 2002), *Lactobacillus rhamnosus* (Poolman, et al. 2004), *Corynebacterium glutamicum* (Kromer, et al. 2006), *Ralstonia eutropha* (Diniz, et al. 2006), and etc. The application of EMA includes the analysis of network robustness and flexibility, the prediction of network functionality and capacity, and the assessment of genetic or environmental perturbations on the cell phenotype. For example, EMA was applied to study the growth of *E. coli* and its metabolic response under different carbon sources and different levels of oxygen stress (Stelling, et al. 2002; Carlson and Srienc 2004; Reed and Palsson 2003). Gene deletions affecting growth in *E. coli* was also predicted using EMA. EMA was applied to direct the efficient strain development for product synthesis such as aromatic compound (Liao, et al. 1996), PHB (Carlson, et al. 2002; Wlaschin, et al. 2006), recombinant protein (Vijayasankaran, et al. 2005), cell biomass (Trinh, et al. 2006), L-methionine (Kromer, et al. 2006), cyanophycin (Diniz, et al. 2006), and ethanol (Trinh, et al. 2008).

3.6 Rational design of efficient strains

Elementary modes represent genetically independent pathways inherent to the overall cellular metabolism (Schuster, et al. 2000; Schuster, et al. 2002). The knowledge of elementary modes allows for the identification of pathway efficiency and the prediction of the effect of pathway elimination on cellular metabolism. The pathway efficiency is evaluated based on the molar yield. The yield information is then used to guide the metabolic engineering of an efficient strain by identifying target pathways to be eliminated for improved cell performance.

Details of the EMA application for rational design of an efficient strain are shown below using the network example given in Fig. 3.2. The presumed objective in this analysis is to optimize the conversion of substrate A into product D. The possible yield of D can be determined from the set of elementary modes consisting of a total of 8 modes. The objective can be fulfilled by eliminating all other modes with no production of D representing a yield of D equal to zero. Analysis of all modes (Fig. 3.4) suggests that deletion of reactions J_3 and Φ_{8r} will leave the system with 2 elementary modes that only produce the product D. If the system is considered a functioning cell, then knocking out genes corresponding to these reactions will force the cell to function only according to these remaining modes. The mutant with these gene knockouts should be able to achieve the most efficient A to D conversion, since all inefficient pathways are eliminated. It should be noted that the cell can be designed with multiple objectives such as to optimize for product synthesis while retaining cell growth or to constrain the cell so the product synthesis and cell growth are coupled. For a large metabolic network, targeted deletion to

enforce the cell to function according to the applied objective can be identified using a sequential reaction deletion strategy as demonstrated in Fig. 3.5.

Implementing multiple-reaction deletions to rationally design an efficient strain is based on the following constraints. Targeted reaction for elimination must result in minimization of the number of remaining modes while still maintaining maximum product yield. The targeted reaction deletion of the first round of screening is applied as a background to determine the targeted reaction deletion in the next round. In this analysis, reaction Φ_{8r} and J_3 are identified as the first and second target reaction deletion respectively. The combination of these reaction deletions results in a strain that can efficiently produce product D. The sequential step of reaction deletion can be continued until no further improvement is achieved. The approach ensures identification of a minimal set of reaction that limits the cell functionality only to the operation of the pathway of interest.

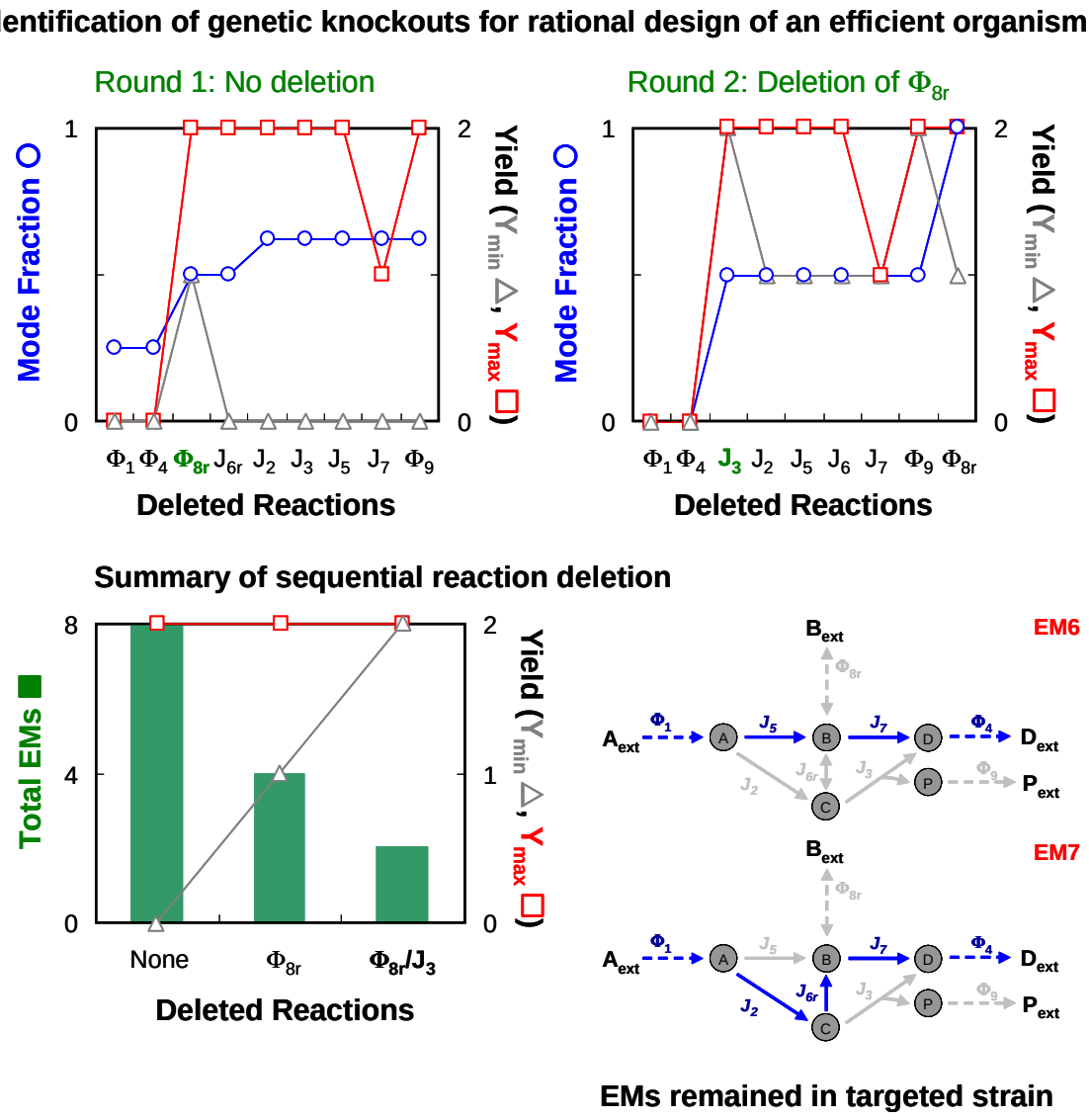


Figure 3.5 Rational design of a mutant strain with efficient conversion of substrate A into product D. (Top) Evaluation of reaction deletion affecting the A to D conversion, maximal and minimal yield of D and the fraction of elementary modes remained. Targeted reaction after each round of evaluation shown in bold was sequentially used as the background for the next round of evaluation. The process was continued until no further improvement could be achieved. (Bottom) The summary of the effect of sequential reaction deletion on the total number of elementary modes and yield range of the product D is included. The elementary mode remained in the strain with targeted reaction deletion is also included.

The example demonstrates the feasibility of the EMA approach for rational strain design for improved cell performance. The strategy was applied to design a minimal cell with efficient functionality for the production of products such as DPA in Chapter 4, LYC in Chapter 5, and ethanol in Chapter 8.

3.7 Metabolic control analysis

Metabolic control analysis (MCA) is a technique that allows for a quantification of the degree of flux control within a cell reaction network. The approach aims to characterize the sensitivity of metabolic responses (the global effects) affected by changes in enzyme activities or parameters (localized perturbations). MCA determines the degree of control that a given reaction (i.e., enzyme) exerts on flux or on the concentration of metabolites of a given pathway or system. Evaluation of the control structure of a metabolic pathway provides insight on how the flux control is distributed within the pathway. The result of MCA can serve as a guideline for optimization and manipulation of the pathway. The basis of MCA for analysis of a metabolic pathway relies on three kinds of coefficients: elasticity, control, and response (Heinrich and Schuster 1998).

The elasticity coefficient refers to the kinetic response of the steady state flux v_i through the pathway after perturbations of metabolite concentration X_j

$$\varepsilon_{X_j}^{v_i} = \frac{\partial v_i}{\partial X_j} \qquad \textbf{(3.8)}$$

To obtain dimensionless and normalized values of the elasticity coefficient, the scaling factor X_j / v_i is applied. The scaled version of the elasticity coefficient is

$$\varepsilon_{X_j}^{v_i} = \frac{X_j}{v_i} \cdot \frac{\partial v_i}{\partial X_j} = \frac{\partial \ln v_i}{\partial \ln X_j} \qquad \textbf{(3.9)}$$

The flux control coefficient measures the degree of control that the reaction rate of a given enzyme exerts on the metabolic flux of a given cell system. The control coefficient $C_{v_i}^{J_X}$ described in Eq (3.10) refers to the response effect of metabolic flux J_X by a perturbation of reaction rate v_i:

$$C_{v_i}^{J_X} = \frac{\partial J_X}{\partial v_i} \qquad \textbf{(3.10)}$$

The control coefficient expressed as a scaled version is

$$C_{v_i}^{J_X} = \frac{v_i}{J_X} \cdot \frac{\partial J_X}{\partial v_i} = \frac{\partial \ln J_X}{\partial \ln v_i} \qquad \textbf{(3.11)}$$

All flux control coefficients are shown to be zero or positive, given that the enzymes were not subjected to substrate inhibition or product activation. The control coefficients of all pathway enzymes and transporters in a metabolic system also add up to one based on the flux summation theorem. The sum of the product of the elasticity coefficients and the flux control coefficients is zero based on the connectivity theorem (Fell 1997).

The response coefficient defines the sensitivity of the pathway J_X to an external effector, P such as inhibitor, activator, or inducer:

$$R_P^{J_X} = \frac{\partial \ln J_X}{\partial \ln P} \qquad \textbf{(3.12)}$$

The response coefficient can be rewritten as the sum of the sensitivity of the pathway to the activity of the enzyme that is affected by the external effector (given by the enzyme's flux control coefficient) and the effect of the external effector on the activity of that enzyme (given by the enzyme's elasticity)

$$R_P^{J_X} = \frac{\partial \ln J_X}{\partial \ln v_i} \cdot \frac{\partial \ln v_i}{\partial \ln P} = C_{v_i}^{J_X} \cdot \varepsilon_P^{v_i} \qquad \textbf{(3.13)}$$

In practice, a measurable noninfinitesimal change of the metabolic flux and the rate of reaction are undertaken. Hence, the control coefficients can be estimated;

$$\begin{aligned} \varepsilon_{X_j}^{v_i} &= \left(\frac{\Delta \ln v_i}{\Delta \ln X_j} \right) \\ C_{v_i}^{J_X} &= \left(\frac{\Delta \ln J_X}{\Delta \ln v_i} \right) \\ R_P^{J_X} &= \left(\frac{\Delta \ln J_X}{\Delta \ln P} \right) \end{aligned} \qquad \textbf{(3.14)}$$

In this dissertation, the flux control coefficient of the metabolic network is examined. The coefficient is applied to determine the rate-limiting reaction within the cell metabolic network. For a reaction that a small variation of the reaction rate promotes a significant change in the metabolic flux, this reaction exerts an elevated flux control, and has a high flux control coefficient. A reaction with the maximal flux control coefficient is presumed to be a rate-controlling step and is a target of manipulation to increase the metabolic flux J_X. The control coefficient can also be used to examine the change between two steady-state systems:

$$C_{v_i}^{J_X} = \frac{\ln J_{X,\,\text{system 2}} - \ln J_{X,\,\text{system 1}}}{\ln v_{i,\,\text{system 2}} - \ln v_{i,\,\text{system 1}}} \qquad \textbf{(3.15)}$$

Quantification of the difference in metabolic fluxes using the control coefficient reflects the degree of control that each reaction in the pathway has on the steady-state flux response J_X. Eq (3.15) represents the degree of control for each reaction during a shift of the system from one metabolic state to another (i.e., system 1 → 2). The analysis of flux control coefficients helps to identify the rate-controlling step in the pathway that is responsible for the metabolic shift.

3.8 Thermodynamic analysis of metabolic pathways

A cell system can be modeled as an open, irreversible process operating at nonequilibrium steady state (Qian, et al. 2003; Qian, et al. 2005). The metabolism of a cell functions by collecting the chemical bond energy from low-entropy substrates and converting them into high-entropy products. Thus, the utilization of pathway choices within the cell can be determined based on the entropy generating properties of the pathways. This has been suggested in the study by Wlaschin, et al. 2006 which shows the determination of fractional contributions of elementary modes to the cellular metabolism based solely on entropy information.

The entropy of reaction for each elementary mode ΔS_i can be computed from the overall reaction stoichiometry of external metabolites in each individual mode

$$\Delta S_i = \sum_{m=1}^{k} \nu_{i,m} \Delta S_{i,m} \qquad \textbf{(3.16)}$$

where ν_m is the stoichiometry coefficient of metabolites m in the overall reaction of elementary mode i containing k reactants, and ΔS_m (kJ/K-mole) is the entropy of formation of metabolites m.

The entropy of formation can be computed using the Gibbs equation

$$\Delta S_m = \frac{\Delta H_m - \Delta G_m}{T} \tag{3.17}$$

where T is the absolute temperature at the condition considered, ΔH_m (kJ/mole) is the molar enthalpy of formation for the metabolites m, and ΔG_m (kJ/mole) is the molar Gibbs free energy of formation of the metabolites m.

The enthalpy and free energy of formation for each metabolite m can be computed from

$$\Delta H_m = \Delta H_m^{\circ} + \int_{298}^{T} C_{P,m} dT \tag{3.18}$$

$$\Delta G_m = \Delta G_m^{\circ} + RT \ln c_m \tag{3.19}$$

where $C_{P,m}$ (kJ/K-mole) is the heat capacity of metabolites m. ΔH_m° and ΔG_m° are the standard enthalpy of formation and the standard free energy of formation for 1 mole at 298 K respectively. R (kJ/K-mole) is the universal gas constant, and c_m is the molar concentration of metabolites m.

Substituting Eq (3.18) and (3.19) into (3.17) gives the entropy of formation of metabolites m

$$\Delta S_m = \frac{1}{T}\left(\Delta H_m^{\circ} + \int_{298}^{T} C_{P,m} dT - \Delta G_m^{\circ} - RT \ln c_m \right) \tag{3.20}$$

The reaction entropy of individual elementary modes thus becomes

$$\Delta S_i = \frac{1}{T} \sum_{m=1}^{k} \nu_{i,m} \left(\Delta H_{i,m}^{\circ} + \int_{298}^{T} C_{P\ i,m} dT - \Delta G_{i,m}^{\circ} - RT \ln c_{i,m} \right) \tag{3.21}$$

The reaction entropy can be estimated from the standard molar values of the enthalpy and free energy since the terms involving the heat capacity and the species concentration are very small compared to the other terms for the compounds considered in this study. The standard values of enthalpy and free energy can be estimated using the degree of reduction correlations given in Sandler and Orbey (1991).

It has been suggested from the previous experimental observation according to Wlaschin, et al. 2006 that the reaction entropy of the i^{th} mode, ΔS_i is linearly correlated with the natural log of the probability of usage of the corresponding mode

$$\Delta S_i = -b \cdot \ln p_i \qquad \textbf{(3.22)}$$

where b is a constant of the linear correlation.

The total entropy production of the overall metabolism, ΔS_{TOT} is defined as the sum of the product of the individual reaction entropy of modes, ΔS_i and its probabilities:

$$\Delta S_{TOT} = \sum p_i \cdot \Delta S_i \qquad \textbf{(3.23)}$$

Combining Eq. (3.22) and (3.23), the overall entropy generation of a cell can be represented based on the usage probability of individual modes:

$$\Delta S_{TOT} = -b \cdot \sum p_i \cdot \ln p_i \qquad \textbf{(3.24)}$$

Eq (3.24) is reminiscent of the Boltzmann entropy that describes the entropy content of a system (macrostate) as a weighted sum of the probability distribution of elementary modes (microstates). This relationship suggests that elementary modes can be interpreted as a discrete ensemble of states representing the overall cell metabolism analogous to an ensemble of quantum states of a macroscopic system in statistical thermodynamics. In the

next section, we will show that the relationship expressed by Eq (3.22) and (3.24) can indeed be derived from the ensemble method of statistical thermodynamics.

3.9 Determination of the total reaction entropy of cells using ensemble method

The ensemble method of statistical thermodynamics is applied for finding the total reaction entropy of a cell system if the reaction consists of the sum of fractional contributions by elementary modes. We consider the entire metabolic network of the cell as a canonical ensemble of n quantum states defined by elementary modes. A schematic representation of the canonical ensemble of the metabolic network of a cell is provided in Fig. 3.6

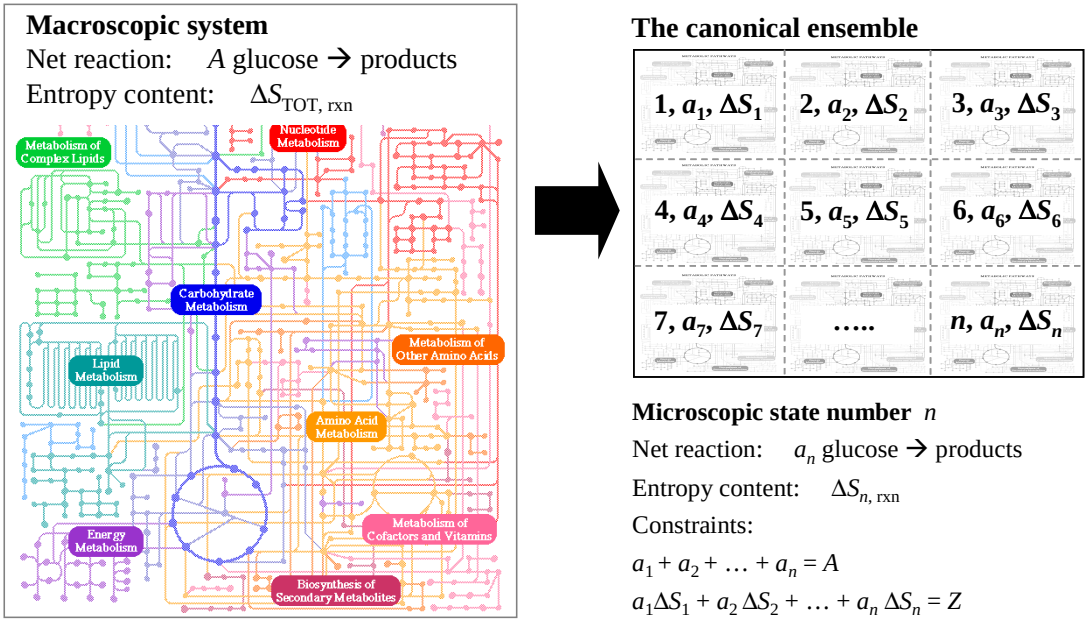


Figure 3.6 A representation of cellular metabolism by the canonical ensemble

If each elementary modes starts with the uptake of glucose molecules, one can characterize each of the n quantum states by the reaction entropy ΔS_i per molecule

glucose reacted and one can observe the number a_i molecules of glucose that are metabolized through each elementary mode. Thus, we can describe any one state of the ensemble by the following:

State No: 1, 2, 3, …, n

Entropy generation: ΔS_1, ΔS_2, ΔS_3, …, ΔS_n

Molecule of glucose utilized: a_1, a_2, a_3, …, a_n

On the basis of mass conservation and entropy generation being extensive, the entire ensemble must satisfy the two constraints:

$$\sum a_i = \mathrm{A} \tag{3.25}$$

$$\sum a_i \Delta S_i = \mathbb{Z} \tag{3.26}$$

where a_i is the number of molecules of glucose metabolized by the i^{th} quantum state (elementary mode), ΔS_i is the reaction entropy generated by this quantum state, A is the total number of molecules of glucose being utilized by the entire ensemble, and $\mathbb{Z}$ represents the entropy formation of the ensemble.

The distribution of the A molecules of glucose such that a_1 is in the 1st quantum state, a_2 are in the 2nd quantum state, and so on can be achieved in many possible ways. The number of ways $W(a_1, a_2, a_3, ...)$ that A identical, distinguishable molecules of glucose can be arranged into groups, each associated with a specific quantum state, is defined as:

$$W(a_1, a_2, a_3, ...) = \frac{\mathrm{A}!}{a_1! a_2! a_3! ...} = \frac{\mathrm{A}!}{\prod a_i!} \tag{3.27}$$

Consequently, the probability p_i that any given molecules of glucose is in the i^{th} quantum state is given by averaging over all the possible distributions

$$p_i = \frac{\langle a_i \rangle}{\mathrm{A}} = \frac{\sum W(a_i) \cdot a_i}{\mathrm{A} \sum W(a_i)} \qquad \textbf{(3.28)}$$

If A is arbitrarily large, the distribution function $W(a_i)$ can be made arbitrarily narrow. The $W(a_1^*, a_2^*, a_3^*, ...)$ where a_i^* is the set of a_i's that maximizes the function $W(a_i)$ will be overwhelmingly huge as compared to $W(a_i)$ for other remaining set of a_i's. Thus, the function $W(a_i)$ can be approximated as $W(a_1^*, a_2^*, a_3^*, ...)$, that is the most probable distribution. We can then write the probability p_i

$$p_i = \frac{\langle a_i \rangle}{\mathrm{A}} = \frac{\sum W(a_i) \cdot a_i}{\mathrm{A} \sum W(a_i)} \simeq \frac{W(a_i^*) \cdot a_i^*}{\mathrm{A} W(a_i^*)} = \frac{a_i^*}{\mathrm{A}} \qquad \textbf{(3.29)}$$

where a_i^* is the set of a_i's associated with the i^{th} quantum state for the most probable distribution $W(a_i)$ of the ensemble.

Eq (3.29) reveals that we can compute the probability of the i^{th} quantum state by determining only the distribution of a set of a_i^* that maximizes a function $W(a_i)$ under the given constraints. The solution of this problem can be obtained with the Lagrange multiplier method.

The function $W(a_i)$ can be simplified by taking the natural logarithm of $W(a_i)$ and with Stirling's approximation $\ln W(a_i)$ becomes

$$\ln W(a_1, a_2, a_3, ...) = \ln \frac{\mathrm{A}!}{\prod a_i !} \simeq \mathrm{A} \ln \mathrm{A} - \sum a_i \ln a_i \qquad \textbf{(3.30)}$$

Applying the method of Lagrange multipliers, we next differentiate Eq (3.30) with respect to a_i, while enforcing the two constraints Eq. (3.25) and (3.26)

For $i = 1, 2, \ldots, n$ $$\frac{\partial}{\partial a_i}\left\{\ln W(a_i) - \alpha \sum a_i - \beta \sum a_i S_i\right\} = 0 \qquad \textbf{(3.31)}$$

where α and β are the Lagrange multipliers.

The set of a_i's that maximizes $W(a_1, a_2, a_3, \ldots)$ can be found from solving Eq (3.31). The solution of a_i's for the most probable distribution for the systems making up the ensemble is

$$a_i^* = e^{-\alpha - 1} \cdot e^{-\beta S_i} \qquad i = 1, 2, \ldots, n \qquad \textbf{(3.32)}$$

Combining Eq (3.25) and (3.32), we obtain

$$\mathrm{A} = e^{-\alpha - 1} \sum e^{-\beta S_i} \qquad \textbf{(3.33)}$$

We can then find an expression for the probability p_i in term of ΔS_i and β by substituting Eq (3.32) and (3.33) into Eq (3.29)

$$p_i = \frac{e^{-\alpha - 1} \cdot e^{-\beta \Delta S_i}}{\mathrm{A}} = \frac{e^{-\alpha - 1} \cdot e^{-\beta \Delta S_i}}{\sum e^{-\alpha - 1} \cdot e^{-\beta \Delta S_i}} = \frac{e^{-\beta \Delta S_i}}{\sum e^{-\beta \Delta S_i}} = \frac{e^{-\beta \Delta S_i}}{Q} \qquad \textbf{(3.34)}$$

where Q is a partition function that has been defined as $Q = \sum e^{-\beta \Delta S_i}$

The expression of Eq (3.34), which is analogous to the Boltzmann distribution law, suggests that the usage probabilities of elementary modes can be determined based on the reaction entropy of individual elementary modes. At a temperature of absolute zero, the total entropy of a system has a value of zero, and the probability of generating zero entropy is 1, according to the Third law of thermodynamics. This leads to

$$Q = \sum e^{-\beta \Delta S_i} = 1 \qquad \textbf{(3.35)}$$

Eq (3.34) and (3.35) allow for an evaluation of the probability of elementary mode p_i and constant β based on the thermodynamic properties of the pathways, particularly the reaction entropies generated by individual elementary modes.

We can rewrite the probability function Eq (3.34) in terms of the reaction entropy of the i^{th} quantum state (elementary mode)

$$\Delta S_i = -\frac{1}{\beta} \ln p_i - \frac{1}{\beta} \ln Q \qquad \textbf{(3.36)}$$

With the constant $1/\beta$ being replaced by b and Q is equal to one, Eq (3.36) thus becomes

$$\Delta S_i = -b \cdot \ln p_i \qquad \textbf{(3.37)}$$

The constant b, which represents the ratio between the associated usage probabilities and the reaction entropies of individual elementary modes, is analogous to Boltzmann constant. However, it is not a universal constant as b would vary depending on the number of family modes present and the reaction entropies of the corresponding modes. Eq (3.37) expresses the reaction entropies of the individual elementary modes as a linear function of the natural log of the usage probabilities of the corresponding elementary modes. This provides the theoretical justification for the relationship that has been previously observed in the experiment (Wlaschin, et al. 2006, Eq (3.22)). The rate of entropy production of individual elementary modes is given by

$$r_{\Delta S_i} = \Delta S_i \cdot r_g \qquad \textbf{(3.38)}$$

where r_g is the rate of glucose uptake (mole/l-hr).

One can also multiply Eq (3.37) with the asymptotic uptake rate of glucose (r_g^*) to obtain the rate of entropy production of individual modes for a fully evolved metabolism (see Chapter 9 for details)

$$r_{\Delta S_i}^* = -b^* \cdot \ln p_i^* \qquad \textbf{(3.39)}$$

where the $r_{\Delta S_i}^*$ is equivalent to $r_g^* \cdot \Delta S_i$, and p_i^* is the pathway probability of elementary modes for a fully evolved metabolism.

Given the probability p_i that a_i molecules of glucose is in the i^{th} quantum state, one can calculate any thermodynamic property of a macroscopic system (a cell system). The thermodynamic property is defined by calculating the ensemble average of that property in the i^{th} quantum state that is associated with the probabilities p_i. Thus, the total reaction entropy of the cell metabolism is

$$\Delta S_{TOT} = \langle S \rangle = \sum p_i \Delta S_i \qquad \textbf{(3.40)}$$

Substituting Eq (3.37) into Eq (3.40), we obtain

$$\Delta S_{TOT} = -b \cdot \sum p_i \ln p_i \qquad \textbf{(3.41)}$$

Eq (3.41) describes the total reaction entropy of the cell metabolism as a weighted sum of the usage probabilities which is similar to the expression for the Boltzmann entropy. The total rate of entropy production of the cell system $r_{\Delta S_{TOT}}$ is then defined as the product of the uptake rate of glucose molecules and the total reaction entropy of the cell system.

$$r_{\Delta S_{TOT}} = r_g \cdot \Delta S_{TOT} \qquad \textbf{(3.42)}$$

Thus, we have shown that the ensemble method of statistical thermodynamics can be applied for finding the usage probabilities of elementary modes Eq (3.34), and the total

reaction entropy of the cell system Eq (3.41). The derivations based on the ensemble method have confirmed that a cell metabolism can be considered an ensemble that is a collection of quantum states expressed by elementary modes.

The developed relationship for the total reaction entropy of cellular metabolism given in Eq (3.41) has led to a theory proposed by Srienc and Unrean (2010) suggesting entropy as a driving force behind the evolution of metabolic pathways. That is, the evolution of metabolic networks is driven to maximize the rate of entropy production. The proposed theory is in agreement with the maximum entropy production principle (Jaynes 1957; Martyushev and Seleznev, 2006; Zhao and Kurata, 2009). From the theoretical prediction, a fully evolved cell has a natural tendency to operate at the state of maximum entropy production where the usage probabilities of individual modes are distributed according to the Boltzmann distribution law. Moreover, for a cell system that is not at the fully evolved state, the evolutionary path towards the maximum entropy generating state can be predicted. An unevolved cell system would naturally evolve to achieve its optimality by spontaneously redistributing the associated usage probabilities of elementary modes towards the probabilities predicted by the Boltzmann distribution law at the maximum rate of entropy production. A validation of the theoretical prediction is presented in Chapter 9.

3.10 Determination of metabolic flux distributions of cell metabolism

The knowledge of elementary modes and their corresponding probabilities provides a useful tool for evaluating the metabolism of a cell in quantitative terms. The overall fluxes for individual reactions within the metabolic networks R_j can be expressed as a

linear, weighted contribution of the rate of individual reactions specified in each elementary mode

$$R_j = \sum p_i \cdot r_{i,j} \quad \textbf{(3.43)}$$

where p_i is mode probability that is the fractional contribution of elementary mode i to the overall metabolic flux R_j, and $r_{i,j}$ represents rate of reaction j derived from elementary mode i. The values of these probabilities can be varied from 0 to 1 and the sum of all probabilities in a given cell system must be one.

The corresponding elementary modes that have the same overall stoichiometry for accumulating external metabolites can be grouped into a family of modes (Wlaschin, et al. 2006). Thus, Eq (3.43) can be written in a reduced form for the net accumulation rates of external metabolites R_j^{ext}:

$$R_j^{ext} = \sum p_k \cdot r_{k,j}^{ext} \quad \textbf{(3.44)}$$

where p_k is probability for family mode k, and $r_{k,j}^{ext}$ is rate of external metabolite derived from the family mode k. Using Eq (3.44), one can determine the family probabilities based on the measured accumulation rates of external metabolites. By assuming an equal contribution of each elementary mode within the family, the probability of elementary mode p_i in the family mode k is

$$p_i = p_k \big/ n \quad \textbf{(3.45)}$$

where n is the number of elementary modes in the family k. Substituting Eq (3.45) into Eq (3.43), a steady-state flux distribution within the cell metabolism can be reconstructed

based solely on the accumulation rates of external metabolites measured in an experiment.

3.11 Conclusion

Analysis of the metabolic network of a cell allows for an examination of how the network is organized and controlled. This information can provide insight into the physiology, robustness, and control of cellular metabolism. Tools for metabolic network analysis are developed on the basis of mass and energy balances. The analysis tools focused in this dissertation are elementary mode analysis (EMA), metabolic control analysis (MCA), and thermodynamic analysis. The structure of metabolic networks identified by EMA provides information about the cellular functionality and capacity. The control of the metabolic network provided by MCA can be used to identify rate-limiting reaction. Thermodynamic analysis of pathways permits a better understanding of the evolution of metabolic networks. Information from these analysis tools has a significant implication for strain improvement based on metabolic engineering. The pathway analysis information can also be incorporated with omics data such as transcriptomic, proteomic and metabolomic to elucidate regulations and interactions within the cellular metabolism.

Chapter 4

Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid[†]

4.1 Chapter summary

The objective of this chapter is to demonstrate the application of EMA for rational strain development. The chapter includes the development and analysis of the recombinant metabolic network model of carotenoid producing *Escherichia coli*. EMA is applied to predict the cell response to gene knockout perturbations and to systematically design of a mutant strain for an enhanced synthesis of the carotenoid diapolycopendioic acid (DPA). Based on the metabolic network analysis, all inefficient carotenoid biosynthesis pathways were eliminated in a strain containing a combination of eight gene deletions. To validate the model prediction, recombinant DNA technology was applied for construction of the designed strain. The performance of the developed strain was evaluated. The designed mutant produced DPA at increased yield and rate in comparison to the wild-type. The consistency between model prediction and experimental results demonstrate that EMA is useful as a guiding tool for the rational strain design for efficient production of desired product. The chapter also provides an approach to design and optimize a fed-batch fermentation of the developed strain to achieve a high titer of the DPA product.

[†]Parts of this chapter are from: Unrean, P., Trinh, CT., Srienc, F. 2009. Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. Metab Eng.

4.2 Introduction

Construction of desirable, better producing strains can be achieved through introduction of pathway modifications that direct the metabolic flux towards the desired products. However, the identifications of the targets of gene manipulations resulting in a specific cellular metabolism to function in a favorable direction is challenging due to the complexity of the interconnected cell reaction network. To address this question rigorously, a quantitative understanding of the cellular metabolic network is needed. In this chapter, we describe an application of elementary mode analysis (EMA) for characterizing the metabolism of a cell and for guiding the rational design of metabolic engineering manipulations.

EMA has been proven to be a powerful tool in metabolic engineering as described in Chapter 3. Previous examples of EMA applications focus on the intermediary metabolism of a cell to optimize cell growth and the formation of primary metabolites such as ethanol (Carlson 2004; Trinh, et al. 2008). This chapter further expands the application of EMA to optimize a recombinant cell containing a complex pathway for a secondary metabolite. To demonstrate the use of EMA in such system with a secondary metabolite-producing pathway, we have applied this method to the metabolic network of transgenic carotenoid expressing *E. coli*.

Carotenoid has received considerable attention in pharmaceuticals and food industry because of their properties as antioxidant, UV-protective compound and colorant (Sandmann, et al. 1999). There are more than 600 carotenoids that have been characterized. Most of them exist as biosynthetic intermediates. Therefore, they occur

only in trace amounts in natural sources and are difficult to extract in sufficient amounts to be useful for applications. Through the help of recombinant DNA technology, *E. coli* has been successfully engineered for expressing various carotenoids including zeaxanthin, lycopene and β-carotene (Albrecht, et al. 1997; Ruther, et al. 1997; Verwaal, et al. 2007). However, carotenoid production levels in recombinant *E. coli* cell are still not high enough to meet industry demand.

Previous attempts of improving carotenoid production in recombinant *E. coli* mostly relied on a non-systematic approach of genetic manipulation. This included over-expression of genes of the carotenoid biosynthesis pathway or deletion of competing central metabolic genes to redirect precursors towards enhanced production of carotenoid (Albrecht, et al. 1999; Wang, et al. 1998, Farmer 2001). However, this strategy is inefficient due to the lack of direction and an extensive amount of time and cost required. In addition, this approach may not capture all genetic modifications needed for an efficient carotenoid formation.

In this chapter, we applied EMA to rationally design an *E. coli* mutant strain for an enhanced production of the carotenoid diapolycopendioic acid (DPA). EMA was systematically used to examine gene deletion effects on DPA production of *E. coli* as well as to identify a set of multiple gene knockouts which results in a higher product yield. The rationally designed mutant was experimentally constructed and the kinetic of DPA production was determined to verify the model prediction. In addition, the developed strain was cultured in an optimized, controlled fed-batch fermentation using growth rate control loop by variation of glucose feeding and dissolved oxygen control

loop by variation of agitation speed which allowed the culture to reach high titer of DPA product.

4.3 Results

4.3.1 Identification of target gene knockouts

We first constructed a DPA expressing *E. coli* metabolic model by adding the biosynthesis pathway of carotenoid DPA into the central metabolic network of *E. coli* MG1655 which was used as the host for synthesizing DPA. The metabolic network of DPA producing *E. coli* considered in this study is based on the network given in Fig 1.1. Elementary mode analysis on the network revealed a total of 29,532 elementary modes. A total of 24,155 modes is aerobic, while 5,377 modes exist under anaerobic conditions. There are 5,923 total modes in which production of biomass and DPA are coupled. The large number of elementary modes illustrates the flexibility and robustness of the cell to adapt itself to particular conditions by using pathways that provides the optimal fitness. Comparison of all elementary modes revealed a maximum DPA yield of 0.83 carbon mole of DPA per carbon mole of glucose. The maximum possible biomass yield was 0.84 carbon mole of biomass per carbon mole of glucose, consistent with previous result (Carlson, et al. 2004).

The knowledge of all pathway options allowed the evaluation of gene knockout effects on cell phenotype. Gene knockouts were simulated by removing the enzymatic reaction corresponding to that gene from the stoichiometric matrix. The phenotype of that specific knockout mutant was then represented by a combination of remaining elementary modes when that reaction was deleted. Specifically, we focused on gene

knockout effects on cell viability (biomass yield), maximal yield of DPA and fraction of remaining modes after each single gene knockout in which a specific gene and its corresponding reaction was eliminated from the metabolic network. The effects of the single gene deletions (Fig. 4.1) were sorted in increasing orders of the fraction of remaining modes. Gene knockouts which result in zero yield of biomass are considered lethal and therefore, these genes cannot be deleted from the cell. The result in Fig. 4.1 assists the process of screening and selecting potential gene knockout targets. Among the set of all individual knockouts, the target genes are identified when elimination of that gene still maintains the maximum possible yield of DPA and retains a reasonable yield of biomass while the largest possible number of elementary modes is eliminated (minimum number of fractional remaining modes).

For example, zwf is a target for deletion since deletion of this gene eliminates more than 60% of identified elementary modes and still supports high yield of DPA and biomass. This approach allows the determination of optimal gene knockout targets which eliminate the largest number of elementary modes without affecting the most efficient pathway for DPA and biomass synthesis. We applied this identification tool in a sequential manner for selecting an optimal combination of multiple gene knockouts as -

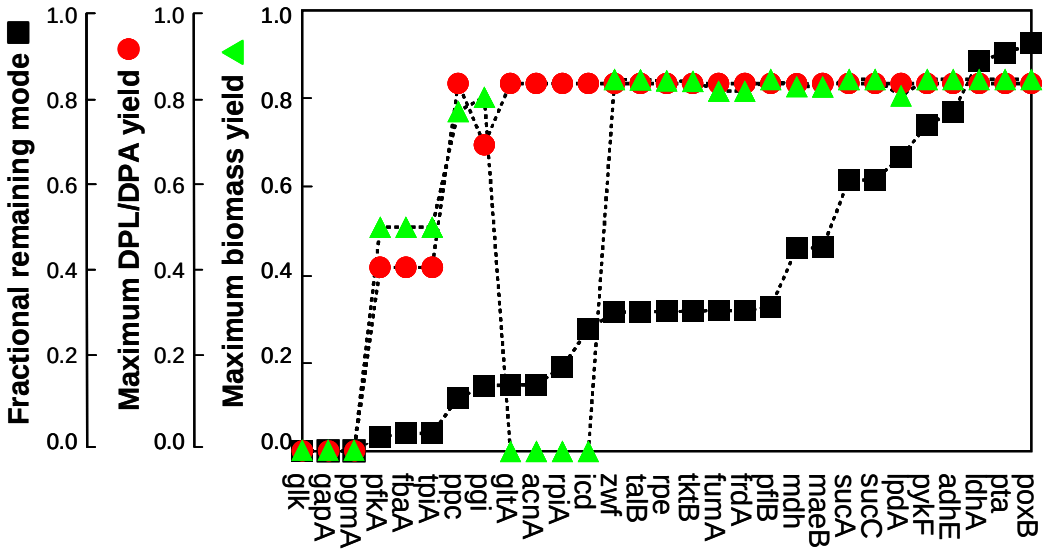


Figure 4.1 Effect of single deleted genes on the synthesis of carotenoid DPA, biomass formation and the fraction of remaining modes. Potential knockout targets are genes whose deletion still support maximum yield of DPA and biomass formation while minimizing the fraction of remaining modes.

described in Chapter 3. The combination of these gene knockout targets, therefore, forces the cell to function according to the remaining efficient DPA-producing elementary modes. Table 4.1 summarizes the number of remaining elementary modes when each gene knockout target was sequentially combined. The reduced number of available elementary modes after sequential gene deletion step restricts the cell's pathway options, hence forcing the cell to operate according to efficient pathways.

Table 4.1 Total elementary modes remaining after sequential deletion of multiple genes. Modes are categorized as aerobic modes (oxygen consuming modes) and anaerobic modes which do not use oxygen. Yield is in mg-diapolycopendioic acid/g-glucose.

Strain	Total modes	Aerobic modes	Anaerobic modes	Predicted Yield[1]
Wild-type	29,532	24,155	5,377	0.0-426
ΔldhAΔfrdAΔpoxBΔpta ΔadhE ΔpykFΔzwfΔmaeB	5	5	0	0.4-426

The identification approach resulted in a combination of eight gene knockouts which predicted a likely over-production of DPA in *E. coli*. The gene deletions included the removal of byproduct synthesis genes for lactate, succinate, acetate and ethanol which were experimentally accomplished by disrupting lactate dehydrogenase (ldhA; reaction R32), fumarate reductase (frdA; reaction R22), pyruvate oxidase (poxB; reaction R31), phosphate acetyltransferase (pta; reaction R35) and alcohol dehydrogenase (adhE; reaction R34) respectively. Other target genes were pyruvate kinase (pykF; reaction R9), glucose-6-phosphate-1-dehydrogenase (zwf; reaction R11) and malate dehydrogenase (maeB; reaction R28). The reaction designation of these eight knockout targets is summarized in Table 4.2. A mutant with the combination of these gene knockouts, named CRT028/pACMNOx, can operate based on the five remaining elementary modes (Fig. 4.2 and 4.3).

Table 4.2 Gene knockout targets for improving the production of diapolycopendioic acid in recombinant *E. coli*

Deleted Gene	Enzyme	Pathway
pykF	Pyruvate kinase	Glycolysis
Zwf	Glucose-6-phosphate-1-dehydrogenase	Pentose phosphate
frdA	Fumarate reductase	Fermentation
maeB	Malate dehydrogenase	Anapleurotic
poxB	Pyruvate oxidase	Fermentation
ldhA	Lactate dehydrogenase	Fermentation
adhE	Alcohol dehydrogenase	Fermentation
Pta	Phosphate acetyltransferase	Fermentation

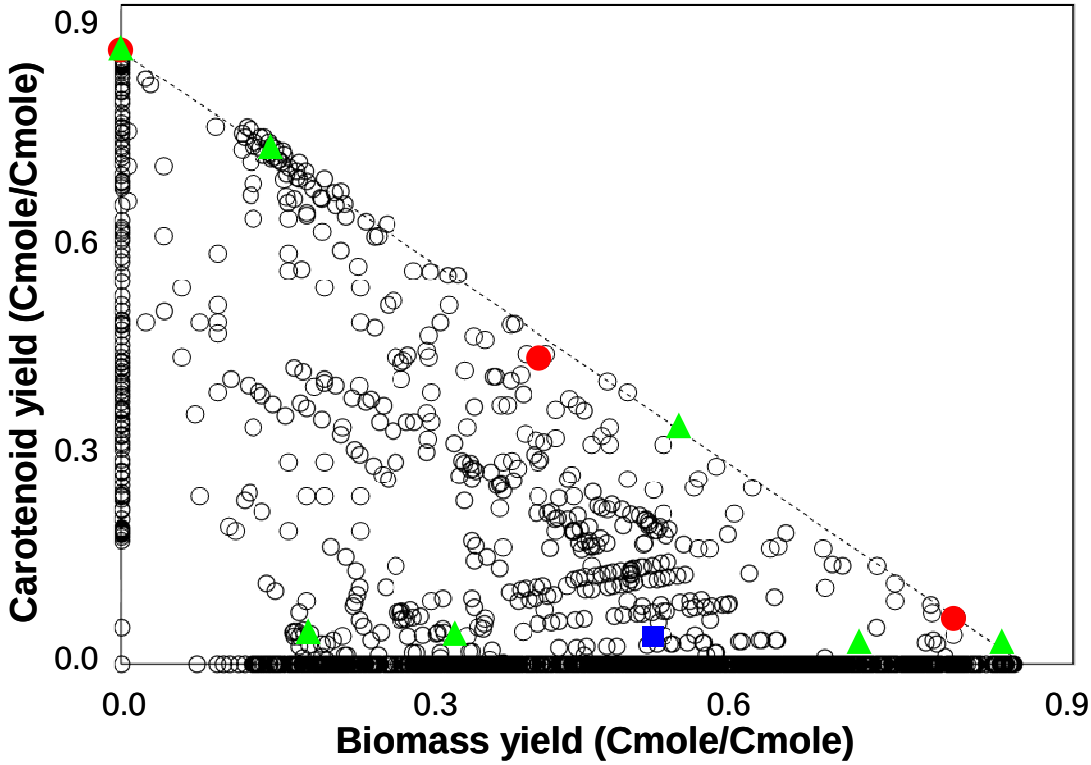


Figure 4.2 Elementary modes of CRT028/pACMNOx (● based on 4 NADP/carotenoid DPA; ▲ based on 8 NADP/carotenoid DPA) as compared with total available elementary modes in the wild-type MG1655/pACMNOx (○). Experimental carotenoid yield of CRT028/pACMNOx is shown in ■. Dashed line connecting the highest-yielding carotenoid mode with the highest-yielding biomass mode represents the yield of linear combinations of the two pathways. The negative slope indicates an inverse relationship between the two products formation.

The overall reaction stoichiometry and detailed pathways of these elementary modes are shown in Table 4.3 and Fig. 4.3 respectively. All of these remaining modes include DPA synthesis. Therefore, the knockout mutant is expected to produce DPA more efficiently than the wild-type which contains many modes that do not produce the product at all. In two of the remaining modes, DPA synthesis is coupled with biomass formation.

Table 4.3 Overall reaction stoichiometry and product yields of each elementary mode remaining in CRT028/pACMNOx

EMs	Overall reaction stoichiometry	Y_P[1]	Y_X[1]
1	Glucose + 0.14 O_2 → 0.17 DPA + 1 CO_2	426	0
2	Glucose + 0.14 O_2 → 0.17 DPA + 1 CO_2	426	0
3	Glucose + 0.14 O_2 → 0.17 DPA + 1 CO_2	426	0
4	Glucose + 0.11 O_2 + 0.53 NH_3 → 0.11 DPA + 0.91 CO_2 + 0.69 x 10^{-3} Biomass	270	0.26
5	Glucose + 0.22 O_2 + 1.30 NH_3 → 1.56 x 10^{-4} DPA + 1 CO_2 + 1.73 x 10^{-3} Biomass	0.41	0.64

[1] Y_P is carotenoid DPA yield (mg-DPA/g-glucose) while biomass yield (g-biomass/g-glucose) is shown in Y_X.

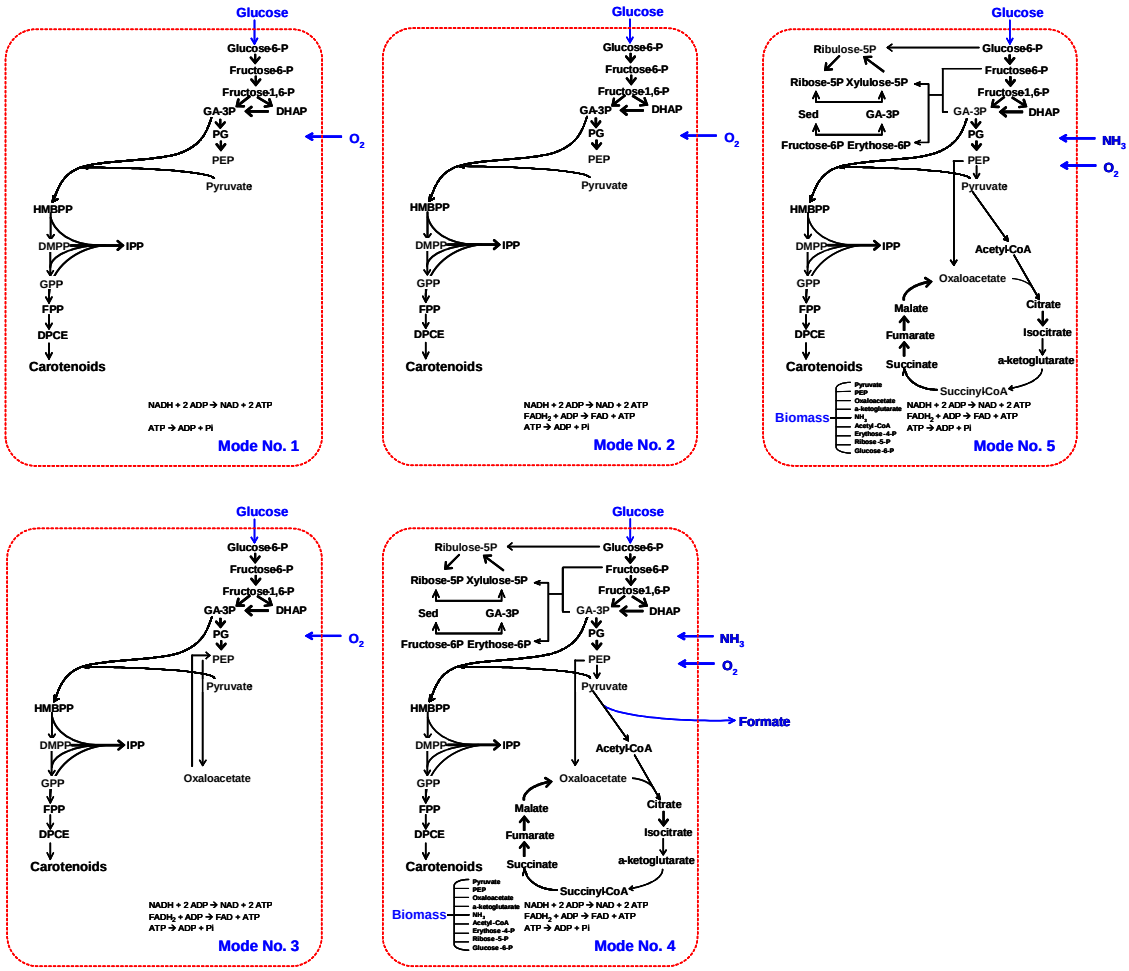


Figure 4.3 Available elementary modes in CRT028/pACMNOx

4.3.2 Effect of NADP on production of diapolycopendioic acid

Since a detailed biosynthesis pathway of diapolycopendioic acid (DPA) has not yet been established, it is unclear how many NADP cofactors are required for the synthesis of one mole of DPA. Therefore, we evaluated the effect of the number of NADP cofactors on the number of available modes and predicted yields of DPA in the host background of CRT028/pACMNOx (Fig. 4.4) that was constructed with the assumption that 4 NADP are required per mole of DPA formed. This number was selected based on

the assumption that for every atom of hydrogen lost in the oxygenation step, one mole of NADP cofactor is required. EMA showed that the change of the number of NADP has no effect on the predicted maximum yield of DPA. However, an increased number of NADP cofactors required for synthesis of one mole of DPA results in an increased number of available elementary modes and a decreased minimum predicted yield of DPA. Interestingly, all elementary modes available in CRT028/pACMNOx are all directed to the production of DPA regardless of the number of NADP chosen. This confirmed that for all numbers of NADP assumed, CRT028/pACMNOx is expected to produce DPA efficiently.

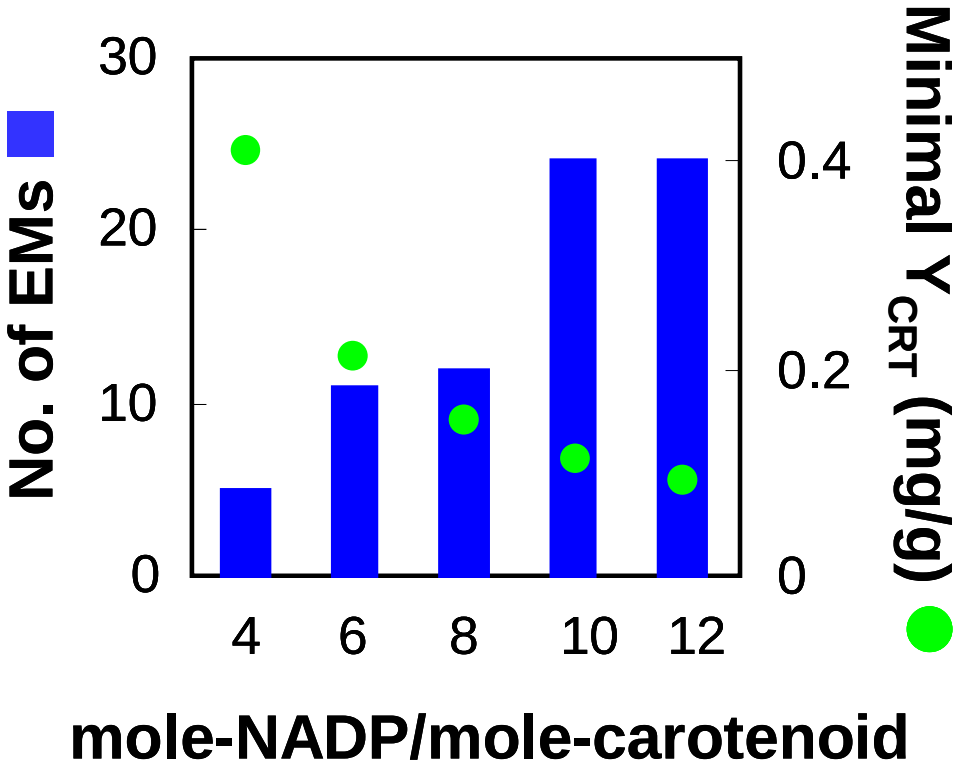


Figure 4.4 Effect of the number of NADP cofactors on the number of available elementary modes and minimum predicted yield of DPA in CRT028/pACMNOx

4.3.3 Quantification of diapolycopendial and diapolycopendioic acid

Three carotenogenic genes crtM, crtN and crtOx were responsible for the synthesis of the end product, diapolycopendioic acid. A previous study has shown that *E. coli* expressing plasmid pACMNOx containing these genes stored the end product as well as its intermediates, diapolycopendial in the cell (Lee et al., 2004). These two products have distinct solubilities. Diapolycopendial is soluble in acetone while diapolycopendioic acid is soluble in an aqueous KOH solution. As a result, we developed two extraction processes to extract each of them from the cell membrane. We first extracted diapolycopendial from the cells using acetone. A 10% KOH solution was then used for the second extraction process to isolate the carotenoid diapolycopendioic acid. Absorption spectra of acetone and KOH extracts using a spectrophotometer are shown in Fig. 4.5. The distinct absorption spectra of the extract solutions confirmed that different carotenoids were being dissolved in each solution. In acetone supernatant, the maximum absorbance was found at a wavelength 506 nm, while maximum absorbance in KOH supernatant was at a wavelength of 490 nm. These observed wavelengths agree with the wavelength reported by in Tao et al. (2005), confirming the presence of carotenoid, diapolycopendial in acetone extract (reported λmax, 507nm) and diapolycopendioic acid in KOH extract (reported λmax, 490nm). Since both carotenoids, DPL and DPA, were found in the cells, carotenoid product measured in later experiments will be given as the sum of both carotenoids.

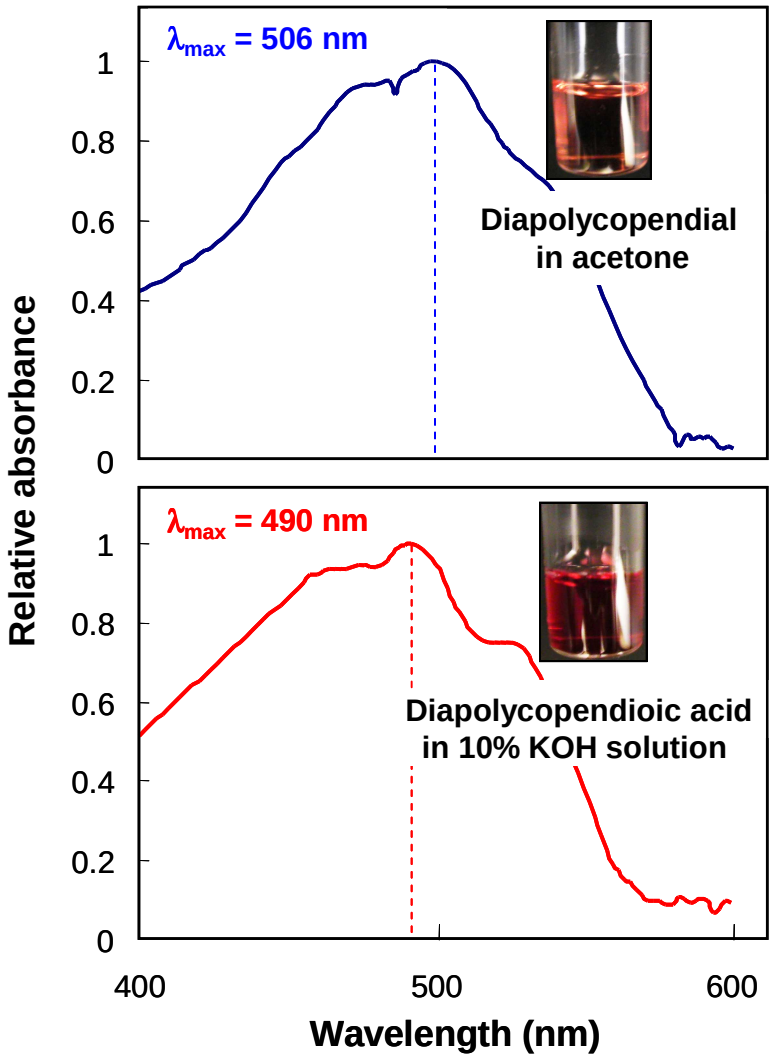


Figure 4.5 Absorption spectra of extract solutions using acetone (top) and 10% KOH (bottom) from *E. coli* cells expressing plasmid pACMNOx. The maximum absorbance was at 506 nm in the acetone extract and at 490 nm in the KOH extract. The absorbance maxima in both extracts are in agreement with Tao et al. (2005) for diapolycopendial in acetone and for diapolycopendioic acid in 10% KOH solution, respectively.

4.3.4 Strain construction

To validate the knockout strain design based on EMA, the designed mutant CRT028, was constructed using established methods (Ausubel, et al. 1995) and verified by PCR

amplification and gel electrophoresis using both inside and outside primers specific to deleted genes. Inside primers specified to the deleted part of the gene were used to ensure that a disrupted gene was not accidentally displaced somewhere else on the chromosome. Outside primers specified to undeleted portions of the gene were used to verify gene disruption at the known locations on the *E. coli* chromosome. Gel pictures (Fig. 4.6) confirmed of the gene disruptions in CRT028 as compared with the wild-type MG1655 containing undisrupted genes. The smaller size PCR product using outside primers and the absence of gene product amplified using inside primers confirmed the deletion of eight genes from the chromosome of the mutant.

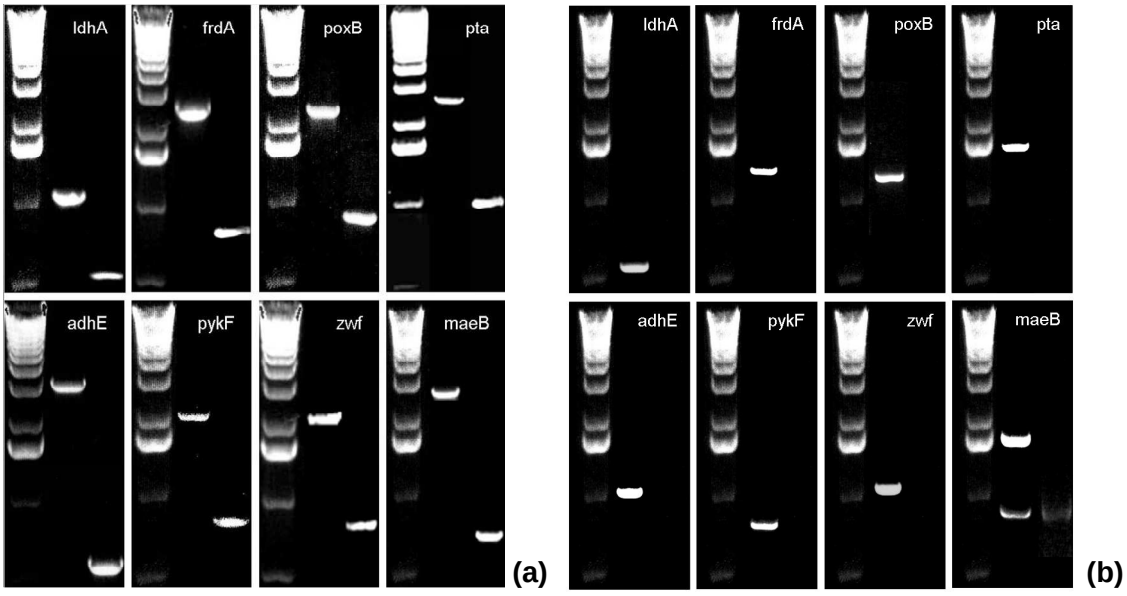


Figure 4.6 Confirmation of gene deletions in mutant CRT028 by PCR and gel electrophoresis using outside (a) and inside (b) primers. For each gene, the first lane is a 1 kbp DNA ladder, the second is the PCR amplification product in the wild-type MG1655 and the third lane is the PCR amplification product in the mutant CRT028.

4.3.5 Strain characterization

According to the strain design by EMA, two of the remaining modes in the mutant are growth-associated DPA producing modes (Table 4.3) which suggests a tight coupling between DPA synthesis and cell growth in the mutant. Therefore, we tested whether or not cell growth and DPA production are indeed coupled in the mutant strain. Growth experiments in aerobic shake flasks showed that the mutant CRT028 grew very slowly without the plasmid pACMNOx and growth was restored when the plasmid was introduced into the mutant (Fig 4.7). The growth rate of the mutant with plasmid was approximately 4-folds faster than that of the mutant without plasmid. Thus, the formation of DPA could provide a strong selection pressure for the plasmid even in the absence of antibiotics.

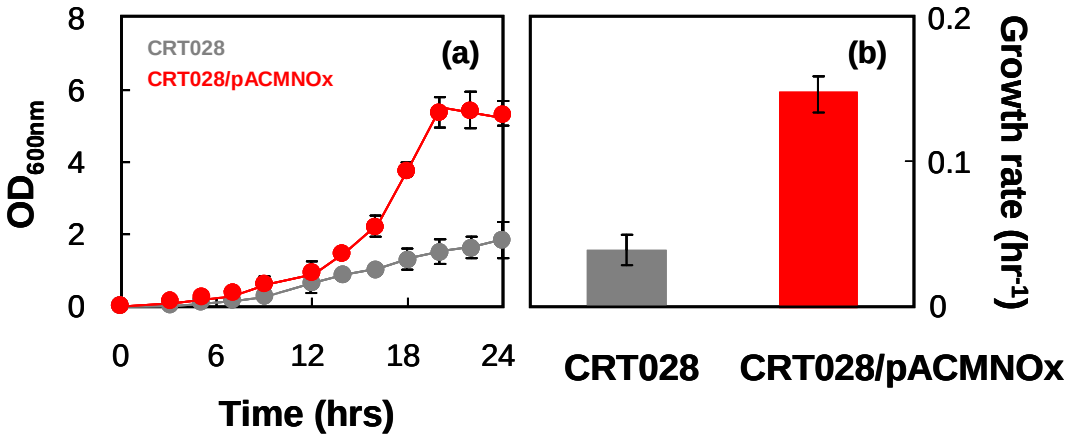


Figure 4.7 Growth characteristics of CRT028 and CRT028/pACMNOx (a) Time-profile of cell growth (b) Specific growth rates

A previous study by Lee, et al. 2004 has shown that *E. coli* expressing plasmid pACMNOx stored the end product DPA as well as its intermediate, diapolycopendial (DPL) in the cell. Thus, the carotenoid product measured in the cell is given as the sum of

both carotenoids. Carotenoid yield (mg of carotenoid produced per g of glucose consumed) of both *E. coli* wild-type MG1655/pACMNOx and mutant strain CRT028/pACMNOx were measured. Fig. 4.8a showed the yield of carotenoid in wild-type and in the constructed knockout strain determined in aerobic shake flask experiments. Carotenoid production in the mutant CRT028/pACMNOx is significantly higher than in the wild-type MG1655/pACMNOx. In aerobic batch bioreactors, the mutant yielded 0.17±0.04 mg-carotenoid/g-glucose compared to 0.04±0.00 mg-carotenoid/g-glucose by the wild-type (Fig. 4.8c). The production profile of total carotenoid in batch bioreactors shown in Fig. 4.8b also revealed that the mutant produced carotenoid at faster production rate than the wild-type. In addition, comparison of the growth curves of MG1655/pACMNOx and CRT028/pACMNOx suggested that the mutant grew slower than the wild-type with the specific cell growth rate reduced by 23%. Growth and carotenoid phenotype of both strains are summarized in Table 4.4. The results from the online off-gas mass spectrometer also revealed in the case of the mutant a decrease in carbon dioxide production rate but a significantly increased oxygen uptake rate as compared to the wild-type.

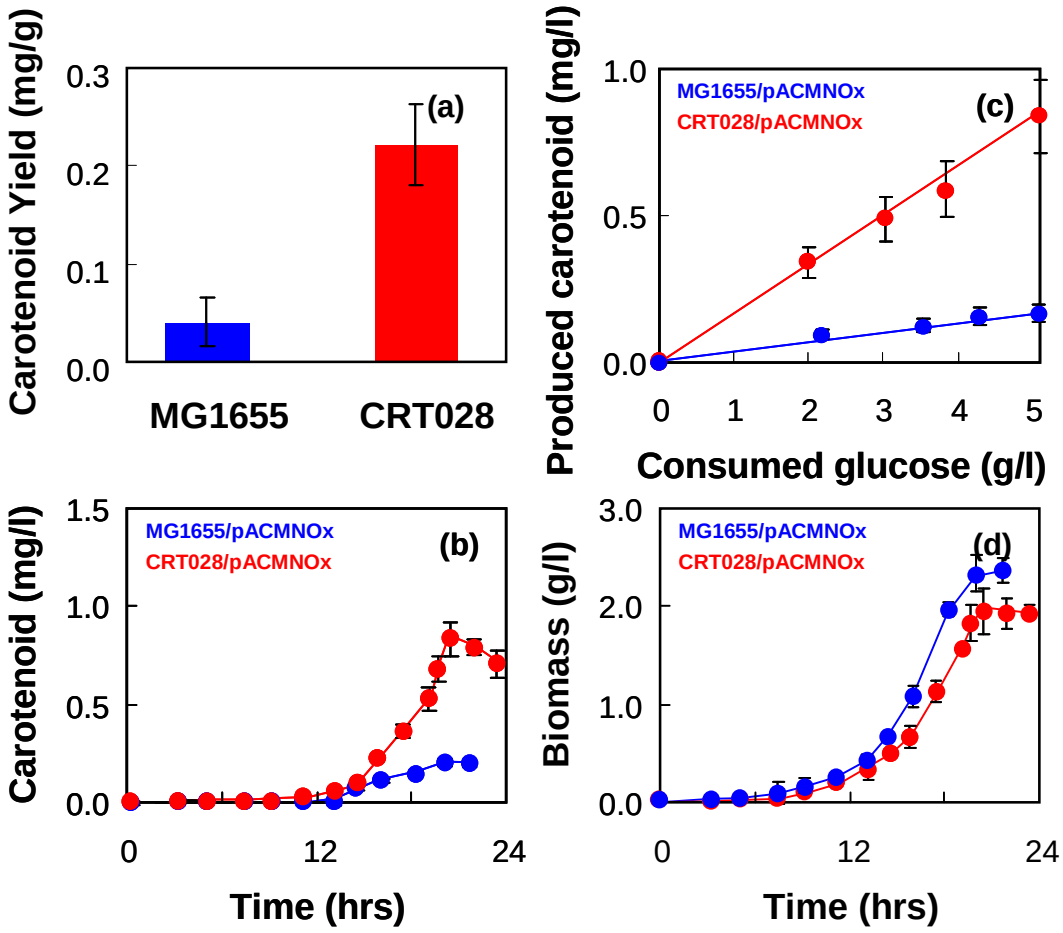


Figure 4.8 Yield of carotenoid of wild-type MG1655/pACMNOx (blue) and mutant CRT028/pACMNOx (red) in aerobic batch shake flasks (a); carotenoid production profile in aerobic batch bioreactors (b); produced carotenoid vs. consumed glucose (c); and time profiles of biomass production (d).

Table 4.4 Comparison of wild-type MG1655/pACMNOx and mutant CRT028/pACMNOx synthesis of carotenoid in aerobic batch bioreactors

Phenotypes	MG1655 /pACMNOx	CRT028 /pACMNOx
Growth rate hr-1	0.17 ± 0.02	0.13 ± 0.01
Carotenoid production mg/l	0.19 ± 0.02	0.83 ± 0.20
Carotenoid yield mg carotenoid/g glucose	0.04 ± 0.00	0.17 ± 0.04

4.3.6 Strain comparison

Several *E. coli* strains have been previously developed to improve production of carotenoid through multiple gene modifications. To compare the differences between our designed strain CRT028/pACMNOx and other mutant *E. coli* strains developed previously, we have applied elementary mode analysis to evaluate the effects of the gene knockouts on the production of DPA and on the reduction of inefficient pathways for DPA production (Table 4.5). The results show that the mutants containing gene knockouts previously identified for enhanced production of carotenoid still contain a significant number of inefficient DPA-producing pathways. These remaining inefficient pathways likely reduce the yield of DPA if the pathways are used. Unlike each of these mutants, CRT028/pACMNOx contains only DPA-producing elementary modes. Therefore, our designed strain is expected to be able to efficiently produce carotenoid diapolycopendioic acid.

Table 4.5 Elementary mode analysis of *E. coli* knockout strains using glucose as a carbon source for production of diapolycopendioic acid under aerobic conditions. DPA is carotenoid diapolycopendioic acid. Yield is in mg-carotenoid/g-glucose.

Knockout/ pACMNOx	Total EMs	DPA yield of total EMs	DPA-prod. EMs	DPA yield of DPA-prod. EMs	References
ΔpykFΔpykA	4,275	0.00-426	929	<0.01-426	Farmer and Liao 2001
ΔaceEΔfdhF	1,049	0.00-426	242	0.10-426	Alper et al 2005
ΔaceEΔtalB	952	0.00-426	352	0.02-426	Alper et al 2005
ΔaceEΔtalBΔfdhF	303	0.00-426	114	0.10-426	Alper et al 2005
ΔldhAΔfrdAΔadhEΔpoxB ΔptaΔpykFΔzwfΔmaeB	5	0.41-426	5	0.41-426	This study

4.3.7 Design of fed-batch fermentation

Like most secondary metabolites, DPA is known as a growth-associated product. Thus, to reach a high product level, a culture of high cell density must be achieved. Fed-batch process is preferred over a batch when growing cells to a high density because of the inhibitory effect of the high initial concentration of substrates toward cell growth (Riesenberg 1991). In this study, the engineered strain CRT028/pACMNOx was applied for the production of DPA in a controlled fed-batch culture.

A controlled fermentation strategy was designed to achieve both high cell density and high product expression of DPA. A previous study by Alper, et al. 2006 has shown that oxygen is critical for carotenoid synthesis likely due to the large energy requirement for carotenoid synthesis. The same study also showed a higher expression of carotenoid product occurs at a reduced growth rate as a result of the trade-off relationship between

the two products. Therefore, the process strategy was aimed at maintaining an adequate supply of oxygen as well as permitting *E. coli* cultivation at a reduced growth rate. To achieve such a goal, we applied a feedback loop control of dissolved oxygen (DO%) with variable agitation speed ensuring sufficient oxygen supply in the fed-batch culture. The DO%(stir)-control loop was to increase the dissolved oxygen concentration in the culture by increasing of agitation rate. The specific growth rate of cell culture was controlled via an exponential feeding rate of glucose corresponding to the fed-batch mathematical model Eq (2.3).

4.3.8 Production of DPA in an O_2-controlled, glucose fed-batch fermentation

A controlled fed-batch fermentation allowed a high cell-DPA concentration to be obtained at a reduced growth rate. The fed-batch fermentation was also controlled to ensure no oxygen supply limitation. We first applied Eq (2.3) to determine the exponential addition rate of glucose feedstock for maintaining a constant growth rate. A low specific growth rate of 0.1 hr^{-1} was selected in this experiment. Other parameters including cell yield, DPA yield and the specific rate used for estimation of feeding rate were obtained from batch culture. Specific maintenance coefficient (m) for *E. coli* was 0.04 g-glucose/g-cell-hr as reported by Andersson (1994). The calculated feeding profile ($M_S(t)$) was entered into the reactor controller which connected to feed pump for controlling glucose feeding rate. To guarantee the sufficient supply of oxygen demanded of the cell, we applied the feedback control system to control the amount of DO% in the culture at 20% by varying agitation speed. Besides the two controlled loops, the pH was partially controlled at 7.0 using NH_4OH which can also be used as a nitrogen source.

A typical time course of the cultivation of CRT028/pACMNOx under application of the described control strategies was shown in Fig. 4.9. During the batch process, the culture grew exponentially after a short lag-phase, whereas the DO% decreased due to an increasing oxygen demand of the growing culture. The process was switched into a fed-batch mode after a sudden rise of DO% which indicated a depletion of glucose. The total consumption of the glucose initially added marked the end of the batch phase with a generated biomass of about 1.7g /l (OD_{600} of approximately 6.0). After the exhaustion of glucose, the fed-batch phase was started by an exponential glucose feeding rate to keep

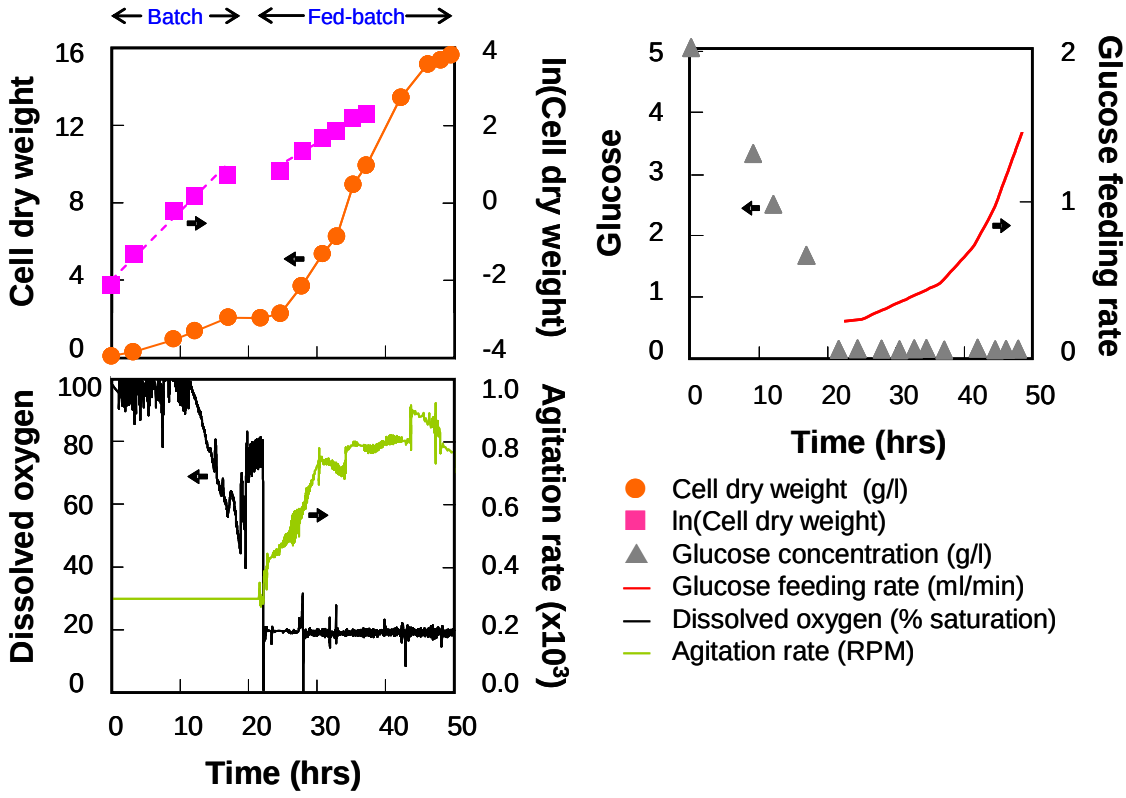


Figure 4.9 Cell growth, glucose concentration, glucose feeding rate, agitation rate and dissolved oxygen in fed-batch fermentation of CRT028/pACMNOx. Cell growth was maintained at a constant rate by a control of glucose feeding. Dissolved oxygen was kept constant through a feedback loop control of agitation rate.

a constant cell growth rate of 0.1 hr^{-1}. The percent dissolved oxygen (DO%) was controlled at 20% saturation to ensure sufficient supply of oxygen for the cell culture. During the whole course of fed-batch culture, the agitation rate was continuously increased as a result of incrasing oxygen consumption of the culture. Throughout the fed-batch process, very little of the glucose was accumulated which suggested a balance between the glucose feeding rate and the glucose consumption rate of the culture.

Time profiles of DPA (mg/l) produced and the product yield obtained during batch and fed-batch process was present in Fig. 4.10. With the optimized feeding strategy based on the fed-batch model, a final cell dry weight of 16.03 g/l (equivalent to OD_{600nm} of about 52.62) and a final DPA concentration of 8.53 mg/l were reached. Yield of DPA during the batch and fed-batch process were 0.38 and 0.53 mg DPA/g-cell dry weight respectively. The higher DPA product yield observed in fed-batch than in the batch process suggested that the glucose feeding control at reduced specific growth rate is appropriate for an optimal synthesis of DPA product. Approximately 30 mg of DPA product was obtained at the end of the culture. High titer of DPA achieved in this study indicated the efficiency of a fed-batch strategy with growth rate control loop and oxygen supply control loop for the production of DPA. The growth characteristics and DPA production phenotype were summarized in Table 4.6.

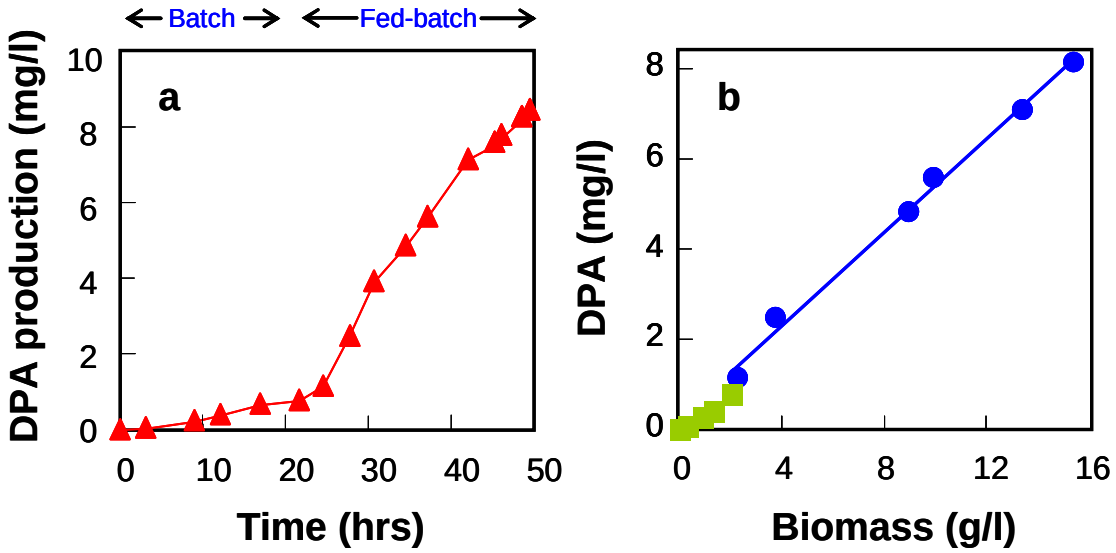


Figure 4.10 Time profiles of DPA production (a) and DPA yield (b) in a controlled fed-batch reactor of CRT028/pACMNOx under aerobic condition with glucose as a feed substrate (see section 2.5.3.3 and 2.5.3.4)

Table 4.6 Summary of cell growth and production of diapolycopendioic acid in a controlled fed-batch culture by CRT028/pACMNOx

Growth and DPA phenotype	CRT028/pACMNOx
Maximum OD_{600}	52.62 ± 2.78
Biomass concentration g-cell/l	16.03 ± 0.28
Total DPA production mg/l	8.53 ± 0.14
DPA yield mg/g-cell	0.53 ± 0.01

4.4 Discussion

Because carotenoid is of considerable industrial and nutritional value, several groups have recently metabolically engineered recombinant cells for enhancing production of carotenoid. The strategies applied, however, were based on overexpression and/or deletion of genes involved in the carotenoid biosynthesis pathway. Although significant

improvement in carotenoid production was observed using this approach, the genetic manipulations in general have not been rationally suggested. Here, we have presented a systematic strategy by using elementary mode analysis (EMA) to design an optimized *E. coli* strain for the most efficient synthesis of carotenoid. In particular, we applied EMA to identify targets for gene deletions resulting in an efficient production of carotenoid diapolycopendioic acid (DPA). The designed strain was obtained through a sequential implementation of eight multiple gene deletions which narrowed the possible pathway space to the efficient DPA producing options.

EMA decomposes a metabolic network into a set of unique and non-divisible pathways that represent all possible physiological states of the cells. Besides EMA, other constraint-based models such as OptKnock or MOMA which are based on the concept of flux balance analysis (see Chapter 3 for details) have been previously used as a guiding tool for metabolic engineering a cell (Alper et al., 2005). However, these models identify only one optimal solution based on a defined objective function. Unlike these constraint-based models, EMA identifies the complete set of possible optimal pathway solutions. As a result, the strain designed by EMA is guaranteed to function according to these optimal pathways. Based on elementary mode analysis, an inverse relationship between the fermentation product formation and DPA production due to shared precursors between these products can be observed. Therefore, removal of the identified fermentation product synthesis genes–ldhA, frdA, poxB, pta, adhE–is expected to direct more flux into the DPA synthesis pathway. A less intuitive target for disruption suggested by EMA is the pyruvate kinase; pykF. The pykF gene encodes for one of the two isoenzymes of

pyruvate kinase which permits the inter-conversion between pyruvate and phosphoenolpyruvate. Deletion of pykF could block the further metabolism of the DPA precursors, pyruvate and glycerol-3-phosphate leading to an increased glycolytic supply for the DPA synthesis pathway. Likewise, the disruption of glucose 6-phosphate-1-dehydrogenase encoded by zwf diverts the carbon flux into the pool of DPA precursors and reduces carbon loss in form of carbon dioxide usually mediated by the pentose phosphate pathway (Sprenger 1995). Deletion of malic enzymes maeB as suggested by EMA likely prevents DPA precursors from draining into the anaplerotic pathway.

Experimental results confirmed that the synthesis of DPA is essential for cell growth in the mutant CRT028 (Fig. 4.7). The coupling between DPA synthesis and cell growth can be used as a natural selection pressure for maintaining DPA synthesis plasmid in the cell. In the designed mutant, the production of DPA was improved as compared with the wild-type under identical growth condition. However, in the mutant, the specific growth rate was reduced. This is consistent with previous studies which showed a higher expression of carotenoid in *E. coli* at reduced growth rate (Alper, et al. 2006). The inverse relationship between production of DPA and biomass is also predicted by EMA. Moreover, the mutant required a significantly higher amount of oxygen in comparison to the wild-type which suggests a critical role of oxygen on DPA synthesis.

Based on the EMA prediction, mutant CRT028/pACMNOx can only function using a combination of the five remaining elementary modes with a predicted minimum DPA yield of 0.41 mg/g-glucose. However, experimental yield observed in the mutant was lower than the predicted value. It is likely that the inconsistency between observed yield

and predicted possible yield may be a result of an incorrect number of NADP cofactor chosen. As shown in Fig. 4.4, the predicted yield of DPA and numbers of available elementary modes depend on number of NADP cofactor assumed for synthesis of one mole of DPA. If eight moles of NADP required per mole of DPA was chosen in the model instead, the predicted minimum yield of DPA in the mutant CRT028/pACMNOx would be 0.15 mg/g-glucose which is consistent with the experimental yield observed.

In addition, there were other factors that could possibly prevent the mutant from reaching a higher yield of DPA. Rate limiting steps in the downstream pathway of DPA biosynthesis could be responsible for preventing the cell from reaching higher product yield. There were several studies showing that overexpression of these downstream genes could lead to an increased yield. For example, production of carotenoid such as lycopene, torulene or β-carotene were found to be improved when the isoprenoid biosynthesis genes such as dxs or idi were overexpressed in *E. coli* (Lee, et al. 2004). The product formation could also be inhibited by some unknown negative effects of the cell regulatory system. Future studies will be needed for investigating these factors to further improve the product yield and productivity. At this point, the improved performance of the mutant in this experimental study has confirmed that elementary mode analysis (EMA) is an efficient tool for rational strain development to enhance secondary metabolite production such as carotenoid.

Chapter 5

Engineering of an efficient lycopene-producing *E. coli* using inverse metabolic engineering and metabolic evolution

5.1 Chapter summary

This chapter demonstrates the use of inverse metabolic engineering and metabolic evolution for the development of efficient strains. The two approaches are applied to develop a mutant *E. coli* strain for an efficient production of lycopene. Inverse metabolic engineering guided by elementary mode analysis (EMA) was employed to predict a set of multiple-gene deletions required for the cell to function according to an efficient lycopene synthesis pathway. By implementing eight gene knockout mutations, the functional space of the cell was reduced to a total of seven pathways that support cell function for growth and lycopene synthesis. Metabolic evolution of the designed strain was applied for co-selection of cell growth and increased lycopene production. The evolved mutant LYC018E1/pACEBI showed improvements in both yield and productivity of lycopene, verifying the applied strategies for efficient strain development. The performance of the mutants also lies within the predicted phenotypic space of available elementary modes which confirms the accuracy of the network model.

5.2 Introduction

Lycopene (LYC) has received considerable attention because of its property as a pigment and, more importantly, its beneficial effect on human health. LYC has been found to prevent some cancers and other degenerative human conditions such as heart disease and immune deficiency (Vershinin, 1999). Owing to its great commercial value,

the *in vivo* synthesis of lycopene has become increasingly important. LYC can be recombinantly produced in *E. coli* through the non-mevalonate pathway using pyruvate and glycerol-3-phosphate as precursors (Pfeiffer et al., 1999). The recombinant *E. coli* accumulates lycopene, but at a much lower level than the native lycopene producers. Previous attempts of engineering *E. coli* for an enhanced LYC formation have been made by overexpression and/or deletion of genes involved in lycopene biosynthesis pathways (Alper et al., 2005; Farmer and Liao, 2000; Farmer and Liao, 2001; Kim and Keasling, 2001). However, pathway optimization by the modification of these genes alone may not result in an optimal phenotype for LYC production. We present here an alternative strategy for strain optimization using the combined approaches of metabolic engineering guided by elementary mode analysis and metabolic evolution. The applied methodologies are demonstrated in engineering an *E. coli* cell for an efficient synthesis of lycopene from glucose under aerobic growth condition.

Elementary mode analysis (EMA) is a metabolic pathway analysis tool that can be applied for optimizing any aspect of cellular metabolism included in the model (Pfeiffer et al., 1999; Schuster et al., 2000). This approach is used to analyze the topology of a metabolic network and to determine all feasible pathway options in the network (see Chapter 3 for details). As demonstrated in Chapter 4, evaluation of all these pathway possibilities permits the rational implementation of only the efficient pathways of interest by removing the inefficient ones through gene knockout mutation, resulting in a cell with its functionality limited to only efficient pathways (Trinh et al., 2006; Trinh et al., 2008; Unrean et al., 2010).

In this chapter, we applied EMA to rationally design an *E. coli* cell with its metabolic functionality modified to efficiently convert glucose into lycopene under aerobic growth conditions. A metabolic evolution approach under a growth-based selection was then implemented for optimizing LYC synthesis in the developed mutant. The metabolic evolution strategy was utilized to capture other possible mutations that may not be identified by the metabolic pathway analysis for an enhanced lycopene synthesis. Like many secondary metabolites, LYC is known as a growth-associated product. The co-production between LYC and cell growth enables the application of metabolic evolution to improve LYC production. Because of this link, an evolved mutant with an improved cell growth would be expected to produce more LYC. The mutant obtained through the combined strategies was characterized and compared with the wild-type under aerobic growth conditions. The performance of the mutant was also compared with the predicted phenotypic range identified by EMA in order to verify the model prediction.

5.3 Results

5.3.1 Design of knockout mutant

EMA was used to rationally design an *E. coli* mutant strain capable of efficiently converting glucose into lycopene. The metabolism of lycopene-producing *E. coli* used in this analysis is given in the Appendix. The net reaction for lycopene synthesis through a non-mevalonate pathway in *E. coli* is (Fraser et al., 1992):

8 Glycerol-3-phosphate + 8 Pyruvate + 12 NADPH + 8 ATP + 4 FAD → 1 Lycopene + 8 CO_2 + 12 NADP + 8 ADP + 4 $FADH_2$.

The metabolic network model is supported by 5,252 elementary modes during growth on glucose under aerobic condition. The analysis also identified a total of 1,215 modes for the production of lycopene under aerobic conditions with a predicted maximum yield of 0.12 mole lycopene per mole glucose (Fig. 5.1). The effects of gene knockouts on the production of lycopene and cell viability based on all pathway options were evaluated. The target gene knockout was the gene which eliminated the largest number of modes without affecting cell viability and the theoretical maximum LYC yield.

The EMA predicted the number of elementary modes that would collapse into the seven remaining efficient LYC-producing elementary modes when a set of eight target genes were eliminated from the intermediary metabolism. The target genes for deletion were identified based on the search algorithm given in Chapter 3. A combination of the knockout mutations corresponding to deletion of pykF, tktB, frdA, pckA, pta, ldhA, adhE, and poxB resulted in a mutant LYC018/pACEBI (Table 5.1). The knockout mutant was predicted to function according to the remaining modes by yielding a high conversion of glucose into lycopene under aerobic growth conditions. Four of the modes remained produce lycopene and two of these coproduce lycopene and biomass. With eight gene knockout mutations, the minimum growth-associated lycopene yield is elevated to 0.06 mole lycopene/mole glucose (Fig. 5.1). Therefore, the mutant operates according to the efficient LYC-producing pathways that co-produced lycopene and cell growth at higher yield than the wild-type.

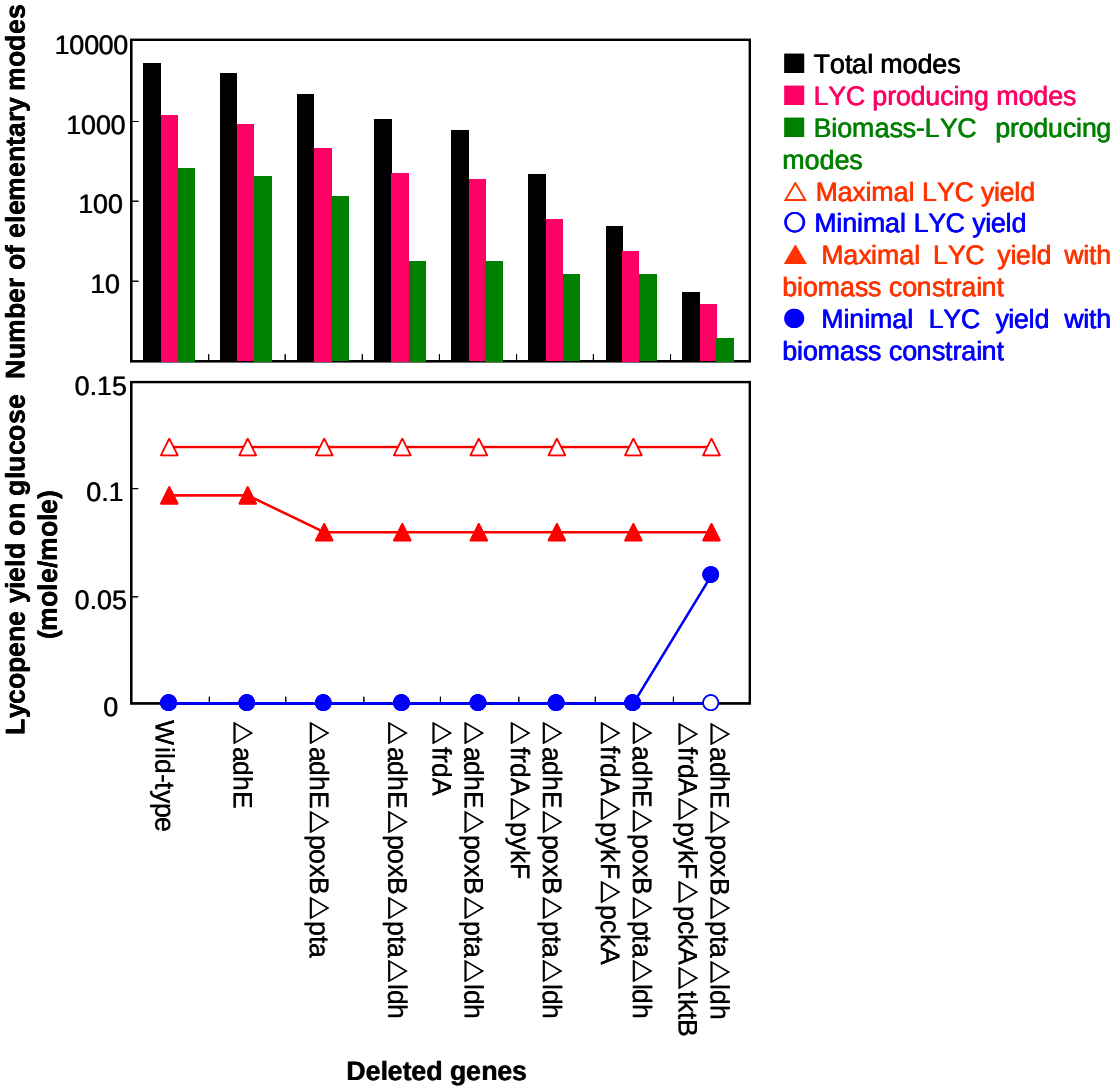


Figure 5.1 Effect of multiple gene knockouts on the number of elementary modes and the yield of lycopene on glucose under aerobic growth condition. Total modes, lycopene producing modes, lycopene-biomass coproducing modes, overall maximum yield of lycopene, overall minimum yield of lycopene, maximum yield of lycopene during growth, and minimum yield of lycopene during growth are shown.

Table 5.1 Gene knockout targets for improving the production of lycopene from glucose in recombinant *E. coli*

Deleted gene	Enzyme	Pathway
tktB	Transketolase	Pentose phosphate
pckA	Phosphoenolpyruvate carboxykinase	Anapleurotic
pykF	Pyruvate kinase	Glycolysis
frdA	Fumarate reductase	Fermentation
poxB	Pyruvate oxidase	Fermentation
ldhA	Lactate dehydrogenase	Fermentation
adhE	Alcohol dehydrogenase	Fermentation
pta	Phosphate acetyltransferase	Fermentation

The predicted phenotypic space, with the lycopene yield expressed as a function of the biomass yield of the designed mutant LYC018/pACEBI, was compared with the wild-type (Fig. 5.2). The phenotypic space is represented by a non-negative, linear combination of available modes. Unlike the wild-type which can operate anywhere in this functional space, the mutant was restricted to the shaded area constructed from the lines connecting the available modes of the mutant.

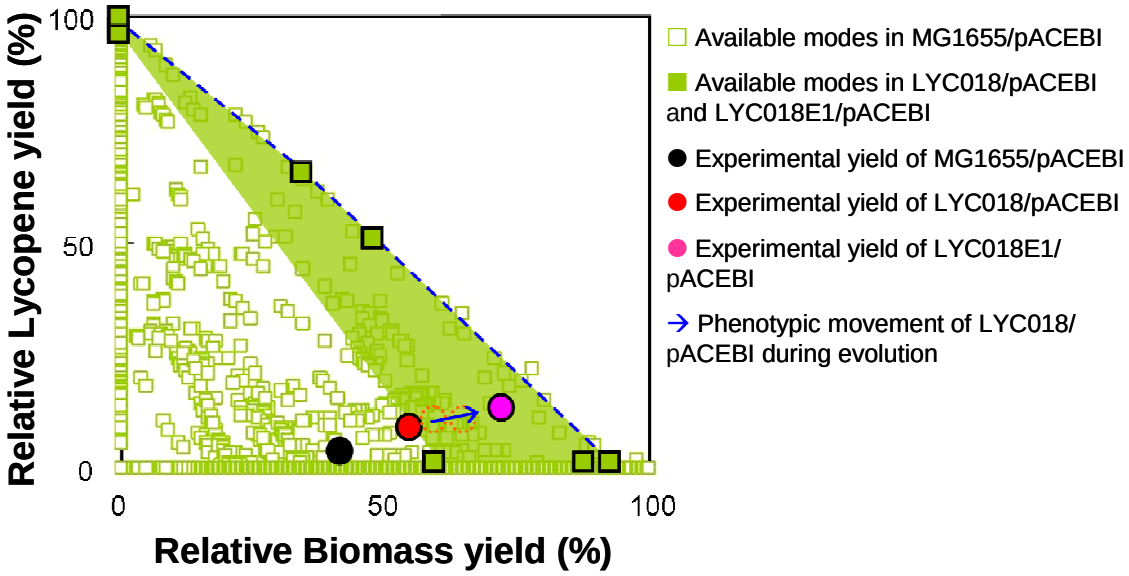


Figure 5.2 *In silico* and *in vivo* phenotype of MG1655/pACEBI, LYC018/pACEBI and LYC018E1/pACEBI. Predicted yields (square) for wild-type and mutant were obtained from elementary mode analysis of a LYC-producing *E. coli* metabolic network model. Experimental yields of MG1655/pACEBI (●), LYC018/pACEBI (●), and LYC018E1/pACEBI (●) obtained from aerobic shake flask batch culture containing NBS and 15 g/l glucose were compared with the predicted yields. The plot is presented in relative yield, which is a yield ratio to the maximum predicted yield based on EMA. The shaded area represents possible phenotypic space of the mutants through a linear combination of available elementary modes. Experimental observation reveals a movement of LYC018/pACEBI towards the metabolic state with higher yield of lycopene and biomass during metabolic evolution in a serial dilution experiment (blue arrow).

5.3.2 Construction of the designed mutant LYC018

The knockout mutant LYC018 that contains target gene deletions as designed by EMA was constructed with the established methods described in Chapter 2. Gene disruptions in the mutant LYC018 were verified using PCR amplification and gel electrophoresis with both inside and outside primers and using the wild-type as a positive control (Fig. 5.3). Outside primers specifying undeleted portions of the gene were used to ensure gene disruption at known location on the chromosome, while inside primers

specifying the deleted parts of the gene were used to ensure that the disrupted gene was not accidentally displaced somewhere else on the chromosome.

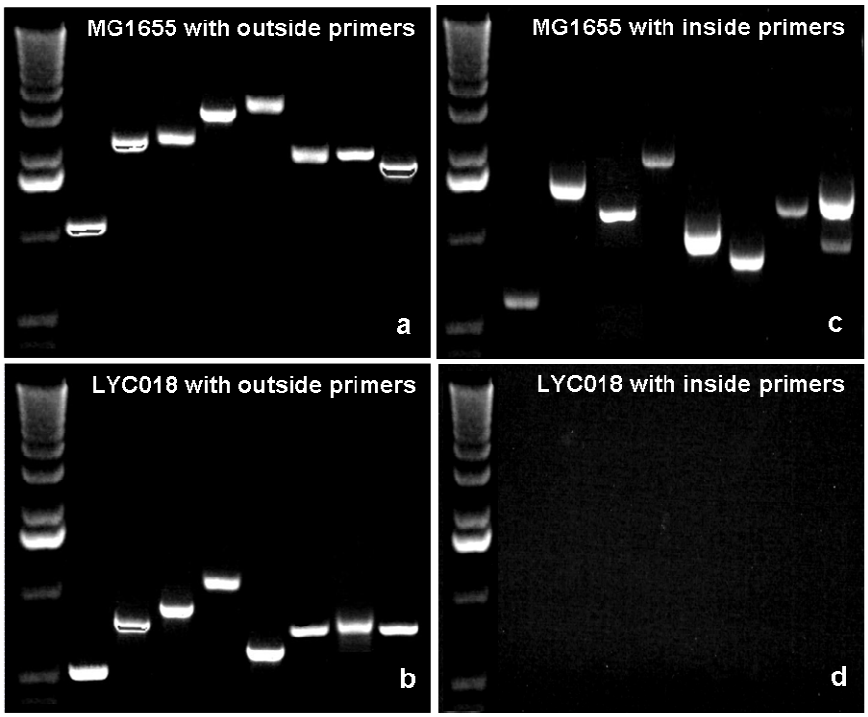


Figure 5.3 Gene deletions in the mutant LYC018 confirmed by PCR and gel electrophoresis using outside and inside primers. (Top) The PCR amplification product in the wild-type MG1655 with outside and inside primers. (Bottom) The PCR amplification product in the mutant LYC018 with outside and inside primer. Gene designations are in the following order: ldhA, frdA, poxB, pta, adhE, pykF, pckA, and tktB.

5.3.3 Strain characterization for production of lycopene

The performance of wild-type MG1655/pACEBI and mutant LYC018/pACEBI was compared. The experiments were conducted in aerobic batch baffled shake flasks containing 100 ml working volume of NBS minimal medium supplemented with 15 g/l glucose. Cell growth and lycopene production by the wild-type and the designed mutant are given in Fig. 5.4. MG1655/pACEBI converted glucose to lycopene at a yield of

0.11±0.02 mg lycopene/g glucose while LYC018/pACEBI reached a yield of 0.29±0.01 mg lycopene/g glucose, which is a 2.6-fold increase in lycopene yield (Table 5.2). Kinetic studies where the production of lycopene was stopped when the cells stopped producing biomass also confirmed lycopene as a growth-associated product.

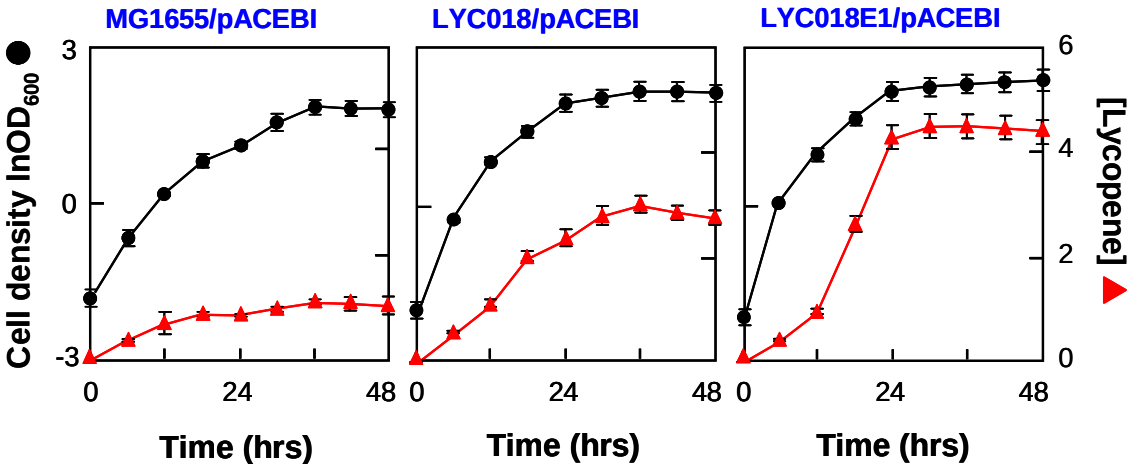


Figure 5.4 Time course of cell growth ($lnOD_{600nm}$) and lycopene (mg/l) production in MG1655/pACEBI, LYC018/pACEBI and LYC018E1/pACEBI aerobic shake flask cultures containing NBS medium supplied with 15 g/l glucose

Table 5.2 Performance summary of MG1655/pACEBI, LYC018/pACEBI and LYC018E1/pACEBI for the production of lycopene from glucose

	MG1655/pACEBI	LYC018/pACEBI	LYC018E1/pACEBI
Lycopene mg/L	1.10 ± 0.06	3.03 ± 0.01	4.47 ± 0.02
Lycopene yield[1] mg/g substrate	0.11 ± 0.02	0.29 ± 0.01	0.39 ± 0.01

[1] Lycopene yield is calculated from total lycopene produced per total glucose consumed.

5.3.4 Metabolic evolution of the designed mutant LYC018

The coupling of cell growth and LYC formation provides a basis for strain improvement through metabolic evolution. Selection for an improved growth would also

co-select for an improved production of LYC. Metabolic evolution with growth-based selection was performed for the designed mutant LYC018 in a serial dilution experiment. Cell growth rate of the mutant in each batch culture under aerobic condition was monitored during the serial cultivation. Initially, the mutant LYC018 started with a specific growth rate on glucose of 0.11±0.03 hr^{-1}. Improvements in specific growth rate were observed after about 65 generations of evolution (Fig. 5.5). The specific growth rate continuously increased with the transfers and become stabilized after approximately 180 generations. At the end of the metabolic evolution experiment of 208 generations, the specific growth rate of the evolved LYC018, renamed LYC018E1 was 0.28±0.02 hr^{-1}, a 2.5-fold higher rate than that of its parent.

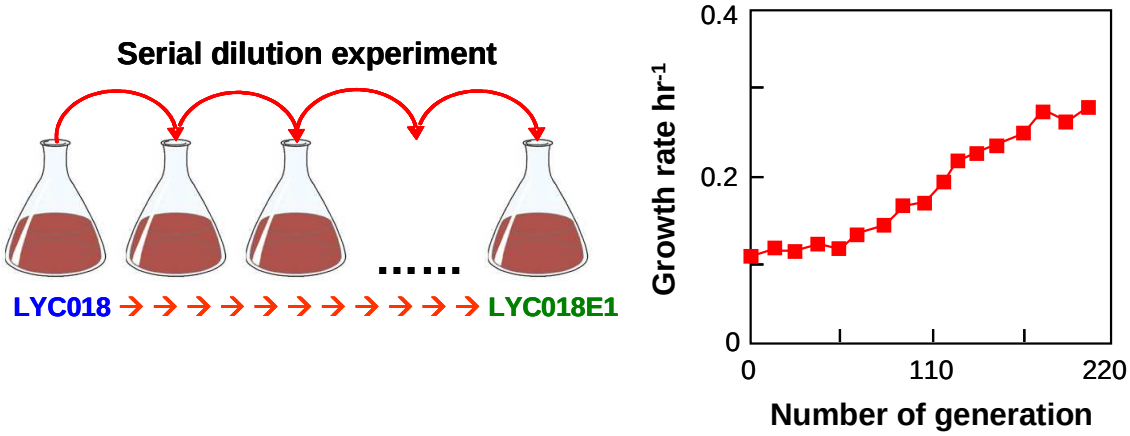


Figure 5.5 Metabolic evolution of LYC018 showing the change in growth rate during a serial dilution experiment. The experiment was conducted in baffled shake flasks containing 50 ml of NBS medium supplied with 2% glucose at 37°C, 200 rpm. The evolved strain with the higher cell growth rate was selected at the end of serial dilution experiment and named LYC018E1.

5.3.5 Production of lycopene by the evolved mutant

Lycopene production performance was characterized in the evolved strain LYC018E1/pACEBI. The mutant LYC018/pACEBI converted glucose into lycopene at a yield of 0.29±0.01 mg lycopene/g glucose after 36 hours, while the evolved mutant achieved a higher lycopene yield of 0.39±0.01 mg lycopene/g glucose around 30 hours. It is interesting to observe that the evolved mutant exhibited improvements in both yield and productivity. LYC018E1/pACEBI achieved approximately a 30% increase in lycopene yield compared to the yield of its parent LYC018/pACEBI and a 3.5-fold increase in lycopene yield compared to that of the wild-type MG1655/pACEBI under aerobic growth conditions. Growth and lycopene phenotypes of all strains are summarized in Table 5.2. Experimental yields of biomass and LYC of the wild-type, the mutant, and the evolved mutant fell into the constrained phenotypic space predicted by EMA (Fig. 5.2).

The probability for all seven elementary modes that contributes to the overall metabolism of LYC018/pACEBI and LYC018E1/pACEBI was computed based on Eq. (3.44) using the net accumulation rates of glucose, biomass and lycopene measured in the experiments and the summation constraint of all probabilities equal to one. Since the system was underdetermined, the probability was obtained using the least square method based on LSQR in MATLAB. Analysis of the associated usage probabilities of the unevolved and evolved mutants suggests that the cell naturally evolves by redistributing the probability of elementary modes to achieve the optimal fitness (Table 5.3). The predicted yields of the mutant LYC018/pACEBI and the evolved mutant

LYC018E1/pACEBI estimated from the product of the seven remaining modes and their probabilities were in agreement with the measured yields (Table 5.4). The results suggest that the EMA approach can also be applied to predict the metabolism of an evolving cell.

Table 5.3 Overall stoichiometric equations and usage probabilities of seven elementary modes in LYC018/pACEBI and LYC018E1/pACEBI for the production of lycopene from glucose. Biomass is shown in a unit of mmole.

EMs	Stoichiometric reactions	p_i of LYC018	p_i of LYC018E1
1	Glucose + 2.5 O_2 + 1.4 NH_3 → 1.8 Biomass + 1.1 CO_2	0.00	0.52
2	Glucose + 2.9 O_2 + 1.3 NH_3 → 1.7 Biomass + 0.4 Formate + 1.4 CO_2	0.00	0.00
3	Glucose + 3 O_2 + 0.9 NH_3 → 1.2 Biomass + 2.9 CO_2	0.93	0.38
4	Glucose + 1.3 O_2 + 0.7 NH_3 → 0.92 Biomass + 0.2 Formate + 0.06 Lycopene + 1.1 CO_2	0.00	0.00
5	Glucose + 0.9 O_2 + 0.5 NH_3 → 0.66 Biomass + 0.08 Lycopene + 1.1 CO_2	0.00	0.00
6	Glucose + 0.8 O_2 → 0.12 Lycopene + 1.3 CO_2	0.07	0.00
7	Glucose + 0.4 O_2 → 0.11 Lycopene + 0.01 Formate + 1.2 CO_2	0.00	0.10

Table 5.4 Measured and predicted yields of biomass and lycopene for LYC018/pACEBI and LYC018E1/pACEBI [1, 2]

Strains	Predicted Y_X	Measured Y_X	Predicted Y_{LYC}	Measured Y_{LYC}
LYC018/pACEBI	0.37	0.39	0.24	0.29
LYC018E1/pACEBI	0.50	0.50	0.36	0.39

[1] Y_X and Y_{LYC} represent yields of biomass (g/g) and lycopene (mg/g) respectively.
[2] Predicted yields are calculated from the sum of the product of the yield of the seven remaining elementary modes and the probability of modes.

5.4 Discussion

In this chapter, we present an alternative way for rational engineering of an efficient strain using the combined approaches of inverse metabolic engineering guided by EMA and metabolic evolution. The strategies are applied for an optimized lycopene production in a recombinant *E. coli* grown on glucose. The resulting strain LYC018E1/pACEBI showed improvements in both yield and productivity of lycopene synthesis.

EMA was applied to rationally design an *E. coli* mutant with minimized metabolic functionality tailored for efficient production of LYC under aerobic growth on glucose. The mutant containing eight gene disruptions of pykF, tktB, frdA, pckA, pta, ldhA, adhE, and poxB was theoretically designed and experimentally constructed to operate according to an efficient production of LYC based on the remaining elementary modes. The optimal combination of target gene knockouts was identified through a sequential search aiming at minimizing the cell's pathway options while retaining a maximum yield of lycopene and enabling cell growth. The kinetic studies demonstrated that the designed mutant LYC018/pACEBI achieved higher LYC yield on glucose than the wild-type MG1655/pACEBI (Fig. 5.4), which is in agreement with the EMA prediction. The combination of gene deletions is not intuitive and the individual contributions to the improved performance could be caused by several factors, including an increased supply of precursors and/or cofactors that are important for synthesis of LYC such as glycerol-3-phosphate, pyruvate, NADPH, or FAD. The disruption of genes involved in fermentative pathways of *E. coli* including frdA, ldhA, poxB, pta and adhE minimizes the production of competitive byproducts such as succinate, lactate, acetate and ethanol. The deletion of the pykF gene encoded for pyruvate kinase results in an increased glycolytic

supply for the LYC synthesis pathway (Al Zaid Siddiquee, et al. 2004). The disruption of transketolase (tktB) in the oxidative pentose phosphate pathway and phosphoenolpyruvate carboxykinase (pckA) in the anaplerotic pathway increases the carbon flux into the pool of lycopene precursors and reduces the carbon loss in form of carbon dioxide (Yang et al., 2003).

The co-production between LYC and cell growth observed in LYC018/pACEBI (Fig. 5.4) facilitates the directed evolution experiment based on a serial dilution of the cell culture. A change in cell growth observed during a serial dilution evolution (Fig. 5.5) suggested an occurrence of a new strain from the parent LYC018. Over the course of metabolic evolution, the mutant with a better cell growth rate would successively displace the parent resulting in a continuous increase in cell growth rate over number of generations. The evolved mutant LYC018E1/pACEBI generated in the present metabolic evolution study not only showed an improvement in cell growth, but was also co-selected for an increase in LYC production (Fig. 5.4). The results confirmed that a metabolic evolution strategy using a growth-based selection can be applied to enhance the production of growth-associated products such as LYC. LYC018E1/pACEBI and LYC018/pACEBI revealed a different time course in lycopene accumulation, suggesting that there were different modes of action and different combinatorial gene mutations between the two strains. In comparison to the parent strain LYC018/pACEBI, the evolved strain LYC018E1/pACEBI has an increased LYC yield and associated specific LYC production rate.

Carbon fluxes pertaining to LYC synthesis are likely controlled by rate-limiting reactions, i.e., ones with low enzyme capacities. These rate limiting steps within the pathways could be responsible for low production of LYC. This present work demonstrates that these reactions can be manipulated naturally through metabolic evolution by the selection of cells having improved cell growth and LYC production. The increased LYC production of the evolved mutant is probably associated with overcoming these rate-limiting steps. The identification of mutations occurring during the evolution that are responsible for the improved cell performance may provide a better understanding of the structure of control and regulation within the lycopene synthesis pathways. This can be achieved through the methods such as comparative genome sequencing and/or genomic microarray, which will be left for future investigation. Metabolic control analysis can also be applied to identify limiting steps within the metabolic network as demonstrated in Chapter 10.

The experimentally measured yields of biomass and LYC for the wild-type MG1655/pACEBI, the mutant LYC018/pACEBI, and the evolved mutant LYC018E1/pACEBI were within the solution space of yields predicted by EMA (Fig. 5.2), confirming the accuracy of the EMA approach. During the course of metabolic evolution, the mutant shifted toward a metabolic state with an increase in cell growth and LYC production. A mutant with increased cell growth was expected to emerge from the evolution process, since cell growth was used as a selection pressure in the evolution experiment. The increase in LYC production during evolution was also expected because

of the growth-associated production of lycopene which allows for a co-selection of the two products in the evolution experiment.

It should be noted that the mutant was restricted to evolve only in the context of the seven remaining designed pathways that enable production of LYC and cell growth more efficiently than those of the wild-type strain. The phenotypic movement of the mutant during evolution also revealed an increase in total entropy production. The overall entropy generation calculated from Eq. (3.23) of the mutant LYC018/pACEBI was 0.089 kJ/K-mole of glucose compared to 0.138 kJ/K-mole of glucose overall entropy produced by the evolved mutant LYC018E1/pACEBI. The observation suggests that the optimization of cellular metabolism during evolution may be driven to maximize the rate of entropy generation. This hypothesis is later confirmed in Chapter 9.

In summary, all the data indicate that LYC018E1/pACEBI operates according to a minimized metabolic network that has been reduced through gene knockout mutation and optimized through metabolic evolution to coproduce cell growth and LYC efficiently. Thus, it is demonstrated that metabolic engineering guided by elementary mode analysis in combination with metabolic evolution can be conducted for efficient strain development. Although in this study we have demonstrated as an example the application of these methodologies for the enhancement of lycopene production in *E. coli*, the approaches are applicable to the production of any desired growth-associated products.

Chapter 6

Development of a robust strain of *Escherichia coli* through metabolic evolution

6.1 Chapter summary

This chapter describes the application of metabolic evolution for optimization of the robustness of a cell. Lignocellulose-based ethanol production requires a microbial host that can effectively convert sugar-derived biomass to ethanol as well as a host that can tolerate inhibitors present in biomass hydrolysates. We have constructed a genetically engineered *E. coli* TCS083/pLOI297 that was capable of fermenting hexoses and pentoses at high yield of ethanol (Trinh et al., 2008). In this chapter, metabolic evolution was applied to select specifically for a robust mutant of the constructed strain. The isolated mutant AFF01/pLOI297 exhibited an improved tolerance against inhibitors, acetic acid and furfural, while maintaining a highly efficient ethanol fermentation of sugars. Under a challenge by the inhibitors, the evolved mutant achieved higher ethanol production efficiency than its parent. The results show that a metabolic evolution approach is applicable for the development of a robust microbial strain suitable for the production of biofuels and biochemicals from lignocellulosic biomass.

6.2 Introduction

The production of ethanol has generated a great deal of interest because ethanol can be used as an automotive fuel or a fuel additive. Ethanol fuel provides a sustainable energy resource and a more environmentally and economically friendly alternative to fossil fuels. Only lignocellulosic biomass is renewable, plentiful and inexpensive enough

to significantly supplement starch as a fermentation feedstock for ethanol production. Lignocellulose can be hydrolyzed into fermentable sugars, usually a mixture of pentoses and hexoses. Several inhibitory substances are also released during the hydrolysis reaction as a consequence of sugar degradation. The major obstacle to the production of ethanol from lignocellulose is a lack of a microbial strain capable of achieving high ethanol production efficiency in presence of these inhibitors.

The amount of inhibitors after hydrolysis reaction depends on the types of biomass feedstock and the methods used (Klinke, et al. 2004). Inhibitors that have received the most attention are acetic acid and furfural since they are major inhibitors commonly found in the lignocellulosic hydrolysates. Acetic acid is generated by the deacetylation of hemicellulose while furfural is produced by the dehydration of pentose sugars (McMillan, 1994; Palmqvist, 1998). These toxins are known to inhibit growth and ethanol production of many ethanologenic organisms including yeasts and bacteria (Zaldivar et al., 1999; Zaldivar and Ingram, 1999; Almeida et al., 2007). Removal of these toxic compounds, however, requires several detoxification steps, which add an additional cost and complexity to the process. Developing a robust organism that can resist the inhibitors is also a challenge. Because different inhibitors exert toxicity on the cell through different mechanisms and these toxic effects appear to be coupled, a complex phenotype of several mutations is likely required for creating a robust mutant strain. Engineering an organism for an improved tolerance to these inhibitors using a rational approach seems unlikely due to the complexity of the toxic effects caused by these inhibitors.

We have previously developed a highly efficient mutant strain of *E. coli*, TCS083/pLOI297 that is capable of fermenting hexose and pentose sugar to ethanol with high conversion efficiency (Trinh et al., 2008). TCS083/pLOI297 contains a total of eight gene deletions from the intermediary metabolism. Its metabolism is supported by six remaining efficient ethanol producing elementary modes during anaerobic growth on glucose or xylose. The metabolic evolution approach is used in this study to develop a robust mutant of TCS083/pLOI297 that is suitable for lignocellulose-based ethanol production. The robust mutant, with its capacity to resist to the inhibitors and to produce ethanol from sugars at high efficiency, was selected through an evolution process in a long-term chemostat culture. The chemostat culture was exposed to a mixture of hexose and pentose sugars and inhibitors. Acetic acid and furfural are inhibitors under examination in this study. The mutant was isolated from chemostat culture and was characterized in comparison to the parent strain for its robustness and for its fermentation performance in sugars under challenge of the inhibitors.

6.3 Results

6.3.1 Metabolic evolution of TCS083

An *E. coli* TCS083 culture was chemically mutagenized with a mutagen, nitrosoguanidine (NTG) in order to increase genetic diversity of the culture before use for inoculation in a chemostat culture. Briefly, an exponential culture of TCS083 grown in LB medium was collected, washed once with ice-cold water, resuspended in 100 μg/ml of NTG and incubated at room temperature for 15 minutes. The mutagenized cell culture was then washed with ice-cold water and used for inoculation. A mutant of TCS083 with

increased resistance to acetic acid and furfural was selected from the chemostat culture, which was continuously fed with 40 g/l mixed glucose and xylose as substrates and various concentrations of acetic acid and furfural as inhibitors. Inhibitory acetic acid concentration in the feed medium was at 5 g/l initially and was increased step-wise to 15 g/l (Fig. 6.1). After 400 hrs of operation, 0.5 g/l of furfural was added to the feed medium in addition to 15 g/l of acetic acid. The concentration of furfural was eventually increased to 2 g/l. A dilution rate of approximately 0.1 hr^{-1} was chosen as the operating condition since it was well below the maximum specific growth rate of TCS083 under the acetic acid challenge (μ_{max} = 0.30±0.02 hr^{-1} at 10g/l acetic acid) and under the furfural challenge (μ_{max} = 0.13±0.01 hr^{-1} at 1g/l furfural).

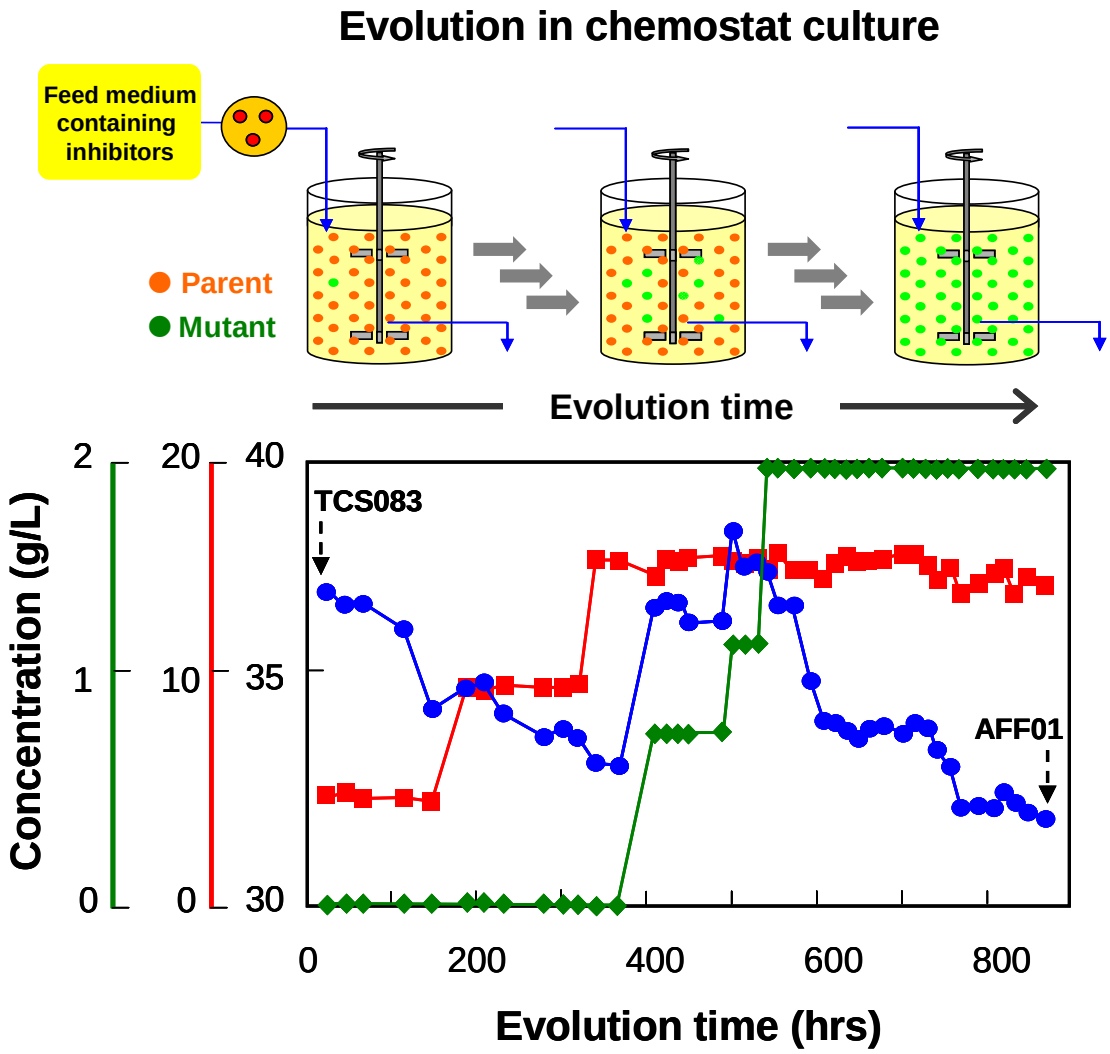


Figure 6.1 Metabolic evolution in a chemostat culture. (Top) Demonstration of metabolic evolution where inhibitors are used as selective pressure for selection of a mutant with improved resistance to the inhibitors. Over evolution time, the mutant strain with better resistance to the inhibitors would eventually replace the parent strain in the chemostat. (Bottom) Time profiles of residual sugar and inhibitors over evolution time in a chemostat culture of TCS083 fed with 40 g/l sugar and inihibitors: acetic acid and furfural. A mutant isolated from the chemostat culture of TCS083 after 880 hours was named AFF01. Symbols represent residual concentration of sugars (●), residual concentration of acetic acid (■), and initial concentration of furfural (◆).

Progress of the selection in the chemostat was monitored by effluent sugar concentration. Time profiles of effluent sugar concentration and inhibitors during the evolution process in the chemostat are shown in Fig 6.1. The culture reached steady state after 48 hrs with a steady-state sugar concentration of 37 g/l, indicating a loss of approximately 3 g/l of sugars in the chemostat. After a prolonged exposure to the acetic acid, the effluent sugar concentration decreased to 33 g/l. Raising the furfural concentration to 0.5 g/l and then to 1 g/l dramatically increased the residual sugar concentration to 37 g/l and 38.5 g/l respectively. Within the following 100 hrs, the culture recovered to almost the residual sugar concentration level prior to the addition of furfural and ultimately reached a plateau at steady-state concentration of 32 g/l after 760 hrs. The reduction of the residual sugar concentration indicated an increased sugar consumption of the chemostat culture. The data demonstrate that the mutant population was capable of utilizing sugars more efficiently than the parent under the challenge of the acetic acid-furfural inhibitors. Over evolution time, the mutant would come to dominate the chemostat population and would eventually replace the parent in the culture resulting in a new steady-state sugar concentration which was lower than that of the parent. At the end of approximately 880 hrs, a single clone was isolated and designated AFF01 for further characterization.

6.3.2 Effect of inhibitors on growth phenotype

The evolved mutant AFF01 and its parent TCS083 were characterized for their growth phenotypes in presence and absence of inhibitors. Both strains grew at a nearly identical specific growth rates (0.55 ± 0.07 hr^{-1} for TCS083 and 0.53 ± 0.02 hr^{-1} for AFF01)

when the cells were cultured without inhibitors. Under the acetic acid challenge, AFF01 showed a substantially higher specific growth rate than TCS083 (Fig. 6.2). While the parent ceased to grow at 30 g/l acetate at pH 7.0, the evolved culture was capable of growth at about 35 g/l acetate at pH 7. It was observed that the IC_{50} value of acetic acid for AFF01 was approximately 40% higher than that of TCS083. Under the challenge of furfural, TCS083 ceased to grow at approximately 2 g/l furfural whereas AFF01 could tolerate up to approximately 3 g/l furfural. AFF01 exhibited an increased furfural resistance with an approximately 70% higher IC_{50} than its parent TCS083. Comparison of resistance capacity against other inhibitors present in hydrolysates also showed AFF01 to have an improved tolerance to other acidic compounds including formic acid, syringic acid, vanillic acid, 3,4-hydroxybenzoic acid, 3,5-hydroxybenzoic acid, and vanillin (Fig. 6.3).

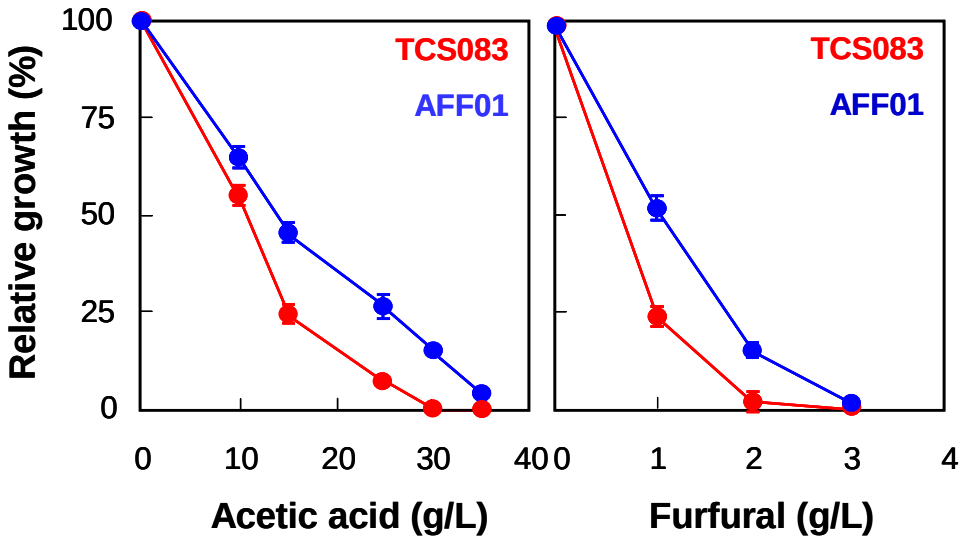


Figure 6.2 Growth inhibitions by acetic acid and furfural for TCS083 (red) and the evolved mutant AFF01 (blue). Relative growth is the ratio of specific growth rate in presence of inhibitor to that in absence of inhibitor.

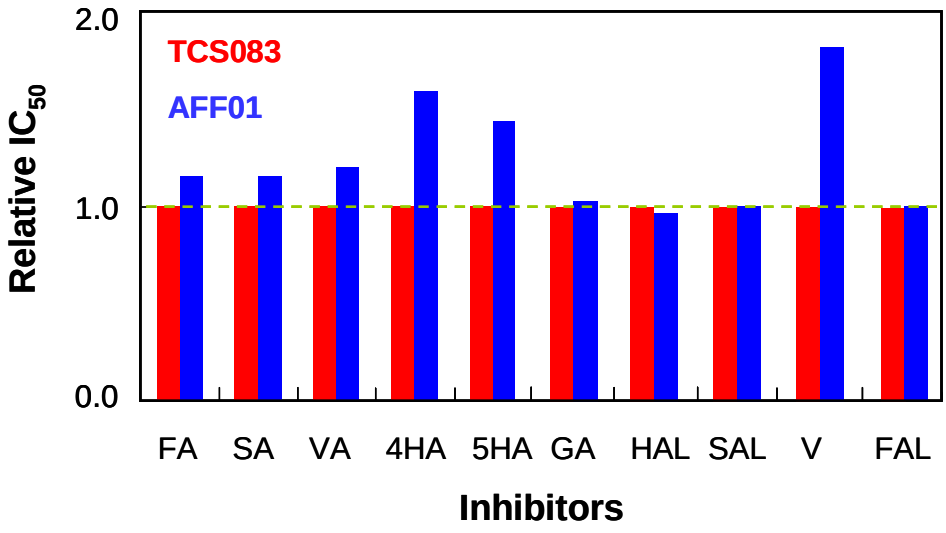


Figure 6.3 Relative IC_{50} value of inhibitors for TCS083 (red) and the evolved mutant AFF01 (blue). Abbreviations for inhibitors: FA: Formic acid; SA: Syringic acid; VA: Vanillic acid; 4HA: 3,4-Hydroxybenzoic acid; 5HA: 3,5-Hydroxybenzoic acid; GA: Gallic acid; HAL: Hydroxybenzaldehyde; SAL: Syringaldehyde; V: Vanillin; FAL: Furfuryl alcohol. Relative IC_{50} is defined as the ratio of inhibitory concentration at 50% reduced growth rate of inhibitors for AFF01 to TCS083.

6.3.3 Effect of inhibitors on ethanol fermentation

Fermentation of ethanol from 100 g/l mixed glucose-xylose in absence of inhibitors by AFF01/pLOI297 and TCS083/pLOI297 was examined initially (Fig. 6.4A and 6.4E). Similar specific growth rates were observed from both strains (0.37±0.02 hr^{-1} for TCS083/pLOI297 and 0.35±0.01 hr^{-1} for AFF01/pLOI207). An ethanol yield of a near theoretical maximum was reached by both cultures (Table 6.1). While the parent TCS083/pLOI297 utilized 81% of total sugars, the evolved mutant AFF01/pLOI297 consumed more than 99% of total sugars resulting in 21% higher final ethanol titer in the evolved mutant as compared to the parent.

The inhibitory effects of acetic acid and furfural on the fermentation performance of the evolved mutant AFF01/pLOI297 and its parent TCS083/pLOI297 were evaluated in Fig. 6.4B and 6.4F for acetic acid and Fig. 6.4C and 6.4G for furfural. The results of cell growth and ethanol production under the challenge of these inhibitors demonstrated that the evolved mutant AFF01/pLOI297 achieved a higher ethanol yield and productivity than its parent TCS083/pLOI297. Under a 15 g/l acetic acid challenge, AFF01/pLOI297 reached an approximately 26% higher ethanol yield and 5-fold faster ethanol productivity than TCS083/pLOI297. All sugars were exhausted within 50 hrs in the culture of AFF01/pLOI297 compared to only 35% of total sugars consumed by the culture of TCS083/pLOI297 over the same period of time. In the presence of 1 g/l furfural, AFF01/pLOI297 produced a 21% higher ethanol yield and an ethanol production rate double that of TCS083/pLOI297. The fermentation kinetics of the evolved mutant AFF01/pLOI297 and the parent TCS083/pLOI297 are summarized in Table 6.1 and compared with other ethanologenic *E. coli* strains reported in other literature in Table 6.2.

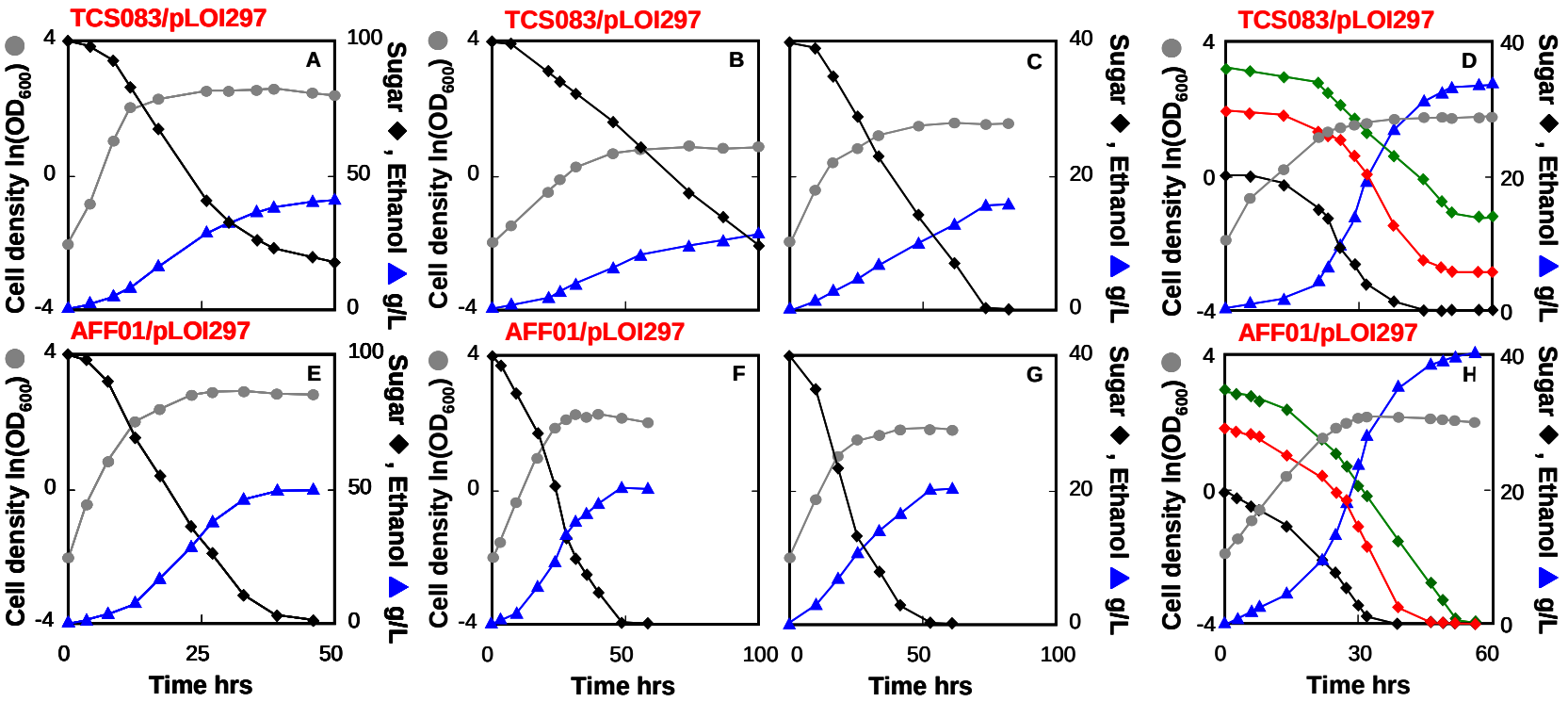


Figure 6.4 Ethanol fermentation by TCS083/pLOI297 (Top) and AFF01/pLOI297 (Bottom) in LB medium supplied with (A, E) 100 g/l sugar, (B, F) 40 g/l sugar and 15 g/l acetic acid, (C, G) 40 g/l sugar and 1 g/l furfural and (D, H) synthetic hydrolysate containing 30 g/l glucose (◆), 35 g/l xylose (◆), 20 g/l arabinose (◆), 5 g/l acetic acid and 0.5 g/l furfural

Table 6.1 Effect of inhibitors on ethanol fermentation by TCS083/pLOI297 and AFF01/pLOI297. Experiments were conducted in LB medium supplemented with sugars and inhibitors. Errors presented are the standard deviation of the duplicate samples.

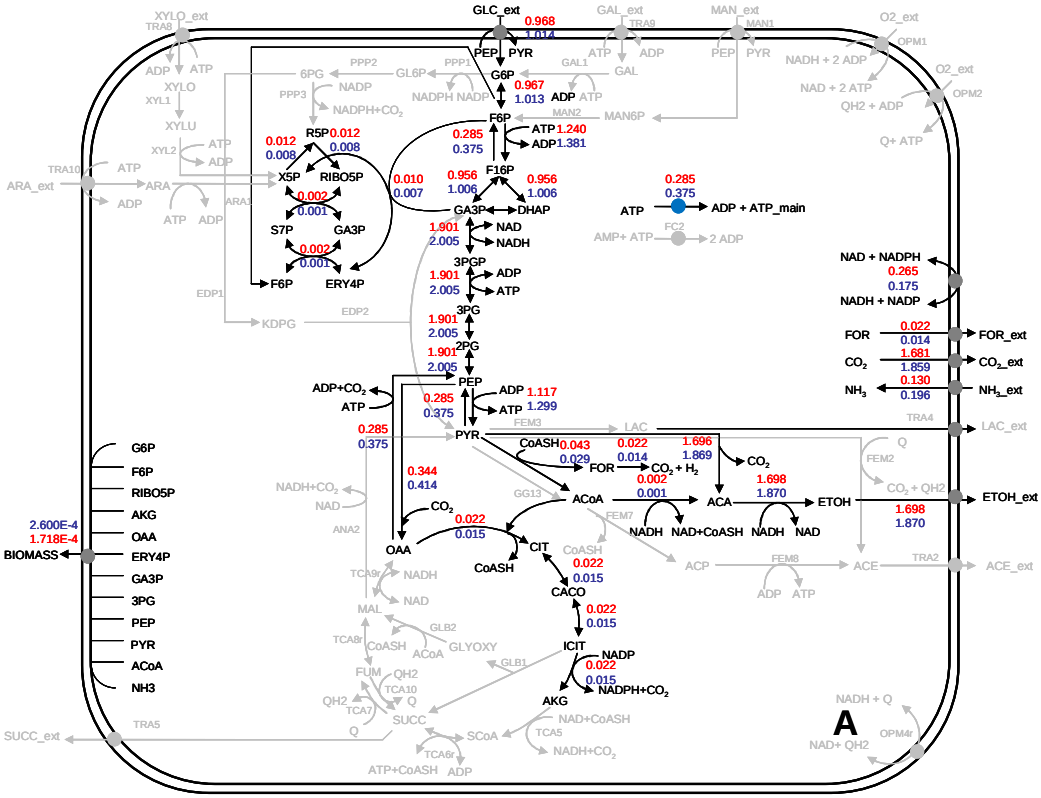
	Ethanol (g/L)	Residual sugars (%)[1]	$Yield_{ETOH}$ (%)[2]	V_{ETOH} (g/L-hr)[3]
No addition of inhibitors				
TCS083/pLOI297	41.2 ± 0.6	18.2 ± 1.0	98.7 ± 0.3	1.37 ± 0.04
AFF01/pLOI297	50.0 ± 0.3	0.6 ± 0.0	98.6 ± 0.0	1.62 ± 0.10
Addition of 15g/l acetic acid				
TCS083/pLOI297	11.4 ± 0.1	23.4 ± 0.4	73.0 ± 0.1	0.12 ± 0.01
AFF01/pLOI297	20.2 ± 0.1	0.0 ± 0.0	99.0 ± 0.0	0.63 ± 0.04
Addition of 1g/l furfural				
TCS083/pLOI297	15.7 ± 0.1	0.0 ± 0.0	76.9 ± 0.7	0.23 ± 0.00
AFF01/pLOI297	20.0 ± 0.2	0.0 ± 0.0	98.0 ± 0.0	0.47 ± 0.00

[1] Residual sugar is shown in percent of total sugar initially added.
[2] Yield based on consumed sugars. Yield was present in percent of theoretical yield of ethanol, 0.51 g ethanol/g consumed sugar.
[3] Volumetric ethanol productivity was calculated from linear regression of ethanol produced vs. time.

Table 6.2 Comparative fermentation performance of ethanologenic *E. coli* strains in LB medium supplemented with sugar. Ethanol yield was present in g-ethanol produced per g-sugar consumed.

Strains	Total sugars %	ETOH (g/L)	Y_{ETOH} (g/g)	References
KO11	9.0	41	0.46	Lindsay et al 1995
SL40	9.0	45	0.50	Lindsay et al 1995
FBR5	8.3	38	0.46	Nichols et al 2001
FBR14	8.7	42	0.48	Nichols et al 2001
FBR16	9.2	41	0.45	Nichols et al 2001
TCS083	10.0	41	0.50	This study
AFF01	10.0	50	0.50	This study

Flux distributions for AFF01/pLOI297 and TCS083/pLOI297 under inhibitor challenges are compared in Fig. 6.5A for acetic acid and Fig. 6.5B for furfural. The intracellular fluxes were computed from a linear combination of six elementary modes available to the strains (Trinh et al., 2008) and their probabilities as given in Table 6.3. The probabilities of each mode were determined from the measured extracellular fluxes of glucose, biomass, and ethanol during the anaerobic glucose fermentation in presence of the inhibitors, 15 g/l acetic acid or 1 g/l furfural. In both cases (acetic acid challenge and furfural challenge), AFF01/pLOI297 showed an increase in fluxes through the glycolytic pathway and the ethanol fermentative pathway when compared to TCS083/pLOI297.


GLC_ext
0.968
1.014
PEP
PYR
G6P
0.967
1.013
ADP
ATP
F6P
ATP 1.240
ADP 1.381
0.285
0.375
F16P
0.956
1.006
GA3P
DHAP
0.010
0.007
R5P 0.012
0.008
X5P
RIBO5P
S7P
ERY4P
0.002
0.001
ATP + ADP → ADP + ATP_main
AMP+ ATP → 2 ADP
NAD + NADPH
0.265
0.175
NADH + NADP
FOR
0.022
0.014
FOR_ext
CO2
1.681
1.859
CO2_ext
NH3
0.130
0.196
NH3_ext
3PGP
1.901
2.005
3PG
2PG
ADP+CO2 → ATP
PEP
ADP 1.117
ATP 1.299
PYR
LAC
LAC_ext
CoASH
0.043
0.029
FOR
CO2 + H2
0.022
0.014
1.696
1.869
CO2
0.002
0.001
ACoA
ACA
1.698
1.870
ETOH
ETOH_ext
1.698
1.870
CO2 + QH2
NADH
NAD+CoASH
NADH
NAD
ACP
ACE
ACE_ext
ADP
ATP
NADH+CO2 → NAD
NAD → CO2
0.344
0.414
OAA
0.022
0.015
CIT
CoASH
CACO
ICIT
NADP
0.022
0.015
NADPH+CO2
AKG
NADH → NAD
MAL
CoASH
ACoA
GLYOXY
FUM
QH2
Q
SUCC
SCoA
NAD+CoASH
NADH+CO2
ATP+CoASH
ADP
SUCC_ext
NADH + Q
NAD+ QH2
A
G6P
F6P
RIBO5P
AKG
OAA
ERY4P
GA3P
3PG
PEP
PYR
ACoA
NH3
2.600E-4
1.718E-4
BIOMASS

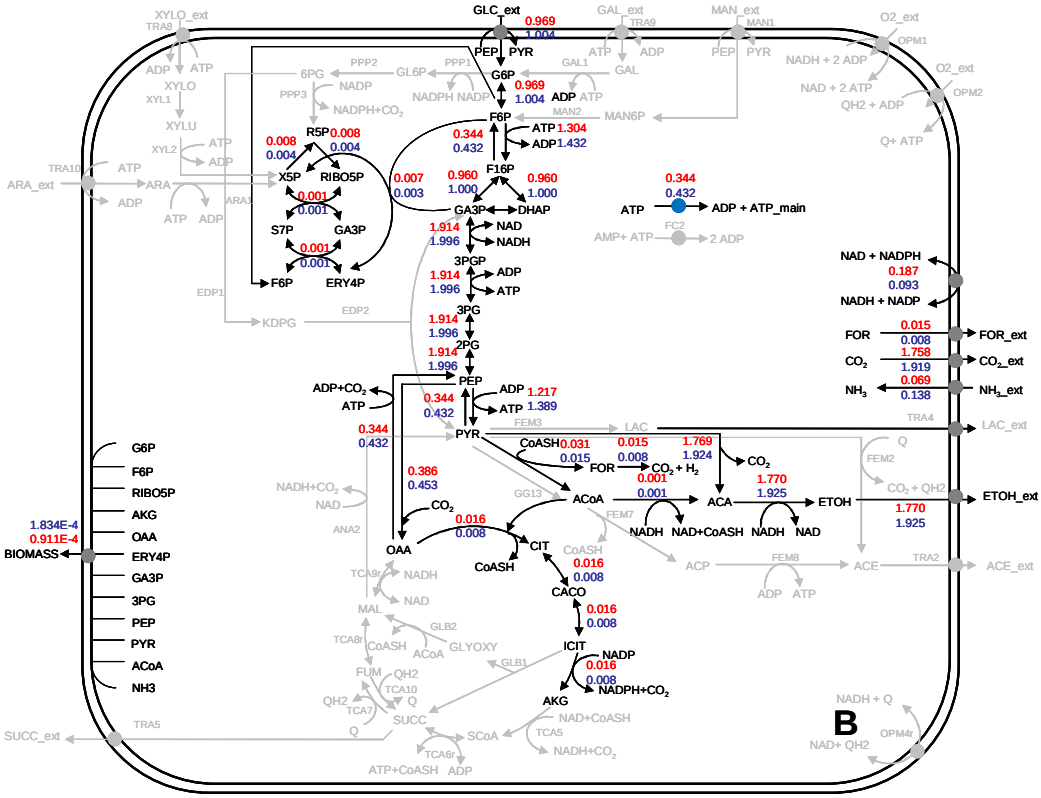

GLC_ext
0.969
1.004
PEP
PYR
G6P
0.969
1.004
ADP
ATP
F6P
ATP 1.304
ADP 1.432
0.344
0.432
F16P
0.960
1.000
GA3P
DHAP
0.007
0.003
R5P 0.008
0.004
X5P
RIBO5P
S7P
ERY4P
0.001
ATP + ADP → ADP + ATP_main
AMP+ ATP → 2 ADP
NAD + NADPH
0.187
0.093
NADH + NADP
FOR
0.015
0.008
FOR_ext
CO2
1.758
1.919
CO2_ext
NH3
0.069
0.138
NH3_ext
3PGP
1.914
1.996
3PG
2PG
ADP+CO2 → ATP
PEP
ADP 1.217
ATP 1.389
PYR
LAC
LAC_ext
CoASH
0.031
0.015
0.015
0.008
FOR
CO2 + H2
1.769
1.924
CO2
0.001
0.001
ACoA
ACA
1.770
1.925
ETOH
ETOH_ext
1.770
1.925
CO2 + QH2
NADH
NAD+CoASH
NADH
NAD
ACP
ACE
ACE_ext
ADP
ATP
NADH+CO2 → NAD
NAD → CO2
0.386
0.453
OAA
0.016
0.008
CIT
CoASH
CACO
ICIT
NADP
0.016
0.008
NADPH+CO2
AKG
NADH → NAD
MAL
CoASH
ACoA
GLYOXY
FUM
QH2
Q
SUCC
SCoA
NAD+CoASH
NADH+CO2
ATP+CoASH
ADP
SUCC_ext
NADH + Q
NAD+ QH2
B
G6P
F6P
RIBO5P
AKG
OAA
ERY4P
GA3P
3PG
PEP
PYR
ACoA
NH3
1.834E-4
0.911E-4
BIOMASS

Figure 6.5 Comparison of metabolic flux maps for TCS083/pLOI297 (top) and AFF01/pLOI297 (bottom) under the challenge of 15 g/l acetic acid (A) and the challenge of 1 g/l furfural (B). The intracellular fluxes were determined from a linear combination of six remaining elementary modes and their probabilities given in Table 6.3 (see Eq (3.44)). Non-zero fluxes and zero fluxes are black and grey correspondingly.

Table 6.3 Stoichiometric equations and probabilities of modes for the mutant TCS083/pLOI297 and the evolved mutant AFF01/pLOI297 grown on glucose under the challenge of acetic acid or furfural inhibitors

EMs	Stoichiometric equations	Acetate challenge		Furfural challenge	
		p_i of TCS083	p_i of AFF01	p_i of TCS083	p_i of AFF01
1	Glucose → 2 Ethanol + 2 CO_2	0.142	0.188	0.172	0.216
2	Glucose → 2 Ethanol + 2 CO_2	0.142	0.188	0.172	0.216
3	Glucose → 2 Ethanol + 2 CO_2	0.142	0.188	0.172	0.216
4	Glucose → 2 Ethanol + 2 CO_2 + 2 ATP	0.142	0.188	0.172	0.216
5	Glucose + 0.49 NH_3 → 1.79 Biomass + 1.40 Ethanol + 0.11 H_2 + 1.41 CO_2	0.199	0.132	0.141	0.070
6	Glucose + 0.49 NH_3 → 1.79 Biomass + 1.40 Ethanol + 0.11 Formate + 1.30 CO_2	0.199	0.132	0.141	0.070

6.3.4 Ethanol fermentation in synthetic hydrolysate

A synthetic hydrolysate that mimics the concentration of sugars and inhibitors present in corn fiber hydrolysate was used in this work for testing the performance of the evolved mutant AFF01/pLOI297 in comparison to that of the parent TCS083/pLOI297. The mixture contains 30 g/l glucose, 35 g/l xylose, 20 g/l arabinose, 5 g/l acetic acid and 0.5 g/l furfural based on O'Brien et al. (2004). In the synthetic hydrolysate, AFF01/pLOI297 reached a higher yield and a higher production rate than TCS083/pLOI297. TCS083/pLOI297 left a significantly higher concentration of sugars than

AFF01/pLOI297 at the same fermentation time. The evolved mutant AFF01/pLOI297 fermented all sugars within 56 hrs with a maximal ethanol concentration of 41 g/l and ethanol yield of 94% of the theoretical maximum (Fig. 6.4D and 6.4H). The fermentation performance of AFF01/pLOI297 in the synthetic hydrolysate compared favorably with the performances for the ethanologenic *E. coli* strains such as KO11 or FBR under similar concentrations of sugars and inhibitors (Table 6.4).

Table 6.4 Comparative fermentation performance of ethanologenic *E. coli* strains in corn fiber hydrolysates. The data were obtained either from real hydrolysates or synthetic hydrolysates. The synthetic hydrolysates contained a mixture of 30 g/l glucose, 35 g/l xylose, 20 g/l arabinose, 5g/l acetic acid and 0.5 g/l furfural which mimic concentration of sugars and inhibitors in corn fiber hydrolysates (O'Brien, et al. 2004). Ethanol yield was present in g-ethanol produced per g-sugar added.

Strains	Total sugars %	ETOH (g/L)	Y_{ETOH} (g/g)	V_{ETOH} (g/l-hr)	References
KO11	5.6	27	0.47	1.05	Moniruzzaman et al 1996
KO11	8.9	35	0.41	1.16	Dien et al 1997
KO11	9.4	42	0.46	0.46	Asghari et al 1996
KO11	10	41	0.41	0.40	Dien et al 2004
SL40	5.6	27	0.46	1.18	Moniruzzaman et al 1996
SL40	8.9	32	0.42	1.12	Dien et al 1997
FBR3	6.6	29	0.47	0.35	Dien et al 1999
FBR5	8.0	35	0.44	0.43	Dien et al 2004
FBR5	8.5	37	0.46	0.77	Dien et al 2000
AFF01	5.7	27	0.46	1.06	This study
AFF01	8.5	41	0.48	0.76	This study
TCS083	8.5	34	0.40	0.55	This study

6.4 Discussion

Given that engineering an efficient and robust microbial strain is a primary challenge for biomass-ethanol conversion, in this chapter, our study attempts to develop a mutant strain of *E. coli* that is highly tolerant of inhibitory substances present in biomass hydrolysates and that can ferment sugar-derived biomass to ethanol at a high efficiency .

Here, metabolic evolution is utilized to select an evolved mutant of TCS083, a strain designed based on elementary mode analysis for the efficient conversion of sugars to ethanol and with high tolerance to inhibitors. A chemostat was used to isolate a mutant of TCS083 with enhanced cell growth under the challenge of acetic acid and furfural, the two major inhibitors in hydrolysates. The chemostat was inoculated with TCS083 and operated for approximately 30 days. Acetic acid and furfural were added to the feed media as a selective pressure for the mutant isolation. Over a prolonged period, microbial populations adapted to the toxic condition by mutations. Through natural selection, the mutant with high tolerance to the inhibitors remained in the culture and eventually replaced the parental strain. This was observed by the decrease in residual sugars in the chemostat culture and the establishment of a new steady-state sugar concentration after a long-term exposure of the culture to the inhibitors (Fig. 6.1). The chemostat culture selection resulted in a mutant that exhibited an improved growth rate in presence of acetic acid and furfural inhibitors as shown in Fig. 6.2.

Growth experiments showed that the evolved mutant AFF01 and its parent TCS083 grew at a similar rate in absence of inhibitors. In presence of inhibitors (acetic acid or furfural), however, AFF01 had an increased tolerance to the inhibitors and grew at a faster rate than TCS083 (Fig. 6.2). The results suggest that the mutations that occurred in the evolved mutant AFF01 specifically affect the mechanisms involved with growth inhibitions. According to the previously proposed mechanism for organic acid toxicity in microorganisms (Palmqvist et al., 1999a; Almeida et al., 2007), the conjugate-neutral form of acetic acid diffuses through the cell membrane into the cytoplasm and dissociates

resulting in an increase of the intracellular concentration of [H^+]. This increase in [H^+] causes the collapse in △pH by uncoupling the proton gradient for the ATP synthesis system. This would lead to a depletion of the supply of ATP which would ultimately result in cell starvation. The acetic acid also causes plasma membrane damage resulting in increased cell leakage. Thus, the acquired resistance to the acetic acid compound of AFF01 may be a result of genetic mutations that are responsible for the synthesis of cell membrane, the control of intracellular △pH and/or the export of the acetic acid. AFF01 also exhibited some improved resistance to other acidic-type inhibitors such as formic acid, syringic acid, vanillic acid and hydroxybenzoic acid. This implies that the mutations may target the toxicity mechanisms that these acidic-type compounds shared in common. It is possible that the increased tolerance to inhibitors of AFF01 is a result of not one gene mutation but several gene mutations combined. Future investigation using comparative genome sequencing or genomic microarray will be needed to identify the genetic modifications and their roles for the increased tolerance to inhibitors of AFF01. Comparing robustness of AFF01 to results by previous studies of other *E. coli* strains showed the strain to be 53% more resistant to acetic acid than the previously developed *E. coli* strain LY01 and 3-fold more resistant than the wild-type *E. coli* K12 (Zaldivar et al., 1999; Lasko et al., 2000).

Besides cell membrane damage, furfural is known to cause a direct inhibition of enzymes in the glycolysis pathway and in the ethanol synthesis pathway (Palmqvist et al., 1999b). As a result, it is likely that in the evolved strain AFF01, mutations also occurred in the enzymes involved in these pathways, causing increases in glycolytic fluxes and

ethanol synthesis fluxes and reducing the toxicity effect of furfural. This assumption is supported by the metabolic flux maps (Fig. 6.5) which showed higher fluxes through the glycolytic pathway and the ethanol synthesis pathway in AFF01/pLOI297 than its parent TCS083/pLOI297. Since the glycolytic pathway is involved in the synthesis of several of biomass precursors and ethanol precursors, the increase in the glycolytic flux of AFF01/pLOI297 may explain the higher growth rate and the higher rate of ethanol synthesis of AFF01/pLOI297 observed in the ethanol fermentation challenged by inhibitors as shown in Fig. 6.4.

Although acetic acid and furfural inhibit ethanol fermentation of both AFF01/pLOI297 and TCS083/pLOI297, the degree of inhibition was much smaller in AFF01/pLOI297 compared to the parent (Fig. 6.4). The fermentation performance of TCS083/pLOI297 in the presence of 15 g/l acetic acid showed a 25% reduction in ethanol yield and 11-folds decrease in ethanol productivity compared to the performance in the absence of acetic acid (Table 6.1). The evolved mutant AFF01/pLOI297, however, showed no ethanol yield reduction and only 2.5-fold reduction in the ethanol production rate under the acetic acid challenge. The fermentation performance of AFF01/pLOI297 surpassed that of *E. coli* KO11 described by Caroline et al. (1999). The evolved mutant AFF01/pLOI297 produced ethanol at rate of 0.63 g/l-hr compared to 0.47 g/l-hr ethanol productivity achieved by KO11 under 15 g/l acetic acid challenge.

A similar trend was observed in the presence of furfural where TCS083/pLOI297 exhibited a 22% reduction in ethanol yield and 6-fold reduction in ethanol productivity, while the evolved mutant AFF01/pLOI297 showed no ethanol yield reduction and only a

3-fold reduction in ethanol production rate. It is interesting to observe that the inhibitors have a greater effect on the production rate of ethanol than on the ethanol yield. Inhibition of ethanol production by acetic acid and furfural was also observed by others for *E. coli* (Zaldivar et al. (1999)) and for *S. cerevisiae* (Banerjee et al. (1981); Palmqvist (1998)).

Ethanol fermentation with the evolved mutant AFF01/pLOI297 in the absence of inhibitors reached a higher ethanol titer than TCS083/pLOI297 suggesting that AFF01/pLOI297 may also have some tolerance for ethanol. Ethanol is known to cause cell membrane damage as previously reported by Ingram (1986). The resistance to ethanol of AFF01/pLOI297 may involve a change in cell membrane composition of the strain. The ethanol yield and titer obtained by AFF01/pLOI297 outperformed the previously reported values of other ethanologenic recombinant *E. coli* strains such as KO11 or FBR as summarized in Table 6.2.

In summary, we demonstrated in this chapter the metabolic evolution approach under growth-based selection for the development of a mutant *E. coli* strain with improved robustness. The engineered strain TCS083 was evolved in a long-term chemostat containing the inhibitors acetic acid and furfural as selective pressure. The evolved mutant AFF01 isolated from the chemostat culture showed improved tolerance against the inhibitors while retaining highly efficient ethanol fermentation from its parent TCS083. The example of improved acetate-furfural tolerance in AFF01 demonstrates that the metabolic evolution approach using a chemostat is valuable for the development of strains for production of biochemicals and biofuels from lignocellulosic biomass.

Chapter 7

Continuous production of ethanol from hexoses and pentoses using immobilized mixed cultures of *Escherichia coli* strains[†]

7.1 Chapter summary

This chapter demonstrates the design and optimization of immobilized mixed cell systems for efficient ethanol fermentation from sugar mixtures. We have developed highly efficient ethanologenic *E. coli* strains that selectively consume pentoses and/or hexoses. Mixed cultures of these strains can be used to selectively adjust the sugar utilization kinetics in ethanol fermentations. Based on the kinetics of sugar utilization of these strains, we have designed and implemented an immobilized cell system for the optimized continuous conversion of sugars into ethanol. The results confirm that immobilized mixed cultures support a simultaneous conversion of hexoses and pentoses into ethanol at high yield and at a faster rate than immobilized homogenous cells. Continuous ethanol production has been maintained for several weeks at high productivity with near complete sugar utilization. The control of sugar utilization using immobilized mixed cultures can be adapted to any composition of hexoses and pentoses by adjusting the strain distribution of immobilized cells. The approach, therefore, holds promise for ethanol fermentation from lignocellulosic hydrolysates where the feedstock varies in sugar composition.

[†]This chapter is from: Unrean, P., Srienc, F. 2010. Continuous production of ethanol from hexoses and pentoses using immobilized mixed cultures of *Escherichia coli* strains. J. Biotechnol.

7.2 Introduction

Ethanol has emerged as an important renewable and sustainable energy source that can reduce our reliance on fossil resources. Inexpensive, abundant and renewable lignocellulosic biomass has become an attractive feedstock for ethanol fuel production by fermentation (Mielenz, 2001; Ragauskas et al., 2006; Wyman, 2001; Zaldivar et al., 2001). The hydrolysis of lignocellulosic materials releases a mixture of hexoses and pentoses including glucose, galactose, mannose, xylose and arabinose. An economic success of lignocellulose based ethanol production would therefore require a simultaneous utilization of these biomass-derived sugars. Finding an ethanologenic organism that is able to ferment efficiently and simultaneously the sugar mixtures present in hydrolysates is a significant hurdle. Most organisms currently used in the production of ethanol either consume hexoses and pentoses sequentially or are unable to utilize pentoses effectively.

Despite many attempts of genetic engineering an organism for co-fermentation of hexoses and pentoses (Becker and Boles, 2003; Ho et al., 1998; Kotter and Ciriacy, 1993; Wisselink et al., 2007), one sugar is often preferred over another by a genetically engineered organism as a result of catabolite repression. This asynchronous consumption of sugar mixtures leads to a longer fermentation time for a complete conversion of all sugars to ethanol. In addition, genetically engineered strains can suffer from limited gene expression as reported for several cases where other products instead of ethanol were produced. For example, *Saccharomyces cerevisiae* engineered for pentose fermentation yields a high amount of undesired side products like xylitol and arabinitol due to limited

enzyme activity in the pentose metabolism pathway (Bruinenberg et al., 1983; Karhumaa et al., 2006; Walfridsson et al., 1995). Furthermore, even if an organism can consume both hexoses and pentoses simultaneously, use of a single organism for the fermentation of sugar mixtures poses a challenge because the organism may not be able to optimally adjust its sugar uptake rate to match fluctuating sugar concentrations in hydrolysates from different sources of biomass. Because of these challenges, a process using a co-culture of multiple strains could be preferred over the culture of a single strain for the fermentation of sugar mixtures. Unlike the single-strain culture, the composition of a co-culture of multiple strains can be conveniently adjusted as the relative consumption rate of each sugar in the mixture depends on the ratio of strains used in the immobilization. Therefore, the co-culture approach should result in a simultaneous consumption of all sugars in a shorter fermentation time. The approach is also expected to be able to optimally handle any given sugar mixtures available in biomass hydrolysates.

We have previously applied a rational strain design in combination with a metabolic evolution step for developing mutant strains of *Escherichia coli* that are capable of fermenting hexoses and/or pentoses to ethanol efficiently and are robust under technologically challenging conditions. This includes the mutant AFF01/pLOI297 which was designed to ferment both hexoses and pentoses and the mutant CT1101/pLOI297 which was designed specifically for the exclusive fermentation of pentoses (Trinh et al., 2008). In this study, we immobilized these mutant strains for a continuous and efficient ethanol fermentation process from mixtures of hexoses and pentoses. Ethanol fermentation using immobilized cells is attractive since it provides several advantages

over freely suspended cells including ease of cell separation and recycling or perfusion as previously reported in several studies (Gilson and Thomas, 1995; Giordano et al., 2008; Nigam, 2000; Takamitsu et al., 1993; Verbelen et al., 2006; Yamada et al., 2002; Zhou et al., 2008). The immobilization technique is simple and low-cost. An easy recovery of immobilized cells helps to reduce process complexity in cell separation. Immobilization is also a suitable process to obtain a high cell density per unit volume which is desirable particularly in a continuous culture. Another benefit of continuous fermentation using immobilized cells is that there is no lag phase and/or turn-around time. Therefore, a higher volumetric ethanol production efficiency can be attained with immobilized cells in comparison to freely suspended cells.

We demonstrate, in this chapter, a continuous co-culture of immobilized AFF01/pLOI297 and CT1101/pLOI297 cells for the fermentation of mixed hexoses and pentoses. A kinetic model was developed and used to design optimal conditions for the co-culture such that the most rapid conversion of hexose and pentose sugars to ethanol is achieved. We also show that the process using immobilized mixed cells is adaptable to any composition of sugar mixtures available in biomass hydrolysates. The high ethanol productivity of immobilized cells was maintained over several weeks of continuous operation without loss of performance.

7.3 Theory

Kinetic model for the simultaneous utilization of sugar mixtures

A kinetic model describing a specific sugar consumption rate of immobilized AFF01/pLOI297 and CT1101/pLOI297 cells in a sugar mixture was developed. The

model is based on the assumption that the sugars are taken up by a transport complex that has a common element for which the sugars compete. One should note that this common element is not necessarily the transport protein itself but rather some other step downstream of the sugar metabolism. The mechanism of sugar uptake can be described by the reaction

$$T + S_{i,ext} \underset{k_{-1,Si}}{\overset{k_{1,Si}}{\rightleftarrows}} TS_i \overset{k_{2,Si}}{\rightarrow} T + S_i \qquad \textbf{(7.1)}$$

where T [1/g-cdw] is the number of transporter complexes per unit cell dry weight, $C_{Si,ext}$ and C_{Si} [mmole/l] are extracellular and intracellular concentration of sugar S_i, and TS_i [1/g-cdw] is the number of transporter-sugar complex per unit cell dry weight.

The uptake rate of sugar S_i is given by

$$q_{Si} = k_{2,Si} TS_i \qquad \textbf{(7.2)}$$

where q_{Si} [mmole/g-cdw-hr] is the specific uptake rate of sugar S_i, TS_i [1/g-cdw] is the number of transporters bound to sugar per unit cell dry weight and $k_{2,Si}$ [mmole/hr] is the sugar specific rate constant.

By applying the quasi-steady state assumption and the conservation of the transporter complex to the balance equation, we obtain

$$TS_i = \frac{k_{1,Si}}{k_{-1,Si} + k_{2,Si}} T \cdot C_{Si} \qquad \textbf{(7.3)}$$

$$T = T_0 - \sum TS_i \qquad \textbf{(7.4)}$$

Substituting Eq (7.3) into the transporter conservation Eq (7.4) yields

$$T = \frac{T_0}{1+\sum K_{m,Si} C_{Si}} \quad \textbf{(7.5)}$$

where $K_{m,Si}$ is $k_{1,Si}/(k_{-1,Si}+k_{2,Si})$.

Substituting Eq (7.3) and Eq (7.5) into the uptake rate of sugar

$$q_{Si} = \frac{k_{2,Si} K_{m,Si} T_0}{1+\sum K_{m,Si} C_{Si}} C_{Si} \quad \textbf{(7.6)}$$

For concentrated sugar concentration $\sum K_{m,Si} C_{Si} >> 1$, the specific consumption rate of each sugar becomes,

$$q_{Si} = \frac{k_{2,Si} K_{m,Si} T_0}{\sum K_{m,Si} C_{Si}} C_{Si} \quad \textbf{(7.7)}$$

By assuming the same $K_{m,Si}$, the sugar uptake rate can be rewritten in a general form as

$$q_{Si} = k_{2,Si} T_0 \frac{C_{Si}}{\sum C_{Si}} = k_{Si} \frac{C_{Si}}{\sum C_{Si}} \quad \textbf{(7.8)}$$

where T_0 [1/g-cdw] is the total number of transporter complexes, C_{Si} is the sugar concentration and k_{Si} [mmole/g-cdw-hr] is the overall rate constant. Eq (7.8) suggests a linear correlation between the specific rate of sugar consumption and the corresponding sugar mole fraction in the mixture. Thus the overall rate constant can be experimentally determined by measuring the specific consumption rates for various mole fractions of the sugar (see Fig. 7.3).

In a glucose-xylose mixture and at a constant mole fraction, the sugars are consumed according to a zero order kinetics described by

$$q_{glc,AF} = k_{glc,AF} \frac{C_{glc,0}}{C_{glc,0}+C_{xyl,0}} \quad \textbf{(7.9)}$$

$$q_{xyl,AF} = k_{xyl,AF} \frac{C_{xyl,0}}{C_{glc,0} + C_{xyl,0}} \quad \textbf{(7.10)}$$

$$q_{xyl,CT} = k_{xyl,CT} \quad \textbf{(7.11)}$$

where Eq (7.9) and (7.10) specify the specific consumption rate by strain AFF01/pLOI297 for glucose and xylose respectively, and Eq (7.11) is the specific xylose consumption rate for strain CT1101/pLOI297 that cannot use glucose. The empirical constant $k_{Si,j}$ represents a maximum specific consumption capacity of sugar S_i by an immobilized cell type j, and $C_{Si,0}$ is the initial sugar concentration.

Batch cultures

Because of the zero order consumption kinetics, the glucose and xylose exhaustion times (the times at which the respective sugars are consumed) for cultures containing immobilized AFF01/pLOI297 at a cell concentration X_{AF} can be computed for glucose

$$t_{glc} = \frac{C_{glc,0}}{q_{glc,AF} X_{AF}} = \frac{C_{glc,0} + C_{xyl,0}}{k_{glc,AF} X_{AF}} \quad \textbf{(7.12)}$$

and for xylose:

$$t_{xyl} = \frac{C_{xyl,0}}{q_{xyl,AF} X_{AF}} = \frac{C_{glc,0} + C_{xyl,0}}{k_{xyl,AF} X_{AF}} \quad \textbf{(7.13)}$$

These times are generally not the same. One can see that for $k_{glc,AF} > k_{xyl,AF}$ the following inequality holds for the sugar consumption times

$$t_{glc} < t_{xyl} \quad \textbf{(7.14)}$$

In a mixed cell culture of immobilized CT1101/pLOI297 and AFF01/pLOI297, the cell concentration of each strain can be written

$$X_{CT} = fX_{TOT} \quad \textbf{(7.15)}$$

$$X_{AF} = (1-f)X_{TOT} \quad \textbf{(7.16)}$$

where X_{TOT} is the total cell concentration [g-cdw/l] and f is the fraction of the total cells consisting of strain CT1101/pLOI297 that can utilize only pentoses.

The exhaustion time for glucose and xylose in a mixture of both sugars by immobilized mixed cultures of the two cell types becomes

$$t_{glc} = \frac{C_{glc,0} + C_{xyl,0}}{k_{glc,AF}(1-f)\cdot X_{TOT}} \quad \textbf{(7.17)}$$

$$t_{xyl} = \frac{C_{xyl,0}(C_{glc,0} + C_{xyl,0})}{k_{xyl,AF}C_{xyl,0}(1-f)X_{TOT} + k_{xyl,CT}(C_{glc,0} + C_{xyl,0})fX_{TOT}} \quad \textbf{(7.18)}$$

Fig. 7.1 shows the exhaustion times for the individual sugars as a function of the cell fraction *f*. One can see that the exhaustion times can be adjusted with the cell composition. A practical process should completely consume both sugars in the shortest possible time. This is achieved when both sugars are exhausted at the same time ($t_{glc} = t_{xyl} = t_{opt}$) which occurs at an optimal cell fraction that can be computed from

$$f_{opt} = \frac{(k_{xyl,AF} - k_{glc,AF})C_{xyl,0}}{(k_{xyl,AF} - k_{glc,AF})C_{xyl,0} - (C_{glc,0} + C_{xyl,0})k_{xyl,CT}} \quad \textbf{(7.19)}$$

With the optimal cell fraction, the exhaustion time of each sugar becomes

$$t_{opt} = \frac{(C_{glc,0} + C_{xyl,0})k_{xyl,CT} - (k_{xyl,AF} - k_{glc,AF})C_{xyl,0}}{k_{glc,AF}k_{xyl,CT}X_{TOT}} \quad \textbf{(7.20)}$$

Fig. 7.1 also illustrates that the time required for exhausting a sugar mixture is shorter in the mixed cell culture than in the single cell culture.

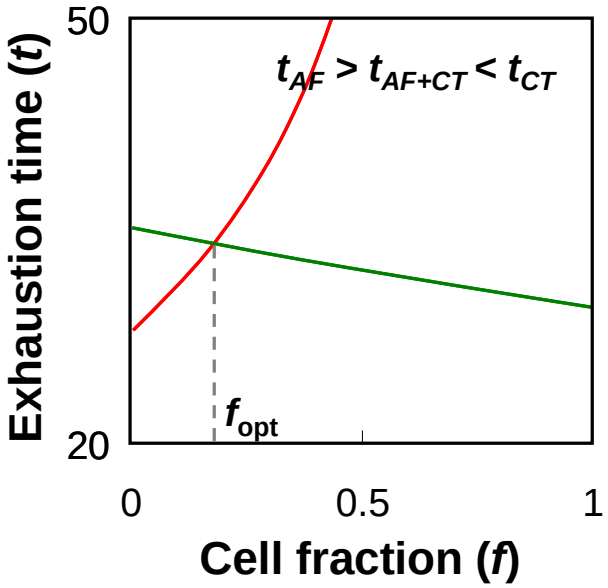


Figure 7.1 Exhaustion time of glucose (red) and xylose (green) as a function of the fraction of CT1101/pLOI297 cells used in the immobilization (see Eq (7.17) and (7.18)). Parameters used for the model are $C_{glc,0}$ = 111 mmole/l; $C_{xyl,0}$ = 400 mmole/l; X_{TOT} = 3.62 g-CDW/l. At cell fraction f_{opt} = 0.21, both sugars are consumed at the same time.

Continuous cultures

The previous result can be readily extended to a continuous culture. In a continuous fermentation of sugar mixtures using immobilized cells of multiple strains, the balance of sugar S_i is

$$\frac{dC_{Si}}{dt} = D_L(C_{Si,F} - C_{Si}) - \sum X_j q_{Si,j} \qquad \textbf{(7.21)}$$

where D_L is dilution rate based on liquid volume of the culture, $C_{Si,F}$ is the sugar concentration in the feed medium, X_j is the cell concentration of strain j, and $q_{Si,j}$ represents a specific uptake rate of sugar S_i by the strain j.

For a complete conversion of each sugar at the same exhaustion time, the optimal residence time $\tau_{L,opt}$ becomes

$$\tau_{L,opt} = \frac{1}{\mathrm{D}_{L,opt}} = \frac{\mathrm{C}_{Si,\mathrm{F}}}{\sum \mathrm{X}_j q_{\mathrm{Si},j}} \quad \textbf{(7.22)}$$

For a steady state with complete utilization of a glucose-xylose mixture by immobilized cells of AFF01/pLI297 and CT1101/pLOI297, the optimal residence time and cell composition are

$$\tau_{L,opt} = \frac{(C_{glc,\mathrm{F}} + C_{xyl,\mathrm{F}})k_{xyl,CT} - (k_{xyl,AF} - k_{glc,AF})C_{xyl,\mathrm{F}}}{k_{glc,AF}k_{xyl,CT}X_{TOT}} \quad \textbf{(7.23)}$$

$$f_{opt} = \frac{(k_{xyl,AF} - k_{glc,AF})C_{xyl,\mathrm{F}}}{(k_{xyl,AF} - k_{glc,AF})C_{xyl,\mathrm{F}} - (C_{glc,\mathrm{F}} + C_{xyl,\mathrm{F}})k_{xyl,CT}} \quad \textbf{(7.24)}$$

which are the same expressions as for the batch culture. Thus, an optimal residence time and cell fraction for a continuous fermentation of mixed sugars by immobilized mixed cells of AFF01/pLI297 and CT1101/pLOI297 can be determined from a known concentration of sugars in the feed.

7.4 Results

7.4.1 Batch fermentation of hexoses and pentoses by immobilized cells

Fermentation of hexose and pentose sugars by immobilized cells of AFF01/pLOI297 and CT1101/pLOI297 was examined as shown in Fig. 7.2. The relation of consumed sugars shows that glucose, xylose and arabinose were utilized simultaneously by the mutant AFF01/pLOI297 compared to a sequential consumption of sugars by its parent MG1655/pLOI297. The mutant CT1101/pLOI297 also shows co-utilization of xylose and arabinose but it does not consume glucose. Ethanol yields by immobilized

AFF01/pLOI297 and immobilized CT1101/pLOI297 in these mixed sugar fermentations exceeded 94% of the theoretical yield.

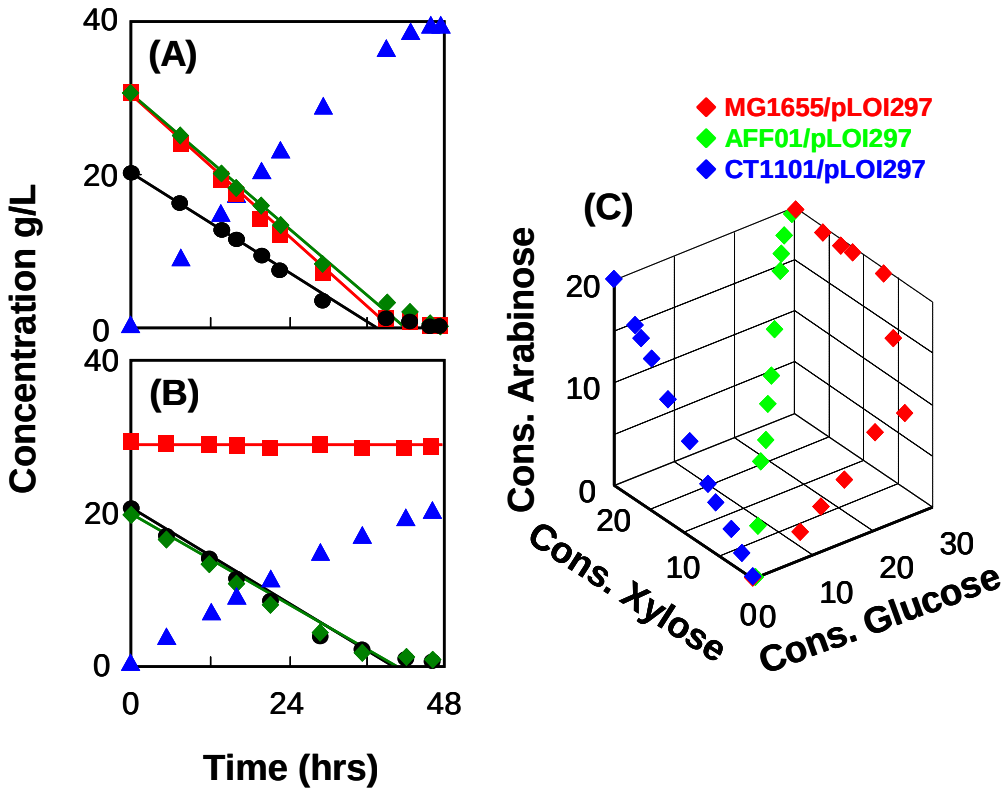


Figure 7.2 Fermentation kinetics of immobilized cultures. The experiments were conducted in anaerobic batch bioreactors containing 0.7 l semi-defined medium supplemented with sugars and with 0.35 l of beads. pH was controlled at 7. The immobilized cell concentration was 11.4 g-CDW/l-gel (see section 2.12). (A) Immobilized AFF01/pLOI297 with an initial concentration of 30 g/l glucose (■), 30 g/l xylose (◆), and 20 g/l arabinose (●) converted into ethanol (▲). (B) Immobilized CT1101/pLOI297 with an initial concentration of 30 g/l glucose, 20 g/l xylose and 20 g/l arabinose. (C) Relation of consumed sugars from a sugar mixture of glucose, xylose, and arabinose for immobilized cells of AFF01/pLOI297 (◆), CT1101/pLOI297 (◆) and wild-type MG1655/pLOI297 (◆).

7.4.2 Kinetics of hexose and pentose utilization by immobilized cells

The kinetics of sugar utilization was measured in more detail in individual batch experiments. Specifically, it was of interest to find out how the consumption rates

depend on the initial sugar composition in the medium. These experiments were not conducted under strict anaerobic conditions but in closed shake flasks. However, due to the high cell density the oxygen concentration in the medium was essentially zero. The results are summarized in Table 7.1 and Fig. 7.3. The data confirm the validity of Eq (7.8) that predicts on the basis of the assumed mechanisms that the zeroth order consumption rates depend on the initial mole fractions of the sugars

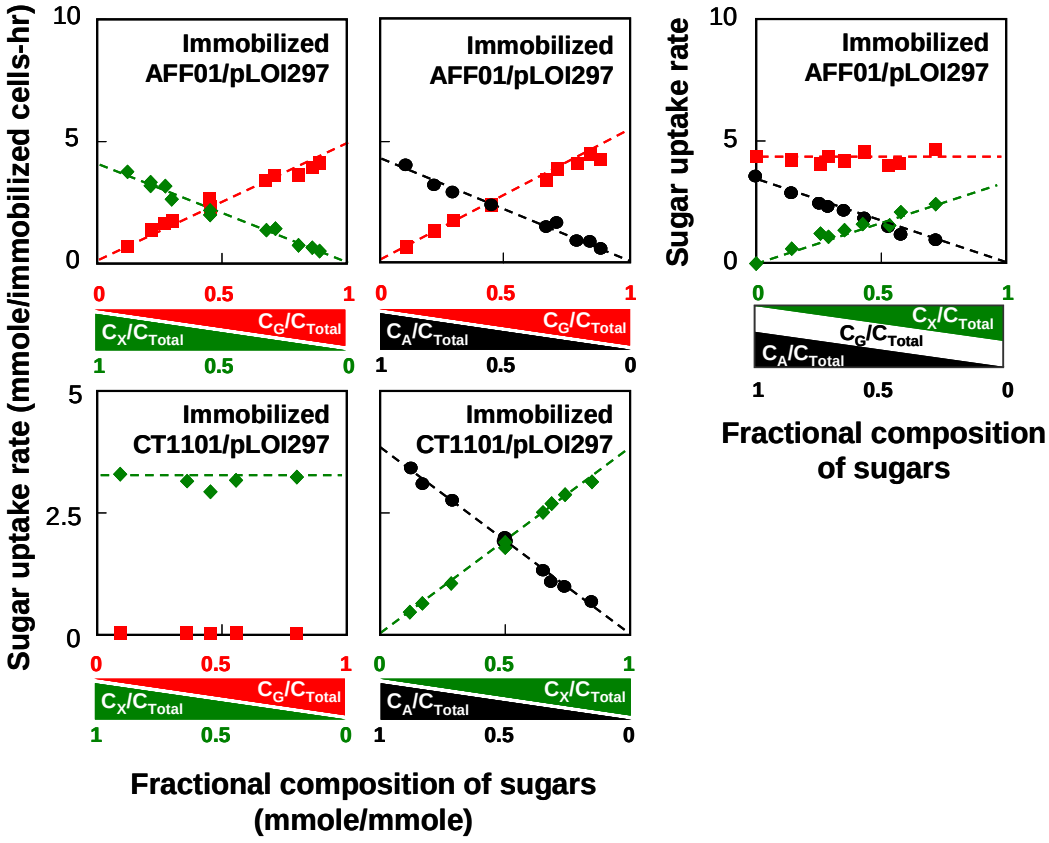


Figure 7.3 Rate of sugar consumption in mixed sugar fermentations of immobilized AFF01/pLOI297 and immobilized CT1101/pLOI297. Specific sugar uptake rates (mmole/g-CDW-hr) as a function of sugar composition in the mixture: glucose (■), xylose (◆), and arabinose (●). The sugar composition is given in mole fraction of total initial sugar concentration (mmole/mmole). The experiments were conducted in capped shake flasks containing 100 ml of semi-defined medium supplied with sugar mixture and with 35 ml beads containing immobilized cells of 16.5 g-CDW/l-

gel. The individual rate constants of each strain in each sugar mixture estimated from the linear regression of these data are summarized in Table 7.1.

Table 7.1 Individual rate constants k_{Si} of immobilized AFF01/pLOI297 and immobilized CT1101/pLOI297 in sugar mixtures. The rate constants are estimated from the linear regression of mixed sugars batch fermentation shown in Fig. 7.3[1]

Immobilized AFF01/pLOI297 in mixture of	$k_{glu,AF}$	$k_{xyl,AF}$	$k_{ara,AF}$
glucose-xylose	5.1 ± 0.1	4.1 ± 0.1	n.a.
glucose-arabinose	5.7 ± 0.1	n.a.	4.2 ± 0.3
glucose-xylose-arabinose	4.0 ± 0.3	3.0 ± 0.1	3.7 ± 0.1
Immobilized CT1101/pLOI297 in mixture of	$k_{glu,CT}$	$k_{xyl,CT}$	$k_{ara,CT}$
glucose-xylose	0.0 ± 0.0	3.4 ± 0.2	n.a.
xylose-arabinose	n.a.	3.8 ± 0.1	3.9 ± 0.1

[1] Rate constant is in unit of mmole/g-CDW-hr. The regression coefficient (R^2) for all plots are greater than 0.9.

7.4.3 Immobilized mixed cell culture for fermentation of hexoses and pentoses

Fig. 7.4A presents the relation of consumed glucose (20g/l) and xylose (60g/l) by immobilized single cells and immobilized mixed cells. The concentration of glucose and xylose chosen reflects the sugar concentrations typically found in lignocellulosic hydrolysates. Due to unequal amounts of glucose and xylose, the two sugars were not consumed at the same rate by the immobilized cell culture of AFF01/pLOI297 and glucose was exhausted first. The use of immobilized mixed cells of AFF01/pLOI297 and CT1101/pLOI297 permits the amount of each sugar consumed to be precisely adjusted by varying the cell composition of the two strains. As demonstrated in Fig. 7.4A,

synchronous consumption of both sugars was achieved when the cell fraction of CT1101/pLOI297 in the immobilized mixed culture was approximately at f = 0.27.

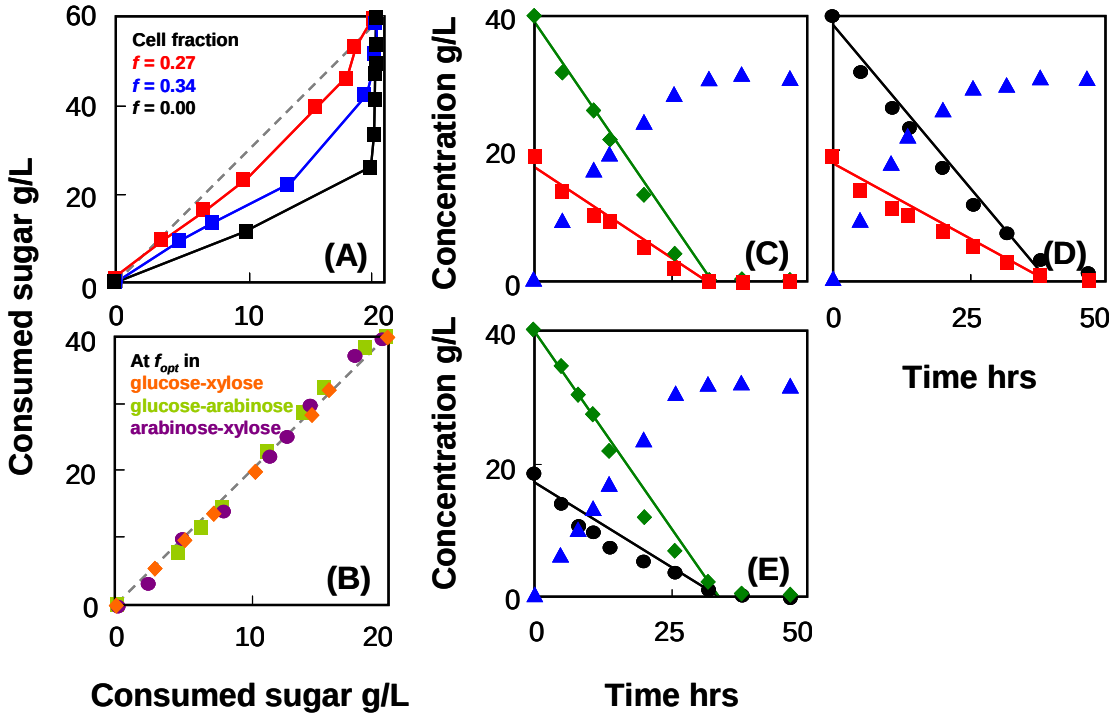


Figure 7.4 Fermentation kinetics by immobilized mixed culture of AFF01/pLOI297 and CT1101/pLOI297. The experiments were performed in capped shake flasks containing 100 ml of semi-defined medium supplied with sugar mixtures and with 35 ml beads. Immobilized cell concentration was 10.2 g-CDW/l-gel. Optimal cell fraction f_{opt} was determined from the kinetic model (see Eq (7.19)). (A) Relation of consumed xylose with an initial concentration of 60 g/l and consumed glucose with an initial concentration of 20 g/l at the fraction of CT1101/pLOI297 cells used in the immobilization: f = 0.27 (■), f = 0.34 (■), f = 0 (■). (B) Relation of consumed sugars in the mixture: glucose-xylose at cell fraction f_{opt} = 0.41 (◆), glucose-arabinose at cell fraction f_{opt} = 0.47 (■), and arabinose-xylose at cell fraction f_{opt} = 0.58 (●). (C) Immobilized mixed culture at the optimal cell fraction in a sugar mixture of glucose (■) and xylose (◆) converted into ethanol (▲). (D) Immobilized mixed cultures at the optimal cell fraction in a sugar mixture of glucose and arabinose (●). (E) Immobilized mixed cultures at the optimal cell fraction in a sugar mixture of xylose and arabinose.

Based on the kinetic model, we can predict with Eq (7.19) an optimal cell fraction f_{opt} of the two immobilized strains that will result in a simultaneous consumption of each sugar in a mixture. Batch fermentation kinetics of immobilized mixed cells of AFF01/pLOI297 and CT1101/pLOI297 operated at the predicted optimal cell fractions in a mixture of glucose-xylose, glucose-arabinose and arabinose-xylose are shown in Fig. 7.4C, 7.4D and 7.4E, respectively. In these immobilized mixed cell cultures, each sugar was consumed from the mixture at a rate proportional to the initial sugar composition in the mixture resulting in the exhaustion of all sugars at an equal fermentation time as predicted by Eq (7.19). The relation of consumed sugars by immobilized mixed cells at the optimal cell fraction in glucose-xylose, glucose-arabinose and arabinose-xylose mixtures is shown in Fig. 7.4B. The ethanol yield achieved by immobilized mixed cells exceeded 96% of theoretical yield.

7.4.4 Immobilized single cells vs. immobilized cell mixtures

The performance of immobilized single cells and immobilized mixed cells in glucose-xylose fermentations was tested and compared in Fig. 7.5. As expected, the immobilized cells of AFF01/pLOI297 show consumption of glucose and xylose at different times and all sugars are completely consumed within about 48 hrs. On the other hand, the immobilized mixed cultures of strains AFF01/pLOI297 and CT1101/pLOI297 simultaneously consumed glucose and xylose with both sugars exhausted at an equal fermentation time of approximately 40 hrs. The experimental results is in agreement with the model prediction in Fig. 7.1 which shows the immobilized mixed cells to consume

sugar mixtures more quickly resulting in a higher ethanol productivity than the immobilized single cells.

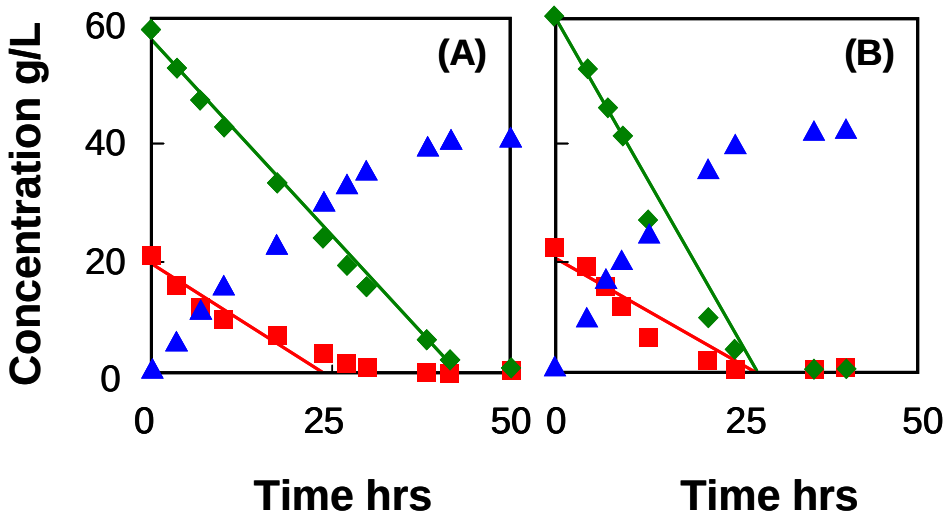


Figure 7.5 Mixed sugar fermentations by immobilized pure culture and immobilized mixed culture. The experiments were conducted in anaerobic batch bioreactors containing 0.7 l semi-defined medium supplied with sugars and with 0.35 l of beads at cell concentrations of 14.8 g-CDW/l-gel. pH was controlled at 7. (A) Immobilized pure culture of AFF01/pLOI297 with a sugar mixture of glucose (■) and xylose (◆) converted to ethanol (▲). (B) Immobilized mixed culture of AFF01/pLOI297 and CT1101/pLOI297 at a CT1101/pLOI297 cell fraction of f = 0.27 with a sugar mixture of glucose and xylose.

7.4.5 Continuous ethanol fermentation by immobilized mixed cells

The kinetic model was also applied to design an immobilized mixed culture of AFF01/pLOI297 and CT1101/pLOI297 for an optimized continuous ethanol fermentation from a mixture of glucose and xylose. The continuous culture of immobilized mixed cells was operated at an optimal residence time and cell fraction based on Eq (7.23) and (7.24). It should be noted that the optimal residence time and cell fraction is a function of sugar composition in the feed medium. The continuous cultures using immobilized mixed cells (Fig. 7.6A and 7.6B) supported a simultaneous conversion of glucose and xylose into

ethanol with a steady-state ethanol titer of approximately 34 g/l and ethanol yield of about 90% of theoretical maximum. Immobilized pure cultures of strain AFF01/pLOI297 (Fig. 7.6C) utilized the same initial sugar mixture less efficiently and at a slower overall rate than the immobilized mixed culture of AFF01/pLOI297 and CT1101/pLOI297. At steady state, only 72.3% of the total sugar was consumed by the immobilized AFF01/pLOI297, compared to 91.7% of the total sugar utilized by the immobilized mixed cells of AFF01/pLOI297 and CT1101/pLOI297. The ethanol productivity achieved by the immobilized pure culture and by the immobilized mixed culture was 1.6 g/l-hr and 2.2 g/l-hr, respectively. The results confirm that for mixed sugar utilization the process based on immobilized mixed cultures is superior to the process based on immobilized single cells. The immobilized mixed cell approach is also adaptable to any given feed composition of sugar mixtures by varying the operating residence time and the cells fraction of each immobilized strain. The kinetic parameters of continuous ethanol fermentation by immobilized single cells and cell mixtures are summarized in Table 7.2 and 7.3.

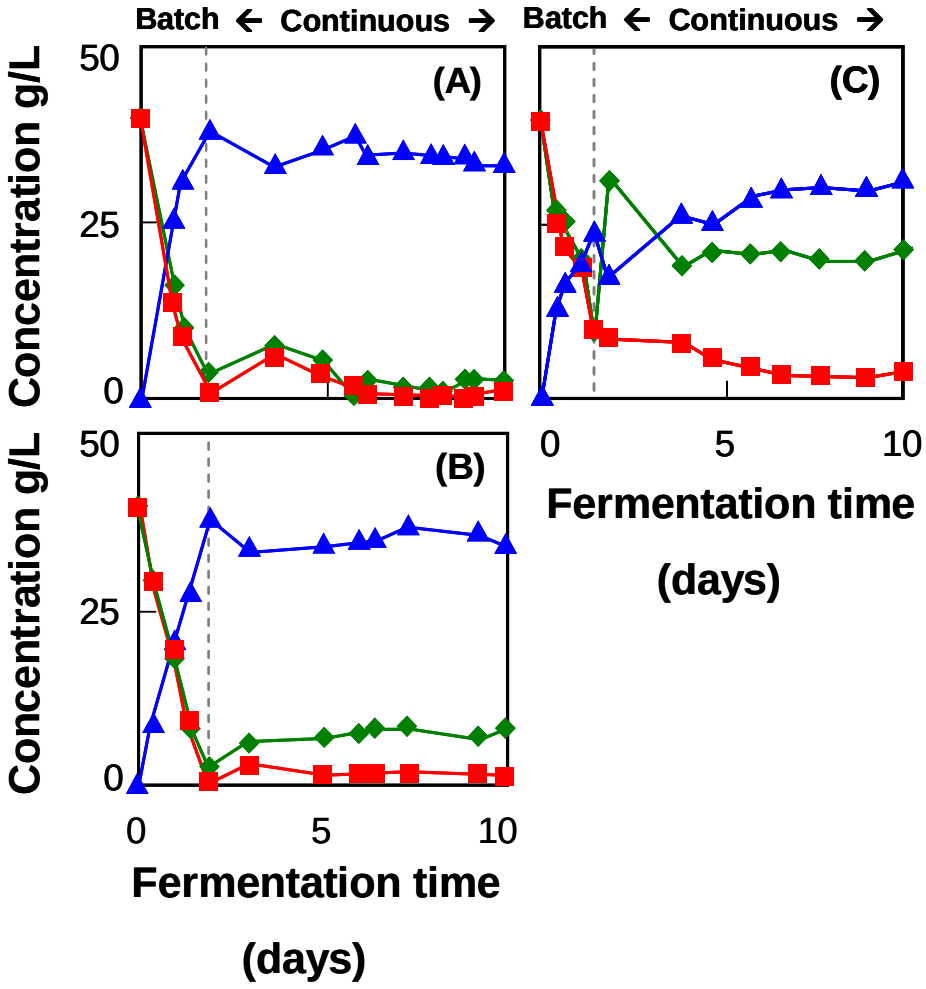


Figure 7.6 Continuous fermentation by immobilized pure culture and immobilized mixed cultures. The experiments were carried out in anaerobic batch bioreactor containing 0.7 l semi-defined medium supplemented with sugars and with 0.35 l of beads. Immobilized cell concentration was 11.4 g-CDW/l-gel. pH was controlled at 7. The immobilized mixed cultures were operated at optimal dilution rate and optimal CT1101/pLOI297 cell fraction (see Eq (7.22) and (7.23)). (A) Immobilized mixed culture of AFF01/pLOI297 and CT1101/pLOI297 at a cell fraction *f* = 0.32 with a feed sugar mixture of 40 g/l glucose (■) and 40 g/l xylose (◆) converted to ethanol (▲) at a dilution rate D = 0.096 hr^{-1}. (B) Immobilized mixed culture of AFF01/pLOI297 and CT1101/pLOI297 at a cell fraction *f* = 0.40 with a feed sugar mixture of 25 g/l glucose and 65 g/l xylose converted to ethanol at a dilution rate D = 0.059 hr^{-1}. (C) Immobilized pure culture of AFF01/pLOI297 with a feed sugar mixture of 25 g/l glucose and 65 g/l xylose converted to ethanol at a dilution rate D = 0.056 hr^{-1}.

Table 7.2 Performance summary of continuous ethanol production by immobilized cells of AFF01/pLOI297 and CT1101/pLOI297[1]

Strain/ pLOI297	Feed Medium g/l [2]	D_L hr^{-1}	% Sugar utilization	[ETOH] g/l	Y_{ETOH} (%)	Cell fraction *f*
AFF01	Xylose (65); Glucose (25)	0.056	72.3	30.1	90.2	-
CT1101+ AFF01	Xylose (65); Glucose (25)	0.059	91.7	36.7	87.2	0.40
CT1101+ AFF01	Xylose (40); Glucose (40)	0.096	96.0	34.0	86.6	0.32

[1] Immobilization conditions were approximately 4mm bead diameter and immobilized cell concentration of 11.4 g-cdw/l-gel.
[2] Feed sugar concentration (g/l) is shown in parenthesis. Cell fraction (*f*) represents a fraction of strain CT1101/pLOI297.

Table 7.3 Comparison of kinetic parameters of immobilized ethanologenic *E. coli* strains in continuous ethanol fermentation[1]

Strain/ pLOI297	Feed Medium g/l [2]	D_L hr^{-1}	[ETOH] g/l	Q_{ETOH} g/l-hr	Y_{ETOH} (%)	% Sugar utilization	Sources/ References
AFF01	Xylose (65); Glucose (25)	0.06	30.1	1.6	90	72	This study
CT1101+ AFF01	Xylose (65); Glucose (25)	0.06	36.7	2.2	87	91	This study
KO11	Xylose(50)	0.05	15.5	0.8	87	70	Zhou et al., 2008
FBR5	Xylose(50)	0.05	19	1.0	76	98	Martin et al., 2006
FBR5	Xylose (100)	0.02	37.1	0.8	98	74	Qureshi et al., 2005
FBR5	Xylose(100)	0.24	18.9	4.5	98	38	Qureshi et al., 2005

[1] Immobilization conditions were approximately 4mm bead diameter and immobilized cell concentration of 11.4 g-cdw/l-gel.
[2] Feed sugar concentration (g/l) is shown in parenthesis.

7.4.6 Operational stability of immobilized cells

For a viable industrial application a long term process should have a stable productivity at the highest possible conversion efficiency. Several recombinant ethanologenic *E. coli* strains have shown loss or diminishing production of ethanol during long-term cultivation (Dumsday et al., 1999; Martin et al., 2006). The stability and the efficiency of conversion of immobilized mixed cultures of AFF01/pLOI297 and CT1101/pLOI297 were tested in a long-term process of repeated batch fermentations. The repeated batch process was operated as batch culture in an anaerobic bioreactor containing LB medium supplied with 20 g/l glucose and 60 g/l xylose. Time profiles of pH value during each batch process were monitored. A sudden rise of the pH profile (data not shown) indicated a depletion of sugars in the culture marking the end of each batch cycle. After each cycle the culture medium was harvested, the immobilized cells were washed with 0.1% $CaCl_2$ solution and resuspended in fresh medium. The kinetics of sugar consumption is shown in Fig. 7.7. The system was stable for 12 batch cycles performed over 3 weeks. It did not show any significant loss in yield or productivity over this time period and could have been continued still longer.

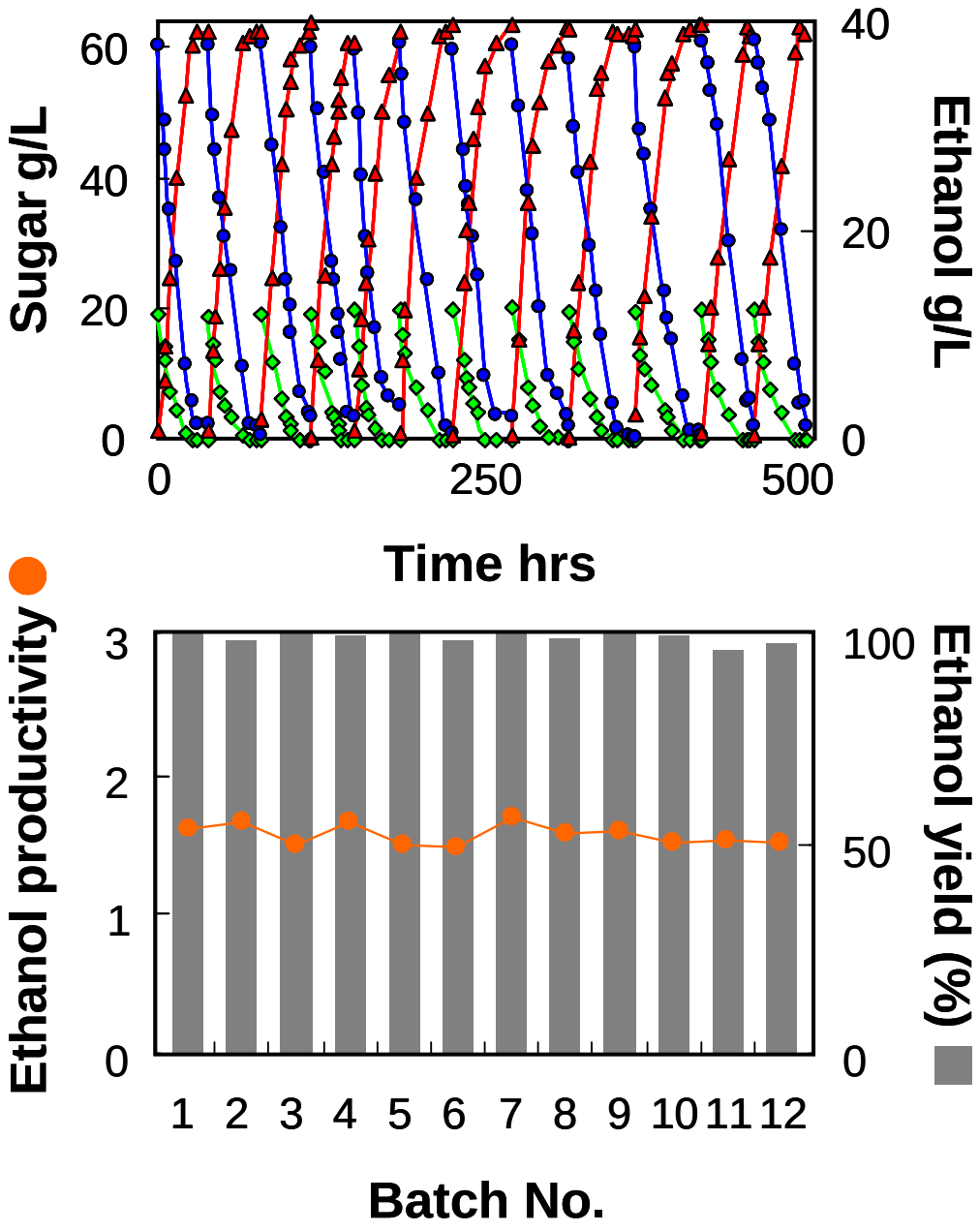


Figure 7.7 Immobilized mixed culture of AFF01/pLOI297 and CT1101/pLOI297 in a repeated batch fermentation. The experiment was conducted in anaerobic batch bioreactor containing 0.7 l of LB medium supplied with an initial concentration of 20 g/l glucose and 60 g/l xylose and with 0.35 l of beads. Immobilized cell contained 11.4 g-CDW/l-gel and a CT1101/pLOI297 cell fraction f = 0.27. (Top) Time profiles of residual glucose (◆), xylose (●), and produced ethanol (▲). (Bottom) Ethanol yield in % of theoretical and ethanol productivity in g/l-hr obtained in each batch run.

7.5 Discussion

We have previously designed and constructed two mutant *E. coli* strains where each strain is selective in the consumption of hexose and/or pentose for ethanol production at yields close to the theoretical maximum (Trinh et al., 2008). In this study, we have implemented an immobilized cell system for the conversion of sugar mixtures into ethanol using the two strains AFF01/pLOI297 and CT1101/pLOI297. Immobilization of mixed cultures of the two strains has proven to be an ideal condition for an efficient and simultaneous conversion of sugar mixtures into ethanol. The advantage of this process is that the two strains can act in concert in consuming mixed sugars resulting in a faster utilization of a sugar mixture than a single strain could convert. The process is adjustable to any sugar composition in the medium by varying the cell fraction of each strain in the immobilization process. This was confirmed in both batch and continuous fermentations of hexose and pentose mixtures where we achieved high ethanol yield and productivity.

The developed kinetic model of mixed sugar utilization and the experimental data show that individual rates of sugar uptake in a sugar mixture is dependent on the mole fraction of the sugar in the medium. This dependence is a consequence of a set of multiple sugars competing for a common step in the initial metabolism. A sugar with higher mole fraction has a higher chance of binding with the transport complex leading to a higher uptake rate than a sugar with lower mole fraction. The uptake rate of each sugar follows a zero-order kinetic since the high sugar concentration used in the experiment approximately 400-500 mM is much higher than the Km value for glucose (1-2 μM) of the transporter protein (Natarajan and Srienc, 1999). The zero-order kinetic results in a

linear relationship between the specific consumption rate of each sugar and its mole fraction in the mixture which is in good agreement with the batch experiments of immobilized AFF01/pLOI297 and immobilized CT1101/pLOI297 as shown in Fig. 7.3. It was also observed that the sugar mole fraction in a sugar mixture remained relatively constant throughout the course of fermentation. Thus, the rates of uptake can be estimated from the mole fraction of the initial sugar concentration. The validity of the kinetic model is demonstrated by the fact that it can be used to accurately predict an optimal cell fraction of immobilized mixed cells for synchronous consumption of sugar mixtures.

The use of minimal media compared to complex media could offer both practical and economic advantages in a continuous ethanol production process. Although we have conducted the repeated batch experiments in LB media supplemented with the individual sugars, we have conducted tests with simple M9 medium without phosphate buffer that can be used in the immobilized cell culture process. The phosphate buffer has been omitted since it causes destruction of alginate gel beads. We observed similar high ethanol yields as with complex LB medium. However, the ethanol productivity is significantly reduced when minimal medium is used. The productivity can be recovered with the addition of 5 g/l yeast extract which is the semi-defined medium we used in the continuous culture experiments.

Controlling the uptake rate of sugar mixtures in a freely-suspended mixed cell culture could be challenging since the cell distribution could change over time as each strain grows at a different rate. In contrast, the cell distribution of an immobilized mixed cell

culture is expected to remain relatively constant since the growth of immobilized cells is restricted. As a result, the process of an immobilized mixed cell culture permits a better control of the consumption rate of each sugar in a mixture than a freely-suspended mixed cell culture. In the conducted experiments the two strains were mixed before immobilization. Thus, each gel bead contained a mixture of both strains. In principle, one could also immobilize pure cultures of the respective strains and then mix desired proportions of the beads containing homogeneous cell populations for use in the conversion reactor. In the latter case one should expect higher local conversion rates in individual beads that could result in mass transfer limitations of individual sugars into the beads.

We have evaluated the internal mass transfer effects through computation of the effectiveness factor (Shuler and Kargi, 1992) which measures the degree of internal mass transfer resistance relative to the reaction rate. It was found that under the condition of immobilization examined in this work (bead size of about 4mm in diameter and cell loading of approximately up to 20 g-CDW/l-gel), the effectiveness factor is close to 1. Thus, internal mass transfer limitation is negligible in our case and the immobilization of pure cultures with subsequent mixing of beads is expected to result in the same performance as the immobilization of cultures that are mixed at the desired proportions.

The gel bead inoculum of 50% (v/v) used in this work represents an optimized condition where high cell loading has been achieved with minimal splitting of gel beads during mixing. During the course of fermentation, little cell leakage was observed. The suspended cell concentration was below $OD_{600} \leq 0.5$ (equivalent to about 0.1 g-CDW/l)

suggesting that most of the cells remain entrapped inside the alginate gel beads. The extent of cell leakage increased if a higher cell loading was used. We observed that the maximum cell loading with minimal cell leakage is approximately 15 g-CDW/l-gel volume. In addition to the fluidized bed reactor, we tested also a packed bed reactor. But this resulted in problems because the evolving CO_2 could not be efficiently removed.

In this work, the experiments were operated at a low dilution rate to maximize utilization of sugar mixture. The process performance of other immobilized-ethnaologenic *E. coli* reported in literatures (Table 7.3) is, however, based on single sugar fermentation. The performance is expected to be reduced if a mixture of sugars is used due to catabolite inhibition. Thus, our immobilized mixed cell system should perform better than these other systems since it is an optimized system for mixed sugar fermentation.

In the continuous operation a lower ethanol yield was observed (about 90% of theoretical yield) compared to the repeated batch operation (> 95% of theoretical yield). This is likely contributed by the continuous exposure of immobilized cells to high concentrations of ethanol during the continuous operation. In contrast, the immobilized cells during repeated batch operation are exposed to high ethanol concentrations only towards the end of each cycle when the sugar source is exhausted. In addition, it is likely that cells can regenerate during the initial time during a repeated batch cycle when cells are exposed to fresh medium. Inhibitory effects of ethanol on the cells could also explain the incomplete utilization of sugar mixtures (Fig. 7.6) in the continuous culture of the immobilized mixed cell cultures. The continuous ethanol fermentation using immobilized

mixed strains culture could potentially be extended to a more complex sugar mixture besides the two-sugar mixture of glucose and xylose demonstrated in this study. The immobilized mixed cell cultures exhibited a high degree of operational stability when it was repeatedly used for several weeks without any loss of viability or ethanologenicity. The observed stability of immobilized cells may be contributed due to the slowly- or non-growing state of cells when cells are immobilized (Barbotin et al., 1998; Chau et al., 2000; Dincbas et al., 1993; Najafpour, 1990). With the restricted growth of immobilized cells, the loss of genes or intracellular enzyme activity is minimized.

In summary, ethanol production from hexoses and pentoses by immobilized mixed cultures of *E. coli* presented in this study represents an example of a case where a cell consortium results in a superior process performance than what can be achieved with homogeneous cell populations. The approach permits a precise adjustment of sugar consumption times which could be useful in the fermentation of sugar mixtures derived from lignocellulosic biomass which is known to fluctuate in the sugar contents. In addition, immobilization offers an effective means for a robust and stable continuous ethanol production process over extended time periods.

Chapter 8

Optimization and evolution of *Thermoanaerobacterium saccharolyticum* metabolic pathways for an efficient production of ethanol

8.1 Chapter summary

In this chapter, a metabolic network model of *Thermoanaerobacterium saccharolyticum* was constructed and analyzed through elementary mode analysis. Specific gene deletions that limit the cell's metabolic functionality to four sets of the most efficient ethanol-producing pathways were identified. The designed strain containing predicted gene deletions would, therefore, produce ethanol at the most efficient yield conversion. Gene deletions were implemented, resulting in a knockout mutant AS411. The constructed mutant secreted ethanol at increased yield compared to the wild-type, confirming the model prediction. The mutant was then evolved in a serial dilution experiment. Using cell growth as selective pressure, we isolated an evolved cell AS411E3 which exhibited an increase in ethanol productivity compared to its parent strain. The experimental observations validate the combined approach of elementary mode analysis and metabolic pathway evolution as an effective metabolic engineering tool in creating strain with improved yield and productivity of product of interest.

8.2 Introduction

An *in silico* model of cellular metabolism has proved to be an effective tool in understanding the cell physiology under perturbations imposed by genetic modifications and/or growth conditions. A large number of works has shown the application of metabolic model and metabolic network analysis to optimize the strain for phenotypic

improvements (Alper et al., 2005; Fong et al., 2005; Wlaschin et al., 2005). One of the most promising tools emerging for the analysis of cell metabolic network is elementary mode analysis (EMA). EMA decomposes the entire metabolism of a cell into a set of unique and indivisible pathways called elementary modes. Each mode represents a stoichiometrically balanced pathway operated under steady state condition (Pfeiffer et al., 1999; Schuster et al., 2000). An elementary mode or a combination of modes represents a phenotypic state of a cell (Carlson and Srienc, 2004). Thus, these elementary flux modes are considered a fundamental determinant of cell physiology. Analysis of the metabolic pathways by EMA permits a quantitative evaluation of the structure and robustness of cellular metabolism (Papin, et al., 2002; Stelling, et al., 2002). As demonstrated in previous chapters, EMA has been used for examining the capabilities of metabolic networks, studying the effect of genetic or environmental perturbations and guiding the rational metabolic engineering of a cell to improve its performance (Carlson and Srienc, 2004; Trinh, et al., 2004; Trinh, et al., 2008; Unrean et al., 2010). In this present work, EMA was applied to analyze metabolic network of *Thermoanaerobacterium saccharolyticum* for the production of ethanol from biomass derived sugars.

Ethanol produced from cellulosic biomass has become the principal candidate for transportation fuels (Ingram et al., 1999; Farrell et al., 2006; Ragauskas et al., 2006). One of the challenges in cellulosic biofuel industry is a high cost of converting biomass to sugars. The use of cellulolytic bacteria for biofuel production could provide a substantial cost advantage for the biomass conversion. *T. saccharolyticum* JW/SL-YS485, which is a thermophilic saccharolytic anaerobe, is a promising biological catalyst for the production

of ethanol from biomass. The organism can ferment the majority of biomass-derived sugars (Liu et al., 1996). *T. saccharolyticum* can be used in a consolidated bioprocessing of lignocellulosic biomass where the production of saccharolytic enzymes, the hydrolysis of biomass to sugars and the fermentation of those sugars to ethanol, is conducted by the organism in a single step (Lynd et al., 2002; Lynd et al., 2005; Lynd et al., 2008).

The metabolic engineering of *T. saccharolyticum* for an efficient cellulosic ethanol production has received extensive attention and effort (Mai and Wiegel, 2000). Previous attempts of improving ethanol production in *T. saccharolyticum* have relied on classical mutagenesis techniques (Lynd et al., 2002) or trivial genetic manipulations, such as deletion of competitive products (Desai et al., 2004; Shaw et al., 2008). This strategy, however, may not capture all genetic modifications needed for an efficient production of ethanol. Identification of the target gene manipulations needed for creating an efficient ethanol producing strain of *T. saccharolyticum* is challenging due to the complexity of the cell metabolic network.

We have pursued, in this study, an approach of inverse metabolic engineering guided by EMA to create a mutant strain of *T. saccharolyticum* that can produce ethanol efficiently. By applying EMA, we have examined the metabolic network structure of *T. saccharolyticum* and identified a set of all unique elementary modes of the entire network. The metabolism of *T. saccharolyticum* is optimized for optimal ethanol production from glucose and/or xylose by limiting its metabolic functionality only to efficient ethanol-producing pathways via gene knockout mutations. A set of five target gene deletions was identified which limited the cell's metabolic functionality to four sets

of efficient ethanol-producing pathways. The performance of the constructed strain AS411 with these gene knockout mutations is presented and discussed in this study. In addition, directed evolution of metabolic pathways was applied for improving ethanol productivity in the constructed mutant.

8.3 Results

8.3.1 Metabolic network analysis of *T. saccharolyticum*

A metabolic network model of *T. saccharolyticum,* including a total of 63 metabolic reactions and 63 metabolites (Fig. 1.2), has been constructed. The model contains a total of 4,336 possible pathways (elementary modes) for growth on glucose and 4,714 pathways for growth on xylose. The large number of elementary modes illustrates the flexibility of the cell to adapt itself to particular conditions by using pathways that provide the optimal fitness. Of these modes, 2,657 modes represent the ethanol synthesis pathways for glucose, and 3,336 modes represent the pathways for xylose (Table 8.1).

Table 8.1 Number of available modes including total modes, ethanol producing modes and biomass-ethanol producing modes, and ethanol yield for wild-type and mutant AS411

Xylose substrate	**Total EMs**	**ETOH EMs**	**Bio-ETOH EMs**	Y_{ETOH}[1]
Wild-type	4,714	3,366	1,000	0-0.51
AS411	42	42	24	0.28-0.51
Glucose substrate	**Total EMs**	**ETOH EMs**	**Bio-ETOH EMs**	Y_{ETOH}
Wild-type	4,336	2,657	1,114	0-0.51
AS411	42	42	24	0.28-0.51

[1] Yield is shown in g-ethanol produced per g-sugar consumed.

Differently than other ethanologenic organisms such as *S. cerevisiae* or *E. coli*, *T. saccharolyticum* synthesizes ethanol through either pyruvate-ferredoxin oxidoreductase (POR) or pyruvate-formate lyase (PFL). The POR uses ferredoxin as the main electron carrier while PFL transfers electrons to formate for the conversion of pyruvate to acetyl-CoA. The cofactor ferredoxin is recycled through hydrogenase reaction (ECH) or NAD(P)H-ferredoxin oxidoreductase reaction (FNOR). An acetyl-CoA precursor is then converted to ethanol via acetaldehyde dehydrogenase (ALD) and alcohol dehydrogenase (ADH), which use either NAD or NADP as electron carriers. The differences in the types of precursors and cofactors of the two ethanol synthesis pathways can affect the final possible yield of ethanol. Based on EMA, achieving maximum ethanol yield requires the POR-FNOR pathway.

8.3.2 Redirecting *T. saccharolyticum* metabolism for optimal ethanol production

EMA is applied to identify efficient pathways for ethanol synthesis from xylose or glucose in *T. saccharolyticum*. The metabolism of *T. saccharolyticum* is, then, restrained to function according to these optimal pathways through implementation of gene knockouts. Based on the knowledge of all pathway options, elementary mode analysis is systematically used to evaluate gene deletion effects on the yield of ethanol. The targeted gene knockouts that result in the maximal yield of ethanol are searched for using a sequential search strategy described in Chapter 3. Such a search is based on the aim at eliminating inefficient ethanol synthesis pathways, minimizing the number of remaining modes while maximizing the yield of ethanol and supporting growth. The analysis assists the process of screening and selecting multiple knockout targets for the maximal

production of ethanol. Based on the search algorithm, the suggested target genes for deletion are PPP1 (*zwf*), MGS1 (*mgs*), LDH (*ldh*), PFP3r (*sbm*), and PTA and ACK (*pta-ack*) as shown in Table 8.2.

Table 8.2 Gene knockout targets and correspondence of deleted genes to deleted reactions

Deleted Reaction	Corresponding gene	Enzyme
LDH	*ldh*	lactate dehydrogenase
PTA	*pta*	phosphate acetyltransferase
ACK	*ack*	acetate kinase
MGS1	*mgs*	methylglyoxal synthase
PPP1	*zwf*	glucose 6-phosphate-1-dehydrogenase
PFP3r	*sbm*	methylmalony CoA mutase

The total number of available modes and ethanol yield after each gene deletion are shown in Fig. 8.1. A combination of these deleted genes limits the metabolism of *T. saccharolyticum* to 42 elementary modes that can efficiently convert glucose or xylose to ethanol with a predicted yield range of 0.28-0.51 g-ethanol/g-sugar (Table 8.3). These 42 pathways of the mutant can also be collapsed into four families of elementary modes based on their overall reaction stoichiometry. The strain with these gene deletions is expected to function by redistributing its pathway fluxes within these four families of modes. Hence, the designed strain should produce ethanol more efficiently than the wild-type, which can operate using modes that do not produce ethanol. It is interesting to observe that the knockout can achieve the same ethanol yield from both xylose and

glucose. All remaining elementary modes in the designed strain also indicate a tight couple between cell growth and ethanol production in the designed strain.

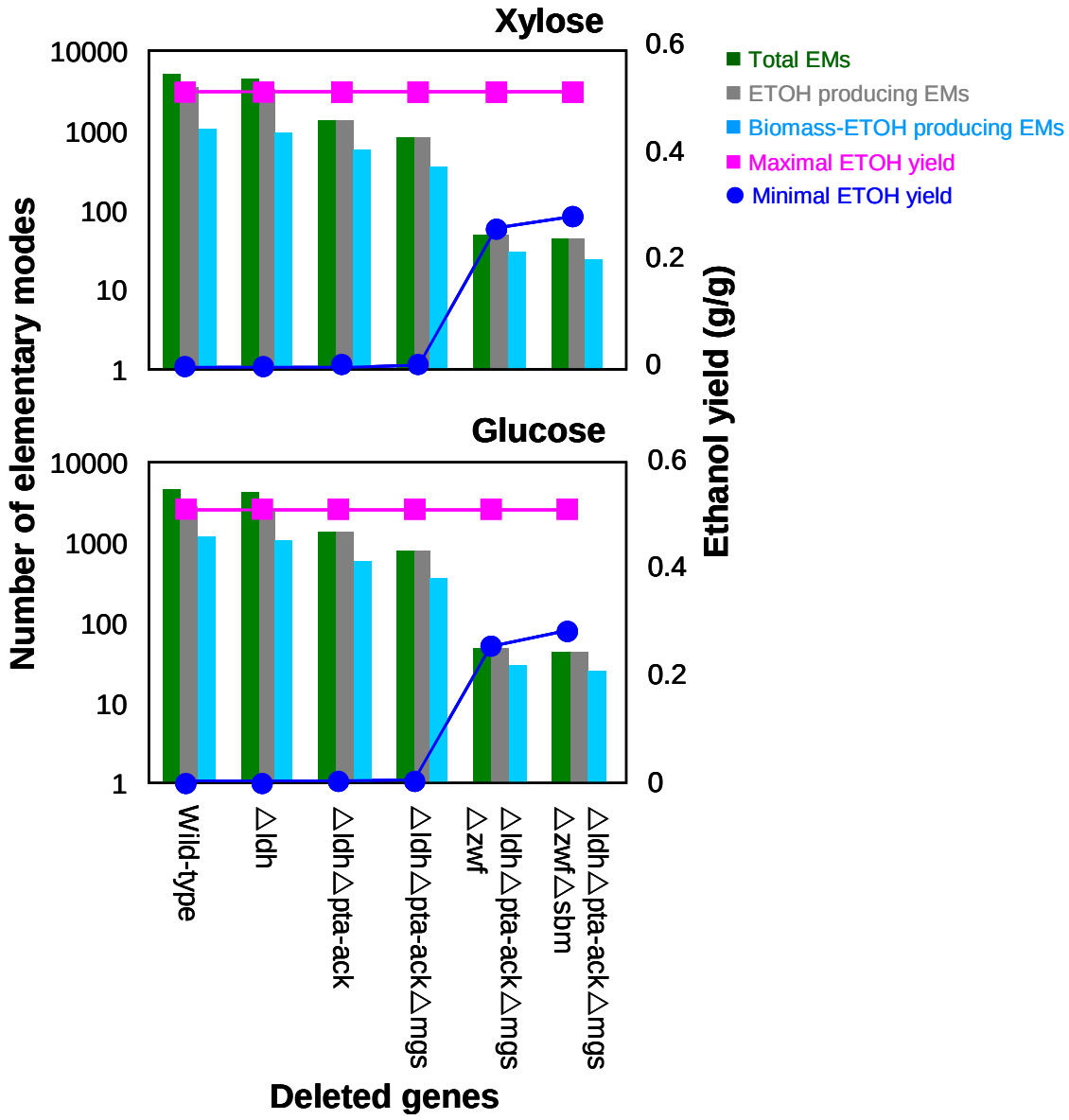


Figure 8.1 Effect of gene deletions on the number of anaerobic elementary modes for growth on xylose (top) and on glucose (bottom). The bars specify the number of elementary modes; total modes; modes that make ethanol; modes that make both biomass and ethanol for strains with deletions of indicated genes. The symbols are the possible maximal and minimal ethanol yields for the strains. The number of modes progressively decrease while the minimal yields are pushed towards the upper theoretical limit with increasing number of deleted genes.

Table 8.3 Stoichiometric equations for AS411 grown on xylose and glucose. All remaining biomass-producing modes have cell growth and ethanol production coupled.

Family	Reaction stoichiometry	No. of EMs	Y_{ETOH}[1]	Y_{BIO}[1]
Xylose substrate				
1	Xylose → 1.67 Ethanol + 1.67 CO_2 + 1.67 ATP	3	0.51	0
2	Xylose → 1.67 Ethanol + 1.67 CO_2	15	0.51	0
3	Xylose + 0.59 NH_3 → 2.40 Biomass + 0.09 H_2 + 0.92 Ethanol + 0.98 CO_2	18	0.28	0.41
4	Xylose + 0.59 NH_3 → 2.40 Biomass + 0.09 Formate + 0.92 Ethanol + 0.89 CO_2	6	0.28	0.41
Glucose substrate				
1	Glucose → 2 Ethanol + 2 CO2 + 2 ATP	3	0.51	0
2	Glucose → 2 Ethanol + 2 CO2	15	0.51	0
3	Glucose + 0.70 NH3 → 2.88 Biomass + 0.11 H2 + 1.10 Ethanol + 1.18 CO2	18	0.28	0.41
4	Glucose + 0.70 NH3 → 2.88 Biomass + 0.11 Formate + 1.10 Ethanol + 1.07 CO2	6	0.28	0.41

[1] Yield is shown in g-product produced per g-sugar consumed.

8.3.3 Construction of efficient strain

The mutant of *T. saccharolyticum* with predicted gene deletions was constructed using the direct single chromosomal gene disruption methods (Mai and Wiegel. 2000; Desai et al., 2005). Briefly, the target gene for deletion is being replaced with kanamycin resistance marker through the first homologous recombination. The kanamycin marker is later removed through the second homologous recombination, and the strain is used as a host for construction of additional gene knockouts. The previously constructed strain M355 containing the deletion of the chromosomal *ldh* gene and *pta-ack* gene (Shaw et

al., 2008) was used as the background host for construction of the designed mutant. The mutant was named AS411, and contained the additional deletion of chromosomal gene *mgs* and *zwf* (Fig. 8.2).

The predicted target gene *sbm*, which is part of the propionic acid synthesis pathway, is assumed to be inactive because propionic acid has not been detected in the fermentation medium of AS411. Therefore it has not been deleted. Gene knockouts in AS411 were verified via PCR amplification using the wild-type as positive control. Outside and inside primers are located on undeleted portions and deleted portions of the structural gene respectively. When compared with the wild-type, the absence of bands with inside primers and reduced band sizes with outside primers for AS411 confirmed the gene knockouts in the strain.

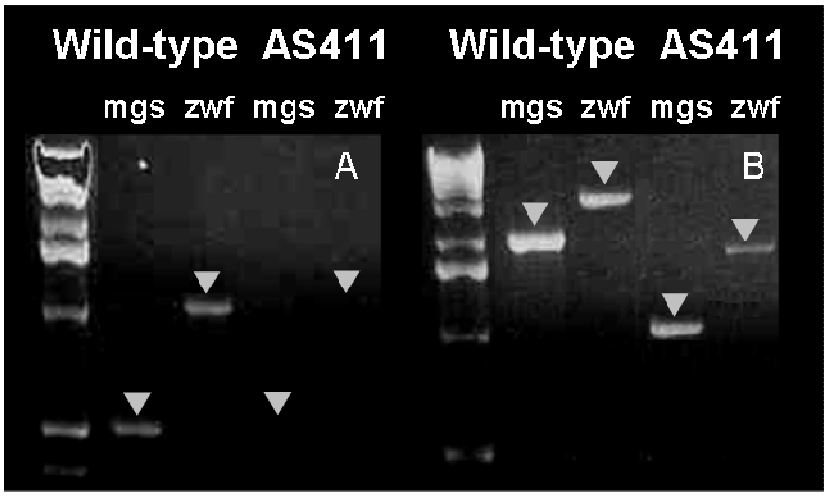


Figure 8.2 PCR test for deleted genes in AS411 using the wild-type as a positive control. Inside primers (A) and outside primers (B) of individual genes are used. Target gene deletion (*mgs* and *zwf*) are listed. The arrow shows the band location of the tested gene. PCR product is absent for the inside primers, if the gene is deleted. A shift to a smaller size of PCR product for the outside primers is observed, if the gene is deleted.

8.3.4 Strain characterization

Anaerobic batch bioreactors (Fig. 8.3) showed that AS411 achieved ethanol yield on glucose of 0.42±0.02 g/g, a 1.4-fold increase compared to the yield obtained by wild-type (0.30±0.01 g/g). For growth on xylose, AS411 reached the ethanol yield of 0.43±0.01 g/g compared to the yield of 0.41±0.02 g/g by the wild-type. Cell yield of the mutant was reduced, compared to the wild-type. The reduction in cell yield is proportional to the increase in ethanol yield, suggesting that the two products are competitive. The growth and ethanol phenotype of both strains is summarized in Table 8.4. The observed yields of both wild-type and mutant are within the predictive yield ranges confirming the accuracy of the metabolic network model.

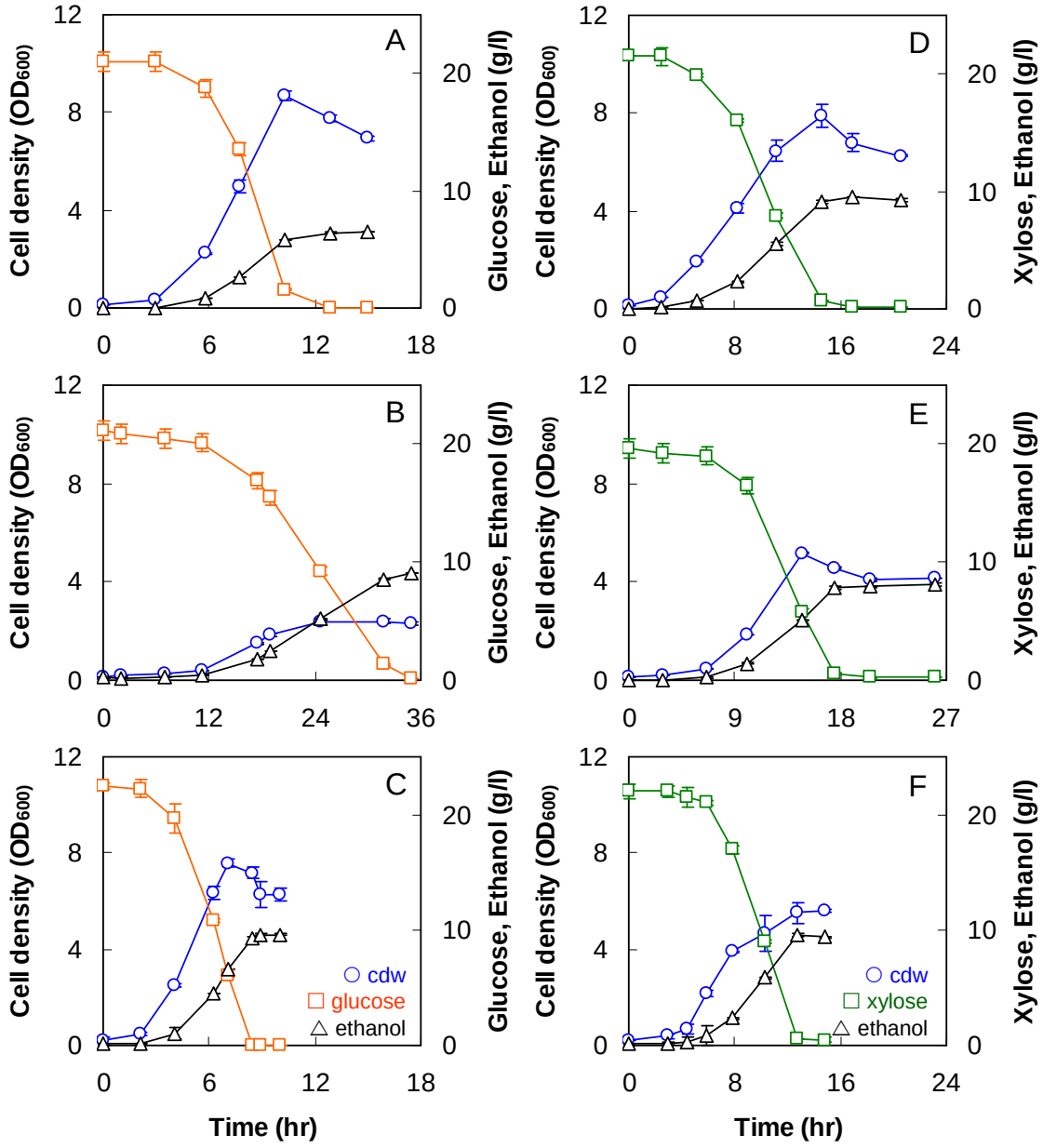


Figure 8.3 Fermentation time profiles for glucose and xylose, cell density and ethanol for the wild-type (A,D), the mutant AS411 (B,E) and the evolved mutant AS411E3 (C,F) in anaerobic pH-controlled bioreactor containing MTC medium and sugar. Symbols are ○ for cell concentration OD_{600}; □ for glucose; □ for xylose; △ for ethanol.

Fermentation kinetics showed the wild-type to ferment all glucose within 13 hrs, while the mutant took 35 hrs. This resulted in a 3.9 fold decrease in productivity by the mutant, compared to that of the wild-type. The wild-type also completely utilized xylose in 17 hrs, compared to 21 hrs for completed xylose utilization by the mutant. It is interesting to observe that AS411 grew more slowly on glucose than on xylose. Unlike the wild-type where mixed fermentative products were observed, the mutant produced ethanol as its major product. Growth and ethanol phenotype for both strains were summarized in Table 8.4. Strains were also tested for the production of ethanol on cellobiose (a disaccharide of glucose) and xylan (a polysaccharide of xylose), which are the products of hydrolysis of cellulose and hemicellulose respectively. Fermentation results (Fig. 8.4) also showed the mutant AS411 to reach higher yield of ethanol than wild-type when cultured in these substrates.

Table 8.4 Summary of cell growth and ethanol synthesis of wild-type, AS411 and AS411E3 cultured in xylose and glucose[1]

	Wild-type	AS411	AS411E3
Xylose substrate			
Growth hr^{-1}	0.42 ± 0.01	0.33 ± 0.00	0.50 ± 0.01
Ye/s g/g	0.41 ± 0.02	0.43 ± 0.01	0.44 ± 0.01
Yx/s OD600/g	0.36 ± 0.03	0.27 ± 0.01	0.26 ± 0.01
Glucose substrate			
Growth hr^{-1}	0.52 ± 0.01	0.20 ± 0.01	0.62 ± 0.04
Ye/s g/g	0.30 ± 0.01	0.42 ± 0.02	0.43 ± 0.01
Yx/s OD600/g	0.41 ± 0.01	0.12 ± 0.00	0.34 ± 0.01

[1] Experiments were conducted in anaerobic bioreactor fermentation. Product yields were calculated based on total sugar consumed. Ye/s and Yx/s represent yield of ethanol and cell mass respectively.

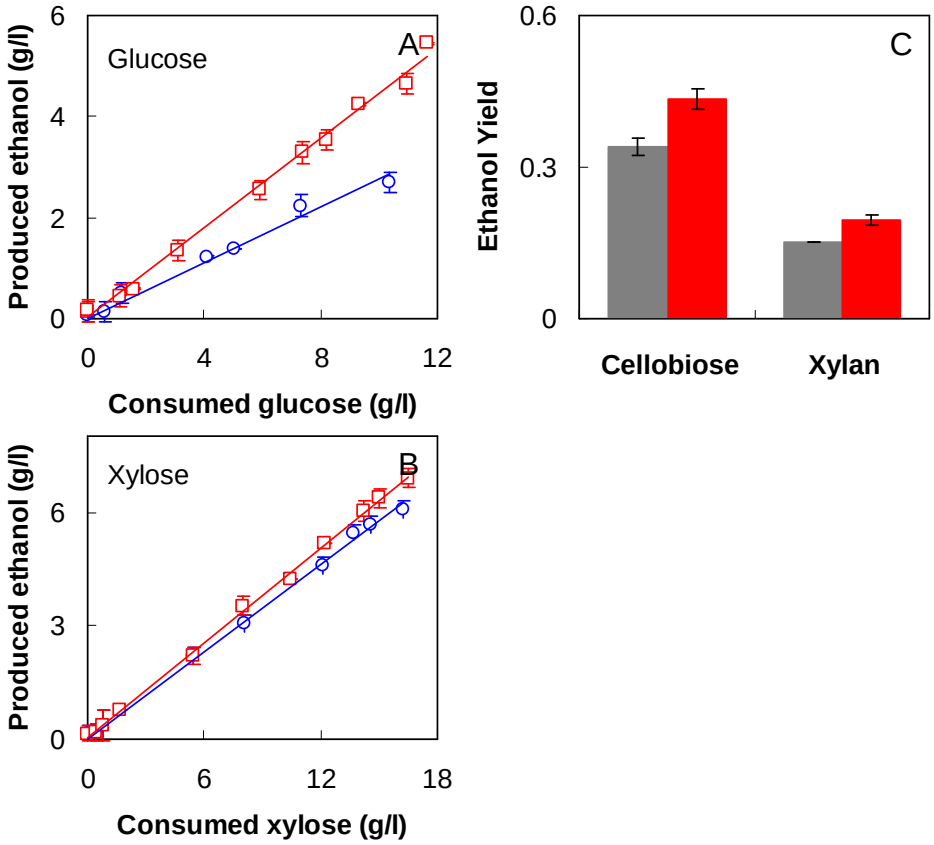


Figure 8.4 Yield of ethanol by the wild-type and AS411 in batch shake-flask experiments. (A, B) Relationship of ethanol vs. consumed sugar of AS411 (red) and wild-type (blue). Slopes of the regression lines define overall ethanol yield on sugar. Overall ethanol yields on glucose were $Y_{E/S}$ = 0.26±0.01 g/g (wild type) and $Y_{E/S}$ = 0.44±0.02 g/g (AS411). Overall ethanol yields on xylose are $Y_{E/S}$ = 0.38±0.01 g/g (wild type) and $Y_{E/S}$ = 0.42±0.01 g/g (AS411). (C) Ethanol yield on cellobiose and on xylan by wild-type (grey) and mutant (red). Yield of ethanol on cellobiose was calculated based on total consumed cellobiose. Yield of ethanol on xylan was calculated based on total xylan added.

8.3.5 Metabolic evolution

In the constructed mutant AS411, the formation of ethanol and cell growth is strictly coupled based on the available modes (Table 8.3). The coupling is, in part, a result of balancing cellular redox potential. In AS411, the reducing and oxidizing equivalents generated from the production of biomass can only be disposed through ethanol synthesis

(Fig. 8.5A). The synthesis of biomass creates the imbalance of cellular redox potential NAD/NADH. EMA indicated that AS411 can only relieve this imbalance by using ethanol as an electron sink. Similarly, the redox equivalent NADP/NADPH generated during cell growth of AS411 is balanced by the production of ethanol using ferredoxin as an electron carrier. This is not the case for the wild-type, since the wild-type can use other products besides ethanol as an electron sink.

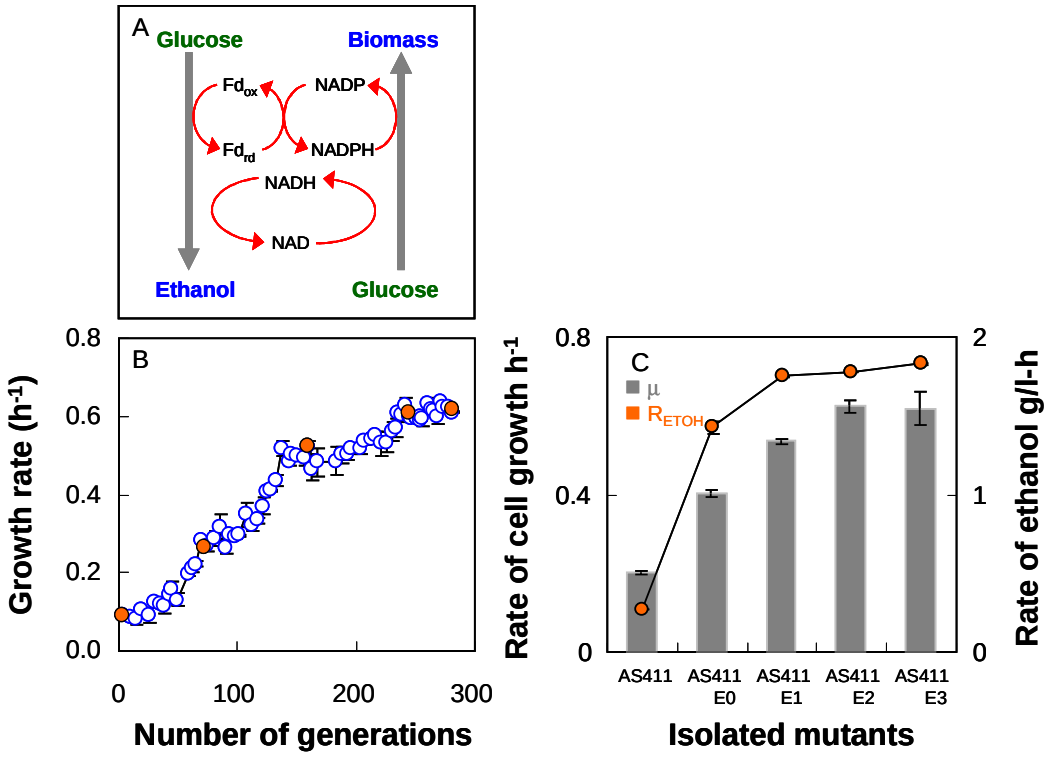


Figure 8.5 Metabolic evolution of AS411 during anaerobic growth on glucose. (A) Overview of co-factor requirements for supporting cell growth and ethanol biosynthesis in AS411 (B) Change in specific growth rate over evolution time during serial dilution. Filled symbols represent evolved cultures AS411E# isolated at different number of generations that have been evaluated in controlled bioreactor experiments. Evolution showing change in growth rate over number of generations during serial dilution. (C) Performance of the evolved cells after an evolution time of 329 hrs – 75 generations (AS411E0); 487 hrs – 159 generations (AS411E1); 600 hrs – 251 generations (AS411E2); 642 hrs- 280 generations (AS411E3). Evolved cells have higher specific growth rates and ethanol production rates in comparison to their parent, AS411.

The necessity of ethanol production for growth of AS411 enables the optimization of flux to ethanol synthesis via metabolic evolution. Due to the strong link between cell growth and ethanol synthesis in the 4 remaining pathways operating in AS411, the evolved mutant with an increased growth rate would also produce ethanol at a faster rate. Thus, a mutant of AS411with a higher ethanol production rate can be selected during the metabolic evolution by using cell growth as selective pressure. Fig. 8.5B shows the change in cell growth rate as a function of the number of generations through a serial dilution experiment using AS411 that had been anaerobically evolved on glucose. At the beginning of metabolic evolution in anaerobic shake tube fermentation, the mutant grew at a specific growth rate of about 0.10 hr^{-1}. The improvements in the growth rate occurred after the 38 generations of evolution and continuously increased until 238 generations. The specific growth rate of the evolved mutant became stabilized at 0.61 hr^{-1} after 280 generations of metabolic evolution in a serial dilution experiment. Faster-growing mutants are expected to be selected during the evolution experiment since strains with faster growth rate will eventually dominate the culture replacing the slower-growing strains. This explains a continuous increase in the specific growth rate of the culture over the number of generations observed.

Four evolved cultures were isolated at different numbers of generations: AS411E0 at 75 generations, AS411E1 at 159 generations, AS411E2 at 251 generations, and AS411E3 at 280 generations. The evolved strains resulted in not only an improved cell growth rate, but also an increase in the ethanol production rate validating the model prediction of the tight coupling of the two products (Fig. 8.5C). Anaerobic batch bioreactor experiments

showed AS411E3 to consume all glucose within 8 hrs, compared to AS411 which utilized the same amount of sugar at around 36 hrs (Fig. 8.3). The evolved mutant showed improvements in both sugar consumption rate and ethanol production rate. The ethanol production rate of AS411E3 was approximately 6-fold higher than that exhibited by AS411. These different fermentation phenotypes between AS411E3 and AS411 indicated the different genotypes between the two strains. The observed ethanol and biomass yields of the evolved mutant AS411E3 (Table 8.4) fell into the predicted yield ranges based on elementary mode analysis, suggesting that the model can accurately predict the phenotype of an evolving cell.

8.4 Discussion

In this study, a system chosen for analysis is the anaerobic production of ethanol in *T. sacharrolyticum*. The metabolism of *T. sacharrolyticum* is reduced by implementing gene knockouts which are predicted based on EMA for an efficient production of ethanol at high yield conversion. The metabolism of the mutant was then subjected to adaptive evolution for further improvement of the kinetics of ethanol production.

Based on inverse metabolic engineering guided by EMA, the combination of targeted gene deletions redirected the carbon metabolism of *T. saccharolyticum* toward the efficient ethanol-producing pathways. The designed strain, AS411, with predicted targeted gene deletions was constructed. In agreement with the model prediction, the mutant AS411, which contained targeted gene deletions, showed an improved ethanol yield compared to the wild-type. The observed improvement in the mutant could be a result of increased precursors and cofactors for ethanol synthesis due to deleting the

targeted genes. The disruption of fermentative pathways through elimination of *pta-ack* for acetic acid and *ldh* and *mgs* for lactic acid leads to more carbon flow toward the ethanol synthesis pathway due to the shared precursors between these fermentative products and ethanol. The disruption of the oxidative pentose phosphate pathway (PPP) by deleting *zwf* forces the cell to catabolize glucose through only glycolysis instead of a possible combination of glycolysis and oxidative PPP. This, therefore, causes the increasing flux through glycolysis and to pyruvate precursor. The disruption of *zwf* also helps to reducethe carbon loss in the form of CO_2 which occurs when glucose is directed to the oxidative PPP.

Although AS411 showed improvements in its ethanol yield compared to that of the wild-type, the rate of ethanol production by AS411 was tremendously reduced. There was approximately a 4-fold decrease in productivity compared to the wild-type. The reduced ethanol productivity of AS411 could be a result of rate-limiting steps causing metabolic imbalance and/or accumulation of metabolites that become toxic to the cell. Based on the available elementary modes, AS411 cannot produce biomass without producing ethanol as a result of balancing cellular redox potential. The tight coupling of cell growth and ethanol pathway in the strain makes it a good candidate to subject to metabolic evolution for overcoming these limitations. Due to a strong link between growth and ethanol synthesis, presumably, the mutant that expresses an improved kinetic of ethanol synthesis will grow faster. During metabolic evolution in a serial dilution experiment, the strain with the faster growth rate will eventually dominate the culture replacing the non-dominant strains. The evolved mutant AS411E3, when isolated from metabolic evolution,

showed improvements in both growth rate and rate of ethanol production, suggesting that mutations occurring in AS411E3 permit the strain to overcome the rate-limiting steps for cell growth and ethanol synthesis.

The application of elementary mode analysis in combination with adaptive evolution has proven to be an effective metabolic engineering tool in creating a strain with improved product yield and productivity, as illustrated in this chapter and in previous chapters. Herein, we have shown the optimization and evolution in metabolic pathways of *T. saccharolyticum* for anaerobic production of ethanol from glucose or xylose. EMA permits a rational identification of gene knockout mutations that redirect the carbon metabolism of *T. saccharolyticum* towards efficient ethanol-producing pathways. Metabolic evolution via natural selection allows a selection of a mutant with an enhanced ethanol production rate.

Chapter 9

Metabolic networks evolve towards states of maximum entropy production

9.1 Chapter summary

A metabolic network can be described by a set of elementary modes representing discrete metabolic states that support cell function. We have shown in Chapter 3 that in the most likely metabolic state the usage probability of individual elementary modes is distributed according to the Boltzmann distribution law while complying with the principle of maximum entropy production (Srienc and Unrean, 2010). To validate that a metabolic network evolves towards such state we examined the flux distribution of *T. saccharolyticum* AS411 during adaptive evolution. The strain operate with a reduced metabolic functionality based on a four set of elementary modes (see Chapter 8 for details). In such reduced metabolic network metabolic fluxes can be conveniently computed from the measured metabolite secretion pattern. Flux analysis based on a combination of elementary modes and their usage probabilities revealed that the metabolic fluxes of the evolving cell are shifted in a direction of increasing rate of entropy production. This experimental observation is further evidence supporting the maximum entropy production principle that a cell metabolism naturally evolves by redistributing its metabolic fluxes towards the maximum entropy point of operation. This state can be predicted using fundamental thermodynamic relations described by the Boltzmann distribution law.

9.2 Introduction

Mutations and genetic rearrangements enable an organism to adapt to environmental conditions. A cell with optimal performance is expected to emerge from an evolutionary process in which fittest cells are selected by the given cell environment (Ibarra, et al. 2002; Fong, et al. 2003; Teusink, et al. 2009). The optimality of a cell is achieved through a redistribution of the flux pattern through the metabolic network by adjusting the expression pattern of catalytic and regulatory proteins. The metabolic state of a fully evolved cell, therefore, reflects the optimal flux distribution. The adaptability of a cell metabolic network in response to evolution can provide a better insight into the evolution of metabolic pathways. A fundamental question in evolution is how the flux distribution of a cell changes in response to selection conditions imposed by the environment.

Understanding the evolutionary optimization of cell metabolism requires knowledge of the steady-state metabolic flux distribution. Such knowledge can be obtained through Elementary Mode Analysis (EMA) which decomposes the overall metabolism of a cell into a set of unique, indivisible pathway fluxes called elementary modes (Pfeiffer, et al. 1999; Schuster, et al. 2000). These pathway fluxes operate based on minimal sets of enzymes and are considered the fundamental components determining cell physiology. The functioning metabolism of a cell can be expressed as a weighted, linear combination of elementary modes (Pfeiffer et al., 1999; Schuster et al., 2000). Thus, analysis of elementary modes permits a quantitative evaluation of the structure, capability and robustness of cell metabolism (Carlson and Srienc, 2004; Papin, et al. 2002; Stelling, et al. 2002; Trinh, et al. 2008; Unrean, et al. 2010). A steady-state flux distribution

corresponding to the state of cellular metabolism can be reconstructed based on a sum of the product of elementary modes and their usage probabilities (Srienc and Unrean, 2010; Wlaschin, et al. 2006, Eq (3.43)). A cell metabolism is expected to adapt to a given condition by varying the probabilities of these elementary modes to result in optimal fitness.

A previous work by Wlaschin, et al (2006) suggested a rational way of estimating the probability of elementary mode based on the thermodynamic properties of the pathway. The theoretical framework given in Chapter 3 showed that elementary modes can be interpreted as a canonical ensemble of metabolic states of overall cell metabolism analogous to an ensemble of quantum states of a macroscopic system in statistical thermodynamics. Using the ensemble concept and the principle of maximum entropy production (Martyushev and Seleznev, 2006; Zhao and Kurata, 2009; Srienc and Unrean, 2010), we hypothesized that a fully evolved cell operates at the most likely state at which entropy production is maximized. In such state, which represents the fully evolved metabolism, the occupation number (usage probability) of the entropy contribution of individual elementary modes is distributed according to the Boltzmann distribution law analogous to the distribution of discrete energy states in a physical system. The maximization of the rate of entropy production represents a critical objective (i.e., a driving force) in the evolution of metabolic pathways supporting cell function. Thus, an evolving cell would spontaneously redistribute its pathway fluxes by optimizing the usage probabilities to ultimately reach the fully evolved state when the rate of total entropy formation is maximized.

We validate here the theoretical predictions by analyzing the changes in metabolic flux distributions of the metabolism of *T. saccharolyticum* AS411 during adaptive evolution in a serial dilution experiment in which the cell culture has been maintained at exponential growth conditions for extended numbers of generations. We show that the evolving mutant of AS411 adjusts the metabolic flux distribution during adaptive evolution in a direction of increasing rate of entropy production supporting the previously proposed assertion that a metabolic network naturally evolves by redistributing its fluxes to maximize the rate of total entropy production.

9.3 Results and discussion

9.3.1 Change in flux distribution during evolution

In Chapter 8, the metabolic network model of *T. saccharolyticum* for anaerobic growth on glucose was constructed and was analyzed by elementary mode analysis to determine all balanced pathways that are inherent in the cell metabolism. Multiple gene knockouts were implemented to limit the metabolic functionality of *T. saccharolyticum* to a set of only 4 classes (families) of efficient ethanol-producing pathways representing 42 elementary modes. Each family of modes has the same overall reaction stoichiometry and the same total reaction entropy (Table 9.1). Note that elementary modes within the same family utilize a different combination of internal reactions, but they have the same overall external reaction stoichiometry that connects substrate to products (Wlaschin et al., 2005). The strain containing the predicted gene deletions was constructed and named AS411.

Table 9.1 Stoichiometric equations and standard reaction entropies of the four family modes in AS411 grown on glucose and the probabilities of mode families for the mutant AS411 and the evolved mutant AS411E3. The standard entropy of reaction per mole glucose of each family (kJ/K-mole) is calculated using Gibbs equation and correlations with the degree of reduction given in Sandler and Orbey (1991).

Family	Reaction stoichiometry	△Si	p_i of AS411	p_i of AS411E3
1	Glucose → 2 Ethanol + 2 CO_2 + 2 ATP	0.547	0.411	0.337
2	Glucose → 2 Ethanol + 2 CO_2	0.503	0.411	0.337
3	Glucose + 0.70 NH_3 → 2.88 Biomass + 0.11 H_2 + 1.10 Ethanol + 1.18 CO_2	0.473	0.078	0.162
4	Glucose + 0.70 NH_3 → 2.88 Biomass + 0.11 Formate + 1.10 Ethanol + 1.07 CO_2	0.453	0.078	0.162

The mutant AS411 was allowed to evolve in a serial dilution experiment for approximately 280 generations equivalent to 650 hrs of anaerobic growth on glucose, resulting in the evolved mutant AS411E3. It is expected that during metabolic evolution, the mutant maximizes its fitness by optimizing the usage of the probabilities of these modes. The minimized metabolic functionality of AS411 permits convenient metabolic flux estimations of the strain from experimentally determined metabolite secretion patterns. The change in pathway flux distributions of the constructed strain in response to adaptive evolution was examined. The probability for each set of pathways contributing to the overall metabolism of the strain was computed from a completely determined algebraic system of mass balance equations based on the measured secretion or consumption fluxes of glucose, ethanol and biomass and the summation constraint of the probabilities (Eq. (3.44)). Analysis of the associated usage probabilities of pathway groups for the mutant AS411 and the evolved mutant AS411E3 suggests an increase in

the probabilities of the biomass-ethanol coproducing modes during the directed evolution (Table 9.1). The known probabilities of families allow for the calculation of the intracellular fluxes as weighted averages of the families. The overall flux distribution for the mutant AS411 and the evolved mutant AS411E3 (Fig. 9.1), which was determined from the sum of the product of the 4 family modes and the usage probabilities, reveals the adaptation capability in the cell metabolism during evolution. The metabolism of AS411E3 has an increase in pentose phosphate pathway (PPP) flux which could be one of the main reasons for improved cell growth observed in the evolved mutant since the PPP is important for the synthesis of several biomass precursors.

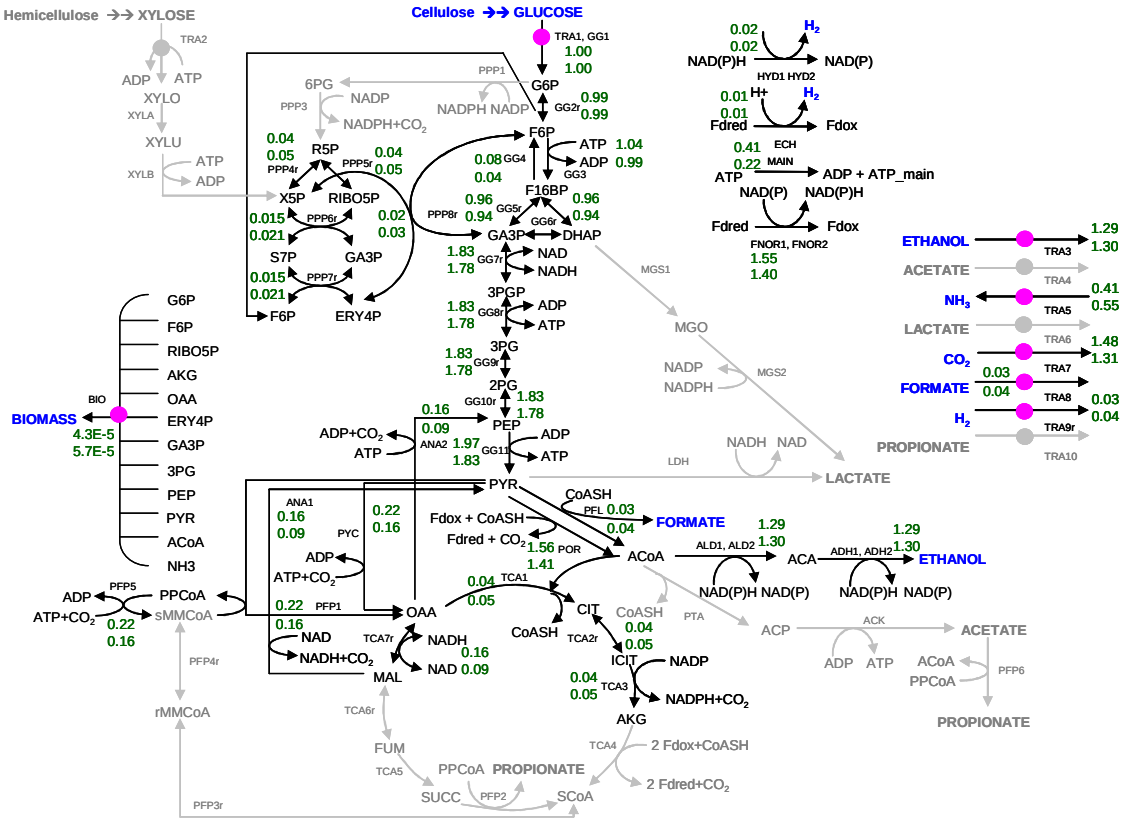


Figure 9.1 Metabolic map of *T. saccharolyticum* central metabolism under anaerobic growth on glucose. Metabolic fluxes in AS411 (Top) and in AS411E3 (Bottom) are listed next to reactions. The fluxes are determined from the sum of the product of the probabilities and the fluxes of the 4 families shown in Table 9.1. All metabolic fluxes are normalized to the uptake rate of glucose. The zero fluxes are shown in grey.

9.3.2 Entropy analysis of AS411 metabolism

The entropy generation of an individual family elementary mode can be calculated from the reaction entropy of the individual family elementary mode as described in Eq 3.21. Fig. 9.2A shows the predicted functional space of the total entropy generation as a function of the usage probability of the individual family modes for AS411 based on Eq (3.23), which is shown in grey plane. The total entropy generation according to

Boltzmann entropy Eq (3.24) is represented by the black plane. The associated usage probability that results in the maximum rate of entropy production (red circle) represents the ultimate state of a fully evolved metabolic network (Srienc and Unrean, 2010). This fully evolved state, which is at the intersection of the two surface plots, is the state where the usage probabilities of the metabolic pathways follow the Boltzmann distribution law at the maximum entropy production.

The overall entropy production for the cell metabolism was computed from the weighted sum of the family mode entropies and the experimentally determined family probabilities (Eq (3.23)). Analysis of the experimentally determined pathway probabilities of the evolving cells suggests that AS411 evolved by redistributing the probabilities of family modes towards the metabolic functional state of the fully evolved cell where the usage probabilities are according to the Boltzmann distribution law (Fig. 9.2A). Likewise, the total rate of entropy production, which was calculated from the product of the uptake rate of glucose and the total reaction entropy (Eq (3.42)) revealed a shift in the metabolic functional state of the evolving cells in a direction of increasing rate of entropy generation (Fig. 9.2B). The results suggest that the cell tends to evolve its metabolic functional state toward the fully evolved state with the maximum rate of entropy generation during evolution. This observation confirms entropy as a driving force of pathway evolution.

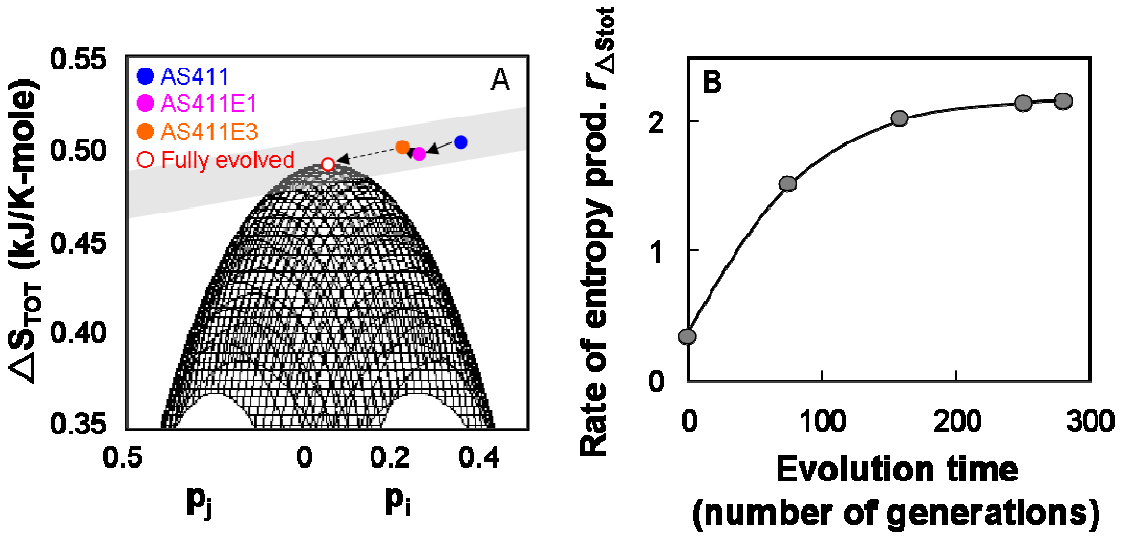


Figure 9.2 Total reaction entropy production for AS411 and the evolved mutants under anaerobic growth on glucose. (A) Experimental entropy generation as a function of probabilities of elementary modes for AS411 (●) and the evolved cell cultures (● AS411E1; ● AS411E3). The plot shows the shift in metabolic state of cells towards the predicted state of a fully evolved system where the entropy is maximized and the usage probabilities are according to the Boltzmann distribution law (see Eq (3.35)). The ○ symbol represents the fully evolved system estimated from the asymptotic value of the fitting functions of probabilities shown in Fig. 9.3 at infinite evolution time. (B) Change in rate of entropy generation, computed from the product of the total reaction entropy and the uptake rate of glucose, as a function of evolution time. Circles represent the experimentally determined rate of entropy production while the line represents the rate of entropy production calculated from the fitting functions of probabilities given in Fig. 9.3. The plot suggests that the cell system has a natural tendency to evolve with time towards an asymptotic state with maximum rate of entropy production.

As shown in Fig. 9.2A, the metabolism of the evolved mutant AS411E3 has not yet reached the predicted state of a fully evolved metabolism, since its experimentally determined point of operation was located away from the predicted maximum entropy point of operation. With a longer evolution time, the mutant would be expected to adjust its flux distribution to eventually reach the fully evolved metabolic state where the pathway fluxes yield the maximum rate of entropy generation and the flux distribution is

according to the Boltzmann distribution law. Hence, to verify that the predicted state of maximum entropy production according to the Boltzmann distribution law represents the metabolic state of a fully evolved cell, we fitted an exponential function to the experimentally determined probabilities over evolution time and estimated the asymptotic values of these probabilities from the fitting function (Fig. 9.3). These asymptotic values represent the probabilities that the cell would ultimately reach at an infinite evolution time. Thus, they represent the usage probabilities of the pathway families in the fully evolved cell system.

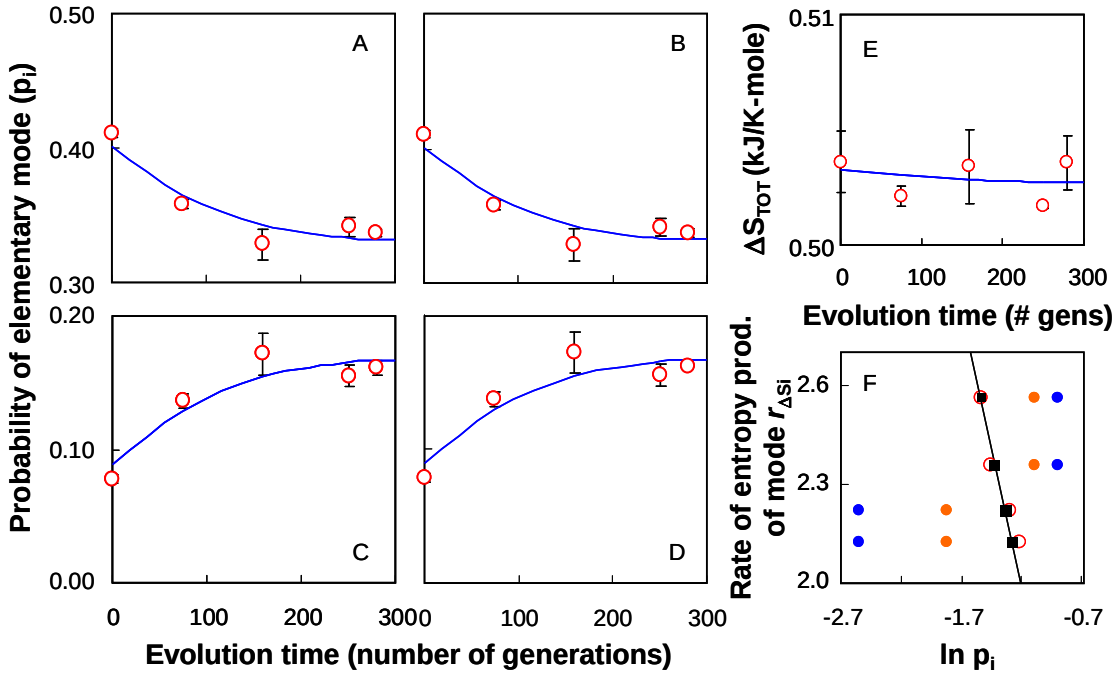


Figure 9.3 Change in usage probabilities of elementary mode families and total reaction entropy production during metabolic pathway evolution. The probabilities of family mode 1 (A), family mode 2 (B), family mode 3 (C), and family mode 4 (D) change over evolution time. Circles represent the experimental values, while the lines represent the exponential fitting functions. The parameters of the function have been determined by a least square method. The fitting functions are $p_1 = 0.188 \cdot e^{(-0.001 \cdot t)} + 0.214$; $p_2 = 0.169 \cdot e^{(-0.001 \cdot t)} + 0.233$; $p_3 = -0.186 \cdot e^{(-0.001 \cdot t)} + 0.273$; $p_4 = -0.212 \cdot e^{(-0.001 \cdot t)} + 0.298$. The total reaction entropy (kJ/K-mole of glucose) over evolution time (E) is computed as a weighted sum of the probabilities and the

reaction entropies of individual modes. (F) The production rate of entropy of individual modes as a function of the natural log of the usage probabilities of modes. The rate of entropy production is computed from the product of reaction entropy of individual modes and the asymptotic rate of glucose uptake (see Eq (3.39)). The experimentally determined usage probabilities of elementary modes after an evolution time of zero generations (●), of 280 generations (●), and the asymptotic values at infinite evolution time (○) are compared to the predicted probabilities based on the Boltzmann distribution law (■). Over the evolution time the probabilities approach the predicted values.

The asymptotic probabilities agree well with the predicted probabilities (Table 9.2) suggesting that the fully evolved cell system functions at the state of maximum entropy production where the distribution of the metabolic pathways follows the distribution law. In addition, the fitting functions of probabilities can accurately predict the entropy generation per mole of glucose and the total entropy production rate over evolution time (Fig. 9.2). Based on the values of asymptotic probabilities, we computed the total reaction entropy for the fully evolved strain to be 0.50 kJ/K-mole, which is in excellent agreement with the predicted maximal entropy production value of 0.49 kJ/K-mole (Table 9.2).

Table 9.2 Usage probabilities of elementary modes and reaction entropy for a fully evolved metabolism of AS411. The p_{Fi} and p_{Ei} represent the probabilities of family modes and of individual modes within the corresponding families respectively. The p_{Ei} value is computed from the p_{Fi} value by dividing by the number of elementary modes in the family. The predicted probabilities are determined from the Boltzmann distribution law. The asymptotic probabilities are estimated from experimental data using the fitting functions given in Fig. 3. Total reaction entropy (kJ/K-mole) is determined from the weighted sum of the probabilities and the reaction entropies of individual modes.

Family	Predicted p_{Fi}	Predicted p_{Ei}	Asymptotic p_{Fi}	Asymptotic p_{Ei}
1	0.214	0.071	0.214	0.071
2	0.242	0.016	0.233	0.016
3	0.264	0.015	0.273	0.015
4	0.279	0.047	0.298	0.050
ΔS_{TOT} (kJ/K-mole)	0.49		0.50	

In summary, we applied elementary mode analysis in combination with thermodynamic analysis to examine the evolution of metabolic pathways. We have demonstrated that the steady-state flux distribution obtained through elementary mode analysis provides a quantitative measure in the evolutionary optimization of cell metabolism. The presented results suggest that the evolution of metabolic pathways is driven by the rate of entropy generation and that the rate of entropy production asymptotically approaches the maximum entropy production rate predicted from the state when the usage probability of individual elementary modes is distributed according to the Boltzmann distribution. Thus, for a cell system that is not at the optimal metabolic flux distributions of the fully evolved metabolism, the evolutionary path towards the maximum entropy generating state can be predicted.

Chapter 10

Identification of rate-controlling enzymes targeted by metabolic pathway evolution using metabolic control analysis

10.1 Chapter summary

In this chapter, elementary mode analysis and metabolic control analysis were applied to examine the control of metabolic pathways during adaptive evolution. An organism presumably achieves the fully evolved state through adaptive changes in the kinetics of rate-controlling enzymes. Metabolic control analysis based on flux control coefficients (FCC) identifies such enzymes. Analysis of changes in steady-state metabolic flux distributions of AS411 towards the fully evolved flux distributions predicted by the Boltzmann distribution law identified phosphoglucose isomerase (PGI) as the rate-controlling enzyme. FCC predicts that an increased activity of PGI would enable the metabolic network to reach the fully evolved state. The prediction was supported by experimental results that showed an increased specific activity of PGI in the evolved mutant AS411E3 compared to the mutant AS411. The results indicate that the evolutionary path is predictable as the strain AS411 adapted by increasing the PGI activity towards the fully evolved state, the state with optimal growth. This experimental finding suggests that cell metabolism maximizes its fitness by overcoming the local control of rate-liming enzymes. Thus, enzymes with the greatest flux control are the targets of metabolic pathway evolution.

10.2 Introduction

The overall metabolic network of an organism is represented by an ensemble of elementary modes where individual modes contribute to the network with certain probabilities. The probabilities of elementary modes provide information of the contribution of each mode to the overall metabolism. The identification of overall fluxes through a linear combination of modes and their probabilities allows us to gain powerful insight into the control and regulation of the cell metabolism. In Chapter 9, we have shown that cell evolves by redistributing the probabilities to reach the fully evolved state that operates using preferentially elementary modes with small reaction entropies (Srienc and Unrean, 2010; Wlaschin, et al. 2006). This fully evolved state represents the metabolic network that provides an optimal cell growth under selective conditions imposed by the culture conditions (Fong and Palsson, 2004; Hua, et al. 2006; Ibarra, et al. 2002; Schuster, et al. 2008). Using the ensemble method of statistical thermodynamics and the principle of maximum entropy production as described in Chapter 3, the fully evolved state can be predicted as the most probable state where the usage probabilities are distributed according to the Boltzmann distribution law at the maximum rate of entropy generation (Srienc and Unrean, 2010; Zhao and Kurata, 2009; Zhao and Kurata, 2010).

During adaptive evolution and natural selection, cell achieves the fully evolved state primarily through genetic modifications that result in adaptive changes in the kinetic parameters of rate-controlling enzymes within metabolic pathways (Cork and Purugganan, 2004; Flowers, et al. 2007; Vitkup, et al. 2006). That is, evolutionary change

at the rate-limiting enzymes by increasing the kinetic parameters is a means of enabling cell to reach the fully evolved state. The identification of these rate-controlling enzymes is a challenging problem that requires knowledge of steady-state flux distributions and of the control structure within metabolic networks. Such knowledge can be obtained through a combination of elementary mode analysis (EMA) and metabolic control analysis (MCA). Both EMA and MCA have been proven to serve as guidelines for the optimization and manipulation of cell metabolic pathways (Moreno-Sanchez, et al. 2008; Trinh et al., 2009). The flux distributions obtained from a weighted, linear combination of elementary modes present and their usage probabilities can be analyzed using MCA to obtain information about the degree of control and regulation that each flux (i.e., enzyme) has within the metabolic pathways (Bordel and Nielsen, 2010; Fell 1997; Heinrich and Schuster, 1996; Heinrich and Schuster, 1998). MCA based on flux control coefficients allows for a measure of metabolic fluxes response affected by changes in enzyme activities or parameters that are directly correlated with the rates of reactions. Rate-controlling enzymes in metabolic networks are identified as the enzyme with maximal flux control coefficients.

In this chapter, the fully evolved metabolism of *T. saccharolyticum* AS411 representing an optimal metabolic network state for anaerobic growth on glucose was predicted based on the Boltzmann distribution law. We applied flux control coefficients (FCC) to analyze changes in the flux distributions of AS411 towards the predicted fluxes of the fully evolved network. FCC which measures relative change in rate of cell growth in response to change in activity of individual enzymes (i.e., rates of enzymatic reactions)

within metabolic pathways identifies rate-controlling enzymes. We hypothesized that these controlling enzymes are targeted by metabolic pathway evolution, and are responsible for the phenotypic change that occurs during evolution. The hypothesis was confirmed by experiments.

10.3 Theory

Metabolic flux distributions for a fully evolved cell

The steady-state flux distribution for the fully evolved state R_j^* can be expressed as a linear, weighted contribution of the rate of individual reactions specified in each elementary mode present as described in Eq (3.43)

$$R_j^* = \sum p_i^* \cdot r_{i,j} \qquad \textbf{(10.1)}$$

where p_i^* is the usage probability of elementary mode i to the overall metabolic fluxes R_j^* in the fully evolved network, and $r_{i,j}$ represents rates of jth reaction per rate of glucose uptake derived from the ith elementary mode.

The probability of elementary mode p_i^* is computed from the Boltzmann distribution law

$$p_i^* = \frac{e^{-\Delta S_i/b}}{\sum e^{-\Delta S_i/b}} \qquad \textbf{(10.2)}$$

where ΔS_i represents the reaction entropy of elementary mode i, and b is a constant which can be calculated from the partition function Q (see Chapter 3 for details).

The predicted probabilities from Eq (10.2) can be used to compute cell yield $Y_{x/s}^*$ and cell growth rate μ^* of the fully evolved cell:

$$Y_{x/s}^* = \sum p_i^* \cdot Y_{x/s,i} \qquad \textbf{(10.3)}$$

$$\mu^* = Y^*_{x/s} \cdot r_g \quad \textbf{(10.4)}$$

where $Y_{x/s,i}$ is the biomass yield of elementary mode i, and r_g is the asymptotic rate of glucose uptake.

Flux control coefficients

Flux control coefficients measure the degree of control that a given enzyme exerts on a metabolic flux of a given network. The control coefficient $C^J_{R_j}$ is defined as a relative change in the metabolic flux J in response to a relative change in enzyme activity which is directly correlated with the rate of reaction R_j (see Chapter 3 for details). A reaction with high flux control coefficient means that a small variation of the reaction rate would cause a significant change in the metabolic flux. Thus, a reaction with maximal flux control coefficient is presumed to be a rate-controlling reaction. In this study, the metabolic flux J represents the metabolic flux for biomass synthesis (i.e., cell growth rate). We applied the flux control coefficients to quantify the degree of control that each rate of reaction in the network of AS411 has on the rate of cell growth during evolution. This was achieved by measuring changes in metabolic fluxes for biomass synthesis (cell growth rate) J_X caused by changes in individual reaction rates between the unevolved and fully evolved networks:

$$C^J_{R_j} = \frac{\Delta \ln J_X}{\Delta \ln R_j} = \frac{\ln J^*_X - \ln J_X}{\ln R^*_j - \ln R_j} \quad \textbf{(10.5)}$$

where J_X and J^*_X represent the flux for biomass synthesis of the unevolved and fully evolved networks respectively, and R^*_j and R_j represent the individual reaction rates in

the unevolved and fully evolved networks respectively. All fluxes used for the calculation of flux control coefficients are given in Fig. 10.1. Comparing the control coefficients of all reactions helps to identify the rate-controlling enzymes in the network that is the target of optimization during metabolic evolution.

10.4 Results and discussion

10.4.1 Analysis of metabolic network for AS411 and the fully evolved cell

T. saccharolyticum AS411 metabolism functions using four family modes under anaerobic growth on glucose (see Chapter 8 for details). Fig. 10.1 shows the metabolic flux distribution of AS411, which was computed as the weighted average of the families and their probabilities, compared to that of the fully evolved cell, which was computed based the Boltzmann distribution law (Eq (10.2)). The distributions of fluxes between the two networks reveal the evolutionary path of AS411. The fully evolved network has a higher flux for biomass synthesis, suggesting that AS411 would evolve to have a higher cell growth. The fully evolved cell is predicted to have a maximal growth rate of 0.87 hr^{-1} and a maximal cell yield of 0.18 g/g based on Eq (10.3) and (10.4). We compared cell growth between the mutant AS411 and the evolved mutant AS411E3 which was adaptively evolved for 280 generations under anaerobic growth on glucose (see Chapter 8 for details). The results showed AS411E3 with approximately 3-folds higher specific growth rate than the parent AS411. Cell yield of AS411E3 was 0.13±0.02 g/g compared to 0.05±0.01 g/g of AS411. The experimental observations were in agreement with the predicted flux distribution that suggests the strain to evolve towards higher cell growth.

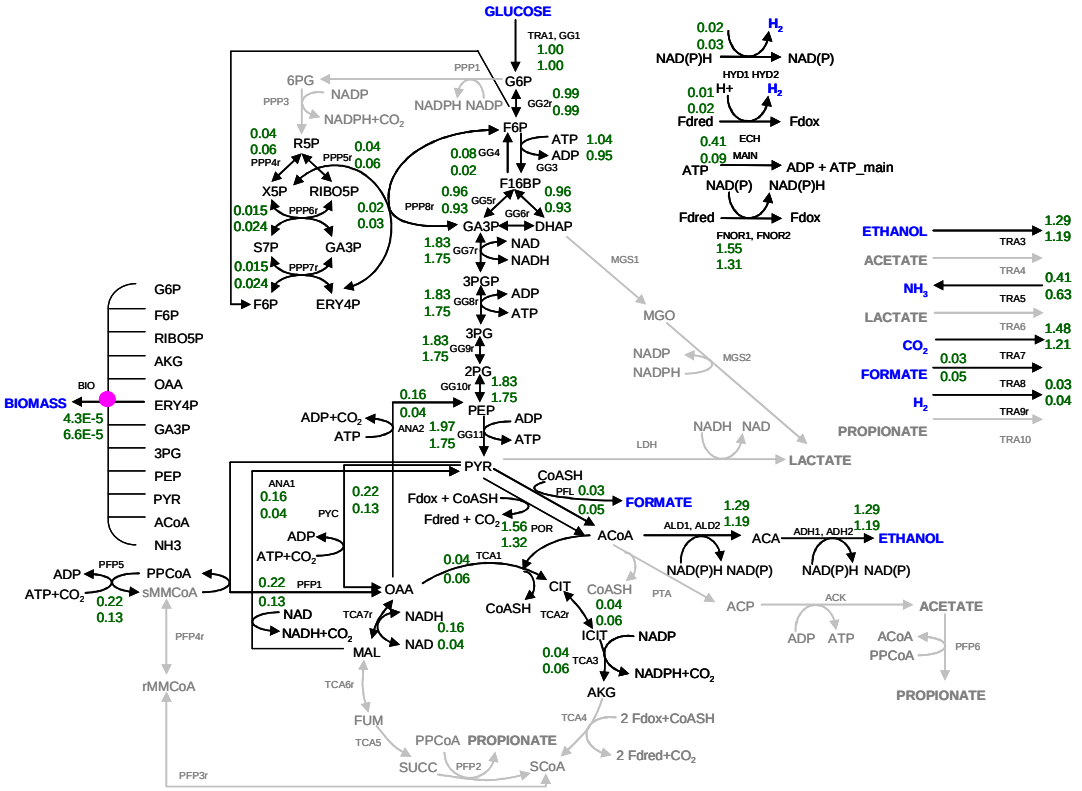


Figure 10.1 Metabolic flux distributions of *T. saccharolyticum* AS411 (Top) and the fully evolved cell (bottom). The fluxes were determined from the sum of the product of four family modes present and their probabilities given in Table 9.1 and 9.2. The usage probabilities of AS411 were determined based on the measured secreting fluxes of external metabolites while those of the fully evolved cell were predicted based on the Boltzmann distribution law (see Eq (3.35)). All fluxes were normalized with the uptake rate of glucose. A comparison of the flux distributions between the two networks determines the evolutionary path of AS411 during metabolic pathway evolution.

10.4.2 Identification of changes in AS411 metabolism during evolution

It is still unclear which enzymes would be the target of manipulation during evolution allowing AS411 to achieve the fully evolved state with optimal cell growth. The fully evolved metabolism is, presumably, achieved by altering the kinetic parameters of rate-controlling enzymes. Therefore, we applied flux control coefficients to quantify the

degree of control that each flux had on cell growth of AS411 during adaptive evolution. Details on flux control coefficients are given in Chapter 3. FCC quantitatively determined how the change in rate (i.e., enzyme activity) of constraining reactions towards the fully evolved cell affects the change in rate of biomass synthesis (i.e., cell growth rate) of AS411. Comparing flux control coefficients of all reactions in the network, we identified reaction GG2r corresponding to phosphoglucose isomerase (PGI) as the enzyme with maximal flux control coefficients (Fig. 10.2). PGI, which catalyzes the reversible isomerization of glucose-6-phosphate (G-6-P) to fructose-6-phosphate (F-6-P), plays a central role in glucose metabolism, both in glycolysis via the Embden-Meyerhof (EM) pathway and in gluconeogenesis.

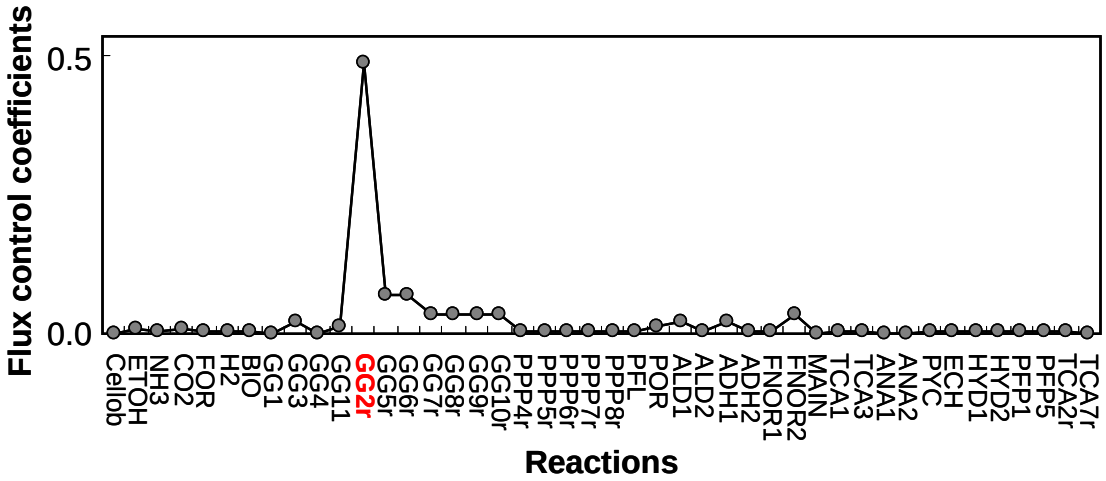


Figure 10.2 Metabolic control analysis of AS411 metabolic network that is shifting towards the fully evolved network. Flux control coefficients of all reactions in the metabolic network were calculated based on Eq (10.7) using the rates of reactions given in Fig. 10.1. The analysis revealed reaction GG2r corresponding to phosphoglucose isomerase (PGI) as rate-controlling reaction for cell growth.

An increase in the PGI activity during evolution would exert the most increase in cell growth of AS411 compared to an increase in other enzymes with zero or small flux

control coefficients. Since PGI is identified as the enzyme with the greatest flux control (i.e., the rate-controlling enzyme), mutations with increased activity at PGI would have the highest growth rate compared to other mutations with increased activity at steps with little or no flux control based on the coefficients. These PGI mutants would outgrow other mutants and eventually take over the culture during metabolic pathway evolution. We, therefore, predicted that PGI is the target of manipulation that is presumably overcome during the evolution of AS411 metabolic pathways.

10.4.3 Activity of phosphoglucose isomerase

To validate that AS411 adapts towards the fully evolved state, the state with maximal cell growth, by altering the PGI activity, we compared the specific activity of PGI between the mutant AS411 and the evolved mutant AS411E3. Enzymatic assays showed higher PGI activity in AS411E3 culture than in AS411 culture (Fig. 10.3A). The specific activity of PGI in AS411E3 was 0.69±0.01 U/mg-protein compared to 0.57±0.01 U/mg-protein in AS411, which is equivalent to a 21% increase in activity. The measured kinetic parameters of PGI shown in Fig. 10.3B revealed AS411E3 with a higher Vmax value (0.76 U/mg-protein in AS411E3 culture and 0.65 U/mg-protein in AS411 culture) and a lower Km value (1.28 mM in AS411E3 culture and 1.61 mM in AS411 culture). The measured flux ratio of the PGI between the mutant AS411 and the evolved mutant AS411E3 was 1.49 (based on the enzyme activity at 10 μM substrate), which is in close agreement with the predicted value of 1.55 based on the flux distributions given in Fig. 9.1.

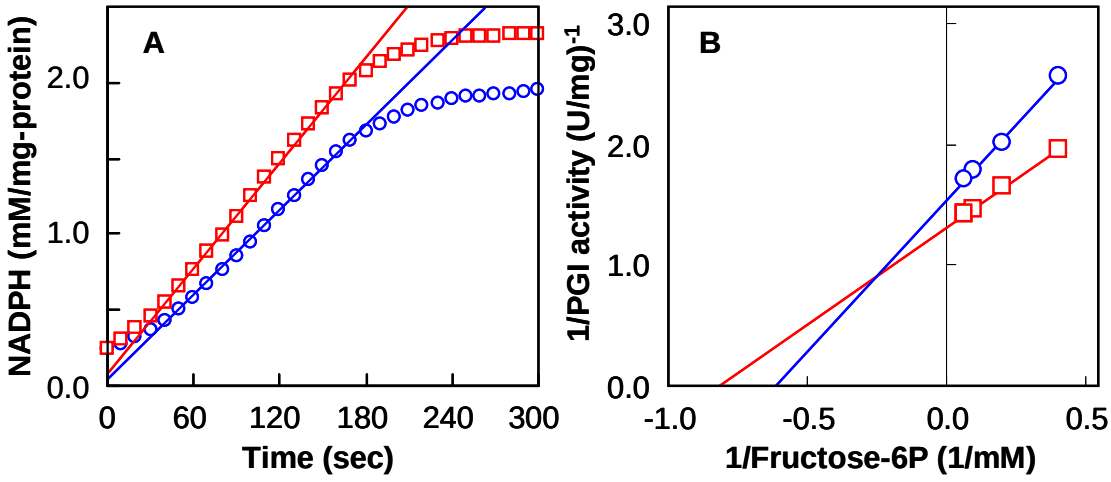


Figure 10.3 The activity of PGI in cell extracts of the mutant AS411 (○) and the evolved mutant AS411E3 (□). (A) PGI activity was assayed by monitoring change in NADPH concentration (mM/mg-protein) over time. Enzyme activity was determined from the slope of the linear region of the plot. The results revealed in the case of AS411E3 an increased PGI activity as compared to AS411. The assays were conducted using 15 mM of fructose-6-phosphate (B) Rate dependence of PGI activity (unit/mg-protein) on the fructose-6-phosphate concentration (mM) at 55°C in cell extracts showed V_{max} value of 0.76 U/mg-protein and K_m value of 1.28 mM for AS411E3 compared to V_{max} value of 0.65 U/mg-protein and K_m value of 1.61 mM for AS411.

The observed increase in PGI activity supports the prediction that AS411 would increase its growth rate by increasing the activity of the rate-controlling enzyme PGI. The activity of hexokinase (GLK) and phosphofructokinase (PFK) enzymes were also compared between the two cultures. These enzymes were selected to verify the prediction of flux control coefficients. We observed no significant change in the activity of the two enzymes between AS411 and AS411E3 culture (Table 10.1) which is expected since both enzymes have zero or small flux control coefficients. The experiments validated the model prediction that the increase in cell growth of the evolved mutant is responsible by the increase in the activity of PGI. Several individual adapted clones isolated from the

evolved culture AS411E3 also showed relatively similar cell growth rate and PGI activity (Table 10.2), suggesting the phenotypic homogeneity in the evolved culture.

Table 10.1 Flux control coefficients and specific activity of hexokinase, phosphoglucose isomerase, and phosphofructokinase enzymes in cell extracts of AS411 and AS411E3 cultures. Specific enzyme activity was measured from cell extracts at 15mM substrate at 55°C. Δ(activity) is the enzyme activity difference between the two cultures. One unit of enzyme is defined as a conversion of 1 mM of substrate per minute.

Reactions (enzymes)	Flux control coefficients (C_{Rj}^{Jx})	Specific enzyme activity (U/mg protein)		Δ(activity)
		AS411	AS411E3	
GG1 (hexokinase)	0.00	0.07 ± 0.01	0.09 ± 0.01	0.02
GG2r (phosphoglucose isomerase)	0.49	0.57 ± 0.01	0.69 ± 0.01	0.12
GG3 (phosphofructokinase)	0.02	0.14 ± 0.01	0.13 ± 0.00	0.01

Table 10.2 Specific growth rate and PGI activity for AS411 compared to individual adapted clones of AS411E3. The specific activity of PGI was measured from cell extracts at 15 mM fructose-6-phosphate.

Cultures	Growth rate (hr^{-1})	PGI activity (U/mg-protein)
AS411	0.20 ± 0.02	0.58 ± 0.02
AS411E3 clone # 1	0.61 ± 0.04	0.68 ± 0.03
AS411E3 clone # 2	0.60 ± 0.02	0.67 ± 0.01
AS411E3 clone # 3	0.59 ± 0.01	0.73 ± 0.03
AS411E3 clone # 4	0.58 ± 0.01	0.71 ± 0.01
AS411E3 clone # 5	0.64 ± 0.05	0.69 ± 0.02
AS411E3 clone # 6	0.62 ± 0.02	0.68 ± 0.01
AS411E3 clone # 7	0.67 ± 0.05	0.70 ± 0.02
AS411E3 clone # 8	0.62 ± 0.02	0.69 ± 0.02
AS411E3 clone # 9	0.58 ± 0.02	0.68 ± 0.04
AS411E3 clone # 10	0.57 ± 0.01	0.66 ± 0.02

In summary, we identified phosphoglucose isomerase (PGI) as the enzyme with the most control on flux towards cell growth of AS411 based on flux control coefficients. This enzyme is predicted as the target of optimization that is responsible for the improved cell growth observed during evolution. The model prediction was confirmed by experimental results indicating the validity of metabolic control analysis approach. The enzymatic assays showed an increase in specific activity of PGI in the evolved culture AS411E3. The results suggest that cell adapts towards the fully evolved state with maximal cell growth by increasing the kinetic parameters of rate-controlling enzymes. That is, evolution and natural selection primarily target the enzymes with the greatest flux control, and the phenotypic change of the evolving cell is correlated qualitatively with

altered enzymatic parameters of the rate-controlling enzymes. The application of elementary mode analysis combining with flux control coefficients demonstrated here could be extended for identification of target overexpression gene for a rational design of an organism that can operate according to the fully evolved metabolic network.

Chapter 11

Research summary and Future work

11.1. Research summary

We have demonstrated in this dissertation the strain improvement by metabolic engineering guided by metabolic pathway analysis and metabolic evolution. The *in silico* metabolic network model and the pathway analysis tool called elementary mode analysis are implemented for characterizing cell metabolism and for designing cells with desired characteristics. The rational strain design is achieved by linking the cellular genotypes and phenotypes via the reconstructed metabolic model. To select for the desirable phenotype, multiple genes targeted for deletion are identified. The deletion of these reactions forces the cell to function only according to the desired pathways, resulting in the strain with the desirable phenotype. The metabolic evolution is also combined with a systems approach of inverse metabolic engineering to develop strains with an enhanced synthesis rate of products of interest or to improve strain robustness against toxic compounds.

This dissertation has demonstrated the application of these approaches in metabolic engineering of *E. coli* and *T. saccharolyticum*. The strains are optimized for an efficient production of primary metabolites such as ethanol and secondary metabolites such as carotenoid. The gene knockout mutant *E. coli* cells CRT028/pACMNOx and LYC018/pACEBI show an enhanced production of Diapolycopendioic acid and Lycopene respectively. The gene knockout mutant *T. saccharolyticum* AS411 is able to ferment sugars into ethanol at an increased yield compared to the wild-type. The mutants

are also tuned up by metabolic evolution, resulting in an evolved strain LYC018E1/pACEBI with further improvement of Lycopene and an evolved strain AS411E3 with an increase in production rate of ethanol. *In vivo* phenotypes of these strains are in agreement with the theoretical predictions by elementary mode analysis. Metabolic control analysis and thermodynamic analysis of the pathways are also applied to examine the control structure of metabolic pathways and the metabolic flux distribution chosen by the cells during evolution.

Furthermore, fermentation strategies are used to further promote the manufacture of desirable products. The controlled fed-batch is utilized for achieving a high titer of Diapolycopendioic acid. The immobilization of mixed cell cultures for an efficient production of ethanol at high yields and productivities from a mixture of sugars are also presented.

11.2. Future work

Chromosomal gene integration

E. coli strains used in this research for the production of carotenoid and ethanol harbor the plasmid expressing foreign genes for the synthesis of the products. Plasmid stability could be problematic in these strains since the recombinant cells could lose the plasmid over the long-term cell cultivation. For the production of carotenoid, the structural instability of the plasmids is also observed. In addition, supplying the culture with antibiotics to maintain the plasmid is not economical for large scale production of the products. Integration of these foreign genes into the chromosome of *E. coli* could resolve the plasmid stability issue. The homologous recombination techniques as

described in Datsenko and Wanner (2000) can be applied for chromosomal gene integration.

Identification of gene mutations

We have isolated strains exhibiting desirable phenotypes through metabolic evolution. In particular, we obtained LYC018E1/pACEBI for enhanced lycopene synthesis, AFF01/pLOI297 for increased tolerance against acetic acid and furfural, and AS411E3 for increased rate of cell growth and ethanol synthesis. Identification of genetic mutations through comparative genome sequencing or genomic microarrays can provide a better understanding of the cellular mechanisms corresponding to the desirable phenotypes acquired.

Accounting for regulation within metabolic network analysis

The metabolic network analysis applied in this dissertation solely relies on the reaction stoichiometry within the cell metabolic network. However, the flux distribution operated by the cell system is also subjected to regulatory mechanisms that determine the kinetic functions of enzymes and the expression of enzymes. Therefore, the accuracy of the metabolic network analysis can be improved by incorporating the kinetic and regulatory function of the network. Such information can be obtained from omics data using C^{13} labeling technique and/or microarray platform (Wiechert 2001).

Strategy for strain improvement

The EMA-guided, inverse metabolic engineering approach demonstrated in this dissertation has focused on identifying gene knockout targets for strain improvement. However, obtaining cells with desirable phenotypes often requires a combination of both

deletion and over- or underexpression of genes. Finding target genes for over- or underexpression is a challenging task since it requires the information of levels of mRNA, protein, enzyme activity, and flux for a particular reaction. Further studies are needed to develop novel strategies and algorithms to incorporate such information into the metabolic models. With these novel strategies and algorithms, a broad range of genetic modification targets that are a suitable optimization strategy for the strain improvements can be identified.

References

Al Zaid Siddiquee, K., Arauzo-Bravo, M. J., & Shimizu, K. (2004). Metabolic flux analysis of pykF gene knockout escherichia coli based on 13C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. *Applied Microbiology and Biotechnology, 63*(4), 407-417.

Albrecht, M., Misawa, N., & Sandmann, G. (1999). Metabolic engineering of the terpenoid biosynthetic pathway of *escherichia coli* for production of the carotenoids b-carotene and zeaxanthin. *Biotechnology Letters, 21*, 791-95.

Albrecht, M., Takaichi, S., Misawa, N., Schnurr, G., Boger, P., & Sandmann, G. (1997). Synthesis of atypical cyclic and acyclic hydroxyl carotenoids in *escherichia coli* transformants. *Journal of Biotechnology, 8*, 177–85.

Alexeeva, S., de Kort, B., Sawers, G., Hellingwerf, K. J., & de Mattos, M. J. (2000). Effects of limited aeration and of the ArcAB system on intermediary pyruvate catabolism in escherichia coli. *Journal of Bacteriology, 182*(17), 4934-4940.

Almeida, J.R.M., Modig, T., Petersson, A., Hahn-Hägerdal, B., Liden, G., Gorwa-Grauslund, M.F. (2007). Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae. Journal of Chemical Technology and Biotechnology*. 82(4), 340-349.

Alper, H., Jin, Y. S., Moxley, J. F., & Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in escherichia coli. *Metabolic Engineering, 7*(3), 155-164.

Alper, H., Miyaoku, K., & Stephanopoulos, G. (2006). Characterization of lycopene-overproducing E. coli strains in high cell density fermentations. *Applied Microbiology and Biotechnology, 72*(5), 968-974.

Alterthum, F., & Ingram, L. O. (1989). Efficient ethanol production from glucose, lactose, and xylose by recombinant escherichia coli. *Applied and Environmental Microbiology, 55*(8), 1943-1948.

Andersson, L., Strandberg, L., Haggstrom, L., & Enfors, S. O. (1994). Modeling of high cell density fed batch cultivation. *FEMS Microbiology Reviews, 14*(1), 39-44.

Asghari, A., Bothast, R. J., Doran, J. B., & Ingram, L. O. (1996). Ethanol production from hemicellulose hydrolysates of agricultural residues using genetically engineered *escherichia coli* strain KO11. *Journal of Industrail Microbiology and Biotechnology, 16*, 42-7.

Atsumi, S., Hanai, T., & Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature, 451*(7174), 86-89.

Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., et al. (1995). *Short protocols in molecular biology* Wiley & Sons, Inc. New York.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of escherichia coli K-12 in-frame, single-gene knockout mutants: The keio collection. *Molecular Systems Biology, 2*, 2006.0008.

Bachmann, B. J. (1996). Derivations and genotypes of some mutant derivatives of *escherichia coli* K-12. In F. C. Neidhardt, et al. (Eds.), *Escherichia coli and salmonella: Cellular and molecular biology* (F. C. Neidhardt, et al Trans.). (pp. 2460-88). Washington: ASM Press.

Bailey, J. E. (1991). Toward a science of metabolic engineering. *Science (New York, N.Y.), 252*(5013), 1668-1675.

Bailey, J. E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W. A., & Tsai, P. S. (2002). Inverse metabolic engineering: A strategy for directed genetic engineering of useful phenotypes. *Biotechnology and Bioengineering, 79*(5), 568-579.

Banerjee, N., Bhatnagar, R., & Viswanathan, L. (1981). Development of resistance in *saccharomyces cerevisiae* against inhibitory effects of browning reaction products. *Enzyme and Microbial Technology, 3*, 24-28.

Barbotin, J. N., Mater, D., Craynest, M., Saucedo, J. N., Truffaut, N., & Thomas, D. (1998). Immobilized cells: Plasmid stability and plasmid transfer. *Progress of Biotechnology, 15*, 591-602.

Bazaraa, M. S., Jarvis, J. J., & Sherali H.D. (2005). *Linear programming and network flows* Hoboken, John Wiley & Sons, Inc.

Becker, J., & Boles, E. (2003). A modified saccharomyces cerevisiae strain that consumes L-arabinose and produces ethanol. *Applied and Environmental Microbiology, 69*(7), 4144-4150.

Bettiga, M., Bengtsson, O., Hahn-Hagerdal, B., & Gorwa-Grauslund, M. F. (2009). Arabinose and xylose fermentation by recombinant saccharomyces cerevisiae expressing a fungal pentose utilization pathway. *Microbial Cell Factories, 8*, 40.

Bettiga, M., Hahn-Hagerdal, B., & Gorwa-Grauslund, M. F. (2008). Comparing the xylose reductase/xylitol dehydrogenase and xylose isomerase pathways in arabinose and xylose fermenting saccharomyces cerevisiae strains. *Biotechnology for Biofuels, 1*(1), 16.

Blattner, F. R., Plunkett, G.,3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., et al. (1997). The complete genome sequence of escherichia coli K-12. *Science (New York, N.Y.), 277*(5331), 1453-1462.

Bordel, S., & Nielsen, J. (2010). Identification of flux control in metabolic networks using non-equilibrium thermodynamics. Metabolic Engineering,

Britton, G. (1995). Structure and properties of carotenoids in relation to function. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology, 9*(15), 1551-1558.

Britton, G. (1998). Overview of carotenoid biosynthesis. In G. Britton (Ed.), *Carotenoids: Biosynthesis and metabolism* (pp. 13-47) Birkhäuser Basel.

Britton, G., Liaaen-Jensen, S., & Pfander H. (1995). Carotenoids. spectroscopy. (pp. 57-61) Birkhäuser Verlag Basel.

Bruinenberg, P. M., Peter, H. M., van Dijken, J. P., & Scheffers, W. A. (1983). The role of redox balances in the anaerobic fermentation of xylose by yeasts. *Eur.J.Appl.Microb.Biotech, 18*, 287-92.

Burgard, A. P., & Maranas, C. D. (2001). Probing the performance limits of the escherichia coli metabolic network subject to gene additions or deletions. *Biotechnology and Bioengineering, 74*(5), 364-375.

Carlson, R., Fell, D., & Srienc, F. (2002). Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnology and Bioengineering, 79*(2), 121-134.

Carlson, R., & Srienc, F. (2004). Fundamental escherichia coli biochemical pathways for biomass and energy production: Creation of overall flux states. *Biotechnology and Bioengineering, 86*(2), 149-162.

Carlson, R., & Srienc, F. (2004). Fundamental escherichia coli biochemical pathways for biomass and energy production: Identification of reactions. *Biotechnology and Bioengineering, 85*(1), 1-19.

Carlson, R., Wlaschin, A., & Srienc, F. (2005). Kinetic studies and biochemical pathway analysis of anaerobic poly-(R)-3-hydroxybutyric acid synthesis in escherichia coli. *Applied and Environmental Microbiology, 71*(2), 713-720.

Chang, D. E., Jung, H. C., Rhee, J. S., & Pan, J. G. (1999). Homofermentative production of D- or L-lactate in metabolically engineered escherichia coli RR1. *Applied and Environmental Microbiology, 65*(4), 1384-1389.

Chau, T. L., Guillan, A., Roca, E., Nunez, M. J., & Lema J.M. (2000). Enhancement of plasmid stability and enzymatic expression by immobilizing recombinant *saccharomyces cerevisiae*. *Biotechnology Letters, 22*, 1247-50.

Clark, D. P. (1989). The fermentation pathways of escherichia coli. *FEMS Microbiology Reviews, 5*(3), 223-234.

Clarke, B. L. (1988). Stoichiometric network analysis. *Cell Biophysics, 12*, 237-253.

Conn, P. F., Lambert, C., Land, E. J., Schalch, W., & Truscott, T. G. (1992). Carotene-oxygen radical interactions. *Free Radical Research Communications, 16*(6), 401-408.

Cork, J. M., & Purugganan, M. D. (2004). The evolution of molecular genetic pathways and networks. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology, 26(5), 479-484.

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America, 97*(12), 6640-6645.

de Graef, M. R., Alexeeva, S., Snoep, J. L., & Teixeira de Mattos, M. J. (1999). The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in escherichia coli. *Journal of Bacteriology, 181*(8), 2351-2357.

Demain, A. L., Davies, J. E., & Atlas, R. M. (Eds.). (1999). *Manual of industrial microbiology and biotechnology*. Washington, D.C.: ASM Press.

Desai, S. G., Guerinot, M. L., & Lynd, L. R. (2004). Cloning of L-lactate dehydrogenase and elimination of lactic acid production via gene knockout in thermoanaerobacterium saccharolyticum JW/SL-YS485. *Applied Microbiology and Biotechnology, 65*(5), 600-605.

Dien, B., Iten, L., & Bothast, R. (1999). Conversion of corn fiber to ethanol by recombinant E. coli strain FBR3. *Journal of Industrial Microbiology & Biotechnology, 22*(6), 575-581.

Dien, B. S., Hespell, R. B., Ingram, L. O., & Bothast, R. J. (1997). Conversion of corn milling fibrous co-products into ethanol by recombinant *escherichia coli* strains K011 and SL40. *World Journal of Microbiology and Biotechnology, 13*, 619-25.

Dien, B. S., Nagle, N., Hicks, K. B., Singh, V., Moreau, R. A., Tucker, M. P., et al. (2004). Fermentation of "quick fiber" produced from a modified corn-milling process into ethanol and recovery of corn fiber. *Applied Biochemistry and Biotechnology, 113-116*, 937-949.

Dien, B. S., Nichols, N. N., O'Bryan, P. J., & Bothast, R. J. (2000). Development of new ethanologenic escherichia coli strains for fermentation of lignocellulosic biomass. *Applied Biochemistry and Biotechnology, 84-86*, 181-196.

Dill, K. A., & Bromberg, S. (2003). *Molecular driving forces: Statistical thermodynamics in chemistry and biology*. New York: Garland Science.

Dincbas, V., Hortacsu, A., & Camurdan, A. (1993). Plasmid stability in immobilized mixed cultures of recombinant escherichia coli. *Biotechnology Progress, 9*(2), 218-220.

Diniz, S. C., Voss, I., & Steinbuchel, A. (2006). Optimization of cyanophycin production in recombinant strains of pseudomonas putida and ralstonia eutropha employing elementary mode analysis and statistical experimental design. *Biotechnology and Bioengineering, 93*(4), 698-717.

Dumsday, G. J., Zhou, B., Yaqin, W., Stanley, G. A., & Pamment, N. B. (1999). Comparative stability of ethanol production by escherichia coli KO11 in batch and chemostat culture. *Journal of Industrial Microbiology & Biotechnology, 23*(1), 701-708.

Edwards, J. S., Ibarra, R. U., & Palsson, B. O. (2001). In silico predictions of escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology, 19*(2), 125-130.

Edwards, J. S., & Palsson, B. O. (2000). Robustness analysis of the escherichia coli metabolic network. *Biotechnology Progress, 16*(6), 927-939.

Edwards, J. S., Ramakrishna, R., Schilling, C. H., & Palsson, B. O. (1999). Metabolic flux balance analysis. In S. Y. Lee, & E. T. Papoutsakis (Eds.), *Metabolic engineering* (pp. 13-57). New York: Marcel Dekker, Inc.

El-Mansi, M. (2004). Flux to acetate and lactate excretions in industrial fermentations: Physiological and biochemical implications. *Journal of Industrial Microbiology & Biotechnology, 31*(7), 295-300.

Farmer, W. R., & Liao, J. C. (2000). Improving lycopene production in escherichia coli by engineering metabolic control. *Nature Biotechnology, 18*(5), 533-537.

Farmer, W. R., & Liao, J. C. (2001). Precursor balancing for metabolic engineering of lycopene production in escherichia coli. *Biotechnology Progress, 17*(1), 57-61.

Farrell, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O'Hare, M., & Kammen, D. M. (2006). Ethanol can contribute to energy and environmental goals. *Science (New York, N.Y.), 311*(5760), 506-508.

Fazekas, Z., Gao, D., Saladi, R. N., Lu, Y., Lebwohl, M., & Wei, H. (2003). Protective effects of lycopene against ultraviolet B-induced photodamage. *Nutrition and Cancer, 47*(2), 181-187.

Fell, D. (1997). In snell k. (Ed.), Understanding the control of metabolism. London and Miami: Portland Press.

Flowers, J. M., Sezgin, E., Kumagai, S., Duvernell, D. D., Matzkin, L. M., Schmidt, P. S., et al. (2007). Adaptive evolution of metabolic pathways in drosophila. Molecular Biology and Evolution, 24(6), 1347-1354.

Fidler, S., & Dennis, D. (1992). Polyhydroxyalkanoate production in recombinant escherichia coli. *FEMS Microbiology Reviews, 9*(2-4), 231-235.

Fong, S. S., Marciniak, J. Y., & Palsson, B. O. (2003). Description and interpretation of adaptive evolution of escherichia coli K-12 MG1655 by using a genome-scale in silico metabolic model. *Journal of Bacteriology, 185*(21), 6400-6408.

Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., et al. (2005). In silico design and adaptive evolution of escherichia coli for production of lactic acid. *Biotechnology and Bioengineering, 91*(5), 643-648.

Fong, S. S., & Palsson, B. O. (2004). Metabolic gene-deletion strains of escherichia coli evolve to computationally predicted growth phenotypes. Nature Genetics, 36(10), 1056-1058.

Fraser, P. D., Misawa, N., Linden, H., Yamano, S., Kobayashi, K., & Sandmann, G. (1992). Expression in escherichia coli, purification, and reactivation of the recombinant erwinia uredovora phytoene desaturase. *The Journal of Biological Chemistry, 267*(28), 19891-19895.

Gagneur, J., & Klamt, S. (2004). Computation of elementary modes: A unifying framework and the new binary approach. *BMC Bioinformatics, 5*, 175.

Gilson, C. D., & Thomas, A. (1995). Ethanol production by alginate immobilized yeast in a fluidized bed bioreactor. *Journal Chemical Technological Biotechnology, 62*, 38-45.

Giordano, R. L., Trovati, J., & Schmidell, W. (2008). Continuous production of ethanol from starch using glucoamylase and yeast co-immobilized in pectin gel. *Applied Biochemistry and Biotechnology, 147*(1-3), 47-61.

Hansen, T., Oehlmann, M., & Schonheit, P. (2001). Novel type of glucose-6-phosphate isomerase in the hyperthermophilic archaeon pyrococcus furiosus. *Journal of Bacteriology, 183*(11), 3428-3435.

Hua, Q., Joyce, A. R., Fong, S. S., & Palsson, B. O. (2006). Metabolic analysis of adaptive evolution for in silico-designed lactate-producing strains. Biotechnology and Bioengineering, 95(5), 992-1002.

Heinrich, R., & Schuster, S. (1996). *The regulation of cellular systems* (pp. 75-111). New York: Chapman and Hall.

Heinrich, R., & Schuster, S. (1998). The modelling of metabolic systems. structure, control and optimality. *Bio Systems, 47*(1-2), 61-77.

Ho, N. W., Chen, Z., & Brainard, A. P. (1998). Genetically engineered saccharomyces yeast capable of effective cofermentation of glucose and xylose. *Applied and Environmental Microbiology, 64*(5), 1852-1859.

Hogsett, D. A. (1995). Cellulose hydrolysis and fermentation by clostridium thermocellum for the production of ethanol. (Doctoral thesis, Dartmouth College).

http://ecocyc.org

Ibarra, R. U., Edwards, J. S., & Palsson, B. O. (2002). Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature, 420*(6912), 186-189.

Ingraham, J. L., Maaloe, O., & Neidhardt, F. C. (1983). *Growth of the bacterial cell*. Sunderland: Sinauer Associates Inc.

Ingram, L.O. (1986). Microbial tolerance to alcohols: Role of the cell membrane. *Trends in Biotechnology*, 4, 40–44.

Ingram, L. O., Aldrich, H. C., Borges, A. C., Causey, T. B., Martinez, A., Morales, F., et al. (1999). Enteric bacterial catalysts for fuel ethanol production. *Biotechnology Progress, 15*(5), 855-866.

Iuchi, S., & Weiner, L. (1996). Cellular and molecular physiology of escherichia coli in the adaptation to aerobic environments. *Journal of Biochemistry, 120*(6), 1055-1063.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physics Review, 106*, 620-630.

Karhumaa, K., Wiedemann, B., Hahn-Hagerdal, B., Boles, E., & Gorwa-Grauslund, M. F. (2006). Co-utilization of L-arabinose and D-xylose by laboratory and industrial saccharomyces cerevisiae strains. *Microbial Cell Factories, 5*, 18.

Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., et al. (2005). EcoCyc: A comprehensive database resource for escherichia coli. *Nucleic Acids Research, 33*(Database issue), D334-7.

Kim, S. W., & Keasling, J. D. (2001). Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in escherichia coli enhances lycopene production. *Biotechnology and Bioengineering, 72*(4), 408-415.

King, V. A., & Zall, R. R. (1985). Ethanol fermentation of whey using calcium alginate entrapped yeasts. *Process Biochemistry, 18*, 17-20.

Klamt, S., & Schuster, S. (2002). Calculating as many fluxes as possible in underdetermined metabolic networks. *Molecular Biology Reports, 29*(1-2), 243-248.

Klamt, S., & Stelling, J. (2002). Combinatorial complexity of pathway analysis in metabolic networks. *Molecular Biology Reports, 29*(1-2), 233-236.

Klinke, H. B., Thomsen, A. B., & Ahring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Applied Microbiology and Biotechnology, 66*(1), 10-26.

Kotter, P., & Ciriacy, M. (1993). Xylose fermentation by *saccharomyces cerevisiae*. *Applied Microbiology Biotechnology, 38*, 776-83.

Kromer, J. O., Wittmann, C., Schroder, H., & Heinzle, E. (2006). Metabolic pathway analysis for rational design of L-methionine production by escherichia coli and corynebacterium glutamicum. *Metabolic Engineering, 8*(4), 353-369.

Kuzuyama, T. (2002). Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Bioscience, Biotechnology, and Biochemistry, 66*(8), 1619-1627.

Lange, B. M., Rujan, T., Martin, W., & Croteau, R. (2000). Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences of the United States of America, 97*(24), 13172-13177.

Lasko, D. R., Zamboni, N., & Sauer, U. (2000). Bacterial response to acetate challenge: A comparison of tolerance among species. *Applied Microbiology and Biotechnology, 54*(2), 243-247.

Lee, P. C., Mijts, B. N., & Schmidt-Dannert, C. (2004). Investigation of factors influencing production of the monocyclic carotenoid torulene in metabolically engineered escherichia coli. *Applied Microbiology and Biotechnology, 65*(5), 538-546.

Lee, P. C., & Schmidt-Dannert, C. (2002). Metabolic engineering towards biotechnological production of carotenoids in microorganisms. *Applied Microbiology and Biotechnology, 60*(1-2), 1-11.

Lee, S. H., Park, S. J., Lee, S. Y., & Hong, S. H. (2008). Biosynthesis of enantiopure (S)-3-hydroxybutyric acid in metabolically engineered escherichia coli. *Applied Microbiology and Biotechnology, 79*(4), 633-641.

Lee, Y. E., Jain, M. K., Lee, C., Lowe, S. E., & Zeikus, J. G. (1993). Taxonomic distinction of saccharolytic thermophilic anaerobes: Description of thermoanaerobacterium xylanolyticum gen. nov., sp. nov., and thermoanaerobacterium saccharolyticum gen. nov., sp. nov.; reclassification of thermoanaerobium brockii, clostridium thermosulfurogenes, and clostridium thermohydrosulfuricum E100-69 as thermoanaerobacter brockii comb. nov., thermoanaerobacterium thermosulfurigenes comb. nov., and thermoanaerobacter thermohydrosulfuricus comb. nov., respectively; and transfer of clostridium thermohydrosulfuricum 39E to thermoanaerobacter ethanolicus. *International Journal of System Bacteriology, 43*, 41-51.

Lendenmann, U., & Egli, T. (1998). Kinetic models for the growth of escherichia coli with mixtures of sugars under carbon-limited conditions. *Biotechnology and Bioengineering, 59*(1), 99-107.

Lendenmann, U., Snozzi, M., & Egli, T. (1996). Kinetics of the simultaneous utilization of sugar mixtures by escherichia coli in continuous culture. *Applied and Environmental Microbiology, 62*(5), 1493-1499.

Liao, J. C., Hou, S. Y., & Chao, Y. P. (1996). Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnology and Bioengineering, 52*(1), 129-140.

Lichtenthaler, H. K. (2000). Non-mevalonate isoprenoid biosynthesis: Enzymes, genes and inhibitors. *Biochemical Society Transactions, 28*(6), 785-789.

Lichtenthaler, H. K., Rohmer, M., & Schwender, J. (1997). Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants. *Physiology Plants, 101*, 643-52.

Lindsay, S. E., Bothast, R. J., & Ingram, L. O. (1995). Improved strains of recombinant escherichia coli for ethanol production from sugar mixtures. *Applied Microbiology and Biotechnology, 43*(1), 70-75.

Liu, Q., Ouyang, S. P., Chung, A., Wu, Q., & Chen, G. Q. (2007). Microbial production of R-3-hydroxybutyric acid by recombinant E. coli harboring genes of phbA, phbB, and tesB. *Applied Microbiology and Biotechnology, 76*(4), 811-818.

Liu, S. Y., Gherardini, F. C., Matuschek, M., Bahl, H., & Wiegel, J. (1996). Cloning, sequencing, and expression of the gene encoding a large S-layer-associated endoxylanase from thermoanaerobacterium sp. strain JW/SL-YS 485 in escherichia coli. *Journal of Bacteriology, 178*(6), 1539-1547.

Liu, S. Y., Rainey, F. A., Morgan, H. W., Mayer, F., & Wiegel, J. (1996). *Thermoanaerobacterium aotearoense* sp. nov., a slightly acidophilic, anaerobic thermophile isolated from various hot springs in new zealand, and emendation of the genus *thermoanaerobacterium*. *International Journal System Bacteriology, 46*, 388-96.

Liu, S. Y., Wiegel, J., & Gherardini, F. C. (1996). Purification and cloning of a thermostable xylose (glucose) isomerase with an acidic pH optimum from thermoanaerobacterium strain JW/SL-YS 489. *Journal of Bacteriology, 178*(20), 5938-5945.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular cell biology*. New York: WH. Freeman & Co.

Lorenz, W. W., & Wiegel, J. (1997). Isolation, analysis, and expression of two genes from thermoanaerobacterium sp. strain JW/SL YS485: A beta-xylosidase and a novel acetyl xylan esterase with cephalosporin C deacetylase activity. *Journal of Bacteriology, 179*(17), 5436-5441.

Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., et al. (2008). How biotech can transform biofuels. *Nature Biotechnology, 26*(2), 169-172.

Lynd, L. R., van Zyl, W. H., McBride, J. E., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: An update. *Current Opinion in Biotechnology, 16*(5), 577-583.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., & Pretorius, I. S. (2002). Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews: MMBR, 66*(3), 506-77.

Mai, V., Lorenz, W. W., & Wiegel, J. (1997). Transformation of *thermoaerobacterium sp* strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance. *FEMS Microbiology Letter, 148*, 163–67.

Mai, V., & Wiegel, J. (2000). Advances in development of a genetic system for thermoanaerobacterium spp.: Expression of genes encoding hydrolytic enzymes, development of a second shuttle vector, and integration of genes into the chromosome. *Applied and Environmental Microbiology, 66*(11), 4817-4821.

Martin, G. J., Knepper, A., Zhou, B., & Pamment, N. B. (2006). Performance and stability of ethanologenic escherichia coli strain FBR5 during continuous culture on xylose and glucose. *Journal of Industrial Microbiology & Biotechnology, 33*(10), 834-844.

Martinez, A., Grabar, T. B., Shanmugam, K. T., Yomano, L. P., York, S. W., & Ingram, L. O. (2007). Low salt medium for lactate and ethanol production by recombinant escherichia coli B. *Biotechnology Letters, 29*(3), 397-404.

Martinez-Morales, F., Borges, A. C., Martinez, A., Shanmugam, K. T., & Ingram, L. O. (1999). Chromosomal integration of heterologous DNA in escherichia coli with precise removal of markers and replicons used during construction. *Journal of Bacteriology, 181*(22), 7143-7148.

Martyushev, L. M., & Seleznev, V. D. (2006). Maximum entropy production principle in physics, chemistry and biology. *Physics Reports-Review Section of Physics Letters, 426*, 1-45.

Matsushika, A., Inoue, H., Murakami, K., Takimura, O., & Sawayama, S. (2009). Bioethanol production performance of five recombinant strains of laboratory and industrial xylose-fermenting saccharomyces cerevisiae. *Bioresource Technology, 100*(8), 2392-2398.

McMillan, J. D. (1994). Conversion of hemicellulose hydrolysates to ethanol. In M. E. Himmel, J. O. Baker & R. P. Overend (Eds.), *Enzymatic conversion of biomass for fuels production* (pp. American Chemical Society p 411-37.). Washington, D.C.: ACS symposium series 566.

Mielenz, J. R. (2001). Ethanol production from biomass: Technology and commercialization status. *Current Opinion in Microbiology, 4*(3), 324-329.

Mijts, B. N., Lee, P. C., & Schmidt-Dannert, C. (2005). Identification of a carotenoid oxygenase synthesizing acyclic xanthophylls: Combinatorial biosynthesis and directed evolution. *Chemistry & Biology, 12*(4), 453-460.

Miller, J. H. (1992). *A short course in bacterial genetics. A laboratory manual and handbook for escherichia coli and related bacteria*. Plainview, N.Y.: Cold Spring Harbor Laboratory Press.

Moes, C. J., Pretorius, I. S., & vanZyl, W. H. (1996). Cloning and expression of the clostridium thermosulfurogenes D-xylose isomerase gene (xylA) in saccharomyces cerevisiae. *Biotechnology Letters, 18*, 269-74.

Moniruzzaman, M., Dien, B. S., Ferrer, B., Hespell, R. B., Dale, B. E., Ingram, L. O., et al. (1996). Ethanol production from afex pretreated corn fiber by recombinant bacteria. *Biotechnology Letters, 18*, 955-990.

Moreno-Sanchez, R., Saavedra, E., Rodriguez-Enriquez, S., & Olin-Sandoval, V. (2008). Metabolic control analysis: A tool for designing strategies to manipulate metabolic pathways. Journal of Biomedicine & Biotechnology, 2008, 597913.

Mori, H., Isono, K., Horiuchi, T., & Miki, T. (2000). Functional genomics of *escherichia coli* in japan. *Research in Microbiology, 151*(2), 121-28.

Najafpour, G. D. (1990). Immobilization of microbial cells for the production of organic acids. *Journal of Science International Research in Iran, 1*, 172–76.

Neidhardt, F. C. (1987). *Escherichia coli* and *salmonella typhimurium*: Cellular and molecular biology. (pp. 1654, xlvii p.). Washington, D.C.: American Society for Microbiology.

Neidhardt, F. C. (1996). *Escherichia coli and salmonella: Cellular and molecular biology*. Washington,D.C.: ASM Press.

Nichols, N. N., Dien, B. S., & Bothast, R. J. (2001). Use of catabolite repression mutants for fermentation of sugar mixtures to ethanol. *Applied Microbiology and Biotechnology, 56*(1-2), 120-125.

Nigam, J. N. (2000). Continuous ethanol production from pineapple cannery waste using immobilized yeast cells. *Journal of Biotechnology, 80*(2), 189-193.

Nishino, H. (1998). Cancer prevention by carotenoids. *Mutation Research, 402*(1-2), 159-163.

O'Brien, D. J., Senske, G. E., Kurantz, M. J., & Craig, J. C.,Jr. (2004). Ethanol recovery from corn fiber hydrolysate fermentations by pervaporation. *Bioresource Technology, 92*(1), 15-19.

Ong, A., & Tee, E. (1992). Natural sources of carotenoids from plants and oils. *Metheods Enzymology, 213*, 142-67.

Ouzounis, C. A., & Karp, P. D. (2000). Global properties of the metabolic map of escherichia coli. *Genome Research, 10*(4), 568-576.

Palmqvist, E. (1998). Fermentation of lignocellulosic hydrolysates: Inhibition and detoxification. (PhD thesis, Department of Applied Microbiology, Lund University).

Palmqvist, E., Grage, H., Meinander, N.Q., Hahn-Hagerdal, B. (1999a). Main and interaction effects of acetic acid, furfural, and *p*-hydroxybenzoic acid on growth and ethanol productivity of yeasts. *Biotechnology & Bioengineering*, 63, 46–55.

Palmqvist, E., Almeida, J.S., Hahn-Hagerdal, B. (1999b). Influence of furfural on anaerobic glycolytic kinetics of *Saccharomyces cerevisiae* in batch culture. *Biotechnology & Bioengineering*, 62, 447–454.

Papin, J. A., Price, N. D., Edwards, J. S., & Palsson, B. B. O. (2002). The genome-scale metabolic extreme pathway structure in haemophilus influenzae shows significant network redundancy. *Journal of Theoretical Biology, 215*(1), 67-82.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J. C., Montero, F., & Schuster, S. (1999). METATOOL: For studying metabolic networks. *Bioinformatics (Oxford, England), 15*(3), 251-257.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J. C., Montero, F., & Schuster, S. (1999). METATOOL: For studying metabolic networks. *Bioinformatics (Oxford, England), 15*(3), 251-257.

Poolman, M. G., Venkatesh, K. V., Pidcock, M. K., & Fell, D. A. (2004). A method for the determination of flux in elementary modes, and its application to lactobacillus rhamnosus. *Biotechnology and Bioengineering, 88*(5), 601-612.

Posfai, G., Koob, M. D., Kirkpatrick, H. A., & Blattner, F. R. (1997). Versatile insertion plasmids for targeted genome manipulations in bacteria: Isolation, deletion, and rescue of the pathogenicity island LEE of the escherichia coli O157:H7 genome. *Journal of Bacteriology, 179*(13), 4426-4428.

Price, N. D., Papin, J. A., & Palsson, B. O. (2002). Determination of redundancy and systems properties of the metabolic network of helicobacter pylori using genome-scale extreme pathway analysis. *Genome Research, 12*(5), 760-769.

Price, N. D., Reed, J. L., & Palsson, B. O. (2004). Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nature Reviews.Microbiology, 2*(11), 886-897.

Prigogine, I. (1967). *Introduction to thermodynamics of irreversible processes*. New York: Interscience Publishers.

Qian, H., & Beard, D. A. (2005). Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophysical Chemistry, 114*(2-3), 213-220.

Qian, H., Beard, D. A., & Liang, S. D. (2003). Stoichiometric network theory for nonequilibrium biochemical systems. *European Journal of Biochemistry / FEBS, 270*(3), 415-421.

Qureshi, N., Dien, B. S., Nichols, N. N., Liu, S., Hughes, S. R., Iten, L. B., et al. (2005). Continuous production of ethanol in high productivity bioreactors using genetically engineered *escherichia coli* FBR5: Membrane and fixed cell reactors *American Institute of Chemical Engineers,* Paper No. 589g.

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., et al. (2006). The path forward for biofuels and biomaterials. *Science (New York, N.Y.), 311*(5760), 484-489.

Reed, J. L., & Palsson, B. O. (2003). Thirteen years of building constraint-based in silico models of escherichia coli. *Journal of Bacteriology, 185*(9), 2692-2699.

Riesenberg, D. (1991). High-cell-density cultivation of escherichia coli. *Current Opinion in Biotechnology, 2*(3), 380-384.

Roels, J. A. (1983). *Energetics and kinetics in biotechnology*. New York: Elsevier Biomedical Press.

Rohmer, M., Knani, M., Simonin, P., Sutter, B., & Sahm, H. (1993). Isoprenoid biosynthesis in bacteria: A novel pathway for the early steps leading to isopentenyl diphosphate. *The Biochemical Journal, 295 (Pt 2)*(Pt 2), 517-524.

Ruther, A., Misawa, N., Boger, P., & Sandmann, G. (1997). Production of zeaxanthin in escherichia coli transformed with different carotenogenic plasmids. *Applied Microbiology and Biotechnology, 48*(2), 162-167.

Sanchez, A. M., Bennett, G. N., & San, K. Y. (2006). Batch culture characterization and metabolic flux analysis of succinate-producing escherichia coli strains. *Metabolic Engineering, 8*(3), 209-226.

Sandler, S., & Orbey, H. (1991). On the thermodynamics of microbial growth processes. *Biotechnology and Bioengineering, 38*(7), 697-718.

Sandmann, G., Albrecht, M., Schnurr, G., Knorzer, O., & Boger, P. (1999). The biotechnological potential and design of novel carotenoids by gene combination in escherichia coli. *Trends in Biotechnology, 17*(6), 233-237.

Schilling, C. H., Edwards, J. S., Letscher, D., & Palsson, B. O. (2000). Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and Bioengineering, 71*(4), 286-306.

Schilling, C. H., Letscher, D., & Palsson, B. O. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology, 203*(3), 229-248.

Schilling, C. H., & Palsson, B. O. (2000). Assessment of the metabolic capabilities of haemophilus influenzae rd through a genome-scale pathway analysis. *Journal of Theoretical Biology, 203*(3), 249-283.

Schuster, S., Fell, D. A., & Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology, 18*(3), 326-332.

Schuster, S., Hilgetag, C., & Fell, D. (1994). Detecting elementary modes of functioning in metabolic networks. *Modern Trends in BioThermoKinetics, 3*, 103-05.

Schuster, S., Hilgetag, C., Woods, J. H., & Fell, D. A. (2002). Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *Journal of Mathematical Biology, 45*(2), 153-181.

Schuster, S., Pfeiffer, T., & Fell, D. A. (2008). Is maximization of molar yield in metabolic networks favoured by evolution? Journal of Theoretical Biology, 252(3), 497-504.

Shanks, R. M., Caiazza, N. C., Hinsa, S. M., Toutain, C. M., & O'Toole, G. A. (2006). Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Applied and Environmental Microbiology, 72*(7), 5027-5036.

Shao, W., Deblois, S., & Wiegel, J. (1995). A high-molecular-weight, cell-associated xylanase isolated from exponentially growing thermoanaerobacterium sp. strain JW/SL-YS485. *Applied and Environmental Microbiology, 61*(3), 937-940.

Shao, W., Obi, S., Puls, J., & Wiegel, J. (1995). Purification and characterization of the (alpha)-glucuronidase from thermoanaerobacterium sp. strain JW/SL-YS485, an important enzyme for the utilization of substituted xylans. *Applied and Environmental Microbiology, 61*(3), 1077-1081.

Shaw, A. J., Jenney, F. E., Adams, M. W., & Lynd, L. R. (2008). End-product pathways in the xylose fermenting bacterium, thermoanaerobacterium saccharolyticum. *Enzyme Microbial Technology, 42*, 453-58.

Shaw, A. J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G., et al. (2008). Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Sciences of the United States of America, 105*(37), 13769-13774.

Shaw, A. J., Podkaminer, K. K., Desai, S. G., Bardsley, J. S., Rogers, S. R., Thorne, P. G., et al. (2008). Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proceedings of the National Academy of Sciences of the United States of America, 105*(37), 13769-13774.

Soni, R., Carmichael, J. P., & Murray, J. A. (1993). Parameters affecting lithium acetate-mediated transformation of saccharomyces cerevisiae and development of a rapid and simplified procedure. *Current Genetics, 24*(5), 455-459.

Sprenger, G. A. (1995). Genetics of pentose-phosphate pathway enzymes of escherichia coli K-12. *Archives of Microbiology, 164*(5), 324-330.

Srienc, F., & Unrean, P. (2010). A Statistical Thermodynamical Interpretation of Metabolism. Entropy, submitted

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., & Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature, 420*(6912), 190-193.

Stephanopoulos, G., Aristidou, A., & Neilson, J. (1998). *Metabolic engineering: Principles and methodologies*. New York: Academic Press.

Takahashi, C. M., Takahashi, D. F., Carvalhal, M. L., & Alterthum, F. (1999). Effects of acetate on the growth and fermentation performance of escherichia coli KO11. *Applied Biochemistry and Biotechnology, 81*(3), 193-203.

Takamitsu, I., Izumida, H., Akagi, Y., & Sakamoto, M. (1993). Continuous ethanol fermentation in molasses medium using zymomonas mobilis immobilized in photo-cross linkable resin gels. *Journal of Fermentation Bioengineering, 75*, 32–35.

Tang, Y. J., Sapra, R., Joyner, D., Hazen, T. C., Myers, S., Reichmuth, D., et al. (2009). Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant geobacillus strain. *Biotechnology and Bioengineering, 102*(5), 1377-1386.

Tao, L., Schenzle, A., Odom, J. M., & Cheng, Q. (2005). Novel carotenoid oxidase involved in biosynthesis of 4,4'-diapolycopene dialdehyde. *Applied and Environmental Microbiology, 71*(6), 3294-3301.

Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R. A., & Smid, E. J. (2009). Understanding the adaptive growth strategy of lactobacillus plantarum by in silico optimisation. *PLoS Computational Biology, 5*(6), e1000410.

Thomas, C. F., Janet, E., Hansen, M. H., JoAnn, H., George, L. L., Laurane, G. M., et al. (1987). Recombinant human insulin-like growth factor II expressed in *escherichia coli*. *BioTechnology, 5*(1047-51)

Tong, I. T., Liao, H. H., & Cameron, D. C. (1991). 1,3-propanediol production by escherichia coli expressing genes from the klebsiella pneumoniae dha regulon. *Applied and Environmental Microbiology, 57*(12), 3541-3546.

Trinh, C. T., Carlson, R., Wlaschin, A., & Srienc, F. (2006). Design, construction and performance of the most efficient biomass producing E. coli bacterium. *Metabolic Engineering, 8*(6), 628-638.

Trinh, C. T., Unrean, P., & Srienc, F. (2008). Minimal escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. *Applied and Environmental Microbiology, 74*(12), 3634-3643.

Trinh, C. T., Wlaschin, A., & Srienc, F. (2009). Elementary mode analysis: A useful metabolic pathway analysis tool for characterizing cellular metabolism. Applied Microbiology and Biotechnology, 81(5), 813-826.

Unrean, P., Trinh, C. T., & Srienc, F. (2010). Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. *Metabolic Engineering*, *12*(2), 112-122.

Unrean, P., & Srienc, F. (2010). Metabolic networks evolve towards states of maximum entropy production, in preparation

Van Dien, S. J., & Lidstrom, M. E. (2002). Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph methylobacterium extorquens AM1, with application to reconstruction of C(3) and C(4) metabolism. *Biotechnology and Bioengineering, 78*(3), 296-312.

Verbelen, P. J., De Schutter, D. P., Delvaux, F., Verstrepen, K. J., & Delvaux, F. R. (2006). Immobilized yeast cell systems for continuous fermentation applications. *Biotechnology Letters, 28*(19), 1515-1525.

Vershinin, A. (1999). Biological functions of carotenoids--diversity and evolution. *BioFactors (Oxford, England), 10*(2-3), 99-104.

Verwaal, R., Wang, J., Meijnen, J. P., Visser, H., Sandmann, G., van den Berg, J. A., et al. (2007). High-level production of beta-carotene in saccharomyces cerevisiae by successive transformation with carotenogenic genes from xanthophyllomyces dendrorhous. *Applied and Environmental Microbiology, 73*(13), 4342-4350.

Vijayasankaran, N., Carlson, R., & Srienc, F. (2005). Metabolic pathway structures for recombinant protein synthesis in escherichia coli. *Applied Microbiology and Biotechnology*, *68*(6), 737-746.

Vitkup, D., Kharchenko, P., & Wagner, A. (2006). Influence of metabolic network structure and function on enzyme evolution. Genome Biology, 7(5), R39.

Voet, D., & Voet, J. (1995). *Biochemistry*. New York: John WIley & Sons Inc.

Von Kamp, A., & Schuster, S. (2006). Metatool 5.0: Fast and flexible elementary modes analysis. *Bioinformatics (Oxford, England), 22*(15), 1930-1931.

Walfridsson, M., Bao, X., Anderlund, M., Lilius, G., Bulow, L., & Hahn-Hagerdal, B. (1996). Ethanolic fermentation of xylose with saccharomyces cerevisiae harboring the thermus thermophilus xylA gene, which expresses an active xylose (glucose) isomerase. *Applied and Environmental Microbiology, 62*(12), 4648-4651.

Walfridsson, M., Hallborn, J., Penttila, M., Keranen, S., & Hahn-Hagerdal, B. (1995). Xylose-metabolizing saccharomyces cerevisiae strains overexpressing the TKL1 and TAL1 genes encoding the pentose

phosphate pathway enzymes transketolase and transaldolase. *Applied and Environmental Microbiology, 61*(12), 4184-4190.

Wang, C. W., Oh, M. K., & Liao, J. C. (1999). Engineered isoprenoid pathway enhances astaxanthin production in escherichia coli. *Biotechnology and Bioengineering, 62*(2), 235-241.

Wiechert, W. (2001). 13C metabolic flux analysis. *Metabolic Engineering, 3*(3), 195-206.

Wieland, B., Feil, C., Gloria-Maercker, E., Thumm, G., Lechner, M., Bravo, J. M., et al. (1994). Genetic and biochemical analyses of the biosynthesis of the yellow carotenoid 4,4'-diaponeurosporene of staphylococcus aureus. *Journal of Bacteriology, 176*(24), 7719-7726.

Wisselink, H. W., Toirkens, M. J., del Rosario Franco Berriel, M., Winkler, A. A., van Dijken, J. P., Pronk, J. T., et al. (2007). Engineering of saccharomyces cerevisiae for efficient anaerobic alcoholic fermentation of L-arabinose. *Applied and Environmental Microbiology, 73*(15), 4881-4891.

Wlaschin, A. P., Trinh, C. T., Carlson, R., & Srienc, F. (2006). The fractional contributions of elementary modes to the metabolism of escherichia coli and their estimation from reaction entropies. *Metabolic Engineering, 8*(4), 338-352.

Wyman, C. E. (2001). Twenty years of trials, tribulations, and research progress in bioethanol technology: Selected key events along the way. *Applied Biochemistry and Biotechnology, 91-93*, 5-21.

Yamada, T., Fatigati, M. A., & Zhang, M. (2002). Performance of immobilized zymomonas mobilis 31821 (pZB5) on actual hydrolysates produced by arkenol technology. *Applied Biochemistry and Biotechnology, 98-100*, 899-907.

Yang, C., Hua, Q., Baba, T., Mori, H., & Shimizu, K. (2003). Analysis of escherichia coli anaplerotic metabolism and its regulation mechanisms from the metabolic responses to altered dilution rates and phosphoenolpyruvate carboxykinase knockout. *Biotechnology and Bioengineering, 84*(2), 129-144.

Yoon, S. H., Kim, J. E., Lee, S. H., Park, H. M., Choi, M. S., Kim, J. Y., et al. (2007). Engineering the lycopene synthetic pathway in *E. coli* by comparison of the carotenoid genes of *pantoea agglomerans* and *pantoea ananatis*. *Applied Microbiology Biotechnology, 74*(1), 131-39.

Zaldivar, J., & Ingram, L. O. (1999). Effect of organic acids on the growth and fermentation of ethanologenic escherichia coli LY01. *Biotechnology and Bioengineering, 66*(4), 203-210.

Zaldivar, J., Martinez, A., & Ingram, L. O. (1999). Effect of selected aldehydes on the growth and fermentation of ethanologenic escherichia coli. *Biotechnology and Bioengineering, 65*(1), 24-33.

Zaldivar, J., Martinez, A., & Ingram, L. O. (2000). Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic escherichia coli. *Biotechnology and Bioengineering, 68*(5), 524-530.

Zaldivar, J., Nielsen, J., & Olsson, L. (2001). Fuel ethanol production from lignocellulose: A challenge for metabolic engineering and process integration. *Applied Microbiology and Biotechnology, 56*(1-2), 17-34.

Zhang, Y., & Lynd, L. R. (2003). Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: Development of an enzyme-linked immunosorbent assay-based method with application to clostridium thermocellum batch cultures. *Analytical Chemistry, 75*(2), 219-227.

Zhang, Y. H., & Lynd, L. R. (2005). Regulation of cellulase synthesis in batch and continuous cultures of clostridium thermocellum. *Journal of Bacteriology, 187*(1), 99-106.

Zhao, Q., & Kurata, H. (2009). Maximum entropy decomposition of flux distribution at steady state to elementary modes. Journal of Bioscience and Bioengineering, 107(1), 84-89.

Zhao, Q., & Kurata, H. (2010). Use of maximum entropy principle with lagrange multipliers extends the feasibility of elementary mode analysis. Journal of Bioscience and Bioengineering,

Zhong, J. J., & Yue, C. J. (2005). Plant cells: Secondary metabolite heterogeneity and its manipulation. *Advances in Biochemical Engineering/Biotechnology, 100*, 53-88.

Zhou, B., Martin, G. J., & Pamment, N. 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), 627-633.

Appendix

A. The *E. coli* metabolic model used for elementary modes computation

Reaction names	Stoichiometric reactions	Genes	Encoded enzymes
	Glycolysis pathway		
R2r	Glucose-6P = Fructose-6P	pgi	phosphoglucose isomerase
R3	Fructose-6P + ATP = Fructose-1,6P + ADP	pfkA; pfkB	6-phosphofrutokinase
R3	Fructose-1,6P = Fructose-6P	glpX; fbp	fructose 1,6-biphosphatase
R5r	Fructose-1,6P = DHAP + GA-3P	fbaB; fbaA	fructose biphosphate aldolase
R6r	GA-3P = DHAP	tpiA	triose phosphate isomerase
R7r	GA-3P + NAD + ADP = 3PG + NADH + ATP	gapA; pgk	glyceraldehyde 3-phosphate dehydrogenase phosphoglycerate kinase
R8r	3PG = PEP	pgml; ytjC; pgmA; eno	phosphoglyceate mutase enolase
R9	PEP + ADP = Pyruvate + ATP	pykA; pykF	pyruvate kinase
R49	Pyruvate + CoASH + NAD = Acetyl-CoA + CO_2 + NADH	lpdA; aceEF	pyruvate dehydrogenase multienzyme complex
	Pentose phosphate pathway		
R11	Glucose-6P + 2 NADP = Ribulose-5P + 2 NADPH + CO_2	zwf; pgl; gnd	glucose 6-phosphate-1-dehydrogenase 6-phosphoglucolactonase 6-phosphogluconate dehydrogenase
R13r	Ribulose-5P = Xylulose-5P	rpe	ribulose phosphate 3-epimerase
R12r	Ribulose-5P = Ribose-5P	rpiA; alsI	ribose-5-phosphate isomerase
R14r	Ribose-5P + Xylulose-5P = Sed-7P + GA-3P	tktAB	transketolase
R15r	GA-3P + Sed-7P = Erythose-4P + Fructose-6P	talAB	transadolase
R16r	Erythose-4P + Xylulose-5P = GA-3P + Fructose-6P	tktAB	transketolase
	Tricarboxylic acid cycle		
R25	Oxaloacetate + Acetyl-CoA = Citrate + CoASH	prpC; gltA	methylcitrate synthase citrate synthase
R17r	Citrate = Caco	acnAB	aconitase
R18r	Caco = Isocitrate	acnAB	aconitase
R19r	Isocitrate + NADP = a-Ketoglutarate + CO_2 + NADPH	icd	isocitrate dehydrogenase
R20	a-Ketoglutarate + NAD + CoASH = NADH + SCoA + CO_2	lpdA; sucAB	2-oxoglutarate dehydrogenase complex
R21r	SCoA + ADP = Succinate + ATP +	sucCD	succinyl CoA synthase

	CoASH		
R22r	Succinate + FAD = Fumarate + FADH	sdhABCD; frdABCD	succinate dehydrogenase fumarate reductase
R23r	Fumarate = Malate	fumABC	fumarase A
R24r	Malate + NAD = Oxaloacetate + NADH	mdh	malate dehydrogenase
	Anapleurotic pathway		
R27	PEP + CO_2 = Oxaloacetate	ppc	phosphoenolpyruvate carboxylase
R28	Malate + NAD = Pyruvate + CO_2 + NADH	sfcA; maeB	malate enzyme
R29	Oxaloacetate + ATP = PEP + ADP + CO_2	pckA	phosphoenolpyruvate carboxylase
	Fermentative acid pathways		
R30	Pyruvate + CoASH = Acetyl-CoA + Formate	pflB; tdcE	pyruvate formate-lyase
R31	Pyruvate = Acetate + CO_2	poxB	pyruvate oxidase
R32	Pyruvate + NADH = Lactate + NAD	ldhA	lactate dehydrogenase
R34	Acetyl-CoA + 1 NADH = ACA + 1 NAD + CoASH	adhP	acetaldehyde dehydrogenase
R41	ACA + 1 NADH = Ethanol + 1 NAD	adhE	alcohol dehydrogenase
R35	Acetyl-CoA = ACP + CoASH	pta	phosphate acetyltransferase
R42	ACP + ADP = Acetate + ATP	ackAB	acetate kinase
	Redox balance and Maintenance		
R37	NADH + 2 ADP + O_2 = NAD + 2 ATP	ndh;nuoAHJKLMNEF; GBCI; atpABCDEFGHI	NADH dehydrogenase I NADH dehydrogenase II ATP synthase
R38	FADH + ADP + O_2 = FAD + ATP	cyoABCD; atpABCDEFGHI	cytochrome bo terminal oxidase ATP synthase
R40r	NAD + NADPH = NADP + NADH	pntAB	pyridine nucleotide transhydrogenase
R39	ATP = ADP + Pi		
	Isoprenoid pathway		
R51	GA-3P + Pyruvate + NADPH + ATP = Hydroxymethylbutenyl-PP + CO_2 + NADP + ADP	dxs; dxr; ispD; ispE; ispF; ispG	1-deoxyxylulose-5-phosphate synthase 1-deoxy-D-xylulose 5-phosphate reductoisomerase 4-diphosphocytidyl-2C-methyl-D-erythritol synthetase 4-diphosphocytidyl-2-C-methylerythritol kinase 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase

			1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase
R52	Hydroxymethylbutenyl-PP + NADPH = Dimethylallyl-PP + Isopentenyl-PP + NADP	ispH	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase
R54r	Isopentenyl-PP = Dimethylallyl-PP	idi	isopentenyl diphosphate isomerase
R55	Dimethylallyl-PP + Isopentenyl-PP = Geranyl-PP	ispA	geranyl diphosphate synthase / farnesyl diphosphate synthase
R56	Geranyl-PP + Isopentenyl-PP = Farnesyl-PP	ispA	geranyl diphosphate synthase / farnesyl diphosphate synthase
Foreign ethanol producing pathway			
R62	Pyruvate = ACA + CO_2	pdc	pyruvate decarboxylase
R41	ACA + NADH = Ethanol + NAD	adhB	ethanol dehydrogenase
Foreign carotenoid producing pathway			
R57	2 Farnesyl diphosphate + 4 FAD = Diapolycopene + 4 FADH	crtM; crtN	dehydrosqualene synthase dehydrosqualene desaturase
R58	Diapolycopene + 4 NADP = Diapolycopen-diacid + 4 NADPH	crtOx	diaponeurosporene oxidase
R53	Farnesyl-PP + Isopentenyl-PP = Geranyl Geranyl-PP	crtE	geranylgeranyl pyrophosphate synthase
R59	2 Geranyl Geranyl-PP + 4 FAD = Lycopene + 4 FADH	crtB; crtI	phytoene synthase phytoene desaturase
Biomass synthesis			
R36	4 Glucose-6P + 52 Ribose-5P + 31 Erythose-4P + 159 PEP + 237 Pyruvate + 73 Acetyl-CoA + 86 a-Ketoglutarate + 142 Oxaloacetate + 2458 ATP + 246 NAD + 1107 NADPH + 755 NH_3 = Biomass + 73 CoASH + 2458 ADP + 246 NADH + 1107 NADP		
Sugars degradation pathways			
R1	Glucose_ext + PEP = Glucose-6P + Pyruvate	ptsG; ptsHI; crr	phosphoenolpyruvate:glucose transferase system
R1	Glucose_ext + ATP = Glucose-6P + ADP	glk	glucose kinase
R60	Xylose = Xylulose	xylA	xylose isomerase
R61	Xylulose + ATP = Xylose-5P + ADP	xylB	xylulokinase
R62	Arabinose + ATP = Xylulose-5P + ADP	araA; araB; araD	L-arabinose isomerase L-ribulokinase L-ribulosephosphate 4-epimerase
Transport reactions			
R60	Xylose_ext + ATP = Xylose + ADP	xylGHF	xylose ABC transporter
R62	Arabinose_ext + ATP = Arabinose + ADP	araFGH	arabinose transporter

TRA4	NH_3_ext = NH_3	*amtB*	ammonium Amt transporter
TRA7	Ethanol = Ethanol_ext		
TRA8	Acetate = Acetate_ext .		
TRA9	Lactate = Lactate_ext		
TRA10	Succinate = Succinate_ext		
TRA6	Formate = Formate_ext		
TRA5	CO_2 = CO_2_ext		

Metabolite nomenclatures DHAP: dihydroxy-acetone-phosphate; GA-3P: glyeraldehyde-3-phosphate; 3PG:3-phosphoglyerate; PEP: phosphoenolpyruvate; CO_2: carbon dioxide; Sed-7P: sedoheptulose-7-phosphate; CoASH: coenzyme A; SCoA: succinyl CoA; ACA: acetaldehyde; ACP: acetylphosphate; ADP: adenosine-5'-diphosphate; ATP: adenosine-5'-triphosphate; FAD: Flavin adenine dinucleotide; FADH: Flavin adenine dinucleotide reduced; NAD: nicotinamide adenine dinucleotide; NADH: dihydronicotinamide adenine dinucleotide; NADP: nicotinamide adenine dinucleotide phosphate; NADPH: dihydronicotinamide adenine dinucleotide phosphate; NH_3: ammonia. Metabolite names containing “ext” are referred to external metabolites.

B. The *T. saccharolyticum* metabolic model used for elementary modes computation

Reaction names	Stoichiometric reactions	Genes	Encoded enzymes
	Xylose degradation pathway		
XYLA	Xylose = Xylulose	xylA	xylose isomerase
XYLB	Xylulose + ATP = Xylose-5P + ADP	xylB	xylulokinase
	Glycolysis pathway		
GG1	Glucose + ATP = Glucose-6P + ADP	glk	glucose kinase
GG2r	Glucose-6P = Fructose-6P	pgi	phosphoglucose isomerase
GG3	Fructose-6P + ATP = Fructose-1,6P + ADP	pfk	6-phosphofructokinase
GG4	Fructose-1,6P = Fructose-6P	glpX; fdp	fructose 1,6-biphosphatase
GG5r	Fructose-1,6P = DHAP + GA3P	fba	fructose biphosphate aldolase
GG6r	GA3P = DHAP	tpiA	triose phosphate isomerase
GG7r	GA3P + NAD = 3PGP + NADH	gapA	glyceraldehyde 3-phosphate dehydrogenase-A complex
GG8r	3PGP + ADP = 3PG + ATP	pgk	phosphoglycerate kinase
GG9r	3PG = 2PG	pgml; gpm	phosphoglycerate mutase
GG10r	2PG = PEP	eno	enolase
GG11	PEP + ADP = Pyruvate + ATP	pyk	pyruvate kinase
	Pentose phosphate pathway		
PPP1	Glucose-6P + NADP = 6PG + NADPH	zwf	glucose 6-phosphate-1-dehydrogenase
PPP3	6PG + NADP = Ribulose-5P + CO_2 + NADPH	gnd	6-phosphogluconate dehydrogenase
PPP4r	Ribulose-5P = Xylulose-5P	rpe	ribulose phosphate 3-epimerase
PPP5r	Ribulose-5P = Ribose-5P	rpiA; alsl	ribose-5-phosphate isomerase
PPP6r	Ribose-5P + Xylulose-5P = Sed-7P + GA3P	tkt	transketolase
PPP7r	GA3P + Sed-7P = Erythose-4P + Fructose-6P	tal	transadolase
PPP8r	Erythose-4P + Xylulose-5P = GA3P + Fructose-6P	tkt	transketolase
	Fermentative acid pathways		
PFL	Pyruvate + CoASH = Acetyl-CoA + Formate	pfl	pyruvate formate-lyase
LDH	Pyruvate + NADH = Lactate + NAD	ldh	lactate dehydrogenase
POR	Pyruvate + CoASH + Fdox = Fdred + Acetyl-CoA + CO_2	por	pyruvate oxidoreductase
PTA	Acetyl-CoA = ACP + CoASH	pta	phosphate acetyltransferase
ACK	ACP + ADP = Acetate + ATP	ack	acetate kinase
ALD1	Acetyl-CoA + NADH = ACA + NAD +	aldA	NAD-acetaldehyde

	CoASH		dehydrogenase
ALD2	Acetyl-CoA + NADPH = ACA + NADP + CoASH	aldB	NADP-acetaldehyde dehydrogenase
ADH1	ACA + NADH = Ethanol + NAD	adhE	NAD-alcohol dehydrogenase
ADH2	ACA + NADPH = Ethanol + NADP	adhE	NADP-alcohol dehydrogenase
MGS1	DHAP = MGO	mgsA	methylglyoxal synthase
MGS2	MGO + NADPH = Lactate + NADP	gre	methylglyoxal reductase
PFP1	Pyruvate + SMMCoA = OAA + PPCoA	n.a.	methylmalony CoA carboxyltransferase
PFP2	Succinate + PPCoA = SCoA + Propionate	scp	propionyl-succinate CoA transferase
PFP3r	SCoA = RMMCoA	sbm	methylmalony CoA mutase
PFP4r	RMMCoA = SMMCoA	n.a.	methylmalony CoA epimerase
PFP5	PPCoA + CO_2 + ATP = SMMCoA + ADP	n.a.	propionyl-CoA carboxylase
PFP6	PPCoA + Acetate = Propionate + Acetyl-CoA	pct	propionate CoA transferase
	Redox balance and Maintenance		
FNOR1	Fdred + NAD = Fdox + NADH	fnor	NAD-ferredoxin oxidoreductase
FNOR2	Fdred + NADP = Fdox + NADPH	fnor	NADP-ferredoxin oxidoreductase
ECH	Fdred = Fdox + H_2	ech	NiFe-hydrogenase
HYD1	NADH = NAD + H_2	hyd	FeFe-hydrogenase
HYD2	NADPH = NADP + H_2	hyd	FeFe-hydrogenase
MAIN	ATP = ADP + ATP_base		
	Tricarboxylic acid cycle		
TCA1	Oxaloacetate + Acetyl-CoA = Citrate + CoASH	prpC; gltA	methylcitrate synthase; citrate synthase
TCA2r	Citrate = Isocitrate	can	aconitase
TCA3	Isocitrate + NADP = a-Ketoglutarate + CO_2 + NADPH	icd	isocitrate dehydrogenase
TCA4	a-Ketoglutarate + 2 Fdox + CoASH = 2 Fdred + SCoA + CO_2	porAB	2-oxoglutarate:ferredoxin oxidoreductase
TCA5	Fumarate + NADH = Succinate + NAD	frd	fumarate reductase
TCA6r	Fumarate = Malate	fum	fumarase
TCA7r	Malate + NAD = Oxaloacetate + NADH	mdh	malate dehydrogenase
	Anapleurotic pathway		
ANA1	Malate + NAD = Pyruvate + CO_2 + NADH	sfc	malate dehydrogenase
ANA2	Oxaloacetate + ATP = PEP + ADP + CO_2	pck	phosphoenolpyruvate carboxykinase
PYC	Pyruvate + CO_2 + ATP = Oxaloacetate + ADP	pyc	pyruvate carboxylase
	Biomass synthesis		

BIO	157 Glucose-6P + 1222 Ribose-5P + 854 a-Ketoglutarate + 1774 Oxaloacetate + 140 Erythose-4P + 568 PEP + 2436 Pyruvate + 1317 GA3P + 1932 Acetyl-CoA + 17497 ATP + 3231 NAD + 13832 NADPH + 9650 NH_3 = BIOMASS + 1932 CoASH + 17497 ADP + 3231 NADH + 13832 NADP
Transport reactions	
TRA1	Cellobiose_ext = Glucose
TRA2	Xylan_ext = Xylose
TRA3	Ethanol = Ethanol_ext
TRA4	Acetate = Acetate_ext
TRA5	NH_3_ext = NH_3
TRA6	Lactate = Lactate_ext
TRA7	CO_2 = CO_2_ext
TRA8	Formate = Formate_ext
TRA9r	H_2 = H_2_ext
TRA10	Propionate = Propionate_ext

Metabolite nomenclatures 2PG: 2-phosphate glycerate; 3PG:3-phosphoglyerate; 3PGP: 3-phospho-glyceroyl-phosphate; 6PG: 6-phosphate gluconate; ACA: acetaldehyde; ACP: acetylphosphate; ADP: adenosine-5'-diphosphate; ATP: adenosine-5'-triphosphate; ATP_base: maintenance energy; FAD: Flavin adenine dinucleotide; FADH: Flavin adenine dinucleotide reduced; Fdox: Ferridoxin oxidized; Fdred: Ferridoxin reduced; RMMCoA: R-methylmalonyl-CoA; SMMCoA: S-methylmalonyl-CoA; PPCoA: propionyl-CoA; ACP: acetylphosphate; ACA: acetaldehyde; MGO: methylglyoxal; PEP: phosphoenolpyruvate; Sed-7P: sedoheptulose-7-phosphate; SCoA: succinyl CoA; AACoA: acetoacetyl CoA; CO_2: carbon dioxide; CoASH: coenzyme A; DHAP: dihydroxy-acetone-phosphate; GA3P: glyeraldehyde-3-phosphate; NAD: nicotinamide adenine dinucleotide; NADH: dihydronicotinamide adenine dinucleotide; NADP: nicotinamide adenine dinucleotide phosphate; NADPH: dihydronicotinamide adenine dinucleotide phosphate; NH_3: ammonia. Metabolite names containing “ext” are referred to external metabolites.

C. Metabolic network of lycopene producing *E. coli.* Reaction designations in the network correspond to the reaction number used in elementary mode analysis. Reactions shown in green are introduced into the cell via genes on plasmid pACEBI. External metabolites are underlined.